# Parking Behaviour Analysis of Shared E-Bike Users Based on a Real-World Dataset - A Case Study in Dublin, Ireland

Sen Yan, Mingming Liu* and Noel E. O'Connor

*Abstract*— **In recent years, an increasing number of shared E-bikes have been rolling out rapidly in our cities. It therefore becomes important to understand new behaviour patterns of the cyclists in using these E-bikes as a foundation for the novel design of shared micromobility services as part of the realisation for next generation intelligent transportation systems. In this paper, we deeply investigate the users' behaviour of shared E-bikes in a case study by using the real-world dataset collected from the shared E-bike company, MOBY, which currently operates in Dublin, Ireland. More specifically, we look into the parking behaviours of users as we know that inappropriate parking of these bikes can not only increase the management costs of the company but also result in other users' inconveniences, especially in situations of battery shortage, which inevitably reduces the overall operational efficacy of these shared E-bikes. Our work has conducted analysis at both bike station and individual level in a fully anonymous and GDPR-Compliant manner, and our results have shown that up to 12.9% of shared E-bike users did not park their bikes properly at the designated stands. Different visualisation tools have been applied to better illustrate our obtained results.**

*Index Terms*—**Shared E-Bikes, Parking Behaviour Analysis, Micromobility**

## I. Introduction

In recent years, electric bicycles (E-bikes) have been widely recognised as an effective and economical alternative to driving especially for people commuting short distances in our cities [1], [2]. By sharing the E-bikes among different users on demand, i.e., shared E-bikes, the economical benefits of daily commuting for citizens can be improved, leading to a more flexible and efficient travel pattern, which forms an essential part of the Mobility-as-a-Service (MaaS) framework in the context of smart cities [3], [4]. To illustrate this point, the research reported in [5] shows that the average distance of weekly commuting trips by E-bikes can range from 9.8km to 17km, and the trips using shared E-bikes can replace up to 76% short trips that were made by private cars. Furthermore, in comparison with other transportation tools using fossil fuels, E-bikes are not only more environmentally friendly but also can give rise to health benefits to a wide range of road users, including both young and old people, as E-bikes require less physical energy input from riders thanks to the mechanism of electrical assistance [6]. A recent study in [7] further indicates that E-bikes can generally motivate riders to cycle faster and over longer distances compared to traditional bikes, where the study points out that E-bike users can cycle 3.0km on average

per day while the pedal bike riders can commute only around 2.6km on average in the same timeline.

With the numerous benefits of E-bikes bringing to our lives, it is expected that an increasing number of shared E-bikes will be rolling out rapidly on roads in the near future. In fact, this already happens in Ireland as the Irish government has recently approved E-bike legislation to allow for these electrified bikes in road traffic in the same way as pedal cycles [1]. However, this new way of transportation will also create new challenges for traffic and operation management in our cities. More specifically, shared E-bikes, unlike pedal cycles, rely on available infrastructure to ensure their batteries are charged at the appropriate level in order to fulfil users' requests for their specific journeys. Therefore, it is essential that careful consideration must be given to these bikes to ensure that they can be parked, charged and managed properly in the desired manner, in a manner that all users can easily access to these bikes as they wish.

To this end, a considerable research effort have been devoted in recent years globally to the analysis of parking behaviour of shared bikes/E-bikes users and the design of efficient parking management strategies for these tools [8]–[13]. For instance, in [8], the authors pointed out some negative effects caused by disorderly parking in Beijing, China, and proposed some management suggestions by using a collaborative governance model to analyse parking behaviour of shared-bikes. In [10], a Mixed-Path Size logit model was devised to analyse the route choice decisions of E-bike users in the Netherlands. The work in [9] carried out analysis of both spatial and temporal characteristics of E-bike user' mobility in Tengzhou City, China. In [11], a trip purpose imputation framework was developed to assist in trip purpose prediction based on the trip records of different micro-mobility tools before and during the COVID-19 pandemic period in Zurich, Switzerland, and finally in [12], a hierarchical clustering method was employed to address the issue of free-floating bike-sharing parking in Beijing, China. However, most of the works have been found either using mathematical models with certain assumptions on users' statistical characteristics of travelling or using realistic data but with a focus on the aggregated pattern. Few works have been found using realistic data but with a focus on the analysis for individual E-bike user, especially in relation to each user's parking behaviour, as reported in this paper.

