Cyclistic - Bike Share Analysis Case Study

saurabh Kamble 31/05/2021

A Capstone Project for Google Data Analytics Professional Certification

Business Task:

With the help of data analysis Understand how we can convert casual riders into annual members and how do casual riders and annual members of Cyclistic - Bike Share use bikes differently.

Stakeholders: 1.Lily Moreno - Director of Marketing 2.Cyclistic marketing analytics team 3.Cyclistic executive team

Asking the right questions

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?

Data set The data set we'll use can be downloaded from here (https://divvy-tripdata.s3.amazonaws.com/index.html (https://divvy-tripdata.s3.amazonaws.com/index.html)) . Thisdataset is publically available under this license (https://www.divvybikes.com/data-license-agreement (https://www.divvybikes.com/data-license-agreement))

Loading Packages

```
install.packages("tidyverse", repos = "https://cran.rstudio.com")
```

```
## Installing package into 'C:/Users/saura/OneDrive/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)
```

```
## package 'tidyverse' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\saura\AppData\Local\Temp\RtmpEtdRor\downloaded_packages
```

```
install.packages("lubridate", repos = "https://cran.rstudio.com")
```

```
## Installing package into 'C:/Users/saura/OneDrive/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)
```

```
## package 'lubridate' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'lubridate'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE):
## problem copying C:\Users\saura\OneDrive\Documents\R\win-
## library\4.0\00LOCK\lubridate\libs\x64\lubridate.dll
## to C:\Users\saura\OneDrive\Documents\R\win-
## library\4.0\lubridate\libs\x64\lubridate.dll: Permission denied
```

```
## Warning: restored 'lubridate'
```

```
##
## The downloaded binary packages are in
##  C:\Users\saura\AppData\Local\Temp\RtmpEtdRor\downloaded_packages
```

```
install.packages("ggplot2", repos = "https://cran.rstudio.com")
```

```
## Installing package into 'C:/Users/saura/OneDrive/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)
```

```
## package 'ggplot2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\saura\AppData\Local\Temp\RtmpEtdRor\downloaded_packages
```

```
install.packages("dplyr", repos = "https://cran.rstudio.com")
```

```
## Installing package into 'C:/Users/saura/OneDrive/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)
```

```
## package 'dplyr' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'dplyr'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE):
## problem copying C:\Users\saura\OneDrive\Documents\R\win-
## library\4.0\00LOCK\dplyr\libs\x64\dplyr.dll to C:
## \Users\saura\OneDrive\Documents\R\win-library\4.0\dplyr\libs\x64\dplyr.dll:
## Permission denied
```

```
## Warning: restored 'dplyr'
```

```
##
## The downloaded binary packages are in
##  C:\Users\saura\AppData\Local\Temp\RtmpEtdRor\downloaded_packages
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(ggplot2)
library(dplyr)
```

CHECKING WORK DIRECTORY

```
getwd()
```

```
## [1] "C:/New folder/cyclist_casestudy_analysis"
```

