**Cyclistic Data Analysis Report**

Cyclistic is a bike-share company in Chicago with over 5,800 bicycles and 600 docking stations. They offer different types of bicycles based on the needs of the rider. Riders can take a bike from one of the stations and return it to the same or a different station in the system at any time.

**Business Task**

Cyclistic wants to develop marketing strategies to convert casual riders to annual members. The key stakeholders are the Director of Marketing, Lily Moreno and the Executive Team.

The business question of our analysis is to find out how casual and annual riders use Cyclistic bikes differently.

**Prepare**

Historical data was chosen for our analysis: the 2020 Q1 data. A zip file with the dataset was downloaded from the Cyclistic database and saved on the computer. The dataset was then extracted as a CSV. file and saved in a folder. The dataset was opened using Excel and saved as an Excel workbook in a different file.

The dataset contained 13 columns and 426,796 rows. It contained information about the bike stations, type of bikes, type of customer, and start and end times. The integrity of the data was checked to ensure it was the correct data to answer our business question. The data was also checked for private information to avoid breaching the privacy of Cyclistic customers.

**Data cleaning and transformation**

The dataset was checked for null values and rows. Two rows had null values and were removed. There were also null values in the latitude and longitude columns but they were left as null values.

Two columns were created in the dataset. The first column to be created was the ride\_length column which was created by subtracting the end\_date from the start\_date column. The second column to be created was the day\_of\_week column, which was created by extracting the day of the week from the start\_date column These columns were critical for our analysis to find out how casual and annual riders differ from each other.

After creating the ride\_length column some cells did not return any results. Upon further investigation, it was discovered that dates in the start\_end and the end\_date columns were interchanged. This was corrected and the right results were returned in the ride\_length column. To ensure the ride\_length column was ready for our analysis rows where ride lengths were equal to 0:00:00 were deleted. The dataset was ready and used in the next process, data analysis.

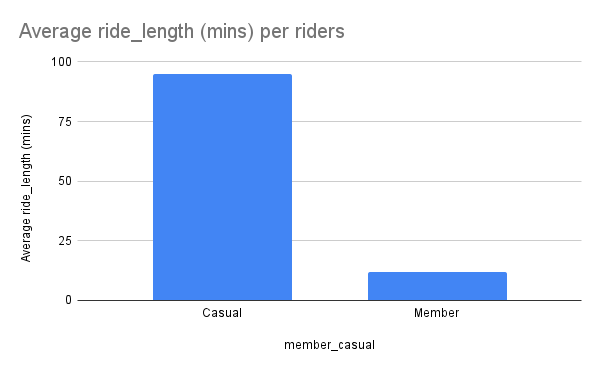
**Analysis summary**

Data analysis was conducted using Google Sheets and SQL in BigQuery. In Google Sheets, a pivot table was used to conduct some descriptive analysis of the data. The results of the analysis show the average ride length was 22 minutes. Most riders of Cyclistic take short rides on average. The average ride length for casual riders was 95 minutes and 12 minutes for annual riders. Casual riders seem to take longer rides compared to the annual riders.

