Case Study: How Does a Bike-Share Navigate Speedy Success?

Ankita Gairola

2023-07-29

About the Company

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime. Until now, Cyclistic's marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders.

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

Goal of Case Study

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

Director has assigned first questions: How do annual members and casual riders use Cyclistic bikes differently?

I have to produce with the following deliverable:

1. A clear statement of the business task
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of your analysis
5. Supporting visualizations and key findings
6. Your top three recommendations based on your analysis

Ask

Business task : Maximise the number of annual membership

Key Stakeholder: The director of marketing, The marketing analysis team, and Cyclistic executive team.

Prepare

Data set: Divvy's, a bike share programme based in Chicago, data from April 2020 - March 2021

Set working directory

getwd()

setwd("E:/Capstone projects_Google/Case Study 1")

Required packages and read the data

install.packages("readxl")

install.packages("tidyverse")

install.packages("lubridate")

install.packages("ggplot2")

install.packages("dplyr")

install.packages("tidyr")

install.packages("geosphere")

install.packages("reprex")

install.packages("skimr")

```
library("tidyverse")
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✓ dplyr     1.1.2     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.2     ✓ tibble    3.2.1
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.1
## ── Conflicts ────────────────────────────────────────
tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
library(lubridate)
```

```
library("ggplot2")
```

```r
library("dplyr")

library("tidyr")

library("readxl")

library("geosphere")
```

```
## The legacy packages maptools, rgdal, and rgeos, underpinning the sp
package,
## which was just loaded, will retire in October 2023.
## Please refer to R-spatial evolution reports for details, especially
## https://r-spatial.org/r/2023/05/15/evolution4.html.
## It may be desirable to make the sf package available;
## package maintainers should consider adding sf to Suggests:.
## The sp package is now running under evolution status 2
##      (status 2 uses the sf package in place of rgdal)
```

```r
library("reprex")

library("skimr")

April_2020 <- read.csv("E:/Capstone projects_Google/Case Study 1/working
data/202004-divvy-tripdata.csv")

May_2020 <- read.csv("E:/Capstone projects_Google/Case Study 1/working
data/202005-divvy-tripdata.csv")

June_2020 <- read.csv("E:/Capstone projects_Google/Case Study 1/working
data/202006-divvy-tripdata.csv")

July_2020 <- read.csv("E:/Capstone projects_Google/Case Study 1/working
data/202007-divvy-tripdata.csv")

August_2020 <- read.csv("E:/Capstone projects_Google/Case Study 1/working
data/202008-divvy-tripdata.csv")

September_2020 <- read.csv("E:/Capstone projects_Google/Case Study 1/working
data/202009-divvy-tripdata.csv")

October_2020 <- read.csv("E:/Capstone projects_Google/Case Study 1/working
data/202010-divvy-tripdata.csv")

Novemeber_2020 <- read.csv("E:/Capstone projects_Google/Case Study 1/working
data/202011-divvy-tripdata.csv")

December_2020 <- read.csv("E:/Capstone projects_Google/Case Study 1/working
data/202012-divvy-tripdata.csv")
```

```
January_2021 <- read.csv("E:/Capstone projects_Google/Case Study 1/working data/202101-divvy-tripdata.csv")

Febraury_2021 <- read.csv("E:/Capstone projects_Google/Case Study 1/working data/202102-divvy-tripdata.csv")

March_2021 <- read.csv("E:/Capstone projects_Google/Case Study 1/working data/202103-divvy-tripdata.csv")
```

For data check

```
colnames(April_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(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(June_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(July_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(August_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(September_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(October_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(Novemeber_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(December_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(January_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(Febraury_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(March_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"
```

