Pollution Map of Indian Cities

AYUSH JHA
BSC COMPUTER SCIENCE AND DATA ANALYTICS
IIT PATNA
ayush_2312res211@iitp.ac.in

GITHUB REPO LINK:https://github.com/Ayush281204/Pollution-Map-of-Indian-Cities/tree/main/Poluation%20city%20in%20india

Introduction

Air pollution has become one of the most serious environmental and health issues across the globe, particularly in developing countries like India. Rapid urbanization, industrial growth, increased vehicular traffic, and construction activities have contributed to deteriorating air quality in many cities. Among the various pollutants, Particulate Matter (PM2.5) is considered a major health hazard because of its ability to penetrate deep into human lungs and bloodstream.

While air pollution data is generally available in tabular or numerical formats, it is difficult to analyze spatial patterns from raw numbers alone. A geospatial visualization of pollution levels provides better clarity by showing how pollution varies across different regions. This project attempts to create a pollution map of Indian cities using geospatial data visualization in Python.

By plotting pollution values on a map of India, the project not only highlights critical hotspots but also serves as a powerful tool for researchers, policymakers, and the general public to understand and address air quality issues.

Problem Statement

Although air pollution data is collected and published by multiple agencies, the data is often limited to **tables, charts, or lists of numbers**. This approach has two major drawbacks:

1. It does not **highlight spatial distribution** – making it hard to see which regions are more affected.

2. It limits **comparative analysis** between cities and regions.

Thus, there is a need to represent air pollution data on a **map-based visualization platform** where cities can be represented as points with varying colors or intensities based on their pollution levels.

This project addresses the problem by integrating **geospatial shapefiles of India** with **pollution datasets** (city-wise PM2.5 values) to create a **visual pollution map**.

Objectives

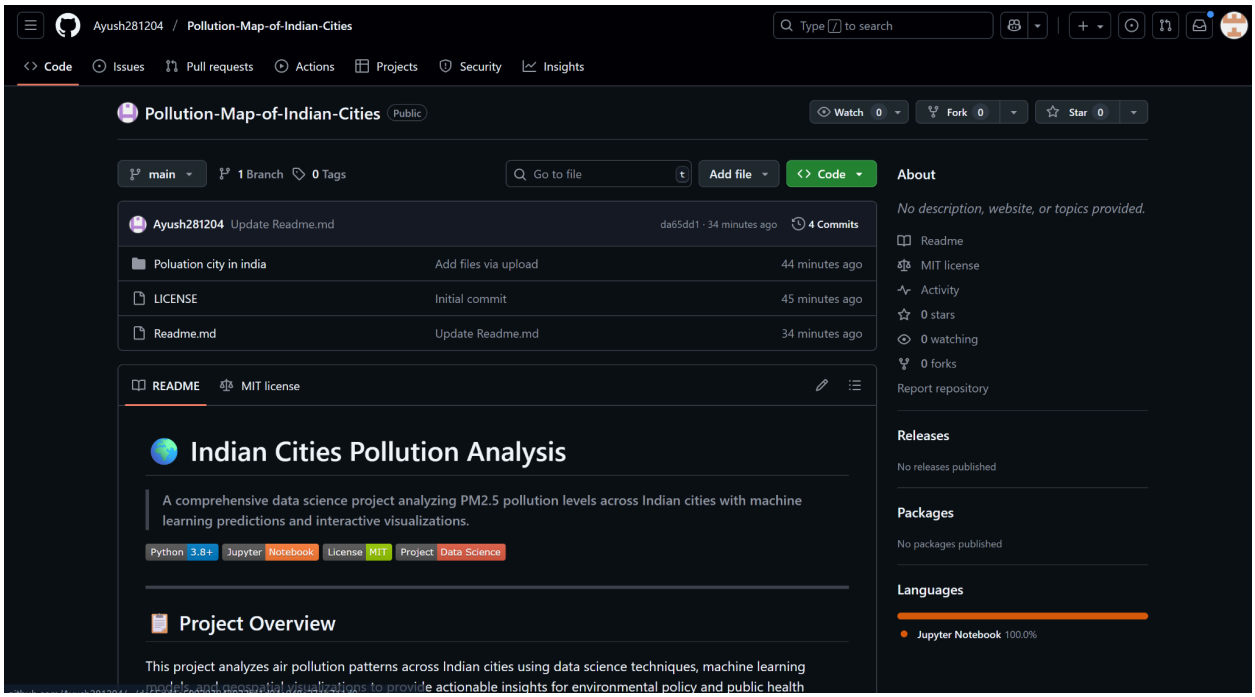
The primary objectives of this project are:

- To create a **geospatial visualization** of air pollution levels across Indian cities.
- To use **Python libraries (GeoPandas, Matplotlib, Pandas)** for handling shapefiles, datasets, and plotting.
- To represent **PM2.5 values** of cities as color-coded points on an India map.
- To highlight **pollution hotspots** and allow **comparisons** between different regions.
- To provide a **framework for future integration** with real-time data sources.

Scope of the Project

The scope of this project covers:

- Visualization of **city-level air pollution data** on a map of India.
- Implementation of data processing and geospatial plotting using Python.
- Focus on **PM2.5 values** as the key indicator of air pollution.
- Provision for **scalability**, where more pollutants (NO2, SO2, CO) and time-series data can be integrated in the future.
- Final output in the form of a **map image**, which can be used in reports, dashboards, or awareness campaigns.

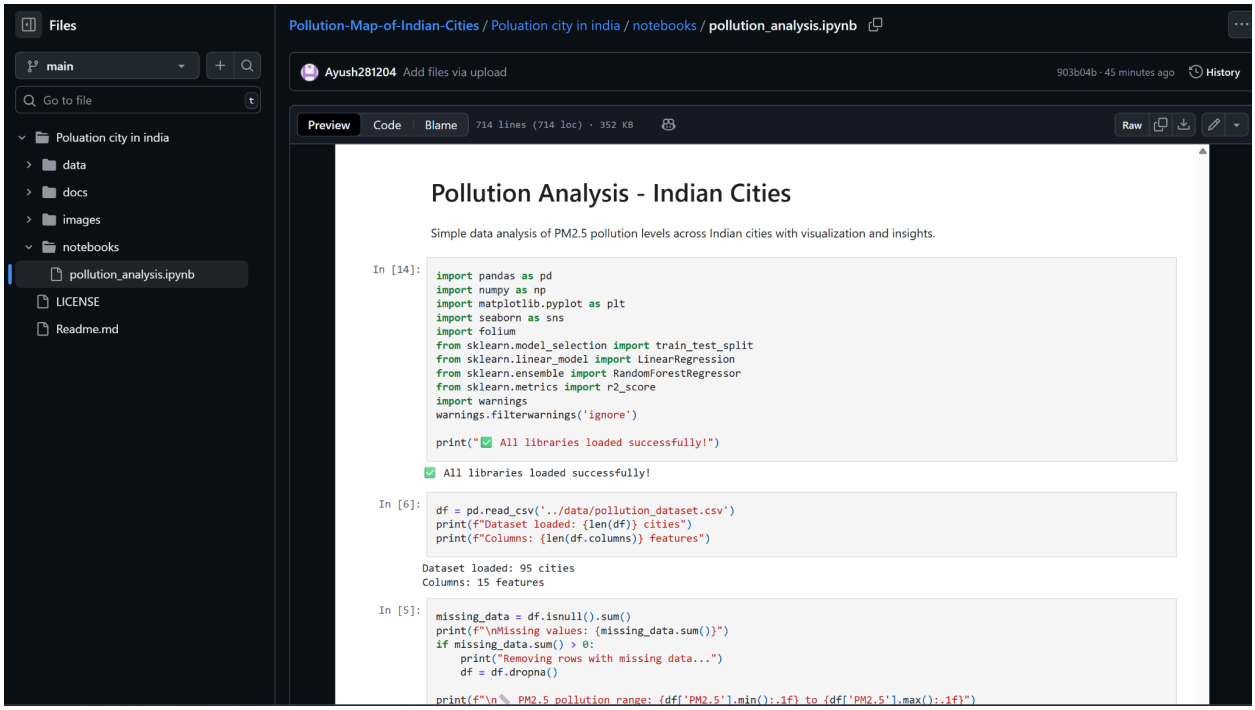


Literature Review / Background

Several studies and projects have attempted to address air pollution visualization:

- **The World Health Organization (WHO)** regularly publishes air quality guidelines, emphasizing PM2.5 as a critical parameter.
- Platforms like **OpenAQ** and **AirVisual** provide real-time pollution maps globally.
- Academic studies show that **geospatial mapping** helps policymakers target interventions in critical regions.

However, many existing solutions are **global or generic**, while this project specifically focuses on **India** and provides a **Python-based, reproducible workflow** for local pollution visualization.



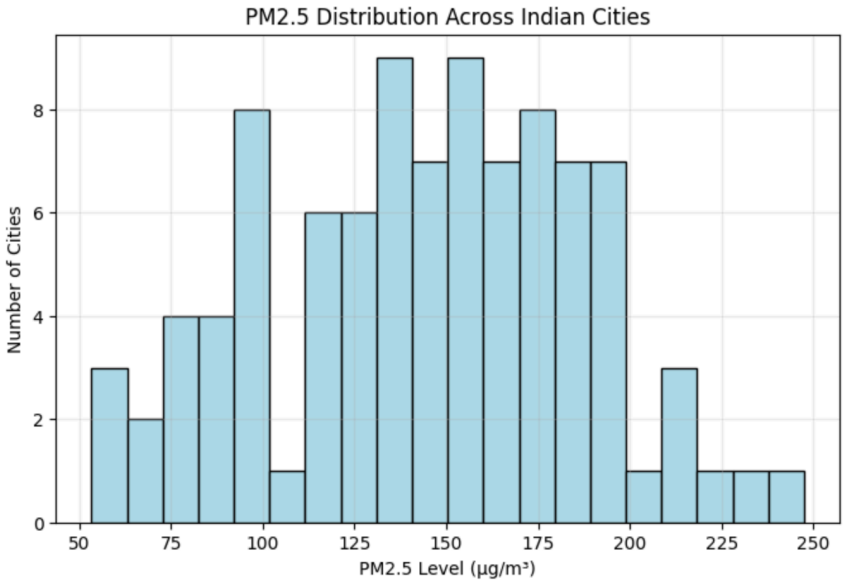
Visualization Results

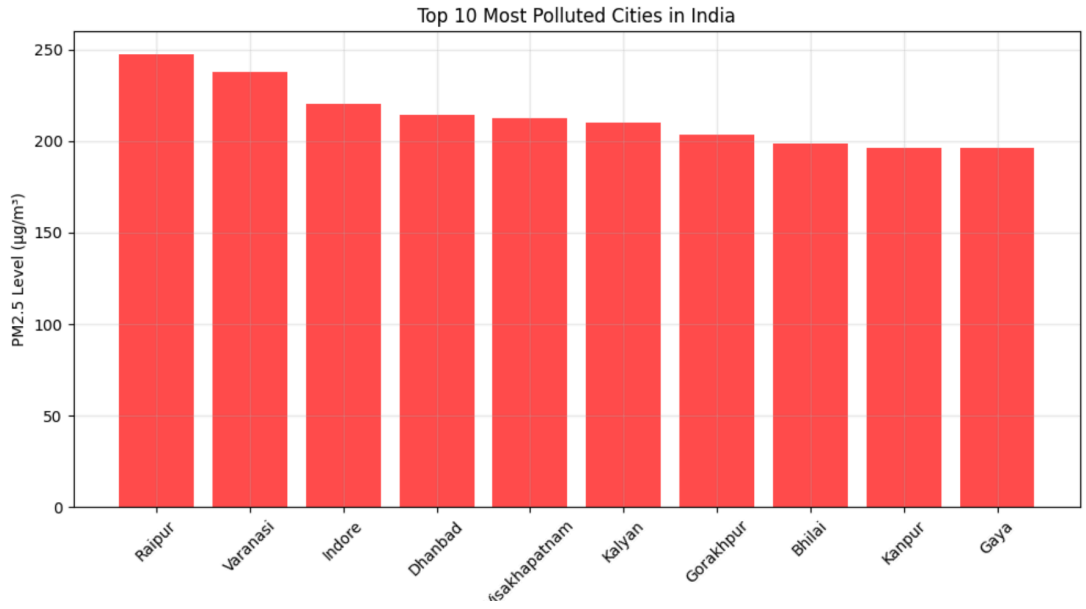
After executing the Python implementation, the project generates a **pollution map of Indian cities**. The visualization highlights how air pollution levels vary spatially across the country.

