Understanding Willingness to Use Dockless Bike Sharing Systems Through Tree and Forest Analytics

Ioannis Politis(✉), Ioannis Fyrogenis, Efthymis Papadopoulos, Anastasia Nikolaidou, and Eleni Verani

Laboratory of Transportation Engineering, Department of Civil Engineering, School of Technology, Aristotle University of Thessaloniki, 541 24 Thessaloniki, Greece
{pol, fyrogeni, efthympg, nikolaid, verani}@civil.auth.gr

Abstract. In this paper we explore factors that affect Bike Sharing System (BSS) usage and how they differentiate between discrete groups of potential users. BSS have known a rampant growth during recent years, through technological advances, re-evaluated business models and reinvention of the mean's utility. Yet, for a realized use of dockless BSS and a successful integration in the urban mobility ecosystem to be achieved, the factors that promote willingness to use them need to be explored. By using a sample of 500 stated preference data, classification trees and random forest models are built for three groups of potential BSS users; car users, bus users and pedestrians. Among the considered factors are BSS cost gains, BSS In Vehicle Time (IVT) and Out of Vehicle Time (OVT) gains, trip frequency, purpose and duration. More specific, it was found that BSS potential, increases for short duration trips of up to 21 min for car users. Bus users and pedestrians were found to be more likely to choose a BSS option for a higher cost up to 0,60 and 0,75 euros respectively. On the other side sociodemographic characteristics such as household income, gender, education level and occupation did not found to be the dominant factors for the mode choice decision. OVT is found only to be relatively important for bus users, while the cost gains are comparatively more significant for bus users and pedestrians.

Keywords: Bike Sharing Systems · Willingness to use · Classification tree · Random forest

1 Introduction

Urbanization has been an increasingly rising trend in recent years. Cities are expected to amass 68% of the world's population by 2050 [1]. This combined with the car-dominant paradigm of recent years, makes adopting socially, environmentally and economically sustainable practices an urgent necessity. The emergence of sharing economies and shared mobility in particularly is widely considered a much promising solution to this predicament. More specifically, increased usage of Bike Sharing Systems (BSS) and cycling in general has been correlated with an abundance of positive

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12250, pp. 784–795, 2020.
https://doi.org/10.1007/978-3-030-58802-1_56

externalities, such as health benefits [2], reduced traffic congestion [3] such as environmental benefits [4].

BSSs' evolutionary history can be summed up in four generations of systems. The first generation of BSSs first appeared in 1965 in Amsterdam and they became known as "White Bikes" [5]. The bikes were randomly allocated in the city center and free for public use, something that made them susceptible to wear, vandalism and theft and resulted in the system's cease of operations. The second generation, known as "Coin Deposit Systems" began operations in Denmark in 1991 consisted of bikes with a lock installed, that users could unlock by a refundable deposit [4]. While this attempted to improve on the first generation's failures, it only partially did, as there was no mechanism in place to limit the bike-share usage times and the users often kept the bicycles for extended periods of time. The third generation of BSSs, that started operating in France in 1998 (known as "IT based systems") was better equipped to handle those problems [5]. It included docking stations with smart technology that made it possible to know when a bike was taken to and from a station [6]. It also incorporated electronic means of payment and high deposits that made it possible to identify the users and deter theft. While the fourth generation of BSS was already defined by many field experts before the rise of dockless BSS, as a system that is demand responsive and a fully integrated part of the city's transportation system, during recent years it has come to be identified with the rise of dockless BSSs [6, 7].

Dockless BSS' explosive growth started in China in 2015 and expanded worldwide. The systems' flexibility allows users to choose the starting and ending point of their trips without depending on rigidly placed docking stations. This is made possible by Global Position System (GPS) sensors that are installed on its bike and the widespread use of smart devices by users that allow them to easily locate, unlock and ride the bicycles in the urban environment [8]. Currently, the operational BSSs around the world are a mix of third and fourth generation systems, that often operate complimentary to one another.

Taken aback by this surge in usage and the sudden emergence of bicycles – and later scooters – in city streets, municipalities and governments lagged behind with regulations, something that created difficulties with formulating frameworks that would enable BSSs to become an integrated part of urban transportation [9].

For BSS and micromobility to reach a fully realized and optimized usage that will make it a part of a robust and resilient urban mobility system, the factors and mechanisms that make it an appealing mode of transport and increase its modal share need to be thoroughly understood.

