

Cyclistic - Bike Share Case Study

Asmita Meshram

25/05/2021

A Capstone Project for Google Data Analytics Professional Certification

Business Task:

Understand how casual riders and annual members of Cyclistic - Bike Share use Cyclistic bikes differently and come up with efficient recommendations to convert casual riders into annual members

Stakeholders

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

Ask Phase of Data Analysis

As it is clear that we want to convert casual members into annual members, we can come up with these SMART 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?

Preparing Our Data

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

Installing the required libraries in R

```
# Libraries are already installed using install.packages()
```

```
library("tidyverse")
```

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

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.1     v dplyr   1.0.5
## v tidyrr   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")
```

```
getwd()
```

```
## [1] "D:/Main Folder/Google Data Analytics Course/Case Study/Cyclistic_Case_Study"
```

Uploading the datasets

```
data_may20 <- read_csv("DataSet/Clean_ds_null_removed/202005-divvy-tripdata.csv")
```

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

```
data_jun20 <- read_csv("DataSet/Clean_ds_null_removed/202006-divvy-tripdata.csv")
```

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

```
data_jul20 <- read_csv("DataSet/Clean_ds_null_removed/202007-divvy-tripdata.csv")
```

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

```
data_aug20 <- read_csv("DataSet/Clean_ds_null_removed/202008-divvy-tripdata.csv")
```

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

```
data_sep20 <- read_csv("DataSet/Clean_ds_null_removed/202009-divvy-tripdata.csv")
```

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

```
data_oct20 <- read_csv("DataSet/Clean_ds_null_removed/202010-divvy-tripdata.csv")
```

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

```
data_nov20 <- read_csv("DataSet/Clean_ds_null_removed/202011-divvy-tripdata.csv")
```

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

```
data_dec20 <- read_csv("DataSet/Clean_ds_null_removed/202012-divvy-tripdata.csv")
```

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

```
data_jan21 <- read_csv("DataSet/Clean_ds_null_removed/202101-divvy-tripdata.csv")
```

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

```
data_feb21 <- read_csv("DataSet/Clean_ds_null_removed/202102-divvy-tripdata.csv")
```

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

```
data_mar21 <- read_csv("DataSet/Clean_ds_null_removed/202103-divvy-tripdata.csv")
```

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

```
data_apr21 <- read_csv("DataSet/Clean_ds_null_removed/202104-divvy-tripdata.csv")
```

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

Processing and Cleaning the datasets

Inspecting dataframes

```
str(data_may20)
```

```

## spec_tbl_df[,15] [199,926 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ ride_id          : chr [1:199926] "02668AD35674B983" "7A50CCAF1EDDB28F" "2FFCDFDB91FE9A5
2" "58991CF1DB75BA84" ...
## $ rideable_type    : chr [1:199926] "docked_bike" "docked_bike" "docked_bike" "docked_bike"
...
## $ started_at       : chr [1:199926] "27-05-2020 10:03" "25-05-2020 10:47" "02-05-2020 14:1
1" "02-05-2020 16:25" ...
## $ ended_at         : chr [1:199926] "27-05-2020 10:16" "25-05-2020 11:05" "02-05-2020 15:4
8" "02-05-2020 16:39" ...
## $ day_of_week      : chr [1:199926] "Wednesday" "Monday" "Saturday" "Saturday" ...
## $ ride_length      : 'hms' num [1:199926] 00:13:00 00:18:00 01:37:00 00:14:00 ...
## ... attr(*, "units")= chr "secs"
## $ start_station_name: chr [1:199926] "Franklin St & Jackson Blvd" "Clark St & Wrightwood Av
e" "Kedzie Ave & Milwaukee Ave" "Clarendon Ave & Leland Ave" ...
## $ start_station_id : num [1:199926] 36 340 260 251 261 206 261 180 331 219 ...
## $ end_station_name : chr [1:199926] "Wabash Ave & Grand Ave" "Clark St & Leland Ave" "Kedzi
e Ave & Milwaukee Ave" "Lake Shore Dr & Wellington Ave" ...
## $ end_station_id   : num [1:199926] 199 326 260 157 206 22 261 180 300 305 ...
## $ start_lat        : num [1:199926] 41.9 41.9 41.9 42 41.9 ...
## $ start_lng        : num [1:199926] -87.6 -87.6 -87.7 -87.7 -87.7 ...
## $ end_lat          : num [1:199926] 41.9 42 41.9 41.9 41.8 ...
## $ end_lng          : num [1:199926] -87.6 -87.7 -87.7 -87.6 -87.6 ...
## $ member_casual    : chr [1:199926] "member" "casual" "casual" "casual" ...
## - attr(*, "spec")=
## .. cols(
## ..   ride_id = col_character(),
## ..   rideable_type = col_character(),
## ..   started_at = col_character(),
## ..   ended_at = col_character(),
## ..   day_of_week = col_character(),
## ..   ride_length = col_time(format = ""),
## ..   start_station_name = col_character(),
## ..   start_station_id = col_double(),
## ..   end_station_name = col_character(),
## ..   end_station_id = col_double(),
## ..   start_lat = col_double(),
## ..   start_lng = col_double(),
## ..   end_lat = col_double(),
## ..   end_lng = col_double(),
## ..   member_casual = col_character()
## .. )

```

