Cyclist Case Study

2023-07-20

Preparing and Processing

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(lubridate)
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(skimr)
```

```
Aug_2022 <- read.csv("202208-divvy-tripdata.csv")
Sept_2022 <- read.csv("202209-divvy-tripdata.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")
Jan_2023 <- read.csv("202301-divvy-tripdata.csv")
Feb_2023 <- read.csv("202302-divvy-tripdata.csv")
Mar_2023 <- read.csv("202303-divvy-tripdata.csv")
Apr_2023 <- read.csv("202304-divvy-tripdata.csv")
May_2023 <- read.csv("202305-divvy-tripdata.csv")
June_2023 <- read.csv("202306-divvy-tripdata.csv")
July_2023 <- read.csv("202307-divvy-tripdata.csv")
```

Combining all the extracted csv files

```
combined_trips <- bind_rows(Aug_2022,Sept_2022,Oct_2022,Nov_2022,Dec_2022,Jan_2023,Feb_2023,Mar_2023,Ap
```

Creating a new column of ride length by using the difference of time between start to end.

```
combined_trips$ride_length <- difftime(combined_trips$ended_at,combined_trips$started_at, units = "min"
combined_trips$ride_length <- round(combined_trips$ride_length, 2)
combined_trips$ride_length <- as.numeric(as.character(combined_trips$ride_length))
```

Separating day of the week, month and year from the date column

```
combined_trips$date <- as.Date(combined_trips$started_at)
combined_trips$month <- format(as.Date(combined_trips$date), "%B")
combined_trips$day <- format(as.Date(combined_trips$date), "%d")
combined_trips$year <- format(as.Date(combined_trips$date), "%Y")
combined_trips$day_of_the_week <- weekdays(combined_trips$date)
```

Analysis

```
combo2 <- drop_na(combined_trips)
combo2 %>%
  group_by (member_casual) %>%
  summarise(number_of_rides=n(), average_ride_length=mean(ride_length))
```

```
## # A tibble: 2 x 3
##   member_casual number_of_rides average_ride_length
##   <chr>                  <int>               <dbl>
## 1 casual               2164281                20.3
## 2 member               3553223                12.0
```

```
combo2%>%
  group_by(member_casual)%>%
  summarise(number_of_rides=n(), min_ride_length=min(ride_length),max_ride_length=max(ride_length),avg_
```

```
## # A tibble: 2 x 6
##   member_casual number_of_rides min_ride_length max_ride_length avg_ride_length
##   <chr>                  <int>           <dbl>           <dbl>           <dbl>
## 1 casual               2164281           -60.2          12136.            20.3
## 2 member               3553223        -10353.            1500.            12.0
## # i 1 more variable: median_ride_length <dbl>
```

extracting data about average ride lengths in months and Days of the week

```
combo2$month <- ordered(combo2$month, levels=c( "August", "September", "October", "November", "December

combo2$day_of_the_week <- ordered(combo2$day_of_the_week, levels=c("Sunday","Monday","Tuesday","Wednesda
```

```
combo2 %>%
  group_by(member_casual, month) %>%
  summarise(number_of_rides=n(), average_ride_length=mean(ride_length))%>%
  arrange (month)
```

```
## `summarise()` has grouped output by 'member_casual'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 24 x 4
## # Groups:   member_casual [2]
##    member_casual month     number_of_rides average_ride_length
##    <chr>         <ord>               <int>               <dbl>
##  1 casual        August             358168                21.4
##  2 member        August             426921                13.1
##  3 casual        September          296077                20.0
##  4 member        September          404550                12.6
##  5 casual        October            208612                18.4
##  6 member        October            349598                11.5
##  7 casual        November           100584                15.5
##  8 member        November           236921                10.9
##  9 casual        December            44791                13.4
## 10 member        December           136887                10.4
## # i 14 more rows
```

```
combo2 %>%
  group_by(member_casual, day_of_the_week) %>%
  summarise(number_of_rides=n(), average_ride_length=mean(ride_length))%>%
  arrange (day_of_the_week)
```

