Cyclistic: Bike-Share

How Does A Bike-share Navigate Speedy Success?

Rishika Gour

Google Data Analytics Cyclistic Capstone

Introduction

- Welcome to the Cyclistic bike-share analysis case study.
- In this case study, I have performed many real-world tasks of a junior data analyst. I worked for a fictional company, Cyclistic, and met different characters and team members. In order to answer the key business questions, I followed the steps of the data analysis process: ask, prepare, process, analyze, share, and act.

Statement of the Business Task

- Analyze the data and identify trends to understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics. Insights from the analysis will be used to design marketing strategies aimed at converting casual riders into annual members.

Key Stakeholders

- Cyclistic: A bike-share program that features more than 5,800 bicycles and 600 docking stations.
- Lily Moreno: The director of marketing and your manager.
- Cyclistic marketing analytics team: A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy.
- Cyclistic executive team: The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

Data Source

- The data used in the analysis is from previous twelve months i.e., from November 2021 to October 2022. It is downloaded from Divvy Data. Each file was originally compressed in a .zip file format. The .zip files were named according to the month in the year the data was collected. For example, the file containing the data from December 2021 was named “202112-divvy-tripdata.zip. The data has been made available by Motivate International Inc. under this data license agreement. The data-privacy issues prohibit from using riders’ personally identifiable information. The data is separated by month, each on its own csv. The population of the dataset is its own clients as bike riders. It has full credibility for the same reason. It is ROCCC because it is reliable, original, comprehensive, current, and cited. All the files have consistent columns and each column has the correct type of data.

Tools

- Used Excel, RStudio and Power BI for the cleaning, analysis and visualization of the data.
- Sorted, filtered and formatted the data using Excel since it is a simple program to use for sorting and filtering.
- Further organized and analyzed data using RStudio.
- Created the visuals using Power BI since it can create compelling visuals in a few simple steps.

Cleaning or Manipulation of data
Step 1: Collect Data

```
X1_202111 <- read_csv("1-202111-divvy-tripdata.csv")
X2_202112 <- read.csv("2-202112-divvy-tripdata.csv")
X3_202201 <- read.csv("3-202201-divvy-tripdata.csv")
X4_202202 <- read.csv("4-202202-divvy-tripdata.csv")
X5_202203 <- read.csv("5-202203-divvy-tripdata.csv")
X6_202204 <- read.csv("6-202204-divvy-tripdata.csv")
X7_202205 <- read.csv("7-202205-divvy-tripdata.csv")
X8_202206 <- read.csv("8-202206-divvy-tripdata.csv")
X9_202207 <- read.csv("9-202207-divvy-tripdata.csv")
X10_202208 <- read.csv("10-202208-divvy-tripdata.csv")
X11_202209 <- read.csv("11-202209-divvy-tripdata.csv")
X12_202210 <- read.csv("12-202210-divvy-tripdata.csv")
```

Step 2: Wrangle Data And Combine Into A Single File

- Compared column names each of the files.
- While the names don't have to be in the same order, they do need to match perfectly before we can use a command to join them into one file.
- colnames(X1_202111) colnames(X2_202112) colnames(X3_202201) colnames(X4_202202) colnames(X5_202203) colnames(X6_202204) colnames(X7_202205) colnames(X8_202206) colnames(X9_202207) colnames(X10_202208) colnames(X11_202209) colnames(X12_202210)
- Renamed columns to make them consistent with X12_202210 (as this will be the supposed going-forward table design for Divvy)
- Inspected the dataframes and look for incongruencies.
- Converted ride_id and rideable_type to character so that they can stack correctly.
- Stacked individual month's data frames into one big data frame.
- Removed lat and long.

