1. OVERVIEW

This case study is based on the Google Data Analytics capstone project.

In this case study the analyst will work for a fictional company, Cyclistic, and meet different characters and team members. In order to answer the key business questions, the analyst will follow the steps of the data analysis process: **ask**, **prepare**, **process**, **analyse**, **share**, and **act**.

2. THE SCENARIO

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

3. THE BUSINESS TASK

To analyse the dataset using various techniques in order to provide insights that assist in answering the question of how annual members and casual riders use Cyclistic bikes differently.

4. THE DATA

The data for this case study was downloaded from the Divvy ride share dataset found at https://divvy-tripdata.s3.amazonaws.com/index.html. The information found on the home page of the company website (https://ride.divvybikes.com/) was used to better understand the dataset.

Twelve tables were downloaded consisting of informing ranging from January 2021 to December 2021. The tables consisted of the following columns;

- The ride identity, column name "ride_id",
- The type of bikes used, column name "rideable_type",
- The date and time when the ride began, column name "started_at",
- The date and time when the ride ended, column name "ended_at",
- The station name where the ride began, column name "start_station_name",
- The station identity where the ride began, column name "start_station_id",
- The station name where the ride ended, column name "ended_station_name",
- The station identity where the ride ended, column name "ended_station_id",

- The latitude point where the ride began, column name "start_lat",
- The longitude point where the ride began, column name "start_lng",
- The latitude point where the ride ended, column name "end_lat",
- The longitude point where the ride ended, column name "end_lng",
- The type of rider (whether casual or member), column name "member_casual"

5. DATA MANIPULATION

BigQuery and RStudio were the main platforms used for analysing the data.

The full code for this case study can be found at https://github.com/Kgothatso-K/Case-Study-Bike-Share.

The twelve tables were combined into one annual dataset consisting of 13 columns and 5,595,063 rows.

At first glance empty spaces were observed in the columns relating to the station names and identities. The dates, months, weekdays and times were extracted from the date-time stamps. The duration of each trip was calculated from the ride start and end times. The distance between where the ride began and where it ended was calculated from the latitude and longitude positions (this was done in BigQuery). Empty spaces were identified in 4,771 rows. The empty spaces were as a result of partial or incomplete trip data.

A new table consisting of 5,590,292 rows and 17 columns was created. The table consisted of the following columns;

- The ride identity, column name "ride_id",
- The type of bikes used, column name "rideable_type",
- The date when the ride began, column name "start_date",
- The month when the ride began, column name "start_month_string",
- The month when the ride began, column name "start_month_int",
- The day when the ride began, column name "start_day_string",
- The day when the ride began, column name "start_day_int",
- The time when the ride began, column name "start_time",
- The date when the ride ended, column name "end_date",
- The month when the ride ended, column name "end_month_string",
- The month when the ride ended, column name "end_month_int",
- The day when the ride ended, column name "end_day_string",
- The day when the ride ended, column name "end_day_int",
- The time when the ride ended, column name "end_time",
- The trip duration in seconds based on the recorded start and end times of the trip, column name, "trip_length",

- The distance in meters between the stations where the trips began and ended, column name "trip_displacement",
- The type of rider (whether casual or member), column name "member_casual"

The column names were not changed as it was not considered material to the analysis. The station names, identities, and coordinate points were also excluded as they too were deemed irrelevant to the analysis.

6. SUMMARY OF ANALYSIS

The analysis was first conducted based on the two types of riders, i.e. member riders and casual riders, and how they used bikes over various time periods. The analysis was expanded to analyse the types of members in relation to the types of bikes used.

7. KEY FINDINGS

These findings will be restricted to start trip data as there is no material difference between the start and end trip data trends.

7.1. The company had 3,064,777 trips from annual members, and 2,525,515 trips from casual riders,

7.2. As shown in figure 7.2.1, both annual members and casual riders favour the classic bikes, but only casual riders favour docked bikes.