For data check

```
str(April_2020)
```

```
## 'data.frame':    84776 obs. of  13 variables:
##  $ ride_id           : chr  "A847FADBBC638E45" "5405B80E996FF60D"
"5DD24A79A4E006F4" "2A59BBDF5CDBA725" ...
##  $ rideable_type     : chr  "docked_bike" "docked_bike" "docked_bike"
"docked_bike" ...
##  $ started_at        : chr  "2020-04-26 17:45:14" "2020-04-17 17:08:54"
"2020-04-01 17:54:13" "2020-04-07 12:50:19" ...
##  $ ended_at          : chr  "2020-04-26 18:12:03" "2020-04-17 17:17:03"
"2020-04-01 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" "Indiana 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(May_2020)
```

```
## 'data.frame':    200274 obs. of  13 variables:
##  $ ride_id           : chr  "02668AD35674B983" "7A50CCAF1EDDB28F"
"2FFCDFDB91FE9A52" "58991CF1DB75BA84" ...
##  $ rideable_type     : chr  "docked_bike" "docked_bike" "docked_bike"
"docked_bike" ...
##  $ started_at        : chr  "2020-05-27 10:03:52" "2020-05-25 10:47:11"
"2020-05-02 14:11:03" "2020-05-02 16:25:36" ...
##  $ ended_at          : chr  "2020-05-27 10:16:49" "2020-05-25 11:05:40"
"2020-05-02 15:48:21" "2020-05-02 16:39:28" ...
##  $ start_station_name: chr  "Franklin St & Jackson Blvd" "Clark St &
Wrightwood Ave" "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" "Kedzie 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(June_2020)
```

```
## 'data.frame':    343005 obs. of  13 variables:
##  $ ride_id           : chr  "8CD5DE2C2B6C4CFC" "9A191EB2C751D85D"
"F37D14B0B5659BCF" "C41237B506E85FA1" ...
```

```
##  $ rideable_type     : chr  "docked_bike" "docked_bike" "docked_bike"
"docked_bike" ...
##  $ started_at        : chr  "2020-06-13 23:24:48" "2020-06-26 07:26:10"
"2020-06-23 17:12:41" "2020-06-20 01:09:35" ...
##  $ ended_at          : chr  "2020-06-13 23:36:55" "2020-06-26 07:31:58"
"2020-06-23 17:21:14" "2020-06-20 01:28:24" ...
##  $ start_station_name: chr  "Wilton Ave & Belmont Ave" "Federal St & Polk
St" "Daley 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(July_2020)
```

```
## 'data.frame':    551480 obs. of  13 variables:
##  $ ride_id           : chr  "762198876D69004D" "BEC9C9FBA0D4CF1B"
"D2FD8EA432C77EC1" "54AE594E20B35881" ...
##  $ rideable_type     : chr  "docked_bike" "docked_bike" "docked_bike"
"docked_bike" ...
##  $ started_at        : chr  "2020-07-09 15:22:02" "2020-07-24 23:56:30"
"2020-07-08 19:49:07" "2020-07-17 19:06:42" ...
##  $ ended_at          : chr  "2020-07-09 15:25:52" "2020-07-25 00:20:17"
"2020-07-08 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" "Clark 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(August_2020)
```

```
## 'data.frame':    622361 obs. of  13 variables:
##  $ ride_id           : chr  "322BD23D287743ED" "2A3AEF1AB9054D8B"
"67DC1D133E8B5816" "C79FBBD412E578A7" ...
##  $ rideable_type     : chr  "docked_bike" "electric_bike" "electric_bike"
"electric_bike" ...
##  $ started_at        : chr  "2020-08-20 18:08:14" "2020-08-27 18:46:04"
"2020-08-26 19:44:14" "2020-08-27 12:05:41" ...
##  $ ended_at          : chr  "2020-08-20 18:17:51" "2020-08-27 19:54:51"
"2020-08-26 21:53:07" "2020-08-27 12:53:45" ...
```

```
##  $ start_station_name: chr  "Lake Shore Dr & Diversey Pkwy" "Michigan Ave
& 14th St" "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" "State 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(September_2020)

## 'data.frame':    532958 obs. of  13 variables:
##  $ ride_id           : chr  "2B22BD5F95FB2629" "A7FB70B4AFC6CAF2"
"86057FA01BAC778E" "57F6DC9A153DB98C" ...
##  $ rideable_type     : chr  "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : chr  "2020-09-17 14:27:11" "2020-09-17 15:07:31"
"2020-09-17 15:09:04" "2020-09-17 18:10:46" ...
##  $ ended_at          : chr  "2020-09-17 14:44:24" "2020-09-17 15:07:45"
"2020-09-17 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(October_2020)

## 'data.frame':    388653 obs. of  13 variables:
##  $ ride_id           : chr  "ACB6B40CF5B9044C" "DF450C72FD109C01"
"B6396B54A15AC0DF" "44A4AEE261B9E854" ...
##  $ rideable_type     : chr  "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : chr  "2020-10-31 19:39:43" "2020-10-31 23:50:08"
"2020-10-31 23:00:01" "2020-10-31 22:16:43" ...
##  $ ended_at          : chr  "2020-10-31 19:57:12" "2020-11-01 00:04:16"
"2020-10-31 23:08:22" "2020-10-31 22:19:35" ...
##  $ start_station_name: chr  "Lakeview Ave & Fullerton Pkwy" "Southport Ave
& Waveland 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" "University 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(Novemeber_2020)
```

```
## 'data.frame':    259716 obs. of  13 variables:
##  $ ride_id           : chr  "BD0A6FF6FFF9B921" "96A7A7A4BDE4F82D"
"C61526D06582BDC5" "E533E89C32080B9E" ...
##  $ rideable_type     : chr  "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : chr  "2020-11-01 13:36:00" "2020-11-01 10:03:26"
"2020-11-01 00:34:05" "2020-11-01 00:45:16" ...
##  $ ended_at          : chr  "2020-11-01 13:45:40" "2020-11-01 10:14:45"
"2020-11-01 01:03:06" "2020-11-01 00:54:31" ...
##  $ start_station_name: chr  "Dearborn St & Erie St" "Franklin St &
Illinois St" "Lake 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" "Federal 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(December_2020)
```

```
## 'data.frame':    131573 obs. of  13 variables:
##  $ ride_id           : chr  "70B6A9A437D4C30D" "158A465D4E74C54A"
"5262016E0F1F2F9A" "BE119628E44F871E" ...
##  $ rideable_type     : chr  "classic_bike" "electric_bike" "electric_bike"
"electric_bike" ...
##  $ started_at        : chr  "2020-12-27 12:44:29" "2020-12-18 17:37:15"
"2020-12-15 15:04:33" "2020-12-15 15:54:18" ...
##  $ ended_at          : chr  "2020-12-27 12:55:06" "2020-12-18 17:44:19"
"2020-12-15 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" ...
```

```
str(January_2021)
```

```
## 'data.frame':    96834 obs. of  13 variables:
##  $ ride_id           : chr  "E19E6F1B8D4C42ED" "DC88F20C2C55F27F"
"EC45C94683FE3F27" "4FA453A75AE377DB" ...
##  $ rideable_type     : chr  "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
##  $ started_at        : chr  "2021-01-23 16:14:19" "2021-01-27 18:43:08"
"2021-01-21 22:35:54" "2021-01-07 13:31:13" ...
##  $ ended_at          : chr  "2021-01-23 16:24:44" "2021-01-27 18:47:12"
"2021-01-21 22:37:14" "2021-01-07 13:42:55" ...
##  $ start_station_name: chr  "California Ave & Cortez St" "California Ave &
Cortez St" "California Ave & Cortez St" "California Ave & Cortez St" ...
##  $ start_station_id  : chr  "17660" "17660" "17660" "17660" ...
##  $ end_station_name  : chr  "" "" "" "" ...
##  $ end_station_id    : chr  "" "" "" "" ...
##  $ start_lat         : num  41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num  -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num  41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num  -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr  "member" "member" "member" "member" ...

str(Febraury_2021)

## 'data.frame':    49622 obs. of  13 variables:
##  $ ride_id           : chr  "89E7AA6C29227EFF" "0FEFDE2603568365"
"E6159D746B2DBB91" "B32D3199F1C2E75B" ...
##  $ rideable_type     : chr  "classic_bike" "classic_bike" "electric_bike"
"classic_bike" ...
##  $ started_at        : chr  "2021-02-12 16:14:56" "2021-02-14 17:52:38"
"2021-02-09 19:10:18" "2021-02-02 17:49:41" ...
##  $ ended_at          : chr  "2021-02-12 16:21:43" "2021-02-14 18:12:09"
"2021-02-09 19:19:10" "2021-02-02 17:54:06" ...
##  $ start_station_name: chr  "Glenwood Ave & Touhy Ave" "Glenwood Ave &
Touhy Ave" "Clark St & Lake St" "Wood St & Chicago Ave" ...
##  $ start_station_id  : chr  "525" "525" "KA1503000012" "637" ...
##  $ end_station_name  : chr  "Sheridan Rd & Columbia Ave" "Bosworth Ave &
Howard St" "State St & Randolph St" "Honore St & Division St" ...
##  $ end_station_id    : chr  "660" "16806" "TA1305000029" "TA1305000034"
...
##  $ start_lat         : num  42 42 41.9 41.9 41.8 ...
##  $ start_lng         : num  -87.7 -87.7 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num  42 42 41.9 41.9 41.8 ...
##  $ end_lng           : num  -87.7 -87.7 -87.6 -87.7 -87.6 ...
##  $ member_casual     : chr  "member" "casual" "member" "member" ...

str(March_2021)

## 'data.frame':    228496 obs. of  13 variables:
##  $ ride_id           : chr  "CFA86D4455AA1030" "30D9DC61227D1AF3"
"846D87A15682A284" "994D05AA75A168F2" ...
##  $ rideable_type     : chr  "classic_bike" "classic_bike" "classic_bike"
"classic_bike" ...
##  $ started_at        : chr  "2021-03-16 08:32:30" "2021-03-28 01:26:28"
```

```
"2021-03-11 21:17:29" "2021-03-11 13:26:42" ...
##  $ ended_at          : chr  "2021-03-16 08:36:34" "2021-03-28 01:36:55"
"2021-03-11 21:33:53" "2021-03-11 13:55:41" ...
##  $ start_station_name: chr  "Humboldt Blvd & Armitage Ave" "Humboldt Blvd
& Armitage Ave" "Shields Ave & 28th Pl" "Winthrop Ave & Lawrence Ave" ...
##  $ start_station_id  : chr  "15651" "15651" "15443" "TA1308000021" ...
##  $ end_station_name  : chr  "Stave St & Armitage Ave" "Central Park Ave &
Bloomingdale Ave" "Halsted St & 35th St" "Broadway & Sheridan Rd" ...
##  $ end_station_id    : chr  "13266" "18017" "TA1308000043" "13323" ...
##  $ start_lat         : num  41.9 41.9 41.8 42 42 ...
##  $ start_lng         : num  -87.7 -87.7 -87.6 -87.7 -87.7 ...
##  $ end_lat           : num  41.9 41.9 41.8 42 42.1 ...
##  $ end_lng           : num  -87.7 -87.7 -87.6 -87.6 -87.7 ...
##  $ member_casual     : chr  "casual" "casual" "casual" "casual" ...
```

