Air Quality Analysis Report

Assessment and Visualization of Urban Air Quality in India (2015–2020)

Supriyo Ghosh
Shuvadip Pal
B.Tech Final Year Students
GitHub Repo: `github.com/ghoshsupriyo/air-quality-india`

July 25, 2025

Abstract

This study presents a data-driven analysis of air quality levels in major Indian cities between 2015 and 2020. Using the dataset from the Central Pollution Control Board (CPCB), pollutants such as PM2.5, PM10, NO2, CO, SO2, and O3 are analyzed to determine temporal and geographical patterns. The study aims to identify pollution trends, compare cities, and evaluate the effectiveness of regulatory measures using data visualization and statistical techniques. Trends are explored across different cities and seasons, and pollutant interdependencies are visualized using correlation heatmaps.

1. Introduction

Air pollution is a critical issue facing many developing nations. In India, industrialization, rapid urbanization, vehicular emissions, and seasonal agricultural burning have contributed to poor air quality across several metropolitan areas.

This report analyzes historical air quality data for Indian cities to evaluate pollutant concentrations and patterns. The project leverages Python-based data analytics and visualization tools to make the findings more insightful and actionable.

2. Dataset Description

Source: Central Pollution Control Board via Kaggle
Dataset File: `city_day.csv`
Data Columns:

- `Date`, `City`
- Pollutants: PM2.5, PM10, NO, NO2, NOx, CO, SO2, O3, NH3
- Volatile Organics: Benzene, Toluene, Xylene
- `AQI`, `AQI_Bucket`: Qualitative AQI classification

The dataset spans from 2015 to 2020 and covers over 25 cities in India. Missing values were interpolated, and any duplicate or corrupted entries were discarded. Data was stored in a structured CSV file.

3. Methodology

- Tools: Python (Pandas, NumPy, Matplotlib, Seaborn)
- Computed daily, monthly, and annual pollutant averages
- Created time-series line graphs and city-wise bar plots
- Used box plots for distribution comparison and heatmaps for correlation
- Tracked AQI bucket transitions by year and region

All analysis was carried out using a Jupyter notebook named `Air_Quality_check.ipynb`, and visualizations were exported directly from code.

4. Findings

4.1 Temporal Trends

- PM2.5 and PM10 levels increase dramatically in northern cities during winter months.
- CO and NO2 levels spike around Diwali and New Year due to fireworks and higher vehicle usage.

4.2 City Comparison

- Delhi showed the highest pollution metrics across PM2.5 and PM10.
- Mumbai and Chennai maintained relatively stable and healthier air due to coastal dispersion.

4.3 AQI Trends

Analysis of AQI buckets shows a trend of degradation across the years. Cities like Delhi shifted from "Poor" to "Very Poor"/"Severe" categories repeatedly. Southern cities stayed in "Moderate" to "Satisfactory" ranges for longer durations.

4.4 Correlation Analysis

- PM2.5 and PM10 have a strong positive correlation.
- CO and NOx also correlate, reflecting vehicular impact.
- O3 showed inverse patterns with some gases.

5. Conclusion

The air quality across urban India continues to be a serious public health concern, especially in North India during winter. The study emphasizes the importance of monitoring, data transparency, and actionable policy interventions. More sustainable transport systems, green city planning, and citizen awareness are critical.

6. Future Work

- Integrate live sensor data with dashboards
- Build machine learning models (LSTM, ARIMA) to predict AQI
- Public engagement via real-time air quality mobile apps

7. References

- **Central Pollution Control Board (CPCB)**: `https://cpcb.nic.in/`
- **Kaggle Dataset**: `https://www.kaggle.com/datasets/rohanrao/air-quality-data-in-indi`
- **WHO Reports on Global Air Quality**