Our objective in this paper is to deeply investigate the parking behaviour of shared E-bike users at individual level by

[1]https://www.gov.ie/en/press-release/12185-government-approves-next-steps-for-escooter-and-ebike-legislation/

analysing the real-world dataset provided by MOBY Bikes [2], one of the main shared E-bikes providers based in Dublin, Ireland, as a case study. In particular, we wish to gain a deep understanding on how users parked their shared E-bikes under the current parking regulations of the company through comprehensive data analysis and visualisation, with an aim to reveal insight regarding how current parking strategies may be improved to encourage more users to better park their bikes for other users' convenience.

For a better context, according to the current regulation of MOBY, any disorderly parking of shared-bikes will incur a small monetary penalty and batteries in their shared E-bikes are required to be manually swapped by a team of operators when they are running low. Clearly, an inappropriately parked shared E-bike with a low battery pack can only be replaced depending on the availability of operators at the given time and space, this will inevitably increase the inconvenience for forthcoming users although a small penalty has to be paid by the previous rider. To further illustrate this point, two subplots from the MOBY App are shown in Figure 1 demonstrating the possibility of a potential forthcoming user with difficulties accessing an E-bike due to bikes with low battery for travel, shown in Figure 1(a), and no bike currently available in the designated station as shown in Figure 1(b).



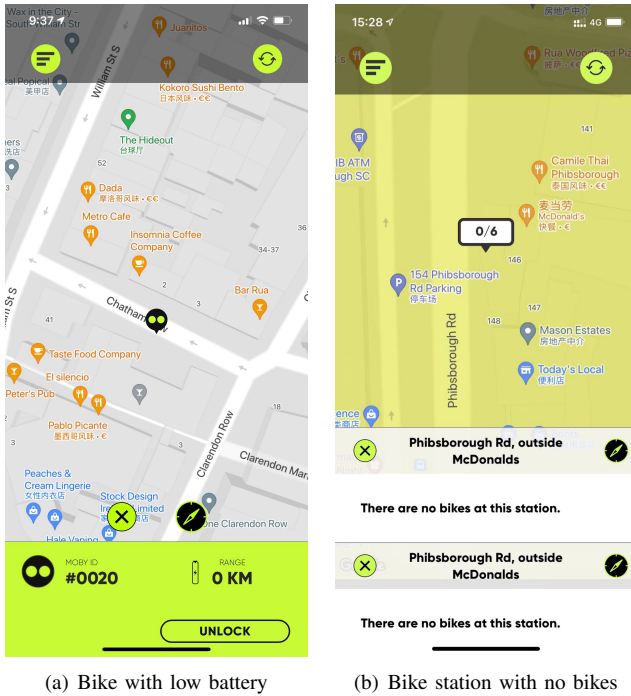(a) Bike with low battery    (b) Bike station with no bikes

Fig. 1. Inconvenience when using the app.

The rest of paper is organised as follows. In Section II, we describe the data obtained from MOBY, introduce the analysis of the dataset and demonstrate several preliminary analytical results based on this. In Section III, we look into user parking behaviours in detail. Finally, we conclude the paper in Section IV and outline some future research directions.

## II. DATA DESCRIPTION AND ANALYSIS

In this section, we describe, analyse and visualise the data collected by MOBY Bikes. It is important to highlight that the dataset received from the company consisting of both user rental reports and GPS reports is purely for academic research purpose as part of the Smart DCU [3] programme and is provided in a fully anonymous and GDPR compliant manner.

### A. Dataset Description

- GPS Information: This part the dataset reflects the GPS coordinates for each bike at different moments with 5 features (i.e., *Bike ID*, *Latitude*, *Longitude*, *Valid* and *Time*) for over 4 million records collected from January 2020 to September 2021.
- Journey Records: This is the other part of dataset showing around 58,000 journey records with 37 features in total. Users are anonymous in the dataset in order to comply with GDPR. The feature set includes information of the user, E-bike (e.g., *Bike ID*, *Bike type*), consumption (e.g. *Total price*, *Commercial conditions*), time (e.g., *Return time*, *Rental duration*), position (e.g., *Rental location*), and distance for each journey made by the user. Some briefly descriptive statistical information for total duration and distance of journey records are reported in Table I.

