Case_Study1

Emre Can Çamcı

2023-04-03

Ask Phase

- Business task is to understand how annual members and casual riders use Cyclistic bikes differently

Prepare Phase

- To analyze the yearly trips from 2022 data has been downloaded from this link and has been made available by Motivate International Inc. under this licence.

Process Phase

Data will contain millions of rows and to process faster I used R.

Importing data to R

- Loading libraries
- Importing the cyclist data of 2022

```
jan_2022 <- read.csv("202201-divvy-tripdata.csv")
feb_2022 <- read.csv("202202-divvy-tripdata.csv")
mar_2022 <- read.csv("202203-divvy-tripdata.csv")
apr_2022 <- read.csv("202204-divvy-tripdata.csv")
may_2022 <- read.csv("202205-divvy-tripdata.csv")
jun_2022 <- read.csv("202206-divvy-tripdata.csv")
jul_2022 <- read.csv("202207-divvy-tripdata.csv")
aug_2022 <- read.csv("202208-divvy-tripdata.csv")
sep_2022 <- read.csv("202209-divvy-publictripdata.csv")
oct_2022 <- read.csv("202210-divvy-tripdata.csv")
nov_2022 <- read.csv("202211-divvy-tripdata.csv")
dec_2022 <- read.csv("202212-divvy-tripdata.csv")
```

- Comparing column names of each data to check if they match perfectly before joining them into a data frame.
- Creating the data frame with the monthly data.
- Dropping columns which will not be used
- Checking the new data frame.

Analyze Phase

- To further analyze the data adding date, weekday, month, and a ride length column.

```
#Adding a ride length column to analyze the duration of the rides in seconds
yearly_trips$ride_length <- difftime(yearly_trips$ended_at,yearly_trips$started_at)
yearly_trips$date <- as.Date(yearly_trips$started_at)
yearly_trips$month <- format(as.Date(yearly_trips$date), "%m")
yearly_trips$day_of_week <- format(as.Date(yearly_trips$date), "%A")
```

