Case Study 1: How does a bike-share navigate speedy success?

Introduction

This exploratory analysis case study is a capstone project required for Google Data analytics Professional Certificate. This case study involves a bike-share company's data of its customer's trip details over a period of 12 months (March 2021 - February 2022). The data has been been made available be Motivate International Inc. under this license.

Scenario:

You are a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

Stakeholders:

- Lily Moreno: Director of marketing
- Cyclistic executive team
- data analytics team

Objective:

Analyse Cyclistic historical bike trip data to understand how annual members and casual riders use Cyclistic bikes differently.

Deliverables:

- Insights on how annual members and casual riders use Cyclistic bikes differently.
- Provide effective visuals and relevant data to support the insights.
- Recommendations to convert casual riders into cyclistic members.

1. Prepare

Data Sources:

A total of 12 data sets have been made available for each month starting from March 2021 to February 2022. Each data set captures the details of every ride logged by the customers of Cyclistic. This data that has been made publicly available and has been scrubbed to omit rider's personal information.

Setting Up The Environment

Install required packages

```
library(tidyverse) #helps wrangle data
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

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

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
## Warning: package 'tibble' was built under R version 4.0.5
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
## Warning: package 'forcats' was built under R version 4.0.5
```

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

```
library(ggplot2)   #helps visualize data
library(lubridate) #helps wrangle date attributes
```

```
## Warning: package 'lubridate' was built under R version 4.0.5
```

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

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

Setting up the working directory

```
getwd()  #displays your working directory
```

```
## [1] "C:/Users/Justin/OneDrive/Documents/R Coding/Cyclistic_data"
```

```
setwd("C:/Users/Justin/OneDrive/Documents/R Coding/Cyclistic_data/") #sets
your working directory to the specified location
```

Loading Datasets

```
tripdata_202103 <- read.csv("202103-divvy-tripdata.csv")
tripdata_202104 <- read.csv("202104-divvy-tripdata.csv")
```

```
tripdata_202105 <- read.csv("202105-divvy-tripdata.csv")
tripdata_202106 <- read.csv("202106-divvy-tripdata.csv")
tripdata_202107 <- read.csv("202107-divvy-tripdata.csv")
tripdata_202108 <- read.csv("202108-divvy-tripdata.csv")
tripdata_202109 <- read.csv("202109-divvy-tripdata.csv")
tripdata_202110 <- read.csv("202110-divvy-tripdata.csv")
tripdata_202111 <- read.csv("202111-divvy-tripdata.csv")
tripdata_202112 <- read.csv("202112-divvy-tripdata.csv")
tripdata_202201 <- read.csv("202201-divvy-tripdata.csv")
tripdata_202202 <- read.csv("202202-divvy-tripdata.csv")
```

Check Column names for each data set for consistency

```
colnames(tripdata_202103)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(tripdata_202104)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(tripdata_202105)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(tripdata_202106)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(tripdata_202107)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(tripdata_202108)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(tripdata_202109)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(tripdata_202110)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(tripdata_202111)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(tripdata_202112)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(tripdata_202201)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(tripdata_202202)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
```

```
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

Check data structures and data types for all data frames

```
str(tripdata_202103)

## 'data.frame':    228496 obs. of  13 variables:
##  $ ride_id           : chr  "CFA86D4455AA1030" "30D9DC61227D1AF3"
"846D87A15682A284" "994D05AA75A168F2" ...
##  $ rideable_type     : chr  "classic_bike" "classic_bike" "classic_bike"
"classic_bike" ...
##  $ started_at        : chr  "2021-03-16 08:32:30" "2021-03-28 01:26:28"
"2021-03-11 21:17:29" "2021-03-11 13:26:42" ...
##  $ ended_at          : chr  "2021-03-16 08:36:34" "2021-03-28 01:36:55"
"2021-03-11 21:33:53" "2021-03-11 13:55:41" ...
##  $ start_station_name: chr  "Humboldt Blvd & Armitage Ave" "Humboldt Blvd
& Armitage Ave" "Shields Ave & 28th Pl" "Winthrop Ave & Lawrence Ave" ...
##  $ start_station_id  : chr  "15651" "15651" "15443" "TA1308000021" ...
##  $ end_station_name  : chr  "Stave St & Armitage Ave" "Central Park Ave &
Bloomingdale Ave" "Halsted St & 35th St" "Broadway & Sheridan Rd" ...
##  $ end_station_id    : chr  "13266" "18017" "TA1308000043" "13323" ...
##  $ start_lat         : num  41.9 41.9 41.8 42 42 ...
##  $ start_lng         : num  -87.7 -87.7 -87.6 -87.7 -87.7 ...
##  $ end_lat           : num  41.9 41.9 41.8 42 42.1 ...
##  $ end_lng           : num  -87.7 -87.7 -87.6 -87.6 -87.7 ...
##  $ member_casual     : chr  "casual" "casual" "casual" "casual" ...

str(tripdata_202104)

## 'data.frame':    337230 obs. of  13 variables:
##  $ ride_id           : chr  "6C992BD37A98A63F" "1E0145613A209000"
"E498E15508A80BAD" "1887262AD101C604" ...
##  $ rideable_type     : chr  "classic_bike" "docked_bike" "docked_bike"
"classic_bike" ...
##  $ started_at        : chr  "2021-04-12 18:25:36" "2021-04-27 17:27:11"
"2021-04-03 12:42:45" "2021-04-17 09:17:42" ...
##  $ ended_at          : chr  "2021-04-12 18:56:55" "2021-04-27 18:31:29"
"2021-04-07 11:40:24" "2021-04-17 09:42:48" ...
##  $ start_station_name: chr  "State St & Pearson St" "Dorchester Ave & 49th
St" "Loomis Blvd & 84th St" "Honore St & Division St" ...
##  $ start_station_id  : chr  "TA1307000061" "KA1503000069" "20121"
"TA1305000034" ...
##  $ end_station_name  : chr  "Southport Ave & Waveland Ave" "Dorchester Ave
& 49th St" "Loomis Blvd & 84th St" "Southport Ave & Waveland Ave" ...
##  $ end_station_id    : chr  "13235" "KA1503000069" "20121" "13235" ...
##  $ start_lat         : num  41.9 41.8 41.7 41.9 41.7 ...
##  $ start_lng         : num  -87.6 -87.6 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num  41.9 41.8 41.7 41.9 41.7 ...
##  $ end_lng           : num  -87.7 -87.6 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr  "member" "casual" "casual" "member" ...
```

```
str(tripdata_202105)

## 'data.frame':    531633 obs. of  13 variables:
##  $ ride_id           : chr  "C809ED75D6160B2A" "DD59FDCE0ACACAF3"
"0AB83CB88C43EFC2" "7881AC6D39110C60" ...
##  $ rideable_type     : chr  "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : chr  "2021-05-30 11:58:15" "2021-05-30 11:29:14"
"2021-05-30 14:24:01" "2021-05-30 14:25:51" ...
##  $ ended_at          : chr  "2021-05-30 12:10:39" "2021-05-30 12:14:09"
"2021-05-30 14:25:13" "2021-05-30 14:41:04" ...
##  $ start_station_name: chr  "" "" "" "" ...
##  $ start_station_id  : chr  "" "" "" "" ...
##  $ end_station_name  : chr  "" "" "" "" ...
##  $ end_station_id    : chr  "" "" "" "" ...
##  $ start_lat         : num  41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num  -87.6 -87.6 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num  41.9 41.8 41.9 41.9 41.9 ...
##  $ end_lng           : num  -87.6 -87.6 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr  "casual" "casual" "casual" "casual" ...

