October
2022

Bike-Share Navigate Speedy Success Capstone Project

ANGGI FIRDIAN SAPUTRA

CONTENT:

Lorem ipsum dolor sit amet, consectetuer adipiscing elit


ASK
Lorem ipsum dolor sit amet, consectetuer adipiscing elit

Background

- In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.

Business Task

- Design marketing strategies aimed at converting casual riders into annual members.

Key Stakeholders

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

Business Objectives

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

Deliverables

1. A clear statement of the business task
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of analysis
5. Supporting visualizations and key findings
6. Top three recommendations based on your analysis


Prepare
Lorem ipsum dolor sit amet, consectetuer adipiscing elit
PREPARE
9

Deliverables

1. The dataset will be used is Cyclistic's historical trip data
2. The datasets have a different name because Cyclistic is a fictional company
3. The data has been made available by
4. Motivate International Inc
5. Data-privacy issues prohibit from using riders' personally identifiable information. This means that would not be able to connect pass purchases to credit card numbers to determine if casual riders live in the Cyclistic service area or if they have purchased multiple single passes.

Is Data ROCCC?

A good data source is **ROCCC** which stands for **R**eliable, **O**riginal, **C**omprehensive, **C**urrent, and **C**ited.

1. **R**eliable - High - Reliable as it has 5595063 data
2. **O**riginal - High - The data has been made available by Motivate International Inc. under license
3. **C**omprehensive - Med - Parameters match most of Cyclistic bikes user parameters
4. **C**urrent – High - Data is 1 year old and it is still relevant
5. **C**ited - High - Data collected from the company Motivate International Inc. and under license

Data Selection For Use

- 202101-divvy-tripdata.csv until 202112-divvy-tripdata.csv


PROCESS IN R
Lorem ipsum dolor sit amet, consectetuer adipiscing elit
PROCESS
13

Install Packages and Insert Library

```
#Install Packages
install.packages('janitor')
install.packages('skimr')
install.packages('here')
install.packages('hablar')
install.packages('chron')
install.packages('DescTools')
install.packages('metR')
```

```
#Calling Library for data analyze
library(tidyverse)
library(janitor)
library(skimr)
library(here)
library(hablar)
library(readxl)
library(data.table)
library(chron)
library(readr)
library(lubridate)
library(magrittr)
library(DescTools)
library(metR)
```

Import Data

```
#Importing Data
Data_01 <- read.csv('PROJECT/GOOGLE ANALYTICS CAPSTONE PROJECT/RAW DATA/202101-divvy-tripdata.csv')
Data_02 <- read.csv('PROJECT/GOOGLE ANALYTICS CAPSTONE PROJECT/RAW DATA/202102-divvy-tripdata.csv')
Data_03 <- read.csv('PROJECT/GOOGLE ANALYTICS CAPSTONE PROJECT/RAW DATA/202103-divvy-tripdata.csv')
Data_04 <- read.csv('PROJECT/GOOGLE ANALYTICS CAPSTONE PROJECT/RAW DATA/202104-divvy-tripdata.csv')
Data_05 <- read.csv('PROJECT/GOOGLE ANALYTICS CAPSTONE PROJECT/RAW DATA/202105-divvy-tripdata.csv')
Data_06 <- read.csv('PROJECT/GOOGLE ANALYTICS CAPSTONE PROJECT/RAW DATA/202106-divvy-tripdata.csv')
Data_07 <- read.csv('PROJECT/GOOGLE ANALYTICS CAPSTONE PROJECT/RAW DATA/202107-divvy-tripdata.csv')
Data_08 <- read.csv('PROJECT/GOOGLE ANALYTICS CAPSTONE PROJECT/RAW DATA/202108-divvy-tripdata.csv')
Data_09 <- read.csv('PROJECT/GOOGLE ANALYTICS CAPSTONE PROJECT/RAW DATA/202109-divvy-tripdata.csv')
Data_10 <- read.csv('PROJECT/GOOGLE ANALYTICS CAPSTONE PROJECT/RAW DATA/202110-divvy-tripdata.csv')
Data_11 <- read.csv('PROJECT/GOOGLE ANALYTICS CAPSTONE PROJECT/RAW DATA/202111-divvy-tripdata.csv')
Data_12 <- read.csv('PROJECT/GOOGLE ANALYTICS CAPSTONE PROJECT/RAW DATA/202112-divvy-tripdata.csv')
```

Make Column and Data Frame

```
#Making Column Names for Each Data Before Join Become a Data Frame
colnames(Data_01)
colnames(Data_02)
colnames(Data_03)
colnames(Data_04)
colnames(Data_05)
colnames(Data_06)
colnames(Data_07)
colnames(Data_08)
colnames(Data_09)
colnames(Data_10)
colnames(Data_11)
colnames(Data_12)
```



```
#Making Data Frame from All Data
All_tripdata <- bind_rows(Data_01,
                          Data_02,
                          Data_03,
                          Data_04,
                          Data_05,
                          Data_06,
                          Data_07,
                          Data_08,
                          Data_09,
                          Data_10,
                          Data_11,
                          Data_12)
```

Cleaning Data

```
#Cleaning Data
colnames(All_tripdata) #To see list of the column
dim(All_tripdata) #To see dimension of All_tripdata Dataframe
head(All_tripdata) #To see first 6 rows of dataframe
str(All_tripdata) #To see list of column and data type
```

Add Column

```
#Making Column Names Such as Date, Month, Day, Day of Week, and Year of Each Ride
All_tripdata$date <- as.Date(All_tripdata$started_at) #The default format is yyyy-mm-dd
All_tripdata$month <- format(as.Date(All_tripdata$date), '%m')
All_tripdata$day <- format(as.Date(All_tripdata$date), '%d')
All_tripdata$day_of_week <- format(as.Date(All_tripdata$date), '%u')
All_tripdata$year <- format(as.Date(All_tripdata$date), '%Y')
```

Change Data Type

```
#Convert Column Name Type of Ride Length, Month and Day Column into Numeric
All_tripdata$ride_lenght <- as.numeric(as.character(All_tripdata$ride_lenght))
All_tripdata$ride_lenght_min <- as.numeric(as.character(All_tripdata$ride_lenght_min))
All_tripdata$month <- as.numeric(All_tripdata$month)
All_tripdata$day <- as.numeric(All_tripdata$day)
is.numeric(All_tripdata$ride_lenght)
is.numeric(All_tripdata$ride_lenght_min)
is.numeric(All_tripdata$month)
is.numeric(All_tripdata$day)

#There are negative value in column ride lenght, may be this is  start_time and end_time were swapped for these rides, or the system simply registered and recorded the rides incorrectly.  So, negative value rides must be excluded.
All_tripdata_Ver1 <- All_tripdata[!( All_tripdata$ride_length < 0),]
```