- Checking the created column

```
str(yearly_trips)
```

```
## 'data.frame':    5667717 obs. of  13 variables:
##  $ ride_id           : chr  "C2F7DD78E82EC875" "A6CF8980A652D272" "BD0F91DFF741C66D" "CBB80ED4191054
##  $ rideable_type     : chr  "electric_bike" "electric_bike" "classic_bike" "classic_bike" ...
##  $ started_at        : chr  "2022-01-13 11:59:47" "2022-01-10 08:41:56" "2022-01-25 04:53:40" "2022-
##  $ ended_at          : chr  "2022-01-13 12:02:44" "2022-01-10 08:46:17" "2022-01-25 04:58:01" "2022-
##  $ start_station_name: chr  "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave" "Sheffield Ave & Fu
##  $ start_station_id  : chr  "525" "525" "TA1306000016" "KA1504000151" ...
##  $ end_station_name  : chr  "Clark St & Touhy Ave" "Clark St & Touhy Ave" "Greenview Ave & Fullerton
##  $ end_station_id    : chr  "RP-007" "RP-007" "TA1307000001" "TA1309000021" ...
##  $ member_casual     : chr  "casual" "casual" "member" "casual" ...
##  $ ride_length       : 'difftime' num  177 261 261 896 ...
##   ..- attr(*, "units")= chr "secs"
##  $ date              : Date, format: "2022-01-13" "2022-01-10" ...
##  $ month             : chr  "01" "01" "01" "01" ...
##  $ day_of_week       : chr  "Thursday" "Monday" "Tuesday" "Tuesday" ...
```

```
head(yearly_trips, 10)
```

```
##             ride_id rideable_type          started_at            ended_at
## 1  C2F7DD78E82EC875 electric_bike 2022-01-13 11:59:47 2022-01-13 12:02:44
## 2  A6CF8980A652D272 electric_bike 2022-01-10 08:41:56 2022-01-10 08:46:17
## 3  BD0F91DFF741C66D  classic_bike 2022-01-25 04:53:40 2022-01-25 04:58:01
## 4  CBB80ED419105406  classic_bike 2022-01-04 00:18:04 2022-01-04 00:33:00
## 5  DDC963BFDDA51EEA  classic_bike 2022-01-20 01:31:10 2022-01-20 01:37:12
## 6  A39C6F6CC0586C0B  classic_bike 2022-01-11 18:48:09 2022-01-11 18:51:31
## 7  BDC4AB637EDF981B  classic_bike 2022-01-30 18:32:52 2022-01-30 18:49:26
## 8  81751A3186E59A6B  classic_bike 2022-01-22 12:20:02 2022-01-22 12:32:06
## 9  154222B86A338ABD electric_bike 2022-01-17 07:34:41 2022-01-17 08:00:08
## 10 72DC25B2DD467EEF  classic_bike 2022-01-28 15:27:53 2022-01-28 15:35:16
##                start_station_name start_station_id              end_station_name
## 1       Glenwood Ave & Touhy Ave              525          Clark St & Touhy Ave
## 2       Glenwood Ave & Touhy Ave              525          Clark St & Touhy Ave
## 3  Sheffield Ave & Fullerton Ave     TA1306000016 Greenview Ave & Fullerton Ave
## 4       Clark St & Bryn Mawr Ave     KA1504000151     Paulina St & Montrose Ave
## 5    Michigan Ave & Jackson Blvd     TA1309000002        State St & Randolph St
## 6          Wood St & Chicago Ave              637       Honore St & Division St
## 7    Oakley Ave & Irving Park Rd     KA1504000158        Broadway & Sheridan Rd
## 8  Sheffield Ave & Fullerton Ave     TA1306000016      Damen Ave & Clybourn Ave
```

```
## 9           Racine Ave & 15th St            13304  Clinton St & Washington Blvd
## 10     LaSalle St & Jackson Blvd     TA1309000004  Clinton St & Washington Blvd
##    end_station_id member_casual ride_length       date month day_of_week
## 1           RP-007        casual    177 secs 2022-01-13    01    Thursday
## 2           RP-007        casual    261 secs 2022-01-10    01      Monday
## 3    TA1307000001        member    261 secs 2022-01-25    01     Tuesday
## 4    TA1309000021        casual    896 secs 2022-01-04    01     Tuesday
## 5    TA1305000029        member    362 secs 2022-01-20    01    Thursday
## 6    TA1305000034        member    202 secs 2022-01-11    01     Tuesday
## 7           13323        member    994 secs 2022-01-30    01      Sunday
## 8           13271        member    724 secs 2022-01-22    01    Saturday
## 9           WL-012        member   1527 secs 2022-01-17    01      Monday
## 10          WL-012        member    443 secs 2022-01-28    01      Friday
```

```
tail(yearly_trips, 10)
```

```
##                  ride_id rideable_type          started_at            ended_at
## 5667708 EB7999FF9DBC0535 electric_bike 2022-12-21 19:08:04 2022-12-21 19:17:14
## 5667709 08BE55EE693263E5 electric_bike 2022-12-06 09:02:21 2022-12-06 09:06:51
## 5667710 05AD87788BCBF206 electric_bike 2022-12-06 17:08:45 2022-12-06 17:14:43
## 5667711 AFF5772FD4185626  classic_bike 2022-12-03 08:50:08 2022-12-03 08:54:17
## 5667712 7BDEDE9860418B53  classic_bike 2022-12-07 06:52:45 2022-12-07 06:56:36
## 5667713 43ABEE85B6E15DCA  classic_bike 2022-12-05 06:51:04 2022-12-05 06:54:48
## 5667714 F041C89A3D1F0270 electric_bike 2022-12-14 17:06:28 2022-12-14 17:19:27
## 5667715 A2BECB88430BE156  classic_bike 2022-12-08 16:27:47 2022-12-08 16:32:20
## 5667716 37B392960E566F58  classic_bike 2022-12-28 09:37:38 2022-12-28 09:41:34
## 5667717 2DD1587210BA45AE  classic_bike 2022-12-09 00:27:25 2022-12-09 00:35:28
##                    start_station_name start_station_id         end_station_name
## 5667708  Michigan Ave & Washington St            13001 Peoria St & Jackson Blvd
## 5667709           Morgan St & Polk St     TA1307000130    Green St & Madison St
## 5667710         Wells St & Hubbard St     TA1307000151    Green St & Madison St
## 5667711 Sangamon St & Washington Blvd            13409 Peoria St & Jackson Blvd
## 5667712 Sangamon St & Washington Blvd            13409 Peoria St & Jackson Blvd
## 5667713 Sangamon St & Washington Blvd            13409 Peoria St & Jackson Blvd
## 5667714       Bernard St & Elston Ave            18016   Seeley Ave & Roscoe St
## 5667715     Wacker Dr & Washington St     KA1503000072    Green St & Madison St
## 5667716 Sangamon St & Washington Blvd            13409 Peoria St & Jackson Blvd
## 5667717  Southport Ave & Waveland Ave            13235   Seeley Ave & Roscoe St
##         end_station_id member_casual ride_length       date month day_of_week
## 5667708          13158        member    550 secs 2022-12-21    12   Wednesday
## 5667709   TA1307000120        member    270 secs 2022-12-06    12     Tuesday
## 5667710   TA1307000120        member    358 secs 2022-12-06    12     Tuesday
## 5667711          13158        member    249 secs 2022-12-03    12    Saturday
## 5667712          13158        member    231 secs 2022-12-07    12   Wednesday
## 5667713          13158        member    224 secs 2022-12-05    12      Monday
## 5667714          13144        member    779 secs 2022-12-14    12   Wednesday
## 5667715   TA1307000120        member    273 secs 2022-12-08    12    Thursday
## 5667716          13158        member    236 secs 2022-12-28    12   Wednesday
## 5667717          13144        casual    483 secs 2022-12-09    12      Friday
```

```
summary(yearly_trips)
```

```
##    ride_id          rideable_type       started_at          ended_at
```

```
##  Length:5667717     Length:5667717     Length:5667717     Length:5667717
##  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:5667717     Length:5667717     Length:5667717     Length:5667717
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  member_casual      ride_length           date               month
##  Length:5667717     Length:5667717    Min.   :2022-01-01   Length:5667717
##  Class :character   Class :difftime   1st Qu.:2022-05-28   Class :character
##  Mode  :character   Mode  :numeric    Median :2022-07-22   Mode  :character
##                                       Mean   :2022-07-19
##                                       3rd Qu.:2022-09-16
##                                       Max.   :2022-12-31
##  day_of_week
##  Length:5667717
##  Class :character
##  Mode  :character
##
##
##
```

