Case Study: Cyclistic Bike-Share

May 8 2023

Vivian To

Google Data Analysis Capstone Project

This case went through the six steps of a core analysis process that I learned in the Google Data Analysis Certificate Program. These six steps or phases include **ask, prepare, process, analyze, share,** and **act.**

Datasets

The datasets I used for this case study included twelve months for the year 2022. You can find the information here Index of bucket "divvy-tripdata".

During the process phase of the analysis, I checked for duplicate rows, added a column to calculate the ride length of each trip, and another column called day_of_week to get the weekday for each trip. I cleaned up some data not make sense for this analysis, which was trips' ride length of less than 1 minute and greater than 1440 minutes or 24 hours. I also cleaned up rows containing the wrong date, such as the started_at date greater than the ended_at date. I could reverse the date columns and calculate the ride length. However, I did not have other sources to reference to make sure they were correct. As a result, I decided to delete them because there were not many rows of data in the wrong columns. I had to exclude these error values and clean up the data before the analysis.

Process and Analyze Phrase

While cleaning up the data, I also checked for the variable and corrected some variable names that were inconsistent with other files.

Although this could complete in Excel, it took more time. So I transfer them to work in SQL. I did the calculation for the ride_length and day_of_week in SQL. I queried the table with aggregation number and group by the factors. Then I export those tables out of SQL, and started to visualize them on the Tableau desktop.

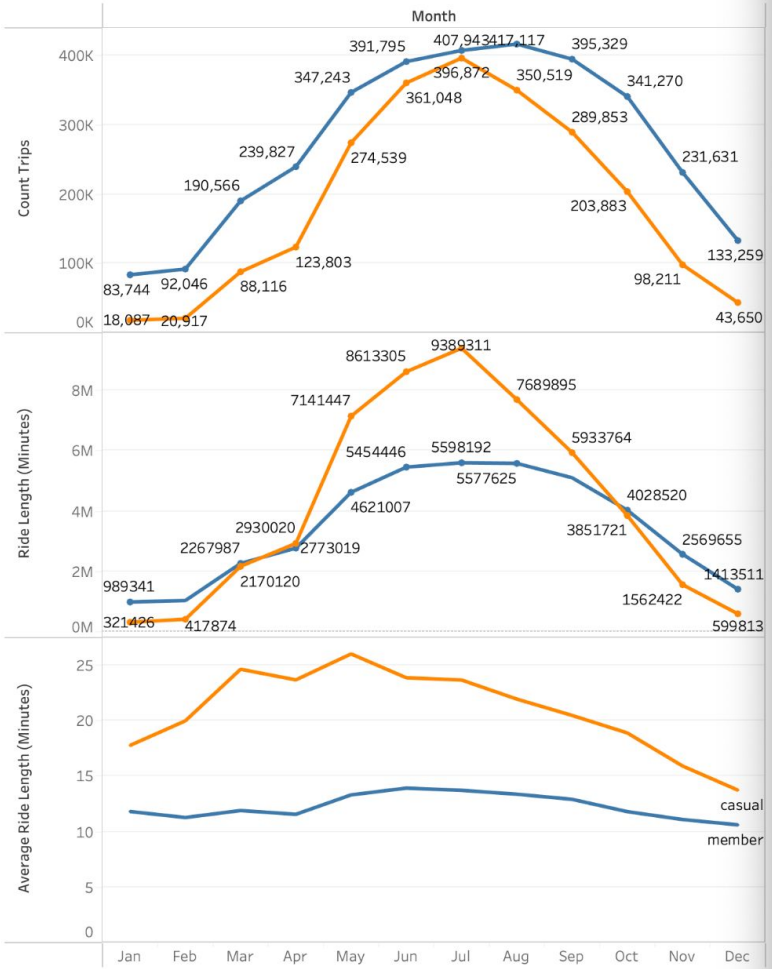
The business task for this case study was to analyze how the member riders differ from the casual riders for Cyclistic company. The goal of this analysis is to find out how Cyclistic can influence those casual riders to convert them into members, which help the Cyclistic Company's growth in the future.

