## **Introduction**

This is a capstone project as part of ***google data analytics professional*** certification. For the analysis I have downloaded the .csv files of cycling data for 12 months from June 2021 to June 2022 (current time period) from the following link mentioned in the capstone project.

[Index of bucket "divvy-tripdata"](https://divvy-tripdata.s3.amazonaws.com/index.html)

Here I have used Excel for doing my primary analysis and then due to large data sets for the analysis finally decided to go ahead with R and R studio IDE .

For the analysis following steps are mainly followed to make the data driven decisions from the huge data .

* ASK
* PREPARE
* PROCESS
* ANALYZE
* SHARE
* ACT

### **Scenario**

You are a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company’s future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

# ASK

In the ask phase as an analyst I have to ask protentional questions which derive the decisions strategies. As per the current scenario explained above followings are the 3 questions, I need to derive the answer from available data.

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.

So from here I can understand that the clear business task is nothing but I have to identify how does annual members and casual riders use Cyclistic bikes differently and build a best marketing strategy to change casual riders to annual members for increasing the revenue .

## Prepare

For the prepare phase of analysis I need to download the previous 12 months of Cyclistic trip data [here](https://divvy-tripdata.s3.amazonaws.com/index.html). (Note: The datasets have a different name because Cyclistic is a fictional company. For the purposes of this case study, the datasets are appropriate and will enable you to answer the business questions. The data has been made available by Motivate International Inc. under this [license](https://ride.divvybikes.com/data-license-agreement).)

### Key tasks

* Download the 12-month data (June 2021 to june2022) and kept in a system folder.
* Identify how its organized like column names.

All trip data is in comma-delimited (.CSV) format. Column names "ride\_id", "rideable\_type", "started\_at", "ended\_at", "start\_station\_name", "start\_station\_id", "end\_station\_name", "end\_station\_id", "start\_lat", "start\_lng", "end\_lat", "end\_lng", "member\_casual" (Total 13 column)

* Filter and sort the data using excel
* Determine the credibility of the data.

For the purposes of this case study, the datasets are appropriate and it will enable me to answer the business questions. But due data-privacy I cannot use rider's personally identification information, and this will prevent me from determining if a single user/rider taken several rides. All ride ids are unique in this data-set.

#### Deliverable

* Description of data source- Main source provided by the **Cyclistic** **company**.

Collect and merge the 12 months data in to single data frame for the analysis phase. For merging phase I am using R studio and reading and binding .csv files.