BMS COLLEGE OF ENGINEERING

(Autonomous College under VTU)
Bull Temple Road, Basavanagudi, Bangalore – 560019

A project report on

"AQI Analysis of Indian Cities"

Submitted in partial fulfillment of the requirements for the award of degree

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE AND ENGINEERING
(DATA SCIENCE)
By

Hemanth Raj M V, 1BM23CD022

Kushal K V, 1BM23CD030

Rohan R, 1BM23CD051

Under the guidance of

Prof. Lavanya Naik, Assistant Professor

Department of Computer Science and Engineering(Data Science)

February-June 2025

BMS COLLEGE OF ENGINEERING

(Autonomous College under VTU)
Bull Temple Road, Basavanagudi, Bangalore – 560019

Department of Computer Science and Engineering
(DATA SCIENCE)

CERTIFICATE

This is to certify that the project entitled "***AQI Analysis of Indian Cities***" is a bona-fide work carried out by **Hemanth Raj MV(1BM23CD022), KushalKV(1BM23CD030), RohanR(1BM23CD051)** in partial fulfillment for the award of degree of Bachelor of Engineering in **CSE(Data Science)** from **Visvesvaraya Technological University, Belgaum** during the year **2024-2025**. It is certified that all corrections/suggestions indicated for Internal Assessments have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the Bachelor of Engineering Degree.

Signature of the Guide	**Signature of the HOD**
Name and Designation	**Name and Designation**

Examiners

Name of the Examiner	**Signature of the Examiner**

1.
2.

Table of Contents

Page No.

Acknowledgment

I would like to express my sincere gratitude to Prof. Lavanya Naik, Assistant Professor, Department of Computer Science and Engineering (Data Science), BMS College of Engineering, for her invaluable guidance, constructive feedback, and continuous encouragement throughout the duration of this project.

Her support at every stage — from understanding the dataset to structuring the analysis and developing insightful dashboards — played a critical role in shaping this report.

I also extend my thanks to the faculty members of the Department of Computer Science and Engineering (Data Science) for creating an academic environment that fosters research, critical thinking, and practical learning.

Additionally, I am thankful to the BMS College of Engineering administration for providing the resources and tools required to carry out this work efficiently.

Finally, I appreciate the collaboration and effort of my team members, whose collective inputs made this project successful.

Abstract

This project focuses on analyzing and visualizing air quality data across major Indian cities using Microsoft Power BI. The dataset includes pollutants like PM2.5, PM10, NOx, and CO, measured across multiple cities and time periods. Through interactive dashboards and visuals, the project highlights pollution trends, identifies high-risk regions, and explores relationships between pollutants and AQI (Air Quality Index).

The goal of the analysis is to support environmental awareness and decision-making by transforming raw environmental data into digestible visual insights. Key findings include the consistent dominance of PM2.5 and NOx in driving AQI values, especially in cities like Delhi and Gurugram. The dashboards enable users to identify seasonal spikes and geographic pollution patterns quickly.

This project demonstrates how business intelligence tools can be applied beyond commercial use cases — in this instance, to support environmental health and urban sustainability efforts.

1. Introduction

1.1 Purpose

Air pollution has become one of the most pressing environmental challenges in modern India, especially in densely populated and industrialized urban regions. The rising levels of airborne pollutants such as particulate matter (PM2.5 and PM10), nitrogen oxides (NOx), and carbon monoxide (CO) have a direct and significant impact on human health, contributing to respiratory illnesses, cardiovascular conditions, and premature deaths.

In light of these concerns, this project was undertaken with the following goals:

- To examine air quality patterns across major Indian cities over a three-year period (2015 to mid-2017).
- To identify the most harmful pollutants and the cities that are most affected by them.
- To transform raw pollution data into meaningful, interactive visualizations using Microsoft Power BI.
- To support environmental analysis through clear visuals that help stakeholders understand city-wise and pollutant-wise trends.

Through this project, we aim to present air pollution data in a way that is not only informative but also easy to interpret. Power BI dashboards provide a platform to monitor how AQI changes over time, how pollution levels vary across regions, and how different pollutants contribute to the overall air quality.