Uploading the Data set

```
m1<- read_csv("Dataset/may_2020.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   start_date = col_character(),
##   started_at = col_time(format = ""),
##   end_date = col_character(),
##   ended_at = col_time(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   member_casual = col_character(),
##   ride_length = col_time(format = ""),
##   day_of_week = col_character()
## )
```

```
m2 <- read_csv("Dataset/Jun_2020.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   start_date = col_character(),
##   started_at = col_time(format = ""),
##   end_date = col_character(),
##   ended_at = col_time(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   member_casual = col_character(),
##   ride_length = col_time(format = ""),
##   day_of_week = col_character()
## )
```

```
m3 <- read_csv("Dataset/july_2020.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   start_date = col_character(),
##   started_at = col_time(format = ""),
##   end_date = col_character(),
##   ended_at = col_time(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   member_casual = col_character(),
##   ride_length = col_time(format = ""),
##   day_of_week = col_character()
## )
```

```
m4 <- read_csv("Dataset/Aug_2020.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   start_date = col_character(),
##   started_at = col_time(format = ""),
##   end_date = col_character(),
##   ended_at = col_time(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   member_casual = col_character(),
##   ride_length = col_time(format = ""),
##   day_of_week = col_character()
## )
```

```
m5 <- read_csv("Dataset/sep_2020.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   start_date = col_character(),
##   started_at = col_time(format = ""),
##   end_date = col_character(),
##   ended_at = col_time(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   member_casual = col_character(),
##   ride_length = col_time(format = ""),
##   day_of_week = col_character()
## )
```

```
m6 <- read_csv("Dataset/oct_2020.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   start_date = col_character(),
##   started_at = col_time(format = ""),
##   end_date = col_character(),
##   ended_at = col_time(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   member_casual = col_character(),
##   ride_length = col_time(format = ""),
##   day_of_week = col_character()
## )
```

```
m7 <- read_csv("Dataset/nov_2020.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   start_date = col_character(),
##   started_at = col_time(format = ""),
##   end_date = col_character(),
##   ended_at = col_time(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   member_casual = col_character(),
##   ride_length = col_time(format = ""),
##   day_of_week = col_character()
## )
```

```
m8 <- read_csv("Dataset/dec_2020.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   start_date = col_character(),
##   started_at = col_time(format = ""),
##   end_date = col_character(),
##   ended_at = col_time(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_character(),
##   end_station_name = col_character(),
##   end_station_id = col_character(),
##   member_casual = col_character(),
##   ride_length = col_time(format = ""),
##   day_of_week = col_character()
## )
```

```
m9 <- read_csv("Dataset/Jan_2021.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   start_date = col_character(),
##   started_at = col_time(format = ""),
##   end_date = col_character(),
##   ended_at = col_time(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_character(),
##   end_station_name = col_character(),
##   end_station_id = col_character(),
##   member_casual = col_character(),
##   ride_length = col_time(format = ""),
##   day_of_week = col_character()
## )
```

```
m10<- read_csv("Dataset/feb_2021.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   start_date = col_character(),
##   started_at = col_time(format = ""),
##   end_date = col_character(),
##   ended_at = col_time(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_character(),
##   end_station_name = col_character(),
##   end_station_id = col_character(),
##   member_casual = col_character(),
##   ride_length = col_time(format = ""),
##   day_of_week = col_character()
## )
```

```
m11 <- read_csv("Dataset/Mar_2021.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   start_date = col_character(),
##   started_at = col_time(format = ""),
##   end_date = col_character(),
##   ended_at = col_time(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_character(),
##   end_station_name = col_character(),
##   end_station_id = col_character(),
##   member_casual = col_character(),
##   ride_length = col_time(format = ""),
##   day_of_week = col_character()
## )
```

```
m12 <- read_csv("Dataset/apr_2021.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   start_date = col_character(),
##   started_at = col_time(format = ""),
##   end_date = col_character(),
##   ended_at = col_time(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_character(),
##   end_station_name = col_character(),
##   end_station_id = col_character(),
##   member_casual = col_character(),
##   ride_length = col_time(format = ""),
##   day_of_week = col_character()
## )
```

