Bike Share Analysis With R

Tumisang Fokase

6/14/2021

Introduction

I am a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, my team wants to understand how casual riders and annual members use Cyclistic bikes differently.

Scenario

Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, the director believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers, she believes there is a very good chance to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Goal

The director of the marketing has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members.

Beginning of the Analysis.Install,load required packages and set up the working directory

```
library(pacman)
pacman::p_load(pacman,tidyverse,lubridate,ggplot2,dplyr,psych)
setwd("C:/Users/T Fokase/Desktop/Bike Share Case Study/Data/12_months_data/")
```

STEP 1: COLLECT DATA

upload Divvy datasets(csv files)

```
tripdata_2020_04 <- read_csv('202004-divvy-tripdata.csv')

## Warning in gzfile(file, mode): cannot open compressed file
'C:/Users/TFOKAS~1/
## AppData/Local/Temp/Rtmpq4ptiy\file61046365af7', probable reason 'No such
file or
## directory'
```

```
##
## -- Column specification --------------------------------------------------
------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   started_at = col_datetime(format = ""),
##   ended_at = col_datetime(format = ""),
##   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()
## )

tripdata_2020_05 <- read_csv('202005-divvy-tripdata.csv')

##
## -- Column specification --------------------------------------------------
------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   started_at = col_datetime(format = ""),
##   ended_at = col_datetime(format = ""),
##   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()
## )

tripdata_2020_06 <- read_csv('202006-divvy-tripdata.csv')

##
## -- Column specification --------------------------------------------------
------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   started_at = col_datetime(format = ""),
##   ended_at = col_datetime(format = ""),
##   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()
## )

tripdata_2020_07 <- read_csv('202007-divvy-tripdata.csv')

##
## -- Column specification --------------------------------------------------
------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   started_at = col_datetime(format = ""),
##   ended_at = col_datetime(format = ""),
##   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()
## )

tripdata_2020_08 <- read_csv('202008-divvy-tripdata.csv')

##
## -- Column specification --------------------------------------------------
------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   started_at = col_datetime(format = ""),
##   ended_at = col_datetime(format = ""),
##   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()
## )
```

```
tripdata_2020_09 <- read_csv('202009-divvy-tripdata.csv')

##
## -- Column specification --------------------------------------------------
------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   started_at = col_datetime(format = ""),
##   ended_at = col_datetime(format = ""),
##   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()
## )

tripdata_2020_10 <- read_csv('202010-divvy-tripdata.csv')

##
## -- Column specification --------------------------------------------------
------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   started_at = col_datetime(format = ""),
##   ended_at = col_datetime(format = ""),
##   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()
## )

tripdata_2020_11 <- read_csv('202011-divvy-tripdata.csv')

##
## -- Column specification --------------------------------------------------
------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   started_at = col_datetime(format = ""),
```

```
##   ended_at = col_datetime(format = ""),
##   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()
## )

tripdata_2020_12 <- read_csv('202012-divvy-tripdata.csv')

##
## -- Column specification --------------------------------------------------
------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   started_at = col_datetime(format = ""),
##   ended_at = col_datetime(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_character(),
##   end_station_name = col_character(),
##   end_station_id = col_character(),
##   start_lat = col_double(),
##   start_lng = col_double(),
##   end_lat = col_double(),
##   end_lng = col_double(),
##   member_casual = col_character()
## )

tripdata_2021_01 <- read_csv('202101-divvy-tripdata.csv')

##
## -- Column specification --------------------------------------------------
------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   started_at = col_datetime(format = ""),
##   ended_at = col_datetime(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_character(),
##   end_station_name = col_character(),
##   end_station_id = col_character(),
##   start_lat = col_double(),
##   start_lng = col_double(),
##   end_lat = col_double(),
##   end_lng = col_double(),
```

```
##   member_casual = col_character()
## )

tripdata_2021_02 <- read_csv('202102-divvy-tripdata.csv')

##
## -- Column specification --------------------------------------------------
------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   started_at = col_datetime(format = ""),
##   ended_at = col_datetime(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_character(),
##   end_station_name = col_character(),
##   end_station_id = col_character(),
##   start_lat = col_double(),
##   start_lng = col_double(),
##   end_lat = col_double(),
##   end_lng = col_double(),
##   member_casual = col_character()
## )

tripdata_2021_03 <- read_csv('202103-divvy-tripdata.csv')

##
## -- Column specification --------------------------------------------------
------
## cols(
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   started_at = col_datetime(format = ""),
##   ended_at = col_datetime(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_character(),
##   end_station_name = col_character(),
##   end_station_id = col_character(),
##   start_lat = col_double(),
##   start_lng = col_double(),
##   end_lat = col_double(),
##   end_lng = col_double(),
##   member_casual = col_character()
## )

tripdata_2021_04 <- read_csv('202104-divvy-tripdata.csv')

##
## -- Column specification --------------------------------------------------
------
## cols(
```

```
##   ride_id = col_character(),
##   rideable_type = col_character(),
##   started_at = col_datetime(format = ""),
##   ended_at = col_datetime(format = ""),
##   start_station_name = col_character(),
##   start_station_id = col_character(),
##   end_station_name = col_character(),
##   end_station_id = col_character(),
##   start_lat = col_double(),
##   start_lng = col_double(),
##   end_lat = col_double(),
##   end_lng = col_double(),
##   member_casual = col_character()
## )
```

Compare the column names of each of the files. Names need to match perfectly before we can join.

```
colnames(tripdata_2020_04)

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

colnames(tripdata_2020_05)

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

colnames(tripdata_2020_06)

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

colnames(tripdata_2020_07)

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

colnames(tripdata_2020_08)
```

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

```
colnames(tripdata_2020_09)
```

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

```
colnames(tripdata_2020_10)
```

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

```
colnames(tripdata_2020_11)
```

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

```
colnames(tripdata_2020_12)
```

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

```
colnames(tripdata_2021_01)
```

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

```
colnames(tripdata_2021_02)
```

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

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

colnames(tripdata_2021_03)

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

colnames(tripdata_2021_04)

##  [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"
```

compare the structure of the table

```
str(tripdata_2020_04)

## spec_tbl_df [84,776 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:84776] "A847FADBBC638E45" "5405B80E996FF60D"
"5DD24A79A4E006F4" "2A59BBDF5CDBA725" ...
##  $ rideable_type     : chr [1:84776] "docked_bike" "docked_bike"
"docked_bike" "docked_bike" ...
##  $ started_at        : POSIXct[1:84776], format: "2020-04-26 17:45:14"
"2020-04-17 17:08:54" ...
##  $ ended_at          : POSIXct[1:84776], format: "2020-04-26 18:12:03"
"2020-04-17 17:17:03" ...
##  $ start_station_name: chr [1:84776] "Eckhart Park" "Drake Ave & Fullerton
Ave" "McClurg Ct & Erie St" "California Ave & Division St" ...
##  $ start_station_id  : num [1:84776] 86 503 142 216 125 173 35 434 627 377
...
##  $ end_station_name  : chr [1:84776] "Lincoln Ave & Diversey Pkwy"
"Kosciuszko Park" "Indiana Ave & Roosevelt Rd" "Wood St & Augusta Blvd" ...
##  $ end_station_id    : num [1:84776] 152 499 255 657 323 35 635 382 359
508 ...
##  $ start_lat         : num [1:84776] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:84776] -87.7 -87.7 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num [1:84776] 41.9 41.9 41.9 41.9 42 ...
##  $ end_lng           : num [1:84776] -87.7 -87.7 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:84776] "member" "member" "member" "member"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
```