Convert data from double to character

```
April_2020 <- mutate(April_2020, start_station_id =
as.character(start_station_id), end_station_id =
as.character(end_station_id))

May_2020 <- mutate(May_2020, start_station_id =
as.character(start_station_id), end_station_id =
as.character(end_station_id))

June_2020 <- mutate(June_2020, start_station_id =
as.character(start_station_id), end_station_id =
as.character(end_station_id))

July_2020 <- mutate(July_2020, start_station_id =
as.character(start_station_id), end_station_id =
as.character(end_station_id))

August_2020 <-mutate(August_2020, start_station_id =
as.character(start_station_id), end_station_id =
as.character(end_station_id))

September_2020 <-mutate(September_2020, start_station_id =
as.character(start_station_id), end_station_id =
as.character(end_station_id))

October_2020 <-mutate(October_2020, start_station_id =
as.character(start_station_id), end_station_id =
as.character(end_station_id))

Novemeber_2020 <- mutate(Novemeber_2020, start_station_id =
as.character(start_station_id), end_station_id =
as.character(end_station_id))

December_2020 <-mutate(December_2020, start_station_id =
as.character(start_station_id), end_station_id =
```

```
as.character(end_station_id))

January_2021 <-mutate(January_2021, start_station_id =
as.character(start_station_id), end_station_id =
as.character(end_station_id))

Febraury_2021 <-mutate(Febraury_2021, start_station_id =
as.character(start_station_id), end_station_id =
as.character(end_station_id))

March_2021 <-mutate(March_2021, start_station_id =
as.character(start_station_id), end_station_id =
as.character(end_station_id))
```

Merge all the data set

```
All_Trip_Data <- bind_rows(April_2020,May_2020,June_2020,July_2020,
August_2020, September_2020, October_2020, Novemeber_2020, December_2020,
January_2021, Febraury_2021, March_2021)
```

Process

Data cleaning before analysis

Number of column

```
ncol(All_Trip_Data)

## [1] 13
```

Number of rows

```
nrow(All_Trip_Data)

## [1] 3489748
```

Column name

```
colnames(All_Trip_Data)

##  [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"
```

Dimensions of data

```
dim(All_Trip_Data)

## [1] 3489748      13
```

For first six rows of data

```
head(All_Trip_Data)

##            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
```

Structure of data i.e (character, numeric, vectors, list etc)

```
str(All_Trip_Data)
```

```
## 'data.frame':    3489748 obs. of  13 variables:
##  $ ride_id           : chr  "A847FADBBC638E45" "5405B80E996FF60D"
"5DD24A79A4E006F4" "2A59BBDF5CDBA725" ...
##  $ rideable_type     : chr  "docked_bike" "docked_bike" "docked_bike"
"docked_bike" ...
##  $ started_at        : chr  "2020-04-26 17:45:14" "2020-04-17 17:08:54"
"2020-04-01 17:54:13" "2020-04-07 12:50:19" ...
##  $ ended_at          : chr  "2020-04-26 18:12:03" "2020-04-17 17:17:03"
"2020-04-01 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  : chr  "86" "503" "142" "216" ...
##  $ end_station_name  : chr  "Lincoln Ave & Diversey Pkwy" "Kosciuszko
Park" "Indiana Ave & Roosevelt Rd" "Wood St & Augusta Blvd" ...
##  $ end_station_id    : chr  "152" "499" "255" "657" ...
##  $ 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" ...
```

Statistical summary of data

```
summary(All_Trip_Data)
```

```
##    ride_id          rideable_type       started_at          ended_at
##  Length:3489748     Length:3489748     Length:3489748     Length:3489748
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  start_station_name start_station_id   end_station_name   end_station_id
##  Length:3489748     Length:3489748     Length:3489748     Length:3489748
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    start_lat       start_lng         end_lat         end_lng
##  Min.   :41.64   Min.   :-87.87   Min.   :41.54   Min.   :-88.07
##  1st Qu.:41.88   1st Qu.:-87.66   1st Qu.:41.88   1st Qu.:-87.66
##  Median :41.90   Median :-87.64   Median :41.90   Median :-87.64
##  Mean   :41.90   Mean   :-87.64   Mean   :41.90   Mean   :-87.64
##  3rd Qu.:41.93   3rd Qu.:-87.63   3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.08   Max.   :-87.52   Max.   :42.16   Max.   :-87.44
##                                   NA's   :4738    NA's   :4738
##  member_casual
##  Length:3489748
##  Class :character
##  Mode  :character
##
##
##
##
```

Add columns of date, month, and year of the ride

```
All_Trip_Data$date <- as.Date(All_Trip_Data$started_at)
All_Trip_Data$month <- format(as.Date(All_Trip_Data$date), "%m")
All_Trip_Data$day <- format(as.Date(All_Trip_Data$date), "%d")
All_Trip_Data$year <- format(as.Date(All_Trip_Data$date), "%Y")
All_Trip_Data$day_of_week <- format(as.Date(All_Trip_Data$date), "%A")
```

Add ‘time’ column after extract time from the ‘started_at’ column

```
All_Trip_Data$time <- as.POSIXct(All_Trip_Data$started_at, format = "%Y-%m-%d
%H:%M:%S")
All_Trip_Data$time <-format(All_Trip_Data$time, format = "%H:%M")
```

Add ‘time2’ column after extract time from the ‘ended_at’ column

```
All_Trip_Data$time2 <- as.POSIXlt(All_Trip_Data$ended_at, format = "%Y-%m-%d
%H:%M:%S")
All_Trip_Data$time2 <- format(All_Trip_Data$time2, format = "%H:%M")
```