### B. Data Preprocessing

- GPS Information: When processing the GPS data, we found some duplicated records in the dataset. In particular, this relates to the feature *Time*, where duplicated records were found for the same timestamp due to the received GPS signals from multiple satellites at the same time. To solve this, we replaced these coordinates to the mean value in the records for every given bike and timestamp. Some redundant features in the dataset were also discarded, one of which includes the feature *Valid* which has the same value "1" in every record. Finally, after removing a small percentage of records with missing values, the cleaned dataset can be used to demonstrate the trajectory information of all bikes geographically.
- Journey Records: Some of the features in this dataset do not contain valuable information (i.e., *Voucher code* and *Status*) in relation to our analysis, so these columns were ignored in the data preprocessing step. A few records with value "null" in some features of interest, such as *Distance* and *Return time*, were removed for simplicity of our analysis. Finally, we also adopted the Z-score to determine and remove outliers from features with extraordinary high/low values in the dataset.

### C. Data Visualization

To illustrate the dataset, a trajectory of the Bike No.40 on a specific date is shown in Figure 2, where the user started on the Chapelizod Road shown in blue, and terminated on the South Circular Road in Dublin 8 shown in red.

TABLE I

SUMMARY OF DESCRIPTIVE STATISTICS FOR DURATION AND DISTANCE FOR ALL E-BIKE JOURNEY RECORDS.

| Features | Minimum | Maximum | Median | Mean | Standard deviation |
|---|---|---|---|---|---|
| Trip Distance (metres) | 1.00 | 7914.00 | 371.00 | 586.66 | 653.90 |
| Trip Duration (minutes) | 0.17 | 4587.18 | 21.62 | 56.53 | 131.93 |



Fig. 2. Trajectory of bike No.40 on a specific date.

In Figure 3, the histogram of trip duration is demonstrated and compared between weekdays and weekends. It is shown that the trip duration for most users is between 600 seconds and 800 seconds, and in particular, the distribution of trip duration on weekdays is similar with that on weekends but with more number of trips being reported.
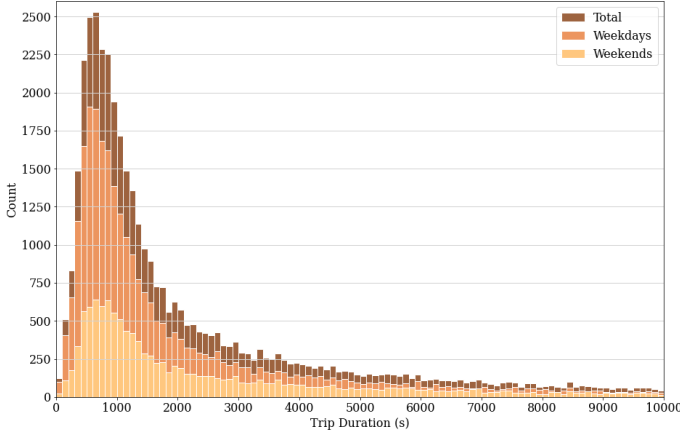


Fig. 3. The histogram of trip duration using the MOBY dataset.

Further, Figure 4 shows sum of records and the percentage values for each E-bike type with respect to different distance internals. It can be seen that the total number of travel records of all E-bikes, shown in the dark line in the figure, keeps increasing sharply till a point when it reaches to the peak level of 3,248 corresponding to the distance interval in the range of 200 to 250 metres. The number of records then falls back gradually to zero when distance interval is increased to a point just beyond 2500 metres. The *DUB-General* type of E-bike users have clearly dominated the results with at least 90.59%