ANALYZE

Lorem ipsum dolor sit amet, consectetuer adipiscing elit

Descriptive Analysis of Ride Length and weekday

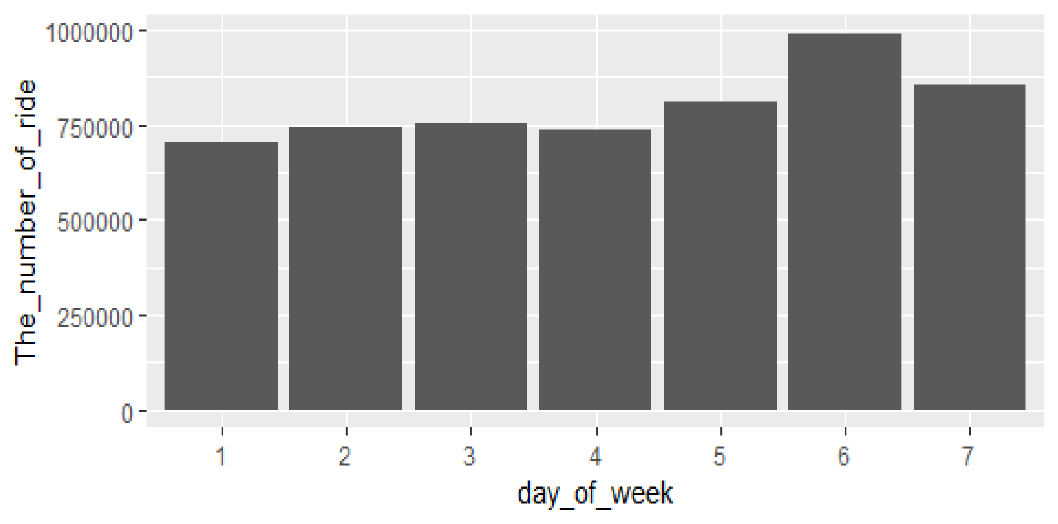
```
All_tripdata %>% summarise(max(ride_lenght_min),  min(ride_lenght_min),
                                    mean(ride_lenght_min))
#The overall average ride length is 21.9 minutes.
```

SHARE

Lorem ipsum dolor sit amet, consectetuer adipiscing elit

Plot 1: The number of Rides vs. Weekday

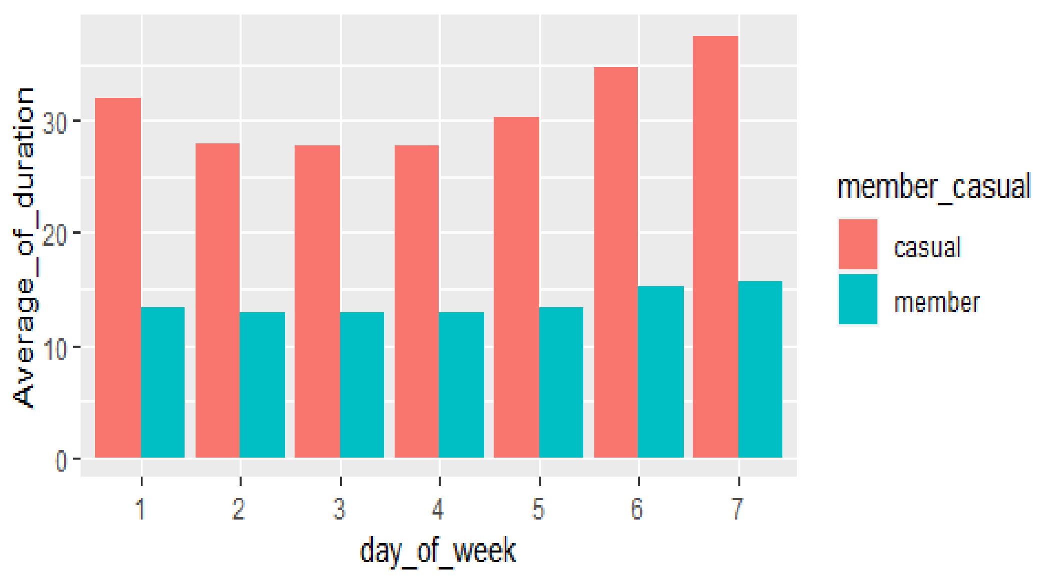
```
All_tripdata %>%
        group_by(day_of_week) %>%
        summarise(The_number_of_ride = n()) %>%
        ggplot(mapping = aes(x =
                        day_of_week,
                        y =
                        The_number_of_ride)) +
        geom_col()
```



The plot shows that most rides were started on Saturday(991047 rides), Saturday(857285 rides) followed by Friday(810508 rides). So most bikes are rented on the weekend.

Plot 2: Ride Length per Day per Rider Type

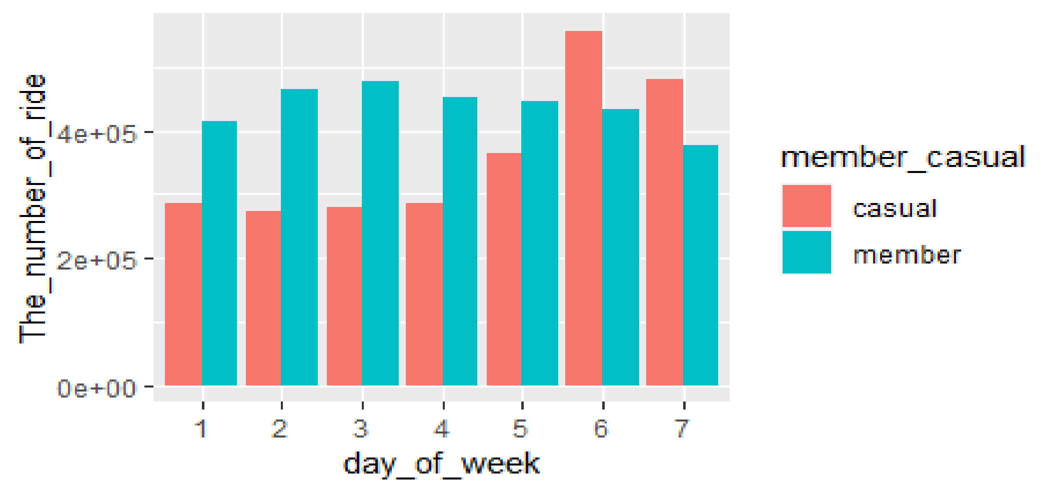
```
All_tripdata %>%
        group_by(member_casual, day_of_week) %>%
        summarise(The_number_of_ride = n(),
                  Average_of_duration =
                  mean(ride_lenght_min)) %>%
        arrange(member_casual, day_of_week) %>%
        ggplot(aes(x = day_of_week, y =
                 Average_of_duration,
                 fill = member_casual)) +
        geom_col(position = 'dodge')
```



The plot demonstrates that casual riders rent bikes for longer durations, especially on Sunday, Saturday, Friday (weekend) and on Monday. Members show a steady riding/using behavior, plus they also tend to ride a little longer on the weekend.