```
str(data_jun20)
```

```

## spec_tbl_df[,15] [342,069 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ ride_id          : chr [1:342069] "8CD5DE2C2B6C4CFC" "9A191EB2C751D85D" "F37D14B0B5659BC
F" "C41237B506E85FA1" ...
## $ rideable_type    : chr [1:342069] "docked_bike" "docked_bike" "docked_bike" "docked_bike"
...
## $ started_at       : chr [1:342069] "13-06-2020 23:24" "26-06-2020 07:26" "23-06-2020 17:1
2" "20-06-2020 01:09" ...
## $ ended_at         : chr [1:342069] "13-06-2020 23:36" "26-06-2020 07:31" "23-06-2020 17:2
1" "20-06-2020 01:28" ...
## $ day_of_week      : chr [1:342069] "Saturday" "Friday" "Tuesday" "Saturday" ...
## $ ride_length      : 'hms' num [1:342069] 00:12:00 00:05:00 00:08:00 00:18:00 ...
## ... attr(*, "units")= chr "secs"
## $ start_station_name: chr [1:342069] "Wilton Ave & Belmont Ave" "Federal St & Polk St" "Dale
y Center Plaza" "Broadway & Cornelia Ave" ...
## $ start_station_id : num [1:342069] 117 41 81 303 327 327 41 115 338 84 ...
## $ end_station_name : chr [1:342069] "Damen Ave & Clybourn Ave" "Daley Center Plaza" "State
St & Harrison St" "Broadway & Berwyn Ave" ...
## $ end_station_id   : num [1:342069] 163 81 5 294 117 117 81 303 164 53 ...
## $ start_lat        : num [1:342069] 41.9 41.9 41.9 41.9 41.9 ...
## $ start_lng        : num [1:342069] -87.7 -87.6 -87.6 -87.6 -87.7 ...
## $ end_lat          : num [1:342069] 41.9 41.9 41.9 42 41.9 ...
## $ end_lng          : num [1:342069] -87.7 -87.6 -87.6 -87.7 -87.7 ...
## $ member_casual    : chr [1:342069] "casual" "member" "member" "casual" ...
## - attr(*, "spec")=
## .. cols(
## ..   ride_id = col_character(),
## ..   rideable_type = col_character(),
## ..   started_at = col_character(),
## ..   ended_at = col_character(),
## ..   day_of_week = col_character(),
## ..   ride_length = col_time(format = ""),
## ..   start_station_name = col_character(),
## ..   start_station_id = col_double(),
## ..   end_station_name = col_character(),
## ..   end_station_id = col_double(),
## ..   start_lat = col_double(),
## ..   start_lng = col_double(),
## ..   end_lat = col_double(),
## ..   end_lng = col_double(),
## ..   member_casual = col_character()
## .. )

```

```
str(data_jul20)
```

```

## spec_tbl_df[,15] [550,044 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ ride_id          : chr [1:550044] "762198876D69004D" "BEC9C9FBA0D4CF1B" "D2FD8EA432C77EC
1" "54AE594E20B35881" ...
## $ rideable_type    : chr [1:550044] "docked_bike" "docked_bike" "docked_bike" "docked_bike"
...
## $ started_at       : chr [1:550044] "09-07-2020 15:22" "24-07-2020 23:56" "08-07-2020 19:4
9" "17-07-2020 19:06" ...
## $ ended_at         : chr [1:550044] "09-07-2020 15:25" "25-07-2020 00:20" "08-07-2020 19:5
6" "17-07-2020 19:27" ...
## $ day_of_week      : chr [1:550044] "Thursday" "Friday" "Wednesday" "Friday" ...
## $ ride_length      : 'hms' num [1:550044] 00:03:00 00:24:00 00:07:00 00:21:00 ...
## ... attr(*, "units")= chr "secs"
## $ start_station_name: chr [1:550044] "Ritchie Ct & Banks St" "Halsted St & Roscoe St" "Lake
Shore Dr & Diversey Pkwy" "LaSalle St & Illinois St" ...
## $ start_station_id : num [1:550044] 180 299 329 181 268 635 113 211 176 31 ...
## $ end_station_name : chr [1:550044] "Wells St & Evergreen Ave" "Broadway & Ridge Ave" "Clar
k St & Wellington Ave" "Clark St & Armitage Ave" ...
## $ end_station_id   : num [1:550044] 291 461 156 94 301 289 140 31 191 142 ...
## $ start_lat        : num [1:550044] 41.9 41.9 41.9 41.9 41.9 ...
## $ start_lng        : num [1:550044] -87.6 -87.6 -87.6 -87.6 -87.6 ...
## $ end_lat          : num [1:550044] 41.9 42 41.9 41.9 41.9 ...
## $ end_lng          : num [1:550044] -87.6 -87.7 -87.6 -87.6 -87.6 ...
## $ member_casual    : chr [1:550044] "member" "member" "casual" "casual" ...
## - attr(*, "spec")=
## .. cols(
## ..   ride_id = col_character(),
## ..   rideable_type = col_character(),
## ..   started_at = col_character(),
## ..   ended_at = col_character(),
## ..   day_of_week = col_character(),
## ..   ride_length = col_time(format = ""),
## ..   start_station_name = col_character(),
## ..   start_station_id = col_double(),
## ..   end_station_name = col_character(),
## ..   end_station_id = col_double(),
## ..   start_lat = col_double(),
## ..   start_lng = col_double(),
## ..   end_lat = col_double(),
## ..   end_lng = col_double(),
## ..   member_casual = col_character()
## .. )

```

```
str(data_aug20)
```

```

## spec_tbl_df[,15] [605,746 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ ride_id          : chr [1:605746] "322BD23D287743ED" "2A3AEF1AB9054D8B" "67DC1D133E8B581
6" "C79FBBD412E578A7" ...
## $ rideable_type    : chr [1:605746] "docked_bike" "electric_bike" "electric_bike" "electric
_bike" ...
## $ started_at       : chr [1:605746] "20-08-2020 18:08" "27-08-2020 18:46" "26-08-2020 19:4
4" "27-08-2020 12:05" ...
## $ ended_at         : chr [1:605746] "20-08-2020 18:17" "27-08-2020 19:54" "26-08-2020 21:5
3" "27-08-2020 12:53" ...
## $ day_of_week      : chr [1:605746] "Thursday" "Thursday" "Wednesday" "Thursday" ...
## $ ride_length      : 'hms' num [1:605746] 00:09:00 01:08:00 02:08:00 00:48:00 ...
## ... attr(*, "units")= chr "secs"
## $ start_station_name: chr [1:605746] "Lake Shore Dr & Diversey Pkwy" "Michigan Ave & 14th S
t" "Columbus Dr & Randolph St" "Daley Center Plaza" ...
## $ start_station_id  : num [1:605746] 329 168 195 81 658 658 196 67 153 177 ...
## $ end_station_name : chr [1:605746] "Clark St & Lincoln Ave" "Michigan Ave & 14th St" "Stat
e St & Randolph St" "State St & Kinzie St" ...
## $ end_station_id   : num [1:605746] 141 168 44 47 658 658 49 229 225 305 ...
## $ start_lat        : num [1:605746] 41.9 41.9 41.9 41.9 41.9 41.9 ...
## $ start_lng        : num [1:605746] -87.6 -87.6 -87.6 -87.6 -87.7 ...
## $ end_lat          : num [1:605746] 41.9 41.9 41.9 41.9 41.9 ...
## $ end_lng          : num [1:605746] -87.6 -87.6 -87.6 -87.6 -87.7 ...
## $ member_casual    : chr [1:605746] "member" "casual" "casual" "casual" ...
## - attr(*, "spec")=
## .. cols(
## ..   ride_id = col_character(),
## ..   rideable_type = col_character(),
## ..   started_at = col_character(),
## ..   ended_at = col_character(),
## ..   day_of_week = col_character(),
## ..   ride_length = col_time(format = ""),
## ..   start_station_name = col_character(),
## ..   start_station_id = col_double(),
## ..   end_station_name = col_character(),
## ..   end_station_id = col_double(),
## ..   start_lat = col_double(),
## ..   start_lng = col_double(),
## ..   end_lat = col_double(),
## ..   end_lng = col_double(),
## ..   member_casual = col_character()
## .. )

```