users in each distance internal especially in comparison to the *Private* type of E-bike users which only appear once in the range of 850 to 900 metres.
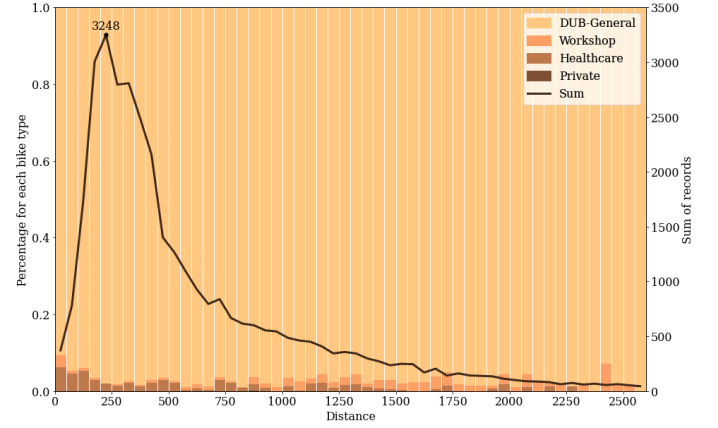


Fig. 4. Percentage and sum of records for each E-bike type.

Since our dataset contains trip records during the COVID-19 pandemic periods, we also briefly analyse the impact of COVID-19 on the E-bike users' travelling behaviours. We do this by referring to the publicly available COVID-19 dataset for Dublin Ireland based on the COVID-19 Data Repository provided by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [14]. Our result is presented in Figure 5, where we can see that the daily number of new cases in COVID-19 shown in yellow is negatively correlated with the number of trip records for MOBY bike users shown in black. However, we note that the rate of change for the number of records made by the shared E-bike users on average is significantly less in comparison to the rate of change in the COVID-related time-series data. To some extent, it demonstrates that sharing E-bike scheme is still attractive for people travelling in Ireland during COVID.

## III. PARKING BEHAVIOUR ANALYSIS

In this section, we explore the shared E-bike users' parking behaviour in detail.

### A. Return Accuracy

As aforementioned, inaccurate returns will significantly increase the difficulty of E-bike management and lead to inconvenience for other users. Thus, it is important to understand which users often return their bikes inconsiderably, and why this may happen. To do this, we need to understand the definition of an inappropriate return and then calculate the return accuracy for a given user.
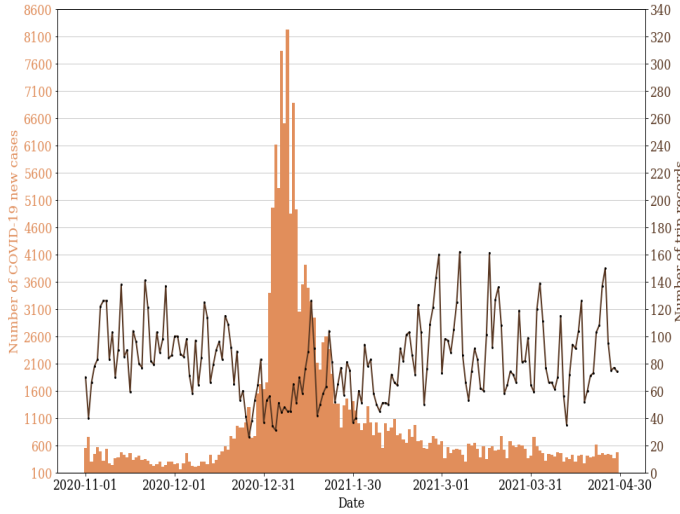
Fig. 5. The number of trip records of MOBY bikes during the COVID-19.



Fig. 6. Sum of accurate and percentage of inaccurate returns.



Fig. 7. The impact of tolerance on user's return accuracy.

