

Maximizing Bike Share Membership: Strategic Marketing Promotions for Peak Summer Engagement

Rontiff A Ha

24 October, 2024

- 1. **Introduction**
- 2. **Ask**
 - 2.0.1 **Analysis Goal**
 - 2.0.2 **Question**
- 3. **Prepare**
 - 3.0.1 **Data Source**
 - 3.0.2 **Credibility of data**
- 4. **Process**
 - 4.0.1 **Tools for the project**
 - 4.0.2 **Setting up environment**
 - 4.0.3 **Import data**
- 5. **Data Cleaning**
- 6. **Data Analysis**
- 7. **Data Visualizations**
 - 7.0.1 **Visualization 1**
 - 7.0.2 **Visualization 2**
 - 7.0.3 **Visualization 3**
 - 7.0.4 **Visualization 4**
- 8. **Act**
 - 8.0.1 **Recommendations**

1. Introduction

The bike share company, **Cyclistic**, is based in Chicago and offers services to its users.

The goal of the project is to find out actionable insight and design marketing strategies aimed at converting casual riders into annual members.

In this project, I will follow the steps of the data analysis process: **ask, prepare, process, analyze, share and Recommendations.**

2. Ask

2.0.1 Analysis Goal

- Design marketing strategies aimed at **converting casual riders into annual members in 2024**, ultimately increase profitability and drive future growth for the company.

2.0.2 Question

finding insights question:

- What is the number of casual rider and annual membership in 2023?
- What is the monthly different of ride Trends by Member Types in 2023?
- What is the average number of Rides per Weekday by Member Type in 2023?
- Where is the top 10 Stations for by Member Type in 2023?

3. Prepare

3.0.1 Data Source

- The data for this analysis was obtained from **Motivate International Inc.** and can be **accessed through the provided link** (<https://divvy-tripdata.s3.amazonaws.com/index.html>). In this project, only data from January 2023 to December 2023 is used.

3.0.2 Credibility of data

- Motivate, Inc. collected the data for this analysis through its management of the Cyclistic Bike Share program for the City of Chicago. The dataset is both comprehensive and consistent, capturing all rides taken by users rather than just a sample. Additionally, the data is up-to-date, as it is released monthly by the City of Chicago and made publicly accessible.

4. Process

4.0.1 Tools for the project

- This project is using **R studio** and **R language**. Given the large size of the dataset for analysis, using spreadsheet tools like Excel or Google Sheets is not ideal. Therefore, I chose to use R for this project. **R can handle large datasets efficiently, perform data cleaning, analysis, and also generate visualizations and reports.** This makes R the perfect tool for this project.

4.0.2 Setting up environment

- **Setting up environment by installing analysis packages in R studio.**

```
#-----#
library(tidyverse)
library(lubridate)
library(janitor)
library(data.table)
library(readr)
library(ggplot2)
library(lubridate)
#-----#
```

4.0.3 Import data

- **Downloaded all the required data from January 2023 to December 2023. Save all the data under the same folder. And import it to R Studio.**

```
#-----#
# Import Data
jan_2023 <- read.csv("data/202301-divvy-tripdata.csv")
feb_2023 <- read.csv("data/202302-divvy-tripdata.csv")
mar_2023 <- read.csv("data/202303-divvy-tripdata.csv")
apr_2023 <- read.csv("data/202304-divvy-tripdata.csv")
may_2023 <- read.csv("data/202305-divvy-tripdata.csv")
jun_2023 <- read.csv("data/202306-divvy-tripdata.csv")
jul_2023 <- read.csv("data/202307-divvy-tripdata.csv")
aug_2023 <- read.csv("data/202308-divvy-tripdata.csv")
sep_2023 <- read.csv("data/202309-divvy-tripdata.csv")
oct_2023 <- read.csv("data/202310-divvy-tripdata.csv")
nov_2023 <- read.csv("data/202311-divvy-tripdata.csv")
dec_2023 <- read.csv("data/202312-divvy-tripdata.csv")
#-----#
```