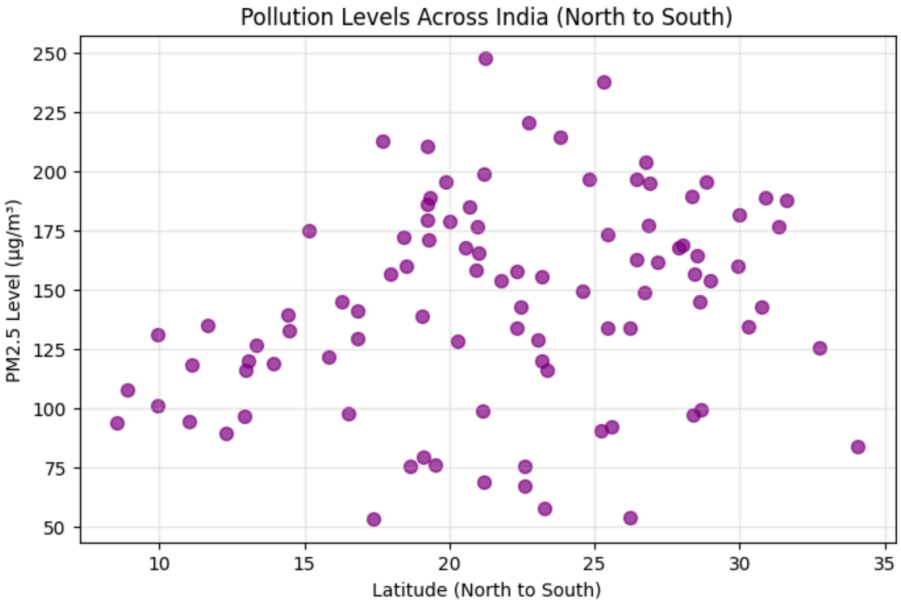
Key Features of the Visualization:

- **Base Map**: India's geographic boundaries are plotted using a shapefile.
- **Pollution Markers**: Each city is represented as a **point** placed according to its latitude and longitude.
- **Color Gradient**:
    - **Green** – Low pollution (cleaner air).
    - **Yellow/Orange** – Moderate pollution.
    - **Red/Dark Red** – Severe pollution levels.
- **City Labels**: Each city is annotated to ensure readability.
- **Legend & Title**: A legend explains the color coding, and the map title provides context.

‘


PM2.5 Distribution Across Indian Cities
Number of Cities
PM2.5 Level (µg/m³)
0
2
4
6
8
50
75
100
125
150
175
200
225
250


Top 10 Most Polluted Cities in India
PM2.5 Level (µg/m³)
0
50
100
150
200
250
Raipur
Varanasi
Indore
Dhanbad
ʻsakhapatnam
Kalyan
Gorakhpur
Bhilai
Kanpur
Gaya


Pollution Levels Across India (North to South)
PM2.5 Level (µg/m³)
Latitude (North to South)
250
225
200
175
150
125
100
75
50
10
15
20
25
30
35


Pollution Map of Indian Cities
Latitude
PM2.5 Level
Kanpur
Gorakhpur
Varanasi
Gaya
Dhanbad
Indore
Raipur
Kalyan
Visakhapatnam
35
30
25
20
15
10
70.0
72.5
75.0
77.5
80.0
82.5
85.0
87.5
225
200
175
150
125
100
75

Applications

The **Pollution Map of Indian Cities** has a wide range of applications across multiple domains. By converting raw air quality data into geospatial visualizations, it provides meaningful insights that can be used by governments, industries, researchers, educators, and the general public.

1. Government & Policy Making

- **Identifying Pollution Hotspots**: Authorities can quickly locate the most polluted regions and prioritize them for intervention.
- **Policy Evaluation**: Helps measure the effectiveness of environmental regulations, such as restrictions on industrial emissions or vehicle usage.
- **Urban Planning**: Supports sustainable city planning by showing areas that require green zones, stricter emission norms, or better traffic management.
- **Resource Allocation**: Enables efficient deployment of air quality monitoring stations and healthcare facilities in highly polluted regions.

2. Environmental Research

- **Data Analysis**: Researchers can analyze how pollution varies geographically and correlate it with population density, industrial activity, or meteorological factors.
- **Trend Monitoring**: The map can be extended into time-series visualizations to study seasonal and long-term pollution trends.

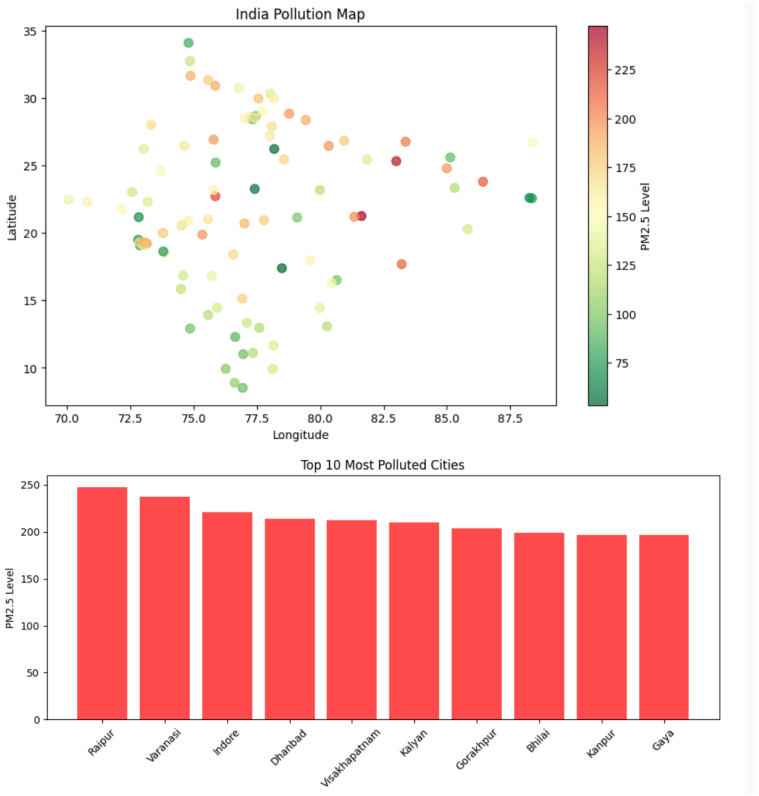
- **Comparative Studies**: Helps compare air quality across multiple cities or states for academic and scientific publications.

3. Public Health & Awareness

- **Health Risk Communication**: Citizens can understand the air quality of their city in a visual, intuitive manner.

- **Awareness Campaigns**: NGOs and environmental activists can use maps in campaigns, posters, and presentations to educate the public.

- **Personal Decision-Making**: Individuals may use pollution maps to decide travel plans, outdoor activities, or even relocation based on air quality.



Future Enhancements

Although the current project provides a static visualization of pollution levels in selected Indian cities, there are several opportunities to **enhance and expand** the system for real-world usage.

1. Real-Time Data Integration

- **Live Pollution Monitoring**: Integrate APIs from platforms like **OpenAQ** or **CPCB (Central Pollution Control Board, India)** to fetch live air quality data.
- **Dynamic Visualization**: Instead of static images, the map can auto-update as new pollution data is recorded.

