

Case Study

How Does a Bike-Share Navigate Speedy Success?



Case Study

How Does a Bike-Share Navigate Speedy Success?

Introduction

This case study is towards Capstone project requirement for the [Google Data Analytics Professional Certificate](#).

Scenario

I am a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, my team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, my team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve my recommendations, so they must be backed up with compelling data insights and professional data visualizations.

Lily Moreno, who is the director of marketing and my manager, has assigned me the question to answer: How do annual members and casual riders use Cyclistic bikes differently?

In order to answer the question, I will use the data analysis process consisting of ask, prepare, process, analyze, share and act, and choose spreadsheet and Tableau which are learned in the course as tools to perform data processing, data analysis and data visualization.

Ask

Objective: Design marketing strategies aimed at converting casual riders into annual members.

Question: How do annual members and casual riders use Cyclistic bikes differently?

Key stakeholders: Lily Moreno, Cyclistic marketing analytic team and Cyclistic executive team

Prepare

The data used are Cyclistic's historical trip datasets and are made available by Motivate International Inc. under this [license](#). I downloaded the previous 12 months of Cyclistic trip data (December 2021 to November 2022) from [here](#), unzipped the files, changed the file names by using appropriate file-naming conventions and saved the files to the appropriate folders.

The data are organized as separate files by month and year and saved as CSV files within .zip folders.

The data include 13 fields as follows:

ride_id: a unique ID per ride.

rideable_type: the type of bicycle used (classic, electric or docked bicycle).

started_at: the date and time when the bicycle was checked out.

ended_at: the date and time when the bicycle was checked in.

start_station_name: the name of the station at the start of the trip.

start_station_id: a unique ID for the start station.

end_station_name: the name of the station at the end of the trip.

end_station_id : a unique ID for the end station.

start_lat: the latitude of the start station.

start_lng: the longitude of the start station.

end_lat: the latitude of the end station.

end_lng: the longitude of the end station.

member_casual: the type of the user taking the bicycle (member or casual).

Process

I used spreadsheet to open each file, and performed the following steps:

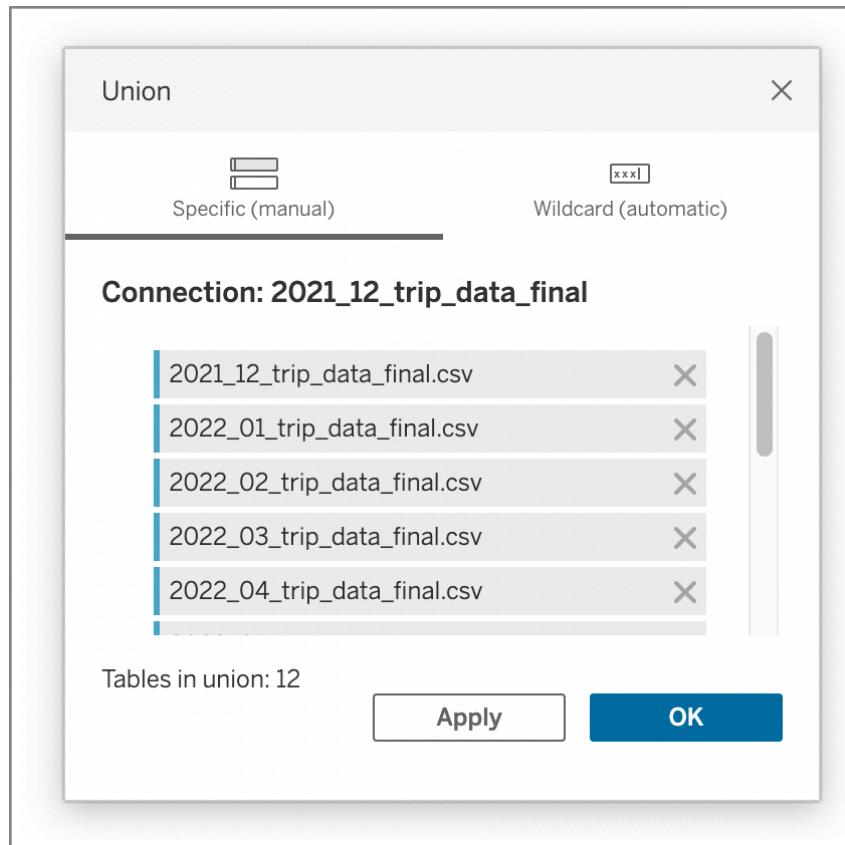
1. Removing duplicates;
2. Deleting the rows where start_station_name & start_station_id fields or end_station_name & end_station_id fields are null by adoption of the filter function as considerable size of data are given;
3. Creating a column called "ride_length (min)", calculating the length of each ride by subtracting the column "start_at" from the column "ended_at" and formatting as HH:MM:SS, converting the time to minutes