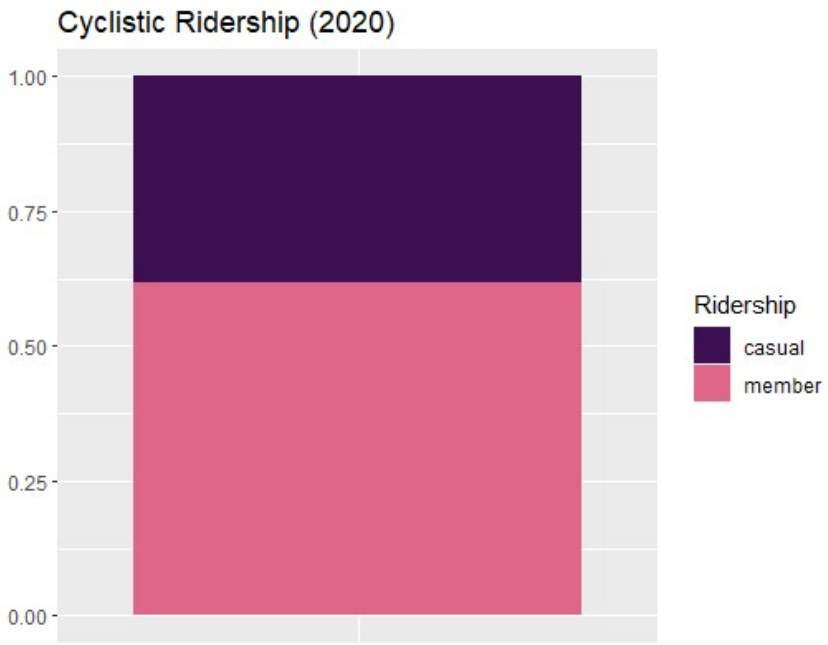
Step 3: Cleaning Up And Adding Data To Prepare For Analysis

- Inspected the new table that has been created.
- colnames(all_trips) #List of column names
- nrow(all_trips) #How many rows are in data frame?
- dim(all_trips) #Dimensions of the data frame?
- head(all_trips) #See the first 6 rows of data frame. Also tail(all_trips)
- str(all_trips) #See list of columns and data types (numeric, character, etc)
- summary(all_trips) #Statistical summary of data
- Reassigned to the desired values.
- Checked to make sure the proper number of observations were reassigned.
- Added columns that list the date, month, day, and year of each ride.
- Added a "ride_length" calculation to all_trips.
- Inspected the structure of the columns.
- Converted "ride_length" from Factor to numeric so we can run calculations on the data
- Removed "bad" data.
- Created a new version of the dataframe (v2) since data is being removed.

Step 4: Conducting Descriptive Analysis

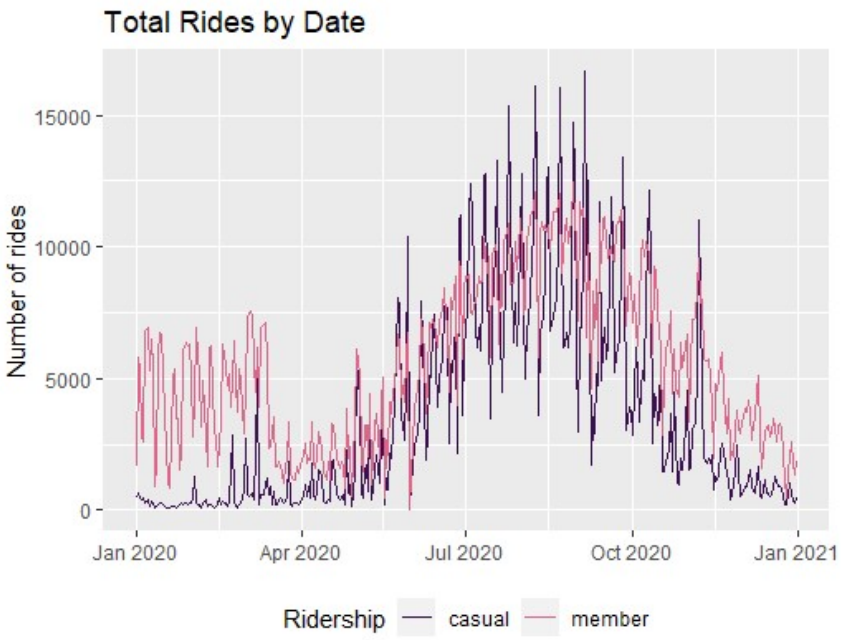
- Conducted Descriptive analysis on ride_length.
- Compared members and casual users.
- Found the average ride time by each day for members vs casual users.
- Analyzed ridership data by type and weekday.
- Created weekday field using wday().
- Calculated the number of rides and average duration .
- Calculated the average duration.