Plot 3: The Number of Rides per Day per Rides For Every Rider Types

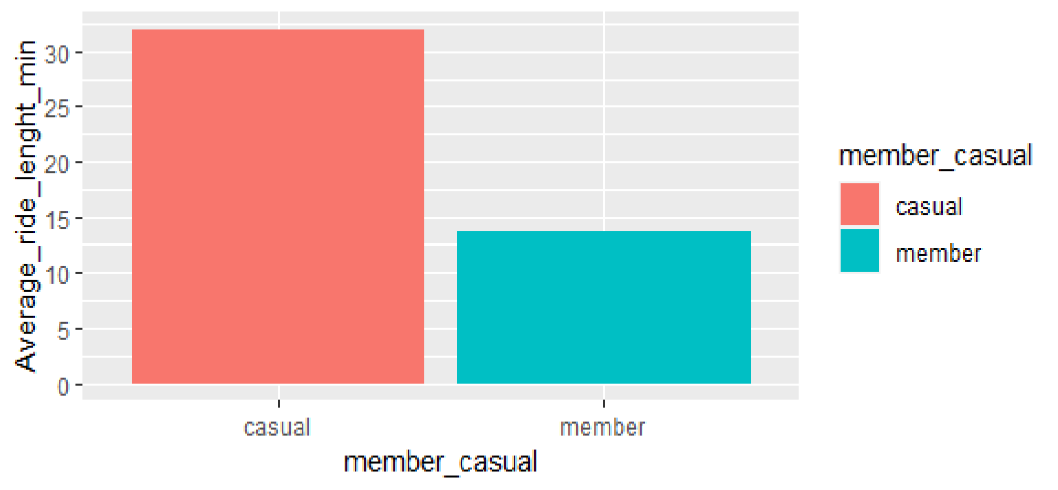
```
All_tripdata %>%
        group_by(member_casual,
        day_of_week) %>%
        summarise(The_number_of_ride = n(),
                  Average_of_duration =
                  mean(ride_lenght_min)) %>%
        ggplot(aes(x = day_of_week, y =
                  The_number_of_ride,
                  fill = member_casual)) +
        geom_col(position = 'dodge')
```



In contrast to the former plot, members begin more rides and thus have higher number of rides on every day of the week except for Saturday and Sunday.

Plot 4: Average of Ride Lenght Min

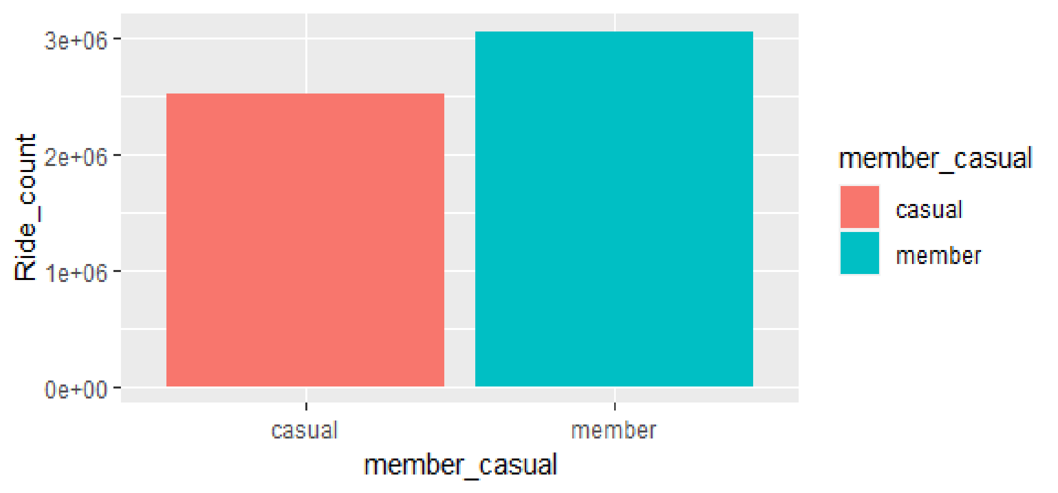
```
All_tripdata %>%
        group_by(member_casual) %>%
        summarise(max(ride_lenght_min),
                  min(ride_lenght_min),
                  Average_ride_lenght_min =
                  mean(ride_lenght_min)) %>%
        ggplot(aes(x = member_casual, y =
                 Average_ride_lenght_min,
                 fill = member_casual)) +
        geom_col() +
        scale_y_continuous(breaks = seq(0, 40, by =
                           5))
```



The result shows that casual riders tend to rent bikes for longer mean durations than members (32 min to 13.6 min), in accordance with plot 2. Members probably use bikes to commute, whereas casual riders may be, among other things, exercising, visiting the city or attending special events.

Plot 5: Overall Rider Count Based on Rider Types

```
All_tripdata %>%
        group_by(member_casual) %>%
        summarise(Ride_count = n()) %>%
        ggplot(aes(x = member_casual, y =
                Ride_count, fill =
                member_casual)) +
        geom_col()
```



The plot indicates that more than half of all riders are member riders.

Plot 6: The Number of Rides by Weekday and Rider Types and Season

```
#Assigning function of season from library metR
All_tripdata$season <- season(All_tripdata$month)

All_tripdata %>%
        group_by(season, day_of_week, member_casual) %>%
        summarise(The_number_of_ride = n(),
                  Average_ride_lenght_min =
                  mean(ride_lenght_min)) %>%
        ggplot() + geom_col(mapping = aes(x = day_of_week, y =
                                  The_number_of_ride,
                                  fill = member_casual),
                                  position = 'dodge') +
        facet_wrap(~season) +
        scale_y_continuous(breaks = seq (0, 400000, by =
                       50000))
```

DJF: Winter, MAM: Spring, JJA: Summer, SON: Fall.