```
## `summarise()` has grouped output by 'member_casual'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 14 x 4
## # Groups:   member_casual [2]
##    member_casual day_of_the_week number_of_rides average_ride_length
##    <chr>         <ord>                   <int>               <dbl>
##  1 casual        Sunday                 331637                23.2
##  2 member        Sunday                 386424                13.3
##  3 casual        Monday                 257367                20.1
##  4 member        Monday                 501368                11.5
##  5 casual        Tuesday                256326                18.4
##  6 member        Tuesday                551845                11.6
##  7 casual        Wednesday              261856                17.4
##  8 member        Wednesday              565797                11.5
##  9 casual        Thursday               288654                18.0
## 10 member        Thursday               569774                11.6
## 11 casual        Friday                 334328                19.8
## 12 member        Friday                 516612                12.0
## 13 casual        Saturday               434113                23.0
## 14 member        Saturday               461403                13.4
```

Looking at popular ride routes (start and end stations combined)

```
combo3 <- (unite(combo2, "ride_routes", start_station_name, end_station_name, sep= " to "))
head (combo3)
```

```
##            ride_id rideable_type          started_at            ended_at
## 1 550CF7EFEAE0C618 electric_bike 2022-08-07 21:34:15 2022-08-07 21:41:46
## 2 DAD198F405F9C5F5 electric_bike 2022-08-08 14:39:21 2022-08-08 14:53:23
## 3 E6F2BC47B65CB7FD electric_bike 2022-08-08 15:29:50 2022-08-08 15:40:34
## 4 F597830181C2E13C electric_bike 2022-08-08 02:43:50 2022-08-08 02:58:53
## 5 0CE689BB4E313E8D electric_bike 2022-08-07 20:24:06 2022-08-07 20:29:58
## 6 BFA7E7CC69860C20 electric_bike 2022-08-08 13:06:08 2022-08-08 13:19:09
##   ride_routes start_station_id end_station_id start_lat start_lng end_lat
## 1          to                                     41.93    -87.69   41.94
## 2          to                                     41.89    -87.64   41.92
## 3          to                                     41.97    -87.69   41.97
## 4          to                                     41.94    -87.65   41.97
## 5          to                                     41.85    -87.65   41.84
## 6          to                                     41.79    -87.72   41.82
##   end_lng member_casual ride_length       date  month day year day_of_the_week
## 1  -87.72        casual        7.52 2022-08-07 August  07 2022          Sunday
## 2  -87.64        casual       14.03 2022-08-08 August  08 2022          Monday
## 3  -87.66        casual       10.73 2022-08-08 August  08 2022          Monday
## 4  -87.69        casual       15.05 2022-08-08 August  08 2022          Monday
## 5  -87.66        casual        5.87 2022-08-07 August  07 2022          Sunday
## 6  -87.69        casual       13.02 2022-08-08 August  08 2022          Monday
```

```
top_routes <- combo3 %>%
  group_by(ride_routes) %>%
  summarise(number_of_rides=n()) %>%
  arrange (desc (number_of_rides))
head (top_routes,10)
```

```
## # A tibble: 10 x 2
##    ride_routes                                                  number_of_rides
##    <chr>                                                                 <int>
##  1 " to "                                                               410862
##  2 "Streeter Dr & Grand Ave to Streeter Dr & Grand Ave"                  10596
##  3 "Ellis Ave & 60th St to University Ave & 57th St"                      7475
##  4 "DuSable Lake Shore Dr & Monroe St to DuSable Lake Shore Dr ~          7372
##  5 "Ellis Ave & 60th St to Ellis Ave & 55th St"                           7020
##  6 "University Ave & 57th St to Ellis Ave & 60th St"                      6879
##  7 "Ellis Ave & 55th St to Ellis Ave & 60th St"                           6642
##  8 "Michigan Ave & Oak St to Michigan Ave & Oak St"                       5253
##  9 "DuSable Lake Shore Dr & Monroe St to Streeter Dr & Grand Av~          5162
## 10 "State St & 33rd St to Calumet Ave & 33rd St"                          4460
```