According to MOBY, the GPS signal from bike is accurate within a range of 10 to 15 metres, and it is possible for this range to be further increased to 50 metres in case there are buildings around. Thus, before calculating the return accuracy, we need to determine some high-use return positions with a clustering algorithm by setting a specific radius as tolerance to accommodate those potential mistakes made by GPS devices on the bikes. To do this, we applied the Canopy cluster algorithm, an unsupervised pre-clustering algorithm that has been widely used in data science [15]. More specifically, Canopy algorithm was applied to conduct a brief clustering for the return coordinates of bikes based on their GPS coordinates. The algorithm uses two thresholds, namely $T_2$ and $T_1$, to determine the minimum and maximum distance between two points to be clustered in the same group. If the exact distance between two points locates in the range of $(T_2, T_1)$, then both points will be attached with the same cluster label, including both accurate stands predefined by MOBY and the inappropriate high-use positions we found using the algorithm. Thus, by setting $T_2 = T_1$ and varying the $T$ values, we can evaluate the return accuracy as a function of threshold, i.e., range tolerance. The results in Figure 6 illustrate that with an increasing level of range tolerance for GPS signals, the number of accurate returns is also increased gradually. According to this analysis, there were 12.9% of MOBY users who did not park their shared-bikes properly with respect to the highest GPS tolerance level set by MOBY, i.e., 50 metres.

After clustering, the return accuracy for each user can be easily calculated, which is shown in Figure 7. In the figure, the bars in different colours represent the percentages of inaccurate and accurate returns with respect to three tolerance thresholds (i.e., 0, 50 and 100 metres). It could be seen that as the tolerance grows, the accuracy for some users (e.g. user 2565 and 2557) obviously improves. However, we also find that there are several customers who always returned their E-bike inaccurately even with tolerance of 50 metres (e.g., user 2570 and 2557).
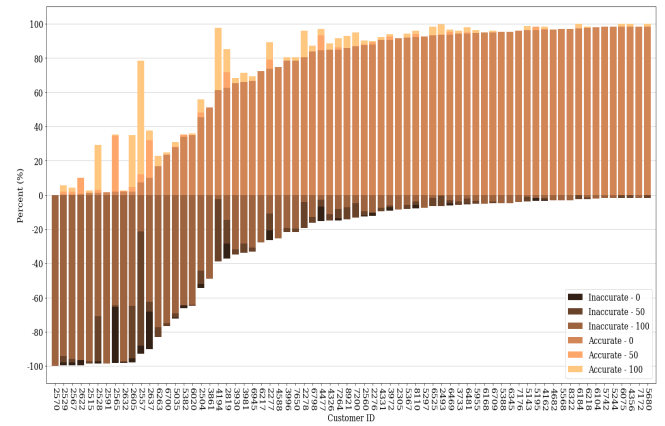
In a similar way, we can also compare the accuracy at the level of bike stands before and after adopting the tolerance. The result is shown in Figure 8, which demonstrates the number of both accurate and inaccurate return records and the return inaccuracy for several stands with and without tolerance of 50 metres, and accordingly it is obvious that the inaccuracy for some stands is remarkably higher compared with others (e.g., stand 25, 150 and 138). Thus, we conducted further analysis on users who frequently return their shared E-bikes at these stands, and the results are shown in Figure 9. It could be found that for different stands, the reasons for the low return accuracy might be different. As shown in Figure 9(a), user 2277 contributed 74.6% of incorrect returns at stand 25, while the other 25.4% is attributed to other users. Thus, it is very likely that the poor return accuracy at the stand 25 is due to the inappropriate parking behaviours of some specific users entering or leaving this building frequently. In contrast, at stand 138, which locates at the Parking Lot near the intersection of "Parnell Street" and "O'Connell Street Upper", and it could be seen in Figure 9(c), the number of inaccurate returns contributed by users are almost identical, and it is clear that the accurate station at "O'Connell Street outside AIB