The number of rides of members is always higher than that of casual riders on every workday in every season. Weekends are still the time where casual riders bike more than members. The only exception to this trend is in the winter months (Dec, Jan, Feb).

SHARE

Plot 7: Ride Lenght by Weekday and Rider Type and Season

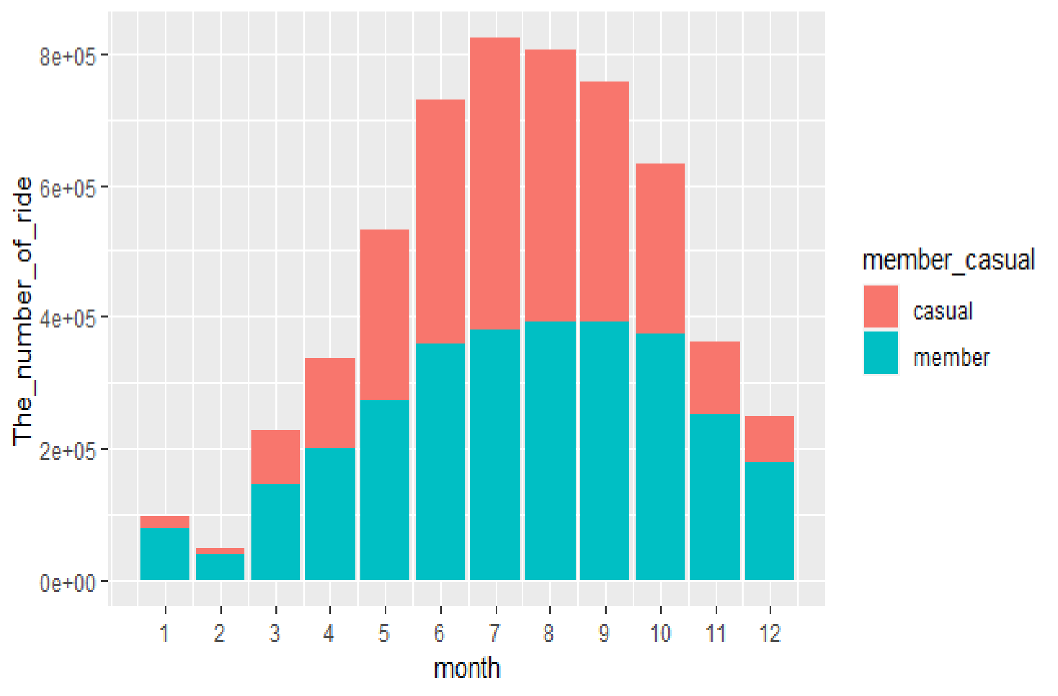
```
All_tripdata %>%
        group_by(season, day_of_week, member_casual)
        %>%
        summarise(The_number_of_ride = n(),
                  Average_ride_lenght_min =
                  mean(ride_lenght_min)) %>%
        ggplot() +
        geom_col(mapping = aes(x = day_of_week, y =
                        Average_ride_lenght_min,
                        fill = member_casual),
                        position = 'dodge') +
        facet_wrap(~season) +
        scale_y_continuous(breaks = seq (0, 50, by = 10))
```

DJF: Winter, MAM: Spring, JJA: Summer, SON: Fall.

The member group has all year long average ride length of about 13.6 minutes. Casual riders use bikes about half an hour long on all days in spring and summer. In winter and fall, the average ride length becomes less than 30 minutes.

Plot 8: The Number of Rides Along the Whole Year

```
All_tripdata %>%
        group_by(month, member_casual) %>%
        summarise(The_number_of_ride = n(),
                  Average_ride_lenght_min =
                  mean(ride_lenght_min)) %>%
        ggplot() +
        geom_col(mapping = aes(x = month, y =
                       The_number_of_ride,
                       fill =
                       member_casual)) +
        scale_x_continuous(breaks = seq(1, 12, by =
                       1))
```



For casual riders or members, ridership peaked around July until August (Summer months being the turning point) and hit the lowest at February before rebounding up swiftly and continuously.

ACTION

Lorem ipsum dolor sit amet, consectetuer adipiscing elit

Trends Identified

1. Casual rides peak during weekends (Plot 3). There is a high probability they are tourists visiting and sightseeing the city, or that they are ordinary Chicago residents who are riding bikes in their leisure time during the weekend. The longer average ride time for casual riders (Plot 2), also peaking at the weekend, provides evidence for this point.
2. Ride length for members are relatively shorter compared to casual riders. This could clarified as such, that most members use the bikes to commute on workdays. This clarification would also explain the short-riding durations of members. They ride from point A to B, namely roughly always the same ride lengths and the same distance
3. Ridership starts to pick up (Plot 8) from February (from Spring through Summer) and starts to decrease in August (from Fall through winter). This correlation is due to seasonal changes. As the weather starts to get warmer and more pleasant in February (the start of Spring), more people start to cycle, and inversely when the weather becomes less warm and cold around September (the start of Fall).
4. More than 50% of the riders are annual members (Plot 5), suggesting that the company has already achieved a certain level of loyalty among its bike users. Based on Plot 1 to Plot 8, casual rides have a number of riders that is always almost more than 50% of the annual member riders. This indicates a positive message, namely that the company is going to be able to convince many casual riders to convert to members and to keep the new members satisfied.

