

# 2017-Ford GoBike Data Analysis

by Vidhyut Prajapati

## Dataset

---

Each trip is anonymized and includes:

- Trip Duration (seconds)
- 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
- User Type (Subscriber or Customer – “Subscriber” = Member or “Customer” = Casual)

There are 155686 records and 13 columns

## Summary of Findings

---

- Trip Duration statistics revealed that 75% of the values were below 1000, so outliers were removed using Z-Score as reference.
- Going further with the trimmed dataset, I noticed that Customers on average spent more time riding than subscribers.
- One of the most interesting observations was the relationship between user type and start hour of the trip. While Subscribers rented it in the mornings and evenings, Customers rented it sporadically, peaking slightly in the afternoon, suggesting that Subscribers used the bikes as a mode of transport for work.

## Key Insights for Presentation

---

- Subscribers rode faster than customers.
- Trips took more time during the 9 AM to 6 PM timeline, irrespective of start and end points.