

Cyclistic_A_Case_Study

Pratheep A

06/05/2021

Introduction.

Cyclistic is a bike-share program that features more than 5,800 bicycles and 600 docking stations. The majority of riders opt for traditional bikes; about 8% of riders use the assertive options. Cyclistic users are more likely to ride for leisure, but about 30% use them to commute to work each day.

Business task.

We are here to understand the usage of two different types of riders that are *Customers* who holds daily or hourly passes and *Subscribers* who has annual membership with the data that's collected.

Data source - <https://divvy-tripdata.s3.amazonaws.com/index.html>.

Data that's been collected belong to the year of 2018.

Let's get down to business.

Loading libraries

```
library(tidyverse)
```

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

```
## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.1.1      v dplyr   1.0.5
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
```

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

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

```
library(ggplot2)
library(readxl)
```

Loading datasets

```
X2018_Q1 = read_excel("E:/Data Analytics/Case studies/Cyclistic - Case study 1/New folder/2018_Q1.xlsx")
X2018_Q2 = read_excel("E:/Data Analytics/Case studies/Cyclistic - Case study 1/New folder/2018_Q2.xlsx")
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## NA inserted for impossible 1900-02-29 datetime
```

```
X2018_Q3 = read_excel("E:/Data Analytics/Case studies/Cyclistic - Case study 1/New folder/2018_Q3.xlsx")
X2018_Q4 = read_excel("E:/Data Analytics/Case studies/Cyclistic - Case study 1/New folder/2018_Q4.xlsx")
```

A peak at the column names to make sure it's uniform.

```
colnames(X2018_Q1)
```

```
## [1] "01 - Rental Details Rental ID"
## [2] "01 - Rental Details Local Start Time"
## [3] "01 - Rental Details Local End Time"
## [4] "01 - Rental Details Bike ID"
## [5] "01 - Rental Details Duration In Seconds Uncapped"
## [6] "03 - Rental Start Station ID"
## [7] "03 - Rental Start Station Name"
## [8] "02 - Rental End Station ID"
## [9] "02 - Rental End Station Name"
## [10] "User Type"
## [11] "Member Gender"
## [12] "05 - Member Details Member Birthday Year"
## [13] "ride_length"
## [14] "day_of_week"
```

```
colnames(X2018_Q2)
```

```
## [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"
## [13] "ride_length"      "day_of_week"
```

```
colnames(X2018_Q3)
```

```
## [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"
## [13] "ride_length"      "day_of_week"
```

```
colnames(X2018_Q4)
```

```
## [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"
## [13] "ride_length"      "day_of_week"
```

Renaming column names of x2018_Q1 to make it consistent with other data frames.

```
X2018_Q1 <- rename(X2018_Q1, "trip_id" = "01 - Rental Details Rental ID",
                    "start_time" = "01 - Rental Details Local Start Time",
                    "end_time" = "01 - Rental Details Local End Time",
                    "bikeid" = "01 - Rental Details Bike ID",
                    "tripduration" = "01 - Rental Details Duration In Seconds Uncapped",
                    "from_station_id" = "03 - Rental Start Station ID",
                    "from_station_name" = "03 - Rental Start Station Name",
                    "to_station_id" = "02 - Rental End Station ID",
                    "to_station_name" = "02 - Rental End Station Name",
                    "usertype" = "User Type", "gender" = "Member Gender",
                    "birthyear" = "05 - Member Details Member Birthday Year")
```

Checking if the column names are updated.

```
colnames(X2018_Q1)
```

```
## [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"
## [13] "ride_length"      "day_of_week"
```