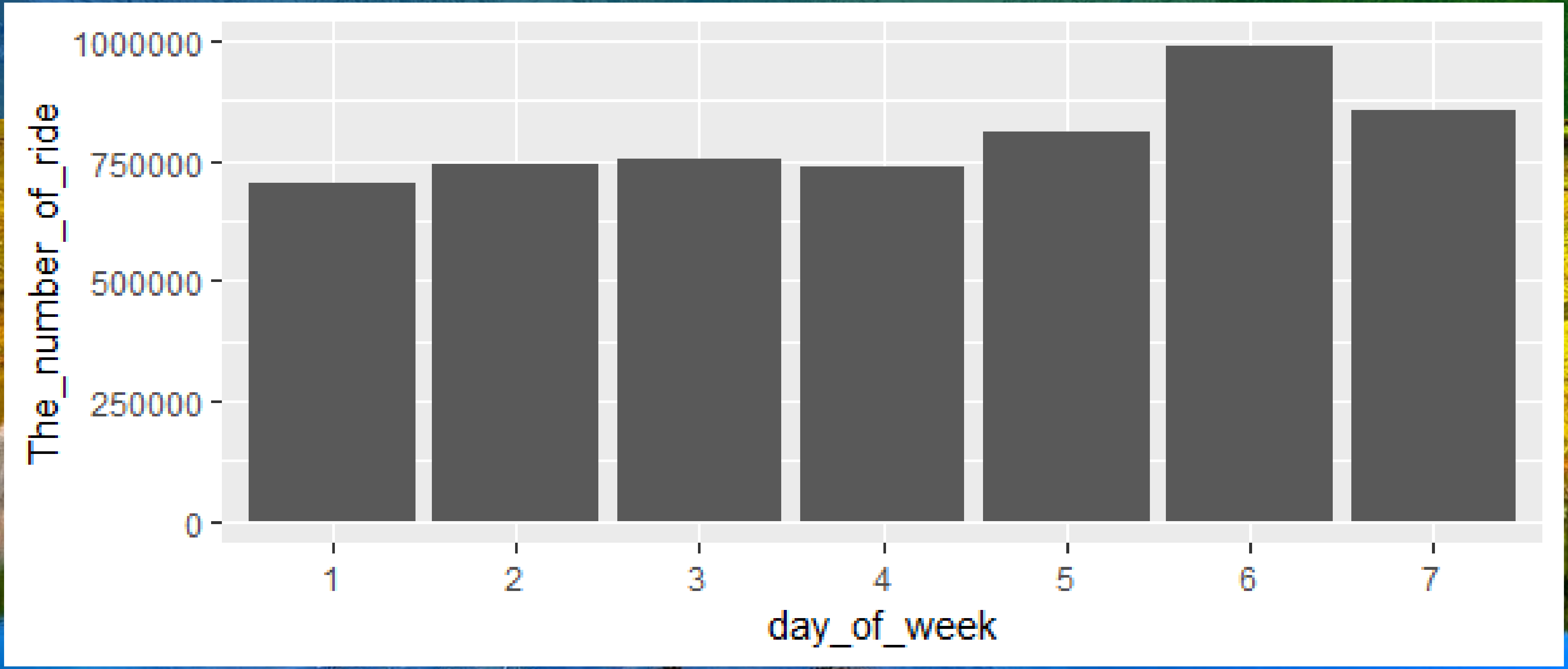
Relationship Between Trends with Cyclistic Marketing Strategy

1. Creating an application, by making an application, all members, both casual and annual, can be reached easily and marketing methods can be easily conveyed by each member.
2. Give a promotion if an annual member join for casual members. Promotion can be in the form of an explanation of the benefits that will be given if you become an annual member.
3. Making banners or other marketing media can be done where in the marketing it is explained about the benefits of cycling for health, recommendations for good use of cycling for health, and so on.
4. On Spring, Summer, and Fall seasons, Cylicstic app can also prompt notification to encourage users to rider the cycle.
5. The marketing team can distribute stickers that can be installed on bicycles, stickers containing the logo of Cyclistic so that everyone can know there is Cyclistic and every ride can feel the joy of getting the sticker which means he is bike lovers. In this way, it can expand and reach all those who do not know about the existence of Cyclistics and can increase the number of annual members because casual members feel proud to be part of Cyclistic members who have a large number of annual members, and this provides a great opportunity to convert casual member become annual member and to keep the new members satisfied.

