

# Considering Context and Dynamics: A Classification of Transit-Oriented Development for New York City

**Abstract:** Transit-Oriented Development (TOD) is a widely recognised planning strategy for encouraging the use of mass and active transport over other less sustainable modes. Typological approaches to TOD areas can be utilised to either retrospectively or prospectively assist urban planners with evidence-based information on the delivery or monitoring of TOD. However, existing studies aiming to create TOD typologies overwhelmingly concentrate input measures around three dimensions of: density, diversity and design; which might be argued as not effectively capturing a fuller picture of context. Moreover, such emphasis on static attributes overlooks the importance of human mobility patterns that are signatures of the dynamics of cities.

This study proposes a framework to address this research gap by enhancing a conventional TOD typology through the addition of measures detailing the spatiotemporal dynamics of activity at transit stations; implemented for the selected case study area, New York City.

**Keywords:** Transit-Oriented Development; TOD Typology; Self-Organising Map; Urban Dynamics; Public Transit

## 1.0 Introduction

Transport Oriented Development (TOD) is considered as a type of sustainable urban development focusing on encouraging transit ridership through providing high density and mixed-use development within walking distance (e.g. 400–800 m; or 5–10-min walk) of public transport facilities (Thomas et al., 2018). The main objective of TOD advocates delivering a favourable environment consisting of urban forms that are highly compact, of mixed-use, pedestrian- and cycling-friendly, and develop neighbourhoods with the vicinity of public transport hubs (i.e. transit stations). Such influences are commonly within a framework referred to as the ‘three-Ds’: namely, high density in development, diversity in land use and good urban design (Cervero and Kockelman, 1997). This development pattern has been widely recognised and accepted as a leading planning strategy by most planning agencies around the world, exemplified by extensive cases in North American and European cities (Lierop et al., 2017; Staricco and Brovarone, 2018), China (Xu et al., 2017), South Korea (Sung and Choi, 2017) and so forth.

TOD principally aims to address common urban transportation challenges associated with automobile dependence, such as traffic congestion and parking difficulties, air quality and noise pollution, excessive greenhouse gas emission, public health and wellbeing-related issues (Rodrigue, 2017; Chavez-Baeza and Sheinbaum-Pardo, 2014; Ettema et al., 2016; Hickman and Banister, 2014; Hynes, 2017; She et al., 2017). Although urban planners have adopted a series of actions aimed at reducing the dependence of private automobile use through encouraging more sustainable alternatives including public transit and active travels (i.e. walking and cycling) (Lee et al., 2013; Winters et al., 2017), TOD presents a focus for more comprehensive planning solutions since it effectively integrates both urban land use and transport system planning (Lee et al., 2013; Taki et al., 2017; Papa et al., 2018).

Although TOD can be argued as consistent in its prescriptions for policy-making and planning, extensive studies have illustrated that for TOD to be successful, there is a necessity to be highly sensitive to local specificities. For the purposes of assisting urban planners in establishing new TOD or evaluating existing TOD, context-based TOD typologies have been implemented to differentiate various station catchment areas (Kamruzzaman et al., 2014; Higgins and Kanaroglou, 2016; Lyu et al., 2016; Papa et al., 2018). Existing studies have overwhelmingly differentiated TOD through measures related to the ‘three-Ds’ such as land use mix, residential and commercial density, and floor area ratio. However, other aspects of context, such as socioeconomic variables are neglected (Higgins and Kanaroglou, 2016). Moreover, such static attributes overlook the dynamic context of TODs,

namely, human's mobility, which as others have shown also plays a vital role in the evolution of urban morphologies and functional regions (Wang et al., 2017; Xia et al., 2018).

It is within this context that we expand upon the existing literature to consider a more comprehensive definition of TOD through a broader range of multidimensional inputs, including their dynamic context. A new analytical framework is implemented here for the case study city of New York City, USA. The paper proceeds first to present a literature review of approaches used to build a TOD typology, followed by a Systematic Literature Review (SLR) designed to identify variables commonly considered as important drivers of differentiation between TOD contexts. General information about the case study area is presented in Section 3, followed by a discussion of the range of station catchment areas and a specification of data pre-processing of the 64 selected candidate variables. In Section 3.2, 472 subway stations are categorised into a four-category TOD typology through the implementation of a proposed methodology framework based on Self-Organising Map (SOM). The groups are named and described according to their salient characteristics. In Section 3.3, subway turnstile data are utilised to capture human mobility patterns, using the same framework to create Temporal Clusters featuring five featured travel patterns. In Section 3.4, the two produced clusters are integrated to explore the interaction between static and dynamic features of the TOD areas. Finally, the paper concludes with a discussion suggesting some future work and limitations to the approach.

## 2.0 Literature Review

There are multiple approaches to building a TOD typology, ranging from the qualitative ascription of idealised TOD contexts (Lyu et al., 2016; Higgins and Kanaroglou, 2016) to more quantitative frameworks utilising models of TOD catchments and associated measures drawn for within these areas (Higgins and Kanaroglou, 2016). TOD contexts within the urban environment have been characterised in the literature through various indicators/variables that are argued to have an effect on (or be a result of) the use of public transport. Given the variable definition of TOD extents, study objectives and locations, the specificity of criteria and indicators selected as influential to TOD characteristics vary between studies.

Following the development of TOD-related research, the concepts of the original 'three Ds' model established by Cervero and Kockelman (1997) has been expanded. For instance, Ewing and Cervero (2010) added Destination accessibility, Distance to transit, and an additional non-environmental variable, i.e. Demographics, to the family of 'D variables', formulating the 'five Ds' concept. These concepts were utilised within a Systematic Literature Review (SLR) to identify those TOD related measures used in the recent literature. We utilised the Scopus<sup>1</sup>, Google Scholar<sup>2</sup>, and Web of Science<sup>3</sup> referencing databases and looked for references published between 2009 and 2018. These databases were queried for research in the broadest sense, including journal articles, official documents, guidelines, and so forth; which either created a TOD typology (or indexed TOD features) for major transit stations or focused on analysing the relationship between multidimensional variables around stations and ridership of public transit more generally. Scopus returned 15 studies, the Web of Science and Google Scholar respectively identified 11 and 6240 results. The studies were checked for diversity, both in terms of geographic context (i.e. the location of the case study) and type (e.g. journal article, governmental documents/policies); and secondly, the quality of the reviewed studies was considered in terms of the influence of the academic studies (measured by the times cited) and the authority of the governmental documents/policies.

Through this process, 29 studies were identified, and from these, a set of common variables selected that are presented in Table 1. Although many align with the 'five Ds', most of the studies are not comprehensive in coverage of all domains of the 'five Ds'. Moreover, some of the variables employed in these studies do not align with the 'five Ds', implying broader or context-specific considerations. The candidate variables could broadly be categorised into four domains, namely, Land

---

<sup>1</sup> <https://www.scopus.com/>

<sup>2</sup> <https://scholar.google.co.uk/>

<sup>3</sup> <https://wok.mimas.ac.uk/>

Use and Built Environment, Location and Accessibility, Socioeconomic and Demographic, and Transit-related.

Input to the Land Use and Built Environment, and Location and Accessibility domains were drawn from a range of sources including the Census and other public survey data, but also Points of Interest (POI) databases either as supplements or alternatives to conventional land-use measures (see Wang et al., 2016; Lyu et al., 2016; Wang et al., 2017). These studies advocated that POI data may capture finer-grained and more up to date information depicting the land use composition and urban facilities. Moreover, other variables, such as the type of dwelling, type of tenure, building height, building age, and average travel time/distance to workplace/transit station, also take a relatively large share of commonly-used variables in these two domains; which may be as a result of their reasonably common availability and broadly understood definitions.

Within the Socioeconomic and Demographic domain, typical variables identified from the literature included the median household income, household vehicle ownership, educational attainment, and occupation type. As for demographic variables, the “seventh D”<sup>1</sup> in D-variables (Ewing and Cervero, 2010, 267), mainly including age composition and household size/type.

Transit-related attributes had high salience in the studies identified; and indicators included measures such as daily/weekly ridership, frequency of metro services, or peak passenger load/frequency in the transit station. Although of utility, such measurements were typically limited in temporal resolution (i.e., weekly ridership) or were somewhat static (e.g., morning peak ridership volume), and therefore only had limited account for actual periodic variation in patterns of use. Given that spatiotemporal data related to transit have become more prevalent, some of the studies, such as Wang et al. (2016), Zhou et al. (2017), Wang et al. (2017) and Kim et al. (2018), utilised attributes from trip transaction data extracted from a smart card system to calibrate more real-time measures. In addition to transit flow data, human mobility was also inferred by Wang et al. (2017) through mobile application data from an online mapping system.

Within the Location and Accessibility domain, travel distance/time from the transit nodes to the main working places are typically adopted by many of the reviewed studies. Moreover, proximity to activities is also a commonly used variable indicating the connectivity between transit nodes and the surrounding environment. Perceived attributes are also employed by some reviewed studies, such as cleanness and safety of the transit station. However, due to the difficulty of quantifying and data availability, most of the reviewed studies do not include these types of variables.