by use of formula, and deleting the rows where values of the ride_length (min) are negative, which indicates that the time of the start_at field is later than the time of the end_at field;

4. Creating a column called “day_of_week”, calculating the day of the week that each ride started using the “WEEKDAY” command, and formatting as a number.

Analyze

I imported the 12 processed files into Tableau and used UNION function to combine them into a single table, and adopted the summary file for further analysis.



A summary of my analysis is as follows:

Total Number of trips:

Yearly Total: 4,410,259

Casual: 1,771,721

Member: 2,638,538

It's obvious that members take more rides than casual riders during the year.

The month with the highest number of trips is July 2022, while the month with the fewest number of trips is January 2022, and both casual riders and members have the similar year-round usage pattern.

Rideable Type Preference:

Classic bikes are mostly used by both casuals and members, followed by electric bikes and docked bikes.

Ride Length:

The max ride length for the year is 34,354 minutes (perhaps the user forgot to check the bike back in?), and there are some ride length of 0 minute, I think this situation may be caused by system error, but these rows will rarely influence our analysis results due to the small size.

The yearly average ride length is 17.17 minutes, 12.46 minutes for member and 24.19 minutes for casual, which is almost double of the average ride length of the member.

May is the month with the longest average ride length, while November is the month with the shortest average ride length, this is in keeping with the tendency that people would like to have longer trips during warm months.

Bike Usage on Days of Week:

The data indicates that casual riders have more rides during weekend, while Sunday is the day with the fewest number of rides for member, which may reflect that casual riders are more likely to ride for leisure, and members are more likely to ride for work on weekdays.

Busiest Stations:

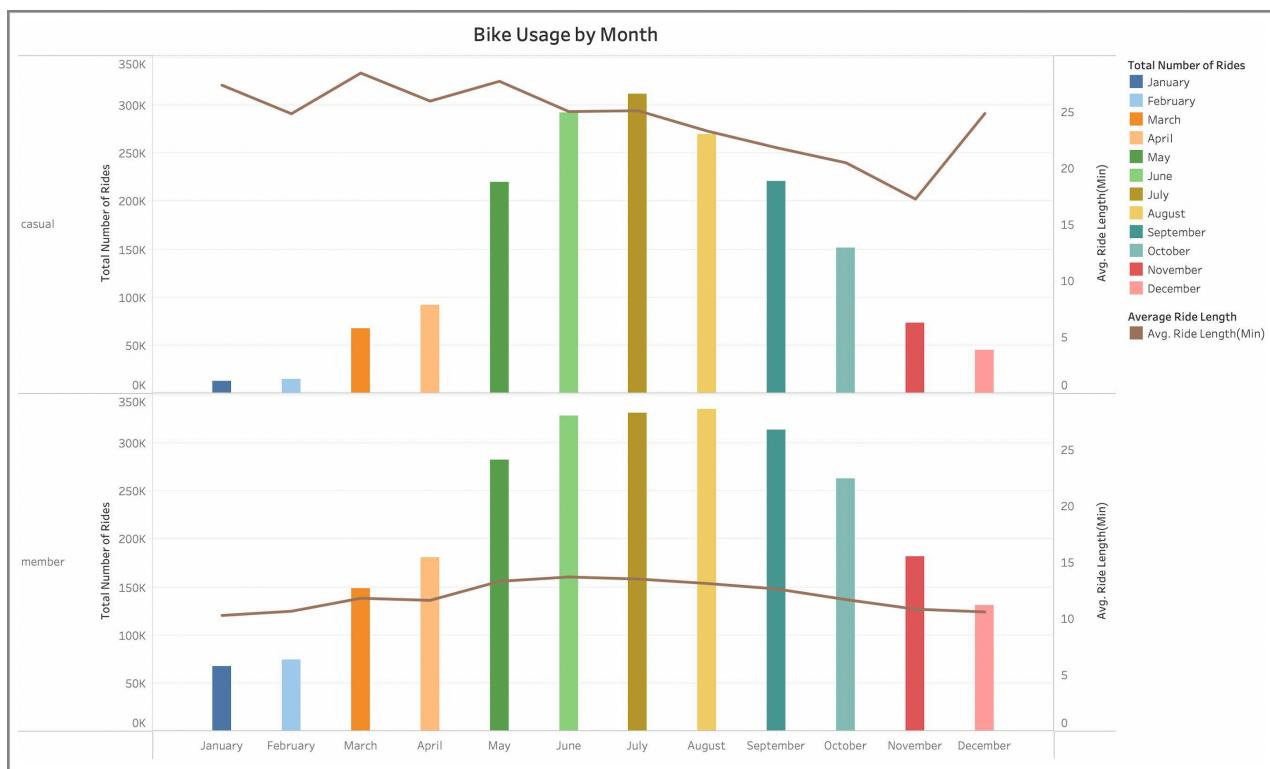
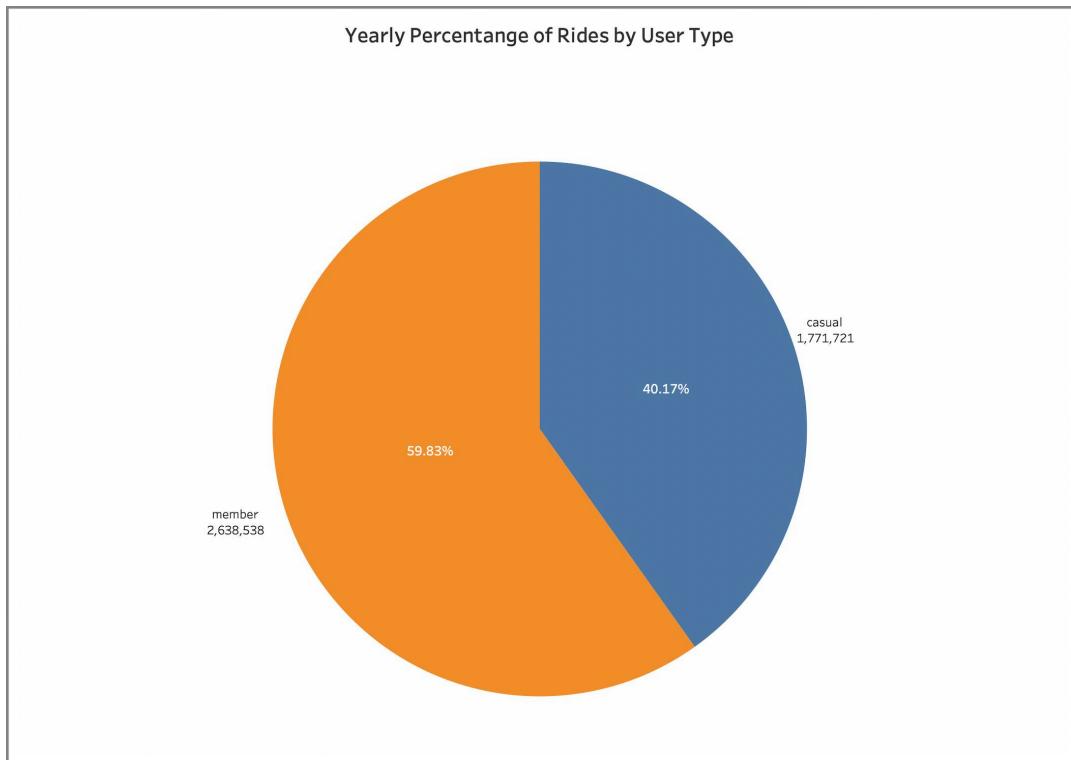
The top 5 start stations and top 5 end stations for casual riders and members are the same, but the rank order is different, for example, Wells St & Concord Ln station is the busiest station for member and the least busy station for casual among the top 5 stations.

