Google Data Analytics Cyclistic Bike Share Data

Aleny Mingear

2024-08-05

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
```

```
## -- 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(conflicted)
library(readr)
conflict_prefer("filter", "dplyr")
```

```
## [conflicted] Will prefer dplyr::filter over any other package.
```

```
conflict_prefer("lag", "dplyr")
```

```
## [conflicted] Will prefer dplyr::lag over any other package.
```

Step 1: Collect Data

```
q1_2019<-read_csv("Divvy_Trips_2019_Q1.csv")
```

```
## Rows: 365069 Columns: 12
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (6): start_time, end_time, from_station_name, to_station_name, usertype,...
## dbl (5): trip_id, bikeid, from_station_id, to_station_id, birthyear
## num (1): tripduration
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
q1_2020<-read_csv("Divvy_Trips_2020_Q1.csv")
```

```
## Rows: 426887 Columns: 13
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (7): ride_id, rideable_type, started_at, ended_at, start_station_name, e...
## dbl (6): start_station_id, end_station_id, start_lat, start_lng, end_lat, en...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Step 2: Wrangle Data and Combine into a single file

```
# compare column names
colnames(q1_2019)
```

```
##  [1] "trip_id"           "start_time"        "end_time"
##  [4] "bikeid"            "tripduration"      "from_station_id"
##  [7] "from_station_name" "to_station_id"     "to_station_name"
## [10] "usertype"          "gender"            "birthyear"
```

```
colnames(q1_2020)
```

```
##  [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"
```

```
# rename columns to make consistent
(q1_2019 <- rename(q1_2019
,ride_id = trip_id
,rideable_type = bikeid
,started_at = start_time
,ended_at = end_time
,start_station_name = from_station_name
,start_station_id = from_station_id
,end_station_name = to_station_name
,end_station_id = to_station_id
,member_casual = usertype
))
```

```
## # A tibble: 365,069 x 12
##     ride_id started_at    ended_at   rideable_type tripduration start_station_id
##       <dbl> <chr>         <chr>              <dbl>        <dbl>            <dbl>
##  1 21742443 1/1/2019 0:04 1/1/2019 ~          2167          390              199
##  2 21742444 1/1/2019 0:08 1/1/2019 ~          4386          441               44
##  3 21742445 1/1/2019 0:13 1/1/2019 ~          1524          829               15
##  4 21742446 1/1/2019 0:13 1/1/2019 ~           252         1783              123
##  5 21742447 1/1/2019 0:14 1/1/2019 ~          1170          364              173
##  6 21742448 1/1/2019 0:15 1/1/2019 ~          2437          216               98
##  7 21742449 1/1/2019 0:16 1/1/2019 ~          2708          177               98
##  8 21742450 1/1/2019 0:18 1/1/2019 ~          2796          100              211
##  9 21742451 1/1/2019 0:18 1/1/2019 ~          6205         1727              150
## 10 21742452 1/1/2019 0:19 1/1/2019 ~          3939          336              268
## # i 365,059 more rows
## # i 6 more variables: start_station_name <chr>, end_station_id <dbl>,
## #   end_station_name <chr>, member_casual <chr>, gender <chr>, birthyear <dbl>
```

```r
# inspect data frames
str(q1_2019)
```

```