---

<sup>1</sup> The “sixth D” in D-variables is considered as demand management (Ewing and Cervero, 2010, 267)

Domain	TOD Indicators used in the reviewed studies	Atkinson-Palombo and Kuby (2011)	Austin et al. (2010)	Bhattacharjee and Goetz (2016)	Center for Transit-Oriented Development (2013)	Chen et al. (2009)	Chorus and Bertolini (2011)	Dirgahayani and Choerunnisa (2018)	Guo et al. (2018)	Higgins and Kanaroglou (2016)	Huang et al. (2018)	Ivan et al. (2012)
<b>Land Use &amp; Built Environment</b>	Average Block Size/ Length		*		*							
	Degree of Functional Mix						*	*				
	Housing Unit /Density											
	Land Cover											
	Mixed-ness of Land Use (Diversity/Entropy)	*		*		*		*	*	*	*	
	Population /Residents Density	*	*	*	*	*	*		*	*	*	*
	Job Density /Business Intensity	*	*	*		*				*	*	
	Floor Area Ratio (FAR)							*				
	Property/Land Values					*				*	*	*
	Street Network/Intersection Density									*	*	*
	Type of Dwelling/Tenure											*
<b>Transit-related</b>	Year Structure Built (Building Age)											
	Attributes of Transit Stations						*					
	Frequency of Metro Services											
	Interchange to Other Transit Modes								*			*
	Number of Directions Served (Bus/Subway)							*	*			*
	Number of Nearby Transit Hubs						*	*	*	*		*
	Parking Facility/Infrastructure	*										*
<b>Location &amp; Accessibility</b>	Utilisation of Transit (Passenger Load/Ridership)					*			*		*	*
	Walkability/Pedestrian Networks/Cyclability				*			*		*	*	
	Accessibility to/ from Station		*					*		*		
	Average Travel Time (to Work/Transit Stations)		*							*		
	Distance to City Centre/CBD		*			*	*	*				
	Perceived Attributes (e.g. Safety, Attractiveness)					*		*		*		
	Proximity of Activities/Amenities at Station					*	*	*		*		
<b>Socioeconomic &amp; Demographic</b>	Ethnic/Age Composition				*	*				*		
	Household Income	*	*			*				*		
	Household Type/Size	*	*									
	Occupation Type/Education Level	*		*			*			*		*
	Transport Mode to Work		*							*		
Vehicles Ownership		*			*							
<b>Case Study Area</b>		Phoenix US	9 Cases US	Denver US	Allegheny County US	New York City US	Tokyo Japan	Jakarta & Bandung Indonesia	Tokyo Japan	Toronto region Canada	Arnhem–Nijmegen MA. Netherlands	Ostrava Czech Republic
<b>Number of Transit Stations (Cases)</b>		27	9	NA	NA	468	99	NA	27	372	22	11
<b>Buffer Distance (metres)</b>		800	800	800	NA	NA	700	NA	1500	800	800	700





## 3.0 Contextualising TOD: New York City

The case study area selected for this study is the New York City (NYC), which is the most densely populated cities within the US with an estimated 8.56 million residents distributed over a land area about 777 km<sup>2</sup> (US Census Bureau, 2017). The city is located at the southern tip of the state of New York on the US eastern seaboard, comprising five boroughs, namely, Brooklyn, Queens, Manhattan, Bronx, and Staten Island. The New York City Subway, first opened in 1904, is a rapid transit system that offers 24/7 service across four of the five boroughs of NYC (i.e. Manhattan, Queens, The Bronx, and Brooklyn), which is controlled by the Metropolitan Transportation Authority (MTA). The system spans 27 lines (665 miles of track) and 472 subway stations, facilitating a major transportation mode for residents and visitors to the city (MTA, 2016). According to the subway ridership statistics provided by MTA, in 2016, an average of 5.65 million passengers used the system daily on weekdays and about 5.75 million at the weekends, making it the largest rapid transit system in the US and the seventh busiest worldwide.

### 3.1: Defining Station Catchment Area and Data Pre-processing

Fundamental to any TOD typology is a definition of the contextual area surrounding the transit stations. For this case study we selected an area of 800 m (approximately 0.5 miles) which mirrored the majority of studies conducted within the US (see Atkinson-Palombo and Kuby, 2011; Austin et al., 2010; Nasri and Zhang, 2014; Bhattacharjee and Goetz, 2016). Although most of these studies employed a Euclidean distance buffer (or circular buffer) to define the catchments of transit station areas, it can be argued that a network distance buffer is more suitable since it “more accurately representing the built environment as experienced by someone walking through it” (Oliver et al., 2007, 8). Fig. 1 illustrates 50 m-trimmed street network-based catchment areas (800 m walking distance) of the NYC subway stations. A zoomed-in inset map, on the upper left corner, shows an example circular buffer and a network buffer at Metropolitan Avenue Station. It is clear that a circular buffer area is less effective representation given the surrounding street density and available paths to walk. Additionally, in some other locations within a circular buffer, the walking distance is longer than 800 m due to more facilitating urban structure, such as block size. A more systematic discussion comparing the influence of these two types of buffer are detailed in Oliver et al. (2007).

Moreover, two inset maps located on the right side of Figure 1 highlight a stretch of census blocks<sup>1</sup> along with their defined station catchment areas. Each catchment contains census blocks, for example, the catchment area of Mets-Willets Point station is formed by seven census blocks; Lefferts Blvd station catchment intersects with 25 census blocks, which are converted into proportion based on their area of overlap. The proportion is subsequently used as a weight ( $w_i$ ) to calculate the weighted average value of selected variables. Eq. (1) illustrates how these weights were used to calculate values attributed to census block where they intersected with the catchment areas.

$$\bar{x} = \frac{\sum_{i=1}^n (x_i * w_i)}{\sum_{i=1}^n w_i}$$

*Eq. 1  $\bar{x}$  is the weighted mean;  $x_i$  is an original value;  $w_i$  is the weight (i.e. the proportion of the area occupied by a specific census block in station catchment area).*

For other variables at the finer spatial resolution, particularly spatial points extracted from the NYCOD and NYCP (e.g. street trees, bus stops), these were first aggregated to the station catchment areas where they were located and either a density or percentage value calculated.

---

<sup>1</sup> Census blocks, the smallest geographic area for which the US Census Bureau collects and tabulates census data (US Census Bureau, 2011).

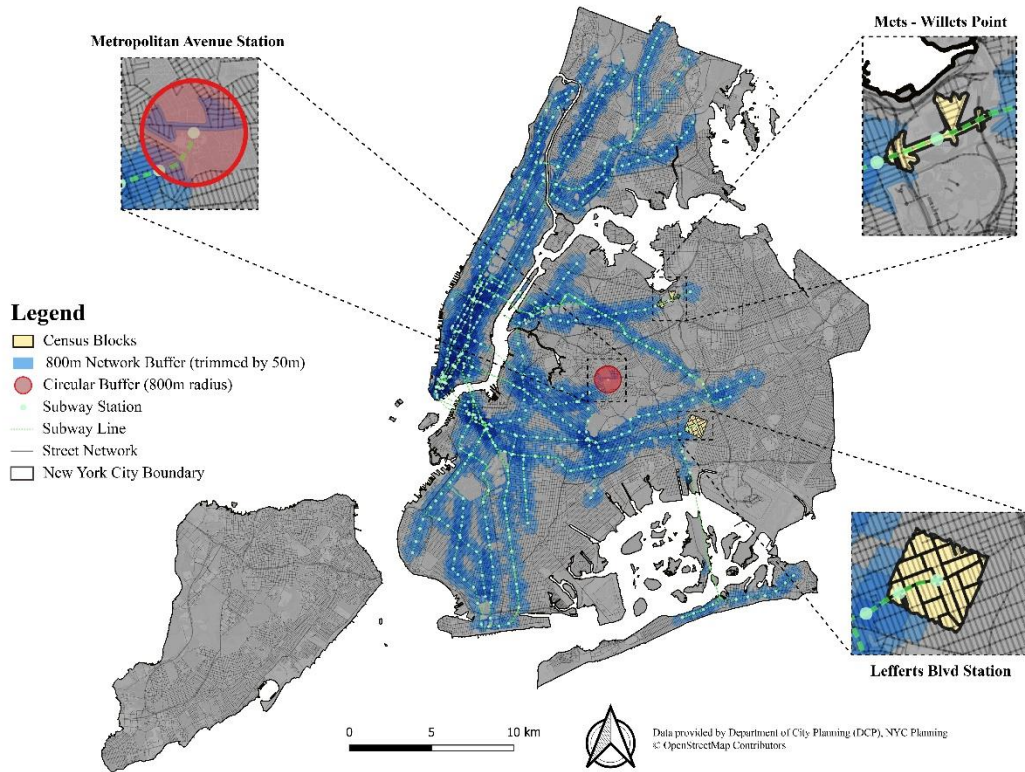


Fig. 1 The New York City subway system and catchment areas (800m walking distance) with highlights of census blocks

The selection of variables was primarily guided by findings from Section 2 alongside further consideration of the quality and availability of potential variables within the case study area. As such, variables selected for this study were categorised into the four previously identified domains: Land Use and Built Environment, Transit-related, Location and Accessibility, and Socioeconomic and Demographic. Variables representative of these domains were extracted from the following seven open data sources: American Community Survey (ACS)<sup>1</sup>, National Walkability Index (NWI)<sup>2</sup>, Smart Location Database (SLD)<sup>3</sup>, NYC Open Data (NYCOD)<sup>4</sup>, NYC Planning (NYCP)<sup>5</sup>, Metropolitan Transportation Authority (MTA)<sup>6</sup>. Table 2 presents the final 64 variables selected for this study alongside a brief description. After the selected variables were assembled for each of the subway station catchment, the Box-Cox transformation (Eq. (2); Box and Cox, 1964) was adopted to transform non-normal variables values to approximate a normal distribution. Furthermore, given that the assembled variables are measured on different scales, z-scores were implemented as a standardisation (Eq. (3)). This frequently used technique creates a transformed variable with a mean of zero and unit of standard deviation.

$$x_i' = \begin{cases} \frac{x_i^\lambda - 1}{\lambda} & , \text{ if } \lambda \neq 0; \\ \log x_i & , \text{ if } \lambda = 0. \end{cases}$$

Eq. 2 where  $x_i'$  is the transformed value;  $\lambda$  ranging from -5 to 5, which can be estimated using the profile likelihood function to achieve 'optimal value'

