Divvy\_Bikeshare\_Analysis

Srikanth Gadiparthi

2022-06-12

## Ask

### 1.1 Background

Cyclistic developed a successful bike-share program in 2016. Since then, the initiative has evolved to a fleet of 5,824 tracked bicycles that are locked into a network of 692 stations around Chicago. The bikes can be unlocked at any time from any station and returned to any other station in the system. Casual riders are customers who purchase single-ride or full-day passes. Cyclistic members are customers who pay yearly memberships. Annual members are far more profitable than casual riders, according to Cyclistic’s financial experts. Rather of developing an all-new marketing program, the director of marketing feels there is a strong potential of converting casual passengers into members.

### Task

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

## Prepare

* This is a public data made available by Motivate International Inc under this [license](https://ride.divvybikes.com/data-license-agreement)
* This raw data was organized in separate files according to months and years, this is good ROCC data.
* For this analysis, the last 12months of data were used.

## Process

For this analysis, R programming tool was used, as the given data sets were to large to process in excel or spreadsheet.

#### Installing Required Packages:

* install.packages(“tidyverse”)
* install.packages(“lubridate”)
* install.packages(“ggplot2”)

#### Loading packages:

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.7 ✔ dplyr 1.0.9  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0  
## ✔ readr 2.1.2 ✔ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library (lubridate)

##   
## Attaching package: 'lubridate'

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

library(ggplot2)

#### Uploading datasets:

month\_06\_2021<- read\_csv("202106-divvy-tripdata.csv")

## Rows: 729595 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

month\_07\_2021<- read\_csv("202107-divvy-tripdata.csv")

## Rows: 822410 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

month\_08\_2021<- read\_csv("202108-divvy-tripdata.csv")

## Rows: 804352 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

month\_09\_2021<- read\_csv("202109-divvy-tripdata.csv")

## Rows: 756147 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

month\_10\_2021<- read\_csv("202110-divvy-tripdata.csv")

## Rows: 631226 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

month\_11\_2021<- read\_csv("202111-divvy-tripdata.csv")

## Rows: 359978 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

month\_12\_2021<- read\_csv("202112-divvy-tripdata.csv")

## Rows: 247540 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

month\_01\_2022<- read\_csv("202201-divvy-tripdata.csv")

## Rows: 103770 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

month\_02\_2022<- read\_csv("202202-divvy-tripdata.csv")

## Rows: 115609 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

month\_03\_2022<- read\_csv("202203-divvy-tripdata.csv")

## Rows: 284042 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

month\_04\_2022<- read\_csv("202204-divvy-tripdata.csv")

## Rows: 371249 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

month\_05\_2022<- read\_csv("202205-divvy-tripdata.csv")

## Rows: 634858 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#### Looking for data integrity and incongruencies:

colnames(month\_06\_2021)

## [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(month\_07\_2021)

## [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(month\_08\_2021)

## [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(month\_09\_2021)

## [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(month\_10\_2021)

## [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(month\_11\_2021)

## [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(month\_12\_2021)

## [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(month\_01\_2022)

## [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(month\_02\_2022)

## [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(month\_03\_2022)

## [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(month\_04\_2022)

## [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(month\_05\_2022)

## [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"

str(month\_06\_2021)

## spec\_tbl\_df [729,595 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:729595] "99FEC93BA843FB20" "06048DCFC8520CAF" "9598066F68045DF2" "B03C0FE48C412214" ...  
## $ rideable\_type : chr [1:729595] "electric\_bike" "electric\_bike" "electric\_bike" "electric\_bike" ...  
## $ started\_at : POSIXct[1:729595], format: "2021-06-13 14:31:28" "2021-06-04 11:18:02" ...  
## $ ended\_at : POSIXct[1:729595], format: "2021-06-13 14:34:11" "2021-06-04 11:24:19" ...  
## $ start\_station\_name: chr [1:729595] NA NA NA NA ...  
## $ start\_station\_id : chr [1:729595] NA NA NA NA ...  
## $ end\_station\_name : chr [1:729595] NA NA NA NA ...  
## $ end\_station\_id : chr [1:729595] NA NA NA NA ...  
## $ start\_lat : num [1:729595] 41.8 41.8 41.8 41.8 41.8 ...  
## $ start\_lng : num [1:729595] -87.6 -87.6 -87.6 -87.6 -87.6 ...  
## $ end\_lat : num [1:729595] 41.8 41.8 41.8 41.8 41.8 ...  
## $ end\_lng : num [1:729595] -87.6 -87.6 -87.6 -87.6 -87.6 ...  
## $ member\_casual : chr [1:729595] "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()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(month\_07\_2021)

