



Politecnico
di Torino

TRANSPORT INNOVATION FOR A SUSTAINABLE, INCLUSIVE, AND SMART MOBILITY

Professor **Cristina Pronello**

REPORT

MICRO-MOBILITY DATA

Presented By:

Saeideh Mohammadikish

S329781

Amir Yarmohamadi

S329783



A.Y.2024/25

Introduction

In this project, Jupyter Notebook was employed as the computational environment for developing and executing the python codes. The analysis relied on a suite of specialized Python libraries: pandas facilitated data manipulation and preprocessing, numpy supported numerical computations, and seaborn along with matplotlib.pyplot enabled the visualization of trends and patterns within the data. For geospatial analysis, geopandas was utilized to handle spatial datasets, while shapely.wkt provided functionality for parsing and manipulating geometric objects. Furthermore, QGIS was used to produce high-resolution maps and conduct advanced spatial analysis, enhancing the interpretation of transport infrastructure and mobility data particularly in visualizing origin and destinations, in part three of the exercise. This integrated approach ensured a rigorous and reproducible methodology for analyzing public transport lines and micro-mobility datasets.

To complete this exercise, two free databases were utilized: the Austin and Chicago micro-mobility databases. The process began with data standardization, followed by data cleaning. Next, patterns and trends within the databases were identified, leading to the resolution of specific, related problems to uncover meaningful insights.

In this report, references to the code are formatted as [EX3_4, Cell 85]. This indicates that the code can be found in the Jupyter Notebook file titled EX3_4, specifically in cell number 85. There are two files EX3_4 and EX1_2. All files can be downloaded at: https://drive.google.com/drive/folders/13Yqz2iLR3LmXvCL9CDeObj2uD2MNzCRS?usp=drive_link

Exercise 1 – Preliminary analysis

Data cleaning [EX1_2, Cell 12]

The initial dataset consisted of **514,767 records**, requiring comprehensive cleaning to ensure data integrity and usability for analysis. The cleaning process addressed various types of "bad data," which included:

Code output:	
[ID	0
Device ID	0
Vehicle Type	0
Trip Duration	0
Trip Distance	0
Start Time	0
End Time	0
Modified Date	0
Month	0
Hour	0
Day of Week	0

Council District (Start)	1099
Council District (End)	1161
Year	0
Census Tract Start	1054
Census Tract End	1051
Start Time (US/Central)	660
End Time (US/Central)	660

1. Malformed Rows [EX1_2, Cell 05]

Rows with fewer than the expected 18 columns were identified as malformed.

These rows, totaling **4 records**, were logged in an error file and removed from the dataset.

Code output:

```
[{"error": true}, {"message": "Internal error"}, {"status": 500}]  
Number of bad lines: 4
```

2. Missing or Invalid Values [EX1_2, Cell 12]

Missing entries were replaced with pd.NA, and common placeholders such as "NULL," "NaN," "none," and "None" were standardized as missing values.

Missing values were found in several columns:

The attributes causing issues in our project when they contain null values are Council District (Start) and Council District (End). Wherever possible, missing values were replaced with default values, such as 0 for categorical fields like council districts.

3. Data Type Standardization [EX1_2, Cell 15]

Relevant columns were converted to appropriate data types, including:

Trip Duration and *Trip Distance*: converted to integers and floats, respectively.

Year: converted to an integer format.

Timestamp fields: formatted as datetime objects for consistency.

4. Duplicate Rows [EX1_2, Cell 31]

No duplicate rows were found in the dataset.

5. Outliers (using the Interquartile Range (IQR) and Z-Score methods) [EX1_2, Cell 35]

Outliers in "*Trip Distance*" and "*Trip Duration*" separately were detected using both the Interquartile Range (IQR) and Z-Score methods. I employ each method separately then fined and outlines the common rows which are detected in both methods. A total of **54 rows** in "*Trip Distance*" and **172 rows** in "*Trip Duration*" were identified

as common outliers between these methods and subsequently removed. These files also saved in a separate “.csv” file to check.

6. Zero-Value Entries [EX1_2, Cell 27]

Records with zero values in critical fields, such as *Trip Duration* and *Trip Distance*, were excluded from the dataset. Also, in some parts of the analysis, null origins or destinations were omitted from the database.

The number of records before and after data cleaning. What were the types of “bad data” that needed to be cleaned?

The dataset initially contained **514,767 records** and was reduced to **498,474 records** through a meticulous data cleaning process. Various issues were addressed, including improperly formatted lines, missing or invalid values, records with zero values in key fields like "Trip Duration" and "Trip Distance," and outliers identified using the IQR and Z-Score methods. This process ensured the removal of unreliable data, resulting in a refined dataset ready for accurate analysis and insights in transport studies.

When did collection of data start (start-date and time) for micro-mobility dataset and what is the most recent date and time available.

The micro-mobility dataset provides a comprehensive temporal coverage, spanning from **August 19, 2021, at 05:30:00 to March 25, 2022, at 01:00:00**. [EX1_2, Cell 25]

<i>Code output:</i>
count 514762.000000 mean 678.592280 std 2917.778137 min 0.000000 25% 234.000000 50% 428.000000 75% 783.000000 max 693899.000000 Name: Trip Duration , dtype: float64 count 5.147620e+05 mean 1.735178e+03 std 2.537801e+04 min 0.000000e+00 25% 5.260000e+02 50% 1.136000e+03 75% 2.108000e+03 max 1.154537e+07 Name: Trip Distance , dtype: float64 count 514762 mean 2021-09-21 01:14:58.048807424

```

min      2021-08-19 05:30:00
25%     2021-08-28 23:30:00
50%     2021-09-06 00:45:00
75%     2021-09-23 03:30:00
max      2022-03-25 01:00:00
Name: Start Time, dtype: object
count          514762
mean   2021-09-21 01:14:58.048807424
min      2021-08-19 05:30:00
25%     2021-08-28 23:30:00
50%     2021-09-06 00:45:00
75%     2021-09-23 03:30:00
max      2022-03-25 01:00:00
Name: End Time, dtype: object
***in all rows, Are all end times after the start time?***
YES, Its Ok

```

Number of records per year and month in micro-mobility dataset – Use bar plots to display patterns using the plot () function of pandas.

1.Patterns in Micro-Mobility Usage [EX1_2, Cell 43]

The analysis of trip durations and distances across different vehicle types reveals key usage trends:

- **Cars** have the longest average trip distance (7,595 meters) and duration (75.99 minutes), indicating their use for longer trips.
- **Bicycles** and **scooters**, which represent the majority of trips, are associated with shorter durations and distances, reflecting their role in urban mobility for shorter distances.
- **Mopeds** have the shortest average trip duration (8.54 minutes), highlighting their use for quick, short-distance trips.

Figure 1 shows the distribution of records per each mobility vehicle. Additionally, The following table presents the average trip duration by vehicle type:

Code output:

Vehicle Type
bicycle 13.428773
car 75.989130
moped 8.541227
scooter 11.499752

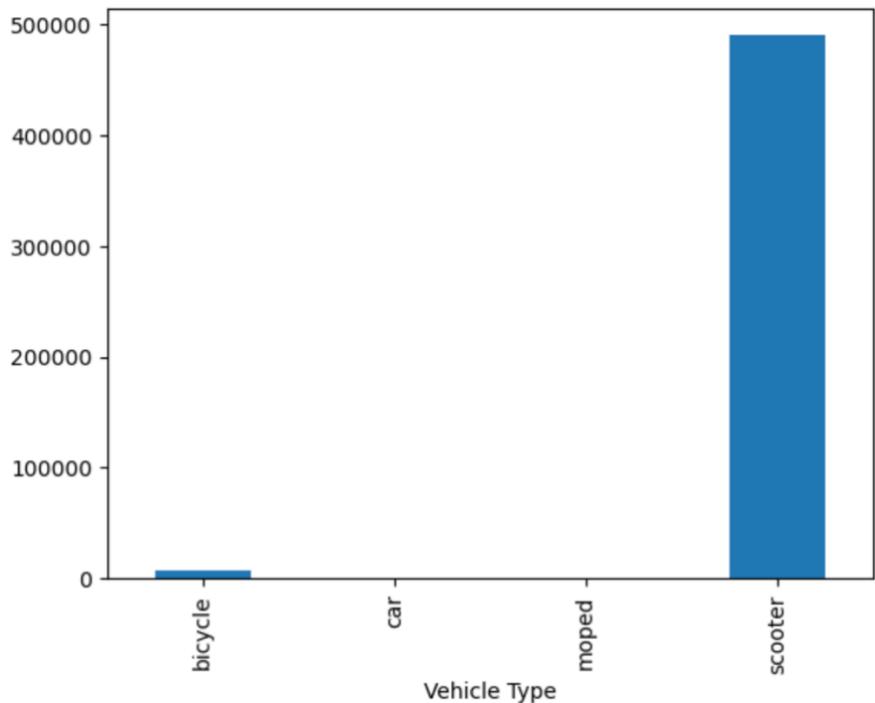


Fig 1. Number of records per vehicle

2. Exclusion of Rare Vehicle Types [EX1_2, Cell 49]

To enhance the relevance and quality of the analysis, vehicle types with a limited number of records, such as cars and mopeds, were removed from the dataset. This step was taken to focus on the more prevalent vehicle types, like scooters and bicycles, which represent the majority of the dataset. As a result, the total number of records decreased from 498,528 to 489,899, reflecting the exclusion of less common vehicle types. This adjustment ensured that the dataset remained robust while prioritizing the most frequently used modes of transportation, thus improving the overall quality of the analysis.

Code output:

```
DataFrame shape Before omitting Cars and Mopeds: (498528, 19)
DataFrame shape: (489899, 19)
```

It should be noted that, in some parts there are some questions regarding records belongs to bycles. This is whay bicycles are not omitted in this part.

3. Month per year included in the database [EX1_2, Cell 51]

The dataset includes data for specific months across two years. For the year 2021, data is available for the months of August (8), September (9), October (10), November (11), and December (12). For the year 2022, data is available for the months of January (1), February (2), and March (3).

Code output:

Months available for each year:

Year

2021 [11, 12, 10, 9, 8]

2022 [1, 2, 3]

4. Seasonal Trends and Monthly Usage Patterns [EX1_2, Cell 53]

An in-depth analysis of trip frequencies by month revealed notable seasonal patterns in vehicle usage. These trends varied significantly across the year and were influenced by different vehicle types.

1. Peak Usage:

The highest trip frequencies occurred during the summer months, specifically in August and September. This period experienced a notable surge in activity, primarily driven by scooters, which saw increased usage likely due to favorable weather conditions and heightened outdoor activities during the summer.

2. Low Usage:

In contrast, the winter months, especially January and February, exhibited a marked decline in trip frequencies. This reduction in activity was observed across all vehicle types, suggesting that colder temperatures and adverse weather conditions significantly impacted the mobility of users.

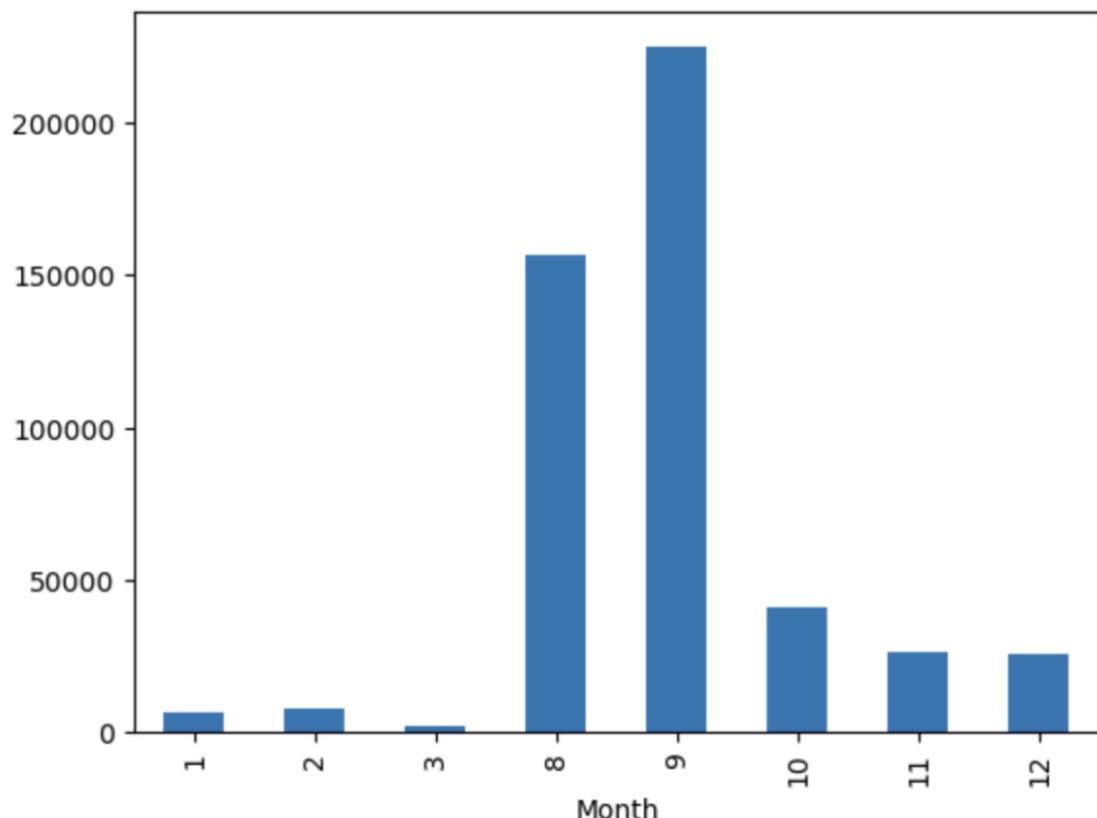


Fig 2. Month record distribution.

To further quantify these seasonal trends, a heatmap was generated, fig. 3, highlighting the number of unique vehicles used per month across different two vehicle types. The heatmap clearly confirmed the seasonal fluctuations, with scooter usage peaking in the summer months and significantly lower activity during the winter period. This visualization reinforces the observed seasonal patterns, providing a more comprehensive understanding of how usage trends are influenced by seasonal factors.

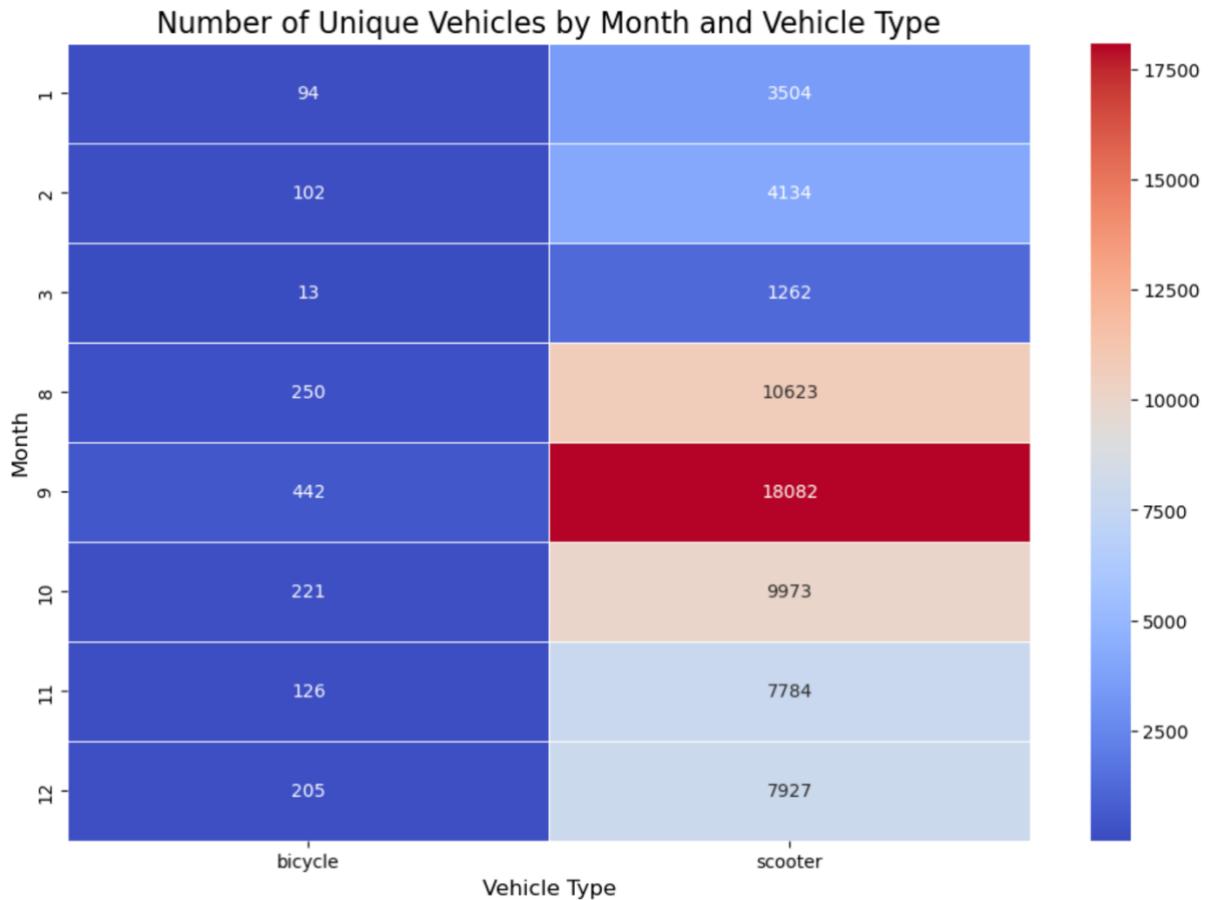


Fig 3. Month vehicle heatmap.

4. Hourly Trip Patterns [EX1_2, Cell 53]

Figure 4 shows the distribution of trips throughout the day. Statistics reveals notable peaks and troughs in activity:

Peak Activity: **Evening Surge (6 PM to 9 PM):** The highest number of trips occurred during evening hours, especially for scooters. This surge can be attributed to the conclusion of the workday and school hours, leading to increased demand for transportation. Limited public transit options and the desire for quick, flexible transportation for social or leisure activities further contribute to the elevated usage during this time.

Midday Increase (10 AM to 5 PM): Trip frequency gradually increased throughout the day, peaking between midday and late afternoon. This increase likely reflects work-related commutes, errands, and leisure activities. During this period, the demand for micro-mobility vehicles rises as people engage in daily activities, such as running errands or commuting to work.

Off-Peak Activity: **Early Morning and Late-Night (2 AM to 6 AM):** These hours recorded the lowest levels of activity, corresponding with minimal transportation needs. The absence of commuters and the unavailability of services during these hours results in a significant drop in trips. Most users are asleep, and the demand for transportation is therefore minimal.