2 Literature Review

A lot of work has already been published that explores both factors and user characteristics that enhance BSS usage. A big amount of the literature focuses on identifying the characteristics of BSS users, comparing them to the non-user population, mainly using revealed preference data, both in form of system-use data and user surveys.

Regarding the sociodemographic characteristics of BSSs users many of the studies found significant difference between bike-sharing users and the general population.

A large number of studies found current BSS users to be younger on average compared to the general population [5, 10, 11]. While the users' income was found to be significantly different by the rest of the population in a number of studies, it was not consistently found to be higher or lower [5, 11–13]. A higher education level was also identified as a contributing factor in some studies [10, 14]. While most studies agree that male users are more likely to prefer BSS, some conclude that females are [10–12, 15].

Other parts of pertinent literature have examined the way environmental factors contribute towards increased BSS usage. Improved infrastructure, such as an extended network of bicycle lanes, mixed land use, docked BSS station placemaking and weather conditions have all been found to significantly affect BSS usage [16–19].

On the contrary, little work has been done to examine users' willingness to switch to BSS usage. Campbell et al. [20] determined that trip characteristics such as trip distance, temperature, precipitation and air quality were factors that mainly effected whether users considered using BSS instead of their currently preferred mode of transport, while sociodemographic characteristics weren't found as important. Li & Kamargianni utilize short-distance, revealed and stated preference data to examine mode choice, with emphasis on bike-sharing. The results showed that choosing BSS heavily depended on air pollution, weather conditions, cost and travel time and less so on socioeconomic characteristics [21].

Based on the existing literature, a question that arises is which factors most heavily incentivize potential BSS users, that currently prefer other modes of transport, to use bike sharing across all urban trip durations:

- Does the proposed population clustering (e.g. car-bus-foot) effectively support different micro-mobility strategies?
- If yes, how and in which manner the various mode and personal characteristics are affecting the mode choice process? What are the threshold values for changing the mode choice behavior towards dockless BSS?
- Is the proposed tools used in the study (classification trees and random forest analytics) appropriate to interpret dockless BSS potential?

3 Survey Design

In order to tackle those questions, a stated preference survey was designed on the Limesurvey platform [22] and deployed on-field by trained interviewers, equipped with tablets. The data collection took place from April 2019 to May 2019 in Thessaloniki, Greece and a sample of 500 answers was collected.

The survey was answered by users of the three currently dominant modes of transport in Thessaloniki; car users, bus users and pedestrians, depending on which of those modes they used for their most frequent trip. The interviewees were asked about the characteristics of their last typical trip, including duration, IVT (In Vehicle Time) and OVT (Out of Vehicle Time), cost etc. Using those answers, competing values of cost, IVT and OVT for the BSS (or total trip time in the case of pedestrians) were calculated by the survey algorithm and presented to the interviewees in the last section of the survey as a set of 9 mode choosing scenarios. In those scenarios the interviewees

had to choose between their current mode of preference, with the cost, IVT and OVT they stated, and the dockless BSS with varying competing combinations of cost, OVT and IVT.

The variables that were collected via the survey, along with their types, levels when applied and descriptions can be seen in Table 1.

Table 1. Variables of the factors affecting willingness to use BSSs that were explored

Variable name	Variable abbreviation	Variable type	Levels/Units	Description
Cost.gain	Cs.	Numeric	Euros	The monetary gain by using BSS compared to the currently preferred mode of transport
IVT.gain	IVT	Numeric	Minutes	The IVT gain by using BSS compared to the currently preferred mode of transport
OVT.gain	OVT	Numeric	Minutes	The OVT gain by using BSS compared to the currently preferred mode of transport
T.gain	T.g	Numeric	Minutes	The total trip time gain by using BSS compared to walking (for pedestrian potential users)
Cost.BSS	C.B	Numeric	Euros	The Cost of the BSS (for pedestrian potential users)
Gender	Gnd	Categorical	Male Female	The gender of the potential user
Age group	AgG	Ordinal	18–24 25–34 35–44 45–54 55–64 65+	The age group of the potential user
Education	Edc	Ordinal	Higher Lower	The education level of the potential user (stratified into higher and lower education)
Occupation	Occ	Binary	Stable Not stable	The current occupation of the potential user (stratified into stable and not stable categories, based on the user's daily schedule)

(*continued*)

