Analysis of Bike Sharing Company: Cyclistic

Wale Adio

2023-08-19

Contents

Data Analysis

Installing and Loading Packages

Analysis commenced by gathering and cleaning data sets from multiple quarters **(2019 q2, q3, q4 & 2020 q1)** to ensure uniform data. The `Tidyverse` package was utilized for data importation and wrangling. The `Lubridate` package was utilized for data handling attributes, while `Dplyr` and `Ggplot2` packages were utilized for data manipulation and visualization respectively.

Setting Directory

First the directory needs to be established for data importation

```
setwd("/Users/Wale/Downloads/Projects/Google Analytics/Cyclistic/Divvy_Project")
```

Importing Datasets

The data set were imported, with the data set capturing bike sharing data from 2019 q2 to 2020 q1.

```r
# Uploading csv files
q2_2019 <- read.csv("Divvy_Trips_2019_Q2.csv")
q3_2019 <- read.csv("Divvy_Trips_2019_Q3.csv")
q4_2019 <- read.csv("Divvy_Trips_2019_Q4.csv")
q1_2020 <- read.csv("Divvy_Trips_2020_Q1.csv")

# Viewing column names and structure

str(q2_2019)
```

```
## 'data.frame':    1108163 obs. of  12 variables:
##  $ X01...Rental.Details.Rental.ID                    : int  22178529 22178530 22178531 22178532 221785
##  $ X01...Rental.Details.Local.Start.Time             : chr  "2019-04-01 00:02:22" "2019-04-01 00:03:0
##  $ X01...Rental.Details.Local.End.Time               : chr  "2019-04-01 00:09:48" "2019-04-01 00:20:3
##  $ X01...Rental.Details.Bike.ID                      : int  6251 6226 5649 4151 3270 3123 6418 4513 3
##  $ X01...Rental.Details.Duration.In.Seconds.Uncapped: chr  "446.0" "1,048.0" "252.0" "357.0" ...
##  $ X03...Rental.Start.Station.ID                     : int  81 317 283 26 202 420 503 260 211 211 ...
##  $ X03...Rental.Start.Station.Name                   : chr  "Daley Center Plaza" "Wood St & Taylor St
##  $ X02...Rental.End.Station.ID                       : int  56 59 174 133 129 426 500 499 211 211 ...
##  $ X02...Rental.End.Station.Name                     : chr  "Desplaines St & Kinzie St" "Wabash Ave &
##  $ User.Type                                         : chr  "Subscriber" "Subscriber" "Subscriber" "St
##  $ Member.Gender                                     : chr  "Male" "Female" "Male" "Male" ...
##  $ X05...Member.Details.Member.Birthday.Year         : int  1975 1984 1990 1993 1992 1999 1969 1991 NA
```

```r
colnames(q2_2019)
```

```
##  [1] "X01...Rental.Details.Rental.ID"
##  [2] "X01...Rental.Details.Local.Start.Time"
##  [3] "X01...Rental.Details.Local.End.Time"
##  [4] "X01...Rental.Details.Bike.ID"
##  [5] "X01...Rental.Details.Duration.In.Seconds.Uncapped"
##  [6] "X03...Rental.Start.Station.ID"
##  [7] "X03...Rental.Start.Station.Name"
##  [8] "X02...Rental.End.Station.ID"
##  [9] "X02...Rental.End.Station.Name"
## [10] "User.Type"
## [11] "Member.Gender"
## [12] "X05...Member.Details.Member.Birthday.Year"
```

```r
str(q3_2019)
```

```
## 'data.frame':    1640718 obs. of  12 variables:
##  $ trip_id          : int  23479388 23479389 23479390 23479391 23479392 23479393 23479394 23479395 23
##  $ start_time       : chr  "2019-07-01 00:00:27" "2019-07-01 00:01:16" "2019-07-01 00:01:48" "2019-07
##  $ end_time         : chr  "2019-07-01 00:20:41" "2019-07-01 00:18:44" "2019-07-01 00:27:42" "2019-07
##  $ bikeid           : int  3591 5353 6180 5540 6014 4941 3770 5442 2957 6091 ...
##  $ tripduration     : chr  "1,214.0" "1,048.0" "1,554.0" "1,503.0" ...
##  $ from_station_id  : int  117 381 313 313 168 300 168 313 43 43 ...
##  $ from_station_name: chr  "Wilton Ave & Belmont Ave" "Western Ave & Monroe St" "Lakeview Ave & Fullo
##  $ to_station_id    : int  497 203 144 144 62 232 62 144 195 195 ...
##  $ to_station_name  : chr  "Kimball Ave & Belmont Ave" "Western Ave & 21st St" "Larrabee St & Webster
##  $ usertype         : chr  "Subscriber" "Customer" "Customer" "Customer" ...
```

