*Avrupa Bilim ve Teknoloji Dergisi*
*Özel Sayı, S. 591-597, Nisan 2020*

***Araştırma Makalesi***

www.ejosat.com ISSN:2148-2683

*European Journal of Science and Technology*
*Special Issue, pp. 591-597, April 2020*

***Research Article***

# Station Preference Analysis of Users in Bike Sharing Systems Big Datasets*

Ahmet Şakir Dokuz [1 †]

[1] Computer Engineering Department, Nigde Omer Halisdemir University, Nigde, Turkey (ORCID: 0000-0002-1775-0954)

(Conference Date: 5-7 March 2020)

(**DOI:** 10.31590/ejosat.araconf71)

**ATIF/REFERENCE:** Dokuz, A.S. (2020). Station Preference Analysis of Users in Bike Sharing Systems Big Datasets. *European Journal of Science and Technology*, (Özel Sayı), 591-597.

**Abstract**

Bike Sharing Systems (BSS) have emerged as an alternative transportation tool for city residents who do not want to prefer conventional transportation systems. By using BSS, city residents could reach their desired destinations while making sports activity in fresh air. BSS became more preferred and prevalent among other transportation systems because of their several benefits, such as environmental friendly, activity enforcing and fresh transportation opportunity. After BSS are being utilized by more users, BSS operators started to collect the BSS datasets to gain insights from these datasets. In the literature, several applications are performed using BSS datasets, including urban pattern analysis. In this study, BSS big dataset is used for analyzing station preferences of different user types. The bike stations and their visits are counted and sorted for each user type, and top-10 preferred bike stations are extracted for each user type as preferred stations. Experimental results show that Customer and Subscriber user types have different station preferences, as hypothesized in this study.

**Keywords:** Bike Sharing Systems; Data mining; Urban pattern analysis; Station preference analysis; Big data mining

# Bisiklet Paylaşımı Büyük Veri Kümelerinde Kullanıcıların İstasyon Tercihlerinin Analizi

**Abstract**

Bisiklet Paylaşım Sistemleri (BPS), geleneksel taşımacılık sistemlerini tercih etmek istemeyen şehir sakinleri için alternative bir taşımacılık sistemi sunan ve son yıllarda yaygınlık kazanan sistemlerdir. BPS kullanan şehir sakinleri, açık havada sportif bir aktivite yaparak varmak istedikleri hedeflerine ulaşabilmektedirler. BPS çevre dostu yaklaşımları, hareketliliğe zorlayıcı yanı ve temiz taşımacılık fırsatı gibi çeşitli avantajları sayesinde diğer taşımacılık sistemlerinden daha yaygın ve tercih edilir hale gelmişlerdir. BPS'ler çok fazla sayıda kullanıcı tarafından tercih edildikçe BPS operatörleri daha iyi bilgiler edinebilmek için kullanıcılarının verilerini toplamaya başlamışlardır. Literatürde, BPS veri kümelerini kullanan ve şehir örüntü analizini de içeren çeşitli çalışmalar bulunmaktadır. Bu çalışmada, BPS büyük veri kümesi kullanılarak farklı kullanıcı türlerinin istasyon tercihlerinin analizi yapılmıştır. Her bir kullanıcı türü için bisiklet istasyonları ve bu istasyonlara yapılan ziyaretler saydırılmış ve her bir kullanıcı türü için en çok tercih edilen ilk 10 istasyon, tercih edilen istasyonlar olarak çıkarılmıştır. Deneysel sonuçlar, Müşteri ve Üye kullanıcı türlerinin farklı istasyon tercihleri olduğunu doğrulamışlardır.

**Anahtar Kelimeler:** Bisiklet Paylaşım Sistemleri, Veri madenciliği, Şehir örüntü analizi, İstasyon tercih analizi, Büyük veri madenciliği.

---

* This paper was presented at the ***International Conference on Access to Recent Advances in Engineering and Digitalization (ARACONF 2020)***.

