# Unveiling Urban Mobility Dynamics: Insights into Ride Durations, Usage Patterns, and Spatial Trends of Divvy

*DATA100 - S12*
*Group 1: Ardales, Pe, Siababa, Yao, Yu*

## I. Introduction

Divvy is a company that provides sharable bike service in the city of Chicago. To avail of its service, the user just needs to download the Divvy application and scan the QR code on the bike to unlock it. Once done, the bike should be returned to any of the Divvy docks. Aside from traditional bikes, Divvy also rents electric bikes and electric scooters. It has four different plans targeting different groups of customers with different prices (Fig.1).

| | **Single Ride**<br>**$1 + $0.17/min**<br>**Get the app →** | **Day Pass**<br>**$16.50/day**<br>**Get a day pass →** | **Divvy**<br>**$130.90/year**<br>**Join →** | **Lyft Pink**<br>**$199/year**<br>**Join →** |
|---|---|---|---|---|
| **Classic bike prices** | $1 unlock + $0.17/min | 3 hours free, then $0.17/min | 45 min free, then $0.17/min | 45 min free, then $0.17/min |
| **Scooter prices** | $1 unlock + $0.42/min | Free unlocks + $0.42/min | Free unlocks + $0.27/min | Free unlocks + $0.27/min |
| **Ebike prices** | $1 unlock + $0.42/min | Free unlocks + $0.42/min | Free unlocks + $0.17/min | Free unlocks + $0.17/min |
| **Bike Angels** | | | ● | ● |
| **Rideshare benefits** | | | | ● |

Figure 1. Per minute rate of different types of transportation from Divvy website

## II. Exploratory Data Analysis

The data this report is working on is limited to the July of 2023. The attributes include the ID for each ride, which is not correlated to any specific account, the type of transportation used, the time and coordinates when and where the ride started and ended, the ID and name of the station where the transportation is at, and the type of user for that ride.

```
(767650, 13)
ride_id                    0
rideable_type              0
started_at                 0
ended_at                   0
start_station_name    122943
start_station_id      122943
end_station_name      130304
end_station_id        130304
start_lat                  0
start_lng                  0
end_lat                 1254
end_lng                 1254
member_casual              0
dtype: int64
```

Figure 2. Shape and null values of the original data

At first look at Figure 2, we could see that it contains 767,650 observations with 13 features. 6 columns contain null values:

start_station_name;
start_station_id;
end_station_name;
end_station_id; end_lat;
and end_lng.

We start preparing the data by handling the missing values. Since we are given the latitude and longitude, we can impute missing station names and ids based on the other data. First, a dictionary is created for each latitude and longitude to the station name and id. Afterwards, we mapped the missing data using those dictionaries.

Next, we converted that data type for the started_at and ended_at columns to datetime so we can better extract specific date or time data such as hour. Two new variables are created, called the peak_hours_start and peak_hours_end. They include the frequency of the ride starting and ending at a certain hour of the day. Based on that, a bar graph is made to visualize the peak hours when the Divvy bikes are borrowed and returned the most (Fig.3).

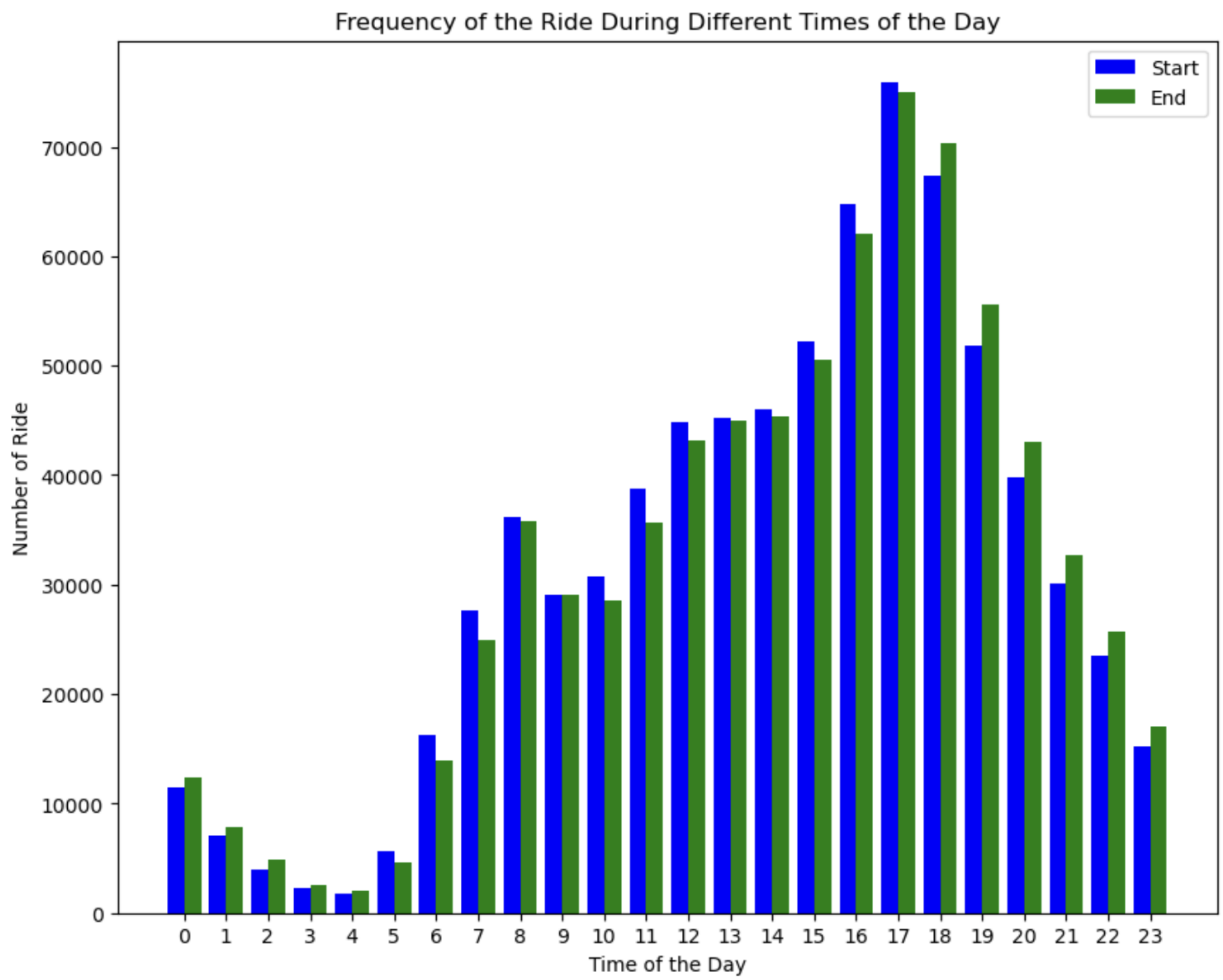


Figure 3. Sum of the number of rides started and ended during each hour of the day

## III. Analysis and Results

Exploration of the dataset revealed intriguing facets warranting deeper examination. This section delves into the ride duration, associated costs, station-to-station distances, and identification of distinct city areas based on station usage patterns. The analysis includes insights into ride duration, revenue distribution, bike docking patterns, and weekday ridership influences. Additionally, a K-means clustering analysis will uncover distinct areas within the city, providing insights into spatial usage patterns and urban mobility trends.