The project is designed for multiple audiences — researchers, policymakers, students, and the general public — who may not have the time or expertise to interpret tabular datasets. By visualizing pollution trends interactively, the project helps uncover high-risk periods, identify regional pollution hotspots, and guide future mitigation strategies.

The long-term intention is to contribute to awareness and inform efforts to improve air quality through effective data presentation.

1.2 Scope

The scope of this project is defined by the dataset used, the pollutants analyzed, and the boundaries of visualization and analysis.

Dataset Overview:

- Total records: Approximately 29,000
- Coverage: Around 25 Indian cities
- Time period: January 2015 to mid-2017
- Fields included: Date, City, PM2.5, PM10, NO, NO2, NOx, CO, O3, AQI, and AQI Bucket
- Key focus cities: Delhi, Gurugram, Patna, Ahmedabad, Talcher, Mumbai

Data Quality:

- Some pollutant fields such as O3 and NO contain missing data for select cities.
- The analysis prioritizes cities and pollutants with more complete and consistent data.
- AQI Buckets were used to classify days into categories: Good, Satisfactory, Moderate, Poor, Very Poor, and Severe.

Visualization Scope:

- Dashboards were created using Microsoft Power BI with visuals including:
    - Line and area charts for AQI and pollutant trends over time
    - Bar and treemap visuals for city comparisons
    - Geographical map visuals to highlight regional AQI
    - Scatter plots to examine relationships between pollutants
- Interactive filters for city, pollutant type, and time period were incorporated to enhance usability.

Analytical Scope:

- The project focuses solely on descriptive and visual analytics; no predictive or real-time modeling was performed.
- While real-time data and advanced forecasting are not included, the current setup lays the groundwork for such future enhancements.

2. Insight on Dataset

The dataset used in this project provides detailed air quality records from **January 2015 to mid-2017**, across various Indian cities. It forms the foundation of the entire analysis and visualization effort.

Key Features of the Dataset:

- Contains **~29,000 records**, with each row representing daily pollution data.
- Covers **~25 cities**, including:
    - Delhi
    - Gurugram
    - Patna
    - Ahmedabad
    - Talcher
    - Mumbai
- Important columns include:
    - **Pollutants**: PM2.5, PM10, NO, NO2, NOx, CO, O3
    - **AQI (Air Quality Index)** and **AQI Bucket** classification
    - **City** and **Date** – enabling both time-series and regional analysis

Observations:

- **Average AQI** across all entries is around **166** (classified as “Moderate to Poor”).
- **PM2.5** and **PM10** levels in some cities exceed **900**, indicating severe pollution.
- Data includes **missing values**, particularly in less-monitored cities or pollutants.
- Focus was maintained on cities and fields with well-populated, clean data.

Usefulness:

- Enabled pollutant trend tracking across time and geography.
- Helped identify cities with consistently high AQI values.

- Supported correlation analysis (e.g., PM2.5 vs PM10).
- Served as a reliable source for creating interactive dashboards in Power BI.

3. Features Explored in Power BI

Power BI was the primary tool used to transform raw air pollution data into intuitive, interactive dashboards. It allowed for the creation of visually rich reports that help explore patterns and trends across cities and pollutants.

Types of Visuals Used:

- **Line Charts**
  - To display time-series trends for AQI and major pollutants.
  - Helped visualize how pollution varied monthly and seasonally.
- **Bar & Column Charts**
  - Used for comparing pollutant levels across cities.
  - Highlighted cities with the highest average PM2.5, PM10, and NOx levels.
- **Donut & Pie Charts**
  - Represented AQI Bucket distributions (e.g., Good, Satisfactory, Poor).
  - Helped understand proportion of days falling into each pollution category.
- **Treemaps**
  - Used to show which cities had the most frequent poor AQI levels.
  - Visually emphasized high-risk areas.
- **Scatter Plots**
  - Displayed relationships between pollutants.
  - Notably showed a strong correlation between PM2.5 and PM10.
- **Map Visuals**
  - Plotted average AQI for each city on the map of India.

    - Enabled easy comparison of geographical pollution patterns.