<sup>1</sup> <https://www.census.gov/programs-surveys/acs/>

<sup>2</sup> <https://catalog.data.gov/dataset/walkability-index>

<sup>3</sup> <https://www.epa.gov/smartgrowth/smart-location-mapping>

<sup>4</sup> <https://opendata.cityofnewyork.us/>

<sup>5</sup> <https://www1.nyc.gov/site/planning/data-maps/open-data.page>

<sup>6</sup> <http://web.mta.info/developers/turnstile.html>



$$z_i = \frac{x_i - \mu}{\sigma}$$

*Eq. 3 where  $z_i$  is the standardised value ,  $x_i$  is an original value,  $\mu$  is the mean of  $x_i$ , and  $\sigma$  is the standard deviation from the mean.*

Database	Code	Domain	Variables Title	Description
ACS	B01001	Socioeconomic & Demographic	Age: 0-4	% of Population Aged between 0 and 4
		Socioeconomic & Demographic	Age: 5-14	% of Population Aged between 5 and 14
		Socioeconomic & Demographic	Age: 15-19	% of Population Aged between 15 and 19
		Socioeconomic & Demographic	Age: 20-24	% of Population Aged between 20 and 24
		Socioeconomic & Demographic	Age: 25-44	% of Population Aged between 25 and 44
		Socioeconomic & Demographic	Age: 45-64	% of Population Aged between 45 and 64
		Socioeconomic & Demographic	Age: 65&above	% of Population Aged 65 and above
	B08303	Location & Accessibility	TTtW: < 5 mins	% of Workers whose Travel Time to Work is less than 5 minutes
		Location & Accessibility	TTtW: 5-14 mins	% of Workers whose Travel Time to Work is between 5 and 14 minutes
		Location & Accessibility	TTtW: 15-29 mins	% of Workers whose Travel Time to Work is between 15 and 29 minutes
		Location & Accessibility	TTtW: 30-44 mins	% of Workers whose Travel Time to Work is between 30 and 44 minutes
		Location & Accessibility	TTtW: 45-59 mins	% of Workers whose Travel Time to Work is between 45 and 59 minutes
		Location & Accessibility	TTtW: > 60 mins	% of Workers whose Travel Time to Work is longer than 60 minutes
	B11016	Socioeconomic & Demographic	HT: 1-person	% of 1-Person Household
		Socioeconomic & Demographic	HT: 2-person	% of 2-Person Household
		Socioeconomic & Demographic	HT: 3-person	% of 3-Person Household
		Socioeconomic & Demographic	HT: 4+-person	% of 4 or more Person Household
	B15003	Socioeconomic & Demographic	EA: No school	% of Population have no qualifications
		Socioeconomic & Demographic	EA: Elementary school	% of Population attained kindergarten to 5th grade
		Socioeconomic & Demographic	EA: Middle school	% of Population attained 6th to 8th grade
		Socioeconomic & Demographic	EA: High school	% of Population attained 9th to 12th grade
		Socioeconomic & Demographic	EA: College / Bachelor	% of Population attained College or Bachelor's degree
		Socioeconomic & Demographic	EA: Master / Doctorate	% of Population attained Master's or Doctorate Degree
	B19013	Socioeconomic & Demographic	Median Income	Household Median Income in the past 12 months
	B25003	Land Use & Built Environment	Tenure: Owner	% of Housing Unit occupied by Owner
		Land Use & Built Environment	Tenure: Renter	% of Housing Unit occupied by Renter
	B25024	Land Use & Built Environment	US: Detached	% of Housing Unit categorised as detached
		Land Use & Built Environment	US: Attached	% of Housing Unit categorised as attached
		Land Use & Built Environment	US: Apartment	% of Housing Unit categorised as apartment (from 2 to 50 units)
	B25034	Land Use & Built Environment	YB: 2010 / Later	% of Building built in 2010 or later
Land Use & Built Environment		YB: 2000 - 2009	% of Building built in between 2000 and 2009	
Land Use & Built Environment		YB: 1980 - 1999	% of Building built in between 1989 and 1999	
Land Use & Built Environment		YB: 1960 - 1979	% of Building built in between 1960 and 1979	
Land Use & Built Environment		YB: 1940 - 1959	% of Building built in between 1940 and 1959	
Land Use & Built Environment		YB: 1939 / Earlier	% of Building built in 1939 or earlier	

		Socioeconomic & Demographic	OT: M.B.S.A.	% of Workers in Management, Business, Science, and Art Occupations
		Socioeconomic & Demographic	OT: S.	% of Workers in Service occupations
	B24010	Socioeconomic & Demographic	OT: S.O.	% of Workers in Sales and office occupations
		Socioeconomic & Demographic	OT: N.C.M.	% of Workers in Natural resources, construction, and maintenance occupations
		Socioeconomic & Demographic	OT: P.T.M.	% of Workers in Production, transportation, and material moving occupations
		Socioeconomic & Demographic	VA: No-vehicle	% of Housing Units have no vehicle
	B25044	Socioeconomic & Demographic	VA: 1-vehicle	% of Housing Units have 1 vehicle
		Socioeconomic & Demographic	VA: 2-vehicle	% of Housing Units have 2 vehicles
		Socioeconomic & Demographic	VA: 3+-vehicle	% of Housing Units have 3 or more vehicles
	B01003	Land Use & Built Environment	Population Density	Population Density
<b>NWI</b>	D4a	Location & Accessibility	D4a	Distance from the population-weighted centroid to the nearest transit stop (meters)
	D1c	Land Use & Built Environment	D1c	Job Density
<b>SLD</b>	D2a_EpHHm	Land Use & Built Environment	D2a_EpHHm	Employment and Household Entropy
	D3a	Land Use & Built Environment	D3a	Road Network Density
	D4d	Location & Accessibility	D4d	Aggregate frequency of transit service per square mile
	CSCL	Land Use & Built Environment	Intersection Density	Street Intersection Density Calculated from the Street Centreline
	STC	Land Use & Built Environment	Tree Density	Street tree density
<b>NYCOD</b>	Bicycle	Land Use & Built Environment	Bike Facilities	Citi-Bike, Bicycle Routes and Parking Shelters density
	Bus	Land Use & Built Environment	Bus Facilities	Bus Stops Density
	Parking	Land Use & Built Environment	Parking Facilities	Parking meters/lots density
	POI	Land Use & Built Environment	POI	Point of Interest Data: contains seven land-use types
<b>NYCP</b>	MapPLUTO	Land Use & Built Environment	Landuse	Land Use: contains seven land-use types
<b>MTA</b>	Turnstile	Transit-related	Turnstile: Entry	Entry Counts for all turnstile data by every 4 hours per day
		Transit-related	Turnstile: Exit	Exit Counts for all turnstile data by every 4 hours per day

Table 2 Final variable selection and basic description

### 3.2: Contextualising TOD

After the assembly of the normalised and standardised input data; similarity in the context of subway stations was explored by the application of a Self-Organising Map (SOM). The Self-Organising Map (SOM), also known as Kohonen Map, is a single layer feedforward artificial neuron network, which is trained by unsupervised, competitive learning as a tool for “visualisation and analysis of high dimensional data” (Bação and Lobo, 2010, 4). The SOM translates high-dimensional inputs into a low-dimensional space, also referred to feature map that is configured by the number of pre-defined neurons arranged on a regular lattice (e.g. a rectangular or hexagonal topology), through ‘fitting’ a grid of nodes to the data over a fixed number of iterations. The resulting map allows a graphical presentation of the data that can be easily interpreted by map-readers, which can be further classified by the machine learning techniques designed for low dimensionality (Bara et al., 2018; Spielman and Folch, 2015; Natita et al., 2016). Numerous studies have highlighted the utility of SOM for visualising complex, nonlinear statistical relationships within high-dimensional data (Yin, 2008; Bação and Lobo, 2010; Das et al., 2016; Miljković, 2017). The method is suitable for this application given the multiplex of measures assembled. Moreover, even after the application of Box-Cox transformation, some variables remained not normally distributed, which may have caused some problems if we directly adopted conventional feature extraction methods such as principal components analysis (PCA), since the underlying assumptions of these techniques are not satisfied (Das et al., 2016). Accordingly, Demartines and Blayo (1992) note that the SOM is not very sensitive to the normal distribution when the input data contain high dimensionality.

Several studies have highlighted the potential applications of SOM in terms of building typology for urban contexts (Jain et al., 2018; Schäfer et al., 2018; Spielman and Thill, 2008; Arribas-Bel and Schmidt, 2013), and specifically within the context of TOD: Sohn (2013) presented an application of SOM for Seoul, South Korea, metro station areas.

Several parameters need to be specified in advance when fitting a SOM, including the number of neurons ( $M$ ), the range of the learning rate and its decline pattern ( $\alpha$ ), the shape/type of the neuron, and the type neighbourhood function (Spielman and Folch, 2015). The first step before training the SOM is to define an appropriate number of neurons that are used to configure the network. A small feature map (i.e. the number of observations far exceeds the number of neurons) results in a generalisation, whereas a large map allows a specific location in geographic space (subway stations, in this study) to be projected to a particular location in the corresponding attribute space, representing specific properties (Spielman and Thill, 2008). A useful ‘rule of thumb’ (Eq. (4)) suggested by Tian et al. (2014), is employed here to determine the number of neurons. Since the 472 stations configure the observations ( $N$ ), 108 ( $M$ ) neurons, projected on a 12 by 9 grid, are accordingly generated to structure the SOM. For the remaining parameters, these were set following an objective of maximising SOM quality through minimisation of the average quantisation error (QE) statistic. We tested the value of QE generated by using various combinations of different SOM parameters following Natita et al. (2016). The results of these experiments are shown in Table 3, with the combination of a rectangular topology, the bubble neighbourhood function and a linear decline in learning rate (ranging from 1.0 to 0.01) resulting in the smallest average QE (3.41). Thus, this combination of parameters is eventually adopted to train the SOM network for creating spatial clusters.

