

Kolkata Neighborhood Clustering and it's relationship with Pollution

Aniketo Ghosh

6th May, 2020.

Introduction: Business Problem

Background: Global warming and climate change are pertinent issues of the current world and the conditions are getting worse by the day. A growing concern among citizens is the ever degrading air quality, especially so in cities. Thanks to accurate sensors of air pollution, we have data readily available for us to understand the real time situation ourselves.

Intended audience: In this project we will try to find the air quality index of Kolkata, India using PM 2.5 readings from across the city and its relationship with the neighbourhoods of the city. Specifically, this report will be targeted to stakeholders interested in climate change and ever increasing concern of degrading air quality ie. Government, NGOs or independent organizations.

Data: Acquisition and Characteristics

Acquisition: The data for this project has been acquired from <https://cleair.io/>. Cleair is a startup which specializes in low cost high accuracy network sensors which gather data about various aspects of air and noise pollution from their numerous sensors across the city.

The neighbourhood data of Kolkata will be collected using the Foursquare API. The venues in the neighbourhood, their types, categories and hence nature will be leveraged to make the study.

Characteristics: The data comes in csv format with the coordinates (latitude and longitude) of the location of the sensors and their average readings over the course of a week.

Feature selection: The features required for this study are Latitude, Longitude and PM 2.5 readings from the Cleair data and the headers we would get from the Foursquare API. Namely, Venue name, location, category, type and their frequency.