Divvy Capstone

Lanai Toledo

2023-07-06

Google Data Analytics Capstone Project: How does a bike-share navigate speedy success?

About the Company

In 2016, Cyclistic launched a successful bike-share offering having a fleet of 5,824 bicycles that are tracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system at any time.

Members are riders who have an annual subscription while casual riders are users who are single-ride or full-day pass and do not have an annual membership plan.

The director of marketing is looking to maximize the number of annual memberships as they are more profitable than single-ride or full-day passes and wants to design marketing strategies to entice casual riders to become annual members.

The questions that need to be answered are:

How do annual members and casual riders use Cyclistic bikes differently?

Why would casual riders buy Cyclistic annual memberships?

How can Cyclistic use digital media to influence casual riders to become members?

Gather the Data

I will be using the public dataset located https://divvy-tripdata.s3.amazonaws.com/index.html . The data has been made available by Motivate International Inc. under this [license]: https://www.divvybikes.com/data-license-agreement .

Key Tasks Followed:

Downloaded data and locally stored copies.

Data is downloaded for the year 2022 and includes January through December.

The data is in CSV (comma-separated values) format, and there are a total of 13 columns.

Set Up Environment

Load packages in R to enable cleaning and data transformation

```
library(tidyverse)
```

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

```
library(lubridate)
library(ggplot2)
getwd()
```

```
## [1] "C:/Users/toled/OneDrive/Desktop/R"
```

Upload data files as CSV

```
jan <- read.csv("C:/Users/toled/OneDrive/Desktop/Case Study 1/2022 CSV/202201-divvy-tripdata.csv")
feb <- read.csv("C:/Users/toled/OneDrive/Desktop/Case Study 1/2022 CSV/202202-divvy-tripdata.csv")
mar <- read.csv("C:/Users/toled/OneDrive/Desktop/Case Study 1/2022 CSV/202203-divvy-tripdata.csv")
apr <- read.csv("C:/Users/toled/OneDrive/Desktop/Case Study 1/2022 CSV/202204-divvy-tripdata.csv")
may <- read.csv("C:/Users/toled/OneDrive/Desktop/Case Study 1/2022 CSV/202205-divvy-tripdata.csv")
jun <- read.csv("C:/Users/toled/OneDrive/Desktop/Case Study 1/2022 CSV/202206-divvy-tripdata.csv")
jul <- read.csv("C:/Users/toled/OneDrive/Desktop/Case Study 1/2022 CSV/202207-divvy-tripdata.csv")
aug <- read.csv("C:/Users/toled/OneDrive/Desktop/Case Study 1/2022 CSV/202208-divvy-tripdata.csv")
sep <- read.csv("C:/Users/toled/OneDrive/Desktop/Case Study 1/2022 CSV/202209-divvy-publictripdata.csv")
oct <- read.csv("C:/Users/toled/OneDrive/Desktop/Case Study 1/2022 CSV/202210-divvy-tripdata.csv")
nov <- read.csv("C:/Users/toled/OneDrive/Desktop/Case Study 1/2022 CSV/202211-divvy-tripdata.csv")
dec <- read.csv("C:/Users/toled/OneDrive/Desktop/Case Study 1/2022 CSV/202212-divvy-tripdata.csv")
```

Inspect data columns before combining into single dataframe

```
colnames(jan)
```

```
##  [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(feb)
```

```
##  [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(mar)
```

```
##  [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(apr)
```

```
##  [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(may)
```

```
##  [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(jun)
```

```
##  [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(jul)
```

```
##  [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(aug)
```

```
##  [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(sep)
```

```
##  [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(oct)
```

```
##  [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(nov)
```

```
##  [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(dec)
```

```
##  [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"
```

