**DSA 6000 – Fall 2019**

**Group 10:** Md Reza, Mujahid Khan, Woosung Jeong, Derick Karolak

**Project Name:** Beijing pm2.5 concentration analysis

**Predict:** Correlation between meteorological factors and pm2.5 concentration level

The objective of this project is to predict how different meteorological factors like temperature, pressure, dewpoints, precipitation, wind speed, wind direction, & season could impact the PM2:5 concentration level as well as increasing the air quality index (AQI) monotonically. According to the U.S. Environmental Protection Agency (EPA), the air quality index (AQI) could be regarded as “Good” if AQI is less than 50. Where our geolocation of interest is Beijing located in temper region of mason climate with a higher air quality index (AQI), between 2014 -2107 on average AQI of Beijing was above 100. Therefore, to identify the root cause we will be leveraging Beijing PM2.5 dataset from the UCI repository.

During different phases of this project, we will be performing the machine-learning algorithm to deal with outliers and missing values, and finally, we will be providing insights on the PM2.5 and AOI trends. We will also be using plot, ggplot, boxplot, sqldf, clustering, linear regression, and decision tree for exploring the correlation between:

* Seasons and pm2.5 - Md Reza
* Wind and pm2.5 - Mujahid Khan
* Precipitation and pm2.5 - Woosung Jeong
* Meteorological factors and pm2.5 - Derick Karolak

Finally, we will be summarizing our findings with visual aids and propose actions that could have improved the Beijing pm2.5 and air quality index.