The Use of Consumer Data to Explore Geographic and Social Variations in Online Gambling

Shunya Kimura[*], Justin van Dijk[†] and Paul A. Longley[‡]

Department of Geography, University College London

GISRUK 2023

Summary

The rapid diffusion of smartphones and Internet use in the past few decades has led to the rise in online gambling, proliferating the opportunities for participation. Together with the recent cost-of-living crisis, it is essential that the extent of the associated harms is clearly understood. Despite the growing consensus on the need to tackle gambling harm as a public health issue, there is limited empirical knowledge on the prevalence of harm and its underlying factors. Here, we aim to reveal any geographic and social patterns of online gambling behaviour at a national scale through exploring the use of consumer data.

KEYWORDS: gambling disorder, public health, consumer data, spatial autocorrelation, geodemographics

1. Introduction

There has been a surge in population studies of gambling behaviours in Great Britain (GB) over the last decade and a half; consistent with the growing consensus on the importance of tackling the associated harms as a potential public health issue (Pickering & Blaszczynski, 2021). However, there is a clear gap in knowledge, particularly around online gambling. This is likely due to the limitations in the conventional methods of data collection, such as the high cost of surveys limiting the sample size resulting in a small sampling fraction. Moreover, low incidence of gambling disorder imposes restrictions on the accuracy of further estimates produced for subdomains, making the identification of geographic and social patterns difficult (Sturgis & Kuha, 2021).

Recent studies have deployed consumer data supplied by an industry to overcome some of these issues (Broda et al., 2008; Forrest and McHale, 2022). Consumer data allows for larger pool of samples at spatially extensive yet granular scales; and sheds light on the contexts and revealed behaviours which would otherwise be hidden (Rains & Longley, 2021). For instance, self-exclusion is where customers voluntarily prevent further gambling by leaving their account indefinitely, which could be used as an indicator of someone experiencing a gambling disorder.

Utilising consumer data, this paper aims to enrich the understanding of the nature of online gambling behaviour by investigating if there is any geographic or social pattern to (a) online gambling and (b) self-exclusion from it.

2. Data

This study utilises consumer data accessed from the world's largest providers of online sports betting and gaming, which runs several major brands in GB. The data incorporates a list of unique account-

[*] shunya.kimura.18@ucl.ac.uk
[†] j.t.vandijk@ucl.ac.uk
[‡] p.longley@ucl.ac.uk

holder identifiers and their home address at a postcode level, stored and accessed through a secure environment. Here, only a 'genuine' gambling population were selected who meet the following conditions: (a) those who have deposited multiple times within a year and (b) those who have played with actual money over multiple days within a year. This was to control for the customers who have registered an account only to play with promotional stakes; to play for a day and leave; or to not play at all. Within the study period between 1st January 2022 and 31st December 2022, circa 250,000 'genuine' customers were sampled. Within them, about 6,000 (2.43%) chose to self-exclude from at least one of their accounts at some point of the year.

The consumer data is supplemented by a geodemographic classification. Geodemographics is an 'analysis of people by where they live' (Brunsdon et al., 2011: 18), which is a statistical approach for dismantling complex and multidimensional data about the society and enhancing the understanding of the population at a small area level (Singleton & Longley, 2015). Here, we use the 2011 Area Classification for Super Output Areas (SOAs). It is a residential classification that builds upon 2011 Census data for SOAs, comprised of Lower Layer Super Output Areas (LSOAs) in England & Wales and Data Zones (DZs) in Scotland (Bates 2018). A range of 60 socio-residential variables were used in the creation of this two-tier hierarchal classification. In this study, the second-tier subgroup is utilised, which consists of 24 types. SOA classification was deemed to be a useful tool to glean some general insights about the endogenous characteristics that may determine the tendency of being an online gambler or self-excluding from it.

3. Methodology

To derive a general understanding of the geographic distribution of online gamblers, we have created three choropleth maps that rank each Local Authority District from the highest (1) to lowest (10), based on the following metrics:

1) Proportion of 'genuine' customers per adult (18+) population
2) Rate of self-exclusion amongst the 'genuine' customers
3) Rate of self-exclusion amongst the adult (18+) population

Moreover, to quantify whether there is any geography to online gambling, we have calculated the Moran I statistic. It is a commonly used spatial autocorrelation statistics for the case of area objects, which evaluates the 'randomness' of the spatial distribution of certain values (Longley et al., 2010; Moran, 1948). To produce a correlation coefficient for the relationship between a variable and its surrounding values, it requires the user to specify how the 'neighbours' are defined. For our analysis, we have used the Queen contiguity, as this approach is known to be "effective when polygons are similar in size and distribution, and when spatial relationships are a function of polygon proximity" (Esri, 2021); and has been tried and tested on the Local Authorities of the UK (Morton et al., 2018). The structural form of Moran I is summarised in the **Equation 1** below.