For calculation of ride length

```
All_Trip_Data$ride_length <- as.double(difftime(All_Trip_Data$ended_at,
All_Trip_Data$started_at, units = "mins"))
```

For confirmation of additional column

```
colnames(All_Trip_Data)

##  [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"      "date"               "month"
## [16] "day"                "year"               "day_of_week"
## [19] "time"               "time2"              "ride_length"

skim(All_Trip_Data)
```

Data summary

Name	All_Trip_Data
Number of rows	3489748
Number of columns	21
_____	
Column type frequency:	
character	15
Date	1
numeric	5
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ride_id	0	1.00	16	16	0	3489539	0
rideable_type	0	1.00	11	13	0	3	0
started_at	0	1.00	19	19	0	3040228	0
ended_at	0	1.00	19	19	0	3027775	0

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
start_station_name	0	1.00	0	53	122175	709	0
start_station_id	83583	0.98	0	35	39218	1260	0
end_station_name	0	1.00	0	53	143242	707	0
end_station_id	98104	0.97	0	35	45599	1260	0
member_casual	0	1.00	6	6	0	2	0
month	0	1.00	2	2	0	12	0
day	0	1.00	2	2	0	31	0
year	0	1.00	4	4	0	2	0
day_of_week	0	1.00	6	9	0	7	0
time	59	1.00	5	5	0	1440	0
time2	0	1.00	5	5	0	1440	0

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
date	0	1	2020-04-01	2021-03-31	2020-08-29	363

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
start_lat	0	1	41.90	0.04	41.64	41.88	41.90	41.93	42.08	▁▁▆▇▁
start_lng	0	1	-87.64	0.03	-87.87	-87.66	-87.64	-87.63	-87.52	▁▁▂▇▁
end_lat	4738	1	41.90	0.04	41.54	41.88	41.90	41.93	42.16	▁▁▇▃▁
end_lng	4738	1	-87.65	0.03	-88.07	-87.66	-87.64	-87.63	-87.44	▁▁▁▇▁
ride_length	0	1	11.99	410.05	-28800.00	0.00	0.00	0.00	59040.00	▁▇▁▁▁

Convert 'ride length from double to numeric

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

## [1] TRUE
```

New Data frame without records that have ride length <=zero minute or >1440 minutes

```
All_Trip_Data_2 <- All_Trip_Data[!(All_Trip_Data$ride_length <=0 |
All_Trip_Data$ride_length < 1440),]
```

Check the new data

```
dim(All_Trip_Data_2)

## [1] 30097    21

summary(All_Trip_Data_2)

##    ride_id           rideable_type       started_at          ended_at
##  Length:30097       Length:30097       Length:30097       Length:30097
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  start_station_name start_station_id   end_station_name   end_station_id
##  Length:30097       Length:30097       Length:30097       Length:30097
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    start_lat       start_lng         end_lat         end_lng
##  Min.   :41.68   Min.   :-87.87   Min.   :41.67   Min.   :-87.84
##  1st Qu.:41.88   1st Qu.:-87.65   1st Qu.:41.88   1st Qu.:-87.65
##  Median :41.89   Median :-87.63   Median :41.89   Median :-87.63
##  Mean   :41.89   Mean   :-87.64   Mean   :41.90   Mean   :-87.64
##  3rd Qu.:41.92   3rd Qu.:-87.62   3rd Qu.:41.92   3rd Qu.:-87.62
##  Max.   :42.06   Max.   :-87.53   Max.   :42.06   Max.   :-87.53
##                                   NA's   :1100    NA's   :1100
##  member_casual           date               month               day
##  Length:30097       Min.   :2020-04-01   Length:30097       Length:30097
##  Class :character   1st Qu.:2020-07-10   Class :character   Class
:character
##  Mode  :character   Median :2020-08-02   Mode  :character   Mode
:character
##                     Mean   :2020-08-18
##                     3rd Qu.:2020-09-12
##                     Max.   :2021-03-31
##
```

```
##      year           day_of_week            time               time2
##  Length:30097       Length:30097       Length:30097       Length:30097
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   ride_length
##  Min.   : 1440
##  1st Qu.: 1440
##  Median : 1440
##  Mean   : 1750
##  3rd Qu.: 1440
##  Max.   :59040
##

view(All_Trip_Data_2)
```

Remove Duplicate ride id

```
All_Trip_Data_3 <- All_Trip_Data_2[!duplicated(All_Trip_Data_2$ride_id),]
dim(All_Trip_Data_3)

## [1] 30097    21

view(All_Trip_Data_3)
```

Calculate the ride distance

```
All_Trip_Data_3$ride_distance <- distGeo(matrix(c(All_Trip_Data_3$start_lng,
All_Trip_Data_3$start_lat), ncol = 2), matrix (c(All_Trip_Data_3$end_lng,
All_Trip_Data_3$end_lat), ncol = 2))

view(All_Trip_Data_3)

summary(All_Trip_Data_3)

##    ride_id           rideable_type       started_at          ended_at
##  Length:30097       Length:30097       Length:30097       Length:30097
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  start_station_name start_station_id   end_station_name   end_station_id
##  Length:30097       Length:30097       Length:30097       Length:30097
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
```

```
##    start_lat       start_lng         end_lat         end_lng
##  Min.   :41.68   Min.   :-87.87   Min.   :41.67   Min.   :-87.84
##  1st Qu.:41.88   1st Qu.:-87.65   1st Qu.:41.88   1st Qu.:-87.65
##  Median :41.89   Median :-87.63   Median :41.89   Median :-87.63
##  Mean   :41.89   Mean   :-87.64   Mean   :41.90   Mean   :-87.64
##  3rd Qu.:41.92   3rd Qu.:-87.62   3rd Qu.:41.92   3rd Qu.:-87.62
##  Max.   :42.06   Max.   :-87.53   Max.   :42.06   Max.   :-87.53
##                                   NA's   :1100    NA's   :1100
##  member_casual          date               month               day
##  Length:30097       Min.   :2020-04-01   Length:30097       Length:30097
##  Class :character   1st Qu.:2020-07-10   Class :character   Class
:character
##  Mode  :character   Median :2020-08-02   Mode  :character   Mode
:character
##                     Mean   :2020-08-18
##                     3rd Qu.:2020-09-12
##                     Max.   :2021-03-31
##
##      year           day_of_week            time              time2
##  Length:30097       Length:30097       Length:30097       Length:30097
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   ride_length    ride_distance
##  Min.   : 1440   Min.   :    0.0
##  1st Qu.: 1440   1st Qu.:    0.0
##  Median : 1440   Median :  844.5
##  Mean   : 1750   Mean   : 1813.4
##  3rd Qu.: 1440   3rd Qu.: 2701.8
##  Max.   :59040   Max.   :25334.5
##                  NA's   :1100
```

Analyze

Ride distribution day of week