```
##  $ gender           : chr  "Male" "" "" "" ...
##  $ birthyear        : int  1992 NA NA NA NA 1990 NA NA NA NA ...
```

```
colnames(q3_2019)
```

```
##  [1] "trip_id"           "start_time"        "end_time"
##  [4] "bikeid"            "tripduration"      "from_station_id"
##  [7] "from_station_name" "to_station_id"     "to_station_name"
## [10] "usertype"          "gender"            "birthyear"
```

```
str(q4_2019)
```

```
## 'data.frame':    704054 obs. of  12 variables:
##  $ trip_id          : int  25223640 25223641 25223642 25223643 25223644 25223645 25223646 25223647 25
##  $ start_time       : chr  "2019-10-01 00:01:39" "2019-10-01 00:02:16" "2019-10-01 00:04:32" "2019-10
##  $ end_time         : chr  "2019-10-01 00:17:20" "2019-10-01 00:06:34" "2019-10-01 00:18:43" "2019-10
##  $ bikeid           : int  2215 6328 3003 3275 5294 1891 1061 1274 6011 2957 ...
##  $ tripduration     : chr  "940.0" "258.0" "850.0" "2,350.0" ...
##  $ from_station_id  : int  20 19 84 313 210 156 84 156 156 336 ...
##  $ from_station_name: chr  "Sheffield Ave & Kingsbury St" "Throop (Loomis) St & Taylor St" "Milwauke
##  $ to_station_id    : int  309 241 199 290 382 226 142 463 463 336 ...
##  $ to_station_name  : chr  "Leavitt St & Armitage Ave" "Morgan St & Polk St" "Wabash Ave & Grand Ave
##  $ usertype         : chr  "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...
##  $ gender           : chr  "Male" "Male" "Female" "Male" ...
##  $ birthyear        : int  1987 1998 1991 1990 1987 1994 1991 1995 1993 NA ...
```

```
colnames(q4_2019)
```

```
##  [1] "trip_id"           "start_time"        "end_time"
##  [4] "bikeid"            "tripduration"      "from_station_id"
##  [7] "from_station_name" "to_station_id"     "to_station_name"
## [10] "usertype"          "gender"            "birthyear"
```

```
str(q1_2020)
```

```
## 'data.frame':    426887 obs. of  13 variables:
##  $ ride_id           : chr  "EACB19130B0CDA4A" "8FED874C809DC021" "789F3C21E472CA96" "C9A388DAC6ABF3
##  $ rideable_type     : chr  "docked_bike" "docked_bike" "docked_bike" "docked_bike" ...
##  $ started_at        : chr  "2020-01-21 20:06:59" "2020-01-30 14:22:39" "2020-01-09 19:29:26" "2020-0
##  $ ended_at          : chr  "2020-01-21 20:14:30" "2020-01-30 14:26:22" "2020-01-09 19:32:17" "2020-0
##  $ start_station_name: chr  "Western Ave & Leland Ave" "Clark St & Montrose Ave" "Broadway & Belmont
##  $ start_station_id  : int  239 234 296 51 66 212 96 96 212 38 ...
##  $ end_station_name  : chr  "Clark St & Leland Ave" "Southport Ave & Irving Park Rd" "Wilton Ave & Be
##  $ end_station_id    : int  326 318 117 24 212 96 212 212 96 100 ...
##  $ start_lat         : num  42 42 41.9 41.9 41.9 ...
##  $ start_lng         : num  -87.7 -87.7 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num  42 42 41.9 41.9 41.9 ...
##  $ end_lng           : num  -87.7 -87.7 -87.7 -87.6 -87.6 ...
##  $ member_casual     : chr  "member" "member" "member" "member" ...
```

```r
colnames(q1_2020)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

Data Cleaning and Wrangling

This section involves renaming columns to ensure uniformity, ensuring data types are consistent, merging data sets and removing irrelevant columns and bad data. Additionally, for data aggregation, new columns would be created.

Renaming Columns

