Dominant Trip Purposes within a Dockless Bicycle Sharing System in London

James Todd*[1], Shunya Kimura†[1], Oliver O'Brien‡[1] and James Cheshire§[1]

[1]Department of Geography, UCL

GISRUK 2023

Summary

Implementing a zero-inflated multilevel negative binomial regression model, this research identifies the relative increase in dockless bicycle sharing journey destinations within close proximity to a range of built environment factors that enable inferences of trip purposes. Results show that among users of the JUMP system in London there is nearly a fourfold increase in bicycle drop-offs in locations within close proximity to train stations, showcasing significant multi-modal first and last mile use, in addition to a significant proportion of users likely to be students using the mode to get to university buildings.

KEYWORDS: bicycle sharing, micromobility, mobility dynamics, sustainable mobility, regression

1 Introduction

Bicycle sharing systems (BSS) are a form of urban micromobility that have observed mass adoption in cities around the world (with over 1,900 currently operational systems Meddin et al. (2022)). Most recent iterations of BSS, dubbed the fifth-generation (Guidon et al., 2019), are characterised by their dockless nature, enabling flexibility in picking-up and dropping-off bicycles within operational areas, whilst also reducing the amount of physical excision required with the implementation electric pedal-assistance. There is limited research surrounding these e-BSS, with researchers such as Becker et al. (2017) suggesting that the differences between previous iterations of BSS being so stark that past findings may not be transferable.

As such, this research aims to extend on current understandings of dockless e-BSS systems by analysing trip purposes within the Uber JUMP e-BSS in London. Unlike prior analyses that have sought to understand the determinants of BSS activity through understanding the effects of built

*james.todd@ucl.ac.uk

†shunya.kimura.18@ucl.ac.uk

‡o.obrien@ucl.ac.uk

§james.cheshire@ucl.ac.uk

environment factors that are within close proximity to journey locations, this research is unique in three ways. Firstly, instead of using journey origin locations (Guidon et al., 2019; Shen et al., 2018), journey destination locations have been elected due to their advantages in being able to infer trip purposes as opposed to the terminus of previous journeys or operator redistributions. Secondly, granular 50-metre radius hexagonal grid cells have been employed over preceding coarse geographical aggregations (Mooney et al., 2019; Guidon et al., 2019; Shen et al., 2018) in order to better identify likely trip purposes given the flexibility of the dockless system and reduce issues of ecological fallacy. Finally, the number of bicycles dropped-off in each hexagonal grid cell were counted in four time blocks - *AM Peak Hours* (weekdays 7am to 11am), *PM Peak Hours* (weekdays 4pm to 8pm), *Off-Peak Hours* (weekdays 8pm to 7am and 11am to 4pm) and *Weekend All* (weekends 12am to 12am). In doing so, it was possible to employ a zero-inflated multilevel negative binomial regression model with interaction effects, typically operationalised in fields such as spatial epidemiology and environmental criminology, to measure the relative increase in bicycle drop-off occurrences for each built environment factor and time block.

2 Data

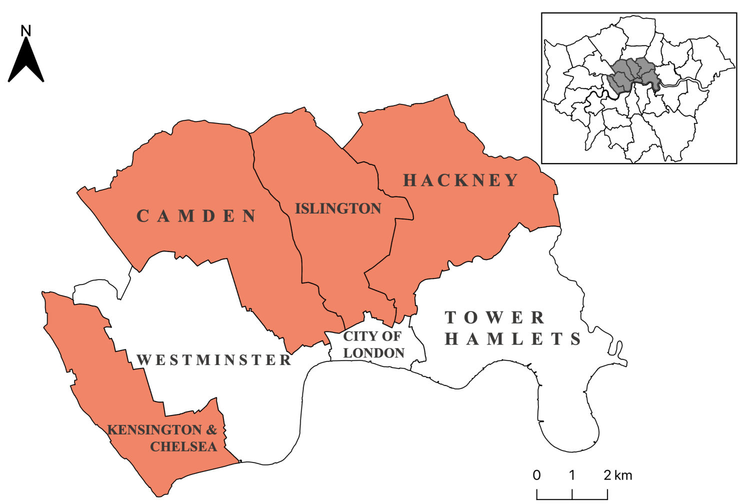

Figure 1: JUMP e-BSS study region [orange areas depict permitted operating boroughs].

