Divvy Bike Share Analysis

Turning Casual Customer Into Full Customers

Introduction

Divvy is a bicycle sharing company in Chicago, Illinois operated by Lyft and owned by the Chicago Department of Transportation. Divvy has several docking stations spread across Chicago and allows users to ride either electric bikes or classic bikes. Each user falls into one of two categories: full member or casual member. A full member is one who pays an annual subscription to use Divvy bikes at any time during the subscription period, while a casual member is one who pays for either a 24-hour day-pass or for a single ride.

Business Task

The purpose of this analysis is to discover how full members and casual members use Divvy bikes differently. This information will ultimately be used to discover how to convert casual members into full members.

Data Source

The data used in this analysis is Divvy's system data, and is provided and collected by Divvy itself, subject to Divvy's Data License Agreement. The system data includes several ZIP folders, but I used only the following for this analysis:

- 202004-divvy-tripdata.zip
- 202005-divvy-tripdata.zip
- 202006-divvy-tripdata.zip
- 202007-divvy-tripdata.zip
- 202008-divvy-tripdata.zip
- 202009-divvy-tripdata.zip
- 202010-divvy-tripdata.zip
- 202011-divvy-tripdata.zip
- 202012-divvy-tripdata.zip
- 202101-divvy-tripdata.zip
- 202102-divvy-tripdata.zip
- 202103-divvy-tripdata.zip
- 202104-divvy-tripdata.zip
- 202105-divvy-tripdata.zip
- 202106-divvy-tripdata.zip
- 202107-divvy-tripdata.zip
- 202108-divvy-tripdata.zip

These folders cover data from April 2020 to August 2021. Each folder holds a CSV file containing the data for that month. The columns in the CSV files are as follows:

- ride_id - ID number for each ride
- rideable_type - classic_bike, electric_bike, or docked_bike. A docked bike is one located in a docking station
- started_at - date and time bike was unlocked
- ended_at - date and time bike was docked
- start_station_name - name of the station from which the bike was unlocked

- start_station_id - ID number of the start station
- end_station_name - name of the station to which the bike was docked
- end_station_id - ID number of the end station
- start_lat - latitude where bike was unlocked
- start_lng - longitude where bike was unlocked
- end_lat - latitude where bike was docked
- end_lng - longitude where bike was docked
- member_casual - casusal or member. The category of customer for that particular ride

Data Cleaning and Manipulation
I used the Python library Pandas to perform the following tasks. For access to the code, see the Appendix.

I added the following columns to each CSV file:

- ride_length - duration of the ride. Difference between ended_at and started_at
- day_of_week - day of the week the ride began, stored as an integer in [0..6], 0 for Monday, 6 for Sunday.

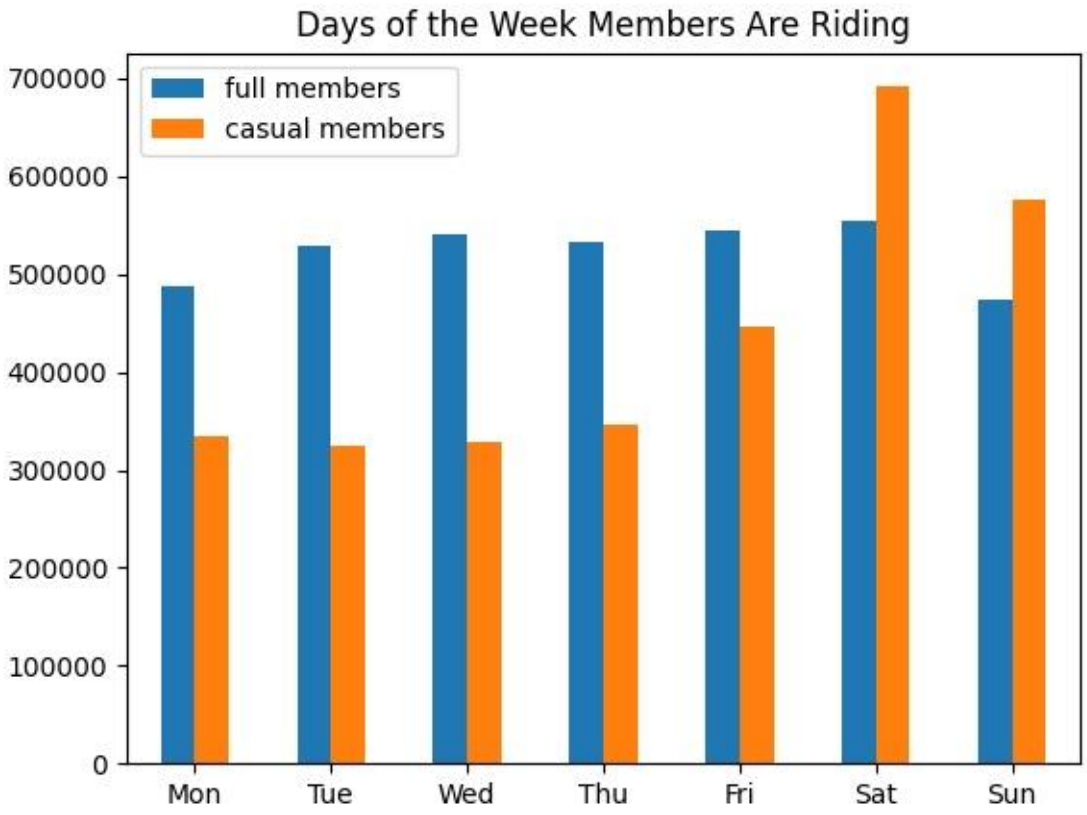
I then combined all the CSV files into a single CSV file. This combined file was the one I used for my analysis. I also imported this file into DB Browser for SQLite to create a DB file so that I could run SQLite queries on it.