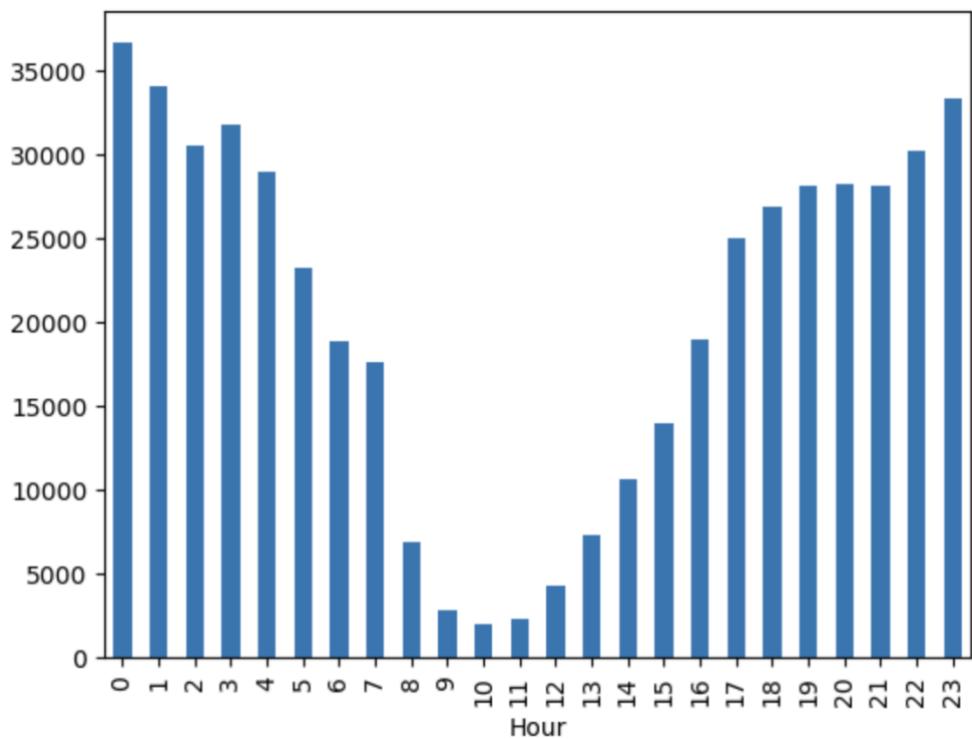


Fig 4. The distribution of trips throughout the day

The dataset shows that micro-mobility usage peaks in the evening, driven by social and work-related factors. Midday trips also show steady activity, likely due to work and personal errands. The lowest activity occurs during the early morning and late-night hours when demand is minimal and services are limited. This pattern highlights the daily rhythms of urban mobility and the influence of public transportation availability and user schedules.

Code output:		
Hour size		
0	0	36618
1	1	34056
2	2	30430
3	3	31731
4	4	28883

5	5	23186
6	6	18807
7	7	17599
8	8	6866
9	9	2782
10	10	1958
11	11	2306
12	12	4276
13	13	7323
14	14	10658
15	15	13906
16	16	18927
17	17	24976
18	18	26825
19	19	28055
20	20	28193
21	21	28055
22	22	30178
23	23	33305

5. Weekly Trip Patterns [EX1_2, Cell 57]

The analysis of trip data by day of the week, starting from Sunday (0), reveals clear patterns in micro-mobility usage. Figure 5 illustrates the trend in the data. The highest trip volumes are recorded on Friday(5), Saturday (6) and Sunday (0), suggesting that micro-mobility vehicles are primarily used for recreational and leisure activities during the weekend. With more free time available, individuals are more likely to use these vehicles for social outings, tourism, or other leisure-related purposes.

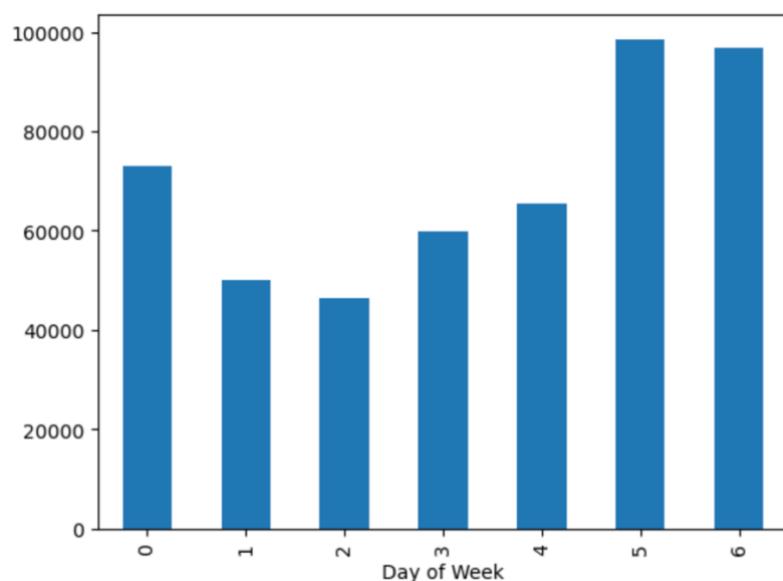


Fig 5. The distribution of trips throughout the week.

In contrast, weekday usage shows moderate activity. Monday (1), Wednesday (3), and Thursday (4) exhibit slightly higher usage, likely driven by commuting and daily errands. Micro-mobility vehicles are often used as convenient transportation options for work or essential activities during the weekdays.

An interesting observation is the decrease in trips on Friday (5). This drop in usage could be due to individuals shifting to other forms of transportation or preparing for the weekend. On Friday, people may be more likely to use traditional transport modes or may already be planning their weekend activities.

In summary, weekend days show the highest demand for micro-mobility vehicles, driven by recreational needs, while weekday usage reflects a more utilitarian pattern, with Friday marking a transition to weekend behaviors.

Code output:

Day of Week size

0	0	73119
1	1	49970
2	2	46365
3	3	59714
4	4	65382
5	5	98514
6	6	96835

Number of unique vehicles by year and vehicle type – identify patterns

Are there more vehicles as years go on ?

The data reveals a significant variation in the number of unique vehicles operating in Austin, Texas, across 2021 and 2022. In 2021, the dataset recorded 24,782 unique vehicles, while in 2022, this number dropped sharply to 7,256. This decline could be attributed to several factors, including seasonal fluctuations in micro-mobility demand, changes in regulations, or operational adjustments by service providers. For instance, stricter regulatory measures or reduced subsidies might have limited fleet expansions, while reduced demand during the COVID-19 recovery phase could have further impacted operations. Another simple reason could be that we had not access to all the data or perhaps the data collection was not complete!

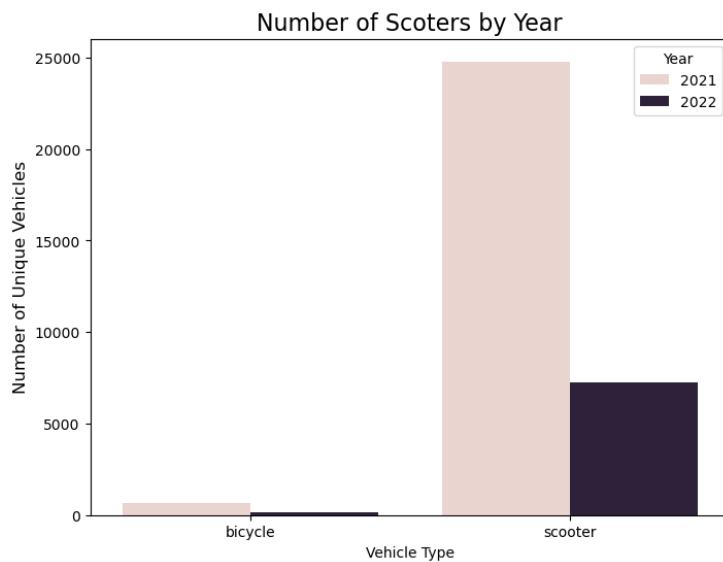


Fig 6. Number of trips per year for scooter and bicycle.

Scooters represented the dominant vehicle type in both years, highlighting their central role in Austin's micro-mobility landscape. Their popularity may stem from their affordability, convenience, and suitability for short urban trips, particularly in a city with relatively mild weather and a spread-out urban layout. Bicycles, while less commonly used, showed a stable presence, likely catering to users seeking a more traditional mode of transport or enjoying the city's outdoor-friendly environment.

Is there some change in usage patterns among different days of the week, months is there a trend – seasonal or weekly?

Weekly and Seasonal Micro-Mobility Usage Patterns [EX1_2, Cell 75]

The analysis of micro-mobility usage in Austin reveals distinct weekly and seasonal patterns, influenced by the city's culture, climate, and transportation needs. The data highlights differences in scooter and bicycle demand across the days of the week and months, reflecting both recreational and practical uses of these vehicles.

Weekly Usage Trends

Based on figure 7 and 8, micro-mobility usage is highest on weekends, particularly Saturdays and Sundays, driven by recreational and social activities in Austin's outdoor culture. Saturdays show the peak usage, combining leisure with some work-related trips. On weekdays, usage is lower and primarily for commuting, errands, and short-distance transport, with a slight increase on Fridays as people prepare for the weekend. Scooters are favored for short, spontaneous trips on weekends, while bicycles are used more consistently throughout the week for commuting and fitness purposes.

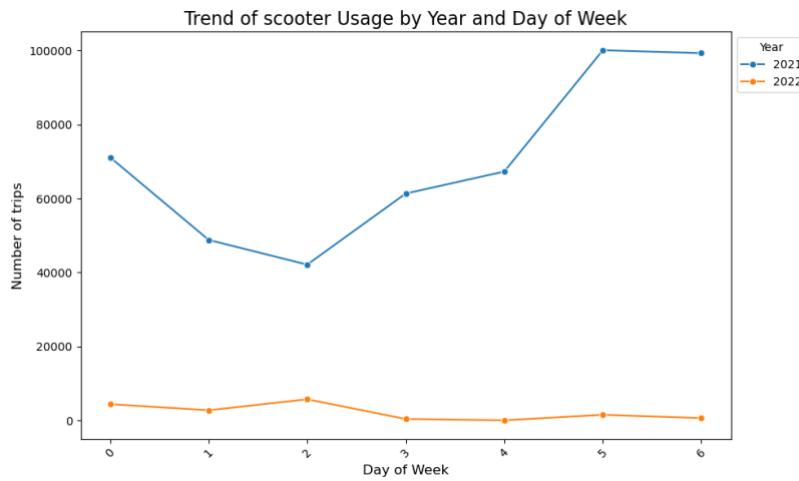


Fig 7. Number of trips for scooters per day of week, 2021 and 2022.

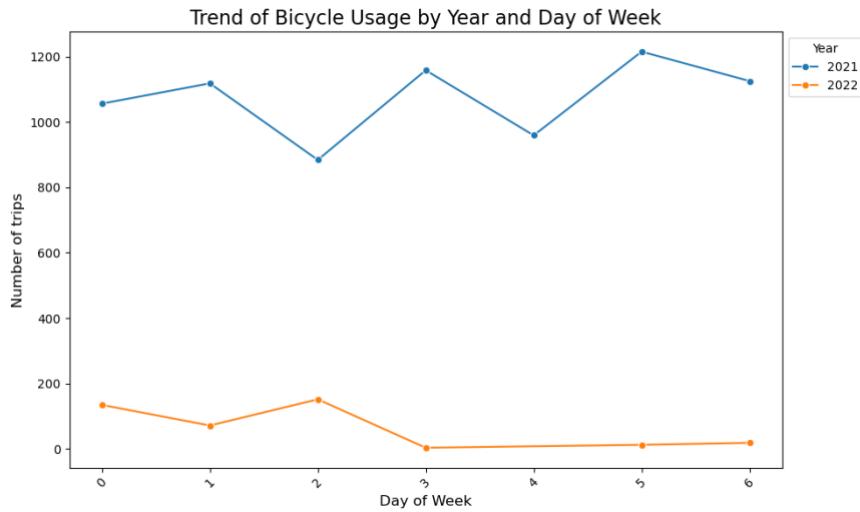


Fig 8. Number of trips for Bicycles per day of week, 2021 and 2022.

Public transportation availability also affects usage patterns. Limited transit options on weekends result in a higher reliance on micro-mobility vehicles, particularly scooters and bicycles.

Seasonal Usage Trends [EX1_2, Cell 67]

Seasonal trends show a clear link between weather and usage. Summer months, especially August and September, experience the highest demand due to favorable weather and outdoor activities, coupled with an increase in tourists and events. In contrast, winter months, particularly January and February, see a significant decline in usage, likely due to cooler temperatures and reduced outdoor activities.

Scooters see greater demand during weekends and summer months for short trips, recreational trips, while bicycles exhibit steadier usage across both weekdays and weekends.

Bicycles are more suited for commuting or fitness-related activities, leading to their consistent usage throughout the week.

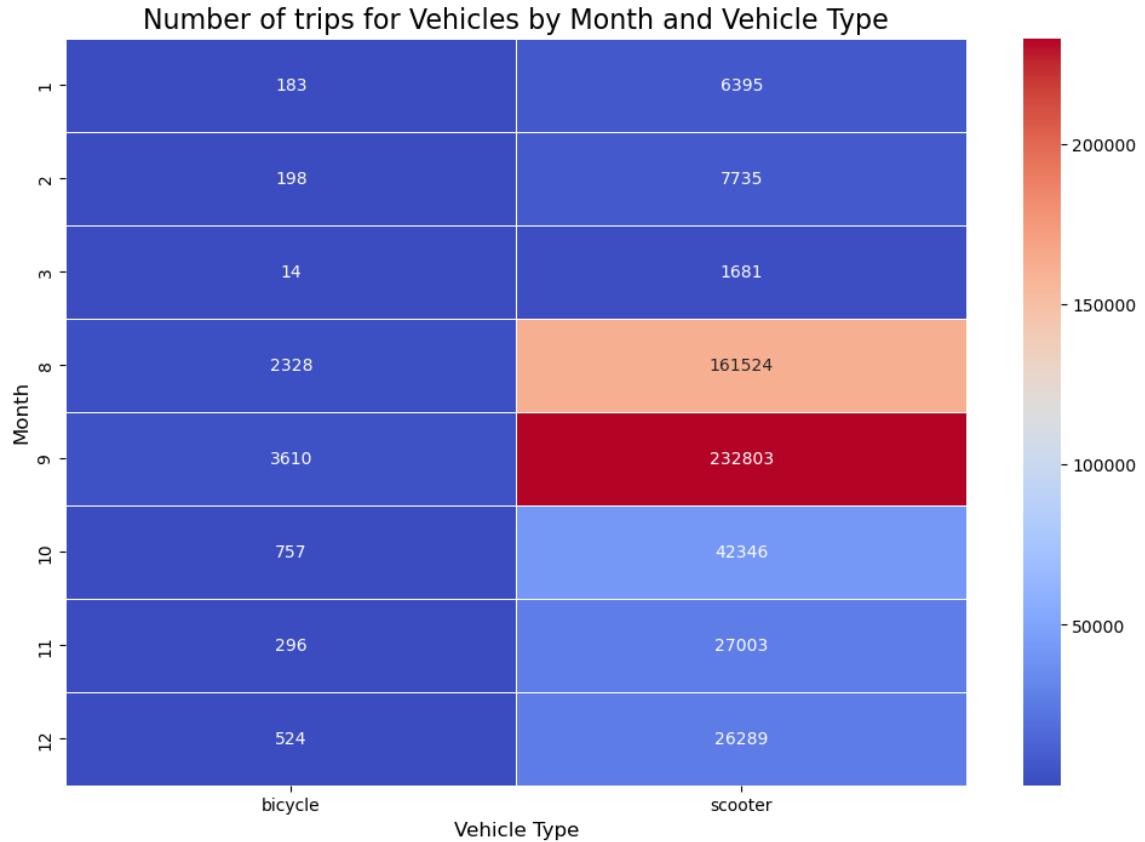


Fig 9. Number of trips for Bicycls per day of week, 2021 and 2022.

Are there any trends based on the gender and age of the user?

NO DATA AVIALABLE IN AUSTIN DATABASE. This part will be investigated in Chicago database.

Calculate the number of trips for each vehicle in the database and produce a histogram for it.

Visualization through bar plots confirmed scooters' dominance, with clear year-over-year declines in unique vehicle counts for both scooters and bicycles. A heatmap further emphasized this pattern, providing a holistic view of the trends across vehicle types and years.

The sharp reduction in unique vehicles between 2021 and 2022 may also reflect economic or operational pressures, such as rising maintenance costs, reduced ridership, or shifts in service provider strategies. These findings underscore the importance of scooters as a primary mode of micro-mobility in Austin while raising questions about the sustainability of fleet sizes under changing urban dynamics. The number of unique vehicles were analyzed by year and vehicle type. The results are shownen in figure and , respectively. [EX1_2, Cell 63]

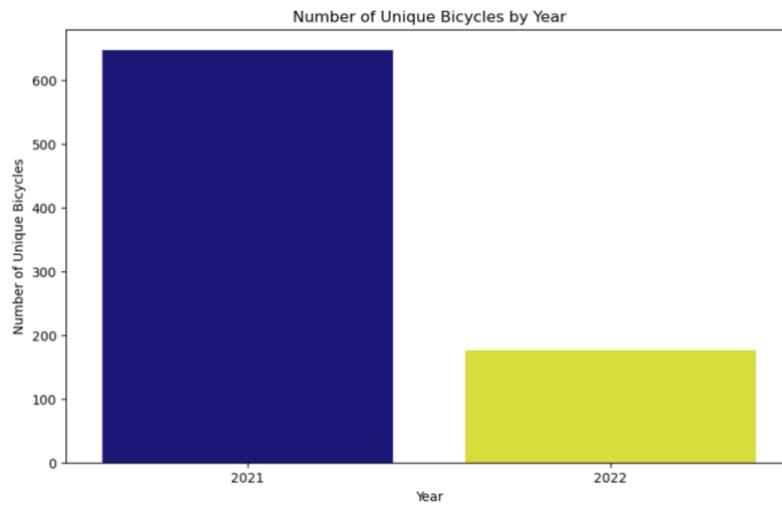


Fig 10. Number of unique Bicycls per day of week, 2021 and 2022.

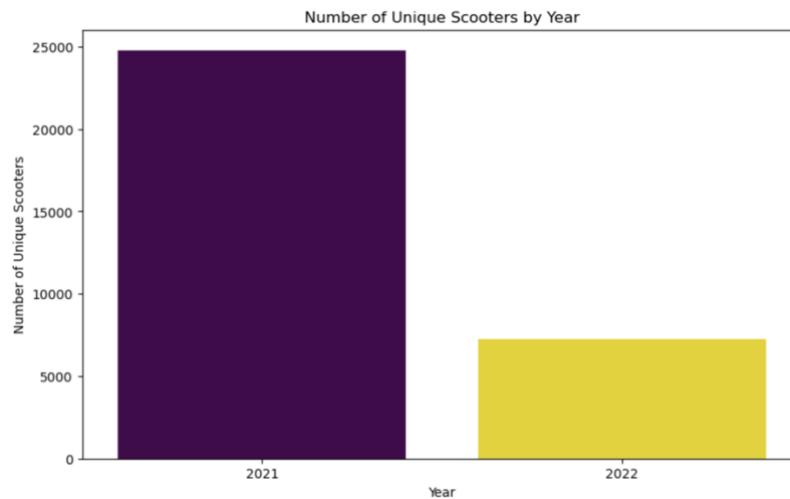


Fig 11. Number of unique Bicycls per day of week, 2021 and 2022.

Calculating and Visualizing the Number of Trips per Vehicle [EX1_2, Cell 81]

To understand the distribution of trips across different vehicles in the dataset, we calculated the number of trips for each vehicle and created a histogram to represent the data. This visualization offers insights into vehicle usage frequency, highlighting patterns of underutilized or overused vehicles.

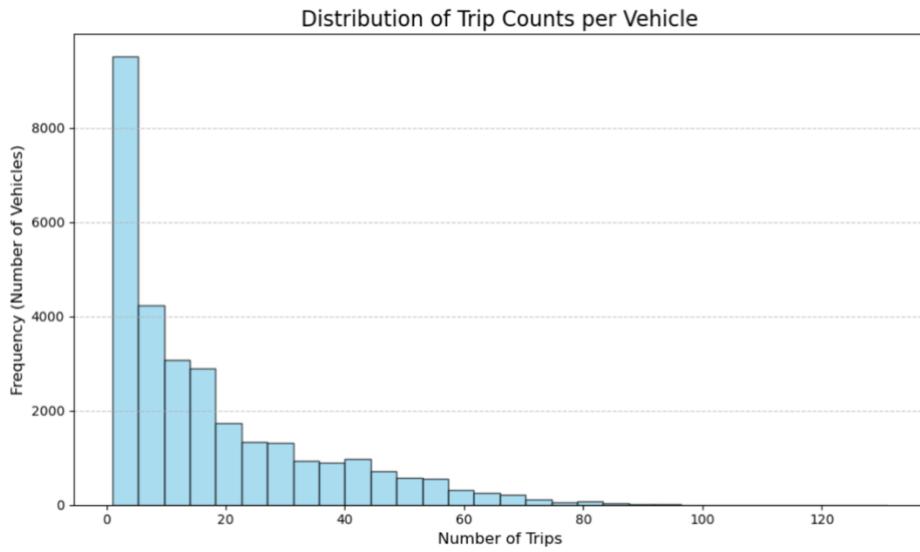


Fig 12. The distribution of trips per unique scooter across different vehicles.