```r
#Columns in q2_2019, q3_2019 & q4_2019 to be renamed to make them consistent with q1_2020
q2_2019 <- rename(q2_2019,
                  ride_id = "X01...Rental.Details.Rental.ID",
                  rideable_type = "X01...Rental.Details.Bike.ID",
                  started_at = "X01...Rental.Details.Local.Start.Time",
                  ended_at = "X01...Rental.Details.Local.End.Time",
                  start_station_name = "X03...Rental.Start.Station.Name",
                  start_station_id = "X03...Rental.Start.Station.ID",
                  end_station_name = "X02...Rental.End.Station.Name",
                  end_station_id = "X02...Rental.End.Station.ID",
                  member_casual = "User.Type")

q3_2019 <- rename(q3_2019,
                   ride_id = "trip_id",
                   rideable_type = "bikeid",
                   started_at = "start_time",
                   ended_at = "end_time",
                   start_station_name = "from_station_name",
                   start_station_id = "from_station_id",
                   end_station_name = "to_station_name",
                   end_station_id = "to_station_id",
                   member_casual = "usertype")

q4_2019 <- rename(q4_2019,
                   ride_id = "trip_id",
                   rideable_type = "bikeid",
                   started_at = "start_time",
                   ended_at = "end_time",
                   start_station_name = "from_station_name",
                   start_station_id = "from_station_id",
                   end_station_name = "to_station_name",
                   end_station_id = "to_station_id",
                   member_casual = "usertype")
```

Converting Data Types

```
##Converting data types in q2_2019, q3_2019 & q4_2019 to match data types in q1_2020
q2_2019 <- mutate(q2_2019, ride_id = as.character(ride_id),
                  rideable_type = as.character(rideable_type))
q3_2019 <- mutate(q3_2019, ride_id = as.character(ride_id),
                  rideable_type = as.character(rideable_type))
q4_2019 <- mutate(q4_2019, ride_id = as.character(ride_id),
                  rideable_type = as.character(rideable_type))
```

Combining Respective Data Sets

```
bike_trips <- bind_rows(q1_2020, q2_2019, q3_2019, q4_2019)
```

Removing Irrelevant Columns

```
#birthyear, gender, start_lat, start_lng, end_lat, end_lng, member.gender, X05...member.details.member.
bike_trips <- bike_trips %>%
  select(-c(birthyear, gender, start_lat, start_lng, end_lat, end_lng, Member.Gender,
            "X05...Member.Details.Member.Birthday.Year", "tripduration", "X01...Rental.Details.Duration
```

Inspecting Combined Data Set

```
summary(bike_trips)
```

```
##    ride_id          rideable_type       started_at          ended_at
##  Length:3879822     Length:3879822     Length:3879822     Length:3879822
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  start_station_name start_station_id end_station_name   end_station_id
##  Length:3879822     Min.   :  1.0    Length:3879822     Min.   :  1.0
##  Class :character   1st Qu.: 77.0    Class :character   1st Qu.: 77.0
##  Mode  :character   Median :174.0    Mode  :character   Median :174.0
##                     Mean   :202.9                       Mean   :203.8
##                     3rd Qu.:291.0                       3rd Qu.:291.0
##                     Max.   :675.0                       Max.   :675.0
##                                                         NA's   :1
##  member_casual
##  Length:3879822
##  Class :character
##  Mode  :character
##
```

