**3. Data Analysis Process**

In tackling this case study, I will utilize the six steps of the data analysis process to address and solve the problem. These six steps consist of:

1. Ask
2. Prepare
3. Process
4. Analyze
5. Share
6. Act

**Step 1: Ask**

My key task involves devising marketing tactics to transition casual riders into Cyclistic members. Specifically, I've been assigned by Lily Moreno, the director of marketing, to delve into the first analysis question: Understanding the distinctions in how annual members and casual riders utilize Cyclistic bikes.

Top of Form

**Step 2: Prepare**

I'll be utilizing Cyclistic's historical trip data (1st quarter data) spanning from January 2022 to March 2022, which is publicly available courtesy of Motivate International Inc. The dataset adheres to the naming convention "YYYYMM-divvy-tripdata," representing data for individual months. Within these files, various details are included such as ride ID, bike type, start and end times, start and end stations, as well as start and end locations. Additionally, information regarding the member status of riders is also provided.

**3. Process**

In order to ensure the cleanliness of the data, I conducted a check to identify any null values in the dataset. The dataset has null values in two columns: 'Member Gender' and '05 - Member Details Member Birthday Year'. All other columns have no missing data.

# Check for null values in the DataFrame

divvy\_trips\_df.isnull().sum()

Following the examination for null values, the next step involves scrutinizing the dataset to identify any duplicate values. The dataset does not contain any duplicate rows.

# Check for duplicate rows in the DataFrame

duplicate\_rows = divvy\_trips\_df.duplicated().sum()

print('Number of duplicate rows:', duplicate\_rows)

Next I will compute the total trip duration, by introducing a new column called "ride\_length" .

Convert the '01 - Rental Details Local End Time' to datetime

divvy\_trips\_df['01 - Rental Details Local End Time'] = pd.to\_datetime(divvy\_trips\_df['01 - Rental Details Local End Time'])

# Calculate the ride length by subtracting the start time from the end time

divvy\_trips\_df['ride\_length'] = (divvy\_trips\_df['01 - Rental Details Local End Time'] - divvy\_trips\_df['01 - Rental Details Local Start Time'])

# Convert the ride length to total seconds for easier analysis

divvy\_trips\_df['ride\_length\_seconds'] = divvy\_trips\_df['ride\_length'].dt.total\_seconds()

# Display the head of the DataFrame to confirm the new column

print(divvy\_trips\_df[['01 - Rental Details Local Start Time', '01 - Rental Details Local End Time', 'ride\_length', 'ride\_length\_seconds']].head())

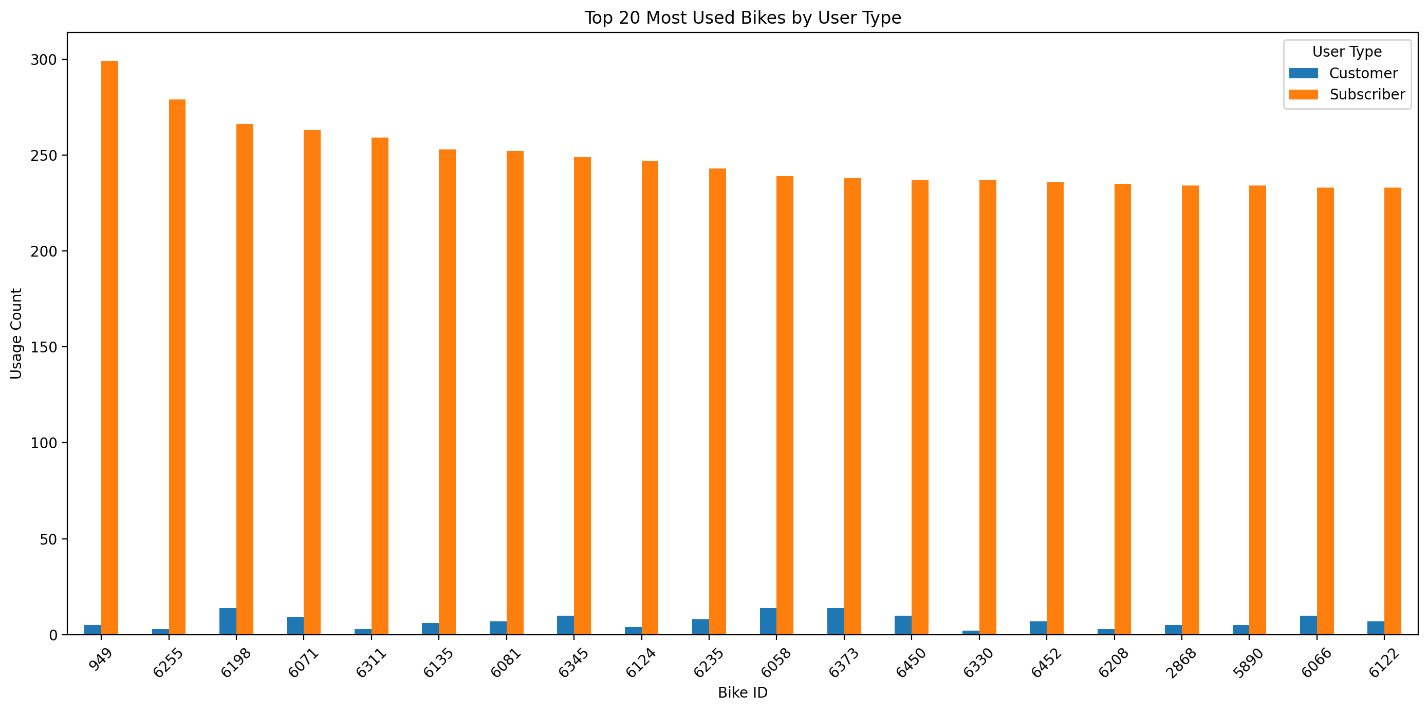
A new column called "ride\_length" has been added to the dataset, representing the duration of each trip. Additionally, the "ride\_length\_seconds" column has been created to express the ride length in seconds for easier analysis. The first few entries have been displayed to confirm the successful addition of these columns.