Table 3 Result of SOM parameter settings for building TOD typology

Test	Topology/Shape	Learning Rate Type	Neighbourhood Type	Average QE	Learning Rate range
1	Hexagon	Linear	Bubble	3.51	1.0-0.01
2	Hexagon	Inverse	Bubble	3.59	1.0-0.01
3	Hexagon	Linear	Gaussian	4.21	1.0-0.01
4	Hexagon	Inverse	Gaussian	4.79	1.0-0.01
5	Rectangle	Linear	Bubble	3.41	1.0-0.01
6	Rectangle	Inverse	Bubble	3.69	1.0-0.01
7	Rectangle	Linear	Gaussian	4.18	1.0-0.01
8	Rectangle	Inverse	Gaussian	4.75	1.0-0.01

$$M \approx 5\sqrt{N}$$

Equation 4 Where  $M$  is the number of neurons, which is an integer close to the result of the right-hand side of the equation and  $N$  is the number of observations.

To reduce the complexity of the computed SOM feature maps further, a hybrid hierarchical k-means (H-K-means) algorithm was applied to aggregate the neurons into groups sharing similar attributes (Chen et al., 2005; Kassambara, 2017). The procedure of this algorithm can be summarised into three steps: firstly, agglomerative hierarchical clustering is applied to the input data and generated tree (i.e. dendrogram), which is cut into  $k$  number of clusters; secondly, the cluster centroids (i.e. the mean value) are computed for each group; and finally, these cluster centroids are utilised as the initial centres for the k-means algorithm (Kassambara, 2017). To select an appropriate number of clusters, a clustergram was created that demonstrates a weighted mean of the first component of a PCA for each cluster centre across a range of tested  $k$  values, where the width of each line represents the number of observations (i.e. neurons in SOM). The detailed rationale of this technique has been discussed elsewhere (see Schonlau, 2002); but generally, the logic is to find the point where the centroids of the clusters are as dissimilar as possible (well-spaced). According to the clustergram shown in Fig. 2, it is easily observed that when the number of clusters reaches four, the difference between cluster centroids (red dots) is maximised (after  $k = 4$ , these centroids are getting close to each other; when  $k = 5$ , two cluster centroids are nearly overlapped, indicating relatively bad clustering results).

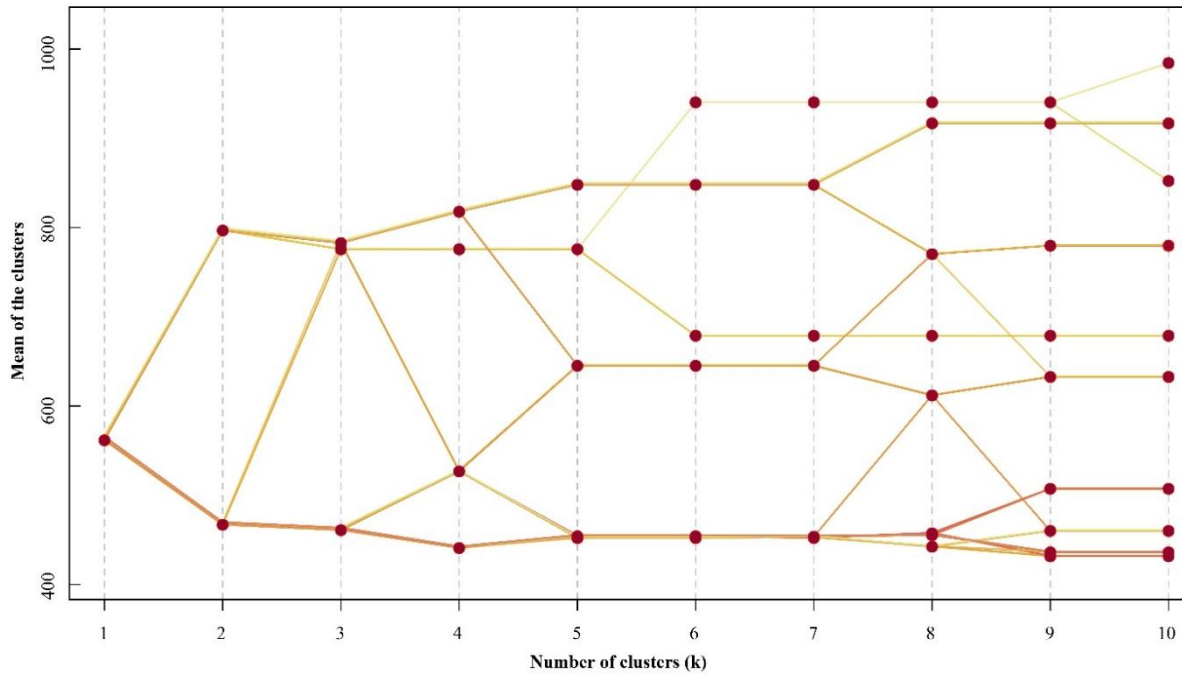


Fig. 2 Clustergram for selecting number of clusters differentiating TOD Typologies.

The result from clustering the 108 neurons is shown in a 2-dimensional plane (Fig. 3) and is mapped in Fig. 4 portraying the geographic distribution of TOD typologies for NYC. The geographic distribution follows a broadly concentric circle-shape, radiating away from the central area of Manhattan.

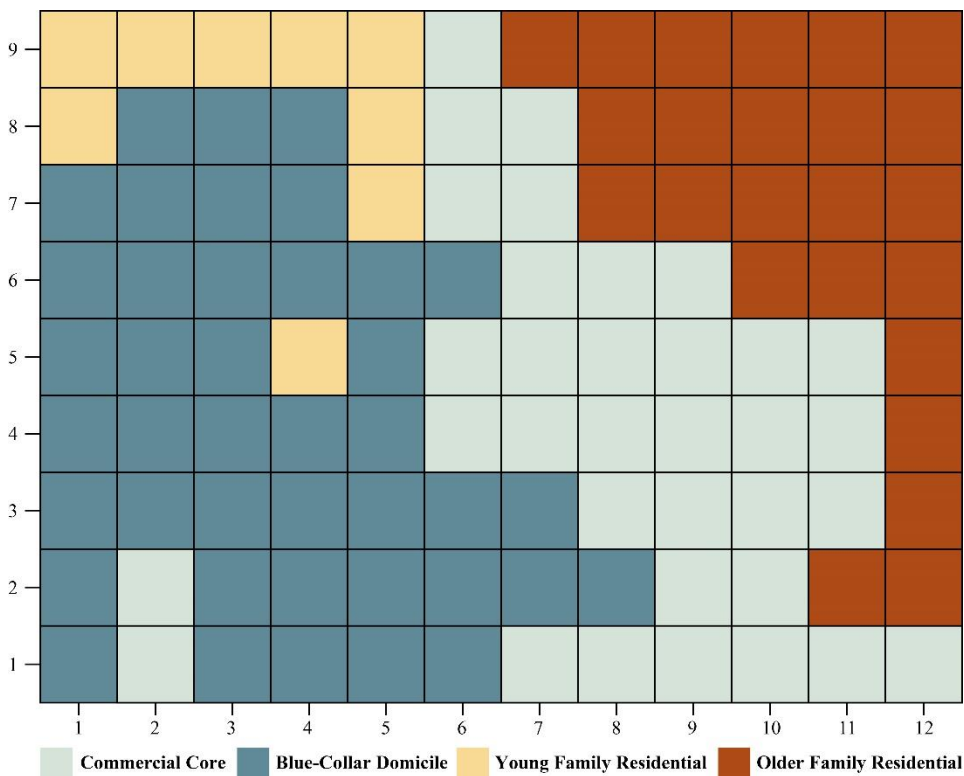


Fig. 3 TOD Typologies by SOM nodes (presented on a 12\*9 grid).

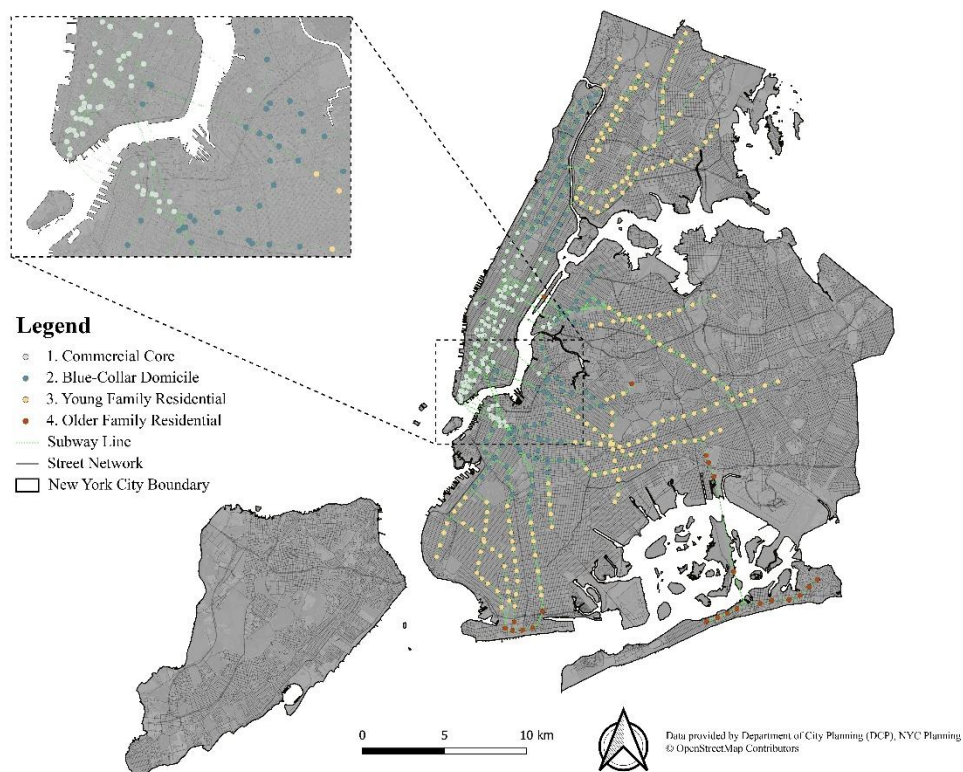


Fig. 4 Geographic Distribution of TOD Typologies

To ascertain the most salient characteristics of the clusters, index scores (i.e.  $x/\bar{x} * 100$ ) were calculated for the input variables and displayed within each cluster in Fig. 5. These scores indicate the (over-) underrepresentation of a target characteristic compared to the regional average value (i.e. a score of 100). An index score 50 would hence equate to a rate that is half the average, and 200 would be double. Utilising both the map and scores, descriptive profiles were generated.

