

Google Analytic: Cyclistic

Aldrich Pinso

2022-10-22

ASK

Introduction

This is part of Google Data Analytics capstone project, sponsored by *Grow with Google Malaysia*. By solving real life situation, I'm able to experience the work of a data analyst and at the same time building a career portfolio.

In this case study, we need to answer business questions posed by Cyclistic bike-share company. I'm assigned to work with the director of marketing and analytics, Lily Moreno and Cyclistic marketing analytics team. To confirm Madam Moreno's hypothesis, which is to maximize the number of annual membership, more data driven insights are required. A new design for marketing strategy, which is to convert casual riders into annual memberships, can only commence after achieving positive approval by Cyclistic executives team.

As of 2016, Cyclistic bike-share program features more than 5,824 bicycles and 692 docking station, we are able to accurately pinpoint these fleet of asset with geotracking devices, it comes with different types of bike, reclining bikes, hand tricycles, and cargo bikes, offering better quality of life to the people in Chicago city.

Backed by Cyclistic's finance analyst, annual members are more profitable than casual rides. Currently, its pricing flexibility helps Cyclistic to attract more customers, however for future growth, Madam Moreno believes in maximising existing casual riders which is familiar with the program that full fill their mobility needs.

Asking the right questions:

- Who are the controlled group?
- What is profitable and what isn't?
- How to convert casual riders into signing up as member?
- What is the future growth of Cyclistics bike-share?

PREPARE

Read and understand the terms and condition from the provided Divvy Bike Data License Agreement (<https://ride.divvybikes.com/data-license-agreement>)

Set up environment: Data uploading

```
install.packages("tidyverse") library(tidyverse) install.packages("lubridate") library(lubridate)
install.packages("readr") library(readr) install.packages("rmarkdown") install.packages("knitr")
```

Upload relevant raw data from Cyclistic bike-share provided by Divvy Bikes.

Data is split into pre-pandemic (2014-2018); pandemic (2019); towards post-pandemic (2020-present). By looking at the historical data in this way, it makes more sense to me, because we get to compare the impact of such crisis.

Please note that this timeline is just a reference. WHO hasn't publicly announced the end of Covid-19 pandemic, although most countries are vaccinated now and people start moving freely again.

Due to time constraint and minimizing the chances of data overloading and crashing, I will only analyse the year of 2020.

```
Q1_2020 <- read.csv("/Volumes/UDISK/Google Analytics Track 1/2013-2022/2020/Divvy_Trips_2020_Q1.csv")
Q2_APR_2020 <- read.csv("/Volumes/UDISK/Google Analytics Track 1/2013-2022/2020/202004-divvy-tripdata.csv")
Q2_MAY_2020 <- read.csv("/Volumes/UDISK/Google Analytics Track 1/2013-2022/2020/202005-divvy-tripdata.csv")
Q2_JUN_2020 <- read.csv("/Volumes/UDISK/Google Analytics Track 1/2013-2022/2020/202006-divvy-tripdata.csv")
Q3_JUL_2020 <- read.csv("/Volumes/UDISK/Google Analytics Track 1/2013-2022/2020/202007-divvy-tripdata.csv")
Q3_AUG_2020 <- read.csv("/Volumes/UDISK/Google Analytics Track 1/2013-2022/2020/202008-divvy-tripdata.csv")
Q3_SEP_2020 <- read.csv("/Volumes/UDISK/Google Analytics Track 1/2013-2022/2020/202009-divvy-tripdata.csv")
Q4_OCT_2020 <- read.csv("/Volumes/UDISK/Google Analytics Track 1/2013-2022/2020/202010-divvy-tripdata.csv")
Q4_NOV_2020 <- read.csv("/Volumes/UDISK/Google Analytics Track 1/2013-2022/2020/202011-divvy-tripdata.csv")
Q4_DEC_2020 <- read.csv("/Volumes/UDISK/Google Analytics Track 1/2013-2022/2020/202012-divvy-tripdata.csv")
```

PROCESS

I have chosen R programming as my main tool for many reasons. Personally, I'd love to be able to go into Data Science area, other than python, R is more intuitive to code and decipher complexity. For this case study, I want to avoid data overloading and crashing, R can handle that swiftly. I will also use SQL to confirm some of my findings here.

In this section, I will merge data, change data type, create new columns, etc.

This is part of my workflow. I will be displaying the data types and internal structure as a reference to identify anomalies or commonalities vice versa.

Set up environment: Data structures & Data types

```
install.packages("here") library("here") install.packages("skimr") library("skimr") install.packages("dplyr")
library("dplyr") install.packages("tibble") library(tibble)
```

