Air Quality Index (AQI) Analysis of Indian Cities

1. Project Overview

Air pollution is a major environmental and public health challenge in India. This project analyzes **city-level air quality data** to understand pollution patterns, seasonal trends, city-wise disparities, and associated health risks.

The analysis combines **Python-based Exploratory Data Analysis (EDA)** with **interactive dashboards (Power BI)** to deliver both analytical depth and business-ready insights.

Objectives

- Analyze AQI trends over time (daily, monthly, yearly)
- Identify the most and least polluted cities
- Understand seasonal patterns in air pollution
- Assess health risk exposure across cities
- Create decision-ready dashboards for comparison and ranking

---

2. Dataset Summary

Dataset Source

- File: `city_day_air_quality.csv`
- Granularity: **Daily city-level air quality records**
- Coverage: **26 Indian cities**
- Time Period: **2015 – 2020**

Key Features

Category	Columns
Location	City
Time	Date, Year, Month
Pollutants	PM2.5, PM10, NO, NO2, NOx, NH3, CO, SO2, O3, Benzene, Toluene
Target Metrics	AQI, AQI_Bucket

Dataset Shape

- Rows: **29,531**
- Columns (final): **19**

---

3. Data Preprocessing & Cleaning

Steps Performed

- Converted `Date` column to datetime format
- Extracted `Year` and `Month` for time-series analysis
- Removed high-missing column (`Xylene`)
- Imputed missing pollutant values using:
    - City-wise median
    - Global median (fallback)

AQI Reconstruction

- Computed **pollutant sub-indices** using **CPCB-defined breakpoints**
- Used **linear interpolation** within breakpoint ranges
- Final AQI computed as the **maximum sub-index** across pollutants
- Missing AQI values filled using calculated AQI

AQI Bucketing (CPCB Standard)

- Good (0–50)
- Satisfactory (51–100)
- Moderate (101–200)
- Poor (201–300)
- Very Poor (301–400)
- Severe (400+)

---

4. Exploratory Data Analysis (EDA)

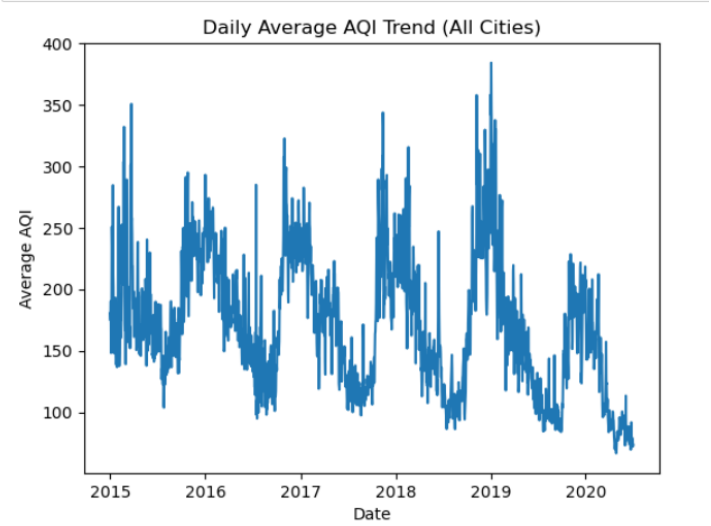
4.1 Overall AQI Statistics

- Mean AQI: **~166**
- Median AQI: **118**
- Indicates majority of days fall in **Moderate** category

4.2 Daily AQI Trend (All Cities)

- High volatility across years
- Extreme spikes observed in northern cities
- Indicates episodic pollution events

