

Investigating Air Quality in Indian Cities

Background:

As the use of fossil fuels and petrol cars increases in our modern society, air quality is becoming more of a concern in daily life. With air pollution having links to numerous respiratory conditions, as well as causing strain to cardiovascular systems resulting in health problems after sustained exposure, we consider investigations into air quality of vital importance.

Due to this, the dataset regarding air quality across Indian cities was of interest to us and warranted a closer look. India is often regarded as having poor air quality, ranking in the top 5 most polluted countries for 2021.

For this project, we used two datasets taken from different authors on Kaggle: one regarding the air quality of Indian cities, and the second regarding background information for those cities, such as population in 2001 and 2011. Although the plan wasn't to use all the data contained in these files, it was more practical to have more data than we required and to decide what we would use once we were able to organise them further. From this, we aimed to look how each city compares in terms of air quality, as well as discover at potential causes for poor air quality.

This project was achieved through the use of Jupyter notebooks, Pandas and PostgreSQL databases, as they seemed the most appropriate for transforming the data we found in the CSV files, which were then used to create a final relational database that was able to be moulded for our purposes.

We set out to answer the following questions with our database:

Does population affect pollution levels?

Are some states more affected by pollution than others?

If there is a correlation, which cities are the most populous and polluted and the least populous and polluted

Findings:

Using the database we created, we found that the more populated a city was, the higher the average PPB (Parts Per Billion) of air pollution was, with Mumbai and Delhi ranking much higher than any other city in both population and rate of air pollution. However, when it came to the less populated cities, there were many instances of a city (such as Buxar or Motihari) having much higher levels of pollution than a significantly more populated city (like Ahmedabad), illustrating that population size is not the only factor we need to take into account. These cases warrant further investigation, which we were not able to do during the scope of this project. It was clear from the database that pollution

level is not even across Indian cities, with vast variations in PPB. So while population does seem to have an effect on pollution levels, the correlation is not by any means wholly established.

Below is a more detailed approach to the data manipulation performed, demonstrating how we arrived at these conclusions and the steps you can take to reproduce the code.