Analyst: Quynh Duong

Project title: Bike Share Customer Analysis/A new marketing strategy

Table of contents:

## I. Introduction:

Cyclistic is a bike-share company in Chicago that features more than 5,800 bicycles and 600 docking stations. 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. Rather than creating a marketing campaign that targets all-new customers, we believe there is a very good chance to convert casual riders into members.

## II. Analysis

1. Business Task:

- A clear goal: Marketing strategy recommendations aimed at converting casual riders into annual members.
- My task: Identify how casual riders and annual members use Cyclistic bikes differently

❖ Key Stakeholders:

❖ Objectives:

2. Preparing data

- The data has been made available by Motivate International Inc. under this license
- One-year trip data from September, 2021 to August, 2022 including a total of 12 CSV files were collected and downloaded from Divvy_trip_data, and they are stored in a folder named Bike_share_trip_2021-2022.
- The datasets collected include 13 attributes which help us answer the business questions: **ride_id (categorical)**, **rideable_type (categorical)**, **started_at (datetime)**, **ended_at (datetime)**, **start_station_name (categorical)**, **start_station_id (categorical)**, **end_station_name (categorical)**, **end_station_id (categorical)**, **start_lat (numeric)**, **start_lng (numeric)**, **end_lat (numeric)**, **end_lng (numeric)**, **member_casual (categorical)**.

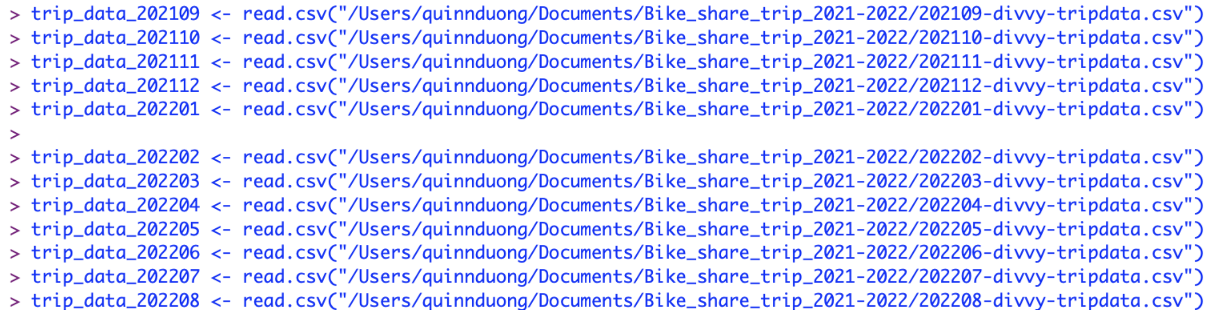
3. Processing data

I use R to clean data and document my cleaning process in RStudio because data cleaning in spreadsheets can be time-consuming and slow compared to R or SQL.

- Installing and loading libraries: Those libraries below will be used during processing and analyzing.

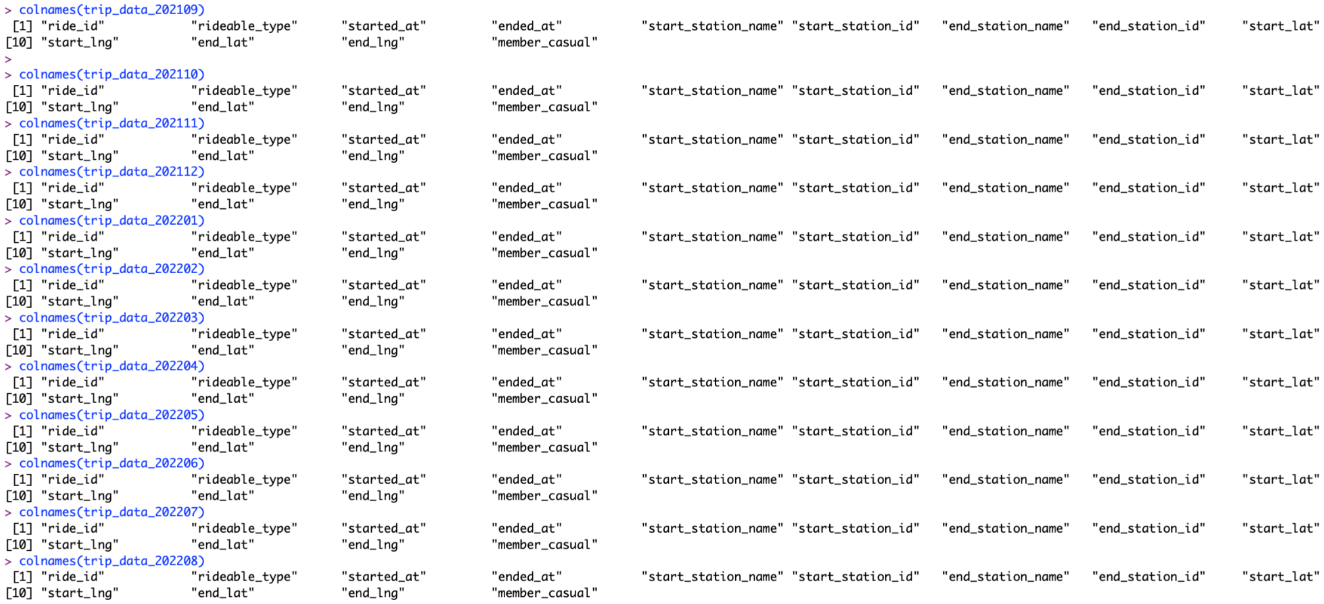
```
>library('tidyverse')
── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
✔ ggplot2 3.3.6      ✔ purrr   0.3.4
✔ tibble  3.1.8      ✔ dplyr   1.0.10
✔ tidyr   1.2.1      ✔ stringr 1.4.1
✔ readr   2.1.2      ✔ forcats 0.5.2
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
> library('lubridate')

Attaching package: ‘lubridate’

The following objects are masked from ‘package:base’:

    date, intersect, setdiff, union
> library('data.table')
data.table 1.14.2 using 8 threads (see ?getDTthreads).  Latest news: r-datatable.com

Attaching package: ‘data.table’

The following objects are masked from ‘package:lubridate’:

    hour, isoweek, mday, minute, month, quarter, second, wday, week,
    yday, year

The following objects are masked from ‘package:dplyr’:

    between, first, last

The following object is masked from ‘package:purrr’:

    transpose
> library("janitor")

Attaching package: ‘janitor’

The following objects are masked from ‘package:stats’:

    chisq.test, fisher.test
```