### Visualizing Ride Durations: Insights from Heatmap Analysis

The span of each ride was computed by directly subtracting the starting time from the ending time and represented in minutes. Noticing the wide distribution and large difference between the frequency of the time used (Fig.4), we decided to visualize the data with heatmaps instead of histograms.

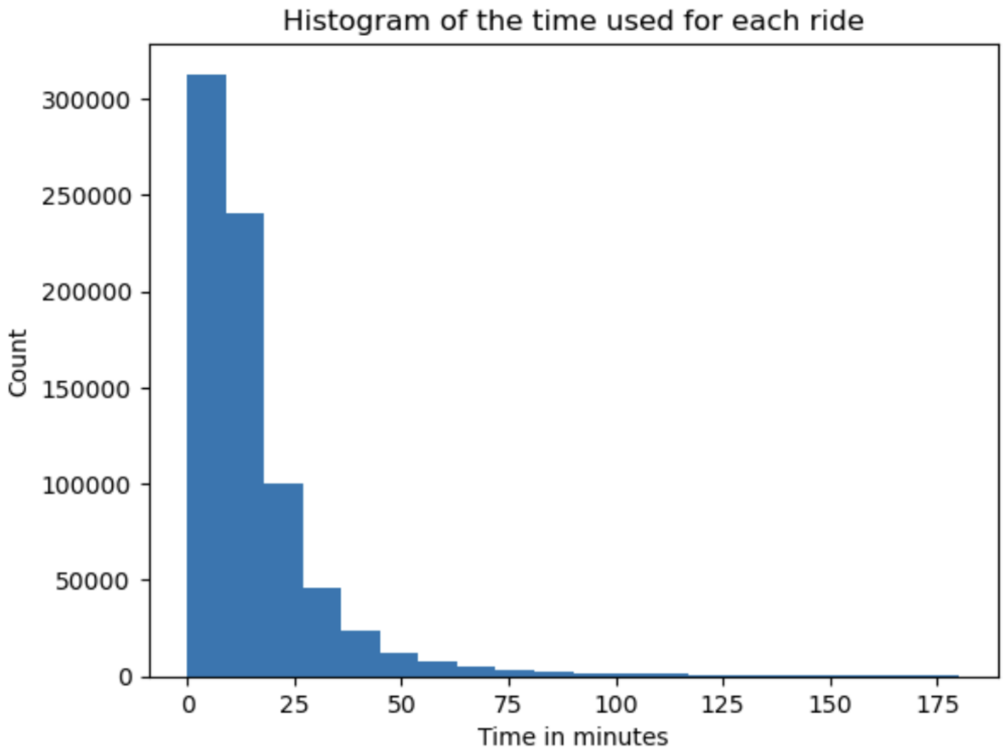


Figure 4. Histogram of the amount of time used for each ride ranging from 0 to 180 minutes

The time range was categorized into 0 to 10 minutes, 10 to 20 minutes, 20 to 45 minutes, 45 minutes to 3 hours, 3 hours to 1 day, and beyond one day. The first two time ranges are selected for us to better understand the urgency of the people using the service, plus, it is also due to the high frequency of those categories. The 45-minute and 3-hour marks are selected due to the specifications of the pricing system of Divvy. Lastly, the 1-day mark is to separate the population that actually needs to use the bike for a long period of time and the possible outliers.

There are three rideable types from the original data: classic bike, electric bike, and docked bike. Interestingly, the docked bike was found to be the old name for the classic bike. Therefore, all the docked bike type are replaced by the classic bike type for th best data visualization and analysis.

Through the use of heat maps, it was found that a disproportionately large amount of revenue was generated from a small set of rides (Fig.5 & Fig.6). In particular, classic bike rides lasting beyond a day made up the plurality of revenue with the longest ride lasting over a month. This is indicative of an issue in which bikes, either accidentally or by choice, are not returned to a docking station, and thus, large bills are charged to the borrowing account. This may point to the need for systems reminding users to return bicycles.


Figure 5. Heatmap displaying the number of rides for different user type and bike type given the different amount of time used

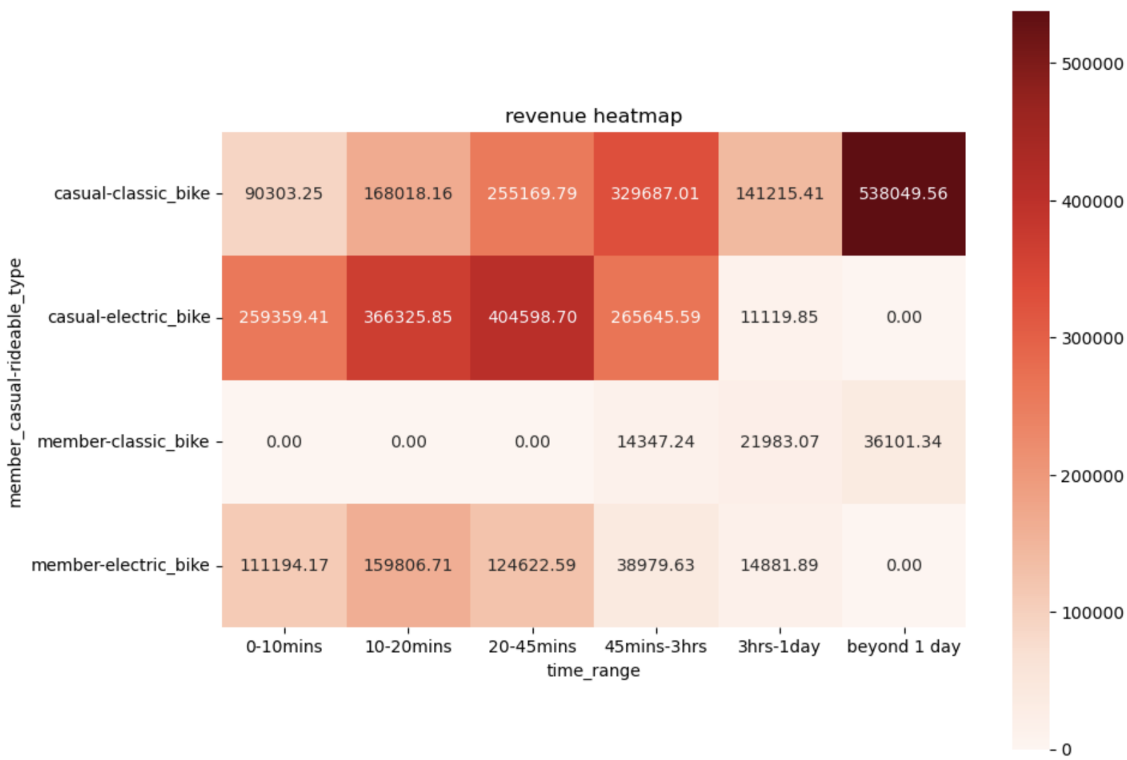


Figure 6. Heatmap displaying the sum of the cost of the rides for different user type and bike type given the different amount of time used

**Deciphering Usage Patterns: Short-Duration Rides and Revenue Trends**

For all further analysis, rides lasting beyond 1 day were omitted from the data set in order to provide a more accurate and reflective analysis of the data. With these values omitted, it becomes more evident that the majority of riders borrow bikes for a short period of time between 0-20 minutes (Fig.7). Overall, electric bikes are more commonly used than their classic counterparts amongst both casual users and members of the service. Additionally, revenue directly generated from rides tends to come from casual users of Divvy (Fig.8), though it is important to note that these heatmaps lack information regarding revenue generated from membership payments.