- **Using summary function preview data structure, such as number of columns, rows, NA value, unwanted columns.**

```
#-----#
# Data Checking (check column, types)
summary(jan_2023)
summary(feb_2023)
summary(mar_2023)
summary(apr_2023)
summary(may_2023)
summary(jun_2023)
summary(jul_2023)
summary(aug_2023)
summary(sep_2023)
summary(oct_2023)
summary(nov_2023)
summary(dec_2023)
#-----#
```

Start of Data Validation:

- Check NA values.
- Check white spaces.
- Check duplicate records
- Correct name of columns

- After data validation, combine data variables to 1 variable for analysis.

```
#-----#
# Combine Data
raw_data <- rbind(jan_2023,feb_2023,mar_2023,
                  apr_2023,may_2023,jun_2023,
                  jul_2023,aug_2023,sep_2023,
                  oct_2023,nov_2023,dec_2023)
#-----#
```

- Save the data frame to 1 CSV file.

```
#-----#
# Save the combined files
write.csv(trip_final,file = "data/trip_final.csv",row.names = FALSE)
#-----#
```

5. Data Cleaning

To do list for data cleaning:

- **have NA value**
- **no duplicate (skip)**
- **Create time_diff_minutes column**
- **Filter rental time error**
- **Filter extreme rental time.**
- **extract datetime to individual variables**
- **filter extreme rental time > 6hrs**
- **Extract datetime to year, month, day, hour, minute**

- **Remove Missing value, empty value, duplicate, delete column**

```
#-----

data_distinct_no_missing <- raw_data %>%
  distinct() %>%
  drop_na() %>%
  remove_missing() %>%
  remove_empty(which = c("rows", "cols"))

#-----#
```

- **Create time_diff_minutes column, filter rental time error, filter extreme rental time.**

```
#-----#

cleaned_datetime <- data_distinct_no_missing %>%
  mutate(
    started_at = ymd_hms(started_at),
    ended_at = ymd_hms(ended_at),
    time_diff_minutes = as.numeric(difftime(ended_at, started_at, units = "mins"))
  ) %>%
  filter(started_at <= ended_at & time_diff_minutes > 0 & time_diff_minutes < 360)

#-----#
```

- **Extract year, month, day, hour, minutes for further analysis.**

```
#-----#
extracted_time_data <- cleaned_datetime %>%
  mutate(
    year = year(started_at),
    month = month(started_at),
    day = day(started_at),
    hour = hour(started_at),
    minute = minute(started_at)
  ) %>%
  drop_na()
#-----#
```

- **Save clean data**

```
#-----#
write.csv(cleaned_data, file = "cleaned_data/cleaned_data.csv", row.names = FALSE)
#-----#
```

6. Data Analysis

- To begin the analysis, **create a new R script, install required packages, and import cleaned data for analysis.**

```
#-----#

library(dplyr)
library(ggplot2)
library(tidyverse)
library(readr)
library(scales)

#import cleaned data
cleaned_data <- read_csv("cleaned_data/cleaned_data.csv")
#-----#
```

- **Number of Ride by Membership in 2023**

```
#-----#
cleaned_data %>%
  group_by(member_casual) %>%
  count(rideable_type) %>%
  ggplot(aes(x = member_casual, y = n, fill = rideable_type)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8), width = 0.7) +
  geom_text(aes(label = comma(n)),
            position = position_dodge(width = 0.8), vjust = -0.5, size = 4) +
  labs(title = "Number of Ride by Membership in 2023",
       x = "Membership Status",
       y = "Count",
       fill = "Rideable Type") +
  theme_minimal() +
  theme(text = element_text(size = 10))
#-----#
```