```
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )

str(tripdata_2020_05)

## spec_tbl_df [200,274 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:200274] "02668AD35674B983"
"7A50CCAF1EDDB28F" "2FFCDFDB91FE9A52" "58991CF1DB75BA84" ...
##  $ rideable_type     : chr [1:200274] "docked_bike" "docked_bike"
"docked_bike" "docked_bike" ...
##  $ started_at        : POSIXct[1:200274], format: "2020-05-27 10:03:52"
"2020-05-25 10:47:11" ...
##  $ ended_at          : POSIXct[1:200274], format: "2020-05-27 10:16:49"
"2020-05-25 11:05:40" ...
##  $ start_station_name: chr [1:200274] "Franklin St & Jackson Blvd" "Clark
St & Wrightwood Ave" "Kedzie Ave & Milwaukee Ave" "Clarendon Ave & Leland
Ave" ...
##  $ start_station_id  : num [1:200274] 36 340 260 251 261 206 261 180 331
219 ...
##  $ end_station_name  : chr [1:200274] "Wabash Ave & Grand Ave" "Clark St &
Leland Ave" "Kedzie Ave & Milwaukee Ave" "Lake Shore Dr & Wellington Ave" ...
##  $ end_station_id    : num [1:200274] 199 326 260 157 206 22 261 180 300
305 ...
##  $ start_lat         : num [1:200274] 41.9 41.9 41.9 42 41.9 ...
##  $ start_lng         : num [1:200274] -87.6 -87.6 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:200274] 41.9 42 41.9 41.9 41.8 ...
##  $ end_lng           : num [1:200274] -87.6 -87.7 -87.7 -87.6 -87.6 ...
##  $ member_casual     : chr [1:200274] "member" "casual" "casual" "casual"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )

str(tripdata_2020_06)

## spec_tbl_df [343,005 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:343005] "8CD5DE2C2B6C4CFC"
"9A191EB2C751D85D" "F37D14B0B5659BCF" "C41237B506E85FA1" ...
##  $ rideable_type     : chr [1:343005] "docked_bike" "docked_bike"
"docked_bike" "docked_bike" ...
##  $ started_at        : POSIXct[1:343005], format: "2020-06-13 23:24:48"
"2020-06-26 07:26:10" ...
##  $ ended_at          : POSIXct[1:343005], format: "2020-06-13 23:36:55"
"2020-06-26 07:31:58" ...
##  $ start_station_name: chr [1:343005] "Wilton Ave & Belmont Ave" "Federal
St & Polk St" "Daley Center Plaza" "Broadway & Cornelia Ave" ...
##  $ start_station_id  : num [1:343005] 117 41 81 303 327 327 41 115 338 84
...
##  $ end_station_name  : chr [1:343005] "Damen Ave & Clybourn Ave" "Daley
Center Plaza" "State St & Harrison St" "Broadway & Berwyn Ave" ...
##  $ end_station_id    : num [1:343005] 163 81 5 294 117 117 81 303 164 53
...
##  $ start_lat         : num [1:343005] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:343005] -87.7 -87.6 -87.6 -87.6 -87.7 ...
##  $ end_lat           : num [1:343005] 41.9 41.9 41.9 42 41.9 ...
##  $ end_lng           : num [1:343005] -87.7 -87.6 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:343005] "casual" "member" "member" "casual"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )

str(tripdata_2020_07)

## spec_tbl_df [551,480 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:551480] "762198876D69004D"
```

```
"BEC9C9FBA0D4CF1B" "D2FD8EA432C77EC1" "54AE594E20B35881" ...
##  $ rideable_type     : chr [1:551480] "docked_bike" "docked_bike"
"docked_bike" "docked_bike" ...
##  $ started_at        : POSIXct[1:551480], format: "2020-07-09 15:22:02"
"2020-07-24 23:56:30" ...
##  $ ended_at          : POSIXct[1:551480], format: "2020-07-09 15:25:52"
"2020-07-25 00:20:17" ...
##  $ start_station_name: chr [1:551480] "Ritchie Ct & Banks St" "Halsted St
& Roscoe St" "Lake Shore Dr & Diversey Pkwy" "LaSalle St & Illinois St" ...
##  $ start_station_id  : num [1:551480] 180 299 329 181 268 635 113 211 176
31 ...
##  $ end_station_name  : chr [1:551480] "Wells St & Evergreen Ave" "Broadway
& Ridge Ave" "Clark St & Wellington Ave" "Clark St & Armitage Ave" ...
##  $ end_station_id    : num [1:551480] 291 461 156 94 301 289 140 31 191
142 ...
##  $ start_lat         : num [1:551480] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:551480] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:551480] 41.9 42 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:551480] -87.6 -87.7 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:551480] "member" "member" "casual" "casual"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )
```

```
str(tripdata_2020_08)
```

```
## spec_tbl_df [622,361 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:622361] "322BD23D287743ED"
"2A3AEF1AB9054D8B" "67DC1D133E8B5816" "C79FBBD412E578A7" ...
##  $ rideable_type     : chr [1:622361] "docked_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : POSIXct[1:622361], format: "2020-08-20 18:08:14"
"2020-08-27 18:46:04" ...
##  $ ended_at          : POSIXct[1:622361], format: "2020-08-20 18:17:51"
"2020-08-27 19:54:51" ...
##  $ start_station_name: chr [1:622361] "Lake Shore Dr & Diversey Pkwy"
```

```
"Michigan Ave & 14th St" "Columbus Dr & Randolph St" "Daley Center Plaza" ...
##  $ start_station_id  : num [1:622361] 329 168 195 81 658 658 196 67 153
177 ...
##  $ end_station_name  : chr [1:622361] "Clark St & Lincoln Ave" "Michigan
Ave & 14th St" "State St & Randolph St" "State St & Kinzie St" ...
##  $ end_station_id    : num [1:622361] 141 168 44 47 658 658 49 229 225 305
...
##  $ start_lat         : num [1:622361] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:622361] -87.6 -87.6 -87.6 -87.6 -87.7 ...
##  $ end_lat           : num [1:622361] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:622361] -87.6 -87.6 -87.6 -87.6 -87.7 ...
##  $ member_casual     : chr [1:622361] "member" "casual" "casual" "casual"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )

str(tripdata_2020_09)

## spec_tbl_df [532,958 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:532958] "2B22BD5F95FB2629"
"A7FB70B4AFC6CAF2" "86057FA01BAC778E" "57F6DC9A153DB98C" ...
##  $ rideable_type     : chr [1:532958] "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : POSIXct[1:532958], format: "2020-09-17 14:27:11"
"2020-09-17 15:07:31" ...
##  $ ended_at          : POSIXct[1:532958], format: "2020-09-17 14:44:24"
"2020-09-17 15:07:45" ...
##  $ start_station_name: chr [1:532958] "Michigan Ave & Lake St" "W Oakdale
Ave & N Broadway" "W Oakdale Ave & N Broadway" "Ashland Ave & Belle Plaine
Ave" ...
##  $ start_station_id  : num [1:532958] 52 NA NA 246 24 94 291 NA NA NA ...
##  $ end_station_name  : chr [1:532958] "Green St & Randolph St" "W Oakdale
Ave & N Broadway" "W Oakdale Ave & N Broadway" "Montrose Harbor" ...
##  $ end_station_id    : num [1:532958] 112 NA NA 249 24 NA 256 NA NA NA ...
##  $ start_lat         : num [1:532958] 41.9 41.9 41.9 42 41.9 ...
##  $ start_lng         : num [1:532958] -87.6 -87.6 -87.6 -87.7 -87.6 ...
```

