Bicycle Sharing Systems: Fast and Slow Urban Mobility Dynamics

James Todd

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Geography
University College London

November, 2022
Doctor of Philosophy
Department of Geography, UCL

I, James Todd confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

.................................................................

Submitted October 2022
Dedicated to my parents,
Jonathan and Keiko Todd
Prologue

Abstract
In cities all around the world, new forms of urban micromobility have observed rapid and wide-scale adoption due to their benefits as a shared mode that are environmentally friendly, convenient and accessible. Bicycle sharing systems are the most established among these modes, facilitating complete end-to-end journeys as well as forming a solution for the first/last mile issue that public transportation users face in getting to and from transit stations. They mark the beginnings of a gradual transition towards a more sustainable transportation model that include greater use of shared and active modes. As such, understanding the way in which these systems are used is essential in order to improve their management and efficiency. Given the lack of operator published data, this thesis aims to explore the utility of open bicycle sharing system data standards that are intended for real-time dissemination of bicycle locations in uncovering novel insights into their activity dynamics over varying temporal and geographical scales.

The thesis starts by exploring bicycle sharing systems at a global-scale, uncovering their long-term growth and evolution through the development of data cleaning and metric creation heuristics that also form the foundations of the most comprehensive classification of systems. Having established the values of these metrics in conducting comparisons at scale, the thesis then analyses the medium-term impacts of mobility interventions in the context of the COVID-19 pandemic, employing spatio-temporal and network analysis methods that highlight their adaptability and resilience. Finally, the thesis closes with the analysis of granular spatial and temporal dynamics within a dockless system in London that enable the identi-
fication of the variations in journey locations throughout different times of the day. In each of these cases, the research highlights the indispensable value of open data and the important role that bicycle sharing systems play in urban mobility.

Impact Statement

This thesis makes several novel contributions in furthering the understanding of openly accessible bicycle sharing system data and their applicability in driving insights into their dynamics. These range from global-scale comparisons to granular hotspot locations in individual systems. Such insights have been drawn through the development of bespoke heuristics that cater to the unique characteristics of these data.

The research highlights how passively produced bicycle sharing system data can be manipulated to create a variety of metrics and datasets that provide valuable insights into their size, use and contextual surroundings that would otherwise be unattainable given the lack of published data sources. Due to their programmatic creation, the methods are scalable and comparable across systems in addition to providing a framework for the analysis of similar data that are produced by other forms of shared transportation such as e-scooters and cars. The utility of these metrics are showcased in the creation of the most comprehensive comparison of bicycle sharing systems globally, presenting novel insights into the sector. This analysis extends on both academic and industry perceptions of the transportation mode, having been published in *Journal of Transport Geography* (Todd et al. 2021) as well as providing a means to validate operator press releases, estimate characteristics of systems for which data are not available and better manage and plan systems.

The analyses also present more granular insights into individual systems, extending on our understanding of changes in cycling dynamics as a result of the COVID-19 pandemic as well as the specific times and places where dockless bicycle sharing systems are used. Network analyses methods that are seldom employed across micromobility modes present a valuable alternative framework to quantify the behavioural adjustments in mobility that have occurred as a result of restrictions
imposed on populations to mitigate the spread of the virus. The results highlight the importance of bicycle sharing systems in equipping urban populations with a flexible and resilient mode of transportation. Similarly, the analysis of a dockless system in London produce methodological contributions in its operationalisation of statistical modelling that are typically deployed in spatial epidemiology and environmental criminology and formulation of unique hotspot detection techniques. Such processes not only showcase their applicability in generating new knowledge within contexts that have yet to be studied but also their ability to help inform operator and governing body decision-making in prioritising investments such as cycling and parking infrastructure to improve the efficiency and safety of systems.

Collectively, this thesis exemplifies the large variety of ways through which passively produced, openly accessible data can be better leveraged to further our perception of mobility in urban environments. Aside from the apparent importance of these data, the results overwhelmingly suggest the importance of shared micromobility modes in aiding the transition towards a more sustainable transportation future.

Acknowledgements

First, I would like to thank my supervisors, Professor James Cheshire and Professor Paul Longley, for their invaluable support and guidance throughout this PhD. I would also like to thank Gerry Casey, my industrial supervisor, for his advice and perspectives from the private sector. Gratitude goes to Oliver O’Brien, who provided me with access to his data in addition to imparting me with his expertise on bicycle sharing systems over countless hours of discussion as well as Dr. Anwar Musah for his time discussing statistical methods with me. A special thanks go to the ESRC and Arup for funding my research. I would also like to thank my friends and colleagues in the Chorley Institute, especially Alfie Long, who have made my time at UCL very memorable. Last, but not least, I would like to give an extra special thank you to my family and Sally whose unwavering support and understanding got me through the most stressful times of my research.
Research Outputs

Publications


Working Publications


Media Publications

Engagement Activities


Contents

1 Introduction 25
   1.1 Research Aims and Objectives .......................... 27
   1.2 Thesis Structure ........................................ 28

2 Literature Review 31
   2.1 Urban Mobility: The Past, The Present and The Future ...... 34
   2.2 The Dynamics of Bicycle Sharing Systems .................. 56
   2.3 Chapter Summary ......................................... 74

3 Bicycle Sharing System Data 77
   3.1 The General Bikeshare Feed Specification ................. 79
   3.2 UCL’s Bicycle Sharing System Data ........................ 85
   3.3 Open Journey Data ....................................... 102
   3.4 Chapter Summary ......................................... 105

4 Manipulating and Validating Bicycle Sharing System Data 107
   4.1 Metric Creation .......................................... 108
   4.2 Exploring Metrics ....................................... 128
   4.3 Chapter Summary ......................................... 135

5 A Global Comparison of Bicycle Sharing Systems 137
   5.1 Slow Dynamics at a Global Scale .......................... 138
   5.2 Studying Bicycle Sharing Systems at Scale ................ 141
   5.3 Employing Metrics ....................................... 143
   5.4 Clustering Bicycle Sharing Systems ....................... 147
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5</td>
<td>The Global Bicycle Sharing System Classification</td>
<td>157</td>
</tr>
<tr>
<td>5.6</td>
<td>Discussion</td>
<td>172</td>
</tr>
<tr>
<td>5.7</td>
<td>Chapter Summary</td>
<td>178</td>
</tr>
<tr>
<td>6</td>
<td>Impacts of COVID-19 on London Bicycle Sharing</td>
<td>181</td>
</tr>
<tr>
<td>6.1</td>
<td>London and COVID-19</td>
<td>183</td>
</tr>
<tr>
<td>6.2</td>
<td>The Impacts of COVID-19 on Bicycle Sharing Systems</td>
<td>185</td>
</tr>
<tr>
<td>6.3</td>
<td>Data</td>
<td>188</td>
</tr>
<tr>
<td>6.4</td>
<td>Exploratory Spatio-Temporal Analysis</td>
<td>192</td>
</tr>
<tr>
<td>6.5</td>
<td>Restriction Compliance and Response Time</td>
<td>201</td>
</tr>
<tr>
<td>6.6</td>
<td>Network Analysis</td>
<td>203</td>
</tr>
<tr>
<td>6.7</td>
<td>Discussion</td>
<td>220</td>
</tr>
<tr>
<td>6.8</td>
<td>Chapter Summary</td>
<td>225</td>
</tr>
<tr>
<td>7</td>
<td>Dockless Bicycle Sharing Mobility Trends</td>
<td>227</td>
</tr>
<tr>
<td>7.1</td>
<td>The JUMP Dockless Bicycle Sharing System in London</td>
<td>229</td>
</tr>
<tr>
<td>7.2</td>
<td>Identifying Journeys</td>
<td>232</td>
</tr>
<tr>
<td>7.3</td>
<td>Journey Hotspots</td>
<td>239</td>
</tr>
<tr>
<td>7.4</td>
<td>Inferring Dominant Trip Purposes</td>
<td>251</td>
</tr>
<tr>
<td>7.5</td>
<td>Discussion</td>
<td>266</td>
</tr>
<tr>
<td>7.6</td>
<td>Chapter Summary</td>
<td>273</td>
</tr>
<tr>
<td>8</td>
<td>Discussions and Conclusions</td>
<td>275</td>
</tr>
<tr>
<td>8.1</td>
<td>The Dynamics of Bicycle Sharing Systems</td>
<td>276</td>
</tr>
<tr>
<td>8.2</td>
<td>Applications and Impacts</td>
<td>279</td>
</tr>
<tr>
<td>8.3</td>
<td>Future Work</td>
<td>281</td>
</tr>
<tr>
<td>8.4</td>
<td>Closing Remarks</td>
<td>282</td>
</tr>
<tr>
<td></td>
<td>Appendices</td>
<td>284</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>298</td>
</tr>
</tbody>
</table>
List of Figures

3.1 The growth of the UCL BSS database . . . . . . . . . . . . . . . . . . 86
3.2 Percentage share of BSS in the UCL BSS database in comparison
to the Meddin Bike-Sharing World Map . . . . . . . . . . . . . . . 89
4.1 Diagram illustrating the basic mechanisms of the journey estimation
methodology . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 109
4.2 First pass journey estimations in comparison to open journey data
in London . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 112
4.3 Diagrammatic representation of the data noise detection methodology113
4.4 Cleaned journey estimations in comparison to open journey data in
London . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 114
4.5 Average hourly journeys in a typical week in the London Santander
Cycles BSS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 130
4.6 Average weekday and weekend entropy in the London Santander
Cycles BSS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 131
4.7 Quarterly evolution of the London Santander Cycles BSS . . . . . . 134
5.1 The evolution of 176 BSS between 2016 and 2018 . . . . . . . . . . 140
5.2 The size and location of the 322 BSS included in the determination
of the global BSS landscape . . . . . . . . . . . . . . . . . . . . . . 145
5.3 Optimal number of k-means clusters as determined by the elbow
method . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 154
5.4 Average weekday use profile for ‘very large, high use BSS’ . . . . . 159
5.5 Average weekend use profiles for ‘very large, high use BSS’ . . . . 160
5.6 Average weekday use profiles for ‘large BSS in major cities’ . . . . 161
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.7</td>
<td>Average weekend use profiles for ‘large BSS in major cities’</td>
<td>162</td>
</tr>
<tr>
<td>5.8</td>
<td>Average weekday use profiles for ‘medium BSS with extensive cycling infrastructure’</td>
<td>164</td>
</tr>
<tr>
<td>5.9</td>
<td>Average weekend use profiles for ‘medium BSS with extensive cycling infrastructure’</td>
<td>165</td>
</tr>
<tr>
<td>5.10</td>
<td>Average weekday use profiles for ‘small to medium efficient BSS’</td>
<td>167</td>
</tr>
<tr>
<td>5.11</td>
<td>Average weekend use profiles for ‘small to medium efficient BSS’</td>
<td>168</td>
</tr>
<tr>
<td>5.12</td>
<td>Average weekday use profiles for ‘small to medium inefficient BSS’</td>
<td>170</td>
</tr>
<tr>
<td>5.13</td>
<td>Average weekend use profiles for ‘small to medium inefficient BSS’</td>
<td>171</td>
</tr>
<tr>
<td>6.1</td>
<td>Comparison of changes in activity levels throughout the first 16 months of the COVID-19 pandemic</td>
<td>193</td>
</tr>
<tr>
<td>6.2</td>
<td>Average weekday entropy in each lockdown and control period</td>
<td>194</td>
</tr>
<tr>
<td>6.3</td>
<td>Average weekend entropy in each lockdown and control period</td>
<td>195</td>
</tr>
<tr>
<td>6.4</td>
<td>Comparison of average hourly weekday and weekend journey counts in each lockdown and control period [a) Lockdown 1, b) Lockdown 2, c) Lockdown 3]</td>
<td>197</td>
</tr>
<tr>
<td>6.5</td>
<td>Percentage change in weekday and weekend journey counts in each lockdown and control period [a) Lockdown 1 weekday, b) Lockdown 1 weekend, c) Lockdown 2 weekday, d) Lockdown 2 weekend, e) Lockdown 3 weekday, f) Lockdown 3 weekend]</td>
<td>198</td>
</tr>
<tr>
<td>6.6</td>
<td>Santander Cycles docking station community assignment and centrality during Lockdown 1 [a) weekday control, b) weekday study, c) weekend control, d) weekend study]</td>
<td>215</td>
</tr>
<tr>
<td>6.7</td>
<td>Santander Cycles BSS docking station community assignment and centrality in Lockdown 2 [a) weekday control, b) weekday study, c) weekend control, d) weekend study]</td>
<td>218</td>
</tr>
<tr>
<td>6.8</td>
<td>Santander Cycles BSS docking station community assignment and centrality in Lockdown 3 [a) weekday control, b) weekday study, c) weekend control, d) weekend study]</td>
<td>219</td>
</tr>
</tbody>
</table>
List of Figures

7.1 JUMP e-BSS study region [orange areas depict permitted operating boroughs] ........................................ 231
7.2 Daily TDB in JUMP e-BSS throughout study period [yellow points depicting weekends] ................................ 236
7.3 Average hourly journeys in a typical week in the JUMP e-BSS . . . 237
7.4 Spatial distribution of JUMP e-BSS journey origins across study region .......................................................... 241
7.5 Spatial distribution of JUMP e-BSS journey destinations across study region ..................................................... 242
7.6 JUMP e-BSS journey origin hotspots and popular flows [a) AM Peak Hours, b) PM Peak Hours, c) Off-Peak Hours, d) Weekend All: grey points depict train stations] ........................................ 248
7.7 JUMP e-BSS journey destination hotspots and popular flows [a) AM Peak Hours, b) PM Peak Hours, c) Off-Peak Hours, d) Weekend All: grey points depict train stations] ........................................ 250
7.8 Frequency distribution of journey destination counts across hexagons 258

A.1 Temporal distribution of London Santander Cycles BSS operator bicycle collection in 2016 ................................. 285
A.2 London Santander Cycles BSS operator bicycle distribution during the average weekday and weekend in 2016 .................... 286

B.1 The evolution of 176 BSS between 2016 and 2018 Source: (Cheshire and Uberti 2021) ............................................. 287

D.1 Public transport access level in London Source: (TfL 2015) .......... 295
D.2 JUMP e-BSS journey origin hotspots ........................................ 296
D.3 JUMP e-BSS journey destination hotspots ................................. 297
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>List of all required files for GBFS feed</td>
<td>83</td>
</tr>
<tr>
<td>3.2</td>
<td>The stored table names in the UCL BSS database</td>
<td>87</td>
</tr>
<tr>
<td>3.3</td>
<td>The stored variable names for dock capacity data and their associated source from the GBFS feed</td>
<td>91</td>
</tr>
<tr>
<td>3.4</td>
<td>The stored variable names for dock location data and their associated source from the GBFS feed</td>
<td>95</td>
</tr>
<tr>
<td>3.5</td>
<td>The stored variable names for dockless bicycle location data and their associated source from the GBFS feed</td>
<td>97</td>
</tr>
<tr>
<td>3.6</td>
<td>Temporal variations in the proportion of BSS by operational type</td>
<td>101</td>
</tr>
<tr>
<td>4.1</td>
<td>Descriptive statistics of journey estimations percentage errors in comparison to operator published journey data</td>
<td>115</td>
</tr>
<tr>
<td>4.2</td>
<td>A summary of variables that influence the use of BSS</td>
<td>123</td>
</tr>
<tr>
<td>4.3</td>
<td>Confounding variable data sources, resolution and buffer distances</td>
<td>124</td>
</tr>
<tr>
<td>4.4</td>
<td>Descriptive statistics of London Santander Cycles metrics over 12 years of operation</td>
<td>129</td>
</tr>
<tr>
<td>5.1</td>
<td>Exploratory analysis of k-means clusters sizes</td>
<td>155</td>
</tr>
<tr>
<td>5.2</td>
<td>First-stage k-means cluster average characteristics</td>
<td>158</td>
</tr>
<tr>
<td>6.1</td>
<td>Definition of study and control lockdown durations</td>
<td>189</td>
</tr>
<tr>
<td>6.2</td>
<td>Average weekday and weekend journey duration for each lockdown and control period (minutes)</td>
<td>200</td>
</tr>
<tr>
<td>6.3</td>
<td>Distribution of SMR results across 339 BSS</td>
<td>203</td>
</tr>
<tr>
<td>6.4</td>
<td>Global network statistics of Lockdown 1 and associated control period</td>
<td>211</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td>Global network statistics of Lockdown 2 and associated control period</td>
<td></td>
</tr>
<tr>
<td>6.6</td>
<td>Global network statistics of Lockdown 3 and associated control period</td>
<td></td>
</tr>
<tr>
<td>7.1</td>
<td>Total and average daily journeys that end in each borough across the study area [bold depicting permitted boroughs]</td>
<td></td>
</tr>
<tr>
<td>7.2</td>
<td>Average journey characteristics in each time block</td>
<td></td>
</tr>
<tr>
<td>7.3</td>
<td>OSM feature keys and values used to identify built environment factors</td>
<td></td>
</tr>
<tr>
<td>7.4</td>
<td>Additional data sources used in regression analysis</td>
<td></td>
</tr>
<tr>
<td>7.5</td>
<td>Descriptive statistics of variables included in regression model</td>
<td></td>
</tr>
<tr>
<td>7.6</td>
<td>Zero-inflated multilevel negative binomial regression results</td>
<td></td>
</tr>
<tr>
<td>A.1</td>
<td>Descriptive statistics of London Santander Cycles BSS operator bicycle collection in 2016</td>
<td></td>
</tr>
<tr>
<td>A.2</td>
<td>Descriptive statistics of London Santander Cycles BSS operator bicycle distribution in 2016</td>
<td></td>
</tr>
<tr>
<td>C.1</td>
<td>Weekday dynamic time warping cluster assignments for ‘very large, high use BSS’</td>
<td></td>
</tr>
<tr>
<td>C.2</td>
<td>Weekend dynamic time warping cluster assignments for ‘very large, high use BSS’</td>
<td></td>
</tr>
<tr>
<td>C.3</td>
<td>Weekday dynamic time warping cluster assignments for ‘large BSS in major cities’</td>
<td></td>
</tr>
<tr>
<td>C.4</td>
<td>Weekend dynamic time warping cluster assignments for ‘large BSS in major cities’</td>
<td></td>
</tr>
<tr>
<td>C.5</td>
<td>Weekday dynamic time warping cluster assignments for ‘medium BSS with extensive cycling infrastructure’</td>
<td></td>
</tr>
<tr>
<td>C.6</td>
<td>Weekend dynamic time warping cluster assignments for ‘medium BSS with extensive cycling infrastructure’</td>
<td></td>
</tr>
<tr>
<td>C.7</td>
<td>Weekday dynamic time warping cluster assignments for ‘small to medium efficient BSS’</td>
<td></td>
</tr>
</tbody>
</table>
C.8 Weekend dynamic time warping cluster assignments for ‘small to medium efficient BSS’ 291
C.9 Weekday dynamic time warping cluster assignments for ‘small to medium inefficient BSS’ 292
C.10 Weekend dynamic time warping cluster assignments for ‘small to medium inefficient BSS’ 293
# List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANPR</td>
<td>Automatic number plate recognition</td>
</tr>
<tr>
<td>API</td>
<td>Application programming interface</td>
</tr>
<tr>
<td>AV</td>
<td>Autonomous vehicles</td>
</tr>
<tr>
<td>BMS</td>
<td>Battery management systems</td>
</tr>
<tr>
<td>BRP</td>
<td>Bicycle-sharing rebalancing problem</td>
</tr>
<tr>
<td>BSPP</td>
<td>Bicycle-sharing service planning problems</td>
</tr>
<tr>
<td>BSS</td>
<td>Bicycle sharing systems</td>
</tr>
<tr>
<td>CBD</td>
<td>Central business district</td>
</tr>
<tr>
<td>CCS</td>
<td>Congestion charge scheme</td>
</tr>
<tr>
<td>COVID-19</td>
<td>Coronavirus Disease 2019</td>
</tr>
<tr>
<td>DBSCAN</td>
<td>Density-based clustering of applications with noise</td>
</tr>
<tr>
<td>DEM</td>
<td>Digital elevation model</td>
</tr>
<tr>
<td>DRC</td>
<td>Democratic Republic of the Congo</td>
</tr>
<tr>
<td>DTW</td>
<td>Dynamic time warping</td>
</tr>
<tr>
<td>e-BSS</td>
<td>e-bicycle sharing systems</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>EV</td>
<td>Electric vehicles</td>
</tr>
<tr>
<td>GBFS</td>
<td>General Bikeshare Feed Specification</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross domestic product</td>
</tr>
<tr>
<td>GDPR</td>
<td>General Data Protection Regulation</td>
</tr>
<tr>
<td>GhG</td>
<td>Greenhouse gas</td>
</tr>
<tr>
<td>GHS</td>
<td>Global human settlement</td>
</tr>
<tr>
<td>GPS</td>
<td>Global positioning system</td>
</tr>
<tr>
<td>GTFS</td>
<td>General Transit Feed Specification</td>
</tr>
<tr>
<td>GUID</td>
<td>Globally unique identifier</td>
</tr>
<tr>
<td>ICT</td>
<td>Information and communication technology</td>
</tr>
<tr>
<td>ID</td>
<td>Identifier</td>
</tr>
<tr>
<td>ITF</td>
<td>International Transport Forum</td>
</tr>
<tr>
<td>KDE</td>
<td>Kernel density estimation</td>
</tr>
<tr>
<td>kmph</td>
<td>Kilometres per hour</td>
</tr>
<tr>
<td>KNN</td>
<td>k-nearest neighbour</td>
</tr>
<tr>
<td>LCA</td>
<td>Life cycle assessment</td>
</tr>
<tr>
<td>LRT</td>
<td>Lag response time</td>
</tr>
<tr>
<td>LTN</td>
<td>Low traffic neighbourhood</td>
</tr>
<tr>
<td>MaaS</td>
<td>Mobility-as-a-service</td>
</tr>
<tr>
<td>MAE</td>
<td>Mean average error</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>MAUP</td>
<td>Modifiable areal unit problem</td>
</tr>
<tr>
<td>MDS</td>
<td>Mobility Data Specification</td>
</tr>
<tr>
<td>MSOA</td>
<td>Middle layer Super Output Area</td>
</tr>
<tr>
<td>NABSA</td>
<td>North American Bikeshare and Scootershare Association</td>
</tr>
<tr>
<td>NPI</td>
<td>Non-pharmaceutical intervention</td>
</tr>
<tr>
<td>OA</td>
<td>Output Area</td>
</tr>
<tr>
<td>OD</td>
<td>Origin-destination</td>
</tr>
<tr>
<td>OMF</td>
<td>Open Mobility Foundation</td>
</tr>
<tr>
<td>OSM</td>
<td>OpenStreetMap</td>
</tr>
<tr>
<td>OxCGRT</td>
<td>Oxford COVID-19 Government response Tracker</td>
</tr>
<tr>
<td>PBSC</td>
<td>Public Bike System Company</td>
</tr>
<tr>
<td>PT</td>
<td>Public transportation</td>
</tr>
<tr>
<td>QoL</td>
<td>Quality of life</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root mean square error</td>
</tr>
<tr>
<td>RR</td>
<td>Risk ratio</td>
</tr>
<tr>
<td>SAH</td>
<td>Stay-at-home</td>
</tr>
<tr>
<td>SD</td>
<td>System dynamics</td>
</tr>
<tr>
<td>SI</td>
<td>Stringency index</td>
</tr>
<tr>
<td>SMR</td>
<td>Similarity in magnitude response</td>
</tr>
<tr>
<td>SoC</td>
<td>State-of-charge</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
</tr>
<tr>
<td>SRTM</td>
<td>Shuttle Radar Topography Mission</td>
</tr>
<tr>
<td>SSE</td>
<td>Sum of squared errors</td>
</tr>
<tr>
<td>TDB</td>
<td>Trips per day per bicycle</td>
</tr>
<tr>
<td>TFL</td>
<td>Transport for London</td>
</tr>
<tr>
<td>TTL</td>
<td>Time to live</td>
</tr>
<tr>
<td>UCL</td>
<td>University College London</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>ULEZ</td>
<td>Ultra Low Emissions Zone</td>
</tr>
<tr>
<td>UNFCCC</td>
<td>United Nations Framework Convention on Climate Change</td>
</tr>
<tr>
<td>USA</td>
<td>United States of America</td>
</tr>
<tr>
<td>VGI</td>
<td>Volunteered geographic information</td>
</tr>
<tr>
<td>WFH</td>
<td>Work from home</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
<tr>
<td>WSS</td>
<td>Within cluster sum of squares</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Cities are the foundations of modern society, accommodating 57% of total population (World Bank 2022) and 80% of global gross domestic product (GDP) (World Bank 2020). They are a culmination of human development, providing concentrated areas to work, trade and live. Although they present innumerable benefits, given the concentration of activity, cities are also the arena for many novel issues. Most notably, the global climate crisis is one that threatens life on all corners of the planet. Cities are the primary culprit, accounting for more than 80% of greenhouse gas (GhG) emissions produced (Hoornweg et al. 2011) and have thus become a focal frontier in the united battle for our future longevity.

Among efforts to tackle climate change by striving towards net-zero emissions, data are an essential tool that have been leveraged to formulate, monitor and manage mitigation strategies. This has been facilitated by the dawn of the information age at the turn of the millennium that transformed the way in which activity and interactions can be studied. The digitisation of day-to-day activities has led to a ‘data deluge’, revolutionising the nature of data collection procedures from active, highly structured and costly procedures to passive by-products. In turn, this has perpetuated discourses surrounding smart cities that aim to leverage these ‘big data’ to help them make more informed decisions that are beneficial for the individual, the city and the planet.

Although all dimensions of emission producing activities require considerable work to reach stringent net-zero targets over the next few decades, some require
significantly more restructuring, investment and time, such as the transition of the energy sector to clean and sustainable means of production. Comparatively, the transportation sector presents a relatively flexible aspect of emissions production due to its disaggregated nature. In addition, urban mobility has been continually plagued by issues surrounding congestion and accessibility that decrease the overall efficiency and productivity of cities. Accounting for 37% of global CO$_2$ emissions (IEA 2022), transportation emerges as an important and significant sector that can make considerable contributions towards such goals. Consequently, cities have begun to develop and implement plans to alleviate the reliance on carbon-based transportation modes by providing competitive and sustainable alternatives. Traditionally, this has been facilitated through investments in public transportation (PT) services, but there has been a growing acceptance that alternative active and shared modes are also necessary to expedite such processes.

Bicycle sharing systems (BSS) are a novel form of urban micromobility that have observed mass adoption in cities around the world due to their potential in alleviating issues surrounding congestion and promoting a modal shift towards shared and sustainable means. They consist of bicycles fleets located around urban environments that can be rented to enable short distance journeys along with increasing the utilisation of PT as they resolve the difficulties that travellers face in getting to and from transit stations. They form part of future visions of urban transportation such as mobility-as-a-service (MaaS), that alleviate the necessity for privately owned vehicles in favour of a network of highly integrated shared modes. Given such prominent practical benefits and their stake in the future of sustainable transportation, it is imperative that we can analyse their dynamics.

Despite the lack of data that are made accessible by micromobility operators, passive data that are produced as a means for user to located bicycles can be collected and repurposed to conduct empirical research. This thesis aims to explore the utility and applicability of these new data sources in generating insights into the dynamics of BSS.
1.1 Research Aims and Objectives

The ‘data trails’ that are produced present valuable opportunities to better understand the dynamics of our cities but also present a myriad of challenges. Firstly, there are substantial ethical and privacy concerns with the use of these data due to the common lack of awareness in the way that they are collected and used. Although awareness and transparency are beginning to increase, with standards and regulations that have been established to ensure their proper storage, dissemination and use, they still introduce possibilities to identify individuals by combining them with other datasets. Furthermore, since they were not intended for such purposes, they lack quality control and are prone to containing errors and biases. Therefore, it is imperative that we evaluate the utility, capabilities and limitations of these data, whilst ensuring the privacy of individuals, so that they can be applied in research and practical settings.

BSS are one of the many digital activities that produce data passively, armed with sensors that log the movement of bicycles which provide crucial operational information to manage them. In addition to the concerns of user privacy, operators of BSS tend to stray away from making all of these data publicly accessible, with risks of wrongful application and loss of competitive advantage that make it economically unjustifiable. Given the highly competitive nature of the current micromobility sector, the availability of granular and clean data sources are typically limited to those systems that have been mandated to release them. Although this is the case, by virtue of BSS applications that provide users with information on the location and availability of bicycles, there are some data that are openly accessible. These data typically come in the form of application programming interface (API) feeds that supply real-time data to operator and third party applications, much like other forms of transportation, that equip users with the information necessary to plan their journeys. These openly accessible BSS data feeds give researchers an opportunity to collect and analyse them in order to better understand the dynamics of these systems and offer suggestions and guidance to their operation.
With these challenges in mind, the overarching aim of this thesis is to explore the value of openly accessible data in presenting novel and actionable insights of BSS dynamics across a variety of temporal and geographical scales. In this context, the thesis works on answering four primary research questions, each contributing to provide a more holistic appreciation of these data:

1. How can BSS data be manipulated to produce clean and accurate metrics?
2. How can BSS metrics be employed to uncover novel insights of the global landscape?
3. What is the utility of BSS journey data in understanding the imposition of mobility restrictions during the COVID-19 pandemic?
4. When, where and why are dockless BSS used?

The motivation for this research began with the collection of the most comprehensive database of openly accessible BSS data. Since the data were initially gathered with the intention of displaying real-time availability of bicycles in shared systems on bikesharemap.com, the true value of these data had not been fully explored. Given its vast depth and breadth, it presents many opportunities that were previously not possible. As a result, this thesis offers a holistic yet granular overview of this potential, being mindful of the unique characteristics these data have when conducting analysis, in hopes of contributing to the growing body of literature on BSS dynamics that may be used for practical applications and improvements for currently operational and future systems.

1.2 Thesis Structure

The thesis starts with a broad and systematic literature review on BSS in Chapter 2. Here, a historical overview of urban transportation is provided in order to situate the rise in micromobility within cities, including the primary benefits and drawbacks that have been identified. The chapter then explores the literature that aims to provide data-driven insights into the mobility dynamics of these systems at different spatial and temporal resolutions, enabling the identification of the opportunities available for further research.
1.2. Thesis Structure

Having established BSS as a promising innovation that provide a sustainable alternative means of urban transportation, Chapter 3 explores various data that pertain to BSS in detail, especially those that are openly accessible. In addition to studying the standards and specifications that are used, the chapter introduces the University College London (UCL) BSS data collection that provides the foundations for the analyses presented in subsequent chapters. It details the process through which they were collected and stored, in addition to acknowledging its limitations. The chapter concludes by presenting an overview of the data collection, providing an indication of its exhaustive extent.

Chapter 4 deals with the heuristics of UCL BSS data that remove the errors to produce clean and accurate metrics. These are created in a scalable, automated, programmatic manner that can be used to provide valuable insights into various aspects of each BSS, including their size and use as well as additional confounding contextual information. The chapter closes with a quick exploration of the metrics that have been created, exemplifying their values in providing an understanding of a single BSS.

Chapter 5 then employs these metrics to conduct analyses at a global-scale, classifying over 300 BSS into distinct groups with unique characteristics. This provides a novel and comprehensive overview of the BSS landscape, exemplifying the variations that exist. This demonstrates the values of the data collection and the metrics that are developed in conducting analyses at scale because of the homogeneity in the ways in which they are calculated, together with an indication of the characteristics of systems that are highly utilised which can help to guide the development of existing and future systems.

Chapter 6 investigates the utility of open BSS journey data in offering medium-term, medium-speed changes in system-scale dynamics within the ‘Santander Cycles’ scheme in London. This has been analysed within the context of the Coronavirus Disease 2019 (COVID-19) pandemic that caused major disruption to mobility dynamics in cities as a result of the various restrictions and policies that were implemented to reduce its spread. This presented a unique opportunity to infer
the impacts of the imposition of such rules through the use of longitudinal data to draw comparisons in activity dynamics. The chapter builds on our understandings of mobility during the pandemic by applying a variety of spatio-temporal and network analysis methods that are informative of the mode’s resilience and adaptive capabilities in facilitating healthy and sustainable means of getting around.

In Chapter 7 we delve deeper into the intricacies of BSS dynamics through the granular temporal and spatial exploration of a dockless system in London. The chapter presents a bespoke methodology that leverages the unique structure and attributes of the data in detailing variations in journey dynamics. The analysis enables the delineation of specific locations and times that exhibit significant use which are invaluable given the lack of pre-existing literature, as well as their practical potential in informing the operation, management and improved efficiency of the system.

Finally, the value of publicly accessible BSS data are considered holistically in Chapter 8; consolidating, discussing and evaluating the principal findings from this work. Key methodological and knowledge contributions are highlighted in the context of BSS dynamics, paying particular attention to the role of these data in future research efforts in providing novel, valuable and actionable insights that can help to perpetuate the positive implementation and use of micromobility modes in our future cities.
Chapter 2

Literature Review

Throughout the literature on urban mobility and smart cities there is a ubiquitous and overarching narrative that identifies increases in urban population as a major factor in mounting pressures within urban environments (Mátrai and Tóth 2016; Khan et al. 2013; Petrolo et al. 2017; Rathore et al. 2018; Cheng et al. 2015). In the period between 2000 and 2021, the global urban population increased from 47% to 57% (World Bank 2022), with projections suggesting that two-thirds of the global population will live in urban areas by 2050 (Ritchie and Roser 2018). Such growth has occurred in conjunction with a dramatic increase in the amount of energy that is being consumed, the majority of which has been fuelled by unsustainable and environmentally degrading fossil fuels (IEA 2021). World consumption of energy increasing from one billion gigawatts in 1990 to around 10 billion gigawatts in 2014 has been attributed to a threefold increase in both world population and average per capita use (Bilgen 2014). Being the fastest growing energy sector, a significant proportion of this escalation has occurred within the transportation sector, expanding by 37% between 1990 and 2005 (Holmberg et al. 2012), accounting for 37% of global CO₂ emissions (IEA 2022) and one of the few industrial sectors where emissions are still growing (WBCSD 2001). This growth is suspected to be a result of the sector’s strong relationship with economic growth as well as helping to improve people’s quality of life (QoL) (Bilgen 2014; Pachauri 2007).

Such rapid increases in energy consumption and their undeniable links to the transportation sector have therefore been a major contributing factor in the exacer-
bation of negative externalities on cities. These issues traverse all scales and impact all aspects of urban life. For example, at the individual level, increases in harmful emissions have been found to have a very strong association with the number of premature deaths in urban residents (Bell et al. 2004; Caiazzo et al. 2013; Dockery et al. 1993; Jerrett et al. 2009). At the societal level, population growth has increased pressures on transportation networks, especially for individuals using private automobiles, that have exacerbated congestion (Çolak et al. 2016). As a result, the loss of productivity from being stuck in traffic has been estimated to cost the United Kingdom (UK) economy £20 billion (Goodwin 2004) and up to $190 billion per year in the United States of America (USA) (Texas A&M Transportation Institute 2021). From a holistic, long-term, global perspective, increases in transportation emissions have been a major contributing factor on climate change (Chapman 2007).

The detrimental impacts of this growth have meant that humanity is currently at a crucial tipping point that necessitates a shift towards more sustainable lifestyles. Cities and urban environments are at the focal point of such changes, since they are responsible for consuming 75% of the world’s resources and generating of 80% of harmful GhG emissions (Mohanty et al. 2016). McMichael (2000) calls for the urgent development of policies which ameliorate existing urban environmental health hazards and larger scale environmental problems through radical social and technological transportation solutions based on low-impact technologies, social enlightenment and sharing.

The sharing economy is a rapidly emerging sector that revolves around the sharing of underutilised physical assets on digital platforms, enabling the interaction of private sellers and private buyers (Allen 2015; Srnicek 2017; Wallsten 2015). Recent technological advancements, such as the proliferation of internet access and smart phones, have not only accommodated the rise in activity within the sector, but also shifted a large proportion of economic activity to the internet, which has since been coined the new ‘mother platform’ of economic activity (Kenney et al. 2015). Two of the biggest names within the sharing economy - Uber and Airbnb - were
valued at over $30 billion each in 2017 (Isaac 2017). Uber is a platform that builds on the framework of traditional taxi services, whilst other sharing platforms within the transportation sector aim to provide on demand access to vehicles, from cars and vans on platforms such as ZipCar, to individual modes of micromobility such as bicycle and scooter sharing systems (Cohen and Kietzmann 2014). Micromobility can be defined in many ways but within this thesis, the International Transport Forum (ITF) definition is taken; ‘vehicles with a mass of no more than 350kg and a design speed no higher than 45 kilometres per hour (kmph), limiting the vehicle’s kinetic energy to 100 times less than the kinetic energy of a compact car at top speed’ (ITF 2020). These new business models are said to have great promise and potential in promoting the necessary shift in collective consumption patterns towards a more sustainable society (Mi and Coffman 2019).

BSS are one of the fastest growing modes of transportation and have been considered to be a part of the sharing economy sector (Chi et al. 2020; Zhao et al. 2018). BSS broadly refer to the provision of bicycles to enable short-term rental within the service area and have been rapidly deployed in many urban environments around the world, increasing from 13 systems in 2004 to over 1850 in operation today (Meddin et al. 2022). It has been identified as a potential solution to many of these important issues which urban environments face today due to the generally positive outlook on their implementation with numerous studies which identify associated reductions in GhG emissions (Zhang and Mi 2018), peak hour congestion (Wang and Zhou 2017) and improvements to the individual and societal health (Otero et al. 2018). They also play a role within growing discourses surrounding smart cities, which is yet to have a universally accepted definition (Al Nuaimi et al. 2015), but commonly refers to the integration of information and communication technology (ICT) and data with traditional infrastructure that aim to improve all aspects of urban environments from efficiency to equity and sustainability (Batty 2012; Kitchin 2014; Mohanty et al. 2016; Kumar Debnath et al. 2011; May et al. 2001; Khan et al. 2013). This rhetoric has increased in popularity as the amount of data that are being produced, estimated at 2.5 million terabytes per day globally (Marr 2018), has led
to a ‘data deluge’ age that exceeds our capacity to analyse and derive insights from them (Baraniuk 2011; Bibri and Krogstie 2018).

This chapter provides a comprehensive review of the literature related to the growing pressures of urban environments and the role of BSS in moving towards a more sustainable transportation future. Section 2.1 provides a holistic account of the changing urban transportation landscape, presenting the historical growth and dominance of private automobiles and the way in which new modes of urban micromobility are facilitating a transition towards more sustainable urban mobility futures. Section 2.2 will then take a deep dive into the current literature surrounding the analysis of BSS dynamics at different geographic and temporal scales. Through such a review, this section highlights the gaps in research that have guided the analyses conducted in this thesis. Finally, this chapter closes by summarising those findings throughout such literature in Section 2.3, helping to situate these studies among broader rhetorics of smart cities and sharing economies.

The review suggests that there are numerous areas of research on BSS dynamics which are yet to be fully understood. There is a need to explore these urban mobility dynamics at different temporal and geographic scales in order to best support their implementation, management and success in building a more sustainable transportation future for urban environments (Midgley 2011; McMichael 2000; Abduljabbar et al. 2021).

2.1 Urban Mobility: The Past, The Present and The Future

This section aims to provide a holistic chronological account of the way in the urban mobility has transformed. Section 2.1.1 starts by providing a brief history of how the transportation sector has become such an issue due to historical dependence on private automobiles. Section 2.1.2 presents the literature surrounding the growth and adoption of BSS in recent years as a means to resolve these issues, drawing particular attention to their merits and limitations on achieving such goals. Section 2.1.3 closes by summarising how these new modes of urban micromobility play
an important role in transitioning towards a more sustainable transportation future under the growing discourses on MaaS.

2.1.1 The Past: The Consequences of Private Automobile Domiance

Increases in urban populations worldwide have coincided with the rapid adoption of private automobiles, with car population in the USA growing three-times faster than its human population between 1970 and 2000 (Khisty and Ayvalik 2003). Mattioli et al. (2020) identify 5 key factors which led to the age of automobile dependence; continuous production from the automobile industry, the provision of car infrastructure, urban planning and urban sprawl, the lack of public transportation, and cultures of car consumption. Each of these factors have been studied in depth. For example, the induced demand of automobiles as an outcome of urban planning and design can be exemplified through an account of the development of Rotterdam, the Netherlands, in comparison to other Dutch cities. Historically, Rotterdam was very similar to other cities in the nation in terms of its design. However, due to its prominence as an important port, it was a target for Nazi bombing during the Second World War. Rotterdam was destroyed, presenting an opportunity for city planners to redesign the city. In a similar fashion to many cities in the USA, urban planners anticipated the rise of the automobile and thus the new urban design catered heavily towards automobile transportation (Nijssen 2016; Hull and O’Holleran 2014). This differentiation has been found to influence mode choice among residents, with Rotterdam showing a negative relationship with cycling as compared to Amsterdam, due to its car oriented design (Ton et al. 2019).

Though this dependence is likely a function of convenience and ease of door-to-door travel (Lesh 2013) and it could be argued to be unavoidable for social inclusion in developed countries (King et al. 2019; Mattioli 2016; Lucas 2012; Lucas et al. 2016), there is a growing necessity to move beyond this dependence due to the unsustainable and inefficient nature of the private automobile.

Congestion is one of the primary and widespread negative externalities associated with heavy automobile dependence, to the extent where it has been considered
to be one of the ‘plagues of modern life’ (Arnott and Small 1994, p. 446). Private automobiles are one of the most inefficient modes of transport (Bigazzi 2019), since the majority of trips made tend to be by individuals, leaving many empty seats (Oliphant et al. 2013) as well as being idle for 95% of the time (Frenken and Schor 2017). This is incredibly space and energy inefficient and as a result, has been associated with four major costs: economic losses due to extra travel time, increased environmental pollution, traffic accidents and fuel consumption (Luo et al. 2007).

Another outcome of such widespread dominance of automobiles lies in the way that they have shaped our cities, especially in relation to urban sprawl. Urban sprawl commonly refers to the tendency of reducing city densities as their footprints expand (Nechyba and Walsh 2004). Although cities exist to eliminate transportation costs for goods, people and ideas, the widespread adoption of automobiles have reduced such costs, making outward expansion much more attractive and viable (Glaeser and Kahn 2004; Mattioli et al. 2020). The development of dense walking cities of the 18th century, like Boston and New York, in contrast with newer cities like Houston and Los Angeles are prime examples of such shifts in the urban planning and design process (Glaeser and Kahn 2004; Wachs 1993). In addition, such ‘autodependence’ have meant that many cities have established minimum parking space requirements for new buildings, further exacerbating the issue (Bart 2010).

Urban sprawl was originally associated with significant improvements to QoL, but more recent thinking have highlighted its impacts of social segregation and environmental damage (Glaeser and Kahn 2004). This has occurred in conjunction with greater awareness of those negative impacts that have been highlighted throughout this section, that have led to claims that the age of automobile dependence has come to an end, with economic, environmental and cultural awakenings that have fuelled such changes in perspective (Newman and Kenworthy 2015). As a result, there has been a collective effort to create policies and incentives that aim to reverse the impacts of this historical auto-dependence.
2.1.1.1 Urban Policies to Combat Automobile Domination and Impacts on the Environment

Though the private automobile still dominates our cities today, there has been a growing catalogue of policies that attempt to disincentivise the use of automobiles in favour of more sustainable modes of transportation, both directly and indirectly and at the international, national and local scale.

At the international level, the formation of the United Nations Framework Convention on Climate Change (UNFCCC) in 1992 has displayed the international prioritisation of climate change among the 197 nations who have signed the treaty. The treaty called for ongoing research on climate change in addition to regular meetings to negotiate future policy agreements to stabilise GgH emissions in the atmosphere. The Kyoto Protocol of 1997 was the first meeting, within which parties agreed to reduce the onset of global warming by reducing GgH concentrations to ‘a level that would prevent dangerous anthropogenic interference with the climate system’ through the monitoring of seven GgH. This has since been updated and clarified by the Paris Agreement that was signed in 2016. The primary outcome was a collective agreement to keep the rise in mean global temperature to below 2°C above pre-industrial levels, which would require emissions to be cut by roughly 50% by 2030 (Rogelj et al. 2016). In addition to the Paris Agreement, the COP26 event held in Glasgow, Scotland in 2021, enhanced commitments towards this goal, which has resulted in many ambitious plans at national and local scales in order to meet these targets.

Amongst these macro-scale climate targets that have been set, there are also more granular policies that have been implemented at international scale in order to directly target vehicle emissions. ICCT (2015) provide a complete overview of these policies that have been implemented among G20 countries, which represent over 90% of global vehicle sales. Here they identify four primary programs that have been adopted, that include: low-sulphur fuel standards, tailpipe emission standards for new vehicles, fuel economy and CO₂ standards for new vehicles and voluntary Green Freight programs. These policies have already been found to have
substantial benefits in pushing vehicle emission levels down (ICCT 2017), but the necessity for more targeted national and local policies have been acknowledged (Fajardy et al. 2019).

Therefore, many countries now have legislative and policy frameworks to monitor and manage GhG emissions. Historically, mitigation strategies and policies have primarily been in the form of GhG emission reductions and adaptive strategies dismissed due to their fatalistic outlook (Biesbroek et al. 2010; Schipper 2006) but recent urgency and necessity has shifted this mentality and adaptation has become a key tool in nations’ arsenal of climate change policy. This has been categorised into broad policy strategies including: information generation and sharing, adoption planning, establishing institutional arrangements, and processes for managing and monitoring adaptation action (Nachmany et al. 2019). Though these policies are a step in the right direction, there is still a gap in policy which include enhanced investment in public goods that go beyond hazard early warning systems, explicit reference to building codes and land use planning and incentives and market-based mechanisms to facilitate adaptation (Nachmany et al. 2019).

At the local level, cities around the world have set ambitious goals in order to facilitate sustainable local strategies, developments and policies in order to move towards the global goal of 2°C. This is true of Paris, which aims to become the greenest city in Europe by 2030 through drastic plans to reinvigorate the city’s iconic areas such as the Champs-Élysées and Eiffel Tower by creating grand parks and dedicated cycle lanes (Oliver 2021). London also has its own initiatives aimed at reaching net-zero by 2030 through the implementation of The London Plan of 2021 (GLA 2021b). There has been a heavy prioritisation towards reducing car traffic by at least 27% by the end of the decade to meet climate change targets (ElementEnergy 2022), with a broader objective to increase the proportion of trips made by walking, cycling or public transportation from 63% in 2018 to 80% in 2041 (GLA 2018). This has been complemented by specific policies aimed at disincentivising automobile trips within the city centre through the implementation of the policies such as London congestion charge scheme (CCS) in February 2003. This requires
vehicles which enter the CCS area to pay an entrance fee, with fines for those who fail to do so (Leape 2006). Beevers and Carslaw (2005) investigate the impact of this scheme on vehicle emissions and find significant reductions in $NO_x$, $CO_2$ and $PM_{10}$, as well as increasing bus ridership, showing the value of such policies in helping to meet demanding climate targets. Similar policies have been implemented since 1975 in Singapore, with various forms of implementation such as licences, toll charges or pay per entry like the London CSS (Liu et al. 2014). Within London, in a bid to further disincentivise the use of the most polluting vehicles, the Ultra Low Emissions Zone (ULEZ) was established in April 2019, which was the world’s most stringent emissions zone (Ma et al. 2021). This has been found to play a role in the downward trend observed within harmful vehicle emissions in London, although it is clear that a multi-faceted set of policies are necessary to observe meaningful impact (Ma et al. 2021). The ULEZ zone was recently expanded in October 2021 to encompass an area 18 times its original implementation (GLA 2021a), exhibiting the continuing efforts and increasing stringency of policies in an effort to meet climate targets. Though a great deal of advancements have been made in relation to policies that aim to reduce private automobile use in cities, the research surrounding these policies, their implementation and the subsequent impacts have been continually monitored and evolved to foster the effective implementation (Gallego et al. 2013; Liu et al. 2014; May 2013; Parry 2009).

In combination with these direct methods of controlling the emissions, indirect methods which encourage the use of more sustainable modes of transportation by making them more accessible and attractive have also seen recent adoption in many nations, especially at the local scale. Built environment factors such as density, land use diversity, pedestrian oriented design, destination accessibility and distance to transit hubs have been found to reduce car use in favour of PT, walking and cycling (Ewing and Cervero 2010). These results suggest that transforming our built environment to cater towards these sustainable individual active modes and mass transport options would likely result in greater uptake and increase demand among these modes. The primary issue with these strategies are that they are time-consuming
to implement, cost vast sums and can cause disruptions to existing flows in the short term in comparison to those direct top-down approaches. Though this is the case, cities like London have had policies which aim to increase the attractiveness of these modes for decades, with the implementation of the London BSS and cycle superhighways in 2010 being a direct result of this policy planning (Buczynski 2018). There have been rising efforts to further foster sustainable modes through the continual expansion of these policies which were fast-tracked in the wake of the COVID-19 pandemic. A £250 million package for pop-up active transport infrastructure was established at the beginning of the pandemic (Reid 2020), resulting in an additional 90 kilometres of cycling infrastructure constructed within a year (O’Malley 2021). Similar rapid construction of cycling infrastructure has occurred across a large collection of cities, including 106 European cities, that resulted in an increase in cycling activity between 11 to 48% and is predicted to generate $1 to $7 billion in health benefits per year (Kraus and Koch 2021).

The COVID-19 pandemic has also promoted the consideration of new urban planning concepts such as the 15-minute city (Moreno et al. 2021; Pozoukidou and Chatziyiannaki 2021; Sisson 2020). The concept was first proposed by Carlos Moreno in 2016 and builds on the themes of smart cities to build human-centric and environmentally sustainable urban futures through the provision of all daily needs including work, housing, food, health, education, culture and leisure, within a 15-minute walk or cycle (Moreno et al. 2021). Such an urban design would eliminate the need for private vehicles and have numerous positive externalities including better health, QoL, greater inclusivity and equitability in addition to boosting local economies. Though there was no initial traction, the pandemic appears to have revitalised and reinvigorated interest, with many cities adopting policies to move towards this model. For example, the concept has been added as a blueprint for post pandemic recovery among C40 cities (a global network of mayors from nearly 100 cities) (C40 2022). The most prominent and notable adoption is in Paris, where mayor Anne Hidalgo collaborated with Carlos Moreno in her successful 2020 re-election campaign (Willsher 2020). This has recently been realised
by the announcement of efforts to make Paris 100% cyclable by 2026 through the construction of 180 kilometres of additional dedicated cycle lanes and tripling the number of bicycle parking spaces (O’Sullivan 2021). Similar efforts have been observed in other cities such as New York, announcing an investment of $904 million over five years to install 250 miles of protected bike lanes by the end of 2026 as part of the NYC Streets Plan (Surico 2022b).

Considering the scale and variety of policies presented here, it becomes apparent that there is a strong urgency and necessity to move away from unsustainable reliance on private automobiles and promote more sustainable mobility options. Increasing the use of PT modes are vital for the future (Ceder 2021) and one way in which we can do so is through the implementation of BSS, which has been found to have numerous benefits, including uptake and efficiency of urban PT networks (Yang et al. 2017).

### 2.1.2 The Present: The Rise of Urban Micromobility

Although BSS have only seen mass adoption in urban environments worldwide within the last 15 years, there is a relatively long history spanning over half a century that depicts a story of technological advancements facilitating a solution to increasing pressures from urban areas in relation to congestion and their environmental impacts. Among the literature that examine the history, development and evolution of BSS, this progression has been classified into distinct generations, each of which have unique characteristics (Beroud and Anaya 2012; DeMaio 2009; Fishman 2016; Parkes et al. 2013).

The first-generation of BSS were initially implemented by a group of activists in Amsterdam in 1965 to combat exacerbating air pollution and consumerism (De-Maio and Gifford 2004; Davis 2014; Zee 2016). These systems were based on a small fleet of bicycles that were painted white (hence dubbed the ‘White Bikes’), unlocked and placed throughout the city to help facilitate journeys to those without access to a bicycle for free. Similar systems were also adopted by other cities such as the La Rochelle system in France in 1974, which were called the ‘Yellow Bikes’, and in Cambridge, UK in 1993, which was called the ‘Green Bike Scheme’
Unfortunately, due to the unrestricted access to these bicycles, these systems were very susceptible to theft and vandalism which made them unsustainable financially and ultimately led to their demise.

The second-generation of BSS developed on the short-comings of the first-generation, aiming to mitigate the issues of theft and vandalism through a coin-deposit system, where bikes were kept locked until a refundable fee was paid (DeMaio and Gifford 2004; Shaheen et al. 2010; Pucher and Buehler 2008). The first of these systems were implemented in Copenhagen, Denmark, in 1995 and were called the ‘Bycyken’ or ‘City Bike’, with 1,100 bicycles and designated city parking racks (Shaheen et al. 2010). Despite the implementation of the coin-deposit system, theft remained a fundamental issue in facilitating the longevity of these systems. This vulnerability was attributed to the lack of a registration process that enabled users to stay anonymous, meaning that operators were not only unable to track down stolen bicycles, but also unable to tie anyone to those stolen bicycles and punish them (Frade and Ribeiro 2014). Although the first- and second-generation of BSS offered an innovative opportunity for urban residents to cycle, the system failed to provide a reliable service, hindering its ability to generate a significant modal shift.

The third-generation of BSS saw the most significant implementations of technology to, again, mitigate against the pitfalls of the preceding generations. The docked BSS we find in many major western cities today, such as the Santander Cycles scheme in London, the Vélib system in Paris, and the Citi Bike scheme in New York are among these third-generation BSS. They are characterised by the implementation of docking stations that require users to pay using bank cards. This enables secure payments and user validation, facilitating more responsible use (Fishman et al. 2013; Trépanier et al. 2004). These ICT developments and increased interest in greener modes of transport in urban areas have seen the implementation of these schemes in many cities around the world (Pucher and Buehler 2008; Bachand-Marleau et al. 2012; Pucher and Buehler 2012). The docking stations within these systems integrate wireless communication technology that enables operators to track the occupancy of these docking stations in near real-time (Midgley
2.1. Urban Mobility: The Past, The Present and The Future

which not only allows operators to locate their fleet, but also provides the opportunity for users to check the availability of bicycles via applications to plan their journeys. This has enabled those BSS to become an integrated part of existing forms of urban mobility (DeMaio 2009; Munkácsy 2017), with many docking stations being located in close proximity to transit hubs to help facilitate multi-modal journeys. Part of their widespread adoption among these cities is attributed to these BSS ability to help resolve the ‘first/last mile issue’, wherein travellers in urban environments using PT are unable to make complete door-to-door journeys due to distances between PT hubs, such as bus and train stations, at the start and end locations of their journeys (Fishman 2016; Saberi et al. 2018; Shaheen et al. 2010).

The fourth-generation of BSS have only recently emerged in 2015, spreading to a large number of cities, especially in China (Sun 2018), and are characterised by their dockless nature (Márai and Tóth 2016; Parkes et al. 2013; Fishman 2019). These systems allow users to pick-up and drop-off bicycles without the necessity to park them in designated docking stations and can be tracked through embedded global positioning system (GPS) receivers in each bicycle, enabling them to be much more flexible in trips compared to third-generation BSS (Shaheen et al. 2013). This has coincided with a rise in implementation of electric pedal-assistance within these bicycles, which has provided further motivation for new users to uptake this modernised mode of micromobility as it reduces the amount of user physical exertion. As a result, there are some debates on whether the deployment of electric bicycles (e-bicycles) necessitate the need to create a new fifth-generation of BSS (Guidon et al. 2019), with BSS researchers such as Becker et al. (2017) commenting that the difference is so stark that the past findings may not be transferable to e-bicycles.