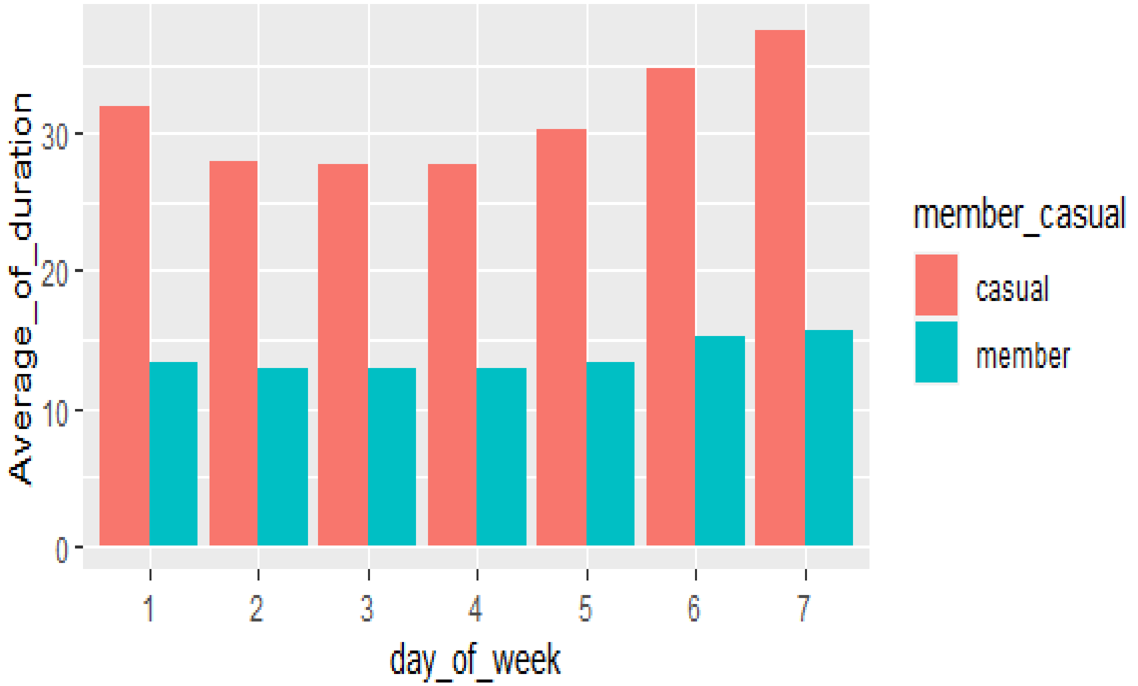
PICTURE DOCUMENTATION

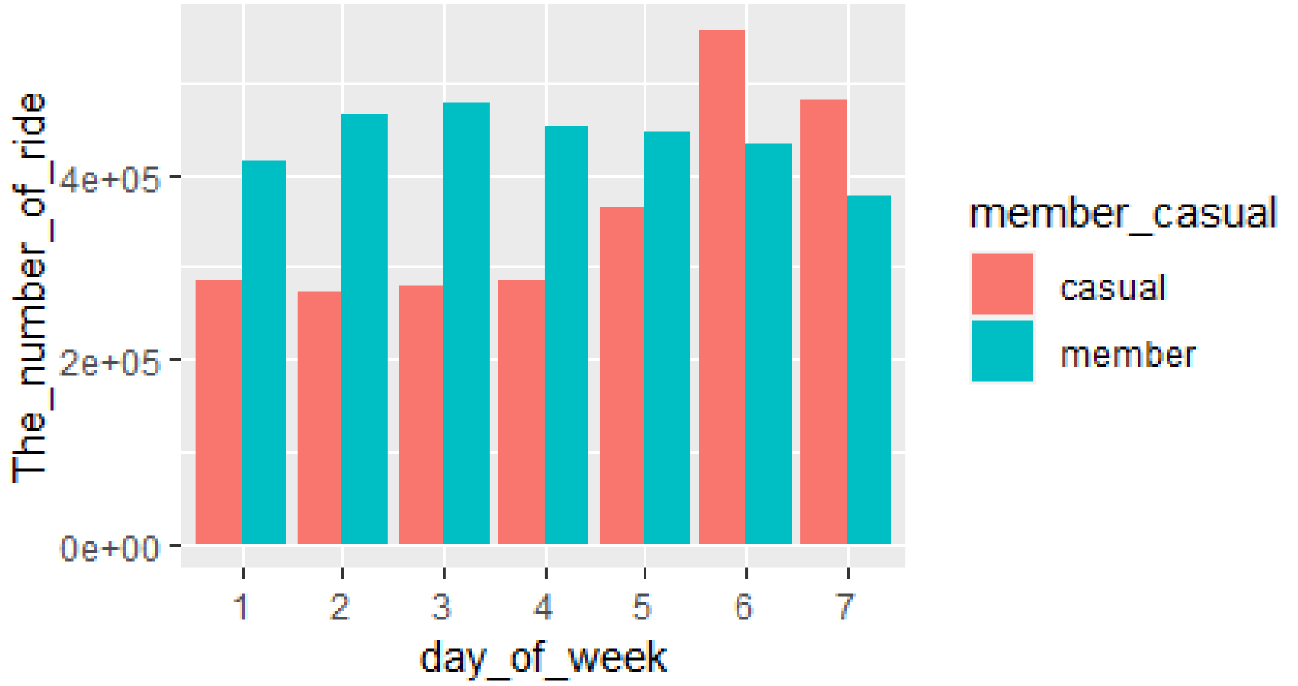
Lorem ipsum dolor sit amet, consectetuer adipiscing elit