A small proportion of vehicles accounted for a majority of trips, demonstrating a skewed distribution. The histogram reveals several key observations about vehicle usage. The majority of vehicles recorded a low to moderate number of trips, indicating occasional or sporadic usage. However, a small subset of vehicles displayed exceptionally high trip counts, suggesting intensive utilization and potentially consistent demand. The data distribution is noticeably right-skewed, characterized by a long tail extending toward higher trip counts, highlighting the disparity in usage patterns among the vehicles.

These findings highlight critical insights into vehicle availability, temporal trends, and usage patterns, supporting the development of strategies for sustainable and inclusive urban mobility systems.

Exercise 2 – OD matrix [EX1_2, Cell 85]

Compute then the O-D matrix. Try to visualize the results in a meaningful way.

Figure 13 shows the origin-destination (O-D) matrix for all the data. There are no time or date constraints in this section. This diagram may not be technically correct, but we used it to provide insight into the data's pattern. This matrix and its accompanying heatmap offer a comprehensive view of mobility patterns across Austin's council districts. The data highlights District 9 as the most active area, with an extraordinary 325,602 intra-district trips. This dominance underscores District 9's role as a central hub, likely driven by its dense mix of residential neighborhoods, business centers, and recreational attractions. Similarly, District 3 shows a high level of activity, reflecting its importance as a secondary hub of movement.

District 1 also emerges as an area of notable activity, with 11,245 intra-district trips. Its connectivity with other districts, particularly Districts 3 and 9, underscores its role in Austin's broader mobility network. As can be seen in figure 14, this activity may be attributed to its diverse land uses, including residential areas and employment opportunities.

The heatmap reveals significant connections between districts, particularly between Districts 1, 3, and 9. These strong mobility links may result from commuting, service access, or recreational travel. Such patterns underline areas with high social or economic interaction, which are critical for understanding transportation needs.

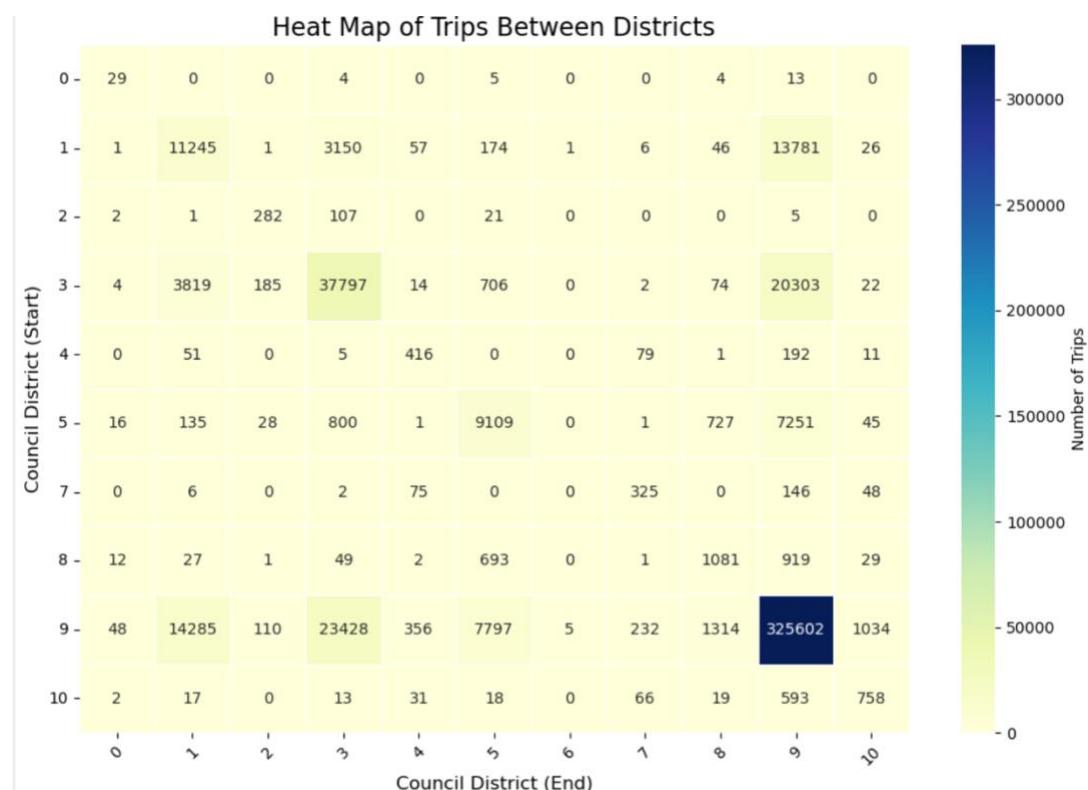


Fig 13. The origin destination matrix heatmap.

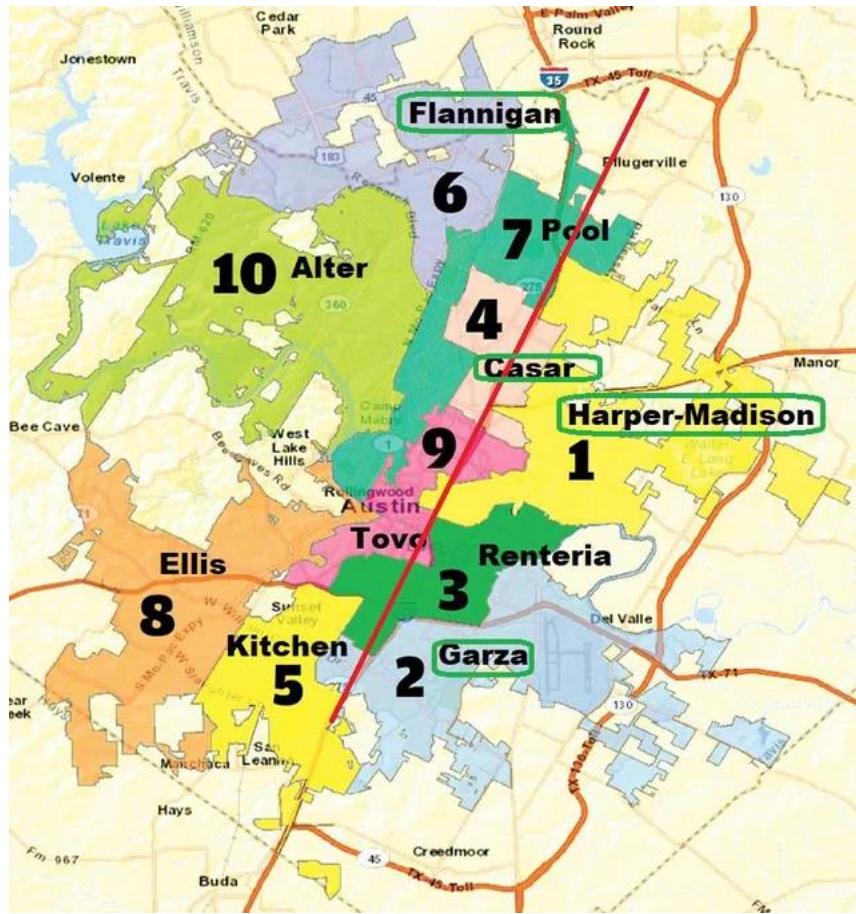


Fig 14. The districts in Austin.

Visualizing these patterns with heatmaps makes it easier to identify key mobility corridors and high-demand zones. This information is essential for city planners, enabling them to prioritize infrastructure development and improve public transit services in areas with the highest demand.

Figures 15 a and b present the origin-destination (O-D) matrix in a 3D format to provide a clearer visualization of trip volumes. In Figure 15, the magnitude of trips within District 9 is so overwhelmingly large that it obscures the patterns and numbers in other areas of the city. To address this issue, District 9 has been excluded in Figure 15 b, allowing for a more effective comparison of trip volumes across the remaining districts. This adjustment offers a more balanced perspective, making it easier to understand mobility patterns and interactions in different parts of the city.

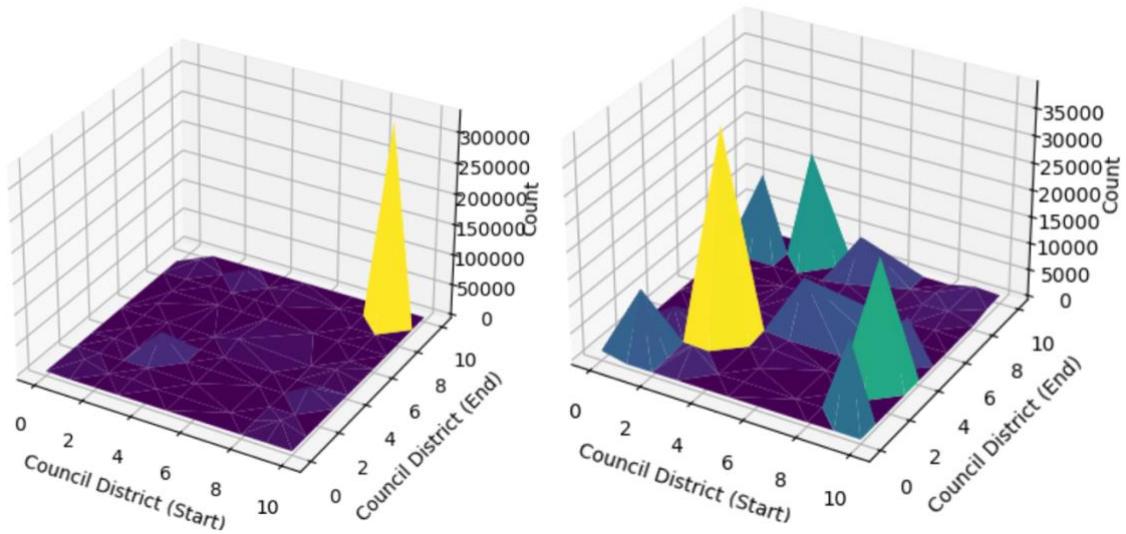


Fig 15. The origin destination matrix_3d view a) all districts b) excluding district 9.

Prepare OD matrices for different hours of the day. Are there any periodicity or trends noticed

The data analysis on scooter usage in District 9 of Austin highlights distinct patterns of travel during different times of a day. The study focused on two time periods: 00:00 AM (peak hour) and 8:00 AM (off-peak hour), comparing how scooter usage varies between these times. [EX1_2, Cell 93]

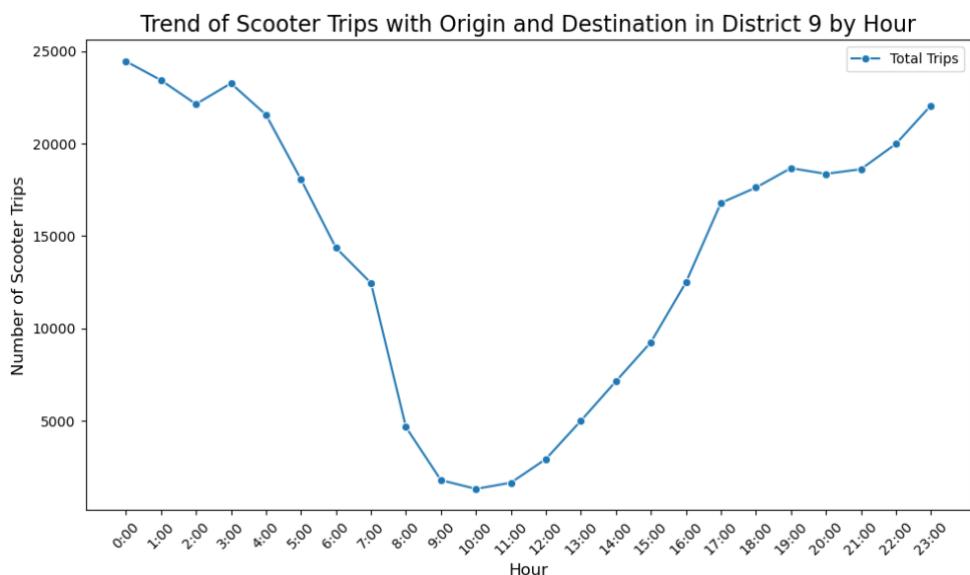


Fig 16. The trend of trips during the day in district 9.

At midnight, there is a noticeable increase in scooter usage. This trend can be attributed to the unavailability of public transportation during late hours, encouraging people to rely on scooters for short, independent trips. The convenience and speed of scooters make them an ideal option for such journeys, particularly in District 9, which is known for its active nightlife and recreational activities.

However, at 8 AM, scooter usage sees a significant drop. This decrease is likely due to the availability of other transportation options, such as public transit and personal vehicles, which are better suited for longer and more time-sensitive commutes. Scooters, being more appropriate for short, intra-city trips, are less appealing during peak morning hours when people generally need to cover greater distances more efficiently.

These observations reflect the impact of time on the demand for scooters in the district, with higher usage in the late-night hours and lower usage during the morning peak.

Figure 17 shows the origin-destination (O-D) matrix as a heatmap, illustrating trip volumes between council districts in Austin during September at 12:00 AM (midnight). The horizontal axis represents the destination districts, while the vertical axis represents the origin districts. The color intensity indicates the number of trips, with darker shades corresponding to higher trip counts.

The most significant observation is the dominance of District 9, which records 11,114 intra-district trips—far exceeding any other district. This highlights District 9's role as a central hub, likely due to its mix of residential, commercial, and recreational areas that remain active late at night. District 9 also has noticeable connections with other districts, including District 1 (471 trips) and District 3 (771 trips), suggesting these areas interact frequently with the city's core.

Outside of District 9, District 3 shows moderate activity with 1,302 intra-district trips, reflecting its importance as a secondary hub. This likely stems from its mix of residential neighborhoods and local businesses. District 1 also records 433 intra-district trips, indicating some localized movement at midnight.

In contrast, most other districts display low trip volumes, as seen in the pale colors of the heatmap. Peripheral districts such as 0, 4, 7, and 10 exhibit minimal activity, which is expected during late-night hours when mobility tends to decrease significantly in less densely populated or suburban areas.

This heatmap highlights the stark differences in mobility patterns across districts. District 9 stands out as the focal point of nighttime activity, while most other districts remain relatively quiet, reflecting typical late-night travel behaviors. These insights are useful for urban planners to prioritize resources and improve late-night transit services in the city.

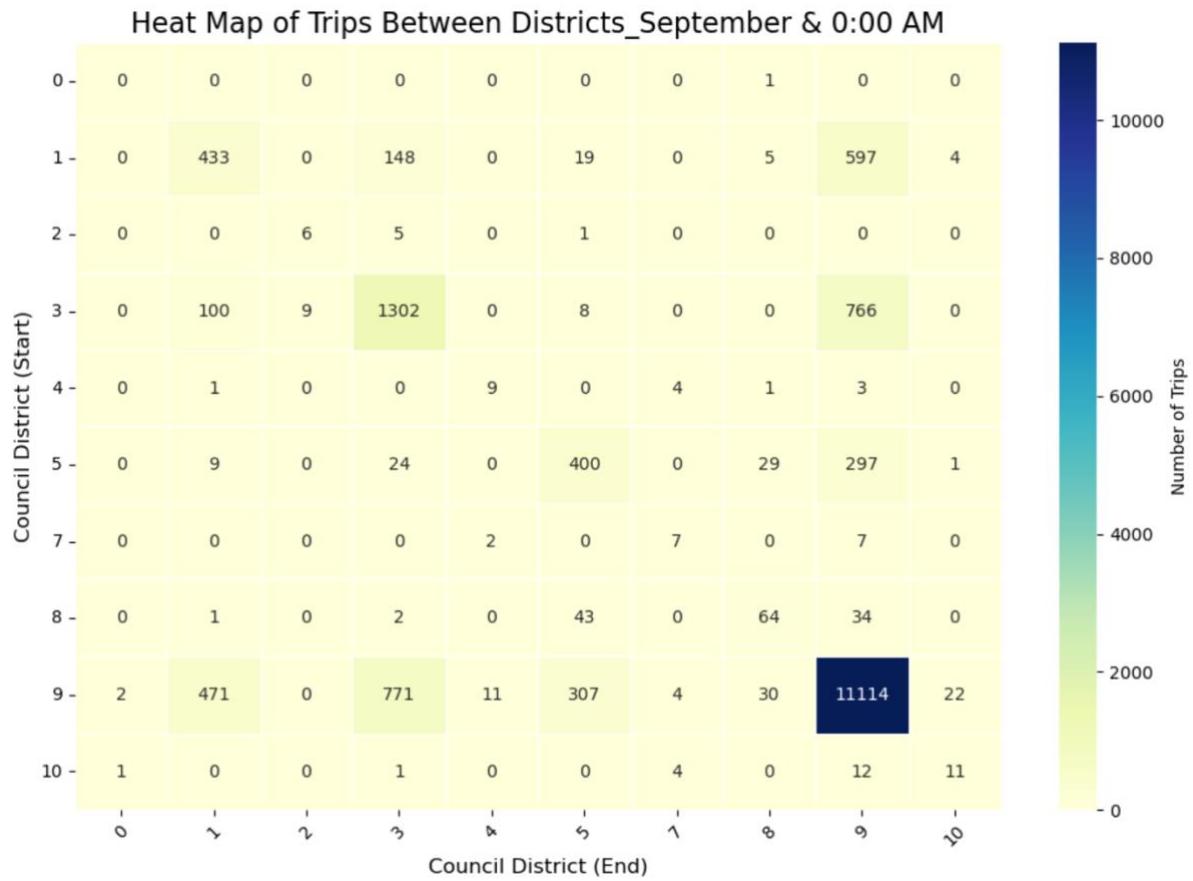


Fig 17. The origin destination matrix heatmap_high Peak seasone:september, high Peak hour:0:00.

Figure 18 presents a heatmap of the origin-destination (O-D) matrix, representing trip volumes between council districts in Austin during September at 8:00 AM. The x-axis indicates the destination districts, and the y-axis represents the origin districts. The color intensity corresponds to the number of trips, with darker colors indicating higher volumes.

At 8:00 AM, District 9 continues to exhibit the highest intra-district activity, with 1,971 trips. This reflects its role as a key urban hub where morning activity, likely associated with commuting and other daily routines, peaks. District 9 also shows moderate connections with District 3 (236 trips) and District 1 (160 trips), emphasizing its strong interaction with neighboring districts during this time.

District 3 displays intra-district trips amounting to 236, reinforcing its role as a secondary hub in the city. This suggests a mix of localized movement and commutes originating within the district. Meanwhile, District 1 shows smaller volumes, such as 73 trips originating within and 52 trips connecting to District 9.

Peripheral districts, such as 4, 7, and 10, exhibit minimal activity, with their light yellow colors indicating very low trip volumes. This is consistent with the typically reduced mobility in these less dense or suburban areas during early morning hours.

Overall, the heatmap highlights a morning peak mobility pattern, with District 9 as the dominant area for both intra- and inter-district trips. The data provides valuable insights for optimizing morning transit services, focusing on high-demand connections like those involving Districts 9 and 3.

