# Communicate Data Findings Project:

## Ford GoBike System Data Set

- **Investigation & dataset overview**

  This data set includes information about individual rides made in a bike-sharing system covering the greater San Francisco Bay area.
  This analysis aims to discover the variables which affect the trip duration of the users by using univariate, bivariate and multivariate exploration.
  There are 183412 ride data in the dataset with 16.
  I'm most interested in figuring out what features affect the duration of the trip.

  - Does the user type (Subscriber or Customer) affect the duration?
  - Does the age of user affect the duration?
  - Does the gender of user affect the duration?

- **Summary of Exploratory Findings:**
  - The duration distribution shows long tailed distribution so I used a log transform and convert the duration to minutes.
  - Under the transformation, the data looked unimodal, with one peak around 10 minutes.
  - Almost 74.6% of the users are males and 23.3% are females.
  - 90.5 of the users are subscribers and only 9.5% are customers.
  - From the member birth year column, I create a new column contains the age of each customer. The age distribution is right skewed with peak value between 30-35 years.

- **Key Insights:**

After performing univariate, bivariate and multivariate exploration, the analysis shows the following:

  - Almost 74.6% of the users are males, 23.3% are females and 2.1% other. There is no significant difference in trip duration between the different genders.
  - 90.5 of the users are subscribers and only 9.5% are customers.
    The customers usually have higher trip duration than the subscribers across different genders. So, the user type affect the trip duration.
  - There is significant decrease in the trip duration with older ages for all genders. The age strongly affect the trip duration.