Cleaning data

Making data types consistent to avoid errors

```
m1 <-  mutate(m1, start_station_id = as.character(start_station_id)
              ,end_station_id  = as.character(end_station_id)
              ,start_date = as.Date(start_date, format= "%d-%m-%Y")
              ,end_date = as.Date(end_date, format= "%d-%m-%Y")
              ,ride_length = as.numeric(ride_length ))
```

```
m2 <-  mutate(m2, start_station_id = as.character(start_station_id)
              ,end_station_id  = as.character(end_station_id )
              ,start_date = as.Date(start_date, format= "%d-%m-%Y")
              ,end_date = as.Date(end_date, format= "%d-%m-%Y")
              ,ride_length = as.numeric(ride_length))
```

```
m3 <-  mutate(m3, start_station_id = as.character(start_station_id)
              ,end_station_id  = as.character(end_station_id )
              ,start_date = as.Date(start_date, format= "%d-%m-%Y")
              ,end_date = as.Date(end_date, format= "%d-%m-%Y")
              ,ride_length = as.numeric(ride_length))
```

```
m4 <-  mutate(m4, start_station_id = as.character(start_station_id)
              ,end_station_id  = as.character(end_station_id )
              ,start_date = as.Date(start_date, format= "%d-%m-%Y")
              ,end_date = as.Date(end_date, format= "%d-%m-%Y")
              ,ride_length = as.numeric(ride_length))
```

```
m5 <-  mutate(m5, start_station_id = as.character(start_station_id)
              ,end_station_id  = as.character(end_station_id )
              ,start_date = as.Date(start_date, format= "%d-%m-%Y")
              ,end_date = as.Date(end_date, format= "%d-%m-%Y")
              ,ride_length = as.numeric(ride_length))
```

```
m6 <-  mutate(m6, start_station_id = as.character(start_station_id)
              ,end_station_id  = as.character(end_station_id )
              ,start_date = as.Date(start_date, format= "%d-%m-%Y")
              ,end_date = as.Date(end_date, format= "%d-%m-%Y")
              ,ride_length = as.numeric(ride_length))
```

```
m7 <-  mutate(m7, start_station_id = as.character(start_station_id)
              ,end_station_id  = as.character(end_station_id )
              ,start_date = as.Date(start_date, format= "%d-%m-%Y")
              ,end_date = as.Date(end_date, format= "%d-%m-%Y")
              ,ride_length = as.numeric(ride_length))
```

```
m8 <-  mutate(m8, start_station_id = as.character(start_station_id)
              ,end_station_id  = as.character(end_station_id )
              ,start_date = as.Date(start_date, format= "%d-%m-%Y")
              ,end_date = as.Date(end_date, format= "%d-%m-%Y")
              ,ride_length = as.numeric(ride_length))
```

```
m9 <-  mutate(m9,start_date = as.Date(start_date, format= "%d-%m-%Y")
              ,end_date = as.Date(end_date, format= "%d-%m-%Y")
              ,ride_length = as.numeric(ride_length))
```

```
m10 <-  mutate(m10,start_date = as.Date(start_date, format= "%d-%m-%Y")
               ,end_date = as.Date(end_date, format= "%d-%m-%Y")
               ,ride_length = as.numeric(ride_length))
```

```
m11 <-  mutate(m11,start_date = as.Date(start_date, format= "%d-%m-%Y")
               ,end_date = as.Date(end_date, format= "%d-%m-%Y")
               ,ride_length = as.numeric(ride_length))
```

```
m12 <-  mutate(m12,start_date = as.Date(start_date, format= "%d-%m-%Y")
               ,end_date = as.Date(end_date, format= "%d-%m-%Y")
               ,ride_length = as.numeric(ride_length))
```

Combining data into a single Dataset

```
all_tripsv2 <- bind_rows(m1, m2, m3, m4 , m5 ,m6 ,m7 ,m8 ,m9 ,m10 ,m11 ,m12 )
```

Formating start_Date into month,year,day and month and year column

```
all_tripsv2$date <- format(as.Date(all_tripsv2$start_date), "%Y-%m-%d")
all_tripsv2$month <- format(as.Date(all_tripsv2$date),"%m")
all_tripsv2$day <- format(as.Date(all_tripsv2$date),"%d")
all_tripsv2$year <- format(as.Date(all_tripsv2$date), "%Y")
all_tripsv2$month_group <- format(as.Date(all_tripsv2$date), "%Y-%m")
```