Across third- and fourth-generation BSS, there are generally four primary types of operator: PT and local authorities, advertising companies, private for-profit firms and non-profit operators (Médard de Chardon et al. 2017; Midgley 2011). Building on the successes of small scale private operations, PT authorities were generally the first to make large investments in BSS, such as the Montreal Bixi
BSS in 2008 and the London Santander Cycles system in 2010. As suggested by the name of the London BSS, in order to maintain operations such systems have since transitioned to using advertising as a means of finance, signing sponsorship agreements with various firms as well as directly to advertising agencies such as JCDecaux and Clear Channel that supply the city with bicycles for free or at a discounted rate in return for advertising privileges (Midgley 2011). The operational model of BSS has since expanded with the success of these large scale systems to private profit seeking companies that have been backed by large sums of capital investment such as Lime and Nextbike.

The recent success of BSS in many cities around the world have also birthed the next iteration of urban shared micromobility from similar profit seeking firms - e-scooter sharing systems. These typically operate in a similar manner to those fourth-generation BSS, due to their dockless nature, providing a novel alternative mode of urban micromobility that does not require user physical exertion. This mode has also seen rapid implementation and adoption globally, especially during the COVID-19 pandemic, sharing many of the benefits that BSS present (Dias et al. 2021). The continual growth and success of this urban transportation sector, currently valued at over $40 billion and projected to reach between $195 billion (Valuates 2021) and $500 billion (McKinsey 2020) by 2030, are a testament to their trajectory and longevity as a revolutionary shift towards more sustainable forms of urban mobility.

2.1.2.1 The Benefits of BSS

The mass adoption of BSS and other forms of micromobility have been attributed to the numerous benefits that are associated with these modes of transportation. Academic studies have sought to quantify these benefits in an effort to better understand the true value of these new modes. Shaheen et al. (2010) summarises them into six broad categories: greater mobility flexibility, a reduction in harmful GhG emissions, individual financial savings, reduced fuel use and congestion, numerous health benefits and greater support for multi-modal transport connections through the facilitation of the first/last mile problem. Here, we can begin to get a sense of
Individual Scale Benefits

When considering the individual perspective, positive health outcomes as a result of increased physical activity are the most dominant within the literature. Though there is no doubt that exercising and being active has positive impacts on health, there is a mounting pool of academic literature that exhibit the negative outcomes of poor air quality on health, with ambient air pollution leading to more than 4 million premature deaths annually worldwide (Hoffmann 2019). These deaths are a result of air pollution’s association with life-threatening health outcomes such as cardiovascular diseases, chronic respiratory diseases and lung cancer, in addition to impacts on unborn children to name a few (Rojas-Rueda et al. 2013; Thurston et al. 2017; Manisalidis et al. 2020). This places a question mark over whether the benefits of physical activity has a net positive impact on health or whether increased exposure to harmful emissions would do greater harm due to the close proximity of BSS activity to road traffic emissions.

Due to this ambiguity, there are a number of papers which aim to analyse the costs and benefits of cycling in cities. Among these studies, there appears to be strong evidence to suggest that the benefits of physical activity outweigh the detrimental effects of air pollution (Bauman et al. 2017; Mueller et al. 2015; Raser et al. 2018), with only those cities in the top 1% of air pollution levels in the World Health Organisation (WHO) ambient air pollution database to have negative impacts associated with active modes of transport (Tainio et al. 2016). Otero et al. (2018) investigate the health benefits of BSS across 12 European cities and discover similar findings, in addition to estimating that if 100% of BSS trips were to replace car trips, up to 73 deaths each year with an economic value of €225 million could be saved. Even in the most conservative scenario at the minimum number of reported car trip substitutions, they estimate that 5 deaths per year with a value of €18 million could be saved. Together, these conclusions depict the value of this new active mode of travel in helping to reduce the negative outcomes of poor air quality on individual
When cycling, the risk of injury and accidents are also an important consideration. By combining hospital injury data with ridership data across North American and European cities, Fishman and Schepers (2016) find that the introduction of BSS in cities was found to be associated with a reduction in cycling injury risk, in addition to the fact that BSS users were less likely than other cyclists to sustain fatal or severe injuries. Similar results are also observed within the London BSS, where Woodcock et al. (2014) were able to show that the risks of injury whilst using the BSS were lower than those estimated for cycling in general, though these benefits were less pronounced among women and older users.

Taken together, it becomes evident that the presence of a BSS has clear benefits for the individual, reducing health risks associated with poor air quality within these environments and making cycling a safer activity in comparison to private cycling activities. In conjunction to these individual health benefits, we find benefits that extend to the wider local populations through the implementation and use of BSS.

**Local Scale Benefits**

These wider local benefits are likely another strong contributing factor in their widespread adoption. One such benefit is the ability for BSS to serve as a mode for first- and last-mile connectivity, helping to increase PT use and reduce road congestion. BSS provide a mode that can not only substitute short private automobile journeys, but also facilitate multi-modal journeys with integration to alternative modes of PT, boasting its merits as a flexible, financially accessible, on-demand urban mobility service. Fishman et al. (2014a) investigate the impacts on BSS on car use within cities and find that there is a reduction in the majority of cities studied except for London. Similarly, Wang and Zhou (2017) examine the influence of BSS on congestion across 96 cities in the USA and find mixed results, though the impacts of reducing peak-hour congestion can be found to be significant, exhibiting the value of the system in helping to facilitate commuting travel and reduce road traffic during the most congested periods. Not only does the mode help to reduce congestion during these hours, BSS are also found to be a faster mode of travel
during these peak hours for short journeys under 3km in New York City compared to taxi services (Faghih-Imani et al. 2017), further demonstrating its merit as an efficient mode of transport.

Fan and Zheng (2020) investigate the impacts of the dockless BSS in Beijing, finding that subway lines with a higher BSS intensity exhibited an 8% larger growth rate in subway ridership compared to ones with lower intensity, in addition to reducing rush hour road congestion by 4% for those stations with BSS trips in the highest quartile. Many studies have also found significant associations between dockless BSS within close proximity to subway stations (Guidon et al. 2019; Shen et al. 2018; Xu et al. 2019), further showcasing BSS ability to help facilitate these multimodal journeys and, in turn, reduce travelling times for those users who would have otherwise walked whilst also removing traffic on the roads from those users who would have otherwise driven.

Not only do BSS help to serve as a ‘feeder mode’ for PT (Fan et al. 2019; Guo et al. 2021), they also serve as a mode to help build transport resilience within cities (Cheng et al. 2021). This term refers to the ability of transport networks to withstand the impacts of disruptive events (Cox et al. 2011), helping to contribute to the society’s economic resilience by building adaptive capabilities to deal with vulnerabilities within the network (Darayi et al. 2019). With disruptions to PT services becoming more common (Lin et al. 2016), BSS can help to serve as an alternative mode of transport during such disruptive events. Yang et al. (2022) and Saberi et al. (2018) study the impacts of London Tube strikes on the use of the BSS, finding significant evidence to suggest the modes’ flexibility in facilitating journeys during such disruptive events, with an 85% increase in journey counts, 88% increase in average journey duration and increase in BSS network connectivity.

Global Scale Benefits

The ability for BSS to help facilitate a reduction in road traffic as well as increase the amount of PT uptake is not only beneficial for the local population, but can also have wider implications for the environment. Since BSS are either powered by the user or have electric motors which assist user pedalling, they do not
emit any GhG upon their use. A shift in modal choice by a significant proportion of the population from unsustainable modes, such as private vehicles, can help to reduce the amount of harmful emissions. In New York, the Citi Bike scheme has been estimated to reduce CO\textsubscript{2} and NO\textsubscript{x} by 493 imperial tons and 1,252 kg respectively between 2014 and 2017 (Chen et al. 2022b). Similarly in Shanghai, the BSS is estimated to have saved 8,258 tonnes of petrol and decreased CO\textsubscript{2} emissions by 25,240 tonnes in 2016 (Zhang and Mi 2018). A recent study by Saltykova et al. (2022) highlights the importance of considering which modes are being substituted by BSS use because presuming the substitution of car journeys may lead to severe over estimations. They demonstrate this when comparing the estimated reductions in CO\textsubscript{2} between PT and car substitution in Chengdu, China. Here, the BSS is estimated to substitute 27.9% of bus journeys and 8.1% of subway journeys, resulting in a net reduction of 2,564 kg compared to an estimated reduction of 4,125.13 kg when substituting car journeys, showing the likely overestimation that have been made in previous studies (Saltykova et al. 2022). In either case, these studies demonstrate the benefits of using BSS in helping to reduce harmful GhG emissions within cites.

In summary, there are a multitude of benefits associated with the implementation and use of BSS which include individual health and safety benefits, local reductions in congestion, accessibility, travel times and transport resilience as well as wider environmental benefits. These examples showcase the quantified benefits associated with BSS that likely play a significant contribution in making them an attractive option in transitioning to a more sustainable form of urban mobility.

2.1.2.2 The Drawbacks of BSS

Although there are numerous benefits associated with BSS, it would be naive to not consider the current limitations of their implementation and the negative externalities that may emerge as a result. Much like those positive externalities, academics have sought to analyse these drawbacks in order to better assess the costs and benefits of such a system. Griffiths et al. (2019) discuss the negative impacts resulting from the relatively novel and unregulated nature of the sharing economy, which includes the BSS sector. Within their discussions, the authors refer to a number of
negative externalities that are associated with BSS, including safety risks from the use and storage of bicycles, in addition to issues surrounding vandalism and bike disposals (Griffiths et al. 2019). Here, they emphasise the necessity to regulate these systems in order to limit any further negative impacts on the natural environment and societal well-being. Additionally, Shui and Szeto (2020) discuss issues pertaining to the bicycle-sharing service planning problems (BSPP) and summarise these issues into eight steps that still require great attention. These areas of attention are: bicycle infrastructure network design, bicycle station design, fleet sizing design, static bicycle relocation, static demand management, inventory level management, dynamic bicycle relocation and dynamic demand management. Taken collectively, it is clear to see that there are still many areas that require greater research and scrutiny in order to help facilitate a more efficient and equitable implementation and use of these BSS.

Implementation Challenges

We can take a closer look at some of these implementation issues by looking at some case studies. The initial implementation of fourth-generation dockless BSS saw many issues that caught the attention of popular media. One such example is the implementation of the MoBike system in Manchester in the summer of 2017. It was a momentous moment, with this system being the 100th city in which MoBike had commenced operations and the first city outside of Asia (Sherriff et al. 2020). Though there was great optimism, the system was quickly forced to suspend operations and withdraw from the city due to the extreme vandalism of the bicycle fleet, with bicycles being stolen and dumped in an obvious attack against its operation (Cox 2018). This came after only 13 months in operation, displaying the necessity to consider the social and geographical fit for each city, rather than assuming a one-size-fits-all policy (Sherriff et al. 2020). There have also been issues in the implementation of dockless BSS in China due to the vast size of bicycle fleets which have been deployed in cities around the country. In 2018, there were approximately 23 million bicycles among the 77 dockless BSS companies operating in China (Han 2021). There is increasing discourse surrounding ‘bike litter’, which
refers to bicycles which have been parked or abandoned in pathways, rivers and other public places which cause visual pollution but also create serious safety and public nuisance hazards (Chen 2019). These examples exhibit the failures of BSS implementation at the cost of the wider society and as a consequence, many BSS operators have seen greater regulation on their systems. One such way operators have motivated better user management of bicycle parking locations is through the use of financial incentives for those users to park within designated geofenced parking areas in combination with fines for the operation or parking of bicycles outside of these areas (Zhang et al. 2019).

Rebalancing Challenges

Even after the implementation of BSS, there are still issues that operators face in order to run a financially and environmentally sustainable operation. One of the largest operational challenges amongst BSS are ensuring that bicycles are located appropriately according to user demands. This is a homogeneous issue among both docked and dockless BSS and has been coined the bicycle-sharing rebalancing problem (BRP) (Dell’Amico et al. 2014). Due to the large financial operational costs of continuously relocating bicycles, a great amount of research into optimising the BRP has been conducted (Vallez et al. 2021). There have been various approaches to find optimal solutions, with the implementation of traditional travelling salesmen problem and vehicle routing algorithms as well as more novel machine learning and network graph analysis methods (Vallez et al. 2021). Although each system has their own characteristics, there are generally two main areas of research that require great attention to solve such an issue: the spatial and temporal forecasting of demand and the routing of rebalancing operations (Alvarez-Valdes et al. 2016). This is an ongoing issue that is continually being researched.

Fishman et al. (2014a) investigate the costs of these rebalancing tasks within London, finding that for each kilometre the BSS reduces in private car use, 2.2km of operator redistribution efforts were necessary. This demonstrates the underlying costs of the operation, that needs to be addressed and optimised in order to achieve a more sustainable system. Though this is the case for London, the authors also
investigate the Minnesota, Melbourne and Washington, D.C. BSS, showing significant reductions in estimated vehicle kilometres travelled.

Although there is a large pool of literature that seeks to optimise BSS operations, Luo et al. (2020) criticises this current framework as it does not optimise the systems based on their sustainability but rather prioritising operational efficiency. The authors suggest a reframing of the BRP to consider the life cycle GhG emissions, from the construction of bicycles and docks to their disposal at the end of their service. In their investigation into the Xiamen BSS in China, they find that only 15% of the current bicycle fleet is necessary to serve the same demand in addition to more frequent and optimised bicycle rebalancing (Luo et al. 2020). A holistic approach in considering the sustainability of BSS can help to reduce the GhG emissions from excessive bicycle manufacturing and inefficient frequency rebalancing which increases fuel consumption. Similar life cycle assessment (LCA) have been conducted comparing the emissions produced by dockless e-bicycles in comparison to docked pedal bicycles, with findings suggesting that 3 times the amount of GhG are produced by dockless e-bicycles over its lifetime, equating to 2.7 times the environmental impact (Bonilla Alicea et al. 2019). This means that an increase in e-bicycle ridership by a factor of 1.8 is necessary to overcome increased environmental impact. This exhibits the costs of manufacturing excessive bicycle fleets, especially among dockless e-bicycles, and the necessity to further optimise BSS in order to facilitate the most sustainable implementation of these systems.

**Equitable Accessibility Challenges**

BSS are also credited to be an equitable and accessible mode of urban mobility for all residents within the service region, but studies which investigate these have found that this may not be the case (Gavin et al. 2016; Médard de Chardon 2019). These studies suggest that BSS typically serve more privileged populations among BSS that were assessed in Europe and North America (Médard de Chardon 2019), with mounting evidence suggesting that BSS typically serve a younger, healthier, wealthier, white male population (Bauman et al. 2017; Buck and Buehler 2012; Fishman et al. 2013; Goodman and Cheshire 2014; Ogilvie and Goodman 2012;
This may be partially a function of the service area of many BSS, which are typically limited to denser urban cores associated to such demographics (Ogilvie and Goodman 2012). As a result, there is a growing plethora of research that aims to identify ways to implement and expand BSS in locations that would facilitate more equitable and inclusive use (Caggiani et al. 2017; Caggiani et al. 2020; Conrow et al. 2018) in combination with using crowdsourcing as a way to identify optimal positioning for docking stations (Griffin and Jiao 2019; Piatkowski et al. 2017; Zhang et al. 2016a).

In summary, although the discourses surrounding BSS have been dominated by their numerous benefits, as highlighted in Section 2.1.2.1, it is important to remain vigilant to issues that arise as a result of their implementation. Operators should be held responsible for such negative externalities, aiming to provide the most efficient, environmentally friendly and equitable services. As a result, it is clear that further work on such issues are necessary to ensure BSS as an enduring staple mode of future urban mobilities.

2.1.3 The Future: A Seamlessly Integrated Transport Model - Mobility-as-a-Service

There are currently many conceptualisations of future urban mobility, which include the uptake and integration of many promising new technologies such as electric vehicles (EV) and autonomous vehicles (AV) as well as great excitement surrounding these technologies due to the multitude of benefits they bring including impressive acceleration, fuelling at home, lower fuel costs, lower maintenance costs, less noise and less vibration and their reduction in GhG emissions (Barton and Schütte 2017). The climate benefits of EV are estimated to increase over time, saving 1.5 billion tonnes of CO₂ in 2050 (Wolfram and Lutsey 2016). Consequently, the uptake of this attractive emerging technology has been supported at the national level in countries such as Norway and Germany, as well as at the state level in the US such as California. These policies include subsidising purchase costs, reducing maintenance and operational costs, such as congestion pricing exemptions, and special privileges including the use of bus lanes (Barton and Schütte 2017). Using data from 2010...
to 2018. Rietmann et al. (2020) predict that EV will constitute 30% of worldwide passenger vehicles by 2032.

Although the use of EV are set to continually grow, given that the majority of energy production is facilitated through the use of finite and environmentally degrading fossil fuels, worldwide CO$_2$ emissions will continue to increase until 2035 (Barton and Schütte 2017). This highlights that the adoption of EV and their ability to impact the environment positively hinges on the transformation of the energy sector to green, sustainable modes of energy production. In combination with these energy production shifts, much like the LCA of BSS vehicles, it is important to consider the production costs of EV, which primarily rely on battery technology. Lithium-ion batteries are the highest commonly used battery technology in most portable consumer electronics, including EV. Peters et al. (2017) conduct a review of LCA studies on these batteries, finding that producing one watt-hour of storage capacity is associated with a cumulative energy demand of 328 watt-hours, causing GhG emissions of 110g CO$_2$ equivalents. In addition to the production costs, such vast scale ups must consider the availability of the natural resources necessary to produce batteries. Though studies have shown that the materials for lithium-ion batteries will likely meet the demand in the near future, there are potential geopolitical issues associated with cobalt (Olivetti et al. 2017), a key material in stabilising batteries and boosting their energy density. These issues primarily result from the monopolisation of the cobalt supply by State-led Chinese investments in cobalt mines in the Democratic Republic of the Congo (DRC) (Searcey et al. 2021). This may hinder the production efficiencies and costs of these batteries outside of China (Olivetti et al. 2017) and failures to account for these offshore production costs of batteries or disposals would mean that emissions are underestimated and displaced globally (Henderson 2020).

Henderson (2020) suggests that we are currently faced with an ‘epochal decision’ on future mobilities and that there is little room for tolerance given the precarious tipping point of our planet’s climate system. Though EV do have great potential to replace internal combustion engine vehicles, there are some clear limitations to
their implementation. Timms et al. (2014) criticises these EV-centric conceptualisations of the future as they fail to consider a socially sustainable view of transport in a future utopia. Similarly, many academics have stated that EV alone cannot provide a sustainable solution to future transportation demands and that there is a necessity to increase the uptake of PT and other shared modes of mobility (Ceder 2021; Henderson 2020; ITF 2018). Kane and Whitehead (2017) suggest that the advent of AV in combination with a shift to the sharing economy has the potential to rapidly change mobility patterns and the make up of urban transportation systems away from private ownership of vehicles. Such developments are currently being trialled by companies such as Cruise in San Francisco, who have recently commenced their fully driverless taxi service in the city (Cruise 2022). Ceder (2021, p. 5) suggests that ‘although we cannot change the direction of the wind... we can adjust the sails to create attractive systems that will naturally shift people from the automobile to other forms of travel’, through the application of new technologies within these modes.

Miskolczi et al. (2021) conducted a literature review of 62 papers on future urban mobilities and have formulated four common scenarios among them. These scenarios are: ‘grumpy old transport’ - private vehicles to remain the dominant mode of mobility, ‘at an easy pace’ - private vehicles are still dominant but progress towards greater use of EV and shared mobilities are on the rise, ‘mine is yours’ - a large shift towards the sharing economy model with significant use of shared mobility modes and EV, and ‘tech eager mobility’ - an intense transition to highly technologically advanced and completely sustainable mobility modes including AV, but PT is expected to be the most effective mode. Among these four potential future scenarios, Miskolczi et al. (2021) suggest that the most likely outcomes are ‘at an easy pace’ or ‘mine is yours’. This has been echoed by many other academics in the field, suggesting that shared mobility futures, including greater modal share of PT, are the most promising (Kamargianni et al. 2016; Kane and Whitehead 2017; Nikitas et al. 2017; Shaheen and Chan 2016; Standing et al. 2019; Smith et al. 2018), with a need to reimagine mobility as a commons-based, collective, non-
individualistic approach to mobility as a shared public good (Nikolaeva et al. 2019).

As a result, among the current conceptualisations of the future of urban mobility the MaaS model has seen a significant amount of attention. There has been a longstanding desire from policymakers for a ‘total journey solution’ (Potter and Skinner 2000), that enables travellers to make on-demand door-to-door journeys through a single interface. This convenience is currently facilitated by private modes but MaaS platforms aim to seamlessly combine different modes of transportation into one service which combine trip planning, reservations and payment (Jittrapirom et al. 2017). In an ideal scenario, MaaS eliminates the necessity for individuals to own private vehicles, as the convenience and cost of MaaS facilitated trips outweigh those of private modes. Micromobility modes such as BSS will likely compose part of these MaaS journeys, either facilitating short distance journeys or help alleviate the first- and last-mile issues as part of longer multi-modal journeys. As a result, practitioners and researchers call for further research to understand the role of these new modes of micromobility within such a framework (Crozet et al. 2019), emphasising their vital importance in these early stages in helping to provide quantified evidence (Kamargianni and Matyas 2017). Currently research into MaaS can be summarised among three broad categories (Utriainen and Pöllänen 2018): the roles of different transport modes and services, the analyses of MaaS pilots and trails and the expected impacts of MaaS. There also appear to be analyses on particular aspects of MaaS services such as improvements to the digital service platforms (Ruutu et al. 2017), fleet optimisation (Dimitriou et al. 2016), surge pricing and labour supply (Zha et al. 2017), and exploration of demand dynamics (Kourti et al. 2017).

There are currently a few small-scale examples of MaaS systems in operation that help to envision future urban mobilities. Whim, a MaaS application created by a Helsinki based firm, MaaS Global, is ‘the first true MaaS operator... [that] provides... a true alternative to car ownership’ (Whim 2022). The platform is currently in operation in six countries and shows promising potential to expand as they make connections with a greater number of transportation operators and further in-
tegrate them into their services. This being said, MaaS is still in its infancy, as highlighted by the recent demise of two high-profile MaaS endeavours, Kutssuplus in Helsinki and Birdj in the USA, both for financial reasons (Djavadian and Chow 2017). There is great potential for MaaS to offer a paradigm shift from transport being fundamentally provider-led to being a fully user-led system whereby the level and type of transport supply continually adjust in response to the desires of travellers. As a result, there is growing research within the field and a call to make policymakers more proactive in supporting and facilitating the growth of this new sector in making connections and leveraging access to finance and other resources as a partner agency (Snellen and Hollander 2017; Miles and Potter 2014).

BSS and other novel forms of shared mobility have therefore transformed the way in which we imagine the future of urban mobilities. Their growing adoption and success globally are the first steps towards a sustainable transport revolution. As a result, having a deep understanding of BSS dynamics across various temporal and geographical scales are paramount in anticipating the adoption of alternative modes of shared mobilities in addition to their contributions towards future MaaS platforms.

2.2 The Dynamics of Bicycle Sharing Systems

System dynamics (SD) refers to a perspective and set of conceptual tools that facilitate an understanding of the structure and dynamics of complex systems that enable the designing of more effective policies and organisations to manage them (Sterman 2001). The seminal work of Forrester (1958) in the late 1950s laid the foundations of these current conceptualisations of SD. Forrester was influenced by the ongoing advancements in computing technologies and simulations in improving our understanding for strategic decision-making (Richardson 2011). Forrester (1958) developed SD as a powerful methodology that built on the frameworks understood in system theory, information science, organisational theory, control theory, tactical decision-making, cybernetics and military games. Drawing from such a board framework, SD has since been re-applied in many of these fields and also
expanded to many other applications including product development (Sterman et al. 1997), business portfolio simulation (Merten et al. 1987) and distribution of body fluids (Hansen and Bie 1987). Forrester built on these ideas of SD and provided frameworks to understand ever greater and more complex systems including industrial dynamics (Forrester 1961), world dynamics (Forrester 1973) and also urban dynamics (Forrester 1970).

Initial conceptualisations of urban dynamics focused on a holistic approach to analyse the interactions within urban environments, with key factors such as businesses, people and houses which would oscillate at first and then settle into an equilibrium (Forrester 1970). This provided a new framework to understand the long term changes among land use and urban morphology which showcased the complex interactions in operation at a large urban scale. Academics began to integrate and build on this framework and the study of urban dynamics began to blossom and develop further. Batty (1976) goes on to define urban dynamics as a representation of change in urban spatial structure through time which embody a myriad of processes at work in cities on different but often interlocking time scales, ranging from life cycle effects in buildings and population to movements over space and time as reflected in spatial interactions. These different temporalities of urban dynamics have often been classified by the pace of change and often reflect particular parts of the urban environment. Wegener (1986) suggests a framework where we can consider fast urban dynamic processes to consider the mobility of individuals and material goods, medium-speed processes reflect those of socioeconomic and technological changes which do not change the physical structure of the city, but change the activities performed within them and slow processes relating to construction of physical structures. Weidlich (1999) also had similar conceptualisations of the temporality of these dynamics from fast and slow processes. He also identified that urban systems consisted of smaller sub-systems have their own rhythms and rates of dynamism, with many intertwined organisational levels starting from microstructures and ending with macrostructures that evolved over different time scales (Weidlich 1999).
Since these initial conceptualisations of urban dynamics over half a century ago, the term has become the general label to refer to the studies on the temporal movement flows within urban environments. Batty (2012) suggests that the paradigm has changed from these systems being centrally organised from the top-down to one where now we consider systems to be constructed from the bottom-up in a decentralised fashion. This has coincided with the rise of notions that cities are far-from-equilibrium and that understanding these nuanced dynamics enable the prediction and design of more efficient, equitable and sustainable cities (Batty 2012). Continuous advancement of computing processing power, approximately doubling every two years under Moore’s Law (Schaller 1997), in combination with the current data deluge (Baraniuk 2011; Bibri and Krogstie 2018) has enabled studies on the dynamism of urban environments to be analysed in ever more granular detail. Urban transportation systems are among these extended applications of urban dynamic models due to their complex nature with multiple variables and nonlinear feedback loops influenced by social, economic and environmental factors (Priester et al. 2014; Wang et al. 2008). Shepherd (2014) reviews various SD models which have been applied in transportation across over 50 papers since 1994. They find SD to be a useful whole system approach to transport planning, allowing for an opportunity to investigate general dynamic tendencies within these systems that can have real-world impact on business and policy.

Within the context of urban dynamics, this thesis will focus on developing an in-depth understanding of the dynamics of BSS at different temporal and spatial scales which build on the conceptualisations suggested by Wegener (1986). First, the analysis will explore the long-term global evolution of BSS, exploring the dynamics of system growth both morphologically and in terms of its use at the international scale. This will consider the construction of physical cycling infrastructure as well as the expansion of existing and construction of new BSS systems. Then, the analysis will explore changing dynamics in the medium-term, focusing on the impacts of COVID-19 on the mobility patterns within BSS, namely the London Santander Cycles BSS. This explores the medium-speed impacts of various poli-
cies and restrictions imposed during the pandemic, again drawing parallels to the medium-speed conceptualisations by Wegener (1986). The thesis will close by exploring the short-term dynamics of a dockless BSS, providing a detailed insight into the temporality and spatiality of its use, as well as exploring the built environment factors that are associated with these granular, individual mobility dynamics. The following sections will review the associated literature to each of these spatial and temporal scales.

2.2.1 Long-Term BSS Dynamics: The Global Growth, Expansion and Comparisons of Systems

Given the novelty and growth of BSS, research has primarily focused on their practical operation and use in the short-term for individual systems as opposed to any larger global-scale long-term analyses. This can be associated with a number of reasons including the lack of routine data release from BSS operators (Mátrai and Tóth 2016) as well as the lack of established standard for the comparison of BSS (Médard de Chardon and Caruso 2015). As a result, research on long-term BSS dynamics is typically limited to those systems that routinely release journey origin-destination (OD) data. Dock capacity data is the most abundant source of BSS data, as they are typically mined and collected from operator websites and API feeds that are used to inform users of bicycle availability. Unfortunately, they rely on collection by researchers and enthusiasts since they are not stored for public use otherwise. Therefore, unless individuals have collected data routinely for long periods of time, these studies are typically limited to short-term analysis of BSS use and operation.

One important consequence of lack of data availability is that it enables operators to overstate the size and success of their BSS in a bid to create positive press and increase potential for investment (Médard de Chardon and Caruso 2015). This skews public perception of BSS, thus greater clarity is necessary to provide an unbiased review of system growth, use and performance.

Comparisons between BSS provide a way of assessing the relationships between different attributes across systems. This enables us to develop an enhanced understanding of the BSS landscape as a whole, as it allows us to see how the vari-
ables impact BSS performance globally. Previous work in this respect has been limited by poor data availability, therefore few regional comparisons and fewer global comparisons have been conducted. This being said, there are a handful of studies that attempt to provide some insights into system comparisons. For example, Bieliński et al. (2019) conduct a comparative study of 56 BSS in Poland, identifying commonalities in local characteristics and BSS operational dynamics which are associated with system performance. Within their analysis they are able to identify an operator which appears to be more effective compared to others. Similar studies include Zhang et al. (2015) who study five BSS in China, Zhao et al. (2014) who study 69 BSS in China based on their urban features and system characteristics, Kou and Cai (2019) who analyse determinants of trip duration and distance across eight USA BSS and Zaltz Austwick et al. (2013) who compare four BSS in the US and the London BSS based on their flows, community clusters and activity patterns across weekdays and weekends. Sarkar et al. (2015) utilise data from 10 systems in order to identify commonalities in docking station activity. Here they are able to identify four main types of docking station activity: morning sinks, morning sources, equally sources/sinks and minor docking stations.

The scope for substantive international comparisons across BSS is constrained by the inconsistency of BSS data collection practices and formats and the limited availability of data for many systems. O’Brien et al. (2014) were among the first researchers to utilise dock capacity data from an international collection of BSS to create an introductory global comparison of systems. The paper compares 38 BSS from all regions of the world and was the most comprehensive global analysis of BSS at the time of its publication based on their geographical footprint and activity dynamics. The research laid the foundations for the analysis of larger scale transport systems by creating a classification of the different systems and seeks to demonstrate that BSS have a lot to offer both as an effective method of transport and a rich source of data. Médard de Chardon et al. (2017) have since conducted the largest analysis of BSS internationally, investigating the common variables that influence the ‘success’ of BSS. Within this analysis, they utilise the trips per day
2.2. The Dynamics of Bicycle Sharing Systems

per bicycle (TDB) metric as a measure of BSS ‘success’, quantifying the relationship between success (TDB) and five categories of variables that were commonly associated with BSS use among the literature, those being: BSS attributes, density and compactness, geography, weather and transportation infrastructure. The use of TDB as a metric, which is simple to calculate using limited dock capacity data, has since been commonly used to provide some understanding of BSS usage rates when comparing systems. Though the authors express TDB as ‘a comparable measure of success’ (Médard de Chardon et al. 2017, p. 203), this thesis will refer to TDB as a measure of system efficiency and rate of use. The primary issue pertaining to this study is the lack of Chinese systems, which constitute the largest market for BSS globally (Gu et al. 2019). Therefore, analysis which investigates BSS at a truly macro-, global-scale are still very limited in their ability to make inferences about the global landscape of BSS.

These regional and global comparison studies provide a great foundational framework to grasp commonalities and differences among systems at a much greater spatial scale. Though this is the case, the majority of these analyses rely on data within the medium-term, typically utilising a few months of data. In order to grasp an understanding of the growth and changes in the morphology of BSS, it is necessary to utilise data across a longer time periods. There are a small selection of studies which analyse the trends and changes observed over multiple years, commonly in association with studies that investigate the impacts of cycling infrastructure implementation, BSS system expansion and general changes in user activity dynamics.

Félix et al. (2020) investigated the long-term modal shifts observed in Lisbon between 2016 and 2018 as a result of increased cycling infrastructure provision and the implementation of a new e-bicycle sharing systems (e-BSS). The analysis of manual ‘pen-to-paper’ counts of cycling activity around the city suggest a 3.5 fold increase in cyclists in the period after the implementation of additional cycling infrastructure, a significant proportion of which was observed to be due to an increase in BSS use. In addition to these increases associated with the construction of cycling
infrastructure, the implementation of the e-BSS was also found to exhibit a 2.5 fold increase in cycling activity. The combined investments in cycling activity within the city was pivotal in helping to mature the city’s cycling culture (Félix et al. 2020). This study demonstrates that longitudinal analysis on BSS can be conducted in the absence of data sourced directly from the BSS, however, this being said, observed counts should be interpreted with some caution as they are likely more susceptible to data errors and biases as a result of observation locations and times.

Zhang et al. (2016b) wanted to explore the changes in users and system usage over long periods due to the lack of research within the field. Utilising data from the Zhongshan BSS in China, they investigate the impacts of an expansion to the BSS, finding that around 45% of all users persisted to use the BSS throughout the study period, with a decrease in system use between 2012 and 2013 despite the system expansion. Though this was the case, the analysis showed no significant difference in trip characteristics between before and after system expansion (Zhang et al. 2016b). In a similar long-term longitudinal study, Jain et al. (2018) explore the trends observed within the Melbourne BSS between 2010 and 2016. Here they stress the importance of studying change over longer periods of time since this area lacks research and the insights are valuable in helping operators adequately respond to changing user needs and helping manage the future evolution of a BSS. Within this investigation, the authors find that whilst the overall use increased marginally, the proportion of casual trips within the system increased from 50% in 2010 to 80% in 2016. This increase was found to be fuelled by the introduction of courtesy helmets and a free tram zone in BSS areas. In the concluding remarks, Jain et al. (2018) emphasise the importance of their findings in helping to inform government and operators’ decision-making with limited funding for system expansion in addition to helping other cities, contexts and systems inform their future policy priority and action. The authors call for more research on longitudinal trends for BSS around the world.

Anaya-Boig et al. (2021) provide a unique holistic, national, long-term investigation into all dock-based BSS ever implemented in Spain. Within the analysis the
authors track the number of bikes and docks in operation in addition to assessing specific features of BSS, including the identification of common patterns in closed systems and the level of use in operating systems. The results show that 62% of BSS in Spain had closed by 2018, which were systems that typically had fewer docks, were located in more economically deprived regions and in cities with smaller populations. City population size and age were found to be significant predictors of BSS survival, along with finding that all BSS larger than 30 stations had a 100% survival rate (Anaya-Boig et al. 2021). This research is immensely valuable in providing indications of the characteristics of the BSS and their localities in predicting long-term BSS survival that can be used to inform successful future implementations of BSS within the nation.

Though the majority of long-term analysis has been conducted looking retrospectively at the historic operation of BSS, Hsieh et al. (2021) propose methodologies to estimate the long-term demand of newly established stations in areas of expansion. This is unique, even among predictive BSS demand models, since these are typically limited to short-term demand estimations in relation to the BRP. This study aims to provide operators and governments with a preliminary estimation of the amount of bike usages in the following half-year in new regions of a city, given just the locations of those new stations. Using data from the New York BSS, the authors show their long term demand advisor is relatively accurate and can be used as a valuable tool to avoid wasting resources when planning the expansion of a BSS.

Taken holistically, it becomes apparent that there is a lack of a truly global and long-term investigation into the growth and use of BSS. Although there are a handful of studies which aim to compare systems internationally and another pool of literature which seeks to investigate the long-term changes in BSS, these are typically isolated. The heuristics and methods presented in Chapter 4 and 5 will aim to build on the studies presented in this section and provide a truly global and comprehensive longitudinal analysis of BSS dynamics. This analysis will utilise the largest collection of BSS dock capacity data to facilitate the creation of valuable comparative heuristics across their operational lifespan, in addition to creating the
most comprehensive global comparison of BSS.

### 2.2.2 Medium-Term BSS Dynamics: Responses and Shifts to COVID-19 Restriction Policies

In March 2020 the WHO declared COVID-19 a pandemic. This highly infectious viral respiratory disease quickly spread worldwide, causing increased pressures on healthcare systems due to the potentially fatal nature of the virus. Early studies of the virus suggested that common symptoms included a fever, cough and fatigue, with 5% of cases resulting in a fatality (Li et al. 2020). With easy transmission between individuals in close proximity (Anderson et al. 2020; Asadi et al. 2020; Morawska and Cao 2020; Teixeira and Lopes 2020) and harmful repercussions, especially among those with weakened immune systems and the elderly (Mueller et al. 2020), governing bodies around the world have implemented various policies and restrictions. These strategies are typically categorised into two main groups: Pharmaceutical Interventions and non-pharmaceutical intervention (NPI). The former primarily rely on the development and administration of a vaccine in order to create long-term herd immunity to the virus. NPI, on the other hand, typically refer to the various social distancing related policies that can be implemented at short notice, such as mandatory face coverings, social distancing, gathering restrictions and retail and workplace closures (Askitas et al. 2021; Bian et al. 2021; Nouvellet et al. 2021). During periods of a rapid increase in the number of cases, lockdown restrictions were imposed on populations, forcing people to quarantine in their homes unless for limited and specific purposes (Ding et al. 2020; Hadjidemetriou et al. 2020; Jeffrey et al. 2020; Sharifi and Khavarian-Garmsir 2020), which has been dubbed the ‘Great Pause’ due to the scale of these restrictions (Wolfe 2020).

NPI have therefore caused major disruptions to all aspects of daily lives throughout their duration. Although there have been a number of papers that empirically investigate the impacts of the pandemic on individuals’ health (Branley-Bell and Talbot 2020; Flynn et al. 2020; Jallow et al. 2021; Rutter et al. 2021), the economy (De Lyon and Dhingra 2021; Keogh-Brown et al. 2020) and the environment (Chen et al. 2021; Liu et al. 2020; Zambrano-Monserrate et al. 2020), there
are fewer studies that investigate the impacts on mobility patterns. These changes in activity can be conceptualised as medium-term changes to the dynamics due to their temporal duration, with the majority changes occurring during the onset of cases in a nation with persisting impacts which have been ebbing and flowing with the number of cases. The pandemic provides a unique opportunity to study how implementation of policy can impact the mobility dynamics in the medium-term, which can not only help to inform future policy decisions in managing mobility for future pandemics, which are likely to increase in severity and frequency (Marani Marco et al. 2021), but also aid in future policy decisions to help transition to a sustainable future of urban mobility.

Conducting a thorough review of the literature surrounding the impacts of COVID-19 on BSS gives some insights into the changing mobility dynamics in the medium-term. As a result of the NPI that have been implemented, the way in which people move around urban areas has seen unprecedented levels of change, with lockdown restrictions resulting in a 25% average reduction in mobility among 128 countries and an associated 18% increase when these restrictions are removed (Palma et al. 2022). Among the emerging literature, there are an assortment of papers which utilise survey data in order to quantify the perceptions of travel modes and the resultant shifts in primary modes of transportation reported during the pandemic period.

Perceptions of PT have been continually evaluated in order to understand the mechanisms which motivate or discourage individuals’ modal choices. In the past, crowding has been shown to have a significant influence on various choice dimensions on modes of urban mobility (Li and Hensher 2011; Tirachini et al. 2013; Hensher et al. 2011; Wardman and Whelan 2011). The pandemic appears to have exacerbated the role of crowding in modal choice, with numerous survey studies which show a rise in the perception of risk in these shared modes. For example, Barbieri et al. (2021) conducted a survey in 10 countries and found that buses and planes are thought to be the riskiest modes of transportation, but there is a general sentiment that people are trying to avoid PT across all study countries. Similarly,
less than 10% of Dutch survey respondents have a positive attitude towards PT whilst attitudes towards private automobiles have improved (Haas et al. 2020). This corroborated with evidence from a latent class choice model study in the Netherlands which shows that PT users are willing to wait an additional 8.75 minutes in an effort to minimise contact with other passengers (Shelat et al. 2022). Bansal et al. (2022) also find similar characteristics among PT users in London, suggesting that mandatory face coverings have a positive effect on the likelihood of taking the underground, though this effect is less pronounced among men under the age of 40 with a monthly income below £10,000. Given the importance of PT in economic recovery and sustainable mobility, there is a necessity to improve travellers’ opinions to foster their uptake in the post-pandemic period (Shelat et al. 2022).

Changes in the outlook on PT have caused a large shift in residents’ modal choices during the pandemic period. In an international survey, PT modes were found to have the highest modal shift, with 73.6% of respondents claiming to have shifted away from PT for commuting purposes (Dingil and Esztergár-Kiss 2021). In a similar modal share study in Budapest, Bucsky (2020) finds bicycle usage to have the greatest growth rate by more than doubling its share. Comparable results are also observed in Lisbon, Portugal, where the percentage of users combining BSS with PT modes saw significant reductions, from 66.5% of trips to 58.9%, whilst BSS trips increased from 33.5% to 41.1% (Teixeira et al. 2022). Evidence such as these striking shifts in modal share has shown these new forms of urban micromobility are an attractive alternative mode of travel to overcrowded PT in the pandemic (Jobe and Griffin 2021) and appears as if they are increasingly establishing themselves as part of the new norm in the aftermath of COVID-19 (Bian et al. 2021).

Changes in perception and modal choice during the pandemic have provided initial insights into the resilience of BSS. There are many similarities in the results of these studies, which suggest that although there is a general decrease in the number of trips that are taken, BSS appear to be a much more resilient mode of transport for urban dwellers during the pandemic period (Nikiforiadis et al. 2020; Hu et al. 2021; Hua et al. 2020; Teixeira and Lopes 2020; Teixeira et al. 2022; Palma et al.
Despite the lack of conclusive evidence regarding the likelihood of COVID-19 infections on public transit (PT) modes compared to other modes (Hörcher et al. 2022), this change is likely due to the fact that cycling is a healthy form of transportation that complies with social distancing rules implemented in many countries (De Vos 2020; Budd and Ison 2020), as opposed to alternative modes of urban transportation, such as public trains and buses, which offer cramped and less ventilated conditions (Teixeira and Lopes 2020).

Although looking at the literature surrounding perceptions and modal shifts enable some insights into the changes that have occurred in cities around the world, they do not facilitate a detailed and holistic understanding of those changes in urban mobility patterns that have occurred during the pandemic period. Investigating data from specific modes of transportation, such as BSS, enable a deeper investigation into the more nuanced changes in activity patterns that have occurred. For example, many BSS have been shown to have a strong commuting user base, such as those with a two-peak weekday journey pattern (O’Brien et al. 2014). This is likely enhanced by the benefits of BSS as a mode of mobility that helps to solve issues surrounding the first/last mile problem (Yang et al. 2019; Fan and Zheng 2020). Due to the large shifts in the population which have been forced to work from home (WFH), around 87% across 10 countries (Barbieri et al. 2021), there has been decreased demand for BSS use to commute. This decrease in commuting utility has been met by an increase in the utility of BSS by leisure and casual users, which was observed in the Chicago BSS, with a significant decrease in the proportion of member trips while the number of non-member trips increased (Hu et al. 2021). This indicates a large shift in the use of BSS within this system, catering to the needs of the surrounding population during the pandemic period.

Similarly, Chai et al. (2020) investigate the impacts of the COVID-19 pandemic on the Beijing BSS. This is one of the few studies which investigates the spatiotemporal dynamics of a BSS, using journey data, weather data, points of interest and the number of cases. In the paper, the authors use a difference-in-difference model to assess the impacts of the pandemic on various aspects of daily life. They
find that BSS journeys in close proximity to subway stations, high-tech companies and shopping plazas fall by 77.9%, 74.2% and 73.6% respectively. They also find that infected residential areas and other residential areas saw a large fall in trips but these two categories were the only ones to have recovered to roughly pre-pandemic levels. Again, this gives a strong indication of shifts in BSS use in Beijing, with significant decreases in commuting trips to business areas as well as non-essential retail shopping trips. Rapid recovery within residential areas gives further evidence to suggest significant shifts away from commuting in the short term.

There also appears to be evidence from several BSS to indicate increases in the average trip length during the pandemic period including London (Heydari et al. 2021), New York City (Teixeira and Lopes 2020) and Boston and Chicago (Padminabhan et al. 2021). These increases in trip length could occur due to the shifts in the modality that were found in other studies (Bucsky 2020; Dingil and Esztergár-Kiss 2021). Since PT modes are perceived to be riskier (Barbieri et al. 2021), those residents without access to private vehicles may resort to these new forms of urban micromobility as a safe alternative mode to make those trips they would previously have taken using PT. Those PT trips would likely be longer distance journeys, therefore causing average trip lengths to increase during this period. Increases in trip length could also arise due to the restricted access to exercise facilities like gyms (Barkley et al. 2020; Simpson and Katsanis 2020). There are many academic papers that have found numerous benefits of regular physical exercise in helping to boost an individual’s immunity, moreover reducing the deleterious effects of stress on immunity (Duggal et al. 2019; Simpson et al. 2015). Having such physical and mental health benefits, lockdown restrictions, such as those in the UK, permitted individuals exercise outdoors every day. Being one of the few permitted activities during lockdown periods, urban dwellers may have used BSS as an alternative form of physical exercise which may have caused the increase in average trip length that has been observed.

Here, the emerging literature gives us a sense that there have been dramatic impacts in mobility dynamics as a result of the pandemic. It is very apparent that
there are some striking shifts in modal choice, both voluntarily, due to the perceived risks of infection, and involuntarily, due to the imposition of NPI restrictions. The lockdown restrictions imposed on populations provide a unique opportunity to study the medium-term changes in urban mobility dynamics. The analysis presented in Chapter 6 of this thesis builds on the literature concerning the impacts of COVID-19, specifically identifying the changes in mobility dynamics within the London BSS during periods of enforced lockdown. This provides a holistic yet granular understanding of changes in mobility dynamics that can be beneficial, not only for future pandemic policy planning, but also in helping to identify the resilience and flexibility that BSS offer as a sustainable transportation mode fit for the ongoing transitions.

2.2.3 Short-Term BSS Dynamics: When, Where and Why BSS are used

Due to the velocity and granularity of data that are being produced as a passive result of the activities that we perform, studies of mobility dynamics has been able to conduct ever more spatially and temporally granular analyses. This is true of BSS, which, since the third-generation, have been implemented with various technologies which enable their tracking in real-time. Whilst docked BSS are typically limited to the analysis of docking station occupancy and journey origins and destinations between fixed dock locations, dockless BSS provide a unique opportunity to study the mobility dynamics that have much greater spatial autonomy and flexibility. This can be achieved through the mining of data generated by the embedded GPS receivers in each bicycle, enabling the quantification of usage dynamics. Although the literature on the dockless BSS that analyses GPS data are currently limited in comparison to docked BSS, there is an emerging pool of literature (McKenzie 2018; Shen et al. 2018; Xu et al. 2019; Guidon et al. 2019; McKenzie 2020).

Spatio-temporal analysis

The increased flexibility of journeys offered by dockless BSS provide a unique ability to derive a detailed understanding of mobility patterns within the systems. Exploiting the GPS traces produced by the bicycles, researchers can analyse the
popularity of different routes across different times of the day. There are a number of studies that explore the dockless BSS in Singapore. Xu et al. (2019) explore the systematic dynamics of the system using an eigendecomposition method to create a deeper understanding of mobility across space and time by plotting the results on a 3D map of grid aggregations, similar to that of the space-time cube model study in Shenzhen used to identify hot and cold spots in time and space (Gao et al. 2022). Here, they were able to show the spatial and temporal variations in trip origins across weekdays, whilst these appeared to remain more static across weekends. Song et al. (2021) investigate the spatio-temporal dynamics of the Singapore dockless system through the use of spatial autocorrelation and community detection. This corroborates the results observed by Xu et al. (2019), with spatial autocorrelation being higher among weekdays, especially during peak community hours. Song et al. (2021) also split the operational area into 300m grid cells and journeys between grid cells across seven time-blocks across a day, with results showing strong associations between weekday peak hour community use in close proximity to metro stations compared to coastal activity in the South on weekends. Similar grid cell analysis has been conducted across a number of dockless BSS studies ranging from 300 metres to 500 meters (Guidon et al. 2019; Shen et al. 2018; Xu et al. 2019; Yang et al. 2019). Guidon et al. (2019) and Shen et al. (2018) simply map a choropleth over the grid enabling the identification that most journeys started around the central business district (CBD) region, but during off-peak hours, bikes were located in the outskirts, especially around the areas associated with recreational activities, illustrating a clear ‘ring-and-core’ structure (O’Brien et al. 2014, p. 263).

Though these studies are able to identify granular spatio-temporal patterns of BSS dynamics, aggregating journeys to grid cells have come under criticism due to issues relating to the modifiable areal unit problem (MAUP) (Gao et al. 2021a), an important issue relating to spatial statistics which is commonly associated with gerrymandering (Buzzelli 2020). Within the context of dockless BSS, the scaling and zoning effects of these grids cause concerns in relation to the assessment of
2.2. The Dynamics of Bicycle Sharing Systems

built environment factors. Using a geographical detector model, Gao et al. (2021a) are able to show the sensitivity of the spatial areal units, with inconsistencies in built environment factors at different scales. This research highlights the importance of being mindful of the spatial scale of any analysis that is conducted, but especially in relation to dockless BSS which have greater freedom in their use.

McKenzie (2018) employed kernel density estimation (KDE) analysis on journey origins to compare the spatial variations of competing dockless BSS in Washington D.C. This provides a way of analysing and comparing the activity dynamics of dockless BSS without the constraints of the MAUP. McKenzie (2020) has since employed similar methods to compare the spatial and temporal differences between a number of scooter-share services and a dockless e-BSS. This spatially indiscriminant method shows the strong, literally overlapping, similarities in operational extent across services, whilst also being able to identify areas of notable difference in where and when those services are used. In addition to being employed as a comparative tool between service dynamics, KDE has also been employed to determine the distribution of bicycles across communities of concern, showing how they provide greater availability of bicycles within these areas in San Francisco (Qian et al. 2020). These examples showcase the capabilities of point density methods to better exploit the granular nature of dockless BSS data in order to better understand their spatial and temporal dynamics.

Taken holistically, despite the analyses being conducted under varying contexts, the fourth-generation bikeshare systems appear to follow a general homogenous spatio-temporal trend, strongly reflecting a notable commuter pattern. Moreover, distinct usage purposes between weekdays and weekends were clearly depicted across the various methods that have been employed to study their mobility dynamics at a granular scale. It should be noted, however, that those analyses which only explored the trip start-points, may be overly dependent on the operator bicycle redistribution rather than the user intentions.
**Built environment factors**

In combination with a detailed spatio-temporal analysis of these systems, the data can be exploited to infer trip purpose through the analysis of journey dynamics in relation to built environment factors. These are typically carried out using various types of regression model in order to gain an understanding of the relationship between activity and points of interest within their vicinity. Studies that analyse the relationship between activity and built environment factors are abundant throughout the analysis of BSS. Guo et al. (2022) conduct a thorough review of those papers which analyse the impacts of built environment factors across 48 papers. The numerous types of variables that have been analysed have been broadly categorised into four main groups: land use, transportation system, urban design and urban form. Within the review it becomes apparent that there are many commonalities between built environment factors and general impacts on system use across systems globally. Nello-Deakin (2020) criticises the analysis of built environment factors, likening the studies that investigate the relationship between cycling infrastructure and cycling uptake to research on the relationship between tobacco and cancer, suggesting that we know enough about the basic facts to take decisive policy measures.

Although this may be true for some variables that have been investigated, like cycling infrastructure, the review shows the importance of regional differences in the use of BSS in relation to the primary purpose for journeys and cycling culture within the regions (Guo et al. 2022). In addition, the use of granular dockless data has proven to be useful in analysing the associations in built environment factors and to provide inferences about trip purposes. Within the papers explored in relation to granular spatio-temporal patterns above, they also seek to understand the relationships to those built environment factors in close proximity journey activities. Guidon et al. (2019) separated the journey activity into five time-blocks before running a negative binomial regression model. The results showed that in Zurich, e-BSS usage complemented the PT service as part of the multi-modal journey throughout the day or week and were more active near recreational facilities, particularly on weekends and off-peak hours. Similarly, Shen et al. (2018) and Xu...
et al. (2019) conducted the same analysis using trip count data for both arrivals and departures in Singapore. Here, the results indicated that the PT services attracted - rather than generated - BSS usage, particularly on weekend mornings. In addition, high residential density generated trips in the mornings and attracted journeys in the evenings whilst commercial land use did not attract bike usage in the mornings. The findings provided interesting evidence on how Singaporean dockless BSS were mostly used to connect people between their homes and PT services in the morning commute hours, or as the ‘first-mile’ solution (DeMaio 2009). Additionally, in comparison to Zurich’s e-BSS system (e.g., (Guidon et al. 2019)), Shen et al. (2018) and Xu et al. (2019) state that the prevalence of cycling infrastructure was less relevant to BSS usage in Singapore. This highlights the contextual differences in how various systems are adopted in different cities (Nikitas 2019; Sherriff et al. 2020), and thus prompts the value of such empirical studies in unravelling such particularities.

The literature presented here, in relation to short-term mobility dynamics of dockless BSS, provides a highly granular and nuanced understanding of the spatio-temporal patterns of operation. Although the analyses surrounding BSS activity dynamics have seemingly been exhausted (Nello-Deakin 2020), conducting such analysis at such high resolutions are valuable in helping to identify specific contextual nuances of those BSS. This provides policymakers with the details they require to make policy decisions regarding those BSS. For example, Song et al. (2021) highlights the importance of their findings in helping urban planners to identify the specific roads and locations that would benefit from the proposed 1,000km of cycling infrastructure by 2026 in Singapore. Using their hotspot identification in combination with community morphology can help to produce a hierarchical cycling path network that serves the current user’s demands. The analysis presented in Chapter 7 provides a unique and detailed perspective of dockless BSS use in London, for which there is currently no granular short-term dynamic analysis. The methods that are employed will build on the techniques identified in the literature and will provide unique insights into the specific dynamics associated within the
context of London.

2.3 Chapter Summary

This concludes the literature review of the thesis. Within this chapter we have been able to provide a holistic overview of urban mobility over the past 80 years and provide an understanding of the current conceptualisations of future urban mobility. Within this timeline, BSS have recently seen significant uptake due to the numerous merits that have been identified in association with their use, ranging from the individual level benefits on health to the macro level benefits for the global environment. BSS also sit in popular discourses on future urban environments and mobility such as the sharing economy, smart city, sustainable transport, 15-minute city and a part of future MaaS platforms. Although there are exciting new technologies surrounding the transformation of fossil-fuel powered vehicles towards EV and AV, their potential is socially imbalanced and hinges on the transformation of energy production to green modes. As a result, it becomes very apparent that BSS and subsequent iterations of novel urban micromobility options, such as e-scooters, are well-timed additions to the arsenal of sustainable mobility options, that are more socially equitable and sustainable.

Exploring the literature surrounding BSS can give a sense of the broad nature of studies which have been conducted, the volume of which has grown alongside the increasing popularity and implementation of the mode across the world. Though there are numerous aspects to BSS which could be considered, the thesis will focus on the dynamics of these BSS across various temporal and spatial scales. These are analysed using system and urban dynamic frameworks, building on the various analyses observed within the literature. At the long-term, global scale, Section 2.2.1 identifies the lack of a comprehensive international comparison of BSS. Exploiting the most exhaustive collection of BSS data, Chapter 4 will detail the metric creation heuristics as an automated process that provide novel and holistic ways to track the growth and evolution of BSS over time. Employing these metrics, Chapter 5 conducts the most comprehensive comparison of BSS around the world, using a
novel two-stage clustering methodology to determine the current global landscape of BSS. The COVID-19 pandemic has provided researchers with an invaluable opportunity to study the impacts of unprecedented policies and restrictions on changes in dynamics. Section 2.2.2 identifies the potential of the imposition of lockdown restrictions as a natural experiment to understand the impacts of restrictions and policies on mobility dynamics over the medium-term. The impacts of the various lockdown events within the UK are explored in Chapter 6, not only to shed light on the way in which activities have changed during these periods, but also to provide guidance for future policies on their role in changing mobility dynamics, whether that be for future pandemics or to help promote the uptake of these sustainable modes in a transition towards future urban mobility realisations. Finally, Section 2.2.3 identifies the utility of granular spatio-temporal data, produced as a passive result fourth-generation BSS operation, in providing a rich understanding of activity patterns within BSS. Building on the limitations of the methods identified, Chapter 7 explores the dockless dynamics in London, an area which has not been explored previously.