More details need to be analyzed in combination with a map in the following section.

Share

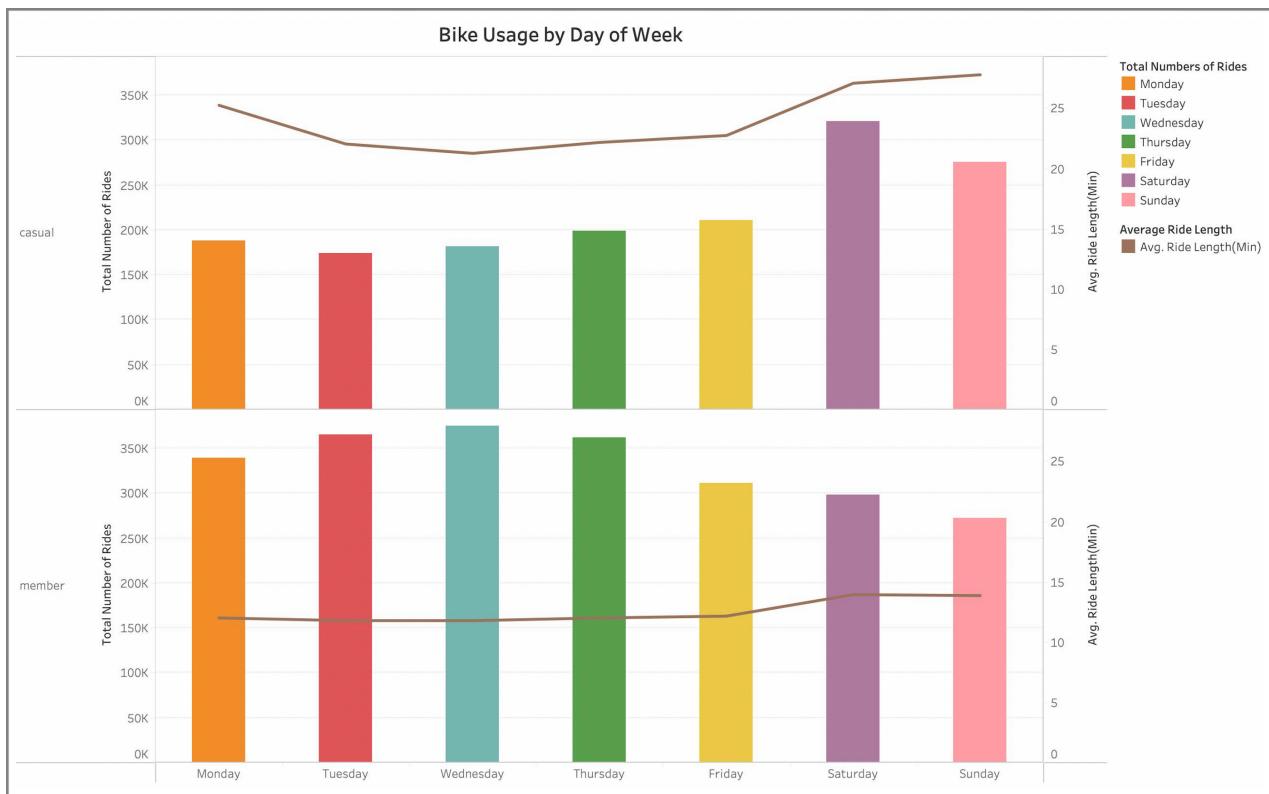
I created my data visualizations on the Tableau and have the findings as follows:

59.83% of all trips are made by members, while the other 40.17% of trips are made by casual riders, which may indicate that there is significant room for converting casual riders into members.