Inspecting dataframes and look for incongruencies.

```
str(X2018_Q1)
```

```
## tibble[,14] [387,145 x 14] (S3: tbl_df/tbl/data.frame)
## $ trip_id      : num [1:387145] 17536702 17536703 17536704 17536705 17536706 ...
## $ start_time   : POSIXct[1:387145], format: "2018-01-01 00:12:00" "2018-01-01 00:41:35" ...
## $ end_time     : POSIXct[1:387145], format: "2018-01-01 00:17:23" "2018-01-01 00:47:52" ...
## $ bikeid       : num [1:387145] 3304 5367 4599 2302 3696 ...
## $ tripduration : num [1:387145] 323 377 2904 747 183 ...
## $ from_station_id : num [1:387145] 69 253 98 125 129 304 164 182 99 99 ...
## $ from_station_name: chr [1:387145] "Damen Ave & Pierce Ave" "Winthrop Ave & Lawrence Ave" "LaSalle
## $ to_station_id   : num [1:387145] 159 325 509 364 205 299 174 142 99 99 ...
## $ to_station_name : chr [1:387145] "Claremont Ave & Hirsch St" "Clark St & Winnemac Ave (Temp)" "T
## $ usertype       : chr [1:387145] "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...
## $ gender         : chr [1:387145] "Male" "Male" "Male" "Male" ...
## $ birthyear      : num [1:387145] 1988 1984 1989 1983 1989 ...
## $ ride_length    : POSIXct[1:387145], format: "1899-12-31 00:05:23" "1899-12-31 00:06:17" ...
## $ day_of_week    : num [1:387145] 2 2 2 2 2 2 2 2 2 2 ...
```

```
str(X2018_Q2)
```

```
## tibble[,14] [1,048,575 x 14] (S3: tbl_df/tbl/data.frame)
## $ trip_id      : num [1:1048575] 1.8e+07 1.8e+07 1.8e+07 1.8e+07 1.8e+07 ...
## $ start_time   : POSIXct[1:1048575], format: "2018-04-01 00:04:44" "2018-04-01 00:06:42" ...
## $ end_time     : POSIXct[1:1048575], format: "2018-04-01 00:13:03" "2018-04-01 00:27:07" ...
## $ bikeid       : num [1:1048575] 3819 5000 5165 3851 5065 ...
## $ tripduration : num [1:1048575] 499 1225 960 434 709 ...
## $ from_station_id : num [1:1048575] 22 157 106 241 228 244 128 130 130 121 ...
## $ from_station_name: chr [1:1048575] "May St & Taylor St" "Lake Shore Dr & Wellington Ave" "State S
## $ to_station_id  : num [1:1048575] 171 190 106 171 219 325 130 69 69 351 ...
## $ to_station_name : chr [1:1048575] "May St & Cullerton St" "Southport Ave & Wrightwood Ave" "Stat
## $ usertype       : chr [1:1048575] "Subscriber" "Subscriber" "Customer" "Subscriber" ...
## $ gender         : chr [1:1048575] "Male" "Male" NA "Male" ...
## $ birthyear      : num [1:1048575] 1994 1965 NA 1998 1983 ...
## $ ride_length    : POSIXct[1:1048575], format: "1899-12-31 00:08:19" "1899-12-31 00:20:25" ...
## $ day_of_week    : num [1:1048575] 1 1 1 1 1 1 1 1 1 1 ...
```