str(tripdata_202106)

## 'data.frame':    729595 obs. of  13 variables:
##  $ ride_id           : chr  "99FEC93BA843FB20" "06048DCFC8520CAF"
"9598066F68045DF2" "B03C0FE48C412214" ...
##  $ rideable_type     : chr  "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : chr  "2021-06-13 14:31:28" "2021-06-04 11:18:02"
"2021-06-04 09:49:35" "2021-06-03 19:56:05" ...
##  $ ended_at          : chr  "2021-06-13 14:34:11" "2021-06-04 11:24:19"
"2021-06-04 09:55:34" "2021-06-03 20:21:55" ...
##  $ start_station_name: chr  "" "" "" "" ...
##  $ start_station_id  : chr  "" "" "" "" ...
##  $ end_station_name  : chr  "" "" "" "" ...
##  $ end_station_id    : chr  "" "" "" "" ...
##  $ start_lat         : num  41.8 41.8 41.8 41.8 41.8 ...
##  $ start_lng         : num  -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num  41.8 41.8 41.8 41.8 41.8 ...
##  $ end_lng           : num  -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr  "member" "member" "member" "member" ...

str(tripdata_202107)

## 'data.frame':    822410 obs. of  13 variables:
##  $ ride_id           : chr  "0A1B623926EF4E16" "B2D5583A5A5E76EE"
"6F264597DDBF427A" "379B58EAB20E8AA5" ...
##  $ rideable_type     : chr  "docked_bike" "classic_bike" "classic_bike"
"classic_bike" ...
##  $ started_at        : chr  "2021-07-02 14:44:36" "2021-07-07 16:57:42"
"2021-07-25 11:30:55" "2021-07-08 22:08:30" ...
```

```
##  $ ended_at          : chr  "2021-07-02 15:19:58" "2021-07-07 17:16:09"
"2021-07-25 11:48:45" "2021-07-08 22:23:32" ...
##  $ start_station_name: chr  "Michigan Ave & Washington St" "California Ave
& Cortez St" "Wabash Ave & 16th St" "California Ave & Cortez St" ...
##  $ start_station_id  : chr  "13001" "17660" "SL-012" "17660" ...
##  $ end_station_name  : chr  "Halsted St & North Branch St" "Wood St &
Hubbard St" "Rush St & Hubbard St" "Carpenter St & Huron St" ...
##  $ end_station_id    : chr  "KA1504000117" "13432" "KA1503000044" "13196"
...
##  $ start_lat         : num  41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num  -87.6 -87.7 -87.6 -87.7 -87.7 ...
##  $ end_lat           : num  41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num  -87.6 -87.7 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr  "casual" "casual" "member" "member" ...

str(tripdata_202108)

## 'data.frame':    804352 obs. of  13 variables:
##  $ ride_id           : chr  "99103BB87CC6C1BB" "EAFCCCFB0A3FC5A1"
"9EF4F46C57AD234D" "5834D3208BFAF1DA" ...
##  $ rideable_type     : chr  "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : chr  "2021-08-10 17:15:49" "2021-08-10 17:23:14"
"2021-08-21 02:34:23" "2021-08-21 06:52:55" ...
##  $ ended_at          : chr  "2021-08-10 17:22:44" "2021-08-10 17:39:24"
"2021-08-21 02:50:36" "2021-08-21 07:08:13" ...
##  $ start_station_name: chr  "" "" "" "" ...
##  $ start_station_id  : chr  "" "" "" "" ...
##  $ end_station_name  : chr  "" "" "" "" ...
##  $ end_station_id    : chr  "" "" "" "" ...
##  $ start_lat         : num  41.8 41.8 42 42 41.8 ...
##  $ start_lng         : num  -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ end_lat           : num  41.8 41.8 42 42 41.8 ...
##  $ end_lng           : num  -87.7 -87.6 -87.7 -87.7 -87.6 ...
##  $ member_casual     : chr  "member" "member" "member" "member" ...

str(tripdata_202109)

## 'data.frame':    756147 obs. of  13 variables:
##  $ ride_id           : chr  "9DC7B962304CBFD8" "F930E2C6872D6B32"
"6EF72137900BB910" "78D1DE133B3DBF55" ...
##  $ rideable_type     : chr  "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : chr  "2021-09-28 16:07:10" "2021-09-28 14:24:51"
"2021-09-28 00:20:16" "2021-09-28 14:51:17" ...
##  $ ended_at          : chr  "2021-09-28 16:09:54" "2021-09-28 14:40:05"
"2021-09-28 00:23:57" "2021-09-28 15:00:06" ...
##  $ start_station_name: chr  "" "" "" "" ...
##  $ start_station_id  : chr  "" "" "" "" ...
##  $ end_station_name  : chr  "" "" "" "" ...
##  $ end_station_id    : chr  "" "" "" "" ...
```

```
##  $ start_lat         : num  41.9 41.9 41.8 41.8 41.9 ...
##  $ start_lng         : num  -87.7 -87.6 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num  41.9 42 41.8 41.8 41.9 ...
##  $ end_lng           : num  -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr  "casual" "casual" "casual" "casual" ...
```

```
str(tripdata_202110)
```

```
## 'data.frame':    631226 obs. of  13 variables:
##  $ ride_id           : chr  "620BC6107255BF4C" "4471C70731AB2E45"
"26CA69D43D15EE14" "362947F0437E1514" ...
##  $ rideable_type     : chr  "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : chr  "2021-10-22 12:46:42" "2021-10-21 09:12:37"
"2021-10-16 16:28:39" "2021-10-16 16:17:48" ...
##  $ ended_at          : chr  "2021-10-22 12:49:50" "2021-10-21 09:14:14"
"2021-10-16 16:36:26" "2021-10-16 16:19:03" ...
##  $ start_station_name: chr  "Kingsbury St & Kinzie St" "" "" "" ...
##  $ start_station_id  : chr  "KA1503000043" "" "" "" ...
##  $ end_station_name  : chr  "" "" "" "" ...
##  $ end_station_id    : chr  "" "" "" "" ...
##  $ start_lat         : num  41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num  -87.6 -87.7 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num  41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num  -87.6 -87.7 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr  "member" "member" "member" "member" ...
```

