Analysis of The PM 2.5 Level of Beijing, China

Mengqi Yuan
Assignment of Week 4
Applied Data Science Capstone Project, Coursera

1. Introduction: Problems and Background

China has one of the largest population in the world. With the rapid increase of middle class and high speed of industrialization, China has never countered such a serious air pollution problem in the past years (Fig 1-1). Beijing, as the capital city of China, has one of the highest population density in the whole country. In this capstone project, I would like to explore the air condition in different boroughs in the city of Beijing.

PM 2.5 is one of the atmospheric aerosol particles - also known as atmospheric particulate matter - which are microscopic solid or liquid matter suspended in the atmosphere of Earth. PM 2.5 are fine particles with a diameter of 2.5 micrometer or less. Particulates are the most harmful form of air pollution due to their ability to penetrate into lungs and blood streams unfiltered, causing DNA mutations, heart attacks, respiratory disease, and premature death.

Fig 1-1: Air pollution in China

The measurement will include three parts:
1. Compare the PM 2.5 LEVEL of different boroughs in the city of Beijing from a point of view in *numerical analysis*.
2. *Visualization* the difference of the PM 2.5 LEVEL of different boroughs in the city of Beijing.
3. *Clustering* the PM 2.5 LEVEL of different boroughs in the city of Beijing and analysis the similarities between each clusters and also the differences among clusters.

2. Data Section

The PM 2.5 LEVEL data can be downloaded on this website and geographical coordinates of each boroughs in the city of Beijing can be found on this website. The PM 2.5 LEVEL will firstly be analysis through numerical analysis, visualization and clustering. Then a new dataframe will be created by merging the PM 2.5 LEVEL dataset and the geographical coordinates dataset. With the geographical coordinates dataset, venues can be called through Foursquare. The relationship between the PM 2.5 LEVEL and the number of venues will be discussed. The result of clustering will be presented on the map of Beijing. Finally, conclusions and recommendations will be summarized according all the results in the previous parts.