```
##  $ end_lat           : num [1:532958] 41.9 41.9 41.9 42 41.9 ...
##  $ end_lng           : num [1:532958] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:532958] "casual" "casual" "casual" "casual"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )
```

```
str(tripdata_2020_10)
```

```
## spec_tbl_df [388,653 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:388653] "ACB6B40CF5B9044C"
"DF450C72FD109C01" "B6396B54A15AC0DF" "44A4AEE261B9E854" ...
##  $ rideable_type     : chr [1:388653] "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : POSIXct[1:388653], format: "2020-10-31 19:39:43"
"2020-10-31 23:50:08" ...
##  $ ended_at          : POSIXct[1:388653], format: "2020-10-31 19:57:12"
"2020-11-01 00:04:16" ...
##  $ start_station_name: chr [1:388653] "Lakeview Ave & Fullerton Pkwy"
"Southport Ave & Waveland Ave" "Stony Island Ave & 67th St" "Clark St & Grace
St" ...
##  $ start_station_id  : num [1:388653] 313 227 102 165 190 359 313 125 NA
174 ...
##  $ end_station_name  : chr [1:388653] "Rush St & Hubbard St" "Kedzie Ave &
Milwaukee Ave" "University Ave & 57th St" "Broadway & Sheridan Rd" ...
##  $ end_station_id    : num [1:388653] 125 260 423 256 185 53 125 313 199
635 ...
##  $ start_lat         : num [1:388653] 41.9 41.9 41.8 42 41.9 ...
##  $ start_lng         : num [1:388653] -87.6 -87.7 -87.6 -87.7 -87.7 ...
##  $ end_lat           : num [1:388653] 41.9 41.9 41.8 42 41.9 ...
##  $ end_lng           : num [1:388653] -87.6 -87.7 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:388653] "casual" "casual" "casual" "casual"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
```

```
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )

str(tripdata_2020_11)

## spec_tbl_df [259,716 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:259716] "BD0A6FF6FFF9B921"
"96A7A7A4BDE4F82D" "C61526D06582BDC5" "E533E89C32080B9E" ...
##  $ rideable_type     : chr [1:259716] "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : POSIXct[1:259716], format: "2020-11-01 13:36:00"
"2020-11-01 10:03:26" ...
##  $ ended_at          : POSIXct[1:259716], format: "2020-11-01 13:45:40"
"2020-11-01 10:14:45" ...
##  $ start_station_name: chr [1:259716] "Dearborn St & Erie St" "Franklin St
& Illinois St" "Lake Shore Dr & Monroe St" "Leavitt St & Chicago Ave" ...
##  $ start_station_id  : num [1:259716] 110 672 76 659 2 72 76 NA 58 394 ...
##  $ end_station_name  : chr [1:259716] "St. Clair St & Erie St" "Noble St &
Milwaukee Ave" "Federal St & Polk St" "Stave St & Armitage Ave" ...
##  $ end_station_id    : num [1:259716] 211 29 41 185 2 76 72 NA 288 273 ...
##  $ start_lat         : num [1:259716] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:259716] -87.6 -87.6 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num [1:259716] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:259716] -87.6 -87.7 -87.6 -87.7 -87.6 ...
##  $ member_casual     : chr [1:259716] "casual" "casual" "casual" "casual"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )

str(tripdata_2020_12)

## spec_tbl_df [131,573 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:131573] "70B6A9A437D4C30D"
"158A465D4E74C54A" "5262016E0F1F2F9A" "BE119628E44F871E" ...
##  $ rideable_type     : chr [1:131573] "classic_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : POSIXct[1:131573], format: "2020-12-27 12:44:29"
"2020-12-18 17:37:15" ...
##  $ ended_at          : POSIXct[1:131573], format: "2020-12-27 12:55:06"
"2020-12-18 17:44:19" ...
##  $ start_station_name: chr [1:131573] "Aberdeen St & Jackson Blvd" NA NA
NA ...
##  $ start_station_id  : chr [1:131573] "13157" NA NA NA ...
##  $ end_station_name  : chr [1:131573] "Desplaines St & Kinzie St" NA NA NA
...
##  $ end_station_id    : chr [1:131573] "TA1306000003" NA NA NA ...
##  $ start_lat         : num [1:131573] 41.9 41.9 41.9 41.9 41.8 ...
##  $ start_lng         : num [1:131573] -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ end_lat           : num [1:131573] 41.9 41.9 41.9 41.9 41.8 ...
##  $ end_lng           : num [1:131573] -87.6 -87.7 -87.7 -87.7 -87.6 ...
##  $ member_casual     : chr [1:131573] "member" "member" "member" "member"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )

str(tripdata_2021_01)

## spec_tbl_df [96,834 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:96834] "E19E6F1B8D4C42ED" "DC88F20C2C55F27F"
"EC45C94683FE3F27" "4FA453A75AE377DB" ...
##  $ rideable_type     : chr [1:96834] "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
```

```
##  $ started_at        : POSIXct[1:96834], format: "2021-01-23 16:14:19"
"2021-01-27 18:43:08" ...
##  $ ended_at          : POSIXct[1:96834], format: "2021-01-23 16:24:44"
"2021-01-27 18:47:12" ...
##  $ start_station_name: chr [1:96834] "California Ave & Cortez St"
"California Ave & Cortez St" "California Ave & Cortez St" "California Ave &
Cortez St" ...
##  $ start_station_id  : chr [1:96834] "17660" "17660" "17660" "17660" ...
##  $ end_station_name  : chr [1:96834] NA NA NA NA ...
##  $ end_station_id    : chr [1:96834] NA NA NA NA ...
##  $ start_lat         : num [1:96834] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:96834] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:96834] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:96834] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr [1:96834] "member" "member" "member" "member"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
```

```
str(tripdata_2021_02)
```

```
## spec_tbl_df [49,622 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:49622] "89E7AA6C29227EFF" "0FEFDE2603568365"
"E6159D746B2DBB91" "B32D3199F1C2E75B" ...
##  $ rideable_type     : chr [1:49622] "classic_bike" "classic_bike"
"electric_bike" "classic_bike" ...
##  $ started_at        : POSIXct[1:49622], format: "2021-02-12 16:14:56"
"2021-02-14 17:52:38" ...
##  $ ended_at          : POSIXct[1:49622], format: "2021-02-12 16:21:43"
"2021-02-14 18:12:09" ...
##  $ start_station_name: chr [1:49622] "Glenwood Ave & Touhy Ave" "Glenwood
Ave & Touhy Ave" "Clark St & Lake St" "Wood St & Chicago Ave" ...
##  $ start_station_id  : chr [1:49622] "525" "525" "KA1503000012" "637" ...
##  $ end_station_name  : chr [1:49622] "Sheridan Rd & Columbia Ave"
"Bosworth Ave & Howard St" "State St & Randolph St" "Honore St & Division St"
...
```

```
##  $ end_station_id    : chr [1:49622] "660" "16806" "TA1305000029"
"TA1305000034" ...
##  $ start_lat         : num [1:49622] 42 42 41.9 41.9 41.8 ...
##  $ start_lng         : num [1:49622] -87.7 -87.7 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num [1:49622] 42 42 41.9 41.9 41.8 ...
##  $ end_lng           : num [1:49622] -87.7 -87.7 -87.6 -87.7 -87.6 ...
##  $ member_casual     : chr [1:49622] "member" "casual" "member" "member"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
```