Interactive Elements:

- **Slicers**
    - Allowed selection of cities, years, and months.
    - Helped users focus on specific regions or time periods.
- **Filters**
    - Applied for pollutant type and AQI range.
    - Made dashboards more dynamic and user-friendly.
- **Custom DAX Measures**
    - Created to calculate:
        - Average pollutant levels
        - Year-wise AQI trends
        - Percentage of days in each AQI category
    - Improved analytical depth and data summarization

Benefits of Using Power BI:

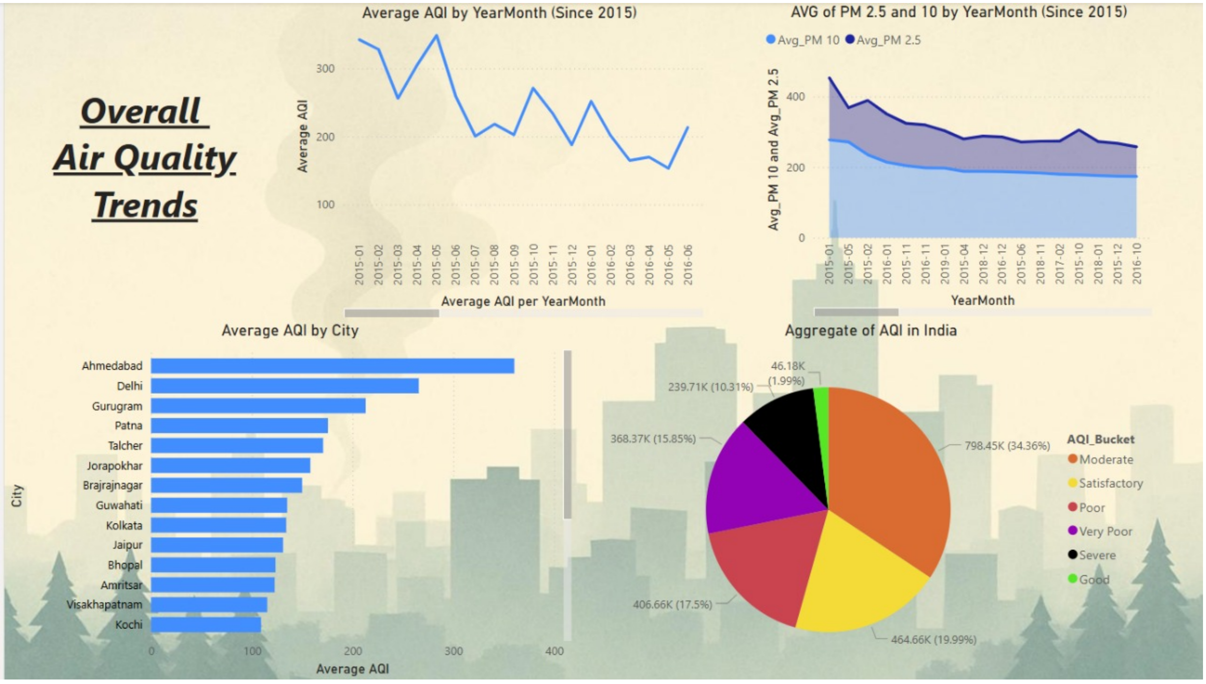
- Drag-and-drop interface made dashboard creation efficient.
- Visuals updated dynamically with filters and slicers.
- Easy export of reports and visuals for documentation and sharing.
- Enabled storytelling through data with clarity and visual appeal.

4. Dashboard

The dashboards created in Power BI serve as the visual foundation of this project. Each one explores different dimensions of the air quality dataset — temporal, spatial, pollutant-specific, and categorical. All visuals are interactive, filterable, and intended to make large datasets easier to interpret.

Below is a breakdown of the dashboards and the visuals used in each.