```
str(data_sep20)
```

```

## spec_tbl_df[,15] [498,263 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ ride_id          : chr [1:498263] "2B22BD5F95FB2629" "57F6DC9A153DB98C" "B9C4712F78C1AE6
8" "0CB5E2496B6F1DF8" ...
## $ rideable_type    : chr [1:498263] "electric_bike" "electric_bike" "electric_bike" "electr
ic_bike" ...
## $ started_at       : chr [1:498263] "17-09-2020 14:27" "17-09-2020 18:10" "17-09-2020 15:1
6" "16-09-2020 21:39" ...
## $ ended_at         : chr [1:498263] "17-09-2020 14:44" "17-09-2020 18:35" "17-09-2020 15:5
2" "16-09-2020 21:53" ...
## $ day_of_week      : chr [1:498263] "Thursday" "Thursday" "Thursday" "Wednesday" ...
## $ ride_length      : 'hms' num [1:498263] 00:17:13 00:25:03 00:36:42 00:13:47 ...
## ... attr(*, "units")= chr "secs"
## $ start_station_name: chr [1:498263] "Michigan Ave & Lake St" "Ashland Ave & Belle Plaine Av
e" "Fairbanks Ct & Grand Ave" "Wells St & Evergreen Ave" ...
## $ start_station_id  : num [1:498263] 52 246 24 291 273 108 511 329 315 138 ...
## $ end_station_name : chr [1:498263] "Green St & Randolph St" "Montrose Harbor" "Fairbanks C
t & Grand Ave" "Broadway & Sheridan Rd" ...
## $ end_station_id   : num [1:498263] 112 249 24 256 273 339 511 268 464 504 ...
## $ start_lat        : num [1:498263] 41.9 42 41.9 41.9 41.9 ...
## $ start_lng        : num [1:498263] -87.6 -87.7 -87.6 -87.6 -87.6 ...
## $ end_lat          : num [1:498263] 41.9 42 41.9 42 41.9 ...
## $ end_lng          : num [1:498263] -87.6 -87.6 -87.6 -87.7 -87.6 ...
## $ member_casual    : chr [1:498263] "casual" "casual" "casual" "casual" ...
## - attr(*, "spec")=
## .. cols(
## ..   ride_id = col_character(),
## ..   rideable_type = col_character(),
## ..   started_at = col_character(),
## ..   ended_at = col_character(),
## ..   day_of_week = col_character(),
## ..   ride_length = col_time(format = ""),
## ..   start_station_name = col_character(),
## ..   start_station_id = col_double(),
## ..   end_station_name = col_character(),
## ..   end_station_id = col_double(),
## ..   start_lat = col_double(),
## ..   start_lng = col_double(),
## ..   end_lat = col_double(),
## ..   end_lng = col_double(),
## ..   member_casual = col_character()
## .. )

```

```
str(data_oct20)
```

```

## spec_tbl_df[,15] [338,298 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ ride_id          : chr [1:338298] "ACB6B40CF5B9044C" "DF450C72FD109C01" "B6396B54A15AC0D
F" "44A4AEE261B9E854" ...
## $ rideable_type    : chr [1:338298] "electric_bike" "electric_bike" "electric_bike" "electr
ic_bike" ...
## $ started_at       : chr [1:338298] "31-10-2020 19:39" "31-10-2020 23:50" "31-10-2020 23:0
0" "31-10-2020 22:16" ...
## $ ended_at         : chr [1:338298] "31-10-2020 19:57" "01-11-2020 00:04" "31-10-2020 23:0
8" "31-10-2020 22:19" ...
## $ day_of_week      : chr [1:338298] "Saturday" "Saturday" "Saturday" "Saturday" ...
## $ ride_length      : 'hms' num [1:338298] 00:18:00 00:14:00 00:08:00 00:03:00 ...
## ... attr(*, "units")= chr "secs"
## $ start_station_name: chr [1:338298] "Lakeview Ave & Fullerton Pkwy" "Southport Ave & Wavela
nd Ave" "Stony Island Ave & 67th St" "Clark St & Grace St" ...
## $ start_station_id : num [1:338298] 313 227 102 165 190 359 313 125 174 114 ...
## $ end_station_name : chr [1:338298] "Rush St & Hubbard St" "Kedzie Ave & Milwaukee Ave" "Un
iversity Ave & 57th St" "Broadway & Sheridan Rd" ...
## $ end_station_id   : num [1:338298] 125 260 423 256 185 53 125 313 635 303 ...
## $ start_lat        : num [1:338298] 41.9 41.9 41.8 42 41.9 ...
## $ start_lng        : num [1:338298] -87.6 -87.7 -87.6 -87.7 -87.7 ...
## $ end_lat          : num [1:338298] 41.9 41.9 41.8 42 41.9 ...
## $ end_lng          : num [1:338298] -87.6 -87.7 -87.6 -87.7 -87.7 ...
## $ member_casual    : chr [1:338298] "casual" "casual" "casual" "casual" ...
## - attr(*, "spec")=
## .. cols(
## ..   ride_id = col_character(),
## ..   rideable_type = col_character(),
## ..   started_at = col_character(),
## ..   ended_at = col_character(),
## ..   day_of_week = col_character(),
## ..   ride_length = col_time(format = ""),
## ..   start_station_name = col_character(),
## ..   start_station_id = col_double(),
## ..   end_station_name = col_character(),
## ..   end_station_id = col_double(),
## ..   start_lat = col_double(),
## ..   start_lng = col_double(),
## ..   end_lat = col_double(),
## ..   end_lng = col_double(),
## ..   member_casual = col_character()
## .. )

```