Together, the analysis presented in this thesis is unique in its ability to combine the large variations in temporal and spatial scales and a pool of methods to help provide a holistic understanding of BSS, a novel, but increasingly vital mode of urban mobility fit for the future.
Chapter 3

Bicycle Sharing System Data

As detailed in the Chapter 2, technological advancements, in conjunction with urban population growth, have created big data of increasing volume, velocity, variety, variability and value (Ishwarappa and Anuradha 2015; Al Nuaimi et al. 2015; Kitchin 2016). This includes the urban mobility sector that has typically relied on ‘active modes’ of data collection, which are purposeful efforts to collect data which require the engagement of participants such as travel diary’s or traditional surveys and questionnaires (Wenz et al. 2019). These methods also include purpose built sensors such as police automatic number plate recognition (ANPR) cameras, which can help to determine mobility volumes, speeds and routing (Wan et al. 2021; Kazagli and Koutsopoulos 2013; Li 2008). Whatever the scenario, the nature of active modes mean that participants are typically highly aware of those data being collected and therefore have some degree of control over what data they share (Wenz et al. 2019). Furthermore, due to their specially engineered designs, active modes of data collection are generally very accurate in collecting data for their intended purposes.

Although these methods have been the primary source of data historically, an increasing proportion of data are collected through opportunistic, passive means that do not require purpose built sensors, but rely on the trails of data that are left behind through the digital activities of individuals. This process has been otherwise referred to as digital footprints (Ztwitter 2014; Golder and Macy 2014), exhausts (Glaeser et al. 2018), or even shadows (Kaisler et al. 2013) due to the level of detail
that can be obtained on each individual. This relies on repurposing and reformatting of alternative data sources, which suffers from its own plethora of data issues (Lazer 2015) but also benefits from the richness of data that can be collected at much higher frequencies than self-reports without burdening the participants (Wenz et al. 2019). Another advantage of passive data is their potential to lead to more accurate estimates than self-reports as a result of issues such as incorrect data input, omission and the Hawthorne effect (Boase and Ling 2013; Evenson et al. 2015; Scherpenzeel 2017).

Although there are numerous benefits to the increasing availability of passive data, there are privacy concerns which require greater attention in order to mitigate any potentially harmful implications to individuals. Due to the passive nature of collection, participants are typically less aware of the trails that they are leaving behind and how those data could be used to identify and track them. Zwitter (2014) calls for greater awareness and education on the implications of such data being collected since individuals may not have specific knowledge or appreciation that data are being collected about them (Nunan and Di Domenico 2013). These data are also being increasingly used as part of the many software-enabled technologies in smart cities, complementing and replacing traditional data-informed urbanism towards data-driven urbanisms (Kitchin 2016). This shows the increasing integration of those data in the functioning of urban environments, from live traffic data being used to determine the fastest routes on travel planning applications (Alvarez et al. 2018; Wahle et al. 2001) to top-down mobility management policies that have been informed using these data-driven insights. This creates a paradox in which individuals can opt out of having their data collected but are likely to result in their exclusion from the increasingly digitally connected world (Nunan and Di Domenico 2013).

As a result, data regulations and standards have been created and updated to better protect the individual and put greater responsibility for the safe and anonymous storage, dissemination and use of data on those data owners (Richards and King 2014; Zwitter 2014; Rubinstein 2013). There is a great amount of discourse surrounding the ethics and privacy of data in this digital age, especially in the wake
of the Cambridge Analytica Scandal (Schneble et al. 2018) that highlighted the real-world impacts of such abuse of passively collected data. The implementation of the General Data Protection Regulation (GDPR) in the European Union (EU) in 2016 and the Clarifying Lawful Overseas Use of Data (CLOUD) Act in the USA in 2018 were direct efforts to control these data. Although not perfect, these regulations have brought greater awareness to privacy issues and have created structures and processes for the safe use of these data (Zarsky 2017). Data standards are homogenised formats that comply with data regulations and are used across a wide variety of industries to enable greater sharing of those data for mutually beneficial purposes. They create a common language so that analysts can spend less time unpacking and understanding the different variables that are contained in the data and instead spend this time making good use of them. Therefore, regulations and standards are an important development that have helped to increase the accessibility of data whilst also minimising the potential risks, enabling researchers and practitioners to employ such data to drive new insights and innovation (Floridi and Taddeo 2016; Richards and King 2014).

This chapter provides a detailed overview of the data that are available on BSS. First, Section 3.1 explores the various data standards that are currently used to disseminate real-time BSS data. Section 3.2 will then detail the way in which these data formats have been accessed and stored within the UCL BSS data collection, which, to the best of our knowledge, is the largest collection of BSS data. Section 3.3 will then discuss the alternative data sources of BSS data that are published by operators and contain granular information on journey origins and destinations. The chapter will be concluded by a summary of these data, discussing the merits and limitations that they pose in uncovering the dynamics of BSS.

### 3.1 The General Bikeshare Feed Specification

The proliferation of new micromobility modes such as BSS and e-scooters have created new opportunities for increased accessibility to sustainable forms of urban transportation (Abduljabbar et al. 2021). Their growth and adoption are a clear
reflection of the gap that they are filling in providing a form of short-distance mobility and a mode to help resolve the first/last mile issue. Although there are clear benefits to these systems, Section 2.1.2.2 highlights the various implementation and management issues that these BSS face. As a result, it is clear that they need to be monitored and analysed in order to better understand the way in which they are used to make more informed decisions on their operation, a sentiment that is echoed throughout academic literature (Xu et al. 2022; Abduljabbar et al. 2021; Midgley 2011; Mateo-Babiano et al. 2016).

Data standards are a set of rules through which data are described and recorded, creating a format and meaning that enable the data to be understood when shared (Zack 2019). These standards vary based on the particular use case and the intended audience, but broadly help people and organisations to publish, access, share and use better quality data (ODI 2022). There are many examples of established data standards such as the Brownfield Site Register Open Data Standard that are used among local planning authorities to share the location and condition of sites suitable for residential development in addition to a plethora of open standards for environmental data used by the US Environmental Protection Agency to reduce the cost of collecting and sharing data (ODI 2022).

Within the transportation sector, the General Transit Feed Specification (GTFS) has become the de facto standard for PT data. It was initially developed in 2007 to integrate various transit feeds into Google Maps (Zack 2019) but was quickly adopted by many transit operators and has enabled the integration of such data to a wide plethora of third party route planning applications with greater ease. They provide a standardised means of sharing data in relation to the location of PT stops, their schedules and more recently have been updated to include real-time locations of PT. The widespread adoption of the GTFS standard has proved hugely beneficial to all parties including the operators, users and local authorities (Wong 2013). These standards also help to facilitate travellers with multi-modal transportation information and provide the foundations to be able to create a MaaS platform with trip planning, booking and payment capabilities (NABSA 2021).
3.1. The General Bikeshare Feed Specification

Data standards conform to the growing rhetoric surrounding the merits of open data, with numerous national and international policies to encourage the sharing of data. Within the UK, the Open Government Data policy and the Data Standards Authority are those policies and institution which aim to encourage safe and open data dissemination.

Rapid growth and adoption of BSS, in addition to other forms of urban micromobility, have necessitated the creation of a data standard for these new modes of urban mobility. The General Bikeshare Feed Specification (GBFS) was first introduced by the North American Bikeshare and Scootershare Association (NABSA) in 2015 to provide a standardised way to ingest, analyse and compare data from micromobility operators (NABSA 2021). Much like GTFS, it was quickly adopted and has since become the primary standard for real-time micromobility data, providing a way to share data on the location and status of micromobility vehicles (including bicycles, scooters, mopeds and cars), in addition to accompanying operational information (such as operational area and parking locations). This ensures that the exchange of information between parties are commonly understood to ease use and application. The GBFS standard has an active Github repository (accessible at: https://github.com/NABSA/gbfs) that is used as a base to detail the specifications of the standard as well as communicate any changes or updates. Users of the standard can also flag issues or suggest changes through the platform’s ‘issue’ and ‘pull request’ functionality respectively. The repository also contains general information, such as the guiding principles where we are able to get a good understanding of the explicit purpose of the standard, that is to provide real- or semi-real-time, read-only data on the availability and location of micromobility vehicles for shared mobility users.

Much like GTFS, the GBFS standard has been championed by Google through the implementation of their Google Micromobility Partner Program. In this program operators are able to apply to get their systems shown as a potential transportation mode on Google Maps. Although any operators can apply, those that have clean GBFS feeds and systems that have large coverage and fleet sizes are generally pri-
oritised. As a platform with more than a billion monthly users (Google 2022), this greatly encourages the adoption of GBFS as it could prove to be a very valuable source of advertisement.

As an open source specification it is constantly evolving to cater for additional features and capabilities. In March 2020, GBFS Version 2.0 was released, which facilitated the support for dockless and hybrid (both docked and dockless) systems, regardless of the types of vehicles, in addition to enabling ’deep links’, which provide seamless connections between third party applications and the provider’s native application for booking and payment purposes (MobilityData 2020). These updates and changes to the data standard also included efforts to safeguard the privacy of the user. The update to GBFS Version 2.0 included the requirement for the rotation of the identifier (ID) variables associated with each vehicle after every trip (NABSA 2021). This reduces the possibility for the feed to be used as a means to infer trip origins and destinations, ensuring that no individual journey can be disclosed. There are currently two ways in which operators ensure this anonymity, namely through random and dynamic ID rotations. Random switching of ID occurs after each unlocking of the vehicle, whilst dynamic rotation refers to the continuous rotation ID values at regular intervals (Xu et al. 2022). These methods have proven to help mitigate the potential identifications of OD trips, but there have been some efforts to use the updated GBFS specifications to help delineate journeys in papers such as Xu et al. (2022), which we will explore in detail in Section 3.2.3.

Within the GBFS standard there are several files that are required, listed in Table 3.1. Each of these files contain valuable information on the current status of the micromobility service, including the real-time location of each vehicle which is contained within station_information.json for those docked systems and free_bike_status.json for those dockless systems. A detailed account of the contents of each of these files are detailed on the NABSA Github repository. These files are typically accessible through public API feeds provided by each operator, the majority of which are also listed on the NABSA Github repository within the systems.csv file. The data feeds are updated every time to live (TTL), measured in seconds. The
### 3.1 The General Bikeshare Feed Specification

Table 3.1: List of all required files for GBFS feed

<table>
<thead>
<tr>
<th>File Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gbfs.json</td>
<td>Auto-discovery file that links to all of the other files published by the system</td>
</tr>
<tr>
<td>system_information.json</td>
<td>Details including system operator, system location, year implemented, URL, contact info, time zone.</td>
</tr>
<tr>
<td>vehicle_types.json</td>
<td>Describes the types of vehicles that system operator has available for rent. REQUIRED of systems that include information about vehicle types in the free_bike_status.json file. If this file is not included, then all vehicles in the feed are assumed to be non-motorized bicycles.</td>
</tr>
<tr>
<td>station_information.json</td>
<td>List of all stations, their capacities and locations. REQUIRED of systems utilizing docks.</td>
</tr>
<tr>
<td>station_status.json</td>
<td>Number of available vehicles and docks at each station and station availability. REQUIRED of systems utilizing docks.</td>
</tr>
<tr>
<td>free_bike_status.json</td>
<td>(as of v2.1) Describes all vehicles that are not currently in active rental. REQUIRED for free floating (dockless) vehicles. OPTIONAL for station based (docked) vehicles. Vehicles that are part of an active rental MUST NOT appear in this feed.</td>
</tr>
</tbody>
</table>

Source: NABSA GBFS Github

The standard does not require a specific value meaning that the TTL varies among operators, though it is likely in their best interests to be as close to real-time in order to facilitate the most accurate data dissemination for vehicle locations for end users.

#### 3.1.1 Other BSS Data standards

Although GBFS is the most commonly adopted micromobility data standard, with 650 system feeds listed on their Github repository (as of June 2022), it is important to note that GBFS is not the only standard adopted by micromobility operators. Other standards are typically synonymous with incumbent operators such as the Public Bike System Company (PBSC) who operate around 45 systems including the London, Barcelona and New York BSS, and JCDecaux who operate around 30 systems including the Paris and Brisbane BSS. In both of the named cases, the operators predate the GBFS standard and have therefore developed their own practices for data sharing and storage which would require significant uproot-
ing and disruption of established internal practices. There are some instances of BSS operated by these companies which are located in cities which mandate the dissemination of open GBFS data, such as the New York Citi Bike BSS under the Open Data Law of 2012 that was passed to make government data open and freely available to the public (Okamoto 2016). Under such circumstances, these operators continue to use their internal data standards for the operation of their systems and supplement/manipulate those data into the GBFS standard, making these publicly available to meet legislative requirements.

There are also some white label companies that offer complete software-as-a-service technology backend solutions to micromobility operators. Examples include Wegoshare, Joyride, Fleetbird and GoUrban. They provide an easy solution for transport companies to digitise their offerings, enabling their monitoring and management. Their presence and the growing popularity of this business model is a testament to the number of new systems looking to utilise such services. Although there are some ways in which these standards differ, the feeds are comparable to GBFS due to the similarity in the standards objectives in helping to locate micromobility availability and share this information on user applications. Since GBFS feeds are generally inconducive for operator monitoring and management of systems, due to the open nature of the data standard which inhibit the dissemination of granular personally disclosive data, white label companies in this space have formulated their own data standards for internal use, similar to that of those incumbent BSS operators. Operators typically have to obfuscate and manipulate their raw, granular data records into the GBFS standard for public consumption. Due to the disclosive nature of the data and the increasingly competitive market of micromobility, operators are typically averse to making such data openly available. As a result, such feeds are frequently limited to those cities, states and countries that mandate the routine release of such data, or to those cities which strongly encourage such activities through investment ties to those modes.

In addition to the other data standards emulating GBFS, it is worth noting the Mobility Data Specification (MDS) which was introduced in 2018. This is a stan-
standard builds on GBFS that was developed by the Open Mobility Foundation (OMF) and, although they have some key similarities, their purposes differ. The GBFS standard was developed to be publicly available in the service of aiding traveller trip planning, while the primary function of the MDS is for non-public use in regulation by cities and other agencies. MDS also includes historical trip and vehicle status, which GBFS does not (NABSA 2021). This makes MDS a much more disclosive data standard and means that access to such data is typically limited to internal use. As a result, the vast majority of academic studies which seek to investigate these micromobility modes are primarily reliant on GBFS and other operator feeds to collect data. This being said, it is also important to note that there are a small handful of systems which regularly provide clean OD journey data, which are very valuable when analysing granular journey data (detailed in Section 3.3).

3.2 UCL’s Bicycle Sharing System Data

Due to the scarcity of routinely released open BSS data, this thesis relies upon the UCL BSS data collection. To the best of our knowledge, this is the largest database of BSS data that exists. It was initially created in August 2010 and has since been continuously maintained and updated to include as many live and publicly available BSS API feeds (GBFS or otherwise) into several standardised tables (Table 3.2). Over the duration of data collection, the database has accumulated data from over 500 BSS across all major continents. The database primarily consists of systems dock-based BSS, although the growing proliferation of dockless and hybrid systems have increased the collection of data from such systems in recent years. Figure 3.1 details the count of active BSS for which data have been collected. The grey section at the end of the graph that indicates a decrease of approximately 80 BSS over 2 months are an artefact of those systems for which collection scripts have broken and require maintenance as opposed to any major closure events and are typical of any snapshot of the database.

The database, which started as part of a visualisation project by Oliver O’Brien at the Centre of Advanced Spatial Analysis (CASA), UCL, uses several Python
collection scripts which call the various BSS API feeds at regular time intervals, manipulating the GBFS (or similar) files into standardised tables (Table 3.2). Approximately 60% of systems within the collection utilise the GBFS, making the collection, processing and storage of data easily scalable, whilst the other 40% of systems rely on adaptations to collection scripts in order for those data to be correctly standardised. For around 70% of systems within the collection bicycle locations and counts are updated every 10 minutes, in addition to a further 20% that are scraped every two minutes. Data on the remaining 10% of systems are collected every 20 minutes or longer, typically due to API call request limitations imposed on those feeds. The difference in collection intervals are primarily a result of storage limitations in the database that arose after surpassing 250 systems in 2015. As a result, BSS for which data were collected prior to 2016 are typically polled at two minute intervals, whilst subsequent BSS for which data have been collected and old BSS that required maintenance due to breaks in the feed were collected at 10 minute intervals. This was employed after Médard de Chardon and Caruso (2015, p. 274) found that 10 minute collection intervals ‘can provide sufficiently good estimates’ of journey occurrences. As such, this enabled the minimisation of redundant data
from smaller and less frequently used BSS whilst collection for large, highly active systems were prioritised to continue collection at two minute intervals in order to maximise the granularity of activity that can be observed within these data.

When considering collections of BSS data, it is very important to note the Meddin Bike-sharing World Map (Meddin et al. 2022). This is the world’s most complete overview of BSS that was initiated by Paul DeMaio in 2007 and was curated by Russell Meddin until his passing in April 2020. It is now maintained by a team of academic and volunteer staff who manually scour the internet for any and all information of BSS. This data feeds into the website (bikesharingworldmap.com) to provide the most complete overview of BSS locations, in addition to details on their current status (whether they are in the planning stages, in operation or have closed, suspended or hibernating during their off seasons), size and use wherever possible. Although there are likely to be some missing systems, this is by far the most frequently updated and detailed accounted of BSS globally. As a result, this has commonly been used to determine and track the global BSS landscape and cited within academia to refer to the total number BSS, with over 260 direct citations on Google Scholar. As of June 2022, there were reported to be over 1,850 open BSS around the world with an additional 210 planed to open in the future (Meddin et al. 2022).

Table 3.2: The stored table names in the UCL BSS database

<table>
<thead>
<tr>
<th>Stored Table Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ind_‘systemname’</td>
<td>Individual dock capacity data, containing the number of bicycles and spaces</td>
</tr>
<tr>
<td></td>
<td>at each docking station during each API poll (Docked BSS)</td>
</tr>
<tr>
<td>sum_‘systemname’</td>
<td>Total dock capacity data, containing the number of bicycles and spaces</td>
</tr>
<tr>
<td></td>
<td>in all docking stations during each API poll (Docked BSS)</td>
</tr>
<tr>
<td>bikelocations</td>
<td>Docking station location data, containing the unique IDs for each station</td>
</tr>
<tr>
<td></td>
<td>and associated coordinates (Docked BSS)</td>
</tr>
<tr>
<td>bik_‘systemname’</td>
<td>Bicycle location data for all dockless bicycles during each API poll</td>
</tr>
<tr>
<td></td>
<td>(Dockless BSS)</td>
</tr>
</tbody>
</table>
The key difference between the Meddin Bike-sharing World Map and the UCL BSS data collection are in their aims and the data that are recorded. Whilst the Meddin World Map aims to provide a complete list of all BSS to have existed, the UCL BSS data collection was designed to archive the distribution of bicycles across BSS that have publicly accessible API feeds. Through the careful cleaning and manipulation of such data it is possible to determine granular insights into the size and activity dynamics of systems. Continual maintenance of such collection processes have fruited data on nearly one third of all active BSS in the world. This data feeds into a publicly accessible web page (bikesharemap.com) that provides an overview of those systems for which data are currently being collected, as well as a demonstration of the real-time distribution of bicycles. The complete collection of UCL BSS data total well over a terabyte of data, and is constantly increasing by over one gigabyte per day.

Figure 3.2 provides an overview of the distribution of all BSS by world regions (sourced from the Meddin World Map) in comparison to the UCL BSS database. Here, we are able to determine that the UCL BSS database is generally representative of the global distribution. The primary disparity that becomes apparent is in the database’s over-representation of BSS in Europe in conjunction with a severe under-representation of Asian BSS. This disparity is primarily a result of data availability, with many systems failing to provide open API feeds. This may also be exaggerated by the fact that China is the largest BSS market globally (Campbell et al. 2016; Zhao et al. 2018) and Chinese regulations of ‘data sovereignty’ have made access to data produced in the country very limited. China also blocks access to sites such as Google, the primary driving force behind the adoption of GBFS, meaning that BSS in China are not incentivised to release such data to make them accessible to third party applications like Google Maps. In the same vein, European and American regulations on data are much more transparent and established, making access to these API feeds much easier. Although there is this disparity in global representatives, having such awareness dissolves any ambiguity and in most cases, the analysis of BSS are typically to an individual or a small collection of BSS as
opposed to any large scale global analyses.

Figure 3.2: Percentage share of BSS in the UCL BSS database in comparison to the Meddin Bike-Sharing World Map

As mentioned, the database consists of standardised tables that are continually updated to include the manipulated data from each collection instance. The various tables that are constructed from GBFS (or similar) BSS data feeds utilising the collection scripts are detailed in Table 3.2. The ind tables (named as such since they contain individual docking station data), also referred to as ‘dock capacity data’, provide the most granular form of docked BSS data, detailing the number of bicycles and number of spaces at each docking station (for those third-generation BSS). The process of converting the GBFS feed to this format is detailed in Section 3.2.1. The sum tables (named as such since they are summary tables, summing values across all docking stations) contain a simple aggregated version of the dock capacity data, detailing the number of bicycles and number of spaces across the entirety of the BSS at each collection instance. The bikelocations table complements the dock capacity data and are dedicated to containing the location of docking stations within the table. Therefore, it is referred to as the ‘dock location data’ and the data collection process for which is detailed in Section 3.2.2. The final major set of tables that make up the UCL BSS database are the bik tables (named as such since they contain individual bike data) which are also referred to as the ‘dockless BSS data’. These tables detail the coordinates at which each dockless bicycle is located during each API poll, the collection process for which is detailed in Section 3.2.3.
3.2.1 Dock Capacity Data

Since the third-generation BSS were the first systems of scale to incorporate tracking technologies, they were also the first to make API feeds for their systems. These were initially created to provide data to the BSS applications that enable users to plan their trips and to check whether bicycles were available at particular docking stations. Those systems, such as the London BSS, that commenced their operations prior to the release of the GBFS standard were pioneers in providing public API feeds, using their own data structures to facilitate these applications. Although this is the case for some systems, this section will focus on the pipeline process that was constructed for the majority of these systems, which rely on the GBFS standard. For those systems that do not conform to the GBFS standard, similar data feeds are available which enabled their data collection, but for the sake of succinctness, the general processes are detailed.

Dock capacity data, within the context of the UCL BSS data collection, refers to the collection of information on the number of bicycles and spaces observed within each docking station during an API poll. This data is sourced from the `station_status.json` file within the GBFS feed. It contains information regarding the occupancy of a docking station in addition to the types of vehicles, their ID, the total number of docks available, whether the dock is accepting rentals and the last reported time of this status (all of the variables are listed on the NABSA GBFS Github repository). For the sake of minimising the number of stored variables, the dock capacity data records which were stored in the database were limited to five variables. Those variables are listed in Table 3.3, which details the stored variable names in association with the raw variable names from the GBFS `station_status.json` file. For each docked BSS, an `ind` table is created, the records within which are continually updated with each API poll.
### Table 3.3: The stored variable names for dock capacity data and their associated source from the GBFS feed

<table>
<thead>
<tr>
<th>Stored Variable Name</th>
<th>Stored Variable Format</th>
<th>GBFS Variable Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tfl_id</td>
<td>smallint</td>
<td>station_id</td>
<td>Unique docking station identification number</td>
</tr>
<tr>
<td>bikes</td>
<td>smallint</td>
<td>num_bikes_available</td>
<td>The number of bikes observed at that docking station</td>
</tr>
<tr>
<td>spaces</td>
<td>smallint</td>
<td>num_docks_available</td>
<td>The number of spaces observed at that docking station</td>
</tr>
<tr>
<td>total_docks</td>
<td>smallint</td>
<td>capacity</td>
<td>Total capacity of docking station</td>
</tr>
<tr>
<td>timestamp</td>
<td>timestamp</td>
<td>API poll timestamp</td>
<td>Timestamp of when the data were collected</td>
</tr>
</tbody>
</table>
The collection of such data provides a detailed and granular account of the changing state of each docking station. These data can also be manipulated in various ways in order to create valuable metrics that can be used to analyse the system, such as estimating the number of journeys, the total number of available bicycles and the imbalance of bicycle distribution within the system. These heuristics have been developed (in Chapter 4) and employed (in Chapter 5) to gain a holistic understanding of the global BSS landscape.

Although the collection of dock capacity data at high frequencies is preferred, due to its value in minimising unobserved bicycle movements, it is a necessary trade-off to minimise the frequency of redundant API polls and data being stored in the database. For this reason, only those largest BSS which have a history of frequent use are collected every 2-minutes. For the majority of systems, 10-minute collection intervals are sufficient to capture most BSS activity, although it is important to be aware that there will be some cases which exhibit no changes between collection instances that experience the same number of journeys starting and finishing between API polls, thus missing such activity within the collected data.

In recent years, an increasing number of docked BSS have began to integrate e-bicycles in addition to systems that integrate the use of dockless bicycles as part of the hybridisation of BSS (Bieliński and Ważna 2018). This has necessitated some updates and adjustments to dock capacity ind tables to facilitate such systems within the data collection. For these systems, an additional column which details the number of e-bicycles within a docking station has been appended to enable the differentiation between those manual pedal bicycles and electric-assisted bicycles. In addition, a ‘floating’ docking station ID is included as a dummy variable used to indicate all those bicycles which are not located within a fixed physical docking station. This enables the simple quantification of dockless bicycles which are available to be rented within those systems. Although the location of each of these dockless bicycles in hybrid systems are not detailed in the dock capacity tables, the locations are collected and stored in a similar to fashion to the dockless BSS data that are detailed in Section 3.2.3 within a supplementary table.
3.2.2 Dock Location Data

To complement the dock capacity data, the location of docking stations are also recorded from the GBFS (or similar) API feeds and stored within a separate table that are also updated with each API poll. This table is to link the dock capacity information to geographical locations in addition to tracking the expansion or contraction of physical BSS infrastructure, whether that be in the form of docking stations opening and closing or the construction of additional parking places within existing docking stations.

Unlike the dock capacity data, dock location data are stored collectively in a single table named `bikelocations`. Details on the various variables that are stored within the table are detailed in Table 3.4. The variables within this table are primarily sourced from the `station_information.json` file, a requirement for GBFS feeds of dock-based BSS. In addition to storing all the data in a single table, the cells within the table are continuously updated as opposed to writing new lines for each API call. This is to minimise the amount of redundant data that are being stored in the database and since the expansion or contraction of BSS occur very infrequently such methods to collect docking station location data would be wasteful.

Therefore, in order to maintain a complete collection of docking station location history the `created_dt` and `updated_dt` variables are used to determine when a docking station was first observed and when the information on the docking station was last updated respectively. This is conducted by first checking for the presence of the same docking station ID in the table. If the same ID exists, it then checks whether the coordinates (location) of the docking station has changed. In the vast majority of cases, the location will not have changed, in which case, the remaining variables are updated, along with the associated timestamp in the `updated_dt` column. For those docking station IDs that do not exist, or have moved, a new entry is created.

It is worth noting that there have been some instances where BSS operators have used the same station ID for docking stations located in vastly different locations. These typically occur when an operator closes one station, making the
station identifier unused, and reuses the same identifier for a new docking station constructed at a future date. There are also instances where the coordinates of a docking station wander or are input incorrectly, thus some data preparation efforts are necessary to ensure clean and accurate dock location data.
### Table 3.4: The stored variable names for dock location data and their associated source from the GBFS feed

<table>
<thead>
<tr>
<th>Stored Variable Name</th>
<th>Variable Type</th>
<th>GBFS Data Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>system</td>
<td>character</td>
<td>named</td>
<td>Named identifier unique to each BSS</td>
</tr>
<tr>
<td>ucl_id</td>
<td>integer</td>
<td>station_id (derived)</td>
<td>Identifier for docking station</td>
</tr>
<tr>
<td>operator_intid</td>
<td>integer</td>
<td>Non-GBFS feeds</td>
<td>Non-GBFS feeds</td>
</tr>
<tr>
<td>operator_altid</td>
<td>character</td>
<td>Non-GBFS feeds</td>
<td>Non-GBFS feeds</td>
</tr>
<tr>
<td>operator_name</td>
<td>character</td>
<td>name</td>
<td>The public name of the station for display purposes</td>
</tr>
<tr>
<td>notes</td>
<td>character</td>
<td>N/A</td>
<td>Manual field noting where and when data was acquired for systems (pre-2015)</td>
</tr>
<tr>
<td>lat</td>
<td>numeric</td>
<td>lat</td>
<td>The latitude coordinate of docking station location</td>
</tr>
<tr>
<td>lon</td>
<td>numeric</td>
<td>lon</td>
<td>The longitude coordinate of docking station location</td>
</tr>
<tr>
<td>initial_bikes</td>
<td>integer</td>
<td>num_bikes_available</td>
<td>Number of bicycles in docking station upon first API poll</td>
</tr>
<tr>
<td>initial_size</td>
<td>integer</td>
<td>capacity</td>
<td>Number of physical docking slots at docking station upon first API poll</td>
</tr>
<tr>
<td>curr_bikes</td>
<td>integer</td>
<td>num_bikes_available</td>
<td>Number of bicycles in docking station at latest API poll</td>
</tr>
<tr>
<td>curr_size</td>
<td>integer</td>
<td>capacity</td>
<td>Number of physical docking slots at docking station upon latest API poll</td>
</tr>
<tr>
<td>createdDt</td>
<td>timestamp</td>
<td>API poll timestamp</td>
<td>The timestamp of the first API poll</td>
</tr>
<tr>
<td>updatedDt</td>
<td>timestamp</td>
<td>API poll timestamp</td>
<td>The timestamp of the latest API poll</td>
</tr>
<tr>
<td>updatedDt_nonzerobikes</td>
<td>timestamp</td>
<td>API poll timestamp</td>
<td>The timestamp of the latest API poll (where curr_bikes does not equal 0)</td>
</tr>
<tr>
<td>pedelec</td>
<td>integer</td>
<td>Non-GBFS feeds</td>
<td>An identifier for whether the docking station accepts electric bicycles</td>
</tr>
<tr>
<td>operator_installflag</td>
<td>integer</td>
<td>is_installed</td>
<td>An identifier for whether the docking station is physically there</td>
</tr>
<tr>
<td>operator_lockflag</td>
<td>integer</td>
<td>is_returning / is_renting</td>
<td>An identifier for whether the docking station is operational</td>
</tr>
<tr>
<td>operator_tempflag</td>
<td>integer</td>
<td>Non-GBFS feeds</td>
<td>An identifier for whether the docking station is temporary</td>
</tr>
</tbody>
</table>
3.2.3 Dockless BSS Data

In addition to the docked BSS data, the growing popularity and presence of dockless BSS, in conjunction with the updated GBFS in March 2020 to officially cater for such systems, has facilitated the mass collection of bicycle location data for such systems. As identified in Section 2.1.2, dockless systems refer to the fourth-generation and most recent iteration of BSS and are flexible by virtue of the lack of physical infrastructure to park each bicycle. As a result of the disparities in the way docked and dockless systems operate, the nature of data collection also differs.

Unlike docked systems, for which the dock capacity and dock location data are stored in separate tables, dockless bicycle locations are stored in a single table under the *bik* format. The data within this table are primarily sourced from the *free_bike_status.json* file, but it is important to note that this was renamed to *vehicle_status.json* as of GBFS version 2.1 in March 2021. The update introduced additional features, such as the type of vehicle (car, scooter, bicycle etc.) and how they are powered (human, electric, combustion etc.), that are yet to be included within these data collection processes. Table 3.5 provides an overview of the way in which the GBFS file have been stored within the database.
Table 3.5: The stored variable names for dockless bicycle location data and their associated source from the GBFS feed

<table>
<thead>
<tr>
<th>Stored Variable Name</th>
<th>Variable Type</th>
<th>GBFS Data Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>character</td>
<td>bike_id</td>
<td>Unique identifier for each bicycle</td>
</tr>
<tr>
<td>name</td>
<td>character</td>
<td>Non-GBFS feeds</td>
<td>Non-GBFS feeds</td>
</tr>
<tr>
<td>veh_type</td>
<td>character</td>
<td>vehicle_type_id</td>
<td>An identifier to differentiate between bicycle types</td>
</tr>
<tr>
<td>lat</td>
<td>double precision</td>
<td>lat</td>
<td>The Latitude coordinate of the bicycle</td>
</tr>
<tr>
<td>lon</td>
<td>double precision</td>
<td>lon</td>
<td>The Longitude coordinate of the bicycle</td>
</tr>
<tr>
<td>hub_id</td>
<td>integer</td>
<td>station_id</td>
<td>The identifier for the docking station (or geo-fenced parking zone) it is currently parked for hybrid systems only, 0 if not in station or dockless system</td>
</tr>
<tr>
<td>address</td>
<td>character</td>
<td>Non-GBFS feeds</td>
<td>Non-GBFS feeds</td>
</tr>
<tr>
<td>battery_level_pc</td>
<td>integer</td>
<td>current_fuel_percent</td>
<td>The current percentage of battery that the bicycle has (if e-BSS)</td>
</tr>
<tr>
<td>in_geofence</td>
<td>integer</td>
<td>Non-GBFS feeds</td>
<td>Non-GBFS Feeds</td>
</tr>
<tr>
<td>is_reserved</td>
<td>smallint</td>
<td>is_reserved</td>
<td>An identifier for whether the bicycle is reserved</td>
</tr>
<tr>
<td>is_disabled</td>
<td>smallint</td>
<td>is_disabled</td>
<td>An identifier for whether the bicycle is disabled</td>
</tr>
<tr>
<td>first_seen</td>
<td>timestamp</td>
<td>API poll timestamp</td>
<td>The timestamp of the first API poll in this location</td>
</tr>
<tr>
<td>last_seen</td>
<td>timestamp</td>
<td>API poll timestamp</td>
<td>The timestamp of the latest API poll in this location</td>
</tr>
<tr>
<td>may_have_gone</td>
<td>integer</td>
<td>N/A</td>
<td>An identifier to determine if the bicycle is still located at this location.</td>
</tr>
<tr>
<td>has_gone</td>
<td>timestamp</td>
<td>API poll timestamp</td>
<td>The timestamp of the last API poll in this location, NULL if in this location</td>
</tr>
</tbody>
</table>
The data collection and storage is very similar to that of the dock location data (Section 3.2.2) in its process of minimising redundant data entries. This is achieved through a series of checks that occur prior to creating a new data entry within the table. Namely, this process first checks for the presence of the same bicycle in the previous API poll by matching the ID values. If the bicycle is present, the script then ensures that the bicycle has not moved by comparing its current location to the last observed location. The GPS coordinates of bicycles are recorded to six decimal places, which offer locations that are precise to approximately 10 centimetres at the equator (Khetarpaul et al. 2011). Though this may seem very precise, it is important to acknowledge the static wandering of GPS coordinates, also known as ‘GPS drift’ (Forno et al. 2020) caused by the diffusion and reflection of GPS signals in dense urban environments with buildings and trees that can create variation in the coordinates returned even whilst the GPS receivers are in situ (Flüchter and Wortmann 2014). Therefore, instead of comparing exact locations, which are likely to see some variation, coordinates are rounded to four decimal places when checking for bicycle movements. In doing so, this alleviates the creation of new data entries that are a likely caused by ‘GPS drift’, and capture those movements over distances greater than approximately 11 metres at the equator. For those bicycles that have not moved and were observed in the previous poll, only the last_seen variable is updated and no new entries are created. For those bicycles that were not observed in the previous poll or those that have moved further than the four decimal place threshold are used to create a new data entry, taking the time at which this observation was made as the first_seen variable.

In conducting these data minimisation strategies, we can use new data entries to infer potential hiring occasions. This then enables the reconstruction of OD records by joining the last_seen location of a bicycle to its first_seen location. Although these data minimisation techniques have been employed in the collection and storage of dockless BSS data, it is important to acknowledge that a significant proportion of data entries are a result of data error as opposed to journeys occurring. These are primarily a result of GPS inaccuracies that are exacerbated due to the
lack of widespread adoption of Assisted-GPS (A-GPS) capabilities, in addition to the generally built up environments within which dockless BSS operate. On top of GPS inaccuracies, the movement of dockless bicycles as rentals are indistinguishable from that of operator management efforts. Combined, the recorded data in the bike tables are likely to contain a high proportion of non-user rental activity and as a result, require careful cleaning and processing procedures in order to remove these data and create an accurate record of valid user journeys. These processes have been explored in Chapter 7 to provide a detailed spatio-temporal and statistical determination of mobility dynamics within the dockless JUMP e-BSS in London.

A final note in relation to dockless data that is important to mention is the updating of GBFS standards to version 2.0. As briefly detailed in Section 3.1, in March 2020 the GBFS standard experienced some significant updates and changes, both to integrate additional features and micromobility modes, in conjunction with making the standards more stringent in relation to the privacy of movements. These privacy concerns were initially identified on the Github repository as an ‘pull request’ that was titled ‘Rotate of bike_id on free_bike_status #147’. To date, this has been one of the most discussed and debated issues on the data standard and was resolved using a community vote. Though there were only 7 participants, the vote included consumers of GBFS such as a representative from Google Maps as well as micromobility operators and producers of GBFS such as Uber (JUMP) and Bird. All votes were in favour of the change, and thus resulted in this change being implemented as part of the GBFS version 2.0 major release. GBFS producers are, therefore, required to obfuscate the static ID for vehicles and instead include a randomised identifier for each vehicle through those random and dynamic ID rotation methods mentioned in Section 3.1 (Xu et al. 2022).

Clarifying the standard in this way makes it impossible to identify linked OD and therefore inhibits the use of the data to reconstruct historical activity and track vehicle movement within the system. This aligns with guiding principles detailed on the GBFS Github repository, which specifies that it is ‘not intended for historical or archival data such as trip records’ (NABSA 2022). Such changes to the standard
have meant that previous data minimisation strategies in the collection of dockless BSS data have broken due to the inability to track individual vehicles. As a result, the collection process forgoes these processes in favour of collecting all data. This ensures that any future studies that may be able to work around such journey obfuscation strategies may be possible. For example, Xu et al. (2022) develop a methodological framework to utilise the new feeds to determine unlinked journey origin and destination locations by aggregating activity to coarse grids. This has proven to be accurate with an $R^2$ larger than 0.9 and a mean average error (MAE) smaller than 2. These methods can be used to understand popular origin and destination locations and times, but mean that it is impossible to determine linked OD journeys, hindering our understanding of journey flows and activity dynamics.

### 3.2.4 UCL BSS Data Summary

The three types of data that have been described in this section detail the BSS records that have been collected from the various open BSS API feeds. These enable the capturing of bicycle movement within both docked and dockless BSS and provide the foundation for the analysis presented within this thesis. The exploration, manipulation and analyses of these data goes some way to detail the various ways in which the open data can be employed to uncover novel, insightful and valuable understandings of system dynamics.

The primary merit of this collection of BSS data is in its scale, both temporally and geographically. As depicted in Figure 3.1, the data have been collected since 2010 and have accumulated data on over 500 BSS. These comprise docked BSS, though the increasing accessibility to data from dockless and hybrid BSS, due to updates to the GBFS feed, have increased this proportion over the last 5 years approximately in line with the availability of such systems as depicted in Table 3.6. Collectively, these data account for approximately a third of operational BSS, the remaining two-thirds for which API feeds are not publicly accessible or are yet to be integrated into collection progresses.

Though the collection has progressively increased over time, it is important to acknowledge that the data for systems are not always complete, containing gaps
where collection scripts have broken, or where operators have ceased to make these feeds publicly available. In such cases, the data that are unable to be recovered and data collection can only recommence once a new API feed or the collection scripts are resolved manually.

**Table 3.6:** Temporal variations in the proportion of BSS by operational type

<table>
<thead>
<tr>
<th>Year</th>
<th>Docked</th>
<th>Dockless</th>
<th>Hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>71.0%</td>
<td>26.3%</td>
<td>2.7%</td>
</tr>
<tr>
<td>2019</td>
<td>66.5%</td>
<td>30.2%</td>
<td>3.3%</td>
</tr>
<tr>
<td>2020</td>
<td>64.3%</td>
<td>31.4%</td>
<td>4.3%</td>
</tr>
<tr>
<td>2021</td>
<td>61.5%</td>
<td>33.7%</td>
<td>4.8%</td>
</tr>
<tr>
<td>2022</td>
<td>60.0%</td>
<td>35.3%</td>
<td>4.7%</td>
</tr>
</tbody>
</table>

Source: Meddin Bike-sharing World Map

Among the possible explanations for the termination of data, the primary reason is due to the closure of BSS. As can be observed on The Meddin Bike-sharing World Map, there are a significant number (over 1230 in June 2022) of BSS which have closed (Meddin et al. 2022). Due to the highly competitive nature of the market, there are numerous instances where operators have withdrawn operations from an area or have deemed profitability to be unlikely, such as the recent closure of Uber owned, JUMP BSS across Europe and being sold to a rival operator, Lime (Dillet 2020). In addition, there have also been instances where the local population have rejected the implementation of BSS, such as the implementation of the MoBike system in Manchester in 2017, as highlighted in Section 2.1.2.2.

In conjunction with missing data, there are also instances of data error which have been identified throughout the process of utilising these data. For example, in the formulation of various heuristics in Chapter 4, it was identified that there were some instances where dock capacity data would flicker between two values. This would lead to the overestimation of journeys and thus had to be programmatically identified and removed in order to obtain more accurate estimations. Similarly, during the maintenance and updating of the database to facilitate the differentiation between e-bicycles and bicycles in 2020, the collection scripts for the ‘ind’,
dock capacity tables were altered to create a new column for these of e-bicycles, as mentioned in Section 3.2.1. In doing so, this created an error which systematically double-counted the number of e-bicycles in a docking station, which was only identified 20 months later. As a result, the data for those systems which recorded the number of e-bicycles had to be reconstructed using a sophisticated Structured Query Language (SQL) script that recalculated the correct number of (e-)bicycles at each API poll. There are also numerous other data cleaning processes that are required in order to remove those data errors, such as in the use of dockless data that are detailed in Chapter 7.

It is important to acknowledge the incomplete and error ridden nature of the data that have been collected in order to ensure that these data are processed and analysed in the most effective manner. It is equally important to commend and admire the persistence and growth of this data collection endeavour, as there is no other database like it. Although these data have been used throughout academic studies, this thesis aims to exploit these data collection to their full potential, exhibiting the various ways in which the collection of open BSS data can be utilised to produce interesting and valuable insights into systems dynamics. With the growth of alternative forms of micromobility that utilise the same (or similar) data standards, these data collection processes can provide a methodological framework for the future studies of such systems.

3.3 Open Journey Data

Alongside BSS data that are accessible through API feeds, there are a small handful of systems that routinely release journey data. These are complete and clean OD records that are published by the operator retrospectively, providing a detailed and accurate account of all journeys that have occurred within a system. They provide significant advantages over data sourced from API feeds, primarily as a result of their reliability, being officially processed, cleaned and published by operators. Systems which release such data include:
• ‘Santander Cycles’ in London, UK
• ‘Bay Wheels’ in San Francisco CA, USA
• ‘Citi Bike’ in New York City NY, USA
• ‘Divvy’ in Chicago IL, USA
• ‘Capital Bikeshare’ in Washington DC, USA
• ‘Bluebikes’ in Boston MA, USA
• ‘Metro Bike Share’ in Los Angeles CA, USA
• ‘Indego’ in Philadelphia PA, USA
• ‘Nice Ride’ in Minneapolis MN, USA
• ‘MiBici’ in Guadalajara, Mexico

The limited availability of journey OD data can be attributed to a number of factors. Firstly, the primary factor in the limited availability is likely to be a lack of obligation to do so. This can be determined since those BSS which release such data are typically located in cities or countries that have open data agreements. For example, as mentioned in Section 3.1.1, under the Open Data Law, governmental data are mandated to be made publicly accessible in New York and published to the ‘Open Data NY’ website. Similar mandates are in place in cities such as Washington DC and Mexico City. In London, the BSS data were made available after Adrian Short, an open data developer, made a freedom of information request for the data (O’Brien 2011), that has subsequently been continued given that these requests would keep coming if the data were not released.

Privacy and ethical concerns are further reasons those non-mandated BSS do not release such data. In this context, there has been a growing awareness surrounding the exploitation of personal data, as highlighted at the beginning of this chapter. Mobility data are part of this and though they may not contain any information in relation to the individual user, the geographic precision of these data could lead to the identification of an individual. This was the primary reason for the obfuscation of GBFS dockless bicycle location data and can also relate to the release of open journey data. This is also the reason why some of hybrid systems only make journey available for those occurring between docking stations journeys. In the case of the
'Bay Wheels’ San Francisco BSS, it appears as though those dockless journeys are included, the accuracy of the coordinates are only made available to two decimal places, instead of the minimum of six decimal places for journeys between docking stations. This approximately reduces the geographic accuracy of dockless journey locations from just 9cm to over a 900m, showcasing the privacy concerns associated with such systems. These data are also generally released in batches throughout the year as opposed to having real- or near-real-time data feeds as such identification concerns are more likely to occur with these data.

Another reason for such a limited availability is due to commercial pressures as BSS seek to obtain sponsorship deals. In the increasingly saturated nature of the micromobility market, there is ever more competition to win market share, offer a better service and gain brand loyalty in comparison to other operators from both local municipalities and end users. Operators are typically in competition to run their systems within an area, therefore it is usually in their best interests to disseminate information which make them seem most favourable. This is also the case for BSS sponsorships, which are a common way for operators to gain additional income from their maintenance. Examples include JCDecaux and Clear Channel, European advertising companies that often provide infrastructure, maintenance and servicing in exchange for rights to commodify municipal public space through advertising bill boards (Médard de Chardon and Caruso 2015). As a result, there have been instances of operator exaggerations to win support from sponsorships over competition (Médard de Chardon and Caruso 2015). For example, one of the largest operational docked BSS, Hangzhou in China, reported to have over 85,000 bicycles in 2017 (Lawson 2017) but through an exploitation of the UCL BSS database, it is apparent that this a severe over exaggeration, with no more than 45,000 bicycle being observed at any one time. Therefore, in making their data open, operators would risk their ability to compete and thus disincentives the release of such data.

Although there are numerous potential reasons for the lack of such data, it is important to encourage the release of historical BSS journey records in order to reduce the reliance on independently collected BSS data, such as the UCL BSS data.
3.4 Chapter Summary

This chapter provides a detailed overview of the BSS data landscape. It starts by exploring the role of data standards generally, then, focusing on BSS, it explores the functionality and nature of the GBFS data standard. This has increasingly become the de facto standard for micromobility open data dissemination. Here, we are able to determine the necessity for the standard in helping to facilitate better integration of micromobility location and availability information to end users through operator and third party applications. GBFS consists of several files which have been continually updated in line with the evolving nature of micromobility modes, in addition to suggestions and issues presented by the community.

Although the GBFS data standard was designed as a means to disseminate real- or near-real-time data on fleet availability within micromobility systems, with no intention for historical or archival trip records (as per the guiding principles), due to the very limited availability of clean and complete journey data (Section 3.3), analysis on the dynamics of BSS have been reliant on exploiting such open

collection, due to their limitations in accuracy, precision and completeness. These open journey data are invaluable in their ability to provide a source of verified, clean and complete journeys to better capture the true nature of their use and growth. Such insights will help to better guide the maintenance and optimisation of BSS and similar micromobility modes in the future. Due to the commercial and personal sensitivity of such data, it is unlikely that there will be such open data availability and thus it is important to make the most of the resources that we have. Thus, the UCL BSS data collection are the primary BSS data that have been employed in this thesis. This being said, the open journey data sources, for those 10 systems for which it is available, have been used as a means to validate the various data cleaning and manipulation techniques employed on the UCL BSS data collection. This is most integral to the journey estimation techniques detailed in Chapter 4, where journey data is used as a tool to compare, tune and optimise scripts to obtain the most accurate estimations.
data feeds from BSS API.

The UCL BSS data collection is the largest and most comprehensive collection of BSS data that has been collected using an assortment of carefully constructed Python collection scripts. These scripts are designed to make API polls from those various BSS open API every 2/10/20+ minute interval, converting the various GBFS (or similar) files into standardised tables within a database. These data have been collected since 2010 and have amassed data on over 500 BSS that are of a docked, dockless or hybrid operating nature. These collection processes are detailed throughout Section 3.2 and highlight the cumbersome task of collecting those data. The section also highlights the shortfalls of such data in its inability to differentiate between true user movements and inclusion of data errors, its tendency to be incomplete for some systems as well as its inability to capture data for all BSS (Figure 3.2) and the recent limitations of dockless journey identification due to the obfuscation of bicycle identifiers (Xu et al. 2022).

Although there are numerous limitations to the UCL BSS data collection, it is vital to acknowledge its importance in facilitating the research presented throughout this thesis. The limitations of these data are able to be overcome through the careful manipulation and cleaning of the records, using the open journey data records as a means to validate those processes. As such, the chapter has provided a transparent overview of the way in which the UCL BSS data have been collected and thus provides a foundational understanding of the data that have been used throughout this thesis.
Chapter 4

Manipulating and Validating Bicycle Sharing System Data

Although the UCL BSS data collection (detailed in Section 3.2) provides static snapshots of the location and availability of bicycles within a system at any single collection instance, the data in their raw state are not conducive to analysis of BSS activity dynamics. Unlike the open BSS journey data (Section 3.3), the BSS data collection from various open API feeds are made available so that real-time location and availability of bicycles can be accessed by applications such as Google Maps. This is hugely beneficial to the operator, functioning as a means to advertise their services to a much wider audience, as well as the end user, enabling seamless integration to the incumbent routing platform, with over a billion monthly users (Russell 2019). As a result, these data are the most abundant source of BSS data. However, these data require careful cleaning, manipulation and validation in order to be transformed into informative, accurate and comparable metrics and variables.

Thus, this chapter will focus on the various heuristics that have been employed in order to convert the UCL BSS data collection into a number of metrics. These metrics include processes to estimate the number of journeys and bicycles in addition to various geographical and contextual attributes of each BSS. Section 4.1 will detail these methods by which the UCL BSS data collection have been manipulated and validated to generate such information. Section 4.2 will then provide a descriptive overview of these metrics through an exploration of their values in offer-
ing system-scale, long-term insights of the London Santander Cycles system, highlighting their depth and breadth. The metrics and the descriptive analysis presented in this chapter provide the foundations for the detailed, quantitative comparison of BSS dynamics at the global-scale in Chapter 5, in addition to metrics that have also been employed in the analysis presented in Chapter 6 and Chapter 7.

4.1 Metric Creation

4.1.1 Journey Estimation

Journey activity is the most fundamental measure of dynamics within BSS, providing an overview of the temporal distribution of use. Having the ability to analyse this activity distribution is invaluable in understanding its dynamics - from the most granular temporal scales, such as the time of day enabling inferences of potential trip purpose, to long term changes in activities, such as seasonal variations and the general growth of a system. Journeys are also a foundational in terms of their importance in creating metrics such as TDB, a metric which has been used throughout BSS literature to determine the efficiency or success of a system (Médard de Chardon et al. 2017; Zhang et al. 2015; Fishman et al. 2014b). Therefore, being able to estimate the occurrence of journeys within BSS was essential. Since the majority systems in the UCL BSS database are third-generation docked BSS, this will outline the process of manipulating the dock capacity data, detailed in Section 3.2.1. Although this section will focus on docked BSS, alternative journey derivation methods have been synthesised for dockless BSS data that are detailed in Section 7.2.

The raw dock capacity data contains only five variables (Section 3.2.1); the docking station ID, the number of bicycles, number of spaces, total dock capacity and the timestamp at which that observation was made. Although these data do not enable the reconstruction of OD data, they do provide a valuable means to estimate the number of journeys that occur within the system. These can be calculated by the exploiting of the fact that the UCL BSS data collection are collected at regular time intervals. As such, ordering the API polls chronologically provides some indication
4.1. Metric Creation

More specifically, by investigating the flow of bicycles in and out of each docking station chronologically, the movement of bicycles can be used to infer the number of journeys that have occurred. This is demonstrated through the illustration in Figure 4.1, which depicts two consecutive observations from a fictional docking station. In this example, the data are collected every two minutes, like those most active BSS within the UCL collection. Between the API polls, the number of bicycles in this docking station decrease, therefore inferring that a bicycle has left this docking station and thus a journey being started.

<table>
<thead>
<tr>
<th>dock_id</th>
<th>timestamp</th>
<th>bikes</th>
<th>spaces</th>
<th>total_docks</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>2018-10-14 14:24:00</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dock_id</th>
<th>timestamp</th>
<th>bikes</th>
<th>spaces</th>
<th>total_docks</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>2018-10-14 14:26:00</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

**Figure 4.1:** Diagram illustrating the basic mechanisms of the journey estimation methodology

This mechanism is employed by codifying this process into a SQL query. The query ensures the data are initially ordered chronologically and grouped by the docking station ID then, using the ‘lag’ functionality, compares the number of bicycles in each row to the previous row. If there is a decrease in the number of bicycles between consecutive observations within the same docking station, each unit decrease is counted as the start of a journey. This process is repeated for all docking station with a system. Depending on the specification of the query, this journey esti-
Information methodology can be used to evaluate the number of journeys at any temporal resolution, all the way down to each collection instance. Although this is possible, this methodology has used to construct journey estimations at the hourly level in order to minimise the amount of redundant data for each BSS. This would also ensure that the journey estimations would be available at granular intervals whilst also being easily aggregated to larger temporal scales should such data be more appropriate. The query was scripted in a parallel process on the ‘ind’ tables of each BSS that were indexed on both timestamp and docking station ID variables. This was essential to improve computation time given the extensive nature of the data collection. These scripts would then simultaneously create and store the manipulated data into new tables that were named ‘je’ in reference to journey estimation. These tables included only three variables; the date, the hour and the estimated number of journeys. These were also indexed on the date and hour columns to ensure their quick retrieval in future queries.

It is important to acknowledge that similar journey estimation methodologies have been previously explored, such as the analysis conducted by Médard de Chardon and Caruso (2015). Here, the authors conduct some exploratory analysis on the various scales of temporal aggregation in order to determine the level at which journey estimations were most accurate. Analysing dock capacity data collected at 10-minute intervals from eight BSS in Europe and North America, the study finds that aggregating estimations to the daily level produced the most accurate root mean square error (RMSE) when validated against open BSS journey data. In addition, the authors find data that are collected more frequently help to produce more accurate estimations. Although the analysis conducted by Médard de Chardon and Caruso (2015) are very similar to that of the journey estimation processes conducted in this thesis, there are areas where the data and processes differ. Therefore, the estimation processes conducted here were subject to further journey validation and cleaning steps, detailed in Section 4.1.2.
4.1.2 Journey Estimation Validation and Cleaning

In order to determine the validity and accuracy of the journey estimation methodology detailed in Section 4.1.1, those estimates were compared to the complete and clean journey data for which data were available (Section 3.3). This was conducted at the daily aggregation level (i.e. midnight to midnight). Although journey OD data were only conducted for 10 systems, this process is paramount in ensuring that estimations are representative not only as a measure of journeys, but also as a variable that would provide the foundations for other BSS metrics. To demonstrate the validation and cleaning process, we follow the example exploring the Santander Cycles London docked BSS. Though this is the only example presented, the following processes were conducted for each BSS for which journey data were available.