```
str(tripdata_2021_03)
```

```
## spec_tbl_df [228,496 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:228496] "CFA86D4455AA1030"
"30D9DC61227D1AF3" "846D87A15682A284" "994D05AA75A168F2" ...
##  $ rideable_type     : chr [1:228496] "classic_bike" "classic_bike"
"classic_bike" "classic_bike" ...
##  $ started_at        : POSIXct[1:228496], format: "2021-03-16 08:32:30"
"2021-03-28 01:26:28" ...
##  $ ended_at          : POSIXct[1:228496], format: "2021-03-16 08:36:34"
"2021-03-28 01:36:55" ...
##  $ start_station_name: chr [1:228496] "Humboldt Blvd & Armitage Ave"
"Humboldt Blvd & Armitage Ave" "Shields Ave & 28th Pl" "Winthrop Ave &
Lawrence Ave" ...
##  $ start_station_id  : chr [1:228496] "15651" "15651" "15443"
"TA1308000021" ...
##  $ end_station_name  : chr [1:228496] "Stave St & Armitage Ave" "Central
Park Ave & Bloomingdale Ave" "Halsted St & 35th St" "Broadway & Sheridan Rd"
...
##  $ end_station_id    : chr [1:228496] "13266" "18017" "TA1308000043"
"13323" ...
##  $ start_lat         : num [1:228496] 41.9 41.9 41.8 42 42 ...
##  $ start_lng         : num [1:228496] -87.7 -87.7 -87.6 -87.7 -87.7 ...
##  $ end_lat           : num [1:228496] 41.9 41.9 41.8 42 42.1 ...
##  $ end_lng           : num [1:228496] -87.7 -87.7 -87.6 -87.6 -87.7 ...
```

```
##  $ member_casual     : chr [1:228496] "casual" "casual" "casual" "casual"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
```

```
str(tripdata_2021_04)
```

```
## spec_tbl_df [337,230 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:337230] "6C992BD37A98A63F"
"1E0145613A209000" "E498E15508A80BAD" "1887262AD101C604" ...
##  $ rideable_type     : chr [1:337230] "classic_bike" "docked_bike"
"docked_bike" "classic_bike" ...
##  $ started_at        : POSIXct[1:337230], format: "2021-04-12 18:25:36"
"2021-04-27 17:27:11" ...
##  $ ended_at          : POSIXct[1:337230], format: "2021-04-12 18:56:55"
"2021-04-27 18:31:29" ...
##  $ start_station_name: chr [1:337230] "State St & Pearson St" "Dorchester
Ave & 49th St" "Loomis Blvd & 84th St" "Honore St & Division St" ...
##  $ start_station_id  : chr [1:337230] "TA1307000061" "KA1503000069"
"20121" "TA1305000034" ...
##  $ end_station_name  : chr [1:337230] "Southport Ave & Waveland Ave"
"Dorchester Ave & 49th St" "Loomis Blvd & 84th St" "Southport Ave & Waveland
Ave" ...
##  $ end_station_id    : chr [1:337230] "13235" "KA1503000069" "20121"
"13235" ...
##  $ start_lat         : num [1:337230] 41.9 41.8 41.7 41.9 41.7 ...
##  $ start_lng         : num [1:337230] -87.6 -87.6 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:337230] 41.9 41.8 41.7 41.9 41.7 ...
##  $ end_lng           : num [1:337230] -87.7 -87.6 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr [1:337230] "member" "casual" "casual" "member"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
```

```
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
```

Convert “start_station_id” and “end_station_id” to numberic so that they can stack correctly.

```
tripdata_2020_04 <- mutate(tripdata_2020_04, start_station_id =
as.character(start_station_id),
                           end_station_id = as.character(end_station_id))
tripdata_2020_05 <- mutate(tripdata_2020_05, start_station_id =
as.character(start_station_id),
                           end_station_id = as.character(end_station_id))
tripdata_2020_06 <- mutate(tripdata_2020_06, start_station_id =
as.character(start_station_id),
                           end_station_id = as.character(end_station_id))
tripdata_2020_07 <- mutate(tripdata_2020_07, start_station_id =
as.character(start_station_id),
                           end_station_id = as.character(end_station_id))
tripdata_2020_08 <- mutate(tripdata_2020_08, start_station_id =
as.character(start_station_id),
                           end_station_id = as.character(end_station_id))
tripdata_2020_09 <- mutate(tripdata_2020_09, start_station_id =
as.character(start_station_id),
                           end_station_id = as.character(end_station_id))
tripdata_2020_10 <- mutate(tripdata_2020_10, start_station_id =
as.character(start_station_id),
                           end_station_id = as.character(end_station_id))
tripdata_2020_11 <- mutate(tripdata_2020_11, start_station_id =
as.character(start_station_id),
                           end_station_id = as.character(end_station_id))
tripdata_2020_12 <- mutate(tripdata_2020_12, start_station_id =
as.character(start_station_id),
                           end_station_id = as.character(end_station_id))
tripdata_2021_01 <- mutate(tripdata_2021_01, start_station_id =
as.character(start_station_id),
                           end_station_id = as.character(end_station_id))
tripdata_2021_02 <- mutate(tripdata_2021_02, start_station_id =
as.character(start_station_id),
                           end_station_id = as.character(end_station_id))
tripdata_2021_03 <- mutate(tripdata_2021_03, start_station_id =
as.character(start_station_id),
                           end_station_id = as.character(end_station_id))
```

```
tripdata_2021_04 <- mutate(tripdata_2021_04, start_station_id =
as.character(start_station_id),
                           end_station_id = as.character(end_station_id))
```

check to see if it worked.

```
str(tripdata_2020_05)
```

```
## spec_tbl_df [200,274 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:200274] "02668AD35674B983"
"7A50CCAF1EDDB28F" "2FFCDFDB91FE9A52" "58991CF1DB75BA84" ...
##  $ rideable_type     : chr [1:200274] "docked_bike" "docked_bike"
"docked_bike" "docked_bike" ...
##  $ started_at        : POSIXct[1:200274], format: "2020-05-27 10:03:52"
"2020-05-25 10:47:11" ...
##  $ ended_at          : POSIXct[1:200274], format: "2020-05-27 10:16:49"
"2020-05-25 11:05:40" ...
##  $ start_station_name: chr [1:200274] "Franklin St & Jackson Blvd" "Clark
St & Wrightwood Ave" "Kedzie Ave & Milwaukee Ave" "Clarendon Ave & Leland
Ave" ...
##  $ start_station_id  : chr [1:200274] "36" "340" "260" "251" ...
##  $ end_station_name  : chr [1:200274] "Wabash Ave & Grand Ave" "Clark St &
Leland Ave" "Kedzie Ave & Milwaukee Ave" "Lake Shore Dr & Wellington Ave" ...
##  $ end_station_id    : chr [1:200274] "199" "326" "260" "157" ...
##  $ start_lat         : num [1:200274] 41.9 41.9 41.9 42 41.9 ...
##  $ start_lng         : num [1:200274] -87.6 -87.6 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:200274] 41.9 42 41.9 41.9 41.8 ...
##  $ end_lng           : num [1:200274] -87.6 -87.7 -87.7 -87.6 -87.6 ...
##  $ member_casual     : chr [1:200274] "member" "casual" "casual" "casual"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )
```

```
str(tripdata_2020_06)
```

```
## spec_tbl_df [343,005 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:343005] "8CD5DE2C2B6C4CFC"
```

