How does Bike-Share navigate speedy success?

The task is to analyse Cyclistic’s historical bike trip data to understand how annual members and casual riders use the bikes differently. By identifying these usage patterns, we would be able to uncover insights that will help design targeted marketing strategies that convert casual riders to annual members.

The data sources used for this analysis include Cyclistic’s historical bike trip data, which contains information about bike rides taken by both annual members and casual riders. The data includes attributes such as trip duration, start and end times, bike types used, and customer type.

Before conducting the analysis, the data required some cleaning and manipulation to ensure its quality and relevance.

* Handling missing or null values: Any missing or null values were addressed appropriately to avoid bias in the analysis.
* Data type conversion: Ensuring that the data types are consistent and accurate for analysis.
* Removing duplicates: Eliminating duplicate records, if any, to avoid duplicate counting in the analysis

The analysis focused on comparing the bike usage patterns between annual members and casual riders based on key metrics such as trip duration, ride frequency, and preferred bike types. This allowed us to identify trends and differences in their behaviour.

I used Cyclistic’s historical bike trip data from the last 12 months, which is available on their website. All data is stored in spreadsheet. There are a total of 12 spreadsheet files.

202207-divvy-tripdata.csv

202208-divvy-tripdata.csv

202209-divvy-publictripdata.csv

202210-divvy-tripdata.csv

202211-divvy-tripdata.csv

202212-divvy-tripdata.csv

202301-divvy-tripdata.csv

202302-divvy-tripdata.csv

202303-divvy-tripdata.csv

202304-divvy-tripdata.csv

202305-divvy-tripdata.csv

202306-divvy-tripdata.csv

After downloading the 12 zip files, I carried the data cleaning process by changing the format of the dates to YYYY-MM-DD, and times to HH:MM:SS.

Once sorting the data on excel was completed, I loaded the packages onto RStudio and created data frames for each of the 12 files.

data1 <- read\_csv(“202207-divvy-tripdata.csv”)

data2 <- read\_csv(“202208-divvy-tripdata.csv”)

data3 <- read\_csv(“202209-divvy-publictripdata.csv”)

data4 <- read\_csv(“202210-divvy-tripdata.csv”)

data5 <- read\_csv(“202211-divvy-tripdata.csv”)

data6 <- read\_csv(“202212-divvy-tripdata.csv”)

data7 <- read\_csv(“202301-divvy-tripdata.csv”)

data8 <- read\_csv(“202302-divvy-tripdata.csv”)

data9 <- read\_csv(“202303-divvy-tripdata.csv”)

data10 <- read\_csv(“202304-divvy-tripdata.csv”)

data11 <- read\_csv(“202305-divvy-tripdata.csv”)

data12 <- read\_csv(“202306-divvy-tripdata.csv”)

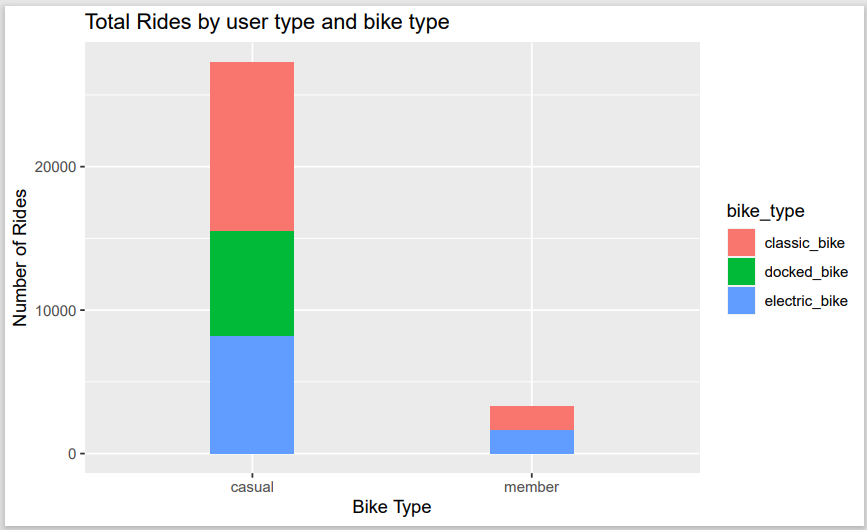
I then combined the 12 dataframes into one using the rbind() function and explored the data using the following functions: select(), nrow(), ncol(), length(), head(), tail(), glimpse(), str(), summary(), names(), rownames(), skim\_without\_charts(), and View().

I added a new columns:

bike\_rides$ride\_length\_min <- round(as.numeric(difftime(bike\_rides$ended\_at, bike\_rides$started\_at, units = “mins”)), 2

I then used this same function to create columns for dates, months, years, days of the week, hours, seasons.

After reviewing the data, I noticed there were some values that were false or invalid. To resolve this issue, I used the select() function to create a data frame with only selected columns. I then removed rides that were less than 60 seconds, where end\_lat & end\_lng are both missing, and rides related to test and repair stations.

