CAPSTONE CYCLISTIC BIKE – GOOGLE DATA ANALYTICS SCIENCE

AHMED EL HAMOUNI

JUNE 11, 2023

GITHUB : ahmedelhamouni / **CAPSTONE-CYCLISTIC-BIKE**

```
R version 4.3.0 (2023-04-21 ucrt) -- "Already Tomorrow"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> # Loading necessary libraries
> library(tidyverse)
── Attaching core tidyverse packages ──────────────────────────────────────── tidyverse 2
✓ dplyr     1.1.2     ✓ readr     2.1.4
✓ forcats   1.0.0     ✓ stringr   1.5.0
✓ ggplot2   3.4.2     ✓ tibble    3.2.1
✓ lubridate 1.9.2     ✓ tidyr     1.3.0
✓ purrr     1.0.1
── Conflicts ────────────────────────────────────────────────────────── tidyverse_conflic
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package to force all conflicts to become errors
> library(lubridate)
> library(ggplot2)
> library(dplyr)
> library(hms) # For converting time values

Attachement du package : ‘hms’

L'objet suivant est masqué depuis ‘package:lubridate’:

    hms

> # Set the scientific notation display option to 100 decimal places
> options(scipen = 100)
> # Changing language to English for weekday names
> Sys.setlocale("LC_TIME", "en_US.UTF-8
+
+ # Importing csv files as data frames using the readr library
+ df_032022 <- read.csv("202203-divvy-tripdata.csv")
Error: unexpected numeric constant in:
"# Importing csv files as data frames using the readr library
df_032022 <- read.csv("202203"
> # Importing csv files as data frames using the readr library
> df_062022 <- read.csv("202206-divvy-tripdata.csv")
> df_072022 <- read.csv("202207-divvy-tripdata.csv")
> df_082022 <- read.csv("202208-divvy-tripdata.csv")
> df_092022 <- read.csv("202209-divvy-tripdata.csv")
> df_102022 <- read.csv("202210-divvy-tripdata.csv")
> df_112022 <- read.csv("202211-divvy-tripdata.csv")
> df_122022 <- read.csv("202212-divvy-tripdata.csv")
> df_032022 <- read.csv("202303-divvy-tripdata.csv")
> df_042023 <- read.csv("202304-divvy-tripdata.csv")
> df_052023 <- read.csv("202305-divvy-tripdata.csv")
> df_012023 <- read.csv("202301-divvy-tripdata.csv")
```

```
> df_022023 <- read.csv("202302-divvy-tripdata.csv")
> # Checking whether all data frames have the same column names before merging
> colnames(df_022023)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_012023)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_122022)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_112022)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_102022)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_092022)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_082022)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_072022)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_062022)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_052023)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_042023)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_032023)
Error: object 'df_032023' not found
> # Importing csv files as data frames using the readr library
> df_062022 <- read.csv("202206-divvy-tripdata.csv")
> df_072022 <- read.csv("202207-divvy-tripdata.csv")
> df_082022 <- read.csv("202208-divvy-tripdata.csv")
> df_092022 <- read.csv("202209-divvy-tripdata.csv")
> df_102022 <- read.csv("202210-divvy-tripdata.csv")
> df_112022 <- read.csv("202211-divvy-tripdata.csv")
> df_122022 <- read.csv("202212-divvy-tripdata.csv")
> df_032023 <- read.csv("202303-divvy-tripdata.csv")
> df_042023 <- read.csv("202304-divvy-tripdata.csv")
> df_052023 <- read.csv("202305-divvy-tripdata.csv")
> df_012023 <- read.csv("202301-divvy-tripdata.csv")
> df_022023 <- read.csv("202302-divvy-tripdata.csv")
> # Checking whether all data frames have the same column names before merging
> colnames(df_062022)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_072022)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_082022)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_092022)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_102022)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
```

```
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_112022)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_122022)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_032023)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_042023)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_052023)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_012023)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> colnames(df_022023)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_la
[13] "member_casual"
> # Before merging rows, check if data in columns are of the same type
> str(df_062022)
'data.frame':  769204 obs. of  13 variables:
 $ ride_id           : chr  "600CFD130D0FD2A4" "F5E6B5C1682C6464" "B6EB6D27BAD771D2" "C9C320375D
 $ rideable_type     : chr  "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...
 $ started_at        : chr  "2022-06-30 17:27:53" "2022-06-30 18:39:52" "2022-06-30 11:49:25" "2
 $ ended_at          : chr  "2022-06-30 17:35:15" "2022-06-30 18:47:28" "2022-06-30 12:02:54" "2
 $ 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 41.8 41.9 ...
 $ start_lng         : num  -87.6 -87.6 -87.7 -87.7 -87.6 ...
 $ end_lat           : num  41.9 41.9 41.9 41.8 41.9 ...
 $ end_lng           : num  -87.6 -87.6 -87.6 -87.7 -87.6 ...
 $ member_casual     : chr  "casual" "casual" "casual" "casual" ...
> str(df_072022)
'data.frame':  823488 obs. of  13 variables:
 $ ride_id           : chr  "954144C2F67B1932" "292E027607D218B6" "57765852588AD6E0" "B5B6BE4431
 $ rideable_type     : chr  "classic_bike" "classic_bike" "classic_bike" "classic_bike" ...
 $ started_at        : chr  "2022-07-05 08:12:47" "2022-07-26 12:53:38" "2022-07-03 13:58:49" "2
 $ ended_at          : chr  "2022-07-05 08:24:32" "2022-07-26 12:55:31" "2022-07-03 14:06:32" "2
 $ start_station_name: chr  "Ashland Ave & Blackhawk St" "Buckingham Fountain (Temp)" "Buckingha
p)" ...
 $ start_station_id  : chr  "13224" "15541" "15541" "15541" ...