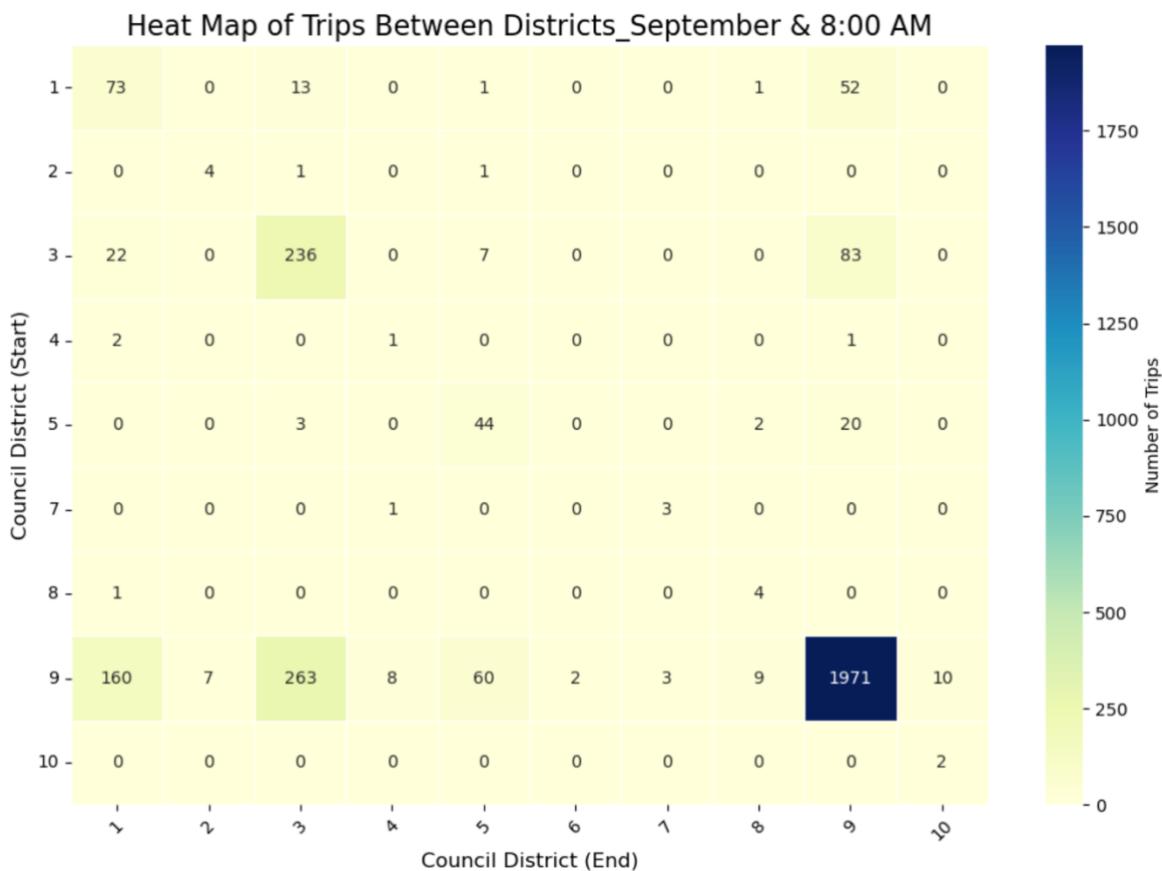


Fig 18. The origin destination matrix heatmap_high Peak seasone:september, low Peak hour:8:00.

Based on your observation, visualise selected OD matrices that show some trends/periodicity on a map.

If a challenge is needed, students can also attempt to create a flowmap for the OD matrices.

To visualize trends and periodicity in the selected origin-destination (OD) matrices, I first processed the council district boundary data of Austin into a GeoDataFrame, allowing for accurate mapping of district centroids using their latitude and longitude coordinates.

After filtering the trip data for valid origin and destination districts, I constructed the OD matrix by counting the trips between district pairs and choose september and 0:00 as the Peak month and hour. Using the Folium library, I created an interactive map with district centroids marked and trip flows represented by lines connecting the origin and destination districts. These lines were color-coded—green for low, blue for moderate, and red for high trip counts—and varied in thickness to reflect the volume of trips.

This map provided a clear, interactive visualization of movement patterns across districts, highlighting areas with high and low trip intensity. Future analysis could incorporate time-

based data to better capture how travel patterns evolve throughout the day, offering insights for more targeted transportation planning and decision-making. [EX1_2, Cell 109]

In this code, the line thickness and color are adjusted based on the flow count (flow_count). The line thickness is calculated using the formula $\max(1, \text{flow_count} / 300)$, ensuring a minimum thickness of 1, while increasing proportionally with the flow count. The line color is determined using the following condition: if the flow count is less than 100, the line color is green; if the flow count exceeds 500, the line color is blue; and otherwise, the line color is red. These adjustments are made to visually represent the intensity of the flow using both line thickness and color.

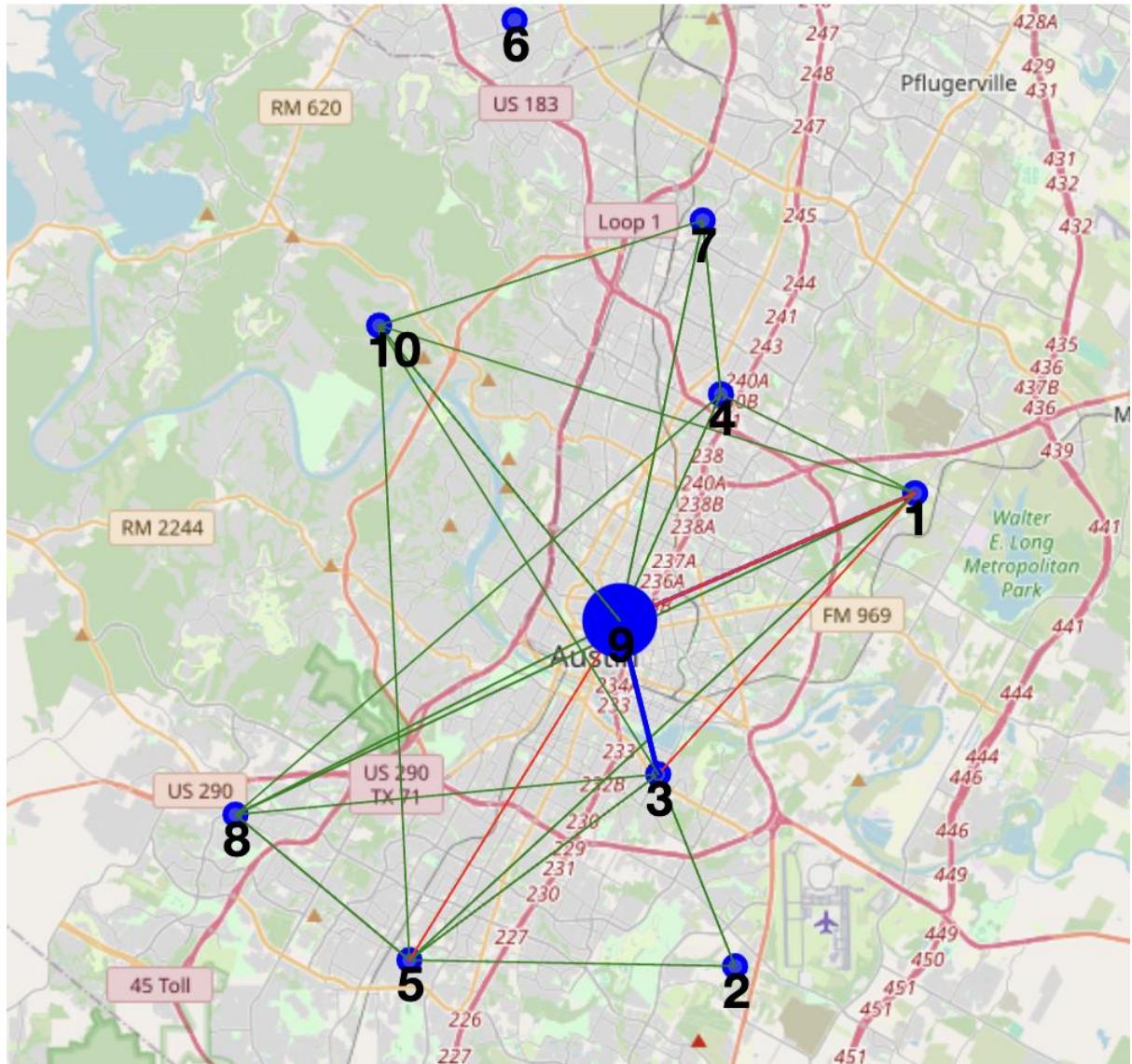


Fig 19. The Austin origin destination flowmap _high Peak seasone:september, high Peak hour:8:00.

Exercise 3 – Relation to Public transport line, Chicago database.

It is recommended to study Exercise 4 first, as the necessary outputs are derived from its results.

Filter the data to consider only one month of data.

Import the public transport lines(as kml files) as one layer on QGIS or ArcGIS, and the Micro-mobility data in csv files as another layer.

Visualising the two layers, record your observations about any patterns on interaction between micro-mobility and public transport.

To analyze the data on Chicago's city map, the base map was first imported into QGIS using Google Maps as a raster file, which also included the wards of the city. Subsequently, subway and bus route maps (Public Transportation lines) were downloaded from the Chicago Open Data Portal and added to the QGIS project as vector layers.

Next, data from the Chicago micro-mobility database, focusing on morning and afternoon peak hours during the highest peak month (July), was imported into the project as four CSV files. “From Latitude”, “From Longitude”, “To Latitude”, and “To Longitude” for trip origins and destinations were then specified in the Layer, Add Delimited Text Layer tab. The CRS also set to EPSG:4326.

Finally, the "Point to Path" tool was used to link the origin and destination points, creating straight-line paths between them. These flow lines were visualized within QGIS, enabling a detailed analysis of micro-mobility patterns across the city.

Figure 20 illustrates micro-mobility trips conducted at 8:00 AM during the busiest months. As observed, the number and density of these routes are notably high, particularly in the city center, where the concentration is even greater. This pattern aligns with the findings from Section 4, where the analysis of OD matrices also highlighted similar trends.

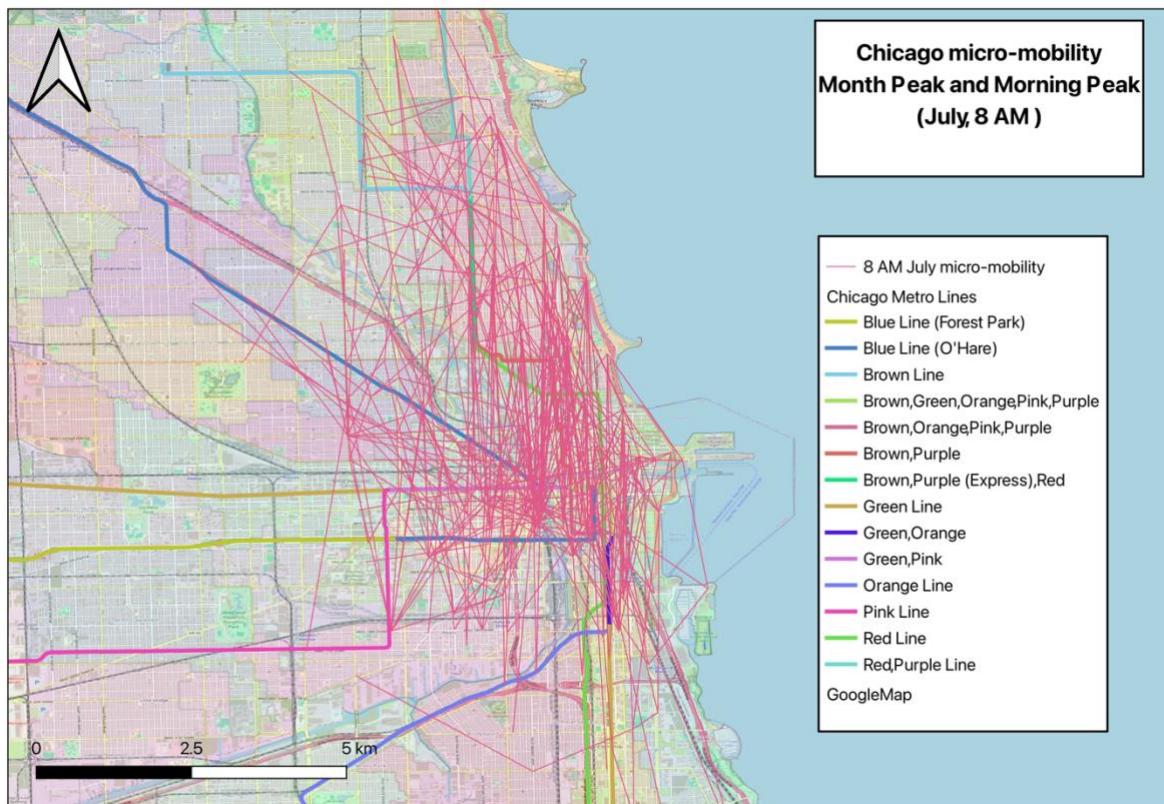


Fig 20. Micro-mobility vs Public Transportatio in Chicago, High peak morning, July, 8:00 AM.

As shown in Figure 21, transportation activity is highly concentrated in the city center. The metro lines also follow similar routes in this area. To avoid overcrowding the map, bus stations and routes have not been included in the visualization; however, their numbers are also substantial. This highlights the dense and multifaceted nature of urban transit in central areas.

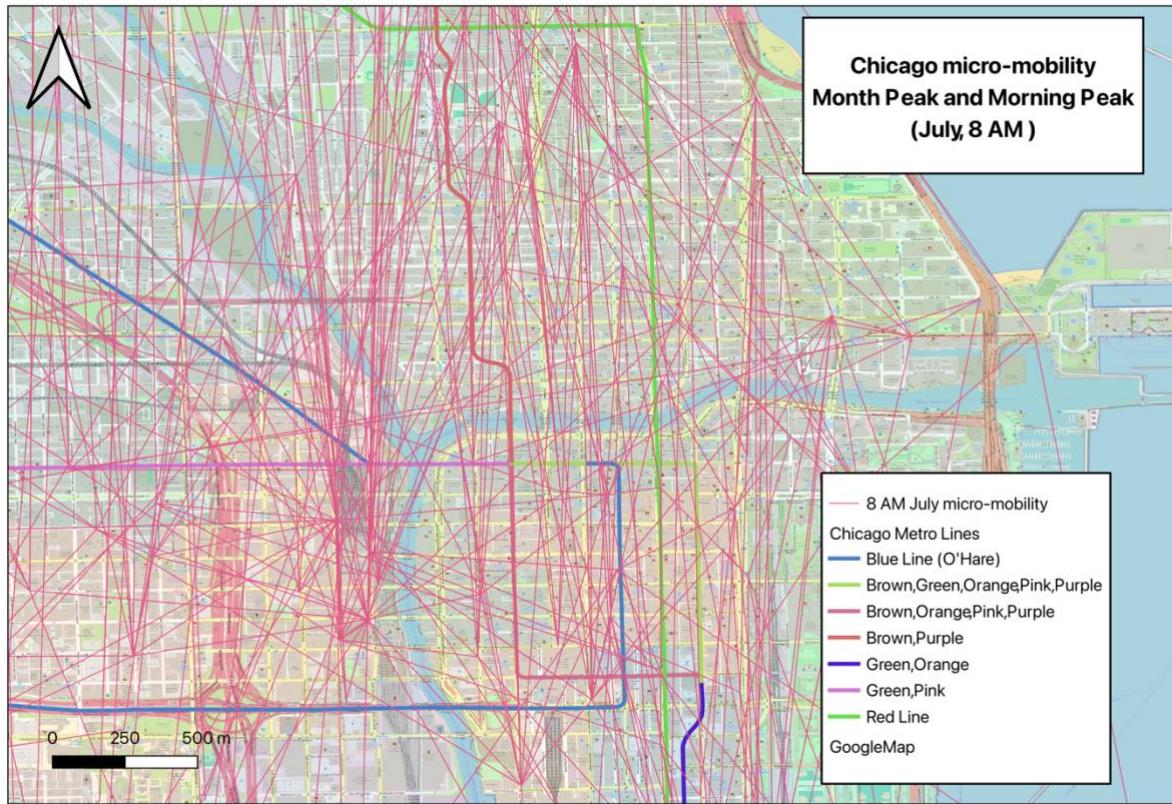


Fig 21. Micro-mobility vs Public Transportatio in Chicago, High peak morning, July, 8:00 AM, lower scale.

Figure 22 illustrates micro-mobility activity during the evening peak hours. The number and density of these routes are higher compared to the morning data. Another notable difference is the broader dispersion of routes, extending more significantly towards peripheral neighborhoods in addition to the city center. This pattern may indicate that many individuals are commuting back to their homes in residential areas located farther from commercial hubs.

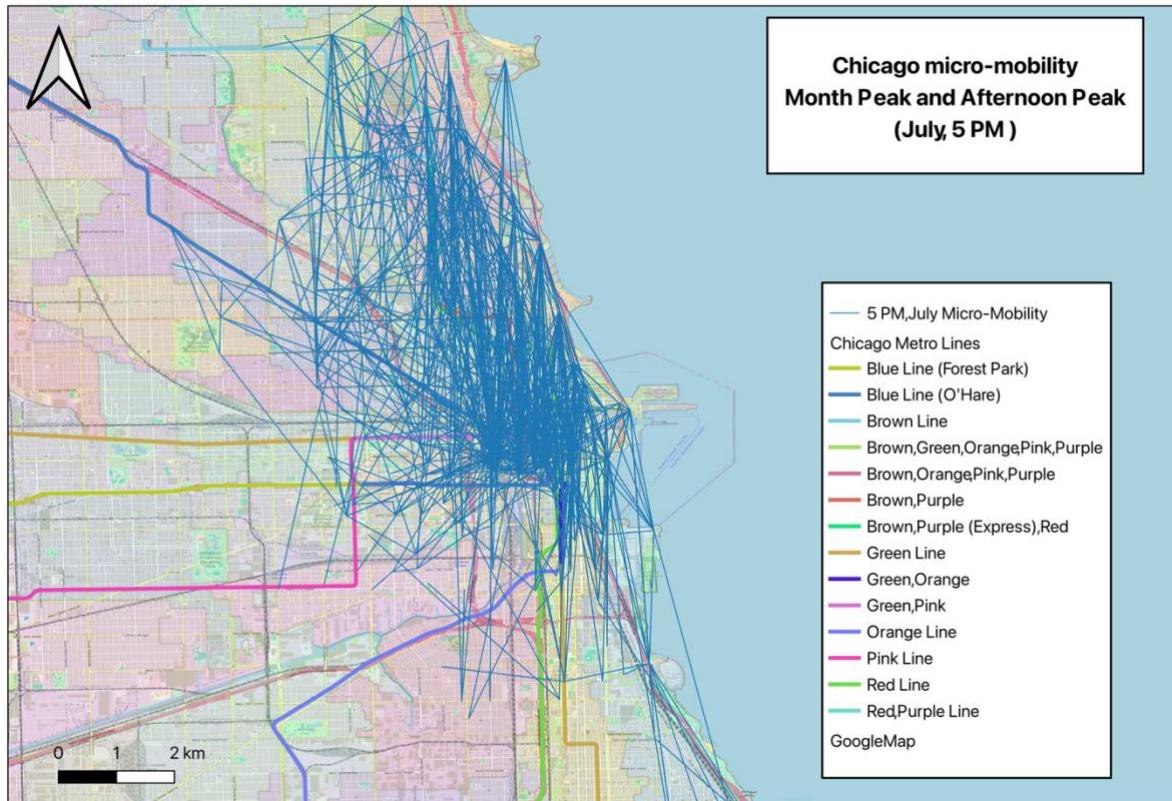


Fig 22. Micro-mobility vs Public Transportatio in Chicago, High Peak afternoon, July, 5:00 PM.

A key question arises: **does micro-mobility in Chicago show a positive contribution to transportation?**

Answering this question with the available limited data is challenging. As we will see in subsequent figures, the city boasts an extensive public transportation system, with a large number of bus routes and numerous stations. If public transit is overcapacity during peak hours and these micro-mobility trips correspond to that excess demand, then shared micro-mobility solutions likely play a supportive role in easing the load on public transport.