```
str(tripdata_202111)
```

```
## 'data.frame':    359978 obs. of  13 variables:
##  $ ride_id           : chr  "7C00A93E10556E47" "90854840DFD508BA"
"0A7D10CDD144061C" "2F3BE33085BCFF02" ...
##  $ rideable_type     : chr  "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : chr  "2021-11-27 13:27:38" "2021-11-27 13:38:25"
"2021-11-26 22:03:34" "2021-11-27 09:56:49" ...
##  $ ended_at          : chr  "2021-11-27 13:46:38" "2021-11-27 13:56:10"
"2021-11-26 22:05:56" "2021-11-27 10:01:50" ...
##  $ start_station_name: chr  "" "" "" "" ...
##  $ start_station_id  : chr  "" "" "" "" ...
##  $ end_station_name  : chr  "" "" "" "" ...
##  $ end_station_id    : chr  "" "" "" "" ...
##  $ start_lat         : num  41.9 42 42 41.9 41.9 ...
##  $ start_lng         : num  -87.7 -87.7 -87.7 -87.8 -87.6 ...
##  $ end_lat           : num  42 41.9 42 41.9 41.9 ...
##  $ end_lng           : num  -87.7 -87.7 -87.7 -87.8 -87.6 ...
##  $ member_casual     : chr  "casual" "casual" "casual" "casual" ...
```

```
str(tripdata_202112)
```

```
## 'data.frame':    247540 obs. of  13 variables:
##  $ ride_id           : chr  "46F8167220E4431F" "73A77762838B32FD"
"4CF42452054F59C5" "3278BA87BF698339" ...
##  $ rideable_type     : chr  "electric_bike" "electric_bike"
"electric_bike" "classic_bike" ...
##  $ started_at        : chr  "2021-12-07 15:06:07" "2021-12-11 03:43:29"
"2021-12-15 23:10:28" "2021-12-26 16:16:10" ...
##  $ ended_at          : chr  "2021-12-07 15:13:42" "2021-12-11 04:10:23"
"2021-12-15 23:23:14" "2021-12-26 16:30:53" ...
##  $ start_station_name: chr  "Laflin St & Cullerton St" "LaSalle Dr & Huron
St" "Halsted St & North Branch St" "Halsted St & North Branch St" ...
##  $ start_station_id  : chr  "13307" "KP1705001026" "KA1504000117"
"KA1504000117" ...
##  $ end_station_name  : chr  "Morgan St & Polk St" "Clarendon Ave & Leland
Ave" "Broadway & Barry Ave" "LaSalle Dr & Huron St" ...
##  $ end_station_id    : chr  "TA1307000130" "TA1307000119" "13137"
"KP1705001026" ...
##  $ start_lat         : num  41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num  -87.7 -87.6 -87.6 -87.6 -87.7 ...
##  $ end_lat           : num  41.9 42 41.9 41.9 41.9 ...
##  $ end_lng           : num  -87.7 -87.7 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr  "member" "casual" "member" "member" ...

str(tripdata_202201)

## 'data.frame':    103770 obs. of  13 variables:
##  $ ride_id           : chr  "C2F7DD78E82EC875" "A6CF8980A652D272"
"BD0F91DFF741C66D" "CBB80ED419105406" ...
##  $ rideable_type     : chr  "electric_bike" "electric_bike" "classic_bike"
"classic_bike" ...
##  $ started_at        : chr  "2022-01-13 11:59:47" "2022-01-10 08:41:56"
"2022-01-25 04:53:40" "2022-01-04 00:18:04" ...
##  $ ended_at          : chr  "2022-01-13 12:02:44" "2022-01-10 08:46:17"
"2022-01-25 04:58:01" "2022-01-04 00:33:00" ...
##  $ start_station_name: chr  "Glenwood Ave & Touhy Ave" "Glenwood Ave &
Touhy Ave" "Sheffield Ave & Fullerton Ave" "Clark St & Bryn Mawr Ave" ...
##  $ start_station_id  : chr  "525" "525" "TA1306000016" "KA1504000151" ...
##  $ end_station_name  : chr  "Clark St & Touhy Ave" "Clark St & Touhy Ave"
"Greenview Ave & Fullerton Ave" "Paulina St & Montrose Ave" ...
##  $ end_station_id    : chr  "RP-007" "RP-007" "TA1307000001"
"TA1309000021" ...
##  $ start_lat         : num  42 42 41.9 42 41.9 ...
##  $ start_lng         : num  -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ end_lat           : num  42 42 41.9 42 41.9 ...
##  $ end_lng           : num  -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ member_casual     : chr  "casual" "casual" "member" "casual" ...

str(tripdata_202202)

## 'data.frame':    115609 obs. of  13 variables:
##  $ ride_id           : chr  "E1E065E7ED285C02" "1602DCDC5B30FFE3"
```

```
"BE7DD2AF4B55C4AF" "A1789BDF844412BE" ...
##  $ rideable_type     : chr  "classic_bike" "classic_bike" "classic_bike"
"classic_bike" ...
##  $ started_at        : chr  "2022-02-19 18:08:41" "2022-02-20 17:41:30"
"2022-02-25 18:55:56" "2022-02-14 11:57:03" ...
##  $ ended_at          : chr  "2022-02-19 18:23:56" "2022-02-20 17:45:56"
"2022-02-25 19:09:34" "2022-02-14 12:04:00" ...
##  $ start_station_name: chr  "State St & Randolph St" "Halsted St &
Wrightwood Ave" "State St & Randolph St" "Southport Ave & Waveland Ave" ...
##  $ start_station_id  : chr  "TA1305000029" "TA1309000061" "TA1305000029"
"13235" ...
##  $ end_station_name  : chr  "Clark St & Lincoln Ave" "Southport Ave &
Wrightwood Ave" "Canal St & Adams St" "Broadway & Sheridan Rd" ...
##  $ end_station_id    : chr  "13179" "TA1307000113" "13011" "13323" ...
##  $ start_lat         : num  41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num  -87.6 -87.6 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num  41.9 41.9 41.9 42 41.9 ...
##  $ end_lng           : num  -87.6 -87.7 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr  "member" "member" "member" "member" ...
```

2. Data wrangling and Data cleaning

```
all_trips=bind_rows(tripdata_202103,tripdata_202104,tripdata_202105,tripdata_
202106,tripdata_202107,tripdata_202108,
tripdata_202109,tripdata_202110,tripdata_202111,tripdata_202112,tripdata_2022
01,tripdata_202202) #Combine all files into a single file
head(all_trips,10)
```

```
##             ride_id rideable_type          started_at            ended_at
## 1  CFA86D4455AA1030  classic_bike 2021-03-16 08:32:30 2021-03-16 08:36:34
## 2  30D9DC61227D1AF3  classic_bike 2021-03-28 01:26:28 2021-03-28 01:36:55
## 3  846D87A15682A284  classic_bike 2021-03-11 21:17:29 2021-03-11 21:33:53
## 4  994D05AA75A168F2  classic_bike 2021-03-11 13:26:42 2021-03-11 13:55:41
## 5  DF7464FBE92D8308  classic_bike 2021-03-21 09:09:37 2021-03-21 09:27:33
## 6  CEBA8516FD17F8D8  classic_bike 2021-03-20 11:08:47 2021-03-20 11:29:39
## 7  297268586B79588B  classic_bike 2021-03-20 14:10:41 2021-03-20 14:22:13
## 8  F39301858B6077DD electric_bike 2021-03-23 07:56:51 2021-03-23 08:05:50
## 9  D297F199D875BABE electric_bike 2021-03-31 15:31:19 2021-03-31 15:35:58
## 10 36B877141175ED7E  classic_bike 2021-03-11 17:37:37 2021-03-11 17:52:44
##              start_station_name start_station_id
## 1  Humboldt Blvd & Armitage Ave            15651
## 2  Humboldt Blvd & Armitage Ave            15651
## 3         Shields Ave & 28th Pl            15443
## 4   Winthrop Ave & Lawrence Ave     TA1308000021
## 5      Glenwood Ave & Touhy Ave              525
## 6      Glenwood Ave & Touhy Ave              525
## 7          State St & Kinzie St            13050
## 8            Shore Dr & 55th St     TA1308000009
## 9          Clinton St & Lake St            13021
## 10       Michigan Ave & Lake St     TA1305000011
##                        end_station_name end_station_id start_lat start_lng
```