 $ end_station_name  : chr  "Kingsbury St & Kinzie St" "Michigan Ave & 8th St" "Michigan Ave & 8
 $ end_station_id    : chr  "KA1503000043" "623" "623" "TA1307000164" ...
 $ start_lat         : num  41.9 41.9 41.9 41.9 41.9 ...
 $ start_lng         : num  -87.7 -87.6 -87.6 -87.6 -87.6 ...
 $ end_lat           : num  41.9 41.9 41.9 41.8 41.9 ...
 $ end_lng           : num  -87.6 -87.6 -87.6 -87.6 -87.7 ...
 $ member_casual     : chr  "member" "casual" "casual" "casual" ...
> str(df_082022)
'data.frame':  785932 obs. of  13 variables:
 $ ride_id           : chr  "550CF7EFEAE0C618" "DAD198F405F9C5F5" "E6F2BC47B65CB7FD" "F597830181
 $ rideable_type     : chr  "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...
 $ started_at        : chr  "2022-08-07 21:34:15" "2022-08-08 14:39:21" "2022-08-08 15:29:50" "2
 $ ended_at          : chr  "2022-08-07 21:41:46" "2022-08-08 14:53:23" "2022-08-08 15:40:34" "2
 $ start_station_name: chr  "" "" "" "" ...
 $ start_station_id  : chr  "" "" "" "" ...
 $ end_station_name  : chr  "" "" "" "" ...
 $ end_station_id    : chr  "" "" "" "" ...
 $ start_lat         : num  41.9 41.9 42 41.9 41.9 ...
 $ start_lng         : num  -87.7 -87.6 -87.7 -87.7 -87.7 ...
 $ end_lat           : num  41.9 41.9 42 42 41.8 ...
 $ end_lng           : num  -87.7 -87.6 -87.7 -87.7 -87.7 ...
 $ member_casual     : chr  "casual" "casual" "casual" "casual" ...
> str(df_092022)
'data.frame':  701339 obs. of  13 variables:
 $ ride_id           : chr  "5156990AC19CA285" "E12D4A16BF51C274" "A02B53CD7DB72DD7" "C82E05FEE8
```

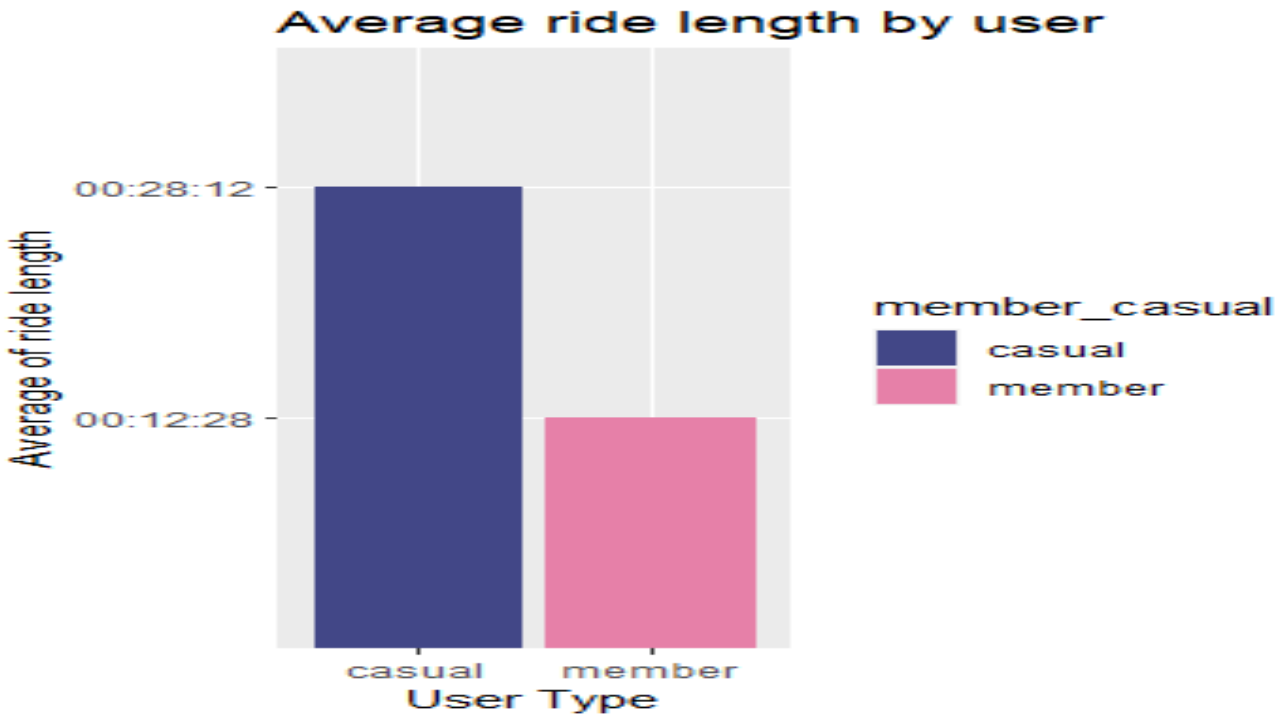
```
 $ rideable_type     : chr  "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...
 $ started_at        : chr  "2022-09-01 08:36:22" "2022-09-01 17:11:29" "2022-09-01 17:15:50" "2
 $ ended_at          : chr  "2022-09-01 08:39:05" "2022-09-01 17:14:45" "2022-09-01 17:16:12" "2
 $ start_station_name: chr  "" "" "" "" ...
 $ start_station_id  : chr  "" "" "" "" ...
 $ end_station_name  : chr  "California Ave & Milwaukee Ave" "" "" "" ...
 $ end_station_id    : chr  "13084" "" "" "" ...
