BIKESHARE USERS ON A BUDGET?
A TRIP CHAINING ANALYSIS OF BIKESHARE USER GROUPS IN CHICAGO

Word count: 6,615 words (text) + 3 tables x 250 words (each) = 7,365 words

Submission Date: February 15, 2019

Siyue Yang
City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong, SAR
Email: siyueyang2-c@my.cityu.edu.hk

Candace Brakewood (Corresponding Author)
University of Tennessee
320 John D. Tickle Building, 851 Neyland Drive, Knoxville, TN, USA 37996
Phone: (865)-974-7706
Email: cbrakewo@utk.edu

Virgile Nicolas
Cerema - Département Laboratoire de Saint-Brieuc
5, rue Jules Vallès, 22015, Saint-Brieuc, France
Email: virgile.nicolas@cerema.fr

Jake Sion
Transit
5333 Avenue Casgrain, Suite 803, Montreal, Quebec H2T 1C2, Canada
Email: jake@transitapp.com

ABSTRACT

This analysis focuses on a smartphone app known as "Transit" that is used to unlock shared bicycles in Chicago. Data from the app were utilized in a three-part analysis. First, Transit app bikeshare usage patterns are compared to system-wide bikeshare utilization using publicly available data. The results reveal that hourly usage on weekdays generally follows classical peaked commuting patterns; however, daily usage reached its highest level on weekends. This suggests that there may be large numbers of both commuting and recreational users. The second part aims to identify distinct user groups via cluster analysis; the results reveal six different clusters: (1) commuters; (2) utility users; (3) leisure users; (4) infrequent commuters, (5) weekday visitors; and (6) weekend visitors. The group unlocking the most shared bikes (45.58% of all Transit app unlocks) was commuters, who represent 10% of Transit app bikeshare users. The third part proposes a trip chaining algorithm to identify "***trip chaining bikers***." This term refers to bikeshare users who return a shared bicycle and immediately check out another, presumably to avoid paying extra usage fees for trips over 30 minutes. The algorithm reveals that 27.3% of Transit app bikeshare users exhibit this type of "***bike chaining"*** behavior, presumably to avoid paying additional usage fees. However, this varies substantially between user groups; notably, 66% of Transit app bikeshare users identified as commuters made one or more bike chaining unlocks. The implications are important for bikeshare providers to understand the impact of pricing policies, particularly to encourage turn-over of bicycles.

INTRODUCTION

Bikesharing has grown rapidly in the recent years. In the United States, only 320,000 trips were made by bikeshare in 2010; by 2017, there were approximately 35 million bikeshare trips *(1).* This increase is the result of significant investment in infrastructure *(2)* as well as the perceived benefits of bikesharing services, including congestion mitigation *(3)*, travel time saving for users *(4),* and health benefits associated with active lifestyles *(5; 6; 7)*.

Although increasing bikesharing can be beneficial, it also brings challenges for management and operations. To effectively manage the system, a good understanding of bikeshare usage patterns is critical *(8)*. Prior research on bikeshare usage patterns often relies on data collected via travel surveys *(2; 9; 10)*; however, the small sample size and cross-sectional nature of most surveys typically do not allow for longitudinal analyses at the individual or system level. Many other prior studies of bikeshare usage have relied on publicly available trip data from bikeshare operators, which usually contain geographical and temporal information for each bikeshare trip. However, these datasets typically lack a unique identifier for each user, so analysis of individual patterns over time is not possible *(11)*. Hence, there is a gap in the prior literature pertaining to longitudinal analysis of individual bikeshare user behavior.

This paper aims to address this by investigating bikeshare user behavior over a three-month period in Chicago. We utilize data from a smartphone app known as "Transit" that can unlock shared bicycles. This dataset advantageously allows for analysis of individual bikeshare user patterns over time because app interactions are assigned an anonymized unique identifier. We conduct a three-part analysis of user behavior. First, Transit app bikeshare usage patterns are compared to system-wide bikeshare utilization in an exploratory analysis of temporal patterns. In the second part, a cluster analysis of Transit app bikeshare users is conducted to identify groups exhibiting similar behaviors. The third part proposes a trip chaining algorithm to identify "***bike chaining***." This term is introduced to refer to bikeshare users who return a shared bicycle and immediately check out another, presumably to avoid paying extra usage fees for trips over 30 minutes; this analysis is motivated by anecdotal evidence of this practice *(12).* The implications are important for bikeshare providers to understand the impact of their pricing policies, particularly to encourage turn-over of shared bicycles.

PRIOR RESEARCH

As previously noted, prior research on bikeshare usage patterns often relies on data collected via travel surveys *(3; 9; 10)*. Recently, there has been an increase in studies examining bikeshare using trip-level data provided by bikeshare operators, which generally includes the time and location of bike unlocks or relocks at stations *(13).* Numerous prior studies have exploited the spatiotemporal nature of bikeshare trip data, and one increasingly common method used to analyze geographic patterns is cluster analysis. For example, one early study using clustering methods on bikeshare trip data was conducted in 2009, and it applied hierarchical clustering by station. The docking stations were classified by usage frequency and three cycling patterns were identified: outgoing (unlocks in the morning and locks in the evening), incoming (unlocks in the evening and locks in the morning), and flat patterns *(14)*. A few more recent studies have conducted similar analyses by grouping stations in London and Paris by their temporal utilization *(15; 16; 17)*. In 2013, the spatial characteristics of bikesharing trip data for Vienna were used to classify stations into communities, providing deeper insight into the relationship between places of activity and bikesharing *(18)*. More recently, various clustering methods have been used to categorize bikeshare flow patterns *(5; 19)* and help with fleet rebalancing *(20).* Among these numerous cluster analysis studies, few if any have considered classifying users at the individual level by trip purpose. Therefore, this research aims to explore trip purpose-related patterns of user groups by exploiting a new, individual-level dataset from a smartphone app.

Two other prior studies motivate this research. The first study analyzed temporal usage patterns of 38 bikeshare systems across the globe. The authors proposed a qualitative framework to categorize cities by their likely type(s) of bikeshare users, which included four groups: commuters, utility users, leisure users, and tourist users *(11)*. ***Commuters*** usually rent bicycles to travel between home and work or between home and transit stations on weekdays during rush hour (6 – 10 a.m. and 4 – 8 p.m.). ***Utility users*** utilize bikeshare on the weekdays for shopping and errands. ***Leisure users*** generally ride on the weekends for fun and exercise. ***Tourist users*** generally ride shared bicycles to destinations such as the beach or to explore the city. This framework will serve as an initial hypothesis for the trip-purpose based classification proposed in this paper, and we will expand on it using cluster analysis.

The second noteworthy study examined the cost sensitivity of bikeshare users in Boston and Washington, D.C. using publicly available trip data. Trip lengths were analyzed, and the results reveal that riders often return bicycles just before additional usage fees are charged (typically at the 30-minute mark). Moreover, registered users tend to return bikes just before the 30-minute payment boundary more frequently than casual users *(21)*. This study inspired our investigation of "***bike chaining***," in which bikeshare users chain trips to avoid paying additional fees.

BACKGROUND

This section provides relevant background information on the bikesharing system in Chicago and the smartphone app that is the focus of this analysis.

