DIVVY_Bikes_Q1_2020

STEP 1: Load Packages and Upload Dataset

Install required packages:

- tidyverse for data import and wrangling
- lubridate for date functions
- ggplot for visualization

```
install.packages("tidyverse")
install.packages("lubridate")
install.packages("ggplot2")
```

Upload Divvy datasets (csv files):

```
q1_2020 <- read.csv("Divvy_Trips_2020_Q1.csv")
```

STEP 2: CLEAN UP AND ADD DATA TO PREPARE FOR ANALYSIS

Add columns that list the date, month, day, and year of each ride.This will allow us to aggregate ride date for each month, day, or year. Before completing this operations we could only aggregate at the ride level.

```
q1_2020$date <- as.Date(q1_2020$started_at)
q1_2020$month <- format(as.Date(q1_2020$date), "%m")
q1_2020$day <- format(as.Date(q1_2020$date), "%d")
q1_2020$year <- format(as.Date(q1_2020$date), "%Y")
q1_2020$day_of_week <- format(as.Date(q1_2020$date), "%A")
```

Add a “ride_length” calculation to all_trips (in seconds)

```
q1_2020$ride_length<- difftime(q1_2020$ended_at,q1_2020$started_at, units="secs")
```

Convert “ride_length” from Factor to numeric so we can run calculations on the data

```
q1_2020$ride_length<-as.numeric(as.character(q1_2020$ride_length))
```

Remove “bad” data. The data frame includes a few hundred entries when bikes were taken out of docks and checked for quality by Divvy or ride_length was negative. We will create a new version of the data frame (v2) since data is being removed:

```
q1_2020_v2<-q1_2020[!(q1_2020$start_station_name == "HQ QR"|q1_2020$ride_length<0),]
```

##STEP 3: CONDUCT DESCRIPTIVE ANALYSIS

Descriptive analysis on ride_length (all figures in seconds):

```
mean(q1_2020_v2$ride_length)
median(q1_2020_v2$ride_length)
max(q1_2020_v2$ride_length)
min(q1_2020_v2$ride_length)
summary(q1_2020_v2$ride_length)
```

Compare members and casual users by ride length:

```
aggregate(q1_2020_v2$ride_length~q1_2020_v2$member_casual, FUN=mean)
aggregate(q1_2020_v2$ride_length~q1_2020_v2$member_casual, FUN=median)
aggregate(q1_2020_v2$ride_length~q1_2020_v2$member_casual, FUN=max)
aggregate(q1_2020_v2$ride_length~q1_2020_v2$member_casual, FUN=min)
```

See the average ride time by each day for members vs casual users:

```
aggregate(q1_2020_v2$ride_length~q1_2020_v2$member_casual+q1_2020_v2$day_of_week, FUN=mean)
```

Notice that the days of the week are out of order. Let’s fix that:

```
q1_2020_v2$day_of_week<- ordered(q1_2020_v2$day_of_week, levels=c("Sunday", "Monday", "Tuesday", "Wednesday", "Th
ursday", "Friday", "Saturday"))
```

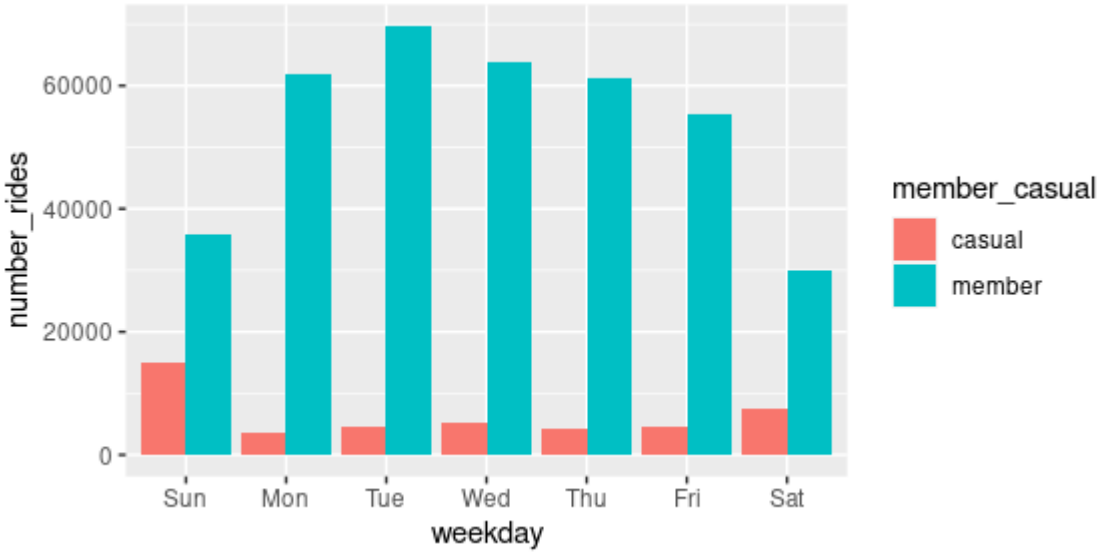
Analyze ridership data by type and weekday:

```
q1_2020_v2 %>%
  mutate(weekday=wday(started_at,label=TRUE)) %>%
  group_by(member_casual,weekday) %>%
  summarize(number_rides=n(),average=mean(ride_length)) %>%
  arrange(member_casual,weekday)
```