- Converting c(ride_length) to numeric so calculations can be executed.

```
yearly_trips$ride_length <- as.numeric(as.character(yearly_trips$ride_length))
is.numeric(yearly_trips$ride_length) # To check the class of column
```

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

```
summary(yearly_trips)
```

```
##    ride_id          rideable_type       started_at          ended_at
##  Length:5667717     Length:5667717     Length:5667717     Length:5667717
##  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:5667717     Length:5667717     Length:5667717     Length:5667717
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  member_casual       ride_length           date               month
```

```
##  Length:5667717     Min.   :-621201   Min.   :2022-01-01   Length:5667717
##  Class :character   1st Qu.:    349   1st Qu.:2022-05-28   Class :character
##  Mode  :character   Median :    617   Median :2022-07-22   Mode  :character
##                     Mean   :   1167   Mean   :2022-07-19
##                     3rd Qu.:   1108   3rd Qu.:2022-09-16
##                     Max.   :2483235   Max.   :2022-12-31
##  day_of_week
##  Length:5667717
##  Class :character
##  Mode  :character
##
##
##
```

- After checking summary, c(ride_length) has some negative values and has to be excluded.

```
yearly_trips_v2 <- subset(yearly_trips, yearly_trips$ride_length >0)
summary(yearly_trips_v2)
```

```
##    ride_id          rideable_type       started_at          ended_at
##  Length:5667186     Length:5667186     Length:5667186     Length:5667186
##  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:5667186     Length:5667186     Length:5667186     Length:5667186
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  member_casual       ride_length           date               month
##  Length:5667186     Min.   :      1   Min.   :2022-01-01   Length:5667186
##  Class :character   1st Qu.:    349   1st Qu.:2022-05-28   Class :character
##  Mode  :character   Median :    617   Median :2022-07-22   Mode  :character
##                     Mean   :   1167   Mean   :2022-07-19
##                     3rd Qu.:   1108   3rd Qu.:2022-09-16
##                     Max.   :2483235   Max.   :2022-12-31
##  day_of_week
##  Length:5667186
##  Class :character
##  Mode  :character
##
##
##
```

Organizing Data

- Reordering and dropping the columns to understand data easier

```
#Removing columns which will not be use to analyze
yearly_trips <- yearly_trips %>%
  select(-c(started_at,ended_at,start_station_name,start_station_id,
            end_station_name,end_station_id))
#Reordering columns
colnames(yearly_trips)
```

```
## [1] "ride_id"       "rideable_type" "member_casual" "ride_length"
## [5] "date"          "month"         "day_of_week"
```

```
col_order <- c("ride_id", "rideable_type","date", "day_of_week","month","ride_length","member_casual")
yearly_trips_v3 <- yearly_trips_v2[, col_order]
```

- Converting c(ride_length) to minutes

```
yearly_trips_v3$ride_length <- yearly_trips_v3$ride_length /60
```

Analyzing the data To understand how the member and casuals use bikes differently checking:

- Ride duration
- Number of rides

Ride duration

- First a descriptive analysis on ride duration.