 $ start_lat         : num  41.9 41.9 41.9 41.9 41.9 ...
 $ start_lng         : num  -87.7 -87.6 -87.6 -87.7 -87.7 ...
 $ end_lat           : num  41.9 41.9 41.9 41.9 41.9 ...
 $ end_lng           : num  -87.7 -87.6 -87.6 -87.7 -87.7 ...
 $ member_casual     : chr  "casual" "casual" "casual" "casual" ...
> str(df_102022)
'data.frame':  558685 obs. of  13 variables:
 $ ride_id           : chr  "A50255C1E17942AB" "DB692A70BD2DD4E3" "3C02727AAF60F873" "47E653FDC2
 $ rideable_type     : chr  "classic_bike" "electric_bike" "electric_bike" "electric_bike" ...
 $ started_at        : chr  "2022-10-14 17:13:30" "2022-10-01 16:29:26" "2022-10-19 18:55:40" "2
 $ ended_at          : chr  "2022-10-14 17:19:39" "2022-10-01 16:49:06" "2022-10-19 19:03:30" "2
 $ start_station_name: chr  "Noble St & Milwaukee Ave" "Damen Ave & Charleston St" "Hoyne Ave &
 $ start_station_id  : chr  "13290" "13288" "655" "KA1504000133" ...
 $ end_station_name  : chr  "Larrabee St & Division St" "Damen Ave & Cullerton St" "Western Ave
T Apts)" ...
 $ end_station_id    : chr  "KA1504000079" "13089" "TA1307000140" "620" ...
 $ start_lat         : num  41.9 41.9 42 41.9 41.9 ...
 $ start_lng         : num  -87.7 -87.7 -87.7 -87.6 -87.6 ...
 $ end_lat           : num  41.9 41.9 42 41.9 41.9 ...
 $ end_lng           : num  -87.6 -87.7 -87.7 -87.6 -87.6 ...
 $ member_casual     : chr  "member" "casual" "member" "member" ...
> str(df_112022)
'data.frame':  337735 obs. of  13 variables:
 $ ride_id           : chr  "BCC66FC6FAB27CC7" "772AB67E902C180F" "585EAD07FDEC0152" "91C4E7ED3C
 $ rideable_type     : chr  "electric_bike" "classic_bike" "classic_bike" "classic_bike" ...
 $ started_at        : chr  "2022-11-10 06:21:55" "2022-11-04 07:31:55" "2022-11-21 17:20:29" "2
 $ ended_at          : chr  "2022-11-10 06:31:27" "2022-11-04 07:46:25" "2022-11-21 17:34:36" "2
 $ start_station_name: chr  "Canal St & Adams St" "Canal St & Adams St" "Indiana Ave & Roosevelt
 $ start_station_id  : chr  "13011" "13011" "SL-005" "SL-005" ...
 $ end_station_name  : chr  "St. Clair St & Erie St" "St. Clair St & Erie St" "St. Clair St & Er
 $ end_station_id    : chr  "13016" "13016" "13016" "13016" ...
 $ start_lat         : num  41.9 41.9 41.9 41.9 41.9 ...
 $ start_lng         : num  -87.6 -87.6 -87.6 -87.6 -87.6 ...
 $ end_lat           : num  41.9 41.9 41.9 41.9 41.9 ...
 $ end_lng           : num  -87.6 -87.6 -87.6 -87.6 -87.6 ...
 $ member_casual     : chr  "member" "member" "member" "member" ...
> str(df_122022)
'data.frame':  181806 obs. of  13 variables:
 $ ride_id           : chr  "65DBD2F447EC51C2" "0C201AA7EA0EA1AD" "E0B148CCB358A49D" "54C5775D2B
 $ rideable_type     : chr  "electric_bike" "classic_bike" "electric_bike" "classic_bike" ...
 $ started_at        : chr  "2022-12-05 10:47:18" "2022-12-18 06:42:33" "2022-12-13 08:47:45" "2
 $ ended_at          : chr  "2022-12-05 10:56:34" "2022-12-18 07:08:44" "2022-12-13 08:59:51" "2
 $ start_station_name: chr  "Clifton Ave & Armitage Ave" "Broadway & Belmont Ave" "Sangamon St &
 $ start_station_id  : chr  "TA1307000163" "13277" "TA1306000015" "KA1503000038" ...
 $ end_station_name  : chr  "Sedgwick St & Webster Ave" "Sedgwick St & Webster Ave" "St. Clair S
 $ end_station_id    : chr  "13191" "13191" "13016" "13134" ...
 $ start_lat         : num  41.9 41.9 41.9 41.8 41.9 ...
 $ start_lng         : num  -87.7 -87.6 -87.7 -87.6 -87.7 ...
 $ end_lat           : num  41.9 41.9 41.9 41.9 41.9 ...
 $ end_lng           : num  -87.6 -87.6 -87.6 -87.7 -87.7 ...
 $ member_casual     : chr  "member" "casual" "member" "member" ...
> str(df_032023)
'data.frame':  258678 obs. of  13 variables:
 $ ride_id           : chr  "6842AA605EE9FBB3" "F984267A75B99A8C" "FF7CF57CFE026D02" "6B61B91603
 $ rideable_type     : chr  "electric_bike" "electric_bike" "classic_bike" "classic_bike" ...
 $ started_at        : chr  "2023-03-16 08:20:34" "2023-03-04 14:07:06" "2023-03-31 12:28:09" "2
 $ ended_at          : chr  "2023-03-16 08:22:52" "2023-03-04 14:15:31" "2023-03-31 12:38:47" "2
 $ start_station_name: chr  "Clark St & Armitage Ave" "Public Rack - Kedzie Ave & Argyle St" "Or
nes St & Kinzie St" ...
 $ start_station_id  : chr  "13146" "491" "620" "TA1306000003" ...