**Step 4 & 5: Analyze and Share**

The data has been carefully organized and readied for in-depth analysis. Pertinent information was extracted from various tables and subsequently presented visually through Tableau. The key aim of this analysis is to comprehend the unique usage patterns of Cyclistic bikes among annual members and casual riders.

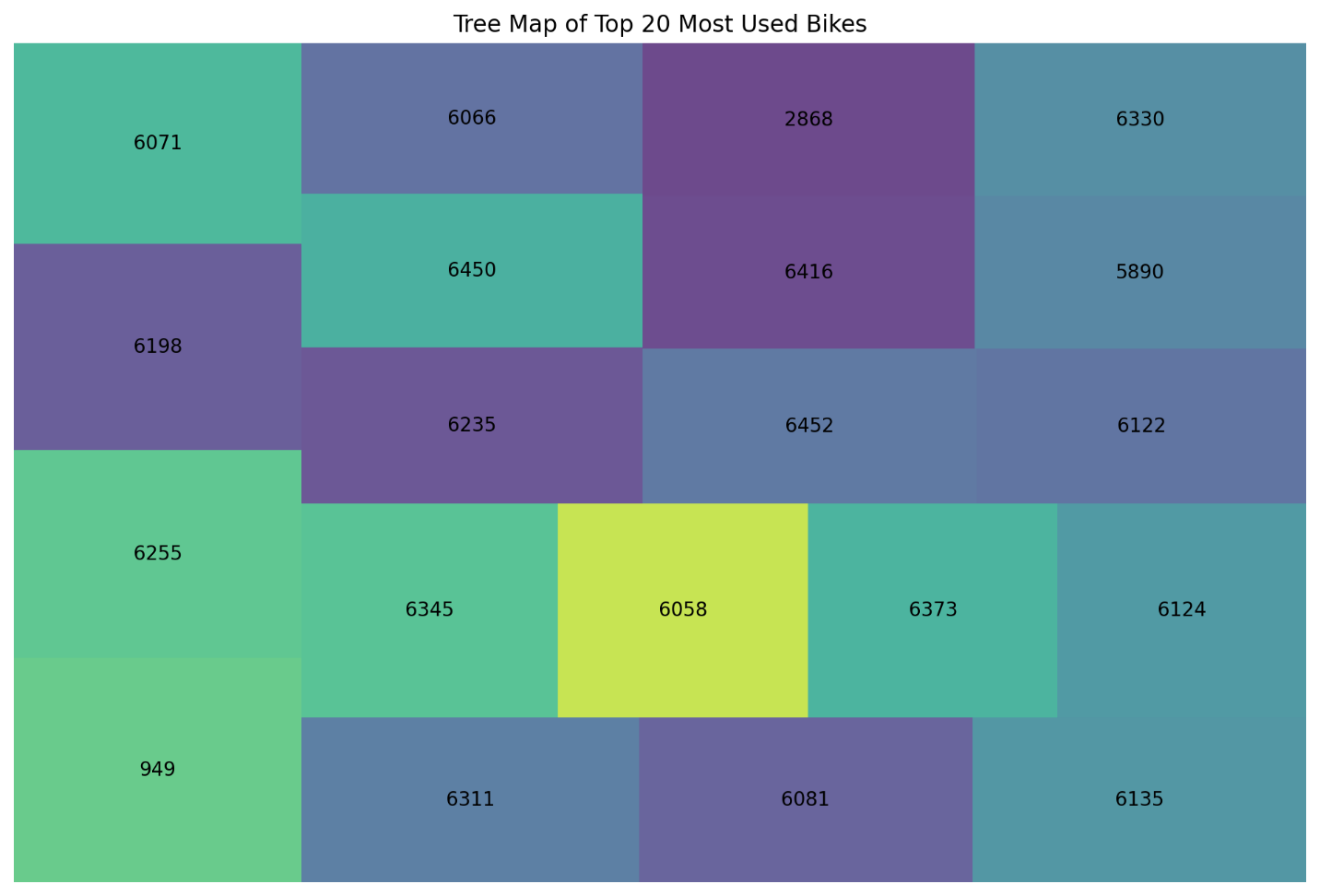
To kickstart this comparison, we meticulously scrutinize and contrast the particular types of bikes that are commonly utilized by each segment, encompassing both annual members and casual riders.