There were several values missing from the start_station_name, end_station_name, start_station_id, and end_station_id columns. There was also a small number of missing values from the end_lat and end_lng columns. However, I did not use these fields in my analysis.

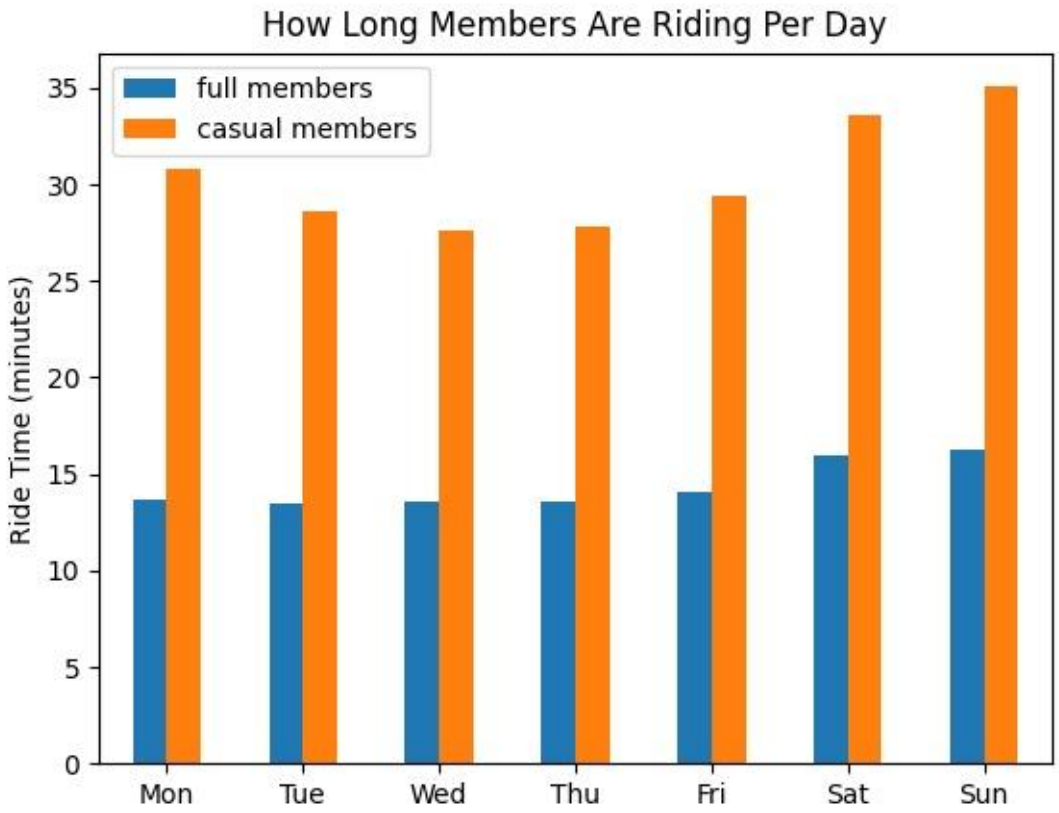
There were approximately 10,000 negative ride_length values, indicating invalid started_at or ended_at times. When creating charts using ride_length, I removed these and other outliers as defined as those values greater than three standard deviations from the mean.