```
str(data_nov20)
```

```

## spec_tbl_df[,15] [221,935 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ ride_id          : chr [1:221935] "BD0A6FF6FFF9B921" "96A7A7A4BDE4F82D" "C61526D06582BDC
5" "E533E89C32080B9E" ...
## $ rideable_type    : chr [1:221935] "electric_bike" "electric_bike" "electric_bike" "electr
ic_bike" ...
## $ started_at       : chr [1:221935] "01-11-2020 13:36" "01-11-2020 10:03" "01-11-2020 00:3
4" "01-11-2020 00:45" ...
## $ ended_at         : chr [1:221935] "01-11-2020 13:45" "01-11-2020 10:14" "01-11-2020 01:0
3" "01-11-2020 00:54" ...
## $ day_of_week      : chr [1:221935] "Sunday" "Sunday" "Sunday" "Sunday" ...
## $ ride_length      : 'hms' num [1:221935] 00:09:40 00:11:19 00:29:01 00:09:15 ...
## ... attr(*, "units")= chr "secs"
## $ start_station_name: chr [1:221935] "Dearborn St & Erie St" "Franklin St & Illinois St" "La
ke Shore Dr & Monroe St" "Leavitt St & Chicago Ave" ...
## $ start_station_id  : num [1:221935] 110 672 76 659 2 72 76 58 394 623 ...
## $ end_station_name : chr [1:221935] "St. Clair St & Erie St" "Noble St & Milwaukee Ave" "Fe
deral St & Polk St" "Stave St & Armitage Ave" ...
## $ end_station_id   : num [1:221935] 211 29 41 185 2 76 72 288 273 2 ...
## $ start_lat        : num [1:221935] 41.9 41.9 41.9 41.9 41.9 ...
## $ start_lng        : num [1:221935] -87.6 -87.6 -87.6 -87.7 -87.6 ...
## $ end_lat          : num [1:221935] 41.9 41.9 41.9 41.9 41.9 ...
## $ end_lng          : num [1:221935] -87.6 -87.7 -87.6 -87.7 -87.6 ...
## $ member_casual    : chr [1:221935] "casual" "casual" "casual" "casual" ...
## - attr(*, "spec")=
## .. cols(
## ..   ride_id = col_character(),
## ..   rideable_type = col_character(),
## ..   started_at = col_character(),
## ..   ended_at = col_character(),
## ..   day_of_week = col_character(),
## ..   ride_length = col_time(format = ""),
## ..   start_station_name = col_character(),
## ..   start_station_id = col_double(),
## ..   end_station_name = col_character(),
## ..   end_station_id = col_double(),
## ..   start_lat = col_double(),
## ..   start_lng = col_double(),
## ..   end_lat = col_double(),
## ..   end_lng = col_double(),
## ..   member_casual = col_character()
## .. )

```

```
str(data_dec20)
```

```

## spec_tbl_df[,15] [113,541 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ ride_id          : chr [1:113541] "70B6A9A437D4C30D" "15F369FDAED4E8E3" "0CFD61DFE00E604
3" "244CB936487039B7" ...
## $ rideable_type    : chr [1:113541] "classic_bike" "electric_bike" "electric_bike" "docked_
bike" ...
## $ started_at       : chr [1:113541] "27-12-2020 12:44" "18-12-2020 13:53" "28-12-2020 17:1
0" "10-12-2020 13:36" ...
## $ ended_at         : chr [1:113541] "27-12-2020 12:55" "18-12-2020 14:01" "28-12-2020 17:1
2" "10-12-2020 14:37" ...
## $ day_of_week      : chr [1:113541] "Sunday" "Friday" "Monday" "Thursday" ...
## $ ride_length      : 'hms' num [1:113541] 00:10:37 00:07:50 00:01:48 01:00:47 ...
## ... attr(*, "units")= chr "secs"
## $ start_station_name: chr [1:113541] "Aberdeen St & Jackson Blvd" "Larrabee St & Armitage Av
e" "Kingsbury St & Kinzie St" "Clark St & Leland Ave" ...
## $ start_station_id : chr [1:113541] "13157" "TA1309000006" "KA1503000043" "TA1309000014"
...
## $ end_station_name : chr [1:113541] "Desplaines St & Kinzie St" "Wells St & Walton St" "Des
plaines St & Kinzie St" "Clark St & Leland Ave" ...
## $ end_station_id   : chr [1:113541] "TA1306000003" "TA1306000011" "TA1306000003" "TA1309000
014" ...
## $ start_lat        : num [1:113541] 41.9 41.9 41.9 42 41.9 ...
## $ start_lng        : num [1:113541] -87.7 -87.6 -87.6 -87.7 -87.6 ...
## $ end_lat          : num [1:113541] 41.9 41.9 41.9 42 41.9 ...
## $ end_lng          : num [1:113541] -87.6 -87.6 -87.6 -87.7 -87.6 ...
## $ member_casual    : chr [1:113541] "member" "member" "member" "casual" ...
## - attr(*, "spec")=
## .. cols(
## ..   ride_id = col_character(),
## ..   rideable_type = col_character(),
## ..   started_at = col_character(),
## ..   ended_at = col_character(),
## ..   day_of_week = col_character(),
## ..   ride_length = col_time(format = ""),
## ..   start_station_name = col_character(),
## ..   start_station_id = col_character(),
## ..   end_station_name = col_character(),
## ..   end_station_id = col_character(),
## ..   start_lat = col_double(),
## ..   start_lng = col_double(),
## ..   end_lat = col_double(),
## ..   end_lng = col_double(),
## ..   member_casual = col_character()
## .. )

```