```
top_routes2 <- combo3 %>%
  group_by(ride_routes, member_casual) %>%
  summarise(number_of_rides=n()) %>%
  arrange (desc(number_of_rides))
```

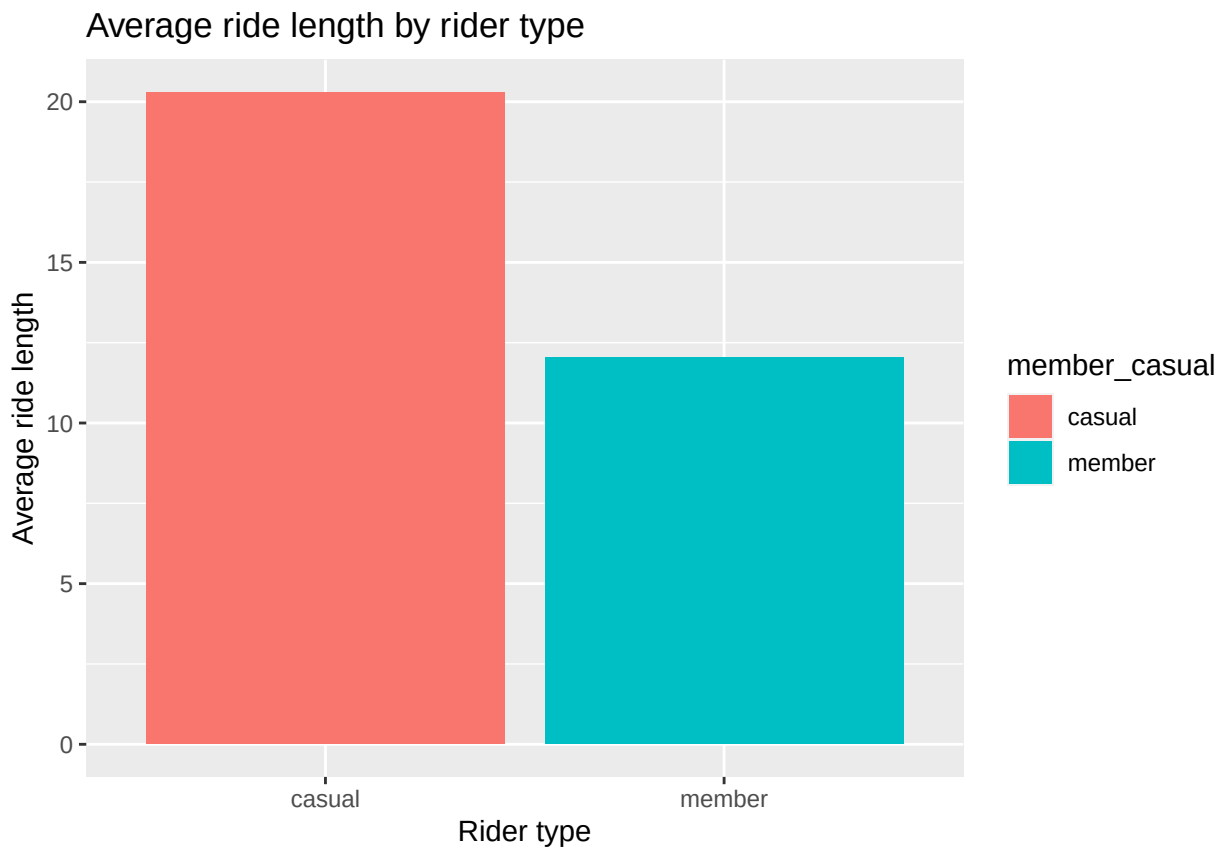
```
## `summarise()` has grouped output by 'ride_routes'. You can override using the
## `.groups` argument.
```