Here is the visualization showing the top 20 most used bikes by user type:

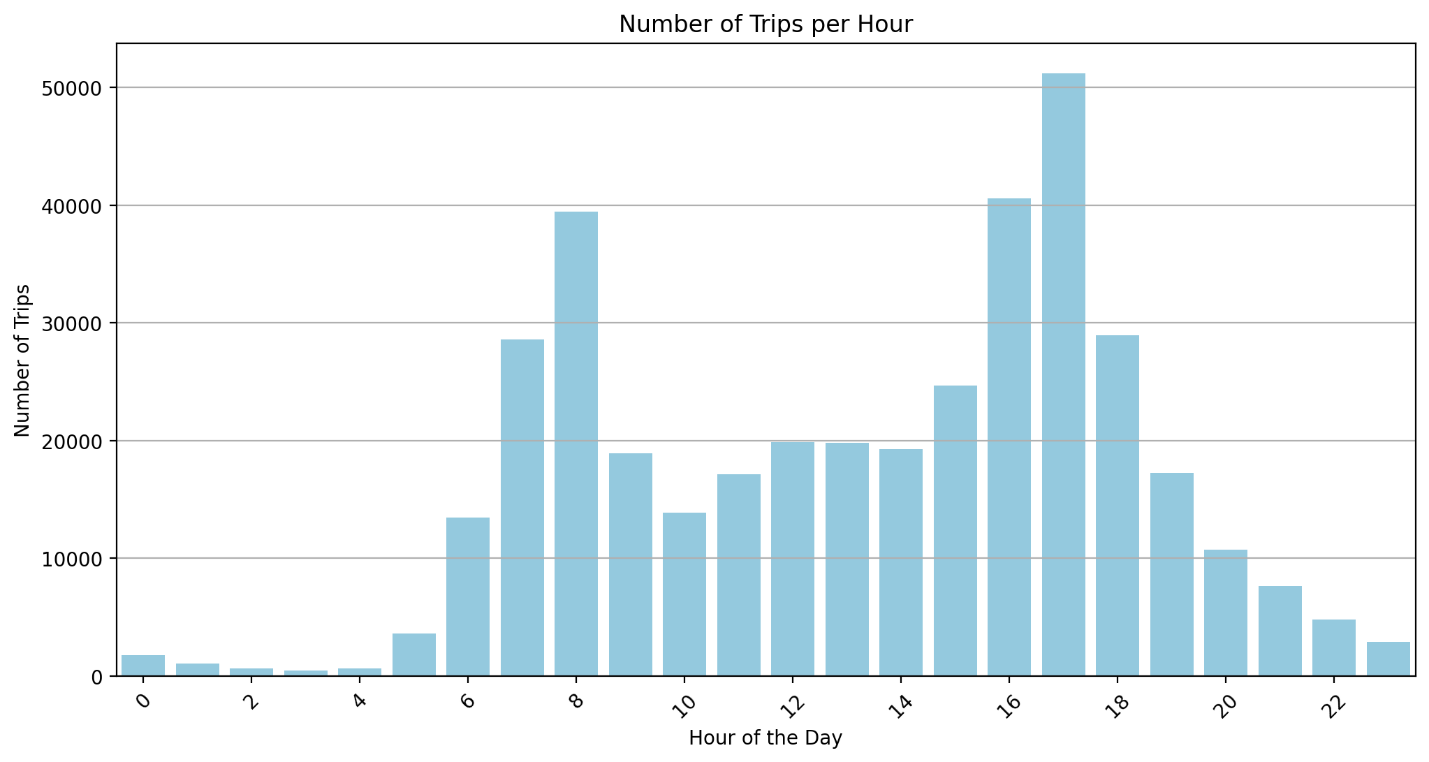


The bar chart visualizes the comparison of bike usage between annual members (Subscribers) and casual riders (Customers). It highlights the top 20 bikes based on the frequency of their use by each user type. The chart provides a clear representation of which bikes are preferred by different types of users, indicating potential patterns or preferences in bike selection among the user groups.

Here is the graph showing the top 20 most used bikes overall:

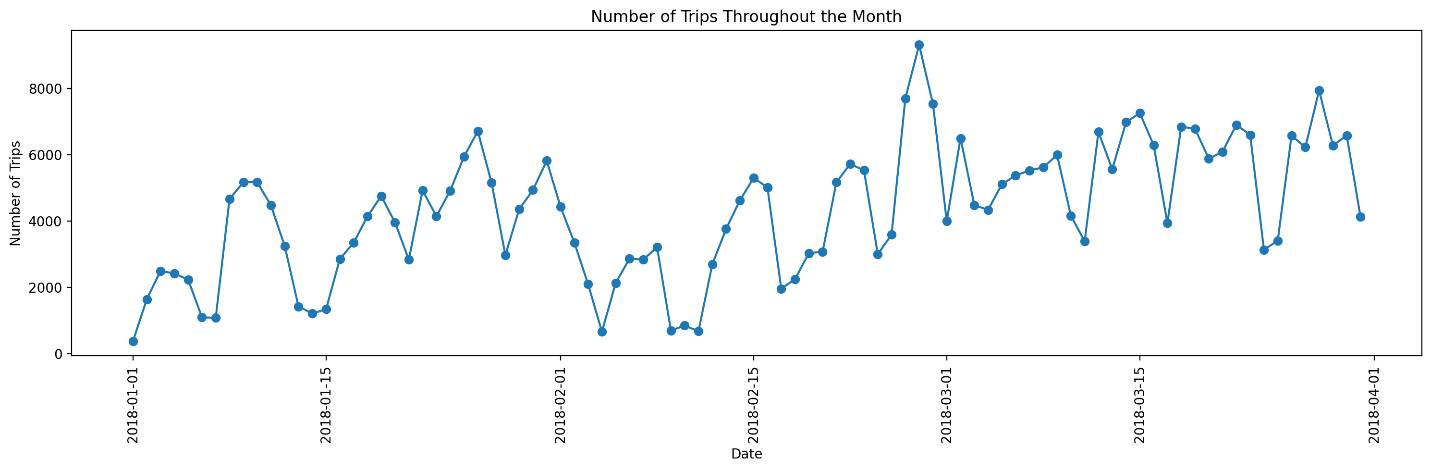


This tree map provides a visual representation of bike usage, with each rectangle corresponding to a different bike ID. The size of each rectangle is proportional to the count of times the bike was used, making it easy to compare at a glance.



The visualization above shows the number of trips per hour throughout the day. It appears that there are two peak periods, likely corresponding to morning and evening rush hours, with the highest number of trips occurring in the late afternoon and early evening. This pattern suggests that the bike-sharing service is commonly used for commuting to and from work.

Here is the visualization showing the number of trips throughout the month:



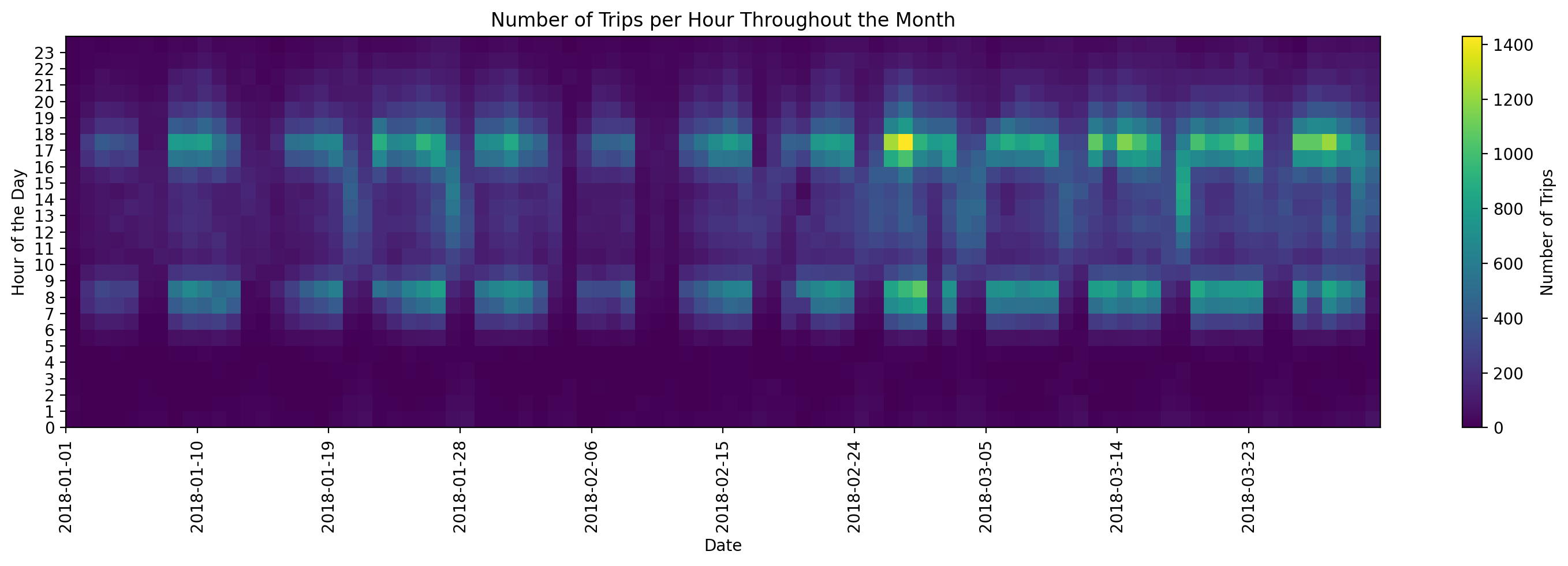
This line chart illustrates the daily trip count over the course of the month, providing insights into the overall usage patterns of the bike-sharing service.

The line chart displayed earlier represents the daily number of trips taken with the bike-sharing service over a month. Here's a brief analysis:

* The x-axis shows the consecutive dates over the month.
* The y-axis represents the number of trips taken each day.
* The line connects points that represent the total trips taken on each date, providing a visual trend of usage over time.

From the chart, we can observe patterns such as which days had the highest and lowest usage, potentially indicating weekly trends or the impact of specific events or weather conditions on bike-sharing usage. If there are peaks or troughs, they may correspond to weekends, holidays, or special events. A consistent pattern across the weeks could suggest regular commuting behavior among users.