### Cluster 1: Commercial Core

This cluster is characterised by commercial areas with a highly educated (Master's or Doctorate) population, aged between 25 and 44, including many of those who are employed in well-paid management, business, science and arts occupations. Residents of such areas are more likely to live in apartments (built after the 1980s) consisting of one- or two-person households. These areas are characterised by an extremely high job density, high level of traffic permeability, and plenty of public services, commercial and mixed-use properties, accompanied with a mature infrastructure for cycling and a high level of accessibility of public transit.

### Cluster 2: Blue-Collar Domicile

Residents of this typology have an age distribution closer to the regional mean, who have increased prevalence to live with family in rented apartments that are situated in areas with high population density, forming a typical three-person household size. Many more of these residents are likely to have occupations within the areas of service and production, transportation, and material moving sectors. Additionally, the annual median income earned by residents of these areas is much lower. The physical environment is characterised by detached properties and apartment constructed in the 1940s and typically linked with adequate parking infrastructures.

### Cluster 3: Young Family Residential

These areas are characterised by residential occupants with (pre-)school-age children. Many residents live in the detached property located in boroughs outside Manhattan. Given the distance of travel to work (more than 60 min), the car dependency of these areas is higher than the regional average, also demonstrated by the high household vehicle availability.

### Cluster 4: Older Family Residential

Populations living within this cluster can be broadly characterised by college-educated middle-age residents (aged between 45 and 64) who are likely to own a detached property built between the 1940s to 1970s and located on the periphery of NYC. Many residents live in relatively large households with dependent children (aged from 5 to 19). Residents of this group show high use of private automobiles for commuting, manifested by high levels of vehicle availability (two or more cars) at more than four times the regional average.

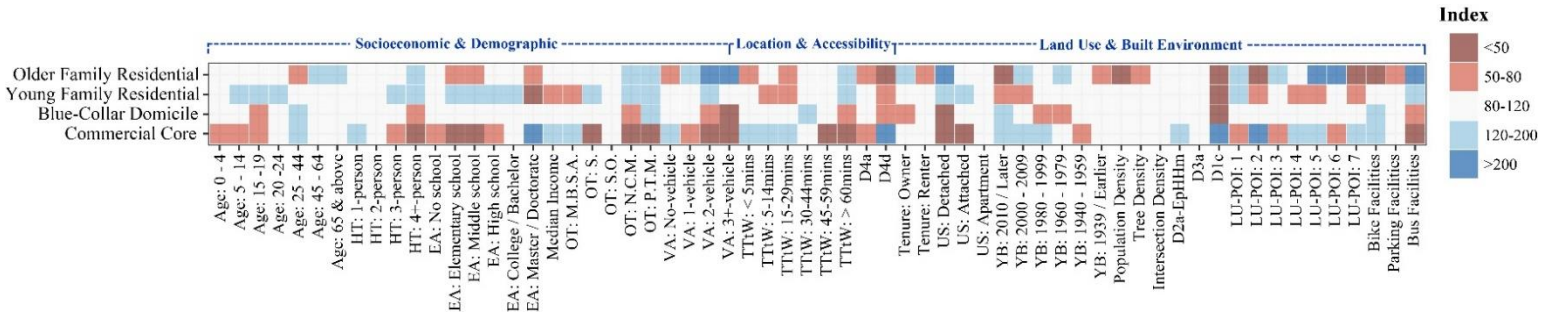


Fig. 5 Index Scores by four TOD Typologies.



### 3.3 Classifying Temporal TOD Dynamics

The space-time dynamics of TOD localities were considered through subway turnstile data supplied by the MTA. This dataset provides a variety of information on subway station entries and exits, organised into four-hourly daily time bands (i.e. six intervals a day); with the period 2015 to 2016 selected. The turnstile data was aggregated by days of the week, and for each station (a station contains several turnstiles) created 30 variables (six-time bands, five working days) for entry counts and a further 30 variables for exits. To provide further insight into those stations sharing similar patterns of transit use, the analytical framework applied earlier to station contextual data was replicated for the subway turnstile ingress and egress. This included the data pre-processing (e.g. normalisation and standardisation), alongside SOM construction and clustering. A clustergram was again used to select an appropriate number of clusters (see Fig. 6), with  $k = 5$  selected. These ‘Temporal Clusters’ are mapped in Fig. 7.

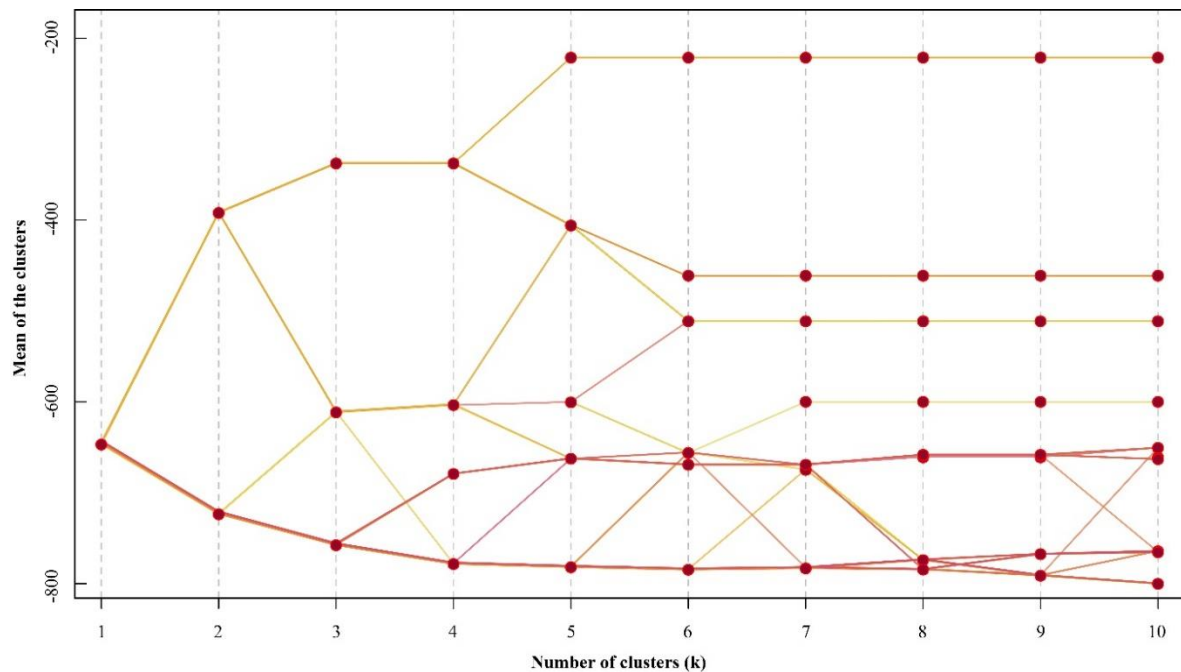


Fig. 6 Clustergram for selecting number of clusters differentiating Temporal Clusters.

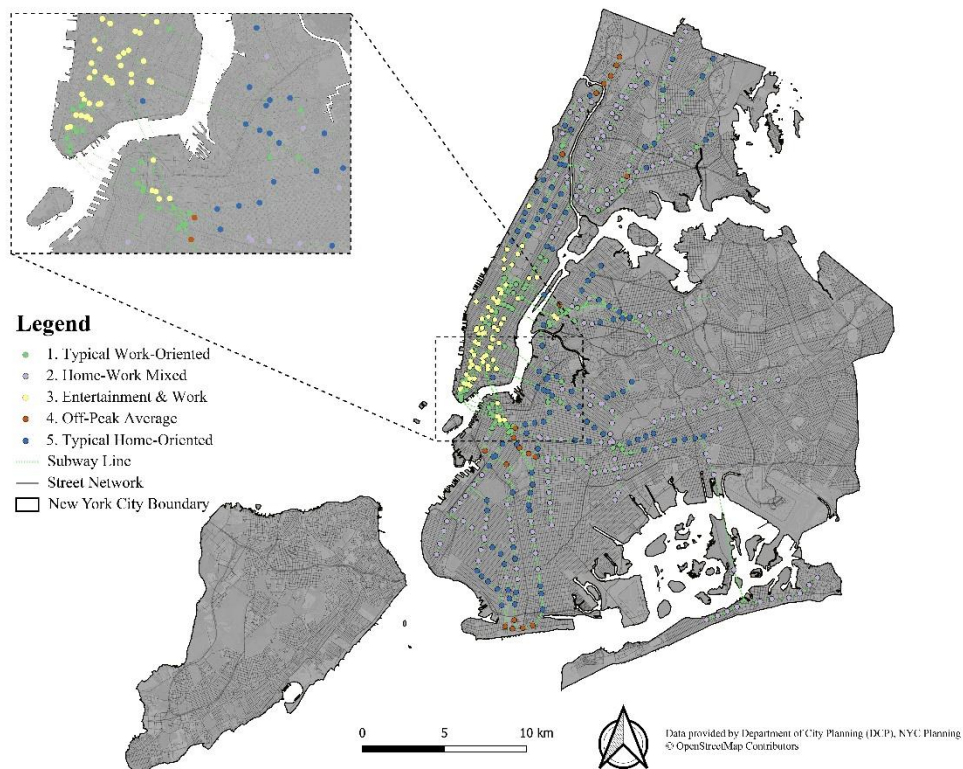


Fig. 7 Geographic Distribution of Temporal Clusters

To ascertain the most salient characteristics of the clusters, a further set of index scores were created for the temporal clusters, using the method previously described in Section 3.2. A series of heatmaps show these scores in Fig. 8 for the five Temporal Clusters. Utilising both the map and scores, descriptive profiles were generated from these insights.