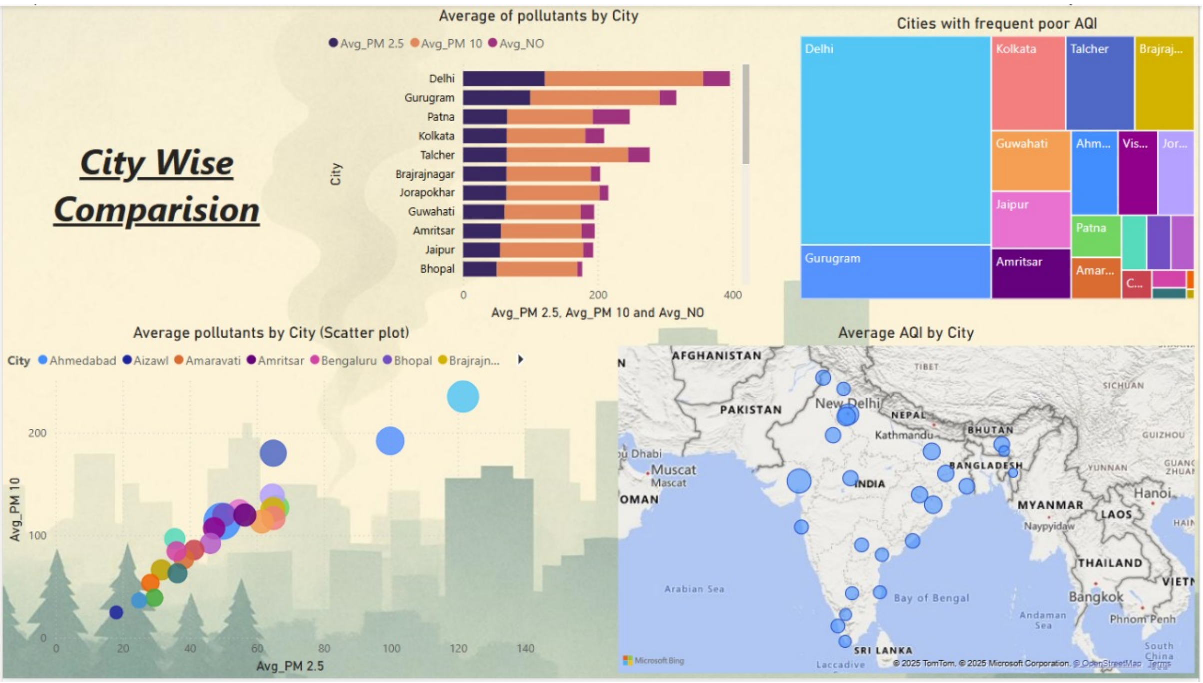
Dashboard 1: Overall Air Quality Trends



This dashboard provides a nationwide overview of AQI behavior over time.

- Line chart (Top-Left): Shows Average AQI by YearMonth from Jan 2015 to mid-2017. There is a noticeable downward trend from 2015 to early 2017, with occasional seasonal spikes.
- Area chart (Top-Right): Displays PM2.5 and PM10 averages across the same period. PM10 levels consistently remain higher than PM2.5, indicating suspended dust or smoke.
- Bar chart (Bottom-Left): Compares Average AQI by City. Cities like Ahmedabad, Delhi, and Gurugram rank highest.
- Pie chart (Bottom-Right): Represents the distribution of AQI Buckets. Over 34% of days fall under "Moderate", followed by "Satisfactory", "Poor", and “Very Poor”.