Summary of Analysis
To discover how full members use Divvy bikes differently from casual members, I created several visualizations of the data. These visualizations showed that full members typically use Divvy on a routine, whereas

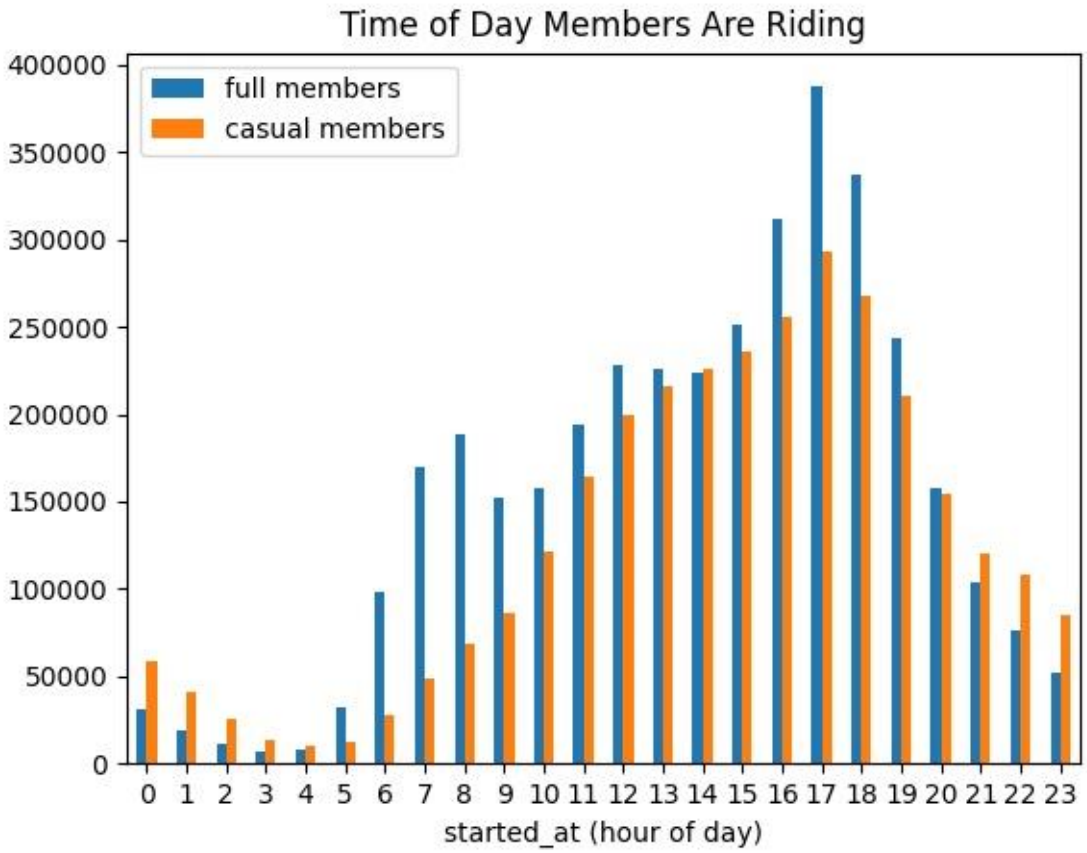
Key Findings
Full members tend to use Divvy at about the same rate every day of the week. Casual members, however, have relatively low usage on weekdays with a sharp increase in usage on the weekend.



Here, we see that casual members ride for about twice as long as full members. Overall, full members ride for about the same amount of time every day, whereas casual member ride times see a dip in the middle of the week with an increase on the weekend.



Rider usage typically peaks between 5:00pm and 6:00pm for both member groups. The key difference here is that full member usage sees a smaller peak from 7:00am to 9:00am.