## spc_tbl_ [365,069 x 12] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : num [1:365069] 21742443 21742444 21742445 21742446 21742447 ...
##  $ started_at        : chr [1:365069] "1/1/2019 0:04" "1/1/2019 0:08" "1/1/2019 0:13" "1/1/2019 0:13
##  $ ended_at          : chr [1:365069] "1/1/2019 0:11" "1/1/2019 0:15" "1/1/2019 0:27" "1/1/2019 0:43
##  $ rideable_type     : num [1:365069] 2167 4386 1524 252 1170 ...
##  $ tripduration      : num [1:365069] 390 441 829 1783 364 ...
##  $ start_station_id  : num [1:365069] 199 44 15 123 173 98 98 211 150 268 ...
##  $ start_station_name: chr [1:365069] "Wabash Ave & Grand Ave" "State St & Randolph St" "Racine Ave &
##  $ end_station_id    : num [1:365069] 84 624 644 176 35 49 49 142 148 141 ...
##  $ end_station_name  : chr [1:365069] "Milwaukee Ave & Grand Ave" "Dearborn St & Van Buren St (*)" "W
##  $ member_casual     : chr [1:365069] "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...
##  $ gender            : chr [1:365069] "Male" "Female" "Female" "Male" ...
##  $ birthyear         : num [1:365069] 1989 1990 1994 1993 1994 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   trip_id = col_double(),
##   ..   start_time = col_character(),
##   ..   end_time = col_character(),
##   ..   bikeid = col_double(),
##   ..   tripduration = col_number(),
##   ..   from_station_id = col_double(),
##   ..   from_station_name = col_character(),
##   ..   to_station_id = col_double(),
##   ..   to_station_name = col_character(),
##   ..   usertype = col_character(),
##   ..   gender = col_character(),
##   ..   birthyear = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```r
str(q1_2020)
```

```
## spc_tbl_ [426,887 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:426887] "EACB19130B0CDA4A" "8FED874C809DC021" "789F3C21E472CA96" "C9A38
##  $ rideable_type     : chr [1:426887] "docked_bike" "docked_bike" "docked_bike" "docked_bike" ...
##  $ started_at        : chr [1:426887] "1/21/2020 20:06" "1/30/2020 14:22" "1/9/2020 19:29" "1/6/2020
##  $ ended_at          : chr [1:426887] "1/21/2020 20:14" "1/30/2020 14:26" "1/9/2020 19:32" "1/6/2020
##  $ start_station_name: chr [1:426887] "Western Ave & Leland Ave" "Clark St & Montrose Ave" "Broadway
##  $ start_station_id  : num [1:426887] 239 234 296 51 66 212 96 96 212 38 ...
##  $ end_station_name  : chr [1:426887] "Clark St & Leland Ave" "Southport Ave & Irving Park Rd" "Wilt
##  $ end_station_id    : num [1:426887] 326 318 117 24 212 96 212 212 96 100 ...
##  $ start_lat         : num [1:426887] 42 42 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:426887] -87.7 -87.7 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:426887] 42 42 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:426887] -87.7 -87.7 -87.7 -87.6 -87.6 ...
##  $ member_casual     : chr [1:426887] "member" "member" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_character(),
```

```
##   ..   ended_at = col_character(),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_double(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_double(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
# convert ride_id and rideable type to character to stack correctly
q1_2019 <- mutate(q1_2019, ride_id = as.character(ride_id)
,rideable_type = as.character(rideable_type))

# stack individual dataframes into one data frame
all_trips <- bind_rows(q1_2019, q1_2020)
```

```
# remove lat, long, birthyear, and gender fields to match data from 2020
all_trips <- all_trips %>%
select(-c(start_lat, start_lng, end_lat, end_lng, birthyear, gender, "tripduration"))
```