Types of data: * characters * numeric * integer * complex * logical

Data Structures: * vector * list * matrix * data frame * array * factor

Object with different attributes: * name * dimension * class

Data processing

```
str(Q1_2020)
```

```
## 'data.frame':    426887 obs. of  13 variables:
##  $ ride_id          : chr  "EACB19130B0CDA4A" "8FED874C809DC021" "789F3C21E472CA9
6" "C9A388DAC6ABF313" ...
##  $ rideable_type     : chr  "docked_bike" "docked_bike" "docked_bike" "docked_bik
e" ...
##  $ started_at        : chr  "2020-01-21 20:06:59" "2020-01-30 14:22:39" "2020-01-0
9 19:29:26" "2020-01-06 16:17:07" ...
##  $ ended_at          : chr  "2020-01-21 20:14:30" "2020-01-30 14:26:22" "2020-01-0
9 19:32:17" "2020-01-06 16:25:56" ...
##  $ start_station_name: chr  "Western Ave & Leland Ave" "Clark St & Montrose Ave"
"Brooklyn Ave & Belmont Ave" "Clark St & Randolph St" ...
##  $ start_station_id  : int   239 234 296 51 66 212 96 96 212 38 ...
##  $ end_station_name  : chr  "Clark St & Leland Ave" "Southport Ave & Irving Park R
d" "Wilton Ave & Belmont Ave" "Fairbanks Ct & Grand Ave" ...
##  $ end_station_id    : int   326 318 117 24 212 96 212 212 96 100 ...
##  $ start_lat         : num   42 42 41.9 41.9 41.9 ...
##  $ start_lng         : num  -87.7 -87.7 -87.6 -87.6 -87.6 ...
##  $ end_lat          : num   42 42 41.9 41.9 41.9 ...
##  $ end_lng          : num  -87.7 -87.7 -87.7 -87.6 -87.6 ...
##  $ member_casual     : chr  "member" "member" "member" "member" ...
```

```
str(Q2_APR_2020)
```

```
## 'data.frame':    84776 obs. of  13 variables:
##  $ ride_id          : chr  "A847FADBBC638E45" "5405B80E996FF60D" "5DD24A79A4E006F
4" "2A59BBDF5CDBA725" ...
##  $ rideable_type     : chr  "docked_bike" "docked_bike" "docked_bike" "docked_bik
e" ...
##  $ started_at        : chr  "2020-04-26 17:45:14" "2020-04-17 17:08:54" "2020-04-0
1 17:54:13" "2020-04-07 12:50:19" ...
##  $ ended_at          : chr  "2020-04-26 18:12:03" "2020-04-17 17:17:03" "2020-04-0
1 18:08:36" "2020-04-07 13:02:31" ...
##  $ start_station_name: chr  "Eckhart Park" "Drake Ave & Fullerton Ave" "McClurg Ct
& Erie St" "California Ave & Division St" ...
##  $ start_station_id  : int   86 503 142 216 125 173 35 434 627 377 ...
##  $ end_station_name  : chr  "Lincoln Ave & Diversey Pkwy" "Kosciuszko Park" "India
na Ave & Roosevelt Rd" "Wood St & Augusta Blvd" ...
##  $ end_station_id    : int   152 499 255 657 323 35 635 382 359 508 ...
##  $ start_lat         : num   41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num  -87.7 -87.7 -87.6 -87.7 -87.6 ...
##  $ end_lat          : num   41.9 41.9 41.9 41.9 42 ...
##  $ end_lng          : num  -87.7 -87.7 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr  "member" "member" "member" "member" ...
```

```
str(Q2_MAY_2020)
```

```
## 'data.frame':    200274 obs. of  13 variables:
## $ ride_id        : chr  "02668AD35674B983" "7A50CCAF1EDDB28F" "2FFCDFDB91FE9A5
2" "58991CF1DB75BA84" ...
## $ rideable_type   : chr  "docked_bike" "docked_bike" "docked_bike" "docked_bik
e" ...
## $ started_at      : chr  "2020-05-27 10:03:52" "2020-05-25 10:47:11" "2020-05-0
2 14:11:03" "2020-05-02 16:25:36" ...
## $ ended_at        : chr  "2020-05-27 10:16:49" "2020-05-25 11:05:40" "2020-05-0
2 15:48:21" "2020-05-02 16:39:28" ...
## $ start_station_name: chr  "Franklin St & Jackson Blvd" "Clark St & Wrightwood Av
e" "Kedzie Ave & Milwaukee Ave" "Clarendon Ave & Leland Ave" ...
## $ start_station_id : int   36 340 260 251 261 206 261 180 331 219 ...
## $ end_station_name  : chr  "Wabash Ave & Grand Ave" "Clark St & Leland Ave" "Kedz
ie Ave & Milwaukee Ave" "Lake Shore Dr & Wellington Ave" ...
## $ end_station_id    : int   199 326 260 157 206 22 261 180 300 305 ...
## $ start_lat         : num   41.9 41.9 41.9 42 41.9 ...
## $ start_lng         : num  -87.6 -87.6 -87.7 -87.7 -87.7 ...
## $ end_lat          : num   41.9 42 41.9 41.9 41.8 ...
## $ end_lng          : num  -87.6 -87.7 -87.7 -87.6 -87.6 ...
## $ member_casual     : chr  "member" "casual" "casual" "casual" ...
```

```
str(Q2_JUN_2020)
```

```
## 'data.frame':    343005 obs. of  13 variables:
## $ ride_id        : chr  "8CD5DE2C2B6C4CFC" "9A191EB2C751D85D" "F37D14B0B5659BC
F" "C41237B506E85FA1" ...
## $ rideable_type   : chr  "docked_bike" "docked_bike" "docked_bike" "docked_bik
e" ...
## $ started_at      : chr  "2020-06-13 23:24:48" "2020-06-26 07:26:10" "2020-06-2
3 17:12:41" "2020-06-20 01:09:35" ...
## $ ended_at        : chr  "2020-06-13 23:36:55" "2020-06-26 07:31:58" "2020-06-2
3 17:21:14" "2020-06-20 01:28:24" ...
## $ start_station_name: chr  "Wilton Ave & Belmont Ave" "Federal St & Polk St" "Dal
ey Center Plaza" "Broadway & Cornelia Ave" ...
## $ start_station_id : int   117 41 81 303 327 327 41 115 338 84 ...
## $ end_station_name  : chr  "Damen Ave & Clybourn Ave" "Daley Center Plaza" "State
St & Harrison St" "Broadway & Berwyn Ave" ...
## $ end_station_id    : int   163 81 5 294 117 117 81 303 164 53 ...
## $ start_lat         : num   41.9 41.9 41.9 41.9 41.9 ...
## $ start_lng         : num  -87.7 -87.6 -87.6 -87.6 -87.7 ...
## $ end_lat          : num   41.9 41.9 41.9 42 41.9 ...
## $ end_lng          : num  -87.7 -87.6 -87.6 -87.7 -87.7 ...
## $ member_casual     : chr  "casual" "member" "member" "casual" ...
```

```
str(Q3_JUL_2020)
```

```
## 'data.frame':    551480 obs. of  13 variables:
## $ ride_id      : chr  "762198876D69004D" "BEC9C9FBA0D4CF1B" "D2FD8EA432C77EC
1" "54AE594E20B35881" ...
## $ rideable_type : chr  "docked_bike" "docked_bike" "docked_bike" "docked_bik
e" ...
## $ started_at   : chr  "2020-07-09 15:22:02" "2020-07-24 23:56:30" "2020-07-0
8 19:49:07" "2020-07-17 19:06:42" ...
## $ ended_at     : chr  "2020-07-09 15:25:52" "2020-07-25 00:20:17" "2020-07-0
8 19:56:22" "2020-07-17 19:27:38" ...
## $ start_station_name: chr  "Ritchie Ct & Banks St" "Halsted St & Roscoe St" "Lake
Shore Dr & Diversey Pkwy" "LaSalle St & Illinois St" ...
## $ start_station_id : int  180 299 329 181 268 635 113 211 176 31 ...
## $ end_station_name : chr  "Wells St & Evergreen Ave" "Broadway & Ridge Ave" "Cla
rk St & Wellington Ave" "Clark St & Armitage Ave" ...
## $ end_station_id   : int  291 461 156 94 301 289 140 31 191 142 ...
## $ start_lat        : num  41.9 41.9 41.9 41.9 41.9 ...
## $ start_lng        : num  -87.6 -87.6 -87.6 -87.6 -87.6 ...
## $ end_lat          : num  41.9 42 41.9 41.9 41.9 ...
## $ end_lng          : num  -87.6 -87.7 -87.6 -87.6 -87.6 ...
## $ member_casual    : chr  "member" "member" "casual" "casual" ...
```

```
str(Q3_AUG_2020)
```

```
## 'data.frame':    622361 obs. of  13 variables:
## $ ride_id      : chr  "322BD23D287743ED" "2A3AEF1AB9054D8B" "67DC1D133E8B581
6" "C79FBBD412E578A7" ...
## $ rideable_type : chr  "docked_bike" "electric_bike" "electric_bike" "electri
c_bike" ...
## $ started_at   : chr  "2020-08-20 18:08:14" "2020-08-27 18:46:04" "2020-08-2
6 19:44:14" "2020-08-27 12:05:41" ...
## $ ended_at     : chr  "2020-08-20 18:17:51" "2020-08-27 19:54:51" "2020-08-2
6 21:53:07" "2020-08-27 12:53:45" ...
## $ start_station_name: chr  "Lake Shore Dr & Diversey Pkwy" "Michigan Ave & 14th S
t" "Columbus Dr & Randolph St" "Daley Center Plaza" ...
## $ start_station_id : int  329 168 195 81 658 658 196 67 153 177 ...
## $ end_station_name : chr  "Clark St & Lincoln Ave" "Michigan Ave & 14th St" "Sta
te St & Randolph St" "State St & Kinzie St" ...
## $ end_station_id   : int  141 168 44 47 658 658 49 229 225 305 ...
## $ start_lat        : num  41.9 41.9 41.9 41.9 41.9 ...
## $ start_lng        : num  -87.6 -87.6 -87.6 -87.6 -87.7 ...
## $ end_lat          : num  41.9 41.9 41.9 41.9 41.9 ...
## $ end_lng          : num  -87.6 -87.6 -87.6 -87.6 -87.7 ...
## $ member_casual    : chr  "member" "casual" "casual" "casual" ...
```

```
str(Q3_SEP_2020)
```

```
## 'data.frame':    532958 obs. of  13 variables:
## $ ride_id      : chr  "2B22BD5F95FB2629" "A7FB70B4AFC6CAF2" "86057FA01BAC778
E" "57F6DC9A153DB98C" ...
## $ rideable_type : chr  "electric_bike" "electric_bike" "electric_bike" "elect
ric_bike" ...
## $ started_at   : chr  "2020-09-17 14:27:11" "2020-09-17 15:07:31" "2020-09-1
7 15:09:04" "2020-09-17 18:10:46" ...
## $ ended_at     : chr  "2020-09-17 14:44:24" "2020-09-17 15:07:45" "2020-09-1
7 15:09:35" "2020-09-17 18:35:49" ...
## $ start_station_name: chr  "Michigan Ave & Lake St" "W Oakdale Ave & N Broadway"
"W Oakdale Ave & N Broadway" "Ashland Ave & Belle Plaine Ave" ...
## $ start_station_id : int  52 NA NA 246 24 94 291 NA NA NA ...
## $ end_station_name : chr  "Green St & Randolph St" "W Oakdale Ave & N Broadway"
"W Oakdale Ave & N Broadway" "Montrose Harbor" ...
## $ end_station_id   : int  112 NA NA 249 24 NA 256 NA NA NA ...
## $ start_lat        : num  41.9 41.9 41.9 42 41.9 ...
## $ start_lng        : num  -87.6 -87.6 -87.6 -87.7 -87.6 ...
## $ end_lat          : num  41.9 41.9 41.9 42 41.9 ...
## $ end_lng          : num  -87.6 -87.6 -87.6 -87.6 -87.6 ...
## $ member_casual    : chr  "casual" "casual" "casual" "casual" ...
```

```
str(Q4_OCT_2020)
```

```
## 'data.frame':    388653 obs. of  13 variables:
## $ ride_id      : chr  "ACB6B40CF5B9044C" "DF450C72FD109C01" "B6396B54A15AC0D
F" "44A4AEE261B9E854" ...
## $ rideable_type : chr  "electric_bike" "electric_bike" "electric_bike" "elect
ric_bike" ...
## $ started_at   : chr  "2020-10-31 19:39:43" "2020-10-31 23:50:08" "2020-10-3
1 23:00:01" "2020-10-31 22:16:43" ...
## $ ended_at     : chr  "2020-10-31 19:57:12" "2020-11-01 00:04:16" "2020-10-3
1 23:08:22" "2020-10-31 22:19:35" ...
## $ start_station_name: chr  "Lakeview Ave & Fullerton Pkwy" "Southport Ave & Wavel
and Ave" "Stony Island Ave & 67th St" "Clark St & Grace St" ...
## $ start_station_id : int  313 227 102 165 190 359 313 125 NA 174 ...
## $ end_station_name : chr  "Rush St & Hubbard St" "Kedzie Ave & Milwaukee Ave" "U
niversity Ave & 57th St" "Broadway & Sheridan Rd" ...
## $ end_station_id   : int  125 260 423 256 185 53 125 313 199 635 ...
## $ start_lat        : num  41.9 41.9 41.8 42 41.9 ...
## $ start_lng        : num  -87.6 -87.7 -87.6 -87.7 -87.7 ...
## $ end_lat          : num  41.9 41.9 41.8 42 41.9 ...
## $ end_lng          : num  -87.6 -87.7 -87.6 -87.7 -87.7 ...
## $ member_casual    : chr  "casual" "casual" "casual" "casual" ...
```

```
str(Q4_NOV_2020)
```

```
## 'data.frame':    259716 obs. of  13 variables:
## $ ride_id      : chr  "BD0A6FF6FFF9B921" "96A7A7A4BDE4F82D" "C61526D06582BDC
5" "E533E89C32080B9E" ...
## $ rideable_type : chr  "electric_bike" "electric_bike" "electric_bike" "elect
ric_bike" ...
## $ started_at   : chr  "2020-11-01 13:36:00" "2020-11-01 10:03:26" "2020-11-0
1 00:34:05" "2020-11-01 00:45:16" ...
## $ ended_at     : chr  "2020-11-01 13:45:40" "2020-11-01 10:14:45" "2020-11-0
1 01:03:06" "2020-11-01 00:54:31" ...
## $ start_station_name: chr  "Dearborn St & Erie St" "Franklin St & Illinois St" "L
ake Shore Dr & Monroe St" "Leavitt St & Chicago Ave" ...
## $ start_station_id : int  110 672 76 659 2 72 76 NA 58 394 ...
## $ end_station_name : chr  "St. Clair St & Erie St" "Noble St & Milwaukee Ave" "F
ederal St & Polk St" "Stave St & Armitage Ave" ...
## $ end_station_id   : int  211 29 41 185 2 76 72 NA 288 273 ...
## $ start_lat        : num  41.9 41.9 41.9 41.9 41.9 ...
## $ start_lng        : num  -87.6 -87.6 -87.6 -87.7 -87.6 ...
## $ end_lat          : num  41.9 41.9 41.9 41.9 41.9 ...
## $ end_lng          : num  -87.6 -87.7 -87.6 -87.7 -87.6 ...
## $ member_casual    : chr  "casual" "casual" "casual" "casual" ...
```

```
str(Q4_DEC_2020)
```

```
## 'data.frame':    131573 obs. of  13 variables:
## $ ride_id      : chr  "70B6A9A437D4C30D" "158A465D4E74C54A" "5262016E0F1F2F9
A" "BE119628E44F871E" ...
## $ rideable_type : chr  "classic_bike" "electric_bike" "electric_bike" "electr
ic_bike" ...
## $ started_at   : chr  "2020-12-27 12:44:29" "2020-12-18 17:37:15" "2020-12-1
5 15:04:33" "2020-12-15 15:54:18" ...
## $ ended_at     : chr  "2020-12-27 12:55:06" "2020-12-18 17:44:19" "2020-12-1
5 15:11:28" "2020-12-15 16:00:11" ...
## $ start_station_name: chr  "Aberdeen St & Jackson Blvd" "" "" "" ...
## $ start_station_id : chr  "13157" "" "" "" ...
## $ end_station_name : chr  "Desplaines St & Kinzie St" "" "" "" ...
## $ end_station_id   : chr  "TA1306000003" "" "" "" ...
## $ start_lat        : num  41.9 41.9 41.9 41.9 41.8 ...
## $ start_lng        : num  -87.7 -87.7 -87.7 -87.7 -87.6 ...
## $ end_lat          : num  41.9 41.9 41.9 41.9 41.8 ...
## $ end_lng          : num  -87.6 -87.7 -87.7 -87.7 -87.6 ...
## $ member_casual    : chr  "member" "member" "member" "member" ...
```

```
colnames(Q1_2020)
```