Firstly, in order to validate daily journey estimation calculations were accurate, estimations counts were compared against journey data. In doing so, it is possible to determine the percentage difference between these figures for each particular day. Figure 4.2 depicts a visual representation of this comparison. Initial inspections of these data identified instances of both under and over estimations of journey counts on any given day. Though the estimations varied in this way, there this process also enabled the identification of several anomalously large days where overestimations were in excess of 10 times that of the observed number of journeys. Such anomalous results warranted a detailed investigation into the causation of such severe overestimations.

As such, inspecting those days on which anomalous estimations occurred, it became apparent that there were major errors within the raw dock capacity data that were collected. On these days dock capacity data would indicate that the number of bicycles were rhythmically fluctuating between two values over extended durations. As the journey estimation methodology identifies these fluctuations in the number of bicycles between API polls as the start of journeys, these periods of data noise were identified as the root cause of large overestimations in the number of journeys. These fluctuations could be positively identified as data noise, as opposed to actual user activity, due to the consistency in the two values they would fluctuate between
in addition to their duration. Although the true source of the data noise are not confirmed, it is hypothesised that these are likely to occur when there are electrical faults within particular docking stations in addition to operator maintenance on API feeds.

As a result, it was necessary to be able to identify these periods of data noise in order to remove those data from the journey estimations. Figure 4.3 depicts a diagrammatic representation of the noise in addition to providing an overview of the methodology that was implemented to detect these instances of data error within the dock capacity data. Here, a script was created in a similar fashion to that of the journey estimation methodology detailed in Section 4.1.1. Unlike the journey estimation methodology, the script compares consecutive rows in order to identify extended periods where the absolute difference in number of bicycles between API polls are the same. Due to the scale of the UCL BSS data collection, it is necessary to determine appropriate threshold levels for the absolute differences and the number of consecutive occurrences that would enable the detection across BSS of
4.1. Metric Creation

varying sizes. Through a rigorous trial-and-error approach, various threshold values were tested and manually inspected to determine the proportion of true positive identifications in comparison to false positive identifications. The priority in this process was to ensure that all true positive occurrences were identified whilst minimising detection of false positives. Due to the presence of small, infrequently used BSS within the collection, it was necessary to set these values relatively low to capture such occurrences within smaller systems. This meant that differences greater than one bicycle over a period longer than six data collection instances were identified as potential periods of noise (Figure 4.3).

![Mass Flicker Detection](Image)

**Figure 4.3:** Diagrammatic representation of the data noise detection methodology

Through this process a total of over 1,500 days were identified to potentially contain instances of data noise. After manual inspections of these data had been undertaken to remove instances of false positives, the final data cleaning effort identified 1,144 days among 125 BSS within which data noise were detected. The results of such cleaning processes for London are depicted in Figure 4.4. The commonality of these data noise instances across systems operated by different companies indicate that this is issue, though anomalous, does appear to occur frequently enough
to warrant the identification of such occurrences, especially under circumstances where these data used to estimate journeys, much like this thesis and Médard de Chardon and Caruso (2015). Although this does produce gaps in the journey estimation data, these cleaning processes were vital in accurately estimating activity.

**Figure 4.4:** Cleaned journey estimations in comparison to open journey data in London

The methodology relied on the identification of data noise in the raw data as opposed to instances of anomalously large overestimations of journeys due to two primary reasons. Firstly, BSS are a flexible mode of transportation that are subject to large and rapid changes in demand. Public holidays (Corcoran et al. 2014; Kim 2018; Kutela and Teng 2019) and PT closures and strikes (Fuller et al. 2012; Saberi et al. 2018; Yang et al. 2022) have been identified to have significant impacts on activity. Thus, identification of data errors in this way would be prone to identifying these genuine occurrences of demand fluctuation as periods of data noise. Secondly, and most importantly, the level of anomalously large overestimations are subject to much greater variance over BSS of different sizes in comparison to the identification of data noise at the docking station level. Therefore, the identification on raw data
was favoured due to its scalability across the UCL BSS data collection.

Once all those days had been removed, journey estimation values were assessed against their published journey data counterparts across 10 systems for which these data were available. Table 4.1 provides a summary of the descriptive statistics of these comparisons between estimated journeys. In each of the validated cases, best efforts were made to employ all of the data that were available, though there are some instances dock capacity data feeds had broken and therefore meant that journeys were unable to be estimated. Assessing the validation of journey estimations, it is apparent that weekends are better estimated in comparison to weekdays. This is likely a result of the reduced amount of activity within the system on these days. There also appears to be some variation in the magnitude errors between systems. Although this is the case, the vast majority of systems showcase the highly accurate nature of the journey estimation methodology, being less than 5% different to validated journey values. Looking more holistically across all the systems, journey estimations are less than 1% off, further highlighting their validity and reliability in accurately estimating journey activity within systems.

<table>
<thead>
<tr>
<th>System</th>
<th>Weekday Mean</th>
<th>Weekday Mean Absolute Deviation</th>
<th>Weekend Mean</th>
<th>Weekend Mean Absolute Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Francisco</td>
<td>-19.87%</td>
<td>10.08%</td>
<td>-16.09%</td>
<td>8.53%</td>
</tr>
<tr>
<td>New York</td>
<td>-15.54%</td>
<td>23.09%</td>
<td>-9.39%</td>
<td>25.28%</td>
</tr>
<tr>
<td>Guadalajara</td>
<td>-4.48%</td>
<td>11.01%</td>
<td>-3.08%</td>
<td>9.09%</td>
</tr>
<tr>
<td>London</td>
<td>-3.62%</td>
<td>4.80%</td>
<td>-2.63%</td>
<td>7.85%</td>
</tr>
<tr>
<td>Chicago</td>
<td>-3.02%</td>
<td>7.62%</td>
<td>1.75%</td>
<td>11.38%</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>0.01%</td>
<td>6.05%</td>
<td>-4.34%</td>
<td>6.74%</td>
</tr>
<tr>
<td>Washington D.C.</td>
<td>1.77%</td>
<td>7.02%</td>
<td>1.33%</td>
<td>8.24%</td>
</tr>
<tr>
<td>Boston</td>
<td>2.33%</td>
<td>8.01%</td>
<td>5.04%</td>
<td>7.43%</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>11.47%</td>
<td>7.74%</td>
<td>6.29%</td>
<td>9.00%</td>
</tr>
<tr>
<td>Minneapolis</td>
<td>15.00%</td>
<td>6.57%</td>
<td>14.64%</td>
<td>7.31%</td>
</tr>
<tr>
<td>All 10 Systems</td>
<td>0.64%</td>
<td>9.92%</td>
<td>0.77%</td>
<td>10.43%</td>
</tr>
</tbody>
</table>

In spite of the overall accuracy, it is important to acknowledge that these journey estimations are not perfect and that depending on the system, there appear to
be some consistent under or overestimations. These errors are likely to occur for a number of reasons that are unavoidable given the nature of the data and the scale of the process.

The principal reason for underestimation is a result of the data collection methodology. Since API polls are made at regular time intervals, there are small windows within which the start and the end of a journey within a docking station can coincide. The collection intervals should be proportional to the activity dynamics of BSS and thus minimise these losses in journey detection, but it is paramount to acknowledge that there will still be such instances of undetectable journeys and therefore underestimations in the number of journeys (Médard de Chardon and Caruso 2015). These occurrences are much more likely in those systems that are consistently observed to have underestimations, such as San Francisco, New York and Guadalajara. In addition, it is important to note that there are some occurrences where data collection fails as a result of breaks in the feed or collection scripts that are quickly rectified. These instances are not penalised in the journey estimation process and may exacerbate underestimation, much similar in nature to those journeys that are unobserved between collection instances. Therefore, it is clear that future BSS API efforts should prioritise frequency of data collection and ensure data that are collected are complete prior to employing these heuristics to avoid such underestimations in the future.

Overestimations are less common but pose a greater threat to journey estimations as a result of their highly indistinguishable and contextual nature. Within each BSS, operator maintenance and redistribution efforts differ based on the specific utility and strategies implemented. Therefore, unlike operator published journey data that have been cleaned and verified to only include user activities, dock capacity data are only able to convey the real-time availability of bicycles, regardless of whether the dropping-off or picking-up of bicycles was conducted by operators or users. These specificities are highlighted by the abundance of academic literature seeking to optimise operator redistribution efforts (Dell’Amico et al. 2014). Thus, identifying such interventions are impractical at the scale of this metric cre-
4.1. Metric Creation

The total fleet size of BSS is another important, yet simple metric that offers valuable insights into the availability of bicycles. Analysing the size of BSS fleets over time can provide indications of whether the system is growing, shrinking or remains stagnant, as well as providing the foundations of the TDB metric. Unlike the journey estimation methodology presented, these processes are much simpler and do not require significant validation and cleaning.

Although journey estimations are susceptible to errors in the dock capacity data, since the fleet size of BSS are not subject to large short term changes, the maximum number of bicycles observed within a day can provide an accurate measure. In order to calculate this, the number of bicycles observed at each API poll are summed. Then, as the name suggests, the largest value of total bicycles are taken to infer the fleet size for each day. This approach has been employed in order to best account for the bicycles that may be in use at any particular collection interval. As a result, in the majority of cases, this maximum observation typically occurs during the middle of the night, when user activity is at a minimum and thus the majority of bicycles parked at docking stations.

Here, it is important to acknowledge that fleet sizes often slightly underrepresent the total number of bicycles that are available in BSS due to the presence of user activity. This is especially true for BSS that are located in large cities which observe activity all throughout the day, though underestimations are likely to be insignificant and negligible due to the lack of activity during the night in the vast majority of systems. There are always likely to be fluctuations in fleet size between days due to bicycle malfunctions and thus daily maximum bicycles aim to provide the most granular and accurate depiction of changes in BSS fleet size.
4.1.4 Trips per Day per Bicycle

As mentioned in Section 4.1.1, TDB is a very valuable metric that has been employed throughout BSS literature to infer a holistic measure of BSS use and efficiency (Zhang et al. 2015; Fishman et al. 2014b) and in some cases has been used to determine the success of systems (Médard de Chardon et al. 2017). In this thesis, TDB is employed as a comparative measure of system efficiency, as it is a simple measure that can be derived for all those BSS for which journeys and fleet size can be estimated using dock capacity data.

It is not taken to refer to the success of a BSS due to the complexity of the term. The success of a BSS is not only determined by the level of activity, but rather in its ability to achieve its aims. These can range from providing an equitable and accessible means of urban transportation to ‘public transport captives’, a term used to refer to individuals who only have access to PT, and are therefore at the mercy of their services (García-Palomares et al. 2013; Beimborn et al. 2003; Ai et al. 2019; Carney et al. 2022), to facilitating greater uptake of PT modes through multimodal journey connections (Guidon et al. 2019; Shen et al. 2018; Xu et al. 2019), in turn reducing the level of harmful GhG emissions from alternative, less efficient modes of urban transportation (Saltykova et al. 2022; Chen et al. 2022b). Efficiency presents a more apt use of TDB, as it is a measure of activity in relation to the fleet size. As an increasing proportion of BSS are operated by profit seeking private firms, it is in their best interests to optimise these systems to appropriately serve and meet the demands of the population whilst also minimising operational costs.

In order to calculate TDB, the total number of estimated journeys calculated in Section 4.1.1 are simply divided by the associated maximum number of bicycles that are detailed in Section 4.1.3 for each day where data were available. TDB values greater than one are taken to indicate relative efficiency as the number of bicycles are proportional to the activity within the system, whilst values smaller than one are therefore indicative of relative inefficiency. Maximising the operational efficiency of these systems are paramount as they represent the widespread and
adoption of systems that also minimise the amount of excess capacity, which is the primary objective of innovation in the sharing economy sector.

4.1.5 Entropy

As highlighted in Section 3.2.1, dock capacity data provide continuous snapshots of the distribution of bicycles within a system upon each API poll. Although these values can be used to produce static depictions of the fleet’s dispersal, much like the visualisations on bikesharemap.com that utilise these data for this exact purpose, a simplistic and holistic measure of bicycle distribution was created to capture the level of imbalance within each BSS. The metric that has been calculated has been coined the ‘entropy’ of BSS, drawing on the term used to measure the state of disorder and randomness within systems among the fields of thermodynamics (Leff 1996), statistical physics (Pressé et al. 2013) and information theory (Carter 2007; Vedral 2002; Gray 2011).

Entropy, within this thesis, develops on the ‘load factor’ measure calculated in O’Brien et al. (2014) and has been calculated by initially computing a perfectly balanced system, with bicycles that are evenly distributed throughout each of the docking stations, proportional to their total capacity. This was constructed by considering the total number of available bicycles (taken from calculations in Section 4.1.3) and the number of docking stations and spaces at each docking station that are available in the bikelocations table (detailed in Section 3.2.2). Using these values, the appropriate proportion of bicycles that should be in each docking station for a perfectly distributed fleet are able to be determined on a daily basis. By taking these balanced values, it is then possible to calculate the level of imbalance in the BSS through the difference between these perfectly distributed bicycle values to the actual bicycle distribution at each docking station. Whilst perfect distributions are calculated for each day, levels of imbalance are determined for each API poll occurrence. Dividing the two values produce entropy measures between zero and one; zero indicating a perfectly balanced distribution of bicycles whilst one indicates circumstances of complete disorder.

This provides a holistic overview of the distribution of bicycles within a system
and, in turn, enables some determination of the dispersal of journeys. For example, journeys of a commuting nature are typically more directed and ordered, with journeys that start or end at a CBD. This means that journeys are more likely to cause the distribution of bicycles to be more imbalanced. In comparison, journeys of leisure and exercise purposes are typically more heterogeneous and present fewer directed origins and destinations. Therefore, bicycles tend to more evenly distributed and thus closer to the optimal distribution of bicycles. In calculating the maximum entropy values observed at hourly intervals, we can gain a general overview of the distribution of journeys throughout a day.

It is important to acknowledge that operators are typically aware of user demands within their given contexts, meaning that systems are never perfectly distributed. Given the BRP, operators are likely to distribute bicycles appropriately to cater for such demands. For example, they may keep bicycles in docking stations that observe a large number of hires at particular times, whilst also keeping spaces in docking stations that observe a large number of drop-offs. As such, entropy is merely used to indicate the general changes in the distribution of bicycles within a system.

4.1.6 Area, Docking Station Density and Journeys per Square Kilometre per Day

In addition to metrics that enable the determination of the size and use of BSS, metrics that provide an indication of the spatial attributes of each BSS are also calculated. These detail valuable information in relation to the spatial extent, the distribution of docking stations and the spatial efficiency of BSS. Here, the methodology to determine the operational area, dock density and journeys per square kilometre per day are outlined.

The operational area of a BSS helps to provide some indication of its spatial extent and are valuable due to their abilities to infer the expansion or contraction of systems that are associated with the number of docking stations. Since the (de)construction of docking stations occur very infrequently and in batches once BSS are in service, the metric was calculated on a quarterly basis. Operational areas
were calculated by exploiting the `bikelocations` table within the UCL BSS data collection, identifying all of the unique docking station IDs that were available in each quarter. This was determined by examining the `created_dt` and `updated_dt` variables, removing all those docking stations which did not exist within the quarter in question. Once acquiring the coordinates of available docking stations for each quarter, the operational areas were calculated by placing a 500 metre radius buffer around each docking station. These buffers are then dissolved to create an operational area polygon. A 500 metre radius buffer was chosen upon examining the literature surrounding the propensity to use BSS, with 500 metres being found to have been the maximum walking distance that users are willing to travel (Bachand-Marleau et al. 2012). Furthermore, studies have found that typical distances between docking stations are around 200-300 metres, meaning that with a 500 metre buffer should capture the entire system (O’Brien et al. 2014).

The area of the BSS provides the foundations for calculating docking station density and journeys per square kilometre per day variables. In order to measure the docking station density, the number of unique docking stations IDs which were identified in the calculation of the operational area were simply divided by the area. This metric complements the operational area measures to provide some insight into the number of docking stations across the system. Similarly, the number of journeys per square kilometre per day divide the number of estimated journeys (detailed in Section 4.1.1) by the quarterly operational area. This metric serves as a measure to indicate the density of activity that are occurring across the BSS. Together, these spatial metrics provide a holistic determination of the geographic density and distribution of the BSS and its activity. These metrics are underutilised among BSS literature and provide additional dimensionality to the analysis that are conducted in this thesis.

### 4.1.7 Confounding Variables

In addition to the spatial metrics, the `bikelocations` table can be used in order to infer additional contextual information in relation to each BSS. The choice of additional confounding variables to be calculated were determined based on two primary fac-
tors. Firstly, the literature surrounding determinants of BSS demand were studied to identify those that were commonly found to have a significant relationship. This analysis of the literature is presented in Table 4.2, providing an overview of the variables, their relationship to BSS demand and the paper within which they were examined. The variables are categorised into groups that are loosely based on the categories identified in Eren and Uz (2020), but also resemble those categories of variables used across studies that analyse the impacts of built and natural environment factors using regression models such as Médard de Chardon et al. (2017) and El-Assi et al. (2017).

As exhibited in Table 4.2, there are a multitude of variables that were found to have significant impacts across BSS and although the calculation of each of these variables for the BSS in the UCL collection were attempted, due to the enormity of the data it was necessary to create a scalable methodology that would produce comparable and consistent metrics. As a result, the use of data with global coverage were essential, thus limiting the number of confounding variables that were able to be calculated. These were primarily in the form of global raster datasets to ensure homogeneity across all systems in the way that each variable was measured and calculated. In an effort to account for the various categories of variables that were found within the literature, best efforts were made to calculate at least one variable in each category. This resulted in the estimation of the population, average precipitation, amount of cycling infrastructure and the topography within each BSS. The source of these data are depicted in Table 4.3.

Population counts were used in the socio-demographic category since population were consistently found to have positive correlations with BSS activity and would help to provide valuable contextual information on the number of people within or in close proximity to each BSS. Similarly, precipitation was consistently observed to have negative impacts on BSS use and therefore was included to include some attribution of the climatic contexts of each system. The topography of an area is also a major contributing factor to the use of a BSS, enabling an insight into its ‘cyclability’, with flatter areas that are typically more conducive to such physical
### Table 4.2: A summary of variables that influence the use of BSS

<table>
<thead>
<tr>
<th>Variable</th>
<th>Relationship</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BSS Factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of docks</td>
<td>Positive</td>
<td>(Faghih-Imani et al. 2014) (Médard de Chardon et al. 2017)</td>
</tr>
<tr>
<td>Dock density</td>
<td>Positive</td>
<td>(Médard de Chardon et al. 2017)</td>
</tr>
<tr>
<td>Dock Capacity</td>
<td>Positive</td>
<td>(Tran et al. 2015) (O’Brien et al. 2014)</td>
</tr>
<tr>
<td>Distance from station</td>
<td>Negative</td>
<td>(Tran et al. 2015) (Bachand-Marleau et al. 2012)</td>
</tr>
<tr>
<td><strong>Socioeconomic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>Positive</td>
<td>(Médard de Chardon et al. 2017) (Faghih-Imani et al. 2014) (Tran et al. 2015)</td>
</tr>
<tr>
<td>Income</td>
<td>Positive</td>
<td>(Fishman et al. 2014b) (Roy et al. 2019) (Woodcock et al. 2014)</td>
</tr>
<tr>
<td>Age</td>
<td>Positive</td>
<td>(Fishman et al. 2014b) (Zhang et al. 2016a)</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>(Fishman et al. 2014b) (Zhang et al. 2016a) (Goodman and Cheshire 2014) (Murphy and Usher 2015)</td>
</tr>
<tr>
<td>Jobs</td>
<td>Positive</td>
<td>(Tran et al. 2015) (Woodcock et al. 2014)</td>
</tr>
<tr>
<td>Education</td>
<td>Positive</td>
<td>(Fishman et al. 2014b) (Shaheen et al. 2013)</td>
</tr>
<tr>
<td>Ethnicity (White)</td>
<td>Positive</td>
<td>(Buck et al. 2013)</td>
</tr>
<tr>
<td><strong>Weather/Climate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rainfall</td>
<td>Negative Insignificant</td>
<td>(Corcoran et al. 2014) (Miranda-Moreno and Nosal 2011)</td>
</tr>
<tr>
<td>Wind speed</td>
<td>Negative</td>
<td>(Corcoran et al. 2014) (Miranda-Moreno and Nosal 2011)</td>
</tr>
<tr>
<td>Air pollution</td>
<td>Negative</td>
<td>(Campbell et al. 2016)</td>
</tr>
<tr>
<td>Temperature</td>
<td>Positive Insignificant</td>
<td>(Faghih-Imani et al. 2014) (Corcoran et al. 2014)</td>
</tr>
<tr>
<td>Humidity</td>
<td>Negative</td>
<td>(Faghih-Imani et al. 2014)</td>
</tr>
<tr>
<td><strong>Topography</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>Negative</td>
<td>(Frade and Ribeiro 2014) (Mateo-Babiano et al. 2016)</td>
</tr>
<tr>
<td>Altitude</td>
<td>Negative</td>
<td>(Tran et al. 2015)</td>
</tr>
<tr>
<td><strong>Infrastructure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cycling Infrastructure</td>
<td>Positive</td>
<td>(Fishman et al. 2014b) (Faghih-Imani et al. 2014) (Buck and Buehler 2012) (Mateo-Babiano et al. 2016)</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public/School Holidays</td>
<td>Insignificant</td>
<td>(Corcoran et al. 2014) (Borgnat et al. 2011) (Brandenburg et al. 2007)</td>
</tr>
<tr>
<td><strong>Temporality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time of day</td>
<td>–</td>
<td>(Faghih-Imani et al. 2017) (O’Brien et al. 2014)</td>
</tr>
</tbody>
</table>
activity. Finally, the amount of cycling infrastructure was calculated as it helps to provide an understanding of the safe cycling facilities that are available to users.

The estimation of each of these confounding variables begin with the same processes that were used to calculate operational areas (as detailed in Section 4.1.6), wherein docking stations within each system were buffered to create polygons. These polygons were then placed on top of each global dataset to identify the values that were contained inside of BSS area that were then aggregated or tallied. In each of these processes, the buffer distance around docking stations varied in order to best account for the resolution of these data and the way in which BSS users would interact with them.

Starting with population and precipitation measures, a one kilometre buffer was chosen in order to best capture the population that would be likely to use the BSS and subsequently be impacted by the effects of rain. Although a 500 metre buffer is used to calculate operational area, as previously mentioned, this was employed to create a continuous polygon to delineate the extent of a BSS due to docking stations typically being between 200 to 300 metres apart. The one kilometre buffer serves to demonstrate the wider population distribution surrounding the BSS that may use systems as part of multimodal journeys. In addition, the precipitation
4.1. Metric Creation

Data were only available at the one kilometre resolution, meaning that aggregating measures in this way would enable the most accurate estimations.

On the other hand, the topography of BSS were determined by aggregating the gradient from the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) using 300 metre buffers. This smaller distance was elected since user cycling activity would likely not go outside of these bounds, given that they have to end journeys within docking stations. 300 metres should therefore capture the general gradient of urban contexts, especially in those systems that are located in valleys where topography varies significantly outside of the bounds of the BSS and therefore enable inferences of cyclability.

Finally, the amount of cycling infrastructure was calculated slightly differently to the other confounding variables due to the lack of a global raster dataset. Here, it was necessary to use OpenStreetMap (OSM) data in order to identify where cycling infrastructure was present at a global scale, setting the feature key to *highway* and feature values to *cycleway*. In this case, convex hulls were used to determine the area within which the length of such infrastructure would be summed. Much like the topography, journeys within docked BSS are limited to cycling between docking stations and thus results in activity which is constrained to these boundaries.

To further validate these subjective decisions that were used to calculate these confounding variables, sensitivity analysis was conducted when such metrics were employed in the global comparison of BSS in Chapter 5. Here, a variety of buffer sizes were tested to ensure the way in which they were calculated did not have significant impacts on our perception of systems. Rerunning the analyses using a variety of buffer distances and combinations between 100 and 500 metres exhibited no significant changes in BSS classification, thus strengthening confidence in the validity of the decisions that were made when estimating confounding variables. Further details of this sensitivity analysis are presented in Section 5.5.1.

4.1.7.1 Confounding Variable Limitations

Since the calculation of the confounding variables relied on additional data sources outside of the UCL BSS data collection, it is important to acknowledge the limita-
tions of such data, as has been highlighted for the BSS data throughout Chapter 3. In the selection of data for confounding variables, best efforts were made to acquire the most granular and up-to-date sources to appropriately capture additional contextual information. As mentioned, it was necessary to choose sources that were of global coverage to ensure consistency in the measures, thus limiting the selection of appropriate data.

In the context of OSM data, the primary limitation is in the way the data are collected. OSM is an open-source collaborative mapping project that provides volunteered geographic information (VGI) (Goodchild 2007). These include various attributes of built and natural environment factors including cycling infrastructure. As a result of the volunteered nature of these data, the quality and coverage of the data have come into question across a number of academic studies, with particular concerns over the Global South (Yeboah et al. 2019) and rural areas (Hagenauer and Helbich 2012; Koukoletsos et al. 2012). Although this is the case, OSM data are increasingly being employed across academic studies, including transportation studies, due to their accessibility (Camargo et al. 2020). Studies have also argued that the coverage of OSM data are more than sufficient and contain minimal inaccuracies (Haklay 2010; Girres and Touya 2010; Zielstra and Zipf 2010), in conjunction with evidence that OSM are a more granular and regularly updated source of data for cycling infrastructure in comparison to proprietary alternatives like Google Maps (Hochmair et al. 2015). As a result, though OSM data are not perfect in coverage and accuracy, they are an invaluable source of VGI that is regularly updated and thus prove to be the most optimal data for the acquisition of cycling infrastructure at global scale.

Similarly, SRTM data are also subject to limitations. The data were collected by a specially modified radar system that flew on board the Space Shuttle Endeavour mission in February 2000. Using such technology, the mission was able to acquire elevation data at one arcsecond (30 metres at the equator) resolution and coverage for all major contents. Although the global coverage and resolution are impressive, the data are not representative of ground level. This is a result of the fact that C-
4.1. Metric Creation

Band (four to eight gigahertz) radar struggles to penetrate dense vegetation (Wendi et al. 2016; Rodríguez et al. 2006; Jordan et al. 1996) and buildings (Yang et al. 2011). Therefore, within the context of urban environments where BSS are located, SRTM data may identify building elevations as opposed to elevation at the ground level (Gamba et al. 2002). This may have been alleviated through the use of Digital Surface Models (DSM) or Light Detection And Ranging (LiDAR) technology, but access to these data of granular global coverage are very scarce. As a result, the SRTM DEM presents the most dependable, frequently used (employed in routing algorithms such as CycleStreets detailed in Section 7.2) and representative way to indicate the urban topographical attributes.

Evaluating the global human settlement (GHS) population data and worldclim.org precipitation data, their limitations primarily lie in the coarse spatial granularity. In the context of population estimates created by the European commission, the data are derived from census and administrative unit data from each country which are then aggregated to grid cells using the distribution and density of built-up areas as mapped in the GHS Settlement Layer. Therefore, the data are limited to larger spatial aggregations as a result of the variations in source data resolution. Likewise, the precipitation estimates are limited to the availability of data collected from weather stations which are then spatially interpolated. Given the sparse nature of data collection, interpolating data to finer spatial resolutions would lead to greater inaccuracies. Although these data are not as granular as the other datasets used to calculate confounding variables, they proved to be the most appropriate given their global coverage. As mentioned, to minimise issues surrounding their resolution, the buffer distances were increased to create more reliable measures of population and precipitation.

Collectively, these data were all subject to issues in their temporality. Though the BSS data are systematically recorded at regular time intervals, enabling an insight into their dynamics, these confounding variables are either static snapshots or long-term averages of the global environment. As a result, the captured data are subject to change over the lifespan of a BSS. This poses the most threat to measures
of cycling infrastructure, which have seen significant investment in recent years due to the endorsement of active and sustainable modes of transportation with new global awareness and efforts to mitigate climate change (Rissel and Rissel 2009; Woodcock et al. 2007), in addition to more recent investments in the wake of the COVID-19 pandemic (O’Malley 2021; Reid 2020; Buehler and Pucher 2021; Kraus and Koch 2021).

Although cycling infrastructure is subject to greater short-term change, the majority of these measures are not likely to exhibit significant variations across the duration of BSS operations. In addition, these metrics merely aimed to provide an indication of contextual surroundings given the constraints that have been mentioned, as opposed to any granular and temporally varying metrics. Thus, it is argued that the data that were chosen to provide such insights have satisfied this objective and are in valuable in providing additional contextual information for each BSS at a global scale.

4.2 Exploring Metrics

Thus far, this chapter has detailed the processes through which the UCL BSS data collection have been manipulated to create a variety of informative and comparable metrics. These aim to provide novel, dynamic insights into the changes in activity and size of each system, in addition to providing static contextual information surrounding the urban environments in which they reside. As a result of the homogenous programmatic manner in which they have been calculated, it is possible to conduct large, global-scale comparisons of BSS. Employing the metrics in this manner are presented in Chapter 5.

Instead, this section will exemplify the values of these metrics through a detailed exploration of their capabilities in providing unparalleled insights into the longitudinal dynamics of the London Santander Cycles BSS. The system was selected to illustrate the values of these metrics for a number of reasons. Firstly, this scheme was the first for which data were collected and was the impetus for the initiation of efforts to collect data from open BSS API feeds. As a result, it is the BSS
for which data are most abundantly available and provides a near-complete history, with only data from the first month of operation (July 2010) not being recorded. Secondly, London is the context within which much of the analyses that are presented in this thesis are based. Therefore, presenting a descriptive overview of the metrics that have been created for this scheme provides an excellent foundation to build on, enabling comparisons to analyses conducted in Chapter 6 and 7. Finally, London is a system for which journey data are available (as detailed in Section 3.3), meaning that metrics can be presented with great confidence having validated them against operator published data.

Table 4.4: Descriptive statistics of London Santander Cycles metrics over 12 years of operation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycles</td>
<td>8,040</td>
</tr>
<tr>
<td>Docking Stations</td>
<td>600</td>
</tr>
<tr>
<td>Daily Journeys</td>
<td>24,113</td>
</tr>
<tr>
<td>TDB</td>
<td>2.9</td>
</tr>
<tr>
<td>Area</td>
<td>97 km$^2$</td>
</tr>
<tr>
<td>Dock Density</td>
<td>7 / km$^2$</td>
</tr>
<tr>
<td>Daily Trip Density</td>
<td>241 / km$^2$</td>
</tr>
<tr>
<td>Population</td>
<td>1,649,890</td>
</tr>
<tr>
<td>Precipitation</td>
<td>54 mm</td>
</tr>
<tr>
<td>Cycling Infrastructure</td>
<td>551 km</td>
</tr>
<tr>
<td>Topography</td>
<td>1°</td>
</tr>
</tbody>
</table>

To start, Table 4.4 provides an overview of the metrics that have been calculated aggregated for the entire 12-year lifespan of the system thus far. Here we get a general indication of the system’s size, its use and its contextual surroundings. It appears as though the system is relatively efficient, with an average of nearly 3 TDB and 24,000 journeys per day. Though valuable in providing an indication of the system’s use throughout its operational history, aggregating the metrics in this way mask many of the interesting intricacies of the system’s dynamics. As such, Figure 4.5 explores the temporal distribution of journeys in the average week. By computing the mean number of journeys per hour for each day of the week, we can gain a detailed insight into the activity dynamics. Here, we get an indication...
that the system exhibits a clear differentiation in use observed during weekdays and weekends. During the weekday, there is a clear bimodal distribution, with activity peaking at 8am and 5pm to nearly 3,000 journeys during those hours. These activity spikes are synonymous with commuting activity, as has been identified across a number of BSS in O’Brien et al. (2014). On the other hand, weekends exhibit a unimodal distribution, with activity peaking at around 2pm with approximately 1,800 journeys. Here, such activity has been identified to be indicative of leisure and recreational use (O’Brien et al. 2014), wherein users are not impelled to use the system for work purposes, but rather have the freedom to use it at their convenience.

**Figure 4.5:** Average hourly journeys in a typical week in the London Santander Cycles BSS

Similarly, by aggregating the entropy values calculated for the system across weekdays and weekends that is depicted in Figure 4.6, we can observe a discrepancy in the temporal distribution of bicycles between weekdays and weekends. During weekdays the system illustrates a significant reduction in bicycle imbalance throughout the day. This decrease is likely caused by two primary reasons. Firstly, the large short-term influx of bicycle hires during peak commuting hours reduce the number of bicycles that are parked in docking stations and increase the number that are being actively used. Therefore, this temporarily decreases the observed number of bicycles that require redistribution in order to achieve a perfectly distributed system. This is especially pronounced during the morning peak, where we
are able to observe a significant decline at 8am. Secondly, due to the directed nature of commuting activity, it is necessary to have an innate imbalance to cater for such demands. In this case, docking stations that are on the periphery of the system and in close proximity to transit hubs are disproportionately occupied by bicycles in comparison to those within the centre of the system and in close proximity to the city’s CBDs during the mornings. These bicycles are then hired and used to cycle towards the centre of the city during the working day, restoring those imbalanced bicycles, that are then returned to such imbalanced state in the evening as a result of commuter return journeys.

Conversely, weekend entropy values depict a more regular wave form. Starting at the early hours of the morning, bicycles appear to steadily become more balanced until around 10am from which point entropy increases. This increase follows the steady rise in bicycle hires that are observed in Figure 4.5 until it peaks at 3pm where bicycles appear to steadily redistribute throughout the late afternoon and evening back towards the mean entropy of 0.54. Therefore, unlike the sharp surges of activity that are observed during weekdays, weekend entropy follows the gradual unimodal distribution of journeys.

![Average Weekday and Weekend Entropy in the London Santander Cycles BSS](image)

**Figure 4.6:** Average weekday and weekend entropy in the London Santander Cycles BSS

Although the increasing imbalance throughout the late morning and early afternoon can be correlated to the activity, the early morning and late afternoon and
evening changes in entropy are unusual given the lack of cycling activities during these times. Similarly, although there is a natural ebb and flow of commuter activity in weekday activity that can justify the trends observed, there are considerable fluctuations in entropy that are unlikely to be a result of user activity. As a result, it is imperative that operator redistribution efforts are considered when exploring a system’s entropy. Such erroneous changes in entropy are able to provide an indication of such non-user activity and are invaluable in attempting to understand the strategies that are employed to cater for user demands. For example, during weekdays it is apparent that the morning 8am peak exhibits an abrupt decline in imbalance that seemingly rectifies itself towards a more gradual decline in subsequent hours. This is likely highly indicative of operator redistribution efforts that aim to mitigate the impacts of the concentrated commuter activity.

We can verify this by exploring data on the redistribution of bicycles within the BSS that was made available following a freedom of information request by Navarrete (2016). From this request we get a sense of the scale of these interventions and how they sync up with those observed in the entropy distributions in Figure 4.6. In the statement by Transport for London (TfL), they highlight the difficulties in managing commuter peaks due to their ‘extremely tidal nature’, especially within docking stations in close proximity to mainline rail stations. As a result, staff are deployed and positioned at ‘key sites’ to manage such influxes by collecting and placing bicycles where necessary (Navarrete 2016). Exploring the data that are provided by this request, it is possible to identify such concentrations of redistribution activities during commuting peaks, especially in the removal of approximately 230 bicycles from docking stations at 9am, significantly greater than any other time of the day. These bicycles appear to be redistributed and placed in docking stations in the subsequent hours, thus helping to explain the gradual rebalancing of the fleet. Though weekend activity does not present a bimodal distribution, redistribution efforts appear to continue to follow this temporal concentration, albeit much more muted with fewer bicycles being moved around. As such, the gradual reduction of entropy during weekend afternoons can be associated with the redistribution of
bicycles that occur during this time. The heterogeneous nature of journeys across weekends due to their change in primary trip purpose are also likely to account for the volumetric decrease in redistribution efforts, with an average of 1,300 and 2,200 bicycles being placed in docking stations and 1,400 and 2,400 bicycles being removed from docking stations on weekends and weekdays respectively. The exploration of these redistribution data are presented in Appendix A.

Although long-term temporal aggregations of estimated journeys to a typical week in Figure 4.5 and entropy values in Figure 4.6 have proved their values in identifying the contrasting temporal distribution of journeys between weekdays and weekends as well as identifying how these journeys are distributed within the system, they lack insights into how the system has evolved and changed over time. Therefore, by aggregating metrics to smaller temporal resolutions we can gain a more holistic overview of the system’s development. Figure 4.7 presents the changes in system size, both in relation to the number of bicycles and its operational area, and its utilisation, in terms of the TDB at quarterly intervals. Exploring the metrics in this manner present a number of interesting insights that would otherwise be unattainable.

Over this 12-year span, the system appears to have experienced three occasions where the its fleet size increased substantially, initially between the third quarter of 2011 and the second quarter of 2012, then a smaller expansion in the beginning of 2013 and finally throughout 2014, growing by 2,000, 500 and 2,000 bicycles respectively. In each case, the operational area of the BSS also increased, indicating the enlargement of the system to include additional docking stations. In the past five years, the system exhibits little variation in its size, but it becomes very evident that there are seasonal fluctuations in the utilisation of the system, where the first quarter of the year consistently depicts much lower levels of TDB in comparison to the third quarter. This is indicative of the increased uptake of cycling as a mode during warmer months in comparison to the colder months at the start and end of each year, highlighting the importance of the weather on cycling activity. Additionally, we are able to identify the anomalously high TDB during the summer months in 2012,
reaching over five TDB. During this time, London hosted the Summer Olympics Games that resulted in a large influx of tourists. The expansion of the system prior to the event were direct efforts to encompass Stratford to the east of the city, the area that hosted the Games which saw significant investment and regeneration. Thus, it is apparent that large events like this have significant impacts on the utilisation of a BSS and can help to provide a flexible mode of transportation to meet such demands.

Collectively, this section highlights the values of the metrics that can be derived from open BSS API feeds in providing novel and valuable insights into their dynamics. Although this exploration is not exhaustive, the variety of ways in which these data can be simply aggregated and descriptively analysed are an invaluable tool to unpack the way in which systems operate and how this has changed over time. Though we focus on exploring the London Santander Cycles BSS, these metrics are calculated for all BSS for which data were available in the UCL data
4.3 Chapter Summary

This chapter details the various ways in which the UCL BSS data collection can be manipulated to create novel, comparable metrics. These processes aim to highlight the way in which open BSS data, that were originally intended to provide static snapshots of bicycle locations and availability, can be exploited to generate accurate measures of system size and use. Due to the enormity of the data collection, these heuristics are created using scripts that enable their programmatic generation. As
a result, should data on additional BSS or other forms of micromobility (like e-scooters that use GBFS) become available, these metric creation procedures can be quickly adapted and employed to produce valuable insights.

Among the metrics that have been created, there are two primary innovations that are important to acknowledge. Firstly, in the journey estimation process, it was possible to identify the occurrence of significant errors within dock capacity data that, unless removed, would lead to severe overestimations. This highlights the importance of the validation that was conducted in ensuring the accurate estimation of journeys within BSS, providing greater confidence in the measure. Secondly, the processes developed in Section 4.1.7 enable the estimation of wider contextual information in each BSS. Leveraging datasets of global scale, the methodology provides estimations of confounding variables that were found to consistently have significant associations with use. In both contexts, these mark novel developments in the processing of these data that have not been explored previously, resulting in more accurate journey estimations and further contextual insights into BSS in a scalable manner.

Section 4.2 provides an overview of how these data can be employed through simple temporal aggregations to generate insights into individual systems. For example, Figure 4.7 depicts the long-term evolution of the London Santander Cycles BSS, enabling the identification system expansion, seasonal variations and anomalous events such as the 2012 London Olympics. Although valuable in this sense, they also provide the foundations for much more exhaustive and sophisticated analyses, such as the most comprehensive global comparison of BSS in Chapter 5, in addition to their values in more granular analyses such as those detailed in Chapter 6 and 7.
A Global Comparison of Bicycle Sharing Systems

BSS have seen widespread global adoption over the last 15 years that can be attributed to the numerous benefits that are associated with their implementation (as highlighted in Section 2.1.2.1). Although there is a growing body of literature that aims to quantify such impacts, these are typically limited to studies that analyse individual systems or a small handful of regional BSS due to limitations surrounding the availability of data. More specifically, the largest comparison of BSS investigate the relationship between various built and natural environment factors across 75 systems (Médard de Chardon et al. 2017), but are limited in the variety of BSS that they study and regions in which they are located. As a result, there are currently no studies that conduct analyses on BSS at a truly global-scale.

Given their growing global presence and diversity, it is imperative that we gain a more holistic understanding. The UCL data collection and the metrics that have been derived from them present a valuable opportunity to leverage the depth and breadth of the data in facilitating such analyses. Therefore, this chapter presents the most comprehensive analyses of BSS to date, providing novel insights into slow, long-term system growth and evolution in Section 5.1 and creating a definitive global classification of BSS based on their size, use and contextual attributes in Section 5.4.

In both cases, the analyses that are presented in this chapter break new ground
in our understanding of the global BSS landscape, providing new perspectives on their current deployment. They highlight the values of passively collected data in providing such insights and also present a framework for employing these data in creating such understandings. Here, a novel two-staged clustering methodology has been developed that leverages the full potential of the calculated metrics in conducting analyses at such scale. These findings are not only beneficial in advancing academic awareness of BSS, but also in facilitating a tool for operators to better plan current or future systems in the image of those that exhibit great efficiency in the way that they operate.

5.1 Slow Dynamics at a Global Scale

Extending on the descriptive analysis presented at the individual system-scale in Section 4.2, the comparable nature of the calculated metrics can be exploited to provide some interesting descriptive insights into systems across long temporal durations. These analyses further highlight the value of such metrics in enabling the analysis of systems at a truly global-scale, an area of BSS that have analysed in great detail.

In a similar manner to the evolution of the London BSS presented in Figure 4.7, the calculated metrics for each BSS can be aggregated to the quarterly level and presented visually. Here, Figure 5.1 presents the evolution of 176 BSS in the period between quarter three (1st July to 30th September) of 2016 and quarter three of 2018. A version of this plot has also been published in ‘Atlas of the Invisible’ by Cheshire and Uberti (2021), that has been included in Appendix B as reference to a professionally edited version. The variables that are displayed also mimic those presented in Figure 4.7, the only differentiations being that colours have been used to depict the continent in which the system operates and size of points were dependent on estimated populations instead of operational area. Here, the presentation of BSS are limited to 176 systems due to the availability of data for systems over the two-year study period.

Taking a closer look at Figure 5.1, it is possible to tease out some interesting
changes that have occurred across BSS globally. In comparison to Figure 4.7 where it is possible to unpack the evolution of a BSS individually, presenting the aggregated metrics across BSS enable the identification of system comparisons, both in terms of their size and use over time. Although analyses relies on descriptive, visual interpretation, they provide a novel way of assessing individual and global trends of BSS.

For example, here it is possible to situate the London BSS in a global context. Although individually it was possible to identify its gradual expansion and change in utility with seasons, here, it appears that though the system has experienced comparatively less significant changes across the two-year study period. We can see other systems that have exhibited much more dramatic changes. Within the Parisian BSS we observe a large decline in both system size and use, with a reduction in fleet size from 20,000 to 8,000 and TDB falling from approximately 5 to 4.5. We observe that these trends are anomalous among BSS globally, the majority of systems that observe increases in both aspects.

Given the abnormality of such trends in the Parisian system prompts further investigations. In this case, research reveals that such changes were indeed anomalous and were the result of significant structural changes within the system. During this period, the Parisian BSS transferred operations from JCDeacaux to Smovengo (Chrisafis 2018). In this transition, the new operator had promised to update and expand the system to include more docking stations as well as update a proportion of the fleet to include e-bicycles. The implementation of this change was met with numerous difficulties and delays, causing an out-roar of public dissatisfaction and anger (Chrisafis 2018).

On the other hand, as mentioned, the majority of systems depict a trend of increasing size and use. More specifically, systems located in cities such as Helsinki, Rio de Janerio, San Francisco and Suzhou exhibit significant improvements. This global trend provides an indication of the increasing popularity and adoption of BSS, with system expansions that cater for previously unmet demand.
Figure 5.1: The evolution of 176 BSS between 2016 and 2018
Although visualising the metrics in this way provide a simplistic and visual depiction of individual BSS in a wider global context that can help to identify global trends and anomalous shifts in individual systems, they are limited to descriptive interpretations. In an effort to provide a more succinct and easily digestible interpretation of the global BSS landscape, the following analyses aim to create the most comprehensive global classification of BSS. This leverages the calculated metrics by employing clustering algorithms that are able to distinguish key attributes into generalised ‘types’ of system that are currently in operation.

5.2 Studying Bicycle Sharing Systems at Scale

As detailed in Chapter 2, BSS literature have increased in conjunction with their adoption. There are numerous studies which conduct holistic overview of literature in the field such as Fishman (2016), Fishman (2019) and Si et al. (2019). These reviews consistently identify academic studies that focus on several aspects of BSS including their benefits, operation and how it can be optimised in addition to the various parameters that are associated with their use. Though there are numerous studies on these various aspects of BSS, their scope are limited to studies on individual systems or regional scale analyses. As a result, there are a very limited number of academic studies which investigate BSS at a global-scale.

The scope for substantive international comparisons across BSS are constrained by the inconsistency of BSS data dissemination practices and formats as well as their limited availability. As a result, when making comparisons between BSS the TDB metric is commonly used as it is straightforward to calculate and enables some understanding of their utility. For example, it is the basis to Médard de Chardon et al. (2017) comparison of 75 systems from around the world in the determination of common variables that influence the use of BSS. Similar to Table 4.3, the authors identify five categories of variable: BSS attributes, density and compactness, geography, weather and transportation infrastructure. Among these variables it becomes apparent that the variables which have previously been identified to have significant impacts on BSS use were found to have similarly consistent
impacts across the systems included in this analysis. For example, helmet require-
ments severely penalise system performance, as was found in Jain et al. (2018),
warmer temperatures increase performance but returns are marginal between 18 and
$33^\circ C$, as identified in Miranda-Moreno and Nosal (2011) and cycling infrastructure
density are also identified to increase use as has been identified in Buck and Buehler
(2012).

Similarly, Bieliński et al. (2019) conduct analyses on 56 BSS located across
Poland, employing regression methods to investigate the determinants of TDB. Though
the study does extend on the analyses conducted by Médard de Chardon et
al. (2017) by supporting findings with survey data, again, the finding make limited
progress. In both cases, these studies are valuable in establishing strong regional
and international stability across attributes associated with BSS use but fall short in
their lack of global representation. More specifically, these studies lack representa-
tion of Asia, and given that the Chinese BSS market is the largest in the world (Gu
et al. 2019), it is imperative that studies that wish to create a more comprehensive
understanding of BSS need to include such systems.

Other studies which aim to analyse multiple BSS include Zaltz Austwick et al.
(2013), Zhang et al. (2015), Sarkar et al. (2015) and Kou and Cai (2019). Zaltz Aust-
wick et al. (2013) are the earliest attempt to conduct analyses comparing five sys-
tems, employing a collection of descriptive, spatial and network analysis techniques
to compare the utilisation among these systems. Zhang et al. (2015) are unique in
the scope of their analysis, investigating commonalities found across five Chinese
BSS. Within this paper, the authors rely on primarily descriptive analyses of system
attributes to determine that Chinese BSS play a multi-faceted role, encompassing
use by commuters, urban dwellers and tourists but do not provide any quantitative
analyses to determine commonalities or comparisons between such systems. Sarkar
et al. (2015) analyse data from 10 BSS employing unsupervised clustering, random
forests and neural network methods to compare the cycling activity patterns within
these systems. Although the analysis is valuable in identifying different types of
docking station and the strategies that should be implemented to optimise the redis-
tribution of bicycles at these types of docking station, the analyses are limited in the accuracy of their data, being sourced from operator maps. Kou and Cai (2019) analyse data from eight BSS, but are limited in their scope, focusing on the comparison of trip duration and distance.

Although these studies have merits and are unique in their contributions to the literature surrounding BSS, it is clear that there are limitations to each which inhibit their contributions at the macro-scale. Thus, there is a clear need for a large global comparison of BSS and one that enables some holistic indication of the BSS landscape, including the similarities and differences in not only their use, but also additional contextual information which have consistently been identified to have significant associations with use.

The research in this chapter therefore marks a major step forward given its foundations on the UCL BSS data collection and the metrics that were calculated in Chapter 4. These metrics are first organised and cleaned (Section 5.3) in order to be processed through a novel two-staged clustering methodology that is detailed in Section 5.4. In the first-stage, a set of static variables are clustered for each BSS, using k-means, in order to provide an overview of the different types of docked BSS that are currently in operation around the world. Then, within each of these initial first-stage clusters, a dynamic time warping (DTW) clustering operation is conducted in order to further distinguish BSS by their typical temporal dynamics. The results of these clustering processes are then outlined in Section 5.5, providing a detailed insight into the unique characteristics and attributes of BSS classifications, that are then discussed and evaluated holistically in Sections 5.6 and 5.6.1.

5.3 Employing Metrics

As mentioned, the analyses in this chapter exploit those metrics created in Chapter 4. They provide excellent foundations to conduct the most comprehensive global classification of BSS as a result of their exhaustive extent in addition to the homogeneity in the way that they have been calculated, ensuring their comparability. Since the metrics are so exhaustive in their raw state, it was necessary to create data
structures that would enable them to be best equipped to serve such purposes, therefore, metrics have been further manipulated and arranged into two formal datasets.

First, it was necessary to define a study period that would enable the analyses of as many BSS as possible. After exploring the UCL data collection, the six-month period between April and September 2018 was identified to be the most contemporary period within which the highest number of BSS could be analysed. This period contained complete data on 322 systems that have been used as the foundations to create this global classification, the global distribution and size of which are depicted in Figure 5.2. The selection of this study period was also valuable since it was not subject to the varying geographical and temporal impacts of COVID-19, ensuring that ‘normal’, pre-pandemic dynamics are represented.

Once the study period was defined, it was then necessary to structure those metrics to enable the classification of BSS in two stages. First, various BSS metrics are aggregated to the six-month study period to create a multivariate data structure that is representative of each system’s size, use and contextual surroundings that provide the foundations for the first-stage of the classification process. Here, four primary BSS metrics have been adopted, those being the estimated number of daily journeys, the maximum number of bicycles, TDB and operational area. Aside from operational area, each of these attributes are separated by weekday and weekend due to the large variations that have been observed across studies of BSS (Faghih-Imani et al. 2014; Faghih-Imani et al. 2017; O’Brien et al. 2014). Aggregating the data in this way enables the differentiation of each system’s average size and use throughout a typical weekday and weekend within the study period.
Figure 5.2: The size and location of the 322 BSS included in the determination of the global BSS landscape.
In combination with those BSS attributes that are aggregated across the six-month study period by weekday and weekend, the confounding variables that are calculated in Section 4.1.7 are also adopted. These are included to provide invaluable insights into the contextual settings of each BSS regarding their population, precipitation, topography and cycling infrastructure. As mentioned, these attributes were initially calculated not only to represent BSS environments, but also due to their merits as variables that have been consistently identified to have significant impacts on system demand and use. These variables are only calculated once for each BSS as a result of their temporal invariability in addition to the limited data availability. Thus, these contextual features are simply appended to those aggregated BSS attributes mentioned above. This produces a multidimensional data structure that are representative of BSS across the six-month study period. This is used to provide the foundations for the first-stage of the classification process in the identification of generalised BSS types.

In the second-stage of clustering, journey estimations are further manipulated to construct a set of time-series that are representative of the temporal distribution of activity within each system. This was valuable for inferring how each system is adopted by users and to extend on the first-stage clusters to further differentiate them in such dimensions. Much like those BSS attributes in the first-stage, the hourly number of estimated journeys in each system are aggregated to the six-month study period and separated between weekdays and weekends. This creates two 24 hour time-series sets for each BSS.

Similar techniques have been employed in studies such as O’Brien et al. (2014), who compare temporal distributions of use across 30 BSS by aggregating the load factor for a typical weekday and weekend across a six-month study period. Within this chapter, the second-stage of clustering aims to extend on these techniques by exploring the similarities in temporal distribution across over 10 times as many systems, employing clustering algorithms to group systems with similar distributions together as opposed to comparing systems individually descriptively.

Collectively, these two data structures that have been derived from the metrics
that were calculated in Section 4.1 exploit the full breadth of BSS data that can be obtained from open API feeds. From the first-stage data structure that enables time-static attributes of BSS including their size, use and contextual setting to be included to provide generalised overviews of global BSS types, to second-stage data structures that provide more detailed insights into the temporal dynamics of such systems, these data exhibit the multiple dimensions of BSS that can be derived and used to create the most comprehensive global classification of systems.

### 5.4 Clustering Bicycle Sharing Systems

Unlike computers, humans find it very difficult to digest large amounts of information quickly. Therefore, we often aim to group information into meaningful structures or collections so that they are more manageable. For example, within human geography, geodemographics is a field that seeks to derive information about people by where they live (Harris et al. 2005; Longley 2005). Given the unique characteristics of individuals and places, it would be an impossible to comprehend where and how different types of people congregate by exploring the data. As a result, classifications are often created to identify areas that are composed of similar demographic attributes. These are typically created using clustering algorithms that are forms of unsupervised machine learning which take large multivariate data and partitions them into meaningful groups. Examples include the Output Area Classification, ACORN and MOSAIC, each of which have been applied across numerous academic disciplines as well as the private sector, in setting like retail planning and marketing (Harris et al. 2005).

Clustering algorithms not only serve to determine groups among unlabelled data, but also as a means of generalisation. For example, when observations within a cluster are missing features, you can infer the missing data from other examples in that cluster. In addition, clustering methods are a way to compress data, using cluster assignments as a replacement for the features, in a similar vein to Principal Component Analysis (PCA), as well as a method of preserving privacy of individuals in clusters for those data that are personally disclosive, like geodemographics.
As a result, classifying large datasets into meaningful clusters have a wide variety of benefits, not only for the researchers who are trying to understand underlying groupings, but also to the private sector and the applicability of data more generally.

In the context of BSS, there are yet to be a global classification of systems. Given their rapid growth and adoption, creating a more digestible global understanding of systems is necessary to provide a holistic overview of the different types of systems that exist and the ways in which they are used. Here, data on BSS are acquired from the UCL collection, employing the variety of metrics that were calculated in Chapter 4 to create a selection of attributes that are representative of their size, use and contextual settings. This multivariate data structure is very similar to those that are used in the creation of geodemographic classifications, being composed of a large variety of variables that are difficult to comprehend by exploring them alone. Therefore, employing clustering algorithms that enable the creation of such classifications are necessary.