#### Cluster 1: Typical work-oriented

Stations within this cluster are mainly located in the Lower and Midtown areas of Manhattan, downtown areas of Brooklyn and Long Island City. They feature a typical ‘double-humped’ (morning and evening) subway travel pattern associated with workplace-oriented usage. In the morning peak, low inbound passenger flow is identified accompanying with a high outbound flow; while during the evening peak, stations have high inbound flow and a low outbound flow. The role of these stations switches during workdays: from a ‘major destination’ in the morning to ‘major origin’ in the evening.

#### Cluster 2: Home-work mixed

Stations classified by this cluster are mainly located outside Manhattan, featuring a mixed subway travel pattern. During the morning peak and even earlier, stations exhibit a high volume of inbound passenger flows and a high volume of outbound flows.

#### Cluster 3: Entertainment and work

Stations within this cluster are predominantly located in either Downtown or Midtown Manhattan, occupying more than half of subway stations in Manhattan. These stations meet a low inbound and high outbound passenger flow during the morning but reverse this pattern during the evening. Moreover, there is additionally a large volume of inbound flows during the midnight-to-late-at-night period, which may be a result of these destination being the popular place of departure from evening events.

#### Cluster 4: Off-Peak Average

Stations of this group are distributed reasonably randomly across New York. This cluster also consists of stations exhibiting moderate levels of passenger flow, which are very close to the average. More

generally, these are less popular stations and experience fewer passengers during commuting peak periods.

### Cluster 5: Typical Home-Oriented

Although a fraction of stations from this group can be identified in the northern part of Central Park, most stations are located outside Manhattan (especially in the periphery of NYC). These stations also experience the ‘double-humped’ travel pattern, however, high inbound and low outbound passenger flow during the morning peak, with the reverse during the evening.

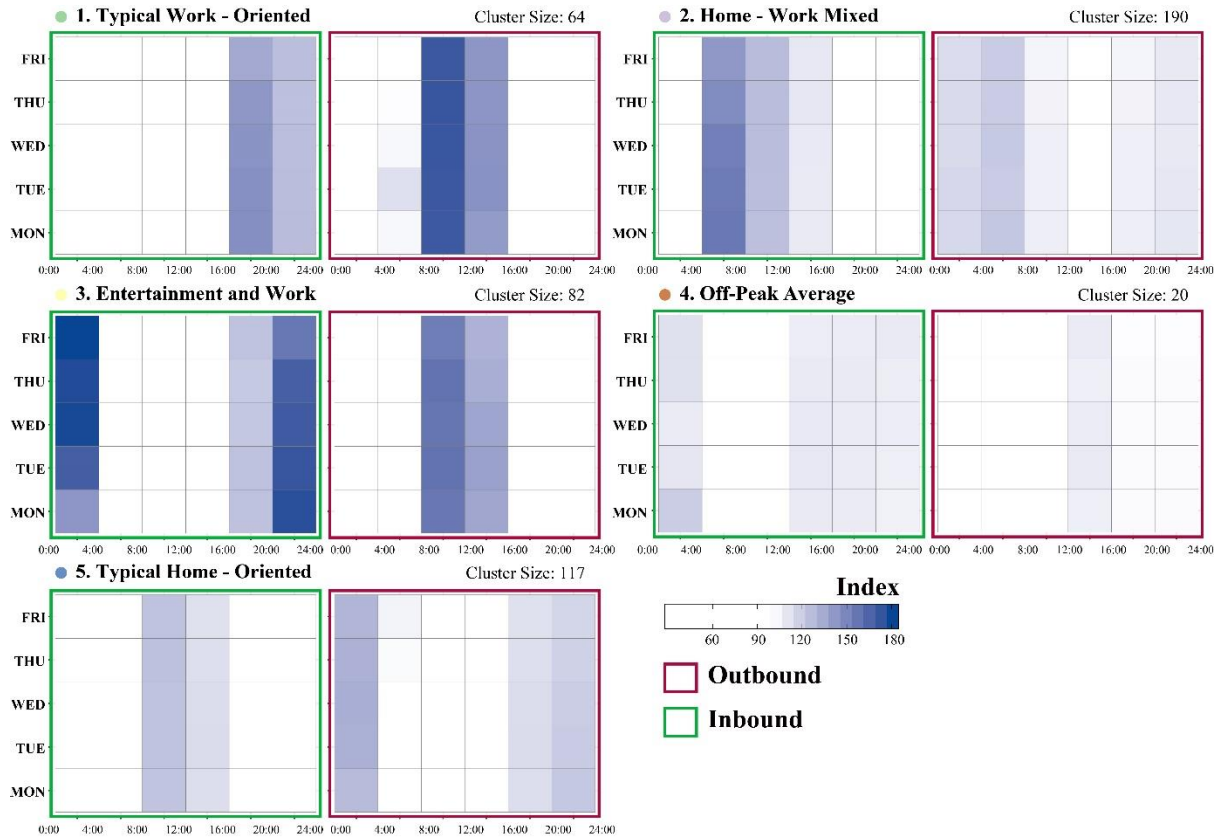


Fig. 8 Inbound and outbound index value of five Temporal Clusters (presented in a ‘weekly travel profile’ manner). In order to achieve better visualisation result, all values less than 100 (less than the mean value) are presented by white.

### 3.4 Integrating Context and Space-Time Dynamics

An overarching purpose of this work has been to extend an existing framework for the creation of TOD typologies to examine both context and dynamics. As such, in this section, we explore the intersection of our two created classifications. Fig. 9 presents an alluvial diagram showing the proportion of subway stations categorised at the intersection of these two classifications for the NYC extent. There is reasonable consistency between these two classifications with some emerging differences.

As might be expected, stations with their context classified as ‘Commercial Core’ predominantly correspond to ‘Typical Work-Oriented’ and ‘Entertainment & Work’ temporal clusters, manifesting typical workplace-oriented function of these TOD areas. Similar temporal patterns can be observed in those stations categorised as ‘Blue-Collar Domicile’ which splits between ‘Typical Home-Oriented’ and ‘Home-Work Mixed’ which might be expected given more residential-oriented usage.

TOD areas categorised as ‘Young Family Residential’ unsurprisingly predominantly correspond with the temporal cluster ‘Typical Home-Oriented’ and ‘Home-Work Mixed’. Stations are both major origins and destinations during peak times, which may be a result of proximity to local employment centres or schools. Given that many of the residences of this cluster are students, the high volume of (early-) morning peak flows may partially be explained by educational establishment opening times.

Within TOD stations classified as ‘Older Family Residential’, there is correspondence to the temporal clusters ‘Home-Work Mixed’ and ‘Off-Peak Average’. There are likely demographic drivers of these patterns alongside a higher rate of private vehicle ownership as a result of their more suburban locations.

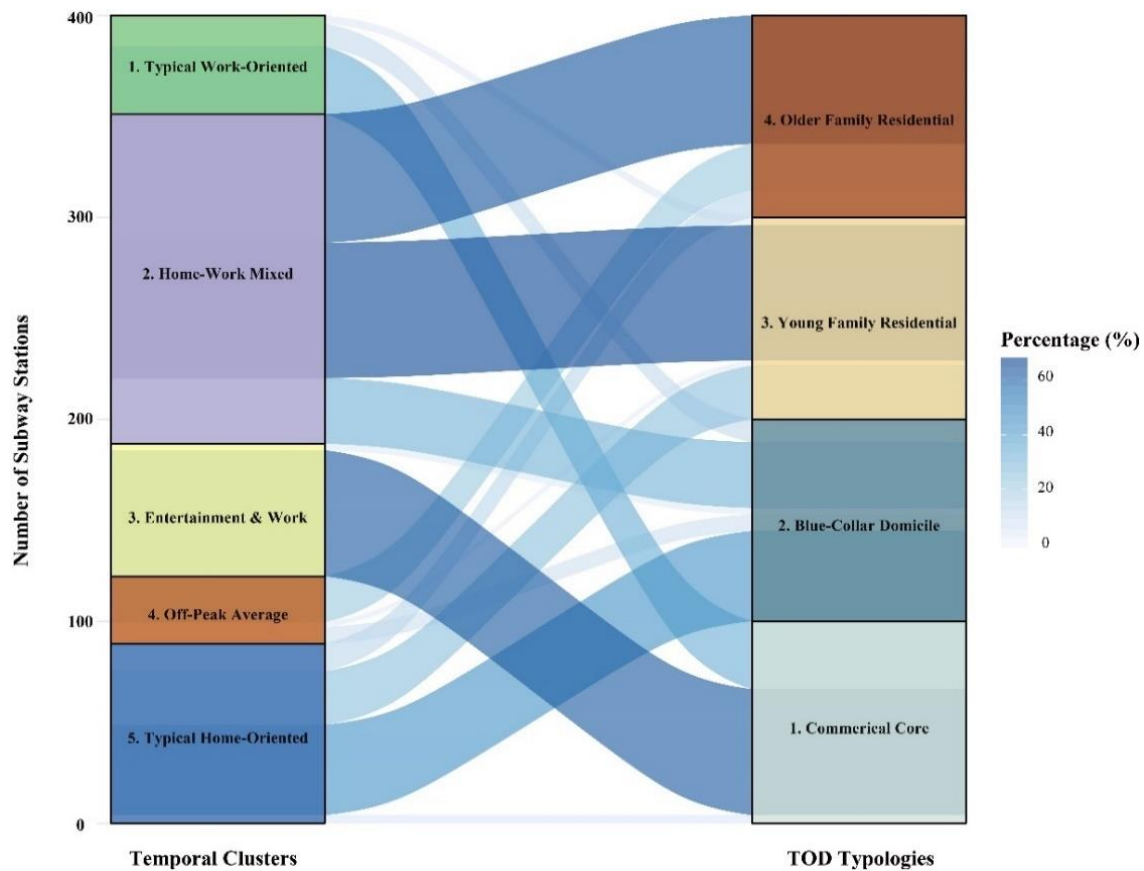


Fig 9. Alluvial Diagram: Percentage strata by Cross-tabulating TOD typologies and Temporal Clusters

## 4.0 Conclusion

Transit-Oriented Development (TOD) is a widely recognised planning method for tackling transport-related challenges. Typological approaches to TOD can be utilised either retrospectively or prospectively to assist urban planners with evidence-based information on the delivery or monitoring of TOD. However, most existing studies creating TOD typologies have overwhelmingly relied upon inputs selected alongside the ‘three Ds’ or the ‘five Ds’ principles, which might be argued as not capturing effectively multidimensional aspects of context alongside dynamics of such areas through human mobility. This study proposed and implemented an analytical framework to address this research gap by enhancing a conventional TOD typology with a wider array of contextual data, while also considering the spatiotemporal dynamics of activity at transit stations.