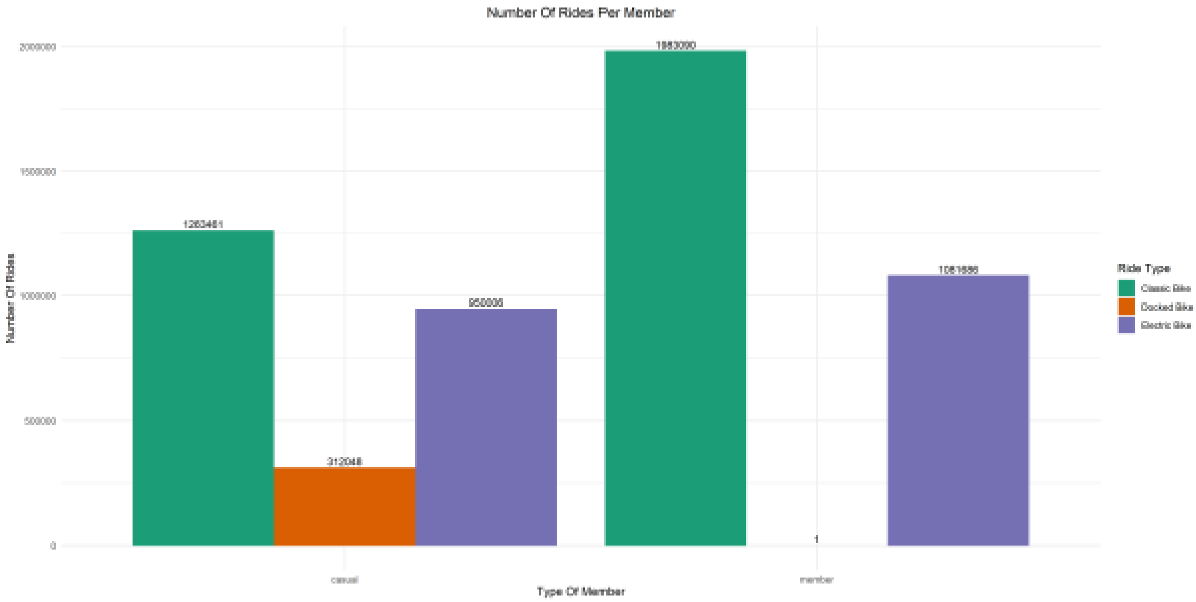


Figure 7.2.1 - Bar Graph Rides Per Member

7.3. Annual members who use classic bikes prefer to take monthly trips between June and October. Most trips for annual classic bike riders took place in August, but most trips for casual classic bike riders took place in July. Annual electric bike riders had an increasing number of trips from February to October with the most trips having occurred in October. Casual electric bike riders had an increasing number of trips from February to July with the most trips having occurred in July. See the below figure 7.3.1.

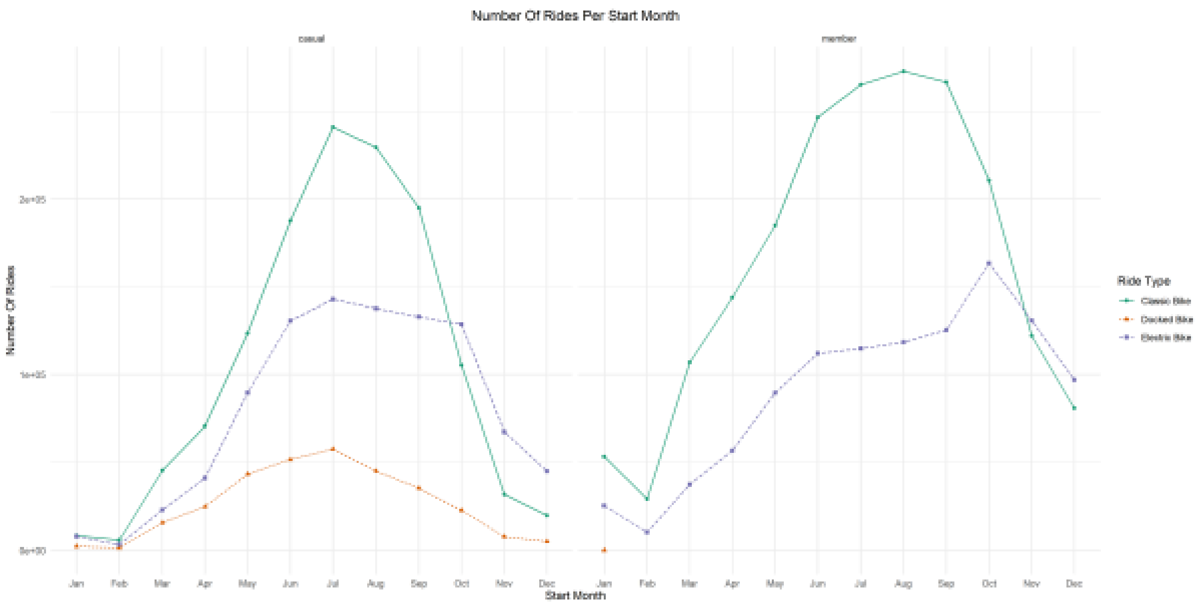


Figure 7.3.1 – Line Graph Rides Per Start Month

Both casual riders and annual members prefer electric bikes for monthly trips covering greater distances. See the below figure 7.3.2.