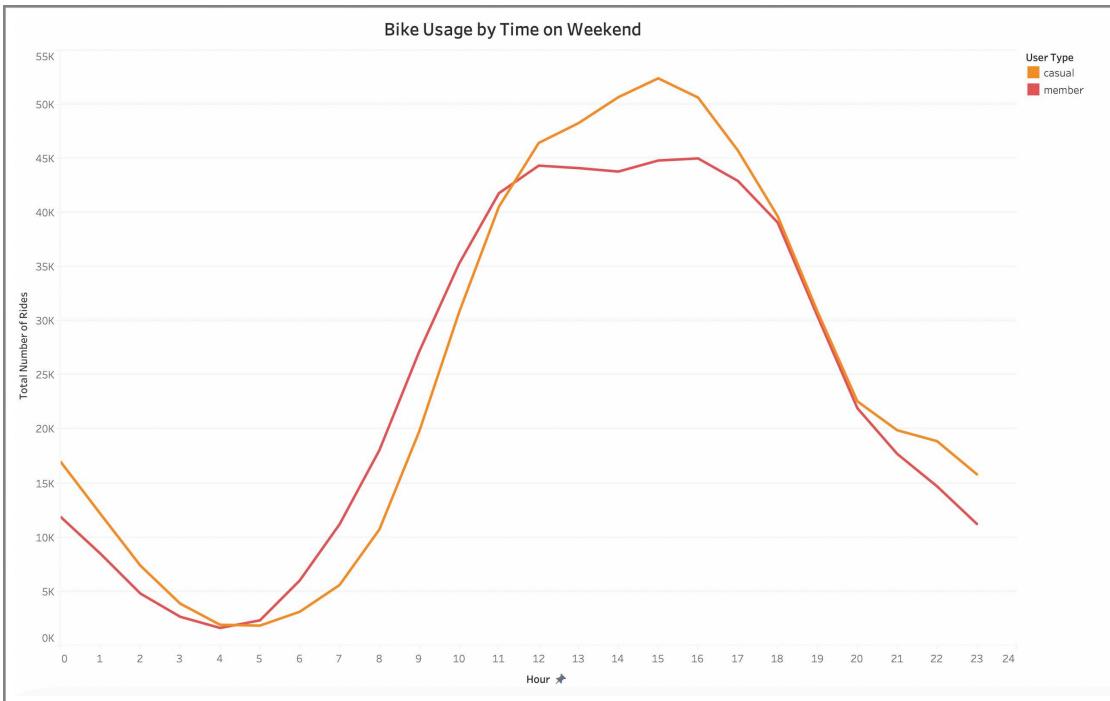
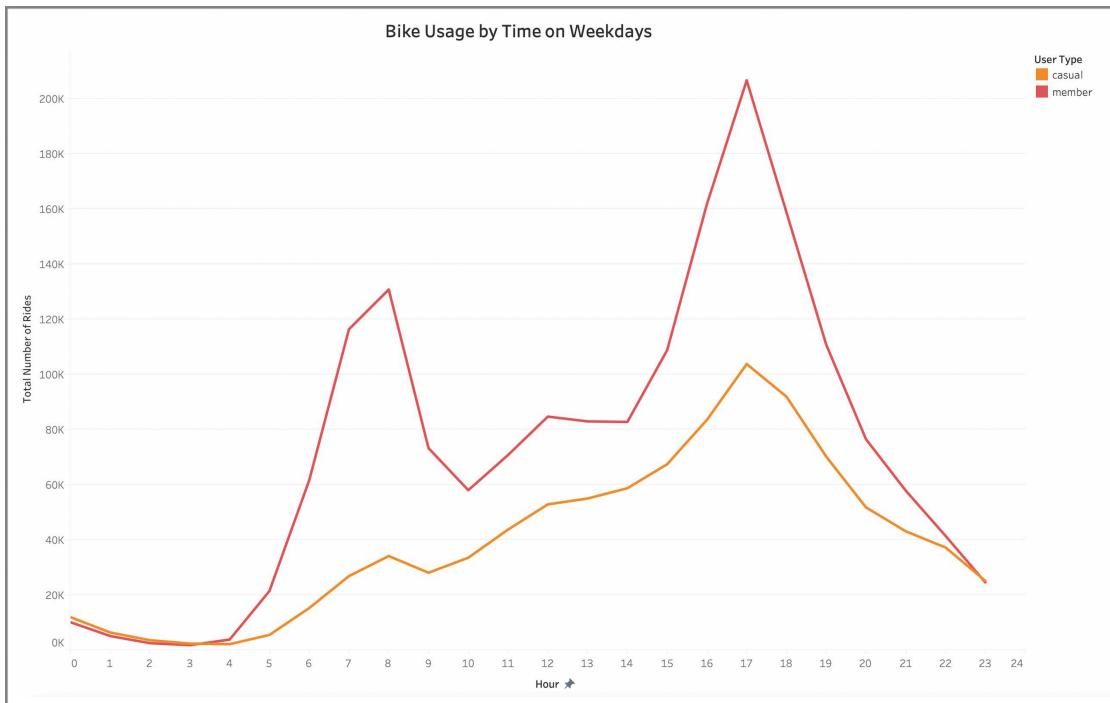
According to the above chart, it is obvious that there are far more rides during warm months for both members and casual riders, while members are much more likely to be persistent in cold months.