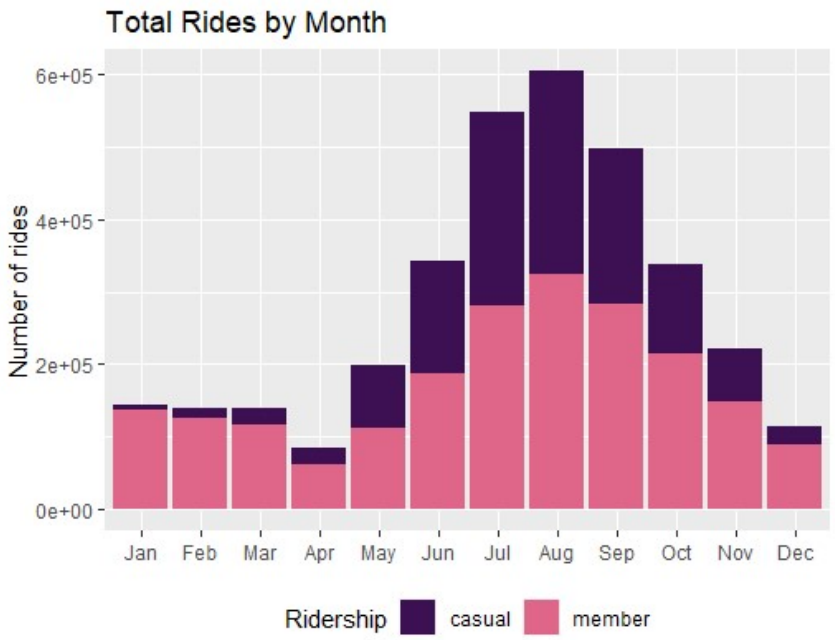
Visualizations and Key Findings
Proportion Of Casual and Member Riders



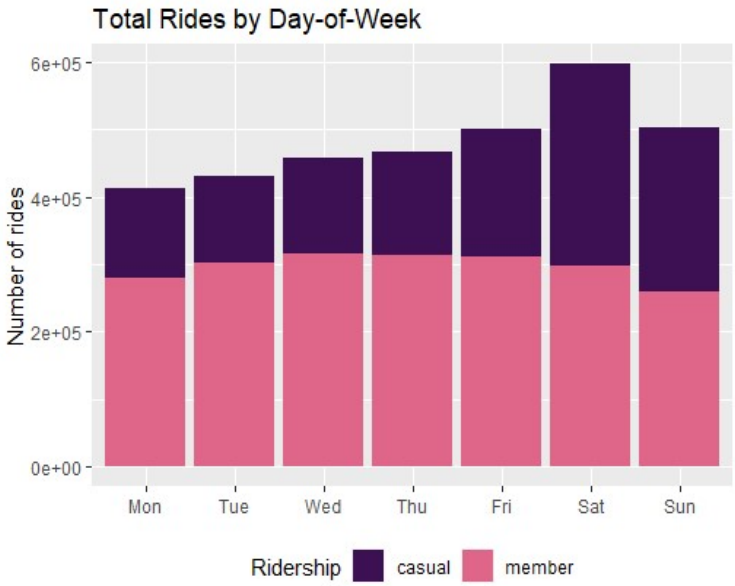
We can see that more than a half of the cyclistic users has chosen the membership option. Members ride more as compared to the casual riders.

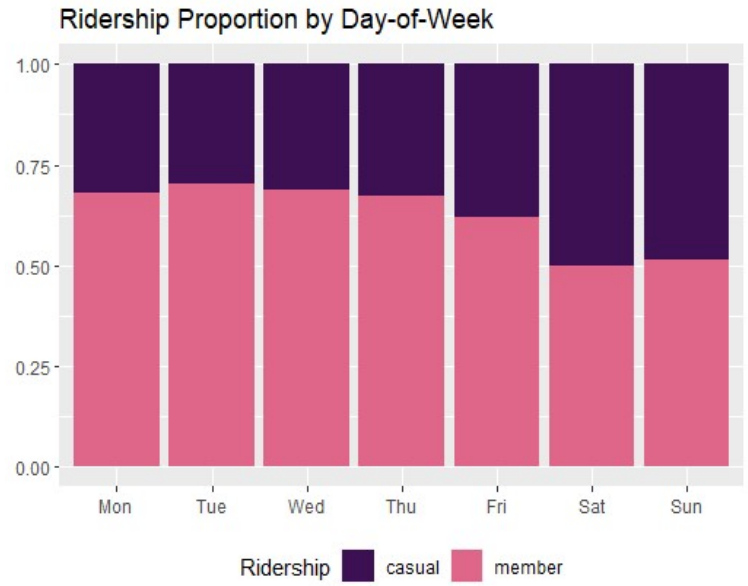
Proportion of Daily Rides and Monthly Rides