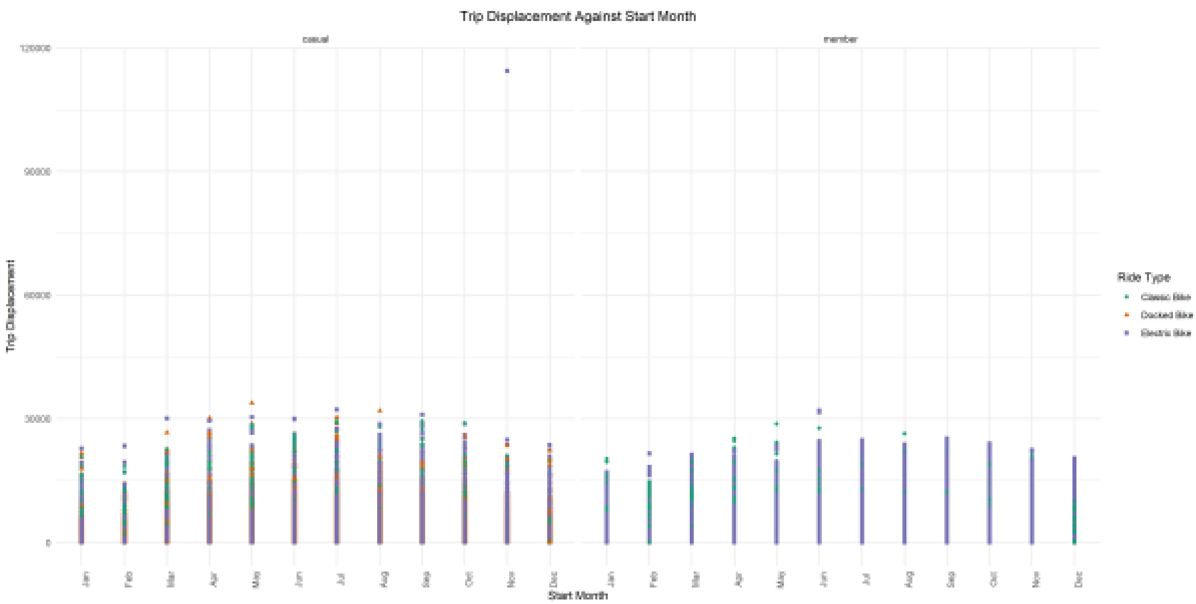


Figure 7.3.2 – Scatterplot Trip Displacement Per Start Month

Casual riders used docked bikes for greater monthly trip lengths while annual members use casual bikes for greater trip lengths. See the below figure 7.3.3.