```
str(X2018_Q3)
```

```
## tibble[,14] [1,048,575 x 14] (S3: tbl_df/tbl/data.frame)
## $ trip_id      : num [1:1048575] 19244622 19244623 19244624 19244625 19244626 ...
## $ start_time   : POSIXct[1:1048575], format: "2018-07-01 00:00:03" "2018-07-01 00:00:13" ...
## $ end_time     : POSIXct[1:1048575], format: "2018-07-01 23:56:11" "2018-07-01 00:06:39" ...
## $ bikeid       : num [1:1048575] 5429 93 2461 2991 2851 ...
## $ tripduration : num [1:1048575] 86168 386 1391 1386 656 ...
## $ from_station_id : num [1:1048575] 140 153 76 76 60 128 168 168 229 229 ...
## $ from_station_name: chr [1:1048575] "Dearborn Pkwy & Delaware Pl" "Southport Ave & Wellington Ave"
## $ to_station_id  : num [1:1048575] 106 250 301 301 166 71 321 321 324 324 ...
## $ to_station_name : chr [1:1048575] "State St & Pearson St" "Ashland Ave & Wellington Ave" "Clark S
## $ usertype       : chr [1:1048575] "Customer" "Subscriber" "Subscriber" "Subscriber" ...
## $ gender         : chr [1:1048575] NA "Male" "Female" "Male" ...
## $ birthyear      : num [1:1048575] NA 1986 1987 1986 1961 ...
## $ ride_length    : POSIXct[1:1048575], format: "1899-12-31 23:56:08" "1899-12-31 00:06:26" ...
## $ day_of_week    : num [1:1048575] 1 1 1 1 1 1 1 1 1 1 ...
```

```
str(X2018_Q4)
```

```
## tibble[,14] [642,686 x 14] (S3: tbl_df/tbl/data.frame)
## $ trip_id      : num [1:642686] 2.1e+07 2.1e+07 2.1e+07 2.1e+07 2.1e+07 ...
## $ start_time   : POSIXct[1:642686], format: "2018-10-01 00:01:17" "2018-10-01 00:03:59" ...
## $ end_time     : POSIXct[1:642686], format: "2018-10-01 00:29:35" "2018-10-01 00:10:55" ...
## $ bikeid       : num [1:642686] 4551 847 6188 6372 1927 ...
## $ tripduration : num [1:642686] 1698 416 534 778 1102 ...
## $ from_station_id : num [1:642686] 85 13 59 328 93 229 148 374 268 125 ...
## $ from_station_name: chr [1:642686] "Michigan Ave & Oak St" "Wilton Ave & Diversey Pkwy" "Wabash Av
## $ to_station_id  : num [1:642686] 166 144 197 419 159 318 11 130 289 175 ...
## $ to_station_name : chr [1:642686] "Ashland Ave & Wrightwood Ave" "Larrabee St & Webster Ave" "Mic
## $ usertype       : chr [1:642686] "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...
## $ gender         : chr [1:642686] "Male" "Female" "Male" "Female" ...
## $ birthyear      : num [1:642686] 1992 1982 1986 1960 1993 ...
## $ ride_length    : POSIXct[1:642686], format: "2013-04-18 00:29:35" "2013-04-28 00:10:55" ...
## $ day_of_week    : num [1:642686] 2 2 2 2 2 2 2 2 2 2 ...
```