 $ end_station_name  : chr  "Larrabee St & Webster Ave" "" "Clark St & Randolph St" "Sheffield A
 $ end_station_id    : chr  "13193" "" "TA1305000030" "13154" ...
 $ start_lat         : num  41.9 42 41.9 41.9 41.9 ...
 $ start_lng         : num  -87.6 -87.7 -87.6 -87.6 -87.7 ...
 $ end_lat           : num  41.9 42 41.9 41.9 41.9 ...
 $ end_lng           : num  -87.6 -87.7 -87.6 -87.7 -87.7 ...
 $ member_casual     : chr  "member" "member" "member" "member" ...
> str(df_042023)
'data.frame':  426590 obs. of  13 variables:
 $ ride_id           : chr  "8FE8F7D9C10E88C7" "34E4ED3ADF1D821B" "5296BF07A2F77CB5" "40759916B7
 $ rideable_type     : chr  "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...
 $ started_at        : chr  "2023-04-02 08:37:28" "2023-04-19 11:29:02" "2023-04-19 08:41:22" "2
 $ ended_at          : chr  "2023-04-02 08:41:37" "2023-04-19 11:52:12" "2023-04-19 08:43:22" "2
```

```
 $ start_station_name: chr  "" "" "" "" ...
 $ start_station_id  : chr  "" "" "" "" ...
 $ end_station_name  : chr  "" "" "" "" ...
 $ end_station_id    : chr  "" "" "" "" ...
 $ start_lat         : num  41.8 41.9 41.9 41.9 41.9 ...
 $ start_lng         : num  -87.6 -87.7 -87.7 -87.7 -87.7 ...
 $ end_lat           : num  41.8 41.9 41.9 41.9 41.9 ...
 $ end_lng           : num  -87.6 -87.7 -87.7 -87.7 -87.6 ...
 $ member_casual     : chr  "member" "member" "member" "member" ...
> str(df_052023)
'data.frame':  604827 obs. of  13 variables:
 $ ride_id           : chr  "0D9FA920C3062031" "92485E5FB5888ACD" "FB144B3FC8300187" "DDEB93BC2C
 $ rideable_type     : chr  "electric_bike" "electric_bike" "electric_bike" "classic_bike" ...
 $ started_at        : chr  "2023-05-07 19:53:48" "2023-05-06 18:54:08" "2023-05-21 00:40:21" "2
 $ ended_at          : chr  "2023-05-07 19:58:32" "2023-05-06 19:03:35" "2023-05-21 00:44:36" "2
 $ start_station_name: chr  "Southport Ave & Belmont Ave" "Southport Ave & Belmont Ave" "Halsted
 $ start_station_id  : chr  "13229" "13229" "13162" "13196" ...
 $ end_station_name  : chr  "" "" "" "Damen Ave & Cortland St" ...
 $ end_station_id    : chr  "" "" "" "13133" ...
 $ start_lat         : num  41.9 41.9 41.9 41.9 42 ...
 $ start_lng         : num  -87.7 -87.7 -87.6 -87.7 -87.7 ...
 $ end_lat           : num  41.9 41.9 41.9 41.9 41.9 ...
 $ end_lng           : num  -87.7 -87.7 -87.7 -87.7 -87.7 ...
 $ member_casual     : chr  "member" "member" "member" "member" ...
> str(df_012023)
'data.frame':  190301 obs. of  13 variables:
 $ ride_id           : chr  "F96D5A74A3E41399" "13CB7EB698CEDB88" "BD88A2E670661CE5" "C90792D034
 $ rideable_type     : chr  "electric_bike" "classic_bike" "electric_bike" "classic_bike" ...
 $ started_at        : chr  "2023-01-21 20:05:42" "2023-01-10 15:37:36" "2023-01-02 07:51:57" "2
 $ ended_at          : chr  "2023-01-21 20:16:33" "2023-01-10 15:46:05" "2023-01-02 08:05:11" "2
 $ start_station_name: chr  "Lincoln Ave & Fullerton Ave" "Kimbark Ave & 53rd St" "Western Ave &
 $ start_station_id  : chr  "TA1309000058" "TA1309000037" "RP-005" "TA1309000037" ...
 $ end_station_name  : chr  "Hampden Ct & Diversey Ave" "Greenwood Ave & 47th St" "Valli Produce
...
 $ end_station_id    : chr  "202480.0" "TA1308000002" "599" "TA1308000002" ...
 $ start_lat         : num  41.9 41.8 42 41.8 41.8 ...
 $ start_lng         : num  -87.6 -87.6 -87.7 -87.6 -87.6 ...
 $ end_lat           : num  41.9 41.8 42 41.8 41.8 ...
 $ end_lng           : num  -87.6 -87.6 -87.7 -87.6 -87.6 ...
 $ member_casual     : chr  "member" "member" "casual" "member" ...
> str(df_022023)
'data.frame':  190445 obs. of  13 variables:
 $ ride_id           : chr  "CBCD0D7777F0E45F" "F3EC5FCE5FF39DE9" "E54C1F27FA9354FF" "3D561E04F7
 $ rideable_type     : chr  "classic_bike" "electric_bike" "classic_bike" "electric_bike" ...
 $ started_at        : chr  "2023-02-14 11:59:42" "2023-02-15 13:53:48" "2023-02-19 11:10:57" "2
 $ ended_at          : chr  "2023-02-14 12:13:38" "2023-02-15 13:59:08" "2023-02-19 11:35:01" "2
 $ start_station_name: chr  "Southport Ave & Clybourn Ave" "Clarendon Ave & Gordon Ter" "Southpo
ourn Ave" ...
 $ start_station_id  : chr  "TA1309000030" "13379" "TA1309000030" "TA1309000030" ...