Background on Chicago's Bikesharing System

The bikesharing system in Chicago is known as Divvy. The Divvy bikeshare system was initially launched in 2013, and over the last five years, it has grown to over 6,000 bikes at 570+ stations in Chicago and the nearby suburb of Evanston *(22).* To use a shared bicycle, travelers can "unlock" a bike from a Divvy station after paying at a kiosk or via their smartphone. During the period of analysis considered in this paper (2016), the only native smartphone application that could be used to unlock Divvy bikes was the Transit app; however, since then, other apps have become available for unlocking Divvy bicycles.

The price to use a Divvy bike is shown in Table 1. Notably, Divvy adjusted its pricing policy in February of 2018 and both the previous pricing policy (2016) and the current pricing policy (2018) are shown in the table. In 2016, which is the timeframe considered for the following analysis, the pricing structure included an annual pass for $99 and a 24-hour pass for $9.95. With either option, riders had 30 minutes to use a shared bike, and after that, they were charged an additional usage fee. Since this analysis was conducted, Divvy has changed their pricing policy, and now annual pass users have 45 minutes to complete a trip and 24-hour pass users have three hours. This adjustment was reportedly made to accommodate high utilization levels by visitors *(23)* and to simplify the pricing policy for additional charges *(24)*.

Background on the Transit Smartphone Application

The Transit app is a free smartphone application providing urban transportation information in more than 175 cities around the world. Its most used feature is real-time transit information. The app also integrates other shared mobility services such as ridehailing, carsharing, or bikesharing in some cities. Since 2015, users can purchase bikeshare passes and unlock bikeshare bicycles directly from the Transit app in some cities in the USA and Canada *(25),* and this feature was launched for Chicago in 2016. To use the app for bikeshare, a user selects a bikeshare station on the map or taps the bikeshare card below the map, as shown in Figure 1a. The first time a user utilizes this feature, s/he has the choice to sign in if s/he already has a bikeshare membership or to purchase a one-day pass directly from the Transit app (as of 2016). To purchase a daily pass, s/he must create an account by completing the form shown in Figure 1b. Once the user has an account and is logged in, s/he can tap the "Get a Bike" icon (shown in Figure 1a), and a 5-digit code will be displayed (Figure 1c). Finally, s/he selects a bike at a bikeshare station and enters the 5-digit code to unlock the bike.

DATASET

The primary dataset for this analysis is from the Transit app. Three months of data (from May 23 to August 21) in 2016 were utilized. The dataset includes, among other things, a device identifier (ID) assigned to each smartphone that has downloaded the app, the location (latitude/longitude) of the station where the bike was unlocked, and a timestamp. Notably, the data does not contain any personally identifiable information to preserve users' anonymity.

The raw dataset was cleaned in the following manner. First, the dataset was processed to keep only bikeshare unlocks. Users can interact with the Transit app in numerous ways, such as signing into a bikeshare account. Therefore, the dataset was filtered to include only interactions corresponding to receiving a code that could then be used to unlock a bicycle. Second, some users requested several codes at the same station within a short time interval. These records might correspond to users who did not reach the bikesharing station within a time limit during which the unlock code must be used (i.e. Divvy unlock codes expire within five minutes if they are not used) or users who obtained a code for a station that did not have any bicycles remaining. To account for this, we removed records at the same station within a 10-minute interval, and only the last unlock was retained since only one unlock likely corresponded to an actual trip. After the data cleaning, 125,570 unlocks made by 11,446 unique Transit app bikeshare users (by device ID) remained.

ANALYSIS

The analysis was conducted in three primary parts. First, an exploratory analysis of temporal bikesharing usage was performed to investigate patterns and compare Transit app data to system-wide data. The second analysis classified individual Transit app bikeshare users into groups using cluster analysis. The third part examined individuals' daily bikeshare usage patterns to identify trip chaining and explore the potential impacts of the bikeshare pricing policy on different user groups.

Part 1: Exploratory Analysis of System Trends

The first analysis assessed trends in bikesharing utilization over time in Chicago. The results of the temporal pattern analyses are shown in Figure 2. Transit app unlocks are shown in orange and system-wide trips from the Divvy operator's website are shown in blue *(25)*. According to the system-wide Divvy trip dataset that is publicly available, there were a total of 1,464,585 bikesharing trips during the three-month study period, and 76% of the trips were taken by bikesharing annual members *(25)*. The use of Transit app in Chicago represents approximately 8.6% of all bikesharing trips during this time period; however, Chicago likely has different Transit app utilization levels now because this dataset is more than two years old.

As shown in Figure 2a, the mean number of unlocked bikes per hour in a day shows classical peaking patterns, which is typical of commuting patterns *(13).* The number of trips per day in the selected three months is shown in Figure 2b, and it exhibits some fluctuations. Some level of fluctuation might correspond to poor weather *(26)*. For instance, it was rainy with mist and haze on May 31, July 24, and August 12, and these days show lower levels of bikeshare use in the graph. Additionally, there are high levels of daily utilization on most Saturdays, particularly July 16, which may correspond to tourism and/or recreational use.

Figure 2c presents average bikeshare use by day of the week during the three-month study period. Two main peaks are observed during weekdays: one in the morning, most likely corresponding to commuting trips and a higher peak in the evening, most likely corresponding to a combination of evening commuting trips and leisure trips. One reason for a smaller morning peak period may be hygiene; the lack of shower access at work could be a barrier to bikesharing in the morning *(27)*. There is also a smaller weekday peak at lunch time. During weekends, ridership seems to be more regular across time, probably corresponding to leisure or recreational trips. Last, the peak on weekends is typically lower than on weekdays.

Overall, these results suggest that trip purpose may vary between bikeshare users. Figure 2c shows high levels of utilization on Saturdays, which likely corresponds to recreational and tourism usage. On the other hand, Figure 2a and Figure 2c have classical weekday peaking patterns appearing in the morning and evening, which is likely for commuting purposes.

Last, the graphs shown in Figure 2 suggest that the smaller Transit app dataset follows similar temporal patterns as the larger, publicly available Divvy trip dataset. To further explore this, the Kolmogorov-Smirnov test was used to compare the two datasets shown in Figure 2 by hours in a day, by day, and also by hours in a week. The corresponding p-values were 0.90, 0.23, and 0.92 (by hours/day, days, and hours/week, respectively). All three p-values were larger than 0.05, and therefore, we cannot reject the hypothesis that the two distributions are the same. Thus, it is reasonable to argue that the Transit app dataset is representative of the Divvy dataset (by hours/day, hours/week, and days).

Part 2: Cluster Analysis to Identify User Groups

The results of the exploratory analysis in Part 1 suggest that bikeshare may be used for different purposes in Chicago (e.g., commuting, recreation). Moreover, the previous literature has found at least four different groups of users in bikeshare systems: commuters, utility users, recreational users, and tourist users *(11)*. Therefore, we hypothesize that there are multiple groups of bikeshare users in Chicago, and to explore this, a cluster analysis was conducted. We designed several variables to describe each Transit app bikeshare user, and the correlation between variables was checked to assure that they were sufficiently different. Then, the following four variables were selected for this analysis:

- **Days of use**: the total number of days in the three months an individual user (by device ID) unlocked bikeshare bicycles using the Transit app. This variable is designed to distinguish frequent users from infrequent users.