```
str(data_jan21)
```

```

## spec_tbl_df[,15] [83,509 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ ride_id           : chr [1:83509] "B9F73448DFBE0D45" "457C7F4B5D3DA135" "57C750326F9FDABE"
## "4D518C65E338D070" ...
## $ rideable_type     : chr [1:83509] "classic_bike" "electric_bike" "electric_bike" "electric_bike" ...
## $ started_at        : chr [1:83509] "24-01-2021 19:15" "23-01-2021 12:57" "09-01-2021 15:28"
## "09-01-2021 15:28" ...
## $ ended_at          : chr [1:83509] "24-01-2021 19:22" "23-01-2021 13:02" "09-01-2021 15:37"
## "09-01-2021 15:37" ...
## $ day_of_week       : chr [1:83509] "Sunday" "Saturday" "Saturday" "Saturday" ...
## $ ride_length       : 'hms' num [1:83509] 00:07:13 00:04:32 00:09:47 00:08:57 ...
## ... attr(*, "units")= chr "secs"
## $ start_station_name: chr [1:83509] "California Ave & Cortez St" ...
## $ start_station_id  : chr [1:83509] "17660" "17660" "17660" "17660" ...
## $ end_station_name  : chr [1:83509] "Wood St & Augusta Blvd" "California Ave & North Ave" "Wood St & Augusta Blvd" "Wood St & Augusta Blvd" ...
## $ end_station_id    : chr [1:83509] "657" "13258" "657" "657" ...
## $ start_lat         : num [1:83509] 41.9 41.9 41.9 41.9 41.9 ...
## $ start_lng         : num [1:83509] -87.7 -87.7 -87.7 -87.7 -87.7 ...
## $ end_lat           : num [1:83509] 41.9 41.9 41.9 41.9 41.9 ...
## $ end_lng           : num [1:83509] -87.7 -87.7 -87.7 -87.7 -87.7 ...
## $ member_casual     : chr [1:83509] "member" "member" "casual" "casual" ...
## - attr(*, "spec")=
## .. cols(
## ..   ride_id = col_character(),
## ..   rideable_type = col_character(),
## ..   started_at = col_character(),
## ..   ended_at = col_character(),
## ..   day_of_week = col_character(),
## ..   ride_length = col_time(format = ""),
## ..   start_station_name = col_character(),
## ..   start_station_id = col_character(),
## ..   end_station_name = col_character(),
## ..   end_station_id = col_character(),
## ..   start_lat = col_double(),
## ..   start_lng = col_double(),
## ..   end_lat = col_double(),
## ..   end_lng = col_double(),
## ..   member_casual = col_character()
## .. )

```

```
str(data_feb21)
```

```

## spec_tbl_df[,15] [42,996 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ ride_id           : chr [1:42996] "89E7AA6C29227EFF" "0FEFDE2603568365" "E6159D746B2DBB91"
## "B32D3199F1C2E75B" ...
## $ rideable_type     : chr [1:42996] "classic_bike" "classic_bike" "electric_bike" "classic_b
ike" ...
## $ started_at        : chr [1:42996] "12-02-2021 16:14" "14-02-2021 17:52" "09-02-2021 19:10"
## "02-02-2021 17:49" ...
## $ ended_at          : chr [1:42996] "12-02-2021 16:21" "14-02-2021 18:12" "09-02-2021 19:19"
## "02-02-2021 17:54" ...
## $ day_of_week       : chr [1:42996] "Friday" "Sunday" "Tuesday" "Tuesday" ...
## $ ride_length        : 'hms' num [1:42996] 00:06:47 00:19:31 00:08:52 00:04:25 ...
## ... attr(*, "units")= chr "secs"
## $ start_station_name: chr [1:42996] "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave" "C
lark St & Lake St" "Wood St & Chicago Ave" ...
## $ start_station_id   : chr [1:42996] "525" "525" "KA1503000012" "637" ...
## $ end_station_name   : chr [1:42996] "Sheridan Rd & Columbia Ave" "Bosworth Ave & Howard St"
## "State St & Randolph St" "Honore St & Division St" ...
## $ end_station_id     : chr [1:42996] "660" "16806" "TA1305000029" "TA1305000034" ...
## $ start_lat          : num [1:42996] 42 42 41.9 41.9 41.8 ...
## $ start_lng          : num [1:42996] -87.7 -87.7 -87.6 -87.7 -87.6 ...
## $ end_lat            : num [1:42996] 42 42 41.9 41.9 41.8 ...
## $ end_lng            : num [1:42996] -87.7 -87.7 -87.6 -87.7 -87.6 ...
## $ member_casual      : chr [1:42996] "member" "casual" "member" "member" ...
## - attr(*, "spec")=
## .. cols(
## ..   ride_id = col_character(),
## ..   rideable_type = col_character(),
## ..   started_at = col_character(),
## ..   ended_at = col_character(),
## ..   day_of_week = col_character(),
## ..   ride_length = col_time(format = ""),
## ..   start_station_name = col_character(),
## ..   start_station_id = col_character(),
## ..   end_station_name = col_character(),
## ..   end_station_id = col_character(),
## ..   start_lat = col_double(),
## ..   start_lng = col_double(),
## ..   end_lat = col_double(),
## ..   end_lng = col_double(),
## ..   member_casual = col_character()
## .. )

```

```
str(data_mar21)
```

```

## spec_tbl_df[,15] [205,689 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ ride_id          : chr [1:205689] "CFA86D4455AA1030" "30D9DC61227D1AF3" "846D87A15682A28
4" "994D05AA75A168F2" ...
## $ rideable_type    : chr [1:205689] "classic_bike" "classic_bike" "classic_bike" "classic_b
ike" ...
## $ started_at       : chr [1:205689] "16-03-2021 08:32" "28-03-2021 01:26" "11-03-2021 21:1
7" "11-03-2021 13:26" ...
## $ ended_at         : chr [1:205689] "16-03-2021 08:36" "28-03-2021 01:36" "11-03-2021 21:3
3" "11-03-2021 13:55" ...
## $ day_of_week      : chr [1:205689] "Tuesday" "Sunday" "Thursday" "Thursday" ...
## $ ride_length      : 'hms' num [1:205689] 00:04:04 00:10:27 00:16:24 00:28:59 ...
## ... attr(*, "units")= chr "secs"
## $ start_station_name: chr [1:205689] "Humboldt Blvd & Armitage Ave" "Humboldt Blvd & Armitag
e Ave" "Shields Ave & 28th Pl" "Winthrop Ave & Lawrence Ave" ...
## $ start_station_id : chr [1:205689] "15651" "15651" "15443" "TA1308000021" ...
## $ end_station_name : chr [1:205689] "Stave St & Armitage Ave" "Central Park Ave & Bloomingd
ale Ave" "Halsted St & 35th St" "Broadway & Sheridan Rd" ...
## $ end_station_id   : chr [1:205689] "13266" "18017" "TA1308000043" "13323" ...
## $ start_lat        : num [1:205689] 41.9 41.9 41.8 42 42 ...
## $ start_lng        : num [1:205689] -87.7 -87.7 -87.6 -87.7 -87.7 ...
## $ end_lat          : num [1:205689] 41.9 41.9 41.8 42 42.1 ...
## $ end_lng          : num [1:205689] -87.7 -87.7 -87.6 -87.6 -87.7 ...
## $ member_casual    : chr [1:205689] "casual" "casual" "casual" "casual" ...
## - attr(*, "spec")=
## .. cols(
## ..   ride_id = col_character(),
## ..   rideable_type = col_character(),
## ..   started_at = col_character(),
## ..   ended_at = col_character(),
## ..   day_of_week = col_character(),
## ..   ride_length = col_time(format = ""),
## ..   start_station_name = col_character(),
## ..   start_station_id = col_character(),
## ..   end_station_name = col_character(),
## ..   end_station_id = col_character(),
## ..   start_lat = col_double(),
## ..   start_lng = col_double(),
## ..   end_lat = col_double(),
## ..   end_lng = col_double(),
## ..   member_casual = col_character()
## .. )

```

