**Statement of Business tasks**

Cyclistic is a bike-share company that operates in the Chicago area. The marketing director has requested an analysis that compares the differences between members and casual riders. Members are defined as those who have an annual pass and casual riders or those who use a single or day pass to use the service. The company goal is to convert casual riders to members. An analysis of the two groups shows that members are more profitable. Based on the findings and recommendations, the marketing team will move forward with a plan.

**Data Sources Used**

The parameters of the study use the most recent 12-month history of their data and were provided in separate files. A review of each file showed a similar data structure. Python is used for performing this analysis because, the size of the data, once combined into one dataset, would be exceptionally large.

*The data sources for this project were provided by the client. A look at the datasets indicate that they are relevant and applicable to use. The data is provided at site* [*https://divvy-tripdata.s3.amazonaws.com/index.html*](https://divvy-tripdata.s3.amazonaws.com/index.html) *and is used under license. For this analysis, the client requested the twelve most recent months. This analysis includes April 2020 through March 2021.*

**Data Cleaning and Data Engineering**

Five fields were added to the dataset to assist the analysis. A **duration** column that calculates the elapsed time from start to end of the ride. **Distance** is a straight-line approximation of length traveled from the start and finish geographical coordinates. The additional columns are **day of week**, **month**, and **hour** that each ride started.

Examination of the data after these fields were added shows some issues. The duration column had negative elapsed times which is not feasible. Determining an effective and accurate way to correct the data was not viable, so I chose to drop those records. Some records had null fields resulting in those records being dropped. The total loss of transactions was less than 6% of the total amount of transaction over the 12-month period. Examination of the records dropped showed no particular month had the bulk of the issues.

**Summary of Analysis**

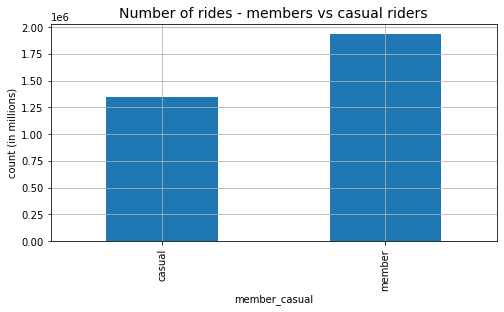
Members make up 58.95% of total rides versus 41.05% for casual riders. This holds true when looking at the data by day of the week. Member riders show more consistent with usage by day while the casual riders’ usage is bigger on weekends. Breaking this down further by hour shows data for member riders has two peaks, one at 9AM and a larger at 6PM. Data for casual riders shows a single peak at 6PM and much lower usage in late evening and morning hours. Data by day of week and hour shows that weekdays have this two peak pattern but display a similar pattern to casual rider usage on weekends.

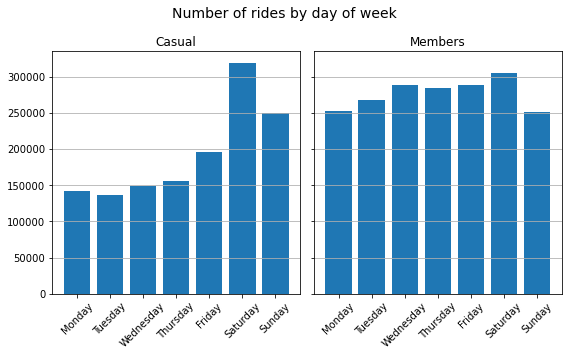
The data also shows that the duration average of casual riders is more than twice as long as the duration average of members. Yet both groups average a similar distance per trip. Data by day of week shows this is consistent for every day of the week. Taking less time to travel a similar distance indicates that members are, on average, traveling at a faster speed on trips. An assumption can indicate there is purpose for most member trips and are designed to get the rider from point A to point B as efficiently as possible. Casual riders’ trips may indicate the pace is more leisurely with different goals and intentions.

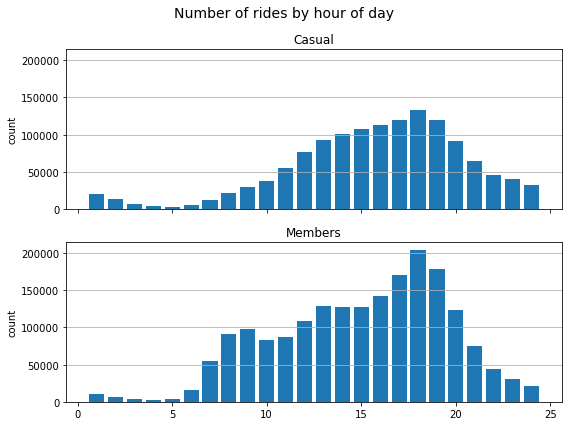
The company offers three types of bikes to their riders, classic, docked, and electric.t Both groups show a strong preference for the docked style. Even when comparing the data by day of the week, there is only a small percent of riders that use another type of bike than docked.