Step 4. Plots

Visualize then number of rides throughout the week by rider type:

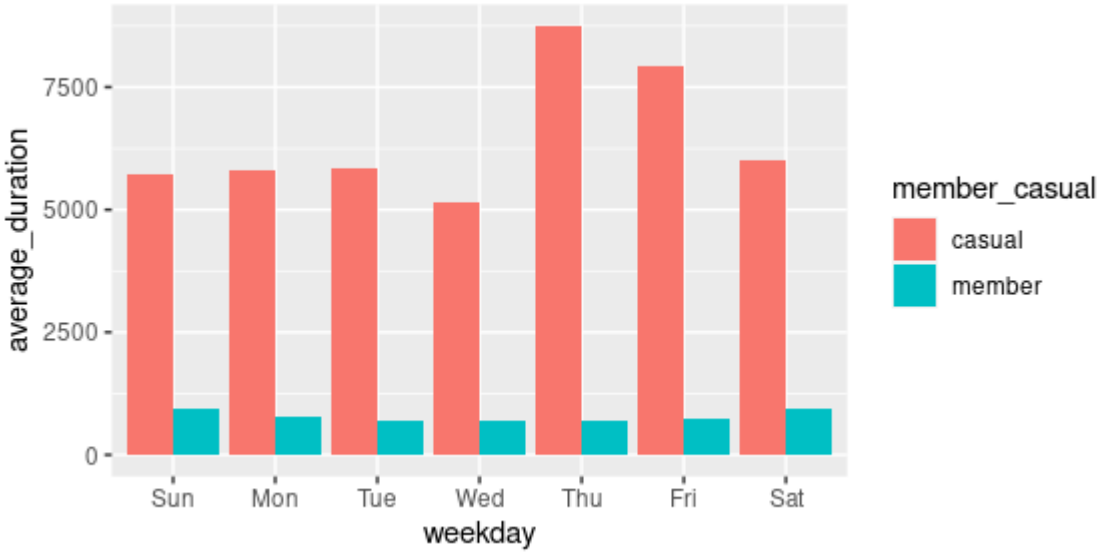
```
q1_2020_v2 %>%
  mutate(weekday=wday(started_at,label=TRUE)) %>%
  group_by(member_casual,weekday) %>%
  summarize(number_rides=n(),average=mean(ride_length)) %>%
  arrange(member_casual,weekday) %>%
  ggplot(aes(x=weekday, y=number_rides, fill=member_casual)) +
  geom_col(position="dodge")
```



Plot #1

Let's create a visualization for the average duration throughout the week by rider type

```
q1_2020_v2 %>%
  mutate(weekday=wday(started_at,label=TRUE)) %>%
  group_by(member_casual,weekday) %>%
  summarize(number_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")
```



Plot #2

CONCLUSION

Casual riders used DIVVY bikes in the first quarter of 2020 only for longer distances, whereas members used DiVVY bikes more often, however, for less period of time. The last can be explained that since members are already paying for the subscription, they were using DIVVY's services for shorter ride lengths to quicker get from point A to point B, even if it was a walking distance. On the other hand, casual riders only were willing to pay for the service if they were heading somewhere far and needed a transportation.