Let’s summarise our new data set

```
str(all_tripsv2)
```

```
## spec_tbl_df [3,231,614 x 18] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:3231614] "02668AD35674B983" "7A50CCAF1EDDB28F" "2FFCDFDB91FE9A52" "58991CF
1DB75BA84" ...
##  $ rideable_type     : chr [1:3231614] "docked_bike" "docked_bike" "docked_bike" "docked_bike" ...
##  $ start_date        : Date[1:3231614], format: "2020-05-27" "2020-05-25" ...
##  $ started_at        : 'hms' num [1:3231614] 10:03:52 10:47:11 14:11:03 16:25:36 ...
##   ..- attr(*, "units")= chr "secs"
##  $ end_date          : Date[1:3231614], format: "2020-05-27" "2020-05-25" ...
##  $ ended_at          : 'hms' num [1:3231614] 10:16:49 11:05:40 15:48:21 16:39:28 ...
##   ..- attr(*, "units")= chr "secs"
##  $ start_station_name: chr [1:3231614] "Franklin St & Jackson Blvd" "Clark St & Wrightwood Ave" "Kedzie
Ave & Milwaukee Ave" "Clarendon Ave & Leland Ave" ...
##  $ start_station_id  : chr [1:3231614] "36" "340" "260" "251" ...
##  $ end_station_name  : chr [1:3231614] "Wabash Ave & Grand Ave" "Clark St & Leland Ave" "Kedzie Ave & Mi
lwaukee Ave" "Lake Shore Dr & Wellington Ave" ...
##  $ end_station_id    : chr [1:3231614] "199" "326" "260" "157" ...
##  $ member_casual     : chr [1:3231614] "member" "casual" "casual" "casual" ...
##  $ ride_length       : num [1:3231614] 777 1109 5838 832 2237 ...
##  $ day_of_week       : chr [1:3231614] "Wednesday" "Monday" "Saturday" "Saturday" ...
##  $ date              : chr [1:3231614] "2020-05-27" "2020-05-25" "2020-05-02" "2020-05-02" ...
##  $ month             : chr [1:3231614] "05" "05" "05" "05" ...
##  $ day               : chr [1:3231614] "27" "25" "02" "02" ...
##  $ year              : chr [1:3231614] "2020" "2020" "2020" "2020" ...
##  $ month_group       : chr [1:3231614] "2020-05" "2020-05" "2020-05" "2020-05" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   start_date = col_character(),
##   ..   started_at = col_time(format = ""),
##   ..   end_date = col_character(),
##   ..   ended_at = col_time(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_double(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_double(),
##   ..   member_casual = col_character(),
##   ..   ride_length = col_time(format = ""),
##   ..   day_of_week = col_character()
##   .. )
```

Dimensions

```
dim(all_tripsv2)
```

```
## [1] 3231614      18
```

No. of rows

```
nrow(all_tripsv2)
```

```
## [1] 3231614
```

```
summary(all_tripsv2)
```

```
##    ride_id          rideable_type        start_date          started_at
##  Length:3231614     Length:3231614     Min.   :2020-05-01   Length:3231614
##  Class :character   Class :character   1st Qu.:2020-07-17   Class1:hms
##  Mode  :character   Mode  :character   Median :2020-08-15   Class2:difftime
##                                        Mean   :2020-09-20   Mode  :numeric
##                                        3rd Qu.:2020-11-16
##                                        Max.   :2021-04-30
##     end_date            ended_at        start_station_name start_station_id
##  Min.   :2020-05-01   Length:3231614     Length:3231614     Length:3231614
##  1st Qu.:2020-07-17   Class1:hms         Class :character   Class :character
##  Median :2020-08-15   Class2:difftime    Mode  :character   Mode  :character
##  Mean   :2020-09-20   Mode  :numeric
##  3rd Qu.:2020-11-16
##  Max.   :2021-05-05
##  end_station_name   end_station_id     member_casual       ride_length
##  Length:3231614     Length:3231614     Length:3231614     Min.   :    0
##  Class :character   Class :character   Class :character   1st Qu.:  498
##  Mode  :character   Mode  :character   Mode  :character   Median :  907
##                                                           Mean   : 1431
##                                                           3rd Qu.: 1646
##                                                           Max.   :82591
##  day_of_week            date              month              day
##  Length:3231614     Length:3231614     Length:3231614     Length:3231614
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      year           month_group
##  Length:3231614     Length:3231614
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

Analyze Phase

```
mean(all_tripsv2$ride_length)
```

```
## [1] 1431.44
```

```
median(all_tripsv2$ride_length)
```

```
## [1] 907
```

```
max(all_tripsv2$ride_length)
```

```
## [1] 82591
```

```
min(all_tripsv2$ride_length)
```

```
## [1] 0
```

```
aggregate(all_tripsv2$ride_length ~ all_tripsv2$member_casual, FUN = mean)
```

```
##   all_tripsv2$member_casual all_tripsv2$ride_length
## 1                    casual               2114.4032
## 2                    member                933.2411
```

```
aggregate(all_tripsv2$ride_length ~ all_tripsv2$member_casual, FUN = median)
```

```
##   all_tripsv2$member_casual all_tripsv2$ride_length
## 1                    casual                    1317
## 2                    member                     709
```

```
aggregate(all_tripsv2$ride_length ~ all_tripsv2$member_casual, FUN = max)
```

```
##   all_tripsv2$member_casual all_tripsv2$ride_length
## 1                    casual                   82591
## 2                    member                   74820
```

```
aggregate(all_tripsv2$ride_length ~ all_tripsv2$member_casual, FUN = min)
```