```r
head (top_routes2, 10)
```

```
## # A tibble: 10 x 3
## # Groups:   ride_routes [9]
##    ride_routes                                      member_casual number_of_rides
##    <chr>                                            <chr>                  <int>
##  1 " to "                                           member                223249
##  2 " to "                                           casual                187613
##  3 "Streeter Dr & Grand Ave to Streeter Dr & Gran~ casual                  9141
##  4 "DuSable Lake Shore Dr & Monroe St to DuSable ~ casual                  6606
##  5 "Ellis Ave & 60th St to University Ave & 57th ~ member                  6120
##  6 "University Ave & 57th St to Ellis Ave & 60th ~ member                  5670
##  7 "Ellis Ave & 60th St to Ellis Ave & 55th St"    member                  5462
##  8 "Ellis Ave & 55th St to Ellis Ave & 60th St"    member                  5203
##  9 "DuSable Lake Shore Dr & Monroe St to Streeter~ casual                  4657
## 10 "Michigan Ave & Oak St to Michigan Ave & Oak S~ casual                  4231
```

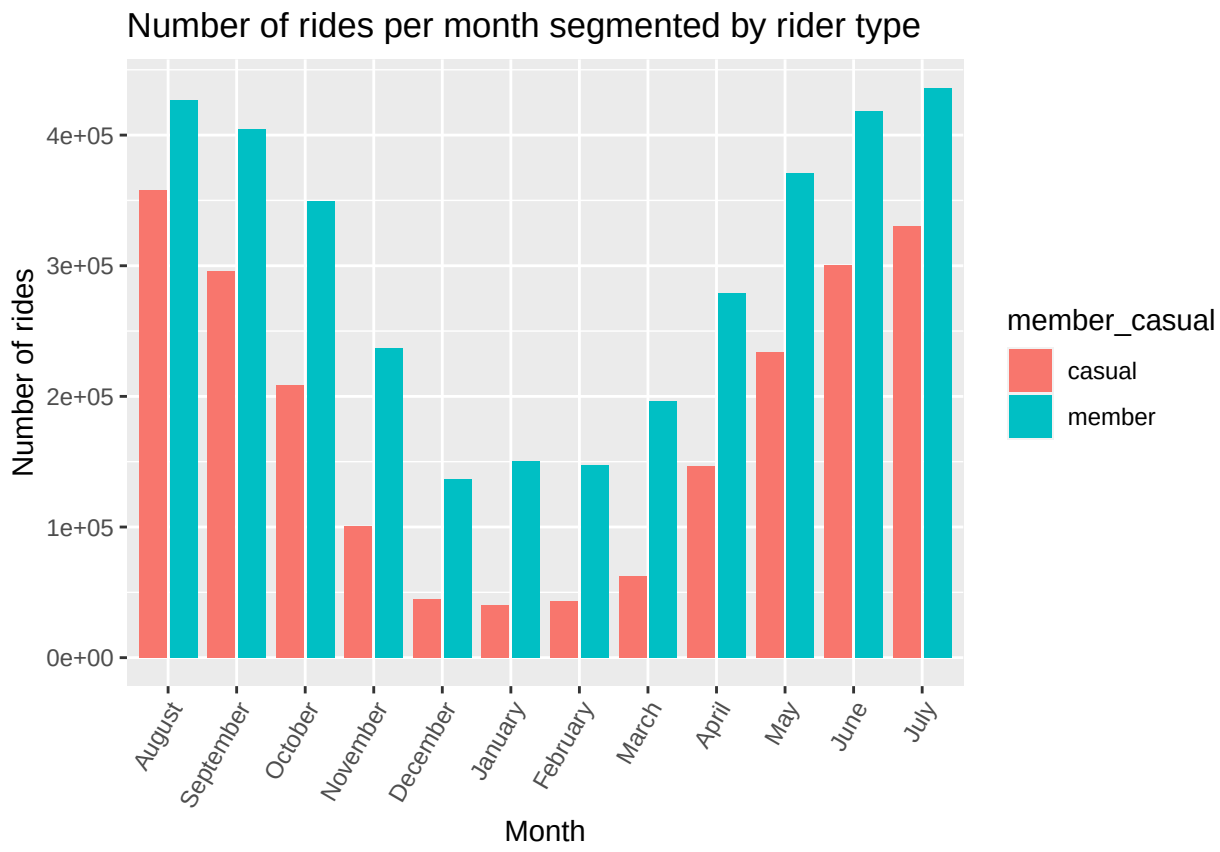
Visualisations