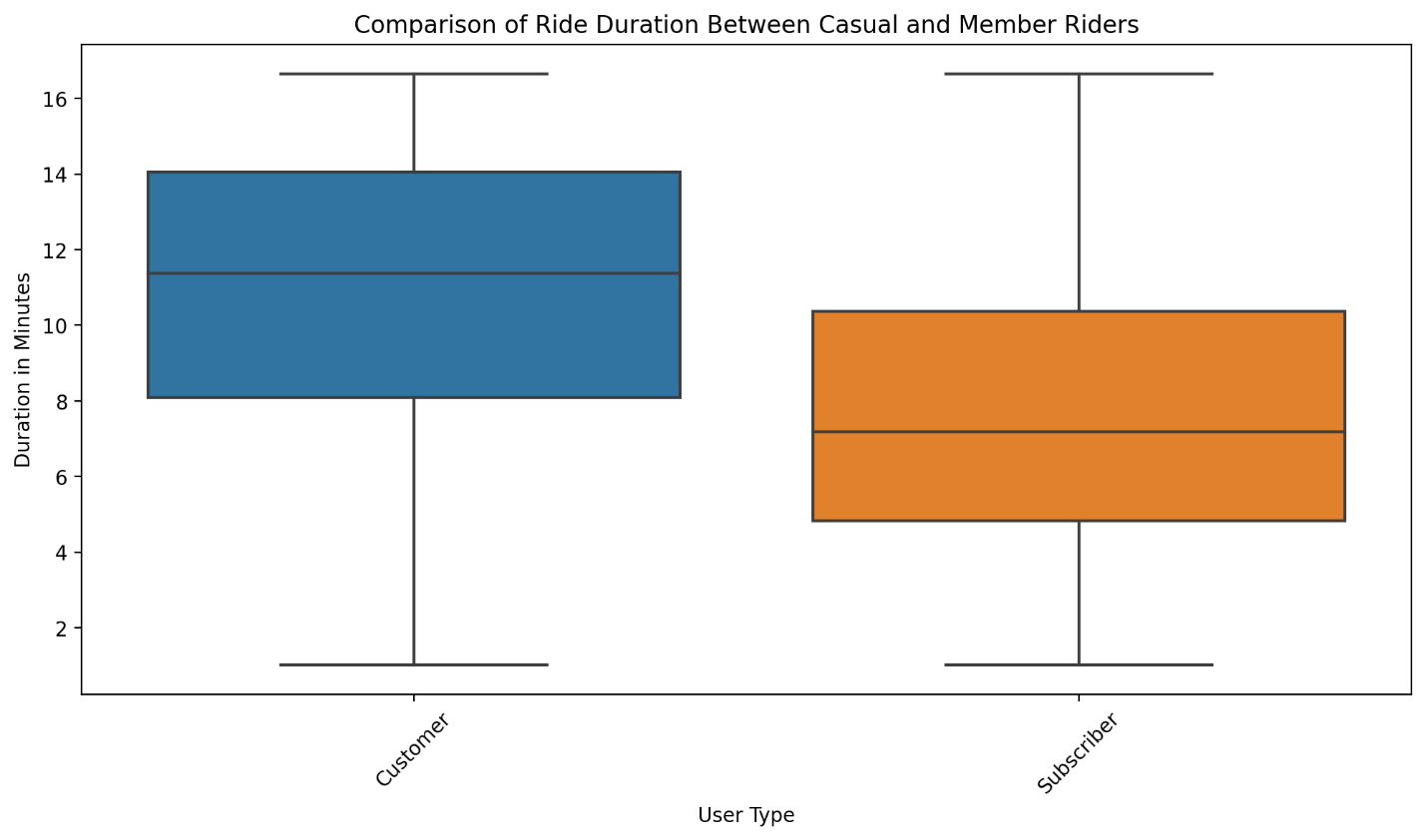
Here is the visualization showing the number of trips per hour throughout the month:



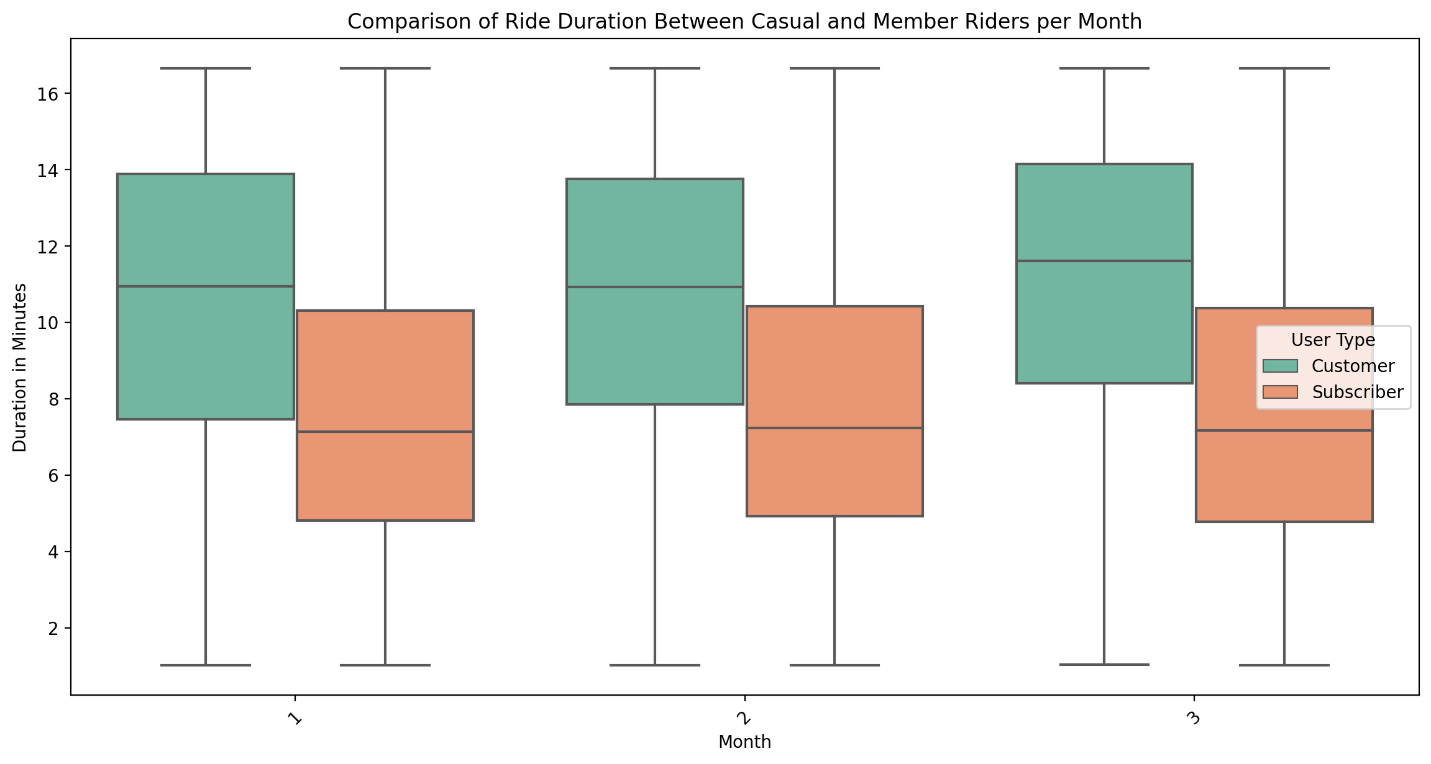
This heatmap provides a visual summary of the trip frequencies, with different hours of the day on the y-axis and dates on the x-axis. The color intensity represents the number of trips, with darker colors indicating higher trip counts.