- **Monthly Ride Trends by Membership in 2023**

```
#-----#
cleaned_data %>%
  group_by(member_casual, month) %>%
  count() %>%
  ggplot(aes(x = as.integer(month), y = n, color = member_casual, group = member_casual))
  +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = 1:12) +
  scale_y_continuous(labels = comma) +
  labs(title = "Monthly Ride Trends by Membership in 2023", x = "Month", y = "Total Ride
s", color = "Member Type") +
  theme_minimal()
#-----#
```

- **Daily Average Number of Rides by Membership in 2023**

```
#-----#
cleaned_data %>%
  mutate(weekday = wday(started_at, label = TRUE, abbr = FALSE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(avg_rides = mean(n()), .groups = 'drop') %>%
  ggplot(aes(x = weekday, y = avg_rides, fill = member_casual)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = round(avg_rides, 1)),
            position = position_dodge(width = 0.9), vjust = -0.5, size = 4) +
  scale_y_continuous(labels = comma) +
  labs(title = "Daily Average Number of Rides by Membership in 2023",
       x = "Weekday",
       y = "Average Number of Rides",
       fill = "Member Type") +
  theme_minimal()
#-----#
```

- **Top 10 Highest usage Stations by Membership in 2023**

```
#-----#

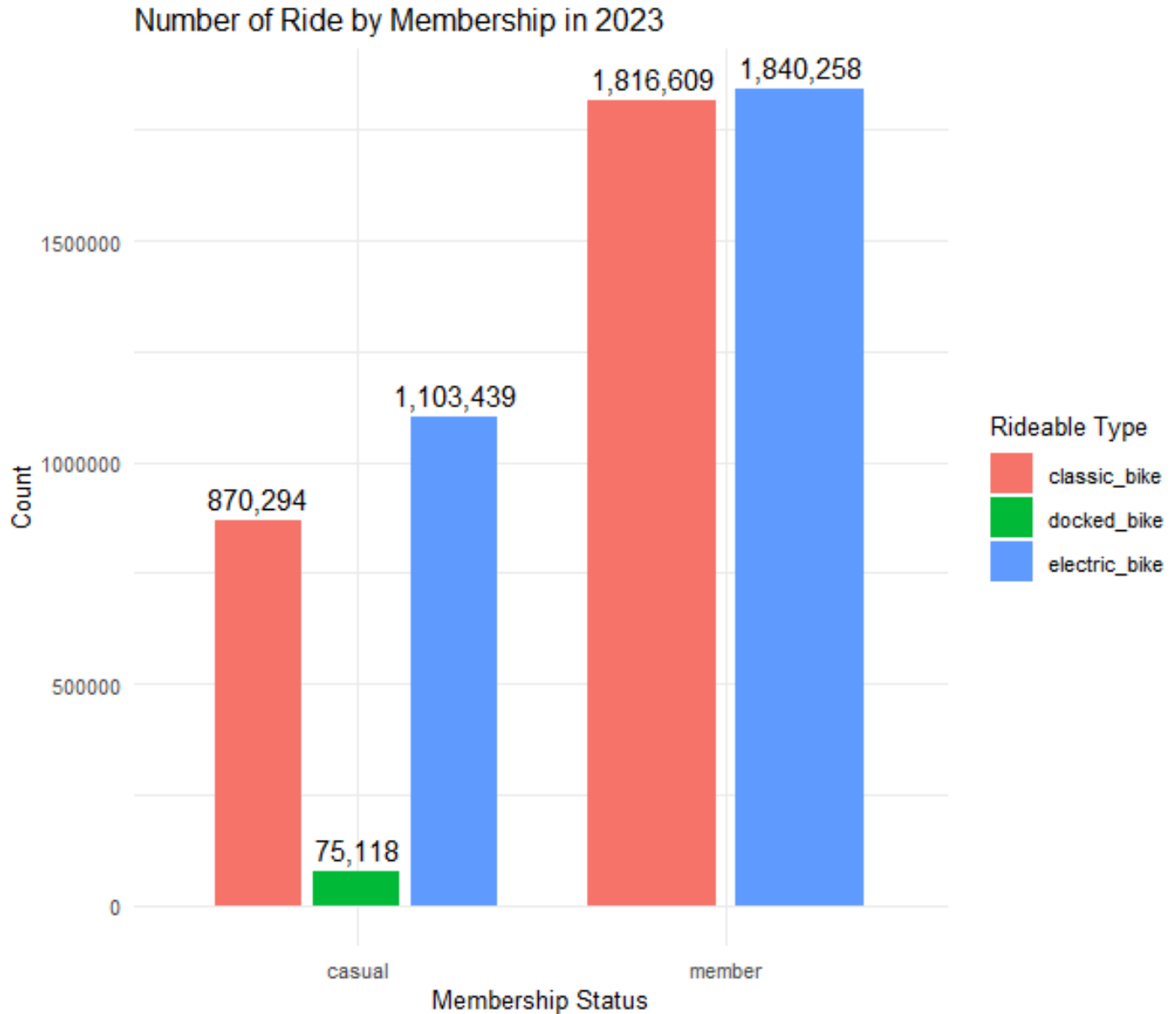
cleaned_data %>%
  filter(!is.na(start_station_name)) %>%
  group_by(start_station_name, member_casual) %>%
  count() %>%
  ungroup() %>%
  group_by(start_station_name) %>%
  summarise(total_rides = sum(n), .groups = 'drop') %>%
  arrange(desc(total_rides)) %>%
  slice_head(n = 10) %>% # Select top 10 stations
  inner_join(cleaned_data, by = "start_station_name") %>%
  group_by(member_casual, start_station_name) %>%
  count() %>%
  ggplot(aes(x = reorder(start_station_name, n), y = n, fill = member_casual)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8), width = 0.7) +
  labs(title = "Top 10 Highest usage Stations by Membership in 2023",
       x = "Start Station Name",
       y = "Count",
       fill = "Member Type") +
  coord_flip() +
  theme_minimal() +
  theme(text = element_text(size = 10))

#-----#
```