```r
combo3 %>%
  group_by(member_casual) %>%
  summarise(Average_ride_length=mean(ride_length)) %>%
  ggplot(aes(x= member_casual, y=Average_ride_length, fill=member_casual)) + geom_col() + labs(title =
```



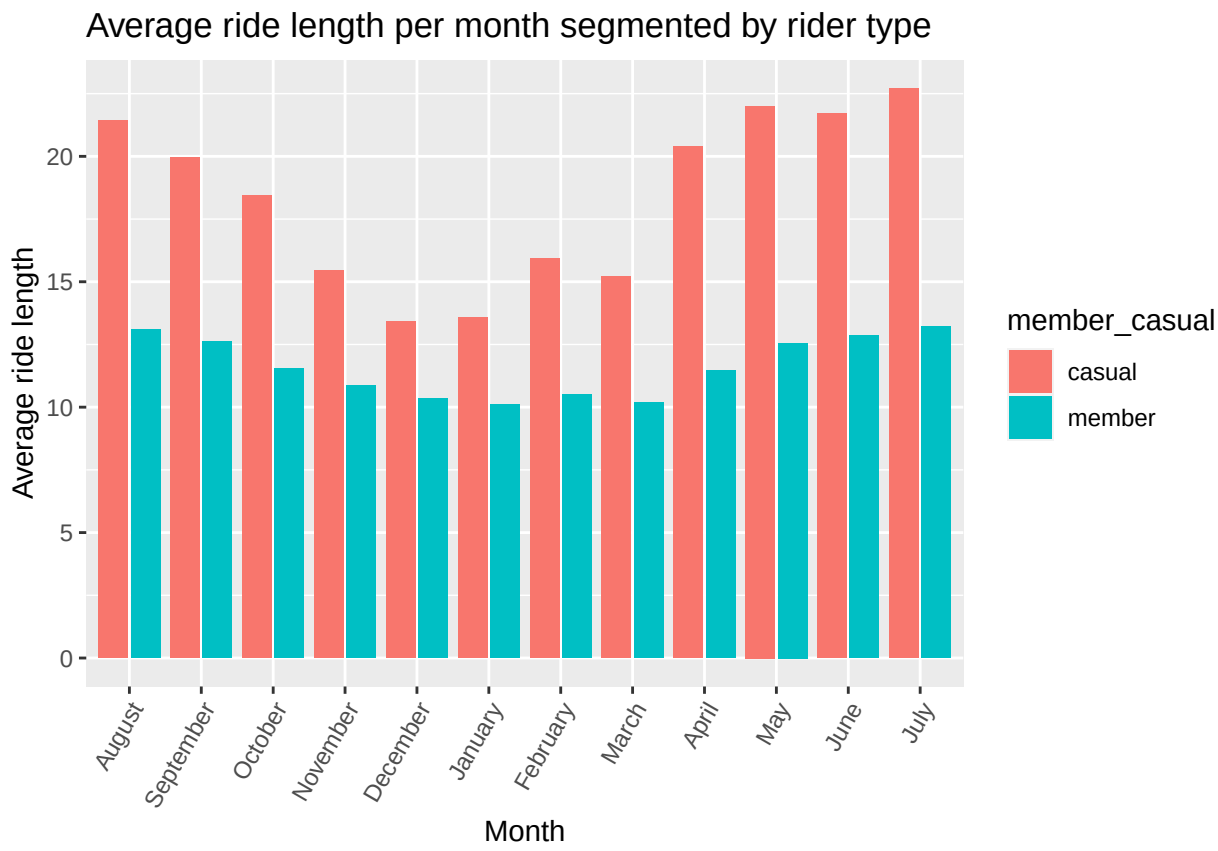
```
combo3 %>%
  group_by(member_casual, month) %>%
  summarise(number_of_rides=n(), average_ride_length=mean(ride_length))%>%
  ggplot (aes(x=month, y=number_of_rides, fill=member_casual)) + geom_col(position= "dodge2") + labs(ti
```