Visualized the analysis data in Tableau

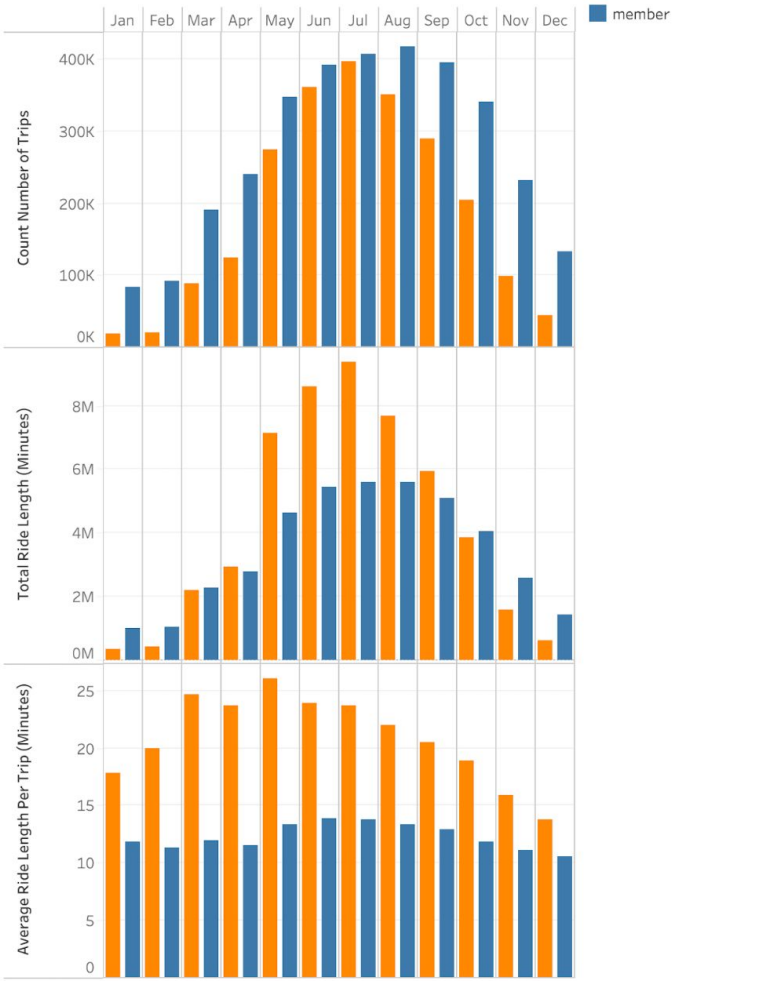
Below graphs show the number of trips for the year 2022 in Chicago. There was a low demand for both groups of riders from December to February, perhaps due to the local weather.

The member group’s average ride length was lower than the casual group.

Type of Riders in year 2022



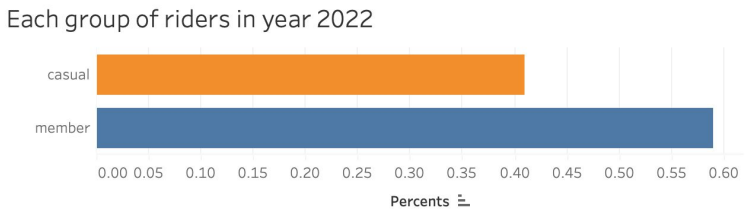
Monthly View for Trips in Year 2022

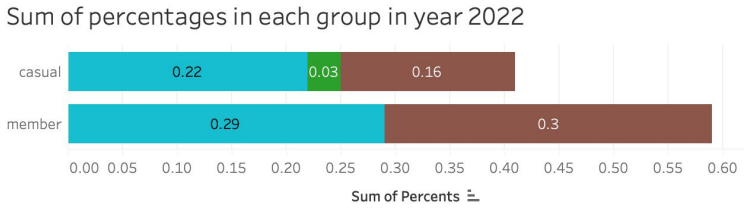


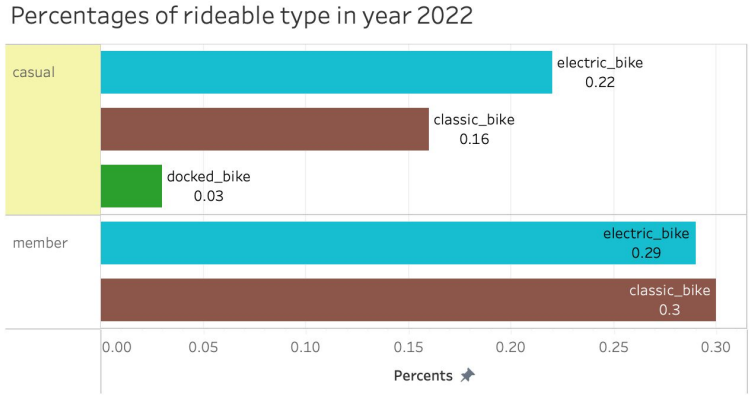
Percentages of both member and casual rider

Casual riders were as much as about 69.5% of the Member riders.

Only 3% of casual group used docked_bike, none of the member group used docked_bike