Within the field of clustering algorithms there are a large number of possible ways in which groups can be assigned. As a result, there is yet to be a complete definition for clustering that have been universally agreed upon (Xu and Tian 2015). Though this is the case, the classical and general definition for clustering follow three rules (Jain and Dubes 1988):

1. Individuals or objects in the same cluster must be similar as much as possible
2. Individuals or objects in different clusters must be different as much as possible
3. The measurement for (dis)similarity between clusters must be defined and have a practical meaning

A consequence of the numerous approaches to clustering is the introduction of a certain degree of subjectivity that persists regardless of chosen methodology. This is a common problem that have been identified among various clustering domains (Cape et al. 2000; Lo Siou et al. 2011; Maharaj 2000; Milligan and Cooper 1987). While there have been attempts to eliminate concerns, the decisions would reemerge in some form even in the newly proposed solutions. For example, Maharaj
(2000) presented a method to programmatically suggest the number of clusters in the hierarchical clustering, however, one still need to subjectively decide upon the significance levels. Therefore, in the majority of clustering approaches, including the classification of BSS, subjectivity arises on multiple levels throughout the clustering process, that generally have four stages, as identified by Xu and Wunsch (2005):

1. Feature selection and extraction: the variables that should be included in the clustering process
2. Clustering algorithm selection: choose the way in which the clusters are assigned, including the measurement for (dis)similarity that are employed in those algorithms
3. Result evaluation: determine whether the results of such clustering processes are valid and are able to create meaningful groups
4. Result explanation: provide practical reasons for the results observed in those clusters that created

Each of these steps can take many different options, presenting an endless combination of possibilities that require the analyst to make subjective decisions. Therefore, clear justifications are necessary in order to clarify them that provide solid rationale for such processes being undertaken. Much akin to the geodemographic classifications of the past, this analyses create a tiered classification of BSS. The first tier (or stage) typically contain a smaller number of groups that provide a good general indication of the different classification types. The second tier is then used to further differentiate those major groups into smaller niches that better capture their unique characteristics. Aside from geodemographics, similar tiered approaches to classification have been employed across academic studies. For example, Watson and Callingham (2003) aimed to classify the statistical literacy among 3,000 school students, limiting the initial number of clusters to six, that are then further differentiated in the second tier to include 20 clusters.

In the context of this analysis, a novel two-tiered clustering approach has been presented for a number of reasons. Firstly, much like classifications of the past,
the two-stages aim to make the identification of groups and structures among BSS easier. Secondly, the two tiers involve the clustering of different types of data, for which there are currently no established methods. Each of these data types require the clustering algorithms to handle them appropriately, therefore this approach is not only valuable to create a tiered classification structure, but also in preserving these data’s structure.

In the first-stage of clustering that are detailed in Section 5.4.1, the general types of BSS are identified. In order to do so, attributes of BSS size, use and contextual characteristics have been gathered into a large multivariate dataset by aggregating such values to the six-month study period. These variables are similar to those population attributes used in geodemographic classifications and therefore similar multivariate clustering methods can be employed. In the second-stage of clustering that is detailed in Section 5.4.2, these general types of BSS are further distinguished by the temporal distribution of journeys across typical weekday and weekend activity observed within each system. In this case, the data are of a time-series format, meaning that the ordering of such variables are essential to their structure. Therefore, it is necessary to employ clustering methods that retain these data structures to ensure that cluster results are meaningful.

5.4.1 First-Stage Clustering: What types of BSS are there?

As detailed, the aim of the first-stage of clustering was to be able to identify the general types of BSS that were in operation in regard to their size, environment and general use. The variables that are employed in these analyses (Section 5.3) are a selection of the metrics that best represent these characteristics of BSS, aggregated to the same period for comparability purposes. Each value represents the average over the six-month study period. Although comparable, it is important to note the vast scale of these data, encompassing BSS with fleet sizes ranging from over 25,000 to fewer than 10. In addition, each of these metrics are measured in different units. Therefore, prior to the implementation of clustering algorithms, it is imperative that these features are of similar magnitude and scale so that (dis)similarity measures are not skewed by the measurement and variation in those variables.
The two most common methods of feature scaling are normalisation and standardisation. Normalisation is the process of scaling that typically constrains the feature to a range between zero and one, also known as ‘min-max scaling’. Here, each feature is scaled using the formula:

\[ N_i = \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \]  

Where:

- \( N_i \) = the normalised value
- \( x_{\text{max}} \) = the maximum values taken by \( x \) in the data
- \( x_{\text{min}} \) = the minimum values taken by \( x \) in the data

Alternatively, standardisation is the practice of feature scaling that centres values around the mean with a unit standard deviation. This means that if we calculate the mean and standard deviation of standardised attributes, their values will be zero and one respectively. The formula for such feature scaling is as follows:

\[ Z_i = \frac{x_i - \mu}{\sigma} \]  

Where:

- \( Z_i \) = the standardised value
- \( \mu \) = the mean of the feature
- \( \sigma \) = the standard deviation of the feature

The choice of feature scaling largely depends on the features and their nature. In this context, due to the large variability in the size of systems and the associated metrics that are calculated, it was necessary to use standardised feature scaling (Equation 5.2). Normalisation is a process that is very sensitive to outliers and although the features are not all normally distributed, standardisation was vital to capture the variability amongst BSS (Almaliki 2020). Standardisation prevents variables with large scales from dominating how clusters are defined, better enabling the determination of BSS types by the composition of all features included within the cluster analysis (Atkinson and Riani 2007).
Following the selection of features and their manipulation in step one, it was then necessary to determine the most appropriate clustering algorithm which should be employed to determine the types of BSS. Jain et al. (1999) categorises clustering algorithms into four broad categories based on the method used to define clusters, namely; partitional, hierarchical, density-based and grid-based algorithms. The latter two methods are most commonly used among spatial data, such as density-based clustering of applications with noise (DBSCAN) (Ester et al. 1996) and KDE (Worton 1989), that are explored in Chapter 7. These categories can be further divided into crisp and fuzzy methods, where crisp ones are adopted to assign each individual or object into a single cluster, whilst multiple clusters can be assigned employing fuzzy methods (Jain et al. 1999). Of these numerous methods, clustering algorithms that employ partitioning methods, such as Partitioning Around Medoids (PAM) and k-means (MacQueen 1967) are considered as these processes are favourable for multivariate data.

The k-means clustering method is one of the most common algorithms employed across various academic studies that aim to group multivariate data (Sinaga and Yang 2020), such as geodemographics (Lansley et al. 2015). As mentioned, there are great similarities in the classification of demographic attributes and BSS attributes, where those features are indicative of observations in different geographic locations. k-means is an iterative, top-down relocation algorithm process that requires the analyst to define the number of clusters prior to the algorithm being run (Harris et al. 2005). The process aims to minimise the sum of squared errors (SSE), that is the distance between each observation and its cluster’s mean. This can be expressed as:

$$SSE = \sum_{j=1}^{k} \sum_{i=1}^{n} |x_{i}^{(j)} - c_{j}|^2$$

Where:

- $x_{i}^{(j)}$ = each observation and
- $c_{j}$ = their respective cluster centre

The algorithm starts by placing k ‘seeds’ within the multidimensional space of
the input data. Each observation is then assigned to the closest seed, creating an initial cluster assignment. Seeds locations are then recalculated and moved to the centre of those assigned observations. Then, observations are reassigned if those new cluster centres are closer than the one they are currently assigned to. This process is repeated until the seed locations no longer move, indicating that the optimum allocation has been reached (Harris et al. 2005). Although this method is very fast, simple and employed across a variety of scenarios throughout academia, it is important to acknowledge its limitations.

Firstly, due to the nature of the algorithm, the clusters that are produced are typically spherical (Jain 2010). This is known to cause issues where clusters do not have isotropic properties and overlap or are densely concentrated. Secondly, as a result of the initial random placement of seeds, the assignment of clusters in the k-means process are stochastic in nature. Finally, the analyst has to predefine the number of clusters that are to be found, meaning that results are subject to the analyst’s decision.

In order to combat these limitations, there are a number of common practices. For example, to eliminate issues surrounding inconsistent cluster assignments, k-means algorithms are commonly implemented numerous times in order to achieve certainty in the assignments. In the context of this analysis, the k-means algorithm goes through 10,000 iterations, that has been suggested to be overly sufficient in ensuring confidence in the results (Singleton and Longley 2009).

The determination of the number of clusters is a much more difficult task due to its highly subjective nature that are unique to each data’s context (Singleton and Longley 2009; Vickers et al. 2005). The analyst has to make a trade-off between a larger number of clusters that would produce tighter, more uniform groups (Singleton and Longley 2009), and a smaller number of clusters that would contain more variation. It has commonly been recognised that classifications with too many groups cease to become useful as this makes the results harder to interpret due to their indistinguishability (Lansley et al. 2015). The balance between cluster homogeneity, complexity and their representativeness of the data being studied are
difficult, but there are a number of tools that have been developed in order to such processes easier and informed based on quantitative measures. For example, the elbow method is the most common way in which the number of clusters are determined. Here, analysts examine a graph that compares the number of clusters on the x-axis to the within cluster sum of squares (WSS) on the y-axis (Kodinariya and Makwana 2013). The ‘elbow’ is typically denoted by the point on the graph which represents a step change in the WSS between cluster allocations and implies an additional cluster does not help as much to differentiate groups. Other methods of determining the number of clusters include the average silhouette method (Rousseeuw 1987), the gap statistic (Tibshirani et al. 2001) and the Calinski-Harabasz Criterion (Caliński and Harabasz 1974), each aiming to guide the analyst in this decision-making process.

Although these methods were employed to help determine the number of clusters, two clusters were consistently suggested as being the optimum, as can be observed in Figure 5.3. Using this suggested number of clusters would only enable the differentiation of 18 of the largest BSS among the 322 that were being analysed. As a result, further exploratory analysis comparing the different number of cluster assignments was conducted. The size of each of the clusters in each assignment are
detailed in Table 5.1. In spite of such strong evidence for two clusters, the decision was taken to cluster the 322 BSS into five clusters as it proved to be an ideal middle ground where each group had clearly defined differentiating features. In comparison, fewer clusters created large amorphous groups whilst increasing the number of clusters appeared to further differentiate already small clusters, depreciating from the generalisability of the clustering process.

Table 5.1: Exploratory analysis of k-means clusters sizes

<table>
<thead>
<tr>
<th>Number of Clusters</th>
<th>Cluster Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>18,304</td>
</tr>
<tr>
<td>3</td>
<td>6,58,258</td>
</tr>
<tr>
<td>4</td>
<td>6,16,64,236</td>
</tr>
<tr>
<td>5</td>
<td>3,5,15,66,233</td>
</tr>
<tr>
<td>6</td>
<td>2,2,5,14,66,233</td>
</tr>
<tr>
<td>7</td>
<td>2,2,5,10,12,61,230</td>
</tr>
<tr>
<td>8</td>
<td>2,2,5,9,11,34,57,202</td>
</tr>
<tr>
<td>9</td>
<td>2,2,5,8,11,20,33,69,172</td>
</tr>
</tbody>
</table>

5.4.2 Second-Stage Clustering: How are these BSS used in a typical day?

Following the first-stage of clustering analysis, second-stage methods were employed to further differentiate BSS types by investigating the distribution of journeys within a typical weekday and weekend. This can help to provide inferences on the way in which systems are used.

As identified in the first-stage of clustering, due to the large variability in the size of BSS included within this analysis, the absolute number of journeys across a typical day had considerable variation. Therefore, it was necessary to scale the time-series’ to ensure their comparability when implementing clustering algorithms. Standardisation (Equation 5.2) of these features were elected over normalisation methods as standardising the time-series’ would enable the feature to capture the relative size of peaks among BSS of varying sizes.

The ordered nature of time-series data are imperative to their understanding. Therefore, clustering algorithms must ensure that this structural component of the
features are preserved. Among methods that cluster time-series features, there are three ways in which groups can be assigned, namely through their similarity in time, shape or change (Bagnall et al. 2006). Although change based time-series analyses are presented in Chapter 6, here, the objective is to determine the typical daily use. As a result, shape-based methods of clustering are most appropriate as they provide a methodology to determine similarity among the temporal variations in BSS journeys.

While comparisons of time-series data using simple Euclidean distances of feature observations at the same point in time are possible, they are very sensitive to distortions in the time-axis (Ratanamahatana and Keogh 2004). Thus, alternative methods that consider shifts in temporality provide a greater opportunity to consider the shape of time-series. The most popular method that incorporates such relaxation in constraints of clustering is dynamic time warping (DTW) (Berndt and Clifford 1994). DTW enables the identification of local similarity in a non-linear manner (Ratanamahatana and Keogh 2005) by minimising the difference within a specified window of each point in time (Ratanamahatana and Keogh 2005). Although Bagnall et al. (2017) identify numerous advances in time-series classification methods that develop on the foundation of DTW, such as Collection of Transformation Ensembles (COTE), these methods are highly computationally intensive. Given the short nature of each time-series’, with only 24 points for each series, as well as their standardised feature scaling, DTW is a quick and effective method to determine tight clusters of BSS with similar temporal dynamics.

Within the DTW methodology, the determination of clusters are highly subject to the size of the window that the algorithm uses to find the shortest distance. Ratanamahatana and Keogh (2005) identify that smaller windows increase accuracy by up to 10%, showing the importance of such warping constraints. Since our series only consist of 24 points, warping was constrained to observe only one adjacent value. Here, the comparison of zero adjacent values would make the DTW algorithm redundant, measuring Euclidean distances, whilst two would be too large given that variability in activity peaks are not subject to significant changes. For
example, morning and evening commuting peaks only vary by an hour or so worldwide, as a result of variations in working culture and the distances individuals have to travel to work (Kung et al. 2013).

Once distance measures were calculated using the DTW algorithm, it was necessary to choose an appropriate method of identifying clusters. Although the k-means algorithm was selected for the first-stage clustering, here, a hierarchical clustering method using the Ward’s method is deployed (Mojena 1977). DTW being adopted for its shape-based methodology lends itself to using a hierarchical method as the number of clusters did not have to be predefined and could be visually assessed at each segmentation of the dendrogram. The Ward’s measure was chosen after testing three alternative measures; namely the single, complete and average methods that failed to produce clusters that were as compact as the Ward’s methods whilst minimising the WSS. These alternative methods were found to have lopsided clusters which were not useful when clustering based on the shape of the time series data (Meesrikamolkul et al. 2012). The number of clusters were decided using the visual aid of a dendrogram to determine the minimum number of clusters needed to separate the major differences between the travel dynamics within each of the first-stage cluster assignments.

### 5.5 The Global Bicycle Sharing System Classification

The centres of each of the first-stage clusters are detailed in Table 5.2. These cluster features have been used to create informative and meaningful names based on their characteristics and unique differentiating factors. They are as follows: ‘very large, high use BSS’, ‘large BSS in major cities’, ‘medium BSS with extensive cycling infrastructure’, ‘small to medium efficient BSS’, and ‘small to medium inefficient BSS’.

Each of these first-stage clusters and their associated second-stage clustering results are described in full detail between Section 5.5.2 and 5.5.6. Appendix C detail the cluster assignments for each of the 322 BSS contained within this analysis that can be explored to identify their intricacies.
Table 5.2: First-stage k-means cluster average characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cluster</th>
<th>Very large, high use BSS</th>
<th>Large BSS in major cities</th>
<th>Medium BSS with extensive cycling infrastructure</th>
<th>Small to medium efficient BSS</th>
<th>Small to medium inefficient BSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of BSS</td>
<td>3</td>
<td>15</td>
<td>5</td>
<td>66</td>
<td>233</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>4,299,517</td>
<td>1,786,311</td>
<td>448,719</td>
<td>350,354</td>
<td>111,302</td>
<td></td>
</tr>
<tr>
<td>Area (km²)</td>
<td>441.3</td>
<td>167.5</td>
<td>73.6</td>
<td>36.1</td>
<td>16.7</td>
<td></td>
</tr>
<tr>
<td>Slope (°)</td>
<td>3.2</td>
<td>2.6</td>
<td>1.4</td>
<td>2.2</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>Precipitation (mm)</td>
<td>176.8</td>
<td>142.3</td>
<td>66.4</td>
<td>82.3</td>
<td>77.0</td>
<td></td>
</tr>
<tr>
<td>Road Length (km)</td>
<td>5,780.2</td>
<td>4,741.4</td>
<td>14,497.3</td>
<td>1,294.1</td>
<td>748.1</td>
<td></td>
</tr>
<tr>
<td>Cycle Length (km)</td>
<td>211.0</td>
<td>394.4</td>
<td>1,533.5</td>
<td>181.5</td>
<td>80.4</td>
<td></td>
</tr>
<tr>
<td>Weekday Journeys</td>
<td>68,659.5</td>
<td>24,902.8</td>
<td>2,708.1</td>
<td>4,447.5</td>
<td>359.2</td>
<td></td>
</tr>
<tr>
<td>Weekday Max Bikes</td>
<td>18,681.1</td>
<td>6,001.5</td>
<td>1,124.9</td>
<td>967.4</td>
<td>279.4</td>
<td></td>
</tr>
<tr>
<td>Weekday TDB</td>
<td>4.6</td>
<td>4.7</td>
<td>1.5</td>
<td>4.9</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Weekend Journeys</td>
<td>65,097.1</td>
<td>21,728.7</td>
<td>2,716.2</td>
<td>3,384.3</td>
<td>305.0</td>
<td></td>
</tr>
<tr>
<td>Weekend Max Bikes</td>
<td>18,839.5</td>
<td>6,015.0</td>
<td>1,121.7</td>
<td>968.7</td>
<td>277.5</td>
<td></td>
</tr>
<tr>
<td>Weekend TDB</td>
<td>4.2</td>
<td>3.9</td>
<td>1.4</td>
<td>3.8</td>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>

5.5.1 First-Stage Cluster Sensitivity Analysis

In an effort to ensure that these first-stage cluster assignments were accurate representations of global BSS types, in addition to running the k-means algorithm 10,000 times, further sensitivity analysis were conducted. This was necessary given the method’s limitation in determining spherical clusters in multivariate feature spaces, with some systems that may be a fuzzy mix between two cluster types. As a result, the subjective decisions that were made in the calculation of confounding variables for BSS in Section 4.1.7 could have influenced the identification of alternative BSS types or cluster assignments.

To alleviate any concerns regarding the subjectivity of these variables, the contextual BSS measures were recalculated using differing buffer sizes ranging between 100 and 500 metres. Employing these measures in place of their chosen counterparts and observing consistent segmentation of BSS would therefore instil greater confidence in such assignments and remove any concerns pertaining to the choices that were made.

In this sensitivity analysis, confounding variables calculated with 500, 400 and 300 metre buffers exhibited no changes in cluster assignments. At 200 metres, the Sarajevo BSS, in Bosnia and Herzegovina, was the only system to be reallocated.
from the ‘small to medium efficient BSS’ cluster to the ‘small to medium inefficient BSS’ cluster. As a result, with only one change in cluster assignment across a variety of different buffer distances, we can have great confidence in the identification of these five major types of BSS, alleviating concerns over their stability and providing further justifications to the decisions that were made in the estimation of such attributes for BSS in Section 4.1.7.

5.5.2 Very Large, High Use BSS

The first-stage cluster ‘very large, high use BSS’ is the smallest cluster, containing only three systems that are all located in East Asia; Taipei, Suzhou and Weifang. As depicted in Table 5.2, these systems are able to differentiate themselves from the other BSS in the analysis due to their size and activity. They are the largest, both in terms of their operational area and population, as well as exhibiting the most activity, with over 42,000 additional daily journeys across both weekdays and weekends in comparison to the second most active group of BSS. This strong correlation between these variables is consistent with those identified across the literature (Médard de Chardon et al. 2017; Faghih-Imani et al. 2014; Tran et al. 2015). Although this is the case, these BSS also appear to exhibit the most precipitation and steepest average gradient. This is unusual since the literature suggests a significant negative relationship between activity and such variables (Corcoran et al. 2014;
Figure 5.5: Average weekend use profiles for ‘very large, high use BSS’

Frade and Ribeiro (2014; Mateo-Babiano et al. 2016). Even in such conditions, it is believed that the population in this geographic region of the world are acclimatised and are less sensitive in comparison to other cyclists. In addition, individuals in East Asia are believed to have built a strong cycling culture due to the inaccessibility to other modes of urban transportation.

Aside from these systems’ dominance in terms of their size and use, these BSS appear to exhibit the least differentiation between weekday and weekend efficiency, with 4.6 and 4.2 TDB respectively. This is indicative of the similarity of the way in which they are used throughout the week, which can be further investigated through the results of the second-stage clustering. As observed in Figure 5.4 and 5.5, it becomes apparent that the bimodal commuter use persists throughout the weekend
5.5. The Global Bicycle Sharing System Classification

for both Suzhou and Weifang. These Chinese systems therefore present a clear
differential from the normal weekend leisure dominated use that have been observed
across many western cities in addition to the Tapei BSS in this cluster. In a study
of BSS users in Suzhou, the authors identify that the system is primarily used by
men with higher incomes who have formal jobs and a college education (Karki
and Tao 2016). These results are based on an in person survey of only 200 users,
making those results highly subject to bias and thus do not aid in the justification of
such weekend commuter activity as such demographics would be unlikely to work
throughout the weekend.

5.5.3 Large BSS in Major Cities

![Figure 5.6: Average weekday use profiles for ‘large BSS in major cities’](image-url)
The first-stage cluster, ‘large BSS in major cities’, contains 15 BSS from major cities around the world, including New York, Paris, London, Barcelona, Seoul and Chicago. These are typically sprawling conurbations with large and densely populated cores; with an average population of around 1.78 million across an average operational area of nearly 170 km$^2$. These systems exhibit the second most active systems among the BSS studied, with an average of 25,000 daily journeys during weekdays and 22,000 during weekends. In comparison to those ‘very large, high use BSS’, the systems in this cluster appear to have much more cycling infrastructure as a proportion of both their size and total road infrastructure, with nearly 400km, which may have aided their strong adoption relative to the population that they serve. These BSS also appear to exhibit slightly lower levels of monthly pre-
5.5. The Global Bicycle Sharing System Classification

5.5.1 High Use BSS

Like those East Asian BSS in the first first-stage cluster, there is a clear difference in weekday and weekend use, which can be identified through a fall in TDB by approximately 0.8. This is further confirmed in the second-stage clusters in Figure 5.6 and 5.7. During the weekday, all systems exhibit a typical bimodal commuter use, with only minor differentiation between the two sub-clusters in their distribution of journeys between the two-peaks. For those systems in sub-cluster 1, commuting peaks are less pronounced, with the morning peak being one standard deviation away from the mean and evening peak being more extended and peaking at 1.6 standard deviations at 6pm. In comparison, the sub-cluster 2 morning peak reaches two standard deviations whilst evening commuting activity is more concentrated, reaching its peak 1.5 at 5pm. Weekend activity dynamics, on the other hand, depict a clear differentiation in use (Figure 5.7). Here, the majority of BSS (11 systems) exhibit one extended peak that are likely driven by leisure and tourism use, but for the remaining four BSS, there appears to be a persistence in the bimodal commuter pattern. These systems are namely the Changshu, Kunshan, Wenzhou and Zhongshan BSS that are all located in China, much similar to those dynamics observed among the ‘very large, high use’ Chinese BSS. This provides further evidence to suggest that Chinese BSS are unique in their tendency to endure similar commuter dominance throughout the weekend.

5.5.4 Medium BSS with Extensive Cycling Infrastructure

The third first-stage cluster, ‘medium BSS with extensive cycling infrastructure’, is a small cluster of five systems that are characterised by an unusually large road and cycling infrastructure network for their size. Although the average size of these systems are around 74km$^2$ and serve a population of around 45,000, it appears that over one in nine kilometres of road have some form of cycling infrastructure. These systems include, three German BSS (Stuttgart, Mannheim and Bremen) and 2 USA BSS (Washington DC and Los Angeles). They are located in areas which are relatively flat in comparison to other BSS types and receive limited amounts of rainfall and hilliness, which may also be some contributing factors to their high use.
that would suggest ideal scenarios for cycling uptake. The literature would suggest that these urban environments would provide ideal conditions for significant BSS demand, but the contrary appears to be the case for such systems, with a very modest amount of activity, with approximately 2,700 daily journeys on both weekdays and weekends. This means that with an average fleet size of 1,100 bicycles, they exhibit 1.5 TDB.

We can explore these trends further in those second-stage clusters that are depicted in Figure 5.8 and 5.9. Much like those ‘very large, high use’ BSS, at face value there appears to be little differentiation in the journey dynamics across weekdays and weekends, with very similar amounts of journeys and TDB. Although
this is the case, four systems appear to exhibit typical weekday bimodal commuter patterns and an extended unimodal distribution during weekends. The primary differentiating factor among these systems of note is in the evening commuting peak, during the weekday, that observe a very significant proportion of such activity, at 2 standard deviations away from the mean at 5pm. There also appears to be a slight increase in use in the early afternoon at 1pm. Such popularity in use towards the evening may be indicative of a greater mixture of BSS use, with a larger proportion of leisure type use during the week in comparison to other BSS.

Aside from those four systems, Stuttgart has a very unique journey distribution that is messy and does not conform to those established use patterns across both
weekdays and weekends. Investigating this BSS in more detail, it has been identified that the system was very small, containing approximately 40 bicycles, and was highly inefficient with only 0.5 TDB. As a result of its size, the system was not of a scale to be adopted by any regular users and thus exhibits a high level of inconsistency in its journey dynamics. Although inefficient, the k-means clustering has grouped the system together with these other BSS due to the fact that during the study period, the operational area of the system was bound to an area that had cycling infrastructure on nearly all of the road network. Although the Stuttgart BSS was very small during the study period, it has since been redeveloped and expanded by a new operator with a fleet of nearly 1,000 bicycles.

### 5.5.5 Small to Medium Efficient BSS

‘Small to medium efficient BSS’ is the fourth first-stage cluster and contains a large group of 66 systems. Since this is such a large group, there is much greater variation across the features but are unified in through the efficient use of their bicycle fleets. Here, systems exhibit nearly five TDB during the weekday that falls to 3.8 TDB during the weekend. This is impressive, given their relatively small nature in comparison to those ‘very large, high use BSS’ and ‘large BSS in major cities’. Here, these BSS serve approximately 350,000 people in an area of $36\text{km}^2$. The cycling infrastructure is of similar proportion to those ‘large BSS in major cities’ and are comparable in terms of their hilliness but differ in the average monthly precipitation, with approximately 82mm.

As identified among the ‘large BSS in major cities’ cluster, initial indications from average daily journeys and TDB suggest a stepwise difference between weekday and weekend use. This is exemplified in the assignment of BSS in the second-stage clusters that can be observed in Figure 5.10 and 5.11. During the weekday, all sub-clusters showcase some variation of bimodal commuter distribution. The eight BSS assigned to sub-cluster 1 have the most distinctive commuting patterns, with peaks in the morning and evening of approximately two standard deviations. Sub-cluster 2 BSS, of which there are 29, appear to have a very distinctive morning peak combined with extended afternoon use. On the other hand, the remaining 29 BSS in
sub-cluster 3 appear to present some indication of elevated use in the morning commuting hours, but use is dominated throughout the afternoon. The distribution of
Figure 5.11: Average weekend use profiles for ‘small to medium efficient BSS’

BSS across these second-stage clusters are likely indicative of differing proportions of commuting users, with the first containing the most, and the third containing the
least. During the weekend, the vast majority of systems are clustered within the first and second sub-cluster, containing 52 and 10 BSS respectively. Here, the primary differentiate between such systems is in the presentation of the extended afternoon peak. Those BSS in sub-cluster 2 appear to exhibit a distinctive decrease in use around 3pm, which is not present in sub-cluster 1. Of those 10 BSS in sub-cluster 2, nine of these are located in Southern Europe, a region that typically takes afternoon naps, most commonly known as a *siestas* from Spain. As such, the decreases in activity during the afternoon are very likely indicative of such napping activity among residents within these BSS. Finally, weekend sub-cluster 3 are indicative of systems that are comprised of commuting users. This is very similar to those Chinese BSS observed with such persistent patterns in the first and second first-stage clusters. Again, those four BSS in this sub-cluster are all Chinese BSS, namely Huangyan, Gaomi, Changyi and Sanxiang, providing stronger evidence for such persistence in commuting activity among Chinese BSS.

### 5.5.6 Small to Medium Inefficient BSS

The final cluster, ‘*small to medium inefficient BSS’*, is the largest first-stage group, with 233 BSS, that are characterised by their inefficient use, exhibiting just one TDB on average. These systems serve an average of 110,000 people across operational areas of approximately $16\text{km}^2$. Being such a large group, the hilliness and level of precipitation are varied, although these are not particularly interesting or differentiating among such systems. Much akin to the vast majority of BSS within this analysis, apart from those ‘*medium BSS with extensive cycling infrastructure’*, the amount of cycling infrastructure are similar proportion to the availability of road infrastructure.

As noted, the efficiency of such systems are the primary differentiating factor, with many such systems failing to achieve the crucial one TDB mark. Exploring the second-stage clusters in Figure 5.12 and 5.13 enable some indication of why this is the case. Among weekday use, only those 73 BSS assigned under sub-cluster 3 exhibit clear bimodal commuting patterns. Those 99 BSS assigned to sub-cluster 2 provide some indication of commuting use, though much like those sub-clusters
Figure 5.12: Average weekday use profiles for ‘small to medium inefficient BSS’
identified in other BSS types, these peaks are much more diluted, with the evening peak far outweighing the morning. Sub-cluster 1 consists of 38 BSS which exhibit
journey dynamics that are typical of leisure users, with more consistent use throughout the late morning and afternoon. Those remaining 123 BSS in sub-cluster 4 are the smallest BSS and do not appear to exhibit any particular daily dynamics. Much like the Stuttgart BSS identified in the ‘medium BSS with extensive infrastructure’ cluster, due to their small nature, these BSS are heavily underutilised and have not been adopted by regular users.

Assessing weekend journey activity among these ‘small to medium inefficient BSS’, the issue of size and regular users becomes even more apparent, with sub-cluster 1 and sub-cluster 3, containing 48 and 12 systems respectively, indicating dynamics that are inconsistent and non-representative of any regular user based activity. This is especially pronounced in sub-cluster 3, though those in sub-cluster 1 do show some similarities to the typical unimodal leisure afternoon use that can be observed in sub-cluster 2. These 173 systems in the second sub-cluster are highly consistent with those observed to have some indication of commuting use during the week. Conducing some further analyses on comparing those 173 systems which exhibit typical weekday commuting and weekend leisure activity against the other BSS, it becomes very apparent that size and efficiency are key differentiating factors between such systems. Those BSS with consistent and regular users serve an average of 13,500 people with approximately 350 bicycles and achieve around 1 TDB on weekdays and weekends. On the other hand, those BSS which are inconsistently used, as identified throughout this second-stage clustering, are significantly smaller, serving an average of 3,900 people with 80 bicycles. It is hypothesised that due to their small scale operation, their TDB are significantly lower, achieving less than 0.5 TDB during weekdays and weekends and thus result in such inconsistent use.

5.6 Discussion

Taken collectively, this analysis has enabled the identification of five prominent types of BSS around the world. This is a significant achievement given the limited availability of data on BSS (Mátrai and Tóth 2016), employing the minimal data that have been extracted from those various BSS API feeds in Chapter 3. In com-
5.6. Discussion

bination with those BSS metrics, utilising global raster data provide a methodology to add contextual information to those BSS that have been found to have significant influence over their demand and utilisation. The k-means clustering algorithm has provided a holistic overview of these BSS types that have subsequently been named based on such characteristics.

‘Very large, high use BSS’ were identified to be the largest and most used systems and are all located in East Asia. ‘Large BSS in major cities’ are similarly a group of highly utilised systems, with their key differentiating factor being their size, experiencing approximately one third of the journeys observed in those largest BSS. ‘Medium BSS with extensive cycling infrastructure’ are smaller still, but are unique in their plentiful availability of cycling infrastructure. Although the literature would suggest that such systems would experience high demand, they appear to be less efficient in comparison to those larger BSS, experiencing around 1.5 TDB. The two largest clusters, ‘small to medium efficient BSS’ and ‘small to medium inefficient BSS’, contain a collection of the smallest system with less than 4,000 bicycles. These systems are differentiated by the efficiency of their operation, with those ‘small to medium efficient BSS’ experiencing in nearly 5 TDB during the weekday, whilst those ‘small to medium inefficient BSS’ are utilised less than once a day.

Although such cluster assignments help to broaden our understanding of the global landscape of BSS, there are some unusual findings which would indicate conflicts between previous research findings. An example of this is regarding the apparent randomness of measures of precipitation, slope and infrastructure length. Where previous literature has found significant relationships between these variables and the use of a BSS, there appears to be little correlation between the size and use of the system to these variables within the clusters (Buck and Buehler 2012; Corcoran et al. 2014; Faghih-Imani et al. 2014; Frade and Ribeiro 2014; Mateo-Babiano et al. 2016; Miranda-Moreno and Nosal 2011; Jurdak 2013). This is due to the fact that the clustering methods are based on (dis)similarity measures and therefore those metrics with larger values are clustered together, regardless of the
impacts that they might have on the use. These variables are included for a con- 
textual understanding of the city, although future research may wish to exploit these 
data and extend on those analyses of built environment factors at a truly global-scale 
(Bieliński et al. 2019; Médard de Chardon et al. 2017)

Amongst these major BSS groups, second-stage clustering aimed to extend on 
our understanding of how these systems are used throughout a typical weekday and 
weekend. Across all of the BSS studied, it appears as though Chinese BSS are 
unique in their dynamics, exhibiting the persistence of a bimodal commuter distri-
bution throughout the weekend. As a result, these systems see little differentiation in 
the temporal distribution of journeys, making them highly efficient, no matter their 
size or associated cluster. This is highly unusual given that most systems exhibit a 
umimodal distribution throughout weekends that are indicative of greater leisure and 
recreational use (O’Brien et al. 2014; Zaltz Austwick et al. 2013; Mateo-Babiano 
et al. 2016; Maas et al. 2021). Conducting analyses on GPS and smart card data 
of buses in Qingdao, China, Shao et al. (2019) are able to identify similar persis-
tence of commuting behaviours among younger people throughout the weekend in 
comparison to the elderly population. This is peculiar given that the spatial analy-
sis indicates a clear reduction in commuting related journeys to areas characterised 
by commercial and office related use. Li et al. (2021b) conduct a study on the 
spatio-temporal distribution of activity in the Shenzhen dockless BSS to help infer 
trip purposes. Here, they are able to determine that weekend trip purposes during 
the morning and evening peaks correlate to both schooling and training institutions 
in addition to work related commuting purposes. Similarly, the Shenzhen metro 
exhibits similar morning and evening peaks, though the weekend evening peaks 
appear to occur slightly later than weekday peaks (Tang et al. 2020).

Although there are no established causal mechanisms for the persistence of 
commuter type journey distribution during weekends in China, the temporal dy-
namics of such activity, combined with evidence from other modes of travel in 
Chinese cities can provide reasonable certainty to the persistence of community 
activity throughout the weekend in such systems. This is novel, providing some
5.6. Discussion

Insight into the potential reasons for the success of BSS in China. The rise of the ‘996.ICU’ movement in 2019 that commenced in retaliation to the unofficial 9am to 9pm 6 day a week work schedule amongst tech workers in China sheds some light on the working culture in the country (Li 2019). Though the commuting patterns observed throughout the weekend are directly correlated with such working activity, the hard-working nature of individuals amongst the various industries may be another potential reason for the persistence of such activity.

Another unique discovery amongst the second-stage clusters are those Southern European BSS in the ‘small to medium efficient BSS’ cluster that display a dip in activity in the early afternoon that are indicative of the siesta tradition. This has previously only been identified within the Seville (Hampshire and Marla 2012; Bean et al. 2021), that is also included within this cluster in this analysis. Therefore, it is clear that such activity dynamics not unique to Seville and are characteristic of cities that also participate in similar traditions when the climate gets too hot. This highlights the merits of the two-staged clustering methodology in the identification of unique BSS temporal dynamics amongst the hundreds that were analysed.

Finally, within those smallest BSS, it is clear that their size is a key determinant and necessity for consistent journey dynamics. This is evident in the second-stage clustering of ‘small to medium inefficient BSS’, where those 173 BSS identified to have typical weekday and weekend journey dynamics were considerably larger than those with much more inconsistent and sporadic dynamics. As a result, it is hypothesised that users are more likely to regularly adopt BSS as a mode of urban transportation in systems of larger scale for two primary of reasons. First of all, if there are more bicycles within a system, the availability and accessibility of the mode are increased. Second of all, systems with more bicycles are likely to operate over larger spatial areas that would enable users to make more effective use of the system. Conversely, systems with only a few bicycles that operate over small areas warrant negligible advantages, if any at all, in comparison to walking. Therefore, it is essential that systems have some sense of scale and size to increase adoption among residents.
Collectively, the two-staged clustering approach detailed in this chapter evidently presents a valuable methodology in the identification of global-scale BSS dynamics. The identification of well-defined clusters, in both the first- and second-stage of clustering are a credit to the well justified clustering decisions. Therefore, these results are presented with great confidence in enabling the most comprehensive comparison of BSS dynamics.

5.6.1 Evaluation

The analysis conducted within this chapter provides novel and nuanced insights into the global BSS landscape, enabling for a greater understanding of the structure and relationships between BSS from all around the world. Although the analysis was conducted with rigour, there are some areas for improvement that need to be acknowledged.

When considering the data, it is important to highlight the limitations identified throughout Chapter 4 in addition to those constraints when employing the metrics at a global-scale. This is primarily an issue in relation to the over-representation of BSS in Europe and under-representation of BSS in Asia, as identified in Figure 3.2. Unfortunately, although best efforts were made to capture as many API feeds as possible, these were significantly more inaccessible for Asian BSS. Although the data are not perfectly representative of the global distribution of systems, the UCL BSS data collection goes some way beyond the existing literature that fail to include any Asian systems or focus exclusively on them. Therefore, future data collection efforts within the UCL collection should prioritise the identification and collection of data from such systems.

Similarly, as has been mentioned before, it is important to highlight and acknowledge that these metrics have been calculated in an automated programmatic manner. This means that BSS attributes may be subject to some inaccuracies. Although these instances have been minimised by identifying potential data errors using the validation techniques detailed in Section 4.1.2, it would be naive to assume that the measures were not subject to some error. This being said, the metrics that have been created are the best quality that are currently available, with metrics
that are homogeneously calculated, ensuring comparability and accountability in the way that they have been created.

In addition to issues surrounding the data, it is also important to note issues surrounding the clustering techniques used to analyse this data. The bespoke two-staged method employed within this chapter is not the optimal means of analysis since the data had to be separated. The results are therefore spread across the two separate analytical methods used, resulting in a large number of clusters. This method was chosen as the clustering techniques used are established and provide useful insights into their respective data types. It is acknowledged that it would be beneficial for the interpretation of a comparison between BSS to include both elements of the data that is available. There are currently very few established methods which enable the clustering of the different types of data, which therefore led to the use of this two-staged method.

The various subjective decisions that were made in the lead up to the presentation of results are also subject to alternative decision-making processes that would have produced differing results. To mitigate such issues, clear justifications have been provided for each decision that has been made. For example, though nine clusters were identified to be the next most optimal partition of the data using traditional identification methods, this was rejected due to the very small nature of clusters that would have been created that did not have strong cluster identities. The results presented aim to provide a good balance between detail and generality, enabling the identification of distinct BSS types and their uses, whilst minimising the number of absolute clusters.

5.6.2 Applications

Although the analyses have achieved their aim in providing novel insights on the global BSS landscape, these results present valuable opportunities for applications outside of academia as well. For example, the lack of data availability on BSS has allowed for unchallenged exaggerations in terms of the published statistics from BSS operators (Médard de Chardon and Caruso 2015). The data and heuristics adopted here could therefore be used to verify the published statistics for those
systems whose operators release highlight statistics as metrics calculated within this research are independent of any operator motivations, they may be considered impartial.

The clustering methodology also provides a means to be able to predict and estimate BSS attributes for those systems that are in operation and do not release data or new BSS that are planning on being opened. This research offers a tool to help those cities better plan systems, should they wish to implement new or improve existing BSS. By comparing the characteristics of the city to those within the analysis, city officials can learn from those more efficient systems. This is timely due to the fast-expanding nature of this mode of micromobility. Alternative modes of micromobility, such as e-scooters and dockless BSS, have a similar data format and are similarly limited in terms of data availability. The methods outlined in this research can easily be extended to other new modes of urban micromobility and help to provide a framework for comparison between system metrics.

5.7 Chapter Summary

This chapter presents a novel two-stage clustering methodology that employ those BSS metrics calculated in Chapter 4 to create the most comprehensive global classification of BSS. The data are manipulated to create a static, long-term overview of system characteristics by aggregating such metrics between April and September of 2018. Using the k-means algorithm, the first-stage of clustering enables an identification of the general types of BSS based on their static attributes, such as their size, use and local demographic, climatic, topographic and infrastructural characteristics. The second-stage clustering employ hierarchical DTW clustering to identify the typical use patterns within each of the first-stage clusters to help identify commonalities among the temporal distribution of journeys.

Such analysis identifies five distinct BSS types that are named after their unique characteristics, those being: ‘very large, high use BSS’, ‘large BSS in major cities’, ‘medium BSS with extensive cycling infrastructure’, ‘small to medium efficient BSS’, and ‘small to medium inefficient BSS’. Within each of these clus-
ters, the majority of systems exhibit a bimodal morning and evening commuting pattern during weekdays, whilst weekends are indicative of greater leisure use with more homogeneous and distributed use throughout the afternoon. Chinese systems appear to be unique in their persistence of commuting patterns throughout the weekend across various sizes, whilst some Southern European BSS appear to exhibit a dip in weekend afternoon activity due to siesta behaviour. The smallest BSS appear to present a lack of consistent journey dynamics that is likely an indication of an issue with scale.

The results demonstrate the exhaustive depth and breadth of the data that have been collected and their value in providing the foundations for very large temporal and geographic scale analysis of BSS that have not been possible before. The novel two-staged clustering methodology on both time constant multivariate features and time-series features provide a methodological framework to best exploit the various aspects of openly accessible micromobility data. The findings can also help to guide future developments of BSS and alternative modes of micromobility, striving towards those systems which are highly utilised and serve the needs of the surrounding population.
Chapter 6

Impacts of COVID-19 on London Bicycle Sharing

The COVID-19 pandemic became a global issue at the start of 2020 and resulted in unprecedented impacts on daily life. These disruptions were primarily a result of the various non-pharmaceutical intervention (NPI) that have been implemented as a way to mitigate its spread (Askitas et al. 2021; Bian et al. 2021; Nouvellet et al. 2021). NPI typically constitute of various restrictions of the movement of populations, aiming to minimise contact and therefore slow and counter the spread of the virus (Yen et al. 2020; Lee et al. 2020; Lin et al. 2020). These included social distancing measures, gathering size restrictions, wearing face coverings, work, education and retail space closures in addition to many others (Askitas et al. 2021; Bian et al. 2021; Nouvellet et al. 2021).

On top of such measures, during periods of significant increases in the number of cases, lockdown restrictions have been imposed, forcing people to quarantine in their homes unless for limited and specific purposes (Ding et al. 2020; Hajjodemetriou et al. 2020; Jeffrey et al. 2020; Sharifi and Khavarian-Garmsir 2020). Early studies on the effectiveness of lockdown restrictions in reducing such spreading has proven their ability to be a powerful tool in helping to contain the pandemic (Alfano and Ercolano 2020; Kharroubi and Saleh 2020; Moris and Schizas 2020). Further evidence of such success can be observed in the number of unnecessary deaths caused by delayed implementations of lockdowns or refusal to implement
such policies. A good example of this is in Sweden, a country that imposed few such restrictions on their population and were also found to have one of the highest levels of excess mortality during the first wave of the pandemic in the world (Sulyok and Walker 2021). As such, what was once considered an authoritarian intervention approach championed by China has become the global standard of response (Ren 2020).

Although lockdown restrictions are a vital tool in the arsenal of NPI to help the world fight against COVID-19, they were also the most constraining on the population’s ability to conduct typical day-to-day activities. As a result, these periods of significant change for populations had the most prominent and disruptive impacts on mobility (Aloi et al. 2020; Askitas et al. 2021; Barbieri et al. 2021; Sulyok and Walker 2021; Warren and Skillman 2020). As highlighted in Section 2.2.2, the study of such mobility changes are generally limited to very holistic and top-level changes, such as estimations of journey volumes (Palma et al. 2022), durations (Heydari et al. 2021; Teixeira and Lopes 2020; Padmanabhan et al. 2021) and modal shift (Dingil and Esztergár-Kiss 2021; Bucsky 2020; Teixeira et al. 2022). Although these are informative of general changes in mobility trends during the pandemic period, they lack detailed insights into more granular mobility dynamic changes in the medium-term.

This chapter aims to explore granular changes in activity dynamics during periods of lockdown in London through an in-depth analysis of the docked Santander Cycles BSS. The analysis of BSS are a great proxy for activity dynamics in cities through the exploitation of open journey data that detail when and where individuals are travelling in the city, especially in relation to those short distance journeys and multi-modal journeys. Within the UK, the national government implemented various combinations of NPI, including periods of enforced lockdown to ‘flatten the curve’ of exponential increases in COVID-19 cases (Davies et al. 2020). To date there have been three national lockdowns of differing lengths that have caused major disruptions economically, socially and environmentally. The analysis compares each of these lockdown periods to a control period taken from the equivalent
6.1 London and COVID-19

The London BSS has been in operation since 2010 and was one of the first large-scale BSS that saw significant use and uptake. Since its conception, the BSS has seen growth, not only in its use but also in size and scale, with the largest expansions occurring prior to the 2012 Summer Olympics to include Stratford on the Eastern periphery of the system (O’Brien 2012). In Chapter 5, the analysis identifies the system to be similar in nature to those of other BSS in major cities such as New York, Paris and Seoul. The system currently consists of a fleet of around 10,000 bicycles distributed across nearly 800 docking stations, operating over an area of $115\text{km}^2$ across inner London boroughs. It is well integrated with other modes of PT, such as the London Underground and bus services and provides a solution to the ‘first/last mile problem’ that commuters face when using PT (Gu et al. 2019; Sari Aslam et al. 2015), as well as opportunities to conduct short utility, leisure or
exercise trips (Zaltz Austwick et al. 2013).

In London, there has been a heavy prioritisation towards reducing car traffic by at least 27% by the end of the decade to meet climate change targets, with a broader objective to increase the proportion of trips made by walking, cycling or PT from 63% in 2018 to 80% in 2041 (TfL 2018b). COVID-19 provided great impetus to accelerate the implementation of green policies such as the announcement of a £250 million package for temporary active transport infrastructure at the beginning of the pandemic (Reid 2020). This also aligns with other policies within the 2021 London Plan (GLA 2021b), such as the Healthy Streets and low traffic neighbourhood (LTN) initiatives that aim to reduce traffic through residential areas and improve local air quality (Aldred et al. 2021). These policies have facilitated the construction of an additional 90 km of cycling infrastructure (O’Malley 2021), which have been found to have very significant and positive impacts on cycling demand and uptake across numerous studies (Buck and Buehler 2012; Buehler and Dill 2016; Faghih-Imani et al. 2017; Mateo-Babiano et al. 2016; Médard de Chardon 2019; Pucher and Buehler 2008; Rixey 2013).

The first case of COVID-19 in the UK was reported on the 31st January 2020 and the first death reported on the 5th March 2020. The increasing severity of the situation promoted the government to implement a COVID-19 Emergency Bill on 8th March 2020, which outlined the national plan to support health and social care sectors. Under such frameworks, it was imperative that PT services should continue to operate as best they could under such circumstances in order to ensure the functioning of society as well facilitating the travel of essential workers. Although such efforts were made, TfL and national rail services experienced reductions in service due to an increasing number of infected individuals. Eventually, the first nation-wide lockdown was enforced on 23rd March 2020, with guidance to avoid all non-essential travel and practise social distancing.

Since the imposition of the first lockdown, the government has continually amended and updated national guidance and policies in conjunction with the spread of the virus. For example, June 1st and 15th saw the phased re-opening of schools
and non-essential retail shops respectively, on August 3rd the ‘Eat Out to Help Out’ scheme was introduced that encouraged people to eat at restaurants with a 50% discount on meals in addition to other relaxations of policy throughout the duration of the pandemic in the UK. Although restrictions were generally eased, there have been additional periods of enforced lockdown that have been implemented under a new tiered system of restrictions that commenced in October 2020. Due to the ‘almost-unprecedented’ and ‘trial-and-error’ nature of policy implementation and behavioural changes resulting from COVID-19 (Cairney 2021), there are numerous opportunities to conduct research to provide detailed insights into their impacts for future policy planning.

6.2 The Impacts of COVID-19 on Bicycle Sharing Systems

The way in which the COVID-19 pandemic has influenced the mobility choices of populations are varied depending on their perceived risk and necessity, as highlighted in Section 2.2.2. Generally, the literature surrounding the mobility of populations suggest a strong overall narrative of severe declines in the number of individuals using PT modes during the onset of the pandemic, whilst private and active modes of transport appear to increase in popularity due to their advantages as modes that minimise contact with potentially infected individuals. Among such active modes, BSS appear to be an increasingly popular mode of transport, facilitating healthy, sustainable and COVID safe journeys.

Song et al. (2021) conducts a summary of recent progress in BSS studies related to COVID-19. Similarly, Chen et al. (2022a) summarises literature on the impacts on cycling behaviour more generally, including the impacts on private cycling behaviour. Throughout such literature, the impacts on other modes of transit, BSS user behaviour, travel demand prediction and changes in the spatio-temporal activity dynamics have been studied. A range of methods have been implemented across BSS from around the world, showing that such impacts are of global interest (Song et al. 2021).
For example, Shang et al. (2021) explore the impacts of COVID-19 on user behaviour and the environment in three dockless BSS in Beijing. Here, the authors find dramatic decreases in typical daily activity during the onset of the pandemic, with over 950,000 fewer journeys. These reductions appear to be concentrated in areas that are associated with higher population densities, towards the centre of the city, with more activity occurring in areas surrounding hospitals. They also estimate reductions in harmful GhG emissions by taking those journeys over 1 km as potential replacement trips of petroleum-based transit methods, suggesting as much as a 14% fall in the Xicheng district. On the other hand, dockless BSS in Singapore appeared to exhibit a 150% increase in total ridership compared to pre-pandemic levels (Song et al. 2021). Although increases in activity were observed, using network analyses methods, the authors were able to identify that these additional journeys were not distributed homogeneously, with more local journeys and in a polycentric pattern. Journeys also appeared to be spatially autocorrelated with dense residential regions or areas with high levels of mixed land-use.

There are also several studies that investigate such impacts in the USA, such as the New York Citi Bike Scheme (Teixeira and Lopes 2020; Wang and Noland 2021; Xin et al. 2022), Chicago Divvy scheme (Hu et al. 2021) and the Washington D.C. Capital Bikeshare Scheme (Chen et al. 2022a). Across these BSS, there appear to be initial reductions in activity, but this activity appears to return much more quickly than other modes of PT. For example, in New York City, Wang and Noland (2021) compare activity dynamics in the subway and BSS to the equivalent day in 2019. Here, the analysis shows initial reductions in journeys by 95% and 70% respectively, that subsequently returned to pre-COVID levels within the BSS whilst subway activity remained 30% below. Studies on the ability for BSS to cater to unmet demand, such as train strike occurrences (Fuller et al. 2019; Saberi et al. 2018; Yang et al. 2022), provide some indication of the modes’ ability to be adaptive and increase resilience during times of restricted travel.

The literature on BSS analysis during COVID-19 also includes Europe, such as Nikiforiadis et al. (2020) who explore the Thessaloniki BSS in Greece and Kubášák
et al. (2021) who explore the Košice BSS in Slovakia. The London BSS is also the subject of a number of studies (Li et al. 2021a; Heydari et al. 2021; Chibwe et al. 2021). Each of these studies employ different methods.

Li et al. (2021a) use segmented regression models and an interrupted time-series approach to identify differences in activity dynamics throughout the duration of the first lockdown, using 2019 as a control, comparative period. Conducting such analyses, they find that low infection boroughs exhibited smaller declines in activity in comparison to boroughs with relatively more infections that are also associated with lower poverty rates. The analysis is also able to identify the docking stations in close proximity to parks were associated with the highest levels of increase, closely followed by hospitals, with docking stations associated with rail stations exhibiting significant declines. These results showcase the merits of the methods’ ability to uncover interesting changes in journey dynamics during the first lockdown.

The analyses conducted by Chibwe et al. (2021) and Heydari et al. (2021) investigate changes in demand using over eight years of historic data. Chibwe et al. (2021) explore more general trends in journey activity since the 2012 London Olympics using generalised negative binomial regression model whilst Heydari et al. (2021) use journey data from 2010 to be able to predict the demand and duration for journeys in 2020 should the pandemic not have occurred, employing a Bayesian second-order random walk time-series model. Although useful in being able to identify periods of significant deviation from such predicted values, these findings do not enable any granular insights into changes in activity dynamics within the system.

As it is evident through this review of literature on BSS during COVID-19, there are a growing body of studies that deploy a variety of methods to provide insights into the changing nature of their use. Although such analyses are valuable in shedding light on such changes, they are still limited in their exploration of various methods and academics have called for more detailed analyses in specific contexts and using alternative methods to justify such claims (Shang et al. 2021; Song et al. 2021). Xin et al. (2022) specifically highlight the lack of complex net-
work analyses among BSS literature, as a valuable methodology to quantify gran-
ular spatio-temporal changes in activity dynamics within a system, something that
Wang and Noland (2021) suggests needs additional work. As a result, the analyses
presented in this chapter aim to extend on those methods presented here, conduct-
ing exploratory spatio-temporal and statistical analysis in combination with more
complex network analysis methods on the docked London BSS.

6.3 Data

6.3.1 COVID-19 Policy and Response Data

The primary independent variables that were used within these analyses are taken
from the Oxford COVID-19 Government response Tracker (OxCGRT) dataset. This
dataset consists of 23 indicators that are representative of various containment and
closure, health system, economic, vaccination and other miscellaneous policies that
have been implemented across 186 countries in response to the pandemic. The
indicators provide the basis for several indices that have been created to capture
more general and holistic characteristics of such policy implementations (Hale et
al. 2021). The stringency index (SI) has become a popular measure of general
restriction stringency that is a weighted combination of various containment and
closure policies in addition to the level of public information campaigns. A detailed
account of the collection and calculation of the OxCGRT dataset is detailed in Hale
et al. (2021).

As a standardised, longitudinal global dataset on the implementation of various
NPI, it has been implemented across a number of quantitative academic studies that
analyse the impacts of such restrictions. For example, Santamaria et al. (2020) anal-
alyse the impacts of containment and closure policies on the movement of European
residents using mobile positioning data, finding that these policies are responsible
for a large proportion of the change in mobility patterns. Similarly, Sulyok and
Walker (2021) and McKenzie and Adams (2020) analyse the relationship between
the SI and Google and Apple COVID-19 mobility datasets, with evidence suggest-
ing strong compliance and quick responses to changes in NPI stringency.
Within this chapter, the OxCGRT is used to define lockdown durations as well as gain an overall level of compliance and response time, much like McKenzie and Adams (2020), through the use of the SI. Lockdown periods were defined using the stay-at-home (SAH) requirement indicator that is measured on an ordinal scale from zero to three, where zero indicates no measures, one indicates recommendations to SAH, two indicates that populations are required to SAH except for essential trips and three indicates populations are required to SAH with minimal exceptions. The definitions of lockdowns are taken as days in which those countries are assigned to levels two and three of the SAH indicator.

This decision was made as the analysis was originally planned to be a part of a larger scale analyses of lockdowns on all BSS within the UCL BSS Data Collection (Section 3.2), that would have provided a homogeneous and scalable methodology for the identification of lockdown periods, much like the confounding variables calculated in Section 4.1.7. The scope of the analysis has since been limited to one BSS in favour of more granular analysis techniques that are difficult to scale. In such a case where more general trends across a large number of BSS are desired, these data and methods can provide the basis for such analysis.

Much like many other studies of BSS during the pandemic, an equivalent control period has been defined using data in 2019 (Li et al. 2021a; Wang and Noland 2021). Table 6.1 detail the duration and intricacies of each of these study and control periods.