In comparison with the average ride length of casual riders, the average ride length of members remains quite constant all the year round, which seems to fit the hypothesis that members use the bikes primarily to commute to work. Casual riders seem to take rides which are twice as long on average as those of members, indicating that casual riders may cover more distance or ride at a leisure pace.



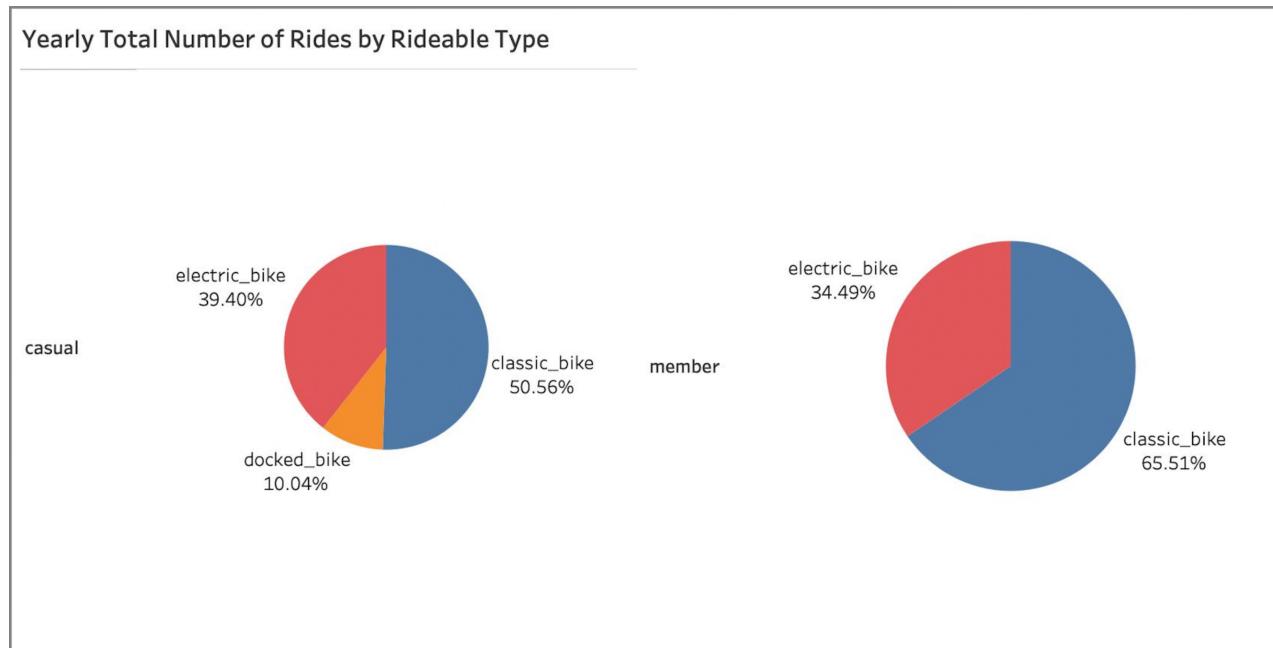
The data shows a significant increase on total number of rides for casual riders on weekend, while there is a slight decrease on total number of rides for members on weekend. The average ride length increases on weekend for both casual riders and members, but the

increase is more obvious for casual riders. This fits the hypothesis that most casual riders are more likely to use the bike for leisure.