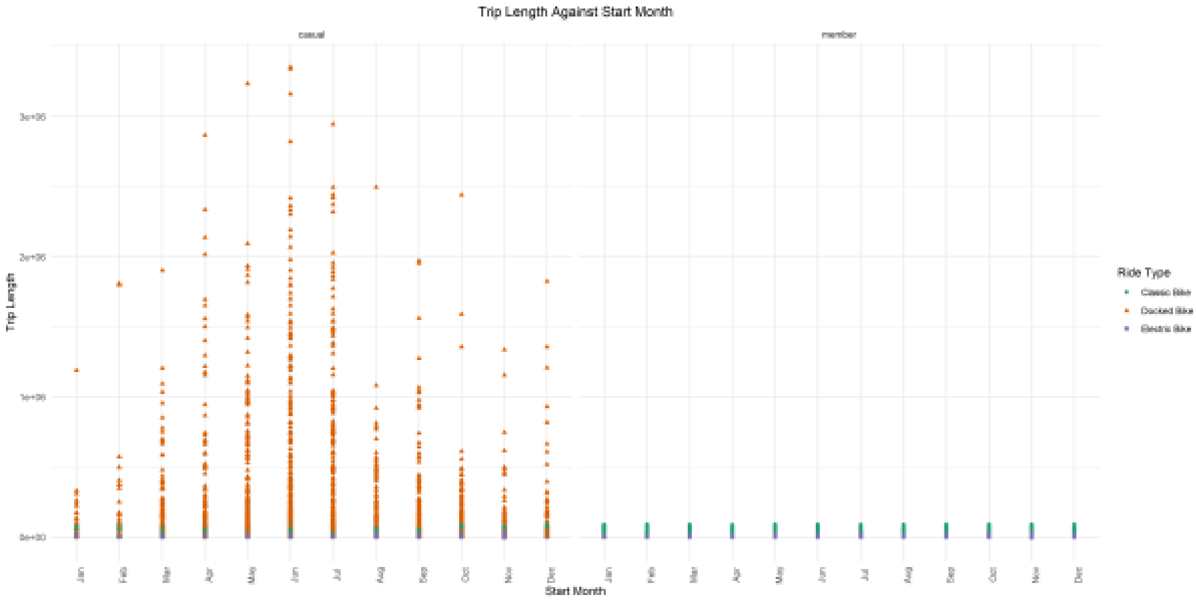


Figure 7.3.3 – Scatterplot Trip Length Per Start Month

7.4. Casual riders went on more daily trips during the weekend, while annual members went on more trips during the week. See the below figure 7.4.1.

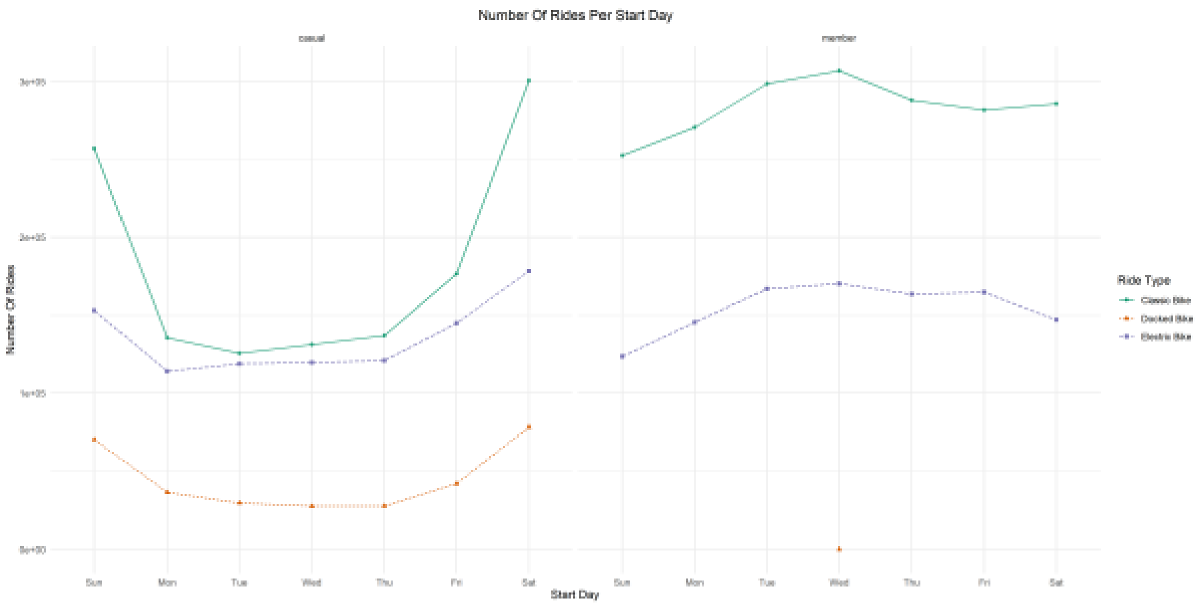


Figure 7.4.1 – Line Graph Rides Per Start Day

Both casual riders and annual members prefer electric bikes for daily trips covering greater distances. See the below figure 7.4.2.