<table>
<thead>
<tr>
<th>Name</th>
<th>Start</th>
<th>End</th>
<th>No. of Weekdays</th>
<th>No. of Weekends</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lockdown 1</td>
<td>23/03/2020</td>
<td>11/05/2020</td>
<td>36</td>
<td>14</td>
</tr>
<tr>
<td>Lockdown 1 Control</td>
<td>25/03/2019</td>
<td>13/05/2019</td>
<td>36</td>
<td>14</td>
</tr>
<tr>
<td>Lockdown 2</td>
<td>05/11/2020</td>
<td>02/12/2020</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>Lockdown 2 Control</td>
<td>07/11/2019</td>
<td>04/12/2019</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>Lockdown 3</td>
<td>19/12/2020</td>
<td>27/03/2021</td>
<td>70</td>
<td>29</td>
</tr>
<tr>
<td>Lockdown 3 Control</td>
<td>21/12/2019</td>
<td>30/03/2019</td>
<td>70</td>
<td>29</td>
</tr>
</tbody>
</table>

1To ensure the same number of weekdays and weekends were taken, the control period was extended to include 30/03/2019 (weekend) whilst removing data from 29/03/2019 (weekday).
6.3.2 London BSS Journey Data

The primary dependent variable in the analysis presented in this chapter is the level of cycling activity within the London BSS. Here, the journey metrics calculated in Chapter 4 are exploited for those analyses that are more easily scalable to future-proof such research, whilst more granular and bespoke analytical methods, namely network analyses, rely on less commonly available journey OD data (Section 3.3).

Within this analysis the journey metrics that were calculated in Chapter 4 were utilised in order to conduct exploratory analyses as well as determine levels of compliance and response time in relation to the OxCGRT SI, much akin to McKenzie and Adams (2020). The journey estimation methodology in Section 4.1.1 are extended to calculate a measure of percentage change in journeys from the equivalent day in 2019 for the clear understanding of relative performance.

Initially, it was vital that the estimated journey data was a complete time-series from January 2019 to June 2021 due to the various time-series analyses that would be conducted using these new metrics. As a result, a simple multiple linear imputation methodology was implemented in order to complete the journey data. Those missing days within the journey estimations were primarily a result of the journey validation and cleaning techniques (Section 4.1.2) that identified and removed data error occurrences in the raw data feed. Missing estimations may have also occurred due to data feed breaks. Within the London BSS journey estimations, only two days were found to have missing journey estimations, demonstrating the minimal imputation and manipulation of this variable.

Although many COVID-19 mobility datasets, such as those from Google and Apple, use a baseline period from the first few weeks of 2020, this methodology is not robust in accounting for seasonal variations in urban mobility patterns. These seasonal variations have been identified throughout BSS literature (Ahmed et al. 2010), as well as important differences that have been found between weekday and weekend patterns (Zaltz Austwick et al. 2013; Faghih-Imani et al. 2014; Faghih-Imani et al. 2017; Lovelace et al. 2020; Mateo-Babiano et al. 2016). As a result, to best capture the relative changes in the number of journeys in 2020 and 2021,
journey data from 2019 were first aggregated to the weekly level, separating weekdays and weekends. These aggregated journey values in 2019 were then used as a baseline from which we were able to calculate the percentage change for each day in 2020/21. By manipulating the data in this way, we are able to best capture such seasonal and weekly changes in activity patterns that fail to be considered when using a short and static baseline period.

In addition to the percentage change in journeys at a daily level, to smooth out any irregularities within this measure, a 7-day moving average - similar to the aggregation in Wang and Zhou (2017) - was calculated. This helps to minimise any large spikes in journeys, enabling insight into some more general changes and trends within journeys during the pandemic period.

The journey data used within this analysis has been downloaded using the bikedata R package (Padgham and Ellison 2017). These data detail each journey’s start and end docking station, with associated timestamps. Within this paper, journey data are utilised to gain an understanding of spatial trends observed during lockdown periods, in addition to network-level changes in journey activity through the creation of graph structures. Dock location data were combined with journey data to enable geographically accurate spatial aggregations and network visualisations.

The OD journey data were initially cleaned to remove all journeys which were shorter than 2 minutes or longer than 6 hours in duration in order to best capture actual BSS trips during COVID-19. Although other studies that utilise journey data apply more stringent cleaning processes, restricting journeys to between 1 and 60 minutes (Feng et al. 2020; Li et al. 2019) or under 180 minutes (Fishman et al. 2014a), for the purposes of this analysis it was important to ensure that all types of leisure activity were captured as those individuals without access to a private bicycle may have hired bicycles for longer durations. Journeys that are shorter than 2 minutes are typically indicative of users returning hire bicycles upon the identification of a defect. Since the pricing strategy implemented among the London BSS encourage users to conduct journeys shorter than 30 minutes in duration (£2
fee for 24 hour access, all trips within 30 minutes of cycle hire are free with a £2 charge for each additional 30 minutes), less than 1% of those trips are removed as a result of the journey data cleaning process.

The journey data were initially segregated to the lockdown periods and associated control periods identified in Table 6.1. Within each study period, journey data were spatially aggregated to Uber’s H3 spatial index at resolution 8 based on the location of the origin docking station location. At the resolution level 8, H3 hexagons are composed of edges that are approximately 450 metres in length. This resolution creates a smooth, continuous surface of the London BSS’ extent, enabling an insight into the general spatial trends across the entire system. For each lockdown period and associated control period, split between weekdays and weekends, the percentage difference in journey origins was calculated.

In addition to the use of journey data for descriptive spatial analysis, journey data have also been utilised to create complex network structures. Similar to the spatial analysis, journey data were initially separated into the lockdown periods and associated control periods, for weekdays and weekends separately. The journeys in each period were then aggregated to create an OD matrix, detailing the number of trips taken between each docking station. This was combined with dock location data to create a directed and weighted graph structure for the purposes of network analysis techniques, detailed in Section 6.6.

6.4 Exploratory Spatio-Temporal Analysis

In order to gain an initial understanding of the general impacts of lockdown restrictions on London BSS activity dynamics, several exploratory methods were employed. Figure 6.1 provides a simple overview of the percentage change in journeys in comparison to the Google COVID-19 mobility data. This enables the identification of trends in BSS activity in relation to activity in different areas including, residential, workplace, retail and park spaces.
Figure 6.1: Comparison of changes in activity levels throughout the first 16 months of the COVID-19 pandemic
Exploring such trends, it becomes very apparent that lockdowns appear to exhibit significant reductions in BSS activity, especially during the onset of such restrictions. This reduction peaks at around 65% during the beginning of the first lockdown, that are less pronounced in subsequent periods. It also appears that towards the end of longer lockdown restrictions, such as Lockdown 1 and 3, the activity appears to recover towards baseline levels prior to the end of such restrictions. Comparing BSS activity to the various activities in the Google mobility data, although activity around transit stations, retail and recreation and workplaces remain significantly lower throughout the duration of the pandemic, typically ranging between a 50 to 80% reduction, BSS activity appears to most closely follow trends observed within parks, that are much more varied and experience extended periods of increased activity. This is likely indicative of the utility of the BSS in facilitating more leisure and exercise activities that are located in such areas, especially outside periods of enforced lockdown.

![Average Entropy Comparisons - Weekday](image)

**Figure 6.2:** Average weekday entropy in each lockdown and control period

Aggregating the entropy metric, that is calculated in Section 4.1.5, to each of the study and control periods provides some additional, system-scale insights into the changing distribution of bicycles. Exploring Figure 6.2 and 6.3, it is clear that lockdown restrictions have had a significant impact on the general distribution of bi-
cycles, especially among weekdays. Control periods tend to exhibit a step change in distribution during the onset of the morning commuting activity at 8am. These are then found to steadily redistributed back towards pre-commuting levels throughout the late morning and early afternoon. Again, the patterns that are observed in the Lockdown 1 and 3 control periods are very similar due to the elongated nature of the restrictions. During study periods, changes in the distribution of bicycles are dramatically muted, with only minor, gradual changes in distribution that are observed during periods of enforced lockdown. The relatively continuous levels of distribution throughout such periods suggest that journeys during lockdowns are much more evenly distributed in nature. This is also exemplified by the clear difference in the average imbalance between control and lockdown periods, with a mean of 48% and 59% respectively. On the other hand, Lockdown 2 exhibit comparatively larger changes in entropy which is likely an artefact of its short temporal duration as well as the lack of compliance during its imposition that can be observed in Figure 6.1.

![Average Entropy Comparisons - Weekend](image)

**Figure 6.3:** Average weekend entropy in each lockdown and control period

Changes in entropy observed during typical weekends (Figure 6.3) depict fewer deviations as compared to weekdays. The primary difference that are exhibited throughout each of the study periods are in the spikes in entropy that are observed in the middle of the afternoon (at 3pm for Lockdown 2 and 3 and 4pm for
Lockdown 1). The similarity signals the consistency in journey distribution between control and lockdown periods, whilst the spikes during the afternoon are likely to be indicative of some changed activity behaviours during such times. The changing distributions of such activities are further explored in Figure 6.4 and 6.5 where distribution of journey volumes are explored temporally, throughout a typical weekday and weekend, as well as spatially, across the system’s operational area, respectively.

First, exploring the changes in journey activity observed in Lockdown 1 (Figure 6.4a) there appears to be a clear contrast between weekdays and weekends. Weekdays appear to present a large reduction in journeys throughout the day, with the most significant reductions occurring during commuting peak hours. On the other hand, weekends appear to demonstrate an increase in journey activity throughout the early afternoon, between 12pm and 5pm. These hours are typically associated with leisure, exercise and touristic activity during the weekends (O’Brien et al. 2014). Outside of these hours, journey activity appears to remain at similar levels to the control period. Taken together, the temporality of these changes across weekdays and weekends provides some indication of significant reductions in commuting activity during the weekday, in conjunction with an uptake of cycling activity on weekends.

These trends are corroborated by the spatial variations in journeys observed in Lockdown 1 (Figure 6.5a,b). During weekdays, those docking stations located within the centre of the BSS experienced the most significant decreases. Between the areas of Kings Cross, Covent Garden and the City of London there were reduction in excess of 75% in comparison to the control period. This central region is synonymous with many non-essential activities, such as retail, workplaces and tourist destinations. Emanating from this core, we find that journeys tended towards the control levels up until those peripheral cells which indicate some minor increases in journey activity, resembling a very monocentric pattern. Weekends, on the other hand, present a general increase in journeys throughout the system. Although the reductions in activity within the centre of the BSS remain persistent, the area is much more concentrated to areas surrounding Soho and those reductions
Figure 6.4: Comparison of average hourly weekday and weekend journey counts in each lockdown and control period [a) Lockdown 1, b) Lockdown 2, c) Lockdown 3]
are generally less than 75%. Similar to the weekday patterns, the peripheral cells present the greatest increase in journeys, with 35 cells depicting journeys in excess of double those observed within the control period and areas such as Brixton and the Isle of Dogs standing out as regions where journey uptake increased in excess.
Looking across Figures 6.4 and 6.5, we can observe these general patterns tend to persist throughout the Lockdown 2 and Lockdown 3. In each of these subsequent lockdown periods, weekdays tend to exhibit a concentrated core of major journey reductions that emanate out towards periphery cells with increases in activity. Weekends exhibit a general increase throughout the BSS, with the largest increases tending to occur within the periphery cells. Similarly, exploring the temporal variations in journey activity are very similar to those observed in Lockdown 1, with weekday reductions throughout the day, which are especially significant during commuting peak hours, as well as weekend increases throughout the afternoon period. These increases in weekend mid-afternoon activity are reflected in the spikes in entropy that are observed in Figure 6.3. Although there are many similarities between the lockdown periods, the differences and evolution of behavioural changes throughout COVID-19 can provide some insight into public perception.

In comparison to Lockdown 1, Lockdown 2 exhibits a much greater increase in activity from the control period, spatially, temporally and across both weekdays and weekends. The differences are especially pronounced within west London neighbourhoods (Figure 6.5c), which are typically associated with high income households such as the areas of Chelsea and Kensington. Unlike Lockdown 1, weekend activity in Lockdown 2 show increases throughout the entire BSS (Figure 6.5d). This can also be observed in Figure 6.4b, with an increase in journeys throughout the day, from 10am until 9pm. This peaks at 2pm with an increase of 132%, compared to increases of 49.2% and 78.3% at the same time during Lockdown 1 and 3 respectively. Similarly, weekday journeys exhibit the least reduction in activity across the lockdowns, with a 56.6% decline during the morning commuting peak (8am) in comparison to an 85.3% and 71.6% reduction in Lockdown 1 and 3 respectively. In addition, weekday activity in Lockdown 2 appears to exhibit an extended period of journey increases in comparison to the control period across the early afternoon between 1pm and 6pm (up to 48.6%) unlike the other lockdown instances. Therefore, Lockdown 2 appears to be different in its spatial and tempo-
eral activity changes in comparison to the control period, exhibiting an increase in activity except during weekday commuting hours.

Finally, investigating Lockdown 3 (Figure 6.4c and Figure 6.5e,f), the trends that are observed appear to resemble Lockdown 1 both spatially and temporally. In terms of the temporality, the patterns of change observed throughout the typical weekday and weekend are very comparable, although the magnitude of these changes differ slightly. Weekday commuting peak hours trend towards pre-pandemic levels whilst weekend peak activity increases from a 49.2% increase in Lockdown 1 to 78.3% increase in Lockdown 3. These increases in activity are also observed spatially, with no cells exceeding a 75% decrease in activity during the weekday and also includes 7 cells which experience over double the number of journeys in relation to the control period in comparison to just 2 in Lockdown 1. The monocentric patterns of Lockdown 1 can still be made out, although the central areas which exhibit reductions are far more concentrated and are littered with cells which exhibit increases in comparison to the control period. These cells which show increases are primarily located in areas in close proximity to green space, while there is also some semblance of commuter use returning with cells containing commuter hub stations such as Victoria Station and Liverpool Street station.

Table 6.2: Average weekday and weekend journey duration for each lockdown and control period (minutes)

<table>
<thead>
<tr>
<th></th>
<th>Weekday Control</th>
<th>Weekday 2020</th>
<th>Weekend Control</th>
<th>Weekend 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lockdown 1</td>
<td>17.4</td>
<td>29.9</td>
<td>24.7</td>
<td>41.0</td>
</tr>
<tr>
<td>Lockdown 2</td>
<td>14.1</td>
<td>19.2</td>
<td>18.0</td>
<td>29.1</td>
</tr>
<tr>
<td>Lockdown 3</td>
<td>15.5</td>
<td>20.4</td>
<td>20.4</td>
<td>28.8</td>
</tr>
</tbody>
</table>

Aggregating journey durations provide some insights about the journeys and how they have changed during periods of lockdown. Heydari et al. (2021) show the changes in journey durations, aggregated by month. Here, Table 6.2 presents the journey durations for each lockdown and associated control period. It becomes apparent that journeys increased in duration across all lockdown periods. This is especially true for those journeys conducted on weekends, which extended in duration.
by nearly 10 minutes across the board. In Lockdowns 2 and 3, journey durations are just under 30 minutes. This is likely a result of the BSS pricing strategy. The weekend activity appears to be on the very cusp of this free initial period which is likely indicative of increased leisure activity, since commuting activity does not tend to increase journey duration unless under circumstances like strike events (Saberi et al. 2018; Yang et al. 2022). An increase in journey duration also occurred during weekday activity in each lockdown period, though these increases were closer to 5 minutes. This is likely due to the fact that weekday activity is primarily composed of commuting activity, leading to a reduced increase in durations.

Taken holistically, this exploratory analysis provides a great foundational understanding of the changes in BSS journey dynamics during periods of lockdown. It becomes apparent that changes observed during the weekday are systematically different to those of the weekend, where weekday changes are much more concentrated temporally, to those morning and evening commuting times, as well as spatially, with the majority of those reductions occurring within the centre of the BSS. This is reflected in the changes observed in the distribution of bicycles, as represented by the average entropy in Figure 6.2 and 6.3. In addition, journeys appear to be longer in duration. These results support those findings in other BSS assessed in the literature, illuminating spatio-temporal intricacies observed within the London Santander Cycles scheme.

6.5 Restriction Compliance and Response Time

Cosine similarity is a measure that is like correlation, enabling a general understanding of the relationship between two time-series variables. In a similar fashion to the analysis conducted by McKenzie and Adams (2020), within this analysis the association between the seven-day rolling percentage change in journeys was analysed in relation to the SI. The analysis produces a measure from 1 to -1, with those positive values indicating a positive association between two time-series measures, and negative if the vectors are diametrically opposed. Since we would like to measure the similarity of journeys in response to the SI, the inverse of the SI was used.
This, therefore, enables a more intuitive interpretation of similarity in magnitude response (SMR), with positive values indicating greater compliance to the increasing stringency of policies. Conducting such analysis, SMR suggests moderate compliance to changes in restriction stringency, with a value of 0.32.

Similar to SMR, lag response time (LRT) aims to capture the compliance of BSS use in relation to the SI. Policy lag is a well-known term in economics that seeks to measure the delay between an economic action and its associated consequence (Bian et al. 2021). Cross-correlation is an analytical method that measures the similarity between two time-series signals as a function of the displacement of one relative to the other. In other words, this signal processing approach enables us to determine the LRT of the seven-day rolling percentage change in journeys in relation to changes observed in the SI. Again, akin to the analysis by McKenzie and Adams (2020), the lag values were calculated through an aggregation method, taking the difference between the number of significant lag days above and below zero. ‘This approach is intended to be more robust than looking for a single highest cross-correlation value as it considers the skewness in the lag distribution’ (McKenzie and Adams 2020, p. 5). The results enable the quantification of the number of days it took for journeys to response to changes in the SI, providing further general indications of user compliance towards COVID restrictions. Results from this analysis suggest that users of BSS respond immediately to changes in restriction stringency, with zero days lag, providing evidence to suggest moderate but rapid compliance to the various NPI implemented among BSS users.

Comparing such results to those countries and activity types analysed in McKenzie and Adams (2020), it becomes more evident that although responses in activity are immediate, compliance is very low. This would make the London BSS the least compliant in comparison to the average SMR across the 129 countries studied, which would cause for concern over the validity of such results. Although this is the case, it is important to highlight the difference in the study period durations that are being analysed. Whilst McKenzie and Adams (2020) conducts analysis using only eight weeks of data, from 15th February to 11th April 2020, the analysis
presented here measure SMR over the duration of 18 months. Therefore, it is expected that the correlation will decrease over greater temporal durations due to the increase in exogenous factors that are likely to influence changes in demand.

As a result, the SMR for all BSS that were available and operational during the pandemic period in UCL BSS data collection were calculated in order to provide some perspective on the relative compliance of activity. The SMR of 339 BSS were able to be conducted, with the summary of the distribution of such values presented in Table 6.3. Here, we get an indication that the activity in London BSS is above the average level of compliance amongst BSS activity globally, but sits in-between the mean and the 75th percentile. This suggests that users are relatively compliant, with reduced activity levels, but changes in restriction stringency are not always reciprocated. This provides further evidence of the observed spatio-temporal patterns in Section 6.4 in addition to providing some wider context in relation to other BSS activity.

### Table 6.3: Distribution of SMR results across 339 BSS

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.99</td>
<td>-0.55</td>
<td>0.14</td>
<td>0.05</td>
<td>0.67</td>
<td>0.94</td>
</tr>
</tbody>
</table>

## 6.6 Network Analysis

The analysis of transportation systems are a difficult task due to their size, complexities and dynamism. Network analysis methods are a valuable tool in the analysis of such complex systems due to their foundations on graph theory that enable the simplification of such structures into graphs. A graph \( G \) consist of nodes \( N \) that are connected by edges \( E \) that are used to represent the relationship between different entities (Bollobas 1998). Both nodes and edges can contain attributes, such as edges containing the weights that relate to the strength of relationships or distances between entities or nodes containing proprieties on those entities. Edges may also have directions to indicate whether the relationships are unidirectional. Using such a framework, it is possible to quantitatively analyse the relationship within BSS.

With the growing popularity of BSS and the availability of such data, network
analysis methods have provided an excellent framework for the granular analysis of new micromobility modes (Zaltz Austwick et al. 2013; Feng et al. 2020; Saberi et al. 2018; Yang et al. 2022). Within the context of this analysis, network analysis methods have been employed in order to conduct a quantitative comparison of journey activity dynamics in the London BSS during periods of lockdown. In order to do so, a weighted and directed graph structure is created for each study and control period (Table 6.1), separating weekdays and weekends, by transforming journey data in those periods into an OD matrix. In each of these graphs, nodes represent docking stations that are connected by edges that are weighted based on the number of journeys that occur between them.

Each of these graph structures are analysed using two primary methods. Firstly, the graphs are analysed using various network characteristics that enable the quantification of various graph properties. By comparing these measures to those in the associated control period, it is possible to derive a detailed understanding into how these journey dynamics have changes. In addition to comparisons of network properties, the graph structures are analysed using community detection methods in order to detail the underlying structures.

### 6.6.1 Network Characteristics

Network characteristics measure and describe different components within the graph as well as overall status of the structure. In the context of transportation systems, these can be used to measure system efficiency (Jiang 2009), identify urban hubs and centres (Huang et al. 2016), understand socio-economic patterns and activities (Strano et al. 2007) and evaluate the resilience of the network structure to node failure (Wilkinson et al. 2012). Within this analysis, a combination of measures have been calculated in order to conduct a granular comparison of changes in the network dynamics.

Firstly, the number of nodes $N$ and edges $E$ are presented as basic information of the graphs that have been constructed. The number of nodes tends to fluctuate between study and control periods due to the opening and closing of docking stations. For example, in September 2020, 3 additional docking stations were con-
structured along a new ‘Cycleway’ in southeast London (TfL 2020). Other simple network characteristics include the total number of journeys $T$, a simple sum of edge weights, to provide volumetric comparisons of activity, as well as the proportion of self-looping journeys $SL$, referring to trips that start and end at the same docking station. These journeys have been found to be synonymous with leisure journeys (Mateo-Babiano et al. 2016), thus providing an indication of such activity purpose changes. On the other hand, reciprocity $R$ is a measure of the likelihood of nodes in a directed network to be mutually linked. This is a valuable measure in the context of analysing changes in BSS dynamics as reciprocated edges are typically indicative of commuting behaviours due to the mirroring of journeys to and from workplaces. Higher values are therefore more likely to suggest greater commuting behaviours.

In addition to basic network characteristics, popular measures that provide a more detailed understanding of structures are also presented. Firstly, the average node degree $\bar{d}$ is calculated. Node degree is a centrality measure for describing the number of edges that are connected to a node, providing some indication of the importance of a docking station in relation to its connectivity. $\bar{d}$ is complemented by the coefficient of variation of node degree $cv(d)$ which is the ratio of the standard deviation to the mean that provides an indication of the variability in node degree in relation to the mean. Higher levels of $cv(d)$ are therefore indicative of a system where there is greater variation in the connectivity of docking stations within the system.

The average edge weight $\bar{w}$ provides an overview of the mean number of journeys that occur between any given pair of docking stations in the network. Again, this is complemented by calculating the coefficient of variation in edge weight $cv(w)$ in order to assess how this varies across edges. In this case, higher variability in $cv(w)$ are suggestive of an increase in the number of new OD pairs which are likely to occur when there are new users or trip purposes.

Network connectivity $\delta$ is a measure that is typically used to understand the completeness of graph structures, with higher values signalling better and more
connected networks. Since BSS are different from other types of PT due to their function as short distance and first/last mile journeys instead of longer, multi-modal journeys, \( \delta \) should not be interpreted as a measure of resilience in the network, but rather used to identify more heterogeneous cycling behaviours as a result of new OD pairs. This is valuable when the number of \( E \) are not directly comparable because of differences in \( N \) between study periods. Here, instead of employing the simplified measure of connectivity that Yang et al. (2022) presents, measures of edge connectivity are measured to quantify such changes in connectivity, taking into consideration the directed nature of journeys.

Finally, transitivity \((t)\) and assortativity \((a)\) are both measures that provide some indication of a nodes’ tendency to cluster. \( t \) is closely related to the clustering coefficient and helps to quantify the overall probability for the network to have adjacent nodes that are interconnected, thus revealing the existence of tightly connected clusters (Schank and Wagner 2005). It is calculated based on triplets of nodes in a graph, that are connected by two or three edges, the former being identified as an open triplet and the latter a closed triplet. It is important to note that here, the directed nature of the graph is not considered. The relative number of closed triplets are compared to the total number of triplets, as follows (Wasserman and Faust 1994):

\[
T = \frac{N_{CT}}{N_{AT}} \quad (6.1)
\]

Where:

- \( N_{CT} \) = the number of closed triplets in the graph
- \( N_{AT} \) = the number of all (closed and open) triplets in the graph

The \( a \) coefficient is a measure of a node’s preferences to connect with other nodes that are similar (Noldus and Van Mieghem 2015). This provides additional depth in the understanding of edge connections that are not possible with node degree, with higher values indicating nodes of similar degree connecting with one another. In the context of BSS this is likely indicative of hub like docking stations connecting, or when values are negative, may reflect more random connections be-
6.6. Network Analysis

between docking stations of all different sizes and connections. It is defined as (Newman 2002):

\[ a = \frac{M^{-1} \sum_i j_i k_i - [M^{-1} \sum_i \frac{1}{2} (j_i + k_i)]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - [M^{-1} \sum_i \frac{1}{2} (j_i + k_i)]^2} \]  

(6.2)

Where:

- \( M \) = the number of edges in the graph
- \( j_i/k_i \) = the degree of the nodes at the start/end of the \( i \)-th edge

6.6.2 Network Communities

In conjunction with network level measures of clustering such as \( t \) and \( a \), the analysis presents more granular and visual depictions of such communities through the deployment of community detection methods. In network analysis, communities refer to those occurrences of groups of nodes in a graph that are more densely connected internally than to nodes outside (Fortunato 2010). Such methods enable the identification of inherent structures and clusters in the network that can be calculated based on the heterogeneity within the network. In graph structures representative of transportation systems, these dense connections tend to occur between nodes that are in close geographical proximity (Saberi et al. 2018; Yang et al. 2019). Therefore, such analyses have some practical application in planning. For example, community detection methods have been used to infer travel to work areas for the whole of the UK, helping to uncover deeper understanding of national labour markets and mobility (Coombes and Bond 2008).

Much like those traditional clustering methods that are discussed in Section 5.4, the determination of communities within networks are difficult due to their exploratory nature. As such, selecting the most appropriate community detection algorithm is a subjective process that requires a deep understanding of the network and the underlying structures that are desired. Several commonly used algorithms include hierarchical community detection methods (Yin et al. 2015; Reichardt and Bornholdt 2006) or the Girvan-Newman algorithm (Girvan and Newman 2002; Duch and Arenas 2005; Despalatovic et al. 2014) that both iteratively remove edges based on their connectivity such as measures of betweenness. These algorithms are
valuable in enabling the analyst to determine the number of clusters but also add
greater subjectivity to such processes.

As a result, algorithms that quantitatively maximise a quality function are typi-
cally favoured as they remove this element of subjectivity and provide a much more
robust means of identifying such structures in graphs. The most popular methods
aim to maximise modularity, a measure that ‘quantifies the deviation of the internal
link density of the clusters from the density one expects to find within the same
group of vertices in random graphs with the same expected degree sequence of the
network at study’ (Lancichinetti and Fortunato 2011, p. 1). In other words, a graph
with high modularity has dense connections between community members whilst
also having sparse connections with nodes in different communities. Optimising
this value theoretically results in the best possible segmentation of nodes in a given
graph structure.

There are various heuristic approaches that are used to solve such maximisa-
tion function, including greedy algorithms, simulated annealing and spectral op-
timisation (Anderson and Dragi´cevi´c 2020). Among such methods, the Louvain
algorithm is one of the most popular approaches. The algorithm works iteratively,
joining adjacent nodes that increase the modularity score until there are no nodes
that can be joined to increase the score (Blondel et al. 2008). Due to the algorithm’s
popularity, efficiency and application in studies of BSS (Yang et al. 2019; Shi et al.
2019; Borgnat et al. 2011; Borgnat et al. 2013; Munoz-Mendez et al. 2018), the
method was initially adopted to determine communities in each study and control
period in this analysis. Analysing such results, communities appeared to be incon-
sistent and were very difficult to interpret or justify. This is likely a result of the
limitations of modularity maximisation algorithms, as identified in Lancichinetti
and Fortunato (2011), with issues pertaining to their sensitivity and ability to un-
cover very small community structures (Anderson and Dragi´cevi´c 2020), as well as
their vulnerability to resolution restrictions, limited to undirected information and
assuming a process of endogenous network formation (Fortunato and Barthélemy
2007). As a result, for the purposes of this analysis the Infomap algorithm (Rosvall
6.6. Network Analysis

and Bergstrom 2008) has been implemented - a method that has been found to be advantageous in comparison to the Louvain algorithm in the context of BSS (Shi et al. 2019).

The Infomap algorithm is another function maximisation method that builds on the shortfalls of the Louvain methodology, acknowledging that the system structure drives the flow in the system, leading to system-wide interdependencies (Munoz-Mendez et al. 2018). Instead of maximising modularity, the algorithm seeks the optimal community structure through the minimisation of the description code length of a random walk on the network. This is implemented by the map equation (Rosvall and Bergstrom 2008):

\[ L(M) = q \rightarrow H(Q) + \sum_{i=1}^{m} p_i \rightarrow H(P_i) \] (6.3)

Where:

\[ q \rightarrow = \text{the probability that the random walker leaves the current module} \]
\[ p_i \rightarrow = \text{the proportion the walker spends in the respective module} \]
\[ H(Q) = \text{the index codebook entropy} \]
\[ H(P_i) = \text{the module codebook entropy (from Shannon (1948))} \]

In practise, the methodology is very similar in nature to the Louvain method, where the algorithm iteratively connects adjacent nodes that reduce \( L(M) \). It aims to mimic the likely travel sphere of agents within the network, based on the direction and weights of edges. This method has been employed in the context of the London BSS in Munoz-Mendez et al. (2018), where results show significant improvements over the Louvain method, being able to identify physical structures such as Hyde Park and Canary Wharf. Therefore, employing this method in order to compare lockdown periods to their equivalent control periods enable the identification of differences in community structures.

In combination with the implementation of the Infomap community detection algorithm, each docking station’s betweenness centrality has been calculated. Betweenness centrality is a measure of a node’s importance in relation to its influence over the flow of entities in a graph (Wasserman and Faust 1994). It is calculated by
examining the shortest paths between all pairs of nodes in a graph, with a node’s betweenness value increasing the more frequently it lies on the shortest path between other nodes. Although this is typically used in reference to social networks (White and Borgatti 1994) or transportation systems that necessitate the interchange or traversing through nodes (Jordán 2008), it is used in this context to provide an indication of a docking station’s importance in relation to all BSS activity as opposed to singular journeys. This is primarily due to the fact that centrality measures in relation to singular journeys, namely degree centrality which measures the importance of nodes by considering its weighted edges, lacks significant variation in such contexts. Therefore, analysing such results obscure the relative changes in docking station importance between study and control periods. The addition of betweenness centrality when analysing community structures provides another facet in order to be able to identify granular changes in BSS activity dynamics, pinpointing the location of docking stations that become or remain integral during periods of lockdown.

6.6.3 Results

The assortment of global network statistics that were calculated for each lockdown and equivalent control period and are presented in Tables 6.4, 6.5 and 6.6. It is important to note that inter-lockdown comparisons should be interpreted with caution due to the varying duration of each study period. This means that any comparisons between study periods are made by analysing the trends and changes in the magnitude of difference between each study and control period as opposed to making any direct comparisons.

Table 6.4 presents these global network statistics for Lockdown 1 compared to the control period defined in Table 6.1, split between weekdays and weekends. Comparing weekday network statistics there appear to dramatic changes in activity patterns. $T$ indicates there were approximately half as many journeys taken in conjunction with a fall in the number of OD pairs ($E$) by 30.7%. Each docking station also appears to experience a significant decrease in $\bar{d}$. Taking these variables together, we can gather that weekday journeys have experienced a significant hit in total activity levels. There appear to be some slight increases in $cv(w)$, which would
suggest many new OD pairs occurring with less popular docking stations. Although there appears to be a volumetric decrease in activity, journeys within lockdown periods appear more varied in nature. Further evidence to support this is shown by the decrease in $a$, from 0.071 to 0.016. This suggests more discretionary trips tend to be made typically between docking stations that are dissimilar in comparison to the control period, much like $cv(w)$. Decreases in $t$ and $\delta$ provide further evidence for a greater heterogeneous activity as a result of a decline in sub-cluster strength.

In comparison, weekend activity changes are vastly dissimilar to those in the weekday. The primary cause of these differences are likely due to the increase in $T$ by 17% in comparison to the control period. Therefore, unlike those weekday trends, docking stations are more interconnected with over 12,000 additional $E$, increasing the average number of OD pairs per dock by 33.8 ($\bar{d}$). This has been reciprocated by an increase in the connectivity of the network ($\delta$), indicating an increase in heterogenous cycling behaviours. These are also likely presented as self-loop journeys ($SL$) that have been observed to increase from 6% of journeys to 15.4%, indicating an increase in the proportion of leisure activity throughout weekends and weekdays in each lockdown period. Although there is a dramatic increase in weekend activity, these activities appear to be distributed, causing increases in both node ($cv(d)$) and edge variation ($cv(w)$). Although there is a contrast between

---

**Table 6.4:** Global network statistics of Lockdown 1 and associated control period

<table>
<thead>
<tr>
<th>Network Measures</th>
<th>Lockdown 1 Weekday</th>
<th>Lockdown 1 Weekend</th>
<th>Lockdown 1 Weekend</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>791</td>
<td>783</td>
<td>791</td>
</tr>
<tr>
<td>$E$</td>
<td>194,540</td>
<td>134,846</td>
<td>105,519</td>
</tr>
<tr>
<td>$T$</td>
<td>1,115,948</td>
<td>556,306</td>
<td>340,926</td>
</tr>
<tr>
<td>$R$</td>
<td>0.704</td>
<td>0.626</td>
<td>0.564</td>
</tr>
<tr>
<td>$SL$</td>
<td>0.028</td>
<td>0.144</td>
<td>0.060</td>
</tr>
<tr>
<td>$d$</td>
<td>491.88</td>
<td>344.43</td>
<td>266.80</td>
</tr>
<tr>
<td>$cv(d)$</td>
<td>0.404</td>
<td>0.402</td>
<td>0.443</td>
</tr>
<tr>
<td>$w$</td>
<td>5.74</td>
<td>4.13</td>
<td>3.23</td>
</tr>
<tr>
<td>$cv(w)$</td>
<td>2.34</td>
<td>2.93</td>
<td>2.76</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.622</td>
<td>0.440</td>
<td>0.337</td>
</tr>
<tr>
<td>$a$</td>
<td>0.071</td>
<td>0.016</td>
<td>0.055</td>
</tr>
<tr>
<td>$t$</td>
<td>0.635</td>
<td>0.517</td>
<td>0.499</td>
</tr>
</tbody>
</table>
weekday and weekend activity volume, the two periods exhibit a decrease in $a$ and $t$. Since negative values of $a$ indicate disassortiveness, where the important ‘hub’ like docking stations connect to low-degree docking stations, this suggests the heterogenisation of journey activities during weekend lockdown periods.

Table 6.5: Global network statistics of Lockdown 2 and associated control period

<table>
<thead>
<tr>
<th>Network Measures</th>
<th>Lockdown 2 Weekday</th>
<th>Lockdown 2 Weekend</th>
<th>Control 2020</th>
<th>Control 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>785</td>
<td>790</td>
<td>785</td>
<td>789</td>
</tr>
<tr>
<td>$E$</td>
<td>134,147</td>
<td>125,906</td>
<td>59,403</td>
<td>91,119</td>
</tr>
<tr>
<td>$T$</td>
<td>542,280</td>
<td>437,415</td>
<td>127,885</td>
<td>236,539</td>
</tr>
<tr>
<td>$R$</td>
<td>0.634</td>
<td>0.613</td>
<td>0.474</td>
<td>0.505</td>
</tr>
<tr>
<td>$SL$</td>
<td>0.012</td>
<td>0.054</td>
<td>0.036</td>
<td>0.084</td>
</tr>
<tr>
<td>$d$</td>
<td>341.78</td>
<td>318.75</td>
<td>151.35</td>
<td>230.97</td>
</tr>
<tr>
<td>$cv(d)$</td>
<td>0.472</td>
<td>0.401</td>
<td>0.482</td>
<td>0.488</td>
</tr>
<tr>
<td>$\bar{w}$</td>
<td>4.04</td>
<td>3.47</td>
<td>2.15</td>
<td>2.61</td>
</tr>
<tr>
<td>$cv(w)$</td>
<td>1.66</td>
<td>1.69</td>
<td>1.25</td>
<td>1.92</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.435</td>
<td>0.403</td>
<td>0.193</td>
<td>0.293</td>
</tr>
<tr>
<td>$a$</td>
<td>0.089</td>
<td>0.057</td>
<td>0.048</td>
<td>-0.004</td>
</tr>
<tr>
<td>$t$</td>
<td>0.555</td>
<td>0.503</td>
<td>0.384</td>
<td>0.447</td>
</tr>
</tbody>
</table>

When inspecting the network properties for Lockdown 2 in Table 6.5, there are some consistencies that can be found to the first lockdown. Although not as significant, during weekdays, there is a decrease in the number of journeys ($T$) by 19.3%. This fall in volume is complemented with a fall in $E$ and $d$ as well as in $a$ and $t$, again suggesting greater variation in trip distribution. Though this does suggest that lockdown restrictions have impacted and changed activity patterns in Lockdown 2, these deviations are markedly smaller and more muted than those observed in Lockdown 1. This is likely indicative of a fall in compliance with these restrictions as well as schools remaining open during the second lockdown, whilst they were forced to close during the first lockdown. This lack of compliance for the full duration of the lockdown can be observed in Figure 6.1, with a delayed initial response to such restrictions as well as a quick return to pre-lockdown trends.

On the other hand, weekend changes appear to be more pronounced, but following those general trends observed in the weekends in Lockdown 1. This implies that although compliance and use within weekdays in Lockdown 2 are more similar
6.6. Network Analysis

to pre-pandemic dynamics, weekend activity appears to be further diverging from those trends. \( T \) increased by 85% in addition to increases in \( L \) and \( \bar{d} \), over double those observed when compared to Lockdown 1 weekend patterns. These additional journeys also are of a more varied nature, with \( \delta \) showing increases in heterogeneous network connectivity. This is further demonstrated by the fall in \( a \) from 0.048 to -0.004, indicating that those previously high-high connections between hubs have transitioned to suggest high-low connections between hub docking stations to those periphery docking stations. Increases in \( t \) likely signify shorter local journeys being made among docking stations in close proximity, again another indication of greater leisure activity within the BSS.

Table 6.6: Global network statistics of Lockdown 3 and associated control period

<table>
<thead>
<tr>
<th>Network Measures</th>
<th>Lockdown 3 Weekday Control</th>
<th>Lockdown 3 Weekend Control</th>
<th>2020/21</th>
<th>2020/21</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>796</td>
<td>796</td>
<td>794</td>
<td>794</td>
</tr>
<tr>
<td>( E )</td>
<td>215,548</td>
<td>189,660</td>
<td>129,039</td>
<td>162,866</td>
</tr>
<tr>
<td>( T )</td>
<td>1,744,259</td>
<td>1,046,866</td>
<td>491,297</td>
<td>634,592</td>
</tr>
<tr>
<td>( R )</td>
<td>0.739</td>
<td>0.693</td>
<td>0.619</td>
<td>0.622</td>
</tr>
<tr>
<td>( \text{SL} )</td>
<td>0.020</td>
<td>0.069</td>
<td>0.049</td>
<td>0.086</td>
</tr>
<tr>
<td>( \bar{d} )</td>
<td>541.58</td>
<td>477.73</td>
<td>324.22</td>
<td>410.24</td>
</tr>
<tr>
<td>( \text{cv}(d) )</td>
<td>0.402</td>
<td>0.357</td>
<td>0.423</td>
<td>0.411</td>
</tr>
<tr>
<td>( \hat{w} )</td>
<td>8.09</td>
<td>5.52</td>
<td>3.81</td>
<td>3.90</td>
</tr>
<tr>
<td>( \text{cv}(w) )</td>
<td>2.25</td>
<td>2.37</td>
<td>2.44</td>
<td>2.78</td>
</tr>
<tr>
<td>( \delta )</td>
<td>0.680</td>
<td>0.602</td>
<td>0.407</td>
<td>0.517</td>
</tr>
<tr>
<td>( a )</td>
<td>0.073</td>
<td>0.047</td>
<td>0.059</td>
<td>-0.007</td>
</tr>
<tr>
<td>( t )</td>
<td>0.664</td>
<td>0.600</td>
<td>0.545</td>
<td>0.570</td>
</tr>
</tbody>
</table>

The global network statistics for Lockdown 3 in Table 6.6 suggest a divergence in activity patterns from the control period on both weekdays and weekends. Though weekday patterns in Lockdown 2 indicated a convergence towards the control period in comparison to Lockdown 1, Lockdown 3 appears to exhibit a marginal return to those diverging trends observed in Lockdown 1. Although Lockdown 3 details the largest volumetric decrease in \( T \), this is a function of the temporal longevity of this lockdown event. Lockdown 3 is nearly double the length of Lockdown 1, making volumetric comparisons between the study periods inequitable. Considering the changes that are observed in Lockdown 3 to the associated control period,
there is a consistent decrease in all network properties except for $cv(W)$. Decreases in $L$ and $\bar{d}$ are a direct result of the volumetric decrease in $T$, but those decreases in $a$, $t$ and $cv(d)$ indicate increases in heterogenous cycling behaviours. The increase in $cv(w)$ provides further supporting evidence for those increases in heterogenous cycling behaviours with previously less popular docking stations.

The weekend changes that are found in Lockdown 3 appear to exhibit further divergence from those trends observed in Lockdown 2. Although this is likely partially a function of the increased length of Lockdown 3, the changes are indicative of increasing uptake of cycling activity of primarily leisure use, as shown by the increases in $SL$ from below 5% to 8.6%. This is supported by an increase in $E$, by 26%, and $\bar{d}$, signifying volumetric increases, with the associated changes in $a$, $t$ and $\delta$ signifying the varying nature of those new journeys.

Looking at the trends that are observed between lockdown and control periods, we get a general sense that weekdays typically exhibit a reduction in activity whilst weekends exhibit an increase in activity. In both cases, journeys tend to be of a more varied nature, with fewer mass directed activities within the system, representing a fall in commuter journeys and an increase in more heterogeneous activity, indicative of the increasing proportion of leisure and exercise use. To complement and provide additional evidence of such findings, Figure 6.6, 6.7 and 6.8 depict the identification of the community structures as well as changes in docking station betweenness centrality in each of the study and control periods.
Figure 6.6: Santander Cycles docking station community assignment and centrality during Lockdown 1 [a) weekday control, b) weekday study, c) weekend control, d) weekend study]
The maps in Figure 6.6 identify changes in those docking station community assignments as well as in docking station centrality in Lockdown 1. Comparing those changes observed in weekday activity (Figure 6.6a,b), the primary difference when considering community assignment is in the size and location of the four communities. In the control period, the Infomap algorithm is able to clearly identify a community in and around Hyde Park as well as a small community of activity in the Queen Elizabeth Olympic Park. The remaining two communities consist of a smaller western cluster that encompasses the high income area south of Hyde Park, such as Hammersmith, Chelsea and Kensington, down to the southern periphery of the BSS, in addition to the largest community of 523 docking stations that dominates the centre and eastern parts of the BSS. In comparison, the journey activities observed in the lockdown are far more evenly distributed as depicted by communities of similar sizes that divide the BSS into quarters, with a northern, southern, eastern and western community. These communities are also generally separated along the River Thames. Such changes in community allocation are suggestive of journeys that tend to occur locally instead of journeys based on particular purpose, such as the identification of the Hyde Park and Queen Elizabeth Olympic Park communities in the control period clearly indicate journeys of leisure and exercise purposes, whilst the large central community is likely indicative of commuting journeys to and from the CBDs in these areas.

Although the number of communities remain consistent during the weekday, weekend communities observe much more dramatic changes in community structure (Figure 6.6c,d). In the control period, there are six communities that are identified, each of which are associated with distinct parts of the city. For example, Hyde Park community observed in weekdays are similarly present in the weekend, but instead of another park based community, the algorithm is able to identify the presence of a community in Canary Wharf, showing similarities to the same community identified in Munoz-Mendez et al. (2018). This is slightly unusual given that Canary Wharf is the financial centre of the city and therefore an area that is unlikely to observe significant weekend activity, but may be an indication of the local
residents who use the BSS as a means to get around and conduct social activities and errands. The remaining area is segmented relatively evenly into communities of the southwestern, northern, southern and eastern parts of the BSS. During the first lockdown, such community structures break down, making them indistinguishable from the control period apart from the presence of a community in Canary Wharf that extends to include docking stations along the northern edge of the river, up to Tower Bridge. The five communities that separated activities that spans the majority of the BSS have been annihilated into one large community of 678 docking stations. This is likely an indication of the heterogenisation of journeys, as indicated by the disassortivity and increase in the network connectivity identified in Table 6.4, as well as longer journeys identified in Table 6.2.

Studying the changes in docking station centrality, we can identify dramatic changes apparent within the central areas of the BSS, with many of those previously highly central docking stations showing a reduction in their importance across both weekdays and weekends. Although this is the case for most docking stations located within the centre of the system, those docking stations in close proximity to green spaces, such as Hyde Park, and bridges along the Thames remain relatively important. Much like the global network statistics, these results are indicative of much greater variation in journeys, with only those docking stations in close proximity to green spaces and the river remaining central to the BSS dynamics. This echoes the similarities exhibited to the park activity observed in Figure 6.1, providing additional evidence that activities during lockdowns tended to occur close to or within green spaces located in the BSS.

Those changes in community structure and docking station centrality observed in Lockdown 1 are generally transferable to the changes that are also presented in Lockdown 2 (Figure 6.7) and 3 (Figure 6.8). The control periods are the only periods that present slight differences between lockdowns, with periods of enforced lockdown that are (for the most part) identical in community structure presentation. Therefore, for the sake of succinctness only those major differences are detailed.
Figure 6.7: Santander Cycles BSS docking station community assignment and centrality in Lockdown 2 [a) weekday control, b) weekday study, c) weekend control, d) weekend study]
Figure 6.8: Santander Cycles BSS docking station community assignment and centrality in Lockdown 3 [a) weekday control, b) weekday study, c) weekend control, d) weekend study]
In Lockdown 2, the only major difference is in the assimilation of the Hyde Park and southwestern communities into one larger community during weekends (Figure 6.7c). This reemerges in Lockdown 3 and therefore can be attributed to the lack of journey data in Lockdown 2 as a result of its shorter duration. In a similar vein, during control weekdays in Lockdown 3, the algorithm is able to differentiate activities centred in Hyde Park in comparison to those on the northwestern periphery of the system (Figure 6.8a), much akin to the communities in Munoz-Mendez et al. (2018). Again, this is likely an artefact of the lockdown’s duration, containing the largest number of journeys and therefore better enabling the algorithm to identify such structures.

Although such changes are present in the control periods, this does not change the interpretation of the changes in community structure observed during enforced lockdowns. This consistency in the presentation of the underlying structure of activity dynamics indicate their persistence throughout the duration of the pandemic, with no indications of a return to pre-pandemic dynamics. This provides further evidence of those trends observed throughout this chapter, with an identification that journey dynamics are more varied and distributed due to the collapse of strong communities and activities are centred around parks and the River Thames, suggesting a significant decrease in commuting activities that have been met by an increase in the proportion of leisure and exercise activities.

6.7 Discussion

Considering the results collectively, we can begin to identify the impacts of COVID-19 lockdowns on the mobility dynamics within the London BSS. Generally, results depict a significant change in activity during periods of enforced lockdown, not only volumetrically, but also temporally and spatially.

Across the three lockdown periods, we observe that weekdays typically exhibit a significant reduction in the number of journeys, by as much as 50% during the first lockdown, whilst weekends typically exhibit large increases. This is likely a result of shifting BSS use from weekday commuting activities to weekend
leisure and exercise activities in compliance with lockdown restrictions inhibiting the vast majority (except those essential workers) from travelling to work (Lock 2020). This becomes very apparent by looking at Figure 6.4 which illustrates a significant reduction in BSS use during the morning and evening peak commuting hours on weekdays contrasted by a general increase in journeys throughout the afternoon on weekends. This can also be visually interpreted through the generally red nature of H3 maps depicted in Figure 6.5 during the weekday and generally green hexagons on the weekend. We can also identify the spatial irregularities in these volumetric changes, with weekday journeys exhibiting the largest reductions within the centre of the system, an area associated with tourist activity, workplaces and retail areas, and the most prominent increases occurring on weekends, primarily located in close proximity to green spaces.

Although weekday and weekend journey changes are volumetrically opposed, the results of global network statistics and community detection algorithms determine that the journeys during lockdown periods tend to be more heterogeneous in nature as shown by a consistent increase in edge variation \( cv(w) \) across all study periods, in addition to decreases in assortativity \( a \). This has also been visualised via the spatial irregularities in journey changes presented in Figure 6.5, most prominently among those weekday journeys. Increasingly heterogeneous cycling activities are likely a result of increases in leisure and exercise activities due to their less directed nature in comparison to commuting trips. In conjunction with increased journey times (Table 6.2), an increased proportion of self-loop journeys and increased relative centrality of docking stations around parks, there is strong evidence to suggest a significant shift in journey activity from commuter use to leisure and exercise use (Zaltz Austwick et al. 2013; Jain et al. 2018; Mateo-Babiano et al. 2016; Padmanabhan et al. 2021). This shows the flexibility and resilience of the BSS in adapting to user needs, not only facilitating COVID safe essential worker commuting trips during lockdown periods but also adapting to accommodate an increased uptake of BSS use as a means of leisure or exercise activity both on weekdays and weekends.
Though the differing lengths of each lockdown period mean that making direct comparisons between lockdowns are not informative, identifying the differences in the magnitude of change between each lockdown and control period across lockdowns can help to provide some indication of the general trajectory of activity throughout the pandemic. These trends show that weekday declines journeys appear to increase towards the control levels whilst increasing beyond control levels across weekend activities. Across weekday trends, changes in network statistics between the control and lockdown period are closer to 0 for most metrics, indicating a tendency towards activities prior to COVID-19, whilst weekend trends increase in magnitude across the majority of metrics, signifying divergence from pre-pandemic activities. This is corroborated by the results from the descriptive spatial and temporal analysis.

The results present a positive indication for the future utility of the BSS, with some indications of a return of weekday commuter use and increased leisure activity use in the system generally, the demand for the London BSS is likely to increase in the future post-pandemic era. This suggests that COVID-19 has highlighted the valuable utility that BSS provide to urban dwellers, enabling a COVID safe, healthy and environmentally friendly mode of urban mobility that provides great flexibility as a mode of future urban mobility (Palma et al. 2022; Marsden and Docherty 2021). Similar to the fast-tracked implementation of temporary cycling infrastructure during the wake of COVID-19 (O’Malley 2021; Reid 2020), it appears that the pandemic has helped to encourage the use of BSS in London. This is not only a positive for the BSS operation, but this will also have wider benefits to London, helping to move towards the tough emission and congestion targets set out in the 2021 London Plan (GLA 2021b).

6.7.1 Future Research

Future research on the impacts of lockdowns should aim to build on the limitations of these analyses. Whilst the methods enable a holistic understanding of the changes in activity patterns observed within the London BSS, comparisons between lockdown periods rely upon descriptive comparisons of the various methods imple-
mented. This is especially pronounced among comparisons between study periods when interpreting global network statistics due to the lack of consistency in lockdown durations. In order to help facilitate more comparable inter-lockdown results, the study periods could have been aggregated to focus on an average weekday or weekend or simply to choose the most representative day within each period. Such refining of the data would help to facilitate equitable comparisons between each of the study periods and therefore better infer the trends observed across lockdowns.

Future research should also aim to consider built environment factors, especially in relation to the construction of new cycling infrastructure between the study and the control period. Since the availability of cycling infrastructure in BSS has been found to have a significant impact on BSS use (Buck and Buehler 2012; Buehler and Dill 2016; Faghih-Imani et al. 2017; Mateo-Babiano et al. 2016; Médard de Chardon 2019; Pucher and Buehler 2008; Rixey 2013) and COVID-19 fast-tracked the construction of 90 km of new cycling lanes (O’Malley 2021), the increases in activity observed within this period have not been able to separate out these effects. Therefore, it is important to acknowledge how this addition would have likely contributed to the increasing use of BSS by new users throughout the study periods. Such analysis would also help to provide further indications of journey purposes through an empirical understanding of the relationship to other built environment factors.

Finally, the BSS data employed within this analysis are limited in that they do not enable any understanding of BSS users regarding their demographic characteristics. Any demographic information would be very valuable in helping to provide a deeper insight into the impacts based on user characteristics such as gender disparities and equitable access. This could be facilitated through the analysis of BSS membership data, but access to these data require direct collaboration with operators. As a result, such analysis would require membership survey data, much like those studies which currently investigate the demographic characteristics of users (Gavin et al. 2016; Hosford et al. 2018; Winters et al. 2019).
6.7.2 Applications

The analyses in this chapter develop our understanding of the impact of lockdown events on urban mobility. Previous research has primarily focused on the use of qualitative survey data and limited to descriptive inferences on trip volumes and durations. As such, through a variety of analytical techniques, the results provide a holistic, system-scale quantification of changes in journey dynamics within the London Santander Cycles BSS.

The primary outcome of this chapter is the identification that BSS serve as an important means of urban transportation during periods of significant change. The resilience, adaptability and flexibility of BSS are vital in ensuring that essential workers can get to their workplaces to mitigate the breakdown of society. This can be observed by the persistence of community peaks, albeit much more muted, during periods of enforced lockdown in Figure 6.4. The results also emphasise the shift in journey purposes to leisure and exercise use during weekends. This versatility has previously been identified during Tube strike events in the city (Saberi et al. 2018; Yang et al. 2022), but also more recently during times of crises such as in the uptake of the New York City Citi Bike Scheme in the aftermath of Hurricane Ida in September 2022 (Surico 2022a). Therefore, it becomes very apparent that BSS serve as a vital mode of urban transportation that should be more widely adopted in cities that are yet to invest.

Methodologically, the analyses highlight the values of network analyses methods in the identification of system-scale changes within micromobility modes. Due to the availability of OD data and the fact that BSS are closed systems, network analysis techniques are ideal for quantifying the changes in activity dynamics. The comparison of network statistics provide a quick and easily interpretable means of empirical differentiation that go beyond simple descriptive statistics. This is demonstrated by the inferences of greater journey heterogeneity across both weekdays and weekends that were also established by changes in community structure that would not be possible through alternative methodologies. Considering these benefits, network analyses methods should be applied more frequently to analyses of micromo-
bility modes, especially of docked nature, to uncover more nuanced understandings of system-scale journey dynamics.

6.8 Chapter Summary

The analysis presented within this chapter facilitates an in-depth understanding of the medium-term changes in mobility dynamics observed within the London BSS volumetrically, spatially, temporally and structurally. This is made possible through a combination of descriptive spatio-temporal data aggregation techniques in combination with time-series and network analysis methods.

The findings provide valuable insights for policymakers to understand how these unprecedented restrictions on human activity have impacted the mobility patterns within this increasingly popular mode of urban micromobility. Although the majority of the western world reached some stability and return to pre-pandemic conditions, there is a possibility of a new variant which would likely result in the re-adoption of these NPI policies, in addition to novel pandemics of a similar scale in the future. This analysis illustrates the resilience that the BSS provides in adapting to user needs, enabling essential workers to travel to their workplaces and provide vital goods and services and non-essential workers access to alternative, COVID safe forms of mobility for increasing demands in leisure and exercise activities. Therefore, it is important to acknowledge the value of BSS in these urban environments, especially as a mode which also facilitates more sustainable mobility. Future work should aim to consider the addition of built environment factors in addition to user demographic characteristics.