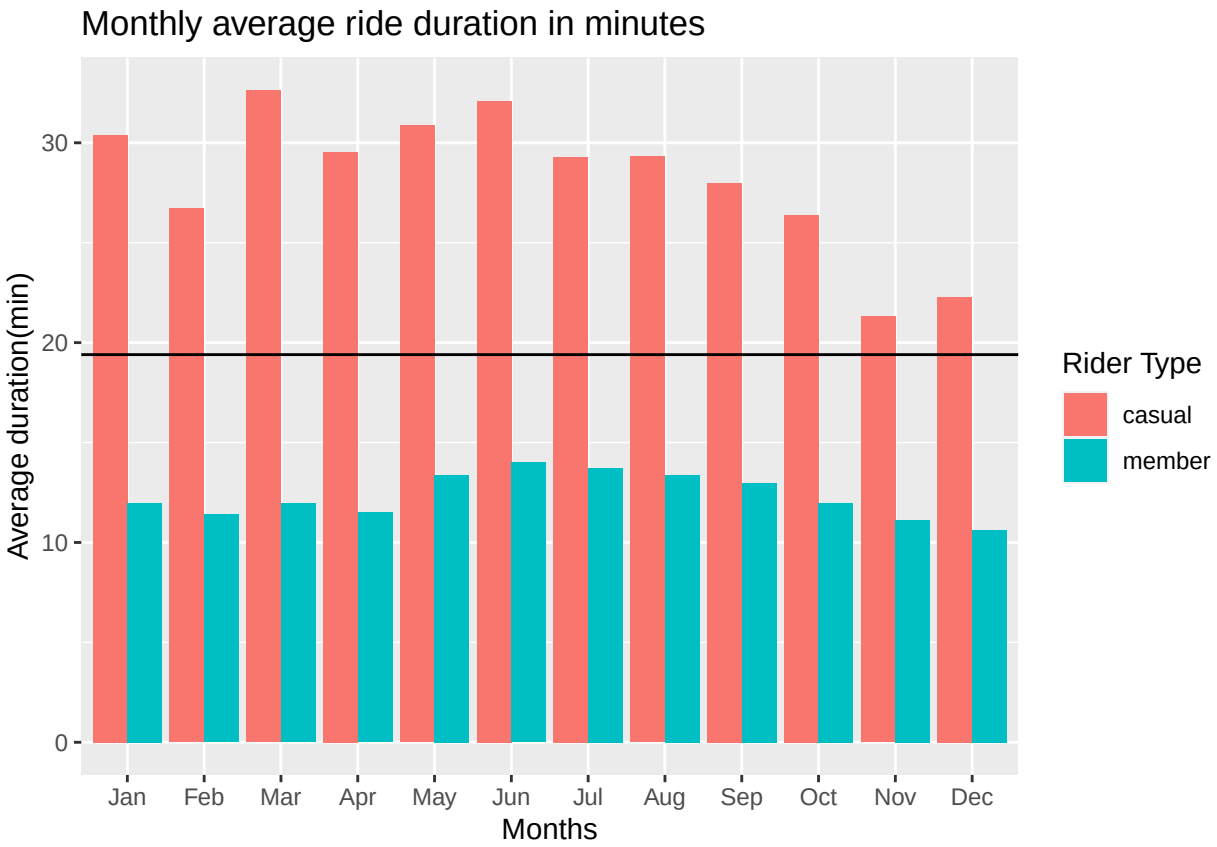
```
yearly_trips_v3 %>%
  summarise(min(ride_length),mean(ride_length),max(ride_length))
```

```
##   min(ride_length) mean(ride_length) max(ride_length)
## 1       0.01666667          19.44743         41387.25
```

We can see that average ride length is 19.4 minutes per rider.

- Comparing members and casual users monthly average ride duration.