Our presented contextual TOD typology was implemented with candidate data inputs gathered through systematically reviewing 29 recent studies related to TOD typology. The ‘five-Ds’ principles were enriched through various measures that were broadly categorised across four domains: Land Use and Built Environment, Transit-related, Location and Accessibility, and Socioeconomic and Demographic. Four salient TOD clusters were generated for the case study city of NYC, by applying a methodology framework formed by the combination of Self-Organising Map (SOM) and hierarchical k-means clustering (H-K-means) to the multidimensional input data. These clusters were further named, described and mapped.

The spatiotemporal dynamics of activity at transit stations was considered through subway turnstile data from the MTA. Through the second application of the proposed framework to the temporal dataset, 472 subway stations were classified into five unique clusters respectively representing different types of travel activity.

The contextual TOD typology was then enhanced through linkage with the classification of aggregate space-time dynamics to illustrate the interaction between context and use. Through cross-validation there was much consistency unveiled, for example, the work-oriented stations are mainly corresponding to the stations located in major employment centres.

One of the main limitations of this study relates to the temporal resolution of the subway turnstile data. The 4-h temporal interval adopted by MTA to aggregate the passenger flows is limited in granularity and may mask valuable details should more disaggregate data be made available. In other contexts, researches such as Liu and Cheng (2018) and El Mahrssi et al. (2014) have utilised smart card data to conduct the travel pattern analysis, which brings finer resolution for both boarding and alighting information. However, such data or similar products were not publicly available from MTA. Secondly, due to data limitations of this contextual area, this study did not consider multimodal journeys which may also offer insight as the ability to interchange to other transit modes has been spotlighted by many of the reviewed studies (Chorus and Bertolini, 2011; Zemp et al., 2011; Dirgahayani and Choerunnisa, 2018). Although the present study employed variables, such as bus stop density, parking facilities, and bike facilities, to attempt to represent the intermodal connectivity, these variables were relatively ‘static’ compared to the data that could infer mode swapping. None-the-less, despite such caveats, this paper has demonstrated a new and powerful technique that implements an innovative methodology to extend a TOD typology to represent both context and dynamics; and will likely be a useful framework for application within other urban contexts..

## Reference

- Arribas-Bel, D., Schmidt, C., 2013. Self-organizing maps and the US urban spatial structure. *Environment and Planning B: Planning and Design*. 40 (2), 362-371.
- Atkinson-Palombo, C. Kuby, M. J., 2011. The geography of advance transit-oriented development in metropolitan Phoenix, Arizona, 2000-2007. *Journal of Transport Geography*, 19 (2), 189–199.
- Austin, M., Belzer, D., Benedict, A., Esling, P., Haas, P., Miknaitis, G., Wampler, E., Wood, J., Young, L. Zimbabwe, S., 2010. *Performance-based transit-oriented development typology guidebook*. Oakland, CA: Center for Transit-Oriented Development. [http://reconnectingamerica.org/assets/Uploads/2010\\_performancebasedtodtypologyguidebook.pdf](http://reconnectingamerica.org/assets/Uploads/2010_performancebasedtodtypologyguidebook.pdf)
- Baçaõ, F., Lobo, V., 2010. *Introduction to Kohonen's Self-Organizing Maps*. <https://pdfs.semanticscholar.org/779d/1280f296d887c502c5bfeac6e7fdb7a7b0ab.pdf>
- Bara, M., Ahmad, N., Modu, M., Ali, H., 2018. *Self-organizing map clustering method for the analysis of e-learning activities*. <https://ieeexplore.ieee.org/document/8363155/>.
- Bhattacharjee, S., Goetz, A., 2016. The rail transit system and land use change in the Denver metro region. *Journal of Transport Geography*. 54, 440-450.
- Box, G. Cox, D., 1964. An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*. 26 (2), 211-243.
- Center for Transit-Oriented Development, 2013. *Transit-Oriented Development Typology Strategy for Allegheny County*. <http://ctod.org/pittsburgh/201302pittsburgh-tod-book-web.pdf>
- Cervero, R., Kockelman, K., 1997. Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*. 2 (3), 199-219.
- Chavez-Baeza, C. Sheinbaum-Pardo, C., 2014. Sustainable passenger road transport scenarios to reduce fuel consumption, air pollutants and GHG (greenhouse gas) emissions in the Mexico City Metropolitan Area. *Energy*. 66 (1), 624-634.
- Chen, B., Tai, P.C. Harrison, R., 2005. Novel hybrid hierarchical-K-means clustering method (HK-means) for microarray analysis. *Computational Systems Bioinformatics Conference, 2005, Workshops and Poster Abstracts. IEEE*, Stanford University, California, USA, 105–108. <https://ieeexplore.ieee.org/abstract/document/1540560>
- Chen, C., Chen, J., Barry, J., 2009. Diurnal pattern of transit ridership: a case study of the New York City subway system. *Journal of Transport Geography*. 17(3), 176-186.
- Chorus, P., Bertolini, L., 2011. An application of the node place model to explore the spatial development dynamics of station areas in Tokyo. *The Journal of Transport and Land Use*. 4 (1), 45-58. <https://www.jtlu.org/index.php/jtlu/article/view/145>
- Das, G., Chattopadhyay, M., Gupta, S., 2016. A comparison of self-organising maps and principal components analysis. *International Journal of Market Research*. 58 (6), 815-834.
- Demartines, P., Blayo, F., 1992. Kohonen self-organizing map: is the normalization necessary? *Complex System*. 6, 105-123 <http://wpmmedia.wolfram.com/uploads/sites/13/2018/02/06-2-2.pdf>
- Dirgahayani, P., Choerunnisa, D. N., 2018. Development of Methodology to Evaluate TOD Feasibility in Built-up Environment (Case Study: Jakarta and Bandung, Indonesia). *IOP Conference Series: Earth and Environmental Science*. 158 <https://iopscience.iop.org/article/10.1088/1755-1315/158/1/012019>
- El Mahrsi, M., Côme, E., Baro, J., Oukhellou, L., 2014. *Understanding Passenger Patterns in Public Transit Through Smart Card and Socioeconomic Data*. <http://www.comeetie.fr/pdfrepos/urbcomp2014.pdf>
- Ettema, D., Friman, M., Gärling, T., Olsson, L., 2016. Travel Mode Use, Travel Mode Shift and Subjective Well-Being: Overview of Theories, Empirical Findings and Policy Implications. In: Wang, D. and He, S. *Mobility, Sociability and Well-being of Urban Living*. Berlin: Springer. 129-150. [https://link.springer.com/chapter/10.1007/978-3-662-48184-4\\_7](https://link.springer.com/chapter/10.1007/978-3-662-48184-4_7)
- Ewing, R., Cervero, R., 2010. Travel and the Built Environment. *Journal of the American Planning Association*. 76 (3), 265-294.
- Guo, J., Nakamura, F., Li, Q., Zhou, Y., 2018. Efficiency Assessment of Transit-Oriented Development by Data Envelopment Analysis: Case Study on the Den-en Toshi Line in Japan. *Journal of Advanced Transportation*. 2018, 1-10. <https://www.hindawi.com/journals/jat/2018/6701484/>
- Hickman, R. Banister, D., 2014. *Transport, Climate Change and the City*. Abingdon: Routledge. <https://www.routledge.com/Transport-Climate-Change-and-the-City-1st-Edition/Hickman-Banister/p/book/9780415660020>