Table 1. (*continued*)

Variable name	Variable abbreviation	Variable type	Levels/Units	Description
Household income	HsI	Ordinal	0–400 € 401–800 € 801–1200 € 1201–1600 € 1601–2000 € 2001–2400 € 2400+ €	The potential user's household income
Private bike ownership	PBO	Binary	Yes No	Whether the potential user owns a private bike
Trip frequency	TrF	Ordinal	Once a day 2–3 times a day 3–5 times a week 3–5 times a month	How often the potential user repeats the specific trip
Trip purpose	TrP	Categorical	Work Education Entertainment Other social reasons	The purpose of the potential user's specific trip
Trip duration	TrD	Numerical	Minutes	The duration of the potential user's specific trip

4 Results

For each of the three groups of potential users (car users, bus users and pedestrians) a classification tree and a random forest model were built. The sample was split into a training and a testing set using a 4:1 split ratio and the models' results were validated using the testing set. 10-fold cross validation was used to fit the models and the hyperparameters of both models were tuned using grid search to select the optimal value in order to increase the model's predictivity. To achieve shorter interpretable trees, the tree models were furtherly pruned for the minimum leaf node size to be consisting of 50 or more observations. The analysis was performed in the R language for statistical computing [23]. The package dplyr [24] was used for data manipulation and the packages rpart [25], randomForest [26] and caret [27] for model fitting. The results were visualized with the rpart.plot [28] and ggplot2 [29] packages.

4.1 Car Users

The classification error of the classification tree shown in Fig. 1 was 18.59%. As can be easily interpreted from the tree, variables related to mode specific features such as duration and cost are closer to the root node (upper part of the tree). Reduced trip duration and especially increased IVT gain appear to be critical factors towards preferring the BSS. The thresholds chosen by the model for those factors are relatively low, as the first split of the dataset happens for trips shorter or longer than 21 min, while car users would require significant IVT benefits in order to make the switch to BSS. On the other hand while reduced cost gain does make preferring the BSS more likely, car users can prefer the BSS even at a slightly higher cost per trip. The combination of factors that seems to increase the likelihood of prefering the BSS over the car is a combination of low trip duration, high IVT gains, relatively younger age groups and work as a trip purpose.

The random forest model that was fitted for car users had a classification error of 10.70%. The relative importance of the variables that were used in the model is shown in Fig. 2. Similarly, to the results of the classification tree the variables that contribute most to decision making are cost gain, trip duration and IVT gain, while household income also stands out.

4.2 Bus Users

The classification error of the classification tree shown in Fig. 3 was 25.86%. As can be seen by the tree, variables relative to mode specific factors are closer to the root node of the three and seem more crucial toward splitting the sample in homogenous groups based on mode choice. Increased cost gain appears to have an important role towards choosing the BSS over the bus. Users with higher education were more likely to prefer the BSS, especially with decreased IVT compared to the bus. The combination of factors that increases the likelihood of preferring the BSS the most is users with higher education that perform shorter duration trips, on a daily basis but only if the BSS is not a much more expensive option.

The random forest model that was fitted for bus users has a classification error of 11.64%. The relative importance of the variables that were used in the model is shown in Fig. 4. Similarly to the results of the classification tree the most important factor in terms of contributing to decision making is the cost gain, while IVT gain, trip duration and OVT (Out of Vehicle Time) gain, as well as household income also are also found to be important.