 $ end_station_name  : chr  "Clark St & Schiller St" "Sheridan Rd & Lawrence Ave" "Aberdeen St &
...
 $ end_station_id    : chr  "TA1309000024" "TA1309000041" "13156" "TA1309000008" ...
 $ start_lat         : num  41.9 42 41.9 41.9 41.8 ...
 $ start_lng         : num  -87.7 -87.6 -87.7 -87.7 -87.6 ...
 $ end_lat           : num  41.9 42 41.9 41.9 41.8 ...
 $ end_lng           : num  -87.6 -87.7 -87.7 -87.6 -87.6 ...
 $ member_casual     : chr  "casual" "casual" "member" "member" ...
> # Merge all data frames into one large data frame using rbind()
> df_cyclistic <- rbind(df_062022, df_072022,
+                       df_082022, df_092022,
+                       df_102022, df_112022,
+                       df_122022, df_032023,
+                       df_042023, df_052023,
+                       df_012023, df_022023)
> # Recreate the df_cyclistic data frame with relevant data and columns for calculations
> df_cyclistic <- df_cyclistic %>%
+   select(ride_id, rideable_type, started_at, ended_at, member_casual)
> # Check if there are missing values in df_cyclistic
> missing_values <- df_cyclistic %>%
+   filter(ride_id == "" | rideable_type == "" |
+            started_at == "" |
+            ended_at == "" | "member_casual" == "")
> View(missing_values) # No data available in table
> # Check for duplicates in the dataframe
> distinct_values <- df_cyclistic %>%
+   distinct(ride_id, rideable_type, started_at, ended_at, member_casual,
+              keep_all = FALSE) %>%
+   filter(duplicated(df_cyclistic)) #No data available in table
> # Columns started_at and ended_at contain date and time information, so I change their format
> df_cyclistic$started_at <- as.POSIXct(df_cyclistic$started_at, format = "%Y-%m-%d %H:%M:%S")
```

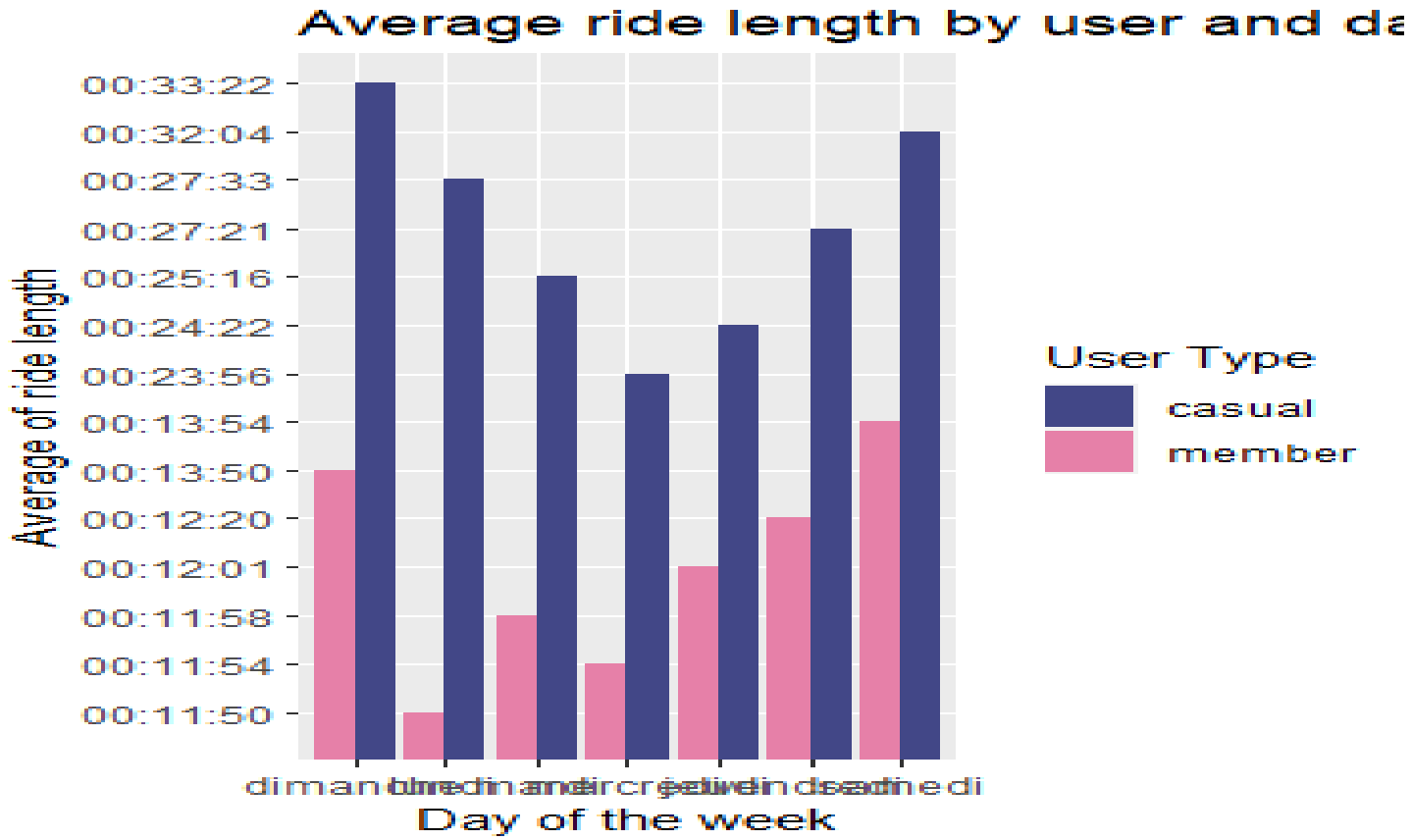
```
> df_cyclistic$ended_at <- as.POSIXct(df_cyclistic$ended_at, format = "%Y-%m-%d %H:%M:%S")
> str(df_cyclistic)
'data.frame':  5829030 obs. of  5 variables:
 $ ride_id      : chr  "600CFD130D0FD2A4" "F5E6B5C1682C6464" "B6EB6D27BAD771D2" "C9C320375DE1D5C
 $ rideable_type: chr  "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...