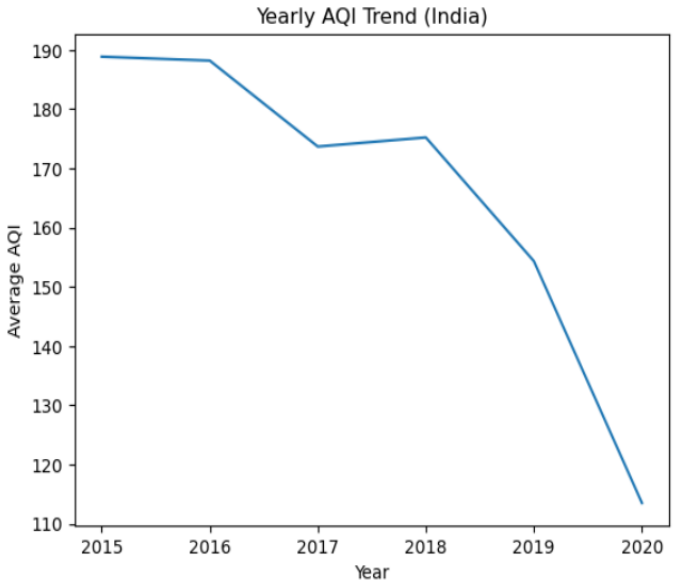
Insight: Air quality fluctuates sharply day-to-day, especially during winter months.



4.3 Yearly AQI Trend (India)

- Overall declining trend from **2015 to 2020**
- Significant improvement after **2018**

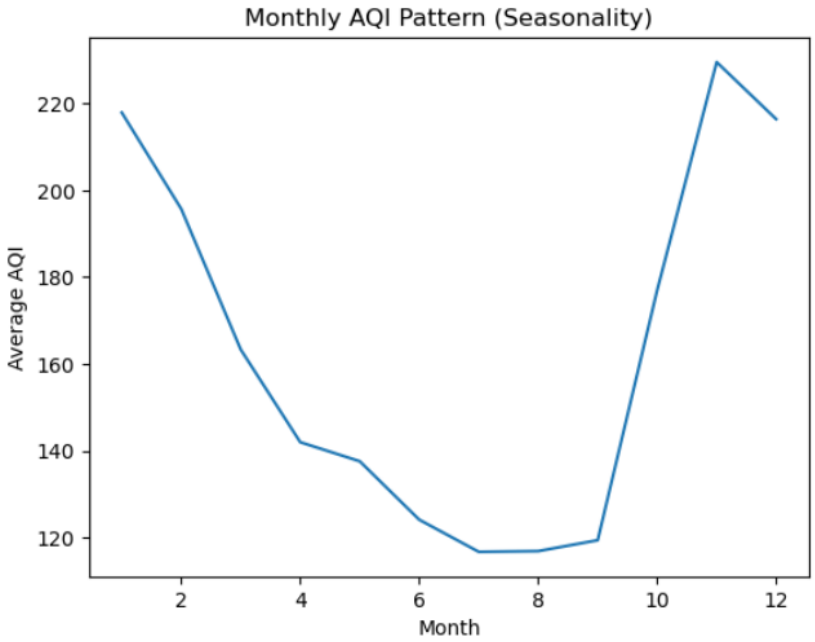
Insight: National-level air quality shows gradual improvement, possibly due to policy interventions.



4.4 Monthly AQI Pattern (Seasonality)

- Worst AQI: **November – January**
- Best AQI: **July – August (Monsoon)**

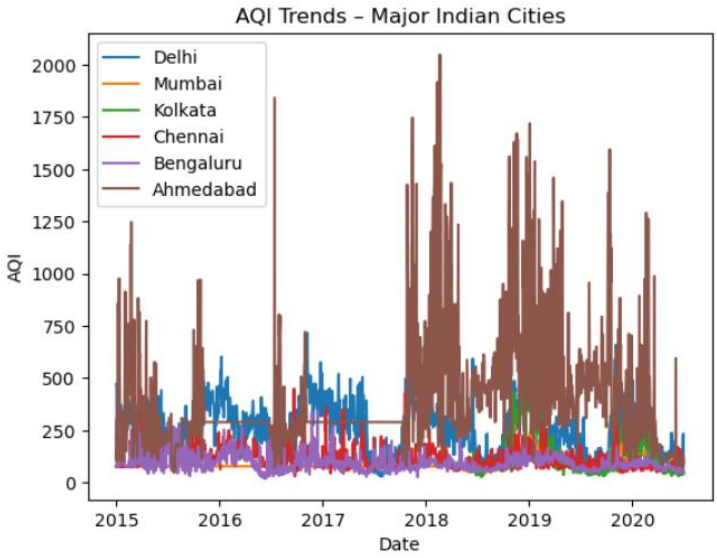
Insight: Strong seasonal dependency driven by weather, crop burning, and emissions.