- **Duration of days**: for every user (by device ID) d, the duration of days $T(d)$ is defined as the time interval between the first day of Transit app bikeshare use $F(d)$ and the last day of use $L(d)$, where the day is expressed as an integer from the start of the year; for example, January 1 is expressed as 1. When a bikeshare users has unlocks throughout the three-month study period, it likely implies that s/he is likely a resident of Chicago. Therefore, we designed this variable to judge whether a user likely lives in the city or may be a visitor.

$$T(d) = L(d) - F(d) \quad (1)$$

- **Weekday usage rate**: for every user (by device ID) d, the weekday usage rate $W(d)$ is calculated by dividing the number of Transit app bikeshare unlocks per weekday $BW(d)$ by the total number of unlocks taken in the three months $U(d)$. This variable is intended to differentiate between potential weekday user groups (e.g., commuters, utility users) and potential weekend user groups (e.g., leisure users, visitors).

$$W(d) = \frac{BW(d)}{U(d)} \quad (2)$$

- **Rush hour usage rate**: for every user (by device ID) d, the rush hour usage rate $R(d)$ is calculated by dividing the number of Transit app bikeshare unlocks during rush hour $BR(d)$ by the total number of Transit app bikeshare unlocks taken in the three months $U(d)$, where rush hour refers to 6 – 10 a.m. and 4 – 8 p.m. on weekdays. This variable will help identify commuters.

$$R(d) = \frac{BR(d)}{U(d)} \quad (3)$$

These four values were calculated for each of the 11,446 Transit app bikeshare users. Then, we applied an unsupervised machine learning algorithm, Hartigan-Wong k-means clustering, to the dataset using the open source software program R. It works as follows: given a number of clusters k, assign the samples into random k clusters and calculate the centroids by Euclidean distance; keep updating the centroids if the centroids change when removing a sample from that cluster; if an update occurs, return to the first cluster and repeat the previous step until all the samples are involved and no additional updates happen *(28)*. The algorithm is commonly used for clustering unlabeled data and has various applications in industry, such as customer relationship management and advertisement *(29)*. Last, to eliminate the potential sensitivity to the sequence of objects and obtain stable clusters, the variables were standardized, and the analysis was run 20 times. The number of clusters k was determined by silhouette analysis *(30),* and finally, k was estimated to be six.

Next, we created numerous visualizations of the clusters to better understand their characteristics. These visualizations are shown in Figure 3. Figure 3(a) has three types of plots to show the statistical features of each cluster: scatterplots along the left side; density plots along the diagonal, and boxplots on the right side. The clustering results were also visualized in a radar graph as shown in Figure 3(b), which was created using Python. Last, the color coding for each cluster is the same for Figure 3(a) and 3(b); the legend is shown on the bottom right side (in Figure 3b) and includes the cluster labels, which are discussed in the following paragraph.

Based on the visualizations of the clusters shown in Figure 3, we assigned labels to each cluster. First, we assigned the four labels from the prior literature first (commuters, utility users, leisure users and tourists) and made a few minor wording changes (‘tourists’ to ‘visitors’; ‘recreational’ to ‘leisure’ users). Next, we defined labels for the remaining two clusters according to their features. We first identified frequent and infrequent users using the variables ‘days of use’ and ‘duration’ of days. Then, we used ‘weekday usage rate’ and ‘rush hour usage rate’ to classify these users.

The six clusters and their corresponding labels can be summarized in the following way:

- **Commuters** (shown in cyan in Figure 3): Commuters used bikeshare on numerous days during the study period (as shown by a high value of the variable 'duration'). They mostly rode during rush hour on weekdays (variable 'rush hour usage' was high). Overall, they were very heavy users.

- **Utility users** (shown in yellow in Figure 3): Utility users may be Chicago residents because this group also had a relatively high value for the variable 'duration'. However, they did not use bikesharing service as much as commuters; their typical number of 'days of use' was not as high as commuters. These users generally cycled on weekdays, potentially for errands or shopping (as indicated by a high value of the variable 'weekday usage' but a low value for 'rush hour usage').

- **Leisure users** (shown in red in Figure 3): Leisure users likely live in Chicago, as indicated by a relatively high value of the variable 'duration'. However, they had a relatively low rate of use on weekdays ('weekday usage' is low). The cyclists in this group likely used shared bicycles to go to the gym, ride to the park, or make other recreational trips.

- **Infrequent commuters** (shown in green in Figure 3): This group of users followed a similar pattern to commuters ('rush hour usage' is relatively high), but they do not use shared bikes very often (both 'days of use' and 'duration' are low).

- **Weekday visitors** (shown in blue in Figure 3): These users mostly cycled on weekdays ('weekday usage' is comparatively high) but they had a low duration value, suggesting that they may not be from the Chicago area.

- **Weekend visitors** (shown in purple in Figure 3): This group of users follows a similar pattern to visitors ('duration' and 'days of use' are low), but they cycled mostly on weekends ('weekday usage' is low).

The number of Transit app users and bikeshare unlocks for each cluster is shown in Table 2. This table reveals that commuters account for 10% of all Transit app bikesharing users, but 45.58% of the trips are taken by this group. This is consistent with a previous study of Montreal, which concluded that regular members make large numbers of bikeshare trips *(31)*. The groups that use bikesharing service infrequently (infrequent commuters, weekday visitors and weekend visitors) in Chicago comprise 55% of Transit app bikesharing users. This is likely because in 2016, only 24-hour passes were sold directly through the Transit app, whereas annual passes must be purchased via other means *(25)*. Those who purchased annual passes could then sign into their account via the Transit app to unlock bikes, but this extra step may explain why there were more infrequent users compared to frequent users.

Part 3: Trip Chaining Algorithm to Identify Bike Chaining Unlocks

As discussed in the previous analysis, the bikesharing users were classified into different groups, and each of the group likely corresponds to different primary trip purposes. To further investigate the differences between the user groups, we examined the potential impacts of the bikeshare pricing policy on each user group.

The prior literature suggests that bikesharing users may exhibit cost sensitivity; a previous study showed that many annual members tended to return shared bikes just before being charged an additional usage fee *(21)*. Furthermore, there is an anecdotal evidence that some cyclists may make longer trips by "***bike chaining***". We coin the term "bike chaining" to refer to the following phenomenon: a bikeshare user returns a bicycle within 30 minutes and immediately checks out another bicycle to continue a trip; this is likely done to avoid paying additional usage fees when a bicycle is rented for more than 30 minutes. Figure 4 provides an example to further explain this concept. It shows bike unlocks of a Transit bikeshare user on a single day. His first unlock of the day, in red, is classified as "unrelated unlock" because it is not followed by another unlock within a 30-minute period. His second unlock, in blue, is classified as the first unlock of a trip that involved "bike chaining." His third and fourth unlocks, in green and purple, seem to be "bike chaining" unlocks because they are within 30 minutes.

The previous literature and anecdotal evidence inspired the following analysis of "bike chaining" unlocks between user groups, especially the differences between annual members (who may be commuters) and 24-hour pass holders. We designed a trip chaining algorithm to detect this unlocking pattern. This was done for each Transit app bikeshare user (by individual device ID) for each day in the three months in 2016 to explore the number of "bike chaining" unlocks.