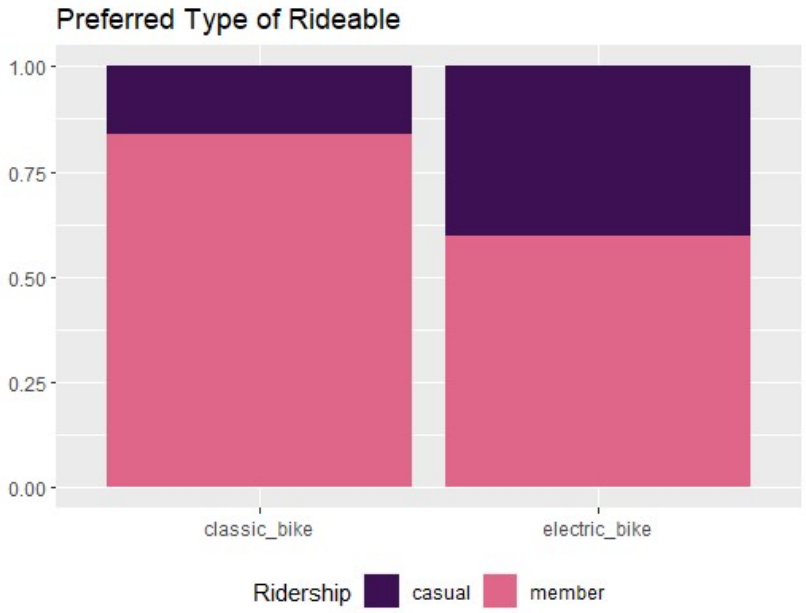
Looking at rides by day of the week, we see that total rides increases during the weekend, especially on Saturday. Members are more consistent in their use of cyclistic during weekdays than casual riders, with 60-70% of total rides per day. During the weekend, casual riders increase their riding activity by 20%.