```
str(data_apr21)
```

```

## spec_tbl_df[,15] [298,203 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ ride_id          : chr [1:298203] "6C992BD37A98A63F" "1E0145613A209000" "E498E15508A80BA
D" "1887262AD101C604" ...
## $ rideable_type    : chr [1:298203] "classic_bike" "docked_bike" "docked_bike" "classic_bik
e" ...
## $ started_at       : chr [1:298203] "12-04-2021 18:25" "27-04-2021 17:27" "03-04-2021 12:4
2" "17-04-2021 09:17" ...
## $ ended_at         : chr [1:298203] "12-04-2021 18:56" "27-04-2021 18:31" "07-04-2021 11:4
0" "17-04-2021 09:42" ...
## $ day_of_week      : chr [1:298203] "Monday" "Tuesday" "Saturday" "Saturday" ...
## $ ride_length      : 'hms' num [1:298203] 00:31:19 01:04:18 22:57:39 00:25:06 ...
## ... attr(*, "units")= chr "secs"
## $ start_station_name: chr [1:298203] "State St & Pearson St" "Dorchester Ave & 49th St" "Loo
mis Blvd & 84th St" "Honore St & Division St" ...
## $ start_station_id : chr [1:298203] "TA1307000061" "KA1503000069" "20121" "TA1305000034"
...
## $ end_station_name : chr [1:298203] "Southport Ave & Waveland Ave" "Dorchester Ave & 49th S
t" "Loomis Blvd & 84th St" "Southport Ave & Waveland Ave" ...
## $ end_station_id   : chr [1:298203] "13235" "KA1503000069" "20121" "13235" ...
## $ start_lat        : num [1:298203] 41.9 41.8 41.7 41.9 41.7 ...
## $ start_lng        : num [1:298203] -87.6 -87.6 -87.7 -87.7 -87.7 ...
## $ end_lat          : num [1:298203] 41.9 41.8 41.7 41.9 41.7 ...
## $ end_lng          : num [1:298203] -87.7 -87.6 -87.7 -87.7 -87.7 ...
## $ member_casual    : chr [1:298203] "member" "casual" "casual" "member" ...
## - attr(*, "spec")=
## .. cols(
## ..   ride_id = col_character(),
## ..   rideable_type = col_character(),
## ..   started_at = col_character(),
## ..   ended_at = col_character(),
## ..   day_of_week = col_character(),
## ..   ride_length = col_time(format = ""),
## ..   start_station_name = col_character(),
## ..   start_station_id = col_character(),
## ..   end_station_name = col_character(),
## ..   end_station_id = col_character(),
## ..   start_lat = col_double(),
## ..   start_lng = col_double(),
## ..   end_lat = col_double(),
## ..   end_lng = col_double(),
## ..   member_casual = col_character()
## .. )

```

We can see in column specifications that start_station_id and end_station_id does not have consistent datatype.
So we will change it to character datatype

```

data_may20 <- mutate(data_may20, start_station_id = as.character(start_station_id)
                     ,end_station_id = as.character(end_station_id))
data_jun20 <- mutate(data_jun20, start_station_id = as.character(start_station_id)
                     ,end_station_id = as.character(end_station_id))
data_jul20 <- mutate(data_jul20, start_station_id = as.character(start_station_id)
                     ,end_station_id = as.character(end_station_id))
data_aug20 <- mutate(data_aug20, start_station_id = as.character(start_station_id)
                     ,end_station_id = as.character(end_station_id))
data_sep20 <- mutate(data_sep20, start_station_id = as.character(start_station_id)
                     ,end_station_id = as.character(end_station_id))
data_oct20 <- mutate(data_oct20, start_station_id = as.character(start_station_id)
                     ,end_station_id = as.character(end_station_id))
data_nov20 <- mutate(data_nov20, start_station_id = as.character(start_station_id)
                     ,end_station_id = as.character(end_station_id))
data_dec20 <- mutate(data_dec20, start_station_id = as.character(start_station_id)
                     ,end_station_id = as.character(end_station_id))
data_jan21 <- mutate(data_jan21, start_station_id = as.character(start_station_id)
                     ,end_station_id = as.character(end_station_id))
data_feb21 <- mutate(data_feb21, start_station_id = as.character(start_station_id)
                     ,end_station_id = as.character(end_station_id))
data_mar21 <- mutate(data_mar21, start_station_id = as.character(start_station_id)
                     ,end_station_id = as.character(end_station_id))
data_apr21 <- mutate(data_apr21, start_station_id = as.character(start_station_id)
                     ,end_station_id = as.character(end_station_id))

```

Combining all datasets into a single dataset

```

trips <- bind_rows(data_may20,data_jun20,data_jul20,data_aug20,data_sep20,data_oct20,data_nov20,
data_dec20,data_jan21,data_feb21,data_mar21,data_apr21)

```

Longitude and latitude columns can be dropped as there is no need of them

```

trips <- trips %>%
  select(-c(start_lat,start_lng,end_lat,end_lng))

```

Formatting data for our convenience

```

trips$date <- as.Date(trips$started_at, "%d-%m-%Y")

trips$month <- format(as.Date(trips$date), "%m")

trips$day <- format(as.Date(trips$date), "%d")

trips$year <- format(as.Date(trips$date), "%Y")

```

Summarizing data for analysis

```

colnames(trips)

```

```
## [1] "ride_id"           "rideable_type"      "started_at"  
## [4] "ended_at"          "day_of_week"        "ride_length"  
## [7] "start_station_name" "start_station_id"    "end_station_name"  
## [10] "end_station_id"     "member_casual"      "date"  
## [13] "month"              "day"                "year"
```