† Corresponding Author: Nigde Omer Halisdemir University, Engineering Faculty, Computer Engineering Department, Nigde, Türkiye, ORCID: 0000-0002-1775-0954, adokuz@ohu.edu.tr

# 1. Introduction

Bike Sharing Systems (BSS) are one of the recent transportation systems which also provide city residents to be active while reaching their desired destinations (Eren and Uz, 2019). People prefer BSS due to their environmental friendly, activity enforcing and fresh transportation opportunity. With the increasing attention to the BSS, BSS operators started to collect the BSS datasets to better service their users and to make better investment decisions.

BSS operators collect the trip information of their users while they are using bikes. The trip information shows the usage characteristics of BSS of city residents. Also, BSS users are divided into two categories, i.e. Customers and Subscribers. Customers are one-pass users which do not have any contract with BSS operators. Contrarily, Subscribers are contracted users of BSS operators who regularly use bike rentals from BSS. By analyzing BSS datasets, many insights could be revealed, such as station usage characteristics, daily and hourly permanently visited stations, and station preference analysis among Customers and Subscribers.

In the literature many studies are performed for the analysis of BSS datasets for different application areas. Several studies are performed for repositioning of bikes for providing BSS users sufficient bikes at every station (Dell'Amico *et al.*, 2018; Yang *et al.*, 2019). Some other studies focus on behavioral analysis of BSS datasets which tries to extract meaningful patterns from trip activities of users (Cheng *et al.*, 2018; Wei *et al.*, 2019). Another study area is urban pattern mining which tries to analyze BSS datasets in spatial and temporal aspects to reveal spatial and temporal urban patterns (Faghih-Imani *et al.*, 2014; Jiménez *et al.*, 2016). The studies related to this study falls under urban pattern mining category.

In the studies related to this study, Vogel et al. (2011) analyzed the use of bike activity patterns of BSS to solve imbalanced bikes problem using clustering algorithms. Faghih-Imani et al. (2014) investigated the effect of different input features on origin and destination station choices in bike station dataset. Faghih-Imani and Eluru (2015) examined the BSS datasets for destination choices using bicyclist attributes, trip attributes and destination attributes with Multinomial Logit Model. Jiménez et al. (2016) considered using data mining for discovering urban mobility patterns using different ratios for bike stations classification. Wergin and Buehler (2017) analyzed spatial information of bike sharing users and found that there are major differences between customers and members of BSSs. Hyland et al. (2018) proposed a hybrid approach for bike stations modelling to find preferences of bike station selection. Li and Zheng (2019) proposed a hierarchical consistency prediction model to predict citywide bike usage in the next period.

In this study, BSS big dataset is used for analyzing station preferences of Customers and Subscribers. The main hypothesis of this study is that Customers and Subscribers have different usage characteristics on origin and destination station preference because of these users have different motivations on using BSS. For analyzing station preferences of different user types, the dataset of Divvy Bikes BSS of Chicago, USA is used. For analyzing station preferences of user types, the dataset is divided into two splits, i.e. Customer dataset and Subscriber dataset. For each dataset, the bike stations and their visits are counted and sorted. Finally, top-10 preferred bike stations are extracted for each user type as preferred stations.

The rest of this study is organized as follows. Section 2 presents the BSS dataset that is used in this study and proposed station preference analysis method. Section 3 presents the experimental results of proposed method and the discussion of the results. Section 4 presents the conclusions and future studies.

# 2. Materials and Methods

In this section, first the dataset, which is used in this study, and the preprocessing steps are introduced, and then the proposed station preference analysis method is presented.

## 2.1. The BSS Dataset

In this study, Divvy Bikes BSS dataset is used as a dataset (Bikes, 2020). Divvy Bikes operates in Chicago, USA and has a vast amount of users and daily usage of bikes vary from few hundreds to tens of thousands. Divvy Bikes has 611 bike stations which are distributed among the city of Chicago.