The algorithm works as follows. First, for each bikeshare unlock, we check if there is a subsequent unlock on the same day, and if so, the ***time*** at which the second unlock occurs is examined to see if it is within 30 minutes of the first one. Then, the ***station*** where the second unlock occurs is checked to make sure that was at a different location from the first unlock. Then, we considered the ***speed*** of the trip made by each Transit app user and compared it with the average speed of all Divvy users. Speed was evaluated as follows: we calculated the Euclidean distance between the two potential bike chaining unlocks (longitude/latitude) by the Haversine Formula *(32)*. The formulas are shown in equations (4) and (5); the minimum distance d between any two points on a spherical body could be calculated given the latitudes φ_1 and φ_2, the change in longitude, $\Delta\lambda$, and the earth's radius R *(33)*.

$$a = sin^2\left(\frac{\Delta\varphi}{2}\right) + cos\varphi_1 \cdot cos\varphi_2 \cdot sin^2(\frac{\Delta\lambda}{2}) \tag{4}$$

$$d = R \cdot 2 \cdot a \cdot tan2(\sqrt{a}, \sqrt{1-a}) \tag{5}$$

Next, the speed s was calculated by dividing the distance d between two consecutive unlocks by the time between unlocks, $t_2 - t_1$, as shown in equation (6).

$$s = \frac{d}{t_2 - t_1} \tag{6}$$

Finally, the speed was compared to the average speed of all Divvy trips, which was calculated by applying the Harversine Formula *(32)* to system-wide trip data from the bikesharing operator. If the estimated speed was within one standard deviation of the average speed of all Divvy users (from 3.25 to 7.84 mph), then this unlock was classified as a "bike chaining" unlock.

Figure 5 summarizes the algorithm to find "bike chaining" unlocks in a flowchart, and this algorithm was run as a script in the open source program R. This process was repeated for all unlocks on each day in the three months in 2016 for each individual bikeshare user in the Transit app dataset.

The results of the trip chaining algorithm are shown in Table 3. As shown in the last row of the table, 5.7% of all Transit app bikeshare unlocks were classified as "bike chaining" unlocks, and 27.3% of the Transit app bikeshare users (by device ID) used this strategy at least once during the three-month study period to avoid paying additional fees.

The trip chaining algorithm was also applied to each cluster from the previous section. The results are shown numerically in Table 3, and the number of "trip chaining bikers" by unique device ID is visualized for each cluster in pie charts in Figure 6.

The first cluster, commuters, made the most "bike chaining" unlocks. Approximately 66% of commuters tried this strategy at least once in the three-month study period, presumably to avoid paying extra fees. This may be because commuters are familiar with bikesharing and local cycling routes; therefore, it is easy for commuters to lock the bikes within 30 minutes and check out another one *(21)*.

In contrast, approximately 9.6% of unlocks made by weekend visitors were "bike chaining" unlocks, which is the same percentage as commuters. However, only 18.9% of weekend visitors were classified as "trip chaining bikers" (based on unique device IDs). Similarly, weekday visitors had 7.9% "trip chaining unlocks" but only 16.7% were "trip chaining bikers". One reason that the percent of visitors exhibiting this behavior is lower than the percent of commuters may be that many visitors are not familiar with the system. Another reason may be that many tourism hotspots in Chicago such as museums are close to each other in the downtown area, so visitors may not need to ride more than 30 minutes.

For utility users, only 5.2% of the trips were classified as "bike chaining" unlocks, which is the smallest portion when compared to the other groups. One explanation could be that the utility users are the people who cycle during weekdays for errands, such as shopping *(11)*; the destinations of these errands may be easy to find within a short distance, such as convenience stores and pharmacies. In total, 34% of the utility users (by unique device ID) made "bike chaining" unlocks in the three months, which was reasonable because they are probably local residents and may be more familiar with the bikesharing stations. Similarly, leisure users are also likely residents, and 32% of them made "bike chaining" unlocks.

The last user group, known as infrequent commuters, also tried the "bike chaining" strategy; 6.9% of their unlocks were classified as "bike chaining" unlocks.

These results suggest that annual members and 24-hour pass holders may respond differently to Divvy's pricing policy. Although the percentages of "bike chaining" unlocks are all between 5% and 10% in each of the groups, the percentages of "trip chaining bikers" vary substantially between different groups. Specifically, the three groups on the left side of Figure 6 (infrequent commuters, weekday visitors, and weekend visitors) follow the same pattern. These users may be more likely to purchase a 24-hour pass due to their low bikeshare usage, and only 16-19% of them were identified as "trip chaining bikers". However, the three clusters shown on the right side of Figure 6 (commuters, utility users, and leisure users) may be more cost sensitive to bikeshare pricing policies. Over 30% of people in each of these clusters tried "bike chaining" at least once during the study period. It is likely that many bikeshare users in these three groups are annual members of Divvy bikeshare.

CONCLUSIONS AND AREAS FOR FUTURE RESEARCH

This study used a new dataset from a smartphone application called "Transit" to analyze bikesharing user patterns in a three-part analysis of Chicago. First, Transit app bikeshare usage patterns were compared to system-wide bikeshare utilization using publicly available data. The results revealed that hourly usage on weekdays generally follows classical peaked commuting patterns; however, daily usage reached its highest level on a weekend. This suggests that there may be large numbers of both commuting and recreational bikeshare users. The second part identified distinct user groups via cluster analysis; the results show six different clusters: (1) commuters; (2) utility users; (3) leisure users; (4) infrequent commuters, (5) weekday visitors; and (6) weekend visitors. The group unlocking the most shared bikes (45.58% of all Transit app unlocks) was commuters, who represented 10% of Transit app bikeshare users. The third part proposed a trip chaining algorithm to identify "***trip chaining bikers***." This term was introduced to refer to bikeshare users who returned a shared bicycle and immediately checked out another bicycle, presumably to avoid paying extra usage fees for trips over 30 minutes. The results of the algorithm revealed that 27.3% of Transit app bikeshare users appear to have made "***bike chaining"*** unlocks to avoid paying additional usage fees. However, this varied substantially between user groups; notably, 66% of Transit app bikeshare users identified as commuters made one or more "bike chaining" unlocks.

The implications of this research are important for bikeshare providers to understand the impact of their pricing policies. In Chicago, the pricing policy has recently changed and now allows for longer trips before additional fees are charged. A key reason for this change was to simplify the pricing policy from the users' perspective. Anecdotal evidence suggests that many bikeshare users would get confused by the additional usage fees for trips, and this may have discouraged potential new users from trying the system. To lower barriers to use and ease customer understanding, the pricing policy was simplified in numerous ways, including extending the 30-minute timeframe before usage fees were charged. In the new pricing policy, annual members pay additional usage fees for trips over 45 minutes and 24-hour (explore) pass holders have 180 minutes, which still encourages turn-over of shared bicycles but gives users additional time to complete longer trips.

There are many areas for improvement and future research that have emerged from this analysis. First, the cluster analysis only considered temporal variables; future research could consider the spatial data associated with the location of bikeshare unlocks. Second, the trip chaining algorithm used the Haversine Formula to approximate "as the crow flies" distance of bikeshare trips. Distance travelled via the roadway network would likely be a more accurate measure, and this has been left for future research. Third, the datasets used in this paper are only three months in duration; however, future research should consider longer time periods (e.g., one year), which could help to more easily identify some user groups such as visitors to Chicago. Fourth, it would be interesting to apply the trip chaining algorithm to newer data since the pricing policy was changed in 2018 to accommodate longer trips before additional fees are charged. Last, the Transit app dataset only includes bikeshare unlocks (i.e., trip origins) but does not contain data about bikeshare locks (e.g., trip destinations). Moreover, the Transit app data represents a subset of all bikeshare users in Chicago. Therefore, if Divvy data are made available with a unique identifier for individual users, future research could consider modifying and applying the clustering and trip chaining analyses to the larger, system-wide dataset that includes both bikeshare unlocks and locks.