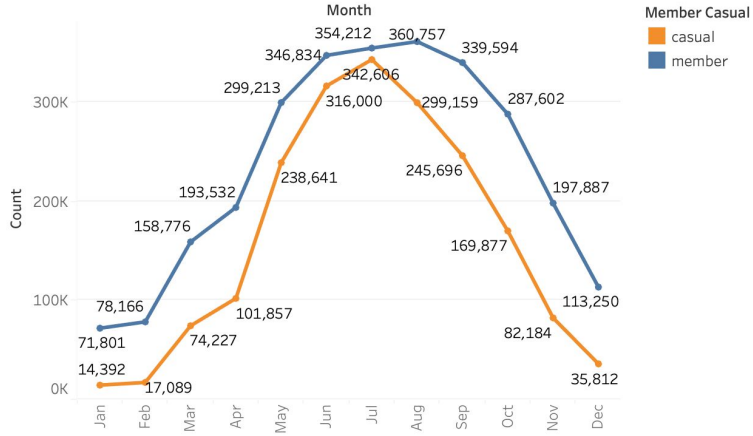


Trips with and without station names

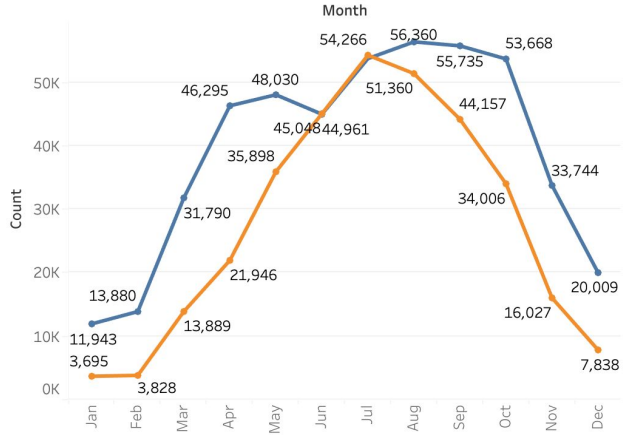
The graph on the top right excluded trips that didn't have the station names for both groups.

The bottom right showed the number of trips excluded from the top right, and those numbers didn't have station names.

Trip counts that have station names



Trip counts that have no station name

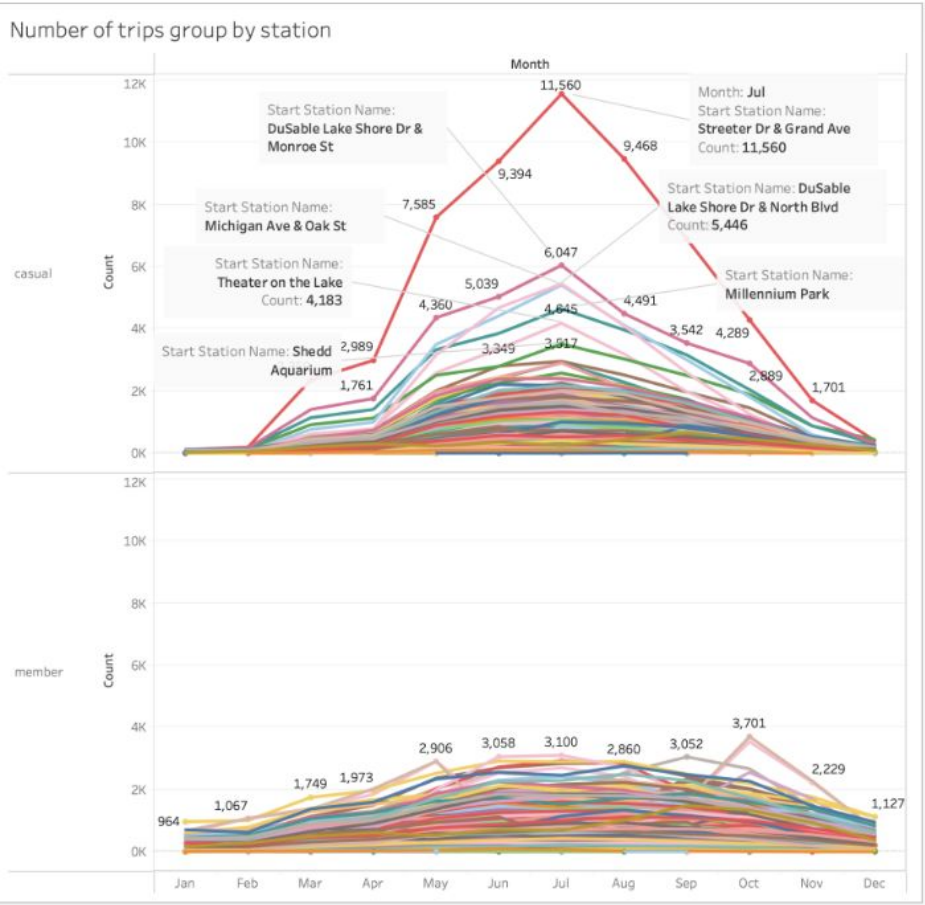


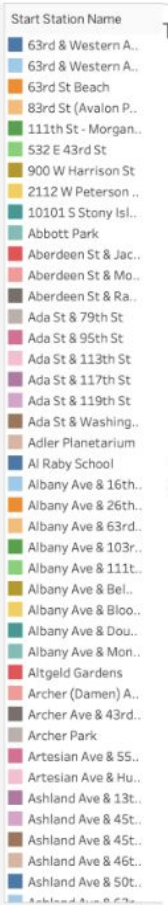
Unusual high trip counts for a few of the stations