```
"9A191EB2C751D85D" "F37D14B0B5659BCF" "C41237B506E85FA1" ...
##  $ rideable_type     : chr [1:343005] "docked_bike" "docked_bike"
"docked_bike" "docked_bike" ...
##  $ started_at        : POSIXct[1:343005], format: "2020-06-13 23:24:48"
"2020-06-26 07:26:10" ...
##  $ ended_at          : POSIXct[1:343005], format: "2020-06-13 23:36:55"
"2020-06-26 07:31:58" ...
##  $ start_station_name: chr [1:343005] "Wilton Ave & Belmont Ave" "Federal
St & Polk St" "Daley Center Plaza" "Broadway & Cornelia Ave" ...
##  $ start_station_id  : chr [1:343005] "117" "41" "81" "303" ...
##  $ end_station_name  : chr [1:343005] "Damen Ave & Clybourn Ave" "Daley
Center Plaza" "State St & Harrison St" "Broadway & Berwyn Ave" ...
##  $ end_station_id    : chr [1:343005] "163" "81" "5" "294" ...
##  $ start_lat         : num [1:343005] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:343005] -87.7 -87.6 -87.6 -87.6 -87.7 ...
##  $ end_lat           : num [1:343005] 41.9 41.9 41.9 42 41.9 ...
##  $ end_lng           : num [1:343005] -87.7 -87.6 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:343005] "casual" "member" "member" "casual"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )
```

```
str(tripdata_2020_07)
```

```
## spec_tbl_df [551,480 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:551480] "762198876D69004D"
"BEC9C9FBA0D4CF1B" "D2FD8EA432C77EC1" "54AE594E20B35881" ...
##  $ rideable_type     : chr [1:551480] "docked_bike" "docked_bike"
"docked_bike" "docked_bike" ...
##  $ started_at        : POSIXct[1:551480], format: "2020-07-09 15:22:02"
"2020-07-24 23:56:30" ...
##  $ ended_at          : POSIXct[1:551480], format: "2020-07-09 15:25:52"
"2020-07-25 00:20:17" ...
##  $ start_station_name: chr [1:551480] "Ritchie Ct & Banks St" "Halsted St
& Roscoe St" "Lake Shore Dr & Diversey Pkwy" "LaSalle St & Illinois St" ...
##  $ start_station_id  : chr [1:551480] "180" "299" "329" "181" ...
```

```
##  $ end_station_name  : chr [1:551480] "Wells St & Evergreen Ave" "Broadway
& Ridge Ave" "Clark St & Wellington Ave" "Clark St & Armitage Ave" ...
##  $ end_station_id    : chr [1:551480] "291" "461" "156" "94" ...
##  $ start_lat         : num [1:551480] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:551480] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:551480] 41.9 42 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:551480] -87.6 -87.7 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:551480] "member" "member" "casual" "casual"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )
```

```
str(tripdata_2020_08)
```

```
## spec_tbl_df [622,361 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:622361] "322BD23D287743ED"
"2A3AEF1AB9054D8B" "67DC1D133E8B5816" "C79FBBD412E578A7" ...
##  $ rideable_type     : chr [1:622361] "docked_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : POSIXct[1:622361], format: "2020-08-20 18:08:14"
"2020-08-27 18:46:04" ...
##  $ ended_at          : POSIXct[1:622361], format: "2020-08-20 18:17:51"
"2020-08-27 19:54:51" ...
##  $ start_station_name: chr [1:622361] "Lake Shore Dr & Diversey Pkwy"
"Michigan Ave & 14th St" "Columbus Dr & Randolph St" "Daley Center Plaza" ...
##  $ start_station_id  : chr [1:622361] "329" "168" "195" "81" ...
##  $ end_station_name  : chr [1:622361] "Clark St & Lincoln Ave" "Michigan
Ave & 14th St" "State St & Randolph St" "State St & Kinzie St" ...
##  $ end_station_id    : chr [1:622361] "141" "168" "44" "47" ...
##  $ start_lat         : num [1:622361] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:622361] -87.6 -87.6 -87.6 -87.6 -87.7 ...
##  $ end_lat           : num [1:622361] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:622361] -87.6 -87.6 -87.6 -87.6 -87.7 ...
##  $ member_casual     : chr [1:622361] "member" "casual" "casual" "casual"
...
##  - attr(*, "spec")=
```

```
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )
```

```
str(tripdata_2020_09)
```

```
## spec_tbl_df [532,958 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:532958] "2B22BD5F95FB2629"
"A7FB70B4AFC6CAF2" "86057FA01BAC778E" "57F6DC9A153DB98C" ...
##  $ rideable_type     : chr [1:532958] "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : POSIXct[1:532958], format: "2020-09-17 14:27:11"
"2020-09-17 15:07:31" ...
##  $ ended_at          : POSIXct[1:532958], format: "2020-09-17 14:44:24"
"2020-09-17 15:07:45" ...
##  $ start_station_name: chr [1:532958] "Michigan Ave & Lake St" "W Oakdale
Ave & N Broadway" "W Oakdale Ave & N Broadway" "Ashland Ave & Belle Plaine
Ave" ...
##  $ start_station_id  : chr [1:532958] "52" NA NA "246" ...
##  $ end_station_name  : chr [1:532958] "Green St & Randolph St" "W Oakdale
Ave & N Broadway" "W Oakdale Ave & N Broadway" "Montrose Harbor" ...
##  $ end_station_id    : chr [1:532958] "112" NA NA "249" ...
##  $ start_lat         : num [1:532958] 41.9 41.9 41.9 42 41.9 ...
##  $ start_lng         : num [1:532958] -87.6 -87.6 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num [1:532958] 41.9 41.9 41.9 42 41.9 ...
##  $ end_lng           : num [1:532958] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:532958] "casual" "casual" "casual" "casual"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )

str(tripdata_2020_10)

## spec_tbl_df [388,653 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:388653] "ACB6B40CF5B9044C"
"DF450C72FD109C01" "B6396B54A15AC0DF" "44A4AEE261B9E854" ...
##  $ rideable_type     : chr [1:388653] "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : POSIXct[1:388653], format: "2020-10-31 19:39:43"
"2020-10-31 23:50:08" ...
##  $ ended_at          : POSIXct[1:388653], format: "2020-10-31 19:57:12"
"2020-11-01 00:04:16" ...
##  $ start_station_name: chr [1:388653] "Lakeview Ave & Fullerton Pkwy"
"Southport Ave & Waveland Ave" "Stony Island Ave & 67th St" "Clark St & Grace
St" ...
##  $ start_station_id  : chr [1:388653] "313" "227" "102" "165" ...
##  $ end_station_name  : chr [1:388653] "Rush St & Hubbard St" "Kedzie Ave &
Milwaukee Ave" "University Ave & 57th St" "Broadway & Sheridan Rd" ...
##  $ end_station_id    : chr [1:388653] "125" "260" "423" "256" ...
##  $ start_lat         : num [1:388653] 41.9 41.9 41.8 42 41.9 ...
##  $ start_lng         : num [1:388653] -87.6 -87.7 -87.6 -87.7 -87.7 ...
##  $ end_lat           : num [1:388653] 41.9 41.9 41.8 42 41.9 ...
##  $ end_lng           : num [1:388653] -87.6 -87.7 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:388653] "casual" "casual" "casual" "casual"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )

str(tripdata_2020_11)
```

```
## spec_tbl_df [259,716 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:259716] "BD0A6FF6FFF9B921"
"96A7A7A4BDE4F82D" "C61526D06582BDC5" "E533E89C32080B9E" ...
##  $ rideable_type     : chr [1:259716] "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : POSIXct[1:259716], format: "2020-11-01 13:36:00"
"2020-11-01 10:03:26" ...
##  $ ended_at          : POSIXct[1:259716], format: "2020-11-01 13:45:40"
"2020-11-01 10:14:45" ...
##  $ start_station_name: chr [1:259716] "Dearborn St & Erie St" "Franklin St
& Illinois St" "Lake Shore Dr & Monroe St" "Leavitt St & Chicago Ave" ...
##  $ start_station_id  : chr [1:259716] "110" "672" "76" "659" ...
##  $ end_station_name  : chr [1:259716] "St. Clair St & Erie St" "Noble St &
Milwaukee Ave" "Federal St & Polk St" "Stave St & Armitage Ave" ...
##  $ end_station_id    : chr [1:259716] "211" "29" "41" "185" ...
##  $ start_lat         : num [1:259716] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:259716] -87.6 -87.6 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num [1:259716] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:259716] -87.6 -87.7 -87.6 -87.7 -87.6 ...
##  $ member_casual     : chr [1:259716] "casual" "casual" "casual" "casual"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )
```

Combine all the individual data frame into one big data frame