However, if public transit is not at capacity, these trips may instead be diverting passengers from public transit—a phenomenon that could be detrimental to the city's overall transportation efficiency. This distinction underscores the importance of perspective. Viewing these data from the standpoint of urban planners aiming to optimize mobility might lead to different conclusions compared to a micro-mobility company assessing profitability. This dual perspective can significantly influence how the impact of micro-mobility is interpreted.

Another important observation which shown in fig. 23 is that the number of trips in the city center is significantly higher than in more distant areas. This is further demonstrated by the longer parking durations in these peripheral neighborhoods. As a result, the micro-mobility provider may focus on expanding bike parking facilities in the city center to increase profitability. However, as mentioned in the previous paragraph, the city center likely already has adequate public transportation options, such as buses and metro lines, which raises questions about the optimal allocation of micro-mobility resources.

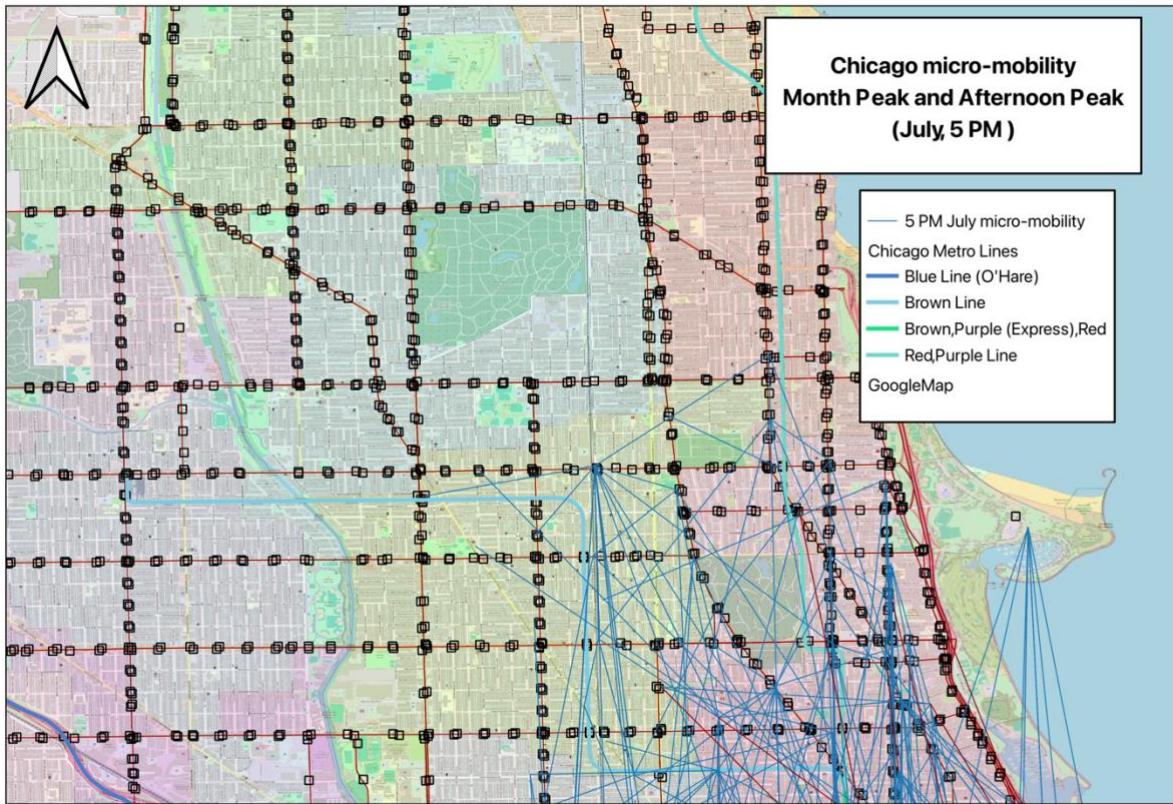


Fig 23. Micro-mobility vs Public Transportatio in Chicago, High Peak afternoon, July, 5:00 PM, zoom to harbor.

It is also positive note that, for accessing certain areas of the city, using bicycles has proven to be highly effective. Figure 24 shows a port that is only accessible by bike and scooter. This highlights the role of micro-mobility in providing connectivity to locations that may not be easily reached by other modes of transport.

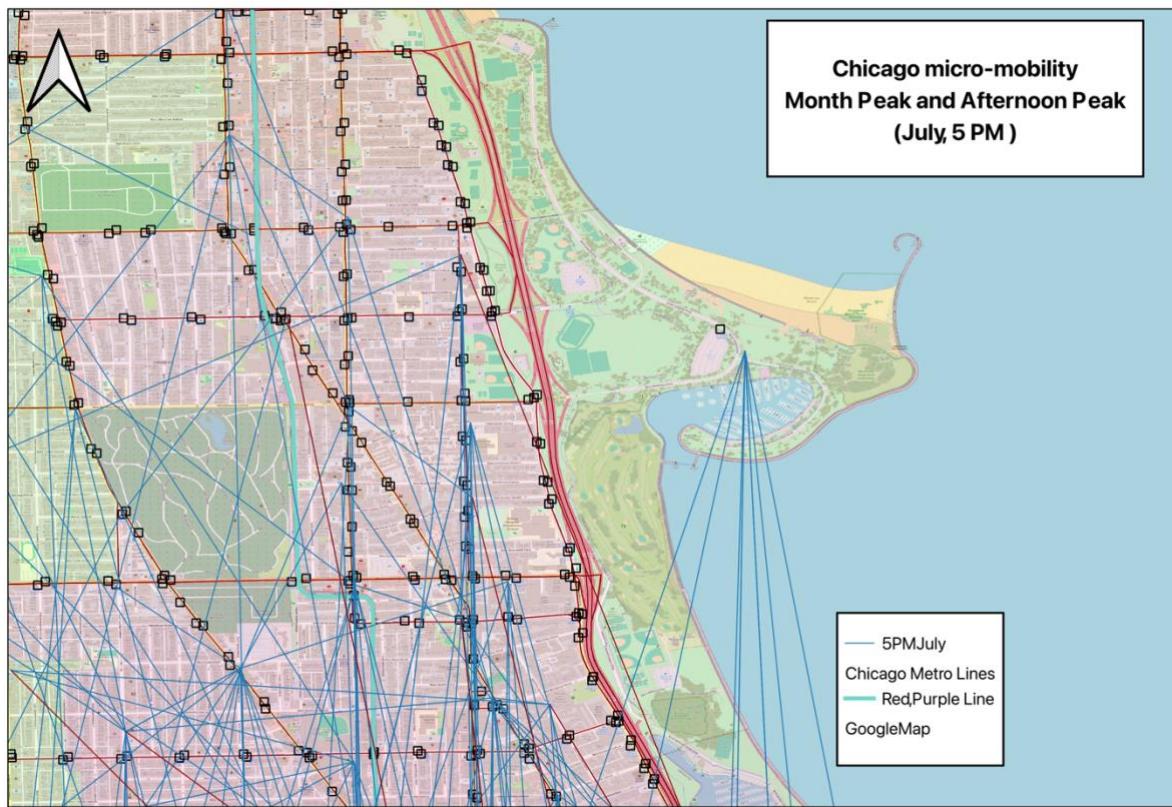


Fig 24. Micro-mobility vs Public Transportatio in Chicago, High Peak afternoon, July, 5:00 PM, lower scale.

Figures 25 and 26 show two key points. First, the 20–30 age group is younger and has a higher ability to use bicycles. Second, this group is typically more active in society due to studies and work commitments. As a result, they exhibit higher trip density in the city center. On the other hand, the 50–60 age group is less prevalent, and their movement patterns extend further into the areas outside the city center.

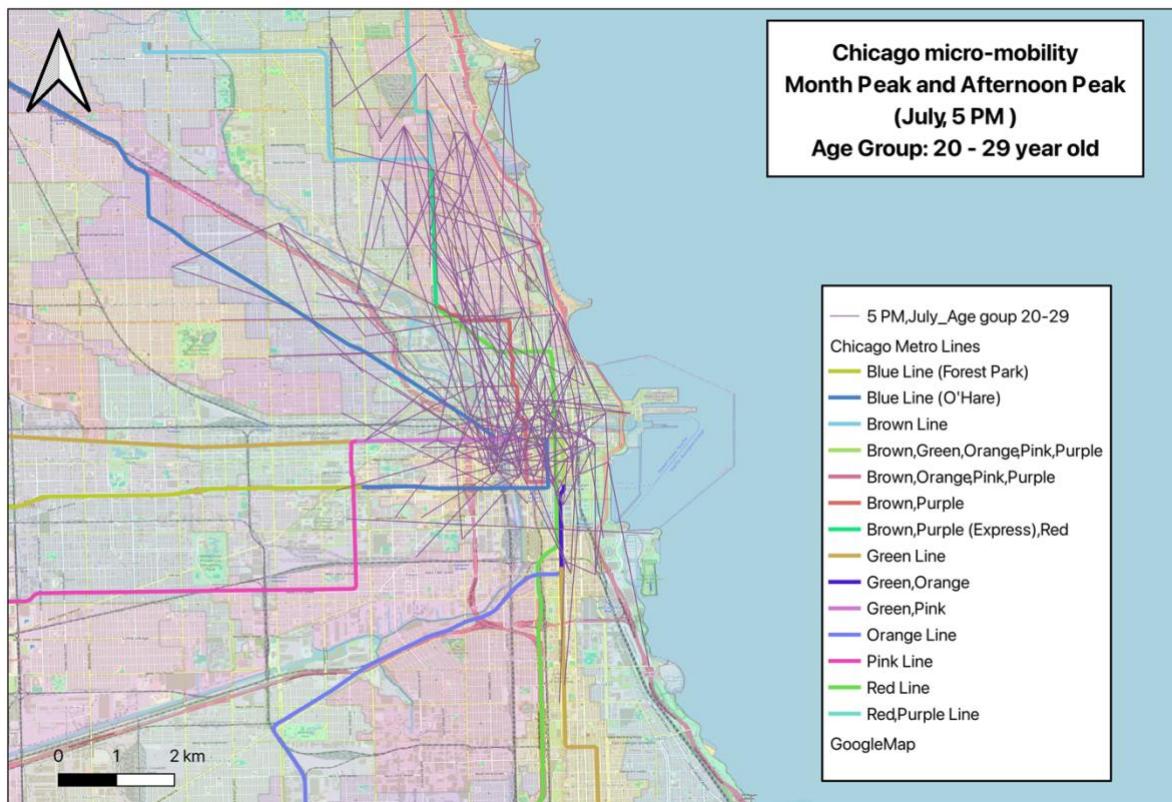


Fig 25. Micro-mobility vs Public Transportatio in Chicago, High Peak morning, July, 5:00 AM_20-29 age group.

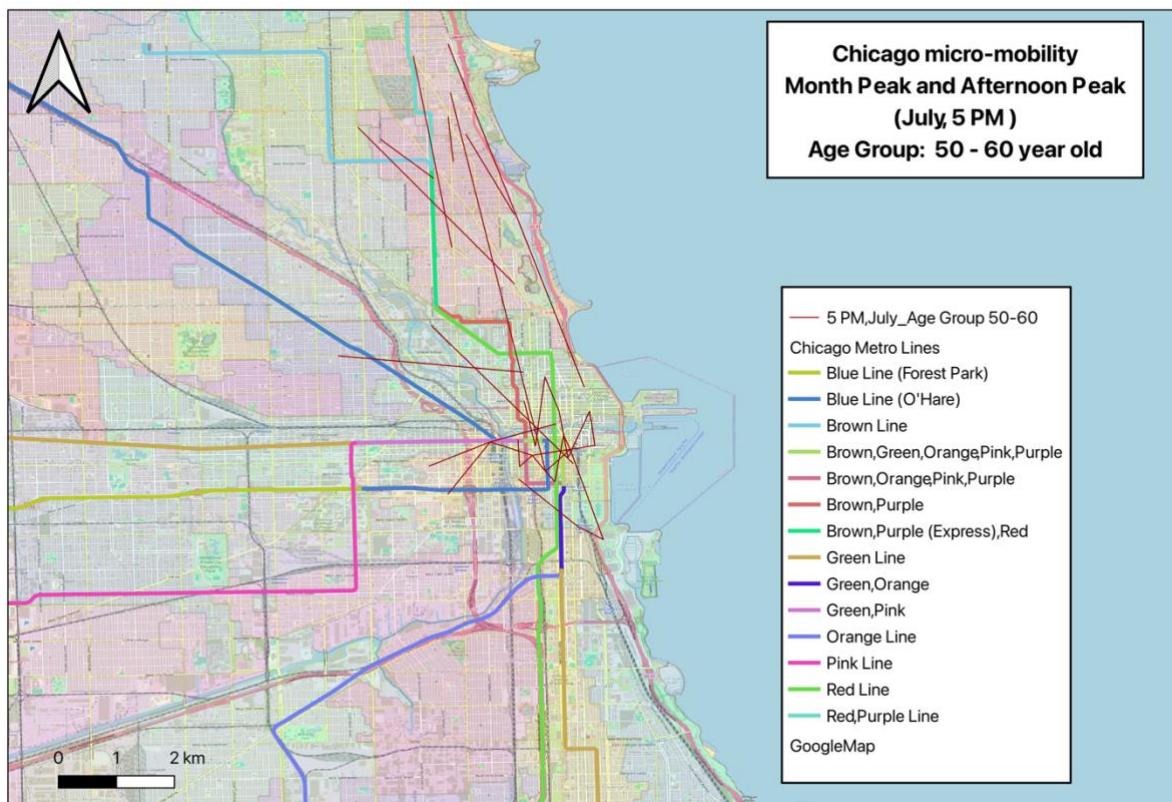


Fig 26. Micro-mobility vs Public Transportatio in Chicago, High Peak afternoon, July, 5:00 PM_50-60 age group.

Exercise 4 – Parking duration and costs

Data Cleaning

The Chicago micro-mobility dataset also underwent a cleaning process to ensure data integrity, consistency, and readiness for in-depth analysis. The following procedures were employed to refine the dataset:

1. Standardization and Type Conversion [EX3_4, Cell 09]

The *BIRTH YEAR* column was converted to a numeric format to facilitate the calculation of users' ages. Missing or invalid entries were replaced with NaN, and the *Age* column was derived by subtracting the *BIRTH YEAR* from the current year.

Date-time columns, including *START TIME* and *STOP TIME*, were parsed into pandas *DateTime* objects using the format `%m/%d/%Y %I:%M:%S %p`. Errors encountered during parsing were set to NaT (Not a Time), ensuring consistency in time representation.

Temporal components such as *START HOUR*, *Month*, *Year*, and *Day of Week* were extracted from the *START TIME* column to enable further analysis of time-based patterns.

2. Handling Missing and Invalid Data [EX3_4, Cell 16]

Several columns, notably *GENDER*, *BIRTH YEAR*, *FROM LOCATION*, and *TO LOCATION*, contained missing values. These gaps were addressed using appropriate imputation methods or, in cases of high data loss, by removing records with incomplete information.

Entries with missing *START TIME* or *STOP TIME* were discarded, as these columns were critical for maintaining the consistency of trip-level data.

Empty strings, NULL values, and NaN placeholders were replaced with `pd.NA` to standardize the dataset and avoid inconsistencies in further analysis.

Code output:	
TRIP ID	0
START TIME	4
STOP TIME	5
BIKE ID	5
TRIP DURATION	5
FROM STATION ID	5
FROM STATION NAME	5
TO STATION ID	5

TO STATION NAME	5
USER TYPE	5
GENDER	589511
BIRTH YEAR	589550
FROM LATITUDE	5
FROM LONGITUDE	5
FROM LOCATION	5
TO LATITUDE	5
TO LONGITUDE	5
TO LOCATION	5
Age	589550
START HOUR	4
Month	4
Year	4
Day of Week	4

3. Trip Duration Adjustments [\[EX3_4, Cell 09\]](#)

The *TRIP DURATION* column, originally in string format with commas, was cleaned and converted to a numeric format. The duration was subsequently converted from seconds to minutes to improve interpretability and facilitate analysis.

4. Outlier Detection and Removal [\[EX3_4, Cell 108\]](#)

Outliers in the *TRIP DURATION* column were identified using both the Interquartile Range (IQR) and Z-Score methods. By cross-referencing the results from both methods, 54 rows were identified as common outliers in both, which were then removed from the dataset. These identified outliers were saved separately in a “.csv” file for further inspection. The removal of outliers helped ensure that extreme values did not unduly influence subsequent analysis.

Spatial Analysis and Geo-Referencing [\[EX3_4, Cell 22\]](#)

Spatial analysis was conducted to examine geographical patterns of micro-mobility trips within Chicago. The following steps were carried out:

1. Importing and Preprocessing Ward Data

Ward boundary data for Chicago was imported from a CSV file containing 50 records with 8 attributes. The *the_geom* column, representing ward boundaries in Well-Known Text (WKT) format, was converted to Shapely MultiPolygon objects for effective spatial operations and analysis.

2. Geo-Referencing Trip Data:

The *FROM LOCATION* and *TO LOCATION* columns, which contained WKT Point geometries, were converted into Shapely Point objects. GeoDataFrames (*df_from_gdf* and *df_to_gdf*) were created to represent these locations spatially, ensuring that the data was ready for geospatial analysis.

The Coordinate Reference System (CRS) for all GeoDataFrames was standardized to EPSG:4326, a global standard, to ensure compatibility across various spatial operations.

3. Spatial Joining:

A spatial join was performed to assign each trip's start and end location to the corresponding ward using the *wards_gdf* dataset. New columns, *FROM WARD* and *TO WARD*, were added to the trip dataset, allowing for detailed spatial analysis of the trips.

Data Distribution Analysis

The dataset was analyzed for patterns related to user demographics, trip frequency, and spatial distribution. Key findings include:

1. Age Distribution [EX3_4, Cell 61]

The age distribution in the Chicago micro-mobility database shows significant differences in usage across age groups. The largest number of users belongs to the 30–39 age group, with 348,766 entries, followed by the 20–29 age group with 335,118 entries, indicating the high popularity of micro-mobility solutions among young adults and early-career professionals. Usage decreases noticeably with age, with 158,361 entries for the 40–49 age group, 95,365 for the 50–59 age group, and just 26,096 for the 60–69 age group. Surprisingly, the 0–19 age group has very low representation, with only 3,152 entries, despite expectations of higher engagement from younger individuals. Notably, the database contains 589,550 records with missing birth year data, which may influence these trends.

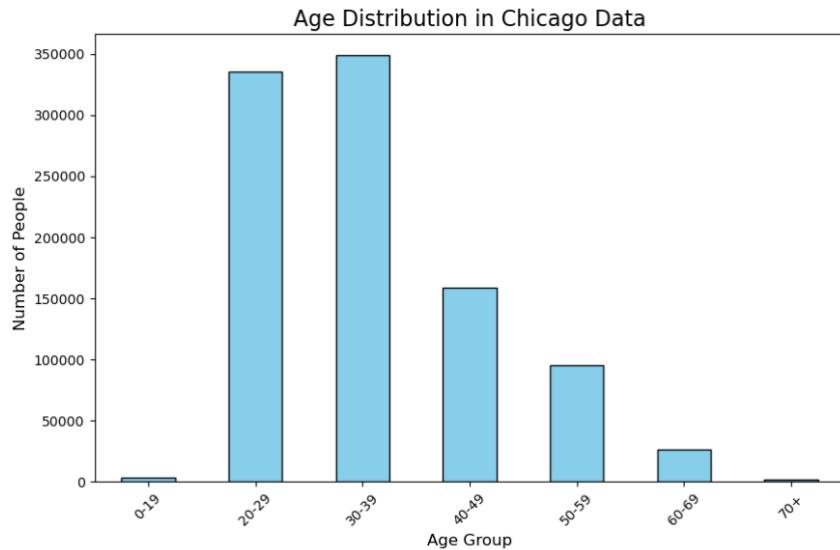


Fig 27. age group distribution.

2. Gender Distribution [EX3_4, Cell 64]

Gender distribution was calculated as a percentage of the total dataset. The analysis highlighted notable gender imbalances, with one gender dominating the dataset. A bar chart was generated, including both absolute counts and percentages, to illustrate these disparities clearly.