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

```
colnames(Q2_APR_2020)
```

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

```
colnames(Q2_MAY_2020)
```

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

```
colnames(Q2_JUN_2020)
```

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

```
colnames(Q3_JUL_2020)
```

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

```
colnames(Q3_AUG_2020)
```

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

```
colnames(Q3_SEP_2020)
```

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

```
colnames(Q4_OCT_2020)
```



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

```
colnames(Q4_NOV_2020)
```

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

```
colnames(Q4_DEC_2020)
```

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

```
head(Q1_2020)
```

```
##           ride_id rideable_type      started_at      ended_at
## 1 EACB19130B0CDA4A  docked_bike 2020-01-21 20:06:59 2020-01-21 20:14:30
## 2 8FED874C809DC021  docked_bike 2020-01-30 14:22:39 2020-01-30 14:26:22
## 3 789F3C21E472CA96  docked_bike 2020-01-09 19:29:26 2020-01-09 19:32:17
## 4 C9A388DAC6ABF313  docked_bike 2020-01-06 16:17:07 2020-01-06 16:25:56
## 5 943BC3CBECCFD662  docked_bike 2020-01-30 08:37:16 2020-01-30 08:42:48
## 6 6D9C8A6938165C11  docked_bike 2020-01-10 12:33:05 2020-01-10 12:37:54
##           start_station_name start_station_id      end_station_name
## 1 Western Ave & Leland Ave           239      Clark St & Leland Ave
## 2 Clark St & Montrose Ave           234 Southport Ave & Irving Park Rd
## 3 Broadway & Belmont Ave           296      Wilton Ave & Belmont Ave
## 4 Clark St & Randolph St            51      Fairbanks Ct & Grand Ave
## 5 Clinton St & Lake St              66      Wells St & Hubbard St
## 6 Wells St & Hubbard St            212      Desplaines St & Randolph St
##           end_station_id start_lat start_lng end_lat  end_lng member_casual
## 1              326      41.9665  -87.6884 41.9671 -87.6674      member
## 2              318      41.9616  -87.6660 41.9542 -87.6644      member
## 3              117      41.9401  -87.6455 41.9402 -87.6530      member
## 4               24      41.8846  -87.6319 41.8918 -87.6206      member
## 5              212      41.8856  -87.6418 41.8899 -87.6343      member
## 6               96      41.8899  -87.6343 41.8846 -87.6446      member
```

```
head(Q2_APR_2020)
```

```
##          ride_id rideable_type      started_at      ended_at
## 1 A847FADBBC638E45   docked_bike 2020-04-26 17:45:14 2020-04-26 18:12:03
## 2 5405B80E996FF60D   docked_bike 2020-04-17 17:08:54 2020-04-17 17:17:03
## 3 5DD24A79A4E006F4   docked_bike 2020-04-01 17:54:13 2020-04-01 18:08:36
## 4 2A59BBDF5CDBA725   docked_bike 2020-04-07 12:50:19 2020-04-07 13:02:31
## 5 27AD306C119C6158   docked_bike 2020-04-18 10:22:59 2020-04-18 11:15:54
## 6 356216E875132F61   docked_bike 2020-04-30 17:55:47 2020-04-30 18:01:11
##          start_station_name start_station_id
## 1                      Eckhart Park           86
## 2          Drake Ave & Fullerton Ave          503
## 3          McClurg Ct & Erie St             142
## 4          California Ave & Division St       216
## 5          Rush St & Hubbard St              125
## 6 Mies van der Rohe Way & Chicago Ave        173
##          end_station_name end_station_id start_lat start_lng end_lat
## 1 Lincoln Ave & Diversey Pkwy           152   41.8964  -87.6610 41.9322
## 2          Kosciuszko Park              499   41.9244  -87.7154 41.9306
## 3  Indiana Ave & Roosevelt Rd           255   41.8945  -87.6179 41.8679
## 4          Wood St & Augusta Blvd         657   41.9030  -87.6975 41.8992
## 5  Sheridan Rd & Lawrence Ave           323   41.8902  -87.6262 41.9695
## 6  Streeter Dr & Grand Ave              35    41.8969  -87.6217 41.8923
##          end_lng member_casual
## 1 -87.6586      member
## 2 -87.7238      member
## 3 -87.6230      member
## 4 -87.6722      member
## 5 -87.6547      casual
## 6 -87.6120      member
```

```
head(Q2_MAY_2020)
```

```
##          ride_id rideable_type      started_at      ended_at
## 1 02668AD35674B983   docked_bike 2020-05-27 10:03:52 2020-05-27 10:16:49
## 2 7A50CCAF1EDDB28F   docked_bike 2020-05-25 10:47:11 2020-05-25 11:05:40
## 3 2FFCDFDB91FE9A52   docked_bike 2020-05-02 14:11:03 2020-05-02 15:48:21
## 4 58991CF1DB75BA84   docked_bike 2020-05-02 16:25:36 2020-05-02 16:39:28
## 5 A79651EFECC268CD   docked_bike 2020-05-29 12:49:54 2020-05-29 13:27:11
## 6 1466C5B39F68F746   docked_bike 2020-05-29 13:27:24 2020-05-29 14:14:45
##          start_station_name start_station_id      end_station_name
## 1 Franklin St & Jackson Blvd           36      Wabash Ave & Grand Ave
## 2  Clark St & Wrightwood Ave           340      Clark St & Leland Ave
## 3 Kedzie Ave & Milwaukee Ave           260      Kedzie Ave & Milwaukee Ave
## 4 Clarendon Ave & Leland Ave           251 Lake Shore Dr & Wellington Ave
## 5  Hermitage Ave & Polk St             261      Halsted St & Archer Ave
## 6  Halsted St & Archer Ave             206      May St & Taylor St
##          end_station_id start_lat start_lng end_lat  end_lng member_casual
## 1          199   41.8777  -87.6353 41.8915 -87.6268      member
## 2          326   41.9295  -87.6431 41.9671 -87.6674      casual
## 3          260   41.9296  -87.7079 41.9296 -87.7079      casual
## 4          157   41.9680  -87.6500 41.9367 -87.6368      casual
## 5          206   41.8715  -87.6699 41.8472 -87.6468      member
## 6          22    41.8472  -87.6468 41.8695 -87.6555      member
```

```
head(Q2_JUN_2020)
```

```
##          ride_id rideable_type      started_at      ended_at
## 1 8CD5DE2C2B6C4CFC   docked_bike 2020-06-13 23:24:48 2020-06-13 23:36:55
## 2 9A191EB2C751D85D   docked_bike 2020-06-26 07:26:10 2020-06-26 07:31:58
## 3 F37D14B0B5659BCF   docked_bike 2020-06-23 17:12:41 2020-06-23 17:21:14
## 4 C41237B506E85FA1   docked_bike 2020-06-20 01:09:35 2020-06-20 01:28:24
## 5 4B51B3B0BDA7787C   docked_bike 2020-06-25 16:59:25 2020-06-25 17:08:48
## 6 D50DF288196B53BE   docked_bike 2020-06-17 18:07:18 2020-06-17 18:18:14
##          start_station_name start_station_id      end_station_name
## 1   Wilton Ave & Belmont Ave          117 Damen Ave & Clybourn Ave
## 2   Federal St & Polk St              41      Daley Center Plaza
## 3   Daley Center Plaza                81      State St & Harrison St
## 4   Broadway & Cornelia Ave           303      Broadway & Berwyn Ave
## 5 Sheffield Ave & Webster Ave          327 Wilton Ave & Belmont Ave
## 6 Sheffield Ave & Webster Ave          327 Wilton Ave & Belmont Ave
## end_station_id start_lat start_lng end_lat end_lng member_casual
## 1             163  41.94018 -87.65304 41.93193 -87.67786      casual
## 2              81  41.87208 -87.62954 41.88424 -87.62963      member
## 3              5  41.88424 -87.62963 41.87405 -87.62772      member
## 4             294  41.94553 -87.64644 41.97835 -87.65975      casual
## 5             117  41.92154 -87.65382 41.94018 -87.65304      casual
## 6             117  41.92154 -87.65382 41.94018 -87.65304      casual
```

```
head(Q3_JUL_2020)
```

```
##          ride_id rideable_type      started_at      ended_at
## 1 762198876D69004D   docked_bike 2020-07-09 15:22:02 2020-07-09 15:25:52
## 2 BEC9C9FBA0D4CF1B   docked_bike 2020-07-24 23:56:30 2020-07-25 00:20:17
## 3 D2FD8EA432C77EC1   docked_bike 2020-07-08 19:49:07 2020-07-08 19:56:22
## 4 54AE594E20B35881   docked_bike 2020-07-17 19:06:42 2020-07-17 19:27:38
## 5 54025FDC7440B56F   docked_bike 2020-07-04 10:39:57 2020-07-04 10:45:05
## 6 65636B619E24257F   docked_bike 2020-07-28 16:33:03 2020-07-28 16:49:10
##          start_station_name start_station_id      end_station_name
## 1   Ritchie Ct & Banks St              180 Wells St & Evergreen Ave
## 2   Halsted St & Roscoe St             299      Broadway & Ridge Ave
## 3 Lake Shore Dr & Diversey Pkwy        329 Clark St & Wellington Ave
## 4   LaSalle St & Illinois St           181      Clark St & Armitage Ave
## 5   Lake Shore Dr & North Blvd          268      Clark St & Schiller St
## 6   Fairbanks St & Superior St         635      Wells St & Concord Ln
## end_station_id start_lat start_lng end_lat end_lng member_casual
## 1             291  41.90687 -87.62622 41.90672 -87.63483      member
## 2             461  41.94367 -87.64895 41.98404 -87.66027      member
## 3             156  41.93259 -87.63643 41.93650 -87.64754      casual
## 4              94  41.89076 -87.63170 41.91831 -87.63628      casual
## 5             301  41.91172 -87.62680 41.90799 -87.63150      member
## 6             289  41.89575 -87.62010 41.91213 -87.63466      casual
```