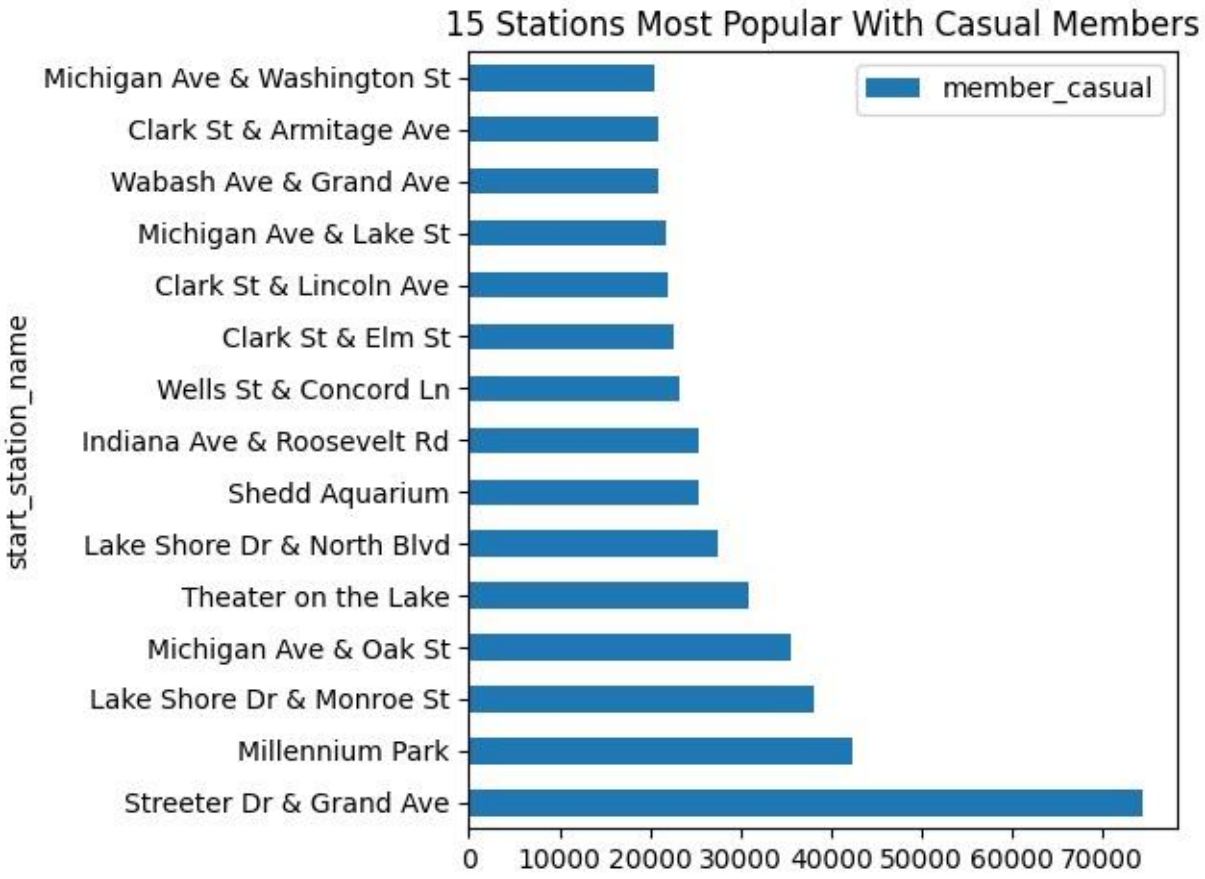


From this information, we see that full members tend to use Divvy more routinely and during standard commute times, but casual members don't have a clear routine in their Divvy usage. The trend among casual members is to use Divvy outside of work hours. From this, we can conclude that full members probably use Divvy for commutes, and casual members use Divvy for leisure.

Recommendations

Based off the results of this analysis, my top three recommendations are as follows:

1. In marketing copy, show casual riders that they can use Divvy to commute to work and that a full membership is the most cost effective option.
2. Emphasize health benefits, environmental benefits, traffic avoidance, and easier parking in marketing copy.
3. The below chart shows the 15 stations most frequented by casual members. Place physical advertisements at these stations.



Appendix

Code and a more detailed description of the analysis can be found here.