7. Data Visualizations

7.0.1 Visualization 1

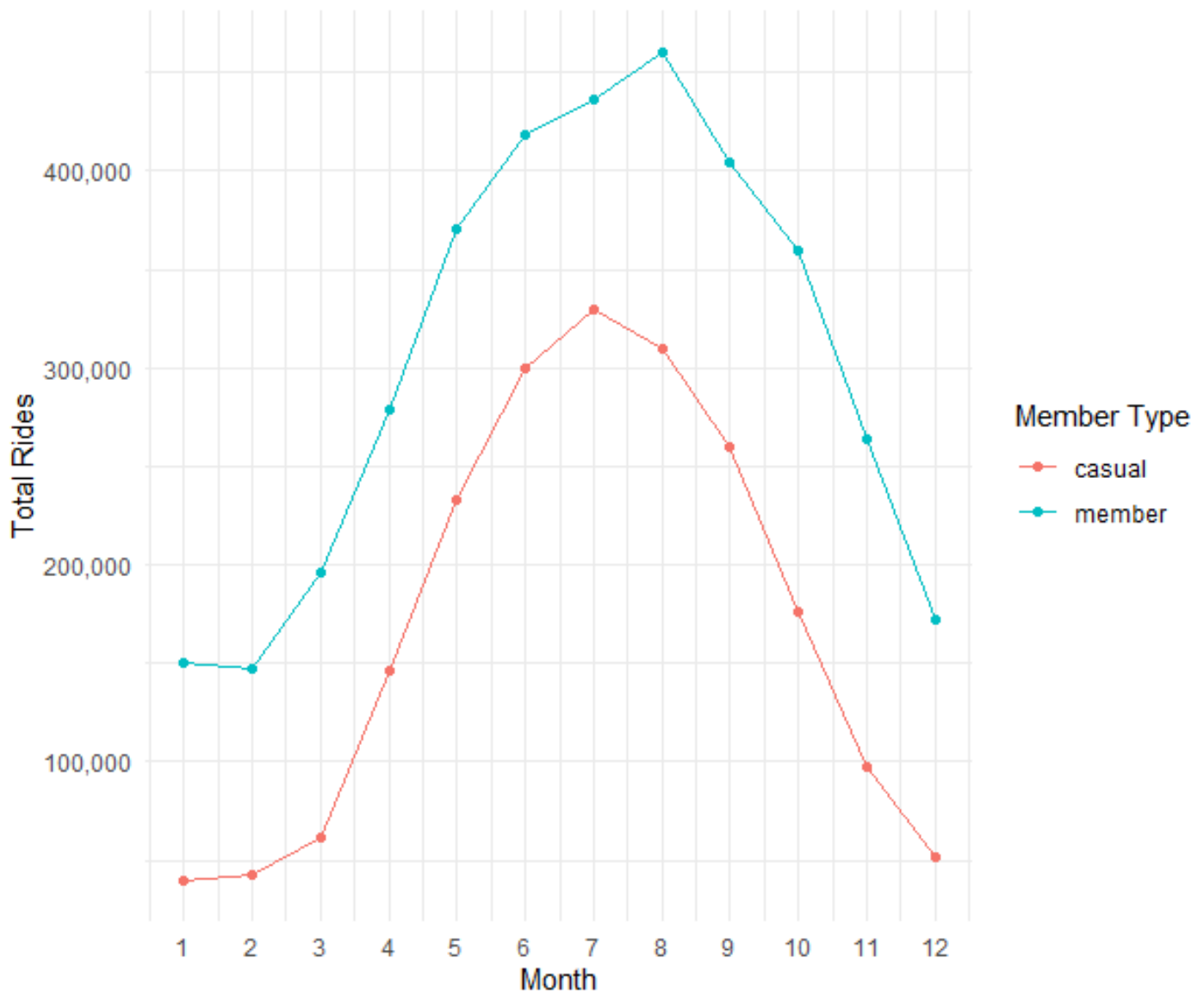
- Visualization 1 shows that both casual and member riders prefer electric bikes, with members logging the most rides. Casual users also use classic and docked bikes, though docked bikes are used much less. Members take more rides overall, indicating they use the service more frequently.



