Battle of the Neighborhoods

An Exploratory Study on Air Pollution in Urban Areas

By

Md. Saiful Bari Siddiqui

6th January, 2021

Introduction

Air Pollution is a burning problem nowadays, especially in urban areas. So It is necessary to examine what is responsible for poor air quality. A good way to do that is to determine what differentiates the cities with Least polluted air from those with Most polluted air. In this project we've used population data and the internal surroundings of 16 such cities to determine what the difference is between the two groups.

This study pretty much concerns every one, Because of the fact that everyone consumes air and is affected by it. **But the primary audience of this study is the City authorities and Urban Planners who are responsible for designing and Planning the structure of a city**, so that they have an insight on what could possibly lead to a City which is Pollution free.

Problem Statement

The core question that we want to answer is **What is responsible for poor air quality in many Cities in the World?** But in this study, we will attempt to address 2 more specific Problems. These are:

1. **Do the internal structure and surroundings of these cities have a significant impact on Air Pollution? If it does, what kind of structures in a city correlates to the city being more polluted?**
2. **What role does population have in case of Pollution? Does Higher Population Density mean More Pollution?**

Data

To understand and explore the following Open Data were used:

1. List of most-polluted cities by particulate matter concentration: https://en.wikipedia.org/wiki/List_of_most-polluted_cities_by_particulate_matter_concentration
2. A Comparative list of the Most and Least Polluted Cities in the World: https://www.rd.com/list/most-and-the-least-polluted-cities/
3. *Foursquare* Developers Access to venue data: https://foursquare.com/
4. Population Data from *Wikipedia*: https://en.wikipedia.org/wiki/List_of_cities_proper_by_population_density
5. Latitude & Longitude of Cities using *geopy* library: https://github.com/geopy/geopy

Using this data will allow exploration and examination to answer the questions. The venue data will be used to properly determine the common internal surroundings of each city and determine if there is some kind of correlation to pollution. The Population data will be used to compare the population density to examine if cities with higher pollution are also more densely populated. The lists of most polluted and least polluted cities will be used as reference. From these, we will use the 8 most polluted and 8 least polluted cities to compare the common venues within the cities retrieved from Foursquare location data and the population densities of the cities got from Wikipedia and other sources. The *geopy* library was used to extract the latitude & longitude values of the cities.