```
##   all_tripsv2$member_casual all_tripsv2$ride_length
## 1                    casual                       0
## 2                    member                       0
```

```
aggregate(all_tripsv2$ride_length ~ all_tripsv2$member_casual + all_tripsv2$day_of_week, FUN = mean)
```

```
##    all_tripsv2$member_casual all_tripsv2$day_of_week all_tripsv2$ride_length
## 1                     casual                  Friday               1941.2059
## 2                     member                  Friday                901.6121
## 3                     casual                  Monday               2034.5942
## 4                     member                  Monday                880.7959
## 5                     casual                Saturday               2278.7315
## 6                     member                Saturday               1035.2096
## 7                     casual                  Sunday               2386.7407
## 8                     member                  Sunday               1045.8114
## 9                     casual                Thursday               1912.8745
## 10                    member                Thursday                884.4855
## 11                    casual                 Tuesday               1978.0553
## 12                    member                 Tuesday                895.9399
## 13                    casual               Wednesday               1914.0463
## 14                    member               Wednesday                878.3114
```

```
all_tripsv2 %>%
  group_by(month_group, member_casual) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length)) %>%
  arrange(month_group, member_casual)
```

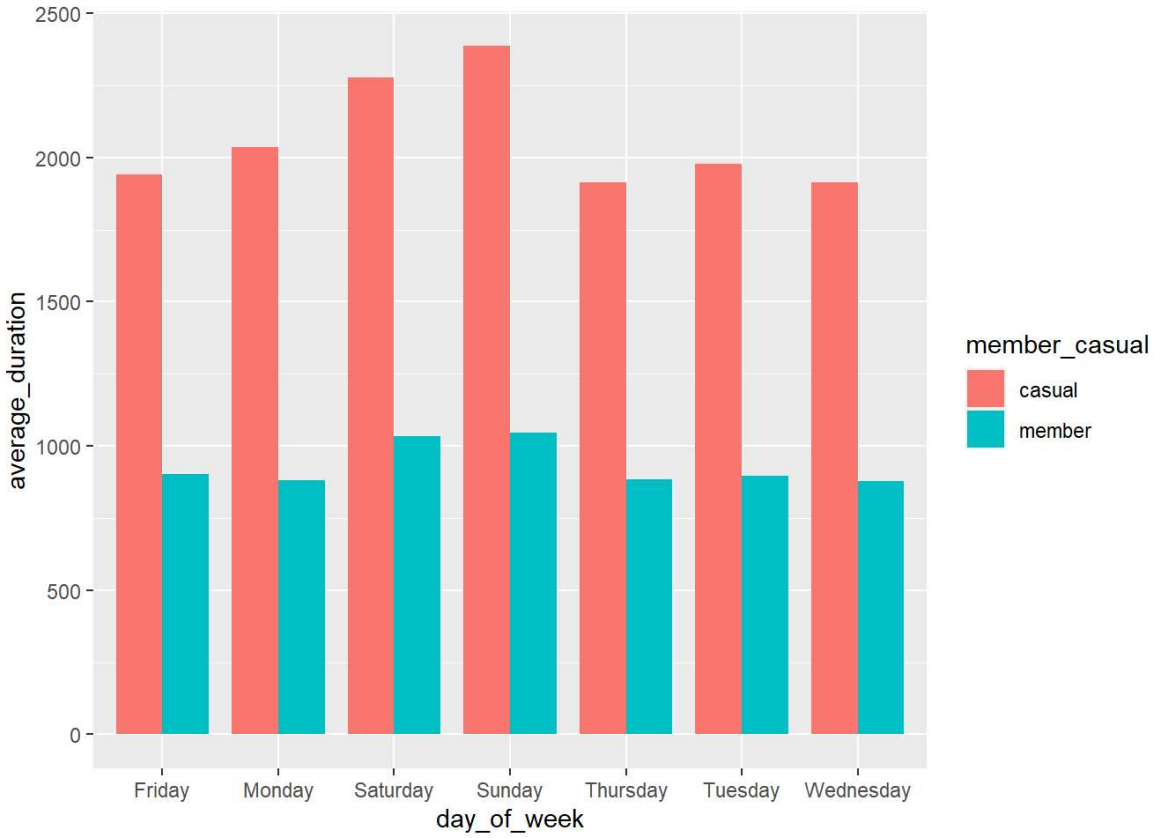
```
## `summarise()` has grouped output by 'month_group'. You can override using the `.groups` argument.
```