Data on the JUMP e-BSS in London were obtained from an openly published General Bikeshare Feed Specification (GBFS) (see NABSA (2022)) feed between 26^{th} December 2019 and 4^{th} March 2020. The raw data undergo a rigorous cleaning process to identify each journey's linked origin-destination (OD) locations and remove all non-user movements, whether that be a result of data anomalies or operator redistribution and maintenance efforts (such as battery swapping). Initially, the dynamically rotating nature of bicycle IDs after the end of each journey are exploited to remove logs of bicycle locations in situ. It is then possible to reconstruct OD records between each pair of bicycle IDs and estimate trip attributes such as their duration, distance, speed and battery use to further process and identify non-user records within these data. Such processes result in a final collection of 58,682 journeys that occurred between 1^{st} January and 3^{rd} March 2020 which form the foundations of this analysis. Figure 1 provides an overview of the study area that have been defined that include non-permitted boroughs due to its disjointed and incomplete geographical nature that reduce but still constitute a significant proportion of total journeys throughout the study period, as identified in Table 1.

Table 1: Total and average daily journeys that end in each borough across the study area.

Borough	**Total Journeys**	**Average No. of Daily Journeys**
Camden	19,974	312.1
City of London	945	14.8
Hackney	8,106	126.7
Islington	23,547	367.9
Kensington and Chelsea	2,309	36.1
Tower Hamlets	2,099	32.8
Westminster	1,702	26.6

Table 2: OSM feature keys and values used to identify built environment factors.

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

In addition, independent built environment factors and population distribution characteristics of the study area were necessary to be derived from a number of data sources. These were primarily obtained from OpenStreetMap (OSM) due to its accessibility and completeness in London. The names of each built environment variable and their corresponding OSM feature names and keys used to extract their locations are detailed in Table 2. Cycle lane locations have been derived from a 2018 Transport for London dataset. Population distribution characteristics that are unattainable from OSM have been gathered from the datasets detailed in Table 3 to provide further contextual information. Combining these data sources together a multilevel data structure consisting of 70,796 observations are created, full details of which are detailed in Section 7.4 of Todd (2022). The distribution of these variables are detailed in Table 4.

Table 3: Additional data sources used in regression analysis.

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

Table 4: Descriptive statistics of variables included in regression model.

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

3 Methods

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, most importantly in the outcome variable of journey counts. In this context, the excess number of zero counts of journey drop-offs in hexagons are caused as a result of the study area, which include non-permitted boroughs and penalties for parking in no-parking zones marked within the application (e.g. parks). As such, linear regression methods are unsuitable due to issues surrounding overdispersion, meaning that variable distributions do not meet the model assumptions and would cause over estimations of desired regression parameters, leading to biased standard errors and p-values (Hilbe, 2011). Therefore, to overcome this data artefact, 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. Accordingly, varying-intercept and varying-slope model with interaction effects between the time blocks and geographic attributes were fitted, the statistical formulation of which is defined as follows:

$$
\begin{aligned}
y_{i,j} &\backsim Pois(\theta_{ij}, \kappa) \\
log(\theta_{ij}) &= \alpha_{j[i]} + \sum_{t=1}^{3} \beta_{j[i]t} T_{ti} + \sum_{t=0}^{3} \sum_{h=0}^{8} \beta_{j[i]th} T_{ti} H_{hi}
\end{aligned}
\tag{1}
$$

Here, θ_{ij}, κ is a Poisson parameter with a log-link function equal to the expected count of bicycle drop-offs in the i^{th} hexagon (i = 1,2,3, ...,70796) in the j^{th} MSOA (j = 1,2,3, ..., 190), where κ 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. The 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, ..., 8), respectively. This means that the regression coefficient $\beta_{j[i]}$ reflects the overall relationship between the h^{th} hexagon-level variable and bicycle count in the reference class (i.e., *AM Peak Hours*) in MSOA j where hexagon i is located and also produce an overall effect when $t \neq 0$. $\alpha_{j[i]}$ represents the average count of bicycles dropped-off at the j^{th} MSOA within the study area. In order to derive a relative Risk Ratio (RR) all $\beta_{j[i]}$ and $\alpha_{j[i]}$ coefficients are hence exponentiated, where a value of 1 indicates null effect and values larger or smaller indicate an increase or decrease from the population average, respectively.

4 Results

The model results in Table 5 suggest that the vast majority of built environment factors are significantly associated with journey destinations but these effects differ both in direction and magnitude.

Shops appear to have the smallest positive association to journey destinations, where, for every additional shop within a hexagon there is expected to be a 8% increase in the rate of bicycle drop-offs (1.08, 95% CI 1.05-1.10). Sustenance amenities observe marginally greater journeys at 16%

Table 5: Zero-inflated multilevel negative binomial regression results.