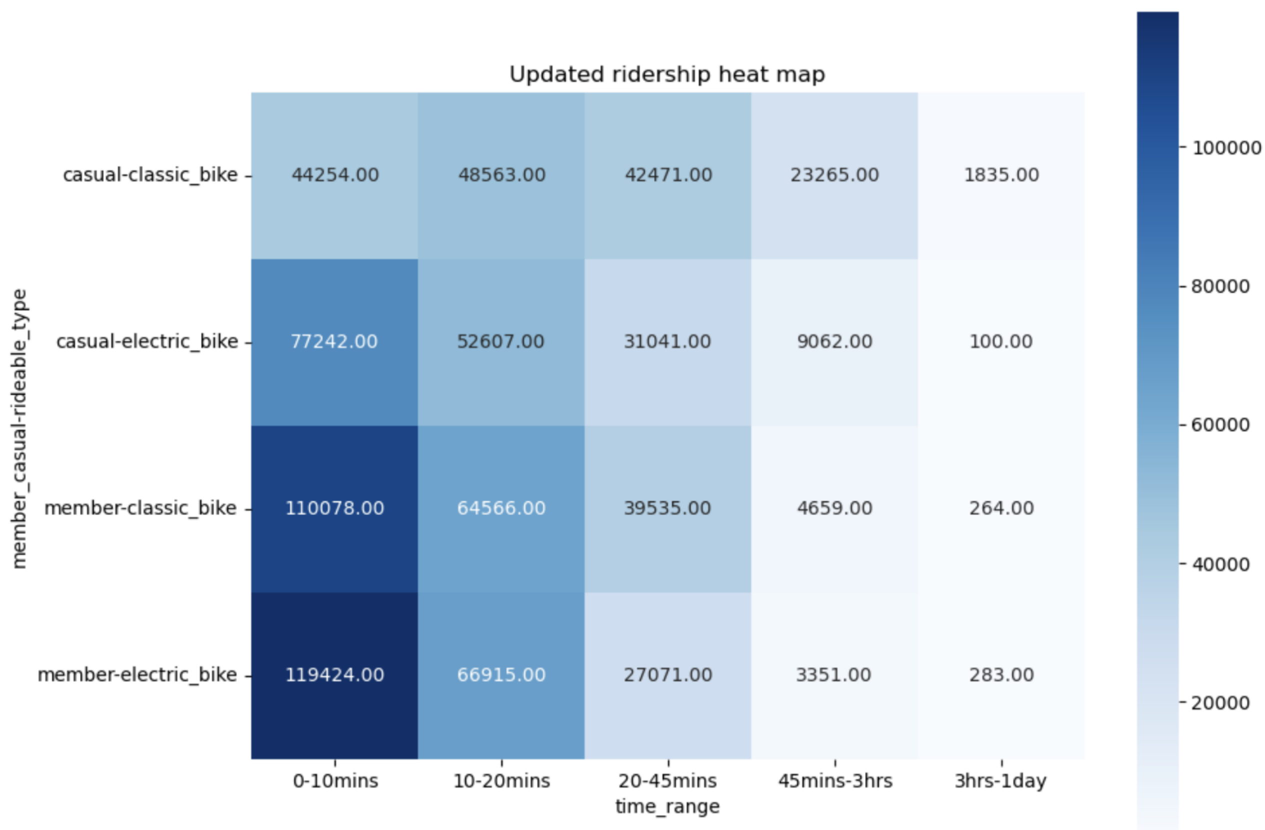


Figure 7. Updated heatmap displaying the number of rides for different user types and bike types given the different amounts of time used that excludes the ride taking more than 1 day

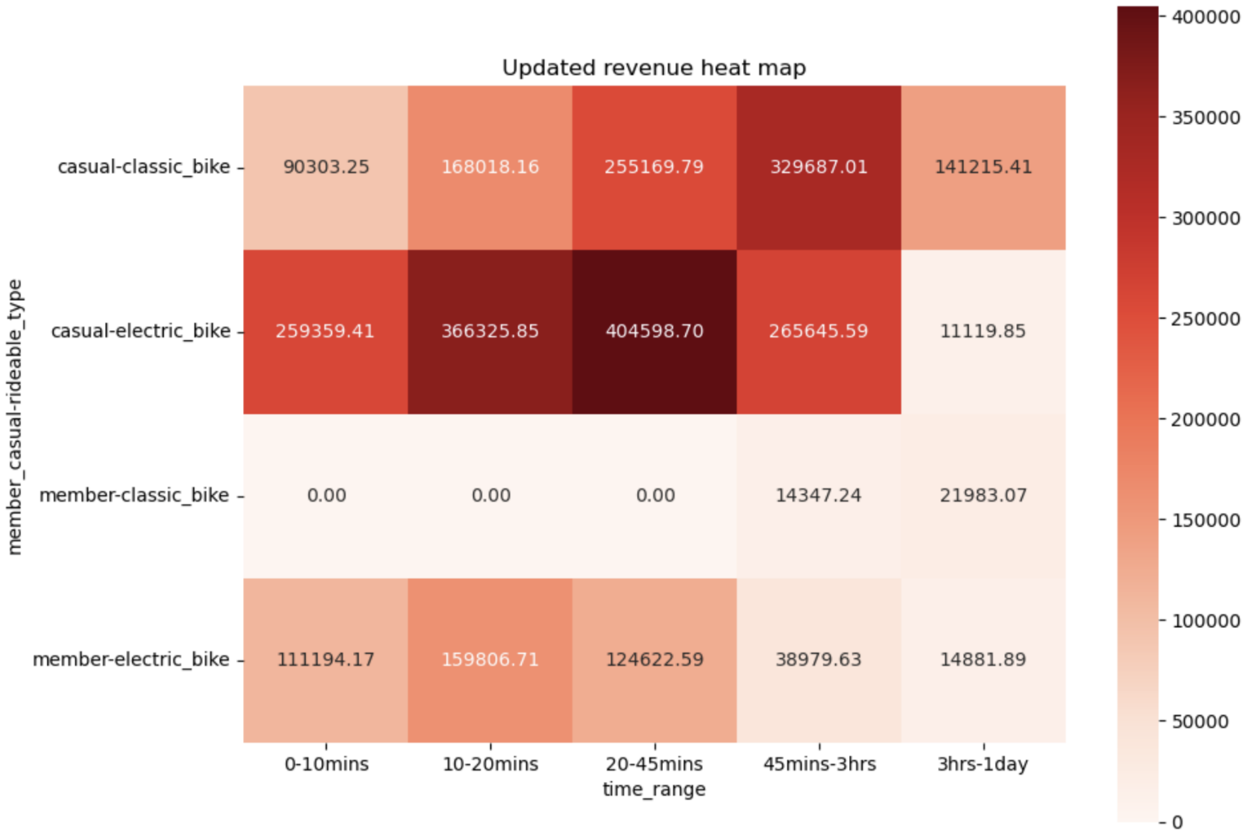


Figure 8. Updated heatmap displaying the sum of the cost of the rides for different user types and bike types given the different amount of time used that excludes the ride taking more than 1 day

**Urban Mobility Insights: Interstation Distance Analysis**

The distance between the starting and ending stations of a ride was calculated using Vicenty's formula for ellipsoids implemented in the geopy library. Despite taking a longer time to compute, this implementation was chosen over the great circle distance method using the haversine formula due to its greater accuracy. These were then turned into a series of heat maps that assessed the interstation distance as it relates to the member and type of bike (Fig.9). From the interstation distance, box plots were generated (Fig.11). It was found that the most common interstation distance traveled was within the 1-2 km range although more generally, rides between 0-5 km are all quite common. An important thing to note is that the usage of electric bikes is more common with longer distances as compared to classic bikes. This makes sense as it is likely that riders would prefer electric bikes for the more physically demanding longer distances.