- Importing all the 12 .csv files into 12 data frames:



```
> trip_data_202109 <- read.csv("/Users/quinnduong/Documents/Bike_share_trip_2021-2022/202109-divvy-tripdata.csv")
> trip_data_202110 <- read.csv("/Users/quinnduong/Documents/Bike_share_trip_2021-2022/202110-divvy-tripdata.csv")
> trip_data_202111 <- read.csv("/Users/quinnduong/Documents/Bike_share_trip_2021-2022/202111-divvy-tripdata.csv")
> trip_data_202112 <- read.csv("/Users/quinnduong/Documents/Bike_share_trip_2021-2022/202112-divvy-tripdata.csv")
> trip_data_202201 <- read.csv("/Users/quinnduong/Documents/Bike_share_trip_2021-2022/202201-divvy-tripdata.csv")
>
> trip_data_202202 <- read.csv("/Users/quinnduong/Documents/Bike_share_trip_2021-2022/202202-divvy-tripdata.csv")
> trip_data_202203 <- read.csv("/Users/quinnduong/Documents/Bike_share_trip_2021-2022/202203-divvy-tripdata.csv")
> trip_data_202204 <- read.csv("/Users/quinnduong/Documents/Bike_share_trip_2021-2022/202204-divvy-tripdata.csv")
> trip_data_202205 <- read.csv("/Users/quinnduong/Documents/Bike_share_trip_2021-2022/202205-divvy-tripdata.csv")
> trip_data_202206 <- read.csv("/Users/quinnduong/Documents/Bike_share_trip_2021-2022/202206-divvy-tripdata.csv")
> trip_data_202207 <- read.csv("/Users/quinnduong/Documents/Bike_share_trip_2021-2022/202207-divvy-tripdata.csv")
> trip_data_202208 <- read.csv("/Users/quinnduong/Documents/Bike_share_trip_2021-2022/202208-divvy-tripdata.csv")
```

- Check column names of each dataset for consistency



```
> colnames(trip_data_202109)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_station_name" "start_station_id"   "end_station_name"   "end_station_id"     "start_lat"
[10] "start_lng"          "end_lat"            "end_lng"            "member_casual"
>
> colnames(trip_data_202110)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_station_name" "start_station_id"   "end_station_name"   "end_station_id"     "start_lat"
[10] "start_lng"          "end_lat"            "end_lng"            "member_casual"
> colnames(trip_data_202111)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_station_name" "start_station_id"   "end_station_name"   "end_station_id"     "start_lat"
[10] "start_lng"          "end_lat"            "end_lng"            "member_casual"
> colnames(trip_data_202112)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_station_name" "start_station_id"   "end_station_name"   "end_station_id"     "start_lat"
[10] "start_lng"          "end_lat"            "end_lng"            "member_casual"
> colnames(trip_data_202201)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_station_name" "start_station_id"   "end_station_name"   "end_station_id"     "start_lat"
[10] "start_lng"          "end_lat"            "end_lng"            "member_casual"
> colnames(trip_data_202202)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_station_name" "start_station_id"   "end_station_name"   "end_station_id"     "start_lat"
[10] "start_lng"          "end_lat"            "end_lng"            "member_casual"
> colnames(trip_data_202203)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_station_name" "start_station_id"   "end_station_name"   "end_station_id"     "start_lat"
[10] "start_lng"          "end_lat"            "end_lng"            "member_casual"
> colnames(trip_data_202204)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_station_name" "start_station_id"   "end_station_name"   "end_station_id"     "start_lat"
[10] "start_lng"          "end_lat"            "end_lng"            "member_casual"
> colnames(trip_data_202205)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_station_name" "start_station_id"   "end_station_name"   "end_station_id"     "start_lat"
[10] "start_lng"          "end_lat"            "end_lng"            "member_casual"
> colnames(trip_data_202206)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_station_name" "start_station_id"   "end_station_name"   "end_station_id"     "start_lat"
[10] "start_lng"          "end_lat"            "end_lng"            "member_casual"
> colnames(trip_data_202207)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_station_name" "start_station_id"   "end_station_name"   "end_station_id"     "start_lat"
[10] "start_lng"          "end_lat"            "end_lng"            "member_casual"
> colnames(trip_data_202208)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_station_name" "start_station_id"   "end_station_name"   "end_station_id"     "start_lat"
[10] "start_lng"          "end_lat"            "end_lng"            "member_casual"
```

4. Analyzing data and data visualization

I am also choosing R for data analyzing and data visualization because it's easier for large and complex datasets, and we also can make data visualizations in the same platform.

## III. Conclusions.