```
## 1              Stave St & Armitage Ave          13266  41.91751 -87.70181
## 2  Central Park Ave & Bloomingdale Ave          18017  41.91751 -87.70181
## 3                 Halsted St & 35th St   TA1308000043  41.84273 -87.63549
## 4               Broadway & Sheridan Rd          13323  41.96881 -87.65766
## 5            Chicago Ave & Sheridan Rd           E008  42.01270 -87.66606
## 6            Chicago Ave & Sheridan Rd           E008  42.01270 -87.66606
## 7           Lake Shore Dr & North Blvd         LF-005  41.88919 -87.62775
## 8                 Ellis Ave & 60th St   KA1503000014  41.79523 -87.58083
## 9           Franklin St & Jackson Blvd   TA1305000025  41.88555 -87.64173
## 10        Racine Ave & Washington Blvd            654  41.88602 -87.62412
##     end_lat   end_lng member_casual
## 1  41.91774 -87.69139        casual
## 2  41.91417 -87.71676        casual
## 3  41.83066 -87.64717        casual
## 4  41.95283 -87.64999        casual
## 5  42.05049 -87.67782        casual
## 6  42.05049 -87.67782        casual
## 7  41.91172 -87.62680        member
## 8  41.78522 -87.60108        member
## 9  41.87729 -87.63616        member
## 10 41.88307 -87.65695        member

str(all_trips)

## 'data.frame':    5667986 obs. of  13 variables:
##  $ ride_id           : chr  "CFA86D4455AA1030" "30D9DC61227D1AF3"
"846D87A15682A284" "994D05AA75A168F2" ...
##  $ rideable_type     : chr  "classic_bike" "classic_bike" "classic_bike"
"classic_bike" ...
##  $ started_at        : chr  "2021-03-16 08:32:30" "2021-03-28 01:26:28"
"2021-03-11 21:17:29" "2021-03-11 13:26:42" ...
##  $ ended_at          : chr  "2021-03-16 08:36:34" "2021-03-28 01:36:55"
"2021-03-11 21:33:53" "2021-03-11 13:55:41" ...
##  $ start_station_name: chr  "Humboldt Blvd & Armitage Ave" "Humboldt Blvd
& Armitage Ave" "Shields Ave & 28th Pl" "Winthrop Ave & Lawrence Ave" ...
##  $ start_station_id  : chr  "15651" "15651" "15443" "TA1308000021" ...
##  $ end_station_name  : chr  "Stave St & Armitage Ave" "Central Park Ave &
Bloomingdale Ave" "Halsted St & 35th St" "Broadway & Sheridan Rd" ...
##  $ end_station_id    : chr  "13266" "18017" "TA1308000043" "13323" ...
##  $ start_lat         : num  41.9 41.9 41.8 42 42 ...
##  $ start_lng         : num  -87.7 -87.7 -87.6 -87.7 -87.7 ...
##  $ end_lat           : num  41.9 41.9 41.8 42 42.1 ...
##  $ end_lng           : num  -87.7 -87.7 -87.6 -87.6 -87.7 ...
##  $ member_casual     : chr  "casual" "casual" "casual" "casual" ...
```

Changing the data type of started_at and ended_at from char to datetime datatype

```
all_trips[['started_at']] <- ymd_hms(all_trips[['started_at']])
all_trips[['ended_at']] <- ymd_hms(all_trips[['ended_at']])
```

Renaming columns to make them consistent and understandable

```
all_trips <- rename(all_trips,
                    trip_id = ride_id,
                    bike_type = rideable_type,
                    start_time = started_at,
                    end_time = ended_at,
                    from_station_name = start_station_name,
                    from_station_id = start_station_id,
                    to_station_name = end_station_name,
                    to_station_id = end_station_id,
                    usertype = member_casual)
```

Inspect the new table that has been created

```
colnames(all_trips) #List of column names
```

```
##  [1] "trip_id"           "bike_type"         "start_time"
##  [4] "end_time"          "from_station_name" "from_station_id"
##  [7] "to_station_name"   "to_station_id"     "start_lat"
## [10] "start_lng"         "end_lat"           "end_lng"
## [13] "usertype"
```

```
nrow(all_trips) #Total no. of rows
```

```
## [1] 5667986
```

```
dim(all_trips)  #Dimensions of the data frame
```

```
## [1] 5667986      13
```

```
head(all_trips,10) #First 10 rows
```

```
##             trip_id     bike_type          start_time            end_time
## 1  CFA86D4455AA1030  classic_bike 2021-03-16 08:32:30 2021-03-16 08:36:34
## 2  30D9DC61227D1AF3  classic_bike 2021-03-28 01:26:28 2021-03-28 01:36:55
## 3  846D87A15682A284  classic_bike 2021-03-11 21:17:29 2021-03-11 21:33:53
## 4  994D05AA75A168F2  classic_bike 2021-03-11 13:26:42 2021-03-11 13:55:41
## 5  DF7464FBE92D8308  classic_bike 2021-03-21 09:09:37 2021-03-21 09:27:33
## 6  CEBA8516FD17F8D8  classic_bike 2021-03-20 11:08:47 2021-03-20 11:29:39
## 7  297268586B79588B  classic_bike 2021-03-20 14:10:41 2021-03-20 14:22:13
## 8  F39301858B6077DD electric_bike 2021-03-23 07:56:51 2021-03-23 08:05:50
## 9  D297F199D875BABE electric_bike 2021-03-31 15:31:19 2021-03-31 15:35:58
## 10 36B877141175ED7E  classic_bike 2021-03-11 17:37:37 2021-03-11 17:52:44
##                from_station_name from_station_id
## 1  Humboldt Blvd & Armitage Ave           15651
## 2  Humboldt Blvd & Armitage Ave           15651
## 3         Shields Ave & 28th Pl           15443
## 4   Winthrop Ave & Lawrence Ave    TA1308000021
## 5      Glenwood Ave & Touhy Ave             525
## 6      Glenwood Ave & Touhy Ave             525
## 7          State St & Kinzie St           13050
## 8            Shore Dr & 55th St    TA1308000009
```