To understand the differences in behavior between casual and member riders, the ride duration of trips is compared in the subsequent analysis.

The boxplot provides a visual summary of the distribution of ride durations for each user type, highlighting differences in their riding behavior.



This boxplot compares the ride duration between casual and member riders for each month, showing how riding behaviors vary between the two user types across different months.



To gain further insights into the differences between casual and member riders, an analysis of the starting and ending stations' locations can be conducted. By filtering and examining stations with the most trips, additional conclusions can be drawn.

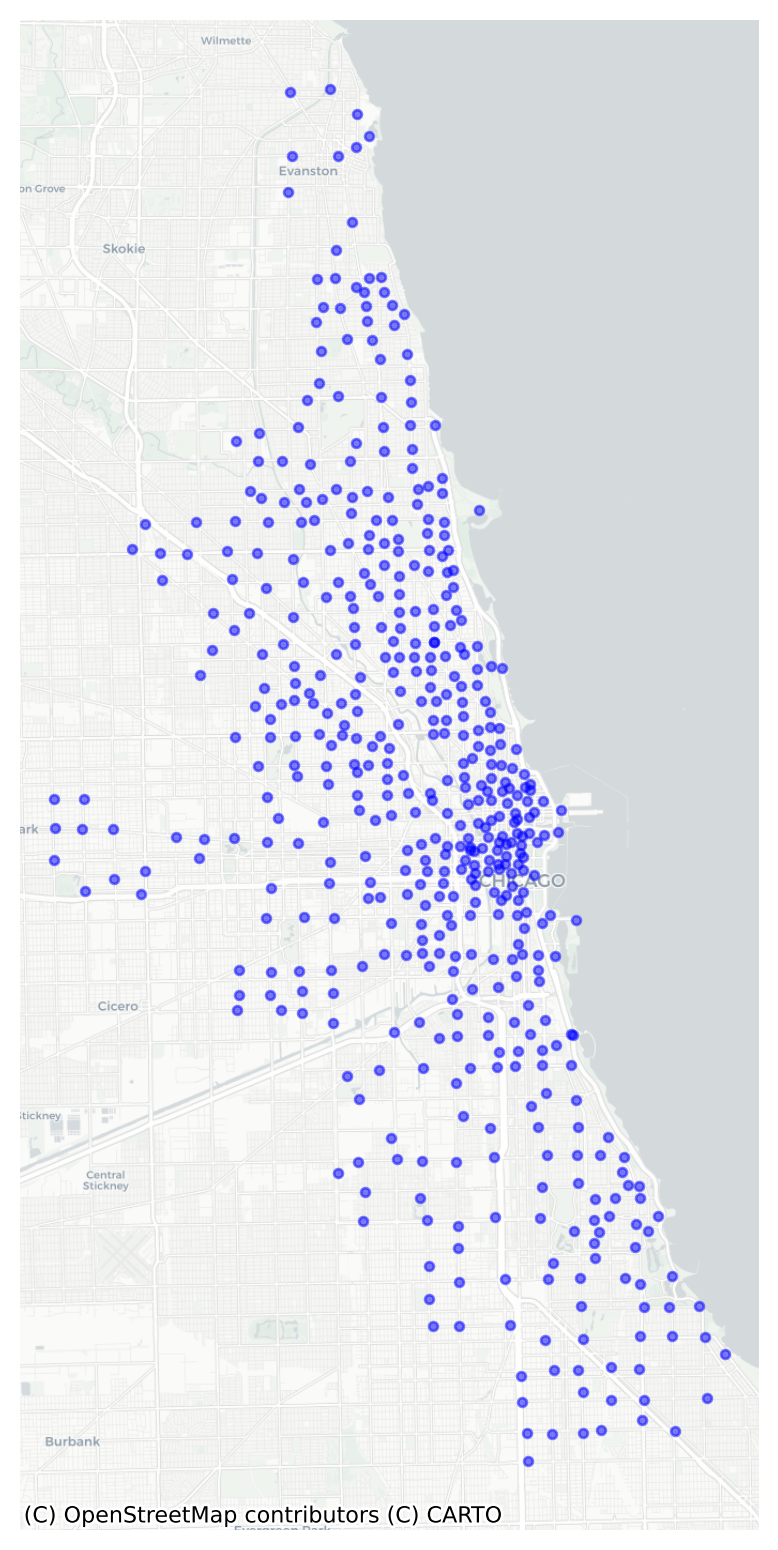
Show Code

|  | **User Type** | **03 - Rental Start Station Name** | **02 - Rental End Station Name** | **Trip Count** |
| --- | --- | --- | --- | --- |
| **166875** | Customer | Lake Shore Dr & Monroe St | Streeter Dr & Grand Ave | 89 |
| **166820** | Customer | Lake Shore Dr & Monroe St | Shedd Aquarium | 66 |
| **247896** | Customer | Shedd Aquarium | Lake Shore Dr & Monroe St | 55 |
| **278916** | Customer | Streeter Dr & Grand Ave | Lake Shore Dr & Monroe St | 51 |
| **166680** | Customer | Lake Shore Dr & Monroe St | Lake Shore Dr & Monroe St | 50 |
| **394597** | Subscriber | Clinton St & Washington Blvd | Michigan Ave & Washington St | 487 |
| **608917** | Subscriber | Wacker Dr & Washington St | Michigan Ave & Washington St | 475 |
| **360757** | Subscriber | Canal St & Adams St | Michigan Ave & Washington St | 466 |
| **520145** | Subscriber | Michigan Ave & Washington St | Clinton St & Washington Blvd | 456 |
| **362449** | Subscriber | Canal St & Madison St | Michigan Ave & Washington St | 442 |

The table above lists the top 5 stations with the most trips for casual (Customer) and member (Subscriber) riders. It shows the start and end stations for these trips along with the trip counts, providing insights into the most popular routes for different types of users.