7.0.2 Visualization 2

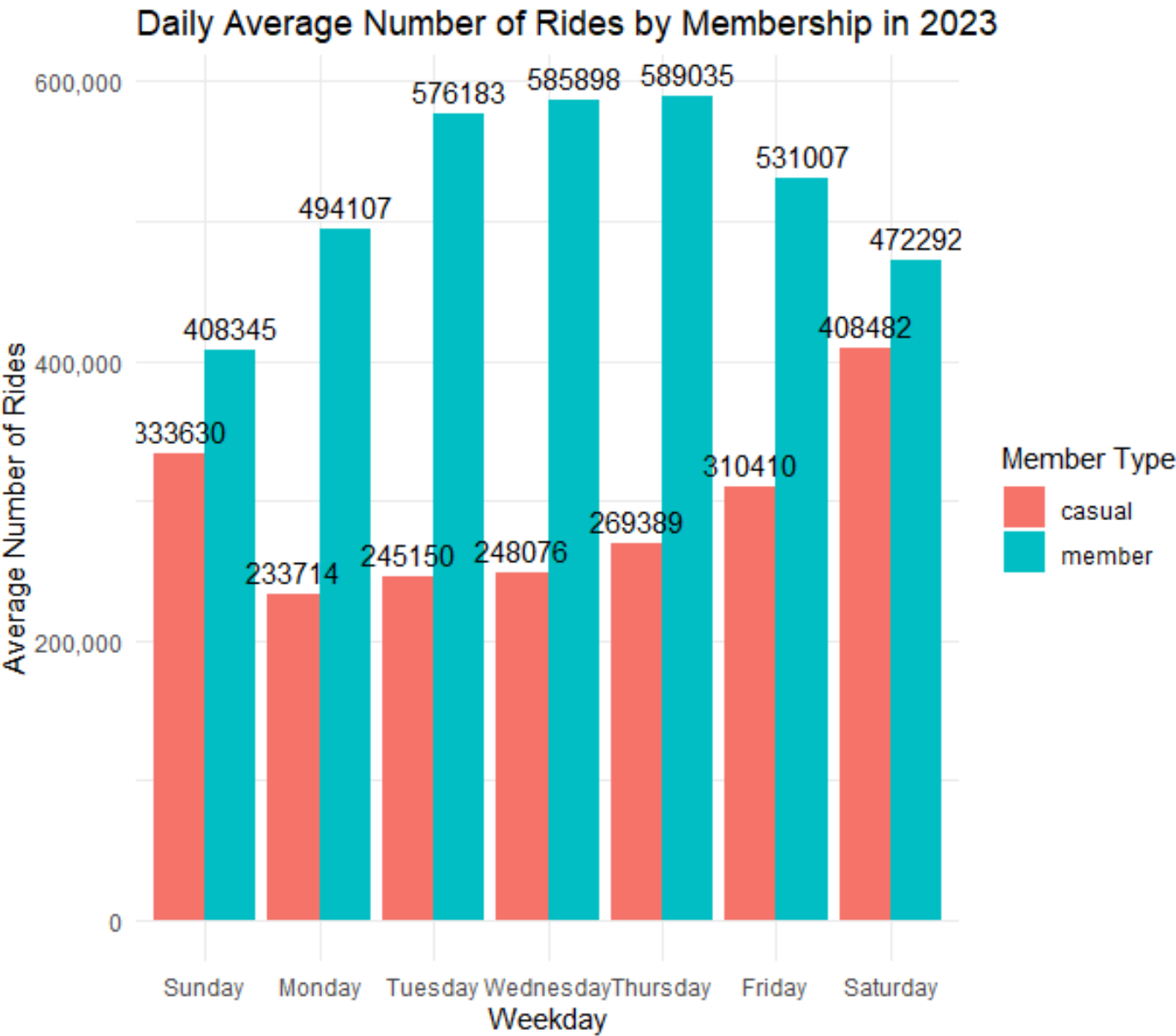
- Visualization 2 shows that both casual and member riders peak in the summer, with the highest rides in July and August. Casual riders drop significantly in winter, while members ride more consistently year-round. Members consistently take more rides than casual users each month.