```
All_Trip_Data_3$day_of_week <- ordered(All_Trip_Data_3$day_of_week, levels =
c('Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday',
'Sunday'))

All_Trip_Data_3 %>%
  group_by(member_casual, day_of_week) %>%
  summarise(number_of_ride = n(), .groups = 'drop') %>%
  arrange(day_of_week)
```

```
## # A tibble: 14 × 3
##    member_casual day_of_week number_of_ride
##    <chr>         <ord>                <int>
##  1 casual        Monday                2458
##  2 member        Monday                 443
```

```
##  3 casual        Tuesday               2260
##  4 member        Tuesday                508
##  5 casual        Wednesday             2366
##  6 member        Wednesday              586
##  7 casual        Thursday              2851
##  8 member        Thursday               669
##  9 casual        Friday                5523
## 10 member        Friday                1209
## 11 casual        Saturday              6093
## 12 member        Saturday              1172
## 13 casual        Sunday                3524
## 14 member        Sunday                 435
```

Ride distribution by month

```
All_Trip_Data_3$month <- ordered(All_Trip_Data_3$month, level = c('05', '06',
'07', '08', '09', '10', '11', '12', '01', '02', '03'))
```

Ride distribution by day of the month

```
All_Trip_Data_3 %>%
  group_by(member_casual, day) %>%
  summarise(number_of_ride = n(),
            .groups = "drop")
```

```
## # A tibble: 62 × 3
##    member_casual day   number_of_ride
##    <chr>         <chr>          <int>
##  1 casual        01               452
##  2 casual        02               664
##  3 casual        03               642
##  4 casual        04               871
##  5 casual        05               898
##  6 casual        06               887
##  7 casual        07               970
##  8 casual        08               959
##  9 casual        09               726
## 10 casual        10               651
## # ℹ 52 more rows
```

whether ride_length can be different depends on rider type.

```
aggregate(All_Trip_Data_3$ride_length ~ All_Trip_Data_3$member_casual +
All_Trip_Data_3$day_of_week, FUN = mean)
```

```
##    All_Trip_Data_3$member_casual All_Trip_Data_3$day_of_week
## 1                         casual                      Monday
## 2                         member                      Monday
## 3                         casual                     Tuesday
## 4                         member                     Tuesday
## 5                         casual                   Wednesday
## 6                         member                   Wednesday
## 7                         casual                    Thursday
## 8                         member                    Thursday
## 9                         casual                      Friday
```

```
## 10                         member                       Friday
## 11                         casual                     Saturday
## 12                         member                     Saturday
## 13                         casual                       Sunday
## 14                         member                       Sunday
##    All_Trip_Data_3$ride_length
## 1                     1811.961
## 2                     1596.027
## 3                     1744.619
## 4                     1536.378
## 5                     1808.216
## 6                     1590.000
## 7                     1875.363
## 8                     1521.794
## 9                     1659.783
## 10                    1467.395
## 11                    1753.826
## 12                    1532.150
## 13                    1988.513
## 14                    1798.069
```

```
All_Trip_Data_3 %>%
  group_by(member_casual, month) %>%
  summarise(average_ride_length = mean(ride_length), groups = 'drop') %>%
  arrange(month)
```

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

```
## # A tibble: 24 × 4
## # Groups:   member_casual [2]
##    member_casual month average_ride_length groups
##    <chr>         <ord>               <dbl> <chr>
##  1 casual        05                  1681. drop
##  2 member        05                  1516. drop
##  3 casual        06                  1780. drop
##  4 member        06                  1554. drop
##  5 casual        07                  1818. drop
##  6 member        07                  1495. drop
##  7 casual        08                  1735. drop
##  8 member        08                  1596. drop
##  9 casual        09                  1742. drop
## 10 member        09                  1570. drop
## # ℹ 14 more rows
```

The number of bike type users

```
All_Trip_Data_3 %>%
  group_by(rideable_type) %>%
  summarise(number_of_ride = n())
```

```
## # A tibble: 3 × 2
##   rideable_type number_of_ride
##   <chr>                 <int>
## 1 classic_bike           1181
## 2 docked_bike           26081
## 3 electric_bike          2835
```

Cycle type usage by user type

```
All_Trip_Data_3 %>%
  group_by(member_casual, rideable_type) %>%
  summarise(number_of_ride = n(),
            .groups = "drop") %>%
  arrange(rideable_type)
```

```
## # A tibble: 6 × 3
##   member_casual rideable_type number_of_ride
##   <chr>         <chr>                 <int>
## 1 casual        classic_bike            600
## 2 member        classic_bike            581
## 3 casual        docked_bike           22606
## 4 member        docked_bike            3475
## 5 casual        electric_bike          1869
## 6 member        electric_bike           966
```

Cycle type and average ride type

```
All_Trip_Data_3 %>%
  group_by(rideable_type) %>%
  summarise(mean = mean(ride_length),
            .groups = "drop") %>%
  arrange(rideable_type)
```

```
## # A tibble: 3 × 2
##   rideable_type  mean
##   <chr>         <dbl>
## 1 classic_bike  1447.
## 2 docked_bike   1797.
## 3 electric_bike 1440.
```

Name of Unique Station

```
stations <- All_Trip_Data_3 %>%
  gather(key, station_name, start_station_name, end_station_name) %>%
  distinct(station_name)

print(paste("Number of station", nrow(stations)))
```

```
## [1] "Number of station 667"
```

Most popular station

```
popular_station <- All_Trip_Data_3 %>%
  gather(key, station_name, start_station_name, end_station_name) %>%
  group_by(station_name) %>%
  summarise(number_trip = n()/2) %>%
```

```
  arrange(desc(number_trip))