$$I = \left(n \sum_i \left[\sum_j w_{ij} (z_i - \bar{z})(z_j - \bar{z}) \right] \right) / \left(\sum_i \left(\sum_j w_{ij} \sum_j \right) \right) (z_i - \bar{z}) \tag{1}$$

Where:

n = the number of geographical units in the study area
z_i = the value of the spatial attribute at unit i
z_j = the value of the spatial attribute at unit j
$\bar{z}$ = the average value of the z's
w_{ij} = the spatial weights matrix

The metric is expressed between -1 (dispersed) and 1 (clustered), whilst 0 indicates Complete Spatial Randomness (CSR).

To explore whether there are any social patterns to online gambling, two groups of customers: (a) 'genuine' and (b) self-excluded, were profiled using the SOA Classification. Accordingly, Index Scores (ISs) were produced for each SOA type by dividing the proportion of a customer group per class by the proportion of adult population in GB per class, as summarised in the **Equation 2** below.

$$IS_i = \frac{\left(\frac{n_i}{n_{gb}}\right)}{\left(\frac{Pop_i}{Pop_{gb}}\right)} \tag{2}$$

Where:

n_i = the number of customers in class i
n_{gb} = the total number of customers in GB
Pop_i = the adult population in class i
Pop_{gb} = the adult population in GB

$IS_i < 1$ indicates a fewer-than-expected presence of the customer group of interest in class i, whilst $IS_i > 1$ indicates a more-than-expected presence of the customer group of interest in class i. To compare the ISs of the 'genuine' and self-excluded customers across each SOA Classification type, the results were then visualised as a Cleveland dot plot (**Figure 2**).

4. Results

As illustrated in Map 1 of **Figure 1**, a North-South divide is apparent, suggesting the presence of a geographical pattern in the uptake of online gambling across GB. This is reflected in the spatial autocorrelation measure, with a moderate-high Moran I statistic of 0.45 (p-value < 0.05). Online gambling seems particularly popular around the Northern Local Authority Districts like County Durham, Sunderland, Scottish Borders and East & South Ayrshire. On the contrary, southern Local Authority Districts around Greater London seem to have a lower uptake overall. A similar geographical distribution is depicted on Map 3 (**Figure 1**), with slightly more dispersion, as reflected in the lower Moran I statistic of 0.29 (p-value < 0.05).

Map 2 (**Figure 1**) depicts the rate of self-exclusion amongst the gambling prone population. Here, geographic patten was not apparent, supported by the Moran I statistic close to 0, indicating a CSR.