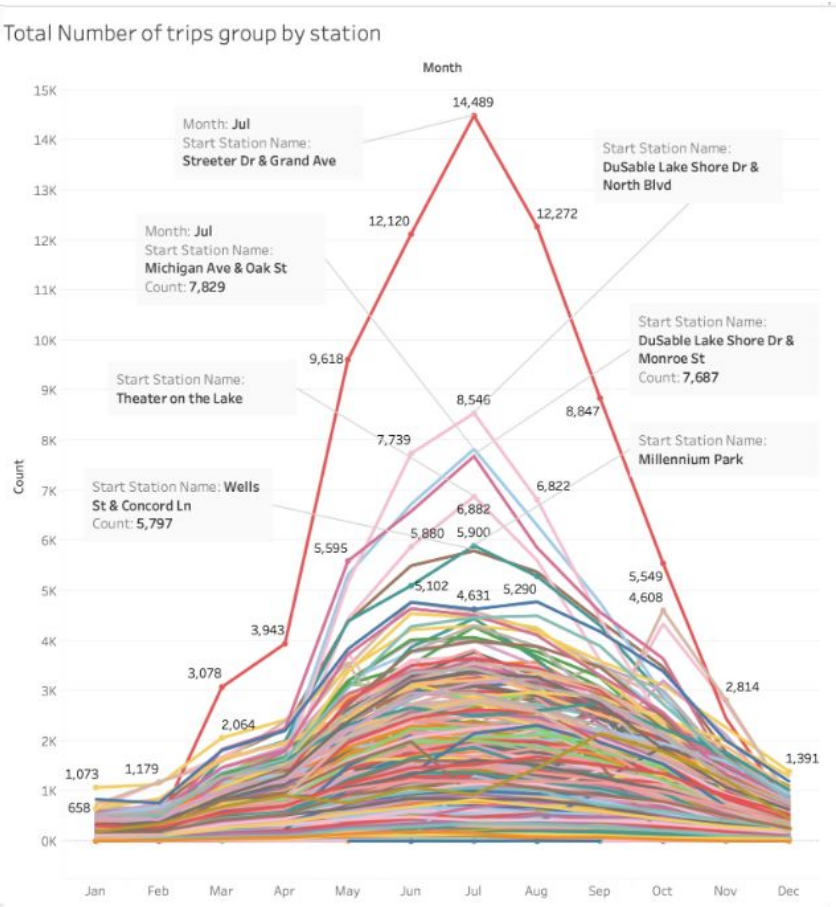
Below graphs showed about seven stations have a spiking number of trips for the casual group. These numbers could be the boom of tourists visiting during summer and fall.

There were no spikes for the member group in these stations.

The graph on the right describes the total number of trips for both groups and the stations


Number of trips group by station
Month
casual
member
Count
Start Station Name: DuSable Lake Shore Dr & Monroe St
Month: Jul
Start Station Name: Streeter Dr & Grand Ave
Count: 11,560
Start Station Name: Michigan Ave & Oak St
Start Station Name: DuSable Lake Shore Dr & North Blvd
Count: 5,446
Start Station Name: Theater on the Lake
Count: 4,183
Start Station Name: Millennium Park
Start Station Name: Shedd Aquarium
11,560
9,468
9,394
7,585
6,047
5,039
4,360
4,491
4,645
3,542
4,289
3,349
3,917
2,989
1,761
2,889
1,701
3,701
3,058
3,100
2,860
3,052
2,906
2,229
1,749
1,973
1,067
964
1,127
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


Start Station Name
63rd & Western A..
63rd & Western A..
63rd St Beach
83rd St (Avalon P..
111th St - Morgan..
532 E 43rd St
900 W Harrison St
2112 W Peterson ..
10101 S Stony Isl..
Abbott Park
Aberdeen St & Jac..
Aberdeen St & Mo..
Aberdeen St & Ra..
Ada St & 79th St
Ada St & 95th St
Ada St & 113th St
Ada St & 117th St
Ada St & 119th St
Ada St & Washing..
Adler Planetarium
Al Raby School
Albany Ave & 16th..
Albany Ave & 26th..
Albany Ave & 63rd..
Albany Ave & 103r..
Albany Ave & 111t..
Albany Ave & Bel..
Albany Ave & Bloo..
Albany Ave & Dou..
Albany Ave & Mon..
Altgeld Gardens
Archer (Damen) A..
Archer Ave & 43rd..
Archer Park
Artesian Ave & 55..
Artesian Ave & Hu..
Ashland Ave & 13t..
Ashland Ave & 45t..
Ashland Ave & 45t..
Ashland Ave & 46t..
Ashland Ave & 50t..