Converting ride_id and bikeid to character so that they can stack correctly.

```
X2018_Q1 <- mutate(X2018_Q1, trip_id = as.character(trip_id), bikeid = as.character(bikeid))
X2018_Q2 <- mutate(X2018_Q2, trip_id = as.character(trip_id), bikeid = as.character(bikeid))
X2018_Q3 <- mutate(X2018_Q3, trip_id = as.character(trip_id), bikeid = as.character(bikeid))
X2018_Q4 <- mutate(X2018_Q4, trip_id = as.character(trip_id), bikeid = as.character(bikeid))
```

Binding individual quarter data to one big data frame

```
all_trips <- bind_rows(X2018_Q1,X2018_Q2,X2018_Q3, X2018_Q4)
```

Removing tripduration , birthyear and gender fields.

```
all_trips <- all_trips %>%
  select(-c("tripduration", "birthyear", "gender" ))
```

Inspecting the data frame

```
colnames(all_trips)
```

```
## [1] "trip_id"          "start_time"       "end_time"
## [4] "bikeid"           "from_station_id"  "from_station_name"
## [7] "to_station_id"    "to_station_name"  "usertype"
## [10] "ride_length"      "day_of_week"
```

```
nrow(all_trips)
```

```
## [1] 3126981
```

```
dim(all_trips)
```

```
## [1] 3126981      11
```

```
head(all_trips)
```

```
## # A tibble: 6 x 11
##   trip_id start_time      end_time      bikeid from_station_id
##   <chr>    <dtm>          <dtm>          <chr>      <dbl>
## 1 17536702 2018-01-01 00:12:00 2018-01-01 00:17:23 3304         69
## 2 17536703 2018-01-01 00:41:35 2018-01-01 00:47:52 5367        253
## 3 17536704 2018-01-01 00:44:46 2018-01-01 01:33:10 4599         98
## 4 17536705 2018-01-01 00:53:10 2018-01-01 01:05:37 2302        125
## 5 17536706 2018-01-01 00:53:37 2018-01-01 00:56:40 3696        129
## 6 17536707 2018-01-01 00:56:15 2018-01-01 01:00:41 6298        304
## # ... with 6 more variables: from_station_name <chr>, to_station_id <dbl>,
## #   to_station_name <chr>, usertype <chr>, ride_length <dtm>,
## #   day_of_week <dbl>
```

```
str(all_trips)
```

```
## tibble[,11] [3,126,981 x 11] (S3: tbl_df/tbl/data.frame)
## $ trip_id      : chr [1:3126981] "17536702" "17536703" "17536704" "17536705" ...
## $ start_time   : POSIXct[1:3126981], format: "2018-01-01 00:12:00" "2018-01-01 00:41:35" ...
## $ end_time     : POSIXct[1:3126981], format: "2018-01-01 00:17:23" "2018-01-01 00:47:52" ...
## $ bikeid       : chr [1:3126981] "3304" "5367" "4599" "2302" ...
## $ from_station_id : num [1:3126981] 69 253 98 125 129 304 164 182 99 99 ...
## $ from_station_name: chr [1:3126981] "Damen Ave & Pierce Ave" "Winthrop Ave & Lawrence Ave" "LaSalle Ave & ..."
## $ to_station_id  : num [1:3126981] 159 325 509 364 205 299 174 142 99 99 ...
## $ to_station_name : chr [1:3126981] "Claremont Ave & Hirsch St" "Clark St & Winnemac Ave (Temp)" " ..."
## $ usertype      : chr [1:3126981] "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...
## $ ride_length    : POSIXct[1:3126981], format: "1899-12-31 00:05:23" "1899-12-31 00:06:17" ...
## $ day_of_week    : num [1:3126981] 2 2 2 2 2 2 2 2 2 2 ...
```

```
summary(all_trips)
```

```
##      trip_id      start_time      end_time
## Length:3126981  Min.   :2018-01-01 00:12:00  Min.   :2018-01-01 00:17:23
## Class :character 1st Qu.:2018-05-17 10:41:59  1st Qu.:2018-05-17 11:02:33
## Mode :character  Median :2018-07-08 15:51:59  Median :2018-07-08 16:27:38
##                      Mean   :2018-07-08 10:20:20  Mean   :2018-07-08 10:43:39
##                      3rd Qu.:2018-08-21 17:12:03  3rd Qu.:2018-08-21 17:29:25
##                      Max.   :2018-12-31 23:59:18  Max.   :2019-01-06 02:36:16
##
##      bikeid      from_station_id from_station_name to_station_id
## Length:3126981  Min.    : 2.0  Length:3126981  Min.    : 2.0
## Class :character 1st Qu.: 76.0  Class :character 1st Qu.: 76.0
## Mode :character  Median :165.0  Mode :character  Median :165.0
##                      Mean   :189.2  Mean   :189.8
##                      3rd Qu.:283.0  3rd Qu.:284.0
##                      Max.   :664.0  Max.   :664.0
##
##      to_station_name      usertype      ride_length
## Length:3126981  Length:3126981  Min.   :1899-12-31 00:01:01
## Class :character  Class :character  1st Qu.:1899-12-31 00:07:52
## Mode :character  Mode :character  Median :1899-12-31 00:15:34
##                      Mean   :1923-05-23 07:19:45
##                      3rd Qu.:1899-12-31 00:52:16
##                      Max.   :2019-01-06 02:36:16
##                      NA's   :1
##
##      day_of_week
## Min.   :1.000
## 1st Qu.:2.000
## Median :4.000
## Mean   :4.015
## 3rd Qu.:6.000
## Max.   :7.000
##
```

Adding columns that list the date, month, day, and year of each ride which will allow us to aggregate the ride data

```
all_trips$date <- as.Date(all_trips$start_time)
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")
```

Calculating ride length for each ride and adding a new column for the same (seconds)

```
all_trips$ride_length <- difftime(all_trips$end_time, all_trips$start_time)
```