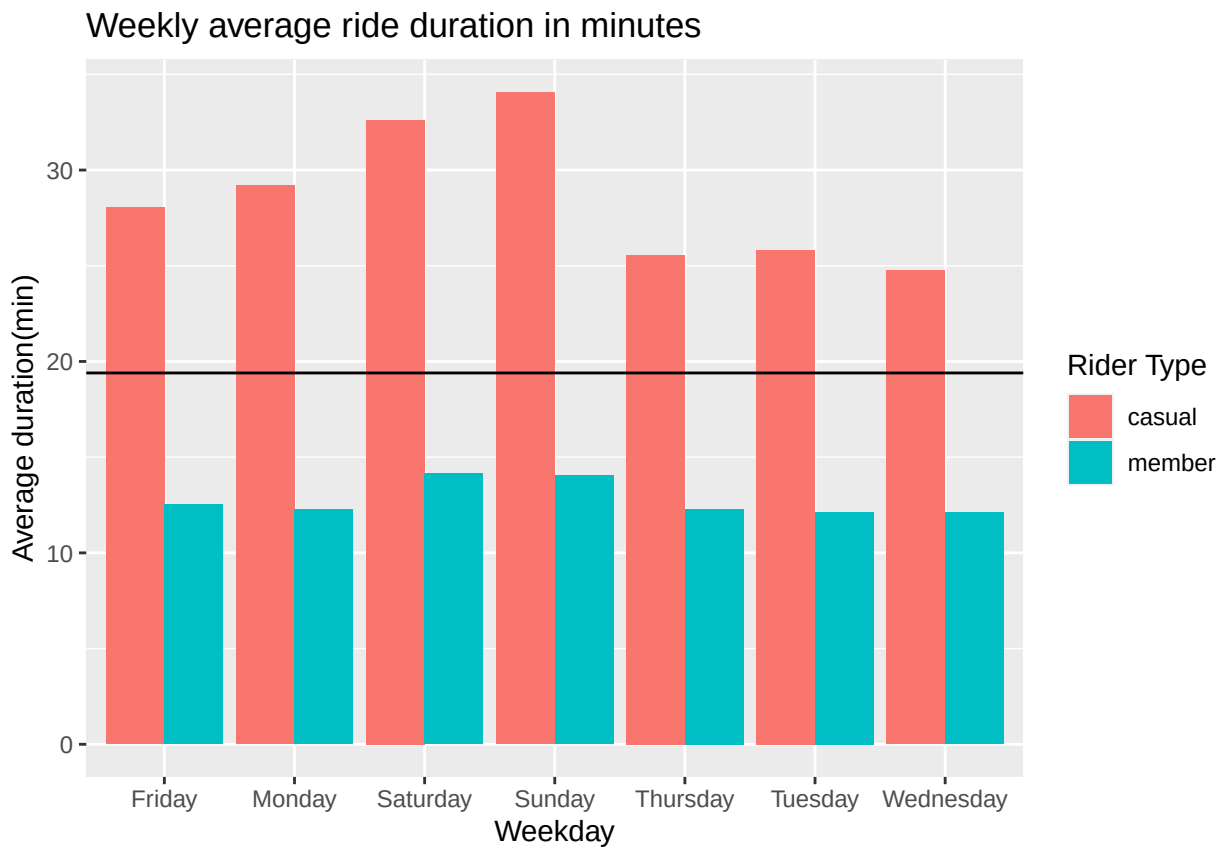
```
yearly_trips_v3 %>%
  group_by(member_casual, month) %>%
  summarise(number_of_rides = n(), average_duration=mean(ride_length)) %>%
  arrange(member_casual,month) %>%
  ggplot(aes(x=month, y=average_duration, fill= member_casual))+
  geom_col(position = "dodge")+
  geom_hline(yintercept = 19.4)+
  xlab("Months")+
  scale_x_discrete(labels = c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug",
                              "Sep","Oct","Nov","Dec"))+
  ylab("Average duration(min)")+
  labs(title = "Monthly average ride duration in minutes" , fill ="Rider Type")
```



The plot shows that casuals tends to ride longer than members, and members ride duration is below average ride duration.