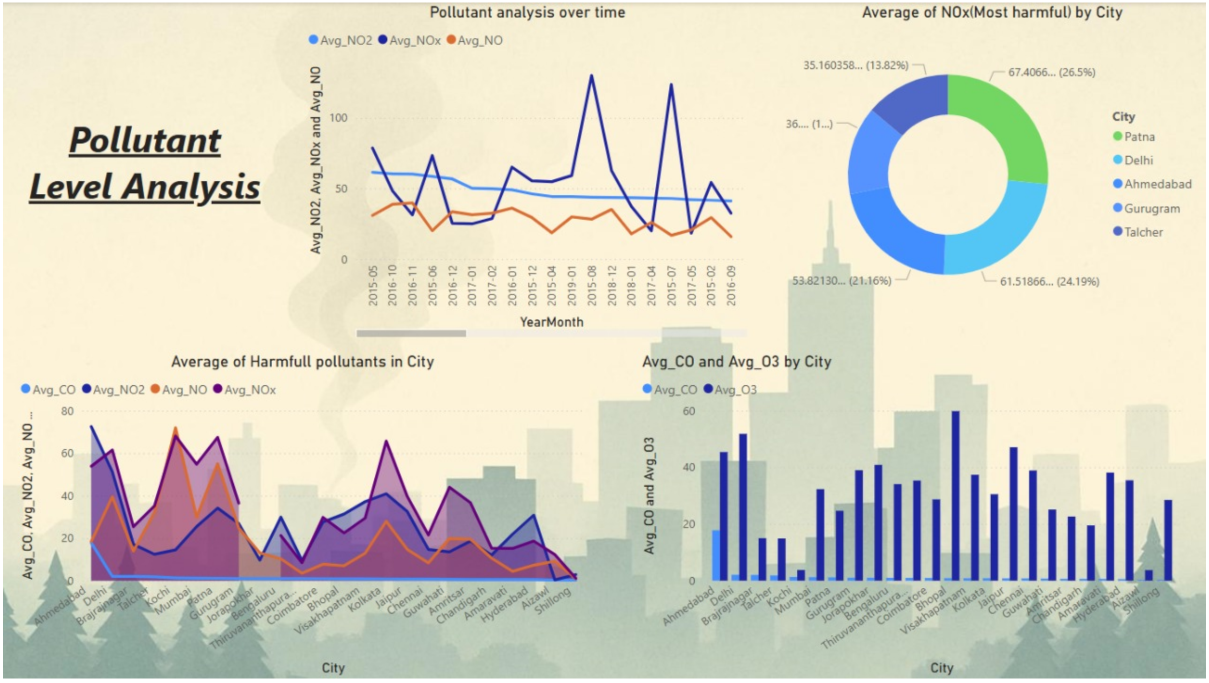
Dashboard 2: City-Wise Comparison



This dashboard highlights spatial comparisons between cities and pollutant behavior.

- Horizontal bar chart (Top-Left): Compares average PM2.5, PM10, and NO values across major cities. Delhi, Gurugram, and Patna show the highest levels.
- Treemap (Top-Right): Emphasizes cities with frequent poor AQI. Delhi dominates in size, followed by Gurugram and Brajrajnagar.
- Map chart (Bottom-Left): Uses a bubble map to visualize average AQI geographically. Cities in North and Central India exhibit larger circles, indicating higher pollution.
- Scatter plot (Bottom-Right): Correlation between PM2.5 and PM10 levels. The linear trend confirms a strong positive correlation — where PM2.5 is high, PM10 is also high.