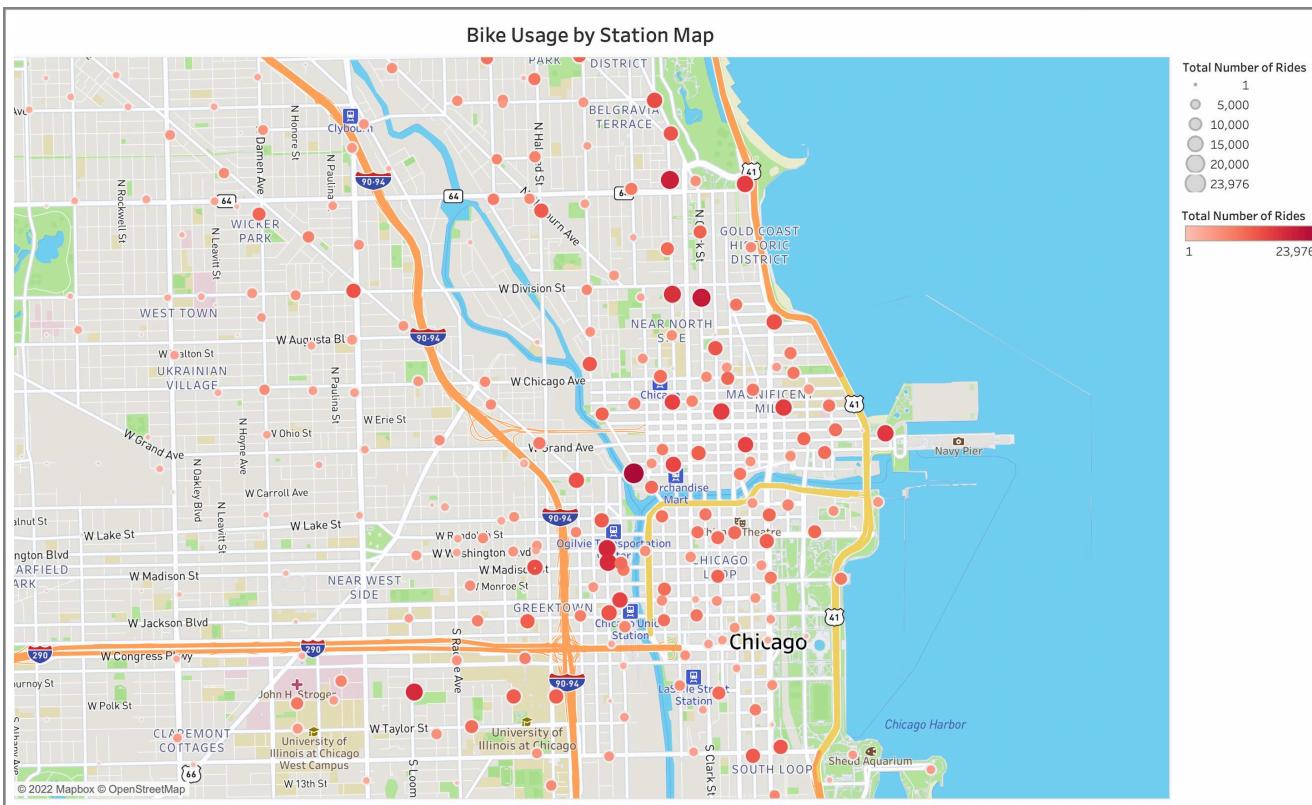
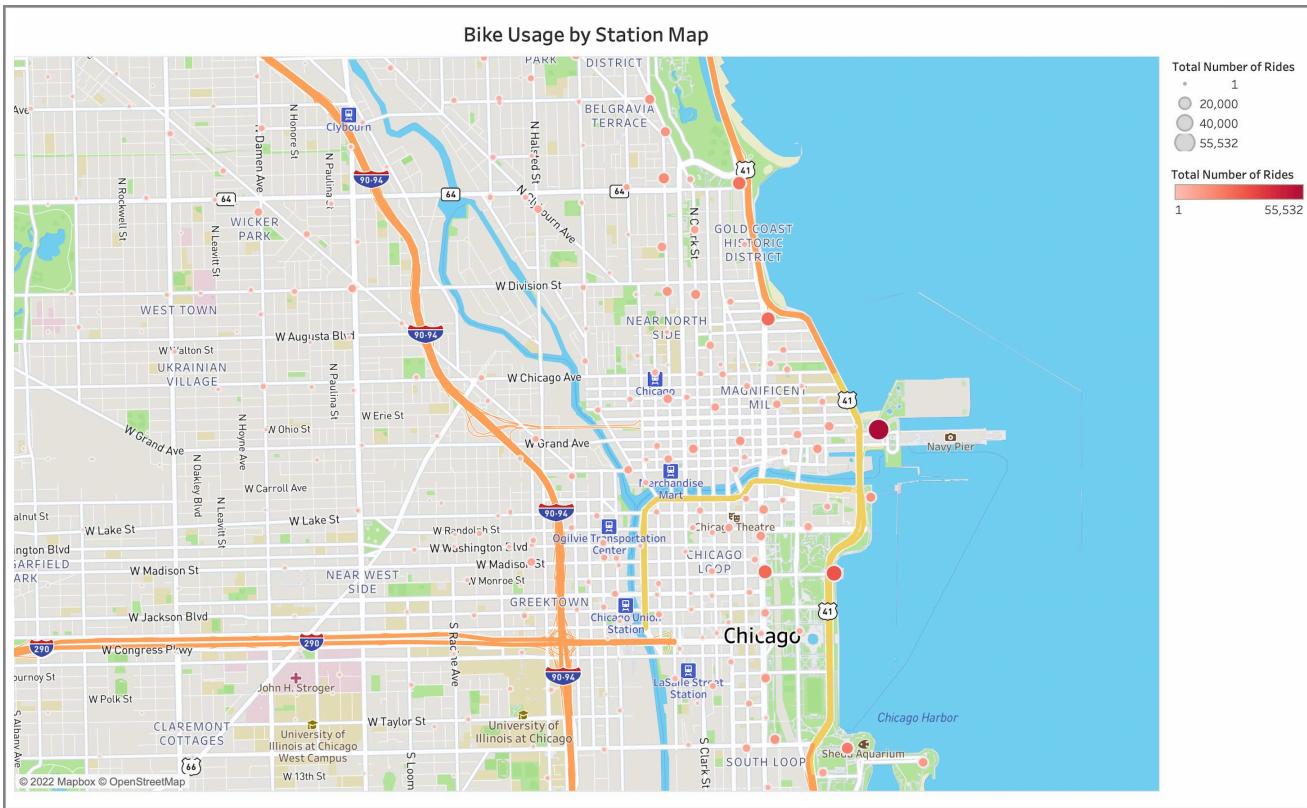