Inspect dataframes to make sure strucure is the same before combining

```r
str(jan)
```

```
## 'data.frame':    103770 obs. of  13 variables:
##  $ ride_id           : chr  "C2F7DD78E82EC875" "A6CF8980A652D272" "BD0F91DFF741C66D" "CBB80ED4191054
##  $ 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-
##  $ ended_at          : chr  "2022-01-13 12:02:44" "2022-01-10 08:46:17" "2022-01-25 04:58:01" "2022-
##  $ start_station_name: chr  "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave" "Sheffield Ave & Fu
##  $ start_station_id  : chr  "525" "525" "TA1306000016" "KA1504000151" ...
##  $ end_station_name  : chr  "Clark St & Touhy Ave" "Clark St & Touhy Ave" "Greenview Ave & Fullerton
##  $ 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" ...
```

```r
str(feb)
```

```
## 'data.frame':    115609 obs. of  13 variables:
##  $ ride_id           : chr  "E1E065E7ED285C02" "1602DCDC5B30FFE3" "BE7DD2AF4B55C4AF" "A1789BDF844412
##  $ 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-
##  $ ended_at          : chr  "2022-02-19 18:23:56" "2022-02-20 17:45:56" "2022-02-25 19:09:34" "2022-
##  $ start_station_name: chr  "State St & Randolph St" "Halsted St & Wrightwood Ave" "State St & Rando
##  $ start_station_id  : chr  "TA1305000029" "TA1309000061" "TA1305000029" "13235" ...
##  $ end_station_name  : chr  "Clark St & Lincoln Ave" "Southport Ave & Wrightwood Ave" "Canal St & Ada
##  $ 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" ...
```

```r
str(mar)
```

```
## 'data.frame':    284042 obs. of  13 variables:
##  $ ride_id           : chr  "47EC0A7F82E65D52" "8494861979B0F477" "EFE527AF80B66109" "9F446FD9DEE3F38
##  $ rideable_type     : chr  "classic_bike" "electric_bike" "classic_bike" "classic_bike" ...
##  $ started_at        : chr  "2022-03-21 13:45:01" "2022-03-16 09:37:16" "2022-03-23 19:52:02" "2022-
##  $ ended_at          : chr  "2022-03-21 13:51:18" "2022-03-16 09:43:34" "2022-03-23 19:54:48" "2022-
##  $ start_station_name: chr  "Wabash Ave & Wacker Pl" "Michigan Ave & Oak St" "Broadway & Berwyn Ave"
##  $ start_station_id  : chr  "TA1307000131" "13042" "13109" "TA1307000131" ...
##  $ end_station_name  : chr  "Kingsbury St & Kinzie St" "Orleans St & Chestnut St (NEXT Apts)" "Broadw
##  $ end_station_id    : chr  "KA1503000043" "620" "15578" "TA1305000025" ...
##  $ start_lat         : num  41.9 41.9 42 41.9 41.9 ...
##  $ start_lng         : num  -87.6 -87.6 -87.7 -87.6 -87.6 ...
##  $ end_lat           : num  41.9 41.9 42 41.9 41.9 ...
##  $ end_lng           : num  -87.6 -87.6 -87.7 -87.6 -87.7 ...
##  $ member_casual     : chr  "member" "member" "member" "member" ...
```

```r
str(apr)
```

```
## 'data.frame':    371249 obs. of  13 variables:
##  $ ride_id           : chr  "3564070EEFD12711" "0B820C7FCF22F489" "89EEEE32293F07FF" "84D4751AEB3188
##  $ rideable_type     : chr  "electric_bike" "classic_bike" "classic_bike" "classic_bike" ...
##  $ started_at        : chr  "2022-04-06 17:42:48" "2022-04-24 19:23:07" "2022-04-20 19:29:08" "2022-
##  $ ended_at          : chr  "2022-04-06 17:54:36" "2022-04-24 19:43:17" "2022-04-20 19:35:16" "2022-
##  $ start_station_name: chr  "Paulina St & Howard St" "Wentworth Ave & Cermak Rd" "Halsted St & Polk S
##  $ start_station_id  : chr  "515" "13075" "TA1307000121" "13075" ...
##  $ end_station_name  : chr  "University Library (NU)" "Green St & Madison St" "Green St & Madison St
##  $ end_station_id    : chr  "605" "TA1307000120" "TA1307000120" "KA1706005007" ...
##  $ start_lat         : num  42 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num  -87.7 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num  42.1 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num  -87.7 -87.6 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr  "member" "member" "member" "casual" ...
```

```r
str(may)
```

```