Monthly Ride Trends by Membership in 2023



7.0.3 Visualization 3

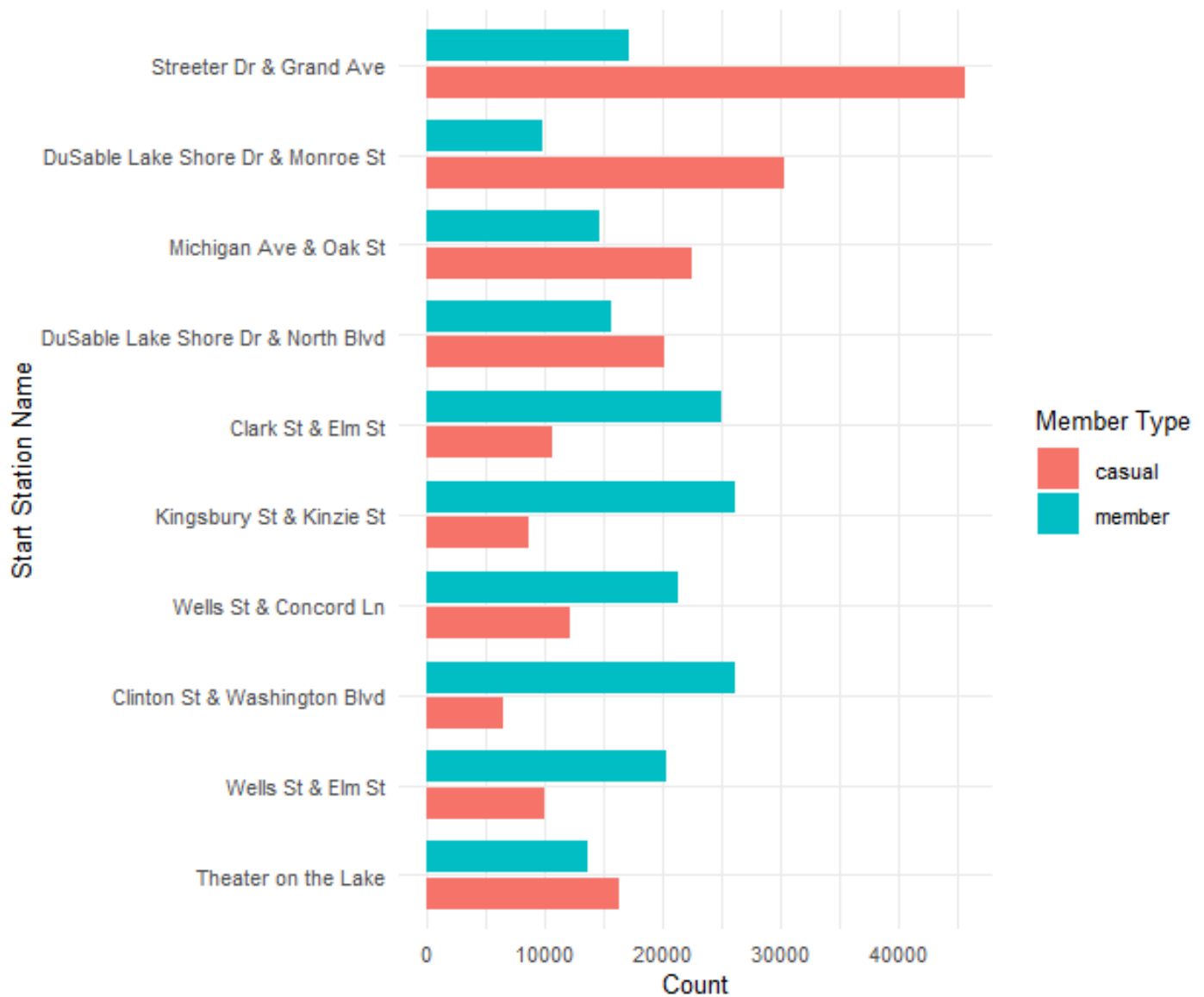
- Visualization 3 shows that casual riders peak on Saturdays (408,482 rides), while members peak on Tuesdays (576,183 rides) and Wednesdays (585,898 rides). Members consistently take more rides than casual users throughout the week, highlighting their higher engagement.



7.0.4 Visualization 4

- Visualization 4 shows that top 10 highest usage stations by membership in 2023. showing that casual rider are more often to ride the bike in Streeter Dr & Grand Ave, DuSable Lake Shore Dr & Monroe St, Michigan Ave & Oak St, and Dusable Lake Shore Dr & North Blvd.

Top 10 Highest usage Stations by Membership in 2023



8. Act

8.0.1 Recommendations

Goal:

- Design marketing strategies aimed at converting casual riders into annual members in 2024, ultimately increase profitability and drive future growth for the company.

Recommendations:

- Targeting Summer Promotions between June and August:** Given that bike usage is highest during the summer months, marketing promotions should be concentrated between June and August to capitalize on peak riding season.

- **Weekend Promotions for Casual Riders:** Casual riders show a significant increase in ridership on Saturdays and Sundays. Therefore, scheduling promotional activities on weekends will attract the largest audience.
 - **Highest Traffic Locations:** The data indicates that casual riders frequently utilize the Streeter Dr & Grand Ave and DuSable Lake Shore Dr & Monroe St locations. Hosting promotional events at these sites will maximize the opportunity to convert casual riders into annual members.
-