 $ started_at   : POSIXct, format: "2022-06-30 17:27:53" "2022-06-30 18:39:52" "2022-06-30 11:49
 $ ended_at     : POSIXct, format: "2022-06-30 17:35:15" "2022-06-30 18:47:28" "2022-06-30 12:02
 $ member_casual: chr  "casual" "casual" "casual" "casual" ...
> # Additional check: Are all values in the ride_id column of the same length?
> df_cyclistic %>%
+   select(ride_id) %>%
+   filter(nchar(ride_id) > 16 | nchar(ride_id) < 16) # 0 rows
[1] ride_id
<0 lignes> (ou 'row.names' de longueur nulle)
> # Additional check: Are all values in the member_casual column 'member' or 'casual'?
> factor(df_cyclistic$member_casual) # Levels: casual member
   [1] casual casual casual casual casual casual casual casual casual casual casual casual casua
  [19] casual casual casual casual casual casual casual casual casual casual casual casual casua
  [37] casual casual casual casual casual casual casual casual casual casual casual casual casua
  [55] casual casual casual casual casual casual casual casual casual casual casual casual casua
  [73] casual casual casual casual casual casual casual casual casual casual casual casual casua
  [91] casual casual casual casual casual casual casual member casual casual casual casual casua
 [109] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [127] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [145] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [163] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [181] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [199] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [217] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [235] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [253] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [271] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [289] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [307] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [325] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [343] casual casual casual casual member casual casual casual casual casual casual casual casua
 [361] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [379] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [397] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [415] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [433] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [451] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [469] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [487] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [505] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [523] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [541] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [559] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [577] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [595] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [613] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [631] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [649] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [667] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [685] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [703] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [721] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [739] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [757] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [775] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [793] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [811] casual casual casual casual casual casual casual casual casual member member member membe