```
all_tripdata <-
bind_rows(tripdata_2020_04,tripdata_2020_05,tripdata_2020_06,tripdata_2020_07
,

tripdata_2020_08,tripdata_2020_09,tripdata_2020_10,tripdata_2020_11,

tripdata_2020_12,tripdata_2021_01,tripdata_2021_02,tripdata_2021_03,
                          tripdata_2021_04)
```

STEP3 : CLEAN UP AND ADD DATA TO PREPARE FOR ANALYSIS

Inspect the new table that has been created

```
colnames(all_tripdata)
```

```
##  [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_tripdata)
```

```
## [1] 3826978
```

```
dim(all_tripdata)
```

```
## [1] 3826978      13
```

```
head(all_tripdata)
```

```
## # A tibble: 6 x 13
##   ride_id rideable_type started_at          ended_at
start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 A847FA~ docked_bike   2020-04-26 17:45:14 2020-04-26 18:12:03 Eckhart
Park
## 2 5405B8~ docked_bike   2020-04-17 17:08:54 2020-04-17 17:17:03 Drake Ave
& Ful~
## 3 5DD24A~ docked_bike   2020-04-01 17:54:13 2020-04-01 18:08:36 McClurg Ct
& Er~
## 4 2A59BB~ docked_bike   2020-04-07 12:50:19 2020-04-07 13:02:31 California
Ave ~
## 5 27AD30~ docked_bike   2020-04-18 10:22:59 2020-04-18 11:15:54 Rush St &
Hubba~
## 6 356216~ docked_bike   2020-04-30 17:55:47 2020-04-30 18:01:11 Mies van
der Ro~
## # ... with 8 more variables: start_station_id <chr>, end_station_name
<chr>,
## #   end_station_id <chr>, start_lat <dbl>, start_lng <dbl>, end_lat <dbl>,
## #   end_lng <dbl>, member_casual <chr>
```

```
str(all_tripdata)
```

```
## spec_tbl_df [3,826,978 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:3826978] "A847FADBBC638E45"
"5405B80E996FF60D" "5DD24A79A4E006F4" "2A59BBDF5CDBA725" ...
##  $ rideable_type     : chr [1:3826978] "docked_bike" "docked_bike"
"docked_bike" "docked_bike" ...
##  $ started_at        : POSIXct[1:3826978], format: "2020-04-26 17:45:14"
"2020-04-17 17:08:54" ...
##  $ ended_at          : POSIXct[1:3826978], format: "2020-04-26 18:12:03"
"2020-04-17 17:17:03" ...
```

```
##  $ start_station_name: chr [1:3826978] "Eckhart Park" "Drake Ave &
Fullerton Ave" "McClurg Ct & Erie St" "California Ave & Division St" ...
##  $ start_station_id  : chr [1:3826978] "86" "503" "142" "216" ...
##  $ end_station_name  : chr [1:3826978] "Lincoln Ave & Diversey Pkwy"
"Kosciuszko Park" "Indiana Ave & Roosevelt Rd" "Wood St & Augusta Blvd" ...
##  $ end_station_id    : chr [1:3826978] "152" "499" "255" "657" ...
##  $ start_lat         : num [1:3826978] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:3826978] -87.7 -87.7 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num [1:3826978] 41.9 41.9 41.9 41.9 42 ...
##  $ end_lng           : num [1:3826978] -87.7 -87.7 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:3826978] "member" "member" "member" "member"
...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )
```

The data can only be aggregated at the ride-level, which is too granular. We will want to add some additional columns of data such as day, month, year – that provide additional opportunities to aggregate the data.

```
all_tripdata$date <- as.Date(all_tripdata$started_at) #The default format is
yyyy-mm-dd
all_tripdata$month <- format(as.Date(all_tripdata$date), "%m")
all_tripdata$day <- format(as.Date(all_tripdata$date), "%d")
all_tripdata$year <- format(as.Date(all_tripdata$date), "%Y")
all_tripdata$day_of_week <- format(as.Date(all_tripdata$date), "%A")
```

Add a “ride_length” calculation to all_trips (in seconds)

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

Inspect the structure of the columns

```
str(all_tripdata)

## spec_tbl_df [3,826,978 x 19] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:3826978] "A847FADBBC638E45"
```

```
"5405B80E996FF60D" "5DD24A79A4E006F4" "2A59BBDF5CDBA725" ...
##  $ rideable_type     : chr [1:3826978] "docked_bike" "docked_bike"
"docked_bike" "docked_bike" ...
##  $ started_at        : POSIXct[1:3826978], format: "2020-04-26 17:45:14"
"2020-04-17 17:08:54" ...
##  $ ended_at          : POSIXct[1:3826978], format: "2020-04-26 18:12:03"
"2020-04-17 17:17:03" ...
##  $ start_station_name: chr [1:3826978] "Eckhart Park" "Drake Ave &
Fullerton Ave" "McClurg Ct & Erie St" "California Ave & Division St" ...
##  $ start_station_id  : chr [1:3826978] "86" "503" "142" "216" ...
##  $ end_station_name  : chr [1:3826978] "Lincoln Ave & Diversey Pkwy"
"Kosciuszko Park" "Indiana Ave & Roosevelt Rd" "Wood St & Augusta Blvd" ...
##  $ end_station_id    : chr [1:3826978] "152" "499" "255" "657" ...
##  $ start_lat         : num [1:3826978] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:3826978] -87.7 -87.7 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num [1:3826978] 41.9 41.9 41.9 41.9 42 ...
##  $ end_lng           : num [1:3826978] -87.7 -87.7 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:3826978] "member" "member" "member" "member"
...
##  $ date              : Date[1:3826978], format: "2020-04-26" "2020-04-17"
...
##  $ month             : chr [1:3826978] "04" "04" "04" "04" ...
##  $ day               : chr [1:3826978] "26" "17" "01" "07" ...
##  $ year              : chr [1:3826978] "2020" "2020" "2020" "2020" ...
##  $ day_of_week       : chr [1:3826978] "Sunday" "Friday" "Wednesday"
"Tuesday" ...
##  $ ride_length       : 'difftime' num [1:3826978] 1609 489 863 732 ...
##   ..- attr(*, "units")= chr "secs"
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   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()
##   .. )
```

Convert “ride_length” from factor to numeric so we can run calculations on the data

```
is.factor(all_tripdata$ride_length)                    #Checks to see if data
type is a factor
```

```
## [1] FALSE

all_tripdata$ride_length <-
as.numeric(as.character(all_tripdata$ride_length)) #Conversion to numeric
is.numeric(all_tripdata$ride_length)       #Checks to see if the data is
numeric