```
## # A tibble: 22 x 4
## # Groups:   month_group [11]
##    month_group member_casual number_of_rides average_duration
##    <chr>       <chr>                  <int>            <dbl>
##  1 2020-05     casual                 85433            2446.
##  2 2020-05     member                112878            1134.
##  3 2020-06     casual                151491            2316.
##  4 2020-06     member                187048            1069.
##  5 2020-07     casual                260035            2423.
##  6 2020-07     member                279589            1016.
##  7 2020-08     casual                554717            2097.
##  8 2020-08     member                646080             953.
##  9 2020-10     casual                   675            1116.
## 10 2020-10     member                     5             769.
## # ... with 12 more rows
```

Ploting of Data

```
all_tripsv2 %>%
  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", width = 0.8)
```

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

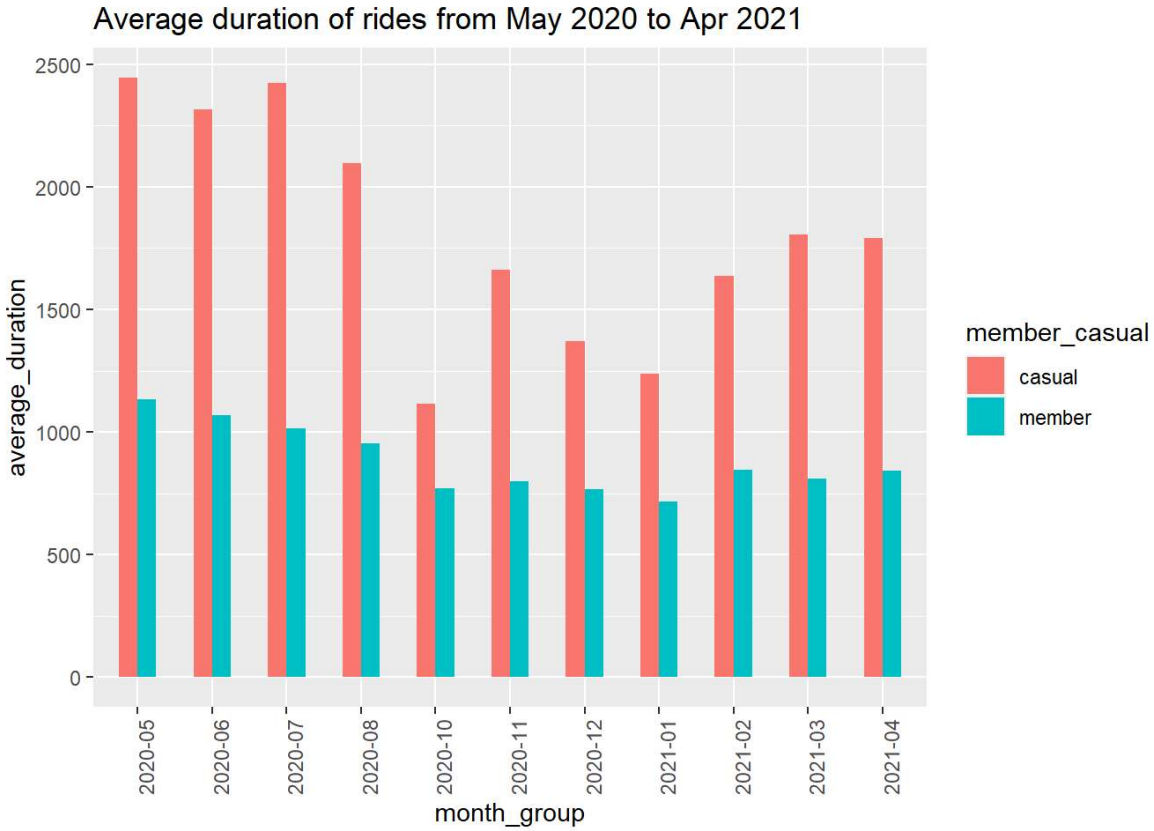