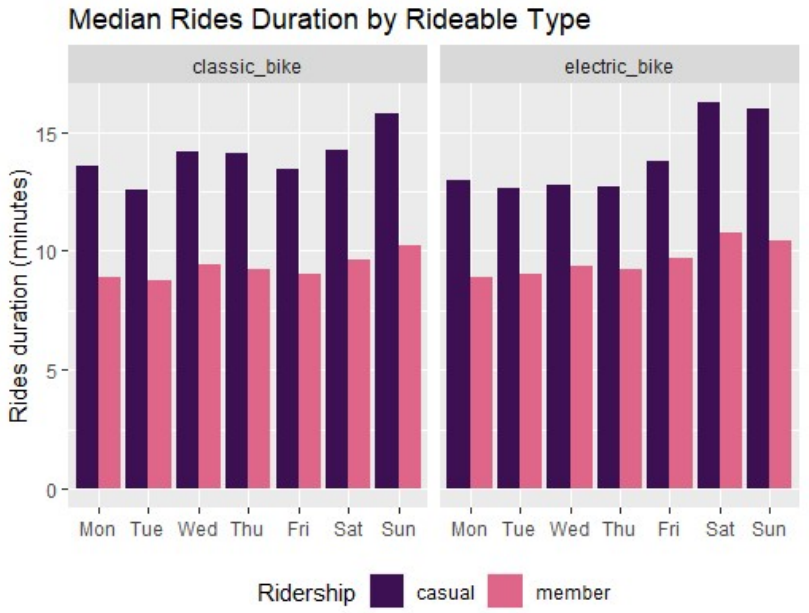




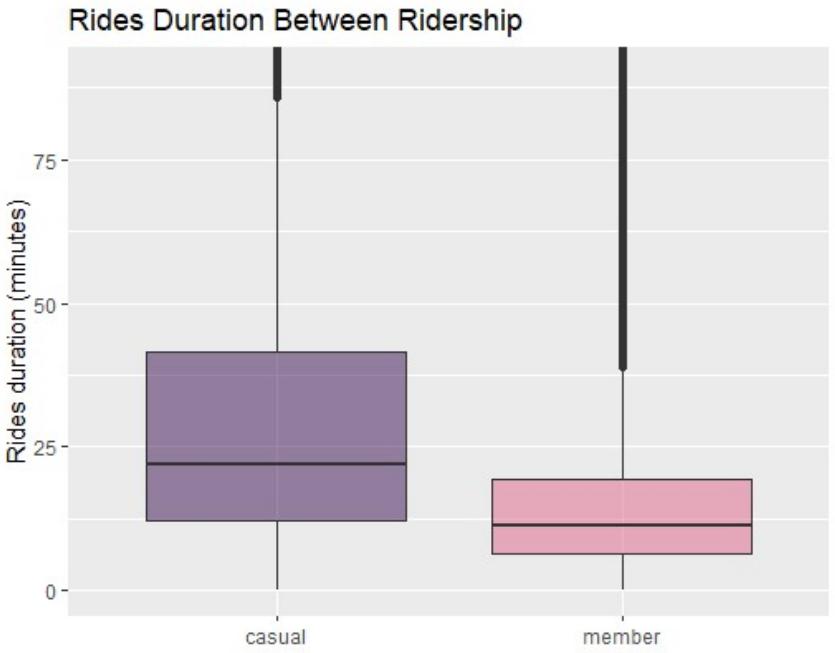
Proportion of Rideables

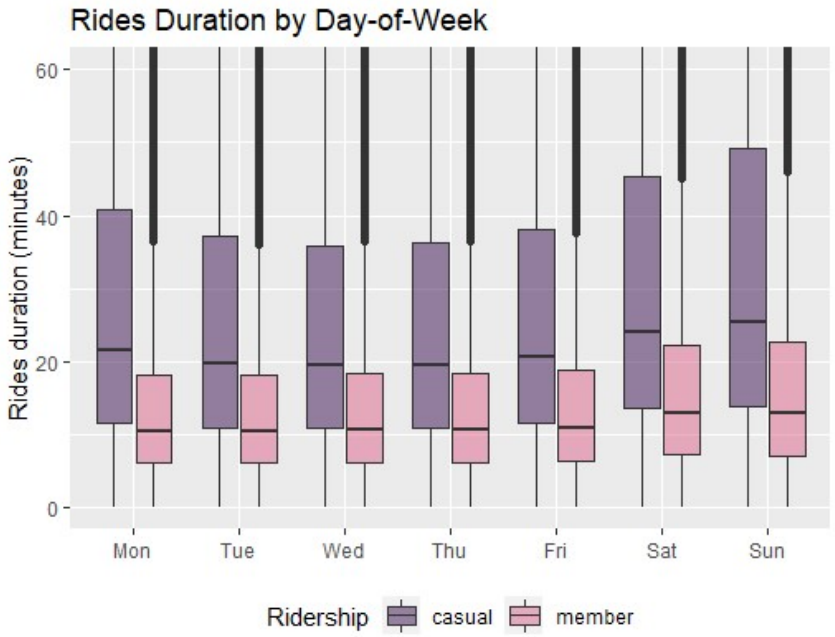
While there is a clear preference for classic bike within the members, the ride length does not seem to be greatly affected by the type of rideables. Casual riders, on the other hand, favor electric bike and tend to take longer rides with it on the weekends.