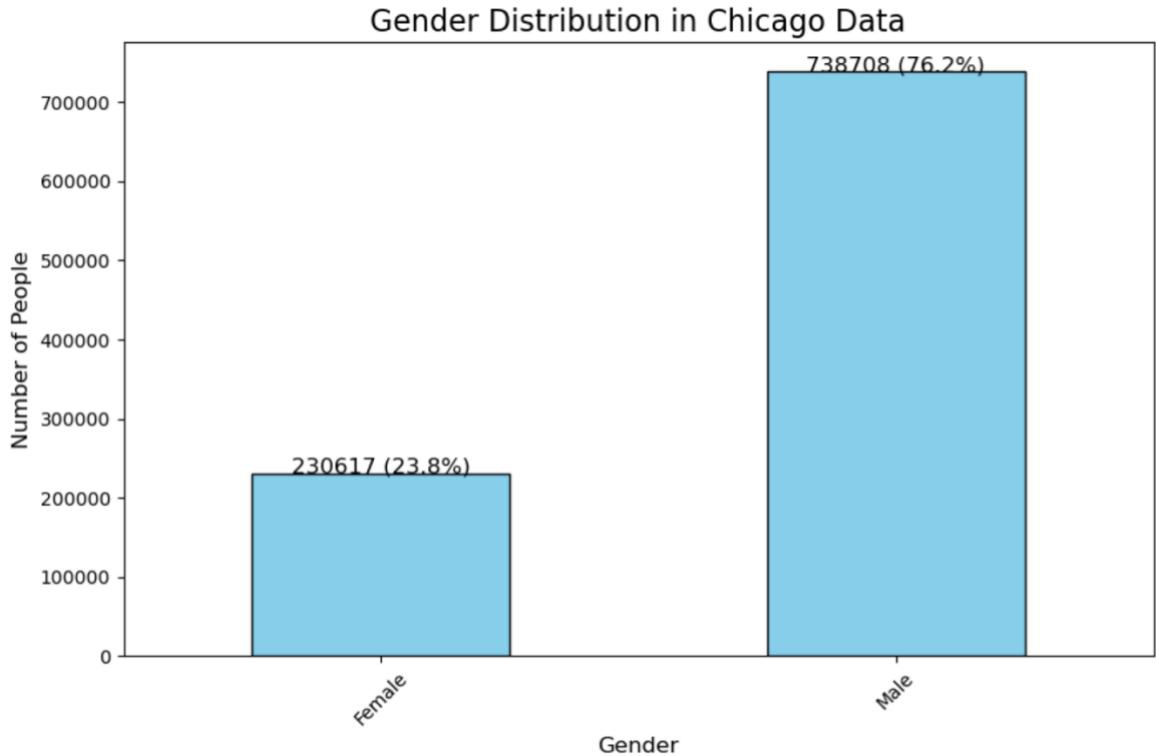


Fig 28. Gender distribution.

3. Yearly Data Distribution [EX3_4, Cell 67]

The data was grouped by year to understand trends over time. The year 2014, which had significant representation in the dataset, was analyzed in depth. A bar chart visualized both the total counts of trips and their corresponding percentages over the years.

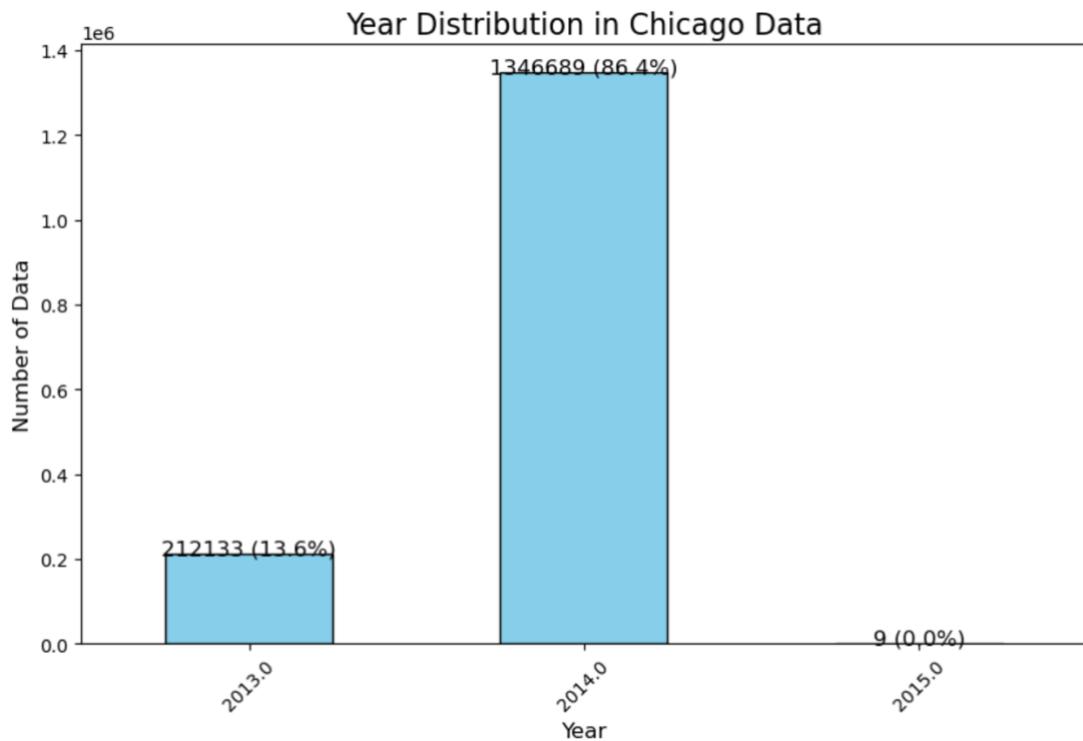


Fig 29. Year distribution.

4. Monthly Distribution in 2014 [EX3_4, Cell 73]

The data shows bike seasonal patterns, with a peak in month 7 (likely summer) and a significant drop in months 8, 11, and 12. Higher usage during warmer months suggests that people use bikes or scooters more when the weather is favorable. The decline in the colder months, especially in November and December, is likely due to less favorable weather and changes in commuting habits, such as reliance on cars or public transit during the winter.

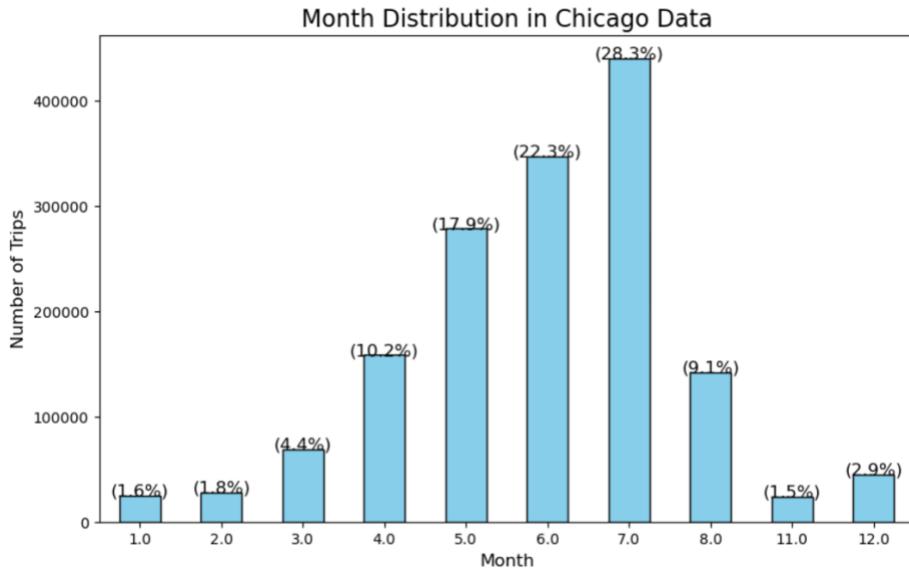


Fig 30. Month distribution in database.

The 2014 dataset was further analyzed by month to uncover seasonal trends in trip frequency. A bar chart was created, representing the distribution of trips across months, highlighting peak and off-peak periods. [EX3_4, Cell 75]

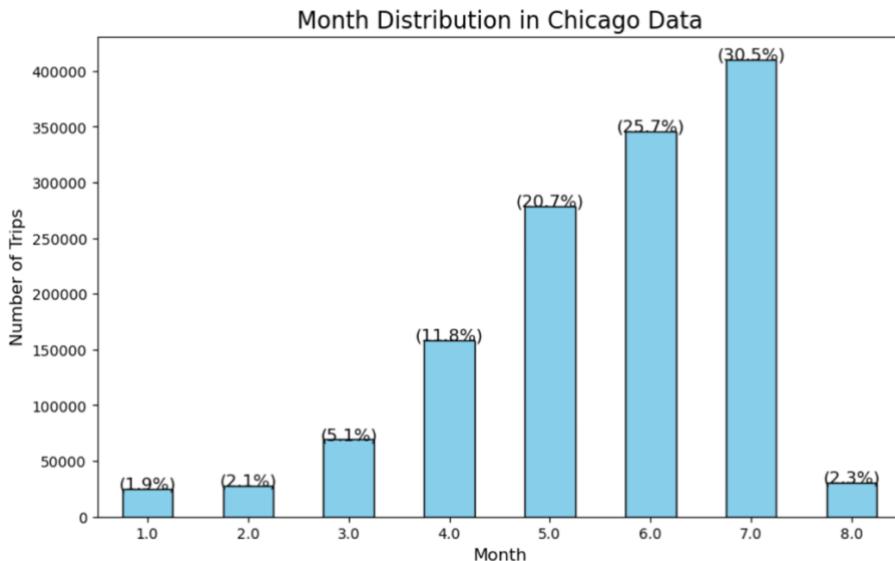


Fig 31. Month distribution in 2024.

5. User Type Distribution [EX3_4, Cell 81]

The *USER TYPE* column was analyzed to understand the distribution of user types. Results indicated that certain user types, such as subscribers, were more prevalent than others, reflecting specific patterns in registration and usage behavior. A bar chart annotated with percentages provided a clear summary. We will focus to these attributes in final steps, the cost secssion.

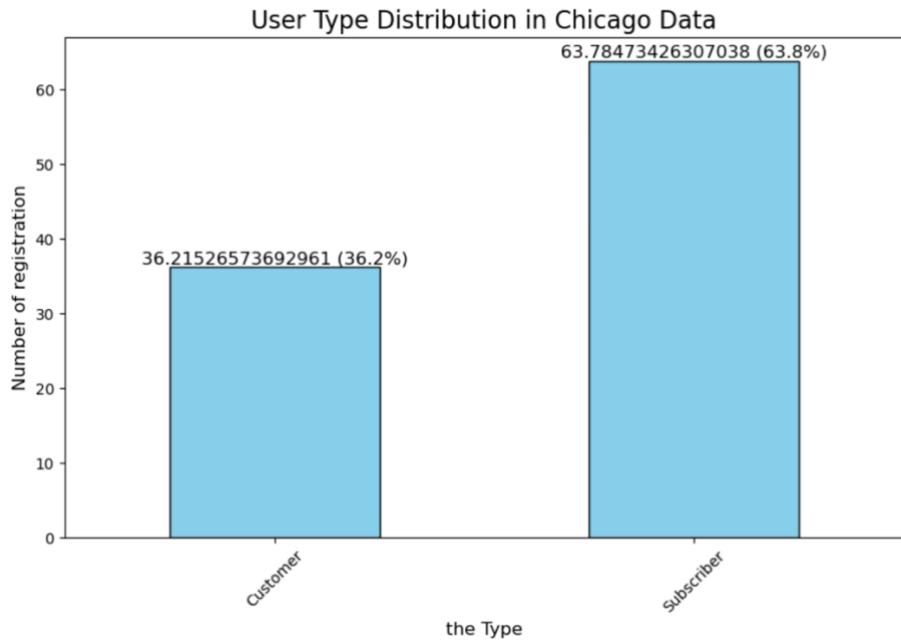


Fig 32. User type distribution.

6. Heatmap Visualization [EX3_4, Cell 84]

Heatmaps were created using the coordinates of trip origins (`from_point`) to visualize the density of trips across Chicago in fig. 29 and trip destination (`To_point`) to visualize the density in fig. 30.

The heatmap, displayed on an interactive Folium map, highlights areas with the highest trip activity.

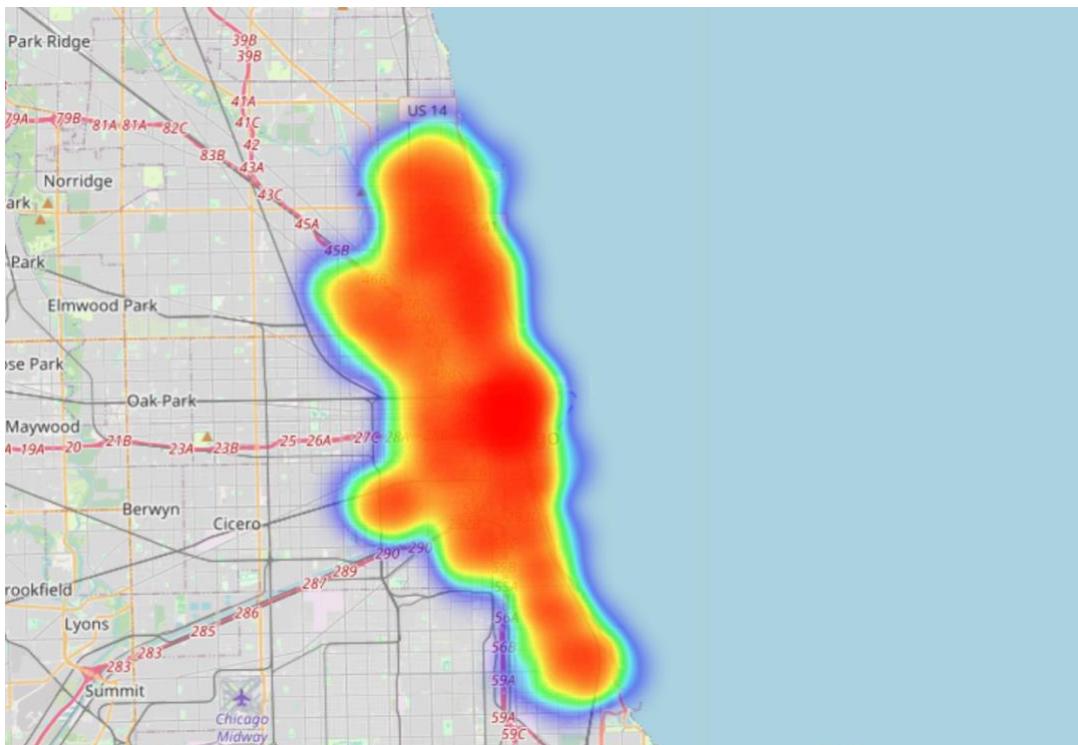


Fig 33. The Origins' heatmap.

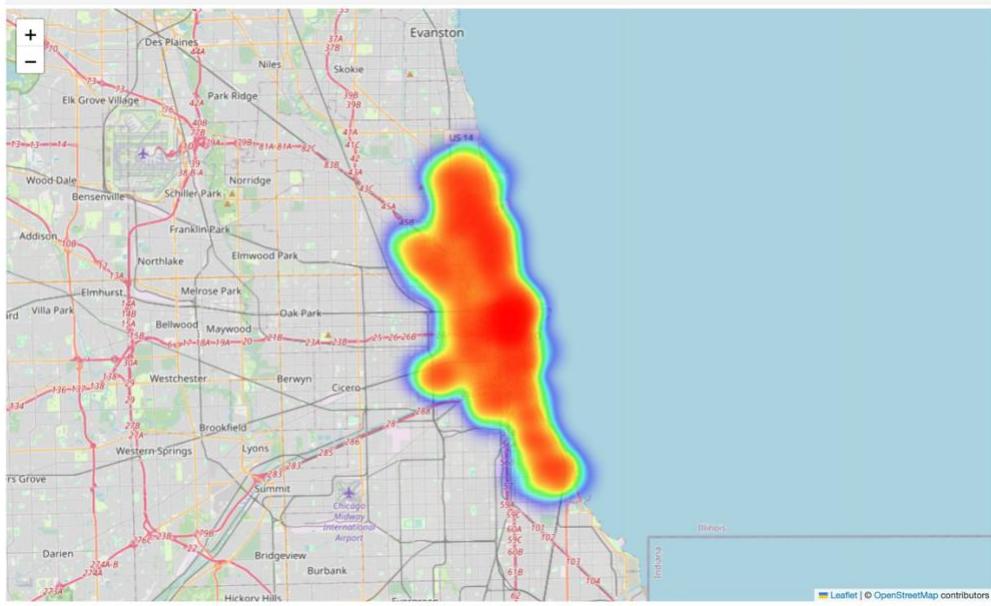


Fig 34. The Destinations' heatmap.

What is the average parking duration of vehicles in different council districts or wards? – Display this info using a map.

Parking Time Calculation [EX3_4, Cell 89]

To analyze the time a bike remains parked between two trips, the following methodology was implemented using the provided code:

1. Data Preparation:

A copy of the cleaned dataset (`df_chicago`) was created to avoid modifying the original data. This ensures that the parking time calculations are isolated and do not interfere with other analyses.

2. Sorting Data:

The dataset was sorted by `BIKE ID` and `START TIME`. Sorting ensures that all trips for each bike are ordered chronologically, which is essential for calculating the parking duration between consecutive trips.

3. Identifying Consecutive Trips:

The `START TIME` of the next trip was determined for each bike using the `.shift(-1)` function within the groupings of `BIKE ID`. This creates a new column, `NEXT START TIME`, that aligns each trip with the subsequent trip's start time.

4. Calculating Parking Time:

Parking time was computed as the difference between the `STOP TIME` of a trip and the `NEXT START TIME` of the following trip. The result, in seconds, was converted into minutes and stored in the `PARKING DURATION` column.

5. Cleaning the Results:

For the last trip of each bike, no subsequent trip exists, resulting in `NaN` values in the `PARKING DURATION` column. These rows were removed to retain only valid parking duration data.

Finding morning and afternoon Peak hours [EX3_4, Cell 104]

To analyze peak usage times for Chicago's bicycle-sharing system, trips were grouped by their start hour to determine the distribution across the day. Using the `groupby` function and the `.size()` method, the dataset was categorized by hour, counting the number of trips. A bar plot was created with start hour on the x-axis and trip count on the y-axis to visualize activity patterns clearly. The analysis revealed two significant peaks: a morning peak between 7:00 AM and 9:00 AM, peaking at 8:00 AM with 84,670 trips, and an evening peak between 4:00 PM and 6:00 PM, with the highest activity at 5:00 PM totaling 156,617 trips. These peaks correspond to traditional commuting hours, highlighting the system's vital role in daily transportation. Periods of minimal activity, such as midnight to 5:00 AM and after 9:00 PM, align with reduced urban activity.

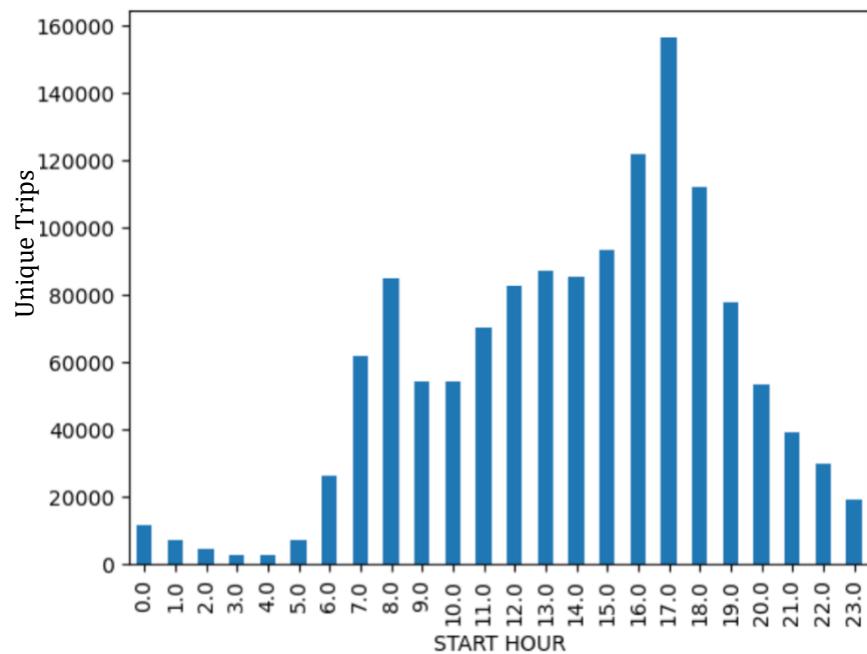


Fig 35. Number of trips accrued in each hour.