## 'data.frame':    634858 obs. of  13 variables:
##  $ ride_id           : chr  "EC2DE40644C6B0F4" "1C31AD03897EE385" "1542FBEC830415CF" "6FF59852924528
##  $ rideable_type     : chr  "classic_bike" "classic_bike" "classic_bike" "classic_bike" ...
##  $ started_at        : chr  "2022-05-23 23:06:58" "2022-05-11 08:53:28" "2022-05-26 18:36:28" "2022-
##  $ ended_at          : chr  "2022-05-23 23:40:19" "2022-05-11 09:31:22" "2022-05-26 18:58:18" "2022-
##  $ start_station_name: chr  "Wabash Ave & Grand Ave" "DuSable Lake Shore Dr & Monroe St" "Clinton St
##  $ start_station_id  : chr  "TA1307000117" "13300" "TA1305000032" "TA1305000032" ...
##  $ end_station_name  : chr  "Halsted St & Roscoe St" "Field Blvd & South Water St" "Wood St & Milwaul
##  $ end_station_id    : chr  "TA1309000025" "15534" "13221" "TA1305000030" ...
##  $ 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 -87.7 -87.6 -87.7 ...
##  $ member_casual     : chr  "member" "member" "member" "member" ...
```

```r
str(jun)
```

```
## 'data.frame':    769204 obs. of  13 variables:
##  $ ride_id           : chr  "600CFD130D0FD2A4" "F5E6B5C1682C6464" "B6EB6D27BAD771D2" "C9C320375DE1D5
##  $ 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" "2022-
##  $ ended_at          : chr  "2022-06-30 17:35:15" "2022-06-30 18:47:28" "2022-06-30 12:02:54" "2022-
##  $ 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(jul)
```

```
## 'data.frame':    823488 obs. of  13 variables:
##  $ ride_id           : chr  "954144C2F67B1932" "292E027607D218B6" "57765852588AD6E0" "B5B6BE443145901
##  $ 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" "2022-0
##  $ ended_at          : chr  "2022-07-05 08:24:32" "2022-07-26 12:55:31" "2022-07-03 14:06:32" "2022-0
##  $ start_station_name: chr  "Ashland Ave & Blackhawk St" "Buckingham Fountain (Temp)" "Buckingham Fou
##  $ start_station_id  : chr  "13224" "15541" "15541" "15541" ...
##  $ end_station_name  : chr  "Kingsbury St & Kinzie St" "Michigan Ave & 8th St" "Michigan Ave & 8th St
##  $ 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(aug)
```

```
## 'data.frame':    785932 obs. of  13 variables:
##  $ ride_id           : chr  "550CF7EFEAE0C618" "DAD198F405F9C5F5" "E6F2BC47B65CB7FD" "F597830181C2E13
##  $ 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" "2022-0
##  $ ended_at          : chr  "2022-08-07 21:41:46" "2022-08-08 14:53:23" "2022-08-08 15:40:34" "2022-0
##  $ 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(sep)
```

```
## 'data.frame':    701339 obs. of  13 variables:
##  $ ride_id           : chr  "5156990AC19CA285" "E12D4A16BF51C274" "A02B53CD7DB72DD7" "C82E05FEE872DF
##  $ 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" "2022-0
##  $ ended_at          : chr  "2022-09-01 08:39:05" "2022-09-01 17:14:45" "2022-09-01 17:16:12" "2022-0
##  $ 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(oct)
```

```
## 'data.frame':    558685 obs. of  13 variables:
##  $ ride_id           : chr  "A50255C1E17942AB" "DB692A70BD2DD4E3" "3C02727AAF60F873" "47E653FDC2D992
##  $ 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" "2022-
##  $ ended_at          : chr  "2022-10-14 17:19:39" "2022-10-01 16:49:06" "2022-10-19 19:03:30" "2022-
##  $ start_station_name: chr  "Noble St & Milwaukee Ave" "Damen Ave & Charleston St" "Hoyne Ave & Balm
##  $ start_station_id  : chr  "13290" "13288" "655" "KA1504000133" ...
##  $ end_station_name  : chr  "Larrabee St & Division St" "Damen Ave & Cullerton St" "Western Ave & Le
##  $ 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(nov)
```

```
## 'data.frame':    337735 obs. of  13 variables:
##  $ ride_id           : chr  "BCC66FC6FAB27CC7" "772AB67E902C180F" "585EAD07FDEC0152" "91C4E7ED3C262F
##  $ 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" "2022-
##  $ ended_at          : chr  "2022-11-10 06:31:27" "2022-11-04 07:46:25" "2022-11-21 17:34:36" "2022-
##  $ start_station_name: chr  "Canal St & Adams St" "Canal St & Adams St" "Indiana Ave & Roosevelt Rd"
##  $ 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 & Erie S
##  $ 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(dec)
```

```
## 'data.frame':    181806 obs. of  13 variables:
##  $ ride_id           : chr  "65DBD2F447EC51C2" "0C201AA7EA0EA1AD" "E0B148CCB358A49D" "54C5775D2B7C91
##  $ 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" "2022-
##  $ ended_at          : chr  "2022-12-05 10:56:34" "2022-12-18 07:08:44" "2022-12-13 08:59:51" "2022-
##  $ start_station_name: chr  "Clifton Ave & Armitage Ave" "Broadway & Belmont Ave" "Sangamon St & Lak
##  $ start_station_id  : chr  "TA1307000163" "13277" "TA1306000015" "KA1503000038" ...
##  $ end_station_name  : chr  "Sedgwick St & Webster Ave" "Sedgwick St & Webster Ave" "St. Clair St & E
##  $ 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" ...
```