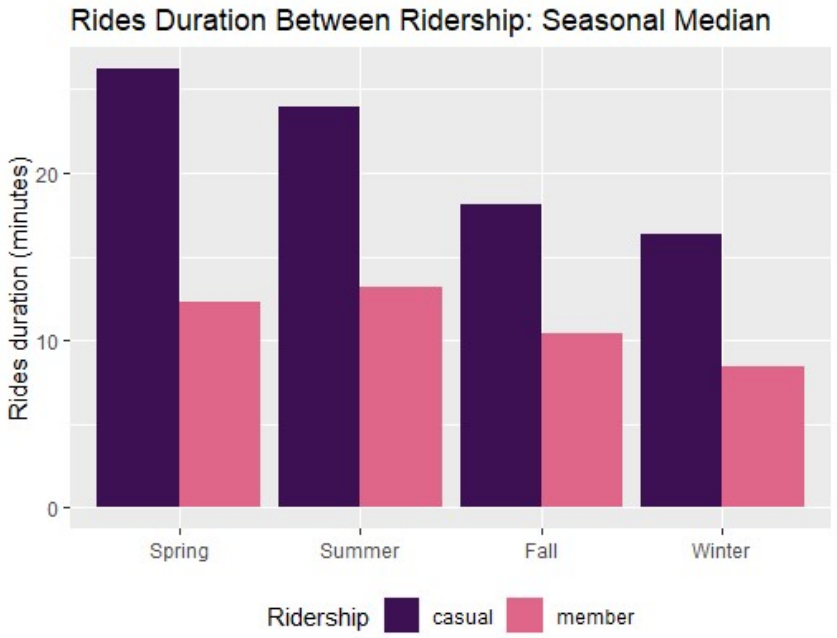


Rides Duration



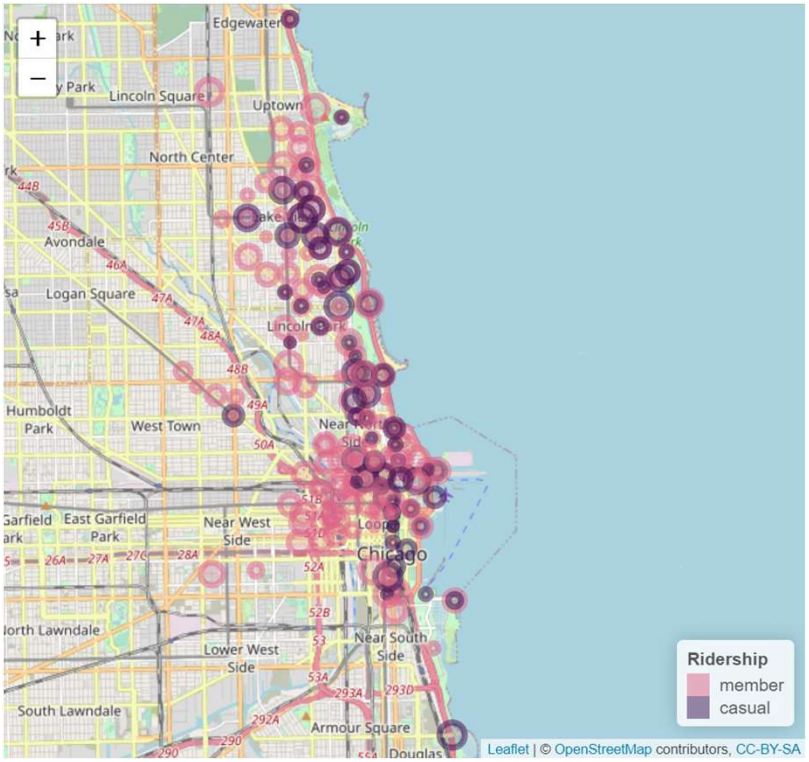


Rides Duration



- Comparing the distribution, we see that casual riders have longer and more variable rides duration: about twice as long than members do. The median duration is 22 minutes for casual riders and 11 minutes for members. For casual riders, 75% of the rides are under 42 minutes and. For members, 75% are under 19 minutes.
- Both of the plots show positive skewness, indicated by the mean that is greater than the median. Mean ride length is 42 minutes for casual riders and 19 minutes for members. This skewness is largely caused by the presence of outliers: rides duration that are relatively very long.

Location Map



- The map below displays 200 stations with the highest number of rides started. The purple circles represent casual riders, and the pink represent members. Radius of the circles illustrate the total rides starting out from a particular station.
- As shown on the map, members are clustered further inland, beginning their rides on docking stations located within the city. Casual riders, on the other hand, tend to start their rides on stations closer to the coastal area.

Summary of the Analysis

1. Casual riders use Cyclistic bikes for a longer duration and ride more frequently on the weekends and warmer seasons. They favor electric bikes and docking stations located close to the lake. These riding patterns suggest that casual riders have a leisure-oriented use of Cyclistic.
2. Users with annual membership, on the other hand, appear to have a more practical use of the bike-share service. They ride more frequently and consistently throughout the weekdays, taking shorter rides and preferring classic bikes. They start their rides from stations located further within the city.

Top Three Recommendations

1. Create seasonal campaigns, e.g., promotional deals for membership sign-up during the peak months (July, August) or on weekends.
2. Offer discounts for shorter rides on weekdays during work commuting hours.
3. Investigate whether adding more docking stations within the city would boost membership sign-ups.