```
## 9          Clinton St & Lake St           13021
## 10       Michigan Ave & Lake St    TA1305000011
##                        to_station_name to_station_id start_lat start_lng
## 1                Stave St & Armitage Ave         13266  41.91751 -87.70181
## 2  Central Park Ave & Bloomingdale Ave         18017  41.91751 -87.70181
## 3                  Halsted St & 35th St  TA1308000043  41.84273 -87.63549
## 4                Broadway & Sheridan Rd         13323  41.96881 -87.65766
## 5             Chicago Ave & Sheridan Rd          E008  42.01270 -87.66606
## 6             Chicago Ave & Sheridan Rd          E008  42.01270 -87.66606
## 7            Lake Shore Dr & North Blvd        LF-005  41.88919 -87.62775
## 8                  Ellis Ave & 60th St  KA1503000014  41.79523 -87.58083
## 9            Franklin St & Jackson Blvd  TA1305000025  41.88555 -87.64173
## 10         Racine Ave & Washington Blvd           654  41.88602 -87.62412
##     end_lat   end_lng usertype
## 1  41.91774 -87.69139   casual
## 2  41.91417 -87.71676   casual
## 3  41.83066 -87.64717   casual
## 4  41.95283 -87.64999   casual
## 5  42.05049 -87.67782   casual
## 6  42.05049 -87.67782   casual
## 7  41.91172 -87.62680   member
## 8  41.78522 -87.60108   member
## 9  41.87729 -87.63616   member
## 10 41.88307 -87.65695   member

tail(all_trips,10) #Last 10 rows

##                  trip_id     bike_type          start_time
end_time
## 5667977 2F8C1FE6298DE76A electric_bike 2022-02-06 05:26:22 2022-02-06
06:05:09
## 5667978 608F7DC6821FE4FF electric_bike 2022-02-01 23:58:44 2022-02-02
00:16:59
## 5667979 70DB19460D085AA0  classic_bike 2022-02-19 23:57:50 2022-02-20
00:02:34
## 5667980 188B462EB7962B3F electric_bike 2022-02-18 09:59:16 2022-02-18
10:05:48
## 5667981 BDEB7AE264C7B778 electric_bike 2022-02-25 23:25:25 2022-02-25
23:29:26
## 5667982 211BE0DC162D85B7 electric_bike 2022-02-23 17:47:49 2022-02-23
18:02:29
## 5667983 D4D53E78000C8CA1 electric_bike 2022-02-04 10:43:47 2022-02-04
10:50:52
## 5667984 9E85F07D2F94492B electric_bike 2022-02-28 09:16:33 2022-02-28
09:28:11
## 5667985 B61B559F81F1D823 electric_bike 2022-02-10 16:55:16 2022-02-10
16:57:53
## 5667986 841C701610CF0609 electric_bike 2022-02-21 16:35:20 2022-02-21
16:42:53
##                      from_station_name from_station_id
```

```
## 5667977
## 5667978
## 5667979 California Ave & Milwaukee Ave           13084
## 5667980
## 5667981
## 5667982
## 5667983
## 5667984          Wood St & Chicago Ave             637
## 5667985
## 5667986
##                      to_station_name to_station_id start_lat start_lng
end_lat
## 5667977                                             41.84000 -87.62000
41.84000
## 5667978                                             41.78000 -87.61000
41.80000
## 5667979 Humboldt Blvd & Armitage Ave         15651  41.92269 -87.69715
41.91751
## 5667980     Leavitt St & Chicago Ave         18058  41.89000 -87.69000
41.89550
## 5667981 Humboldt Blvd & Armitage Ave         15651  41.92000 -87.69000
41.91751
## 5667982     Leavitt St & Chicago Ave         18058  41.88000 -87.63000
41.89550
## 5667983     Leavitt St & Chicago Ave         18058  41.91000 -87.68000
41.89550
## 5667984          Canal St & Adams St         13011  41.89571 -87.67221
41.87926
## 5667985          Canal St & Adams St         13011  41.88000 -87.63000
41.87926
## 5667986         Larrabee St & Oak St  KA1504000116  41.88000 -87.65000
41.90022
##           end_lng usertype
## 5667977 -87.62000   member
## 5667978 -87.59000   member
## 5667979 -87.70181   member
## 5667980 -87.68202   member
## 5667981 -87.70181   member
## 5667982 -87.68202   member
## 5667983 -87.68202   member
## 5667984 -87.63990   member
## 5667985 -87.63990   member
## 5667986 -87.64299   member

str(all_trips)

## 'data.frame':    5667986 obs. of  13 variables:
##  $ trip_id          : chr  "CFA86D4455AA1030" "30D9DC61227D1AF3"
"846D87A15682A284" "994D05AA75A168F2" ...
##  $ bike_type        : chr  "classic_bike" "classic_bike" "classic_bike"
```

```
"classic_bike" ...
##  $ start_time       : POSIXct, format: "2021-03-16 08:32:30" "2021-03-28
01:26:28" ...
##  $ end_time         : POSIXct, format: "2021-03-16 08:36:34" "2021-03-28
01:36:55" ...
##  $ from_station_name: chr  "Humboldt Blvd & Armitage Ave" "Humboldt Blvd &
Armitage Ave" "Shields Ave & 28th Pl" "Winthrop Ave & Lawrence Ave" ...
##  $ from_station_id  : chr  "15651" "15651" "15443" "TA1308000021" ...
##  $ to_station_name  : chr  "Stave St & Armitage Ave" "Central Park Ave &
Bloomingdale Ave" "Halsted St & 35th St" "Broadway & Sheridan Rd" ...
##  $ to_station_id    : chr  "13266" "18017" "TA1308000043" "13323" ...
##  $ start_lat        : num  41.9 41.9 41.8 42 42 ...
##  $ start_lng        : num  -87.7 -87.7 -87.6 -87.7 -87.7 ...
##  $ end_lat          : num  41.9 41.9 41.8 42 42.1 ...
##  $ end_lng          : num  -87.7 -87.7 -87.6 -87.6 -87.7 ...
##  $ usertype         : chr  "casual" "casual" "casual" "casual" ...
```

Add columns that list the date, month, day, and year of each ride

```
all_trips$date <- as.Date(all_trips$start_time)
all_trips$month <- format(as.Date(all_trips$date), '%B')
all_trips$day <- format(as.Date(all_trips$date), "%d")
all_trips$year <- format(as.Date(all_trips$date), "%Y")
all_trips$day_of_week <- format(as.Date(all_trips$date), "%A")
```

Add a "ride_length" calculation to all_trips (in minutes)

```
all_trips$ride_length <-
as.numeric(difftime(all_trips$end_time,all_trips$start_time,units = 'mins'))
```

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

```
all_trips_v2 <- all_trips[!(all_trips$from_station_name == "HQ QR" |
all_trips$ride_length<0),]
```

3. Descriptive analysis

```
summary(all_trips_v2)  #Statistical summary of data

##    trip_id            bike_type           start_time
##  Length:5667841     Length:5667841     Min.   :2021-03-01 00:01:09
##  Class :character   Class :character   1st Qu.:2021-06-13 11:41:27
##  Mode  :character   Mode  :character   Median :2021-08-07 19:11:51
##                                        Mean   :2021-08-10 07:32:40
##                                        3rd Qu.:2021-10-02 14:15:39
##                                        Max.   :2022-02-28 23:58:44
##
##     end_time                    from_station_name  from_station_id
##  Min.   :2021-03-01 00:06:28   Length:5667841     Length:5667841
```