Factors influencing these trends include work and school schedules, station distribution, cycling infrastructure, weather conditions, and costs. For example, favorable weather increases ridership, while adverse conditions reduce it. Additionally, areas with concentrated commercial and administrative activities see higher demand during working hours. Insights from this analysis emphasize the importance of aligning resource allocation and maintenance with peak usage hours, supported by spatial analysis and predictive modeling to optimize

station placement and system efficiency. This ensures better integration into Chicago's urban mobility network.

In the Austin data, we observed that the usage trend throughout the day was very different from that of Chicago. One reason we believe for this is the colder climate of Chicago, along with its unique geographical characteristics and urban structure, which differ from the warmer region of Texas. These factors likely influence how people use micro-mobility services, leading to different patterns in each city.

Do an overlapping visualisation with OD matrices and average parking durations across the different parts of the city for different hours of the day.

The Visualization of Parking Duration / The parking locations with most duration [EX3_4, Cell 113]

Before any further analyzes, it is crucial to study the Chicago city better. Based on ChatGPT, In Chicago, the concept of "city centers" typically refers to areas of economic, cultural, or administrative importance. Key city center wards include The Loop (Ward 42), the central business district known for its financial and commercial activity, along with theaters, museums, and parks; the Near North Side (Wards 2, 3, 43, and 44), which includes areas like River North and the Gold Coast, known for high-end retail and residential areas along the Magnificent Mile; the West Loop (Ward 27), a rapidly gentrifying area popular for restaurants, tech startups, and commercial spaces; the South Loop (Ward 4), which features a mix of residential, cultural, and business centers including Soldier Field and Museum Campus; Lincoln Park (Ward 43), known for its parks, the Lincoln Park Zoo, and DePaul University; and Uptown (Ward 46), an area with dense residential, commercial, and entertainment offerings, as well as diverse cultural institutions. These wards are central in business, culture, and tourism, forming the core of Chicago's urban life.

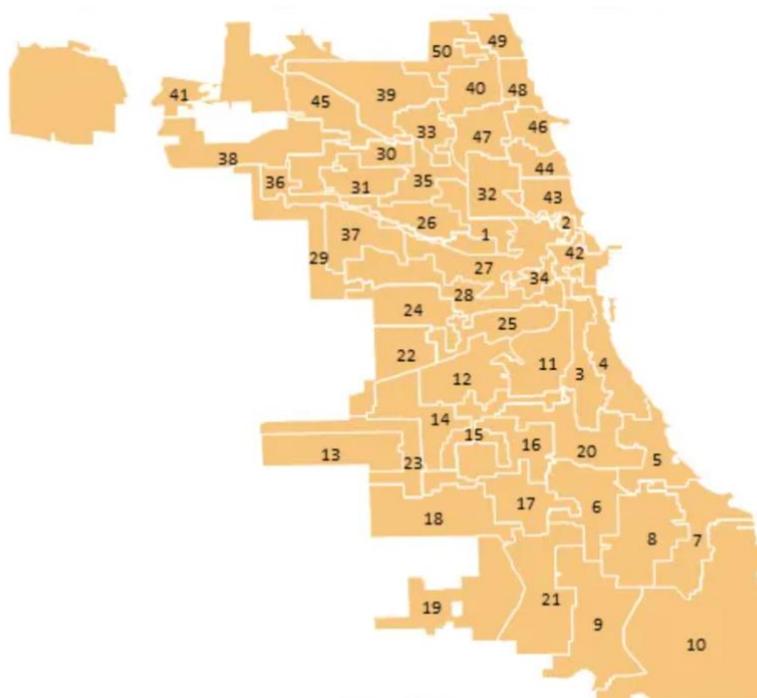


Fig 36. Wards in Chicago.

The dataset is first grouped by the "TO WARD" to calculate the average parking duration, as well as the average latitude and longitude of each ward. The dataset is then cleaned and renamed for clarity. In the next step, a map is created using the Folium library, centered on the city of interest. The locations are marked using circular markers where the color and size of each marker represent the parking duration, with shorter durations marked in green, longer durations in red, and moderate durations in orange. The size of each circle is determined by the average parking duration, capped at a maximum size to maintain a clear visual scale.

The average parking duration is categorized into three groups based on the time values. If the average parking duration is less than 700, it is considered short, and the color green is assigned. If the average parking duration is greater than 1500, it is categorized as mid-range, and the color red is used. For values between 700 and 1500, the parking duration is classified as long, and the color orange is applied. The results indicate that bikes were used more frequently and parked for shorter durations as they got closer to the city center.

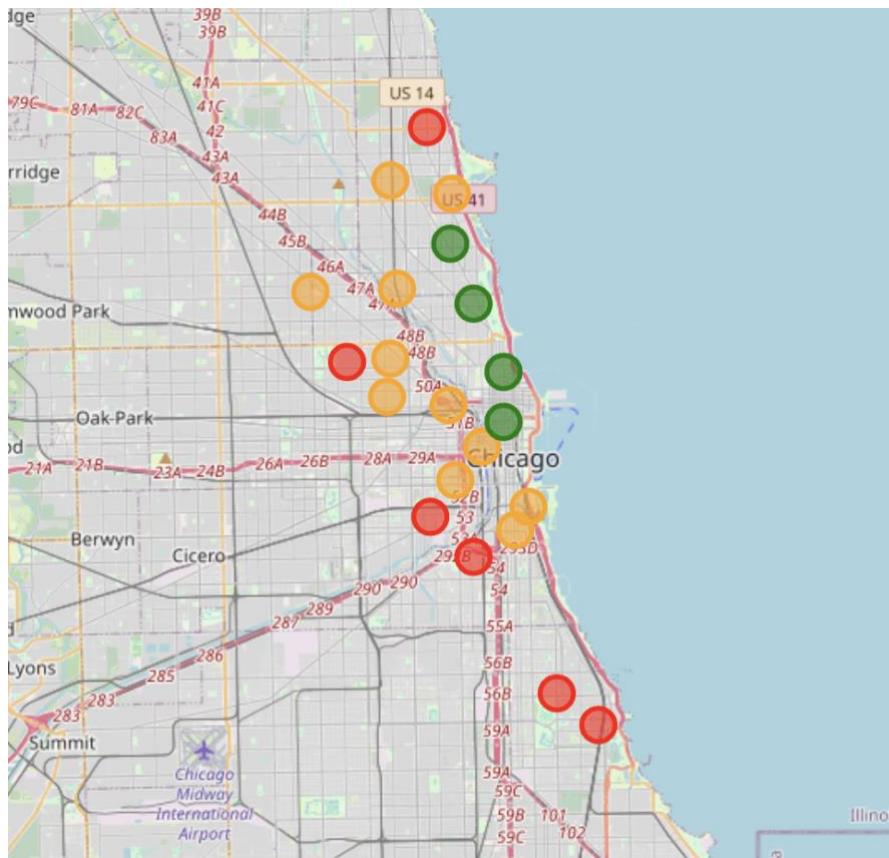


Fig 37. Parking duration in each ward.

To enhance the visualization, the Folium map includes interactive features like the Fullscreen plugin for better map exploration. The result is a dynamic map where each ward's average parking duration is represented, providing insights into areas with short, medium, or long parking times. By adding the OD matrix data for different hours of the day, the map can also highlight how parking behavior varies across the city at different times, offering a more detailed understanding of traffic flow and parking demand. The overall shape of the surface

suggests that parking demand is not uniformly distributed across the city. There are certain areas (e.g., the yellow peak) where parking demand is concentrated, while other areas have lower demand. [EX3_4, Cell 136]

3D Surface of Average Parking Duration by Start Hour and from Ward

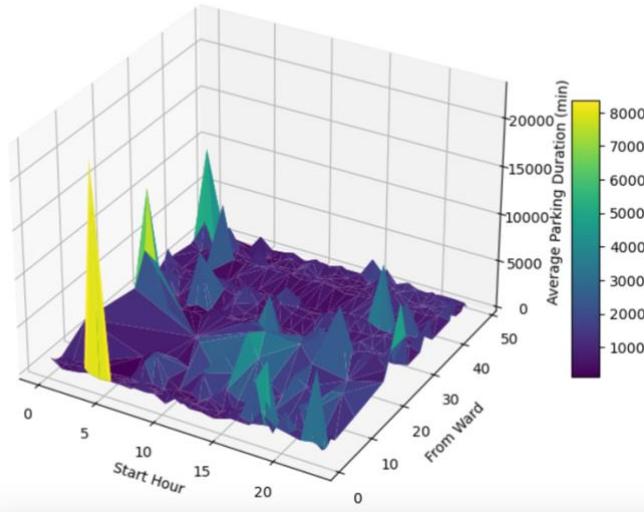


Fig 38. OD matrix and Parking duration, 3D view, all data.

The chart reveals distinct patterns in average bike parking duration throughout the day. As shown in figure 31, two significant peaks occur during 7-9 AM and 4-7 PM, corresponding to peak commuting hours when individuals typically begin and end their work or school day. In contrast, 6 AM marks the lowest parking duration, as this time is primarily associated with active commuting rather than extended parking. A decline in parking duration is observed between 10 AM and 4 PM and during late-night hours, reflecting reduced urban activity. Interestingly, the longest parking durations occur between midnight and 4 AM, a pattern consistent with nighttime rest periods when bike usage is minimal. These trends highlight the interplay between urban activity cycles and bike-sharing system utilization. [EX3_4, Cell 116]

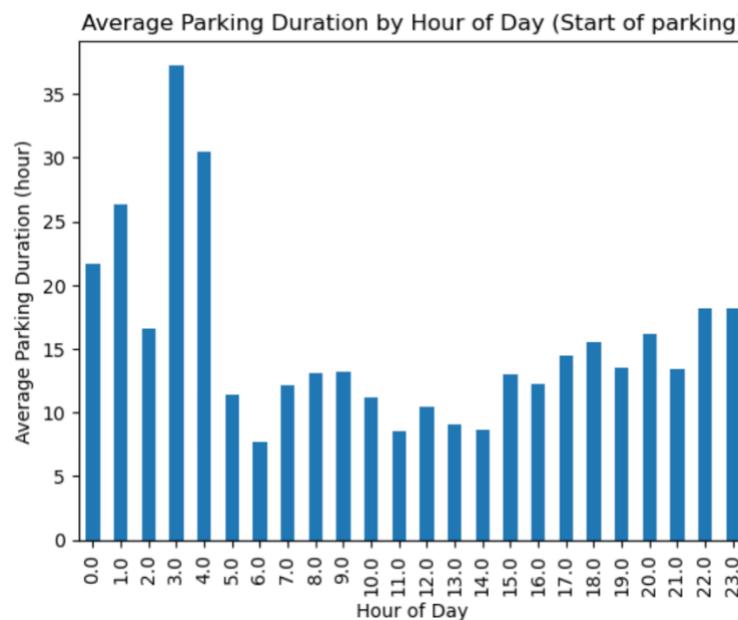


Fig 39. Parking duration during the day.

Comprehensive Analysis of Average Bike Parking Duration in Different Areas of Chicago [EX3_4, Cell 120]

The chart presents the average parking duration for shared bikes across various wards in Chicago, with Ward 20 displaying a significantly higher mean of 55.2 hours. Ward 20 might have some places for bicycle storage or repair locations. Additionally, This outlier suggests the presence of a distinct parking facility or a higher density of shared bike usage within this ward, potentially associated with major transportation hubs or recreational zones, where bikes are left parked for prolonged periods. Wards 5, 11, 25, and 26 exhibit higher average parking durations, ranging from 32 to 39 hours, likely due to the concentration of commercial, educational, or recreational infrastructure, which fosters longer bike usage for activities such as commuting, study, or leisure. Conversely, wards 2, 42, and 43 show comparatively shorter parking durations (approximately 8 to 9 hours), which may be attributed to factors such as enhanced access to public transit, shorter travel distances, and more transient, short-term use of shared bikes.

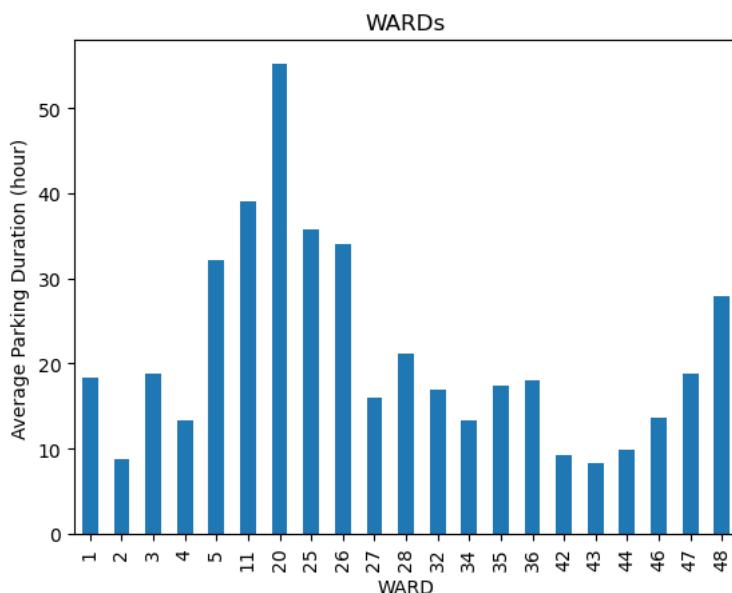


Fig 40. Average Parking duration for the wards.

The OD Matrix in Peak month, Peak Morning Hour (July, 8 am) [EX3_4, Cell 142]

The Origin-Destination (OD) Matrix for 8 AM in July reveals distinct patterns in the flow of bike-sharing trips across different wards in Chicago. High-traffic wards like Ward 42 and Ward 34 emerge as central hubs, with Ward 42 receiving 145 inbound trips and maintaining strong connections with Ward 34 (43 trips). This significant activity likely stems from urban factors such as the presence of commercial centers, office districts, or major transit hubs, which position these wards as pivotal nodes in the transportation network.

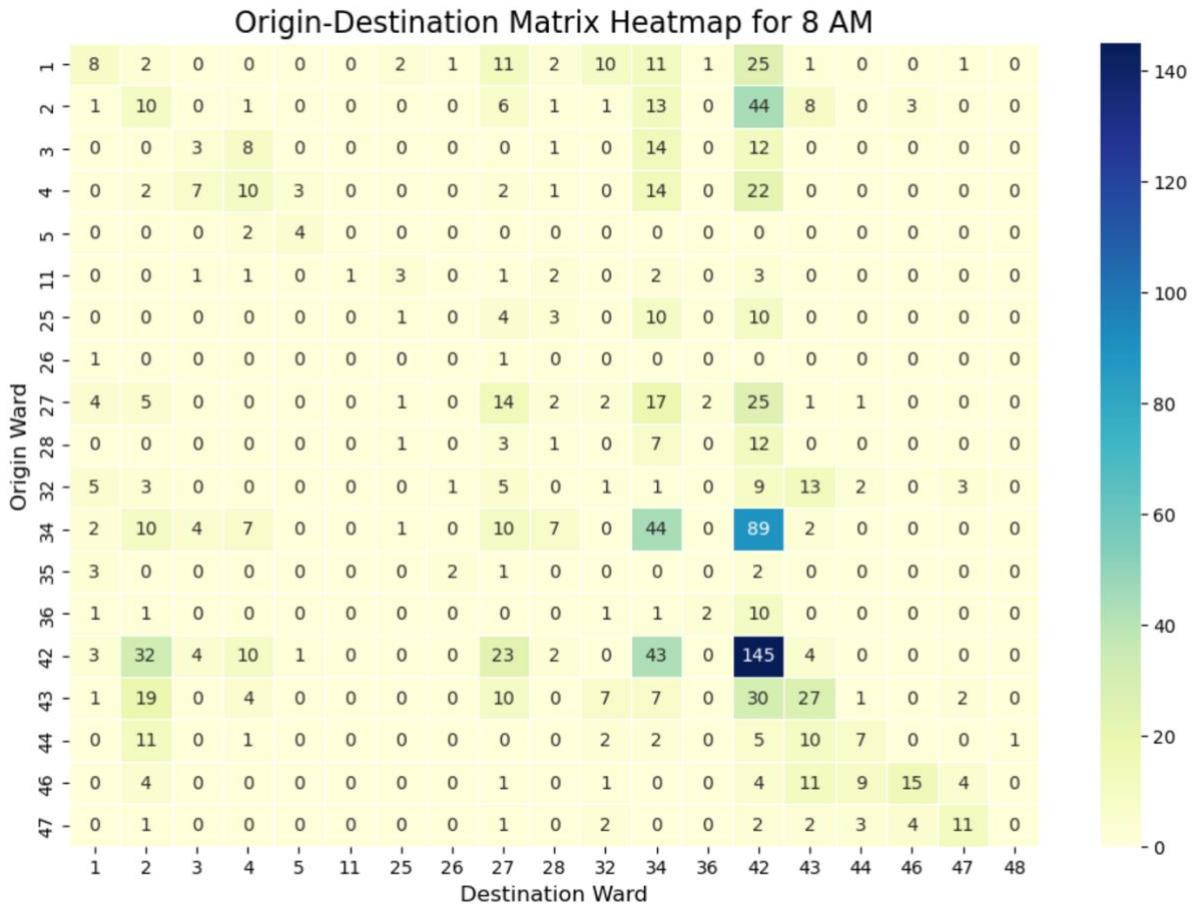


Fig 41. OD matrix, Peak month, Peak Hour (July, 8 am).

Conversely, low-traffic wards such as Ward 5 and Ward 35 exhibit minimal trip activity. This could be attributed to factors such as limited cycling infrastructure, lower population density, or land use characteristics that make these areas less attractive as trip origins or destinations. Such wards may also experience longer parking durations and less efficient utilization of shared bikes.

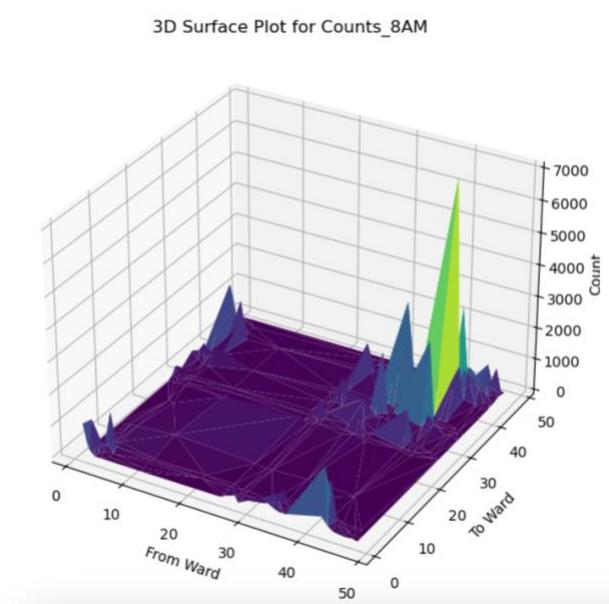


Fig 42. OD matrix, Peak month, Peak Hour (July, 8 am), 3D view.

