# Project 1: US Bikeshare

## Question/need:

* Looking at the data, we need to know which trips are the most popular, to increase the number of the Bikes in start stations to minimize the lack of bikes in that stations.
* We need to know when the most common month, day and hour for all trips are,

to take the opportunity in the free traffic time for performing maintenance on the bikes.

* For the purpose of advertisement, we need to know the average age for bike’s riders and what is their gender.
* We need to know which trips are the longest trips and from which start stations to provide more bikes into those stations by taking in consideration that the bike whetherit is returned into the same station or returned on other station.

## Data Description:

#### The data used in this project represent a bike share system provider for many major cities in the United States, to uncover bike share usage patterns. The data provided is only for New York city. The data used in this project in intended for Citibike company and advertisement company. The goal of this project is to help these company to decide for their work process based on the project result. For example, an advertisement company can determine their target customer based on the average age and gender of the bike’s rider.

#### Finally, We have multiple features in this date as the following:

|  |  |
| --- | --- |
| Feature | Example |
| **Start Time** | **(e.g., 2017-01-01 00:07:57)** |
| **End Time** | **(e.g., 2017-01-01 00:20:53)** |
| **Trip Duration** | **(in seconds - e.g., 776)** |
| **Start Station** | **(e.g., Broadway & Barry Ave)** |
| **End Station** | **(e.g., Sedgwick St & North Ave)** |
| **User Type** | **(Subscriber or Customer)** |

## Tools:

#### I will use python to tackle then questions mentioned above with the following libraries:

#### Panda: For data Processing and data cleaning.

#### Matplotlib: For visualization, the relation between features.

## MVP Goal:

By looking on below figure, we can see that most of our customers age are in the range of 20 to 40. And we have a good bunch of customers in the range of 40 to 60. Hence , we need to do more investigation on the data to know which stations are the most crowded to aim for them.

Chart, histogram

Description automatically generated