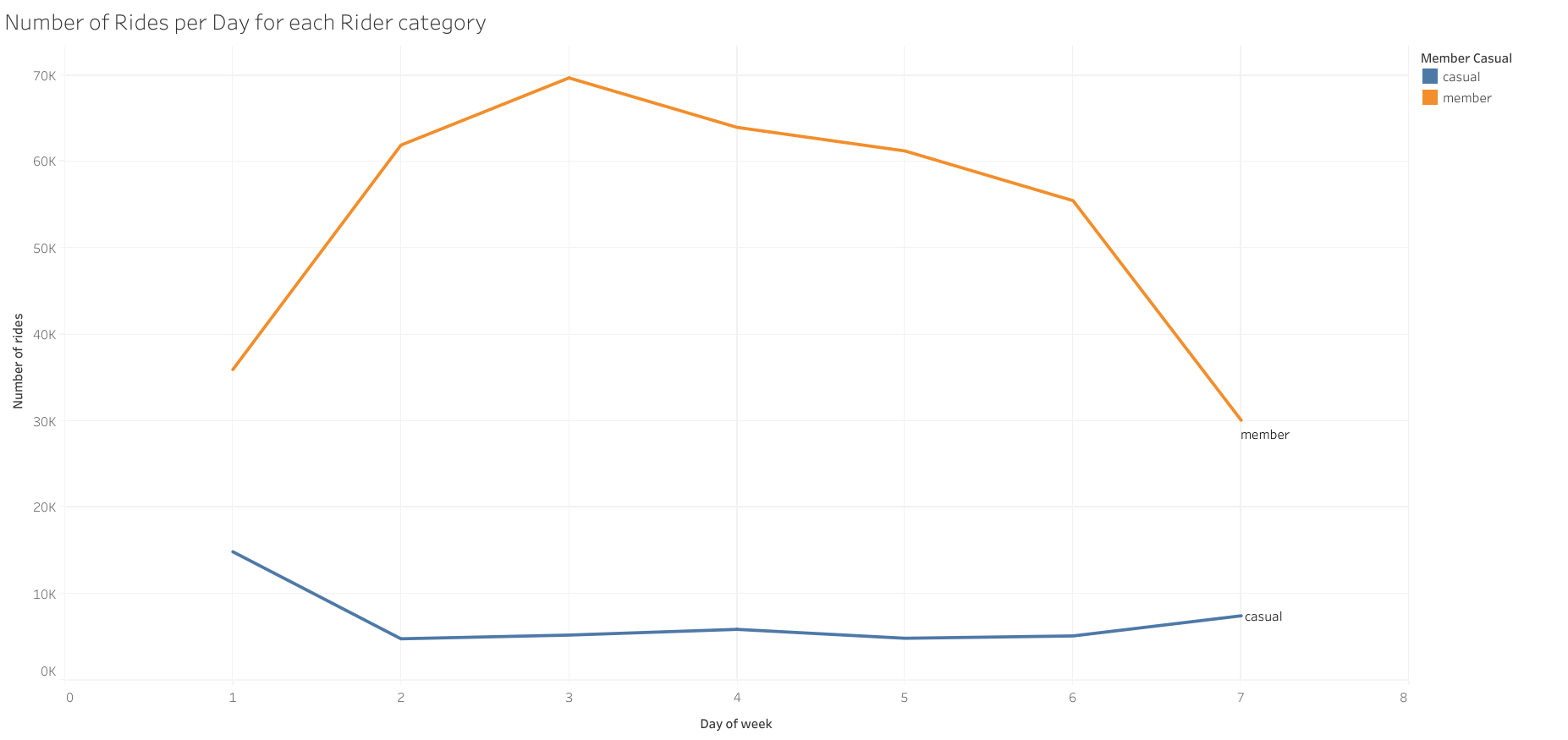
The results of the analysis show that casual riders take a lot of rides on Saturday and Sunday with 7480 and 14486 rides respectively. This shows casual riders prefer to take rides on weekends preferably for leisure. The results also show that annual riders take a lot of rides on weekdays with Tuesday being the highest with 69,697 rides. This shows annual riders are most likely taking rides to their work. The analysis also showed that most bicycles are taken on Tuesdays with a total of 74,948 rides. These findings helped answer the question of how casual and annual riders differ.

**Visualizations**

Google Sheets and Tableau were used to create visualizations for this analysis.



The figure above shows the average ride length of casual and annual riders.



The figure above shows the number of riders per day for each rider category.

**Recommendations**

Based on the analysis the following are the recommendations:

1. Give weekend memberships to casual members since most of them take rides on the weekends.
2. Offer discounted annual memberships to casual riders who take longer rides.
3. Offer one month free on annual memberships to all casual riders to attract them to become annual riders.

To enhance this analysis further additional data on the riders could be helpful. Such additional data could be the distance covered by the riders and the pricing of the different services offered by the company.

**Limitations**

The analysis had a few limitations:

1. The dataset used was only for 2020 Q1 hence might not give a picture of how riders are different in the whole year.
2. The dataset had more records of annual riders compared to casual riders. This may have affected the average ride lengths of the two rider categories and the overall average ride lengths.

Links

For the dataset click [here](https://docs.google.com/spreadsheets/d/1DN7kXBr_fP1DtZq7EoqZNTvR80t_5OAropjHI7FURbY/edit?usp=sharing)