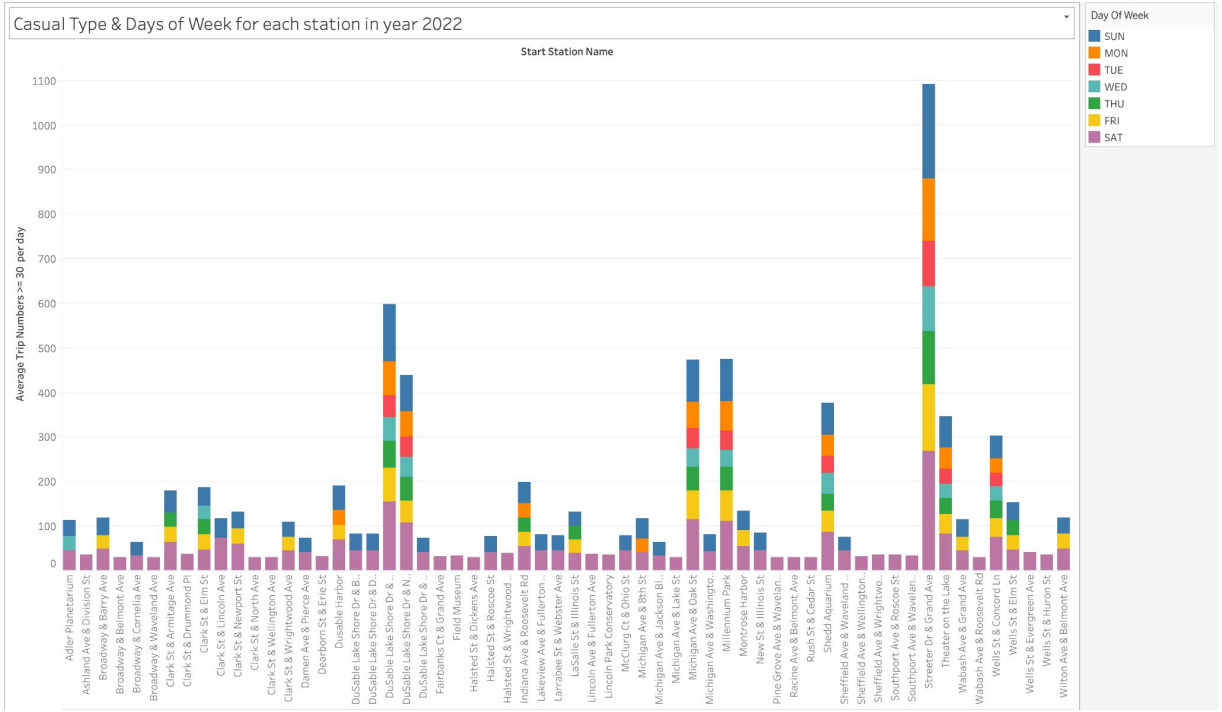

Total Number of trips group by station
Month
Count
Month: Jul
Start Station Name: Streeter Dr & Grand Ave
Start Station Name: DuSable Lake Shore Dr & North Blvd
Month: Jul
Start Station Name: Michigan Ave & Oak St
Count: 7,829
Start Station Name: DuSable Lake Shore Dr & Monroe St
Count: 7,687
Start Station Name: Theater on the Lake
Start Station Name: Millennium Park
Start Station Name: Wells St & Concord Ln
Count: 5,797
14,489
12,120
12,272
9,618
8,546
8,847
7,739
6,822
6,882
5,880
5,900
5,595
5,102
4,631
5,290
5,549
4,608
3,943
3,078
2,814
2,064
1,391
1,073
1,179
658
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

Average Weekday Riders in the year 2022

The Casual group shows large numbers of trips on average per day that were >= 30 for each station that happened to be on Saturday. Also, a few spikes for Monday to Sunday are station names that matched the tourist area in previous graphs. There were a few riders from Monday to Sunday if the average rider per day was below five or equal to one. It meant there weren't many riders for the Casual from Monday to Sunday versed many riders for the member.

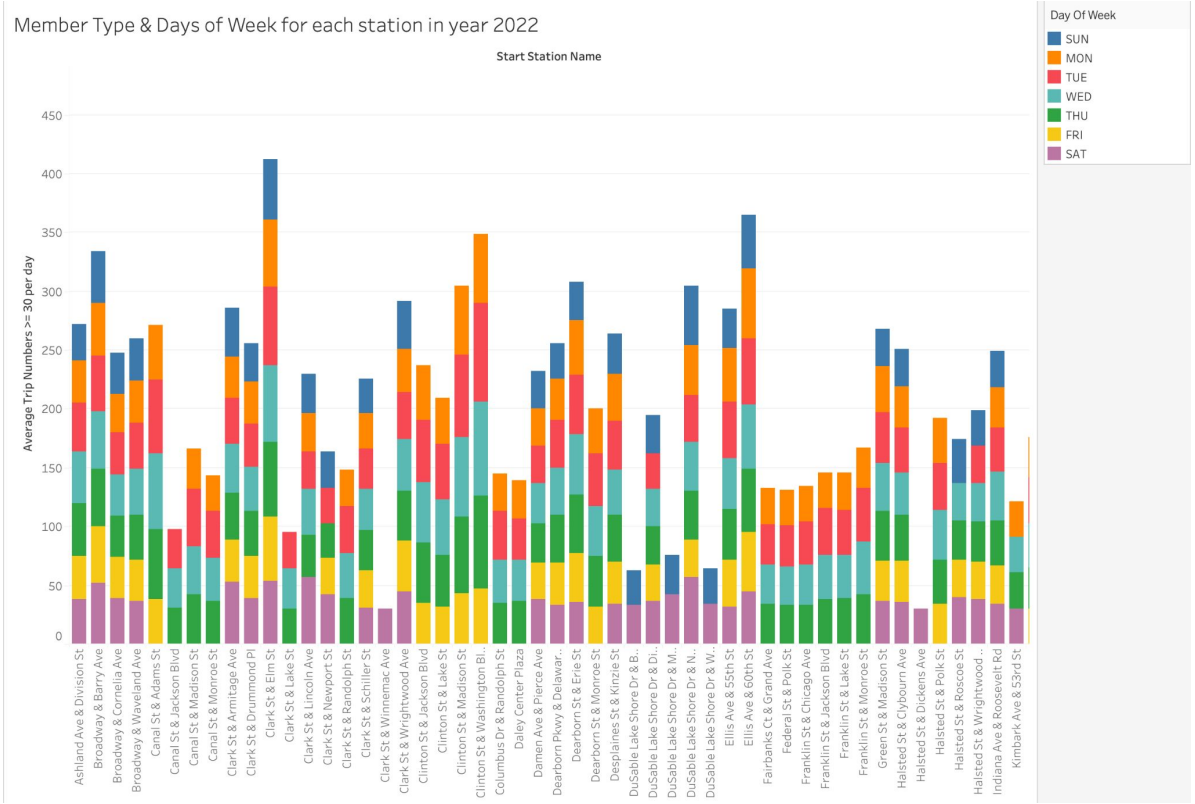
Casual riders:

A large number of the Casual riders on most Stations happened on Saturday.

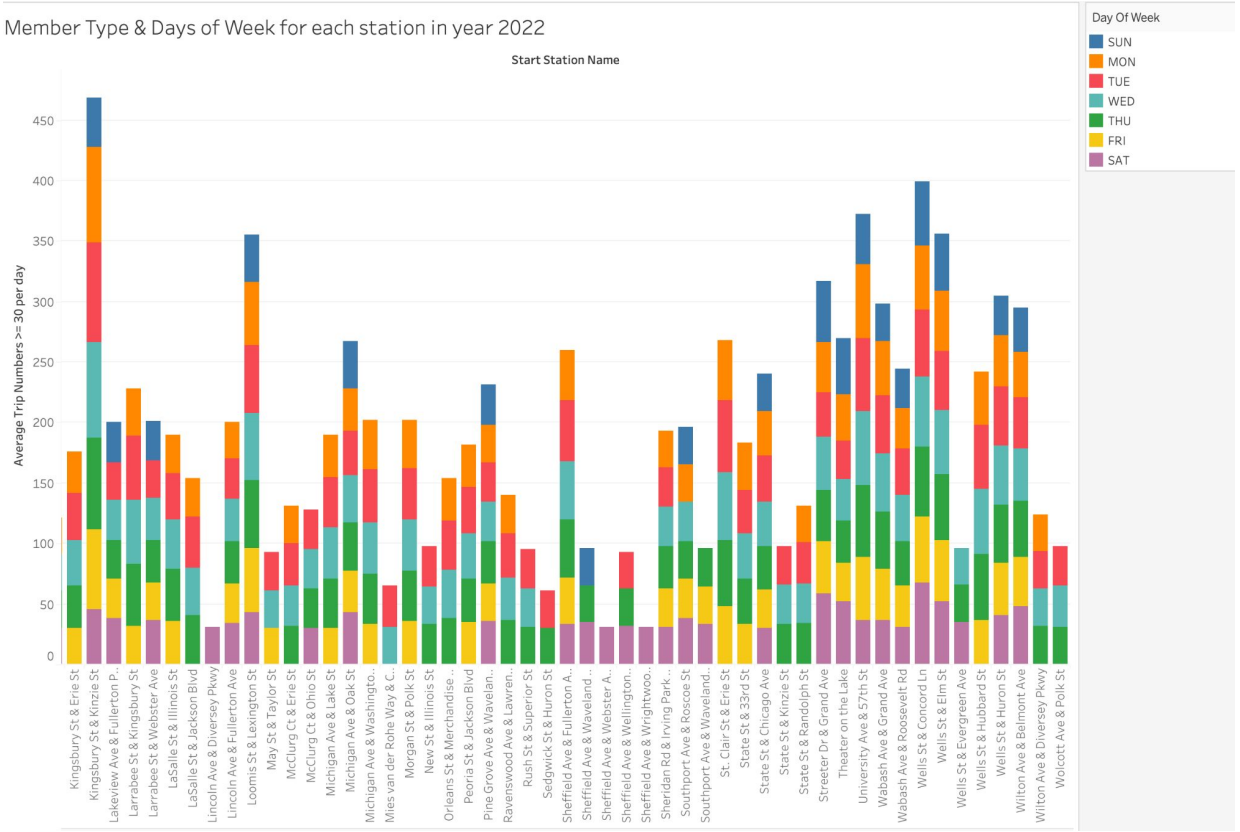