Inspecting the structure

```
str(all_trips)
```

```
## tibble[,15] [3,126,981 x 15] (S3: tbl_df/tbl/data.frame)
## $ trip_id      : chr [1:3126981] "17536702" "17536703" "17536704" "17536705" ...
## $ start_time   : POSIXct[1:3126981], format: "2018-01-01 00:12:00" "2018-01-01 00:41:35" ...
## $ end_time     : POSIXct[1:3126981], format: "2018-01-01 00:17:23" "2018-01-01 00:47:52" ...
## $ bikeid       : chr [1:3126981] "3304" "5367" "4599" "2302" ...
## $ from_station_id : num [1:3126981] 69 253 98 125 129 304 164 182 99 99 ...
## $ from_station_name: chr [1:3126981] "Damen Ave & Pierce Ave" "Winthrop Ave & Lawrence Ave" "LaSalle Ave & ...
## $ to_station_id  : num [1:3126981] 159 325 509 364 205 299 174 142 99 99 ...
## $ to_station_name : chr [1:3126981] "Claremont Ave & Hirsch St" "Clark St & Winnemac Ave (Temp)" " ...
## $ usertype       : chr [1:3126981] "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...
## $ ride_length    : 'difftime' num [1:3126981] 5.383333333333333 6.283333333333333 48.4 12.45 ...
## ..- attr(*, "units")= chr "mins"
## $ day_of_week     : chr [1:3126981] "Monday" "Monday" "Monday" "Monday" ...
## $ date            : Date[1:3126981], format: "2018-01-01" "2018-01-01" ...
## $ month           : chr [1:3126981] "01" "01" "01" "01" ...
## $ day            : chr [1:3126981] "01" "01" "01" "01" ...
## $ year            : chr [1:3126981] "2018" "2018" "2018" "2018" ...
```

Converting ride_length from factor to number to perform calculations.

```
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
```

Removing bad data as few entries in ride length were showing in negative with a new data frame.

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

Time for some descriptive analysis.

```
mean(all_trips_v2$ride_length)
```

```
## [1] 23.31394
```

```
median(all_trips_v2$ride_length)
```

```
## [1] 11.1
```

```
max(all_trips_v2$ride_length)
```

```
## [1] 239000.7
```

```
min(all_trips_v2$ride_length)
```

```
## [1] 1.016667
```

Comparing subscribers vs customers.

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

```
##   all_trips_v2$usertype all_trips_v2$ride_length
## 1           Customer           63.82699
## 2           Subscriber           14.23280
```

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

```
##   all_trips_v2$usertype all_trips_v2$ride_length
## 1           Customer           28.0
## 2           Subscriber           9.5
```

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

```
##   all_trips_v2$usertype all_trips_v2$ride_length
## 1           Customer      239000.7
## 2           Subscriber      226020.3
```

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

```
##   all_trips_v2$usertype all_trips_v2$ride_length
## 1           Customer           1.016667
## 2           Subscriber           1.016667
```

Average ride time by each day for subscribers vs customers.

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



```
##   all_trips_v2$usertype all_trips_v2$day_of_week all_trips_v2$ride_length
## 1      Customer      Friday      67.97818
## 2      Subscriber    Friday      13.72580
## 3      Customer      Monday      61.81996
## 4      Subscriber    Monday      13.37815
## 5      Customer      Saturday     62.80487
## 6      Subscriber    Saturday     15.89929
## 7      Customer      Sunday      68.70193
## 8      Subscriber    Sunday      15.89729
## 9      Customer      Thursday     60.63596
## 10     Subscriber    Thursday     14.15888
## 11     Customer      Tuesday     63.16416
## 12     Subscriber    Tuesday     13.70862
## 13     Customer      Wednesday    58.17234
## 14     Subscriber    Wednesday    14.30379
```

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

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

```
##   all_trips_v2$usertype all_trips_v2$day_of_week all_trips_v2$ride_length
## 1      Customer      Sunday      68.70193
## 2      Subscriber    Sunday      15.89729
## 3      Customer      Monday      61.81996
## 4      Subscriber    Monday      13.37815
## 5      Customer      Tuesday     63.16416
## 6      Subscriber    Tuesday     13.70862
## 7      Customer      Wednesday    58.17234
## 8      Subscriber    Wednesday    14.30379
## 9      Customer      Thursday     60.63596
## 10     Subscriber    Thursday     14.15888
## 11     Customer      Friday      67.97818
## 12     Subscriber    Friday      13.72580
## 13     Customer      Saturday     62.80487
## 14     Subscriber    Saturday     15.89929
```