```
head(Q3_AUG_2020)
```

```
##          ride_id rideable_type      started_at      ended_at
## 1 322BD23D287743ED   docked_bike 2020-08-20 18:08:14 2020-08-20 18:17:51
## 2 2A3AEF1AB9054D8B   electric_bike 2020-08-27 18:46:04 2020-08-27 19:54:51
## 3 67DC1D133E8B5816   electric_bike 2020-08-26 19:44:14 2020-08-26 21:53:07
## 4 C79FBBD412E578A7   electric_bike 2020-08-27 12:05:41 2020-08-27 12:53:45
## 5 13814D3D661ECADB   electric_bike 2020-08-27 16:49:02 2020-08-27 16:59:49
## 6 56349A5A42F0AE51   electric_bike 2020-08-27 17:26:23 2020-08-27 18:07:50
##          start_station_name start_station_id      end_station_name
## 1 Lake Shore Dr & Diversey Pkwy           329   Clark St & Lincoln Ave
## 2      Michigan Ave & 14th St             168   Michigan Ave & 14th St
## 3      Columbus Dr & Randolph St           195   State St & Randolph St
## 4      Daley Center Plaza                 81     State St & Kinzie St
## 5      Leavitt St & Division St           658 Leavitt St & Division St
## 6      Leavitt St & Division St           658 Leavitt St & Division St
## end_station_id start_lat start_lng end_lat end_lng member_casual
## 1             141  41.93259 -87.63643 41.91569 -87.63460      member
## 2             168  41.86438 -87.62368 41.86422 -87.62344      casual
## 3              44  41.88464 -87.61955 41.88497 -87.62757      casual
## 4              47  41.88409 -87.62964 41.88958 -87.62754      casual
## 5             658  41.90299 -87.68377 41.90300 -87.68384      casual
## 6             658  41.90302 -87.68373 41.90309 -87.68363      casual
```

```
head(Q3_SEP_2020)
```

```
##          ride_id rideable_type      started_at      ended_at
## 1 2B22BD5F95FB2629   electric_bike 2020-09-17 14:27:11 2020-09-17 14:44:24
## 2 A7FB70B4AFC6CAF2   electric_bike 2020-09-17 15:07:31 2020-09-17 15:07:45
## 3 86057FA01BAC778E   electric_bike 2020-09-17 15:09:04 2020-09-17 15:09:35
## 4 57F6DC9A153DB98C   electric_bike 2020-09-17 18:10:46 2020-09-17 18:35:49
## 5 B9C4712F78C1AE68   electric_bike 2020-09-17 15:16:13 2020-09-17 15:52:55
## 6 378BBCE1E444EB80   electric_bike 2020-09-17 18:37:04 2020-09-17 19:23:28
##          start_station_name start_station_id      end_station_name
## 1      Michigan Ave & Lake St             52   Green St & Randolph St
## 2      W Oakdale Ave & N Broadway           NA W Oakdale Ave & N Broadway
## 3      W Oakdale Ave & N Broadway           NA W Oakdale Ave & N Broadway
## 4 Ashland Ave & Belle Plaine Ave          246      Montrose Harbor
## 5      Fairbanks Ct & Grand Ave            24   Fairbanks Ct & Grand Ave
## 6      Clark St & Armitage Ave             94
## end_station_id start_lat start_lng end_lat end_lng member_casual
## 1             112  41.88669 -87.62356 41.88357 -87.64873      casual
## 2              NA  41.94000 -87.64000 41.94000 -87.64000      casual
## 3              NA  41.94000 -87.64000 41.94000 -87.64000      casual
## 4             249  41.95606 -87.66892 41.96398 -87.63822      casual
## 5              24  41.89186 -87.62101 41.89135 -87.62032      casual
## 6              NA  41.91826 -87.63636 41.88000 -87.62000      casual
```

```
head(Q4_OCT_2020)
```

```
##          ride_id rideable_type      started_at      ended_at
## 1 ACB6B40CF5B9044C electric_bike 2020-10-31 19:39:43 2020-10-31 19:57:12
## 2 DF450C72FD109C01 electric_bike 2020-10-31 23:50:08 2020-11-01 00:04:16
## 3 B6396B54A15AC0DF electric_bike 2020-10-31 23:00:01 2020-10-31 23:08:22
## 4 44A4AEE261B9E854 electric_bike 2020-10-31 22:16:43 2020-10-31 22:19:35
## 5 10B7DD76A6A2EB95 electric_bike 2020-10-31 19:38:19 2020-10-31 19:54:32
## 6 DA6C3759660133DA electric_bike 2020-10-29 17:38:04 2020-10-29 17:45:43
##          start_station_name start_station_id      end_station_name
## 1 Lakeview Ave & Fullerton Pkwy          313      Rush St & Hubbard St
## 2 Southport Ave & Waveland Ave          227 Kedzie Ave & Milwaukee Ave
## 3 Stony Island Ave & 67th St           102 University Ave & 57th St
## 4 Clark St & Grace St                  165      Broadway & Sheridan Rd
## 5 Southport Ave & Wrightwood Ave        190      Stave St & Armitage Ave
## 6 Larrabee St & Division St             359      Wells St & Huron St
## end_station_id start_lat start_lng end_lat end_lng member_casual
## 1          125  41.92610 -87.63898 41.89035 -87.62607      casual
## 2          260  41.94817 -87.66391 41.92953 -87.70782      casual
## 3          423  41.77346 -87.58537 41.79145 -87.60005      casual
## 4          256  41.95085 -87.65924 41.95281 -87.65010      casual
## 5          185  41.92886 -87.66396 41.91778 -87.69143      casual
## 6           53  41.90353 -87.64335 41.89440 -87.63431      casual
```

```
head(Q4_NOV_2020)
```

```
##          ride_id rideable_type      started_at      ended_at
## 1 BD0A6FF6FFF9B921 electric_bike 2020-11-01 13:36:00 2020-11-01 13:45:40
## 2 96A7A7A4BDE4F82D electric_bike 2020-11-01 10:03:26 2020-11-01 10:14:45
## 3 C61526D06582BDC5 electric_bike 2020-11-01 00:34:05 2020-11-01 01:03:06
## 4 E533E89C32080B9E electric_bike 2020-11-01 00:45:16 2020-11-01 00:54:31
## 5 1C9F4EF18C168C60 electric_bike 2020-11-01 15:43:25 2020-11-01 16:16:52
## 6 7259585D8276D338 electric_bike 2020-11-14 15:55:17 2020-11-14 16:44:38
##          start_station_name start_station_id      end_station_name
## 1 Dearborn St & Erie St           110      St. Clair St & Erie St
## 2 Franklin St & Illinois St        672      Noble St & Milwaukee Ave
## 3 Lake Shore Dr & Monroe St         76      Federal St & Polk St
## 4 Leavitt St & Chicago Ave          659      Stave St & Armitage Ave
## 5 Buckingham Fountain              2      Buckingham Fountain
## 6 Wabash Ave & 16th St             72 Lake Shore Dr & Monroe St
## end_station_id start_lat start_lng end_lat end_lng member_casual
## 1          211  41.89418 -87.62913 41.89443 -87.62338      casual
## 2           29  41.89096 -87.63534 41.90067 -87.66248      casual
## 3           41  41.88098 -87.61675 41.87205 -87.62955      casual
## 4          185  41.89550 -87.68201 41.91774 -87.69139      casual
## 5           2  41.87650 -87.62036 41.87645 -87.62034      casual
## 6           76  41.86029 -87.62581 41.88099 -87.61677      casual
```

```
head(Q4_DEC_2020)
```

```
##          ride_id rideable_type      started_at      ended_at
## 1 70B6A9A437D4C30D  classic_bike 2020-12-27 12:44:29 2020-12-27 12:55:06
## 2 158A465D4E74C54A  electric_bike 2020-12-18 17:37:15 2020-12-18 17:44:19
## 3 5262016E0F1F2F9A  electric_bike 2020-12-15 15:04:33 2020-12-15 15:11:28
## 4 BE119628E44F871E  electric_bike 2020-12-15 15:54:18 2020-12-15 16:00:11
## 5 69AF78D57854E110  electric_bike 2020-12-22 12:08:17 2020-12-22 12:10:59
## 6 C1DECC4AB488831C  electric_bike 2020-12-22 13:26:37 2020-12-22 13:34:50
##          start_station_name start_station_id      end_station_name
## 1 Aberdeen St & Jackson Blvd      13157 Desplaines St & Kinzie St
## 2
## 3
## 4
## 5
## 6
##      end_station_id start_lat start_lng end_lat end_lng member_casual
## 1      TA1306000003  41.87773 -87.65479 41.88872 -87.64445      member
## 2
##      41.93000 -87.70000 41.91000 -87.70000      member
## 3
##      41.91000 -87.69000 41.93000 -87.70000      member
## 4
##      41.92000 -87.70000 41.91000 -87.70000      member
## 5
##      41.80000 -87.59000 41.80000 -87.59000      member
## 6
##      41.80000 -87.59000 41.78000 -87.60000      member
```