head(popular_station, 10)
```

```
## # A tibble: 10 × 2
##    station_name                    number_trip
##    <chr>                                 <dbl>
##  1 ""                                    1284
##  2 "Millennium Park"                      718.
##  3 "Michigan Ave & Lake St"               548
##  4 "Michigan Ave & 8th St"                470.
##  5 "Michigan Ave & Washington St"         446.
##  6 "Wabash Ave & 9th St"                  414
##  7 "Fairbanks Ct & Grand Ave"             394.
##  8 "Buckingham Fountain"                  380
##  9 "Wabash Ave & Roosevelt Rd"            379
## 10 "Wabash Ave & Grand Ave"               361
```

Share

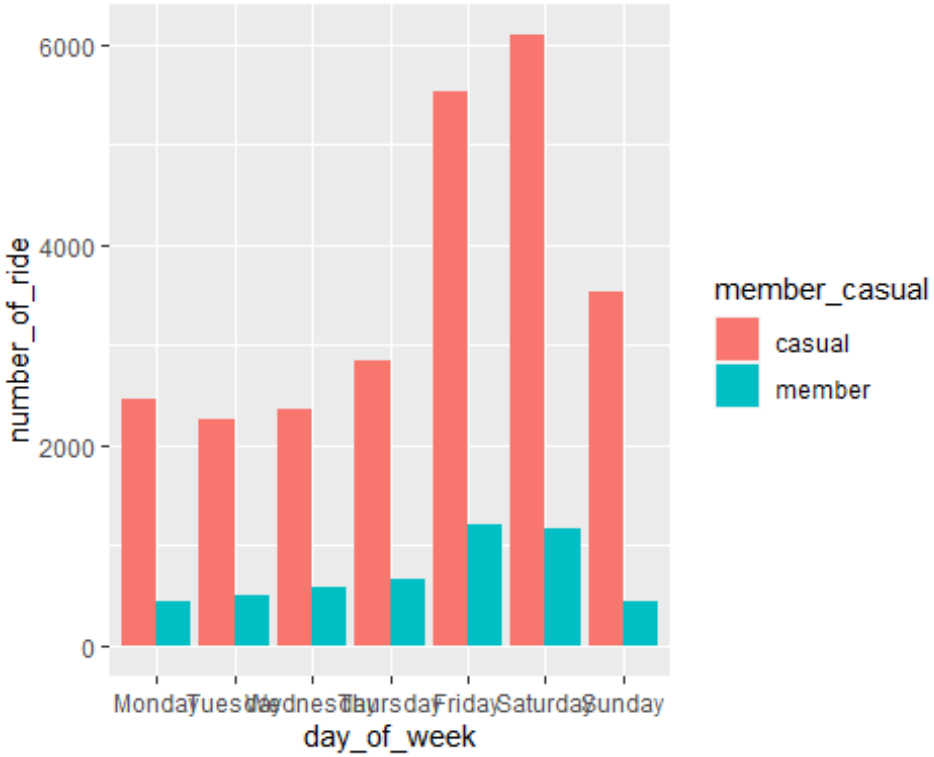
Ride distribution day of week

```
All_Trip_Data_3 %>%
  group_by(member_casual, day_of_week) %>%
  summarise(number_of_ride = n(),
            .group = "drop") %>%

ggplot(aes(x = day_of_week,y =  number_of_ride, fill = member_casual)) +
  geom_col(position = "dodge", start = "identity")
```

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

## Warning in geom_col(position = "dodge", start = "identity"): Ignoring
unknown
## parameters: `start`
```




```
  labs (title = "The Number of rides by Weekday", x = "Weekday", y = "Number
of rides")
```

```
## $x
## [1] "Weekday"
##
## $y
## [1] "Number of rides"
##
## $title
## [1] "The Number of rides by Weekday"
##
## attr(,"class")
## [1] "labels"
```

Ride distribution by month

```
All_Trip_Data_3 %>%
  group_by(member_casual, month) %>%
  summarise(number_of_ride = n(),
            .groups = 'drop')
```

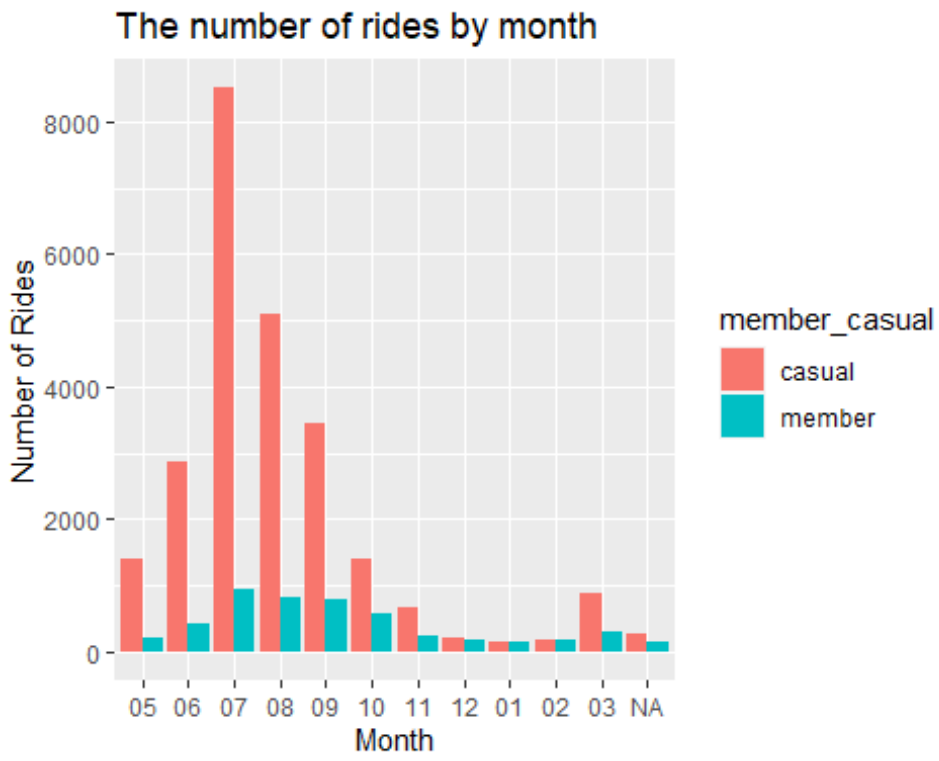
```
## # A tibble: 24 × 3
##    member_casual month number_of_ride
##    <chr>         <ord>          <int>
##  1 casual        05              1402
##  2 casual        06              2880
##  3 casual        07              8508
##  4 casual        08              5102
```

```
##  5 casual        09              3441
##  6 casual        10              1396
##  7 casual        11               670
##  8 casual        12               202
##  9 casual        01               145
## 10 casual        02               169
## # ℹ 14 more rows
```

```
All_Trip_Data_3 %>%
  group_by(member_casual, month) %>%
  summarise(number_of_ride = n(),
            .group = "drop") %>%

ggplot(aes(month, number_of_ride, fill = member_casual)) +
  geom_col(position = "dodge") +
  labs(title = "The number of rides by month", x = "Month", y = "Number of
Rides")
```