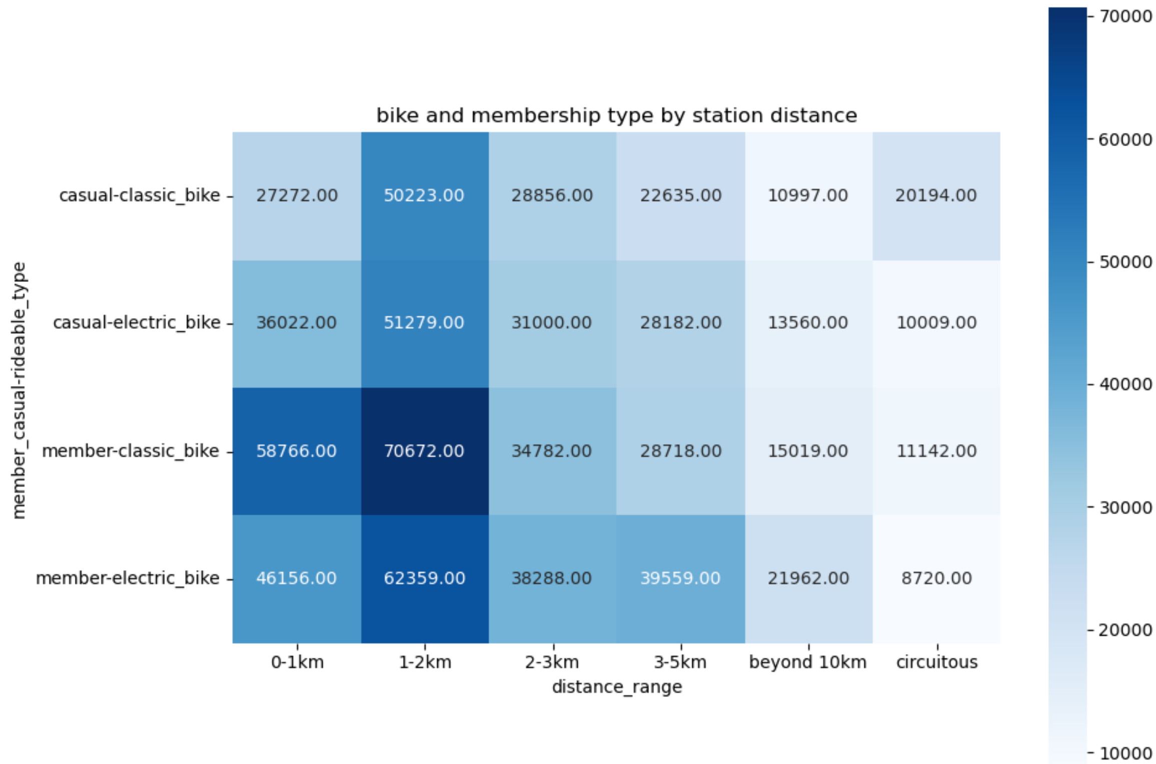


Figure 9. Heatmap displaying the number of rides for different user types and bike types given the distance between the starting and ending stations

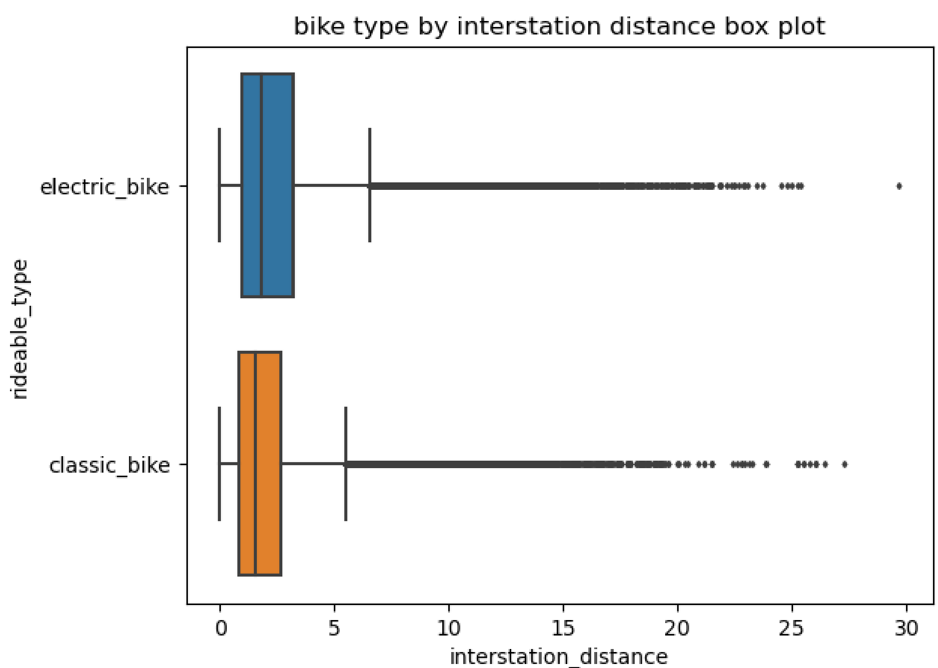


Figure 10. Box plot for the comparison between the interstation distance traveled by the riders using electric bikes and classic bikes

As seen in the box plot (Fig.10), it was found that electric bikes tend to be docked at stations further away from the starting station when compared to classic bikes. On average, both types of bikes are used to travel to stations between 2-3 kilometers from their starting station. This also suggests that the majority of rides are short in duration as at an average biking speed, this distance can be covered in approximately 10 minutes.

**Weekday and Weekend Ridership: Impact on Usage Trends**

The day of the week has also shown an influence on the number of riders using the service. This is visualized using both a bar chart (Fig.11) and heat maps (Fig.12 & 13). It can be seen that the most activity was had on Saturday. Besides this, a relatively steady level of ridership is maintained throughout the week with a dip on Wednesdays.