By analyzing the 24-hour total number of rides on weekdays and weekend for casual riders and members respectively, I find that there are two peaks for members on weekdays, one is at 8 a.m. and the other one is at 5 p.m., which are the time for members to commute to work. It's worth noting that there is still a part of casual riders who use the bikes to commute to work on weekdays as well, and this part of causal riders can be converted into members.

On the other hand, casual riders and members have the similar usage pattern on weekend, we can see a much more gradual curve from 11 a.m. to 6 p.m., which may suggest that weekend riders are using the bikes for leisure.



Classic bike is the most popular option for both casual riders and members, followed by electric bike, and members rarely use docked bike at all.



According to the above two maps used for representing distribution of total number of rides on different stations respectively for casual riders and members, I find that stations used by casual riders are mostly concentrated in the beach area, while stations used by members are more evenly distributed around the city and concentrated around areas such as subway stations.

These findings also support the hypothesis that most members use bikes for commute to work while most casual riders use bikes for leisure.

Act

Based on the analysis above, my top three recommendations are as follows:

1. As a part of casual riders still use bikes for work during weekdays, advertising campaigns should target those casual riders and focus on the benefits of using the service for work commutes;
2. With a surge in the number of casual riders during the months of summer, a large summer campaign should be carried out around the most used stations on weekend, in order to reach a high number of casual riders;
3. As casual riders take longer trips, a plan for accumulating points according to the ride length can be provided, wherein the points can be used for getting a discount on the fee of annual member.