Analyzing ridership data by usertype and weekday

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

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

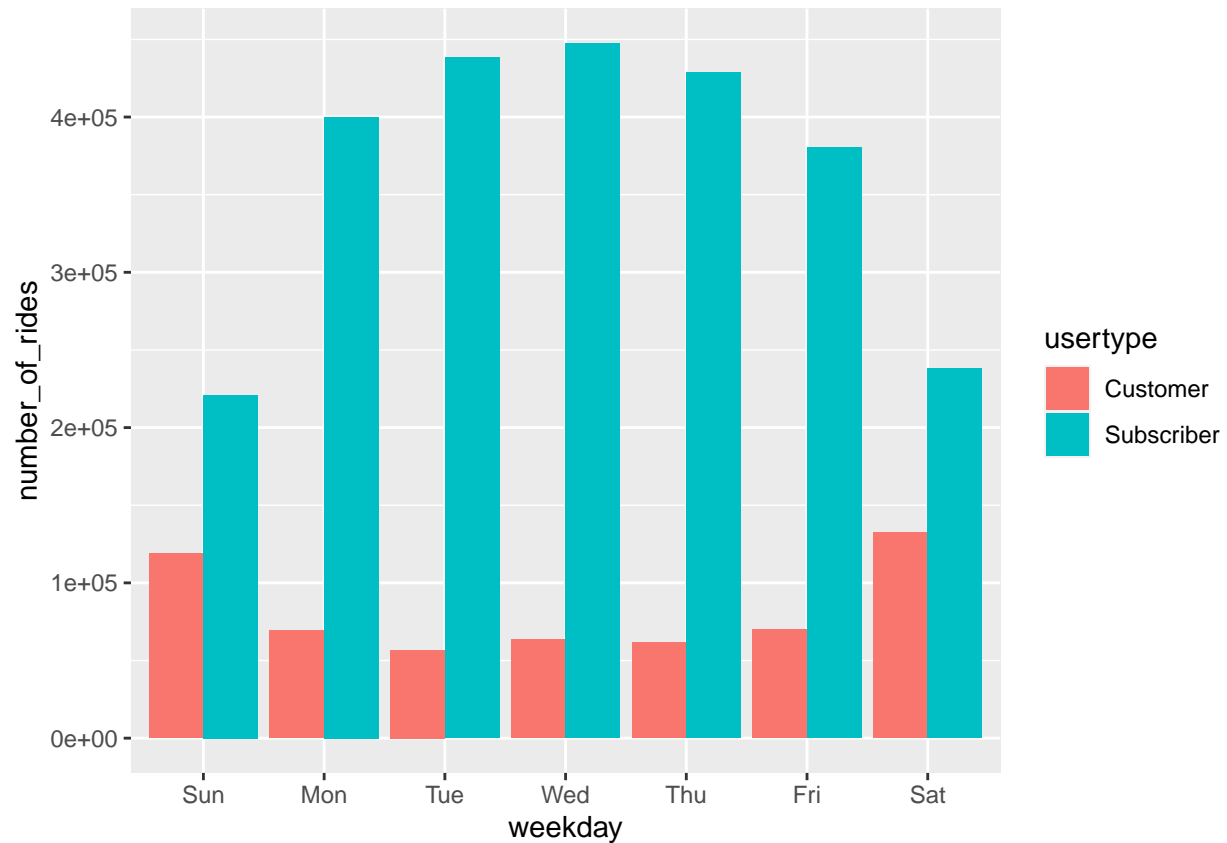
```
## # A tibble: 14 x 4
## # Groups:   usertype [2]
##   usertype weekday number_of_rides average_duration
##   <chr>      <ord>          <int>          <dbl>
```