```
##  1st Qu.:2021-06-13 12:08:45   Class :character   Class :character
##  Median :2021-08-07 19:34:17   Mode  :character   Mode  :character
##  Mean   :2021-08-10 07:54:25
##  3rd Qu.:2021-10-02 14:38:40
##  Max.   :2022-03-01 08:55:17
##
##  to_station_name    to_station_id        start_lat       start_lng
##  Length:5667841     Length:5667841     Min.   :41.64   Min.   :-87.84
##  Class :character   Class :character   1st Qu.:41.88   1st Qu.:-87.66
##  Mode  :character   Mode  :character   Median :41.90   Median :-87.64
##                                        Mean   :41.90   Mean   :-87.65
##                                        3rd Qu.:41.93   3rd Qu.:-87.63
##                                        Max.   :45.64   Max.   :-73.80
##
##     end_lat         end_lng         usertype             date
##  Min.   :41.39   Min.   :-88.97   Length:5667841     Min.   :2021-03-01
##  1st Qu.:41.88   1st Qu.:-87.66   Class :character   1st Qu.:2021-06-13
##  Median :41.90   Median :-87.64   Mode  :character   Median :2021-08-07
##  Mean   :41.90   Mean   :-87.65                      Mean   :2021-08-09
##  3rd Qu.:41.93   3rd Qu.:-87.63                      3rd Qu.:2021-10-02
##  Max.   :42.17   Max.   :-87.49                      Max.   :2022-02-28
##  NA's   :4617    NA's   :4617
##     month               day               year          day_of_week
##  Length:5667841     Length:5667841     Length:5667841     Length:5667841
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   ride_length
##  Min.   :    0.00
##  1st Qu.:    6.67
##  Median :   11.87
##  Mean   :   21.75
##  3rd Qu.:   21.57
##  Max.   :55944.15
##
```

```
table(all_trips$usertype) #Checking how many casual riders and members are
there
```

```
##
##  casual  member
## 2540693 3127293
```

```
summary(all_trips_v2$ride_length) #Descriptive analysis on ride_length
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##     0.00     6.67    11.87    21.75    21.57 55944.15
```

Compare members and casual users

```
aggregate(all_trips_v2$ride_length ~ all_trips_v2$usertype, FUN = mean)
```

```
##   all_trips_v2$usertype all_trips_v2$ride_length
## 1                casual                 31.92153
## 2                member                 13.48515
```

```
aggregate(all_trips_v2$ride_length ~ all_trips_v2$usertype, FUN = median)
```

```
##   all_trips_v2$usertype all_trips_v2$ride_length
## 1                casual               15.900000
## 2                member                9.483333
```

```
aggregate(all_trips_v2$ride_length ~ all_trips_v2$usertype, FUN = max)
```

```
##   all_trips_v2$usertype all_trips_v2$ride_length
## 1                casual               55944.150
## 2                member                1559.933
```

```
aggregate(all_trips_v2$ride_length ~ all_trips_v2$usertype, FUN = min)
```

```
##   all_trips_v2$usertype all_trips_v2$ride_length
## 1                casual                        0
## 2                member                        0
```

The days of the week are out of order. Let's fix that.

```
all_trips_v2$day_of_week <- ordered(all_trips_v2$day_of_week,
levels=c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
"Saturday"))
```

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

```
aggregate(all_trips_v2$ride_length ~
all_trips_v2$usertype+all_trips_v2$day_of_week, FUN = mean)
```

```
##    all_trips_v2$usertype all_trips_v2$day_of_week all_trips_v2$ride_length
## 1                 casual                   Sunday                 37.50135
## 2                 member                   Sunday                 15.48980
## 3                 casual                   Monday                 31.80985
## 4                 member                   Monday                 13.05340
## 5                 casual                  Tuesday                 27.84800
## 6                 member                  Tuesday                 12.67914
## 7                 casual                Wednesday                 27.67719
## 8                 member                Wednesday                 12.65123
## 9                 casual                 Thursday                 27.84429
## 10                member                 Thursday                 12.67423
## 11                casual                   Friday                 30.17155
## 12                member                   Friday                 13.21802
## 13                casual                 Saturday                 34.54571
## 14                member                 Saturday                 15.09962
```

Analyze ridership data by type and weekday