Step 3: Clean Up Data and Add Data to Prepare for analysis

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

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

```
nrow(all_trips) #how many rows in the data frame
```

```
## [1] 791956
```

```
dim(all_trips) # dimensions of data frame
```

```
## [1] 791956      9
```

```
head(all_trips) # see the first 6 rows
```

```
## # A tibble: 6 x 9
##   ride_id  started_at ended_at rideable_type start_station_id start_station_name
##   <chr>    <chr>      <chr>    <chr>                   <dbl> <chr>
## 1 21742443 1/1/2019 ~ 1/1/201~ 2167                      199 Wabash Ave & Gran~
## 2 21742444 1/1/2019 ~ 1/1/201~ 4386                       44 State St & Randol~
## 3 21742445 1/1/2019 ~ 1/1/201~ 1524                       15 Racine Ave & 18th~
## 4 21742446 1/1/2019 ~ 1/1/201~ 252                       123 California Ave & ~
## 5 21742447 1/1/2019 ~ 1/1/201~ 1170                      173 Mies van der Rohe~
## 6 21742448 1/1/2019 ~ 1/1/201~ 2437                       98 LaSalle St & Wash~
## # i 3 more variables: end_station_id <dbl>, end_station_name <chr>,
## #   member_casual <chr>
```

```
str(all_trips) # see list of columns and data types
```

```
## tibble [791,956 x 9] (S3: tbl_df/tbl/data.frame)
```

```
##  $ ride_id           : chr [1:791956] "21742443" "21742444" "21742445" "21742446" ...
##  $ started_at        : chr [1:791956] "1/1/2019 0:04" "1/1/2019 0:08" "1/1/2019 0:13" "1/1/2019 0:13
##  $ ended_at          : chr [1:791956] "1/1/2019 0:11" "1/1/2019 0:15" "1/1/2019 0:27" "1/1/2019 0:43
##  $ rideable_type     : chr [1:791956] "2167" "4386" "1524" "252" ...
##  $ start_station_id  : num [1:791956] 199 44 15 123 173 98 98 211 150 268 ...
##  $ start_station_name: chr [1:791956] "Wabash Ave & Grand Ave" "State St & Randolph St" "Racine Ave &
##  $ end_station_id    : num [1:791956] 84 624 644 176 35 49 49 142 148 141 ...
##  $ end_station_name  : chr [1:791956] "Milwaukee Ave & Grand Ave" "Dearborn St & Van Buren St (*)" "
##  $ member_casual     : chr [1:791956] "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...
```

```
summary(all_trips) # statistical summary of data
```

```
##    ride_id           started_at          ended_at         rideable_type
##  Length:791956      Length:791956      Length:791956      Length:791956
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  start_station_id start_station_name end_station_id  end_station_name
##  Min.   :  2.0    Length:791956      Min.   :  2.0   Length:791956
##  1st Qu.: 77.0    Class :character   1st Qu.: 77.0   Class :character
##  Median :174.0    Mode  :character   Median :174.0   Mode  :character
##  Mean   :204.4                       Mean   :204.4
##  3rd Qu.:291.0                       3rd Qu.:291.0
##  Max.   :675.0                       Max.   :675.0
##                                      NA's   :1
##  member_casual
##  Length:791956
##  Class :character
##  Mode  :character
##
##
##
##
```

```
table(all_trips$member_casual)
```

```
##
##     casual   Customer     member Subscriber
##      48480      23163     378407     341906
```

```
# reassign to 2020 Labels
all_trips <- all_trips %>%
mutate(member_casual = recode(member_casual
,"Subscriber" = "member"
,"Customer" = "casual"))
```

```
table(all_trips$member_casual)
```

```
##
## casual member
##  71643 720313
```

```
# add columns that list date, month, day, and year of each ride
all_trips <- all_trips %>%
```

```
mutate(member_casual = recode(member_casual
,"Subscriber" = "member"
,"Customer" = "casual"))
table(all_trips$member_casual)
```

```
##
## casual member
##  71643 720313
```

```
all_trips$started_at <- as.Date(all_trips$started_at, "%m/%d/%Y")
```

```
all_trips$ended_at <- as.Date(all_trips$ended_at, "%m/%d/%Y")
```

```
all_trips$date <- as.Date(all_trips$started_at)
```

```
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")
str(all_trips)
```

```