##	1	Customer	Sun	118879	68.7
##	2	Customer	Mon	69525	61.8
##	3	Customer	Tue	56798	63.2
##	4	Customer	Wed	63747	58.2
##	5	Customer	Thu	61540	60.6
##	6	Customer	Fri	69792	68.0
##	7	Customer	Sat	132296	62.8
##	8	Subscriber	Sun	221074	15.9
##	9	Subscriber	Mon	399984	13.4
##	10	Subscriber	Tue	438511	13.7
##	11	Subscriber	Wed	447462	14.3
##	12	Subscriber	Thu	428830	14.2
##	13	Subscriber	Fri	380553	13.7
##	14	Subscriber	Sat	237983	15.9

Adding visualization

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

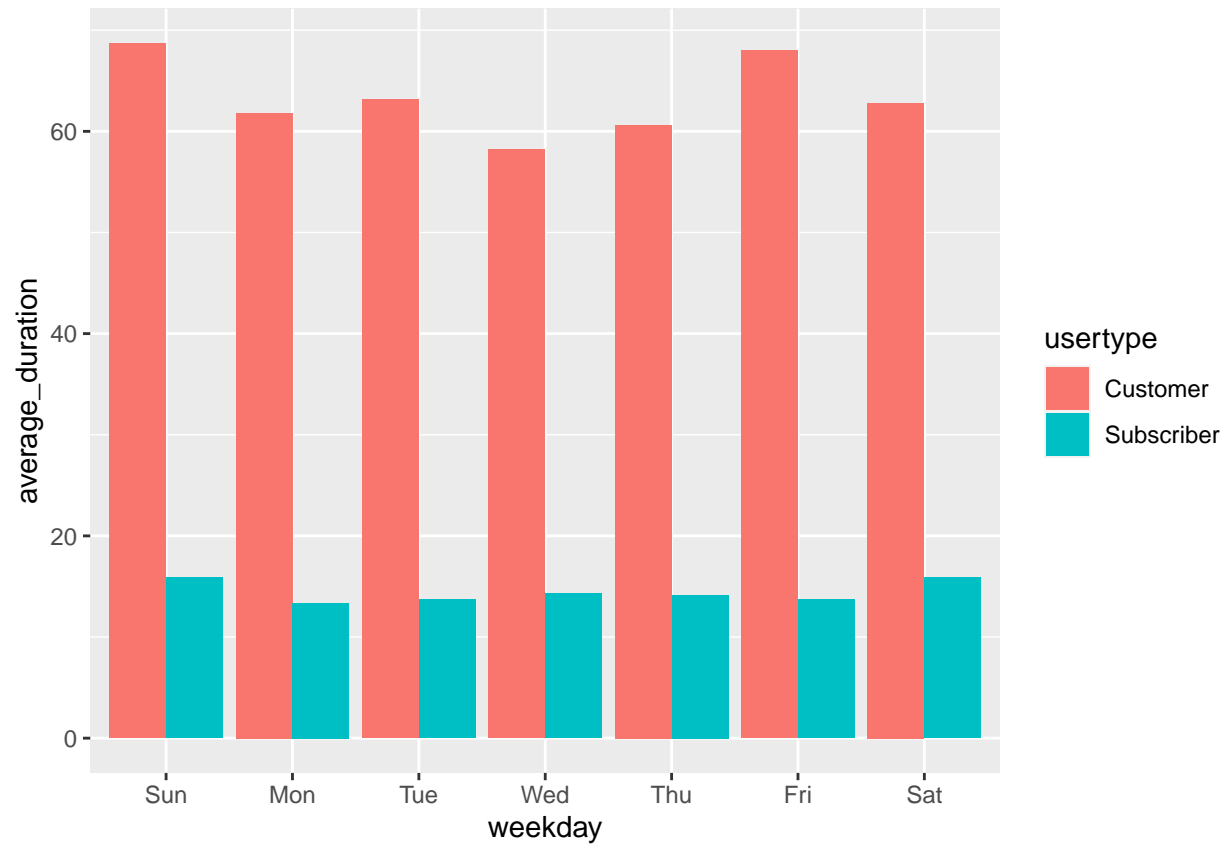
'summarise()' has grouped output by 'usertype'. You can override using the '.groups' argument.