The BSS dataset of Divvy bikes are stored as quarter-year files resulting four files for one-year period. In this study, one quarter dataset is used as data source which starts from 1 April 2019 to 30 June 2019. 1.108.165 bike trips from the BSS users are recorded within selected dataset. The distribution of the dataset among user types is presented in Fig. 1.

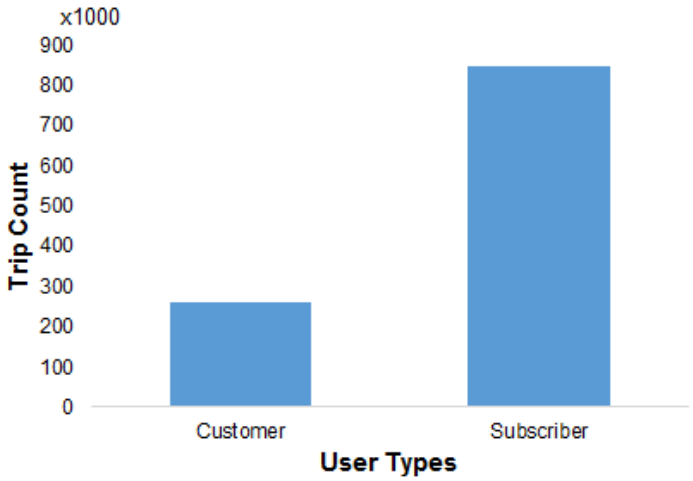


*Fig. 1 Dataset distribution among user types*

The dataset contains 12 columns that are about the bike trips of each individual record. In this study, user type, start station, and end station are used for analyzing station preferences of Subscribers and Customers.

As a preprocessing step, the quotes are removed from the dataset to obtain more viable dataset. Also, the trips that last more than 12 hours are excluded from the dataset because these trips do not indicate real bike usage of users.

### 2.2. Station Preference Analysis Method

In this study, the station preference analysis of different user types, i.e. Subscriber or Customer, is performed using a big dataset of BSS. At the first step, the dataset is read into the memory. Second, preprocessing steps are applied on the dataset. Third, the dataset is divided into two parts, such as Customer and Subscriber. Fourth, the mostly preferred stations are extracted by using occurrence count of each user type for each bike stations. Fifth, the stations are sorted with their preference counts. Finally, sixth most preferred 10 origin and destination bike stations are discovered for each user types. The method of this study is presented in Fig. 2.

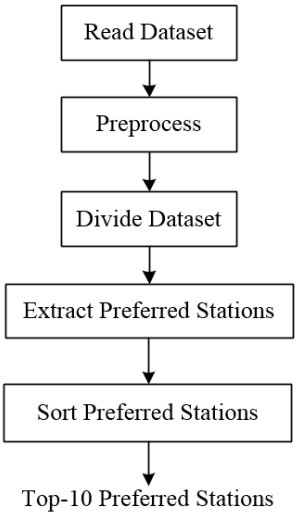


*Fig. 2 Station preference analysis method*

As can be seen in Fig. 2, the first three steps are used for preparing the dataset for analysis. The most important sections in Fig. 2 are fifth and sixth steps of Extract Preferred Stations and Sort Preferred Stations. In preferred station extraction, each record in the dataset is traversed and the user type and start station and end station are investigated. For each user type start station and occurrence in the start station are counted. Similarly, for each end station and the occurrence in the end station are counted. Finally, the stations are listed based on their occurrence counts. Then in sixth step, the bike stations are sorted using the stations and their occurrence counts. And finally, top-10 preferred origin and destination stations are displayed for both Customers and Subscribers.

## 3. Results and Discussion

In this section, the results of station preference analysis method are presented. First, the results of each user type are presented, and then discussion of the results of both user types are presented.

### 3.1. Results of Customer Users

Table 1 presents top-10 preferred origin and destination bike stations of Customer users with their proportion to the total number of records of Customer users. As can be seen in the table, origin and destination stations are same stations with little fluctuations in their preference order. Another important thing to note is the origin and destination proportion of the stations are different which means that the preference as origin or destination differ in Customer users.