ACKNOWLEDGEMENTS

The authors thank the Transit app for sharing data. This research was sponsored in part by the National Science Foundation (NSF) through a Research Experience for Undergraduates (REU) award and the Extreme Science and Engineering Discovery Environment, with additional support from the Joint Institute of Computational Sciences at University of Tennessee, Knoxville. The authors are grateful to Dr. Kwai Wong and Dr. Cheng Liu for their guidance through the NSF REU program. This research was also supported in part by a 2015 City University of New York Collaborative Incentive Research Grant and a 2016 University Transportation Research Center faculty-initiated grant. The authors would like to acknowledgement contributions made to an early draft of this paper by former City College of New York student Niloofar Ghahramani, and they would like to thank Professor Jonathan Peters at the College of Staten Island for reviewing an early draft of this paper.

AUTHOR CONTRIBUTIONS

The author confirm contribution to the paper as follows: study conception and design: Yang, Nicolas, Brakewood; data collection: Yang, Nicolas, Sion; analysis and interpretation of results: Yang, Nicolas, Brakewood; draft manuscript preparation: Yang, Brakewood, Nicolas, Sion. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. National Association of City Transportation Officials. *Bike Share in the U.S.: 2017.* https://nacto.org/bike-share-statistics-2017/. Accessed July 31, 2018.
2. Zheng, Y., L. Capra, O. Wolfson, and H. Yang. Urban Computing: Concepts, Methodologies, and Applications. *ACM Transaction on Intelligent Systems and Technology,* 2014. 5(3): 38:1-38:55.
3. Shaheen, S. A., E. W. Martin, and A. P. Cohen. Public Bikesharing and Modal Shift Behavior: A Comparative Study of Early Bikesharing Systems in North America. *International Journal of Transportation*, 2013. 1(1): 33-54.
4. Jäppinen, S., T. Toivonen, and M. Salonen. Modelling the Potential Effect of Shared Bicycles on Public Transport Travel Times in Greater Helsinki: An Open Data Approach. *Applied Geography*, 2013. 43: 13-24.
5. Bargar, A., A. Cupta, S. Gupta, and D. Ma, Interactive Visual Analytics for Multi-City Bikeshare Data Analysis. *The 3rd International Workshop on Urban Computing*, 2014.
6. Guell, C., J. Panter, N. R. Jones, and D. Ogilvie. Towards a Differentiated Understanding of Active Travel Behaviour: Using Social Theory to Explore Everyday Commuting. *Social Science & Medicine*, 2012. 75(1): 233-239.
7. Goodman, A., and J. Cheshire. Inequalities in the London Bicycle Sharing System Revisited: Impacts of Extending the Scheme to Poorer Areas but then Doubling Prices. *Journal of Transportation Geography*, 2014. 41: 272-279.
8. Chang, X., J. Shen, X. Lu, and S. Huang. Statistical Patterns of Human Mobility in Emerging Bicycle Sharing Systems. *PLoS ONE*, 2018. 13(3): e0193795. https://doi.org/10.1371/journal.pone.0193795.
9. Shaheen, S. A., H. Zhang, E. Martin, and S. Guzman. China's Hangzhou Public Bicycle: Understanding Early Adoption and Behavioral Response to Bikesharing. *Transportation Research Record: Journal of the Transportation Research Board*, 2011. 2247: 33-41. https://doi.org/10.3141/2247-05.
10. Martin, E. W., and S. A. Shaheen. Evaluating Public Transit Modal Shift Dynamics in Response to Bikesharing: A Tale of Two U.S. Cities. *Journal of Transportation Geography*, 2014. 41: 315-324.
11. O'Brien, O., J. Cheshire, and M. Batty. Mining Bicycle Sharing Data for Generating Insights into Sustainable Transport Systems. *Journal of Transportation Geography*, 2014. 34: 262-273.
12. DNAinfo. *Divvy Late Fees Top $4 Million As More Than A Million Rides Go Too Long*. May 3, 2017. https://www.dnainfo.com/chicago/20170503/wicker-park/divvy-bike-share-four-years-21-million-trips-how-many-trips-over-30-minutes-overage-fees/. Accessed July 31, 2018.
13. Fishman, E. Bikeshare: A Review of Recent Literature. *Transport Reviews*, 2016. 36(1): 92-113.
14. Froehlich, J., J. Neumann, and N. Oliver. Sensing and Predicting the Pulse of the city through Shared Bicycling. *Proceedings of the 21st international joint conference on Artificial intelligence*, 2009.
15. Lathia, N., S. Ahmed, and L. Capra. Measuring the Impact of Opening the London Shared Bicycle Scheme to Casual Users. *Transportation Research Part C: Emerging Technologies*, 2012. 22: 88-102.
16. Côme, E., and L. Oukhellou. Model-based Count Series Clustering for Bike-sharing System

Usage Mining, a Case Study with the Velib System of Paris. *ACM Transactions on Intelligent Systems and Technology*, 2014. 5(3): 39:1- 39:21.
17. Vogel, P., T. Greiser, and D. C. Mattfeld. Understanding Bike-sharing Systems using Data Mining: Exploring Activity Patterns. *Procedia – Social and Behavioral Sciences*, 2011. 20: 514-523.
18. Austwick, M. Z., O. O'Brien, E. Strano, and M. Viana.c. *PLoS ONE*, 2013. 8(9): e74685. https://doi.org/10.1371/journal.pone.0074685.
19. Zhou, X. Understanding Spatiotemporal Patterns of Biking Behavior by Analyzing Massive Bike Sharing Data in Chicago. *PLoS ONE*, 2015. 10(10): e0137922. https://doi.org/10.1371/journal.pone.0137922.
20. Schuijbroek, J., R. Hampshire, and W. V. Hoeve. Inventory Rebalancing and Vehicle Routing in Bike Sharing Systems. *European Journal of Operational Research*, 2017. 257(3): 992-1004.
21. Jurdak, R. The Impact of Cost and Network Topology on Urban Mobility: A Study of Public Bicycle Usage in 2 U.S. Cities. *PLoS ONE*, 2013. 8(11): e79396. https://doi.org/10.1371/journal.pone.0079396.
22. Divvy, https://www.divvybikes.com/. Accessed July 31, 2018.
23. Curbed Chicago. *Divvy will offer $3 single-ride passes and more*. Jan 25, 2018. https://chicago.curbed.com/2018/1/25/16929478/divvy-bike-new-pass-price-options. Accessed July 31, 2018.
24. Chicago Tribune. *Divvy to Lower Daily Prices and Add Longer-ride Option.* Jan 24, 2018. http://www.chicagotribune.com/news/local/breaking/ct-met-divvy-prices-20180124-story.html. Accessed July 31, 2018.
25. Transit, http://transitapp.com/. Accessed July 31, 2018.
26. Singhal, A., C. Kamga, and A. Yazici. Impact of Weather on Urban Transit Ridership. *Transportation Research Part A: Policy and Practice*, 2014. 69: 379-391.
27. Webster, K., and C. Cunningham. Preparing for Bike-Sharing: Insight from Focus Groups and Surveys, Chattanooga, Tennessee. *Health Promotion Practice*, 2013. 14: 62-68.
28. Hartigan, J. A., and M. A. Wong. A K-means Clustering Algorithm. *Applied Statistics*, 1979. 28: 100-108.
29. Ngai, E., L. Xiu, and D. Chau. Application of Data Mining Techniques in Customer Relationship Management. *Expert Systems with Applications*, 2009. 36: 2592–2602.
30. Rousseeuw, P. J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, 1987. 20: 53-65.
31. Morency, C., M. Trepanier, and F. Godefroy. Insights into Montreal's Bikesharing System. *Transportation Research Board 90th Annual Meeting*, 2011.
32. Conservation Ecology, *Distance between two latitude-longitude points in R.* June 30, 2013. https://conservationecology.wordpress.com/2013/06/30/distance-between-two-points-in-r/. Accessed July 31, 2018.
33. Chopde, N., and M. Nichat. Landmark Based Shortest Path Detection by Using A* and Haversine Formula. *International Journal of Innovative Research in Computer and Communication Engineering*, 2013. 1: 298-302.