Bank" is very close to stand 138, so a potential reasoning could be that the low return accuracy of the stand might be related to the high traffic flow on the "Parnell Street", where the cyclists must move across the street if they want to return their bikes properly to the "O'Connell Street". Finally, for the stand 150 shown in Figure 9(b), which locates closely to the "AIB Bank on Richmond Street South", the number of users making inappropriate returns is significantly higher compared with the other two stands. The nearest accurate return station, which locates at the location "Richmond Row, outside Atlas Language School", is around 200 metres away from this stand. Thus, a reasonable assumption here could be that the return accuracy at the stand 150 was affected by a lot of people who often cannot find a parking space in the designated parking area near the language school.
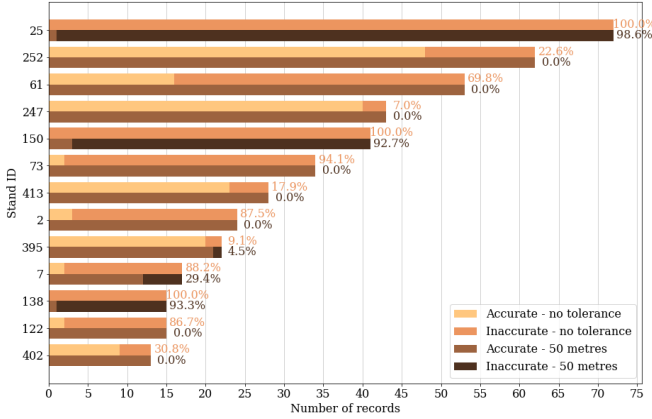


Fig. 8. The impact of tolerance on return accuracy for stands.

### B. Inference from the starting point

Up to now, we have analysed our data and reported the results primarily based on the information of return stands and their coordinates. These results are helpful for us to understand the users' parking behaviours at the end of their journeys, but these may not be good enough to infer the behaviours of users proactively. To do this, we also need to leverage the information from users' trip records based on their starting points. Mathematically, let $S := \{1, 2, \ldots, n\}$ be the indexing set of stands with a capacity $n$. Let $P_{a,b}^i$ denote the probability for user $i$ moving from stand $a \in S$ to stand $b \in S$, then it can be calculated by:

$$P_{a,b}^i = \frac{N_{a,b}^i}{N_a^i} \qquad (1)$$

where $N_{a,b}^i$ is the number of travel records of the user $i$ with stand $a$ as the starting point and $b$ as the ending point, and $N_a^i$ represents the total number of travel records of the user $i$ with stand $a$ as the starting point. Similarly, let $A_{a,b}^i$ be the return accuracy for the user $i$ with stand $a$ as the starting point and stand $b$ as the destination, then it can be represented by the formula below:

$$A_{a,b}^i = \frac{T_{a,b}^i}{T_{a,b}^i + F_{a,b}^i} \qquad (2)$$

where $T_{a,b}^i$ is the number of travel records with stand $a$ and $b$ as the starting and ending point when the E-bike is properly returned, while $F_{a,b}^i$ denotes the number of records for improperly returned trips of the user. Now, we define

$$P_a^i := \sum_{j=1}^n P_{a,j}^i \cdot A_{a,j}^i \qquad \forall i \qquad (3)$$

which represents the score of the user $i$, i.e., a value between 0-1, properly returns his/her E-bikes starting from the specific stand $a \in S$. Given this, we can obtain the prior probability of a user at the beginning of his/her journey. Our numerical results for the dataset are illustrated in Figure 10, where the six most inconsiderate E-bike users are summarised with respect to their respective starting stands presented in the pie charts.

## IV. CONCLUSION AND FUTURE WORK

In this paper, we have investigated the shared E-bike users' parking behaviour by using a realistic dataset provided by MOBY bike. Our results have shown that up to 12.9% users were not parking their bikes properly with respect to the highest technical toleration of GPS devices set by MOBY. A variety of visualisation and statistical analysis on the given dataset have been included to demonstrate the users' parking behaviours as per the company's current parking management strategy from different perspective. Given the insight from the data, our future work will focus on the design of a parking assistance system to encourage those inconsiderate users to park their shared E-bikes properly. Incentive mechanisms will be designed and performance analysis for return accuracy will also be carried out as part of our future work.

## REFERENCES

[1] M. McQueen, J. MacArthur, and C. Cherry, "The e-bike potential: Estimating regional e-bike impacts on greenhouse gas emissions," *Transportation Research Part D: Transport and Environment*, vol. 87, p. 102482, 2020.

[2] R. Nocerino, A. Colorni, F. Lia, and A. Lue, "E-bikes and e-scooters for smart logistics: environmental and economic sustainability in pro-e-bike italian pilots," *Transportation research procedia*, vol. 14, pp. 2362–2371, 2016.