```
all_trips_v2 %>%
  group_by(usertype, day_of_week)  %>%
  summarise(number_of_rides = n()    #calculates the number of rides and
average duration
            ,average_duration = mean(ride_length))  # calculates the average
duration
```

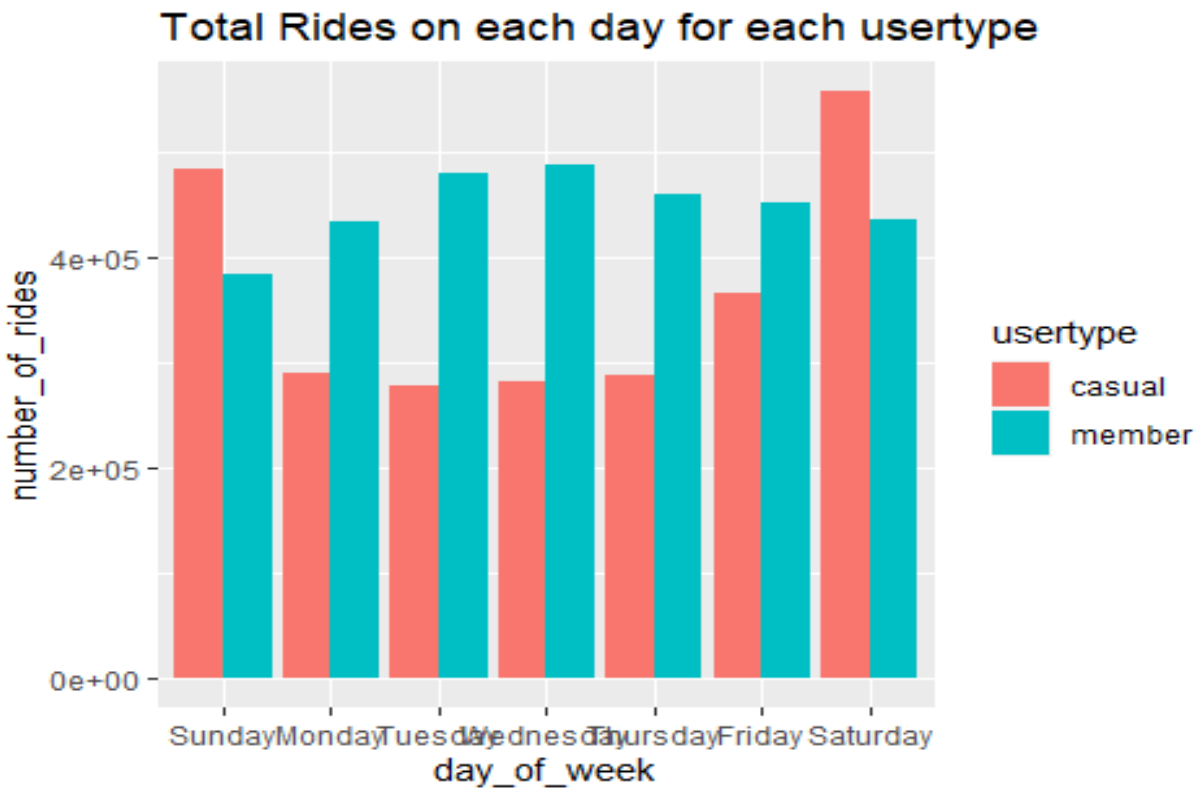
```
## `summarise()` has grouped output by 'usertype'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 14 x 4
## # Groups:   usertype [2]
##    usertype day_of_week number_of_rides average_duration
##    <chr>    <ord>                <int>            <dbl>
##  1 casual   Sunday              483568             37.5
##  2 casual   Monday              290534             31.8
##  3 casual   Tuesday             276680             27.8
##  4 casual   Wednesday           280743             27.7
##  5 casual   Thursday            287096             27.8
##  6 casual   Friday              364898             30.2
##  7 casual   Saturday            557115             34.5
##  8 member   Sunday              383934             15.5
##  9 member   Monday              432813             13.1
## 10 member   Tuesday             479226             12.7
## 11 member   Wednesday           486730             12.7
## 12 member   Thursday            459167             12.7
## 13 member   Friday              450496             13.2
## 14 member   Saturday            434841             15.1
```

Number of rides by weekday for each user type

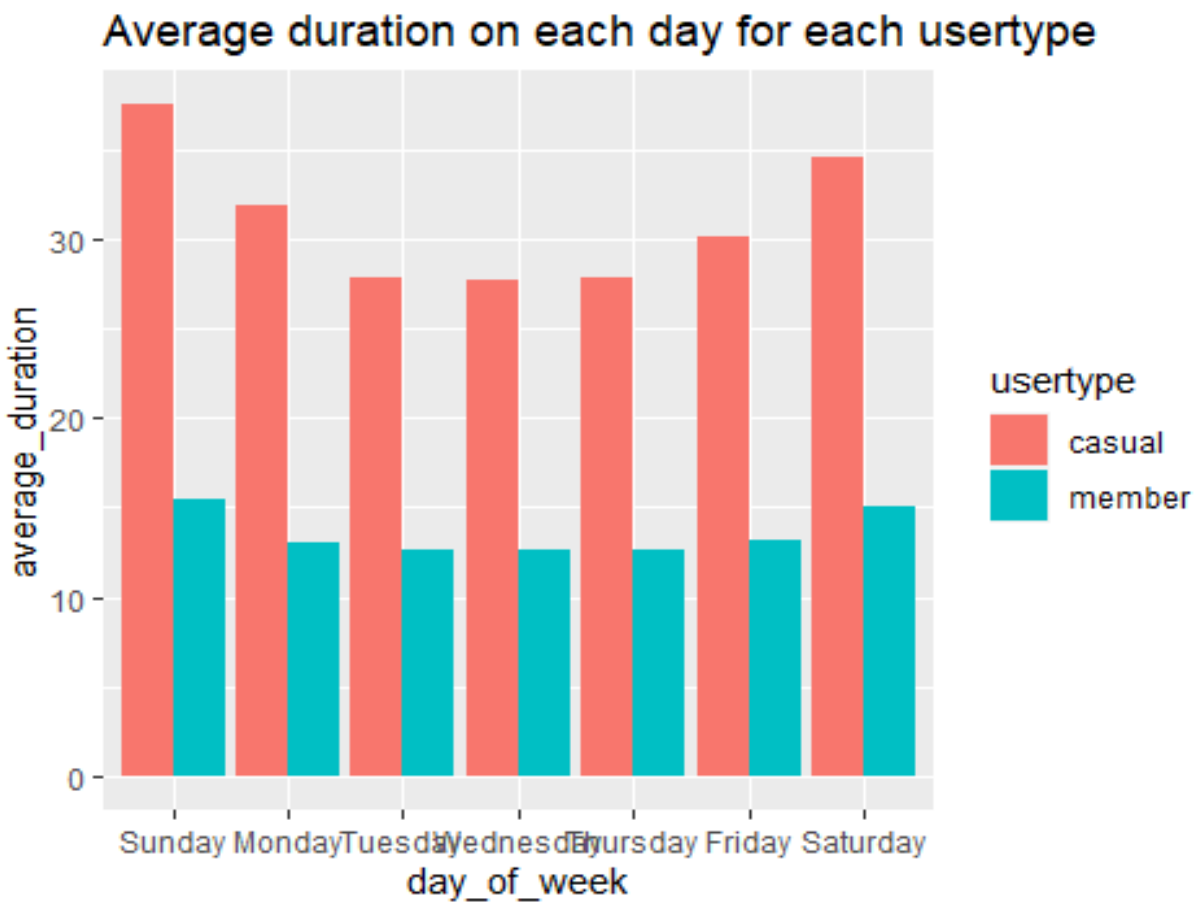
```
all_trips_v2 %>%
  group_by(usertype, day_of_week) %>%
  summarise(number_of_rides = n()) %>%
  ggplot(aes(x = day_of_week, y = number_of_rides, fill = usertype)) +
  geom_col(position = "dodge") + labs (title = "Total Rides on each day for
each usertype")
```

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



Average duration by weekday

```
all_trips_v2 %>%
  group_by(usertype, day_of_week) %>%
  summarise(average_duration = mean(ride_length)) %>%
  ggplot(aes(x = day_of_week, y = average_duration, fill = usertype)) +
  geom_col(position = "dodge") + labs (title = "Average duration on each day
for each usertype")