```
##
##
##
```

```
str(bike_trips)
```

```
## 'data.frame':    3879822 obs. of  9 variables:
##  $ ride_id           : chr  "EACB19130B0CDA4A" "8FED874C809DC021" "789F3C21E472CA96" "C9A388DAC6ABF3
##  $ rideable_type     : chr  "docked_bike" "docked_bike" "docked_bike" "docked_bike" ...
##  $ started_at        : chr  "2020-01-21 20:06:59" "2020-01-30 14:22:39" "2020-01-09 19:29:26" "2020-
##  $ ended_at          : chr  "2020-01-21 20:14:30" "2020-01-30 14:26:22" "2020-01-09 19:32:17" "2020-
##  $ start_station_name: chr  "Western Ave & Leland Ave" "Clark St & Montrose Ave" "Broadway & Belmont
##  $ start_station_id  : int  239 234 296 51 66 212 96 96 212 38 ...
##  $ end_station_name  : chr  "Clark St & Leland Ave" "Southport Ave & Irving Park Rd" "Wilton Ave & Be
##  $ end_station_id    : int  326 318 117 24 212 96 212 212 96 100 ...
##  $ member_casual     : chr  "member" "member" "member" "member" ...
```

```
head(bike_trips)
```

```
##            ride_id rideable_type          started_at            ended_at
## 1 EACB19130B0CDA4A   docked_bike 2020-01-21 20:06:59 2020-01-21 20:14:30
## 2 8FED874C809DC021   docked_bike 2020-01-30 14:22:39 2020-01-30 14:26:22
## 3 789F3C21E472CA96   docked_bike 2020-01-09 19:29:26 2020-01-09 19:32:17
## 4 C9A388DAC6ABF313   docked_bike 2020-01-06 16:17:07 2020-01-06 16:25:56
## 5 943BC3CBECCFD662   docked_bike 2020-01-30 08:37:16 2020-01-30 08:42:48
## 6 6D9C8A6938165C11   docked_bike 2020-01-10 12:33:05 2020-01-10 12:37:54
##         start_station_name start_station_id              end_station_name
## 1 Western Ave & Leland Ave              239          Clark St & Leland Ave
## 2  Clark St & Montrose Ave              234 Southport Ave & Irving Park Rd
## 3   Broadway & Belmont Ave              296       Wilton Ave & Belmont Ave
## 4   Clark St & Randolph St               51       Fairbanks Ct & Grand Ave
## 5     Clinton St & Lake St               66          Wells St & Hubbard St
## 6    Wells St & Hubbard St              212    Desplaines St & Randolph St
##   end_station_id member_casual
## 1            326        member
## 2            318        member
## 3            117        member
## 4             24        member
## 5            212        member
## 6             96        member
```

Ensuring consistency in member_casual column

```
bike_trips <- bike_trips %>%
  mutate(member_casual = recode(member_casual,
         "Subscriber" = "member",
         "Customer" = "casual"))
```

Creating New Columns

New columns **(Day, Month & Year)** were created to enable data aggregation

```
bike_trips$date <- as.Date(bike_trips$started_at)
bike_trips$month <- format(as.Date(bike_trips$date),"%m")
bike_trips$day <- format(as.Date(bike_trips$date),"%d")
bike_trips$year <- format(as.Date(bike_trips$date),"%Y")
bike_trips$day_of_week <- format(as.Date(bike_trips$date),"%A")

#Creating new column to calculate duration of each ride
bike_trips$ride_length <- difftime(bike_trips$ended_at,bike_trips$started_at)

is.factor(bike_trips$ride_length)
```

```
## [1] FALSE
```

```
bike_trips$ride_length <- as.numeric(as.character(bike_trips$ride_length))
is.numeric(bike_trips$ride_length)
```