## spec\_tbl\_df [822,410 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:822410] "0A1B623926EF4E16" "B2D5583A5A5E76EE" "6F264597DDBF427A" "379B58EAB20E8AA5" ...  
## $ rideable\_type : chr [1:822410] "docked\_bike" "classic\_bike" "classic\_bike" "classic\_bike" ...  
## $ started\_at : POSIXct[1:822410], format: "2021-07-02 14:44:36" "2021-07-07 16:57:42" ...  
## $ ended\_at : POSIXct[1:822410], format: "2021-07-02 15:19:58" "2021-07-07 17:16:09" ...  
## $ start\_station\_name: chr [1:822410] "Michigan Ave & Washington St" "California Ave & Cortez St" "Wabash Ave & 16th St" "California Ave & Cortez St" ...  
## $ start\_station\_id : chr [1:822410] "13001" "17660" "SL-012" "17660" ...  
## $ end\_station\_name : chr [1:822410] "Halsted St & North Branch St" "Wood St & Hubbard St" "Rush St & Hubbard St" "Carpenter St & Huron St" ...  
## $ end\_station\_id : chr [1:822410] "KA1504000117" "13432" "KA1503000044" "13196" ...  
## $ start\_lat : num [1:822410] 41.9 41.9 41.9 41.9 41.9 ...  
## $ start\_lng : num [1:822410] -87.6 -87.7 -87.6 -87.7 -87.7 ...  
## $ end\_lat : num [1:822410] 41.9 41.9 41.9 41.9 41.9 ...  
## $ end\_lng : num [1:822410] -87.6 -87.7 -87.6 -87.7 -87.7 ...  
## $ member\_casual : chr [1:822410] "casual" "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()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(month\_08\_2021)

## spec\_tbl\_df [804,352 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:804352] "99103BB87CC6C1BB" "EAFCCCFB0A3FC5A1" "9EF4F46C57AD234D" "5834D3208BFAF1DA" ...  
## $ rideable\_type : chr [1:804352] "electric\_bike" "electric\_bike" "electric\_bike" "electric\_bike" ...  
## $ started\_at : POSIXct[1:804352], format: "2021-08-10 17:15:49" "2021-08-10 17:23:14" ...  
## $ ended\_at : POSIXct[1:804352], format: "2021-08-10 17:22:44" "2021-08-10 17:39:24" ...  
## $ start\_station\_name: chr [1:804352] NA NA NA NA ...  
## $ start\_station\_id : chr [1:804352] NA NA NA NA ...  
## $ end\_station\_name : chr [1:804352] NA NA NA NA ...  
## $ end\_station\_id : chr [1:804352] NA NA NA NA ...  
## $ start\_lat : num [1:804352] 41.8 41.8 42 42 41.8 ...  
## $ start\_lng : num [1:804352] -87.7 -87.7 -87.7 -87.7 -87.6 ...  
## $ end\_lat : num [1:804352] 41.8 41.8 42 42 41.8 ...  
## $ end\_lng : num [1:804352] -87.7 -87.6 -87.7 -87.7 -87.6 ...  
## $ member\_casual : chr [1:804352] "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()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(month\_09\_2021)

## spec\_tbl\_df [756,147 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:756147] "9DC7B962304CBFD8" "F930E2C6872D6B32" "6EF72137900BB910" "78D1DE133B3DBF55" ...  
## $ rideable\_type : chr [1:756147] "electric\_bike" "electric\_bike" "electric\_bike" "electric\_bike" ...  
## $ started\_at : POSIXct[1:756147], format: "2021-09-28 16:07:10" "2021-09-28 14:24:51" ...  
## $ ended\_at : POSIXct[1:756147], format: "2021-09-28 16:09:54" "2021-09-28 14:40:05" ...  
## $ start\_station\_name: chr [1:756147] NA NA NA NA ...  
## $ start\_station\_id : chr [1:756147] NA NA NA NA ...  
## $ end\_station\_name : chr [1:756147] NA NA NA NA ...  
## $ end\_station\_id : chr [1:756147] NA NA NA NA ...  
## $ start\_lat : num [1:756147] 41.9 41.9 41.8 41.8 41.9 ...  
## $ start\_lng : num [1:756147] -87.7 -87.6 -87.7 -87.7 -87.7 ...  
## $ end\_lat : num [1:756147] 41.9 42 41.8 41.8 41.9 ...  
## $ end\_lng : num [1:756147] -87.7 -87.7 -87.7 -87.7 -87.7 ...  
## $ member\_casual : chr [1:756147] "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()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(month\_10\_2021)

## spec\_tbl\_df [631,226 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:631226] "620BC6107255BF4C" "4471C70731AB2E45" "26CA69D43D15EE14" "362947F0437E1514" ...  
## $ rideable\_type : chr [1:631226] "electric\_bike" "electric\_bike" "electric\_bike" "electric\_bike" ...  
## $ started\_at : POSIXct[1:631226], format: "2021-10-22 12:46:42" "2021-10-21 09:12:37" ...  
## $ ended\_at : POSIXct[1:631226], format: "2021-10-22 12:49:50" "2021-10-21 09:14:14" ...  
## $ start\_station\_name: chr [1:631226] "Kingsbury St & Kinzie St" NA NA NA ...  
## $ start\_station\_id : chr [1:631226] "KA1503000043" NA NA NA ...  
## $ end\_station\_name : chr [1:631226] NA NA NA NA ...  
## $ end\_station\_id : chr [1:631226] NA NA NA NA ...  
## $ start\_lat : num [1:631226] 41.9 41.9 41.9 41.9 41.9 ...  
## $ start\_lng : num [1:631226] -87.6 -87.7 -87.7 -87.7 -87.7 ...  
## $ end\_lat : num [1:631226] 41.9 41.9 41.9 41.9 41.9 ...  
## $ end\_lng : num [1:631226] -87.6 -87.7 -87.7 -87.7 -87.7 ...  
## $ member\_casual : chr [1:631226] "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()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(month\_11\_2021)

## spec\_tbl\_df [359,978 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:359978] "7C00A93E10556E47" "90854840DFD508BA" "0A7D10CDD144061C" "2F3BE33085BCFF02" ...  
## $ rideable\_type : chr [1:359978] "electric\_bike" "electric\_bike" "electric\_bike" "electric\_bike" ...  
## $ started\_at : POSIXct[1:359978], format: "2021-11-27 13:27:38" "2021-11-27 13:38:25" ...  
## $ ended\_at : POSIXct[1:359978], format: "2021-11-27 13:46:38" "2021-11-27 13:56:10" ...  
## $ start\_station\_name: chr [1:359978] NA NA NA NA ...  
## $ start\_station\_id : chr [1:359978] NA NA NA NA ...  
## $ end\_station\_name : chr [1:359978] NA NA NA NA ...  
## $ end\_station\_id : chr [1:359978] NA NA NA NA ...  
## $ start\_lat : num [1:359978] 41.9 42 42 41.9 41.9 ...  
## $ start\_lng : num [1:359978] -87.7 -87.7 -87.7 -87.8 -87.6 ...  
## $ end\_lat : num [1:359978] 42 41.9 42 41.9 41.9 ...  
## $ end\_lng : num [1:359978] -87.7 -87.7 -87.7 -87.8 -87.6 ...  
## $ member\_casual : chr [1:359978] "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()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(month\_12\_2021)

## spec\_tbl\_df [247,540 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:247540] "46F8167220E4431F" "73A77762838B32FD" "4CF42452054F59C5" "3278BA87BF698339" ...  
## $ rideable\_type : chr [1:247540] "electric\_bike" "electric\_bike" "electric\_bike" "classic\_bike" ...  
## $ started\_at : POSIXct[1:247540], format: "2021-12-07 15:06:07" "2021-12-11 03:43:29" ...  
## $ ended\_at : POSIXct[1:247540], format: "2021-12-07 15:13:42" "2021-12-11 04:10:23" ...  
## $ start\_station\_name: chr [1:247540] "Laflin St & Cullerton St" "LaSalle Dr & Huron St" "Halsted St & North Branch St" "Halsted St & North Branch St" ...  
## $ start\_station\_id : chr [1:247540] "13307" "KP1705001026" "KA1504000117" "KA1504000117" ...  
## $ end\_station\_name : chr [1:247540] "Morgan St & Polk St" "Clarendon Ave & Leland Ave" "Broadway & Barry Ave" "LaSalle Dr & Huron St" ...  
## $ end\_station\_id : chr [1:247540] "TA1307000130" "TA1307000119" "13137" "KP1705001026" ...  
## $ start\_lat : num [1:247540] 41.9 41.9 41.9 41.9 41.9 ...  
## $ start\_lng : num [1:247540] -87.7 -87.6 -87.6 -87.6 -87.7 ...  
## $ end\_lat : num [1:247540] 41.9 42 41.9 41.9 41.9 ...  
## $ end\_lng : num [1:247540] -87.7 -87.7 -87.6 -87.6 -87.6 ...  
## $ member\_casual : chr [1:247540] "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()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(month\_01\_2022)

## spec\_tbl\_df [103,770 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:103770] "C2F7DD78E82EC875" "A6CF8980A652D272" "BD0F91DFF741C66D" "CBB80ED419105406" ...  
## $ rideable\_type : chr [1:103770] "electric\_bike" "electric\_bike" "classic\_bike" "classic\_bike" ...  
## $ started\_at : POSIXct[1:103770], format: "2022-01-13 11:59:47" "2022-01-10 08:41:56" ...  
## $ ended\_at : POSIXct[1:103770], format: "2022-01-13 12:02:44" "2022-01-10 08:46:17" ...  
## $ start\_station\_name: chr [1:103770] "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave" "Sheffield Ave & Fullerton Ave" "Clark St & Bryn Mawr Ave" ...  
## $ start\_station\_id : chr [1:103770] "525" "525" "TA1306000016" "KA1504000151" ...  
## $ end\_station\_name : chr [1:103770] "Clark St & Touhy Ave" "Clark St & Touhy Ave" "Greenview Ave & Fullerton Ave" "Paulina St & Montrose Ave" ...  
## $ end\_station\_id : chr [1:103770] "RP-007" "RP-007" "TA1307000001" "TA1309000021" ...  
## $ start\_lat : num [1:103770] 42 42 41.9 42 41.9 ...  
## $ start\_lng : num [1:103770] -87.7 -87.7 -87.7 -87.7 -87.6 ...  
## $ end\_lat : num [1:103770] 42 42 41.9 42 41.9 ...  
## $ end\_lng : num [1:103770] -87.7 -87.7 -87.7 -87.7 -87.6 ...  
## $ member\_casual : chr [1:103770] "casual" "casual" "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\_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()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(month\_02\_2022)

## spec\_tbl\_df [115,609 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:115609] "E1E065E7ED285C02" "1602DCDC5B30FFE3" "BE7DD2AF4B55C4AF" "A1789BDF844412BE" ...  
## $ rideable\_type : chr [1:115609] "classic\_bike" "classic\_bike" "classic\_bike" "classic\_bike" ...  
## $ started\_at : POSIXct[1:115609], format: "2022-02-19 18:08:41" "2022-02-20 17:41:30" ...  
## $ ended\_at : POSIXct[1:115609], format: "2022-02-19 18:23:56" "2022-02-20 17:45:56" ...  
## $ start\_station\_name: chr [1:115609] "State St & Randolph St" "Halsted St & Wrightwood Ave" "State St & Randolph St" "Southport Ave & Waveland Ave" ...  
## $ start\_station\_id : chr [1:115609] "TA1305000029" "TA1309000061" "TA1305000029" "13235" ...  
## $ end\_station\_name : chr [1:115609] "Clark St & Lincoln Ave" "Southport Ave & Wrightwood Ave" "Canal St & Adams St" "Broadway & Sheridan Rd" ...  
## $ end\_station\_id : chr [1:115609] "13179" "TA1307000113" "13011" "13323" ...  
## $ start\_lat : num [1:115609] 41.9 41.9 41.9 41.9 41.9 ...  
## $ start\_lng : num [1:115609] -87.6 -87.6 -87.6 -87.7 -87.6 ...  
## $ end\_lat : num [1:115609] 41.9 41.9 41.9 42 41.9 ...  
## $ end\_lng : num [1:115609] -87.6 -87.7 -87.6 -87.6 -87.6 ...  
## $ member\_casual : chr [1:115609] "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()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(month\_03\_2022)

## spec\_tbl\_df [284,042 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:284042] "47EC0A7F82E65D52" "8494861979B0F477" "EFE527AF80B66109" "9F446FD9DEE3F389" ...  
## $ rideable\_type : chr [1:284042] "classic\_bike" "electric\_bike" "classic\_bike" "classic\_bike" ...  
## $ started\_at : POSIXct[1:284042], format: "2022-03-21 13:45:01" "2022-03-16 09:37:16" ...  
## $ ended\_at : POSIXct[1:284042], format: "2022-03-21 13:51:18" "2022-03-16 09:43:34" ...  
## $ start\_station\_name: chr [1:284042] "Wabash Ave & Wacker Pl" "Michigan Ave & Oak St" "Broadway & Berwyn Ave" "Wabash Ave & Wacker Pl" ...  
## $ start\_station\_id : chr [1:284042] "TA1307000131" "13042" "13109" "TA1307000131" ...  
## $ end\_station\_name : chr [1:284042] "Kingsbury St & Kinzie St" "Orleans St & Chestnut St (NEXT Apts)" "Broadway & Ridge Ave" "Franklin St & Jackson Blvd" ...  
## $ end\_station\_id : chr [1:284042] "KA1503000043" "620" "15578" "TA1305000025" ...  
## $ start\_lat : num [1:284042] 41.9 41.9 42 41.9 41.9 ...  
## $ start\_lng : num [1:284042] -87.6 -87.6 -87.7 -87.6 -87.6 ...  
## $ end\_lat : num [1:284042] 41.9 41.9 42 41.9 41.9 ...  
## $ end\_lng : num [1:284042] -87.6 -87.6 -87.7 -87.6 -87.7 ...  
## $ member\_casual : chr [1:284042] "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()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(month\_04\_2022)

## spec\_tbl\_df [371,249 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:371249] "3564070EEFD12711" "0B820C7FCF22F489" "89EEEE32293F07FF" "84D4751AEB31888D" ...  
## $ rideable\_type : chr [1:371249] "electric\_bike" "classic\_bike" "classic\_bike" "classic\_bike" ...  
## $ started\_at : POSIXct[1:371249], format: "2022-04-06 17:42:48" "2022-04-24 19:23:07" ...  
## $ ended\_at : POSIXct[1:371249], format: "2022-04-06 17:54:36" "2022-04-24 19:43:17" ...  
## $ start\_station\_name: chr [1:371249] "Paulina St & Howard St" "Wentworth Ave & Cermak Rd" "Halsted St & Polk St" "Wentworth Ave & Cermak Rd" ...  
## $ start\_station\_id : chr [1:371249] "515" "13075" "TA1307000121" "13075" ...  
## $ end\_station\_name : chr [1:371249] "University Library (NU)" "Green St & Madison St" "Green St & Madison St" "Delano Ct & Roosevelt Rd" ...  
## $ end\_station\_id : chr [1:371249] "605" "TA1307000120" "TA1307000120" "KA1706005007" ...  
## $ start\_lat : num [1:371249] 42 41.9 41.9 41.9 41.9 ...  
## $ start\_lng : num [1:371249] -87.7 -87.6 -87.6 -87.6 -87.6 ...  
## $ end\_lat : num [1:371249] 42.1 41.9 41.9 41.9 41.9 ...  
## $ end\_lng : num [1:371249] -87.7 -87.6 -87.6 -87.6 -87.6 ...  
## $ member\_casual : chr [1:371249] "member" "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\_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()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(month\_05\_2022)

## spec\_tbl\_df [634,858 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:634858] "EC2DE40644C6B0F4" "1C31AD03897EE385" "1542FBEC830415CF" "6FF59852924528F8" ...  
## $ rideable\_type : chr [1:634858] "classic\_bike" "classic\_bike" "classic\_bike" "classic\_bike" ...  
## $ started\_at : POSIXct[1:634858], format: "2022-05-23 23:06:58" "2022-05-11 08:53:28" ...  
## $ ended\_at : POSIXct[1:634858], format: "2022-05-23 23:40:19" "2022-05-11 09:31:22" ...  
## $ start\_station\_name: chr [1:634858] "Wabash Ave & Grand Ave" "DuSable Lake Shore Dr & Monroe St" "Clinton St & Madison St" "Clinton St & Madison St" ...  
## $ start\_station\_id : chr [1:634858] "TA1307000117" "13300" "TA1305000032" "TA1305000032" ...  
## $ end\_station\_name : chr [1:634858] "Halsted St & Roscoe St" "Field Blvd & South Water St" "Wood St & Milwaukee Ave" "Clark St & Randolph St" ...  
## $ end\_station\_id : chr [1:634858] "TA1309000025" "15534" "13221" "TA1305000030" ...  
## $ start\_lat : num [1:634858] 41.9 41.9 41.9 41.9 41.9 ...  
## $ start\_lng : num [1:634858] -87.6 -87.6 -87.6 -87.6 -87.6 ...  
## $ end\_lat : num [1:634858] 41.9 41.9 41.9 41.9 41.9 ...  
## $ end\_lng : num [1:634858] -87.6 -87.6 -87.7 -87.6 -87.7 ...  
## $ member\_casual : chr [1:634858] "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()  
## .. )  
## - attr(\*, "problems")=<externalptr>

#### Convert ride\_id and rideable\_type to character so that they can stack correctly:

month\_06\_2021 <- mutate(month\_06\_2021, ride\_id = as.character(ride\_id),rideable\_type = as.character(rideable\_type))  
month\_07\_2021 <- mutate(month\_07\_2021, ride\_id = as.character(ride\_id),rideable\_type = as.character(rideable\_type))  
month\_08\_2021 <- mutate(month\_08\_2021, ride\_id = as.character(ride\_id),rideable\_type = as.character(rideable\_type))  
month\_09\_2021 <- mutate(month\_09\_2021, ride\_id = as.character(ride\_id),rideable\_type = as.character(rideable\_type))  
month\_10\_2021 <- mutate(month\_10\_2021, ride\_id = as.character(ride\_id),rideable\_type = as.character(rideable\_type))  
month\_11\_2021 <- mutate(month\_11\_2021, ride\_id = as.character(ride\_id),rideable\_type = as.character(rideable\_type))  
month\_12\_2021 <- mutate(month\_12\_2021, ride\_id = as.character(ride\_id),rideable\_type = as.character(rideable\_type))  
month\_01\_2022 <- mutate(month\_01\_2022, ride\_id = as.character(ride\_id),rideable\_type = as.character(rideable\_type))  
month\_02\_2022 <- mutate(month\_02\_2022, ride\_id = as.character(ride\_id),rideable\_type = as.character(rideable\_type))  
month\_03\_2022 <- mutate(month\_03\_2022, ride\_id = as.character(ride\_id),rideable\_type = as.character(rideable\_type))  
month\_04\_2022 <- mutate(month\_04\_2022, ride\_id = as.character(ride\_id),rideable\_type = as.character(rideable\_type))  
month\_05\_2022 <- mutate(month\_05\_2022, ride\_id = as.character(ride\_id),rideable\_type = as.character(rideable\_type))

#### Individual quarter’s data frames into one big data frame:

total\_trips <- bind\_rows(month\_06\_2021, month\_07\_2021, month\_08\_2021, month\_09\_2021, month\_10\_2021, month\_10\_2021, month\_11\_2021, month\_12\_2021, month\_01\_2022, month\_02\_2022, month\_03\_2022, month\_04\_2022, month\_05\_2022)

#### Remove lat, long:

total\_trips <-total\_trips %>%   
select(-c(start\_lat, start\_lng, end\_lat, end\_lng))

## Clean up and add data to prepare for analysis

colnames(total\_trips)

## [1] "ride\_id" "rideable\_type" "started\_at"   
## [4] "ended\_at" "start\_station\_name" "start\_station\_id"   
## [7] "end\_station\_name" "end\_station\_id" "member\_casual"

nrow(total\_trips)

## [1] 6492002

dim(total\_trips)

## [1] 6492002 9

head(total\_trips)

## # A tibble: 6 × 9  
## ride\_id rideable\_type started\_at ended\_at start\_station\_n…  
## <chr> <chr> <dttm> <dttm> <chr>   
## 1 99FEC9… electric\_bike 2021-06-13 14:31:28 2021-06-13 14:34:11 <NA>   
## 2 06048D… electric\_bike 2021-06-04 11:18:02 2021-06-04 11:24:19 <NA>   
## 3 959806… electric\_bike 2021-06-04 09:49:35 2021-06-04 09:55:34 <NA>   
## 4 B03C0F… electric\_bike 2021-06-03 19:56:05 2021-06-03 20:21:55 <NA>   
## 5 B9EEA8… electric\_bike 2021-06-04 14:05:51 2021-06-04 14:09:59 <NA>   
## 6 62B943… electric\_bike 2021-06-03 19:32:01 2021-06-03 19:38:46 <NA>   
## # … with 4 more variables: start\_station\_id <chr>, end\_station\_name <chr>,  
## # end\_station\_id <chr>, member\_casual <chr>

str(total\_trips)

## tibble [6,492,002 × 9] (S3: tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:6492002] "99FEC93BA843FB20" "06048DCFC8520CAF" "9598066F68045DF2" "B03C0FE48C412214" ...  
## $ rideable\_type : chr [1:6492002] "electric\_bike" "electric\_bike" "electric\_bike" "electric\_bike" ...  
## $ started\_at : POSIXct[1:6492002], format: "2021-06-13 14:31:28" "2021-06-04 11:18:02" ...  
## $ ended\_at : POSIXct[1:6492002], format: "2021-06-13 14:34:11" "2021-06-04 11:24:19" ...  
## $ start\_station\_name: chr [1:6492002] NA NA NA NA ...  
## $ start\_station\_id : chr [1:6492002] NA NA NA NA ...  
## $ end\_station\_name : chr [1:6492002] NA NA NA NA ...  
## $ end\_station\_id : chr [1:6492002] NA NA NA NA ...  
## $ member\_casual : chr [1:6492002] "member" "member" "member" "member" ...

summary(total\_trips)

## ride\_id rideable\_type started\_at   
## Length:6492002 Length:6492002 Min. :2021-06-01 00:00:38.00   
## Class :character Class :character 1st Qu.:2021-08-03 18:25:40.25   
## Mode :character Mode :character Median :2021-10-03 17:19:50.00   
## Mean :2021-10-25 02:33:18.61   
## 3rd Qu.:2021-12-14 12:43:49.75   
## Max. :2022-05-31 23:59:56.00   
## ended\_at start\_station\_name start\_station\_id   
## Min. :2021-06-01 00:06:22.00 Length:6492002 Length:6492002   
## 1st Qu.:2021-08-03 18:45:38.00 Class :character Class :character   
## Median :2021-10-03 17:43:13.00 Mode :character Mode :character   
## Mean :2021-10-25 02:53:50.66   
## 3rd Qu.:2021-12-14 12:57:14.00   
## Max. :2022-06-02 11:35:01.00   
## end\_station\_name end\_station\_id member\_casual   
## Length:6492002 Length:6492002 Length:6492002   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##

#### Begin by seeing how many observations fall under each usertype:

table(total\_trips$member\_casual)

##   
## casual member   
## 2817099 3674903

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

#### This will allow us to aggregate ride data for each month, day, or year … before completing these operations we could only aggregate at the ride level:

total\_trips$date <- as.Date(total\_trips$started\_at)  
total\_trips$month <- format(as.Date(total\_trips$date), "%m")  
total\_trips$day <- format(as.Date(total\_trips$date), "%d")  
total\_trips$year <- format(as.Date(total\_trips$date), "%Y")  
total\_trips$day\_of\_week <- format(as.Date(total\_trips$date), "%A")

#### Add a “ride\_length” calculation to total\_trips (in seconds):

total\_trips$ride\_length <- difftime(total\_trips$ended\_at, total\_trips$started\_at)

#### Convert “ride\_length” from Factor to numeric so we can run calculations on the data:

is.factor(total\_trips$ride\_length)

## [1] FALSE

total\_trips$ride\_length <-as.numeric(as.character(total\_trips$ride\_length))  
is.numeric(total\_trips$ride\_length)

## [1] TRUE

#### Remove “bad” data

#### The dataframe includes a few hundred entries ride\_length was negative:

total\_trips\_v2 <- total\_trips[!(total\_trips$ride\_length<0),]

## Analysis

#### Descriptive analysis on ride\_length (all figures in seconds):

summary(total\_trips\_v2$ride\_length)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0 379 675 1232 1225 3356649

#### Compare members and casual users:

aggregate(total\_trips\_v2$ride\_length~ total\_trips\_v2$member\_casual, FUN=mean)

## total\_trips\_v2$member\_casual total\_trips\_v2$ride\_length  
## 1 casual 1822.6851  
## 2 member 779.3509

aggregate(total\_trips\_v2$ride\_length~ total\_trips\_v2$member\_casual, FUN=median)

## total\_trips\_v2$member\_casual total\_trips\_v2$ride\_length  
## 1 casual 907  
## 2 member 544

aggregate(total\_trips\_v2$ride\_length~ total\_trips\_v2$member\_casual, FUN=max)

## total\_trips\_v2$member\_casual total\_trips\_v2$ride\_length  
## 1 casual 3356649  
## 2 member 93594

aggregate(total\_trips\_v2$ride\_length~ total\_trips\_v2$member\_casual, FUN=min)

## total\_trips\_v2$member\_casual total\_trips\_v2$ride\_length  
## 1 casual 0  
## 2 member 0

#### Average ride time by each day for members vs casual users:

aggregate(total\_trips\_v2$ride\_length~ total\_trips\_v2$member\_casual+total\_trips\_v2$day\_of\_week, FUN=mean)

## total\_trips\_v2$member\_casual total\_trips\_v2$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  
## total\_trips\_v2$ride\_length  
## 1 1717.1352  
## 2 763.6777  
## 3 1813.5540  
## 4 752.7667  
## 5 1996.1543  
## 6 873.7515  
## 7 2127.9669  
## 8 885.0652  
## 9 1641.1163  
## 10 741.2748  
## 11 1554.2345  
## 12 733.5182  
## 13 1578.3658  
## 14 735.4067

#### Notice that the days of the week are out of order. Let’s fix that:

total\_trips\_v2$day\_of\_week <- ordered(total\_trips\_v2$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(total\_trips\_v2$ride\_length~ total\_trips\_v2$member\_casual+total\_trips\_v2$day\_of\_week, FUN=mean)