## tibble [791,956 x 14] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:791956] "21742443" "21742444" "21742445" "21742446" ...
##  $ started_at        : Date[1:791956], format: "2019-01-01" "2019-01-01" ...
##  $ ended_at          : Date[1:791956], format: "2019-01-01" "2019-01-01" ...
##  $ rideable_type     : chr [1:791956] "2167" "4386" "1524" "252" ...
##  $ start_station_id  : num [1:791956] 199 44 15 123 173 98 98 211 150 268 ...
##  $ start_station_name: chr [1:791956] "Wabash Ave & Grand Ave" "State St & Randolph St" "Racine Ave &
##  $ end_station_id    : num [1:791956] 84 624 644 176 35 49 49 142 148 141 ...
##  $ end_station_name  : chr [1:791956] "Milwaukee Ave & Grand Ave" "Dearborn St & Van Buren St (*)" "W
##  $ member_casual     : chr [1:791956] "member" "member" "member" "member" ...
##  $ date              : Date[1:791956], format: "2019-01-01" "2019-01-01" ...
##  $ month             : chr [1:791956] "01" "01" "01" "01" ...
##  $ day               : chr [1:791956] "01" "01" "01" "01" ...
##  $ year              : chr [1:791956] "2019" "2019" "2019" "2019" ...
##  $ day_of_week       : chr [1:791956] "Tuesday" "Tuesday" "Tuesday" "Tuesday" ...
```

```
# Add ride_length calculation to all_trips in seconds
all_trips$ride_length <- difftime(all_trips$ended_at,all_trips$started_at)
```

```
str(all_trips)
```

```
## tibble [791,956 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:791956] "21742443" "21742444" "21742445" "21742446" ...
##  $ started_at        : Date[1:791956], format: "2019-01-01" "2019-01-01" ...
##  $ ended_at          : Date[1:791956], format: "2019-01-01" "2019-01-01" ...
##  $ rideable_type     : chr [1:791956] "2167" "4386" "1524" "252" ...
##  $ start_station_id  : num [1:791956] 199 44 15 123 173 98 98 211 150 268 ...
##  $ start_station_name: chr [1:791956] "Wabash Ave & Grand Ave" "State St & Randolph St" "Racine Ave &
##  $ end_station_id    : num [1:791956] 84 624 644 176 35 49 49 142 148 141 ...
##  $ end_station_name  : chr [1:791956] "Milwaukee Ave & Grand Ave" "Dearborn St & Van Buren St (*)" "W
##  $ member_casual     : chr [1:791956] "member" "member" "member" "member" ...
##  $ date              : Date[1:791956], format: "2019-01-01" "2019-01-01" ...
##  $ month             : chr [1:791956] "01" "01" "01" "01" ...
##  $ day               : chr [1:791956] "01" "01" "01" "01" ...
```

```
##  $ year              : chr [1:791956] "2019" "2019" "2019" "2019" ...
##  $ day_of_week       : chr [1:791956] "Tuesday" "Tuesday" "Tuesday" "Tuesday" ...
##  $ ride_length       : 'difftime' num [1:791956] 0 0 0 0 ...
##   ..- attr(*, "units")= chr "secs"
```

```
str(all_trips)
```

```
## tibble [791,956 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:791956] "21742443" "21742444" "21742445" "21742446" ...
##  $ started_at        : Date[1:791956], format: "2019-01-01" "2019-01-01" ...
##  $ ended_at          : Date[1:791956], format: "2019-01-01" "2019-01-01" ...
##  $ rideable_type     : chr [1:791956] "2167" "4386" "1524" "252" ...
##  $ start_station_id  : num [1:791956] 199 44 15 123 173 98 98 211 150 268 ...
##  $ start_station_name: chr [1:791956] "Wabash Ave & Grand Ave" "State St & Randolph St" "Racine Ave &
##  $ end_station_id    : num [1:791956] 84 624 644 176 35 49 49 142 148 141 ...
##  $ end_station_name  : chr [1:791956] "Milwaukee Ave & Grand Ave" "Dearborn St & Van Buren St (*)" "W
##  $ member_casual     : chr [1:791956] "member" "member" "member" "member" ...
##  $ date              : Date[1:791956], format: "2019-01-01" "2019-01-01" ...
##  $ month             : chr [1:791956] "01" "01" "01" "01" ...
##  $ day               : chr [1:791956] "01" "01" "01" "01" ...
##  $ year              : chr [1:791956] "2019" "2019" "2019" "2019" ...
##  $ day_of_week       : chr [1:791956] "Tuesday" "Tuesday" "Tuesday" "Tuesday" ...
##  $ ride_length       : 'difftime' num [1:791956] 0 0 0 0 ...
##   ..- attr(*, "units")= chr "secs"
```

```
# convert ride_length from factor to numeric
is.factor(all_trips$ride_length)
```