Member riders

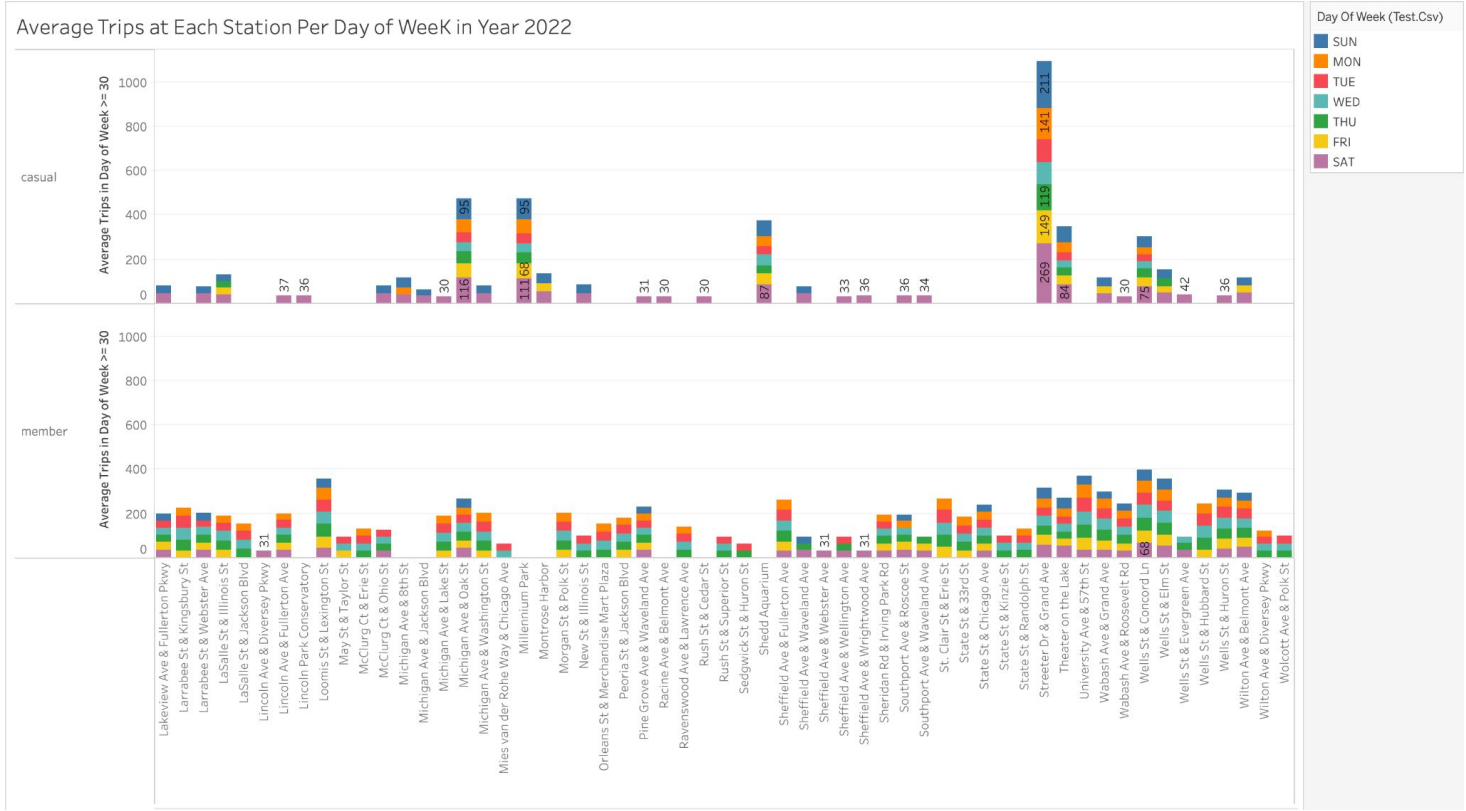
Member riders tend to have a large number of trips from Monday to Sunday.