FIGURES AND TABLES

FIGURE 1 Transit App Screenshots

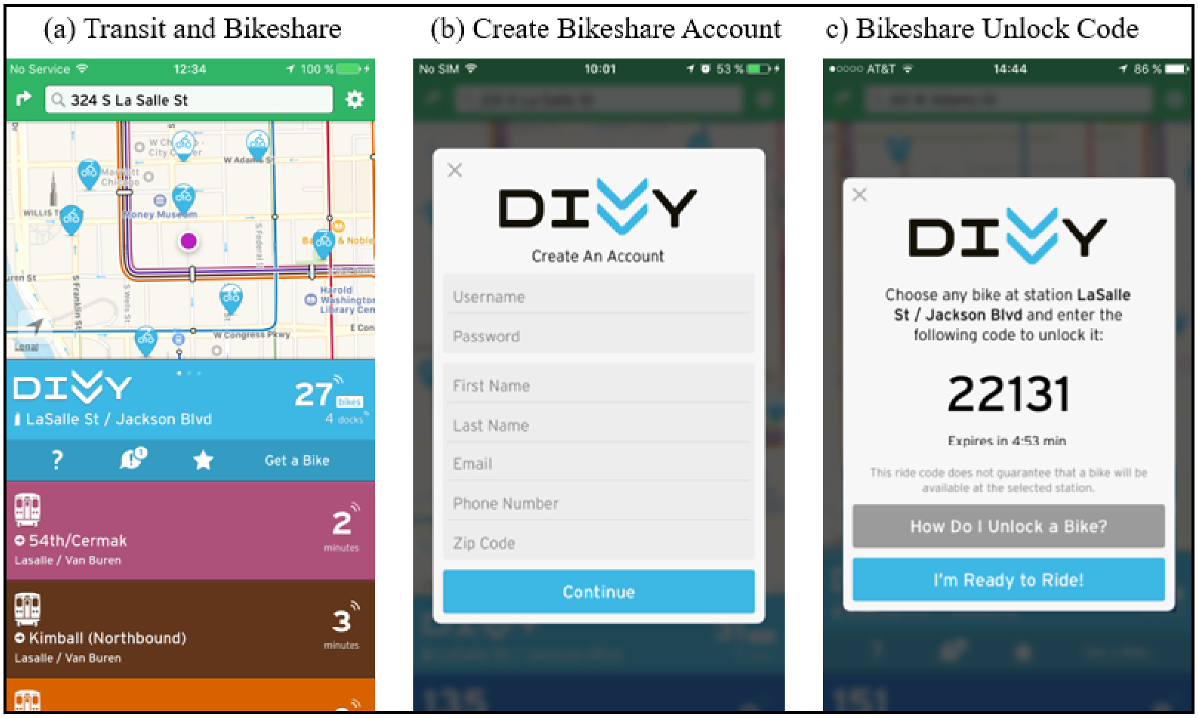

FIGURE 2 Results of Temporal Pattern Analyses

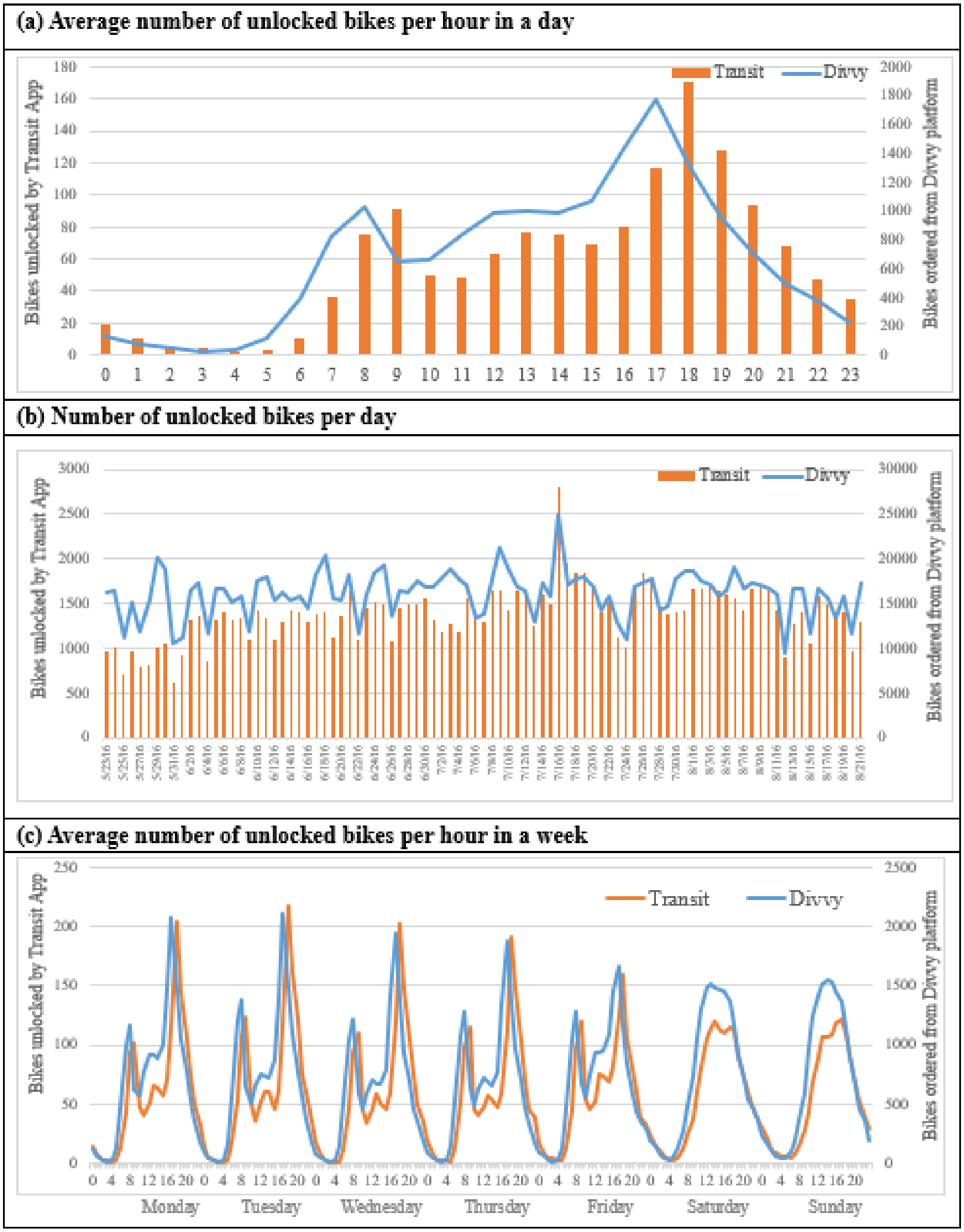

FIGURE 3 Characteristics of Each Group of Users

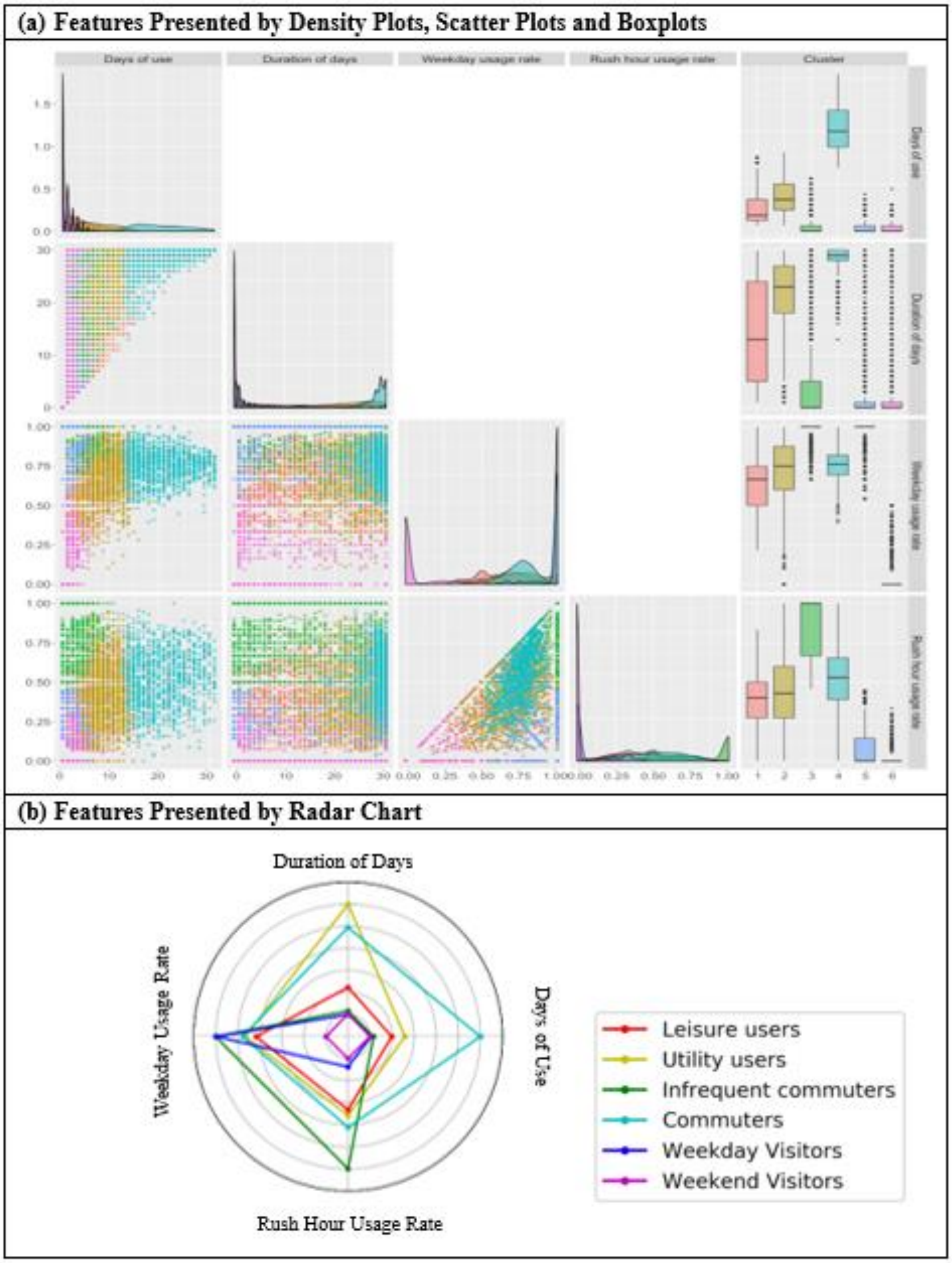

FIGURE 4 Individual User with Bike Chaining Unlocks

FIGURE 5 Trip Chaining Algorithm Identifying Bike Chaining Unlocks

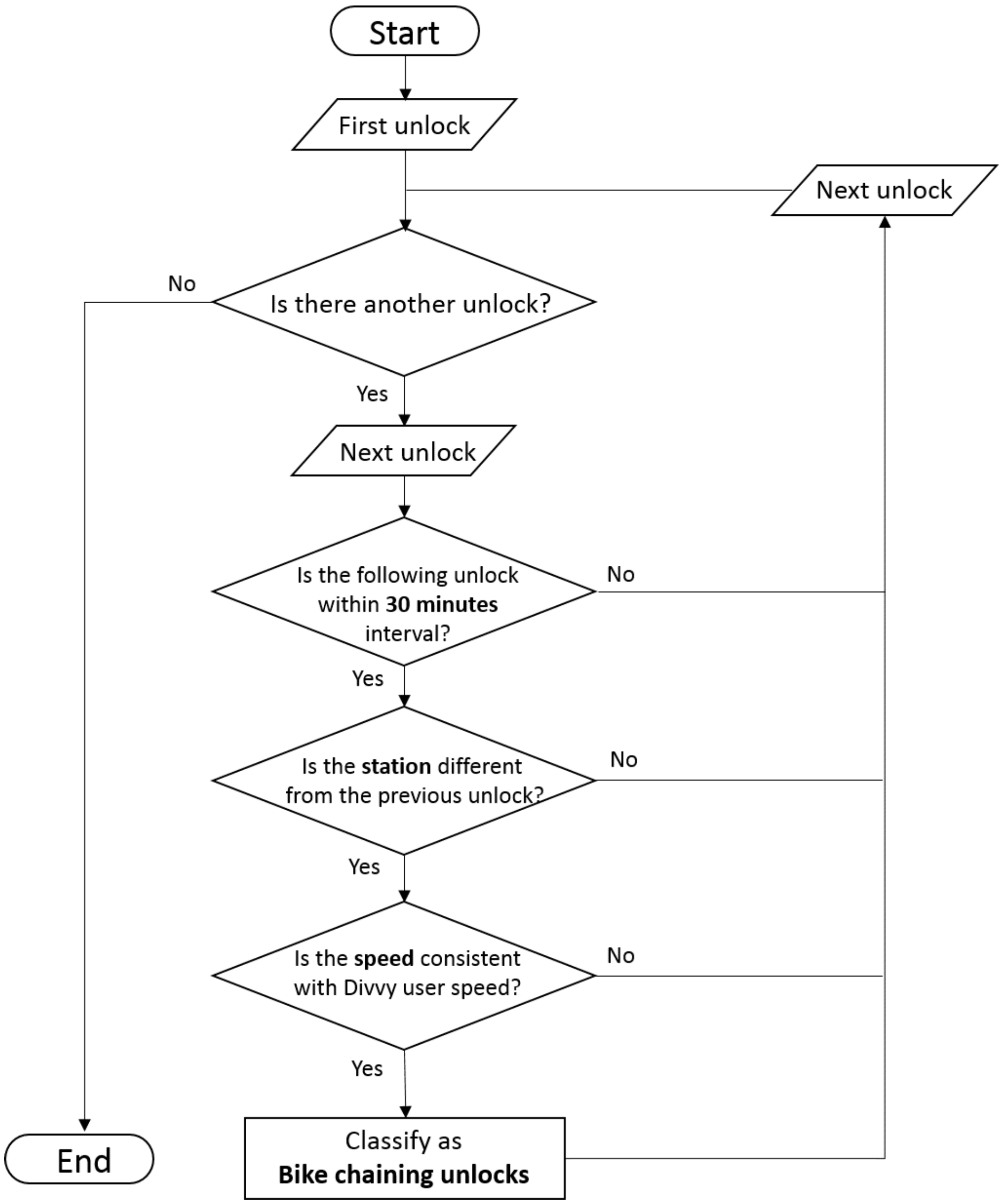

FIGURE 6 Percent of Trip Chaining Bikers per Cluster

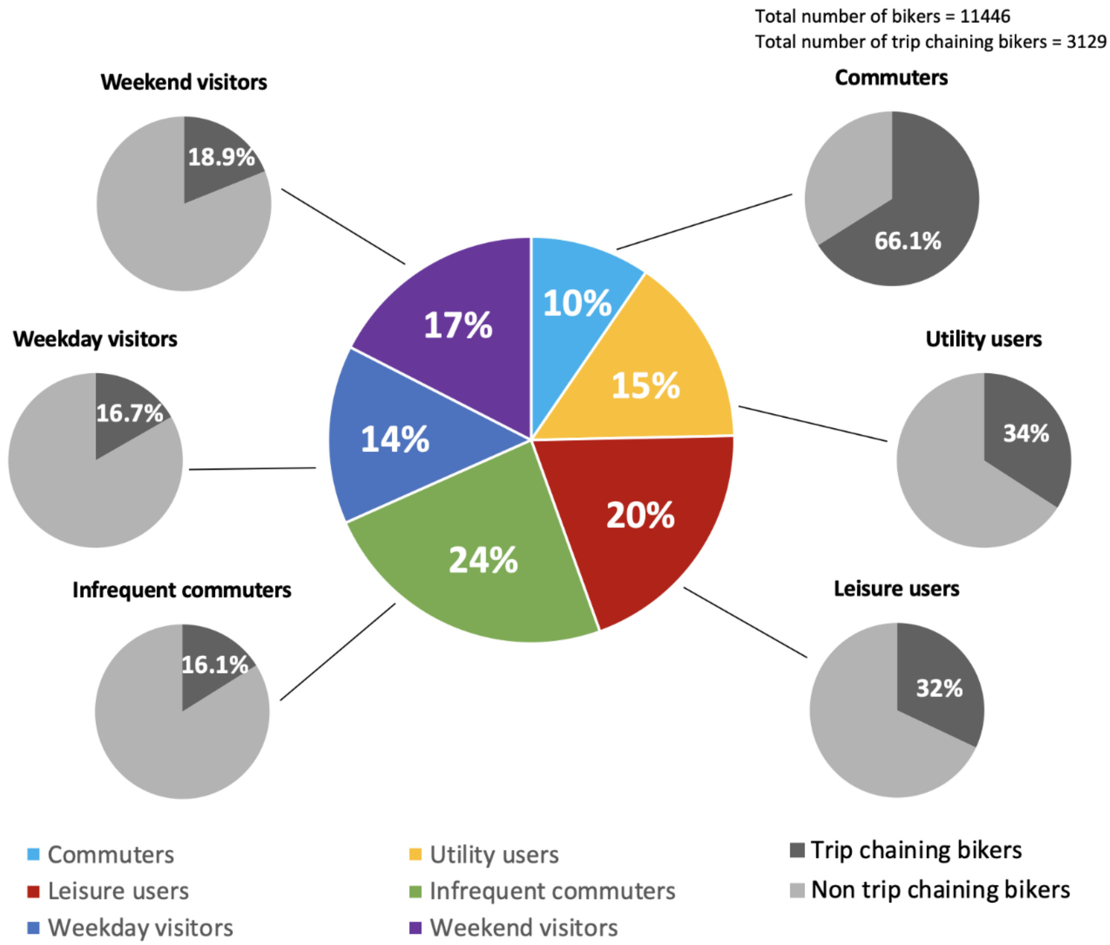

TABLE 1 Pricing Structure for the Divvy Bikesharing System

<table>
<tr><th></th><th colspan="2">2016</th><th colspan="3">2018</th></tr>
<tr><th></th><th>Annual Membership</th><th>24-Hour Pass</th><th>Annual Membership</th><th>Explore (24-Hour) Pass</th><th>Single Ride</th></tr>
<tr><td>Fees</td><td>$99</td><td>$9.95</td><td>$99</td><td>$15</td><td>$3</td></tr>