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

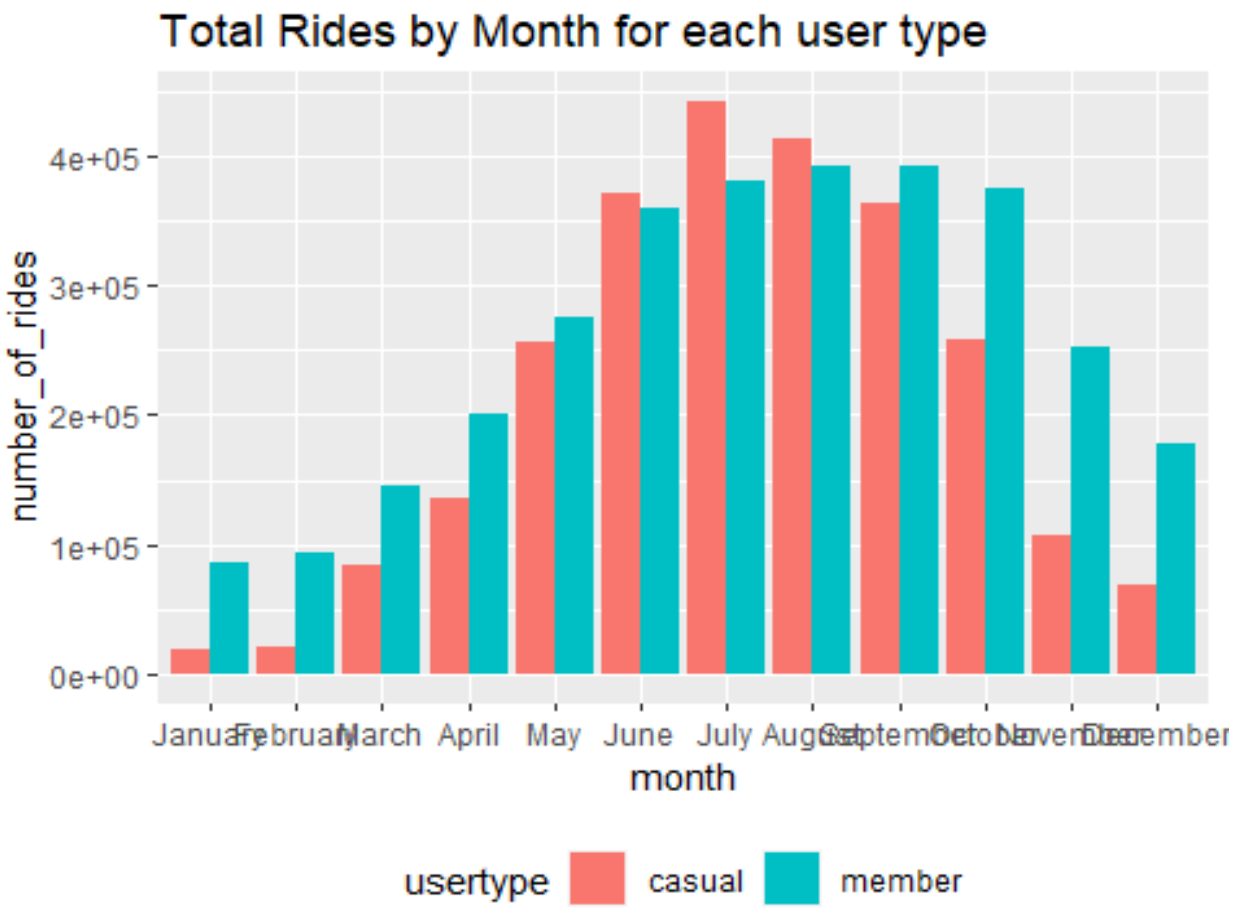


Ordering the months

```
all_trips_v2$month <- ordered(all_trips_v2$month,
levels=c("January","February","March","April","May","June","July","August","S
eptember","October","November","December"))
```

Total Rides by months

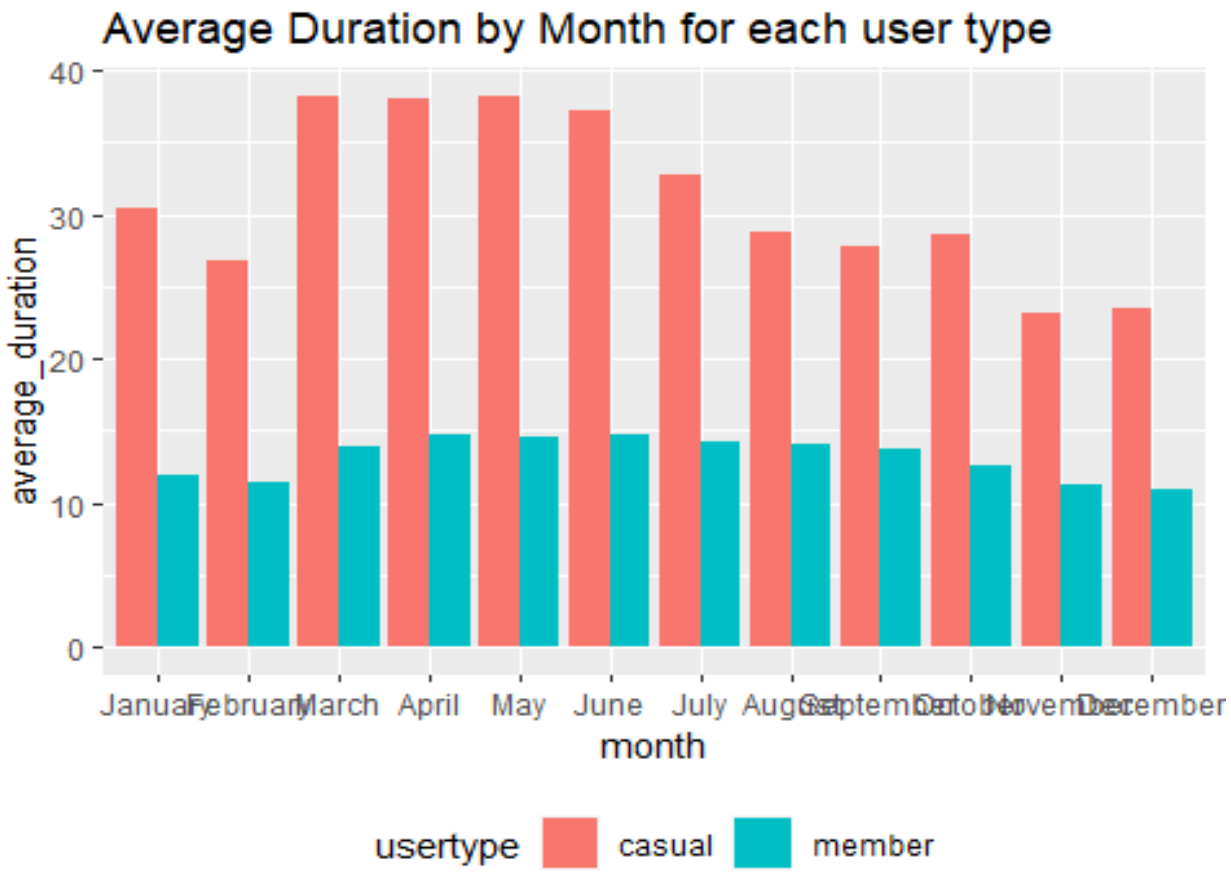
```
all_trips_v2 %>%
  group_by(usertype, month) %>%
  summarise(number_of_rides = n()) %>%
  ggplot(aes(x = month, y = number_of_rides, fill = usertype)) +
geom_bar(position = "dodge", stat='identity') +
  labs(
    title = "Total Rides by Month for each user type") +
  theme(legend.position = "bottom")

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



Average Duration by months

```
all_trips_v2 %>%
  group_by(usertype, month) %>%
  summarise(average_duration = mean(ride_length)) %>%
  ggplot(aes(x = month, y = average_duration, fill = usertype)) +
geom_bar(position = "dodge", stat='identity') +
  labs(title = "Average Duration by Month for each user type") +
  theme(legend.position = "bottom")
```

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



Exporting summary file for further analysis

```
File <- aggregate(all_trips_v2$ride_length ~
all_trips_v2$usertype+all_trips_v2$day_of_week, FUN = mean)
write.csv(File, file = "C:/Users/Justin/OneDrive/Documents/R
Coding/Cyclistic_data/avg_length.csv")
write.csv(all_trips_v2, file = "C:/Users/Justin/OneDrive/Documents/R
Coding/Cyclistic_data/All_trips_data.csv")
```