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

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

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

```
# remove bad data and create new data frame
all_trips_v2 <- all_trips[!(all_trips$start_station_name == "HQ QR" | all_trips$ride_length<0),]
```

Step 4 Conduct Descriptive Analysis (all figures in seconds)

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

```
## [1] 532.7453
```

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

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

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

```
## [1] 10627200
```

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

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

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

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
```

```
##        0        0        0      533        0 10627200
```

```
# Compare members and casual users
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                3861.9503
## 2                     member                 219.0251
```

```
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                        0
## 2                     member                        0
```

```
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                10627200
## 2                     member                 6134400
```

```
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                5078.5474
## 2                      member                   Friday                 248.3190
## 3                      casual                   Monday                3477.0166
## 4                      member                   Monday                 237.8484
## 5                      casual                 Saturday                3501.4028
## 6                      member                 Saturday                 475.5323
## 7                      casual                   Sunday                3117.4780
## 8                      member                   Sunday                 338.7278
## 9                      casual                 Thursday                7168.7701
## 10                     member                 Thursday                 135.2285
## 11                     casual                  Tuesday                3202.6262
## 12                     member                  Tuesday                 174.8605
## 13                     casual                Wednesday                2865.0195
## 14                     member                Wednesday                 122.6165
```

```
# Correcting Days of the Week so they are in order
all_trips_v2$day_of_week <- ordered(all_trips_v2$day_of_week, levels=c("Sunday", "Monday",
"Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))
```

```
# average ride time by 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                3117.4780
## 2                      member                   Sunday                 338.7278
## 3                      casual                   Monday                3477.0166
```