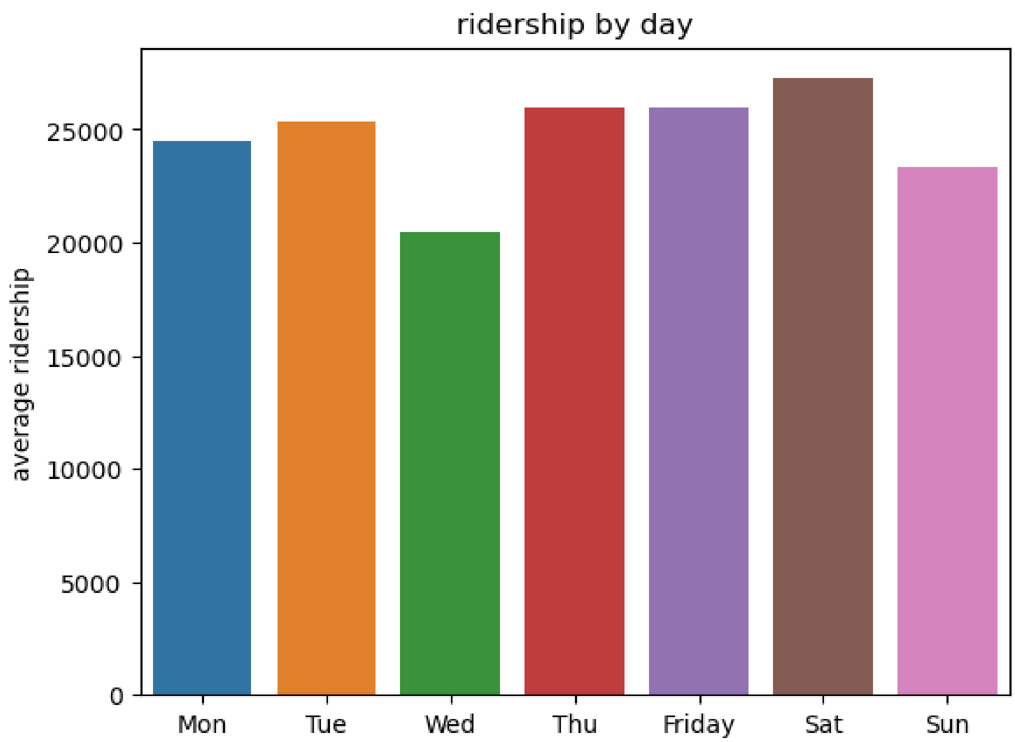


Figure 11. The number of rides during different days of the week is represented in a bar graph

For a more in-depth look at the day of the week's influence, we turn to the heat maps. These heat maps reveal that ridership on the weekends displays a higher frequency of longer-duration rides. This may be interpreted as people riding for leisure or to more distant locations rather than shorter routes on the weekdays. Additionally, circuitous rides are also seen to increase during the weekends which may be interpreted as riders taking short journeys to local businesses like a local grocery or cafe before returning home. The shorter duration rides of the weekdays may be indicative of a rider's short and direct commute to work.

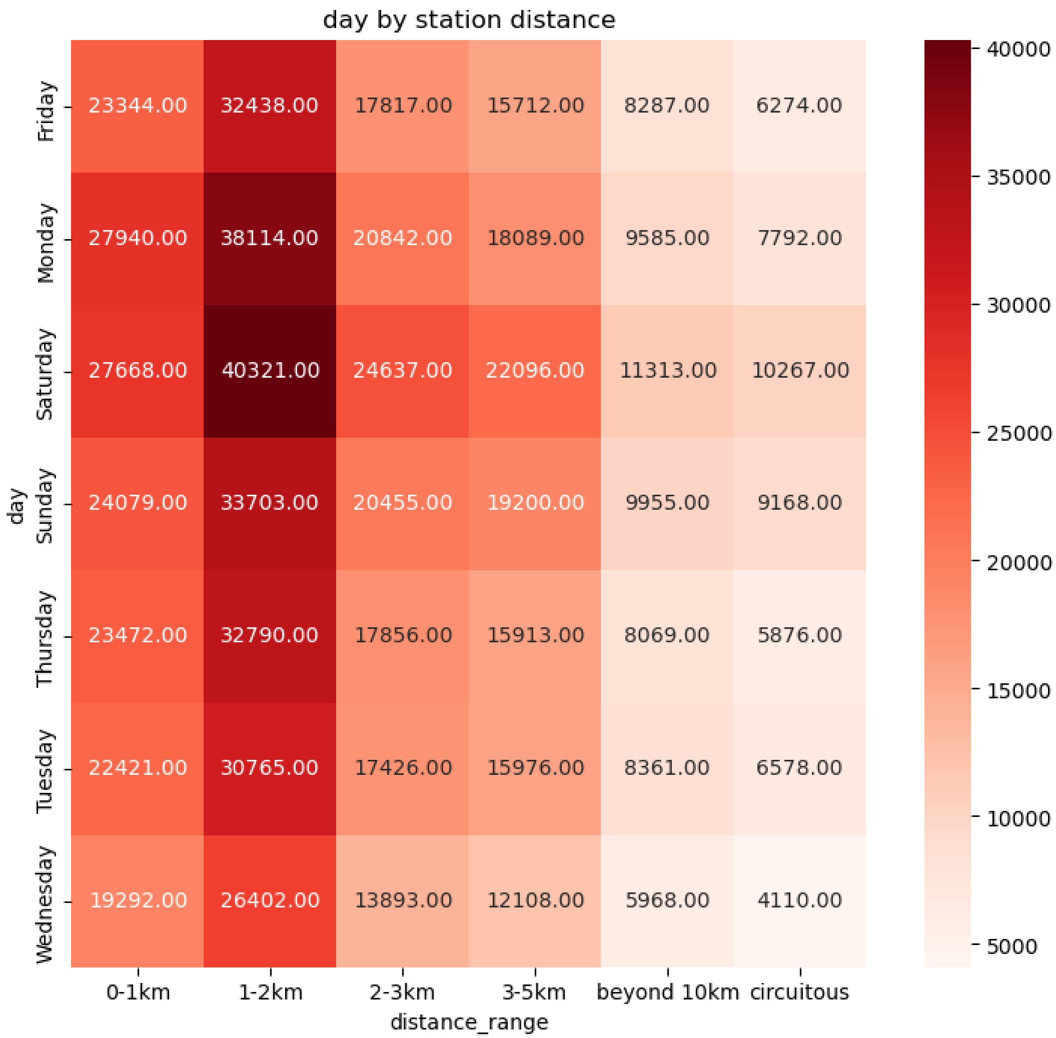


Figure 12. Heatmap displaying the number of rides on different days of the week given the distance between the starting and ending stations