*Table 1. Customer origin/destination preferred stations with their proportions*

| Order | Origin | Prop. | Destination | Prop. |
|---|---|---|---|---|
| 1 | 35 | 0.058 | 35 | 0.080 |
| 2 | 76 | 0.048 | 76 | 0.038 |
| 3 | 3 | 0.025 | 90 | 0.028 |
| 4 | 85 | 0.024 | 85 | 0.026 |
| 5 | 90 | 0.023 | 268 | 0.024 |
| 6 | 268 | 0.020 | 177 | 0.022 |
| 7 | 43 | 0.019 | 3 | 0.020 |
| 8 | 177 | 0.017 | 43 | 0.018 |
| 9 | 341 | 0.015 | 341 | 0.013 |
| 10 | 2 | 0.014 | 2 | 0.013 |

Fig. 3 presents origin and destination bike stations of Customer users which are shown in a map. As can be seen in the figure, Customer bike users prefer the bike stations that are near the shore which have a good sight.

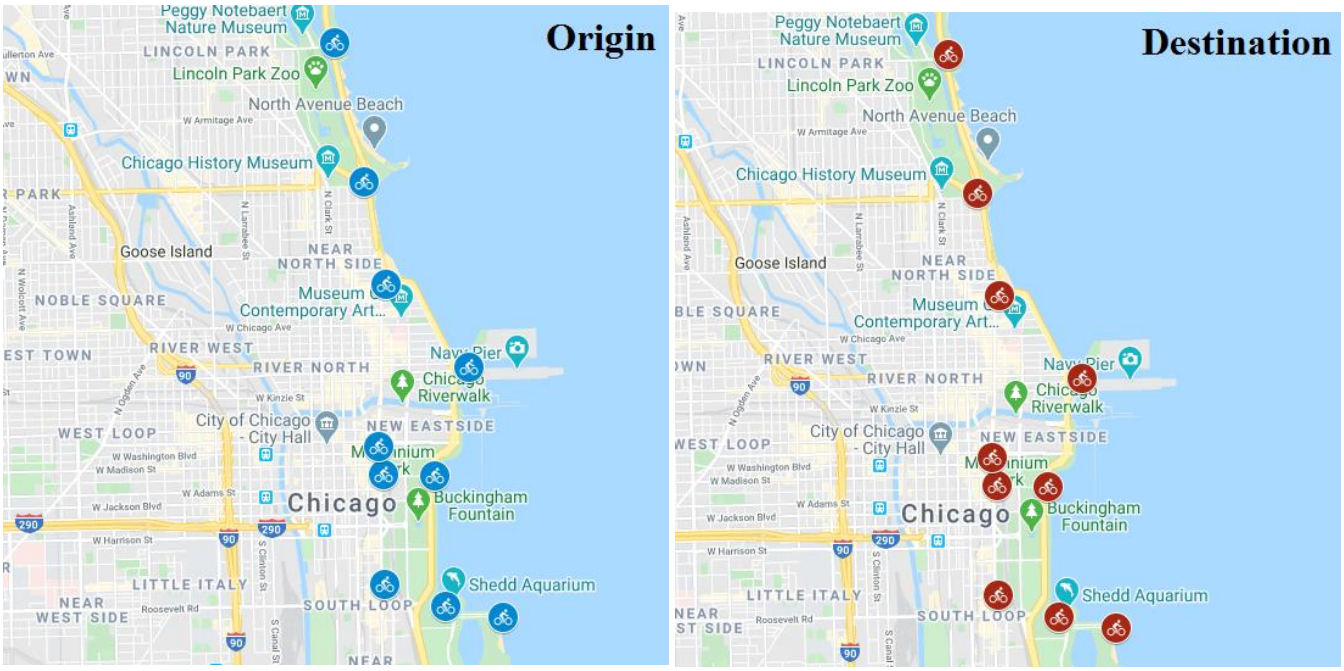


*Fig. 3 Customer origin and destination bike stations*

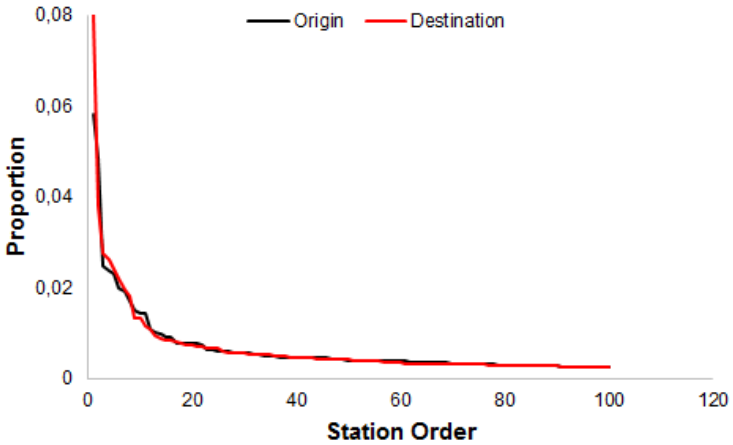


*Fig. 4 Customer preference proportion of top-100 stations*

Fig. 4 presents Customer preference proportion of top-100 stations. As can be seen in the figure, top-20 stations have the main proportion of the Customer users, which means that these stations are the mostly preferred stations for Customer users. Other stations do not have a significant preference for Customer users because these stations are not attractive.

The station preferences of Customer users show that these users choose bike sharing systems for enjoying, having a good time, and as a hobby, because these users prefer stations that have good sight and are remote from the central of the Chicago city.