[3] H. Becker, M. Balac, F. Ciari, and K. W. Axhausen, "Assessing the welfare impacts of shared mobility and mobility as a service (maas)," *Transportation Research Part A: Policy and Practice*, vol. 131, pp. 228–243, 2020.

[4] Z. H. Farahmand, K. Gkiotsalitis, and K. T. Geurs, "Mobility-as-a-service as a transport demand management tool: A case study among employees in the netherlands," *Case Studies on Transport Policy*, vol. 9, no. 4, pp. 1615–1629, 2021.
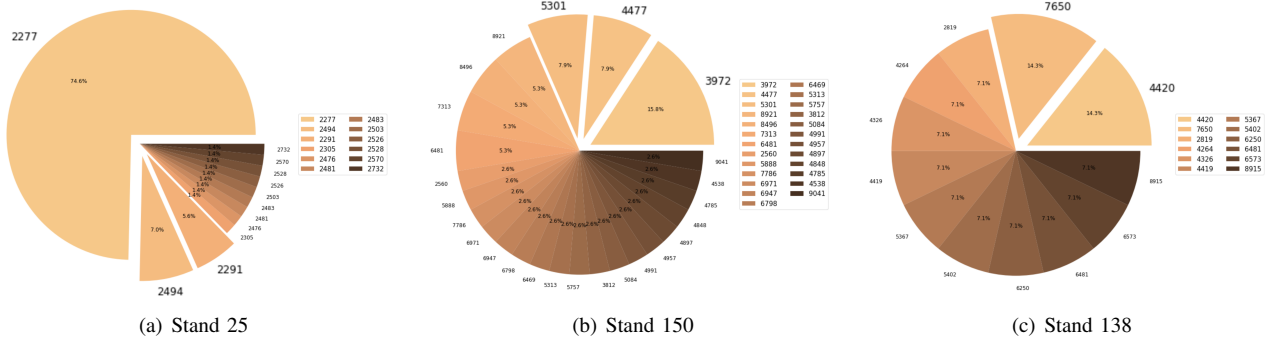
(a) Stand 25

(b) Stand 150

(c) Stand 138

Fig. 9. Users returning shared E-bikes at stands with low return accuracy.