```
## 4                    member                Monday                237.8484
## 5                    casual               Tuesday               3202.6262
## 6                    member               Tuesday                174.8605
## 7                    casual             Wednesday               2865.0195
## 8                    member             Wednesday                122.6165
## 9                    casual              Thursday               7168.7701
## 10                   member              Thursday                135.2285
## 11                   casual                Friday               5078.5474
## 12                   member                Friday                248.3190
## 13                   casual              Saturday               3501.4028
## 14                   member              Saturday                475.5323
```

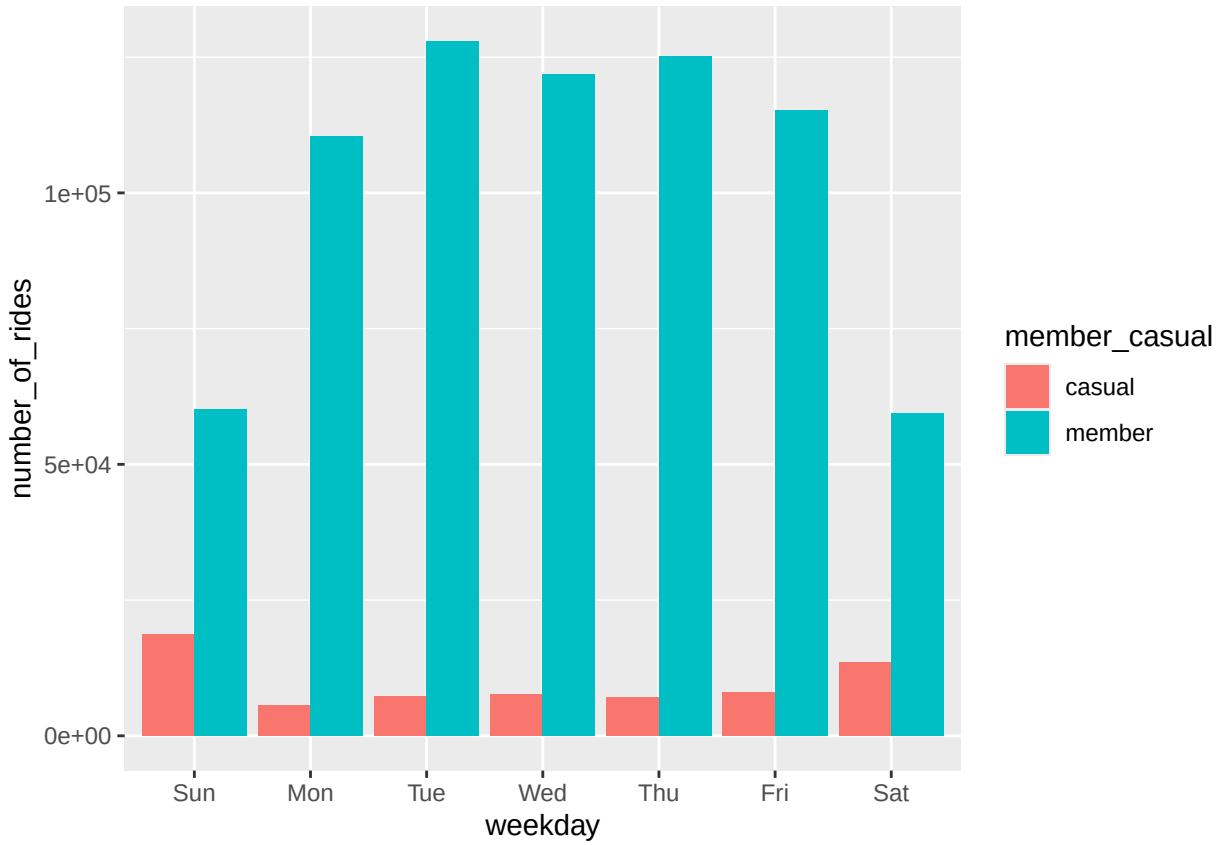
```r
# analyze ridership data by type and weekday
all_trips_v2 %>%
mutate(weekday = wday(started_at, label = TRUE)) %>%
group_by(member_casual, weekday) %>%
summarise(number_of_rides = n()
,average_duration = mean(ride_length)) %>%
arrange(member_casual, weekday)
```