```
## `summarise()` has grouped output by 'member_casual'. You can override using the
## `.groups` argument.
```



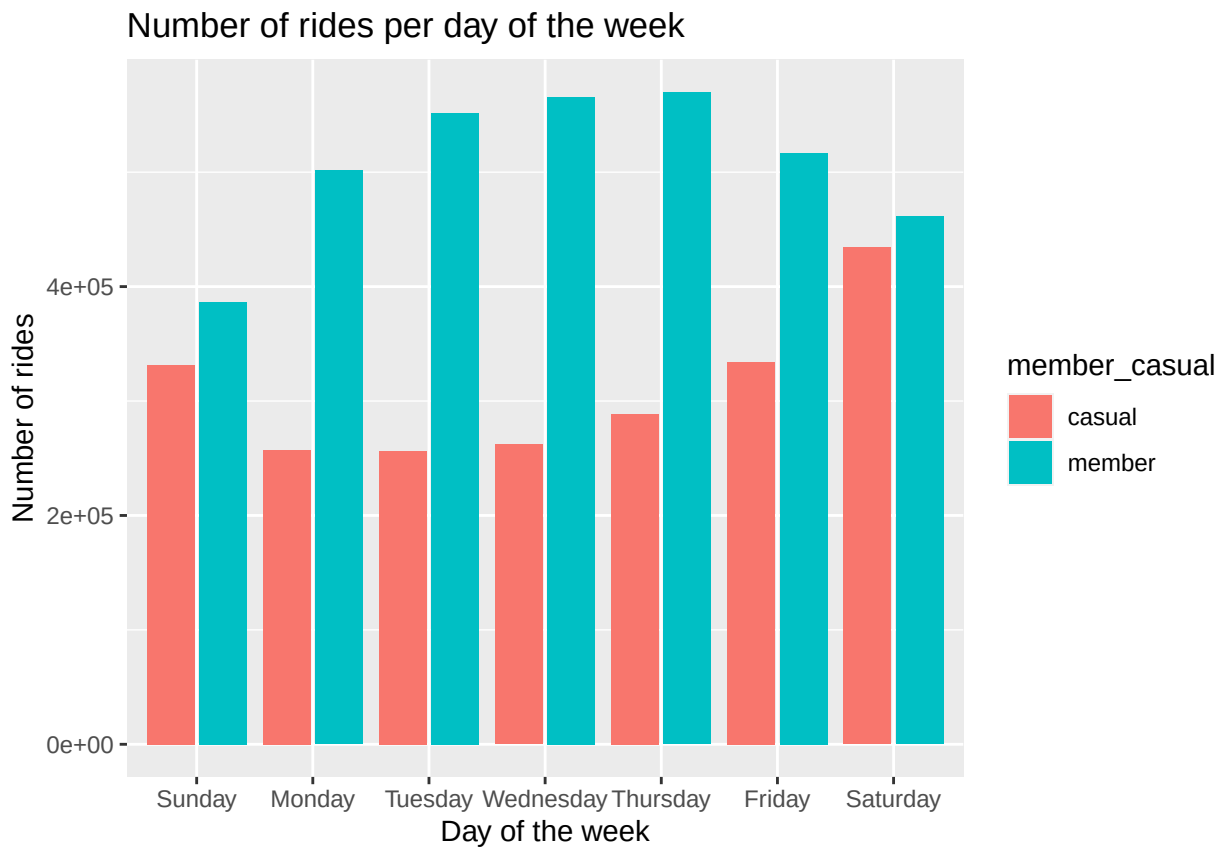

```
combo3 %>%
  group_by(member_casual, month) %>%
  summarise(number_of_rides=n(), average_ride_length=mean(ride_length))%>%
  ggplot (aes(x=month, y=average_ride_length, fill=member_casual)) + geom_col(position= "dodge2") + labs
```