Combine all months data into one dataframe

```
all_trips <- bind_rows(jan, feb, mar, apr, may, jun, jul, aug, sep, oct, nov, dec)
```

Inspect the new table that has been created

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

```
##  [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"
```

```
nrow(all_trips)  #How many rows are in data frame?
```

```
## [1] 5667717
```

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

```
## [1] 5667717      13
```

```
head(all_trips)  #See the first 6 rows of data frame.
```

```
##            ride_id rideable_type          started_at            ended_at
## 1 C2F7DD78E82EC875 electric_bike 2022-01-13 11:59:47 2022-01-13 12:02:44
## 2 A6CF8980A652D272 electric_bike 2022-01-10 08:41:56 2022-01-10 08:46:17
## 3 BD0F91DFF741C66D  classic_bike 2022-01-25 04:53:40 2022-01-25 04:58:01
## 4 CBB80ED419105406  classic_bike 2022-01-04 00:18:04 2022-01-04 00:33:00
## 5 DDC963BFDDA51EEA  classic_bike 2022-01-20 01:31:10 2022-01-20 01:37:12
## 6 A39C6F6CC0586C0B  classic_bike 2022-01-11 18:48:09 2022-01-11 18:51:31
##              start_station_name start_station_id              end_station_name
## 1      Glenwood Ave & Touhy Ave              525          Clark St & Touhy Ave
## 2      Glenwood Ave & Touhy Ave              525          Clark St & Touhy Ave
## 3 Sheffield Ave & Fullerton Ave     TA1306000016 Greenview Ave & Fullerton Ave
## 4      Clark St & Bryn Mawr Ave     KA1504000151     Paulina St & Montrose Ave
## 5   Michigan Ave & Jackson Blvd     TA1309000002        State St & Randolph St
## 6         Wood St & Chicago Ave              637       Honore St & Division St
##   end_station_id start_lat start_lng  end_lat   end_lng member_casual
## 1         RP-007  42.01280 -87.66591 42.01256 -87.67437        casual
## 2         RP-007  42.01276 -87.66597 42.01256 -87.67437        casual
## 3   TA1307000001  41.92560 -87.65371 41.92533 -87.66580        member
## 4   TA1309000021  41.98359 -87.66915 41.96151 -87.67139        casual
## 5   TA1305000029  41.87785 -87.62408 41.88462 -87.62783        member
## 6   TA1305000034  41.89563 -87.67207 41.90312 -87.67394        member
```

```r
str(all_trips)  #See list of columns and data types (numeric, character, etc)
```

```
## 'data.frame':    5667717 obs. of  13 variables:
##  $ ride_id           : chr  "C2F7DD78E82EC875" "A6CF8980A652D272" "BD0F91DFF741C66D" "CBB80ED4191054
##  $ 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-
##  $ ended_at          : chr  "2022-01-13 12:02:44" "2022-01-10 08:46:17" "2022-01-25 04:58:01" "2022-
##  $ start_station_name: chr  "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave" "Sheffield Ave & Fu
##  $ start_station_id  : chr  "525" "525" "TA1306000016" "KA1504000151" ...
##  $ end_station_name  : chr  "Clark St & Touhy Ave" "Clark St & Touhy Ave" "Greenview Ave & Fullerton
##  $ 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" ...
```

```r
summary(all_trips)
```

```
##    ride_id          rideable_type       started_at          ended_at
##  Length:5667717     Length:5667717     Length:5667717     Length:5667717
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  start_station_name start_station_id   end_station_name   end_station_id
##  Length:5667717     Length:5667717     Length:5667717     Length:5667717
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    start_lat       start_lng         end_lat         end_lng
##  Min.   :41.64   Min.   :-87.84   Min.   : 0.00   Min.   :-88.14
##  1st Qu.:41.88   1st Qu.:-87.66   1st Qu.:41.88   1st Qu.:-87.66
##  Median :41.90   Median :-87.64   Median :41.90   Median :-87.64
##  Mean   :41.90   Mean   :-87.65   Mean   :41.90   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63   3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :45.64   Max.   :-73.80   Max.   :42.37   Max.   :  0.00
##                                   NA's   :5858    NA's   :5858
##  member_casual
##  Length:5667717
##  Class :character
##  Mode  :character
##
##
##
##
```