* For casual riders, the most trips occur between Lake Shore Dr & Monroe St and Streeter Dr & Grand Ave.
* For member riders, the most frequent trips are between Clinton St & Washington Blvd and Michigan Ave & Washington St.

This information could be used to optimize bike availability and maintenance schedules at these high-traffic stations.



This map shows the geographical distribution of the Divvy bike rental start stations across the city, with each point representing a station. The visualization can help in understanding the coverage and density of the bike-sharing service.

Summary:

Casual Rider

Casual riders display a preference for using bikes throughout the day, with heightened activity on weekends. travel distances are roughly double those of members, yet they undertake fewer trips overall.

Additionally, casual riders often start and end their journeys near various recreational sites like parks, museums, coastal areas, and other leisure destinations.

Member Rider

Member riders favor bikes primarily for commuting, especially on weekdays during the summer and spring, with peak usage during typical commute hours (8 am and 5 pm). They engage in more frequent but shorter rides compared to casual riders, with trip durations approximately half as long.

Step 6: Act Following the analysis of differences between casual and member riders, there's an opportunity to formulate marketing strategies encouraging casual riders to transition into members. The following recommendations are proposed:

Considering the heightened activity of casual riders on weekends and during summer and spring, introducing seasonal or weekend-only memberships could prove effective. Given that casual riders tend to cover longer distances, offering discounts for extended ride durations might incentivize casual riders and potentially attract members to longer rides as well.

Top of Form