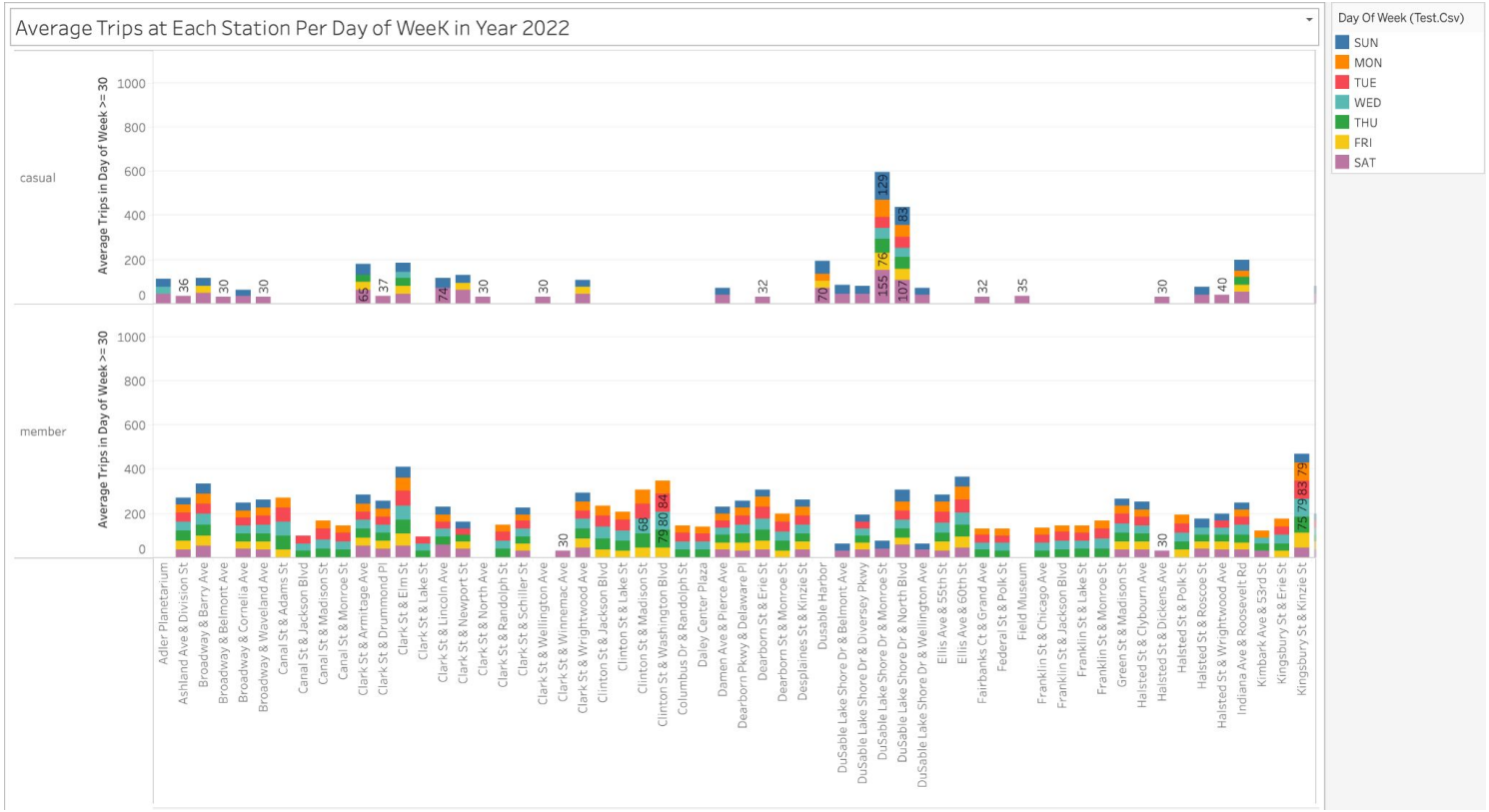
Member riders



Both groups align in the same stations

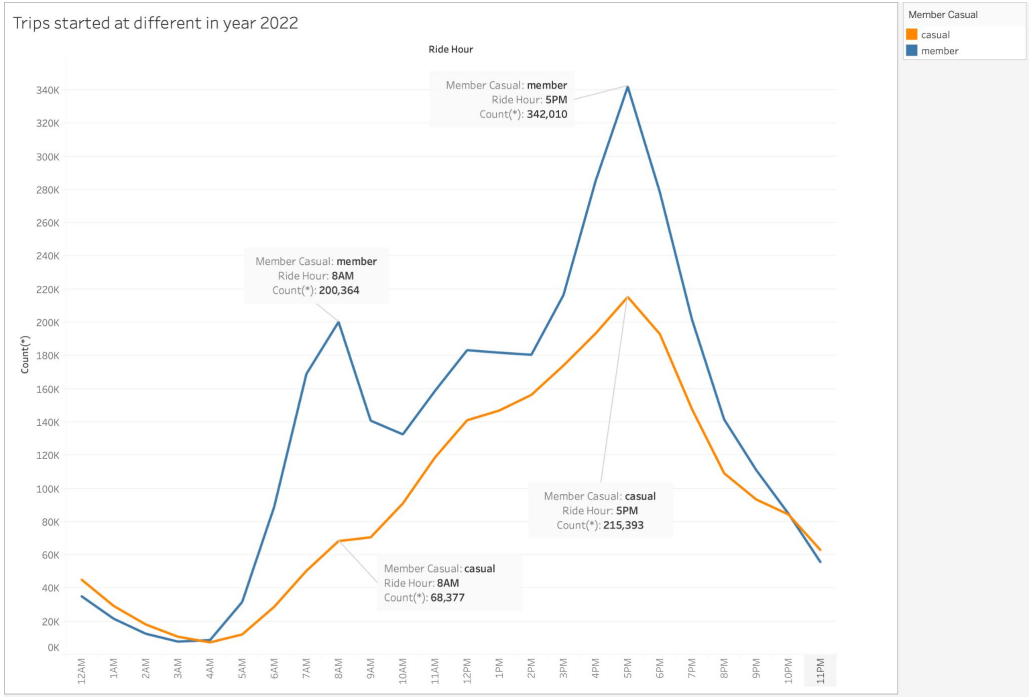


Both groups align in the same stations



Trips with station names

The member group had a higher number of trips began at 8 AM and 5 PM. It could be a lot of the Members were using the bike to commute as a tool for daily basis.



Conclusion

Based on this analysis, there was evidence that the Casual group made up many trips from tourists visiting, so it would not make sense to influence them to buy into members. The Casual group didn’t have a large number of riders on a weekday basis, which is from Monday to Sunday. But they tended to have longer ride lengths than the member group.

Recommendation

To confirm the above assumption was correct. The next step would be to conduct a small survey for the Casual from the seven stations where they displayed unusually high riders to prove that they weren't the resident. It would be hard to influence them to buy into the Cyclistic membership.

So it is possible to use a digital marketing strategy to influence the small number of Casual riders that live locally or visit locally often to buy into members.