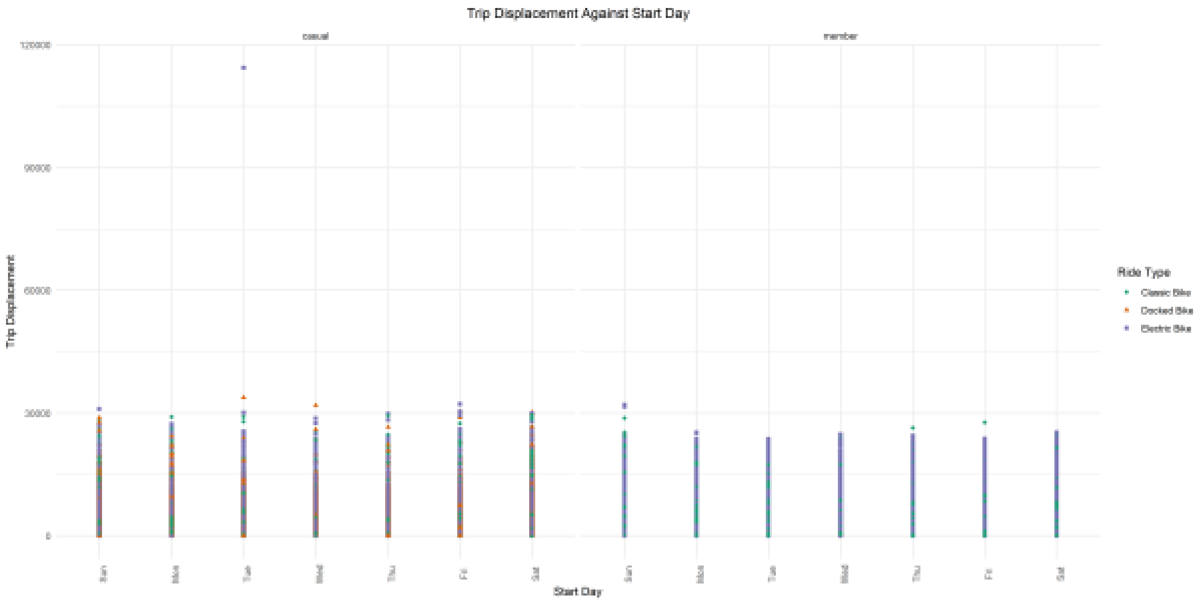


Figure 7.4.2 – Scatterplot Trip Displacement Per Start Day

Casual riders used docked bikes for greater daily trip lengths while annual members use casual bikes for greater trip lengths. See the below figure 7.4.3.