4.5 City-wise AQI Trends (Major Cities)

Cities Analyzed: - Delhi - Mumbai - Kolkata - Chennai - Bengaluru - Ahmedabad

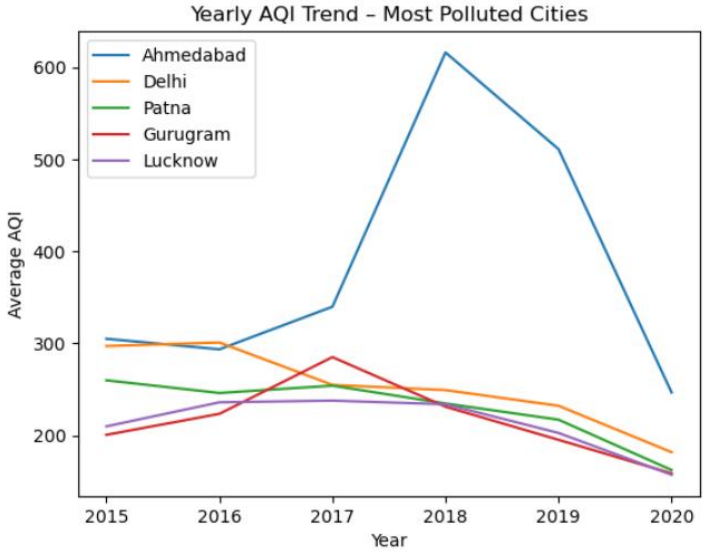
Observations: - Ahmedabad & Delhi show extreme volatility - Coastal cities (Mumbai, Chennai) maintain relatively stable AQI



4.6 Most Polluted Cities (Average AQI)

Top contributors: 1. Ahmedabad 2. Delhi 3. Patna 4. Gurugram 5. Lucknow

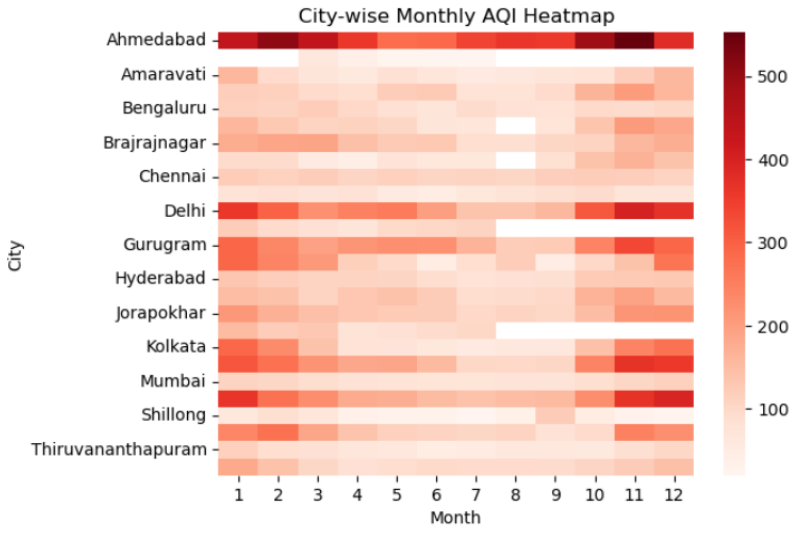
Insight: Inland and northern cities dominate pollution rankings.



4.7 City-wise Monthly AQI Heatmap

- Northern cities show intense winter pollution
- Southern & coastal cities show moderate seasonality

Insight: Geography and climate strongly influence AQI behavior.