```
## [1] TRUE
```

Removing Bad Data

```
bike_trips_v2 <- bike_trips[!(bike_trips$start_station_name == "HQ QR" | bike_trips$ride_length<0),]
```

Arranging Days of the Week in Order

```
bike_trips_v2$day_of_week <- ordered(bike_trips_v2$day_of_week, levels=c("Sunday", "Monday", "Tuesday",
```

Descriptive Analysis

```
mean(bike_trips_v2$ride_length)
```

```
## [1] 1479.139
```

```
median(bike_trips_v2$ride_length)
```

```
## [1] 712
```

```
max(bike_trips_v2$ride_length)
```

```
## [1] 9387024
```

```
min(bike_trips_v2$ride_length)
```

```
## [1] 1
```

```r
summary(bike_trips_v2$ride_length)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       1     412     712    1479    1289 9387024
```

Comparing Members vs Casual Riders

```r
aggregate(bike_trips_v2$ride_length ~ bike_trips_v2$member_casual, FUN = mean)
```

```
##   bike_trips_v2$member_casual bike_trips_v2$ride_length
## 1                      casual                 3552.7502
## 2                      member                  850.0662
```

```r
aggregate(bike_trips_v2$ride_length ~ bike_trips_v2$member_casual, FUN = median)
```

```
##   bike_trips_v2$member_casual bike_trips_v2$ride_length
## 1                      casual                      1546
## 2                      member                       589
```

```r
aggregate(bike_trips_v2$ride_length ~ bike_trips_v2$member_casual, FUN = max)
```

```
##   bike_trips_v2$member_casual bike_trips_v2$ride_length
## 1                      casual                   9387024
## 2                      member                   9056634
```

```r
aggregate(bike_trips_v2$ride_length ~ bike_trips_v2$member_casual, FUN = min)
```

```
##   bike_trips_v2$member_casual bike_trips_v2$ride_length
## 1                      casual                         2
## 2                      member                         1
```

Comparing Average Ride Times by Day for Members vs Casual Rider

```r
aggregate(bike_trips_v2$ride_length ~ bike_trips_v2$member_casual + bike_trips_v2$day_of_week, FUN = mea
```

```
##    bike_trips_v2$member_casual bike_trips_v2$day_of_week
## 1                       casual                    Sunday
## 2                       member                    Sunday
## 3                       casual                    Monday
## 4                       member                    Monday
## 5                       casual                   Tuesday
## 6                       member                   Tuesday
## 7                       casual                 Wednesday
## 8                       member                 Wednesday
## 9                       casual                  Thursday
## 10                      member                  Thursday
```