Intermediate wards, such as Ward 27, demonstrate a balanced mix of local and inter-ward trips. This indicates moderate travel activity, likely driven by well-connected cycling infrastructure or neighborhood-level mobility patterns.

The OD Matrix in Peak Month, Peak Afternoon Hour (July, 5 pm) [EX3_4, Cell 148]

The OD matrix reveals significant disparities in bike-sharing activity across Chicago's wards during the 5 PM peak hour. Certain wards, such as Ward 42 and Ward 34, stand out as critical hubs of activity. Ward 42 exhibits the highest internal flows (231 trips) and strong connectivity with Ward 34 (105 trips) and other surrounding wards. This pattern likely reflects the presence of dense commercial zones, transit hubs, or areas of high employment, driving both internal and inter-ward movements. Similarly, Ward 34 shows substantial activity with Ward 42 (91 trips) and other adjacent wards like Ward 43, emphasizing its role as a secondary hub in the city's bike-sharing network.

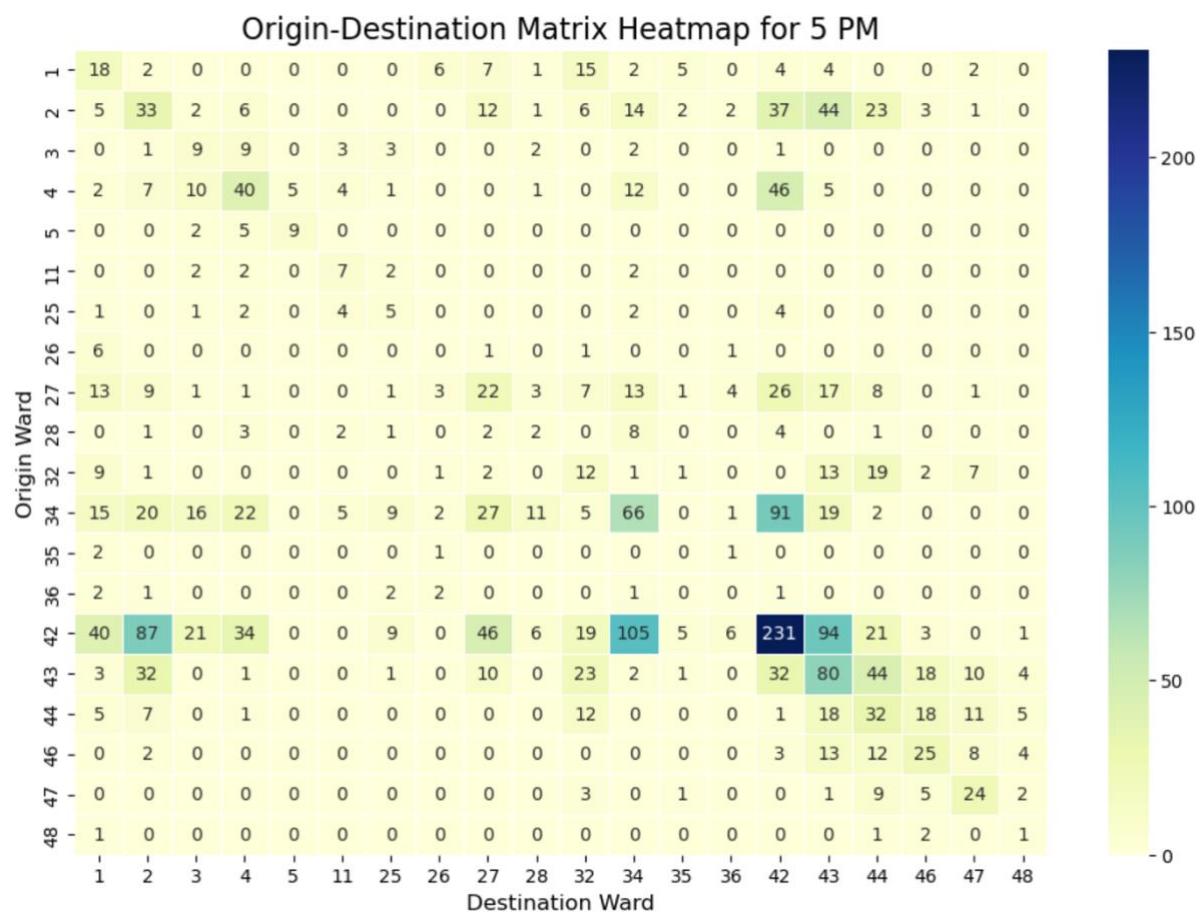


Fig 43. OD matrix, Peak month, Peak afternoon Hour (July, 5 pm).

In contrast, wards like 5, 35, and 36 show minimal bike-sharing interactions, both internally and with other wards. This low activity might stem from limited bike-sharing infrastructure, fewer dedicated bike lanes, or lower population density in these regions. Parking durations in such areas are likely longer, indicating underutilized resources. Conversely, wards with high

trip volumes, such as 42 and 34, likely experience shorter parking durations due to rapid turnover, driven by high demand during peak commuting hours.

Intermediate wards, like Ward 27, show a mix of moderate internal flows and connectivity to key hubs like Ward 42. These patterns suggest localized commuting and neighborhood-level trips, supported by moderately available infrastructure. The overall spatial dynamics captured in the OD matrix highlight the critical interplay between infrastructure placement, socio-economic factors, and urban mobility behaviors.

Figure 40 reveals better the differences.

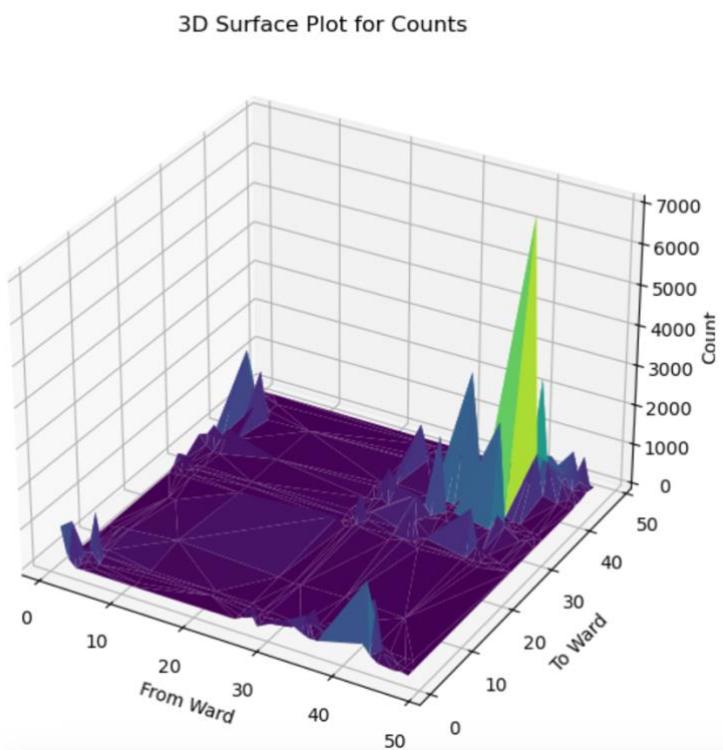


Fig 44. OD matrix, Peak month, Peak afternoon Hour (July, 5 pm), 3D view.

The comparison between the peak morning and peak afternoon reveals distinct differences in travel patterns. During the morning peak, travel demand is more evenly distributed, with high flows observed between wards 42, 43, and 34, suggesting movement toward central or economic areas. In contrast, the afternoon peak shows significantly higher travel volumes, particularly between central wards like 42, 34, and 43, indicating a return commute from work or school. Overall, the afternoon is marked by more concentrated traffic, especially in central areas, while the morning sees a more widespread flow of people heading out from residential wards. These differences highlight the contrasting mobility demands throughout the day and also shown in figures 20 and 22.

Calculate the revenues (different streams of revenue for a micro-mobility company) and try to calculate the costs (both capital and operational) to understand the business model.

Distribution of Vehicles Utilization Percentage [EX3_4, Cell 117]

The analysis of vehicle utilization for the micro-mobility company highlights the overall usage patterns of its fleet over the course of a year. By calculating the total trip duration for each vehicle and comparing it to the maximum possible availability (24 hours a day for 365 days), the utilization percentage provides insight into how frequently each vehicle is used relative to its potential availability. The average trip duration per vehicle is approximately 9775 minutes per year, which translates to about 6.79 full days of use annually. Similarly, the median trip duration is slightly higher at 10053 minutes per year, or approximately 6.98 full days. These values suggest that while vehicles are in use for a substantial portion of the year, their availability is far from fully optimized.

The mean utilization percentage across the fleet is 1.86%, indicating that, on average, each vehicle is being used for less than 2% of its total available time. The median utilization is slightly higher at 1.91%, but it still reflects a very low level of fleet activity. This suggests that while the fleet has potential for more utilization, the majority of vehicles are underutilized throughout the year. This low utilization rate can be a critical factor for the company to consider, as it may imply inefficiencies in fleet management, demand distribution, or operational strategy.

The histogram, fig. 41, of utilization percentage further emphasizes the distribution of vehicle usage across the fleet. It shows that the majority of vehicles fall into the lower range of utilization, reinforcing the observation that only a small proportion of the fleet is being actively used. This information can be useful for identifying areas where operational improvements or strategic adjustments might be necessary to increase vehicle utilization and overall fleet efficiency. The low utilization rates also suggest that the company could explore methods for improving demand forecasting, vehicle redistribution, or promotional efforts to enhance the efficiency of its micro-mobility service.

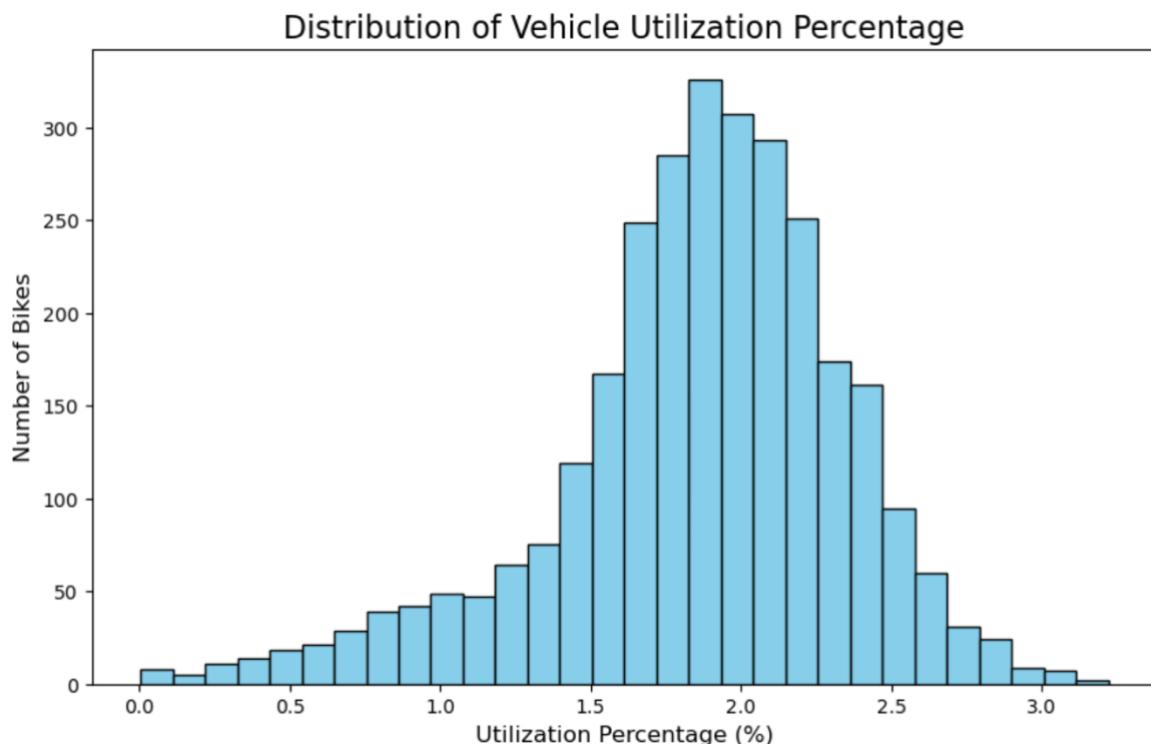


Fig 45. The distribution of the bike usage in percentage.

Monthly Utilization [EX3_4, Cell 120]

The monthly vehicle utilization analysis for the micro-mobility fleet reveals substantial fluctuations in utilization throughout the year. The results show that the average monthly utilization across all months is relatively low, with a mean utilization of 2.01%. The median utilization is even lower at 0.68%, indicating that a significant portion of the fleet is underutilized in most months. This suggests that the vehicles are not being actively used for a large proportion of their available time, reflecting potential inefficiencies in fleet operation or demand distribution.

The monthly breakdown of utilization shows notable peaks in the warmer months, particularly in June and July, where the utilization percentages rise to 5.46% and 6.85%, respectively. These months are likely to correspond with increased demand due to favorable weather conditions, seasonal tourism, or other factors that encourage more frequent use of micro-mobility services. Conversely, the months of September and October show zero utilization, likely indicating either a lack of trips recorded during those months or a significant drop in demand, which could be an important area to investigate further.

The variation in monthly utilization suggests that the company may benefit from more targeted operational strategies, such as increasing vehicle availability during high-demand months or exploring ways to stimulate usage during slower periods. Strategies could include better demand forecasting, seasonal promotions, or introducing flexible vehicle distribution models to ensure more consistent utilization throughout the year. The low overall utilization rate points to room for improvement in the business model, especially in terms of optimizing fleet management and increasing the service's attractiveness to users year-round.

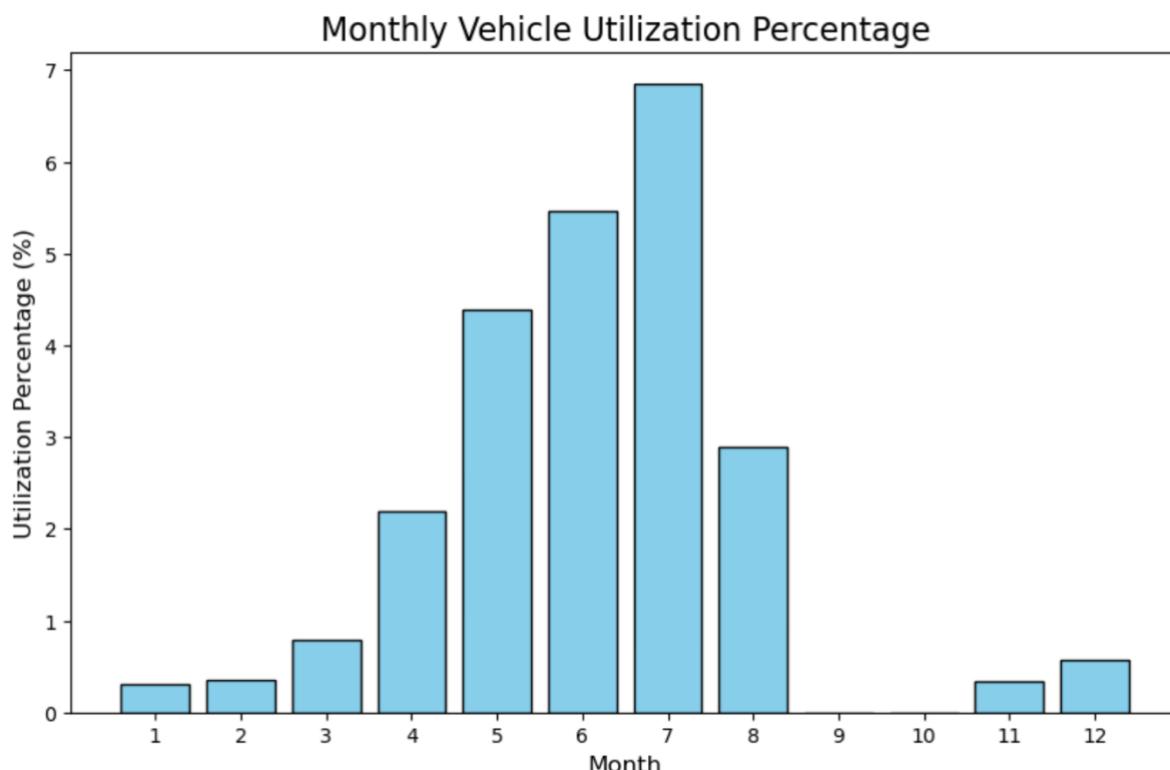


Fig 46. The distribution of the bike usage in percentage for each month in database.

Revenue Calculation [EX3_4, Cell 44]

To estimate the total revenue generated from both customers and subscribers, a simplified model was used due to the lack of user-specific identifiers. Based on the data provided by the official website of Divvy Bikes, customers are assumed to pay \$1 to unlock a trip and \$0.44 per minute for its duration. For subscribers, the annual subscription fee of \$134.90 is divided by an estimated usage pattern—10 trips per week—spread across the year. This calculation considers 12 months, 4 weeks per month, and 10 trips per week, resulting in a prorated cost per trip. Additionally, a per-minute charge of \$0.44 is applied to subscribers' trip durations to align with the dataset structure.

	Single Ride	Day Pass	Divvy	Lyft Pink
	\$1 + \$0.18/min	\$18.10/day	\$143.90/year	\$199/year
	Get the app →	Get a day pass →	Join →	Join →
Classic bike prices	\$1 unlock + \$0.18/min	3 hours free, then \$0.18/min	45 min free, then \$0.18/min	45 min free, then \$0.18/min
Scooter prices	\$1 unlock + \$0.44/min	Free unlocks + \$0.44/min	Free unlocks + \$0.29/min	Free unlocks + \$0.29/min
Ebike prices	\$1 unlock + \$0.44/min	Free unlocks + \$0.44/min	Free unlocks + \$0.18/min	Free unlocks + \$0.18/min

Fig 47. Subscriptions [divvyybikes.com].

The total revenue is computed by summing the revenue contributions from both user types. The code implementation ensures accurate revenue allocation based on the described assumptions. The final calculated total revenue is **\$13,687,624.40**.

Code output:
Total Revenue from Customers and Subscribers: \$13687624.40

Cost Calculation [EX3_4, Cell 49]

To calculate the costs associated with the bike-sharing service, the analysis includes fixed and variable cost components:

1. Fixed Costs:

- **Bike Purchase Costs:** Each bike is assumed to cost \$500, and the total cost is calculated based on the number of unique bikes in the dataset.
- **Bike Maintenance Costs:** An annual maintenance cost of \$50 per bike is applied.
- **Operational Costs:** These include staff salaries (\$200,000 annually) and office infrastructure costs (\$50,000 annually).

The total fixed costs are calculated by summing these components, amounting to **\$1,890,100.00**.

2. Variable Costs:

- **Per Trip Costs:** Each trip incurs a \$1 cost, representing expenses such as customer support and basic trip-related operations.
- **Energy Costs:** A \$0.05 energy cost per trip-minute is considered, calculated based on the total trip duration across all trips. The total variable costs, combining these components, are **\$2,786,093.19**.

The implementation ensures a structured cost breakdown. The calculated total costs provide a clear view of the financial requirements for operating the system effectively.

Code output:
Total Fixed Costs: \$1890100.00
Total Variable Costs: \$2786093.19

The Profit / loss calculation [EX3_4, Cell 52]

Profit Calculation: Process and Outcome

The profit or loss is determined by subtracting the total costs (fixed and variable) from the total revenue (from both customers and subscribers).

- **Total Revenue** is obtained by summing up the individual revenue values calculated for both customers and subscribers.
- **Total Costs** is the sum of fixed costs (including bike purchase, maintenance, and operational costs) and variable costs (associated with each trip, such as energy and customer support).



Fig 48. The profit or loss schematic view.

By subtracting the total costs from the total revenue, the resulting profit is \$9,011,431.21, indicating a profitable outcome for the bike-sharing service. However, this profit is relatively small considering the size of a company like Divvy Bikes. It's important to note that this profit/loss calculation is based on my assumptions, as there is a lack of subscriber data.

Code output:

Profit/Loss: + \$9011431.21