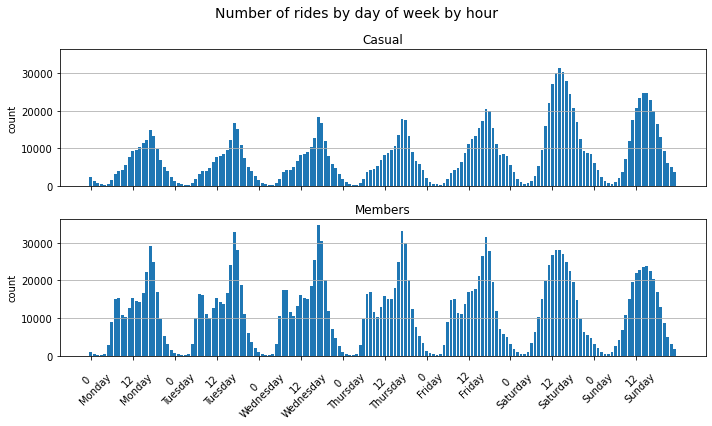
Finally, a review of the top 25 stations by number of rides by members and casual riders shows that more of the most popular stations for members are north of the Loop while most of the popular stations for casual riders are located near more urban areas and the city center. The data shows that many of the popular stations for members appear in areas that appear more residential.

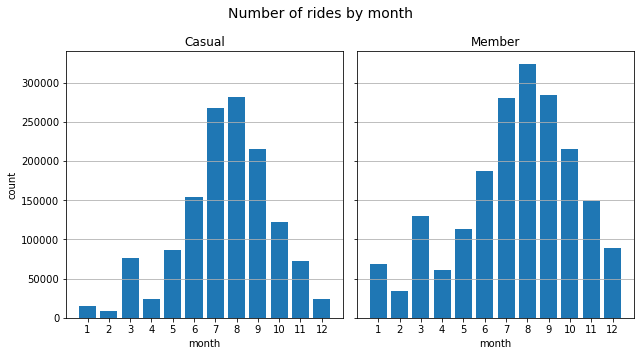
**Visuals**

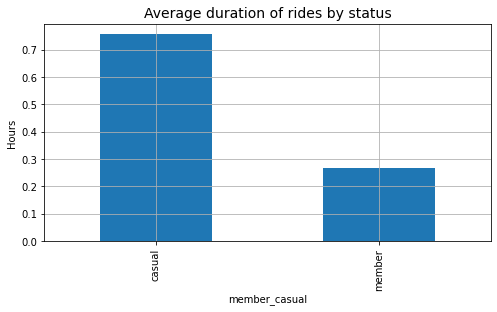


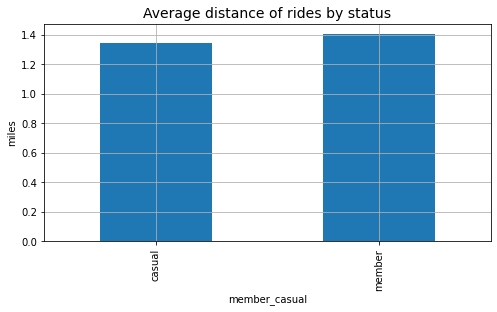


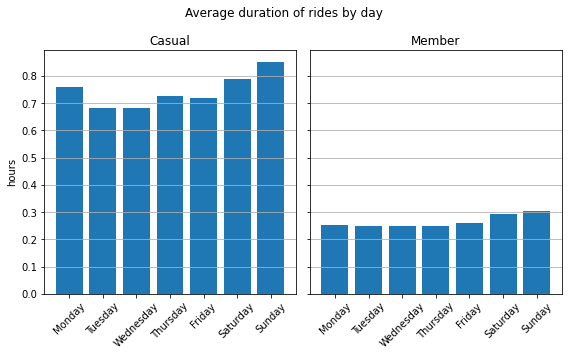


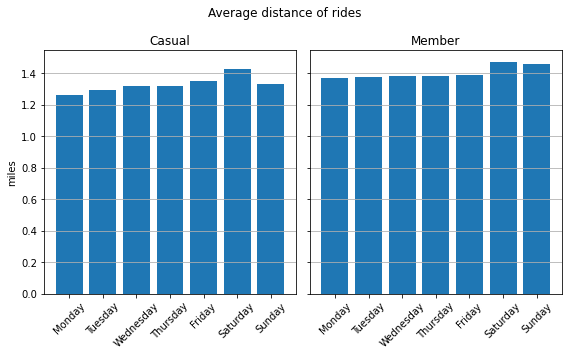


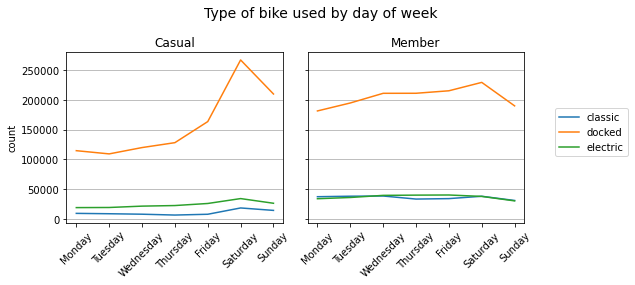












Map

Description automatically generated with low confidence

Top of map

A picture containing map

Description automatically generated

Bottom of map

**Recommendations**

1. Determine the percent of casual riders that are local, in the service area, versus those that do not live in the target area. This could include commuters coming from outside the target area on another means of transportation such as the ‘L’.
2. Find the casual local riders that originate rides at the more popular casual stations and determine the closest station to their home. See if some of these riders are aware of all the stations available to them.
3. What is the reason for the lower electric bike demand? Could pricing on electric bikes not be in line with what some riders are willing to pay for the type of bike that better suits their need? For example, if electric bikes cost less to use, would riders in business attire (suits, slacks, dresses, etc.) be willing to use the service? Does getting over heated or sweaty during the warm months limit regular use of the service for some casual riders? Would a rack or pouch to store items like jackets or purses add value? What price might entice some casual riders to convert to members with alternative annual offers?