Set up environment: Data Cleaning

```
install.packages("tidyr") library(tidyr) install.packages("janitor") library("janitor")
```

```
colnames(Q1_2020) == colnames(Q2_APR_2020)
```

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

```
colnames(Q2_APR_2020) == colnames(Q2_MAY_2020)
```

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

```
colnames(Q2_MAY_2020) == colnames(Q2_JUN_2020)
```

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

```
colnames(Q2_JUN_2020) == colnames(Q3_JUL_2020)
```

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

```
colnames(Q3_JUL_2020) == colnames(Q3_AUG_2020)
```

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

```
colnames(Q3_AUG_2020) == colnames(Q3_SEP_2020)
```

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

```
colnames(Q3_SEP_2020) == colnames(Q4_OCT_2020)
```

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

```
colnames(Q4_OCT_2020) == colnames(Q4_NOV_2020)
```

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

```
colnames(Q4_NOV_2020) == colnames(Q4_DEC_2020)
```

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

```
colnames(Q4_DEC_2020) == colnames(Q1_2020)
```

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

Data Validation:

- double checked: data frame [ALL SAME FORMAT],
- double checked: colnames(abc) == colnames(xyz) [ALL RETURN TRUE; no anomalies in the column],
- double checked: rename() [NOT REQUIRED]

In case of data type compatibleness, use this code: `xyz <- mutate() {xyz, abc =as.integer(abc), def = as.characters(def)}`

Make sure data types are the same before merging data.

```
compare_df_cols(Q1_2020, Q2_APR_2020, Q2_MAY_2020, Q2_JUN_2020, Q3_JUL_2020, Q3_AUG_2020,
Q3_SEP_2020, Q4_OCT_2020, Q4_NOV_2020, Q4_DEC_2020)
```

- double checked: datatypes; mutate() [NOT REQUIRED], data is good for merging.

Merge Data for 2020 (12 files in total) into one data set

```
Q1234_2020 <- rbind(Q1_2020, Q2_APR_2020, Q2_MAY_2020, Q2_JUN_2020, Q3_JUL_2020,
Q3_AUG_2020, Q3_SEP_2020, Q4_OCT_2020, Q4_NOV_2020, Q4_DEC_2020)
```

```
View(Q1234_2020)
```

Integration successful!

What else is needed to answer our analysis? To answer time related concern, started_at and ended_at is currently "chr". Clearly, Date-Time format is vital, a conversion is needed for further calculation.

Set up environment: Data Cleaning: R Date-Time format

```
install.packages("lubridate") library(lubridate)
```

```
Q1234_2020started_at <- as_datetime(Q1234_2020started_at)
Q1234_2020ended_at <- as_datetime(Q1234_2020ended_at)
```

How to find the order of days in a week on R?

```
Q1234_2020weekday <- ordered(Q1234_2020date_weekday, level=c("Sunday", "Monday", "Tuesday",
"Wednesday", "Thursday", "Friday", "Saturday"))
```

```
Q1234_2020 %>% mutate(weekday = wday(started_at, label = TRUE)) %>% group_by(member_casual,
weekday) %>% arrange(member_casual, weekday)
```

```
date_dayname <- data.frame(date = as.Date(c("2020-01-01", "2020-01-02", "2020-01-03", "2020-01-04"))+1,
"%A")
weekday <- weekdays(date_dayname$date)
print(weekday)
```

```
Q1234_2020 %>% mutate(weekday = wday(started_at, label = TRUE)) %>% group_by(member_casual,
weekday) %>% arrange(member_casual, weekday)
```

```
Q1234_2020date <- as.Date(Q1234_2020started_at) #Q1234_2020
year <- format(as.Date(Q1234_2020date), "%Y") #Q1234_2020
month <- format(as.Date(Q1234_2020date), "%m") #Q1234_2020
day <- format(as.Date(Q1234_2020date), "%d") #Q1234_2020
weekday <- format(as.Date(Q1234_2020started_at), "%A")
```

```
View(Q1234_2020)
```

```
##Check date vs week days Sunday = 1
```

- wday(ymd("2019-12-29")) Sunday = 1
- wday(ymd("2019-12-30")) Monday = 2
- wday(ymd("2019-12-31")) Tuesday = 3
- wday(ymd("2020-01-01")) Wednesday = 4
- wday(ymd("2020-01-02")) Thursday = 5
- wday(ymd("2020-01-03")) Friday = 6
- wday(ymd("2020-01-04")) Saturday = 7
- wday(ymd("2020-01-21")) Code = 3, therefore it's Tuesday!
- yday("2020-12-31") Which day in 2020?