## 3.2. Results of Subscriber Users

Table 2 presents top-10 preferred origin and destination bike stations of Subscriber users with their proportion to the total number of records of Subscriber users. As can be seen in the table, some origin stations are not present in destination stations and vice versa. This means that origin and destination station preferences of Subscriber users are more distributed among the bike stations, contrary to Customer users. Also, the proportions of the origin and destination stations are lower, which means that all bike stations have closer preference with respect to Customer users.

*Table 2. Subscriber origin/destination preferred stations with their proportions*

| Order | Origin | Prop. | Destination | Prop. |
|---|---|---|---|---|
| 1 | 192 | 0.016 | 91 | 0.016 |
| 2 | 91 | 0.015 | 192 | 0.015 |
| 3 | 77 | 0.015 | 77 | 0.015 |
| 4 | 195 | 0.011 | 81 | 0.010 |
| 5 | 287 | 0.011 | 133 | 0.010 |
| 6 | 81 | 0.011 | 43 | 0.009 |
| 7 | 174 | 0.010 | 174 | 0.009 |
| 8 | 133 | 0.009 | 287 | 0.009 |
| 9 | 43 | 0.008 | 176 | 0.008 |
| 10 | 100 | 0.007 | 66 | 0.007 |

Fig. 5 presents origin and destination bike stations of Subscriber users which are shown in a map. As can be seen in the figure, Subscriber bike users prefer the bike stations that are within the center of the city for shortening their urban transportation.

