

# ReadMe

## Dataset

I chose the Ford Go Bike System Data. The data consisted of 239111 bike ride records. 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)
- Member Year of Birth
- Member Gender

In order to have the most current view, I used the data from April 2019 data titled ‘201904-fordgobike-tripdata.csv’.

## Summary of Findings

- Bike stations are in Northern California (San Jose area).
- There are 4520 unique bikes.
- More trips are taken by subscribers than by customers. (203196 by subscribers, 35914 by customers)
- More trips are taken by men than by women and other. (55498 by females, 168139 males, and 4274 by other)
- One-time customers take longer trips than regular subscribers. (Customers median trip is 879 sec, subscribers median trip is 521 sec)
- Females trips are longer than male or other trips. Female median is 623 sec, male median is 532 sec, and other median is 586)
- None of the continuous variables ('start\_station\_latitude', 'start\_station\_longitude', 'end\_station\_id', 'end\_station\_latitude', 'end\_station\_longitude', 'bike\_id', 'member\_birth\_year', 'intercept') significantly explain response variable 'duration\_sec'
- Older people don't take long trips, there is more variation in younger generation.

- There are more young riders than older riders.

The highlighted findings are the key insights included in the explanatory analysis. The numbering links to the corresponding slide in the explanatory analysis.

### **Key Insights**

In my presentation I focused on bike station locations, user demographics, and user trip durations. One main insight was that the bike stations are in the bay area in an interesting triangular shape.

Another insight is that more trips are taken by subscribers than by customers, which makes sense because a subscriber probably has a deal that incentivizes use. More trips are taken by men than my women.

The last insight is on trip duration. Customer trips are longer than subscriber trips which is possible because customers may want to enjoy the rare experience since they do not ride as often. Female trips are longer than male trips. Maybe women do not pedal as quickly as men do, but who knows as it is 2019. ;)

### **Resources**

1. Ford GoBike Website <https://www.fordgobike.com/system-data>
2. GoogleMaps <https://www.google.com/maps>
3. Toward Data Science <https://towardsdatascience.com/>
4. Stack Overflow <https://stackoverflow.com/>
5. Pandas Documentation <https://pandas.pydata.org/pandas-docs/stable/>