 [829] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [847] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [865] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [883] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [901] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [919] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [937] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [955] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [973] casual casual casual casual casual casual casual casual casual casual casual casual casua
 [991] casual casual casual casual casual casual casual casual casual
 [ reached getOption("max.print") -- omitted 5828030 entries ]
Levels: casual member
> # Creating a new column ride_length that calculates the duration of the ride in seconds
> df_cyclistic <- df_cyclistic %>%
+   mutate(ride_length = hms(seconds_to_period(as.numeric(difftime(ended_at, started_at, units =
> # Finding incorrect values
> incorrect_data <- df_cyclistic %>%
+   filter(ride_length < 00:00:00)
```

```
> # Removing incorrect values
> df_cyclistic <- anti_join(df_cyclistic, incorrect_data, by="ride_id")
> # Creating a new column day_of_week that names the day when the ride started using the started
> df_cyclistic <- df_cyclistic %>%
+   mutate(day_of_week = wday(started_at, label = TRUE, abbr = FALSE))
> # Calculating the average ride length of all users and converting the numeric to char
> avg_of_ride_length <- df_cyclistic %>%
+   summarize(avg_of_ride_length = substr(
+     format(hms(seconds_to_period(mean(ride_length, na.rm = TRUE))),
+            format = "%H:%M:%S.%OS"), 1, 8))
> print(avg_of_ride_length)
  avg_of_ride_length
1           00:18:42
> # Calculate the average ride length by member_casual column and convert the numeric to char
> avg_of_ride_length_by_user <- df_cyclistic %>%
+   group_by(member_casual) %>%
+   summarize(avg_of_ride_length = substr(
+     format(hms(seconds_to_period(mean(ride_length, na.rm = TRUE))),
+            format = "%H:%M:%S.%OS"), 1, 8))
> print(avg_of_ride_length_by_user)
# A tibble: 2 × 2
  member_casual avg_of_ride_length
  <chr>         <chr>
1 casual        00:28:12
2 member        00:12:28
> # Calculate the average ride length by user type and day of the week
> avg_of_ride_length_by_day_of_week <- df_cyclistic %>%
+   group_by(member_casual, day_of_week) %>%
+   summarize(avg_of_ride_length_by_day_of_week = substr(
+     format(hms(seconds_to_period(mean(ride_length, na.rm = TRUE))),
+            format = "%H:%M:%S.%OS"), 1, 8))
`summarise()` has grouped output by 'member_casual'. You can override using the `.groups` argume
> # Rename columns in avg_of_ride_length_by_day_of_week
> colnames(avg_of_ride_length_by_day_of_week) <- c("User Type", "Day of Week", "Average Trip Len
> # Remove NA values from avg_of_ride_length_by_day_of_week
> avg_of_ride_length_by_day_of_week <- na.omit(avg_of_ride_length_by_day_of_week)
