Investigating the Global Impact of Air Pollution on Public Health: A Multi-City Analysis Using Air Quality and Meteorological Data

Abstract

Air pollution is a growing problem in many cities around the world. It can seriously affect people's health, especially when it comes to breathing and heart problems. This project looks at air quality data from cities across different countries between 2015 and 2025. The goal is to understand how pollution levels — like PM2.5, PM10, NO_2, SO_2, CO, and O_3 — are connected to health issues.

The dataset also includes weather details like temperature, humidity, and wind speed, which help provide a broader view. I will clean the data, study the numbers, and create charts to show how pollution levels change over time and across cities. In the end, I hope to identify the cities most at risk and share insights that could support better air quality rules and protect public health.

Research Question & Problem Statement

Many people around the world are exposed to polluted air every day, especially in big cities. Breathing in pollutants like PM2.5, PM10, and harmful gases can lead to serious health problems such as asthma, lung disease, and heart conditions. But how strong is the connection between air pollution and these illnesses? Which cities are more affected than others?

This project aims to explore the link between air quality and health by using a dataset that tracks pollution levels in cities around the world from 2015 to 2025. By understanding which cities have the cleanest or most polluted air, and how pollution levels relate to possible health risks, I hope to bring attention to this issue and support better health and environmental policies.

Methodology

To explore how air pollution affects public health in cities around the world, this project will use a dataset containing daily air quality measurements from 2015 to 2025. The dataset includes key pollutant levels (PM2.5, PM10, NO_2, SO_2, CO, and O_3), along with additional information such as temperature, humidity, wind speed, date, city, and country.

The following steps will be taken:

1. **Data Cleaning**

   I'll begin by preparing the dataset. This includes checking for missing or incorrect values and fixing them using simple techniques like filling with average values or removing unusable rows.

2. **Data Organization**

   I will group the data by city and year to look for changes in pollution levels over time and compare trends across locations.

3. **Descriptive Analysis**

   I will calculate averages, maximums, and minimums for each pollutant to get a clear view of pollution levels in each city. This step helps summarize the data in a meaningful way.

4. **Visualization**

   Charts, graphs, and heatmaps will be used to make the data easier to understand. This will include comparisons between cities and visual patterns over time.

5. **Correlation Study**

   Where possible, I'll examine if there's any visible link between high pollution levels and health issues such as respiratory or heart problems, based on secondary research or known studies.

6. **Business and Policy Relevance**

   The findings may offer insights useful for government bodies, public health agencies, or city planners looking to reduce pollution and improve public health outcomes.

Tools and Resources

I will use Microsoft Excel and Python (with tools like Pandas and Matplotlib) for analysis and visualization. The dataset size is manageable, so the work can be done on a standard laptop or cloud tools like Google Colab — no special hardware is needed.

Expected Results

By the end of this project, I expect to identify which cities in the dataset have the highest and lowest levels of air pollution, based on key pollutants like PM2.5, PM10, NO_2, SO_2, CO, and O_3. I also aim to find trends over time — such as whether pollution is getting better or worse in certain areas.

I hope to discover meaningful differences between cities that could be connected to climate, policy, or location. Using clear visuals and basic analysis, I will try to highlight how pollution levels might be linked to health concerns, based on information from trusted studies. In the end, I want this project to offer useful insights that raise awareness and support better decisions around clean air and public health.

References

[1] B. R. Gurjar, K. Ravindra, and A. S. Nagpure, “Air pollution trends over Indian megacities and their local-to-global implications,” Atmospheric environment (1994), vol. 142, pp. 475–495, 2016, doi: 10.1016/j.atmosenv.2016.06.030.

[2] M. Greenstone and R. Hanna, “Environmental Regulations, Air and Water Pollution, and Infant Mortality in India,” *The American economic review*, vol. 104, no. 10, pp. 3038–3072, 2014, doi: 10.1257/aer.104.10.3038.

[3] M. Fahim, Md. E. Uddin, R. Ahmed, Md. R. Islam, and N. Ahmed, “A Machine Learning Based Analysis Between Climate Change and Human Health: A Correlational Study,” in *2022 International Conference on Computer and Applications (ICCA)*, IEEE, 2022, pp. 1–6. doi: 10.1109/ICCA56443.2022.10039484.

[4] S. M. T. Hasan, “Global Urban Air Quality Index Dataset (2015–2025),” Kaggle, [Online]. Available: https://www.kaggle.com/datasets/syedmtalhahasan/global-urban-air-quality-index-dataset-2015-2025