(a) User 2529

(b) User 5035

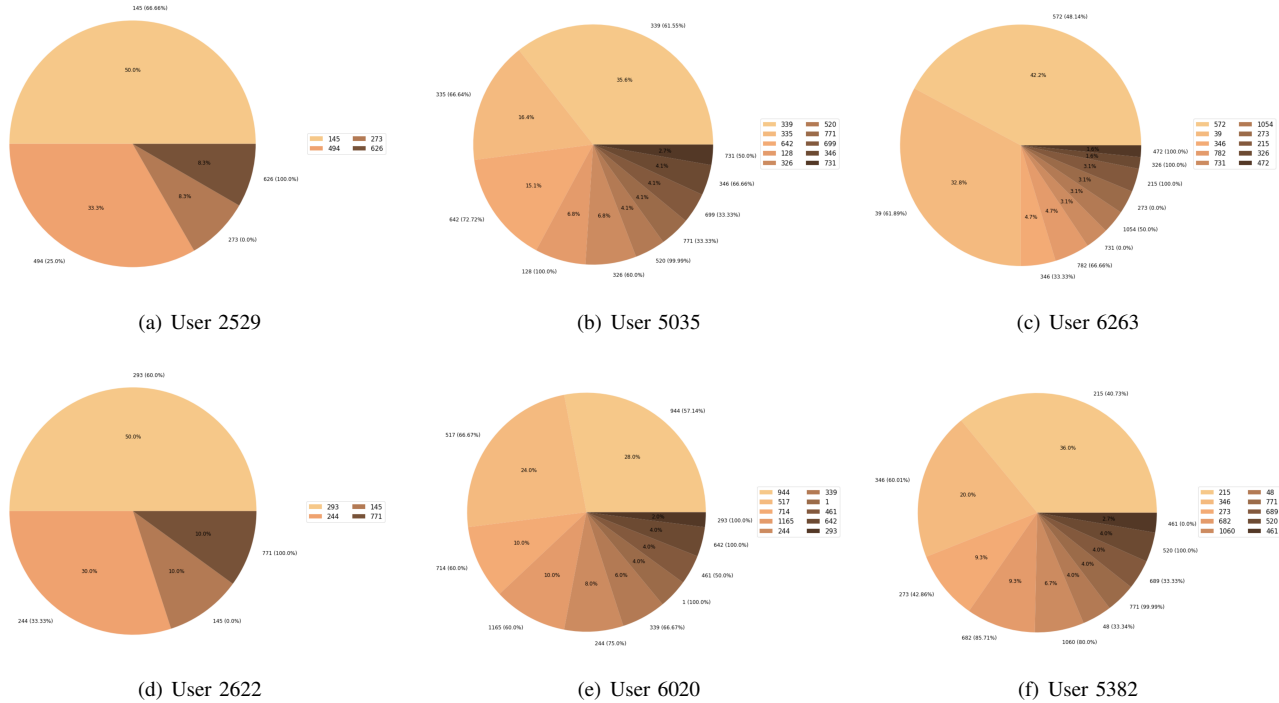(c) User 6263

(d) User 2622

(e) User 6020

(f) User 5382

Fig. 10. Performance of users with low return accuracy for each stand.

[5] S. Cairns, F. Behrendt, D. Raffo, C. Beaumont, and C. Kiefer, "Electrically-assisted bikes: Potential impacts on travel behaviour," *Transportation Research Part A: Policy and Practice*, vol. 103, pp. 327–342, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0965856415301865

[6] Y. Gu, M. Liu, M. Souza, and R. N. Shorten, "On the design of an intelligent speed advisory system for cyclists," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 3892–3897.

[7] M. Kroesen, "To what extent do e-bikes substitute travel by other modes? evidence from the netherlands," *Transportation Research Part D: Transport and Environment*, vol. 53, pp. 377–387, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361920916304837

[8] D. Zhao and D. Wang, "The research of tripartite collaborative governance on disorderly parking of shared bicycles based on the theory of planned behavior and motivation theories—a case of beijing, china," *Sustainability*, vol. 11, no. 19, 2019. [Online]. Available: https://www.mdpi.com/2071-1050/11/19/5431

[9] Y. Li, Z. Dai, L. Zhu, and X. Liu, "Analysis of spatial and temporal characteristics of citizens' mobility based on e-bike gps trajectory data in tengzhou city, china," *Sustainability*, vol. 11, no. 18, 2019. [Online]. Available: https://www.mdpi.com/2071-1050/11/18/5003

[10] G. Dane, T. Feng, F. Luub, and T. Arentze, "Route choice decisions of e-bike users: Analysis of gps tracking data in the netherlands," in *Geospatial Technologies for Local and Regional Development*, P. Kyriakidis, D. Hadjimitsis, D. Skarlatos, and A. Mansourian, Eds. Cham: Springer International Publishing, 2020, pp. 109–124.

[11] A. Li, P. Zhao, H. He, and K. W. Axhausen, "Understanding the variations of micro-mobility behavior before and during covid-19 pandemic period," *Arbeitsberichte Verkehrs- und Raumplanung*, vol. 1547, 2020.

[12] D. Xu, Y. Bian, J. Rong, J. Wang, and B. Yin, "Study on clustering of free-floating bike-sharing parking time series in beijing subway stations," *Sustainability*, vol. 11, no. 19, 2019. [Online]. Available: https://www.mdpi.com/2071-1050/11/19/5439

[13] Z. Chen, H. Wu, N. E. O'Connor, and M. Liu, "A comparative study of using spatial-temporal graph convolutional networks for predicting availability in bike sharing schemes," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021, pp. 1299–1305.

[14] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track covid-19 in real time," *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 533–534, 2020.

[15] A. Kumar, Y. S. Ingle, A. Pande, and P. Dhule, "Canopy clustering: a review on pre-clustering approach to k-means clustering," *Int. J. Innov. Adv. Comput. Sci.(IJIACS)*, vol. 3, no. 5, pp. 22–29, 2014.