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

This will allow us to aggregate ride data for each month, day, or year

```
all_trips$date <- as.Date(all_trips$started_at) #The default format is yyyy-mm-dd
all_trips$month <- format(as.Date(all_trips$date), "%m")
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 seconds)

```
all_trips$ride_length <- difftime(all_trips$ended_at,all_trips$started_at)
```

Review the structure of all the columns that were just created

```
str(all_trips)
```

```
## 'data.frame':    5667717 obs. of  19 variables:
##  $ ride_id           : chr  "C2F7DD78E82EC875" "A6CF8980A652D272" "BD0F91DFF741C66D" "CBB80ED4191054
##  $ 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-
##  $ ended_at          : chr  "2022-01-13 12:02:44" "2022-01-10 08:46:17" "2022-01-25 04:58:01" "2022-
##  $ start_station_name: chr  "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave" "Sheffield Ave & Fu
##  $ start_station_id  : chr  "525" "525" "TA1306000016" "KA1504000151" ...
##  $ end_station_name  : chr  "Clark St & Touhy Ave" "Clark St & Touhy Ave" "Greenview Ave & Fullerton
##  $ 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" ...
##  $ date              : Date, format: "2022-01-13" "2022-01-10" ...
##  $ month             : chr  "01" "01" "01" "01" ...
##  $ day               : chr  "13" "10" "25" "04" ...
##  $ year              : chr  "2022" "2022" "2022" "2022" ...
##  $ day_of_week       : chr  "Thursday" "Monday" "Tuesday" "Tuesday" ...
##  $ ride_length       : 'difftime' num  177 261 261 896 ...
##   ..- attr(*, "units")= chr "secs"
```

Convert "ride_length" to numeric so we can run calculations on the data

```
is.numeric(all_trips$ride_length)
```

```
## [1] FALSE
```

```r
all_trips$ride_length <- as.numeric(as.character(all_trips$ride_length))
is.numeric(all_trips$ride_length)
```

```
## [1] TRUE
```

Ensure all "bad" data is removed (bikes taken out of docks for inspection and neg ride_length)

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

Descriptive analysis on ride_length (all figures in seconds)

```r
mean(all_trips_v2$ride_length) #straight average (total ride length / rides)
```

```
## [1] 1166.794
```

```r
median(all_trips_v2$ride_length) #midpoint number in the ascending array of ride lengths
```

```
## [1] 617
```

```r
max(all_trips_v2$ride_length) #longest ride
```

```
## [1] 2483235
```

```r
min(all_trips_v2$ride_length) #shortest ride
```

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

```r
summary(all_trips_v2$ride_length)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0     349     617    1167    1108 2483235
```

Compare members and casual users

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

```
##   all_trips_v2$member_casual all_trips_v2$ride_length
## 1                     casual                1748.8022
## 2                     member                 762.8632
```

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

```
##   all_trips_v2$member_casual all_trips_v2$ride_length
## 1                     casual                      780
## 2                     member                      530
```

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

```
##   all_trips_v2$member_casual all_trips_v2$ride_length
## 1                     casual                  2483235
## 2                     member                    89998
```

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

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

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

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

```
##    all_trips_v2$member_casual all_trips_v2$day_of_week all_trips_v2$ride_length
## 1                      casual                   Friday                1682.6659
## 2                      member                   Friday                 751.8463
## 3                      casual                   Monday                1751.2287
## 4                      member                   Monday                 736.2065
## 5                      casual                 Saturday                1956.7994
## 6                      member                 Saturday                 848.3793
## 7                      casual                   Sunday                2043.8469
## 8                      member                   Sunday                 842.0978
## 9                      casual                 Thursday                1532.9450
## 10                     member                 Thursday                 737.5637
## 11                     casual                  Tuesday                1549.3584
## 12                     member                  Tuesday                 727.7694
## 13                     casual                Wednesday                1485.0376
## 14                     member                Wednesday                 726.2934
```

Reorder the days of the week

```
all_trips_v2$day_of_week <- ordered(all_trips_v2$day_of_week, levels=c("Sunday", "Monday", "Tuesday", "
```

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

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