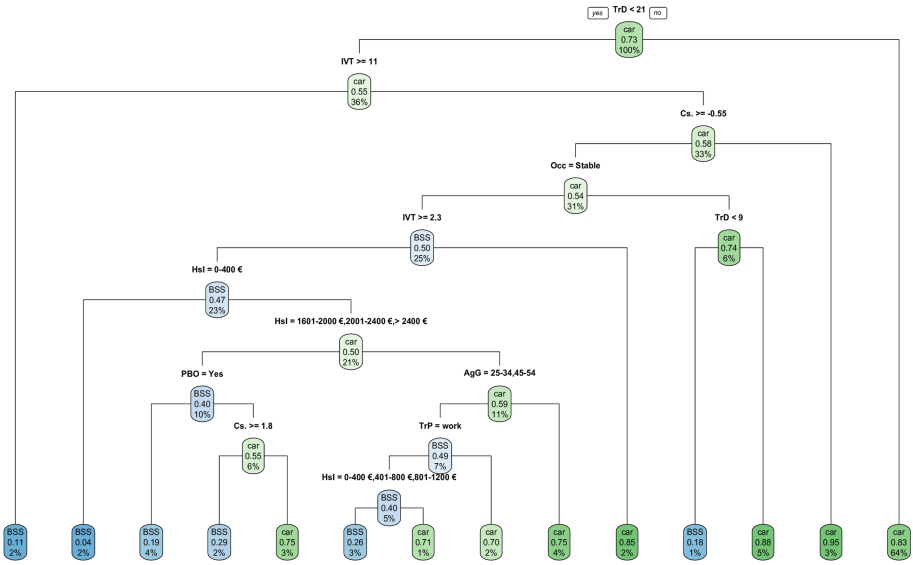

Fig. 1. Classification tree of mode choice for car users

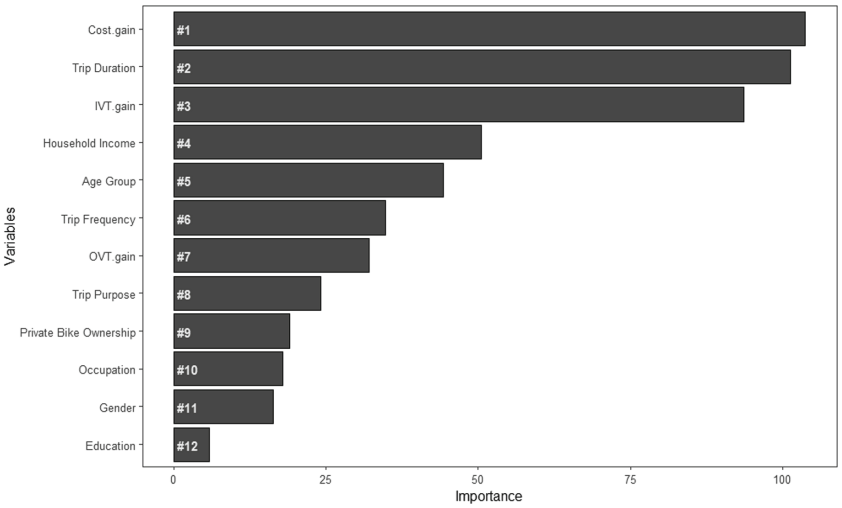

Fig. 2. Variable importance derived from the random forest model for car users

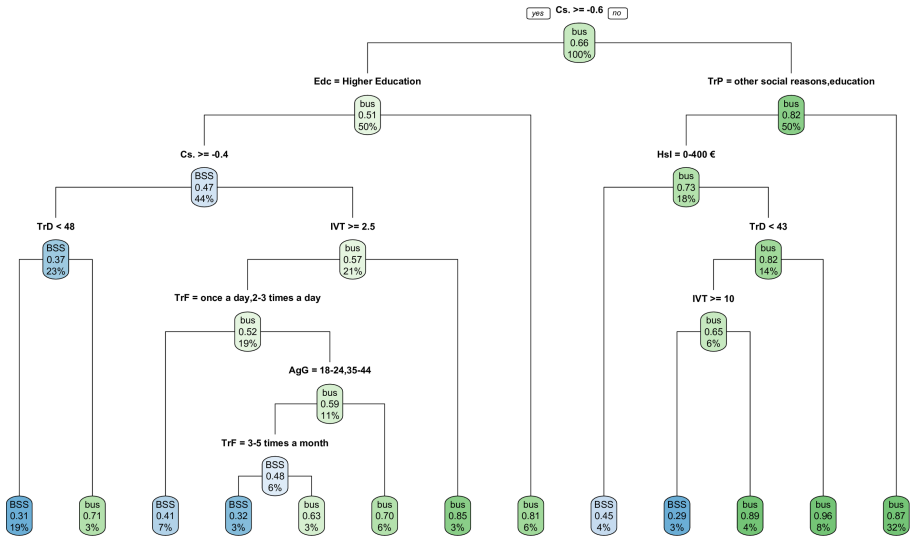

Fig. 3. Classification tree of mode choice for bus users

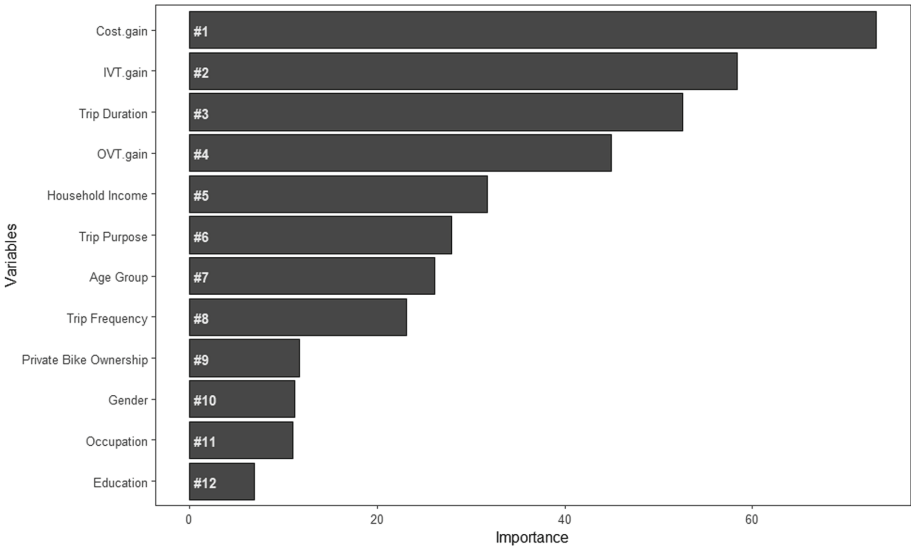

Fig. 4. Variable importance derived from the random forest model for bus users

4.3 Pedestrians

The classification error of the classification tree shown in Fig. 5 was 22.76%. As can be observed by the tree, variables relative to mode specific factors are closer to the tree's root and are more significant towards splitting the data in homogenous groups based on their choices. Decreased BSS cost appears to be crucial towards preferring the BSS, as 90% of the pedestrians preferred walking if the cost was higher than 0.75 Euros. Time gains is not found to be an important factor for decision in case of high BSS cost. Although, time gains from trips with the BSS also increase the chances of preferring the BSS over walking. The combination of factors that increases the likelihood of preferring the BSS the most is low BSS cost and increased time gains compared to walking.

The random forest model that was fitted for pedestrians had a classification error of 10.57%. The relative importance of the variables that were used in the model is shown in Fig. 6. Similarly, to the results of the classification tree the most important factor in terms of contributing to decision making is the cost of the BSS with a significant difference from the second, that is time gain from preferring the BSS. Household income, trip duration and the age group of the users follow.