Fixed Effects	**RR**	**(95% CI)**	**p-value**
Intercept	0.04	(0.02 - 0.09)	< 0.001 ***
Time block: *PM Peak Hours*	2.09	(1.79 - 2.44)	< 0.001 ***
Time block: *Off-Peak Hours*	2.43	(2.11 - 2.79)	< 0.001 ***
Time block: *Weekend All*	2.04	(1.71 - 2.44)	< 0.001 ***
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: reference category = *AM Peak Hours*; '***' < 0.001, '**' < 0.01, '*' < 0.05, '·' < 0.1

(1.16, 95% CI 1.12-1.20), followed by tourist attractions by 37% (1.37, 95% CI 1.21 - 1.56). Each of these do not show any significant differences between time blocks. On average, cycle lanes observe an 80% increase (1.80, 95% CI 1.66-1.95) but *AM Peak Hours* are consistently observed to have 12% greater rates of drop-off to all other time blocks that is likely an artefact of cycle lane and bicycle converge towards the City. The presence of university buildings are found nearly double the RR (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 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*. Parks within the study area appear to to exhibit no significant relationship. Finally, considering JUMP activity in close proximity to transit related features indicate some of the largest increases in RR, showcasing the importance of e-BSS in facilitating multimodal journeys. The presence of a bus stop is found to increase the 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). Across both public transit 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 during *Off-Peak Hours* that are likely due to a lack of commuters.

5 Conclusions

The research presented here provide a holistic indication of the most dominant built environment factors that are associated with e-BSS use in London. The results highlight the importance of such systems in facilitating multi-modal journeys and the considerable use by university students whilst also identifying the lack of journeys to green spaces due to the parking restrictions. The interaction terms between built environment factors and time blocks enable a more granular understand of how such relationships change depending on the time of day, with some unexpected findings, including cycle lanes during *AM Peak Hours* and the converse relationship to train stations. These are likely to be an artefact of the segregated nature of permitted operating areas as well as the geographical arrangement of the study area.

6 Acknowledgements

This research was partly funded by the Economic and Social Research Council (ESRC).

The dataset used for this research was released by the Consumer Data Research Centre (CDRC), an ESRC data investment, under the project ES/L011840/1; ES/L011891/1. The dataset is safeguarded by the CDRC.

References

Becker, H., Ciari, F., & Axhausen, K. W. (2017). Comparing car-sharing schemes in Switzerland: User groups and usage patterns. *Transportation Research Part A: Policy and Practice*, *97*, 17–29, https://doi.org/10.1016/j.tra.2017.01.004.

Guidon, S., Becker, H., Dediu, H., & Axhausen, K. W. (2019). Electric Bicycle-Sharing: A New Competitor in the Urban Transportation Market? An Empirical Analysis of Transaction Data. *Transportation Research Record*, *2673*(4), 15–26, https://doi.org/10.1177/0361198119836762.

Hilbe, J. M. (2011). *Negative Binomial Regression*. Cambridge: Cambridge University Press, 2 edition.

Meddin, R., DeMaio, P., O'Brien, O., Rabello, R., Yu, C., Seamon, J., & Benicchio, T. (2022). The Meddin Bike-sharing World Map https://bikesharingworldmap.com/.

Mooney, S., Hosford, K., Howe, B., Yan, A., Winters, M., Bassok, A., & Hirsch, J. (2019). Freedom from the station: Spatial equity in access to dockless bike share. *Journal of Transport Geography*, *74*, 91–96, https://doi.org/10.1016/j.jtrangeo.2018.11.009.

NABSA (2022). General Bikeshare Feed Specification https://github.com/NABSA/gbfs.

Shen, Y., Zhang, X., & Zhao, J. (2018). Understanding the usage of dockless bike sharing in Singapore. *International Journal of Sustainable Transportation*, *12*(9), 686–700, https://doi.org/10.1080/15568318.2018.1429696.

Todd, J. (2022). Bicycle sharing systems: Fast and slow urban mobility dynamics https://discovery.ucl.ac.uk/id/eprint/10161120/.

Biography

Dr. James Todd is a Research Data Scientist at the CDRC at UCL. He recently defended his PhD thesis, entitled '*Bicycle Sharing Systems: Fast and Slow Urban Mobility Dyanmics*', which analysed the values of passively sourced micromobility data in deriving mobility dynamic insights across a wide variety of spatial and temporal scales.

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.

Oliver O'Brien is a urban data researcher, data scientist and software developer at the CDRC at UCL, where he specialises in visualising city datasets. He has created a number of software products including Bike Share Map, CityDashboard, DataShine, CensusProfiler, Named, and CDRC Maps. He edits Mapping London.

James Cheshire is the Director of the Social Data Institute at UCL and is Professor of Geographic Information and Cartography in the Department of Geography. His research focuses on visualisation and analysis of new form of data for the study of social science.