```
## `summarise()` has grouped output by 'member_casual'. You can override using the
## `.groups` argument.
```



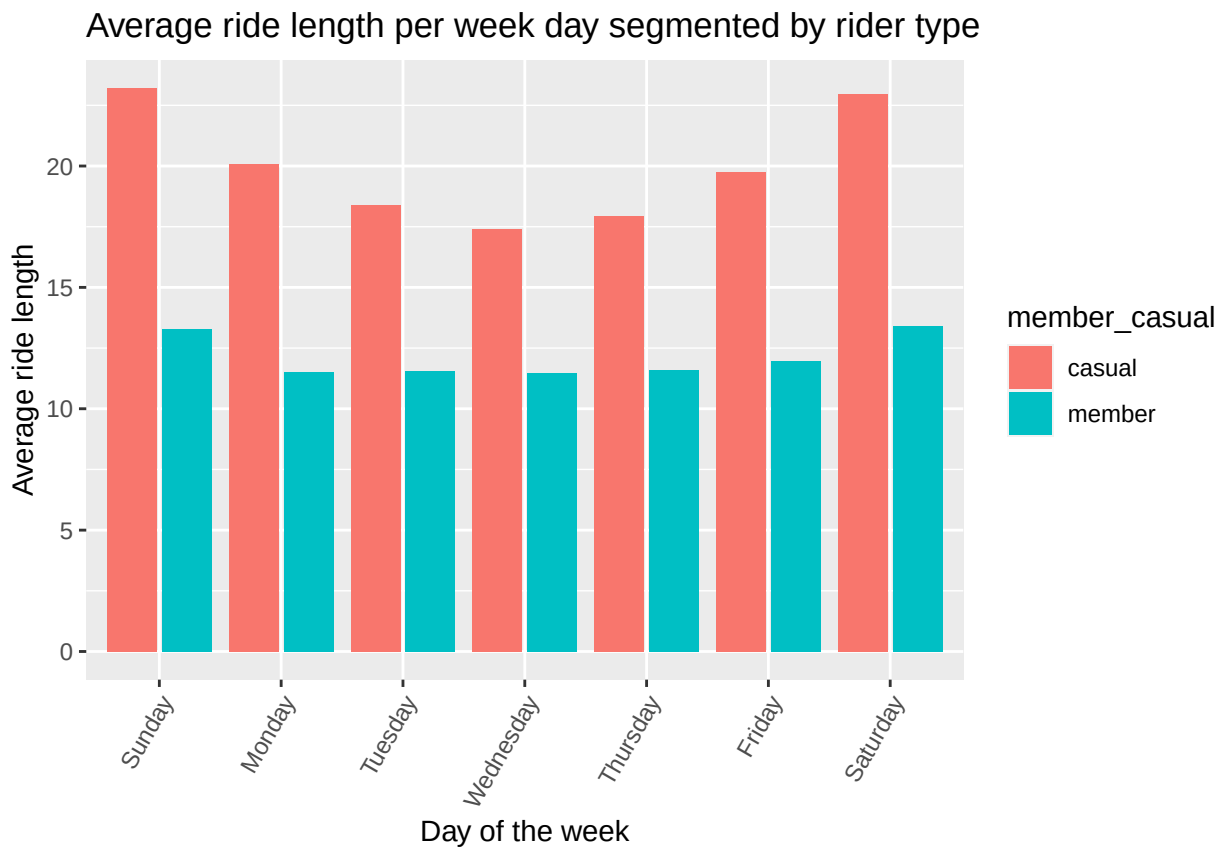

```
combo3 %>%
  group_by(member_casual, day_of_the_week) %>%
  summarise(number_of_rides=n(), average_ride_length=mean(ride_length))%>%
  ggplot (aes(x=day_of_the_week, y=number_of_rides, fill=member_casual)) + geom_col(position= "dodge2")
```