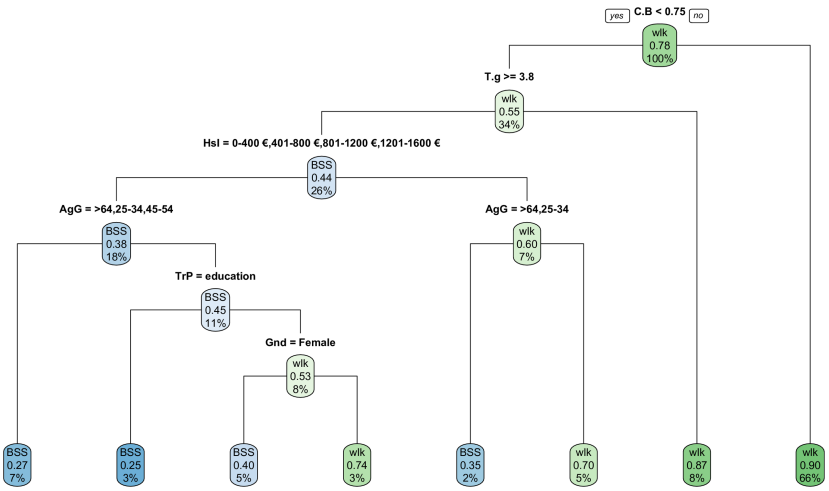

Fig. 5. Classification tree of mode choice for pedestrians

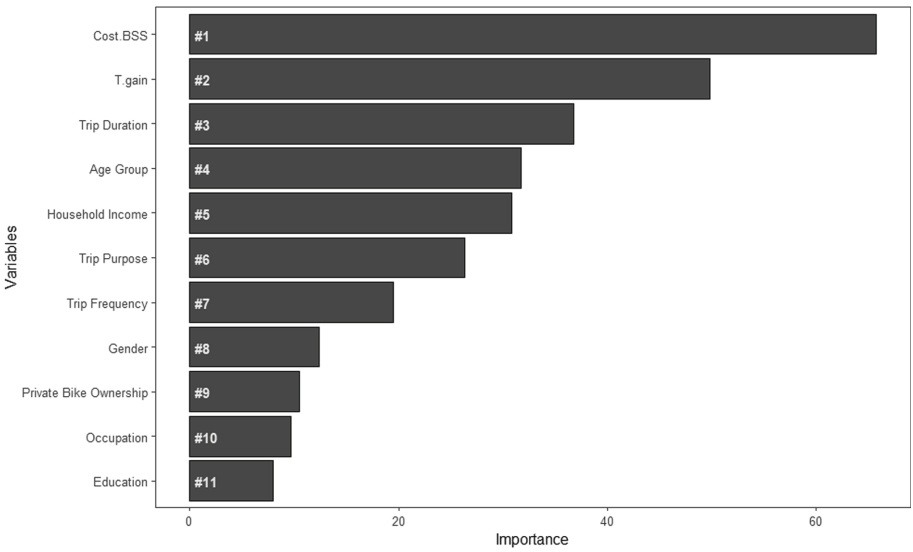

Fig. 6. Variable importance derived from the random forest model for pedestrians

5 Discussion and Conclusions

Taking into consideration all the results that were presented in Sect. 4, re-emerging or common themes, as well as discrete differences can be noticed. Cost gain, trip duration and IVT gain are the three variables with the greatest importance for all three user groups. Both car and bus users would be willing to use the BSS even if its cost is slightly higher compared to their mode of preference but only if using it is accompanied with IVT gains, while pedestrians would consider using BSS if the cost is low and it offers them trip time gains. OVT gains can be a significant factor towards using BSS for bus users, something that is true to a much lesser degree for car users. If the times it takes to locate a shared bike and get to its location is significantly less than the time it takes to walk to the bus stop and wait for the bus, this time gain is more likely to make them consider using the BSS. Furthermore, the relative importance of the possible cost gain that comes with using the BSS is higher for bus users and pedestrians compared to car users, indicating that for those groups of users keeping the BSS costs as low as possible is a more crucial factor.

The above results can have a meaningful impact on shaping the landscape of future urban mobility, in a way that it incorporates a modal share that makes the most out of each mode's strengths. Gaining a deeper understanding of the mechanisms that drive decision making in mode choice can become a powerful tool for policy makers and mobility stakeholders, allowing them to formulate future management policies, operational and business plans in a way that promotes three-pronged sustainability and maximizes each mode's positive externalities. Future research can include further analysis of the willingness to use BSS, by quantifying the effect of the incentives on different user groups and for different trip lengths.

Acknowledgements. This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-04582).