```
## `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               18652            3117.
##  2 casual        Mon                5591            3477.
##  3 casual        Tue                7311            3203.
##  4 casual        Wed                7690            2865.
##  5 casual        Thu                7147            7169.
##  6 casual        Fri                8013            5079.
##  7 casual        Sat               13473            3501.
##  8 member        Sun               60197             339.
##  9 member        Mon              110430             238.
## 10 member        Tue              127974             175.
## 11 member        Wed              121902             123.
## 12 member        Thu              125228             135.
## 13 member        Fri              115168             248.
## 14 member        Sat               59413             476.
```

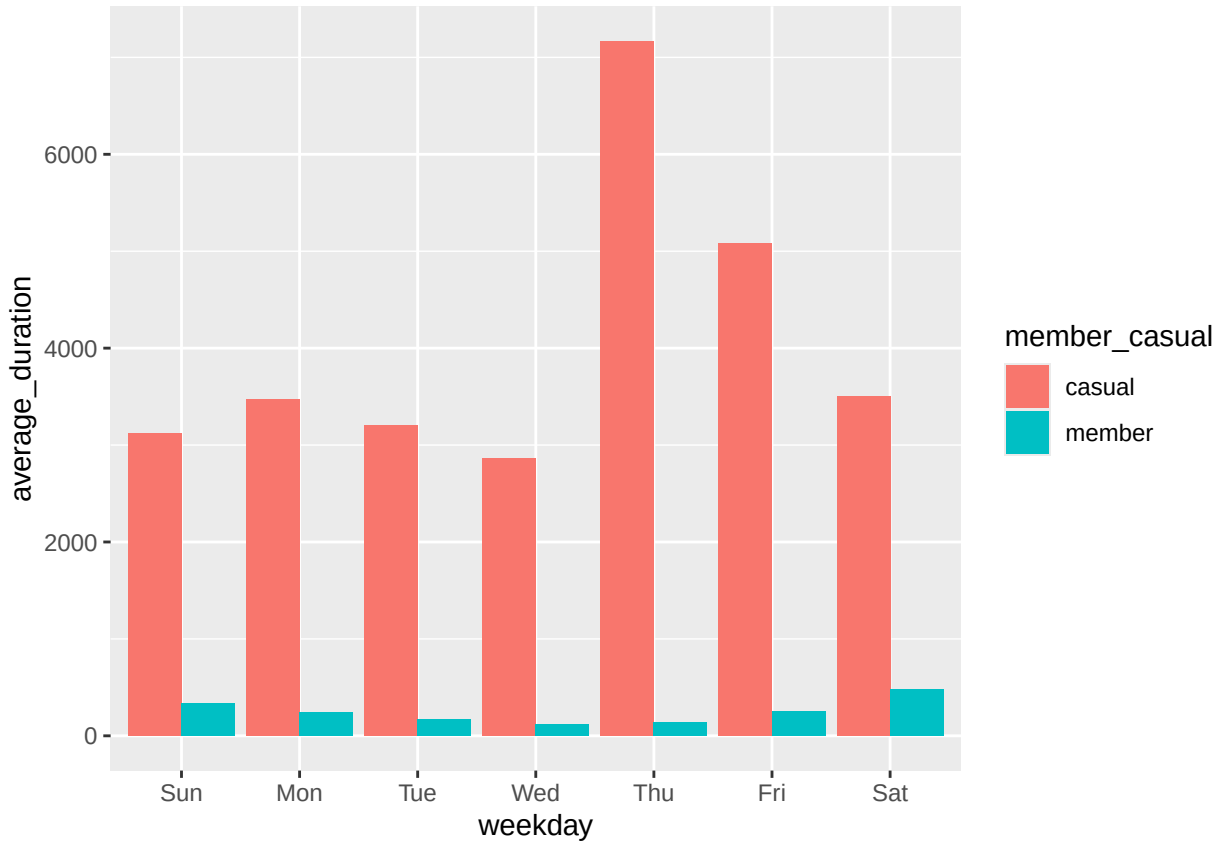
```r
## Visualization
# number of rides by rider type
all_trips_v2 %>%
mutate(weekday = wday(started_at, label = TRUE)) %>%
group_by(member_casual, weekday) %>%

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.
```




```
# 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.
```



Step 5: Export Summary File for Further Analysis

```
print(all_trips_v2)
```

```
## # A tibble: 788,189 x 15
##    ride_id  started_at ended_at   rideable_type start_station_id
##    <chr>    <date>     <date>     <chr>                    <dbl>
##  1 21742443 2019-01-01 2019-01-01 2167                       199
##  2 21742444 2019-01-01 2019-01-01 4386                        44
##  3 21742445 2019-01-01 2019-01-01 1524                        15
##  4 21742446 2019-01-01 2019-01-01 252                        123
##  5 21742447 2019-01-01 2019-01-01 1170                       173
##  6 21742448 2019-01-01 2019-01-01 2437                        98
##  7 21742449 2019-01-01 2019-01-01 2708                        98
##  8 21742450 2019-01-01 2019-01-01 2796                       211
##  9 21742451 2019-01-01 2019-01-01 6205                       150
## 10 21742452 2019-01-01 2019-01-01 3939                       268
## # i 788,179 more rows
## # i 10 more variables: start_station_name <chr>, end_station_id <dbl>,
## #   end_station_name <chr>, member_casual <chr>, date <date>, month <chr>,
## #   day <chr>, year <chr>, day_of_week <ord>, ride_length <dbl>
```

```
write.csv(all_trips_v2,"C:\\Users\\Aleny\\OneDrive\\Desktop\\Cyclistic Analysis.csv", row.names=FALSE)
```

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

```
write_csv(counts,"C:\\Users\\Aleny\\OneDrive\\Desktop\\avg_ride_length.csv")
```