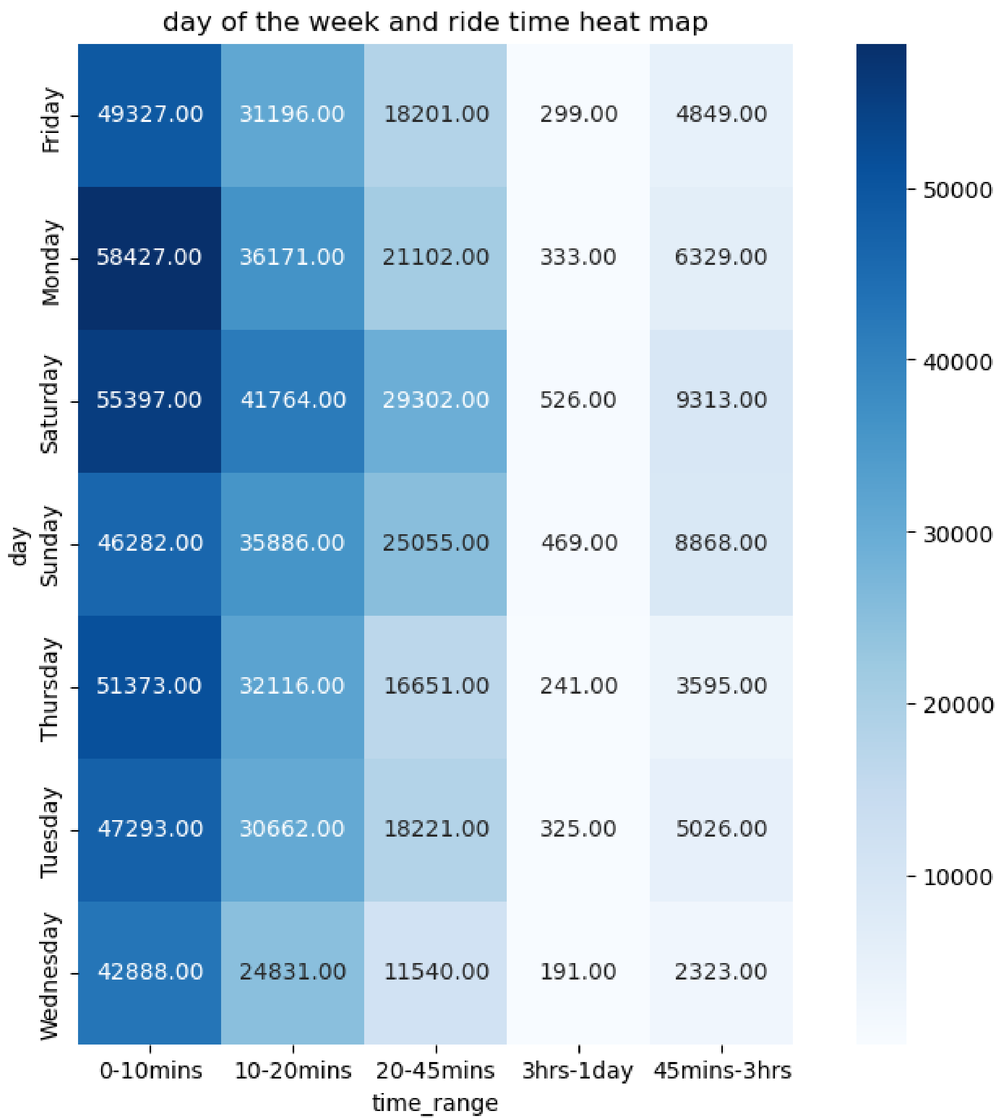


Figure 13. Heatmap that shows the number of rides on different days of the week given the time used

**Spatial Mapping: Defining Urban Zones through K-Means Analysis**

Since the city is relatively large and has numerous stations, we can further divide the city into distinct clusters or regions based on the starting station coordinates. To do this, the team utilized the usage of spatial mapping through the K-Means algorithm. In this analysis, regions can be classified into having high, medium, and low ridership.

As a first step towards implementing the K-Means analysis, the optimal value for K will be identified through the use of an easy and heuristic approach which is the Elbow Method.

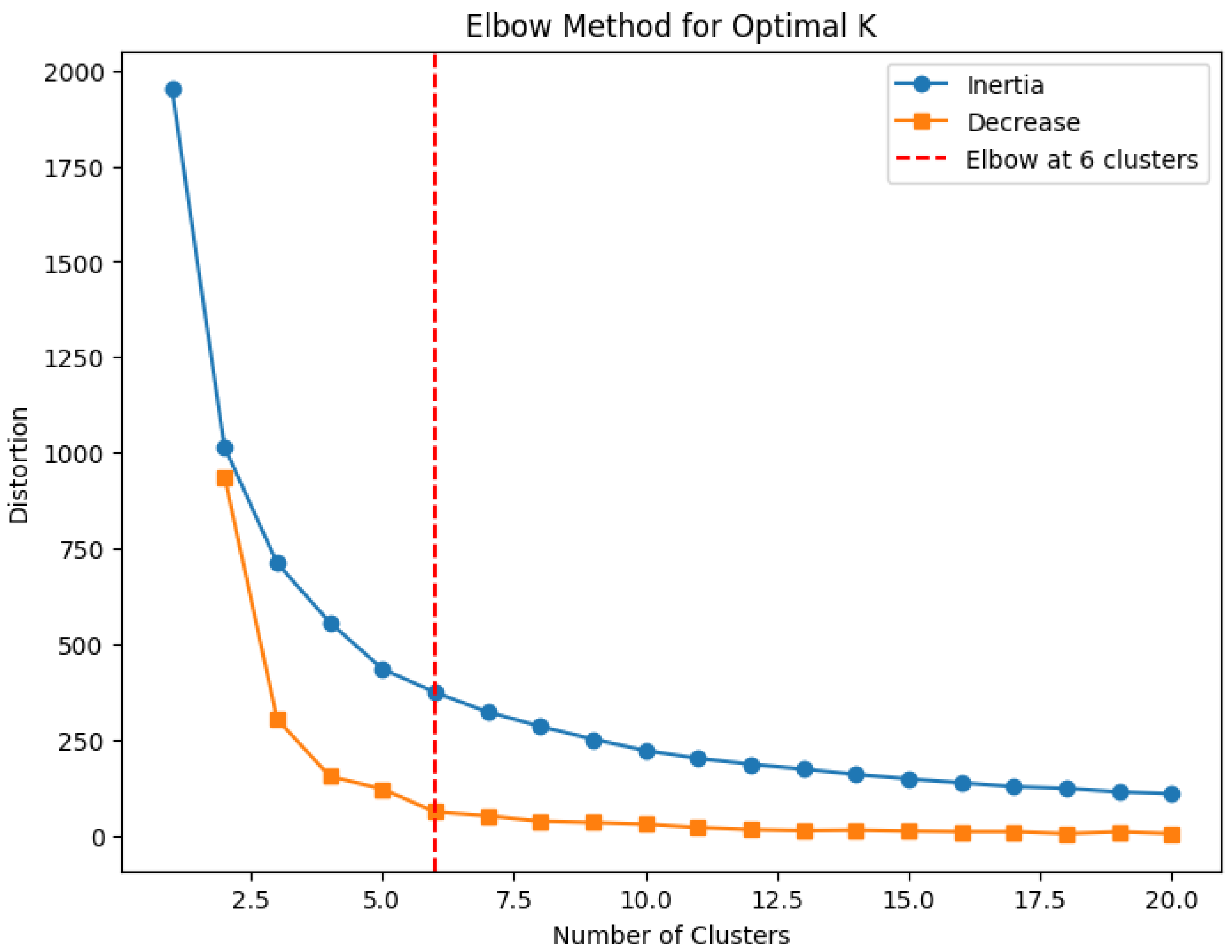


Figure 14. Elbow method for finding the optimal K-value

As seen in the line plot for the elbow method (Fig.14),  the “elbow” or the optimal value for K can be seen when there are 6 clusters, as going lower than 6 would mean that the distortion or inertia of the K-Means model would be higher and going higher than 6 would not yield a significant decrease in the inertia of the K-Means model.

After identifying the optimal value for K, the K-means model will be trained and the data is fitted to the model. Through unsupervised learning, the K-means algorithm would return distinct clusters that can be used for further analysis.

As seen in the scatter plot of the distinct clusters (Fig.15), the coordinates are bounded inside the Chicago Metropolitan Area with Cluster 1 having a high density of ridership and distribution of docking stations. It can also be seen that Cluster 1 is at the city center.