ANALYZE

Overall population statistic (such as mean, midpoint, average, min, max)

- nrow(Q1234_2020) # Rough total Population
- dim(Q1234_2020)
- head(Q1234_2020)
- tail(Q1234_2020)
- summary(Q1234_2020) # Show total population (Need further calculation!)

Check duration of rides data type before calculation, change its data type from "chr" into "numeric" to ease calculation.

```
Q1234_2020ride_duration_s <- difftime(Q1234_2020ended_at, Q1234_2020started_at) # in seconds!
```



```
Q1234_2020ride_duration_s <- as.numeric(as.character(Q1234_2020ride_duration_s)) # change "chr" to
"num" is.numeric(Q1234_2020ride_duration_s) #logical query, should return TRUE to validate we've
succesfully change the data types.
```

Remove unnecessary data columns for further analysis: start_lat, start_lng, end_lat, end_lng

- `Q1234_2020_cleaned1 <- Q1234_2020[!(Q1234_2020$ride_duration_s <=0),]` # Remove 0 or NIL seconds data, probably some failed rides, could be used in the future.
- `Q1234_2020_cleaned1 <- na.omit(Q1234_2020_cleaned1)` # Remove potential NA values
- `Q1234_2020_cleaned1 <- Q1234_2020_cleaned1 %>% distinct()` # Remove potential duplicates
- `Q1234_2020_cleaned1 <- Q1234_2020_cleaned1 %>% filter(!(is.na(start_station_name) | start_station_name == "")) %>% filter(!(is.na(end_station_name) | end_station_name == ""))` # Remove blank results
- `Q1234_2020_cleaned1 <- Q1234_2020 %>% select(-c(start_lat, start_lng, end_lat, end_lng))` #latitude and longitude might be helpful to plot data onto a map, i.e pin location of stations, and probably use different color to show it's frequency.
- `str(Q1234_2020_cleaned1)`

Arrange cleaned data according to dates

```
Q1234_2020_cleaned1started_at <- as.POSIXct(Q1234_2020_cleaned1started_at, format = "%m/%d/%Y
%l:%M:%S %p", tz = "GMT") class(Q1234_2020_cleaned1$started_at) Q1234_2020_cleaned1 <-
Q1234_2020_cleaned1[do.call(order, Q1234_2020_cleaned1), ]
```

Total population (calculate mean, midpoint, average, min, max)

```
nrow(Q1234_2020_cleaned1) Q1234_2020_cleaned1 %>% group_by(member_casual) %>%
summarise(ride_freq = length(ride_id))
```

```
Q1234_2020_cleaned1 <- Q1234_2020_cleaned1 %>% select(-c(start_lat, start_lng, end_lat, end_lng))
str(Q1234_2020_cleaned1) View(Q1234_2020_cleaned1)
```

Ride duration statistics

Total riding duration

- `sum(Q1234_2020_cleaned1$ride_duration_s) ##### Mean`
- `mean(Q1234_2020_cleaned1$ride_duration_s) ##### Midpoint`
- `median(Q1234_2020_cleaned1$ride_duration_s) ##### Longest ride`
- `max(Q1234_2020_cleaned1$ride_duration_s) ##### Shortest ride`
- `min(Q1234_2020_cleaned1$ride_duration_s) #Why is this happening? ##### Standard Deviation`
- `sd(Q1234_2020_cleaned1$ride_duration_s) ##### Variance`
- `var(Q1234_2020_cleaned1$ride_duration_s) ### Riding duration by type of customer`
- `setNames(aggregate(ride_duration_s ~ member_casual, Q1234_2020_cleaned1, sum), c("customer_type", "ride_duration_s"))` # Casual riders more than member!!!! Almost doubled!

Rideable type statistic

Rideable type population

- `table(Q1234_2020_cleaned1$rideable_type)` # Docked > Electric > Classic ##### Customer rideable preferences

- `str(Q1234_2020_cleaned1)`

From here, we can also see the bike preferences for both member and casual riders. ##### Change rideable type from “num” to “chr” * `Q1234_2020_cleaned1 %>% group_by(Q1234_2020_cleaned1 member_casual)count(Q1234_2020_cleaned1rideable_type)`

Find the population for members (calculate mean, midpoint, average, min, max)

- `aggregate(Q1234_2020_cleaned1ride_duration_s Q1234_2020_cleaned1 member_casual, FUN = sum)`
- `aggregate(Q1234_2020_cleaned1ride_duration_s Q1234_2020_cleaned1 member_casual, FUN = mean)`
- `aggregate(Q1234_2020_cleaned1ride_duration_s Q1234_2020_cleaned1 member_casual, FUN = median)`
- `aggregate(Q1234_2020_cleaned1ride_duration_s Q1234_2020_cleaned1 member_casual, FUN = max)`
- `aggregate(Q1234_2020_cleaned1ride_duration_s Q1234_2020_cleaned1 member_casual, FUN = min)`
- `aggregate(Q1234_2020_cleaned1ride_duration_s Q1234_2020_cleaned1 member_casual, FUN = sd)`
- `aggregate(Q1234_2020_cleaned1ride_duration_s Q1234_2020_cleaned1 member_casual, FUN = var)`
- `aggregate(Q1234_2020_cleaned1ride_duration_s Q1234_2020_cleaned1 member_casual, FUN = range)`
- `aggregate(Q1234_2020_cleaned1ride_duration_s Q1234_2020_cleaned1 member_casual, FUN = quantile)`

```
Q1234_2020_cleaned1 %>% group_by(ride_duration_s) %>% summarize(sum = sum(ride_duration_s, na.rm = TRUE), mean = mean(ride_duration_s, na.rm = TRUE), median = median(ride_duration_s, na.rm = TRUE), max = max(ride_duration_s, na.rm = TRUE), min = min(ride_duration_s, na.rm = TRUE), sd = sd(ride_duration_s, na.rm = TRUE), var = var(ride_duration_s, na.rm = TRUE), range = diff(range(ride_duration_s, na.rm = TRUE)), quantile = list(quantile(ride_duration_s, probs = seq(.1, 1, by = .1), na.rm = TRUE))) %>% unnest_wider(quantile)
```

Find the population for casuals (calculate mean, midpoint, average, min, max)

- `mean(3732614925/2)`
- `median((3732614925+1952034518)/2)`

SHARE

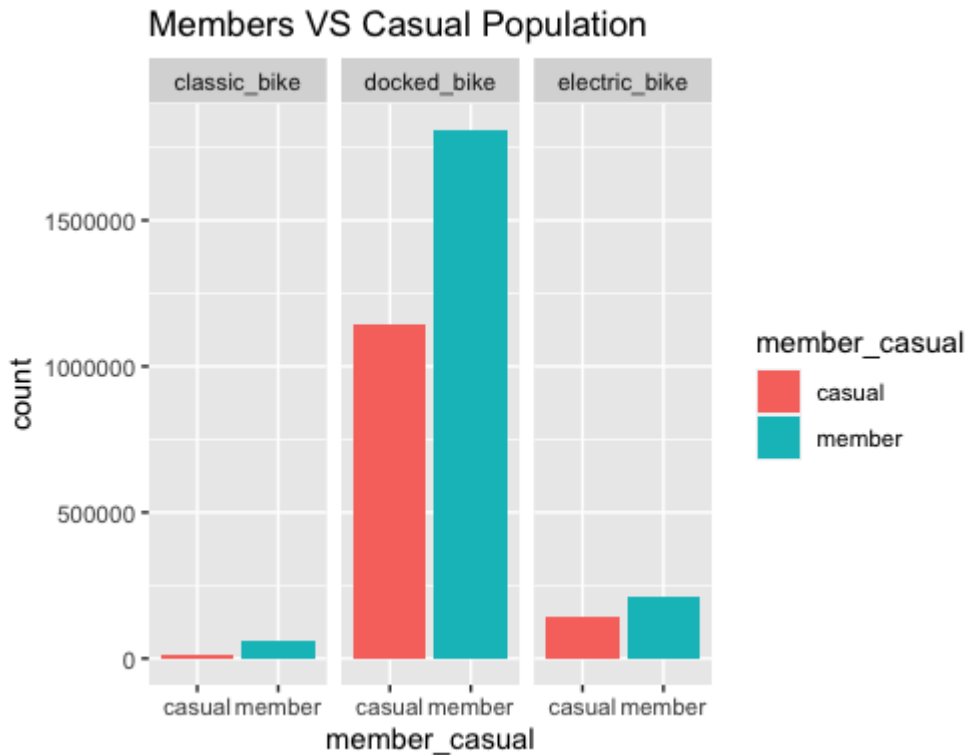
Set up environment: Data Visualization

```
install.packages("ggplot2") library(ggplot2) install.packages("lubridate") library(lubridate)
install.packages("ggrepel") library(ggrepel) installed.packages("rmarkdown")
```