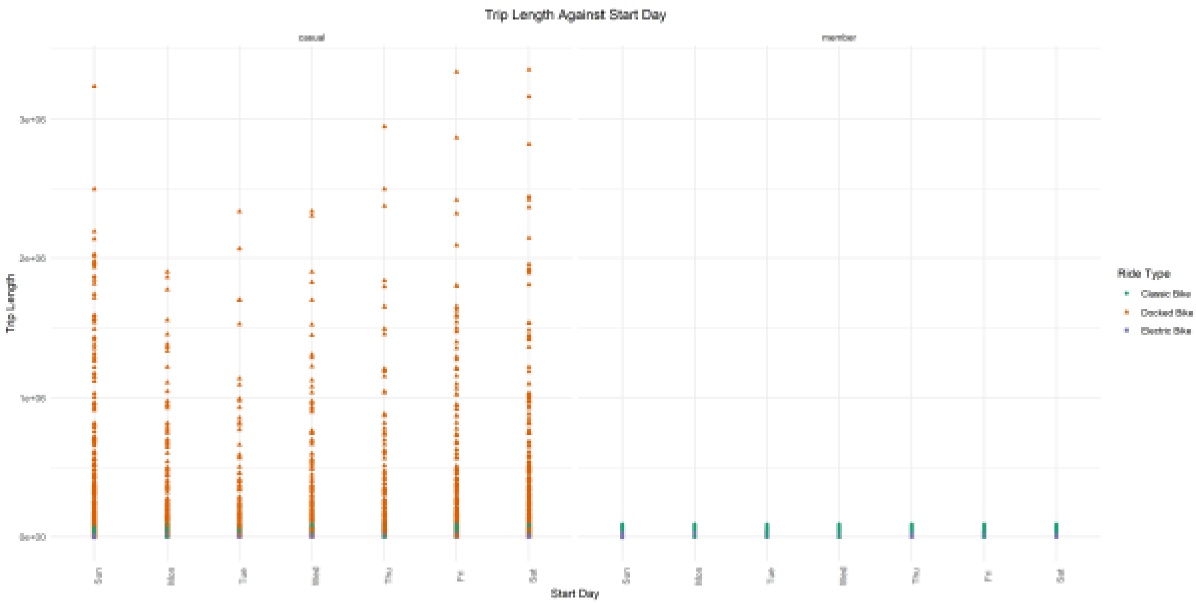
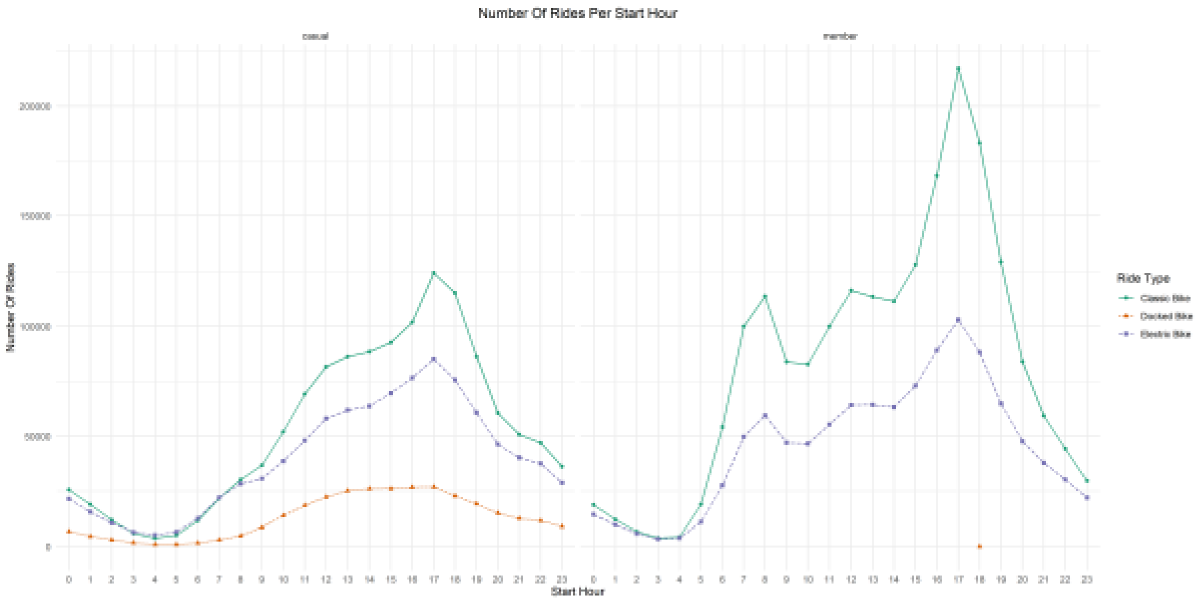
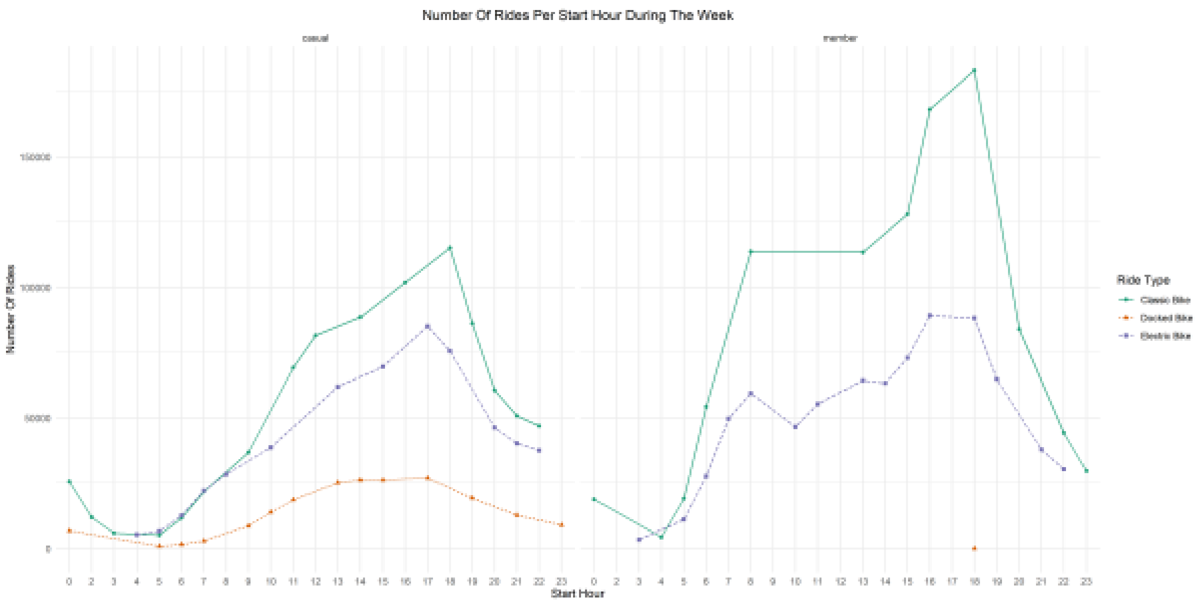


