**Data Science Case Study – Problem Statement**

**Overview:**

Your client, The Mayor of New York City, needs a better understanding of Citi Bike ridership. He wants an Operating Report for **January 2017** on his desk by the end of the week. Based on previous engagements we know the mayor is a big fan of visualizing data in charts.

Luckily, Citi Bike publishes quarterly trip data available for you to download and analyze. The data includes:

* Trip Duration (seconds)
* Start Time and Date
* Stop Time and Date
* Start Station Name
* End Station Name
* Station ID
* Station Lat/Long
* Bike ID
* User Type
  + (Customer = 24-hour pass or 3-day pass user; Subscriber = Annual Member)
* Gender (Zero=unknown; 1=male; 2=female)
* Year of Birth

Data can be downloaded here:

https://www.citibikenyc.com/system-data

The client wants to see a variety of data visualizations to answer the following questions:

1. Top 5 stations with the most starts (showing # of starts)
2. Trip duration by user type
3. Most popular trips based on start station and stop station)
4. Rider performance by Gender and Age based on avg trip distance (station to station), median speed (distance traveled / trip duration)
5. What is the busiest bike in NYC in 2017? How many times was it used? How many minutes was it in use?

Additionally, the Mayor has an idea that he wants to pitch to Citi Bike and needs your help proving its feasibility. He would like Citi Bike to add a new feature to their kiosks:

“Enter a destination and we’ll tell you how long the trip will take”.

We need you to **build a model that can predict how long a trip will take given a starting point and destination**. You will need to get creative about the factors that will predict travel time. Include model evaluation statistics and discussion of predictive features.

**Submission Details**

1. Presentation or Report: this is the version of your results you would present to a client. Include your visualizations, answers all 5 questions, and present the results of your model.

2. Your code: This should be where you actually did you work, in whatever tool you used (Jupyter Notebook, RStudio code, etc). This should be well commented so a reader could understand what you did without needing to run it.