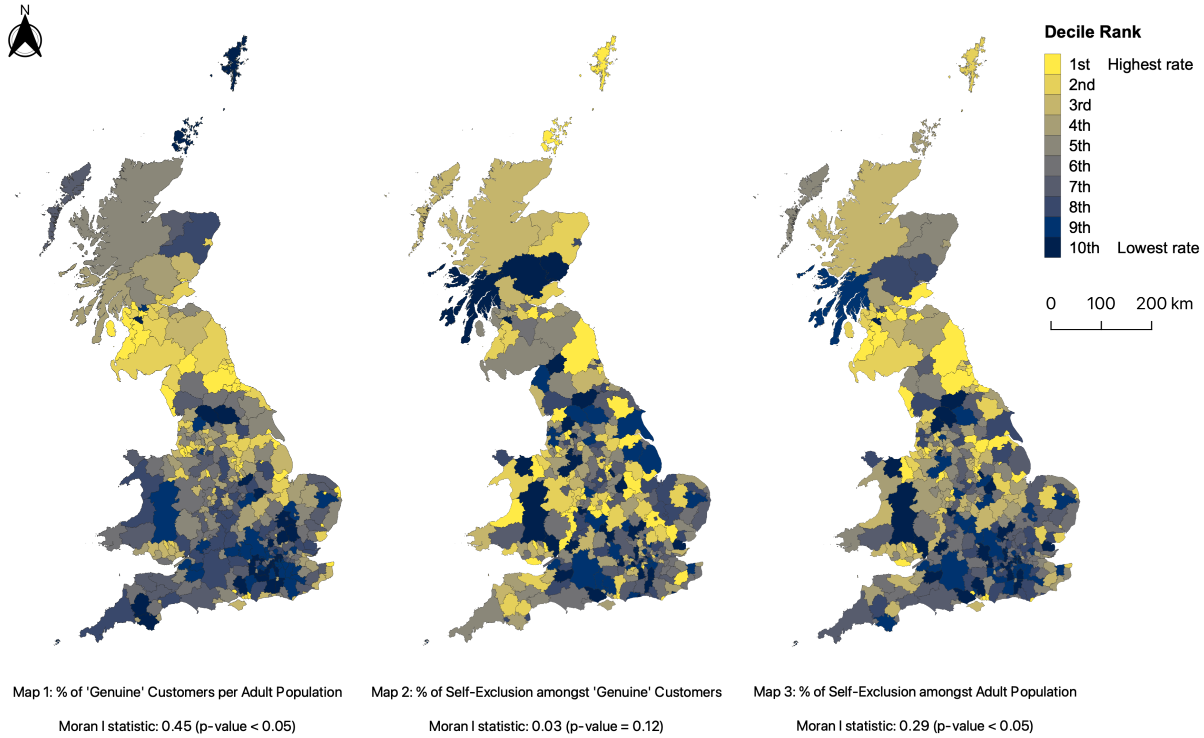

Figure 1 Choropleth maps of the three metrics expressed in decile rankings across Local Authority Districts in Great Britain

Figure 2 presents the results of the profile analysis of the 'genuine' and self-excluded customers according to the SOA Classification Types. Overall, the geodemographic distribution of the two customer groups were similar. The top three SOA types in which majority of the customers ('genuine' and self-excluded) were classified were *4A: Challenged white communities* (IS=1.68 and IS=2.11); *4B: Constrained renters* (IS=1.77 and IS=2.01); and *4D: Hard-pressed flat dwellers* (IS=2.08 and IS=1.96). These areas typically have relatively high level of unemployment rates, low qualification attainments and high rates of divorce or separation, compared to the national average.

In general, economically challenged groups were seen to have higher tendency of self-exclusion in relative to the online gambling population. However, there were no obvious ethnic, age nor urban/rural dimension that differentiates the two groups of customers.

