**Project:** (Ford GoBike Data Exploration)

**by Abdullah Abusabba**

## Dataset

I have used Ford GoBike Dataset, which are a dataset contains 2207546 bike with 16 feature most of the variables are numeric, and some of them categorical variables and time series.

. This data set includes information about individual rides made in a bike-sharing system for 2019 covering the greater San Francisco Bay area.

# The Data

- **Trip Duration**
- **Start Time and Date**
- **End Time and Date**
- **Start Station ID**
- **Start Station Name**
- **Start Station Latitude**
- **Start Station Longitude**
- **End Station ID**
- **End Station Name**
- **End Station Latitude**
- **End Station Longitude**
- **Bike ID**
- **Distance_km**
- **day**
- **Bike Share for All Trip**

> which tracks members who are enrolled in the Bike Share for All program for low-income residents.

- **User Type**

> (Subscriber or Customer – "Subscriber" = Member or "Customer" = Casual)

**Summary of Findings**

- **Univariate Exploration:**

1. the majority in the dataset are not enrolled in "Bike Share for All Program" and the user type majority are subscriber.

2. The majority of bike ride are in the weekdays mostly Thursday 17.3% whereas on Saturday 9% and Sunday 8.1% as we can see a major drop of bike ride users and in terms of duration, they fall between 3-7 min. Most distance are between 0.6-1.5 km.

**Bivariate Exploration:**

1. The higher the station id name the more consistence the bikers will ride their bikes in terms of distance from slightly below 1 km up to 3.5 km.

2. We noticed that the distance of bike users is higher if the user started in San Francisco Building (Harry Brings Plazza).

**Multivariate Exploration:**

1. Trip Duration in different months taking into account the user type, now we can see the Customer has much higher trip duration in average.

2. There is not a single user who's a customer and enrolled in bike for all program in 2019.We noticed that the subscribers have the greatest number of trips on October and on the other side where the customers have the

**Key Insights for Presentation :**

Many visualizations on (Ford GoBike System Data characteristics will be created to study their relationship on duration_min The main characteristics are duration_min,distance_km,day,bike_share_for_all_trip,start_time, start_station_name, start_station_name , user_type ,start_station_latitude , start_station_longitude , end_station_latitude , end_station_longitude  to help my go through the investigation part in our dataset