```
##    all_trips_v2$member_casual all_trips_v2$day_of_week all_trips_v2$ride_length
## 1                      casual                   Sunday                2043.8469
## 2                      member                   Sunday                 842.0978
## 3                      casual                   Monday                1751.2287
## 4                      member                   Monday                 736.2065
```

```
## 5                    casual              Tuesday              1549.3584
## 6                    member              Tuesday               727.7694
## 7                    casual            Wednesday              1485.0376
## 8                    member            Wednesday               726.2934
## 9                    casual             Thursday              1532.9450
## 10                   member             Thursday               737.5637
## 11                   casual               Friday              1682.6659
## 12                   member               Friday               751.8463
## 13                   casual             Saturday              1956.7994
## 14                   member             Saturday               848.3793
```

Analyze ridership data by type and weekday

```
all_trips_v2 %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n()
            ,average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday)
```

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

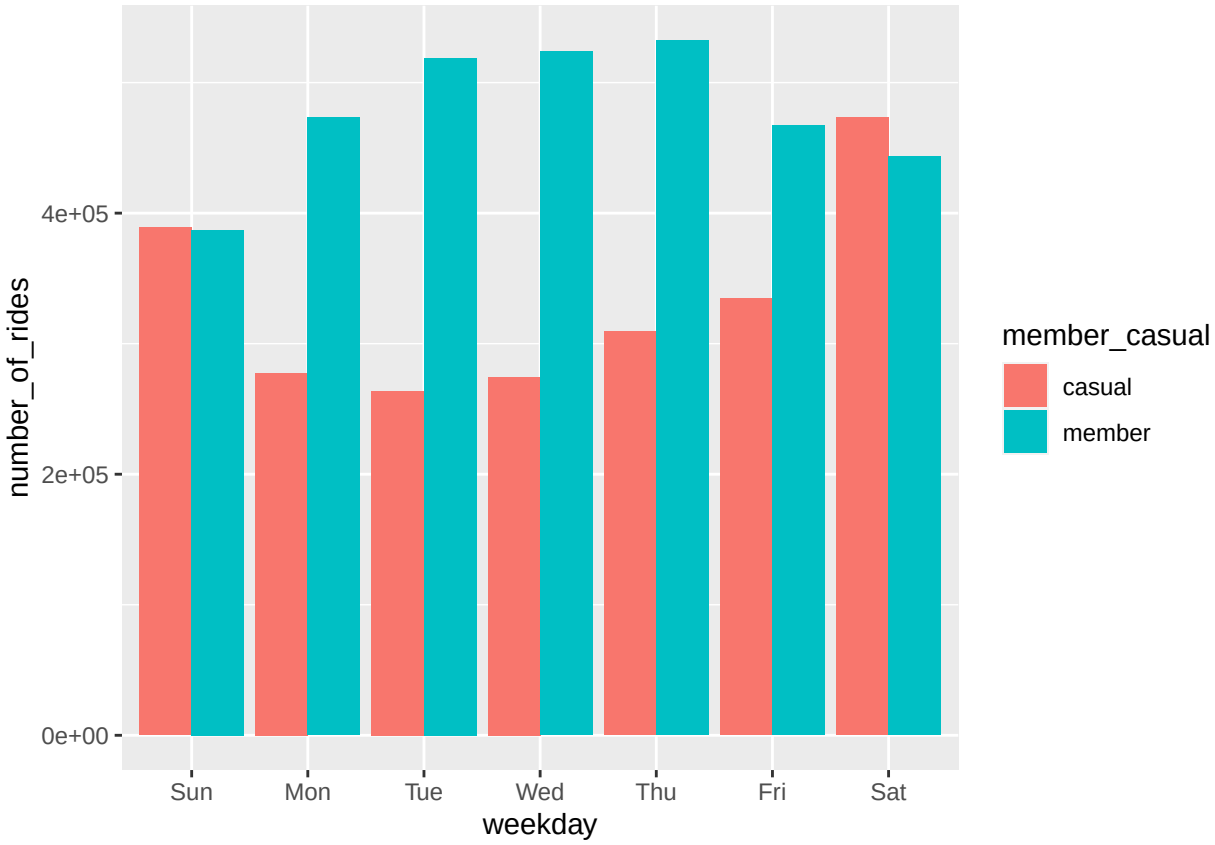
```
## # A tibble: 14 x 4
## # Groups:   member_casual [2]
##    member_casual weekday number_of_rides average_duration
##    <chr>         <ord>             <int>            <dbl>
##  1 casual        Sun              389011            2044.
##  2 casual        Mon              277671            1751.
##  3 casual        Tue              263731            1549.
##  4 casual        Wed              274354            1485.
##  5 casual        Thu              309327            1533.
##  6 casual        Fri              334698            1683.
##  7 casual        Sat              473185            1957.
##  8 member        Sun              387208             842.
##  9 member        Mon              473335             736.
## 10 member        Tue              518618             728.
## 11 member        Wed              523867             726.
## 12 member        Thu              532255             738.
## 13 member        Fri              467083             752.
## 14 member        Sat              443274             848.
```

Visualize the number of rides by rider type