## [1] TRUE
```

Remove “bad” data The dataframe includes a few hundred entries when bikes were taken out of docks and checked for quality by Divvy or ride_length was negative

```
summary(all_tripdata$ride_length)

##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -1742998      468      861     1483     1583  3523202
```

We will create a new version of the dataframe (v2) since data is being removed

```
all_trip_data_v2 <- all_tripdata[!(all_tripdata$start_station_name == "HQ QR"
| all_tripdata$ride_length<0),]
```

STEP 4: CONDUCT DESCRIPTIVE ANALYSIS

Descriptive analysis on ride_length (all figures in seconds)

```
mean(all_trip_data_v2$ride_length, na.rm = TRUE)  #straight average (total
ride length / rides)

## [1] 1685.571

median(all_trip_data_v2$ride_length, na.rm = TRUE)#midpoint number in the
ascending array of ride length

## [1] 874

max(all_trip_data_v2$ride_length, na.rm = TRUE)        #longest ride

## [1] 3523202

min(all_trip_data_v2$ride_length, na.rm = TRUE)        #shortest

## [1] 0
```

You can condense the four lines above to one line using summary() on the specific attribute

```
summary(all_trip_data_v2$ride_length, na.rm = TRUE)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##       0     479     874    1686    1602 3523202  148183
```

Compare members and casual users

```
aggregate(all_trip_data_v2$ride_length ~ all_trip_data_v2$member_casual, FUN
= mean)
```

```
##   all_trip_data_v2$member_casual all_trip_data_v2$ride_length
## 1                         casual                    2717.6094
## 2                         member                     966.1052
```

```
aggregate(all_trip_data_v2$ride_length ~ all_trip_data_v2$member_casual, FUN
= median)
```

```
##   all_trip_data_v2$member_casual all_trip_data_v2$ride_length
## 1                         casual                         1278
## 2                         member                          689
```

```
aggregate(all_trip_data_v2$ride_length ~ all_trip_data_v2$member_casual, FUN
= max)
```

```
##   all_trip_data_v2$member_casual all_trip_data_v2$ride_length
## 1                         casual                      3341033
## 2                         member                      3523202
```

```
aggregate(all_trip_data_v2$ride_length ~ all_trip_data_v2$member_casual, FUN
= min)
```

```
##   all_trip_data_v2$member_casual all_trip_data_v2$ride_length
## 1                         casual                            0
## 2                         member                            0
```

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

```
aggregate(all_trip_data_v2$ride_length ~ all_trip_data_v2$member_casual +
all_trip_data_v2$day_of_week,
          FUN = mean)
```

```
##    all_trip_data_v2$member_casual all_trip_data_v2$day_of_week
## 1                          casual                       Friday
## 2                          member                       Friday
## 3                          casual                       Monday
## 4                          member                       Monday
## 5                          casual                     Saturday
## 6                          member                     Saturday
## 7                          casual                       Sunday
## 8                          member                       Sunday
## 9                          casual                     Thursday
## 10                         member                     Thursday
## 11                         casual                      Tuesday
## 12                         member                      Tuesday
## 13                         casual                    Wednesday
## 14                         member                    Wednesday
##    all_trip_data_v2$ride_length
## 1                     2615.8820
## 2                      945.1852
## 3                     2715.2761
## 4                      921.1990
## 5                     2817.0078
```

```
## 6                      1070.2137
## 7                      3053.4045
## 8                      1095.5406
## 9                      2562.8716
## 10                      908.3416
## 11                     2477.4782
## 12                      909.7689
## 13                     2466.9220
## 14                      915.2750
```

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

```
all_trip_data_v2$day_of_week <- ordered(all_trip_data_v2$day_of_week,
levels=c("Sunday", "Monday","Tuesday", "Wednesday", "Thursday", "Friday",
"Saturday"))
```

Now, let's run the average ride time by each day for members vs casual users

```
aggregate(all_trip_data_v2$ride_length ~ all_trip_data_v2$member_casual +
all_trip_data_v2$day_of_week,
          FUN = mean)

##    all_trip_data_v2$member_casual all_trip_data_v2$day_of_week
## 1                          casual                       Sunday
## 2                          member                       Sunday
## 3                          casual                       Monday
## 4                          member                       Monday
## 5                          casual                      Tuesday
## 6                          member                      Tuesday
## 7                          casual                    Wednesday
## 8                          member                    Wednesday
## 9                          casual                     Thursday
## 10                         member                     Thursday
## 11                         casual                       Friday
## 12                         member                       Friday
## 13                         casual                     Saturday
## 14                         member                     Saturday
##    all_trip_data_v2$ride_length
## 1                     3053.4045
## 2                     1095.5406
## 3                     2715.2761
## 4                      921.1990
## 5                     2477.4782
## 6                      909.7689
## 7                     2466.9220
## 8                      915.2750
## 9                     2562.8716
## 10                     908.3416
## 11                    2615.8820
## 12                     945.1852
```

```
## 13                    2817.0078
## 14                    1070.2137
```

analyze ridership data by type and weekday

```
all_trip_data_v2 %>%
      mutate(weekday = wday(started_at, label = TRUE)) %>% #creates weekday
field using wday()
      group_by(member_casual, weekday) %>%                 #groups by
usertype and weekday
      summarise(number_of_rides = n()                     #calculates the
number of rides and average duration
      ,average_duration = mean(ride_length)) %>%          # calculates the
average duration
      arrange(member_casual, weekday)
```

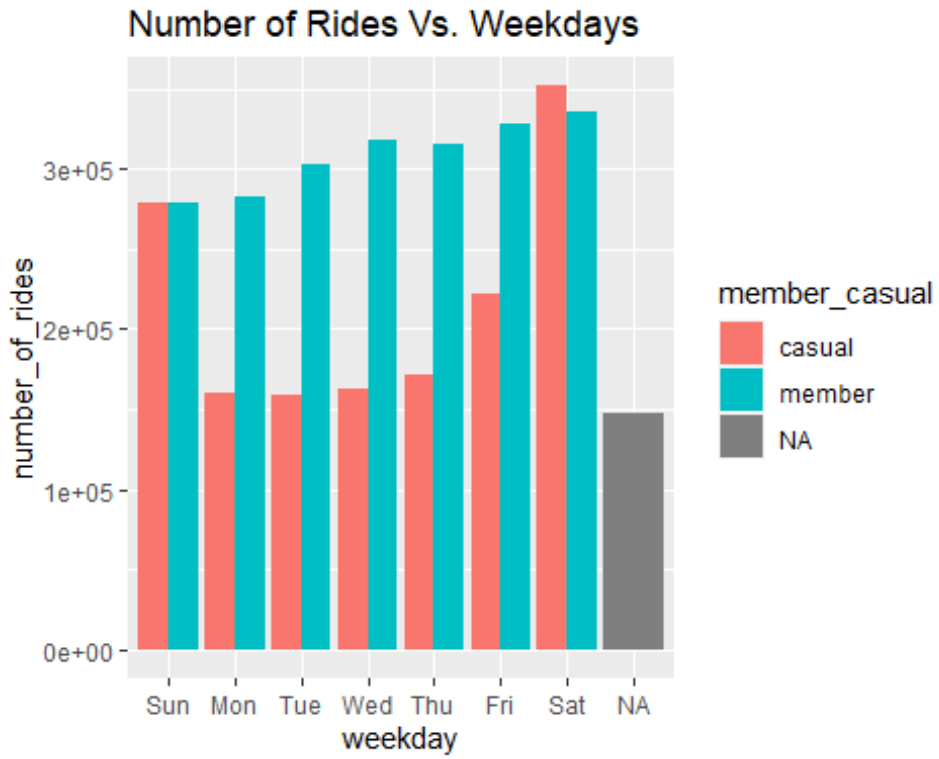
```
## `summarise()` has grouped output by 'member_casual'. You can override
using the `.groups` argument.
```