```
  ggtitle("Average duration of rides by week days")
```

```
## $title
## [1] "Average duration of rides by week days"
##
## attr(,"class")
## [1] "labels"
```

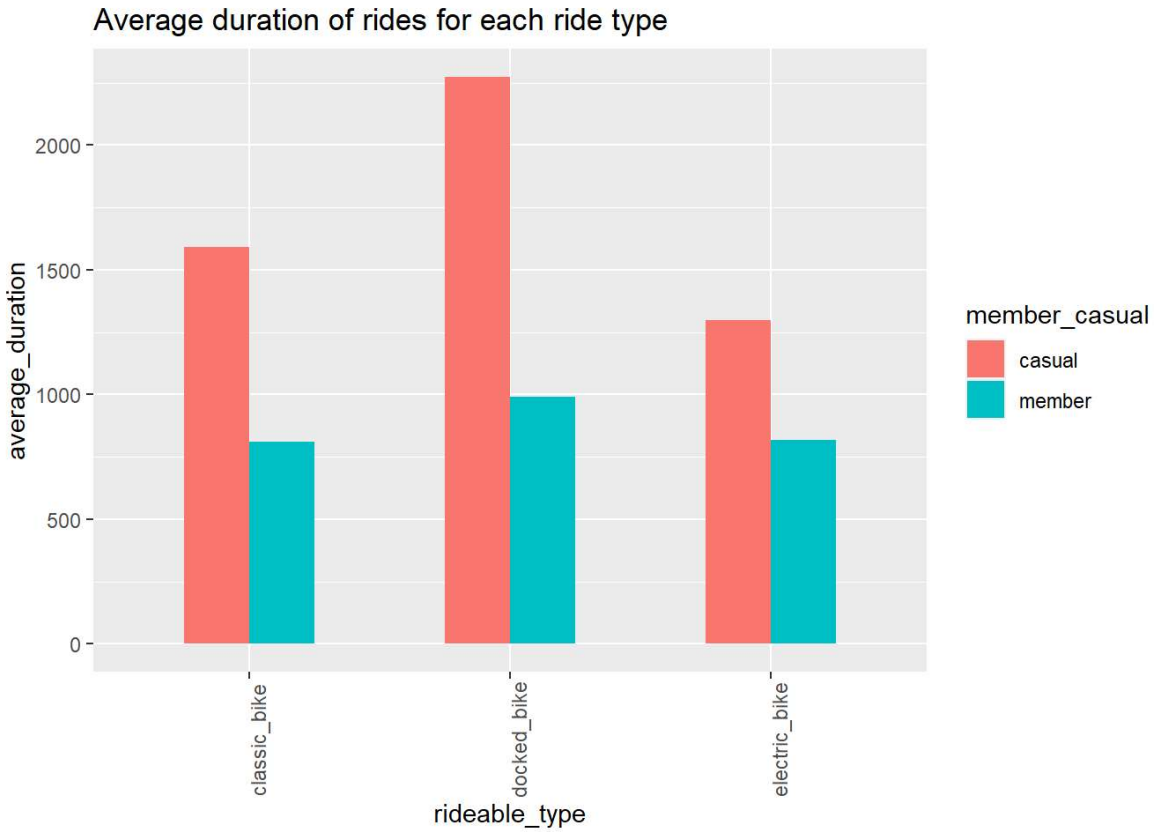
```
all_tripsv2%>%
  group_by(month_group, member_casual) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length)) %>%
  arrange(month_group, member_casual) %>%
  ggplot(aes(x = month_group, y = average_duration, fill = member_casual)) +
  geom_col(position = "dodge", width = 0.5) +
  theme(axis.text.x = element_text(angle = 90)) +
  ggtitle("Average duration of rides from May 2020 to Apr 2021")
```

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




```
all_tripsv2 %>%
  group_by(rideable_type, member_casual) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length)) %>%
  arrange(rideable_type, member_casual) %>%
  ggplot(aes(x = rideable_type, y = average_duration, fill = member_casual)) +
  geom_col(position = "dodge", width = 0.5) +
  theme(axis.text.x = element_text(angle = 90)) +
  ggtitle("Average duration of rides for each ride type")
```

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



Exporting Data

for further Analysis

#write.csv(agg_data, file = C:/New folder/cyclist_casestudy_analysis/agg_data.csv')

```
#agg_data = aggregate(all_tripsv2$ride_length ~ all_tripsv2$member_casual + all_tripsv2$day_of_week, FUN =
 mean)
```