```
## 11                      casual                    Friday
## 12                      member                    Friday
## 13                      casual                  Saturday
## 14                      member                  Saturday
##    bike_trips_v2$ride_length
## 1                  3581.4054
## 2                   919.9746
## 3                  3372.2869
## 4                   842.5726
## 5                  3596.3599
## 6                   826.1427
## 7                  3718.6619
## 8                   823.9996
## 9                  3682.9847
## 10                  823.9278
## 11                 3773.8351
## 12                  824.5305
## 13                 3331.9138
## 14                  968.9337
```

Analyzing Ridership Data by Type & Weekday

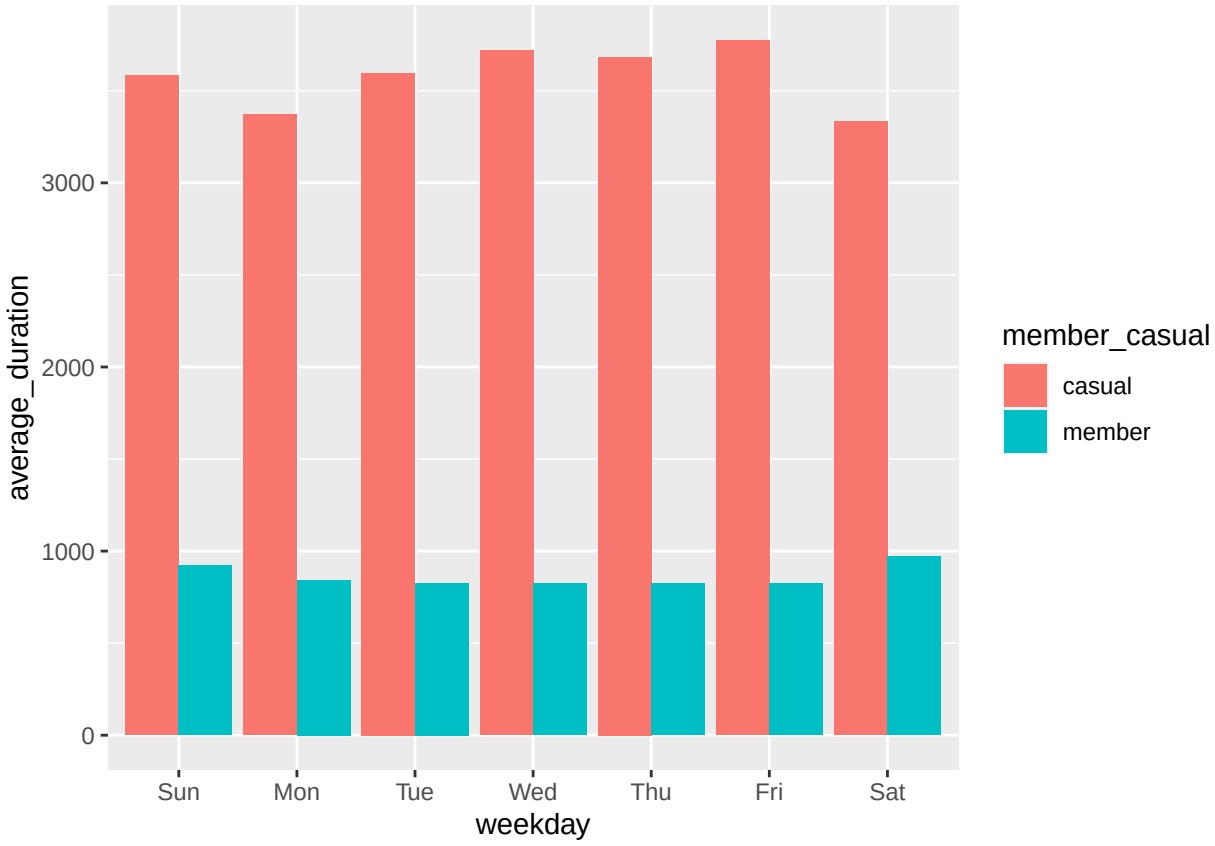
```
bike_trips_v2 %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday)
```