2. Multi-Pollutant Support

- Extend the visualization to include other pollutants such as **NO2, SO2, CO, and Ozone**.
- Use **layered maps** or **multi-colored markers** to represent different pollutants.
- Provide an option to toggle between pollutants for comparative study.

3. Time-Series & Trend Analysis

- Add the ability to visualize **pollution changes over time** (daily, monthly, or yearly).
- Use **animated maps** that show how pollution spreads or reduces during specific periods (e.g., Diwali, winter smog, lockdown periods).

4. Predictive Modeling (AI/ML)

- Apply **machine learning models** to predict future pollution levels based on historical data, weather conditions, and traffic patterns.

- Enable **early warnings** for critical pollution events, such as smog in Delhi during winter.

5. Interactive Dashboards

- Build a **web-based dashboard** using **Plotly Dash, Streamlit, or Flask** where users can explore pollution data interactively.

- Features may include:

    - Search by city/state

    - Filter by pollution type

    - Graphical reports alongside maps

Machine Learning

In [10]:

```
X = df[['PM10', 'NO2', 'AQI', 'Industry_Score', 'Traffic_Score']]
y = df['PM2.5']

print(f"Using better features: {list(X.columns)}")

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

linear_model = LinearRegression()
linear_model.fit(X_train, y_train)
linear_pred = linear_model.predict(X_test)

forest_model = RandomForestRegressor(n_estimators=50, random_state=42)
forest_model.fit(X_train, y_train)
forest_pred = forest_model.predict(X_test)

linear_r2 = r2_score(y_test, linear_pred)
forest_r2 = r2_score(y_test, forest_pred)

print(f"\nLinear Regression Accuracy: {linear_r2:.3f} ({linear_r2*100:.1f}%)")
print(f"Random Forest Accuracy: {forest_r2:.3f} ({forest_r2*100:.1f}%)")

if forest_r2 > linear_r2:
    print("✅ Random Forest is better!")
    winner = "Random Forest"
    best_score = forest_r2
else:
    print("✅ Linear Regression is better!")
    winner = "Linear Regression"
    best_score = linear_r2

print(f"\nBest model: {winner}")
print(f"Best accuracy: {best_score:.3f} ({best_score*100:.1f}%)")
print("✅ Much better results with proper features!")
```

Conclusion

The **Pollution Map of Indian Cities** project successfully demonstrates how geospatial visualization can transform raw environmental data into actionable insights. By combining pollution datasets with geospatial shapefiles, the project bridges the gap between **numerical data representation** and **spatial awareness**, making the information more intuitive and impactful.

Through this project, we have shown how **PM2.5 values** across different Indian cities can be effectively visualized using **Python libraries such as GeoPandas, Pandas, and Matplotlib**. The resulting map highlights pollution hotspots in a way that tabular data alone cannot achieve. Cities with extreme pollution levels, such as Delhi, stand out clearly in the visualization, while comparatively cleaner cities, such as Chennai, are also easily identifiable.

The project underscores several key learnings:

- **Geospatial Analysis Value**: Visual mapping allows for easier interpretation and comparison of environmental data across regions.

- **Data-Driven Awareness**: The visualization provides a strong basis for public awareness campaigns and policy interventions.

- **Scalable Framework**: The methodology can be easily extended to include additional pollutants, more cities, or even real-time streaming data.

At the same time, the project acknowledges certain **limitations**. The dataset used for demonstration is small and static, and pollution levels within a city can vary across neighborhoods. Additionally, shapefile resolution may lack fine granularity. Despite these limitations, the framework remains highly adaptable for future enhancements.

Thank You

This project on *Pollution Map of Indian Cities* has been an enriching learning experience, combining data analysis with geospatial visualization to highlight the issue of air pollution across different regions of India. I am grateful to all open-source contributors and resources that made this work possible.

For source code, dataset, and detailed implementation, please visit my GitHub repository: https://github.com/Ayush281204/Pollution-Map-of-Indian-Cities