Figure 7.4.3 – Scatterplot Trip Length Per Start Day

7.5. Casual riders predominantly go on trips in the late afternoon to early evening while annual members go on trips in the early morning but mostly in the late afternoon to early evening. See the below figure 7.5.1 and figure 7.5.2.



[*Figure 7.5.1 – Line Graph Rides Per Start Hour*]



[*Figure 7.5.2 – Line Graph Rides Per Start Hour - Midweek*]

On weekends, classic bike casual riders stop going on trips from 15h00 while annual members stop going on trips from 21h00. See the below figure 7.5.3.

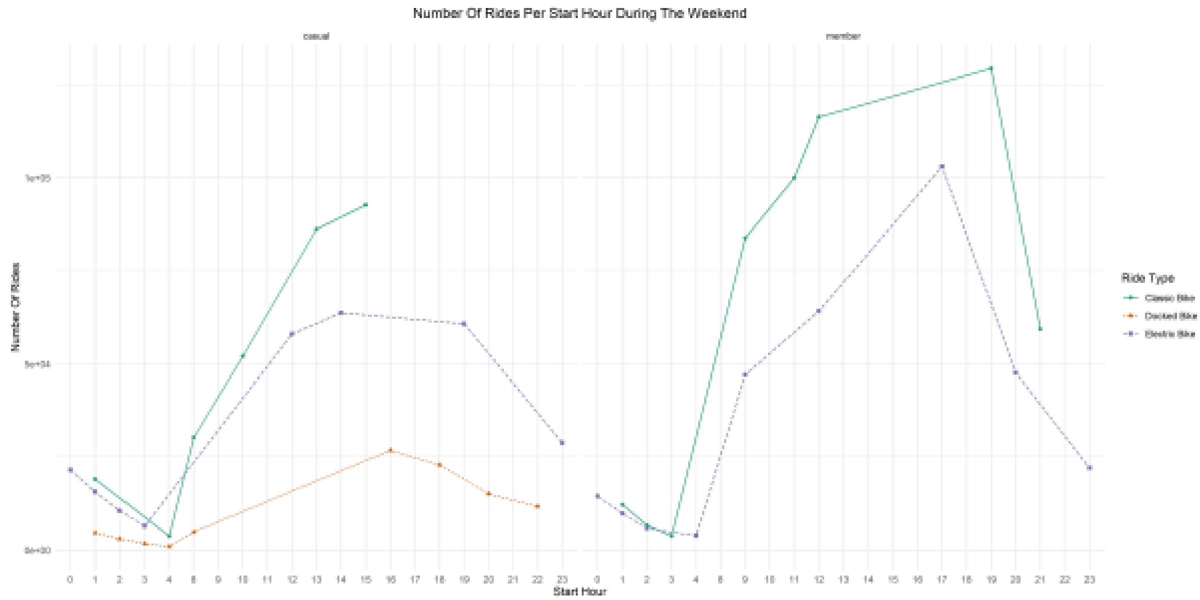


Figure 7.5.3 – Line Graph Rides Per Start Hour - Midweek

7.6. On average, casual riders take longer same-day trips than annual members. Casual riders also drop-off their bikes further from their pick-up stations.

7.7. On average casual riders take longer multi-day trips than annual members. Annual members however drop-off their bikes further from their pick-up stations.

7.8. Ultimately there exists no linearity between the variables.

8. RECOMMENDATIONS

8.1. Obtain qualitative data explaining what riders like and dislike about the docked bike. The data may be used to improve the designs of the electric and classic bikes. The campaign should highlight the benefits of these new designs while structuring a package for annual members.

8.2. The campaign should highlight the benefits of using electric bikes to cover greater distances.

8.3. The campaign should highlight the advantages of using bikes for early morning commute and late afternoon commute, along with the advantages of using the bikes for weekend commutes.

8.4. The campaign should target peak seasons when more riders are using bikes, these are the warmer seasons of the year, i.e. spring and summer. Casual riders are more likely to be reached by campaigns during these seasons.