- If we look at weekly average ride duration

```
yearly_trips_v3 %>%
  group_by(member_casual, day_of_week) %>%
  summarise(number_of_rides = n(), average_duration=mean(ride_length)) %>%
  arrange(member_casual,day_of_week) %>%
  ggplot(aes(x=day_of_week, y=average_duration, fill= member_casual))+
  geom_col(position = "dodge")+
  geom_hline(yintercept = 19.4)+
  xlab("Weekday")+
  ylab("Average duration(min)")+
  labs(title = "Weekly average ride duration in minutes" , fill ="Rider Type")
```

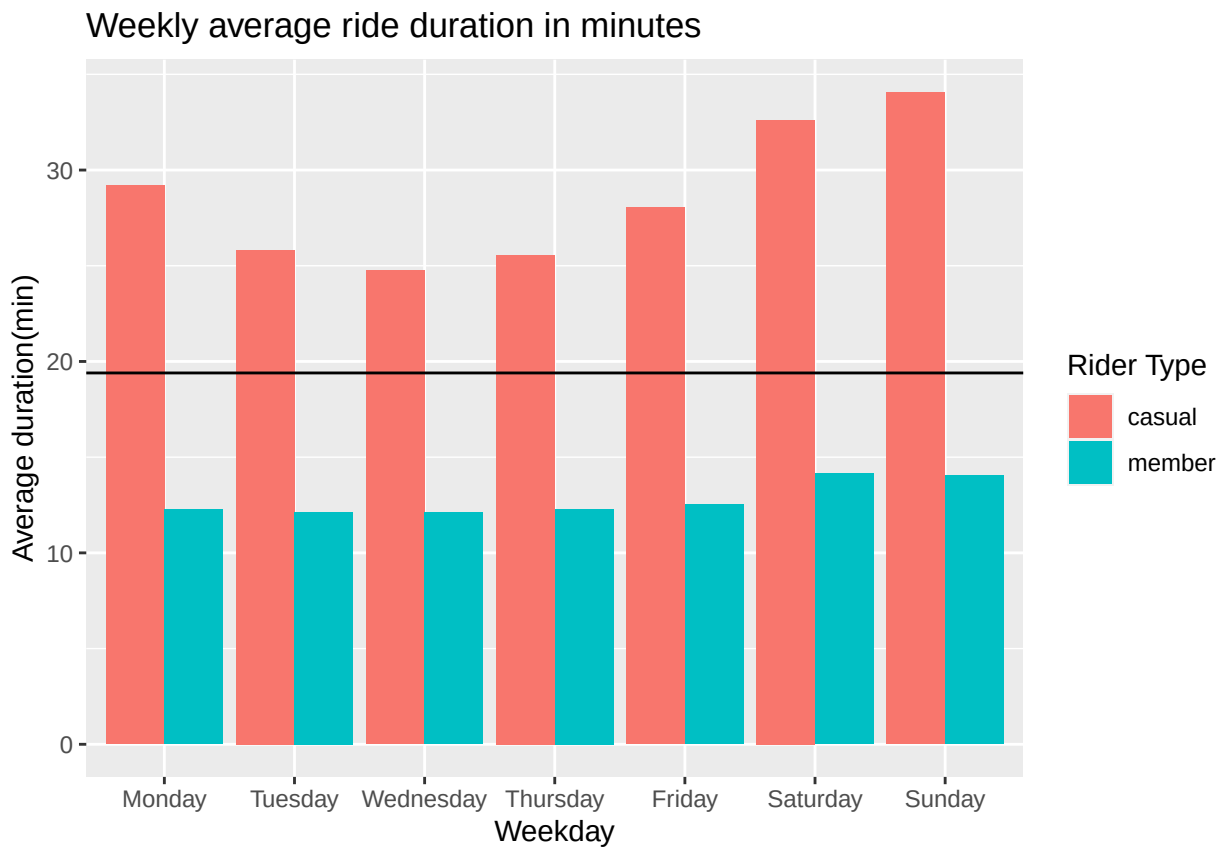


Days aren't ordered to fix it

```
yearly_trips_v3$day_of_week <- ordered(yearly_trips_v3$day_of_week, levels=
                                         c("Monday","Tuesday", "Wednesday",
                                           "Thursday", "Friday", "Saturday",
                                           "Sunday"))
```

After fixing it

```
yearly_trips_v3 %>%
  group_by(member_casual, day_of_week) %>%
  summarise(number_of_rides = n(), average_duration=mean(ride_length)) %>%
  arrange(member_casual,day_of_week) %>%
  ggplot(aes(x=day_of_week, y=average_duration, fill= member_casual))+
  geom_col(position = "dodge")+
  geom_hline(yintercept = 19.4)+
  xlab("Weekday")+
  ylab("Average duration(min)")+
  labs(title = "Weekly average ride duration in minutes" , fill ="Rider Type")
```



Number of rides

- Lets look at yearly trips.