```
nrow(trips)
```

```
## [1] 3500219
```

```
dim(trips)
```

```
## [1] 3500219      15
```

```
head(trips)
```

```
## # A tibble: 6 x 15  
##   ride_id   rideable_type started_at   ended_at day_of_week ride_length  
##   <chr>     <chr>       <chr>       <chr>      <chr>       <time>  
## 1 02668AD35674~ docked_bike 27-05-2020 1~ 27-05-2020 ~ Wednesday 00:13  
## 2 7A50CCAF1EDD~ docked_bike 25-05-2020 1~ 25-05-2020 ~ Monday   00:18  
## 3 2FFCDFDB91FE~ docked_bike 02-05-2020 1~ 02-05-2020 ~ Saturday 01:37  
## 4 58991CF1DB75~ docked_bike 02-05-2020 1~ 02-05-2020 ~ Saturday 00:14  
## 5 A79651EFECC2~ docked_bike 29-05-2020 1~ 29-05-2020 ~ Friday   00:38  
## 6 1466C5B39F68~ docked_bike 29-05-2020 1~ 29-05-2020 ~ Friday   00:47  
## # ... with 9 more variables: start_station_name <chr>, start_station_id <chr>,  
## #   end_station_name <chr>, end_station_id <chr>, member_casual <chr>,  
## #   date <date>, month <chr>, day <chr>, year <chr>
```

```
str(trips)
```

```
## # tibble[,15] [3,500,219 x 15] (S3: tbl_df/tbl/data.frame)
## $ ride_id          : chr [1:3500219] "02668AD35674B983" "7A50CCAF1EDDB28F" "2FFCDFDB91FE9A5
2" "58991CF1DB75BA84" ...
## $ rideable_type    : chr [1:3500219] "docked_bike" "docked_bike" "docked_bike" "docked_bik
e" ...
## $ started_at       : chr [1:3500219] "27-05-2020 10:03" "25-05-2020 10:47" "02-05-2020 14:1
1" "02-05-2020 16:25" ...
## $ ended_at         : chr [1:3500219] "27-05-2020 10:16" "25-05-2020 11:05" "02-05-2020 15:4
8" "02-05-2020 16:39" ...
## $ day_of_week      : chr [1:3500219] "Wednesday" "Monday" "Saturday" "Saturday" ...
## $ ride_length      : 'hms' num [1:3500219] 00:13:00 00:18:00 01:37:00 00:14:00 ...
## ... attr(*, "units")= chr "secs"
## $ start_station_name: chr [1:3500219] "Franklin St & Jackson Blvd" "Clark St & Wrightwood Av
e" "Kedzie Ave & Milwaukee Ave" "Clarendon Ave & Leland Ave" ...
## $ start_station_id : chr [1:3500219] "36" "340" "260" "251" ...
## $ end_station_name : chr [1:3500219] "Wabash Ave & Grand Ave" "Clark St & Leland Ave" "Kedz
ie Ave & Milwaukee Ave" "Lake Shore Dr & Wellington Ave" ...
## $ end_station_id   : chr [1:3500219] "199" "326" "260" "157" ...
## $ member_casual    : chr [1:3500219] "member" "casual" "casual" "casual" ...
## $ date             : Date[1:3500219], format: "2020-05-27" "2020-05-25" ...
## $ month            : chr [1:3500219] "05" "05" "05" "05" ...
## $ day              : chr [1:3500219] "27" "25" "02" "02" ...
## $ year             : chr [1:3500219] "2020" "2020" "2020" "2020" ...
```

```
summary(trips)
```

```

##   ride_id      rideable_type      started_at      ended_at
## Length:3500219  Length:3500219  Length:3500219  Length:3500219
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##
##   day_of_week      ride_length      start_station_name start_station_id
## Length:3500219  Length:3500219  Length:3500219  Length:3500219
## Class :character Class1:hms      Class :character Class :character
## Mode  :character Class2:difftime Mode  :character Mode  :character
##                           Mode  :numeric
##
##
##   end_station_name      end_station_id      member_casual      date
## Length:3500219  Length:3500219  Length:3500219  Min.   :2020-05-01
## Class :character Class :character Class :character 1st Qu.:2020-07-20
## Mode  :character Mode  :character Mode  :character Median :2020-09-04
##                           Mean    :2020-09-27
##                           3rd Qu.:2020-11-08
##                           Max.   :2021-04-30
##
##   month          day          year
## Length:3500219  Length:3500219  Length:3500219
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##
##

```

The ride_length cannot be summarized as it is in HH:MM:SS format. so let's convert it to numeric representing seconds only

```

trips$ride_length <- hms(trips$ride_length)
trips$ride_length <- as.numeric(trips$ride_length)

```

```

(trips <- rename(trips,
                 ride_length_in_sec = ride_length))

```

```

## # A tibble: 3,500,219 x 15
##   ride_id    rideable_type started_at    ended_at day_of_week ride_length_in_~
##   <chr>      <chr>        <chr>       <chr>      <chr>           <dbl>
## 1 02668AD35~ docked_bike 27-05-2020 ~ 27-05-202~ Wednesday          780
## 2 7A50CCAF1~ docked_bike 25-05-2020 ~ 25-05-202~ Monday           1080
## 3 2FFCDFDB9~ docked_bike 02-05-2020 ~ 02-05-202~ Saturday          5820
## 4 58991CF1D~ docked_bike 02-05-2020 ~ 02-05-202~ Saturday          840
## 5 A79651EFE~ docked_bike 29-05-2020 ~ 29-05-202~ Friday           2280
## 6 1466C5B39~ docked_bike 29-05-2020 ~ 29-05-202~ Friday           2820
## 7 2500D7957~ docked_bike 20-05-2020 ~ 20-05-202~ Wednesday         3300
## 8 ED42D3E06~ docked_bike 06-05-2020 ~ 06-05-202~ Wednesday         2760
## 9 23AFBD962~ docked_bike 30-05-2020 ~ 30-05-202~ Saturday          1140
## 10 52C0D13F6~ docked_bike 23-05-2020 ~ 23-05-202~ Saturday         1800
## # ... with 3,500,209 more rows, and 9 more variables: start_station_name <chr>,
## #   start_station_id <chr>, end_station_name <chr>, end_station_id <chr>,
## #   member_casual <chr>, date <date>, month <chr>, day <chr>, year <chr>

```

Analyse Phase

```
mean(trips$ride_length_in_sec)
```

```
## [1] 1452.345
```

```
median(trips$ride_length_in_sec)
```

```
## [1] 857
```

```
max(trips$ride_length_in_sec)
```

```
## [1] 86397
```

```
min(trips$ride_length_in_sec)
```

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

```
aggregate(trips$ride_length_in_sec ~ trips$member_casual, FUN = mean)
```

```

##   trips$member_casual trips$ride_length_in_sec
## 1                 casual             2219.4635
## 2                 member            912.6057

```

```
aggregate(trips$ride_length_in_sec ~ trips$member_casual, FUN = median)
```

```
##   trips$member_casual trips$ride_length_in_sec
## 1           casual          1260
## 2         member           664
```