Although such findings are not especially novel, the chapter extends on the existing literature that analyses the impacts of such lockdown events by providing more granular insights on the London BSS. The novelty in the analysis presented lies in the depth and breadth of methods that have been employed, showcasing how they can complement each other in the identification of changes in the dynamics of micromobility modes. This highlights the value of network analysis methods as a very appropriate method for such modes due to the granularity in the data that
are available and the structure of these systems. Employing only openly available BSS data, the methods are also able to provide a unique perspective in showcasing medium-speed, medium-term changes in SD by using data from 2019, signalling their value in being able to identify changes at the system and docking station levels over such time periods.
Chapter 7

Dockless Bicycle Sharing Mobility Trends

Thus far, the analysis presented in this thesis explores long-term, global-scale changes and comparisons of BSS in Chapter 5 and medium-term, system-scale changes in Chapter 6. In both cases, these analyses are centred on third-generation docked BSS. Although these modes still dominate the market, newer fourth-generation dockless BSS have been increasing in popularity since their deployment in 2015, as identified through the analysis of Meddin Bike-sharing World Map in Table 3.6. The advantage of a dockless model over traditional docked BSS lies primarily in their flexibility, being unconstrained to starting and finishing journeys at designated stations within the system (Mátrai and Tóth 2016; Parkes et al. 2013; Fishman 2019; Chen et al. 2020). The successes of the dockless model have since paved the way for the deployment of novel e-scooter sharing schemes that have also seen mass adoption in cities around the world (Schellong et al. 2019).

The growth of dockless BSS has also coincided with the increasing popularity of e-bicycle systems. These pedal-assisted bicycles increase the accessibility of cycling to users who would otherwise avoid using such active modes as it significantly decreases the amount of physical exertion required by the user (Langford et al. 2017). Bourne et al. (2020) conduct a literature review on the effects of e-bicycles on travel behaviour, finding that users were more likely to increase their number of cycling journeys, distance and duration, in addition to the likelihood of a
modal shift away from motorised modes. Collaborative Mobility UK (CoMoUK), a shared-transport charity, created a report of 11 e-BSS in the UK and show similar findings, as well as identifying that such systems encourage new users to begin cycling (CoMoUK 2016). These results highlight the variation in e-bicycle activity that have incited the call for a fifth-generation of BSS to be defined (Guidon et al. 2019).

The intersection of such developments in BSS in recent years have created a new frontier in the literature surrounding their dynamics. Although dockless BSS have the potential to be analysed in much greater detail and can be used to infer trip purposes more accurately due to their dockless nature - enabling users to drop-off bicycles in much closer proximity to their desired destinations - the sensitivity of such data mean that access is very limited. This is highlighted in Section 3.3, where dockless journeys in the hybrid San Francisco Bay Wheels BSS are much more coarse in terms of their spatial resolution. As a result, the analysis of such systems are very limited, especially in those countries and cities that do not mandate the release of these data.

This chapter aims to address this gap in the literature by analysing the spatio-temporal dynamics and built environment factors that are associated with the use of the dockless Uber JUMP e-BSS in London using openly available GPS data that were recorded prior to the obfuscation of the dockless GBFS feed in March 2020 (Section 3.1). Although some efforts were made to identify journey occurrences in the collection of dockless BSS data, as highlighted in Section 3.2.3, the data still contain a significant number of errors that require rigorous cleaning efforts that are detailed in Section 7.2. These cleaned journey OD data provide the foundations for detailed spatio-temporal analyses in Section 7.3 that enable the identification of journey hotspots during different times of day within the service area. These findings are then complimented with statistical analysis of journey destination locations in association with surrounding built environment factors in Section 7.4, providing an insight into the journey purposes at different times of day.

Whilst much of the previous research has relied on surveys and interviews
which are limited to small sample sizes, this research is unique in its deployment of comprehensive GPS data to empirically examine the geospatial and temporal patterns of e-BSS use. This enables a holistic yet granular understanding of how the micromobility system is adopted in London, a city that is yet to be studied in such contexts. The results are able to identify specific locations, times of day and relations to points of interest around the city that see significant activity, providing the tools and evidence to operators and policymakers to ensure its optimisation and appropriate management. The analyses are also able to highlight the value of open GBFS data feeds in being able to identify each journey OD in dockless BSS. Although such data are not available in the same form any more, this showcases the potential in how alternative OD data from such micromobility modes, including e-scooters, can be analysed to derive valuable insights to better understand their role and adoption in cities.

7.1 The JUMP Dockless Bicycle Sharing System in London

The JUMP dockless e-BSS was originally founded as Social Bicycles Inc. (SoBi) in 2010 and set out on a mission to make bicycles more accessible to everyone as an alternative to the car (Canales and Rapier 2020). The company grew with the boom of dockless systems in 2016, differentiating themselves with their impressive e-bicycle design that was praised for its intuitive design and smooth ride compared to other models at the time (Canales and Rapier 2020). Observing the growth of the shared micromobility sector and the success of JUMP, Uber bought JUMP in 2018 for $200 million as a means to expand into the market. Upon its acquisition, the Uber JUMP scheme started expanding rapidly, using its established ride-hailing service as a platform to grow on, deploying in 37 cities across 9 countries (O’Brien 2022b). The micromobility services would be advertised on the incumbent’s application, with over 100 million users (Iqbal 2022), enabling the service to be quickly viewed as a direct alternative to Uber taxi services for short distances.

Although the initial roll-out was relatively successful, with Uber having its first
profitable year in 2018 after going public (Canales and Rapier 2020; Iqbal 2022),
the service came under pressure in subsequent years as internal management is-
sues surfaced highlighting its inefficient operation and unprofitability, leading to its
ultimate demise. In 2020 Uber made a deal with Lime, a growing micromobility op-
erator that were direct competitors to JUMP bicycle and scooter sharing services, to
acquire the business along with a $170 million funding round led by Uber that also
gave them the option to acquire Lime between 2022 and 2024 (Canales and Rapier
2020). This transferred the operations of services to Lime but meant that JUMP
bicycle fleets were scrapped and removed from operation. This caused a public
uproar after numerous news articles surfaced, reporting pictures depicting tens of
thousands of JUMP bicycles piled up and wasting away (Hawkins 2020; Blunden
2020; Shead 2020). Though there were some initial backlash, since this merger
occurred Lime has seen great successes in the market, reporting the first profitable
quarter in 2020 (Musulin 2020), surpassing over 250 million rides, making it the
most used shared micromobility operator (Lime 2021) and being named as one of
the 2021 Time100 most influential companies (Van Houten 2021). Lime is currently
operating in over 200 cities in nearly 30 countries (Lime 2022a) and continues to
operate those previous JUMP e-BSS fleets, including the London scheme.

The London Uber JUMP scheme commenced its operations in May 2019, start-
ing with a trial fleet of 350 bicycles that eventually tripled to over 1,000 bicycles by
the beginning of 2020 (O’Brien 2019a). Being such an early adopter, it became the
second-largest BSS provider in the city (O’Brien 2022b). The scheme was compet-
itive with its pricing, costing £1 to unlock and 12 pence for every minute after the
first five minutes of use; being advertised on the Uber application gave it an edge
over other operators who had to rely on enthusiasts to download their specific appli-
cation (O’Brien 2022b). JUMP joined in a period of fierce competition in London,
where there were a total of eight BSS fighting to win market share (O’Brien 2019b),
which has since dwindled down to five, showing the successes of the scheme and
its longevity in comparison to other operators in the city.

Though bicycle sharing is not specifically regulated in the UK and can legally
operate without a permit, JUMP, in addition to other dockless BSS in London, have agreements with councils that they will not remove their bicycles under the obstruction of the public highway legislation. Therefore, operators enforce use within these areas to not upset other councils that they may hope to come to similar agreements with in the future. During its operation, the JUMP scheme was given permission to operate in four London boroughs: Camden, Hackney, Islington and Kensington and Chelsea (O’Brien 2022b), the extent of which are depicted in Figure 7.1. If bicycles were parked outside of the designated operating areas, users were fined up to £25.

\[\text{Figure 7.1: JUMP e-BSS study region [orange areas depict permitted operating boroughs]}\]

The disjointed and incomplete geographical nature of the permitted operational area was an important issue due to their constraints and hindrances to user activity. For example, the lack of permission in Westminster meant that many desirable cycling locations such as Hyde Park and Regents Park, popular retail, leisure and tourist activity areas such as Soho and Covent Garden as well as many major commuter rail stations such as Victoria and Paddington were inaccessible. Similarly, the City of London and Tower Hamlets were not included in this operating area. The
City of London is especially problematic, due its importance as a major business and financial centre, with the Bank of England headquartered there. In practice, this means that although there is a very low resident population, there are over half a million employees working in the area (City of London 2022), highlighting the immense amount of commuting activity that occurs to the City on a daily basis. Such constraints force users to park on the edge of permitted operating areas and walk to the desired final destination (nullifying the advantages of BSS in solving the first/last mile issue) or ignore such rules and incur a fine. As a result, these geographical constraints are very important to be acknowledged and considered when interpreting results of both spatio-temporal analyses and built environment factors. In order to capture such behaviour, those boroughs adjacent and in-between permitted boroughs have been included within this analysis, as defined in Figure 7.1.

7.2 Identifying Journeys

Data on the JUMP e-BSS was obtained from an openly published GBFS feed between 26th December 2019 and 4th March 2020 - a day prior to the operator halting the feed due to issues relating to privacy concerns in line with the update to GBFS Version 2.0. As detailed in Section 3.2.3, the `free_bike_status` file from the data standard were polled and subject to some initial checks in order to minimise the amount of redundant data that were collected. In this process, the location of each bicycle was checked against its location in the previous poll and if it was observed to move over a distance greater than the equivalent of 11 metres, a new record was created. However, since the analyses in this chapter aims to understand the granular movements of user journeys, it was imperative that the data were filtered and cleaned to ensure they were representative of true user journeys.

It is important to highlight that among the data collected on the JUMP e-BSS, the `name` variable from the `bik` table, detailed in Table 3.5, was used to record the globally unique identifier (GUID) that were included in the `free_bike_status` file. This variable is very similar to the dynamic ID rotation methodology detailed in (Xu et al. 2022), but instead of being randomised continually at regular time intervals,
7.2. Identifying Journeys

the GUID was only randomised after a bicycle was locked. This was unique to the
JUMP feed at the time and provided a great proxy and starting point to identify
valid user journeys. In combination with the chronological ordering of the data by
each bicycle ID to identify any anomalous errors in the data, such as multiple hires
of the same bicycle occurring simultaneously, of the 1.9 million records of detected
bicycle movements in the UCL BSS data collection, removing all those records
between each GUID change resulted in a 97.5% reduction in the total number of
recorded movements.

Once all of these false bicycle movements were removed, journey histories
were reconstructed by linking the OD pairs for each GUID. This enabled the es-
timation of trip attributes such as duration, distance, speed and battery use, in or-
der to process, identify and eliminate non-user journey movements from the data.
Much like docked BSS, these non-user movements are typically synonymous with
operator system management efforts, such as the BRP or removing bicycles that
were damaged and required maintenance. In addition to such operator intervention,
e-BSS present the extra requirement for operator maintenance as a result of the bat-
teries that are used to power the pedal-assistance. Therefore, operators frequently
need to swap those batteries to ensure that they have enough charge to facilitate
future journeys. In this process bicycles are unlocked to free the wheels, which
would inadvertently be counted as a trip. These non-user journeys were identified
and removed by setting threshold values for such measures that were guided by
McKenzie (2020), Shen et al. (2018) and Xu et al. (2019), who have undergone
similar processes in the context of the USA and Singapore.

Firstly, trips that lasted less than one minute or longer than one hour were re-
moved. Excessively short journeys imply an accidental hire or a faulty bicycle being
discovered and those longer than one hour indicate a significant time the bicycle was
likely static, meaning it may have been stored in the operator’s truck for some time
before redistribution or used for multiple journeys within a single hire. Once such
trips were removed, the CycleStreets routing engine was employed in order to be
able to best estimate the likely route the user would have taken between each OD.
CycleStreets is a UK-wide cycling journey planner system that leverages OSM data to construct a graph structure that is used to perform various route planning algorithms that can be optimised based on preferences a cyclist may have (CycleStreets 2022a). The graphs are populated with nodes and edges that contain various attributes that enable the routing engines to determine routes fit for a cyclist. These consider factors such as the length and type of streets, riding surface quality, traffic signals and many more that are taken from OSM. Since cycling is a mode that requires physical exertion it is imperative that these routing algorithms also consider the impacts of topography on route choice. Therefore, the CycleStreets algorithm takes SRTM data (that was also used in Chapter 4) in order to derive the elevation at each node in the graph. The algorithm then calculates the time delay or saving when going up or down a hill respectively, by apply a cycling version of Naismith’s rule (CycleStreets 2022c). William Naismith was a Scottish mountaineer that devised a rule of thumb to help estimate how much longer it would take to walk up a gradient, that has also been found to be useful when estimating cycling journeys (Scarf 2007).

Taking all of these factors into consideration, the CycleStreets routing algorithms can then be used to determine three types of route, namely the fastest, quietest or a balanced route (CycleStreets 2022b). These can be conducted on journeys that are longer than four metres and shorter than 300 km, which are used to further remove such anomalous journeys within the data. In this context, the algorithm has been used in order to best estimate the likely routes that users of the JUMP e-BSS have taken in an effort to validate journey speed and time characteristics. It is well known that cyclists do not tend to take the most efficient route, but rather a complex judgement based on safety and efficiency (Broach et al. 2012; Scott et al. 2021). As a result, the balanced routing algorithm has been employed due to its considerations for both efficiency and safety, being a mix of the fastest and quietest routes (CycleStreets 2022b). The average speed of these journeys were taken into consideration and were used to further identify valid user journeys. Trips with an average speed slower than three kmph were removed as they are considerably
slo\lower{1ex}ower than typical walking speed and unrealistic given the pedal-assisted nature of the e-bicycles. Similarly, trips with an average speed faster than 25kmph were removed as they likely indicated operator redistribution or recall in a truck, given that e-bicycles are legally limited to 25kmph on UK roads.

Finally, trips that ended with more than 20\% increase in battery levels were removed as they indicate battery swapping and maintenance efforts from the operator as opposed to user journeys. 20\% was taken as the threshold value because the exploration of the journey data indicated significant variations in battery levels. Within batteries, battery management systems (BMS) are used to monitor, manage and optimise usage and also utilise such data to estimate the current state-of-charge (SoC) (also referred to as battery level). Although there are a variety of different approaches in calculating SoC, many older and inexpensive methods typically rely on voltage to estimate battery levels (Pop et al. 2005). These are highly inaccurate as changes in temperature, discharge rates and ageing cause fluctuations in such estimates (Pop et al. 2005). Due to the variations in levels observed within the journey data, it can be assumed that JUMP e-bicycle BMS rely on similar methods to determine SoC. As a result, it was necessary to set threshold values for battery level changes between journeys over 20\% to ensure that changes were definitively the result of operator maintenance.

Through this rigorous cleaning process that considers all aspects of journeys to remove non-user movements, 25.6\% of reconstructed journeys were deemed to be valid user journeys which amounted to 61,477 trips. Of these journeys, only those that took place within the JUMPs permitted boroughs and surrounding boroughs were considered, as defined in Figure 7.1. This resulted in a final collection of 58,682 journeys that occurred between 1\textsuperscript{st} January and 3\textsuperscript{rd} March 2020 that have provided the foundations of the analysis of short-term and granular dockless e-BSS dynamics.

\section*{7.2.1 Exploratory Analysis Of JUMP Journeys}

On an average day between 1\textsuperscript{st} January and 3\textsuperscript{rd} March 2020 there were 1,130 JUMP bicycles available to hire on the streets of London. These were used for approxi-
approximately 975 journeys per day, resulting in a mean TDB of 0.81. This indicates the inefficient operation of the system, with a significant proportion of bicycles deployed that were not utilised or remained unused for a great majority of the day. Figure 7.2 illustrates the fluctuation in this utilisation rate (TDB) across the 63 days. Here, we are able to identify that TDB increases throughout the first month of the year and eventually peaks between the 5\textsuperscript{th} and 7\textsuperscript{th} February at 1.22, 1.12 and 1.15 TDB respectively. This is likely due to the seasonality of the study period, wherein the holiday period is drawing to a close with schools and universities that tend to remain closed for the first week of the year. As the gradual return to work occurs, utilisation tends to stabilise and fluctuate around the average TDB. There are also clear cyclical dips in TDB that can be observed on 26\textsuperscript{th} January and 2\textsuperscript{nd}, 9\textsuperscript{th} and 16\textsuperscript{th} February that were all Sundays, indicating fewer journeys as well as the persistence of fleet numbers throughout the weekend.

Altogether, only 19\% days recorded over one TDB. This is significantly lower than the TDB observed in the JUMP system in Washington D.C. (2.27\textsuperscript{1}) (McKenzie 2020) as well as the Smide dockless e-BSS in Zurich (1.5\textsuperscript{2}) (Guidon et al. 2019). This is also significantly lower than the three TDB in the docked London BSS in

\textsuperscript{1}Assuming that all of the 200 e-BSS in the system were available every day
\textsuperscript{2}39,112 journeys made in 122 days when 141 bicycles were available each day
Chapter 4. Due to its size and relative inefficiency, the JUMP e-BSS in London would likely be classified as a small to medium inefficient BSS within the major types of BSS that were identified in Chapter 5.

Although this is the case, it is again important to highlight the temporality of the study duration in comparison to the other studies considered that would have reduced the average utilisation. In addition, the system was less than a year old and still in the process of growing and expanding, with reports showing significant increases in fleet size to 3,000 as of August 2022 (O’Brien 2022a) and ridership over doubling from 2019 to 2020 throughout the duration of the pandemic in comparison to the study period (Lime 2022b).

Figure 7.3 further explores these journeys by aggregating journey ending times to each hour of the day throughout a typical week. Here, the distribution of journeys are indicative of those bimodal commuting patterns during weekdays and extended unimodal leisure and exercise use on weekends that have been observed throughout BSS in this thesis. It is possible to observe the significant reductions in activity that are identified on Sundays in Figure 7.2 in comparison to Saturdays, where early afternoon journeys between 12pm and 3pm exhibit an average of over 600 journeys per hour, whilst only peaking at just over 500 journeys for one hour at 2pm.

Table 7.1 details a descriptive spatial overview of the total frequency of jour-
By counting the total number of journeys and average number of daily journeys that end in each borough, it is possible to gain a coarse understanding of the effects of the segmented and incomplete permitted operating area. The boroughs of Camden and Islington are very clearly the centre of activity, with 74.2% of total journeys. Hackney is the next most active borough but exhibits significantly fewer journeys at 13.8%. Being geographically isolated from the cluster of permitted northern inner London boroughs, Kensington and Chelsea observe similar levels of activity to those boroughs where JUMP is not permitted to operate, such as Tower Hamlets and Westminster. This suggests that in spite of fines to deter such activity, the illusion of a more expansive operating area have meant that the JUMP system users, although considerably less than the core permitted area, were almost equally likely to use the system in these peripheral boroughs.

Table 7.1: Total and average daily journeys that end in each borough across the study area

<table>
<thead>
<tr>
<th>Borough</th>
<th>Total Journeys</th>
<th>Average No. of Daily Journeys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camden</td>
<td>19,974</td>
<td>312.1</td>
</tr>
<tr>
<td>City of London</td>
<td>945</td>
<td>14.8</td>
</tr>
<tr>
<td>Hackney</td>
<td>8,106</td>
<td>126.7</td>
</tr>
<tr>
<td>Islington</td>
<td>23,547</td>
<td>367.9</td>
</tr>
<tr>
<td>Kensington and Chelsea</td>
<td>2,309</td>
<td>36.1</td>
</tr>
<tr>
<td>Tower Hamlets</td>
<td>2,099</td>
<td>32.8</td>
</tr>
<tr>
<td>Westminster</td>
<td>1,702</td>
<td>26.6</td>
</tr>
</tbody>
</table>

7.2.2 Evaluating The Validity And Limitations Of Journeys

Although a thorough cleaning process was conducted to ensure the removal of all non-user movements, it is important to acknowledge the potential pitfalls of these processes. Firstly, the aim of these processes was to remove all non-user movements whilst retaining as many valid user journeys as possible. In order to ensure this, threshold values for the removal of journeys based on their duration, average speed, battery usage and location erred on the side of caution by setting relatively stringent and inflexible limits. Though these are well justified, there are likely to be instances that valid user journeys fail to meet these criteria and are therefore removed from the
7.3 Journey Hotspots

In a bid to understand the granular spatio-temporal patterns of JUMP e-BSS activity, the journey data were initially segmented into distinct time blocks, much akin to the temporal divisions that have been employed in similar analyses (Guidon et al. 2019; Li et al. 2021b). Here, four time blocks have been created and used to identify important differentiations in activity patterns that have been observed throughout this thesis as well as in the exploratory analysis of journeys in Section 7.2.1, enabling a simplified yet granular insight into activity dynamics within the system.

The first two time blocks are named AM Peak Hours and PM Peak Hours in association with the bimodal commuting patterns and have been differentiated due to the distinct in-flow towards CBD areas in the morning and out-flows away from these areas in the evening. AM Peak Hours are defined to be those journeys that take place between 7am and 11am, whilst PM Peak Hours are those journeys between 4pm and 8pm. In each case, these four hour blocks are taken to include the analysis that is presented in this chapter. These may be characterised by journeys that start and end in the same location or longer-term rentals that include multiple legs of a journey.

In an ideal scenario, these processes would have been validated against clean journey data, much like the journey estimation methods conducted in Section 4.1.1. Unfortunately, the lack of such data inhibits such validation processes but the parallels that are found in similar dockless journey data cleaning in: McKenzie (2020), Shen et al. (2018) and Xu et al. (2019), as well as the exploration of the final journey data in Section 7.2.1 help to ease such concerns. The journeys that remain are distributed temporally and spatially in a manner that conform to the patterns observed in other dockless BSS (McKenzie 2018; McKenzie 2020; Guidon et al. 2019; Shen et al. 2018; Xu et al. 2019) as well as mirroring the patterns that are observed in the docked London BSS in Chapter 6. As a result, the subsequent analyses on such journeys can be interpreted with great confidence in their representation of valid user activities.
start and end of each peak, as observed in Figure 7.3. The final two time blocks are named \textit{Off-Peak Hours} and \textit{Weekend All} and are used to capture the activity that occurs outside of these peak commuting times and are differentiated by weekday and weekend respectively. The \textit{Off-Peak Hours} block capture all of the journeys in a weekday that do not occur in the \textit{AM Peak Hours} and \textit{PM Peak Hours}, therefore including journeys throughout the night between 8pm and 7am and also between peaks from 11am to 4pm. The \textit{Weekend All} time block simply take all journeys that occur on a weekend. These activities have been grouped together as activity during these times are significantly reduced in comparison to those of peak commuting hours and thus require greater temporal aggregations in order to capture aggregate trends which is reflected in the similarity in total number of journeys in each time block with an average of 14,671. Table 7.2 provides a descriptive overview of the time blocks that have been used and exhibit the significant reductions that can be observed during these periods, further justifying the larger temporal aggregations. It is also interesting to note that JUMP users typically paid around £2.50 for journeys, matching the price of a single journey within a single zone of the London Underground but being slightly more expensive than the cost of a 24-hour hire in the docked BSS (£2) or a single bus journey (£1.50), exhibiting its competitiveness.

\begin{table}[H]
\centering
\begin{tabular}{|l|c|c|c|c|c|}
\hline
\textbf{Time Block} & \textbf{Avg. Trips per Hour} & \textbf{Avg. Trip Distance (km)} & \textbf{Avg. Trip Duration (mins)} & \textbf{Avg. Trip Speed (kmph)} & \textbf{Avg. Trip Cost (£)} \\
\hline
\textit{AM Peak Hours} (7am-11am) & 3,608.5 & 2.62 & 12.89 & 12.18 & 2.55 \\
\textit{PM Peak Hours} (4pm-8pm) & 3,659.5 & 2.47 & 12.74 & 11.75 & 2.53 \\
\textit{Off-Peak Hours} (8pm-7am + 11am-4pm) & 1,012.4 & 2.22 & 11.76 & 11.64 & 2.53 \\
\textit{Weekend All} (12am-12am) & 588.8 & 2.28 & 12.53 & 11.53 & 2.50 \\
\hline
\end{tabular}
\caption{Average journey characteristics in each time block}
\end{table}
7.3.1 Exploring The Spatial And Temporal Distribution Of Journeys

First, the journey origins and destinations in each time block are aggregated to hexagonal geometries, much like the journeys in the docked BSS in Section 6.4, in order to gain an initial overview of the spatial distribution of bicycles. Due to the dockless nature of the JUMP system, journey locations can be analysed in much greater granularity and as a result it was possible to create bespoke hexagonal aggregations using the \texttt{sf} package in R that enables the specification of edge lengths. In Figures 7.4 and 7.5, journeys have been aggregated to hexagons with edges of approximately 150 metres to maximise the spatial resolution whilst also minimising the number of empty cells across the study area.

\begin{figure}[ht]
\centering
\includegraphics[width=\textwidth]{figure7.4}
\caption{Spatial distribution of JUMP e-BSS journey origins across study region}
\end{figure}

Figure 7.4 depicts the spatial distribution of journey origins. Here, it is possi-
ble to identify the dichotomy that were expected between AM Peak Hours and PM Peak Hours. During AM Peak Hours, large volumes of journeys tend to originate within the middle of the boroughs of Camden and Islington, with some distinct areas exhibiting a high concentration of activity. On the other hand, PM Peak Hours identify that the majority of journeys still tend to occur within these boroughs but are located on the southern borders in close proximity to the City of London. By comparing these trends to journey destinations in Figure 7.5, it becomes very apparent that these journeys are a clear reflection of commuting activities since the distribution of journeys are concentrated in those same areas, albeit reversed.

**Figure 7.5:** Spatial distribution of JUMP e-BSS journey destinations across study region

Unlike the clear commuting dynamics that can be identified during peak hours, although the majority of journeys still do tend to occur within Camden and Islington, the distribution of journey origins and destinations in Off-Peak Hours and Week-
are much more spatially dispersed. This is especially pronounced in the Weekend All time block, which depicts a vast spatial spread of journeys with only one hexagonal cell in Islington with more than 140 journeys. Unlike peak hours, aggregating the data in this way does not enable the identification of popular OD flows between concentrated areas due to their seemingly randomly distributed nature.

Taken collectively, although visualising the data in this way enables the identification of the general commuting flows centred around the City of London, the distributed nature of off-peak journeys makes it difficult to identify particular hotspots of activity. As a result, a more sophisticated and granular approach to determining journey hotspots are established in Section 7.3.2 that employs a bespoke spatial clustering technique to better delineate such areas.

### 7.3.2 Better Ways to Identify Journey Hotspots

As mentioned in Section 5.4.1, among various methods to cluster unlabelled data there are a subset of methods that enable the quantification of geographical groups. Within this analysis, the exact GPS coordinates of each journey’s origin and destination are leveraged, employing a combination of DBSCAN and KDE to accurately delineate the spatial extent of such hotspots across the study area. The analysis is conducted using the four time blocks defined in Section 7.3 to determine how these locations may vary during different times of the day and week.

#### 7.3.2.1 Clustering Point Data

First, much like the clustering of non-geographical data, it is necessary to determine the most appropriate methods to cluster the locations based on the characteristics of the data and the aims of the analysis. Here, the aim is to better identify journey hotspot locations using point-density clustering methods. As depicted in Table 7.1 and Figures 7.4 and 7.5, there was significant heterogeneity in the distribution of journeys across the study area. Even among those central areas that observe the majority of the activity, the irregular spatial extent of journey origins and destinations mean that methods which cluster points homogeneously over a study area are not appropriate as they would only enable the identification of the general areas within
which there are a dense number of points.

DBSCAN is an unsupervised clustering method that uncovers groups in spatial data (Ester et al. 1996). The algorithm searches for a set of data points that are ‘dense’, which is defined using two parameters that are determined by the analyst; the minimum number of points that should be considered as a cluster and the maximum distance between points for them to be considered a cluster, which is more commonly known as \( \epsilon \) (Arribas-Bel et al. 2021; Ester et al. 1996). Using these parameters, the algorithm works iteratively and assigns each point to a cluster if it satisfies the defined criteria, that are otherwise referred to as ‘core’ points, and identifies those points that are on the periphery of these clusters as ‘border’ points. If the points do not meet either of these criteria, they are classified as outliers, or ‘noise’ points, that are not assigned to any clusters. Therefore, unlike other clustering methods that require the analyst to specify the number of groups, DBSCAN relies upon the threshold values set to identify the optimal number of clusters.

KDE is another density-based clustering method that can be used to identify areas in which there is a high concentration of points (Worton 1989; Worton 1995). Unlike DBSCAN that offers the analyst some agency over the identification of clusters, KDE identifies these locations by overlaying a grid on the study area, where the centroids of each grid cell are assigned a density value based on the number of points and their proximity. It does so by placing a symmetrical surface, called the ‘kernel’, over each grid cell where its shape (or method of interpolation) and size (or bandwidth) require analyst specification (Azzalini and Torelli 2007). As Worton (1989, p. 166) illustrates, the optimum bandwidth is typically calculated using the \emph{ad hoc} method which uses a bivariate normal distribution as follows:

\[
\begin{align*}
  h &= \sigma n^{-\frac{1}{6}} \\
  \sigma^2 &= 0.5(var(x) + var(y))
\end{align*}
\]

Where:
\( h \) = the bandwidth
\( n \) = the number of points
\( x/y \) = the coordinates of each point

The advantage of DBSCAN over KDE is in its ability to delineate ‘clusters of any arbitrary shape and size’ that better conforms to points that are not distributed homogeneously throughout a study area (Khan et al. 2014, p. 232). This is achieved by calculating density locally at each point location, as opposed to globally across the study area, such as the continuous estimate of the probability density function by KDE. In addition, as KDE produces a continuous measure of density, the identification of distinct clusters are difficult and subjective as they require the analyst to choose a particular contour set that depict areas of equal probability as the definition of clusters (Seaman and Powell 1996). In such contexts, if the analyst sets the threshold too high in favour of identifying areas where points are densest, the process may fail to identify smaller areas that may be considered local hotspots and conversely if set too low, will fail to identify more granular areas. Therefore, in the context of journey ODs within the JUMP e-BSS, DBSCAN present a more appropriate method to initially determine distinct hotspot areas due to their spatial irregularity across the study area.

As alluded to, although DBSCAN are initially used to define whether journey ODs are part of a dense cluster, these areas can extend over large geographical areas, thus wasting the value of the high spatial resolution of the point data. In order to best leverage the data and pinpoint locations within each hotspot area, KDE is applied to each cluster identified by DBSCAN. This bespoke methodology has been used as the deployment of KDE within an area that has already been defined as being dense is much more appropriate in analysing the particularities in hotspot locations.

Therefore, it was necessary to establish the parameters of DBSCAN that would be used to define the initial dense hotspot areas. Since DBSCAN is a well-established methodology, there are processes that help to establish the most appropriate levels for the data that are trying to be analysed. For example, in the determination of \( \text{epsilon} \), the \( k \)-nearest neighbour (KNN) distance is a commonly used heuristic...
(Campos et al. 2016). However, Everitt et al. (2011) suggest that the method has a low true positive rate for datasets with varying densities, as has been identified in this case, failing to identify their idiosyncrasies. In fact, applying the suggested epsilon from KNN was found to generate a single large cluster since the majority of journeys were disproportionately concentrated in those areas surrounding the City of London, as identified in Figures 7.4 and 7.5.

Due to the heterogeneity in JUMP e-BSS journeys that are further complicated by their segmentation to the four time blocks, it was necessary to finely tune the parameters to complement the data. Through the visual inspection of numerous iterations of parameter values, it was determined that the minimum number of points should be set to 100 and the maximum distance between points set to 150 metres. 100 bicycles was determined to be a reasonable threshold to classify a dense area given the limited fleet size and temporality of the study duration. Similarly, 150 metres was selected as the epsilon value after considering values up to 300 metres as it was assumed that dockless users would be unwilling to walk over larger distances to access bicycles or destinations. Likewise, for KDE, the ac hoc method has been used due to its applicability across areas identified as being dense. These kernels have been placed across 1,000 grid cells that divide each DBSCAN cluster. The consistency in the number of grid cells were elected over proportional values as clusters should be similarly dense and would enable overly sufficient spatial granularity given the limited study area.

The following results depict the hotspots across the four time blocks for both origins and destinations, with colours to distinguish DBSCAN identified clusters. The lighter shades outline the 95% home-ranges and are used to identify the general area of DBSCAN assignments, whilst the darker shades specify the 25% home-ranges to visualise the granular spatial extents of the densest 25% of points. To further explore the role of these hotspot locations, popular OD flows have been visualised. Journeys that start or end in the DBSCAN clusters have been spatially aggregated to the same 150 metre hexagonal grids used to depict journey distributions in Figures 7.4 and 7.5 in order to simplify and clarify visualisation efforts in
addition to mitigating ethical privacy issues concerning the identification of individuals. Similarly, aggregated flows with less than 20 journeys are not displayed to further ensure anonymity. The opacity of lines are used to depict the strength of flows, where bolder lines show more journeys occurring between the centre of each cluster and a given hexagonal cell. Visualising the hotspots in this manner is valuable, not only in identifying granular spatial extents within each cluster, but also the areas users most typically travel to or from which provides additional details that can be analysed and interpreted.

7.3.3 Results

To supplement the figures presented here that include popular journey OD flows that can obscure the centres of hotspot locations, Appendix D contains visualisations of hotspots without flows. These should be used to further inspect the granular spatial extents of hotspot locations within each cluster that can be used to identify specific street segments where the majority of journey origin and destinations occur. These locations should be used by policymakers and operators to aid users by providing dedicated parking locations to reduce congestion along pavements.

7.3.3.1 Origin Hotspots

A large proportion of journeys during AM Peak Hours (Figure 7.6a) tended to flow south towards the edges of City of London and Westminster boroughs, particularly from the areas surrounding well-connected train stations such as King’s Cross St Pancras, Angel and Highbury & Islington. Many of the users from the Kings Cross area (red), however, appeared to have travelled northwards as well, into an area where a university campus is located. Flows from the smaller-sized clusters (on the southern and eastern border of Islington) were not presented, indicating that there were no particular destination locations with strong links.

On the contrary, during PM Peak Hours (Figure 7.6b), the majority of the flows were going north-bound, most notably back towards Kings Cross and Angel Stations. All clusters at the southern border of Camden and Islington, which are predominantly commercial and employment districts, were strongly linked with
the Kings Cross area. There were also some substantial flows originating from the southern areas between Holborn and Shoreditch into Highbury & Islington Station. Flows from the Shoreditch area were particularly distinct, reaching destinations in Hackney and northern Islington, showing the occurrence of regular, longer distance journeys from this area. Here, the presence of clusters around the major train stations persisted.

As Figure 7.6c depicts, many trips were made locally during Off-Peak Hours. This is clear from the quiet flow map with the exception of the green and red clusters in Camden. The former cluster, located around the West End, London’s major late night entertainment district, had a strong link with Old Street Station. The latter cluster, located around Kings Cross, had links to destinations in all directions, such as Euston Square Station to the west and Angel Station to the east. The spatial distributions of the clusters were almost identical to that of PM Peak Hours, albeit that
the clusters in southern Islington were dispersed into smaller patches. Moreover, a cluster south of Angel Station in Clerkenwell coincided with the premises of a university.

Similarly, most trips were made locally on weekends (Figure 7.6d). Here, the area around Angel had strong links with distant destinations, going as far as Finsbury Park Station to the north and near Tottenham Court Road Station to the south. Most of the trips were taken northbound with few going into inner-London. As indicated by the more prominent line, there was a particularly strong flow between Angel and Highbury & Islington Stations. Given the similar bicycle counts across all time blocks, the usage during Weekend All is by far the most spatially dispersed. This is clear from the few clusters identified, which tended to be close to leisure destinations. Three out of seven clusters were located near King’s Cross Central, a recently redeveloped major retail and leisure destination. Similarly, just north of Angel station is a shopping centre. Furthermore, all the other clusters were in close proximity to green spaces with cycle routes, including Highbury Fields Park to the north of Highbury & Islington Station, Finsbury Park to the north of Finsbury Park Station and Regents Canal just below the easternmost cluster. Interestingly, no cluster was identified in the periphery of the City of London CBD.

7.3.3.2 Destination Hotposts

Here, the end-point of each line represents the origin of the journeys that were made into the popular destination area. Overall, the spatial distributions of popular trip origins and destinations during AM Peak Hours and PM Peak Hours are inverted with respect to each other, whilst they are homogenous on Off-Peak Hours and Weekend All.

It is clear from Figure 7.7a that during AM Peak Hours, many JUMP users who ended their journeys around the southern border of Camden, Islington and Hackney travelled from the north. Furthermore, within the northernmost red cluster around Finsbury Park Station, a strong link was found with a particular eastern origin in Hackney. This may be indicative of common first-mile journeys occurring where the operator has relocated the bikes to maximise the utility rate during these times.
For PM Peak Hours, as depicted in Figure 7.7b, the locations of the destination clusters were almost identical to that of the origin clusters depicted in AM Peak Hours (Figure 7.6a). Despite the presence of additional popular destination hotspots towards the centre of the city, such as Farrington and Tottenham Court Road Station clusters, the clusters persisted to be generally smaller spatial extents that formed around train stations. Many of the journeys that originated at the edge of City of London flowed towards Angel Station. Furthermore, Highbury & Islington Station had a strong link with the area around Angel Station, as depicted by the bold line between them.

Figure 7.7c illustrates the popular trip destinations during weekdays intra-peak hours, where the journey flows generally reflected the findings from Figure 7.6c. Many trips that originated in the West End ended at Old Street Station, whilst some shorter journey flows were also identified around King’s Cross. However, due to
the lack of flows that have been visualised, in combination with the short nature of lines that are presented reflect the heterogeneity identified during Off-Peak Hours and those that occurred commonly tended to be short, local journeys.

Similarly, on the weekend flow map (Figure 7.7d), the lines depicted from the majority of destination clusters were short, indicating that the areas of origin with strong linkage were particularly local. Nonetheless, many trips that ended at Angel Station came from all directions, mostly from the north and east. There were moderate links from the area to the south, such as origins near Tottenham Court Road Station. Furthermore, a particularly strong link is illustrated with Highbury & Islington Station. This was the time block with the fewest number of clusters identified by DBSCAN, indicating that the JUMP bicycle usage was the most dispersed on the weekends.

### 7.4 Inferring Dominant Trip Purposes

Many studies have sought to understand the determinants of BSS activity by identifying the trends in points of interest that occur within close proximity to journey locations. Unlike docked BSS, dockless BSS enable much more granular insights into these relationships due to their unrestricted nature that allow users to finish journeys right next to their intended destinations. Much like the studies that have been conducted in the past (Sections 2.2.3 and 4.1.7), this analysis employs a regression model in order to quantify these relationships.

To leverage the granularity of these data, this analysis starts by constructing a multilevel data structure in Section 7.4.1, identifying individual, small area points of interest within the vicinity of journey destinations in addition to contextual, larger area demographic characteristics. This structure enables the identification of both local built environment factors that are associated with these locations whilst controlling for broader population distributions that are likely to influence user destinations. Section 7.4.2 will then detail the specification of a zero-inflated multilevel negative binomial regression that has been operationalised to take advantage of the data structure to infer the relationships between these variables coupled with inter-
action effects to each of the four time blocks defined in Section 7.3. Constructing
the regression model in this way enables the differentiation in the magnitude of re-
lationships with built environment factors at different times of the day and week,
therefore allowing inferences of potential changes in trip purpose. Subsequently,
Section 7.4.3 outlines the results of such analysis and their appropriate interpreta-
tion.

7.4.1 Constructing a Multilevel Data Structure

This analysis is unique from the past approaches in three ways: the spatial aggre-
gations that are used to identify local points of interest (built environment factors)
are much smaller, journey destination locations are used rather than origin points
and the number of journeys were analysed using interaction effects with four time
blocks.

Previous studies of built environment factors on dockless BSS typically ag-
gregate points of interest to coarse geographical resolutions, such as predefined
large area neighbourhoods in Mooney et al. (2019), 300 metre grids in Guidon et
al. (2019) and Shen et al. (2018) and 500 metre grids in Xu et al. (2019). Unlike
these studies, the analysis that is presented employs a much more granular 50 me-
tre radius hexagonal grid. This has been constructed using the same methodology
employed throughout this chapter and has been used again to minimise issues sur-
rounding ecological fallacy. Since the dockless nature of the JUMP e-BSS provides
great flexibility in choosing where to end journeys, users are highly likely to park
as close to their desired destinations as possible to minimise walking distances. As
a result, it is safe to assume that users will finish journeys within 100 metres of
their intended destinations, an assumption that is corroborated by studies that show
this reduction in walking distance compared to docked BSS (Pal and Zhang 2017).
Thus, justifying the appropriateness of the 50 metre radius hexagonal grids that have
been employed.

Secondly, contrary to Guidon et al. (2019) and Shen et al. (2018), journey des-
tinations have been used to determine relationships with built environment factors
as they better enable inferences of trip purpose. Whilst journey origin locations are
fixed, requiring users to incur the cost of locating and travelling to bicycles, journey
destinations are much more flexible and informed by user choices that are highly
likely to be in close association with their planned purpose. Similarly, origin loca-
tions are typically indicative of where previous journeys have ended or the results of
operator redistribution that were observed to constitute a significant proportion of
activity given the removal of nearly 75% of the journey records detailed in Section
7.2, thus subject to bias led by the compound treatment effect.

Finally, the number of bicycles dropped-off in each hexagonal grid cell were
counted within each of the four defined time blocks. As a result, interaction terms
could be included in the model between these time block dummies and the built
environment factors. In doing so, the temporal intricacies of relationships to built
environment factors can be determined. This is highly valuable and extends on the
methods that are presented in Guidon et al. (2019) by integrating the temporality
into a single model rather than disaggregating them, enabling a holistic and statisti-
cally robust means of comparing the impacts over time.

The built environment factors that are aggregated to the hexagonal grid cells
are obtained from OSM due to its accessibility and completeness in highly popu-
lated cities in the Global North (Girres and Touya 2010; Haklay 2010; Zielstra and
Zipf 2010) and their dominant use across studies that analyse relationships to built
environment factors on BSS. The names of each built environment variable and
their corresponding OSM feature names and keys are detailed in Table 7.3. Across
the study region, the number of shops, food/drink places (sustenance amenities) and
tourist attractions were counted for each hexagonal cell, creating continuous vari-
ables across the study region. Whilst shops are used to indicate the presence of
businesses that have stocked goods for sale, sustenance facilities are used to iden-
tify locations where individuals can go out to eat or drink. The tourist variable has
been specified to only include major attractions such as galleries, museums and at-
tractions, excluding touristic accommodation such as hotel and hostels due to their
ubiquity across the study region and to better infer journeys to touristic activities as
opposed to their temporary residences. In addition, the presence of a train station,
bus stop, park, university building or cycle lane were identified as a binary dummy variable as these were identified using polygons and lines as opposed to individual points. In order to gain a more accurate understanding of the distribution of cycling infrastructure, their locations have been sourced from a 2018 TfL dataset to appropriately reflect their availability during the study period given the heavy investment that cycling infrastructure received since the start of the pandemic.

Table 7.3: OSM feature keys and values used to identify built environment factors

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>OSM Feature Key</th>
<th>OSM Feature Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shops</td>
<td>shop</td>
<td>all</td>
</tr>
<tr>
<td>Sustenance</td>
<td>amenity</td>
<td>pub, bar, restaurant, fast_food, café, food_court, ice_cream</td>
</tr>
<tr>
<td>Tourism</td>
<td>tourism</td>
<td>artwork, attraction, gallery, museum, zoo, viewpoint</td>
</tr>
<tr>
<td>Train</td>
<td>railway building</td>
<td>subway_entrance, train_station</td>
</tr>
<tr>
<td>Bus</td>
<td>highway</td>
<td>bus_stop</td>
</tr>
<tr>
<td>Park</td>
<td>leisure</td>
<td>park</td>
</tr>
<tr>
<td>University</td>
<td>building</td>
<td>university</td>
</tr>
</tbody>
</table>

In addition, population distribution characteristics that are unattainable from OSM have been gathered from the datasets detailed in Table 7.4 to provide further contextual information, much like those in the global classification of BSS in Chapter 5. First, the distribution of residential populations are determined by gathering Middle layer Super Output Area (MSOA) level data across the study area. MSOAs are statistical geographical units in the UK that are created to have a relatively consistent number of residents and households for comparative purposes, with an average of 8,288 residents in the 2020 mid-year population estimates. They are derived from the 2011 Census and provide an annual update to population characteristics by accounting for the number of births, deaths, international and internal migrations.

Although measuring population distribution in this way provides a good indication of residential night-time population density, they lack inferences on day-time workplace population distributions. In this pre-pandemic period WFH was a very uncommon phenomena and therefore meant that population distributions during the
7.4. Inferring Dominant Trip Purposes

day were significantly different and are much more heterogeneous due to the concentration of workplace activities in CBDs. Therefore, creating a separate variable for these population distributions are valuable in controlling for the effects of workplaces and residential areas on journey purposes. In order to do so, employee estimates in workplace zone geographies have been leveraged to provide an accurate indication of such distributions. Workplace zones are bespoke geographical units that were created to better reflect the distribution of workplaces and are constructed based on Output Area (OA) geographies. OAs are the smallest statistical geographical unit in the UK and are nested within MSOAs, constructed in a similar manner to contain a comparable number of residents and households using postcode building blocks. Although they are consistent for residential populations, OAs do not reflect the distribution of workplaces and employees appropriately, given that over 75% have fewer than 100 workers. In order to circumvent these issues, OAs have been modified to have a minimum of three workplace postcodes and a working population of at least 200, merging units that were under these thresholds and splitting OAs that contained more than 625 workers (ONS 2014).

Table 7.4: Additional data sources used in regression analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Spatial Resolution</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>Official source of population sizes inbetween censuses</td>
<td>MSOA</td>
<td>Office for National Statistics (2020)</td>
</tr>
<tr>
<td>Employee</td>
<td>Workplace and employee estimations using the Inter-Departmental Business Register</td>
<td>Workplace Zones</td>
<td>Office for National Statistics (2015)</td>
</tr>
<tr>
<td>Cycle Lane</td>
<td>Location of cycling lanes and tracks</td>
<td>Linestring JSON</td>
<td>Transport for London (2018)</td>
</tr>
</tbody>
</table>

In order to create a consistent spatial resolution for population distribution characteristics, workplace zone employee estimates were aggregated to the MSOA level. This has been adopted over more granular spatial units due to the constraints of workplace zone units that do not align with any other, more granular geographies. Therefore, MSOA level population attributes were the smallest spatial resolution
that would enable the identification of these population distribution characteristics over the study area and enable the model to account for them. For both measures, the estimated number of individuals were aggregated proportional to the area of each MSOA. This provides a variable that indicates the number of people per kilometre squared, better reflecting a measure of population density. These values have been assigned to each hexagonal cell in the study area based on their centroids. This is valuable as it enables an indication of the estimated population during the day and night for each cell that are employed in the model to control for the differences in population distribution between time blocks.

Table 7.5 provides a descriptive overview of the resulting data structure across all time blocks that have been used as the basis for the regression analysis. There were a total of 190 MSOAs and 17,699 hexagons across the study area, resulting in a total of 70,796 observations as a result of the time block aggregations. This includes both continuous and binary variables. In spite of the large proportion of zero observations across the study area, it is important to remember that the purpose of this analysis is to best infer trip purposes. Different levels of spatial aggregation were tested and were found to make inferences of purposes more difficult due to the issues surrounding ecological fallacy, where larger areas would include multiple potential purposes that were difficult to untangle.

Table 7.5: Descriptive statistics of variables included in regression model

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Min.</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Max.</th>
<th>Proportion of Zeros</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journeys</td>
<td>0</td>
<td>0.83</td>
<td>2.36</td>
<td>243</td>
<td>75.9%</td>
</tr>
<tr>
<td>Shops</td>
<td>0</td>
<td>0.45</td>
<td>1.72</td>
<td>48</td>
<td>85.9%</td>
</tr>
<tr>
<td>Sustenance</td>
<td>0</td>
<td>0.33</td>
<td>1.06</td>
<td>18</td>
<td>85.3%</td>
</tr>
<tr>
<td>Tourism</td>
<td>0</td>
<td>0.05</td>
<td>0.27</td>
<td>8</td>
<td>96.1%</td>
</tr>
<tr>
<td>Population</td>
<td>0</td>
<td>13,879</td>
<td>6,273</td>
<td>29,450</td>
<td>1.0%</td>
</tr>
<tr>
<td>Employees</td>
<td>0</td>
<td>16,650</td>
<td>28,475</td>
<td>125,204</td>
<td>2.7%</td>
</tr>
<tr>
<td>Train</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>97.7%</td>
</tr>
<tr>
<td>Bus</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>87.8%</td>
</tr>
<tr>
<td>Park</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>75.1%</td>
</tr>
<tr>
<td>University</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>98.8%</td>
</tr>
<tr>
<td>Cycle Lane</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>72.5%</td>
</tr>
</tbody>
</table>
7.4.2 Model Specification

The data structure raised several challenges regarding the model specification due to the hierarchical nature of spatial units and the considerable proportion of zero values. This can be observed across the built environment factors in Table 7.5, but most importantly in the outcome variable in Figure 7.8 that depicts the zero-inflated nature of hexagonal journey destination counts. Given the significant proportion of zero counts, the figure struggles to depict cell counts much beyond 10. Though this is the case, it is imperative to note that there are 1,140 hexagons with more than 10 journeys drop-offs, accounting for over a third (21,448) of total journeys, highlighting the importance of these infrequent but concentrated locations. This means that simple linear regression methods are unsuitable due to issues surrounding overdispersion, meaning that variable distributions do not meet the model assumptions (Hilbe 2011). If employed, this would cause over estimations of desired regression parameters, leading to biased standard errors and p-values, which in turn would produce spurious significant coefficients (Hilbe 2011). Therefore, to overcome this data artefact, a generalised regression model, namely a zero-inflated multilevel negative binomial regression was conducted, allowing for both levels to be included in a single model whilst accounting for the overdispersion (Gelman and Hill 2006). Accordingly, varying-intercept and varying-slope model with interaction effects between the time blocks and geographic attributes were fitted.

The proposed model is appropriate for this research context because it is ideal for dealing with spatially-referenced count data in assessing hexagonal-level patterns with a significant proportion of zeros (Agarwal et al. 2002; Atkins and Gallop 2007). Within zero-inflated models, there are two types of zero values; true zeros and excess zeros. True zeros are those that we wish to analyse within our model, whilst excess zeros are those that are erroneous and should be appropriately controlled for. For example, when investigating the distribution of income among 18 to 25 year-olds there is likely to be a high proportion of individuals who earn nothing. Although this is the case, there are two primary explanations for these phenomena. Firstly, a significant proportion of this population are likely to be students, thus un-
able to get a job given that they are still in full-time education. On the other hand, there are also likely to be individuals who are not students that are unemployed. Therefore, if a researcher were trying to understand the determinants of income among this population it is paramount that they control the excess zeros that are a result of the students within the population modelling the relationship to true zero appropriately.

In the context of this analysis, the excess number of zero counts within these data are caused as a result of the study area and the operational limitations of the JUMP e-BSS. As highlighted in Figure 7.1, the study area includes non-permitted operational boroughs. Although users are penalised for such activity, removing these journey records from the data would obscure the true nature of mobility dynamics within the system and would be the equivalent of removing Kensington and Chelsea (a permitted borough) given the frequency of journeys that are observed (Table 7.1). As a result, the inclusion of these non-permitted boroughs within the study area are a significant contributing factor to the number of excess zero counts in hexagons observed that need to be controlled for. Similarly, within the operational area, JUMP users were penalised for ending in no-parking zones that are
marked in the application, including hospitals, stadiums, canals, Royal Parks and protected areas (Katwala 2019). Again, this constraint on user activity further inflated the number of zero observations within these data. For these reasons, it is paramount that our model appropriately control for these excess zeros so that the relationship between built environment factors and journey destinations are correctly determined.

Zero-inflated negative binomial regression models have been applied and operationalised across a number of epidemiological and criminological studies to predict the burden of disease or crime in an area (Chen et al. 2018; Jafarzadeh et al. 2014; Luo et al. 2022; Park 2015; Siegel et al. 2020; Simons et al. 2006). Given their spatial infrequency and heterogeneity, they share similar distributions to those observed in the context of this analysis. The primary outcome measure for this frequentist approach is the relative risk ratio (RR) that is used to quantify whether there is an increase or decrease in the level of disease or crime in an area, whereby, an area could have a higher or lower risk than the average risk in the standard population. This concept from spatial epidemiology and environmental criminology has been employed in this study, whereby RR are estimated as measures of the occurrence of JUMP e-BSS journey drop-offs for each built environment factor.

RR are estimated where an RR of 1 is deemed as a null value (i.e. no effect), and a RR that is less than 1 means a low number of journey drop-offs in close proximity to those built environment factors. Conversely, if a built environment factor has a RR greater than 1, then the journey drop-off rate is higher than the standard hexagon that does not contain that point of interest. Note that built environment factors with RR greater than 2 or 3 and so on, are deemed as having two times or three times and so on, higher journey drop-off rates respectively. The RR reported for built environment factors are deemed statistically significant when the null value of one is excluded from its corresponding 95% credible intervals.
The statistical formulation of the model is defined as follows:

\[ y_{i,j} \sim \text{Pois}(\theta_{i,j}, \kappa) \]

\[ \log(\theta_{i,j}) = \alpha_{j[i]} + \sum_{t=1}^{3} \beta_{j[i]t} T_{ti} + \sum_{h=0}^{8} \sum_{t=0}^{3} \beta_{j[i]th} T_{ti} H_{hi} \]  

(7.1)

In addition to Equation 7.1, the group-level errors are described in Equation 7.2 and 7.3:

\[ \alpha_j = \gamma_{00}^\alpha + \sum_{k=1}^{2} \gamma_{0k}^\alpha G_k + \eta_j^\alpha \]  

(7.2)

\[ \beta_j = \gamma_{10}^\beta + \sum_{k=1}^{2} \gamma_{1k}^\beta G_k + \eta_j^\beta \]  

(7.3)

In Equation 7.1, \( \theta_{i,j}, \kappa \) is a Poisson parameter equal to the expected count of bicycle drop-offs in the \( i \)th hexagon \( (i = 1, 2, 3, \ldots, 70796) \) in the \( j \)th MSOA \( (j = 1, 2, 3, \ldots, 190) \) within the study area. Here, the logarithm is used as the link function for our \( \theta_{i,j}, \kappa \), where \( \kappa \) is the zero-inflated parameter that accounts for the overdispersion in the data by considering whether the \( i \)th hexagon is located in non-permitted borough or a no-parking zone. In Equation 7.1, the level-1 predictors are denoted as \( T_t, H_h \), which in turn corresponds to the \( t \)th time block dummy indicator \( (t = 0, 1, 2, 3) \) and \( h \)th hexagonal-level characteristic \( (h = 1, 2, 3, \ldots, 8) \), respectively; and the interaction is simply the joint production of \( T_t \) and \( H_h \) (i.e., \( T_t H_h \)). Note that by forcing \( T_0 = 1 \) when \( t = 0 \), this formula expands the summations for the main effects. This means that the regression coefficient \( \beta_{j[i]} \) (our primary interests) would reflect the overall relationship between the \( h \)th hexagon-level variable and bicycle count in the reference class (i.e., \( AM \) Peak Hour) in MSOA \( j \) where hexagon \( i \) is located. When \( t \neq 0 \), the summation also produces the interaction effects between the \( t \)th time block dummy indicator and the \( h \)th hexagonal-level variable, where \( \beta_{j[i]} \) would indicate the overall effect at the \( t \)th time block in MSOA \( j \) where hexagon \( i \) is located. In order to derive the relative RR, all \( \beta_{j[i]} \) coefficients are hence exponentiated.
In addition, Equations 7.2 and 7.3 are two separate models for expressing the varying-intercept and varying-slope components of the formulation in Equation 7.1. In Equation 7.2, the parameter $\alpha_j$ is the intercept for each MSOA, which contains a common component $\gamma_{00}^\alpha$, a random intercept $\eta_j^\alpha$ and level-2 predictors $G_k (k = 1,2)$. It represents the average count of bicycles dropped-off at the $j^{th}$ MSOA within the study area. In Equation 7.3, the term $\beta_j$ is slope for each MSOA, composed of a common component $\gamma_{10}^\beta$, a random intercept $\eta_j^\beta$ and level-2 predictors $G_k$. Accordingly, $\gamma_{00}^\alpha$ and $\gamma_{10}^\beta$ are regression coefficients, which, in turn can be utilised to assess the average impact of the MSOA-level characteristics on bicycle counts within the study area.