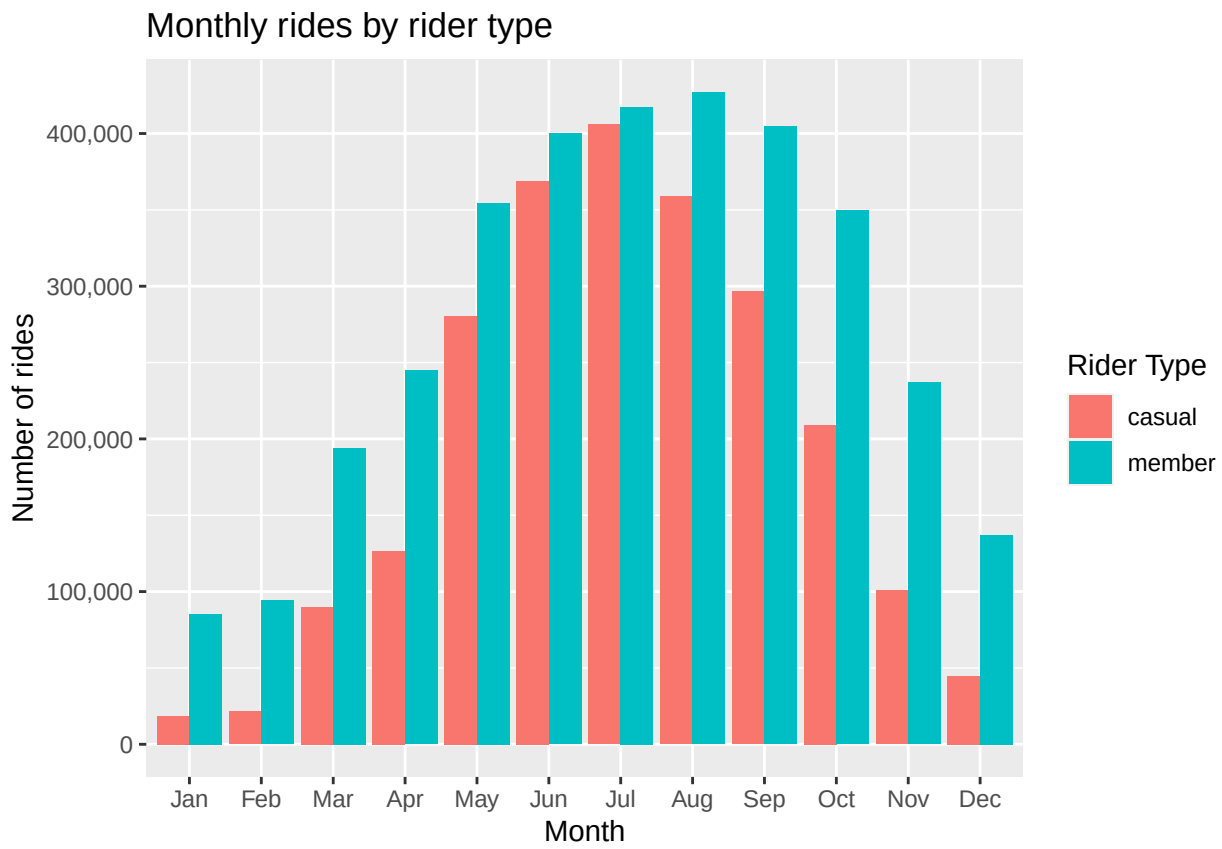
```
yearly_trips_v3 %>%
  summarise(number_of_rides = n())
```

```
##   number_of_rides
## 1         5667186
```

This year total of 5.667.186 trips made by our members and casuals.

- If we look at our monthly trips.

