# (Ford GoBike System Dataset Exploration)

## by (NANCHI GUMUT)

## Dataset Investigation Overview

> Provide basic information about your dataset in this section. If you selected your own dataset, make sure you note the source of your data and summarize any data wrangling steps that you performed before you started your exploration.

(Dataset of Ford Gobike System 2019 Exploration)

The Dataset in this README was provided by Udacity and it is Ford GoBike System Dataset.

Ford GoBike Sysytem is a company that shares bike to the public covering the greater part of San Francisco Bay area. This dataset explores information containing biketrips of individauls from Ford GoBike system. This dataset has been uploaded to get some data wrangling in order to make it tidy for analysis,exploration and a good viualisation.

This data set includes information about individual(Customers/Subscribers) rides made in a bike-sharing system covering the greater San Francisco Bay area .The dataset of this company can be viewed and assessed in two parts.

The first part is the exploratory analysis of Ford GoBike System Data. Where I used NumPy and Pandas along with visualization libraries with other necessary libraries to explore the dataset’s variables and understand the dataset structure and relationships. The analysis in this part should be structured, going from simple univariate relationships to bivariate relationships up through multivariate relationships.

The second part is my main findings from the exploration part and convey them to others through an explanatory analysis. To this end, I create a slide deck that leverages and polished the explanatory visualizations to communicate my results on the relationships.

The Ford GoBike System Dataset has 183412 rows and 16 columns

There are 11 numeric variables (duration\_sec, start\_time, end\_time, start\_station\_id, start\_station\_latitude, start\_station\_longitude, end\_station\_id, end\_station\_latitude, end\_station\_longitude, bike\_id, member\_birth\_year) and 5 categoric variables (start\_station\_name, end\_station\_name, user\_type, member\_gender,bike\_share\_for\_all\_trip)

## Summary of Findings

> Summarize all of your findings from your exploration here, whether you plan on bringing them into your explanatory presentation or not.

In Ford GoBike Sysytem Dataset the number of Distribution of Gender, Period of the day is very much dependable on the age of Riders.In investigating the factors affecting duration of the trips, I have plotted 3 distributions (gender, period and age) to know their tally against the entire dataset

Also Male gender feature most prominently in the dataset, most rides start in the morning and evenings and large block of the riders are between 23 and 38 years old and the count steeply decreases as bikers age increase beyond age 40.

There are a lot more subscriber usage than customers overall counts. Subscriber features prominently the dataset with a count of 163544 and Customer has the least count of 19868.

## Key Insights for Presentation

> Select one or two main threads from your exploration to polish up for your presentation. Note any changes in design from your exploration step here.