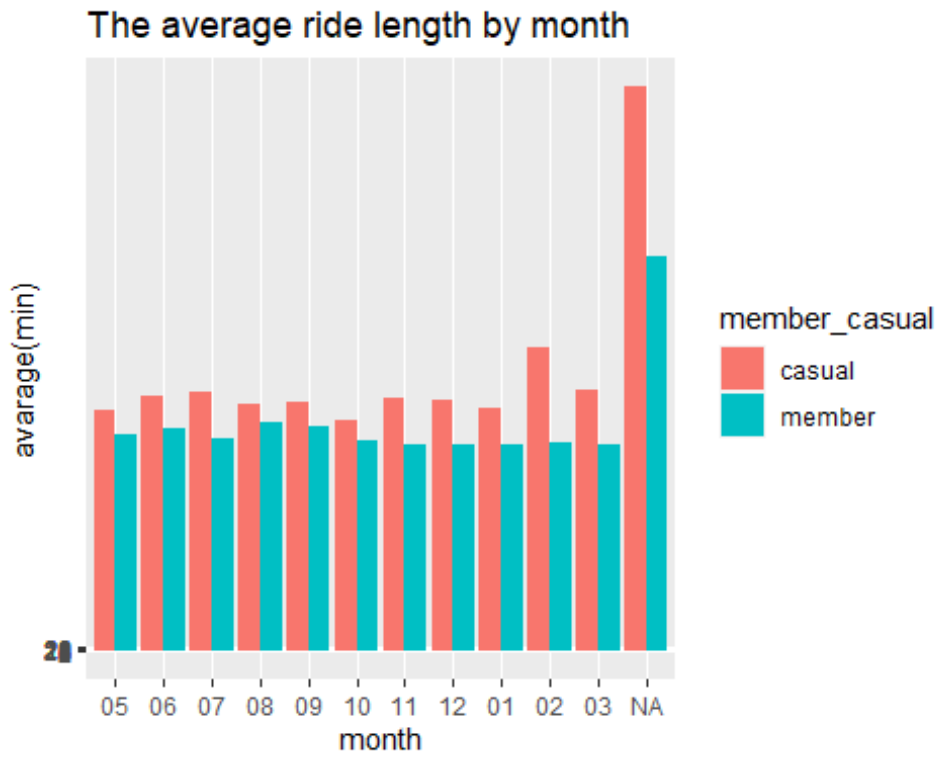
```
## `summarise()` has grouped output by 'member_casual'. You can override
using the
## `.groups` argument.
```



whether ride_length can be different depends on rider type.

```
All_Trip_Data_3 %>%
  group_by(member_casual, month) %>%
  summarise(mean = mean(ride_length),
            .groups = "drop") %>%
  ggplot(aes(month, mean, fill = member_casual)) +
```

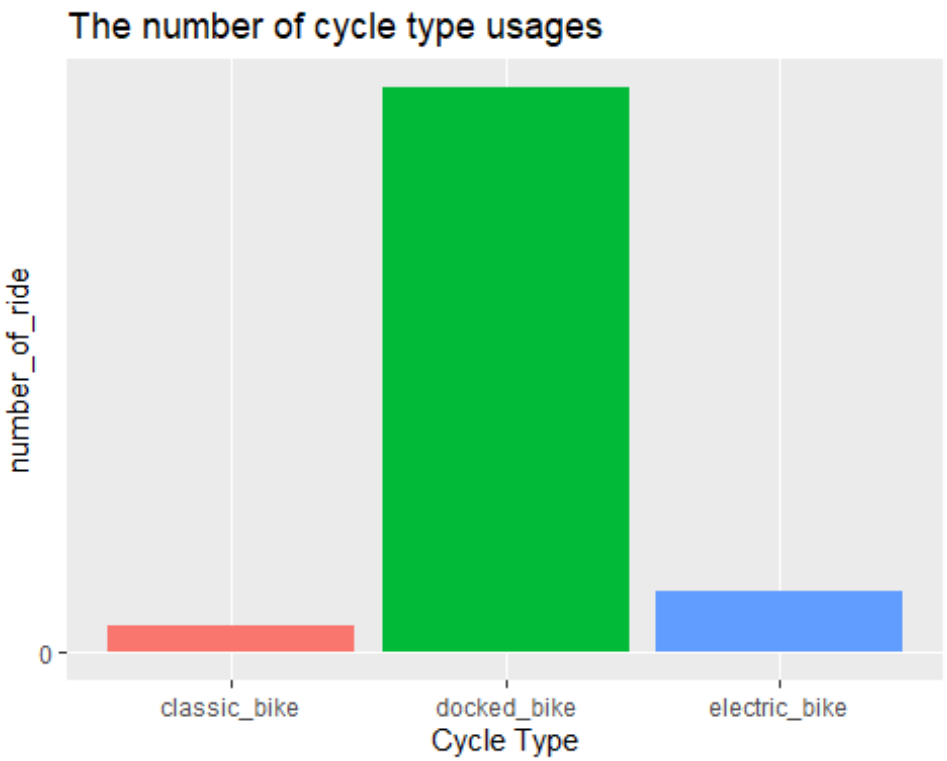
```
  geom_col(position = "dodge") +
  scale_y_continuous(breaks = seq(0, 28, 2)) +
  labs(title = "The average ride length by month", x = "month", y =
"avarage(min)")
```