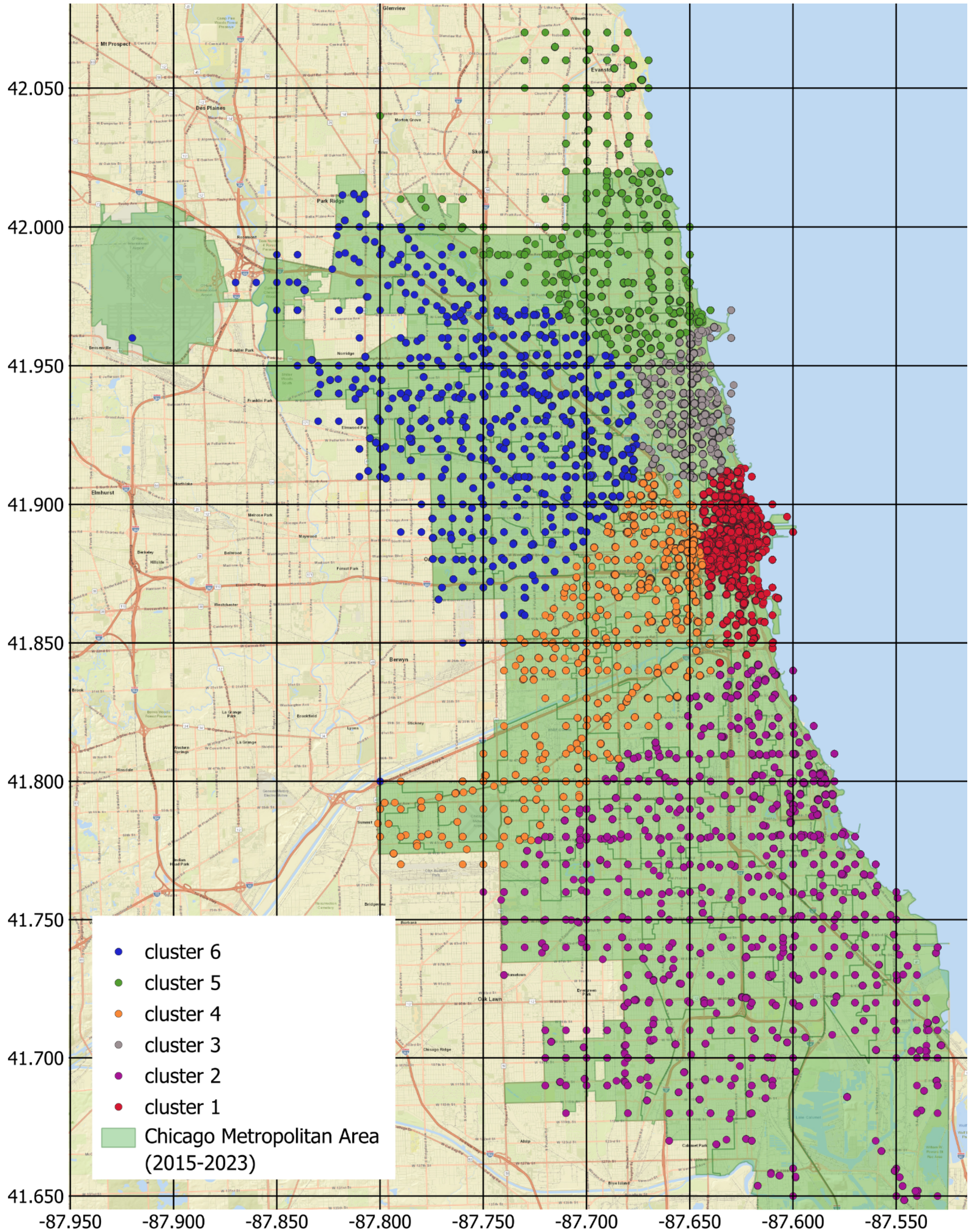


Figure 15. Scatter plot of distinct clusters in Chicago

To further improve the visualization of the distinct regions, the clusters will be replaced with polygons covering their respective regions (Fig.16). Through this, distinct regions within the city based on station usage patterns will be clearly defined.

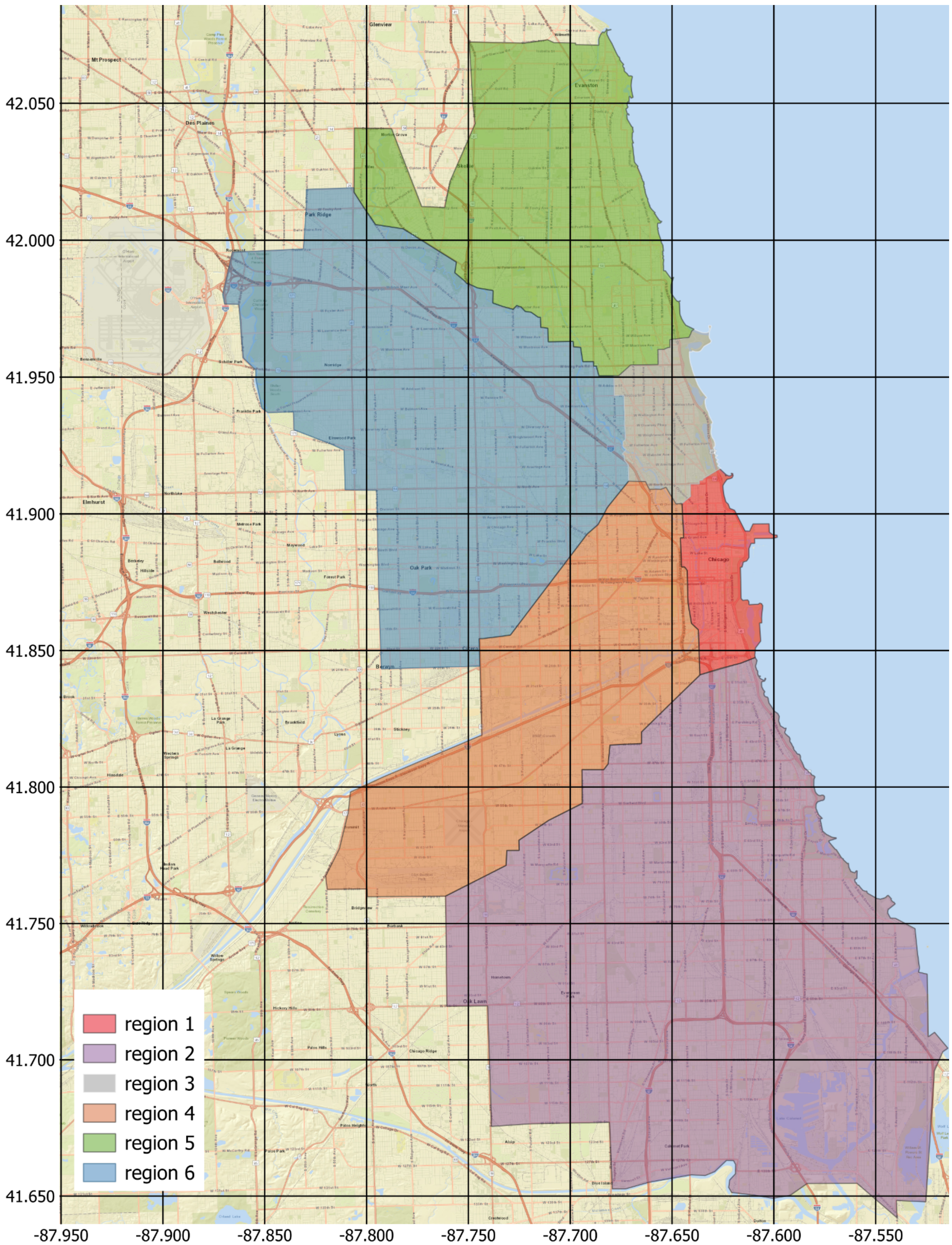


Figure 16. Spatial map of distinct regions in Chicago

For more visualization of the frequency of ridership in Chicago, a bar graph will be utilized (Fig.17). In this bar graph, the frequency of ridership for each region is shown. The graph is also divided into three using the upper and lower quantile of the data to identify the regions that have high, medium, and low frequency of ridership.