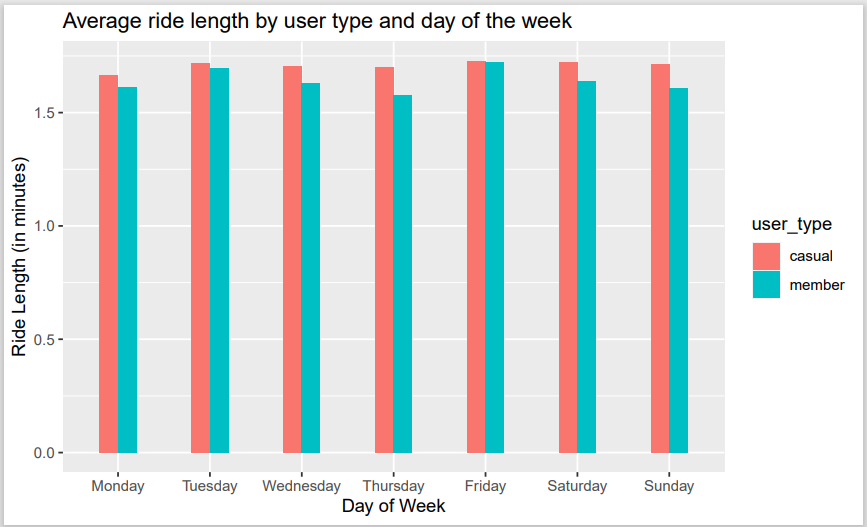


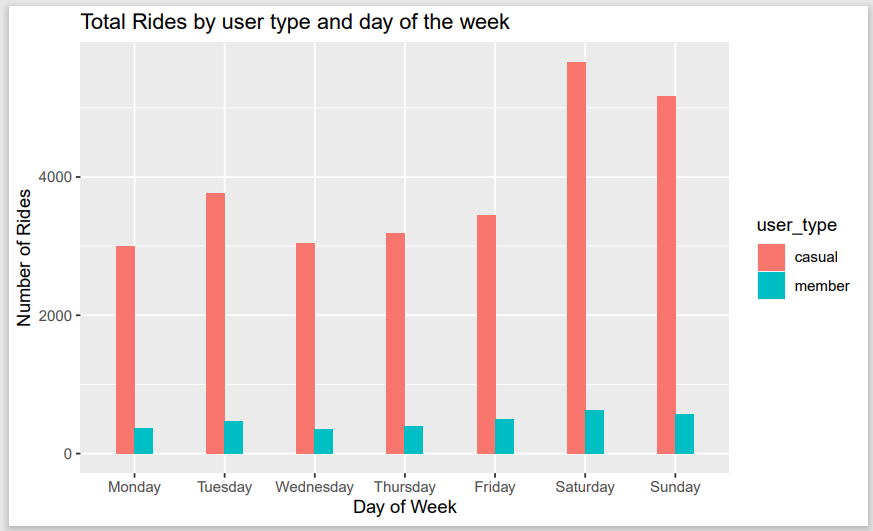
A graph of a number of different colored squares

Description automatically generated

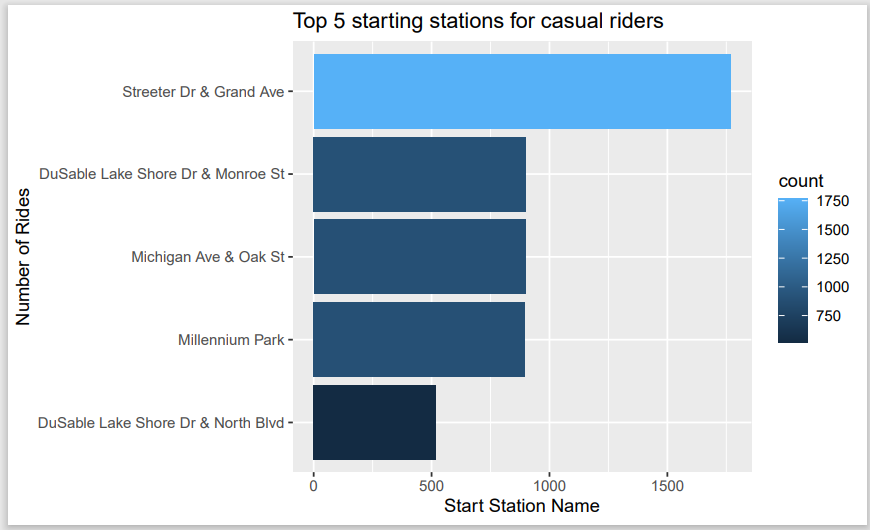
A graph of a number of different colored bars

Description automatically generated with medium confidence









Based on the graphs, it can be noted that casual users ride bikes more often than the member. However, the average ride length for both user types is more than 1.5 in all days of the week. It can also be noted the most popular starting station for casual riders is Streeter Dr & Grand Ave.