- Higgins, C. D., Kanaroglou, P. S., 2016. A latent class method for classifying and evaluating the performance of station area transit-oriented development in the Toronto region. *Journal of Transport Geography*, 52, 61–72.
- Huang, R., Grigolon, A., Madureira, M., Brussel, M., 2018. Measuring transit-oriented development (TOD) network complementarity based on TOD node typology. *The Journal of Transport and Land Use*. 11 (1), 304-324.
- Hynes, M., 2017. At a Crossroads: Investigating Automobility and Its Implications for Local Urban Transport Policy Design. *Urban Science*. 1 (2), 14. <https://www.mdpi.com/2413-8851/1/2/14/htm>
- Ivan, I., Boruta, T., Horák, J., 2012. Evaluation of railway surrounding areas: the case of Ostrava city. In: Longhurst, J.W.S., Brebbia, C.A. (Eds.), *Urban Transport XVIII—Urban Transport and the Environment in the 21st Century*. WITPress, 141–152. <https://pdfs.semanticscholar.org/6921/8414caf6fd21b696bd38410e4ae169101135.pdf>
- Jain, D. K., Dubey, S. B., Choubey, R. K., Sinhal, A., Arjaria, S. K., Jain, A., Wang, H., 2018. An approach for hyperspectral image classification by optimizing SVM using self organizing map. *Journal of Computational Science*. 25, 252-259.
- Jun, M., Choi, K., Jeong, J., Kwon, K., Kim, H., 2015. Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul. *Journal of Transport Geography*. 48, 30-40.
- Kamruzzaman, Md., Baker, D., Washington, S., Turrell, G., 2014. Advance transit-oriented development typology: Case study in Brisbane, Australia. *Journal of Transport Geography*, 34, 54–70.
- Kassambara, A., 2017. *Hierarchical K-Means Clustering: Optimize Clusters*. Available: <https://www.datanovia.com/en/lessons/hierarchical-k-means-clustering-optimize-clusters/>.
- Kim, M., Kim, S., Heo, J., Sohn, H., 2017. Ridership patterns at subway stations of Seoul capital area and characteristics of station influence area. *KSCE Journal of Civil Engineering*. 21 (3), 964-975.
- Kim, M., Kim, S., Sohn, H., 2018. Relationship between Spatio-Temporal Travel Patterns Derived from Smart-Card Data and Local Environmental Characteristics of Seoul, Korea. *Sustainability*. 10 (3), 1-18.
- Lee, S., Yi, C., Hong, S., 2013. Urban structural hierarchy and the relationship between the ridership of the Seoul Metropolitan Subway and the land-use pattern of the station areas. *Cities*. 35, 69-77.
- Lierop, D. V., Maat, K., El-Geneidy, A., 2017. Talking TOD: learning about transit-oriented development in the United States, Canada, and the Netherlands. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*. 10 (1), 49-62. <https://www.tandfonline.com/doi/abs/10.1080/17549175.2016.1192558?journalCode=rjou20>
- Liu, Y. Cheng, T., 2018. Understanding public transit patterns with open geodemographics to facilitate public transport planning. *Transportmetrica A: Transport Science*. 16 (1) <https://doi.org/10.1080/23249935.2018.1493549>
- Lyu, G., Bertolini, L., Pfeffer, K., 2016. Developing a TOD typology for Beijing metro station areas. *Journal of Transport Geography*. 55, 40-50.
- Metropolitan Transportation Authority, 2016. *Introduction to Subway Ridership*. <http://web.mta.info/nyct/facts/ridership/>.
- Metropolitan Transportation Authority, 2016. *Subways*. <http://web.mta.info/nyct/facts/ffsubway.htm>.
- Miljković, D., 2017. *Brief Review of Self-Organizing Maps*. Available: <https://ieeexplore.ieee.org/document/7973581/>
- Monajem, S., Nosrati, F. E., 2015. The evaluation of the spatial integration of station areas via the node place model; an application to subway station areas in Tehran. *Transportation Research Part D*. 40, 14-27.
- Nasri, A., Zhang, L., 2014. The analysis of transit-oriented development (TOD) in Washington, D.C. and Baltimore metropolitan areas. *Transport Policy*. 32, 172-179.
- Natita, W., Wiboonsak, W., Dusadee, S., 2016. Appropriate Learning Rate and Neighborhood Function of Self-organizing Map (SOM) for Specific Humidity Pattern Classification over Southern Thailand. *International Journal of Modeling and Optimization*. 6 (1), 61-65. <http://www.ijmo.org/vol6/504-M08.pdf>
- Oliver, L., Schuurman N., Hall, A. 2007. Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. *International Journal of Health Geographics*. 6 (41). <https://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-6-41>
- Papa E., Carpentieri G., Angiello G., 2018. A TOD Classification of Metro Stations: An Application in Naples. In: Papa R., Fistola R., Gargiulo C. (Eds.) *Smart Planning: Sustainability and Mobility in the Age of Change. Green Energy and Technology*. Springer, Cham, 285-300 [https://link.springer.com/chapter/10.1007/978-3-319-77682-8\\_17](https://link.springer.com/chapter/10.1007/978-3-319-77682-8_17)
- Pollack, S., Gartsman, A., Boston, M., Benedict, A., Wood, J., 2014. *Rating the performance of station areas for effective and equitable transit oriented development*. <http://docs.trb.org/prp/14-0392.pdf>

- Rodrigue, J., 2017. Urban Transport Challenges. In: Rodrigue, J., Comtois, C. and Slack, B. *The Geography of Transport Systems*. 4th ed. New York: Routledge.
- Schäfer, P., Hostert, P., Leser, U., 2018. *Classifying land cover from satellite images using time series analytics*. <http://ceur-ws.org/Vol-2083/paper-02.pdf>
- Schonlau, M., 2002. The clustergram: a graph for visualizing hierarchical and non-hierarchical cluster analyses. *The Stata Journal*. 2 (4), 391-402.
- She, Z., King, D., Jacobson, S., 2017. Analyzing the impact of public transit usage on obesity. *Preventive Medicine*. 99, 264-268.
- Singh, Y.J., Lukman, A., Flacke, J., Zuidgeest M., Van Maarseveen M.F.A.M., 2017. Measuring TOD around transit nodes—towards TOD policy. *Transport Policy*. 56, 96–111.
- Sohn, K., 2013. Feature Mapping the Seoul Metro Station Areas Based on Self-Organizing Map. *Journal of Urban Technology*. 20 (4), 23-42.
- Song, J., Deguchi, A., 2013. Evaluation and typology of railway station areas in a 30 km circumference surrounding Central Tokyo from view of transit-oriented development. *Journal of Architecture and Planning (Transactions of AIJ)*. 78 (684), 413–420. [https://www.jstage.jst.go.jp/article/aija/78/684/78\\_413/article](https://www.jstage.jst.go.jp/article/aija/78/684/78_413/article)
- Spielman, S., Folch, D., 2015. Social Area Analysis and Self-Organizing Maps. In: Brunson, C. and Singleton, A. *Geocomputation: A Practical Primer*. London: SAGE Publications Ltd. 152-168.
- Spielman, S., Thill, J., 2008. Social area analysis, data mining, and GIS. *Computers, Environment and Urban Systems*. 32(2), 110-122.
- Staricco, L., Brovarone, E., 2018. Promoting TOD through regional planning. A comparative analysis of two European approaches. *Journal of Transport Geography*. 66, 45-52.
- Sung, H., Choi, C., 2017. The link between metropolitan planning and transit-oriented development: An examination of the Rosario Plan in 1980 for Seoul, South Korea. *Land Use Policy*. 63, 514-522.
- Taki, H., Maatouk, M., Qurnfulah, E., Aljoufie, M., 2017. Planning TOD with land use and transport integration: a review. *Journal of Geoscience, Engineering, Environment, and Technology*. 2 (1), 84-94. <https://www.neliti.com/publications/155016/planning-tod-with-land-use-and-transport-integration-a-review>
- Thomas, R., Pojani, D., Lenferink, S., Bertolini, L., Stead, D., Krabben, E., 2018. Is transit-oriented development (TOD) an internationally transferable policy concept? *Regional Studies*. 52 (9), 1201-1213.
- Tian, J. Azarian, M., Pecht, M., 2014. Anomaly detection using self-organizing maps-based k-nearest neighbour algorithm. *Proceedings of the European Conference of the Prognostics and Health Management Society*. 5. <https://www.phmsociety.org/node/1300>
- United States Census Bureau., 2011. *Census Blocks*. <https://www.census.gov/newsroom/blogs/random-samplings/2011/07/what-are-census-blocks.html>
- United States Census Bureau., 2017. *QuickFacts*. <https://www.census.gov/quickfacts/newyorkcitynewyork>
- Vale, D. S., 2015. Transit-oriented development, integration of land use and transport, and pedestrian accessibility: combining node-place model with pedestrian shed ratio to evaluate and classify station areas in Lisbon. *Journal of Transport Geography*. 45, 70–80.
- Wang, J. Kong, X., Rahim, A., Tolba, A., Al-Makhadmeh, Z., 2017. IS2Fun: Identification of Subway Station Functions Using Massive Urban Data. *IEEE Access*. 5, 27103-27113. <https://ieeexplore.ieee.org/document/8082741>
- Wang, S., Sun, L., Rong, J., Hao, S. Luo, W., 2016. Transit trip distribution model considering land use differences between catchment areas. *Journal of Advanced Transportation*. 50(8), 1820-1830.
- Winters, M., Buehler, R., Götschi, T., 2017. Policies to Promote Active Travel: Evidence from Reviews of the Literature. *Current Environmental Health Reports*. 4 (3), 278-285.
- Xia, F., Wang, J., Kong, X., Wang, Z., Li, J., Liu, C., 2018. Exploring Human Mobility Patterns in Urban Scenarios: A Trajectory Data Perspective. *IEEE Communications Magazine*. 56 (3), 142-149. <https://ieeexplore.ieee.org/abstract/document/8316783>
- Xu, W., Guthrie, A., Fan, Y., Li, Y., 2017. Transit-oriented development in China: Literature review and evaluation of TOD potential across 50 Chinese cities. *Journal of Transport and Land Use*. 10 (1), 743-762.



Yin, H., 2008. The Self-Organizing Maps: Background, Theories, Extensions and Applications. In: Fulcher, J. and Jain, L.C. *Computational Intelligence: A Compendium. Studies in Computational Intelligence*. Berlin, Heidelberg: Springer. 715-762. [https://link.springer.com/chapter/10.1007/978-3-540-78293-3\\_17](https://link.springer.com/chapter/10.1007/978-3-540-78293-3_17)

Zemp, S., Stauffacher, M., Lang, D. J., Scholz, R.W., 2011. Classifying railway stations for strategic transport and land use planning: context matters! *Journal of Transport Geography*. 19 (4), 670–679.

Zhou, Y., Fang, Z., Zhan, Q., Huang, Y., Fu, X., 2017. Inferring Social Functions Available in the Metro Station Area from Passengers' Staying Activities in Smart Card Data. *ISPRS International Journal of Geo-Information*. 6 (12), 394.