PICTURE DOCUMENTATION

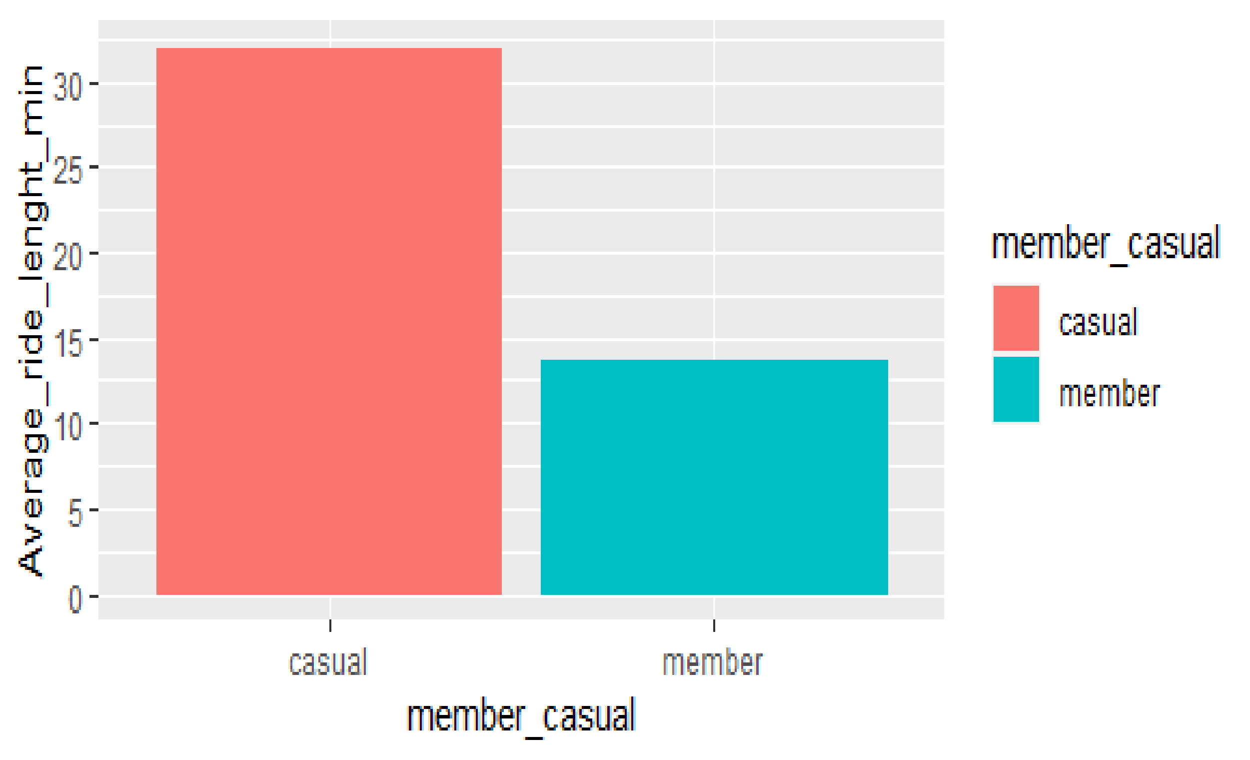


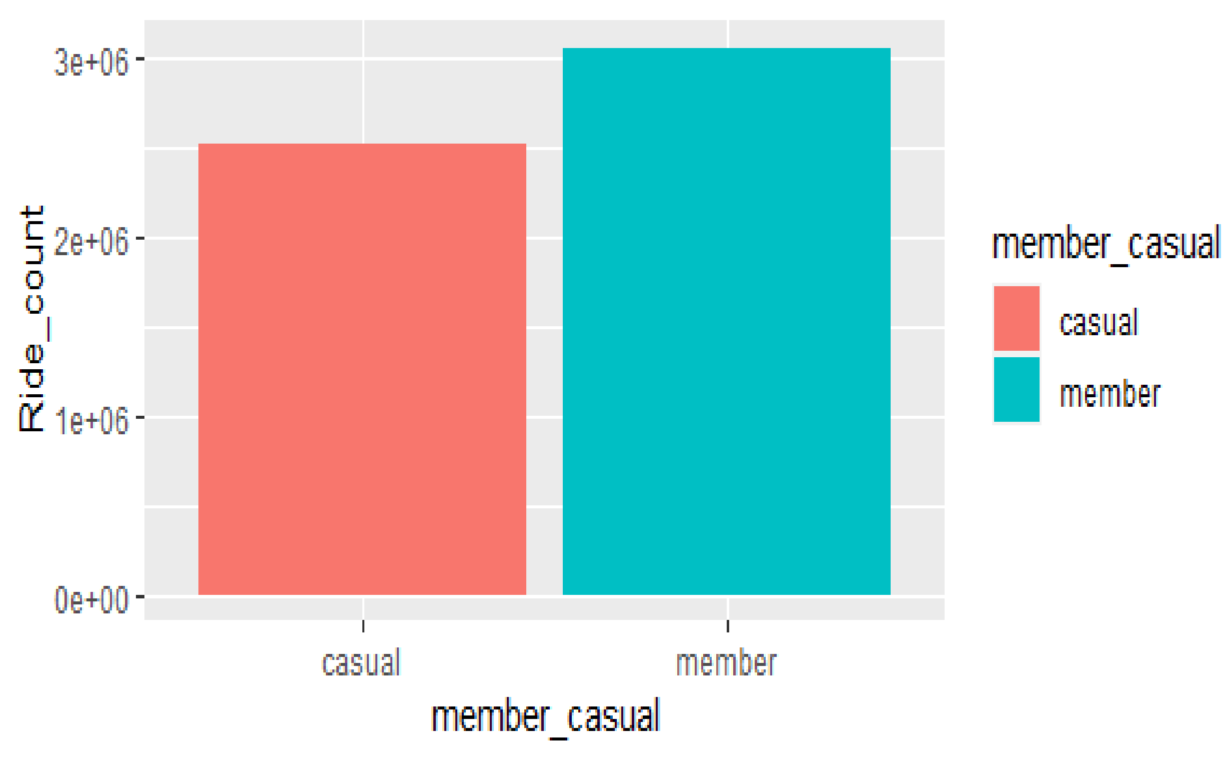
PICTURE DOCUMENTATION



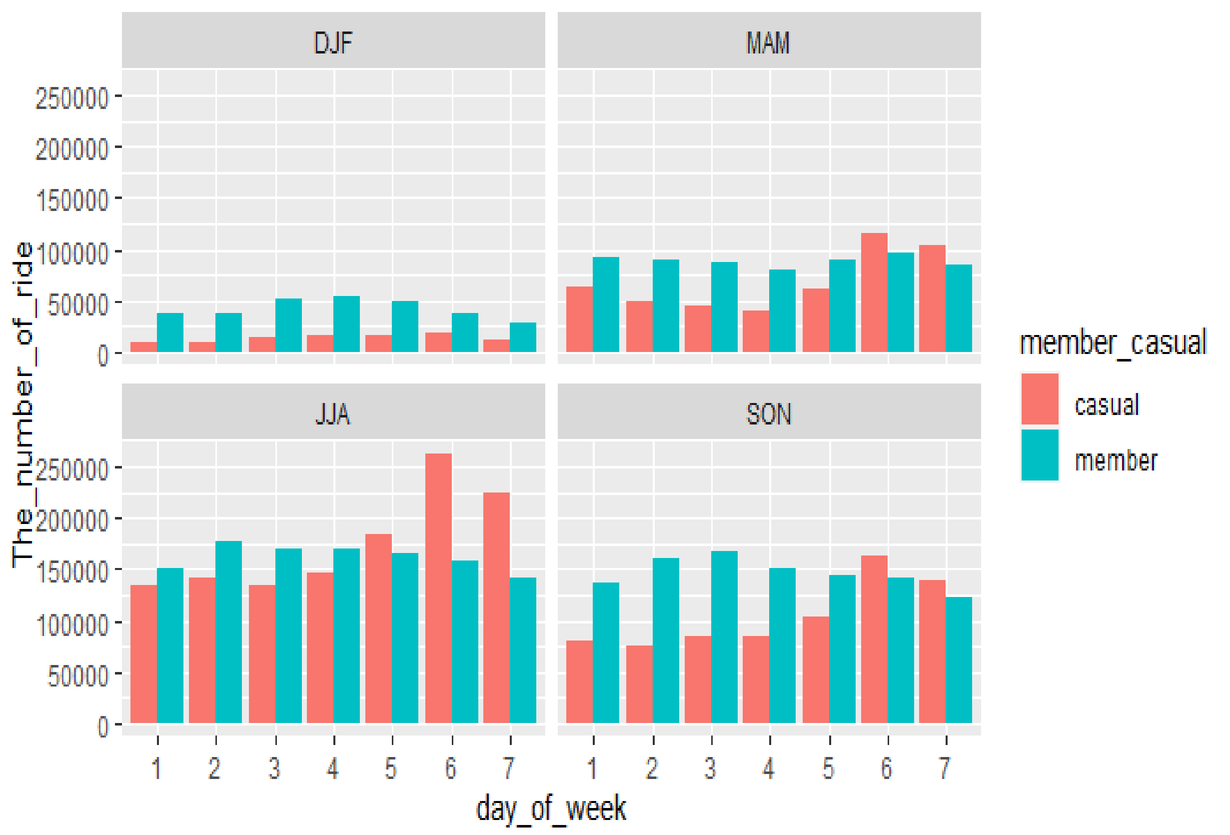


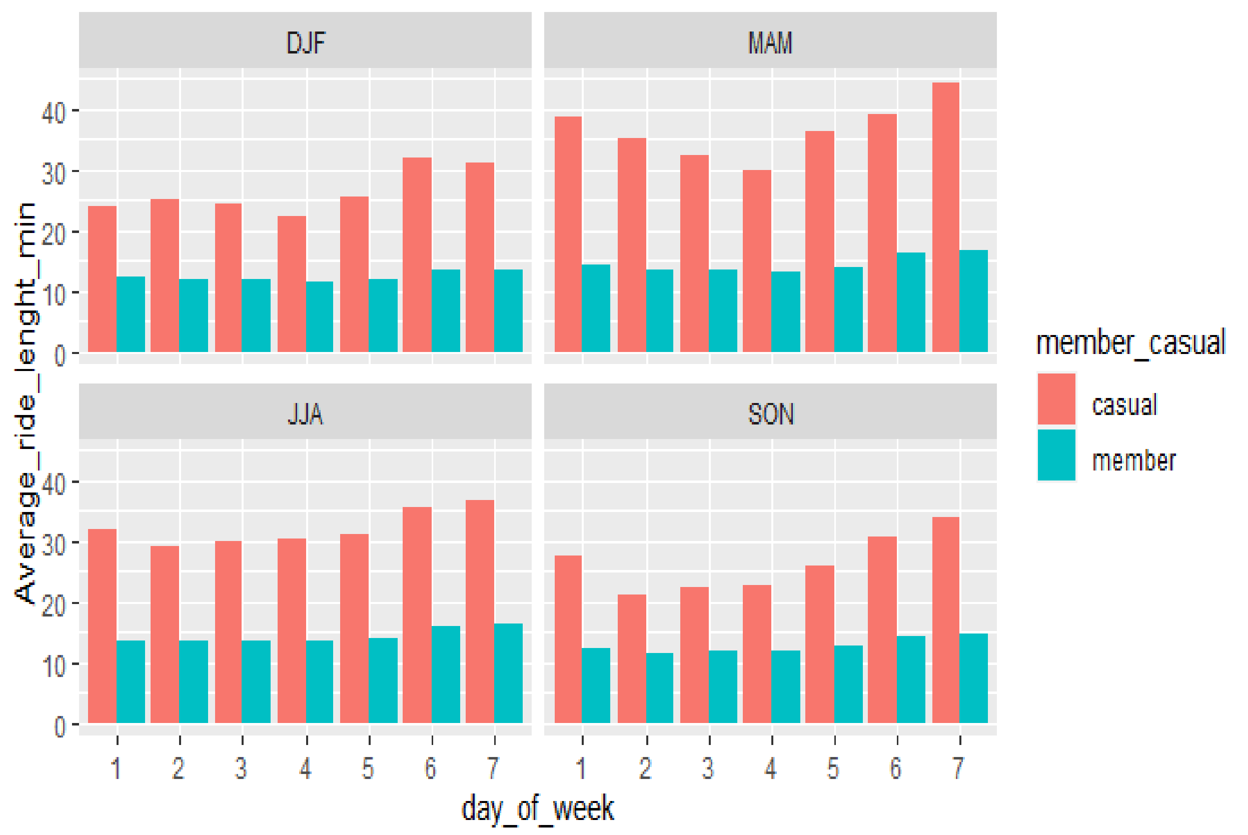