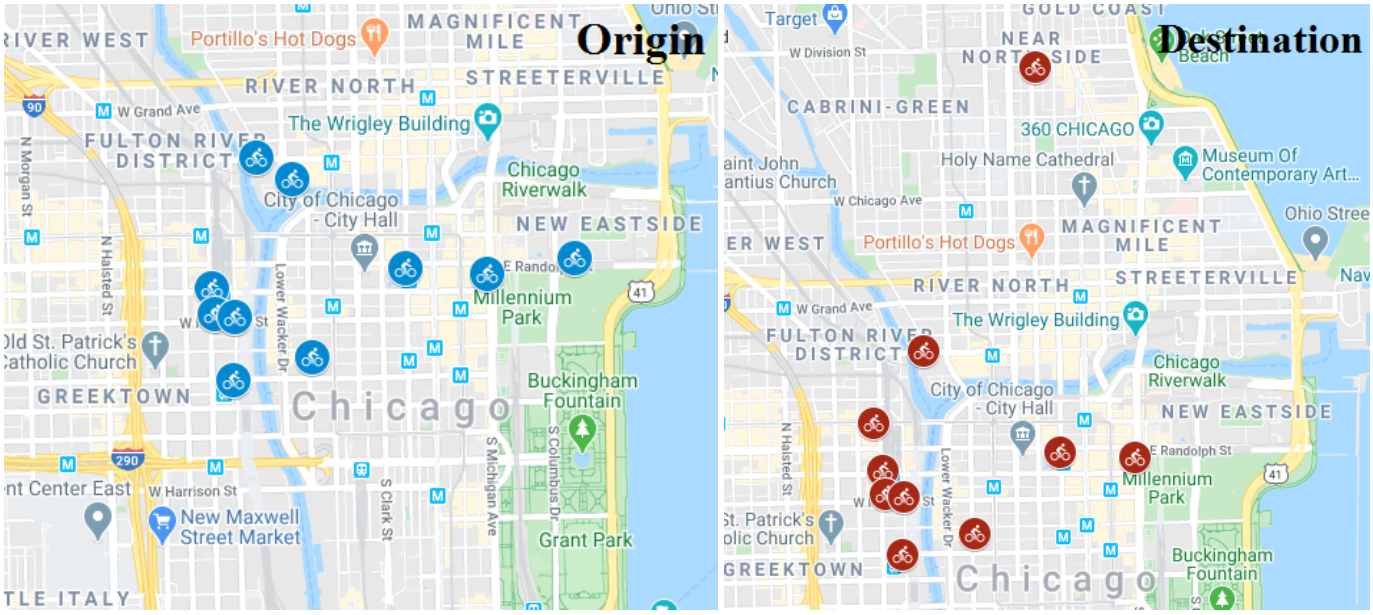


*Fig. 5 Subscriber origin and destination bike stations*

Fig. 6 presents Subscriber preference proportion of top-100 stations. As can be seen in the figure, both origin and destination station proportions are lower and the stations have closer and similar proportion, with the exception of top-5 or top-6 stations. The figure shows that Subscriber users prefer to use stations that are closer to their desired destination.

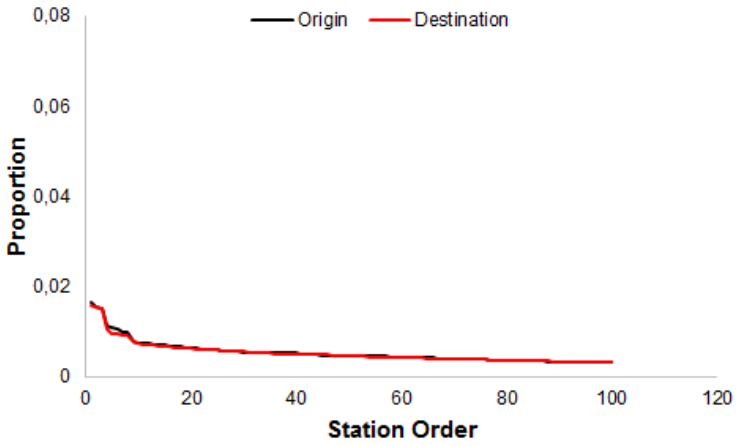


*Fig. 6 Subscriber preference proportion of top-100 stations*

The station preferences of Subscriber users show that Subscriber users primarily prefer bike sharing systems as a transportation tool. Also these users might prefer such systems for hobby or sports purposes, too.