> # Total number of rides per user
> total_number_of_rides_by_user <- df_cyclistic %>%
+   group_by(member_casual) %>%
+   summarize(total_number_of_rides_by_user = n())
> # Total number of rides per user and day of the week
> total_number_of_rides_by_user_and_day <- df_cyclistic %>%
+   group_by(member_casual, day_of_week) %>%
+   summarize(total_number_of_rides_u_d = n())
`summarise()` has grouped output by 'member_casual'. You can override using the `.groups` argume
> # Removing NA values
> total_number_of_rides_by_user_and_day <- na.omit(total_number_of_rides_by_user_and_day)
> # Total number of rides per month
> total_number_of_rides_by_month <- df_cyclistic %>%
+   group_by(month = month(started_at)) %>%
+   summarize(total_number_of_rides_by_month = n())
> # Removing NA values
> total_number_of_rides_by_month <- na.omit(total_number_of_rides_by_month)
```

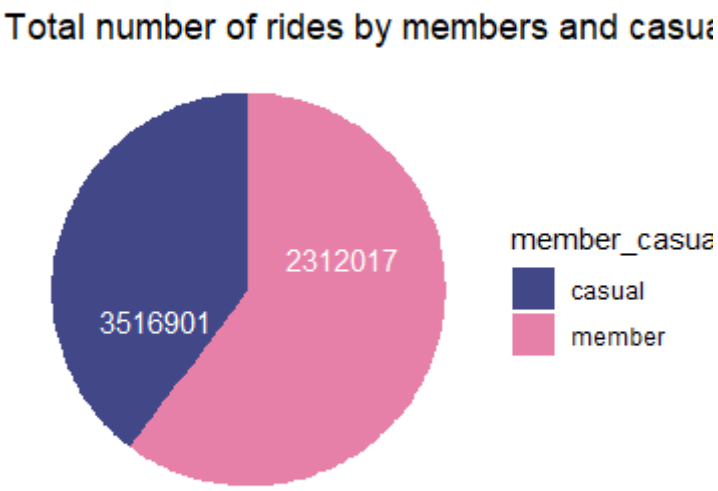
```
> # Visualize the average ride length by user type

> ggplot(data = avg_of_ride_length_by_user, aes(x = member_casual, y = avg_of_ride_length, fill
+   geom_bar(stat = "identity") +
+   scale_fill_manual(values = c("#424787", "#E680A8")) +
+   labs(title = "Average ride length by user", x = "User Type", y = "Average of ride length")
```



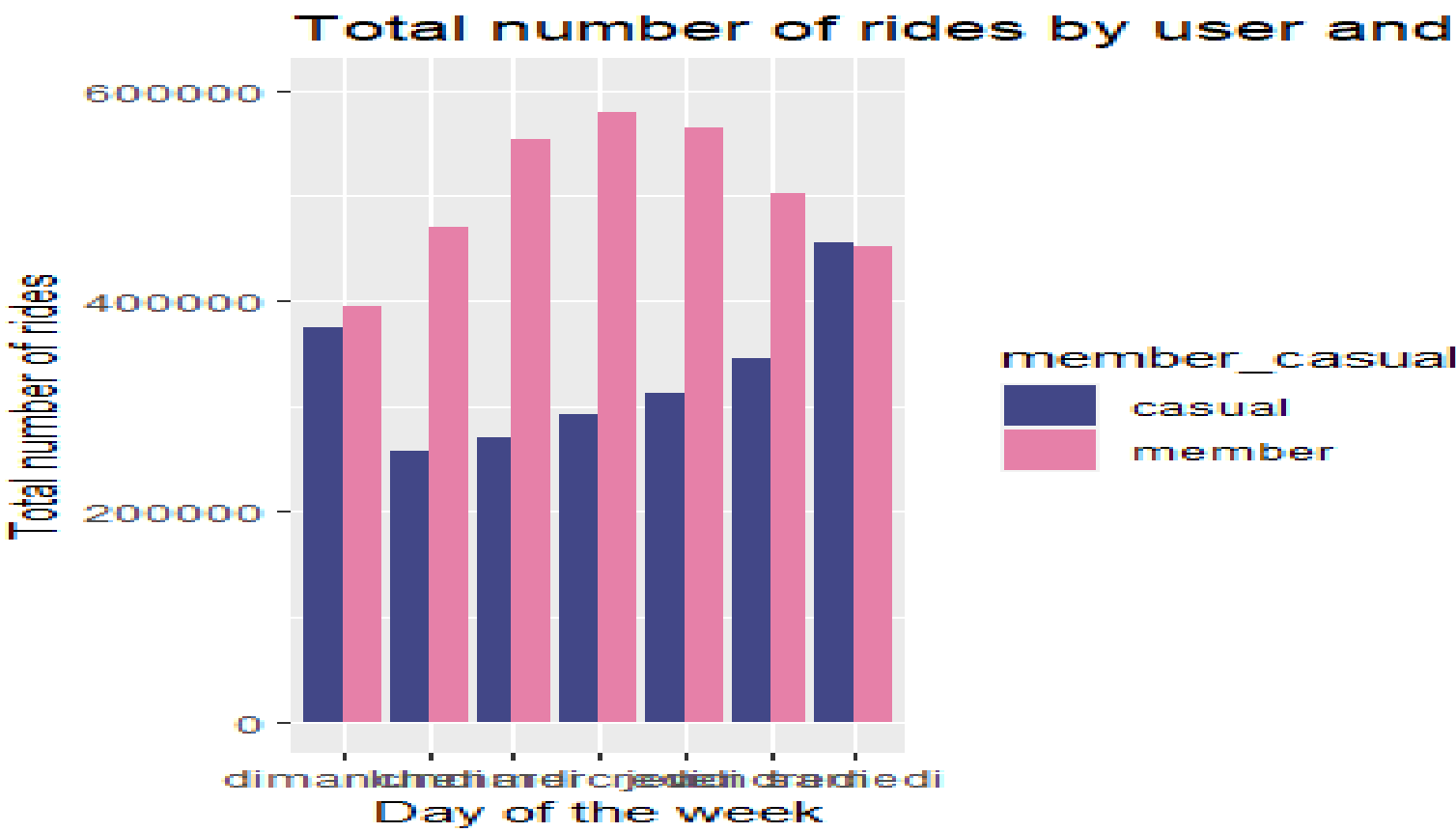

```
> # Visualize the average ride length by user type and day of the week in seconds
> ggplot(data = avg_of_ride_length_by_day_of_week, aes(x = `Day of Week`, y = `Average Trip Leng
+   geom_bar(position="dodge", stat = "identity") +
+   scale_fill_manual(values = c("#424787", "#E680A8")) +
+   labs(title = "Average ride length by user and day of the week", x = "Day of the week", y = "
```



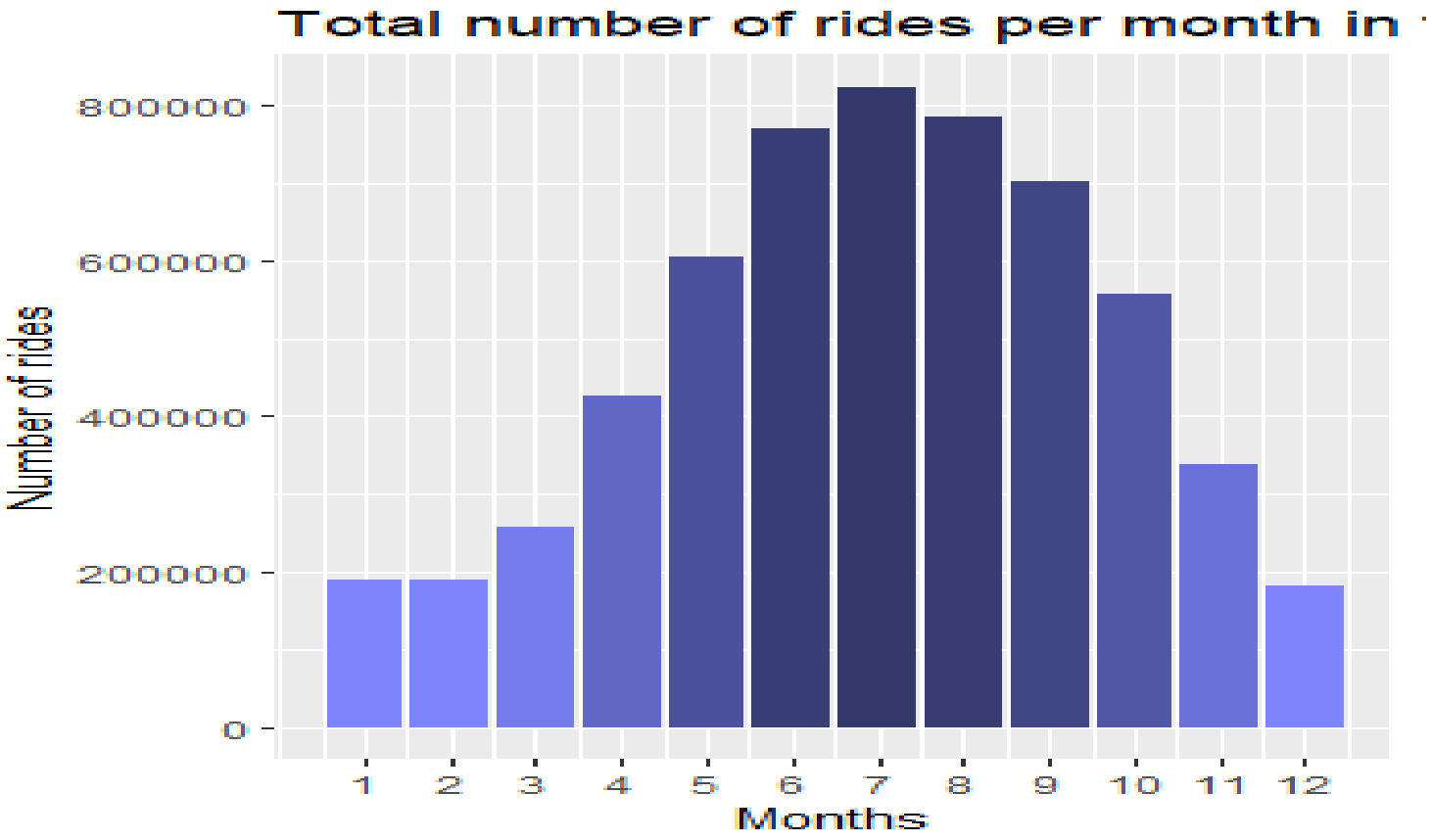
```
> # Visualization of the total number of rides per user, pie chart
> ggplot(data = total_number_of_rides_by_user, aes(x = "", y = total_number_of_rides_by_user, fi
+   geom_bar(width = 1, stat = "identity") +
+   scale_fill_manual(values = c("#424787", "#E680A8")) +
+   labs(title = "Total number of rides by members and casual users", x = "", y = "") +
+   geom_text(aes(y = total_number_of_rides_by_user/2 + c(0, cumsum(total_number_of_rides_by_use
+                 label = total_number_of_rides_by_user),
+             color = "white") +
+   coord_polar(theta = "y", start=0) +
+   theme_void()
```




```
> # Visualization of the total number of rides per user and day of the week
> ggplot(data = total_number_of_rides_by_user_and_day, aes(x = day_of_week, y = total_number_of_
+   geom_bar(position="dodge", stat = "identity") +
+   scale_fill_manual(values = c("#424787", "#E680A8")) +
+   labs(title="Total number of rides by user and day of the week", x="Day of the week", y="Tota
+   scale_y_continuous(limits = c(0, 600000))
```




```
> # Visualization of the total number of rides per month
> ggplot(data = total_number_of_rides_by_month, aes(x = month, y = total_number_of_rides_by_mont
+   geom_bar(stat="identity") +
+   scale_fill_gradient(low = "#7D86FF", high = "#34386B") +
+   labs(title="Total number of rides per month in the year 2022-2023", x="Months", y="Number of
+   scale_x_continuous(breaks = 1:12, labels = 1:12) +
+   theme(legend.position = "none")
```




```
>
```