PICTURE DOCUMENTATION



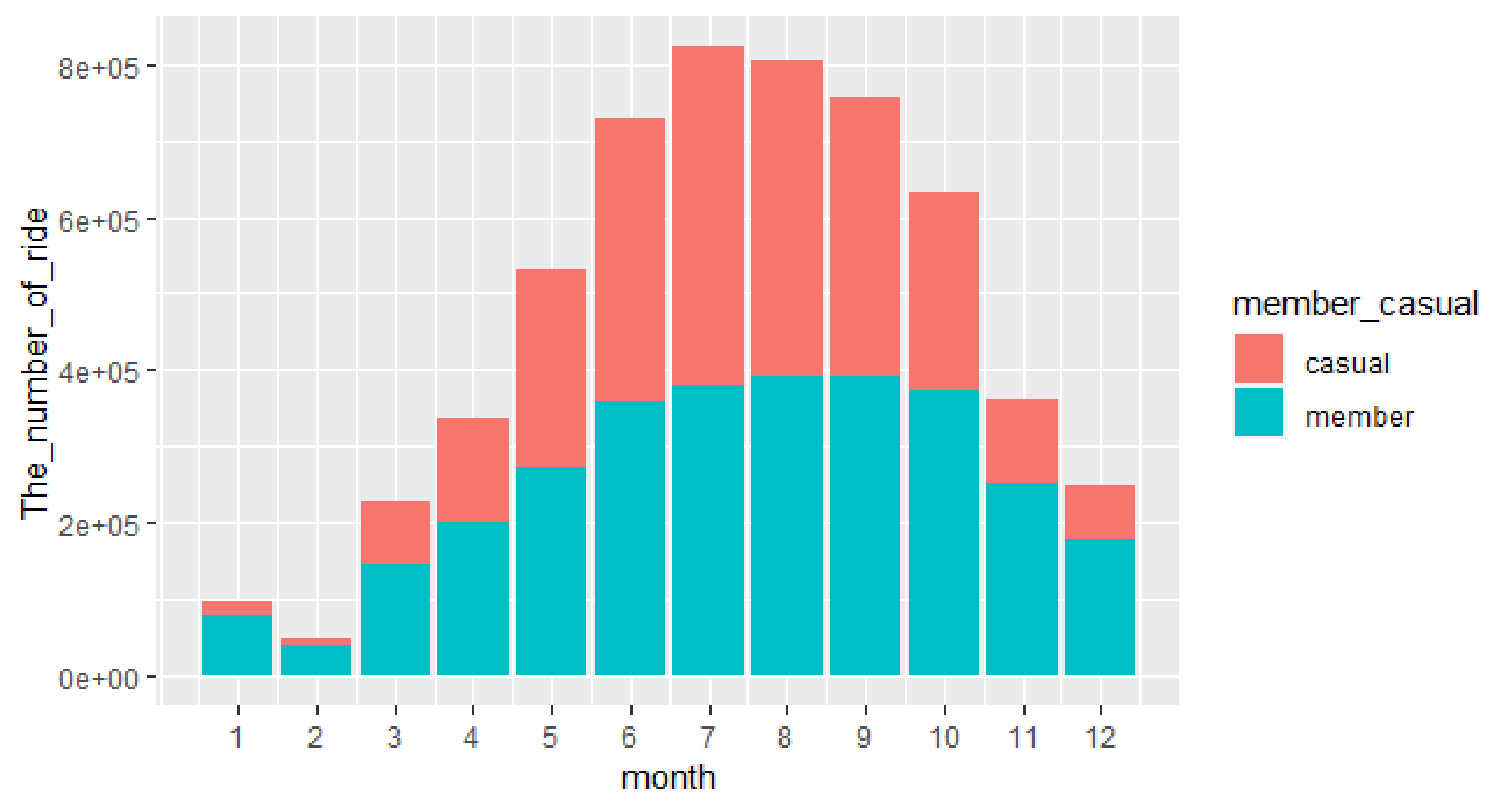


PICTURE DOCUMENTATION





PICTURE DOCUMENTATION



THANK YOU!

ANGGI FIRDIAN SAPUTRA

Phone

(+62)878-5305-3237

Email

afersaputra1@gmail.com