Plot distribution total population (Membership VS Casual)

```
Q1234_2020_cleaned1 %>% group_by(member_casual) %>% summarise(ride_freq = length(ride_id)) %>%
arrange(member_casual) %>% ggplot(aes(x=member_casual,y=ride_freq,fill=member_casual)) + labs(title
="Members VS Casual Population")
```

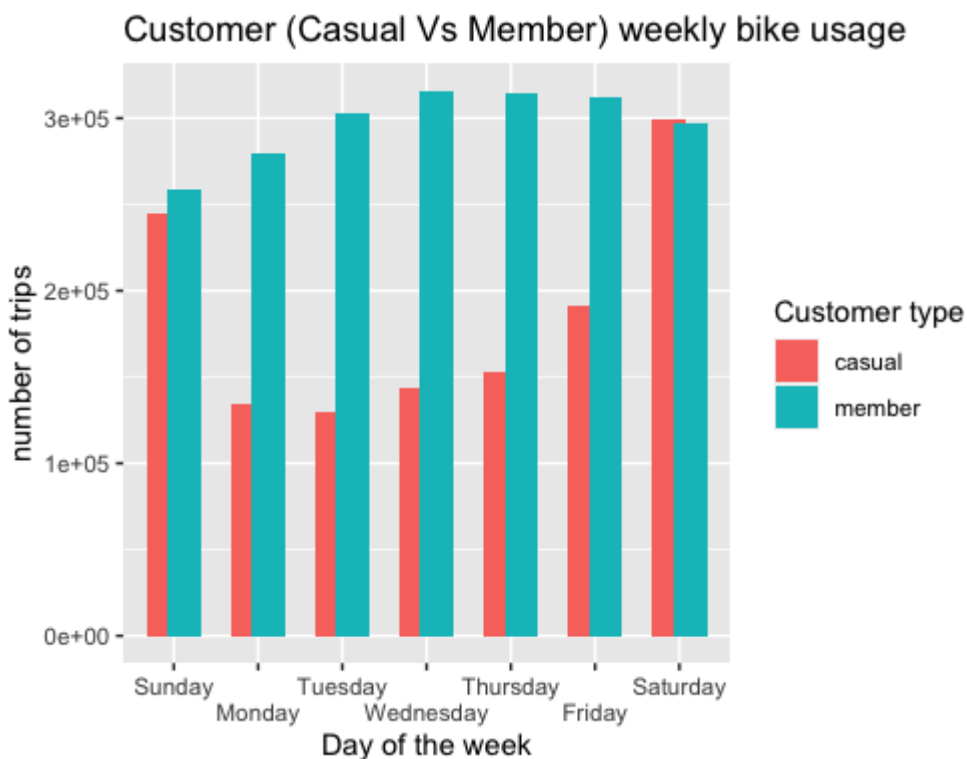
```
ggplot(data = Q1234_2020_cleaned1) + geom_bar(mapping=aes(x=member_casual,fill=member_casual)) +
labs(title = "Members VS Casual Population") + facet_wrap(~rideable_type)
```



Plot Member Vs Casual Population

Casual VS Member weekly bike usage

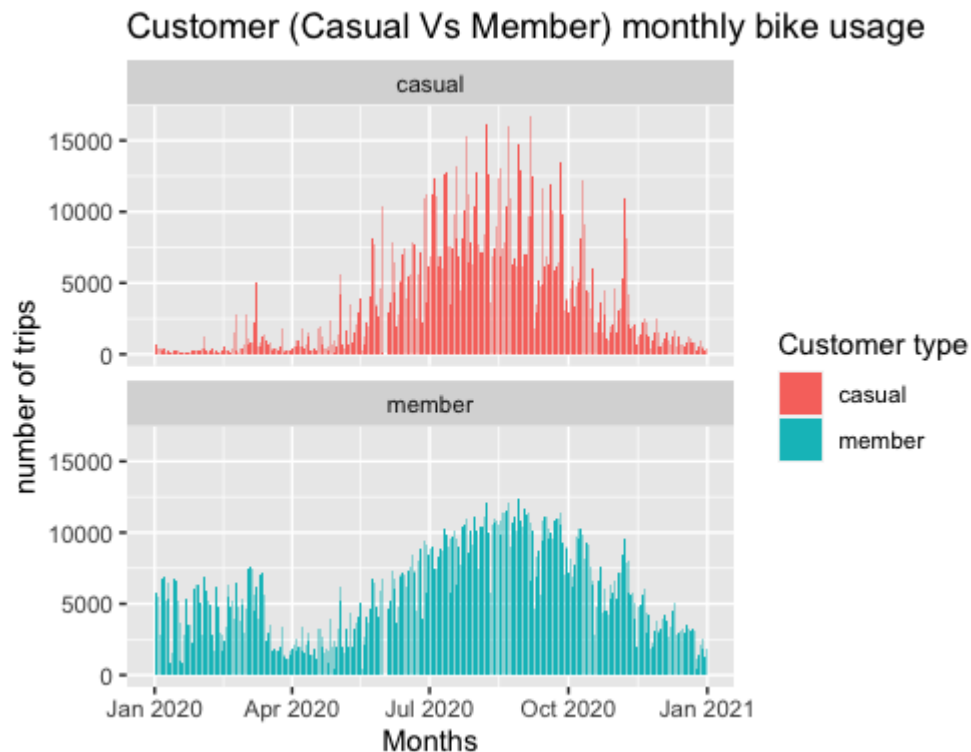
```
Q1234_2020_cleaned1 %>% group_by(member_casual, week_name) %>% summarise(ride_freq = n()) %>%
  arrange(member_casual, week_name) %>% ggplot(mapping = aes(x = factor(week_name, c("Sunday",
    "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday")), y=ride_freq, fill=member_casual)) +
  labs(title = "Customer (Casual Vs Member) weekly bike usage", x="Day of the week", y="number of trips",
    fill="Customer type") + geom_col(width = 0.8, position = position_dodge(width=0.5)) + scale_x_discrete(guide
    = guide_axis(n.dodge=2))
```



Plot Customer(Casual VS Member) weekly bike usage

Casual VS Member monthly bike usage

```
Q1234_2020_cleaned1 %>% group_by(member_casual, date) %>% summarise(ride_freq = n()) %>%
  arrange(member_casual, date) %>% ggplot(mapping = aes(x = date, y=ride_freq, fill=member_casual)) +
  labs(title = "Customer (Casual Vs Member) monthly bike usage", x="Months", y="number of trips",
  fill="Customer type") + geom_col(width = 0.6, position = position_dodge(width=0.6)) +
  facet_wrap(~member_casual, Q1234_2020_cleaned1$date) # + scale_x_continuous(guide =
  guide_axis(n.dodge=2))
```



Plot Casual VS Member monthly bike usage

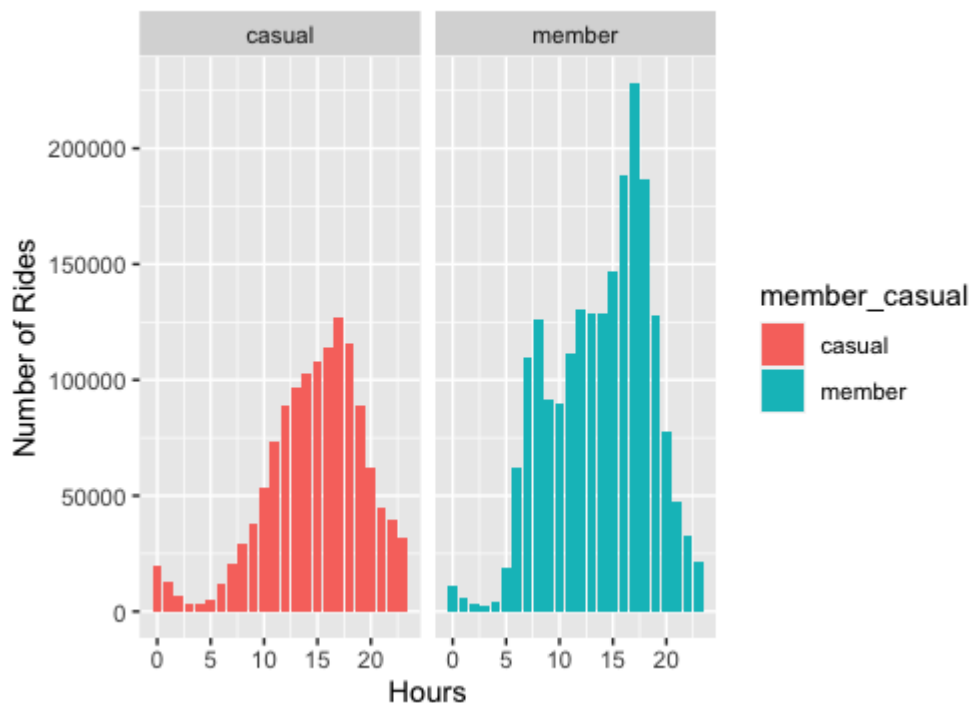
Popular Riding Hours

```
ggplot(data = Q1234_2020_cleaned1) + geom_bar(mapping = aes(x = hour(started_at), fill= member_casual)) +
  facet_wrap(~member_casual) + theme(axis.text.x = element_text(angle = 0)) + labs(title="Divvy ride trip hourly
  trend for Casual and Member rider", x="Hours", y="Number of Rides")
```

```
ggplot(data = Q1234_2020_cleaned1) + geom_freqpoly(mapping = aes(x = hour(started_at), color=
  member_casual)) + facet_wrap(~Day_of_week) + theme(axis.text.x = element_text(angle = 45)) +
  labs(title="Hourly distribution of casual rides and member rides for each day of week", x="Hours", y="Number
  of Rides")
```

```
hour <- c("0:00", "1:00", "2:00", "3:00", "4:00", "6:00", "7:00", "8:00", "9:00", "10:00", "11:00", "12:00",
  "13:00", "14:00", "15:00", "16:00", "17:00", "18:00", "19:00", "20:00", "21:00", "22:00", "23:00") # Can't seem
  to visualise this format!
```

