## Visual Analysis of Ford Gobike System Dataset

## By: Kazeem Adeyanju

The dataset for this project is the Ford Gobike system data. Ford Gobike is a model for equitable bike share access in the US and an exploration has been carried out on the dataset. The dataset contains the following columns before exploration:

- 1. duration\_sec
- 2. start time
- 3. end time
- 4. start\_station\_id
- 5. start\_station\_name
- 6. start\_station\_latitude
- 7. start\_station\_longitude
- 8. end\_station\_id
- 9. end\_station\_name
- 10. end\_station\_latitude
- 11. end\_station\_longitude
- 12. bike\_id
- 13. user\_type
- 14. member\_birth\_year
- 15. member\_gender
- 16. bike\_share\_for\_all\_trip

## **Summary**

The datatype for a number of the variables were changed to the appropriate datatype e.g the time variables were changed from string to datetime. Our exploration required adding additional time variables like start\_weekday, end\_weekday, start\_year\_month, end\_year\_month etc. This was done for deeper exploration of our dataset. The visual exploration followed the "Question-Visualisation-Observation" framework and the following are some of the findings as identified from the visual exploration of our dataset:

- 1. The dataset is a bike trip records for February 2019
- 2. There are more trips record for males than females.
- 3. There are more trip records for those who subscribe to the trip service than customers
- 4. The customers embarked on longer trips compared to those who subscribed to the service.
- 5. People within the age of 30 to 35 have the highest trips record.
- 6. There is a noticeable sharp decline in patronage from age 35 and above.
- 7. There is a similar age distribution between the subscribers and customers.
- 8. The weekend has the lowest trips record compared to other days of the week

## **Key Insights for Explanatory visualization.**

- 1. Distribution of age by gender and user type for the top 10 start station
- 2. Trip duration in seconds by User type for the top 10 start stations.
- 3. Age distribution and Trip duration for the two user types for the top 10 Start stations