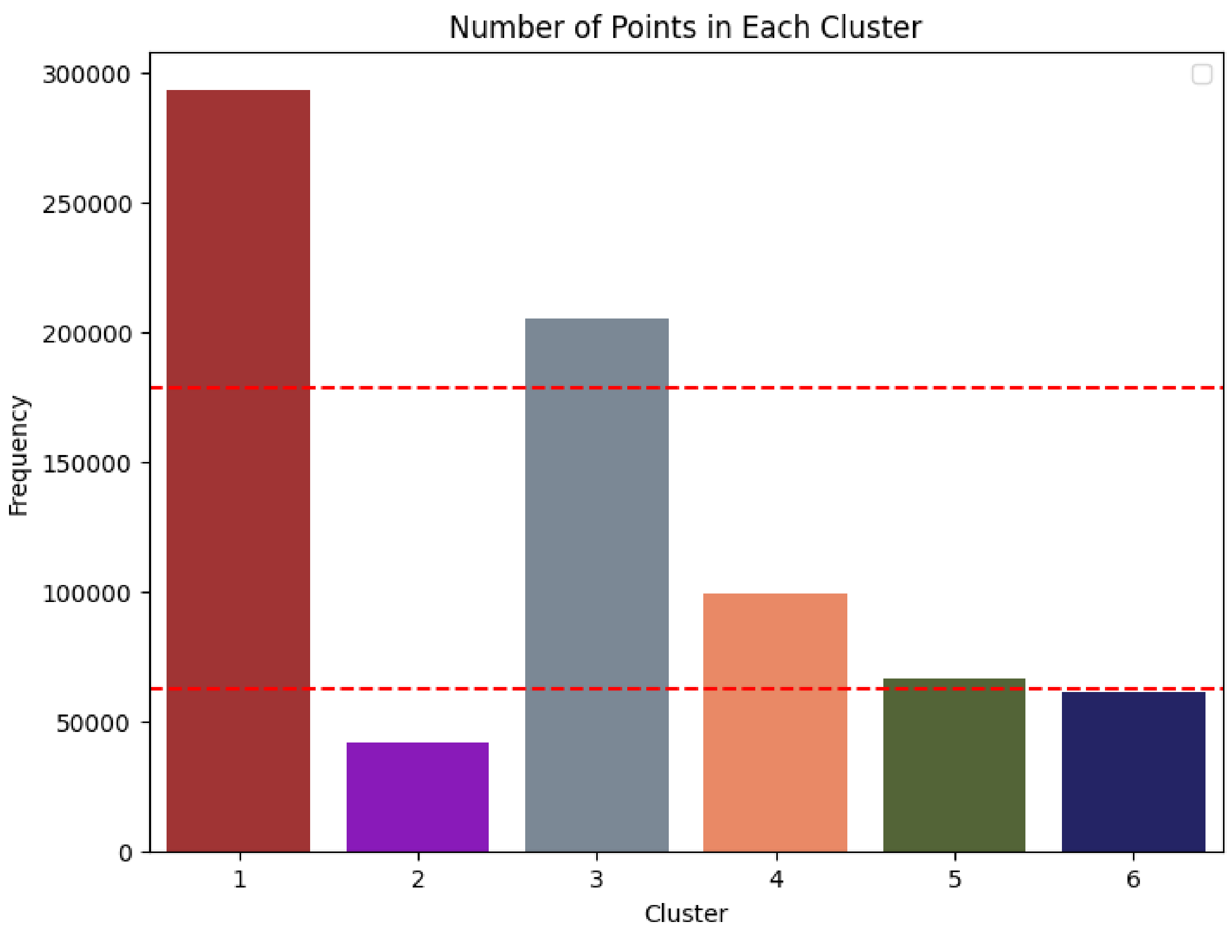


Figure 17. Ridership frequency for each region

As seen in the bar graph, Region 1 has the highest frequency of ridership among all regions. Furthermore, Regions 1 and 3 can be classified as having a high frequency, Regions 4 and 5 as having a medium frequency, and Regions 2 and 6 as having the lowest frequency.

## IV. Recommendations

The team proposes the following recommendations in order to improve Divvy:

1. **Bike Return Reminders and Bike Stolen Report Features**

   In Figures 5 and 6, it was found that almost half a million dollars in revenue was generated by only 918 users who had a renting time of over a day. The longest renting time recorded in July of 2023 was even more than a month with the bill costing around 8000 dollars. These irregularities are attributed to either the rider's unawareness of returning the bike or the possibility of the bike being stolen. In either case, the large costs spent are not due to the user's intention, which may tremendously affect their likelihood of continuing using the service provided by Divvy. Thus, notifications regarding returning the bike or feature that allows the user to report on lost bikes may help this situation. Regarding the lost bikes, maybe Divvy can charge the user a certain amount since they are also accountable for the loss, but this is definitely better than a 8,000 dollar bill.

2. **Data Collection**

According to Figure 1, there are 4 different types of users and three different types of transportation. However, the data collected only include 2 types of users and 2 types of bikes. The lack of such data will hinder the data analysts to provide accurate interpretations of the performance of the company, which ma directly affect its management strategies and development.

3. **Targeted Advertising and Promotions**

As can be seen in Figures 7-12, there exist particular patterns in the usage of different types of bicycles depending on a number of factors. These patterns may be used by the company to further reinforce usage along said patterns. For example, it is seen that longer, more leisurely rides are more prevalent on the weekends. As such, advertising can be made with this in mind; emphasizing Divvy as a service providing its users with a leisurely, stress-free experience to enjoy the weekends with. As another example, the type of bike can instead be emphasized in promotional media such as advertising Divvy's electrical bikes as a fast and cheap option for commuters traveling to work on a weekday.

4. **Station and Bike Upkeep Practices**

Through both the clustered geospatial data (Figures 15 & 16) and information regarding the common interstation distances of rides depending on bike type (Figures 9 & 10), a set of "best practices" for bike and station upkeep may be identified. In regards to the clusters, certain regions have a much more dense distribution of docking stations in comparison to regions further from the city center. Besides this, it is noted that the different types of bikes have different strengths with classic bikes being used for shorter distances while electric bikes are often used for long distances. With this in mind, it would be wise to stock the more station-dense regions with a greater number of classic bikes while electric bikes are more numerous in sparser regions. At the start of the day, riders from the sparser regions commute into the city center utilizing electric bikes. For transportation within the city center, the cheaper classic bikes are used. Finally, for the commute back home at the end of the day, electric bikes which were docked in the city center by commuters in the morning are used in the commute home. Station and bike upkeep practices should serve to ensure the smooth operation of these daily activities.