<tr><td>0-30 minutes</td><td colspan="2">$0</td><td rowspan="2">$0</td><td rowspan="5">$0</td><td>$0</td></tr>
<tr><td>31-45 minutes</td><td rowspan="2">$1.50</td><td rowspan="2">$2</td><td rowspan="5">+ $3 for each additional 30 minutes</td></tr>
<tr><td>45-60 minutes</td><td rowspan="4">+ $3 for each additional 30 minutes</td></tr>
<tr><td>61-90 minutes</td><td>$4.50</td><td>$6</td></tr>
<tr><td>91-180 minutes</td><td rowspan="2">+ $6 for each additional 30 minutes</td><td rowspan="2">+ $8 for each additional 30 minutes</td></tr>
<tr><td>180+ minutes</td><td>+ $3 for each additional 30 minutes</td></tr>
</table>

TABLE 2 Results of Classification of Transit App Users and their Bikesharing Unlocks

Category	**Number of users (% of users)**	**Number of unlocks (% of unlocks)**	**Mean days of use**	**Mean duration of days**	**Mean weekday usage rate**	**Mean rush hour usage rate**
Commuters	1094 (10%)	57,232 (45.58%)	20.73	44.81	0.75	0.52
Utility users	1729 (15%)	22,737 (18.1%)	7.51	56.20	0.73	0.44
Leisure users	2271 (20%)	23,749 (18.9%)	5.19	15.47	0.64	0.39
Infrequent commuters	2727 (24%)	10,267 (8.18%)	2.01	4.06	0.98	0.83
Weekday visitors	1627 (14%)	5107 (4.06%)	1.54	2.23	0.97	0.07
Weekend visitors	1998 (17%)	6478 (5.16%)	1.41	2.47	0.06	0.01
Total	**11,446 (100%)**	**125,570 (100%)**	**5.09**	**17.56**	**0.69**	**0.40**

TABLE 3 Number of Bike Chaining Unlocks and Trip Chaining Bikers by Category

Category	**Bike chaining unlocks (% of bike chaining unlocks in each category)**	**Trip chaining bikers (% of trip chaining bikers in each category)**
Commuters	2,627 (9.6%)	723 (66.1%)
Utility users	1,176 (5.2%)	590 (34%)
Leisure users	1,588 (6.7%)	727 (32%)
Infrequent commuters	711 (6.9%)	439 (16.1%)
Weekday visitors	402 (7.9%)	272 (16.7%)
Weekend visitors	602 (9.6%)	378 (18.9%)
Total	**7,106 (5.7%)**	**3,129 (27.3%)**