The model is specified in this way because it allows the effects of built environments on JUMP bicycle usage to differ across the time blocks. The purpose of dockless BSS was expected to differ depending on the time period, as has been identified in previous studies (Guidon et al. 2019; Xu et al. 2019) and the spatio-temporal analyses in Section 7.3. Therefore, interaction effects would provide a powerful tool to evaluate these findings in the context of e-BSS in London. Furthermore, the model ensured that the zero-inflated nature of these data were properly accounted for by identifying whether hexagon was located in those areas where activity were penalised. This final model specification is presented after rigorous consideration of alternative approaches and formulations that were found to be less appropriate given the distinctive nature of these data.

### 7.4.3 Relationships to the Built Environment

The model results are reported in Table 7.6 and describe the relationships between each built environment factor and the number of JUMP e-BSS drop-offs, as well as how these differ between different time blocks in comparison to AM Peak Hours (i.e. the reference category). Exploring the model holistically, it appears as though the vast majority of built environment factors are significantly associated with journey destinations but these effects differ both in direction and magnitude.

First, shops appear to have the smallest positive association to journey destinations. For every additional shop within a hexagon, there is expected to be a
Table 7.6: Zero-inflated multilevel negative binomial regression results

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>RR</th>
<th>(95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.04</td>
<td>(0.02 - 0.09)</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>Time block: PM Peak Hours</td>
<td>2.09</td>
<td>(1.79 - 2.44)</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>Time block: Off-Peak Hours</td>
<td>2.43</td>
<td>(2.11 - 2.79)</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>Time block: Weekend All</td>
<td>2.04</td>
<td>(1.71 - 2.44)</td>
<td>&lt; 0.001 ***</td>
</tr>
</tbody>
</table>

**Hexagon-level Characteristics**

| Shops                       | 1.08 | (1.05 - 1.10)    | < 0.001 *** |
| Shops × PM Peak Hours       | 1.00 | (0.98 - 1.04)    | 0.743     |
| Shops × Off-Peak Hours      | 1.00 | (0.97 - 1.03)    | 0.904     |
| Shops × Weekend All         | 1.02 | (0.98 - 1.05)    | 0.333     |
| Sustenance                  | 1.16 | (1.12 - 1.20)    | < 0.001 *** |
| Sustenance × PM Peak Hours  | 0.97 | (0.92 - 1.02)    | 0.192     |
| Sustenance × Off-Peak Hours | 0.98 | (0.94 - 1.03)    | 0.503     |
| Sustenance × Weekend All    | 0.99 | (0.94 - 1.04)    | 0.587     |
| Tourism                     | 1.37 | (1.21 - 1.56)    | < 0.001 *** |
| Tourism × PM Peak Hours     | 0.85 | (0.71 - 1.02)    | 0.078     |
| Tourism × Off-Peak Hours    | 0.86 | (0.72 - 1.03)    | 0.101     |
| Tourism × Weekend All       | 0.85 | (0.71 - 1.02)    | 0.083     |
| Train                       | 3.92 | (3.27 - 4.70)    | < 0.001 *** |
| Train × PM Peak Hour        | 0.67 | (0.52 - 0.87)    | 0.002 **  |
| Train × Off-Peak Hour       | 0.63 | (0.49 - 0.81)    | < 0.001 *** |
| Train × Weekend All         | 0.70 | (0.54 - 0.90)    | 0.005 **  |
| Bus                         | 1.75 | (1.59 - 1.93)    | < 0.001 *** |
| Bus × PM Peak Hours         | 0.97 | (0.85 - 1.11)    | 0.693     |
| Bus × Off-Peak Hours        | 0.86 | (0.76 - 0.99)    | 0.031 *   |
| Bus × Weekend All           | 0.96 | (0.84 - 1.10)    | 0.565     |
| Park                        | 0.98 | (0.89 - 1.08)    | 0.715     |
| Park × PM Peak Hours        | 0.88 | (0.77 - 1.01)    | 0.066     |
| Park × Off-Peak Hours       | 0.83 | (0.73 - 0.95)    | 0.006 **  |
| Park × Weekend All          | 1.00 | (0.87 - 1.14)    | 0.973     |
| University                  | 1.97 | (1.52 - 2.55)    | < 0.001 *** |
| University × PM Peak Hours  | 0.65 | (0.45 - 0.93)    | 0.018 *   |
| University × Off-Peak Hours | 0.96 | (0.67 - 1.35)    | 0.798     |
| University × Weekend All    | 0.72 | (0.50 - 1.05)    | 0.085     |
| Cycle lane                  | 1.80 | (1.66 - 1.95)    | < 0.001 *** |
| Cycle lane × PM Peak Hours  | 0.88 | (0.79 - 0.98)    | 0.019 *   |
| Cycle lane × Off-Peak Hours | 0.88 | (0.79 - 0.98)    | 0.017 *   |
| Cycle lane × Weekend All    | 0.88 | (0.79 - 0.99)    | 0.027 *   |

**MSOA-level Characteristics**

| Population (÷ 1000)         | 1.05 | (1.00 - 1.09)    | 0.052     |
| Employees (÷ 1000)          | 1.01 | (1.00 - 1.02)    | 0.049     |

Notes: baseline (reference) category = AM Peak Hours; ‘***’ < 0.001, ‘**’ < 0.01, ‘*’ < 0.05, ‘.’ < 0.1
8% increase in the rate of bicycle drop-offs (1.08, 95% CI 1.05-1.10). This effect does not appear to differ significantly between time blocks, suggesting the persistent marginal increases in activity for shopping purposes. Similarly, hexagons that contain sustenance (food and drink) amenities are also found to increase the rate of journeys, with each additional facility observed to increase by 16% (1.16, 95% CI 1.12-1.20). Therefore, this implies that users are more likely to use the system to eat or drink as opposed to going shopping. This makes sense intuitively given that using the system with the intention to shop would require the user to make return journeys with goods that are bought which would be cumbersome since there is a very limited amount of storage on the bicycle and would require greater physical exertion. For example, in an analysis of cycling potential in London, TfL (2017) find that users were not willing to cycle when they had to carry large bags and luggage, especially among professions that require heavy or bulky work equipment such as electricians, plumbers and carpenters, as well as women who are more likely to make shopping trips. On the other hand, using BSS as a means to get food and drink do not encumber users, thus better suiting such purposes that are reflected in the 8% increase in RR in comparison to shops. Again, these effects do not differ significantly between time blocks.

Analysing the relationship between journeys and touristic destinations are found to increase the rate of drop-offs by 37% (1.37, 95% CI 1.21 - 1.56) for each additional point of interest in a hexagon. Such significant relations imply the use of the JUMP system by tourists, visiting the various galleries, museums and attractions located within the study area. Much like shops and sustenance amenities, there appear to be no significant differences between time blocks at the 5% significance level, but journeys rates are found to decrease slightly by 15% at the 10% level during PM Peak Hours and Weekend All. This suggests that there may be less activity to such destinations during these times, but given that touristic activities are not hypothesised to differ significantly across the week due to their unconstrained temporality, the results are able to reflect such hypotheses.

Although cycling infrastructure are not explicitly an indication of journey pur-
pose, assessing its association with journey drop-off rates are imperative given that they have consistently been found to have a significant positive relationship. The results of this model suggest no differently, with an 80% increase in the rate of journey terminations within close proximity to cycle lanes (1.80, 95% CI 1.66-1.95). In this context, the substantial association is evidence that JUMP users were highly likely to use these resources as a means to traverse the city safely. Though AM Peak Hours are consistently observed to have greater rates of drop-offs, with 12% decreases in all other time blocks, this is likely an artefact of the accumulation bicycles on the border of the City of London that are observed in Figure 7.7 which is also where the vast majority of cycling infrastructure converge. Therefore, as opposed to suggesting greater use during AM Peak Hours, these differences are merely a signal of such concentrations of journeys towards the centre of the city in the morning.

As highlighted in Figure 7.6, a significant proportion of journeys appear to flow towards university buildings within the study area. Here, the statistical analyses confirm these observations with results exhibiting an increase in the drop-off rate by nearly 2 times (1.97, 95% CI 1.52-2.55), the second largest of all built environment factors. Therefore, it is safe to assume that a considerable proportion of JUMP e-BSS users are students. The only significant difference in temporality at the 5% level are observed during PM Peak Hours, where journey rates are found to decrease by 35% (0.65, 95% CI 0.45-0.93) in comparison to AM Peak Hours. At the 10% level, Weekend All journeys towards universities are found to decrease by 28%. These results highlight the flexibility that students have in comparison to workers who are generally constrained to travelling during peak hours, whereby the lack of significant differences during Off-Peak Hours may be an indication of students travelling for early afternoon engagements. Similarly, the lack of commitments during PM Peak Hours are clearly identified by significant decreases in journey terminations.

Considering JUMP activity in relation to transit related features suggest some of the largest increases in RR, showcasing the importance of the dockless e-BSS in facilitating multimodal journeys. The presence of a bus stop is found to increase
the rate of journey terminations by 75% (1.75, 95% CI 1.59-1.93) whilst train stations exhibit nearly four times higher drop-off rates (3.92, 95% CI 3.27-4.70). This demonstrates that the system is used as a means to resolve the first/last mile issue primarily among rail users in comparison to bus users. Unexpectedly, AM Peak Hours are found to observe the highest association with train stations. This is unusual given that during the morning commute, users were expected to pick-up bicycles from these locations and drop them off in close proximity to their workplaces, whilst during the evening commute users were expected to return bicycles to transit hubs and therefore increase drop-off rates during PM Peak Hours. Across both PT features Off-Peak Hours exhibit the largest decreases in journey terminations at 14% (0.86, 95% CI 0.76-0.99) and 37% (0.63, 95% CI 0.49-0.81) in close proximity to bus stops and train stations respectively. This suggests that there are fewer multimodal connections facilitated by the e-BSS during Off-Peak Hours. This may be justified by more local, door-to-door journeys since commuters are already in the city or at home during these times. Unlike bus stops that exhibit no significant differences during PM Peak Hours and Weekend All, journeys within 100m of train stations are found to decrease by 33% (0.67, 95% CI 0.52-0.87) and 30% (0.70, 95% CI 0.54-0.90) respectively. These decreases may be an artefact of the greater spatial heterogeneity during these times in comparison to AM Peak Hours, where journeys primarily flow towards the centre of the city, areas that are synonymous with the highest public transport access level (PTAL) (TfL 2015) (reference map included in Appendix D).

Finally, when investigating those journeys that end in close proximity to parks and green spaces within the study area there appears to be no significant relationship. This is most likely an artefact of the fact that many of the largest parks that JUMP users would be inclined to cycle to were part of the no-parking zones identified within the application. This includes Green Park, St James’s Park and the areas surrounding The Mall and Buckingham Palace as well as Hyde Park and Regents Park. Although hexagons on these park boundaries and smaller public green spaces and squares were considered as journeys with the intention of using parks, it is clear
that these operational limitations had an undeniable impact on the journey destinations, with users constrained to using the system as a means to get to the other points of interest analysed in this model.

### 7.5 Discussion

The analysis presented in this chapter provide a holistic, yet granular understanding of the activity dynamics of the JUMP e-BSS. Generally, the system appears to be inefficient and underutilised, with an average daily utilisation rate of less than one TDB, a figure significantly lower than the London docked BSS that exhibit nearly three TDB (as detailed in Section 4.2) and other dockless BSS (as detailed in Section 7.2.1). Although this is the case, through a combination of spatio-temporal analysis and a zero-inflated negative binomial regression it becomes very clear that weekday and weekend use of the system is significantly different.

During the weekday, the system exhibits activity that is illustrative of multimodal commuter use, especially during peak hours. This is initially demonstrated through the identification of origin and destination clusters on weekday AM Peak Hours (Figure 7.6a and Figure 7.7a) and PM Peak Hours (Figure 7.6b and Figure 7.7b), revealing a dense region of bicycle counts towards the centre of the city. The morning flows showcase the use of JUMP as a mode to solve the last-mile problem (Davis 2014). Conversely, in the evenings, many of the trips that originated from the CBD had ended around the train stations, indicating first-mile use. Such multimodal use of the system is further corroborated through statistical evidence, with the most influential built environment factors being train stations and bus stops (Table 7.6), confirming the synergies between the dockless e-BSS and PT services in London (Saberi et al. 2018; Yang et al. 2022; Goodman and Cheshire 2014). Therefore, it is clear that JUMP is primarily used as a means to resolve the first/last mile problem, with flows emanating towards and out of, but not inside, the City of London, due to limitations of the system’s permitted operational area.

On the other hand, weekends tend to exhibit much greater homogeneity in use throughout the day, with activity that is distributed throughout the afternoon (Figure
These journeys tend to be of a similar duration in spite of typically shorter distances (Table 7.2) and also appear to occur much less frequently, with significant decrease in TDB on weekends. Spatially, journeys across the weekend tend to be similar in nature to those Off-Peak Hour journeys during the weekday, with activity clustering around Angel Station (Figure 7.6c and Figure 7.7c). Statistically, Off-Peak Hours yielded the least journeys to transit hubs, indicating a reduction in the amount of multi-modal journeys taken during these times. This is likely due to the fact that the bulk of commuting multi-modal activity occurs outside of peak hours, with more heterogeneous door-to-door journeys occurring. On the other hand, journeys taken to university buildings show no statistically significant differences during Off-Peak Hours, identifying the flexibility in journey times by students in comparison to other workers. The primary differentiating factors between weekends and built environment factors were in the persistence of leisure and utilitarian activities, such as shops, sustenance and tourism. Though these were also found to show no statistical differences between any time blocks, this suggests the continued use of the system for such purposes throughout the week. Interestingly, train stations observe the least reduction in activity during Weekend All, whilst bus stops show no significant differences. These suggest the continued use of the system as a means to facilitate multi-modal journeys throughout the weekend. Therefore, weekend use of the JUMP system exhibits significant, temporal, spatial and statistical differences from those of weekday journeys.

This being said, there are some interesting commonalities that emerge between weekday and weekend activity, such as the relative importance of cycling infrastructure, corroborating the findings of many other studies (Nello-Deakin 2020; Fishman et al. 2014b; Faghih-Imani et al. 2014; Buck and Buehler 2012; Mateo-Babiano et al. 2016), but contrasts the utility of dockless BSS in Singapore (Shen et al. 2018; Xu et al. 2019). Similarly, shops, sustenance and tourist areas were found to increase the rate of journeys, much akin to previous analyses of dockless BSS (Bielinski and Wąża 2020; McKenzie 2018), albeit very marginally, but differed in the lack of a relationship to parks and green spaces that have been found in studies
of BSS in China (Du et al. 2019; Guo and He 2020; Zhao and Li 2017). These differences in the magnitude and direction of built environment determinants of use between systems highlight the importance of analysing these particularities in facilitating a more nuanced understanding to better guide operator and policymaker management.

Collectively, the overwhelming statistical significance of all built environment factors (apart from parks that were a result of operational constraints) with journey destinations signify that trips were indeed made with a particular purpose, most notably in association with train stations and university buildings.

### 7.5.1 Evaluation and Limitations

Firstly, it is important to acknowledge the limitations of the data, with inaccuracies in the GPS traces and cleaning processes. In relation to GPS inaccuracies, it is understood that the accuracy depends on the number of satellite signals that reach the GPS receiver – the more the better (Van Dijk and De Jong 2017). However, these signals can bounce, reflect or diffuse in the environment, which can create variation in the coordinates returned by up to hundreds of meters even whilst the GPS receivers are in situ (Flüchter and Wortmann 2014; Yin and Wolfson 2004). This is especially so in urban areas with dense high-rise buildings (Van Dijk and De Jong 2017), like London. This issue has been taken into account by only selecting the initial location of each signal during the state of the GPS wandering. However, it is not certain whether the first signal ping was the most accurate, as it may have improved after the signals have bounced around for a while (Van Dijk and De Jong 2017). The nature of the data collection process also meant that user and non-user movements could not be distinguished with perfect certainty. Thus, sets of rules were defined based on some assumptions and guided by prior research to filter-out the non-user movements, which accounted for a substantial proportion (nearly 75%) of the total movement records. During the process, round trips were reluctantly excluded, which means the analysis conducted in this study overlooks this dimension of e-BSS usage, as detailed in Section 7.2.2.

There was also the issue of modifiable areal unit problem (MAUP) that was
a result of limitations on the study area, e-BSS operations and data disclosure. In visualising the flows of journeys in or out of each cluster within Section 7.3, the corresponding origins or destinations were aggregated into 150 metre radius hexagonal cells to conceal the individual journey information, having taken into account ethical considerations of anonymisation. Consequently, the results were subject to the MAUP due to the aggregation of data into the selected geographical scale. This methodology was also employed to determine the association to built environment factors for the statistical analysis in Section 7.4, that present more prominent issues in relation to the interpretation of results. In particular, those journeys that reside on the periphery of hexagonal cells could be incorrectly classified, leading to issues surrounding misclassification bias (Copeland et al. 1977), whereby neighbouring cells may contain built environment factors that are closer to the exact drop-off location and therefore better represent journey purposes. The effects of MAUP have been examined in the context of BSS and built environment determinants by Gao et al. (2021a), finding significant deviations in the relative ranking of points of interest across different levels of spatial aggregation. Although significant deviations were found in this case, such considerations were made prior to the presentation of the final model in this chapter. Comparing hexagonal cells from radii of 50 metres to 100 metres at 25 metre increments, no significant differences were observed in the relative ranking of built environment factors as well as larger spatial aggregations muddying the effects due to the presence of a greater number and variety of factors. As a result, the model presents the most granular geographical aggregations of any built environment analyses of dockless BSS to more accurately identify the most likely journey purposes where larger scales may obscure the true heterogeneity of activity dynamics. Therefore, best efforts were made to mitigate the harmful impacts of MAUP, but future studies may wish to consider the use of distance based measures to each destination location instead of spatial aggregations to better alleviate such issues.

In conjunction to issues of spatial aggregation within the statistical modelling, it is also important to acknowledge the artificial constraints of JUMP’s operational
area. Although the effects of non-permitted boroughs and no-parking zones were accounted for by the overdispersion parameter within the zero-inflated model, the true purpose of journeys are obscured whilst also inflating the effects of variables such as cycling infrastructure, touristic destinations and parks. These have been carefully considered when interpreting the results, however, it is argued that the spatio-temporal hotspot analyses presented in this chapter help to complement the statistical analysis and sheds light on this blind spot, indicating use just outside of the City of London and park borders.

Finally, it is also important to note that given the ecological study design using a cross-sectional framework, though there is high internal validity of specified study area and duration, there is limited externally validity. For example, this study was conducted using data during the infancy of the system’s operation and included the New Year holiday period that would have likely reduced its utilisation and adoption. Similarly, it was prior to the widespread and long-lasting impacts of COVID-19. As a result, the use of dockless e-BSS in London may have exhibited significant changes in their activity dynamics over the duration of the pandemic, in addition to the post-pandemic period. The resilience of BSS as a mode of urban mobility has been identified across numerous studies that employ qualitative survey data to determine changes in modal choice (Bucsky 2020; Dingil and Esztergár-Kiss 2021; Teixeira et al. 2022). Though flexible and resilient, a lack of commuter activity during periods of enforced lockdown, with reductions of around 87% across 10 countries (Barbieri et al. 2021), will have caused significant changes in the activity dynamics and their relations to built environment factors. Similarly, the merging of JUMP and Lime dockless e-BSS in London, in addition to their growth in fleet size and operational area present further deviations from the present system. As a result, future research should focus on obtaining more contemporary dockless e-BSS data from the pandemic and post-pandemic period in order to determine the short- and long-term changes in mobility dynamics in comparison to those in the pre-pandemic period investigated in this chapter.
7.5.2 Applications

The analyses in this chapter makes several contributions to academic perceptions of dockless e-BSS, having identified the hotspots and built environment factors that are associated with user activities in London, a city for which such analyses have yet to be conducted. In addition, the methodologies and subsequent results present a number of potential practical applications outside of academia, providing operators and governing bodies with data-led evidence of the way in which this system is used.

For example, the granularity of hotspot locations that have been identified at different times of the day in Figure 7.6 and Figure 7.7 can be used to identify prospective areas for physical or virtual parking spaces. Since dockless bicycles have been increasingly associated with issues surrounding ‘bike litter’ (Chen 2019), it is important that users have access to dedicated parking spaces to help alleviate such issues. Given the significant amount of negative press associated with such issues, dockless operators have increasingly prioritised user drop-off behaviour management through various techniques including geo-fences that depict virtual parking spaces, mandatory pictures of parked bicycles and monetary incentives and penalties for good and bad parking practices respectively, as per the ‘Dockless bike share code of practice’ (TfL 2018a). Although such endeavours have been made, research finds that it is imperative parking locations are as close as possible to users intended destinations to ensure that these reward-penalty schemes are effective (Gao et al. 2021b). Therefore, using the bespoke hotspot location identification techniques detailed in this chapter and the subsequently pinpointed locations, dockless operators should work collaboratively with boroughs in order to create physical parking spaces for micromobility vehicles at such locations to better facilitate user demands and alleviate issues of inconsiderate and unsafe parking behaviours.

Similarly, given the differentiations in built environment factor effects and hotspot locations in the time blocks that have been analysed, dockless operators should use such insights to better prioritise and optimise redistribution efforts. In the context of the JUMP e-BSS it is clear that during weekdays there is a cyclical
inflow in the morning towards the centre of the city and an outflow northwards in the evenings. As a result, ensuring bicycle availability at these locations at the appropriate time of day are important in increasing the utilisation of the system. Such redistribution efforts are should occur during Off-Peak Hours, when the system’s activity is at a minimum during weekdays. Unlike weekdays, weekend activity locations are much more dispersed, but since hotspot locations are still possible to be identified, as well as the continuation of strong and positive associations with transit stations within the statistical analysis, bicycles should be redistributed to these locations on weekends.

From the perspective of borough officials, the results provide further evidence that investment in shared and active modes of transportation are essential due to their increasing popularity and importance in facilitating first/last mile journeys for commuters and university students alike. First, the persistence of a strong and significant relationship with cycling infrastructure, that have been identified both in Section 7.4 as well as numerous other studies, bolster evidence that greater investments should be made to make cycling activities safer in the city. Consequently, routing algorithms on journey OD data, such as the CycleStreets engine detailed in Section 7.2, should be used as a tool to prioritise cycling infrastructure construction along these roads. Secondly, the undeniable benefits of the dockless e-BSS in facilitating commuter and student multimodal journeys provide further evidence that officials should make greater efforts to facilitate such use. This should be prioritised in those boroughs that did not give permission for the system to operate, as evidence from this chapter suggests that such restrictions inhibited the uptake of journeys within these areas. Furthermore, since it is evident that the JUMP system exhibits considerable demand and activity in the northern areas of Camden, Islington and Hackney, there are unmet demands for micromobility modes in regions that are not serviced by the docked Santander Cycles system. As a result, permitting the use of dockless BSS in those neighbouring boroughs, as well as though south of the River Thames, would likely foster greater adoption of these shared and active modes of transportation among users that are discouraged by the lack of service in
Finally, as alternative dockless systems and forms of micromobility share similar issues of data availability and characteristics, the methodologies that have been developed in this chapter provide a great framework to conduct granular analyses of activity dynamics. Therefore, it is suggested that future research should consider employing analogous analyses to determine the intricacies of individual systems in specific contexts to better guide their management and operation, much like those detailed here.

7.6 Chapter Summary

This chapter set out to break new ground in applying GPS data in understanding an e-BSS system in London. The availability of precise GPS data on the London Uber JUMP system has enabled granular analyses of its use, a city for which there are currently no research. Through a combination of bespoke spatio-temporal hotspot analyses and novel multilevel regression modelling that include temporal interaction effects, it has been possible to identify the exact locations and times of day and week that are associated with journey dynamics. The findings presented across these various methodologies complement each other, helping to identify significant commuter use as part of a multimodal journey during morning and evening peak hours across the weekday with significant flows to and from the City of London. During weekends and intra-peak weekday hours, journey activity appears to be much more distributed and consist of journeys at a local scale, suggesting the dominance of leisure and recreational activity during such times.

More specifically, this has been demonstrated by the spatial concentration of journeys, as identified in the hotspot analyses, as well as statistical analysis that identify significant associations with PT services. The adoption of the e-BSS by university students is also apparent, with the most activity observed during AM Peak Hours and Off-Peak Hours which are demonstrative of their relative flexibility in travel times unlike office workers that work a typical 9am to 5pm schedule. Although activity within close proximity to parks were found to have no signifi-
cant effects on bicycle drop-offs, this highlights the importance of considering the artificial constraints on user activity by the operator but also showcase their successes in minimising activity in such areas. The adoption and use of the system even though such constraints were imposed on users are a testament to their value in facilitating the first/last mile problem and a signal of their potential given their wider acceptance by boroughs.

Beyond those findings in direct association with the JUMP e-BSS in London, the analyses in this chapter highlight the value of granular journey OD within the context of dockless systems in deepening our understanding of such systems. Although potentially disclosive, with rigorous considerations of user privacy there are numerous possibilities to infer valuable and actionable insights. This ability to delineate linked OD data has been inhibited with the introduction of GBFS Version 2.0, meaning that analyses are limited to coarse inferences of unlinked origins and destinations (Xu et al. 2022). As a result, access to these granular OD data are invaluable and necessary in order to derive informative insights that can help to drive data-informed decisions to optimise their adoption, operation and management. Thus, future research should aim to access such data by working collaboratively with dockless operators with the broader aim of increasing the sustainability, accessibility and equitability of transportation in our future cities.
Chapter 8

Discussions and Conclusions

The rapid adoption and implementation of BSS in cities around the world mark the beginnings of a sustainable mobility revolution that embark to a mission to create healthier and more sustainable cities. Their effectiveness in facilitating a smooth transition towards an emissions free future are very promising due to their familiarity as means of transport that have existed for two centuries. They also offer a number of competitive advantages over alternative means of reaching these goals, being significantly more cost effective. For example, EV require individuals to make large investments only to remain unused for the majority of the time and the expansion of PT services incur vast financial and temporal costs as a result of large scale infrastructure construction. BSS on the other hand, can be rented by users as and when they are required, removing costs associated with ownership, as well as relying on existing infrastructure that can be retrofitted at considerably lower costs. Therefore, it is imperative to understand the way in which BSS have grown and been adopted by users, not only for the future of BSS, but also for the future of urban transportation more generally.

This thesis has aimed to uncover the values of passively produced, openly accessible BSS data in broadening our understandings of the dynamics of these new modes. Through a thorough exploration of their utility at different temporal and geographical resolutions, this work offers new heuristics and frameworks that best leverage the unique characteristics of these data to provide crucial, novel insights into their dynamics that can also be employed in similar emerging and future modes
This chapter concludes the thesis by consolidating the findings and contributions in advancing academic perceptions of BSS dynamics at different scales in Section 8.1. Applications of these methods and results are discussed more generally from the perspectives of government and micromobility operators in Section 8.2. Section 8.3 acknowledges the limitations of this work and subsequently identifies areas that require further research. Finally, Section 8.4 summarises the thesis with some closing remarks.

8.1 The Dynamics of Bicycle Sharing Systems

By exploring these dynamics through a variety of lenses, it has been possible to uncover the value of openly accessible data in discovering new insights that are hoped to better guide current and future implementations of BSS in addition to similar modes of urban micromobility.

Chapter 4 explores the value of docked BSS data in generating scalable and comparative metrics that can offer insights into the dynamics of individual systems but also offer the basis for much greater temporal and geographical scale analyses. As docked BSS are the longest established form that still dominate the market, leveraging these data was essential. The data presented a number of challenges to ensure the most accurate estimations of cycling activity that required pragmatic steps with a vast number of considerations. This included the formalisation of existing knowledge across the domains of travel behaviour, human mobility and BSS to create an algorithm that could detect the hiring of a bicycle. Validation processes that employed the limited availability of journey data enabled the identification of data errors that caused significant overestimations. This highlighted the necessity to create additional algorithms that would identify and remove such occurrences which, when implemented, resulted in much more accurate estimations of journey activity.

Similar procedures were necessary for the cleaning and manipulation of dockless BSS data in Chapter 7, ensuring the careful consideration of the unique char-
acteristics of e-bicycles and system operational extent. Due to the lack of reference data to validate such processes, the filtering of data required some subjective decisions, however, each were rigorously designed, justified and tested to ensure their utmost accuracy. In both cases, these procedures proved to be vital in order to enable the identification of BSS dynamics and demonstrated that although the data were originally intended to provide real-time information on the location and availability of bicycles, they could be repurposed for empirical analyses on system dynamics.

Chapter 5 explores the values of the metrics that can be derived from docked BSS in enabling global-scale comparisons and classifications that have yet to be analysed at such extent. As a result, the full breadth of the UCL BSS data collection could be exploited. Descriptive visualisations enabled preliminary identification of BSS growth and changes across a two-year period, depicting trends of growth with size and efficiency across 176 BSS in Figure 5.1. The chapter then employs a bespoke two-staged clustering methodology to identify the general types of BSS that exist. These are produced only using metrics that were derived from Chapter 4, therefore ensuring homogeneity and comparability. Although a broader number of variables would evidently produce more descriptive classifications, the aim of this analysis was to investigate the utility of comparable metrics in static global-scale representations of the landscape given its limited scope and appraisal both within and outside of academia. Even with such limitations, the methodology enabled the identification of distinct BSS characteristics, such as the persistence of commuting patterns across weekends in Chinese BSS as well as systems in countries that practise the siesta. This is a testament to the metrics’ and methodologies’ capabilities in uncovering unparalleled insights over extensive ranges.

The scope of the analysis was reduced in Chapter 6 to investigate the unprecedented impacts of COVID-19 on the medium-term activity dynamics within the London Santander Cycles BSS. This provides a novel framework that enables the quantification of changes in mobility across a number of dimensions in comparison to the previous years’ activity. The use of spatio-temporal techniques identifies the
heterogeneity in adaptations that users have made in line with the implementation of restrictions on movement. Network analysis methods extended on these techniques by exploiting the closed nature of the system to provide empirical evidence of such changes. The combination of these methods highlighted the importance of BSS in facilitating a safe and sustainable means of essential travel whilst also adapting to the requirements of those who were forced to quarantine in their homes, supporting the adoption of the mode as a means of exercise and leisure. This demonstrates the versatility and resilience of BSS that are unattainable with alternative means of PT, extending the already lengthy portfolio of benefits that urban micromobility modes offer.

Chapter 7 presents even more fine-grained spatial and temporal system-scale BSS dynamics through an in-depth exploration of the London JUMP dockless e-BSS. The development of the fourth-generation of BSS combined with electric pedal-assistance in e-bicycles showcase the continual developments that further increase accessibility to the mode. They also introduce new opportunities for research that can leverage their unrestricted nature to better infer journey locations and purposes. Through a bespoke hotspot detection methodology that combines point-density clustering methods to delineate specific areas and street segments that observe a large proportion of activity, we detect the ebb and flow of commuting behaviour concentrated around the boundaries of the City of London and train stations, such as King’s Cross and Angel, during peak hours. Off-peak and weekend activity exhibit significant deviations from the directed commuting patterns, with journeys that are much more distributed in nature. This is quantified by employing a statistically robust regression method that was drawn from the fields of epidemiology and criminology due to the similarities in data distribution. These heuristic processes provide a valuable framework to uncover the spatial and temporal particularities of dockless BSS that are distinct to each systems’ context.

Together, these empirical chapters demonstrate the indispensable value that open BSS data offer in furthering our understanding of BSS dynamics across a range of temporal and spatial scales. Although the use of accurate and validated operator
published data would be preferable, their limited availability inhibits the continual
generation of new insights and understanding of these systems. Thus, these real-
time BSS API feeds ensure the availability of substitute data sources that can be
collected, cleaned and analysed for such purposes. Building on the growing body of
literature that analyses BSS, these analyses contribute to discourses surrounding the
use of new sources of urban mobility data within the fields of transport geography,
system dynamics and urban planning.

8.2 Applications and Impacts

In conjunction with those novel methodological and analytical contributions to aca-
demic disciplines, the research presented in this thesis have some practical applica-
tions in the planning, operation and management of BSS.

Firstly, as these modes of micromobility are still in their infancy, many cities
are hesitant to widely adopt and invest in them. Analyses, such as those in Chapter
6 and 7, highlight the benefits of including these systems within cities, helping to
increase the adoption of healthy and sustainable cycling activity and build greater
transport resilience. The limitations that governing bodies impose are especially
evident in the context of the London JUMP dockless e-BSS, where the limited and
disjointed nature of the system’s operational extent inhibit more widespread adop-
tion and efficiency. Evidence from the global-scale analyses in Chapter 5 also show
that size is a key factor in increasing the use and efficiency of BSS. Widespread
acceptance of these modes would not only help to facilitate better connections to
desirable locations but also increase uptake and synergies with existing forms of
PT that would further reduce the burdens of road congestion and subsequent envi-
ronmental and air quality concerns. Therefore, the research presented in this thesis
adds to the growing pool of evidence that can be used to advise cities and governing
officials to foster these modes with open arms to induce further demand.

The results can also be applied in more specific contexts to inform data-driven
investments to increase the uptake of BSS. As is evident from the statistical anal-
ysis of built environment factors in Chapter 7 that echo the findings across numerous
other studies; dedicated cycling infrastructure and PT services are highly correlated with increasing BSS uptake. Using the heuristics developed in the analysis of the dockless JUMP system, it is possible to delineate popular flows and hotspots of activity that can be used to target investments for particular routes and areas. This is especially valuable in identifying locations to build physical or virtual micromobility parking spots that can meet the demands of users whilst alleviating issues surrounding the improper and potentially dangerous parking of vehicles. Similarly, using routing engines between popular OD flows, cities can appropriately prioritise investments in cycling infrastructure to improve the safety and uptake of these modes.

In the context of BSS data, metric creation processes developed in this thesis can be used to validate operator publications that have previously been found to overstate size and use. Whilst being valuable in ensuring that reporting is accurate, they also offer advertisers a means of estimating the value of these modes prior to investing in them, and governing bodies a way of ensuring that systems do not breach contractual rules and obligations. Global classification of BSS based on these metrics also provide a basis for estimating attributes of systems for which API feeds are not accessible. Such estimations can be extended to systems that are yet to exist, guiding the planning and implementation proceedings to follow the practices of highly utilised systems.

Finally, although this thesis focuses on BSS, many alternative modes of shared mobility that are emerging share similarities in data structure due to the likeness in the way that they operate. These include, but are not limited to, e-scooters, mopeds and car services. Collectively, these shared modes provide the basis for future visions of MaaS and a pathway to reduce reliance on privately owned vehicles. Therefore, the heuristics that are presented in this thesis can be easily adapted and transferred to analyse data from these alternative emerging modes to create a more holistic understanding of shared modes of urban mobility that are vital in further progressing their adoption and use.
8.3 Future Work

This thesis presents a broad range of insights on BSS dynamics that not only extend on academic understandings of micromobility modes, but also have some practical applications in helping to increase their adoption and efficiency. Despite this, as with any analysis, this work has its limitations. In this case, these are primarily a result of the fundamental lack of data on these modes.

Firstly, as established, the data contain a number of uncertainties and inaccuracies due to their provenance. Although best efforts have been made to accurately repurpose the data, processes of metric creation and journey identification are still prone to errors. This means that although results are a great representation and tool to conduct such analyses, they are not perfect. Thus, we urge operators to work collaboratively with researchers to provide greater access to these validated data sources to give further confidence in the findings presented in this thesis.

Given the limitations with the data, analysis of BSS user characteristics have not been possible. This is a common limitation throughout the existing literature on BSS and are justified due to their highly disclosive nature. Although this is the case, conducting analyses on user socio-economic attributes offer more granular insights that may be pivotal in further unpacking and understanding the dynamics of BSS. Heterogeneity in the uptake of BSS has been identified across different age groups, income levels and sex, but these are typically based on survey methods that lack scale and accuracy. Therefore, future research should aim to consider these demographic dimensions of BSS dynamics in further detail where access to such has been permitted, ensuring utmost ethical considerations and rigour when doing so.

In addition to inaccessible demographic attributes of users, there are certain unquantifiable measures that are likely to have some influence on the dynamics of BSS. Most notably, cultural factors are believed to have an intrinsic association with the uptake and use of BSS. Although the global-scale comparison of BSS in Chapter 5 offer some insights into these cultural differences, they are by no means exhaustive. The establishment of historical cycling cultures in countries such as the
Netherlands and Japan, may arise due to traditions and practices that are immeasurable. Therefore, acknowledging the existence of these unaccountable influences of BSS are a limitation that require further, more contextual and qualitative research.

In the context of analysing BSS dynamics, this thesis works retrospectively, creating a deeper understanding of the way in which they have been used but lack considerations of how these data could be used to model future dynamics. As mentioned, advancements in computing power have opened opportunities for more computationally intensive operations. The field of machine learning is a rapidly growing research area and offers new tools that can leverage the extensive collection of BSS data to model future demands and dynamics. There is a limited but growing body of literature that endeavours to employ these algorithms to tackle practical issues that require foresight, such as the BRP. Similarly, agent based models that simulate BSS operation and their environments offer another potential methodological framework to uncover novel future insights. The construction of vast digital twins of cities are an innovative way in which future scenarios can be realised and analysed. In these contexts, simulations are a valuable tool in creating alternative scenarios, such as the expansion of BSS, construction of new infrastructures or imposition of new policies/restrictions and estimate the expected outcomes on an individual agent basis. In future research, the UCL BSS data collection offers great potential in serving such methodological developments that can further improve their implementation and operational efficiency.

8.4 Closing Remarks

In summary, open and passively produced BSS data have proven to be an invaluable resource in analysing the system dynamics across a range of scales. Although the data pose some challenges and limitations, these can be overcome through rigorous cleaning and manipulation processes that have been developed in line with previous research and specific contextual information, with metrics being validated where possible. The findings offer a very positive outlook on the future of urban mobilities, with the implementation of BSS having a largely positive reception and adoption by
cities and individuals alike. They appear to facilitate a significant uptake of cycling for a variety of trip purposes that can aid in the transformation of the transportation sector to tackle issues arising in cities as a result of increasing populations.

This being said, there are still some very apparent areas to improve, such as many smaller to medium BSS, including the dockless JUMP system, exhibiting inefficient use. These should be analysed in greater detail, using these novel data sources and methodologies to help them reach their full potential in facilitating greater urban transportation integration and advancing the adoption of sustainable, active and shared modes. With greater understanding of these modes, individuals, micromobility operators and cities can make more informed choices and decisions.

Thinking more broadly about the collective efforts that need to be made across various sectors of the economy, the values of open and passively produced BSS data are representative of the countless other sources of data that are being produced in cities today. It is imperative that we explore the values of these data to create smarter cities that leverage them to better plan the appropriate steps necessary to monitor progress and achieve goals that minimise the impacts of the global climate crisis. Given the wide-ranging utility of these data that have been exemplified in this thesis, it is abundantly clear that greater efforts should be made to make better use of these data to take action to improve the health and longevity of our cities.
Appendix A

Exploration of Santander Cycles
2016 Redistribution Data

Figure A.1: Temporal distribution of London Santander Cycles BSS operator bicycle collection in 2016

Table A.1: Descriptive statistics of London Santander Cycles BSS operator bicycle collection in 2016

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Quartile</th>
<th>Mean</th>
<th>3rd Quartile</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekday</td>
<td>697</td>
<td>2,194</td>
<td>24.23</td>
<td>2,696</td>
<td>3,112</td>
</tr>
<tr>
<td>Weekend</td>
<td>552</td>
<td>1,119</td>
<td>1,425</td>
<td>1,739</td>
<td>2,223</td>
</tr>
</tbody>
</table>
Figure A.2: London Santander Cycles BSS operator bicycle distribution during the average weekday and weekend in 2016

Table A.2: Descriptive statistics of London Santander Cycles BSS operator bicycle distribution in 2016

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Quartile</th>
<th>Mean</th>
<th>3rd Quartile</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekday</td>
<td>469</td>
<td>1,992</td>
<td>2,196</td>
<td>2,436</td>
<td>2,975</td>
</tr>
<tr>
<td>Weekend</td>
<td>171</td>
<td>1,057</td>
<td>1,394</td>
<td>1,564</td>
<td>2,108</td>
</tr>
</tbody>
</table>
Figure B.1: The evolution of 176 BSS between 2016 and 2018

Source: (Cheshire and Uberti 2021)
Appendix C

Global Classification System

Assignments

K-Means Cluster 1 (Very large, High use [3])

Table C.1: Weekday dynamic time warping cluster assignments for ‘very large, high use BSS’

<table>
<thead>
<tr>
<th>‘Very large, high use BSS’ Weekday DTW Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 [3]</td>
</tr>
</tbody>
</table>

Table C.2: Weekend dynamic time warping cluster assignments for ‘very large, high use BSS’

<table>
<thead>
<tr>
<th>‘Very large, high use BSS’ Weekend DTW Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 [2]</td>
</tr>
<tr>
<td>2 [1]</td>
</tr>
</tbody>
</table>
Appendix C. Global Classification System Assignments

K-Means Cluster 2 (Large systems in major cities [15])

Table C.3: Weekday dynamic time warping cluster assignments for ‘large BSS in major cities’

<table>
<thead>
<tr>
<th>‘Large BSS in major cities’ Weekday DTW Clusters</th>
</tr>
</thead>
</table>

Table C.4: Weekend dynamic time warping cluster assignments for ‘large BSS in major cities’

<table>
<thead>
<tr>
<th>‘Large BSS in major cities’ Weekend DTW Clusters</th>
</tr>
</thead>
</table>

K-Means Cluster 3 (Medium systems with extensive cycling infrastructure [5])

Table C.5: Weekday dynamic time warping cluster assignments for ‘medium BSS with extensive cycling infrastructure’

<table>
<thead>
<tr>
<th>‘Medium BSS with extensive cycling infrastructure’ Weekday DTW Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 [1] Stuttgart</td>
</tr>
</tbody>
</table>

Table C.6: Weekend dynamic time warping cluster assignments for ‘medium BSS with extensive cycling infrastructure’

<table>
<thead>
<tr>
<th>‘Medium BSS with extensive cycling infrastructure’ Weekend DTW Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 [1] Stuttgart</td>
</tr>
</tbody>
</table>
### K-Means Cluster 4 (Small to medium efficient BSS [66])

**Table C.7:** Weekday dynamic time warping cluster assignments for ‘small to medium efficient BSS’

<table>
<thead>
<tr>
<th>‘Small to medium efficient BSS’ Weekday DTW Clusters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2 [29] Bergen, Bordeaux, Buenos Aires, Castellon, Changyi, Gaomi, Girona, Guadalajara, Honolulu, Huangyan, Lille, Ljubljana, Lyon, Madrid, Marseille, Nantes, Nice, Palma, Pisa, Querétaro, Santos, Sanxiang, Saragossa, Seville, Toulouse, Turin, Valencia, Vila Velha, Vilnius</td>
<td></td>
</tr>
<tr>
<td>3 [29] Bialystok, Brighton, Bydgoszcz, Cardiff, Daejeon, Gliwice, Helsinki, Kalisz, Kanazawa, Katowice, Kołobrzeg, Kraków, La Rochelle, León, Łódź, Lublin, Ordu, Oslo, Porto Alegre, Poznan, Radom, Recife, Rio, Salvador, São Paulo, Sorocaba, Szczecin, Vienna, Wrocław</td>
<td></td>
</tr>
</tbody>
</table>

**Table C.8:** Weekend dynamic time warping cluster assignments for ‘small to medium efficient BSS’

<table>
<thead>
<tr>
<th>‘Small to medium efficient BSS’ Weekend DTW Clusters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2 [10] Castellon, Daejeon, Girona, León, Madrid, Milan, Palma, Pisa, Seville, Valencia</td>
<td></td>
</tr>
<tr>
<td>3 [4] Changyi, Gaomi, Huangyan, Sanxiang</td>
<td></td>
</tr>
</tbody>
</table>
## K-Means Cluster 5 (Small to medium inefficient BSS)

Table C.9: Weekday dynamic time warping cluster assignments for ‘small to medium inefficient BSS’

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Cities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 [38]</td>
<td>Vor Wien, Aruba, Bath, Bottrop, Columbia, Exeter, Greenville, Gütersloh, Hailey/Ketchum, Hamburg, Jurmala, Karlovac, Kent State University, Kingston (UK), Konya, Las Vegas, Lincoln (UK), Makarska, Munich, Northampton, Nottingham, Oberhausen, Quickborn, Richmond (CA, USA), Romerland, Schenectady, Serfaus, Sisak, Tubingen, Tulln, Tuzla, Vannes, Velikagorica, Vukovar, West Palm Beach, Wienerwald, Wiener Neustadt, Zadar</td>
</tr>
<tr>
<td>2 [99]</td>
<td>Abu Dhabi, Albany, Almaty, Aspen, Astana, Atlanta, Augsburg, Austin, Batumi, Belfort, Berlin, Bochum, Boise, Boulder, Broward County, Budapest, Buffalo, Calais, Charlotte, Chattanooga, Cincinnati, Clermont-Ferrand, Cluj, Cologne, Columbus, Des Moines, Detroit, Dortmund, Dresden, Duisburg, Düsseldorf, Essen, Fort Worth, Frankfurt, Gelsenkirchen, Grodzisk Mazowiecki, Hamilton, Houston, Innsbruck, Kaiserslautern, Kansas City, Kazan, Klagenfurt, Kranj, Leipzig, Liverpool, Long Beach (CA, USA), Long Beach (NY, USA), Louisville, Lviv, Madison, Marburg, Marrakech, Melbourne, Memphis, Miami Beach, Milwaukee, Minneapolis, Monash University, Montevideo, Montpellier, Mülheim an der Ruhr, Nashville, New Orleans, Nitra, Nuremberg, Omaha, Opole, Orlando, Park City, Phoenix, Pioneer Valley, Pittsburgh, Portland, Potsdam, San Antonio, San Diego, San Jose, Santa Monica, Santander, Shengzezhen, Shymkent, Stalowa Wola, Saint-Étienne, Stirling, St Petersburg (FL, USA), Szamotuly, Tallinn, Tampa, Tel Aviv, Timisoara, Topeka, Trondheim, Tucson, University of Warwick, Usedom, Vancouver, Virginia, Wiesbaden</td>
</tr>
<tr>
<td>3 [73]</td>
<td>Aksu, Amiens, Amstetten, Anqiu, Aral, Avignon, Besançon, Brinje, Brisbane, Brussels, Bucharest, Changle, Colorado Springs, Dayton, Denver, Derby, Dijon, Drammen, Fenzhuzhen, Flensburg, Glasgow, Gospić, Gothenburg, Groß-Enzersdorf, Hoboken, Hollabrunn, Indianapolis, Karlsruhe, Kona, Laa an der Thaya, Lillestrøm, Limassol, Lincoln (NE, USA), Linctu, Lund, Lunz am See, Luxembourg City, Lucerne, Maastricht, Malta, Marchfeld, Metković, Milton Keynes, Mödling, Mulhouse, Namur, Nancy, Nantong, Offenbach am Main, Oakland/Berkeley, Offenburg, Cergy, Petrolina, Philadelphia, Puebla, Reading, Rennes, Reykjavik, Riga, Rouen, Rüsselsheim am Main, Salt Lake City, Slavonskibrod, Slough, St Polten, Taibaozhuang, Toyama, Lower Traisental, Traisen, University of South Florida, Valence, Victoria, Zagreb</td>
</tr>
</tbody>
</table>
Table C.10: Weekend dynamic time warping cluster assignments for ‘small to medium inefficient BSS’

<table>
<thead>
<tr>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 [48]</td>
</tr>
<tr>
<td>10 Vor Wien, Auckland, Bottrop, Byblos, Christchurch, Derby, Drammen, Exeter, Flensburg, Gelsenkirchen, Gibraltar, Hailey/Ketchum, Hamburg, Hamm, Herne, Jackson, Jurmala, Kent State University, Kingston (UK), Lake Neusiedl, Las Vegas, Lund, Mödling, Oberhausen, Offenburg, Créteil, Prague, Quickborn, Rennes, Saratoga Springs, Schenectady, Šibenik, Sisak, St Polten, Sursee, Topeka, Tuzla, University of South Florida, Vannes, Velagorica, Virginia, Vukovar, Wiener Neustadt, Wurzburg, Yangzi, Zadar</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 [173]</td>
</tr>
<tr>
<td>Abu Dhabi, Aksu, Albany, Almaty, Amiens, Anqiu, Aral, Aruba, Aspen, Astana, Atlanta, Augsburg, Austin, Avignon, Bath, Batumi, Belfort, Berlin, Besançon, Bhopal, Bochum, Boise, Boulder, Brisbane, Broward County, Brussels, Bucharest, Budapest, Buffalo, Calais, Changle, Charlotte, Chattanooga, Cincinnati, Clermont-Ferrand, Cluj, Cologne, Colorado Springs, Columbia, Columbus, Dayton, Denver, Des Moines, Detroit, Dijon, Dortmund, Dresden, Duisburg, Essen, Fenhuzhen, Fort Worth, Frankfurt, Glasgow, Gothenburg, Greenville, Grodzisk Mazowiecki, Hamilton, Hoboken, Houston, Indianapolis, Innsbruck, Kaiserslautern, Kansas City, Karlovac, Karlsruhe, Kazan, King Abdullah Economic City, Klagenfurt, Kona, Konstancin-Jeziorna, Konya, Kranj, Leipzig, Lillestrøm, Limassol, Lincoln (UK), Lincoln (NE, USA), Linqu, Liverpool, Long Beach (CA, USA), Long Beach (NY, USA), Louisville, Lunz am See, Luxembourg City, Lucerne, Lviv, Maastricht, Madison, Makarska, Malta, Marburg, Marchfeld, Marrakech, Melbourne, Memphis, Miami Beach, Milton Keynes, Milwaukee, Minneapolis, Monash, Montevideo, Montpellier, Mülheim an der Ruhr, Mulhouse, Munich, Namur, Nancy, Nantong, Nashville, New Orleans, Nitra, Northampton, Nottingham, Nuremberg, Oakland/Berkeley, Omaha, Opole, Orlando, Cergy, Park City, Petrolina, Philadelphia, Phoenix, Pioneer Valley, Pittsburgh, Porec, Portland, Potsdam, Puebla, Reading, Reykjavik, Richmond (CA, USA), Riga, Romenland, Rouen, Rüsselsheim am Main, Salt Lake City, San Antonio, Santander, Serfaus, Shengzezhen, Shymkent, Slavonski Brod, Stalowa Wola, Saint-Étienne, Stirling, St Petersburg (FL, USA), Szamotuly, Taibaozhuang, Tallinn, Tampa, Tel Aviv, Timisoara, Toyama, Traisental, Trondheim, Tubingen, Tucson, Tulln, University of Warwick, Usedom, Valence, Vancouver, Victoria, Wachau, West Palm Beach, Wienerwald, Wiesbadenm Zagreb</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 [12]</td>
</tr>
<tr>
<td>Amstetten, Brinje, Dajiawa, Gospić, Groß-Enzersdorf, Gütersloh, Holnbrunn, Ivanić-Grad, Laa an der Thaya, Metković, Slough, Lower Traisental</td>
</tr>
</tbody>
</table>
Appendix D

JUMP Hotspot Locations

Figure D.1: Public transport access level in Greater London

Source: TfL 2015
Figure D.2: JUMP e-BSS journey origin hotspots

- **a) AM Peak Hours**
- **b) PM Peak Hours**
- **c) Off-Peak Hours**
- **d) Weekend All**
Figure D.3: JUMP e-BSS journey destination hotspots


BIBLIOGRAPHY


Canales, K. and G. Rapier (2020). The rise and fall of the Jump bike, the electric machine that wooed Uber into a $200 million sale and is now being discarded by the thousands. en-US. URL: https://www.businessinsider.com/what-happened-to-jump-bike-rise-and-fall-uber-2020-6 (visited on 08/29/2022).


2015 IEEE International Congress on Big Data. ISSN: 2379-7703, pp. 592–599. DOI: 10.1109/BigDataCongress.2015.91.


Cruise (2022). We’re Cruise, a self-driving car service designed for the cities we love. en. URL: https://www.getcruise.com/about-us (visited on 08/03/2022).


Davis, L. (2014). “Rolling along the last mile: Bike-sharing programs blossom nationwide”. In: 80, pp. 10–16.


Félix, R., P. Cambra, and F. Moura (2020). “Build it and give ‘em bikes, and they will come: The effects of cycling infrastructure and bike-sharing system in Lisbon”. In: *Case Studies on Transport Policy* 8.2, pp. 672–682. DOI: 10.1016/j.cstp.2020.03.002.


nologies for Automotive (AEIT AUTOMOTIVE), pp. 1–6. DOI: 10.23919/AEITAUTOMOTIVE50086.2020.9307411.


Hampshire, R. C. and L. Marla (2012). “An analysis of bike sharing usage: Explaining trip generation and attraction from observed demand”. In: *Proceedings of


Huang, X., Y. Zhao, C. Ma, J. Yang, X. Ye, and C. Zhang (2016). “TrajGraph: A Graph-Based Visual Analytics Approach to Studying Urban Network Central-


level mobility data from mobile phones suggests that initial compliance with COVID-19 social distancing interventions was high and geographically consistent across the UK”.


— (2012). So How Big Was the Big London Bike Share Expansion? en-US. Suprageography. URL: https://oobrien.com/2012/03/so-how-big-was-the-london-bike-share-expansion/ (visited on 05/30/2022).


— (2019b). Then There Were Eight – Suprageography. en-US. URL: https://oobrien.com/2019/07/then-there-were-eight/ (visited on 08/29/2022).
— (2022a). @fluctuo @0jhl @joedunckley @ibikebrighton @MayorofLondon @SebDance @nuttyxander @tier_mobility @humanforest.uk Currently 9000 Santander Cycles (it was 11600 in April), 3000 Lime ebikes, 1300 HumanForest ebikes, 500 TIER ebikes, 500 Dott ebikes, 8 Beryl Cargo ebikes. Plus 4100 escooters (TIER/Lime/Dott). 500 Santander Cycle ebikes soon. 200 HumanForest eMopeds for delivery drivers only. en. Tweet. URL: https://twitter.com/oobr/status/1554077277654323200 (visited on 08/30/2022).


BIBLIOGRAPHY


Binomial Regression”. In: The American Journal of Drug and Alcohol Abuse 32.1, pp. 41–53. DOI: 10.1080/00952990500328539.


Sustainable Development, p. 188. URL: http://docs.wbcsd.org/2001/12/Mobility2001_FullReport.pdf.


Yang, Y., A. Heppenstall, A. Turner, and A. Comber (2019). “A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban