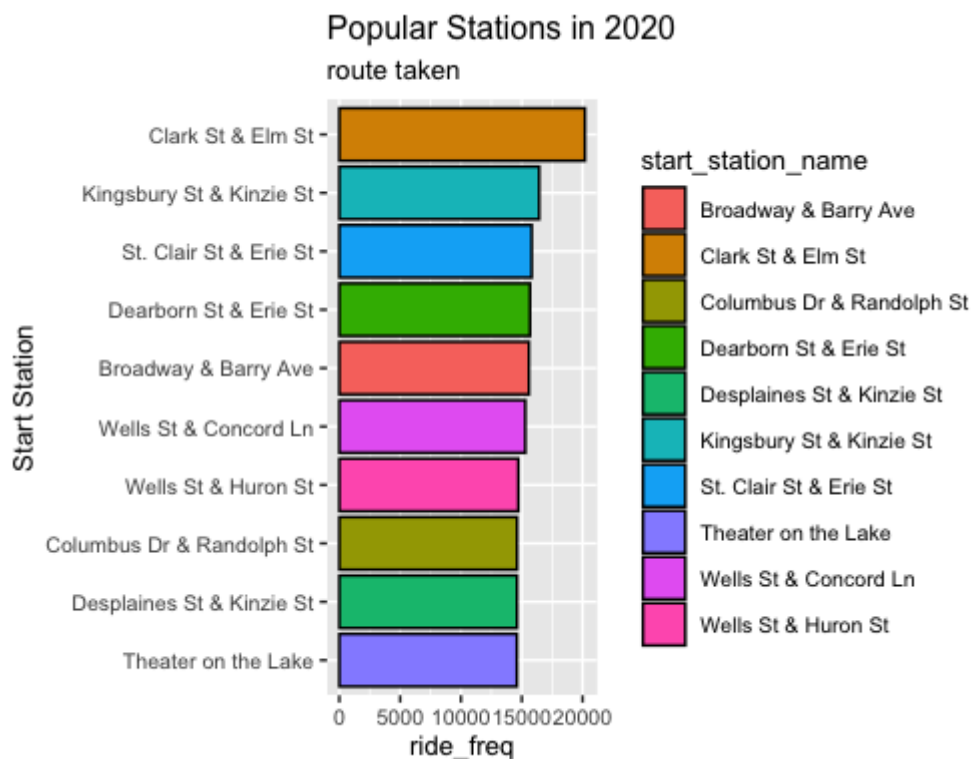
Divvy ride trip hourly trend for Casual and Member ride



Plot Popular Riding Hour

Popular Routes

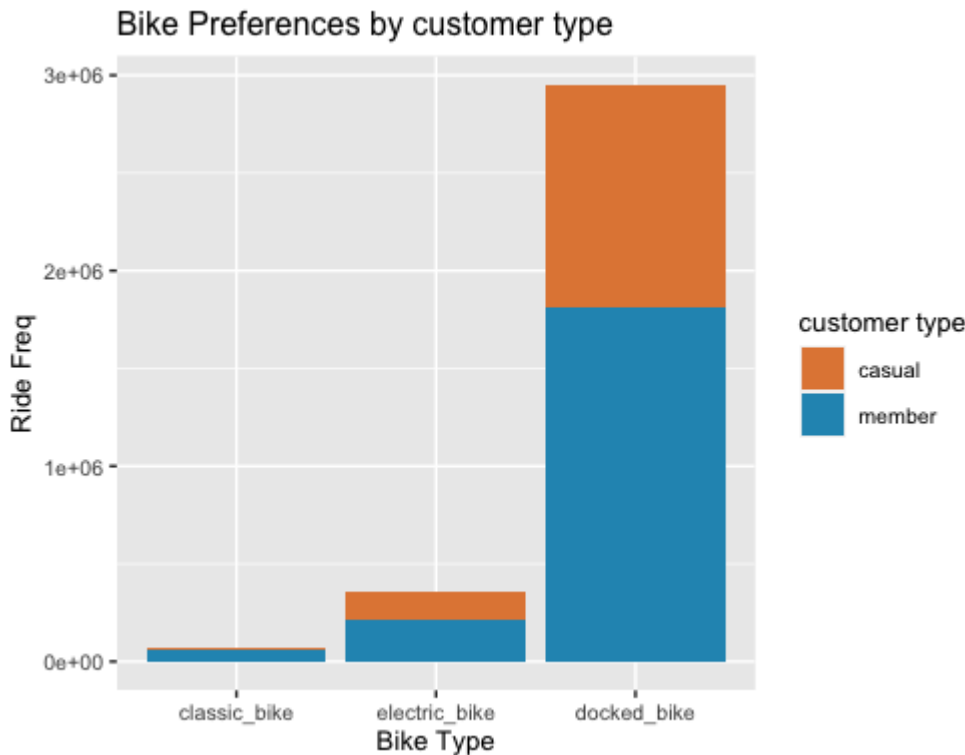
```
Q1234_2020_cleaned1 %>% filter(member_casual == "member" & start_station_name != "missing") %>%
  group_by(start_station_name) %>% summarise(ride_freq = n()) %>% arrange(-ride_freq) %>% head(10) %>%
  ggplot(aes(x = ride_freq, y = reorder(start_station_name, ride_freq), fill = start_station_name)) +
  #geom_col(alpha = .3) + geom_col(color = "black") + theme(text = element_text(size = 10)) + labs(title = "Popular
  Stations in 2020", subtitle = "route taken", y = "Start Station")
```



Plot Popular Route

Bike preferences by Casual and Member

```
Q1234_2020_cleaned1 %>% group_by(rideable_type, member_casual) %>% summarise(ride_freq = n(),
.groups = "drop") %>% arrange(ride_freq) %>% ggplot(aes(reorder(rideable_type, ride_freq), ride_freq,
fill=member_casual)) + geom_bar(stat="identity") + labs(title = "Bike Preferences by customer type", x = "Bike
Type", y="Ride Freq", fill="customer type") + theme(text= element_text(size=10)) + scale_y_continuous() +
scale_fill_manual(values = c("casual" = "#e28743", "member" = "#2596be"))
```



Plot Bike Preferences

Conclusion:

From our plots, we know a little bit more about our customer. These data allow us to read the decision made by Cyclistic's customers. They've preferred docked bike most of the time, probably because of the easy access and usage. For classic bike and electric bike, it is less profitable, we need more data on this. Surprisingly, bike usage dropped in April 2020, we need more data on the weather changes and new wave of Covid-19 outbreak.

There were many limitations in this analysis, such as the time constraint and also the massive 9 years worth of data. We were not able to fully understand the whole situation, hence this particular analysis is only good for sampling and archiving, the wealth of information if uncovered will surely provide better insight for targeting specific users, that includes diving into sensitive data that got hidden away.

ACT

- Financial incentives for casual riders during working days, a campaign is needed to entice casual riders opting for healthier and cost effective lifestyle. They can even save time avoiding unnecessary slow traffic.
- Automatic upgrade to membership for free, casual riders exceeding a certain number of usage could enjoy this.
- A weekly pass, monthly pass, annual pass should be priced so that the option with longer usage become cheapest, this might entice and allow money conscious casual riders to gear towards better options.
- More perks should be given to members, such as getting discounted coffee or breakfast at popular routes.

- Social media is a great tool, influencer on YouTube, Instagram, Facebook, could be featured to promote better lifestyle through bike share

Resources

Github, RStudio, Kaggle, Medium, rpubs, Stackoverflow

What's next?

- Merge data from 2013 to 2022 for more comprehensive insight.
- Use Google Map to pin all the stations, the color intensity of each individual pin should indicate the volume of usage.

Personal note:



Chicago docked bike 2022

I'm a big supporter of city bikes, I have enjoyed my ride around Chicago city although the weather was brutally cold especially for a tropical North-Bornean.

Being a bike enthusiast, it was definitely a win-win being able to solve real world problem related to my interest. I'm very excited to get onto the next data related task!