## total\_trips\_v2$member\_casual total\_trips\_v2$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  
## total\_trips\_v2$ride\_length  
## 1 2127.9669  
## 2 885.0652  
## 3 1813.5540  
## 4 752.7667  
## 5 1554.2345  
## 6 733.5182  
## 7 1578.3658  
## 8 735.4067  
## 9 1641.1163  
## 10 741.2748  
## 11 1717.1352  
## 12 763.6777  
## 13 1996.1543  
## 14 873.7515

#### Analyze ridership data by type and weekday:

total\_trips\_v2 %>%   
 mutate(weekday = wday(started\_at, label = TRUE)) %>%  
 group\_by(member\_casual, weekday) %>%  
 summarise(number\_of\_rides = n() ,average\_duration = mean(ride\_length)) %>%   
 arrange(member\_casual, weekday)

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

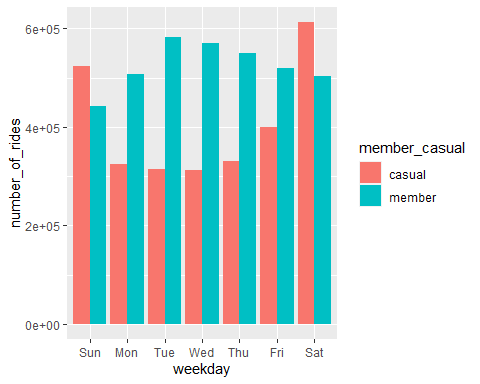
## # A tibble: 14 × 4  
## # Groups: member\_casual [2]  
## member\_casual weekday number\_of\_rides average\_duration  
## <chr> <ord> <int> <dbl>  
## 1 casual Sun 522799 2128.  
## 2 casual Mon 324125 1814.  
## 3 casual Tue 314100 1554.  
## 4 casual Wed 312628 1578.  
## 5 casual Thu 330943 1641.  
## 6 casual Fri 399538 1717.  
## 7 casual Sat 612905 1996.  
## 8 member Sun 442824 885.  
## 9 member Mon 507772 753.  
## 10 member Tue 582368 734.  
## 11 member Wed 570326 735.  
## 12 member Thu 548928 741.  
## 13 member Fri 519594 764.  
## 14 member Sat 503013 874.

## Share.

#### Let’s visualize the number of rides by rider type:

total\_trips\_v2 %>%   
mutate(weekday = wday(started\_at, label = TRUE)) %>%  
 group\_by(member\_casual, weekday) %>%  
 summarise(number\_of\_rides = n() ,average\_duration = mean(ride\_length)) %>%   
 arrange(member\_casual, weekday) %>%   
 ggplot(aes(x = weekday, y = number\_of\_rides, fill = member\_casual)) +  
 geom\_col(position = "dodge")

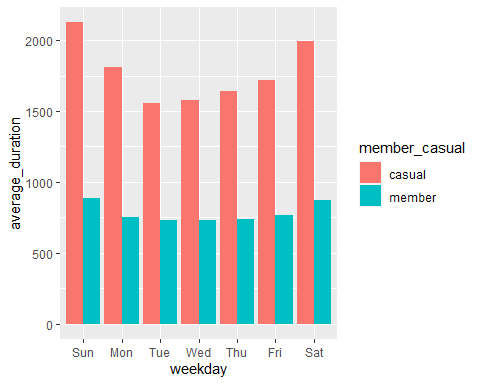
## `summarise()` has grouped output by 'member\_casual'. You can override using the  
## `.groups` argument.



#### Let’s create a visualization for average duration:

total\_trips\_v2 %>%   
mutate(weekday = wday(started\_at, label = TRUE)) %>%  
 group\_by(member\_casual, weekday) %>%  
 summarise(number\_of\_rides = n() ,average\_duration = mean(ride\_length)) %>%   
 arrange(member\_casual, weekday) %>%   
 ggplot(aes(x = weekday, y = average\_duration, fill = member\_casual)) +  
 geom\_col(position = "dodge")

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



## Act

### Conclusion

Based on analysis, casual riders are different to the membership riders with-

* Where casual riders uses the bike for higher duration with less number of rides
* Casual riders having more number of rides during weekends compared to weekdays.
* Membership riders having more number of rides during weekdays and with less average ride duration compared to casual.

### Recommendation

Within the membership

* Providing more number of rides per day than current limit on number of rides per day.
* Reward points on weekend rides duration.
* unlimited number of rides on weekends