Visualization with average duration

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

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



Final findings.

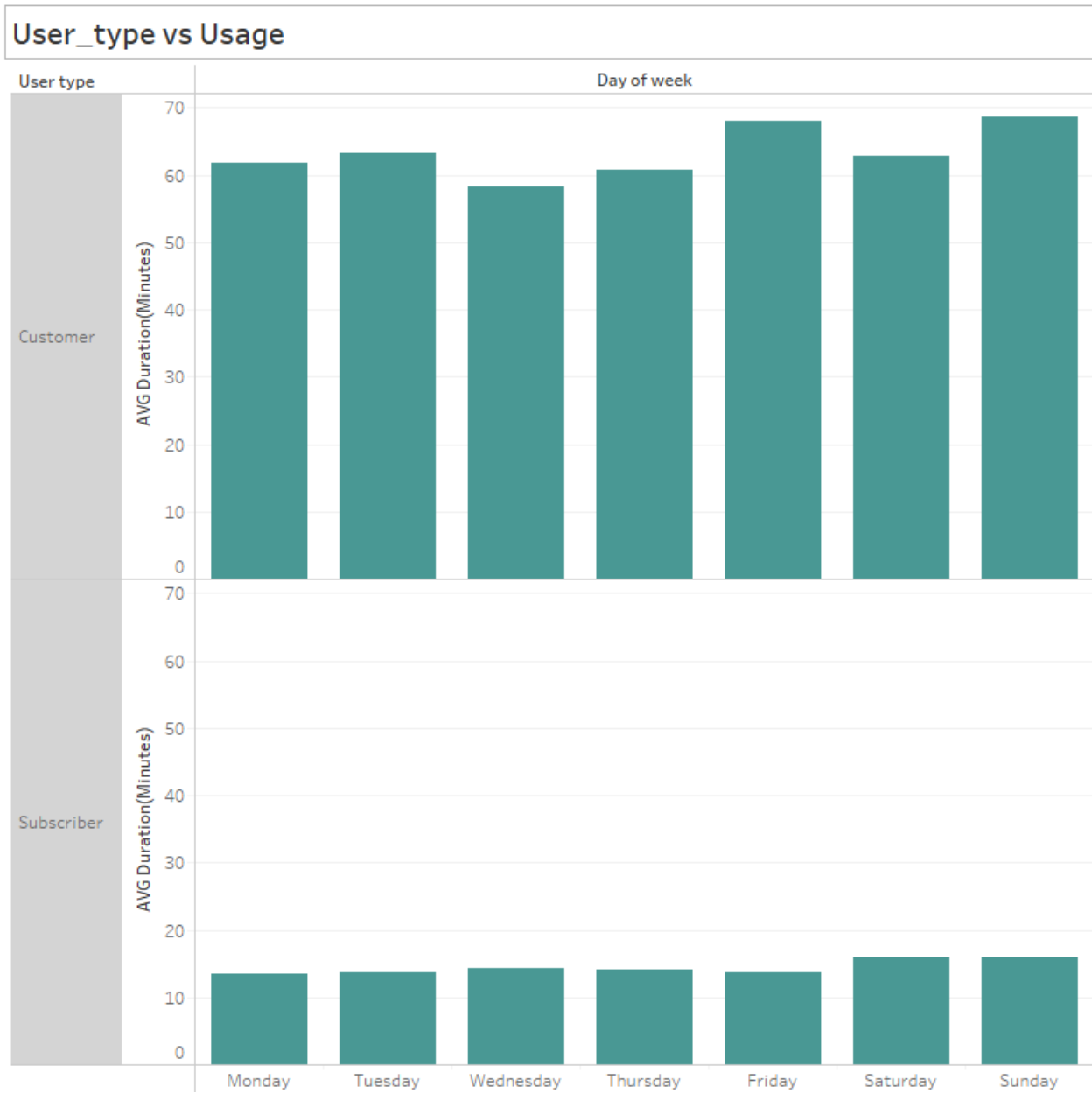


Figure 1: Customers V Subscribers