```
## # A tibble: 14 x 4
## # Groups:   member_casual [2]
##    member_casual weekday number_of_rides average_duration
##    <chr>         <ord>             <int>            <dbl>
##  1 casual        Sun              181293            3581.
##  2 casual        Mon              103296            3372.
##  3 casual        Tue               90510            3596.
##  4 casual        Wed               92457            3719.
##  5 casual        Thu              102679            3683.
##  6 casual        Fri              122404            3774.
##  7 casual        Sat              209543            3332.
##  8 member        Sun              267965             920.
##  9 member        Mon              472196             843.
## 10 member        Tue              508445             826.
## 11 member        Wed              500329             824.
## 12 member        Thu              484177             824.
## 13 member        Fri              452790             825.
## 14 member        Sat              287958             969.
```

Visualizing Findings

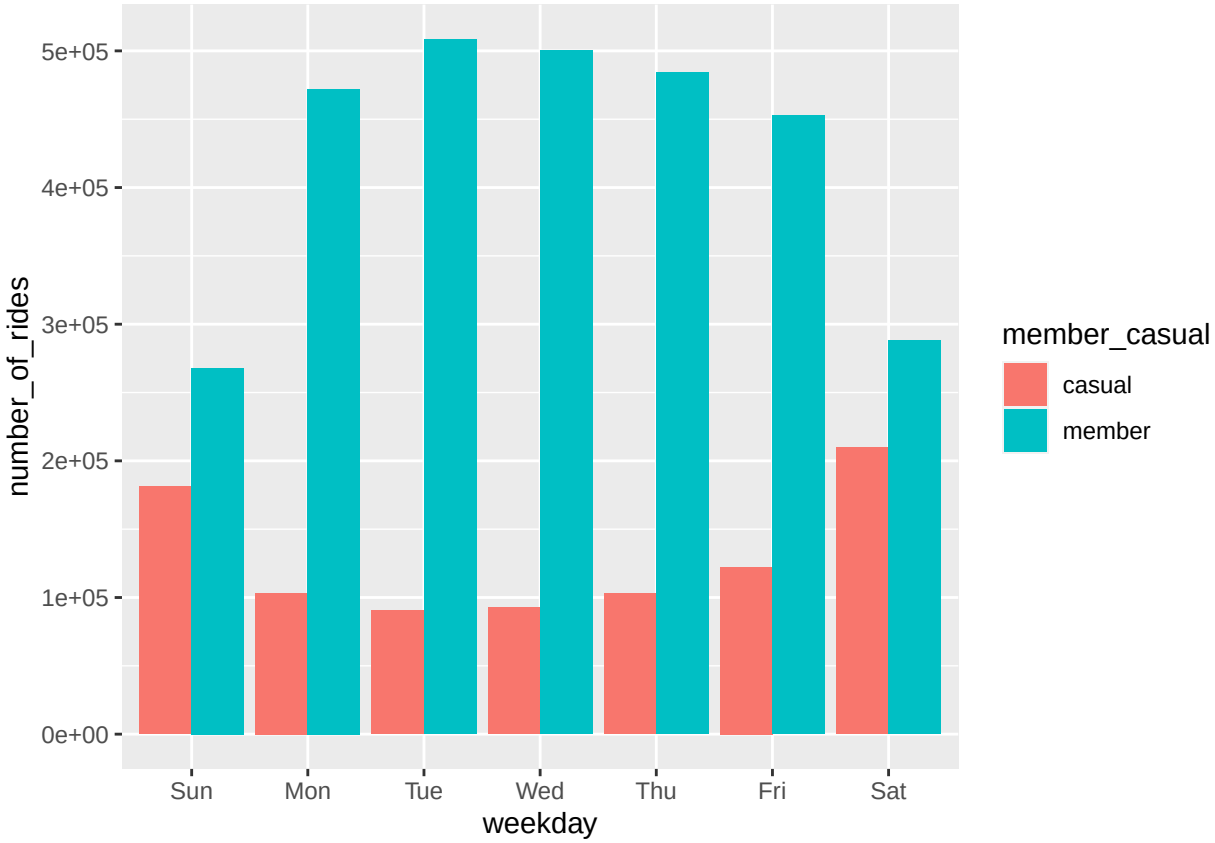
Visualizing Average Duration

```
bike_trips_v2 %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n(),
            average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday) %>%
  ggplot(aes(x = weekday, y = average_duration, fill = member_casual)) +
  geom_col(position = "dodge")
```



Visualizing by Rider Type

```
bike_trips_v2 %>%
   mutate(weekday = wday(started_at, label = TRUE)) %>%
   group_by(member_casual, weekday) %>%
   summarise(number_of_rides = n(),
             average_duration = mean(ride_length)) %>%
   ggplot(aes(x = weekday, y = number_of_rides, fill = member_casual)) +
   geom_col(position = "dodge")
```


5e+05
4e+05
3e+05
2e+05
1e+05
0e+00
number_of_rides
Sun
Mon
Tue
Wed
Thu
Fri
Sat
weekday
member_casual
casual
member