```
aggregate(trips$ride_length_in_sec ~ trips$member_casual, FUN = max)
```

```
##   trips$member_casual trips$ride_length_in_sec
## 1           casual        86397
## 2         member        86160
```

```
aggregate(trips$ride_length_in_sec ~ trips$member_casual, FUN = min)
```

```
##   trips$member_casual trips$ride_length_in_sec
## 1           casual            0
## 2         member            0
```

```
aggregate(trips$ride_length_in_sec ~ trips$member_casual + trips$day_of_week, FUN = mean)
```

```
##   trips$member_casual trips$day_of_week trips$ride_length_in_sec
## 1           casual      Friday       2099.0218
## 2         member      Friday        896.4971
## 3           casual     Monday      2210.4806
## 4         member     Monday        866.3278
## 5           casual    Saturday     2373.1811
## 6         member    Saturday     1016.1211
## 7           casual     Sunday     2475.6447
## 8         member     Sunday      1021.4206
## 9           casual   Thursday     2009.2141
## 10        member   Thursday      865.4825
## 11           casual   Tuesday     2064.0539
## 12        member   Tuesday      861.1518
## 13           casual Wednesday     1993.8069
## 14        member Wednesday      862.7429
```

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

```
## `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 weekday number_of_rides average_duration
##   <chr>        <ord>     <int>            <dbl>
## 1 casual        Sun      211113            2240.
## 2 casual        Mon      211701            2238.
## 3 casual        Tue      218726            2279.
## 4 casual        Wed      200605            2202.
## 5 casual        Thu      203009            2189.
## 6 casual        Fri      203160            2205.
## 7 casual        Sat      197295            2175.
## 8 member        Sun      296640             913.
## 9 member        Mon      298088            924.
## 10 member       Tue      294955            931.
## 11 member       Wed      287474            905.
## 12 member       Thu      293157            906.
## 13 member       Fri      295582            912.
## 14 member       Sat      288714            897.

```

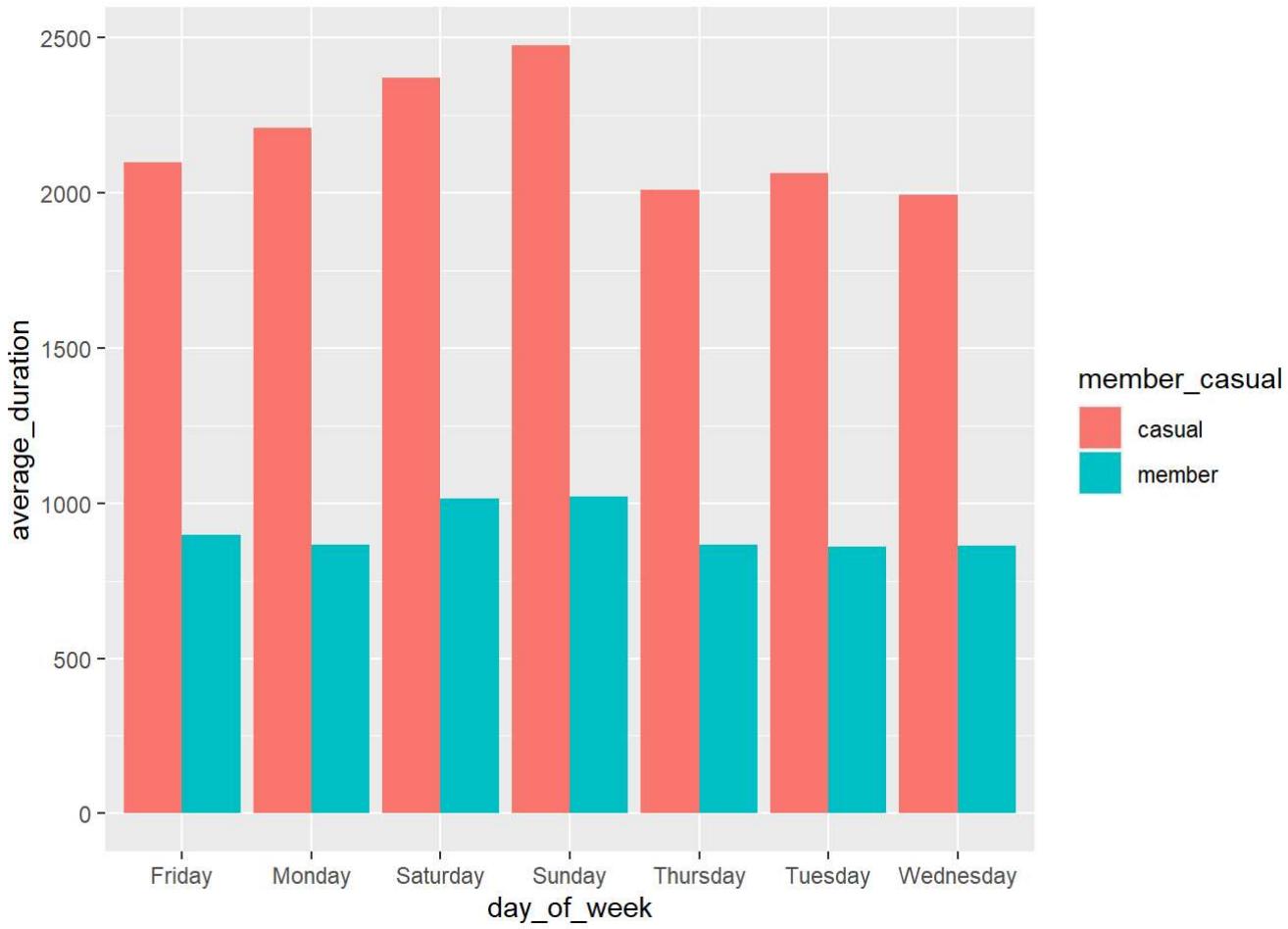
Plotting our data

```

trips %>%
  group_by(member_casual, day_of_week) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length_in_sec)) %>%
  arrange(member_casual, day_of_week) %>%
  ggplot(aes(x = day_of_week, y = average_duration, fill = member_casual)) +
  geom_col(position = "dodge")

```

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



Export data for further Analysis

```
agg_data = aggregate(trips$ride_length_in_sec ~ trips$member_casual + trips$day_of_week, FUN = mean)
# write.csv(agg_data, file = 'D:\\Main Folder\\Google Data Analytics Course\\CaseStudy\\Cyclists_Case_Study\\DataSet\\agg_data.csv')
```