### 3.3. Discussion

In this study, station preference analysis of Customer and Subscriber users of BSS are evaluated. The results show that station preferences and utilization change among Customer and Subscriber users. As can be seen in Tables 1 and 2, Customer and Subscriber users prefer different stations from each other, but the origin and destination station preference in each user type do not change dramatically. This result shows that Customer and Subscriber users prefer some of the bike stations for both origin and destination purposes. Also, as can be seen in Figures 3 and 5, Customer users primarily prefer BSS for hobby or sports purposes, and prefer shore bike stations which have good sight, while Subscriber users primarily prefer BSS for transportation, and prefer bike stations that are within city center to reach their desired destinations.

Also, the preference proportion of top-100 bike stations of Customers and Subscribers differ which can be seen in Figures 4 and 6. Customer users prefer little bike stations with a high proportion, while Subscriber users have more smooth preference among all bike stations. This result shows that Customer users prefer stations that are comfortable and have good sight, however, Subscriber users prefer stations that would ease their transportation.

## 4. Conclusion

In this study, station preferences of different BSS user types are analyzed. In particular, Customer and Subscriber users are evaluated whether they have different origin and destination bike station preferences. For this purpose, a method is proposed which analyzes each user type separately to extract top-10 preferred stations of user types. Experimental results show that Customer and Subscriber users have different preferences on using BSS. It's clear from the experimental results that Customer users prefer stations that are comfortable and have good sight, however, Subscriber users prefer stations that are within the center of the city and would ease their transportation.

For the future studies, the effects of hour and day information will be investigated on station preferences of BSS users which could have impact on station choices. Also, spatial and temporal analysis of station preferences of BSS users could be performed.

## References

Bikes, D. (2020) Divvy Bike Sharing Dataset. Available: https://www.divvybikes.com/ Access Date: 01.02.2020.

Cheng, P., Hu, J., Yang, Z., Shu, Y., & Chen, J. (2018). Utilization-aware trip advisor in bike-sharing systems based on user behavior analysis. IEEE Transactions on Knowledge and Data Engineering, 31(9), 1822-1835.

Dell'Amico, M., Iori, M., Novellani, S., & Subramanian, A. (2018). The bike sharing rebalancing problem with stochastic demands. Transportation research part B: methodological, 118, 362-380.

Eren, E., & Uz, V. E. (2019). A Review on Bike-Sharing: The Factors Affecting Bike-Sharing Demand. Sustainable Cities and Society, 101882.

Faghih-Imani, A., Eluru, N., El-Geneidy, A. M., Rabbat, M., & Haq, U. (2014). How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. Journal of Transport Geography, 41, 306-314.

Faghih-Imani, A., & Eluru, N. (2015). Analysing bicycle-sharing system user destination choice preferences: Chicago's Divvy system. Journal of transport geography, 44, 53-64.

Hyland, M., Hong, Z., de Farias Pinto, H. K. R., & Chen, Y. (2018). Hybrid cluster-regression approach to model bikeshare station usage. Transportation Research Part A: Policy and Practice, 115, 71-89.

Jiménez, P., Nogal, M., Caulfield, B., & Pilla, F. (2016). Perceptually important points of mobility patterns to characterise bike sharing systems: The Dublin case. Journal of Transport Geography, 54, 228-239.

Li, Y., & Zheng, Y. (2019). Citywide bike usage prediction in a bike-sharing system. IEEE Transactions on Knowledge and Data Engineering,

Vogel, P., Greiser, T., & Mattfeld, D. C. (2011). Understanding bike-sharing systems using data mining: Exploring activity patterns. Procedia-Social and Behavioral Sciences, 20, 514-523.

Wei, X., Luo, S., & Nie, Y. M. (2019). Diffusion behavior in a docked bike-sharing system. Transportation Research Part C: Emerging Technologies, 107, 510-524.

Wergin, J., & Buehler, R. (2017). Where Do Bikeshare Bikes Actually Go?: Analysis of Capital Bikeshare Trips with GPS Data. Transportation research record, 2662(1), 12-21.

Yang, Z., Chen, J., Hu, J., Shu, Y., & Cheng, P. (2019). Mobility modeling and data-driven closed-loop prediction in bike-sharing systems. IEEE Transactions on Intelligent Transportation Systems.