References

1. Departemen of Economic and Social Affairs United Nation: World Urbanization Prospects 2018 (2018)
2. Cavill, N., Foster, C., Oja, P., Martin, B.W.: An evidence-based approach to physical activity promotion and policy development in Europe: contrasting case studies. Promot. Educ. **13**, 104–111 (2006). https://doi.org/10.1177/10253823060130020104
3. Hamilton, T.L., Wichman, C.J.: Bicycle infrastructure and traffic congestion: evidence from DC's Capital Bikeshare. J. Environ. Econ. Manag. **87**, 72–93 (2018). https://doi.org/10.1016/j.jeem.2017.03.007
4. DeMaio, P.: Bike-sharing: history, impacts, models of provision, and future. J. Public Transp. **12**, 41–56 (2009). https://doi.org/10.5038/2375-0901.12.4.3
5. Shaheen, S.A., Zhang, H., Martin, E., Guzman, S.: China's Hangzhou public bicycle: understanding early adoption and behavioral response to bikesharing. Transp. Res. Rec. J. Transp. Res. Board. **2247**, 33–41 (2011). https://doi.org/10.3141/2247-05
6. Shaheen, S.A., Guzman, S., Zhang, H.: Bikesharing in Europe, the Americas, and Asia. Transp. Res. Rec. J. Transp. Res. Board. **2143**, 159–167 (2010). https://doi.org/10.3141/2143-20
7. Gu, T., Kim, I., Currie, G.: To be or not to be dockless: empirical analysis of dockless bikeshare development in China. Transp. Res. Part A Policy Pract. **119**, 122–147 (2019). https://doi.org/10.1016/j.tra.2018.11.007
8. Institute for Transportation & Development Policy: The Bikeshare Planning Guide. Institute for Transportation & Development Policy (2018)
9. Zarif, R., Pankratz, D., Kelman, B.: Small is Beautiful (2019). https://doi.org/10.1049/me:19900212
10. LDA Consulting: Capital Bikeshare 2011 Member Survey Report, Washington, DC (2012)
11. Buck, D., Buehler, R., Happ, P., Rawls, B., Chung, P., Borecki, N.: Are bikeshare users different from regular cyclists? Transp. Res. Rec. J. Transp. Res. Board. **2387**, 112–119 (2013). https://doi.org/10.3141/2387-13
12. Ogilvie, F., Goodman, A.: Inequalities in usage of a public bicycle sharing scheme: socio-demographic predictors of uptake and usage of the London (UK) cycle hire scheme. Prev. Med. **55**, 40–45 (2012). https://doi.org/10.1016/j.ypmed.2012.05.002
13. Rixey, R.A.: Station-level forecasting of bikesharing ridership. Transp. Res. Rec. J. Transp. Res. Board. **2387**, 46–55 (2013). https://doi.org/10.3141/2387-06
14. Fuller, D., et al.: Use of a new public bicycle share program in Montreal. Canada. Am. J. Prev. Med. **41**, 80–83 (2011). https://doi.org/10.1016/j.amepre.2011.03.002
15. Martin, E.W., Shaheen, S.A.: Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two U.S. cities. J. Transp. Geogr. **41**, 315–324 (2014). https://doi.org/10.1016/j.jtrangeo.2014.06.026
16. Bachand-Marleau, J., Lee, B.H.Y., El-Geneidy, A.M.: Better understanding of factors influencing likelihood of using shared bicycle systems and frequency of use. Transp. Res. Rec. J. Transp. Res. Board. **2314**, 66–71 (2012). https://doi.org/10.3141/2314-09

17. Buck, D., Buehler, R.: Bike lanes and other determinants of capital bikeshare trips. In: 91st Transportation Research Board Annual Meeting (2012)
18. Cervero, R., Duncan, M.: Walking, bicycling, and urban landscapes: evidence from the San Francisco bay area. Am. J. Public Health **93**, 1478–1483 (2003). https://doi.org/10.2105/AJPH.93.9.1478
19. Zhao, J., Deng, W., Song, Y.: Ridership and effectiveness of bikesharing: the effects of urban features and system characteristics on daily use and turnover rate of public bikes in China. Transp. Policy **35**, 253–264 (2014). https://doi.org/10.1016/j.tranpol.2014.06.008
20. Campbell, A.A., Cherry, C.R., Ryerson, M.S., Yang, X.: Factors influencing the choice of shared bicycles and shared electric bikes in Beijing. Transp. Res. Part C Emerg. Technol. **67**, 399–414 (2016). https://doi.org/10.1016/j.trc.2016.03.004
21. Li, W., Kamargianni, M.: Providing quantified evidence to policy makers for promoting bike-sharing in heavily air-polluted cities: a mode choice model and policy simulation for Taiyuan-China. Transp. Res. Part A Policy Pract. **111**, 277–291 (2018). https://doi.org/10.1016/j.tra.2018.01.019
22. Limesurvey. https://www.limesurvey.org/. Accessed 20 July 2019
23. R Core Team: R: A language and environment for statistical computing (2020). https://www.r-project.org/
24. Wickham, H., Francois, R., Henry, L., Muller, K.: dplyr: A Grammar of Data Manipulation (2019). https://cran.r-project.org/package=dplyr
25. Therneau, T., Atkinson, B.: rpart: Recursive Partitioning and Regression Trees (2019). https://cran.r-project.org/package=rpart
26. Liaw, A., Wienar, M.: Classification and Regression by randomForest. R News **2**, 18–22 (2002)
27. Kuhn, M.: caret: Classification and Regression Training (2019). https://cran.r-project.org/package=caret
28. Milborrow, S.: rpart.plot: Plot "rpart" Models: An Enhanced Version of "plot.rpart" (2019). https://cran.r-project.org/package=rpart.plot
29. Wickham, H.: ggplot2: Elegant Graphics for Data Analaysis. Springer, New York (2016). https://doi.org/10.1007/978-3-319-24277-4