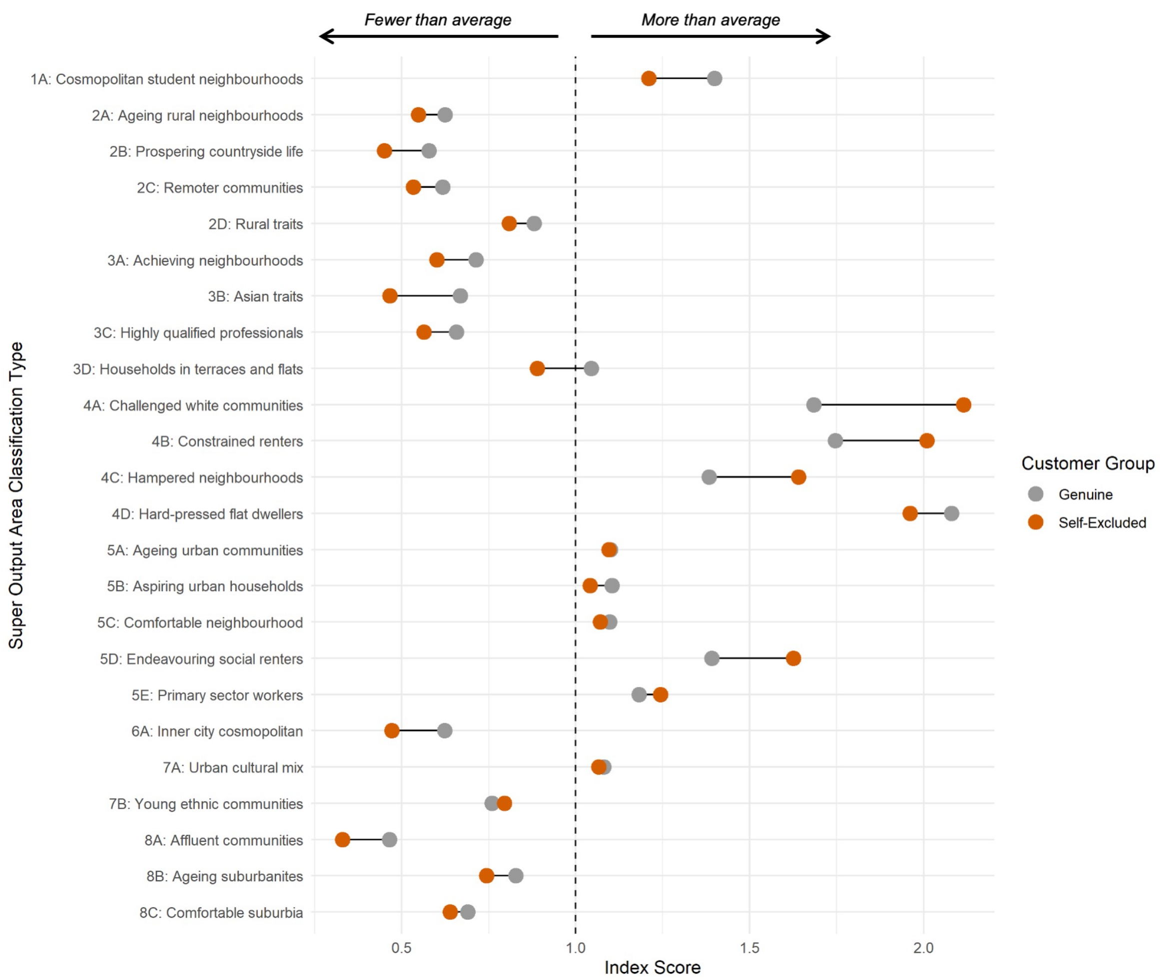

Figure 2 Comparison of the index scores calculated for each Super Output Area Classification type between 'genuine' and self-excluded customer groups

5. Conclusion

We have attempted to explore a GB-wide spatial patterns to online gambling and to self-exclusion from it upon linking consumer data with a geodemographic classification. We have found that there is indeed a geography to online gambling behaviour, with higher market penetration in the Northern regions. However, the rate of self-exclusion amongst the gambling population does not seem to vary geographically. Given that their endogenous characteristics were also similar, this may suggest that anyone who gambles online are at-risk of developing a problematic gambling behaviour. Nonetheless, social selectivity in gambling disorder is apparent, especially amongst the economically challenged groups of people. As there are no obvious physical indicators of the prevalence of online gambling disorder, these findings may be of value to policy makers and charity organisations who aim to devise a more targeted, evidence-based, public health strategy in prevention and treatment of the associated harms. For a successful public health intervention, perhaps more than the socio-economic dimensions ought to be used to identify people with gambling disorders, such as their patterns of play. This will be our next focus of our studies.

Acknowledgements

This work is funded by the ESRC in collaboration with GambleAware.

References

Bates A G (2018). 2011 Area Classification for Super Output Areas. Available at: https://www.ons.gov.uk/releases/2011areaclassificationforsuperoutputareas (accessed 20 January).

Borda A, LaPlante D A, Nelson S E, LaBrie R A, Bosworth L B and Shaffer H J (2008). Virtual harm reduction efforts for Internet gambling: Effects of deposit limits on actual Internet sports gambling behavior. *Harm Reduction Journal*, 5(1), 27.

Brunsdon C, Longley P A, Singleton A and Ashby D (2011). Predicting participation in higher education: a comparative evaluation of the performance of geodemographic classifications. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(1), 17–30.

Esri (2021). Modelling spatial relationships. Available at: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/modeling-spatial-relationships.htm#GUID-729B3B01-6911-41E9-AA99-8A4CF74EEE27 (accessed 20 January).

Forrest D and McHale I (2022). Patterns of Play Extended Executive Summary Report, GambleAware, available at: https://www.begambleaware.org/sites/default/files/2022-06/Patterns%20of%20Play_Summary%20Report_final%5B2281%5D_1.pdf (accessed 20 January).

Morton C, Anable J, Yeboah G and Cottrill C (2018). The spatial pattern of demand in the early market for electric vehicles: Evidence from the United Kingdom. *Journal of Transport Geography*, 72, 119-130.

Pickering D and Blaszczynski A (2021). Paid online convenience samples in gambling studies: Questionable data quality. *International Gambling Studies*, 1–21.

Rains T and Longley P A (2021). The provenance of loyalty card data for urban and retail analytics. *Journal of Retailing and Consumer Services*, 63, 102650.

Singleton A D and Longley P A (2015). The internal structure of Greater London: a comparison of national and regional geodemographic models. *Geo: Geography and Environment,* 2 (1). 69-87.

Sturgis, P., and Kuha, J., (2021) Methodological factors affecting estimates of the prevalence of gambling harm in the United Kingdom: A multi-survey study, GambleAware, available at: https://www.begambleaware.org/sites/default/files/2021-05/Methodology_Report_(FINAL_14.05.21).pdf (accessed 20 January).

Biographies

Shunya Kimura is a second year PhD student in Human Geography at UCL, funded by the UBEL DTP in collaboration with GambleAware. His research aims to create a geodemographic typology of online gambling behaviour in Great Britain utilising large-scale consumer data from a multi-national sports betting and gaming group.

Paul A. Longley is Professor of Geographic Information Science at UCL and Director of the UK Consumer Data Research Centre at UCL. His research focuses on the application of geographic information science, with a strong emphasis on the development and deployment of geo-temporal data infrastructures developed from Big or Open Data.

Justin van Dijk is a Lecturer in Social and Geographic Data Science in the Department of Geography at University College London. His primary research interests are grouped around the analysis and visualisation of large-scale spatial data, urban mobility, socio-spatial inequalities, and geospatial data in general.