```
all_trips_v2 %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n()
            ,average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday)  %>%
```

```
  ggplot(aes(x = weekday, y = number_of_rides, fill = member_casual)) +
  geom_col(position = "dodge")
```

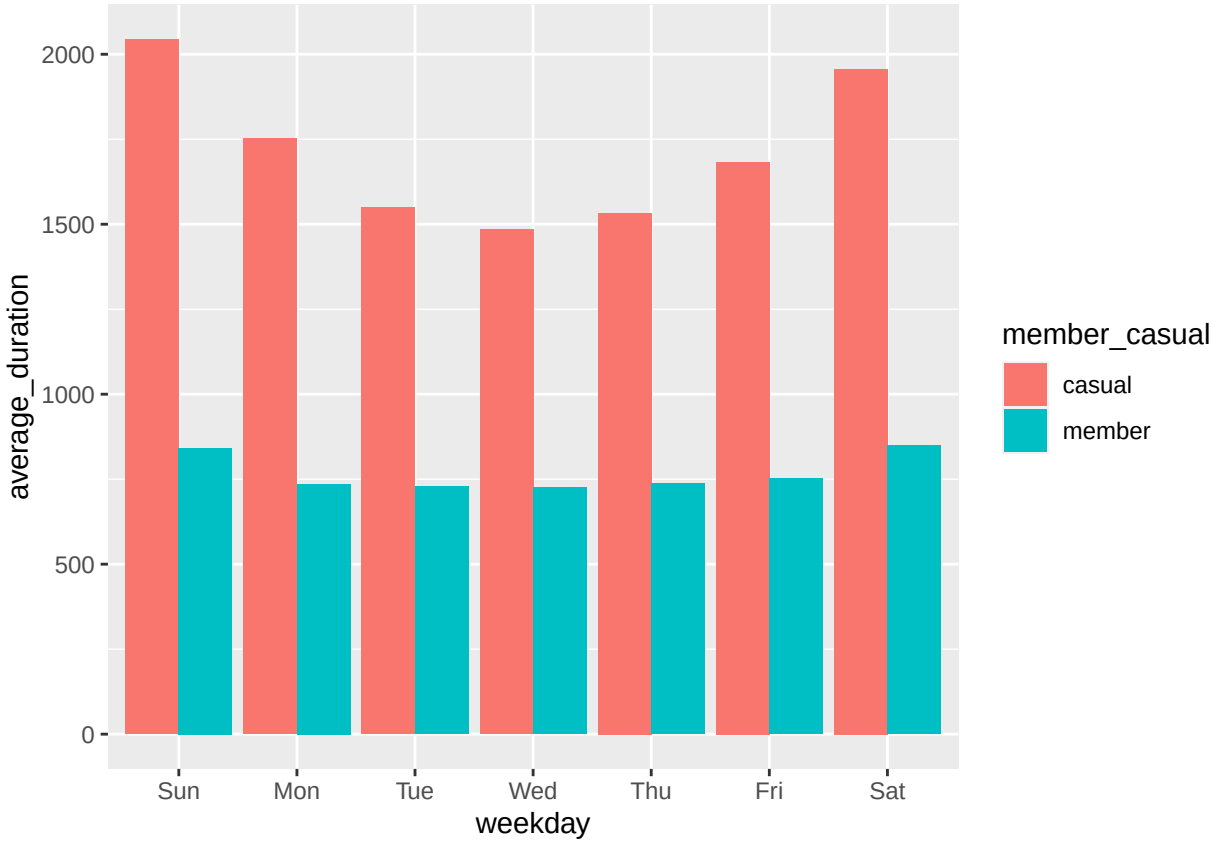
```
## `summarise()` has grouped output by 'member_casual'. You can override using the
## `.groups` argument.
```



Create a visualization for average duration

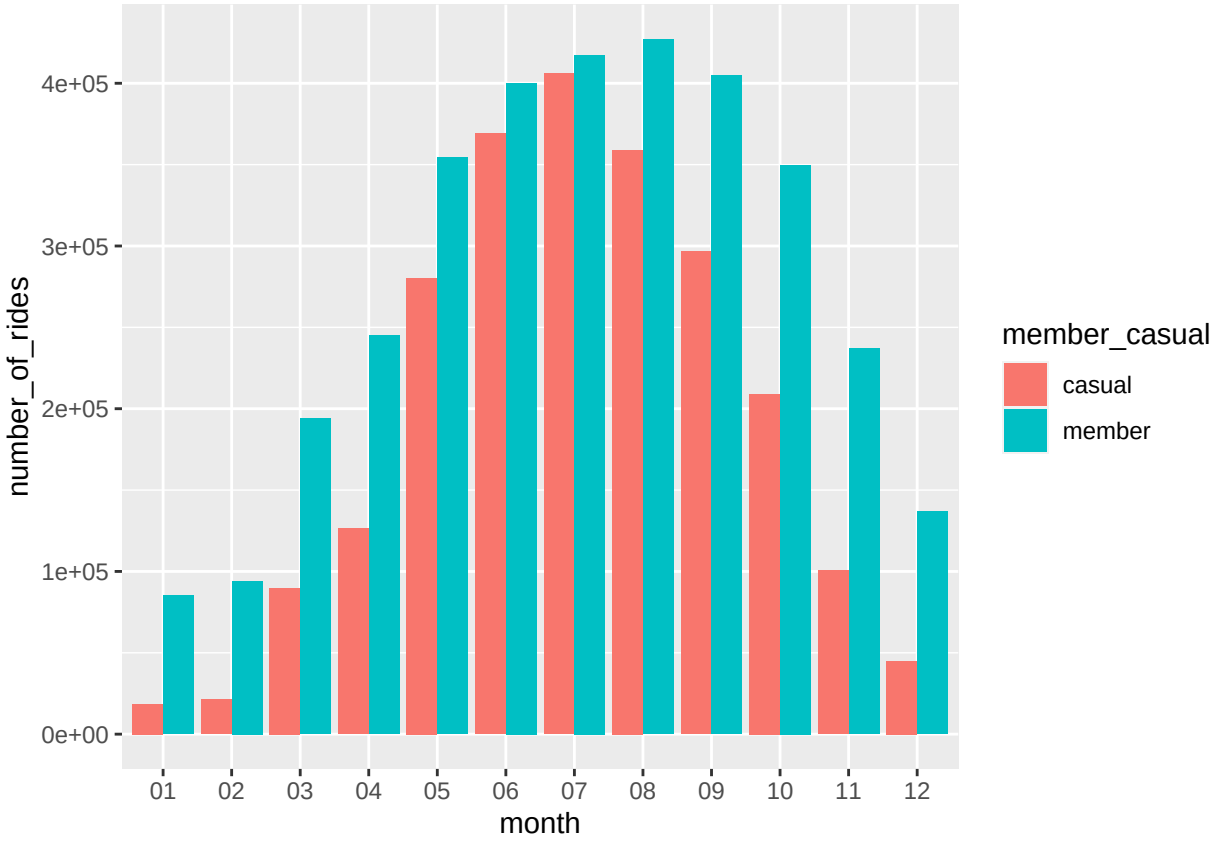
```
all_trips_v2 %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n()
            ,average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday)  %>%
  ggplot(aes(x = weekday, y = average_duration, fill = member_casual)) +
  geom_col(position = "dodge")
```

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




```
all_trips_v2 %>%
  group_by(member_casual, month)%>%
  summarise(number_of_rides = n()
            ,average_duration = mean(ride_length)) %>%
  ggplot(aes(x = month, y = number_of_rides, fill = member_casual)) +
  geom_col(position = "dodge")
```

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



Create a csv file that we can be further analyzed in Tableau

Summary of Analysis

Casual riders tend to take fewer trips than members.

Casual riders tend to take longer trips than members.

Casual riders take more rides on weekends than during the week.

Casual ridership is higest during the summer months.

Recommendations

Offer discounts on annual membership during spring and summer months to entice causl riders into becoming annual members.

Increasing weekend single ride and singe day pricing to incentivize membership.

Increasing cost for trips lasting more than a certain time for non-members to incentivize membership.