The number of bike type users

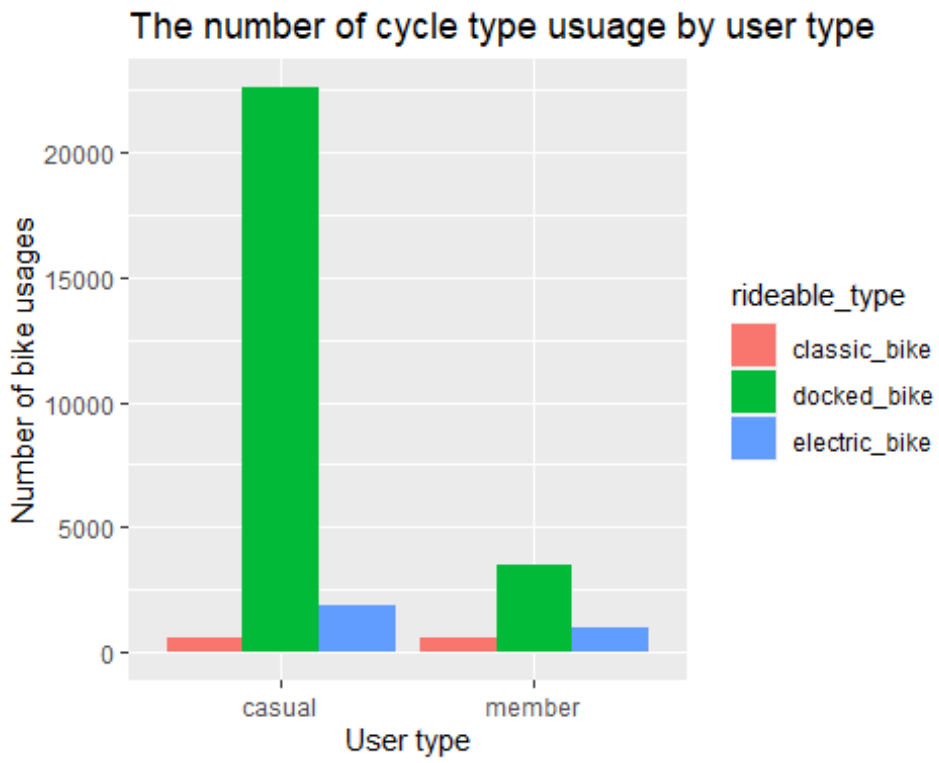
```
All_Trip_Data_3 %>%
  group_by(rideable_type) %>%
  summarise(number_of_ride = n ()) %>%
  ggplot(aes(rideable_type, number_of_ride, fill = rideable_type)) +
  geom_col(position = "dodge", show.legend = FALSE) +
  scale_y_continuous(labels = scales::label_number_si(),
                     breaks = seq(0, 3000000, 1000000)) +
  labs(title = "The number of cycle type usages", x = "Cycle Type", Y =
"Number of cycle usages")
```

```
## Warning: `label_number_si()` was deprecated in scales 1.2.0.
## ℹ Please use the `scale_cut` argument of `label_number()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



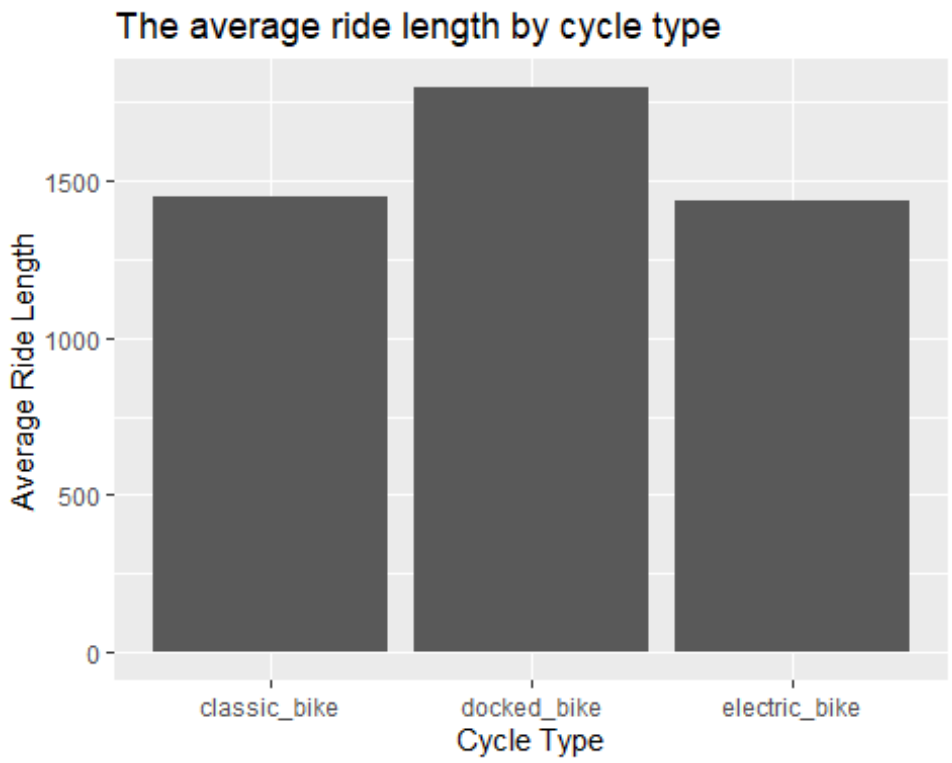
Cycle type usage by user type

```
All_Trip_Data_3 %>%
  group_by(member_casual, rideable_type) %>%
  summarise(number_of_ride = n(),
            .groups = "drop") %>%
  ggplot(aes(member_casual, number_of_ride, fill = rideable_type)) +
  geom_col(position = "dodge") +
  labs(title = "The number of cycle type usuage by user type", x = "User
type", y = "Number of bike usages")
```



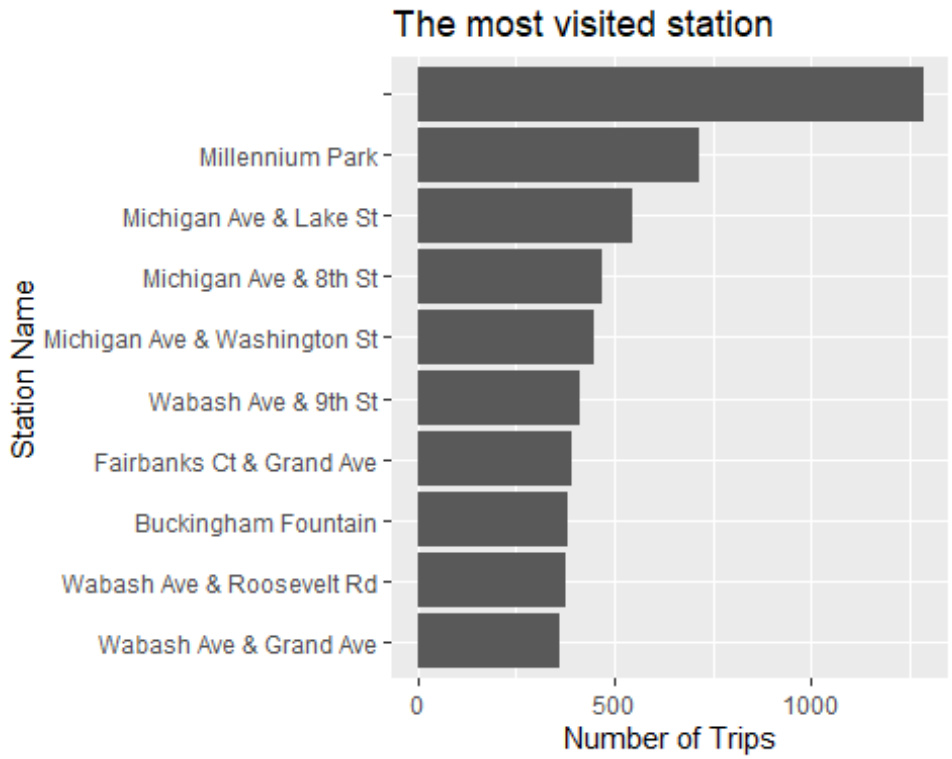
Cycle type and average ride type

```
All_Trip_Data_3 %>%
  group_by(rideable_type) %>%
  summarise(mean = mean(ride_length),
            .groups = "drop") %>%
  ggplot(aes(rideable_type, mean)) +
  geom_col(position = "dodge") +
  labs(title = "The average ride length by cycle type", y = "Average Ride
Length", x = "Cycle Type")
```



Most popular station

```
popular_station %>%
  slice(1:10) %>%
  ggplot(aes(number_trip, reorder(station_name, number_trip))) +
  geom_col() +
  labs(title = "The most visited station", x = "Number of Trips", y =
"Station Name")
```


The most visited station
Station Name
-
Millennium Park
Michigan Ave & Lake St
Michigan Ave & 8th St
Michigan Ave & Washington St
Wabash Ave & 9th St
Fairbanks Ct & Grand Ave
Buckingham Fountain
Wabash Ave & Roosevelt Rd
Wabash Ave & Grand Ave
0
500
1000
Number of Trips