---

5. Health Risk Analysis

AQI Category Distribution (%)

- Moderate: **34.4%**
- Satisfactory: **34.0%**
- Poor or Worse: **~26%**

High-Risk Exposure (Poor + Very Poor + Severe)

- Ahmedabad: Highest number of high-risk days
- Delhi & Patna follow closely

Public Health Insight: Even a small percentage of Severe AQI days poses serious health risks.

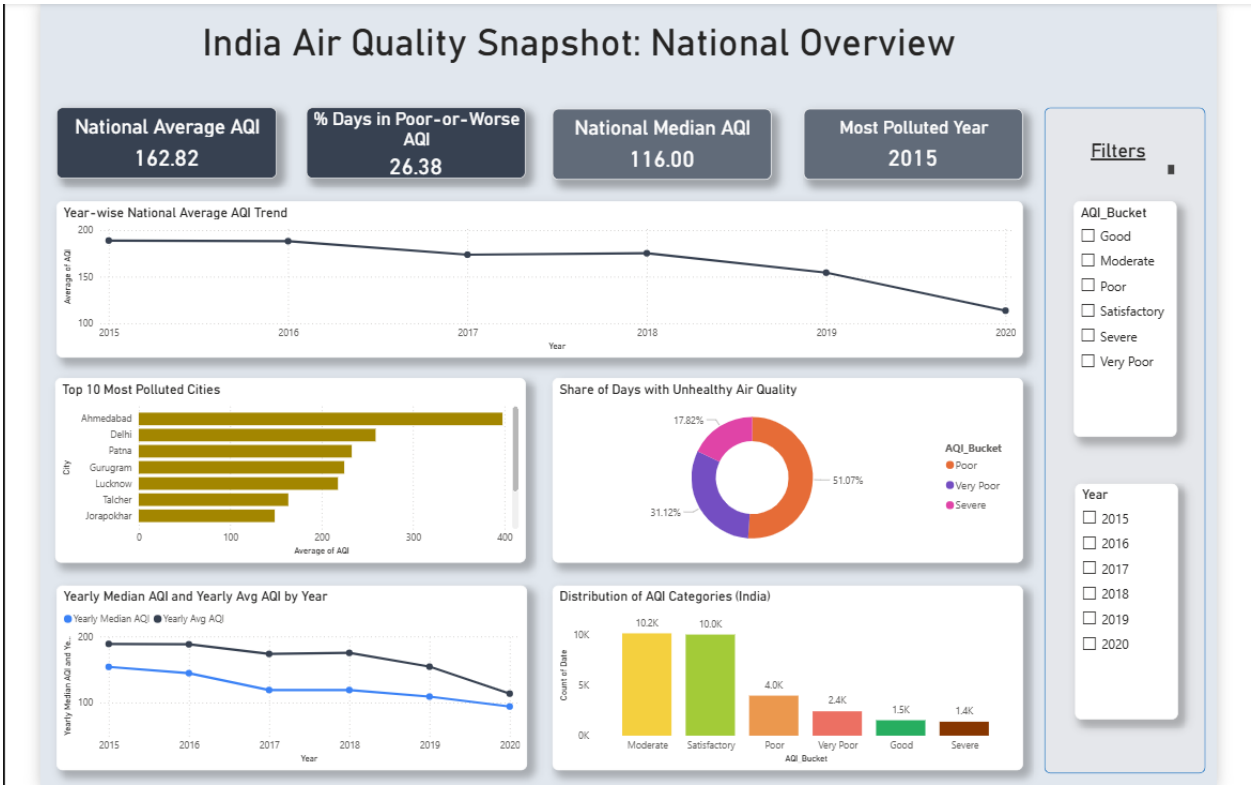
---

6. Dashboard Integration (Power BI)

Dashboard 1: National AQI Overview

- AQI distribution
- Yearly & monthly trends

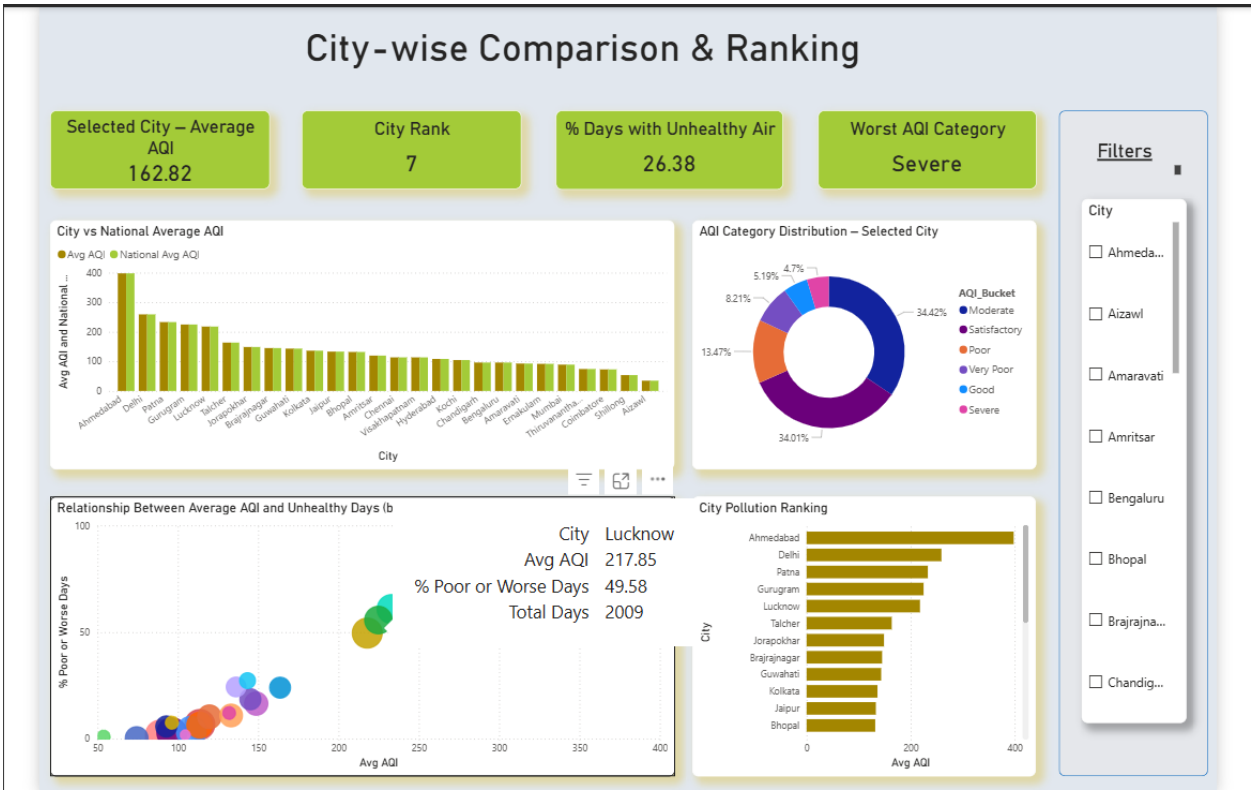
- National health risk exposure



Dashboard 2: City-wise Comparison & Ranking

- City AQI ranking
- City vs national average
- % unhealthy days

- Worst AQI category (KPI-driven)



---

7. Key Insights & Conclusion

- Air pollution in India is **highly seasonal and location-dependent**
- Northern inland cities face **persistent and severe pollution**
- National AQI shows **gradual improvement**, but city-level risks remain
- Data-driven dashboards enable **targeted policy and health planning**

---

8. Tools & Technologies Used

- **Python**: Pandas, Matplotlib, Seaborn
- **Power BI**: Interactive dashboards, DAX measures
- **Statistical Methods**: Median imputation, ranking, correlation

---

9. Future Scope

- Predictive AQI modeling (ML)
- Policy impact analysis
- Integration with weather data
- Real-time AQI monitoring dashboards