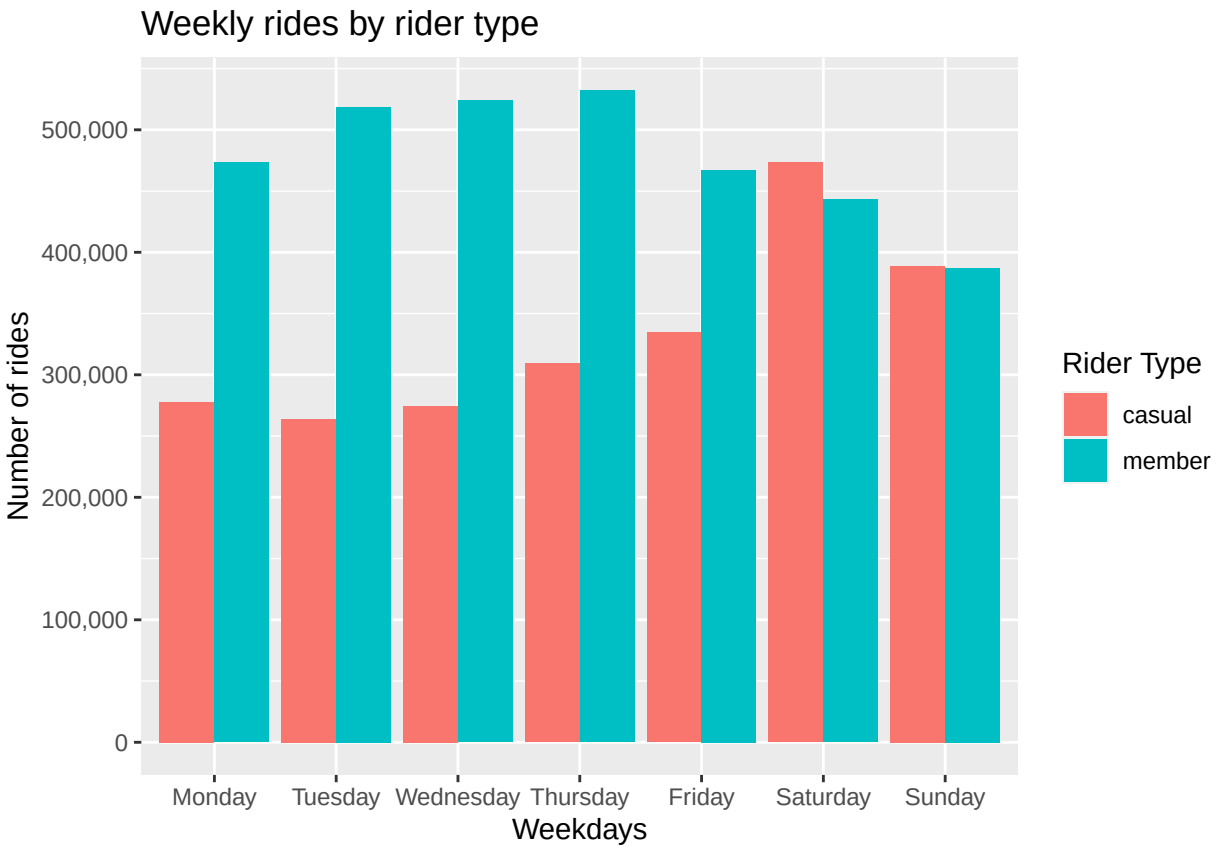
```
yearly_trips_v3 %>%
  group_by(month, member_casual) %>%
  summarise(number_of_rides = n()) %>%
  ggplot(aes(x=month,y=number_of_rides, fill=member_casual))+
  geom_col(position = "dodge")+
  xlab("Month")+
  scale_x_discrete(labels = c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug",
                              "Sep","Oct","Nov","Dec"))+
  ylab("Number of rides")+
  scale_y_continuous(labels = comma)+
  labs(title="Monthly rides by rider type", fill= "Rider Type")
```



From the plot we can see that casuals tends to rent more bike in summer.

- If we look at weekly trips.

```
yearly_trips_v3 %>%
  group_by(day_of_week, member_casual) %>%
  summarise(rider_count = n()) %>%
  ggplot(aes(x=day_of_week,y=rider_count, fill=member_casual))+
  geom_col(position = "dodge")+
  xlab("Weekdays")+
  ylab("Number of rides")+
  scale_y_continuous(labels = comma, breaks = seq(0,500000, by =100000))+
  labs(title="Weekly rides by rider type", fill= "Rider Type")
```



In contrast to ride duration our members rent more bike and casual rent more bike at weekends.

Share Phase

After the analysis:

- Casuals ride longer than members.
- Members rent more bike than casuals.
- Casuals tend to rent more bike in summer and weekends.

Recommendations

- A discount campaign can be launched for membership in the summer since the casual usage rate increases.
- Discount coupons can be issued according to the length of their trips.