```
## `summarise()` has grouped output by 'member_casual'. You can override using the
## `.groups` argument.
```



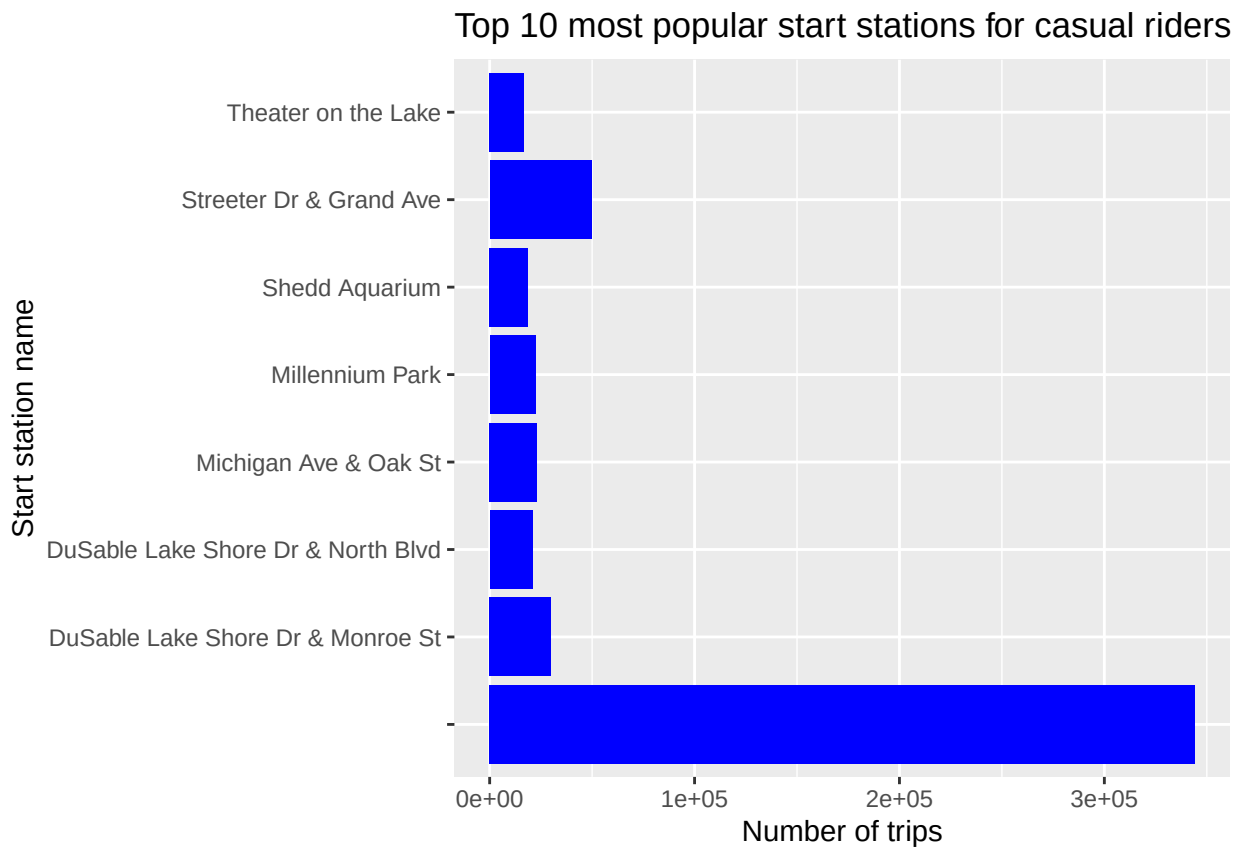

```
combo3 %>%
  group_by(member_casual, day_of_the_week) %>%
  summarise(number_of_rides=n(), average_ride_length=mean(ride_length))%>%
  ggplot (aes(x=day_of_the_week, y=average_ride_length, fill=member_casual)) + geom_col(position= "dodg
```

```
## `summarise()` has grouped output by 'member_casual'. You can override using the
## `.groups` argument.
```




```
combo2 %>%
  group_by(start_station_name, member_casual) %>%
  summarise(number_of_trips=n()) %>%
  arrange(desc (number_of_trips)) %>%
  filter(member_casual== "casual", number_of_trips >= 15460) %>%
  select(start_station_name, number_of_trips) %>%
  ggplot(aes(x=start_station_name, y=number_of_trips)) + geom_col(fill="blue") + coord_flip() + labs(ti
```

```
## `summarise()` has grouped output by 'start_station_name'. You can override
## using the `.groups` argument.
```



importing processed csv for more visualisations

```
write.csv(top_routes,"C:\\Users\\sacha\\OneDrive\\Desktop\\data a\\data\\top_routes.csv", row.names=FALS
write.csv(combo3,"C:\\Users\\sacha\\OneDrive\\Desktop\\data a\\data\\combo3.csv", row.names=FALSE)
```