```
## # A tibble: 15 x 4
## # Groups:   member_casual [3]
##    member_casual weekday number_of_rides average_duration
##    <chr>         <ord>             <int>            <dbl>
##  1 casual        Sun              278706            3053.
##  2 casual        Mon              160437            2715.
##  3 casual        Tue              158627            2477.
##  4 casual        Wed              163295            2467.
##  5 casual        Thu              171461            2563.
##  6 casual        Fri              222434            2616.
##  7 casual        Sat              351844            2817.
##  8 member        Sun              278734            1096.
##  9 member        Mon              282837             921.
## 10 member        Tue              303177             910.
## 11 member        Wed              317753             915.
## 12 member        Thu              315424             908.
## 13 member        Fri              327945             945.
## 14 member        Sat              335564            1070.
## 15 <NA>          <NA>             148183              NA
```

Let's visualize the number of rides by rider type

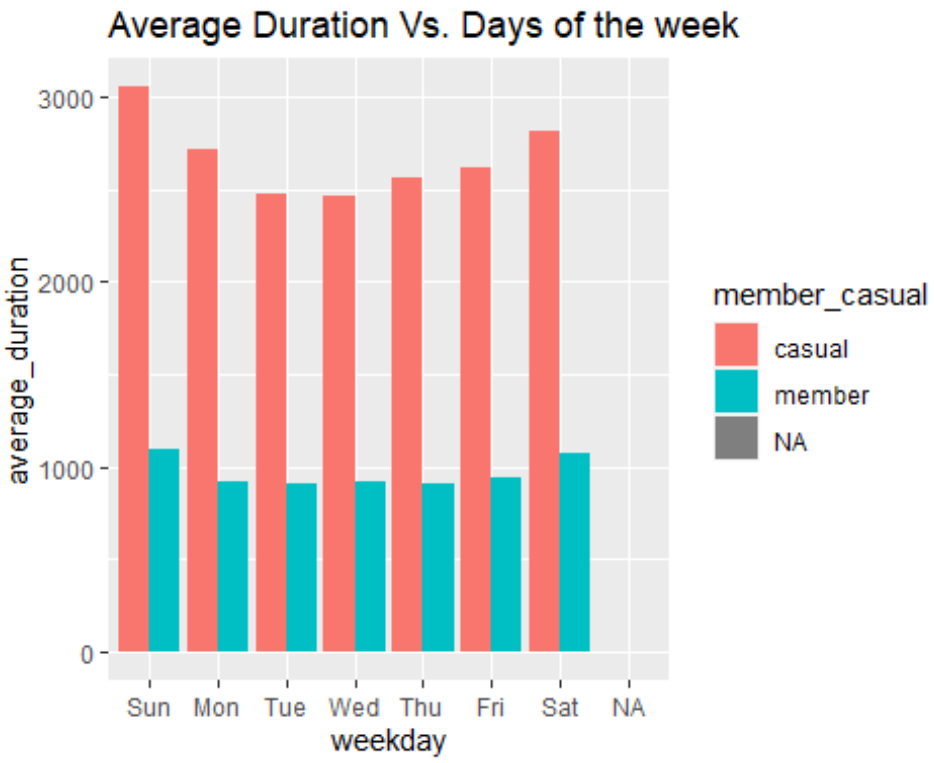
```
all_trip_data_v2 %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n()
            ,average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday) %>%
  ggplot(aes(x = weekday, y = number_of_rides, fill = member_casual)) +
  geom_col(position = "dodge") +
  labs(title ="Number of Rides Vs. Weekdays")
```

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



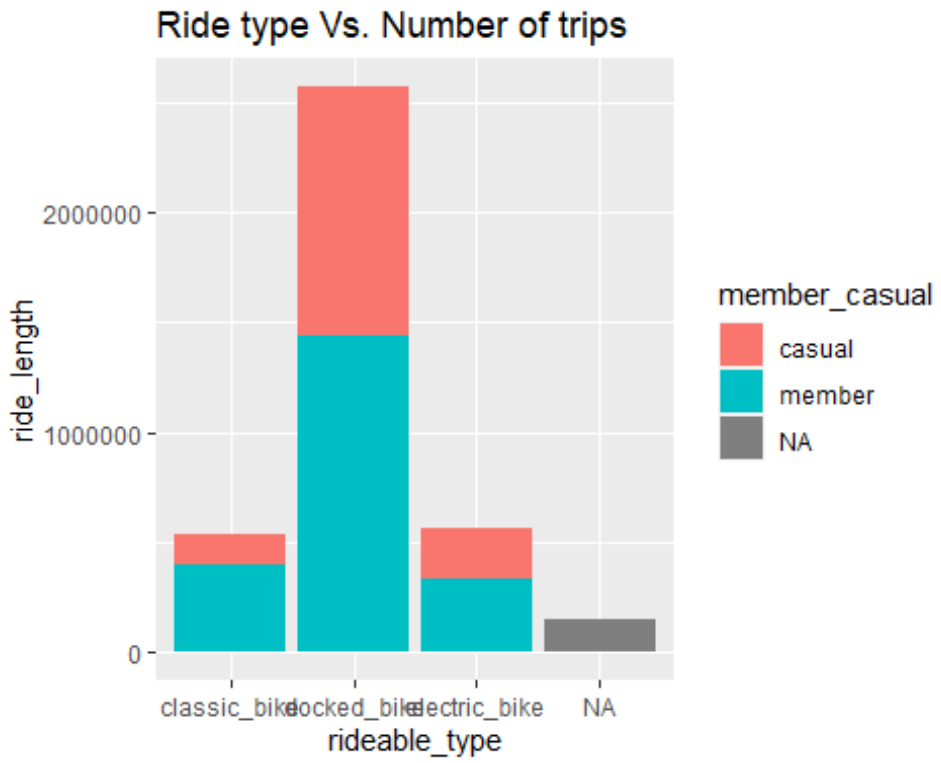
Let's create a visualization for average duration

```
all_trip_data_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")+
  labs(title ="Average Duration Vs. Days of the week")

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

## Warning: Removed 1 rows containing missing values (geom_col).
```



Visualization of ride type Vs. number of trips by customer type

```
all_trip_data_v2 %>%
  group_by(rideable_type, member_casual) %>%
  summarise(ride_length = n()) %>%
  ggplot(aes(x= rideable_type, y=ride_length, fill= member_casual))+
              geom_bar(stat='identity') +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE)) +
  labs(title ="Ride type Vs. Number of trips")

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



Key Takeaways:

- Members use bikes consistently throughout the week while casual customers use them more during weekends.
- Average trip duration of casual riders is more for casual customers than member riders.
- Members prefer classic bikes while casual rider prefer docked bikes.

Recommendations:

- Offer any promotions on Thursdays and Saturdays as bicycles are under-utilized on these days.
- Offer discounted membership fee for renewals after the first year. It might nudge casual riders to take up membership.
- Offer discounted pricing during non-busy hours/days (Thursday) so that casual riders might choose to use bikes more often and level out demand over the day

Additional data that could expand scope of the analysis:

- Age and gender profile. This data could be used to study the category of riders who can be targeted for attracting new members.
- Use latitude/longitude data to create map-like visuals to examine trip traffic by station.