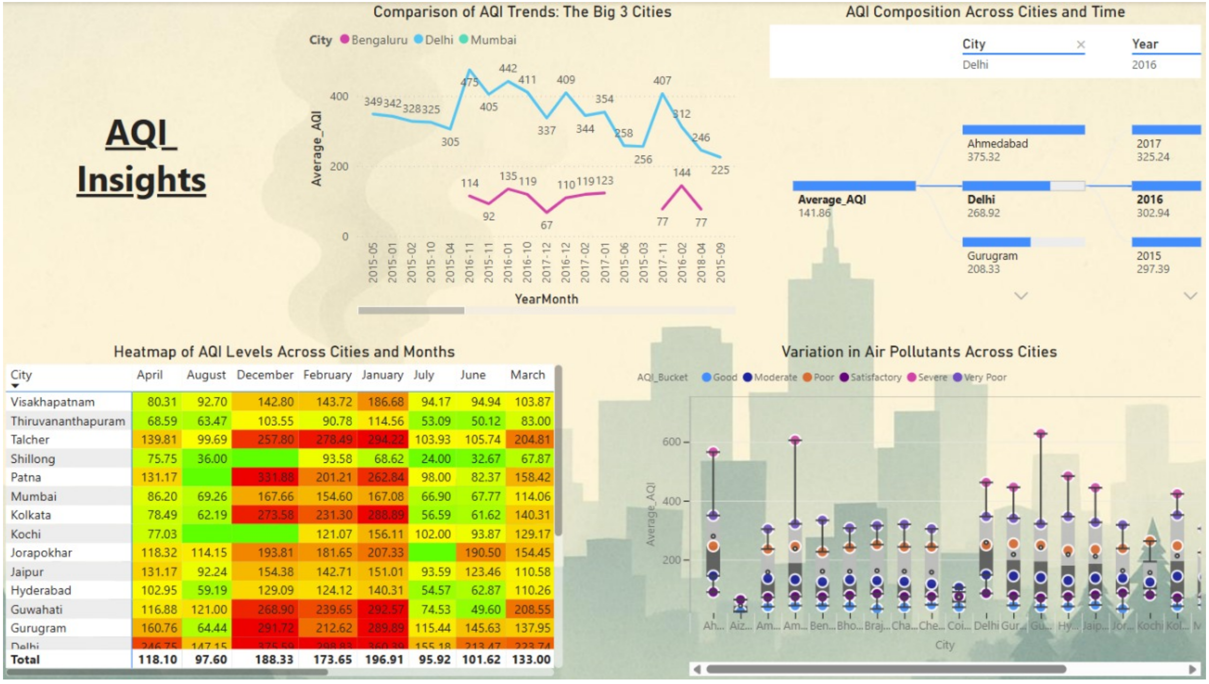
Dashboard 3: Pollutant Level Analysis



This dashboard focuses on pollutant-specific analysis over time and by city.

- Line chart (Top-Left): Shows temporal trends of NO, NO2, and NOx. Spikes are seen in late 2015 and early 2016, aligning with festival and crop-burning seasons.
- Donut chart (Top-Right): Displays the most affected cities by NOx. Patna and Delhi top the list, indicating harmful nitrogen emissions.
- Stacked area chart (Bottom-Left): Visualizes the average values of CO, NO, NO2, and NOx across cities. Delhi, Gurugram, and Talcher exhibit dense pollutant overlap.
- Bar chart (Bottom-Right): Compares Avg_CO and Avg_O3 by City. Cities like Ahmedabad, Chennai, and Lucknow have high ozone presence, suggesting photochemical smog.

Dashboard 4: Overall AQI Trends and City-Wise Insights



This dashboard highlights AQI variation across major Indian cities by time, severity, and pollutant category.

- Line chart (Top-Center): Shows AQI trends in Delhi, Mumbai, and Bengaluru from 2015 to 2017.
  ▪ Delhi records the highest and most fluctuating AQI levels, especially in winter months.
  ▪ Bengaluru consistently has the lowest AQI among the three cities.

- Sankey chart (Top-Right): Compares average AQI of Delhi, Ahmedabad, and Gurugram across three years.
  ▪ Delhi's AQI declined from 302.94 (2016) to 268.92 (2017).
  ▪ Ahmedabad had the highest value at 375.32 in 2017.

- Heatmap (Bottom-Left): Displays monthly AQI levels across cities.
  ▪ Patna, Talcher, and Delhi show extreme AQI in winter (red zones).
  ▪ Cities like Kochi and Visakhapatnam remain in the safe "green" range most months.

- Bubble/bar chart (Bottom-Right): Categorizes cities by AQI buckets (Good to Very Poor).
  ▪ Delhi, Talcher, and Kolkata frequently fall in "Severe" and "Very Poor" zones.
  ▪ Shillong and Kochi stay largely within "Good" and "Satisfactory" levels.

5. Analysis

This section outlines the major findings and interpretations drawn from the visual dashboards. By exploring pollutant levels, AQI patterns, and city-wise variations, the analysis provides insight into India's air quality trends during the period of 2015 to 2017.

5.1 City-Wise AQI Observations

- Cities like Delhi, Gurugram, Patna, and Talcher reported consistently high AQI values.
- Ahmedabad showed concerning spikes in PM2.5 and PM10, despite not being among the top polluted cities by reputation.
- Coastal cities such as Kochi and Visakhapatnam exhibited better air quality, likely due to natural wind flow and lower industrial density.

5.2 Seasonal and Temporal Trends

- AQI levels peaked during the winter months (October to January), particularly in northern cities.
- Possible contributing factors included:
  • Crop residue burning in the Indo-Gangetic plain
  • Lower wind speeds and atmospheric inversion during winter
  • Diwali-related emissions due to firecrackers and household heating
- Monsoon and summer months recorded comparatively lower AQI levels.

5.3 Pollutant Behavior and Comparison

- PM2.5 and PM10 consistently recorded the highest concentrations across cities.
- The scatter plot confirmed a strong positive correlation between PM2.5 and PM10 levels.
- NOx concentrations were especially high in Delhi and Patna, indicating a link to vehicular and combustion sources.
- Cities such as Lucknow and Ahmedabad had elevated CO and O3 levels, reflecting urban traffic and smog formation.

5.4 AQI Bucket Distribution

- Over 34% of the recorded days fell under the “Moderate” AQI category.
- “Satisfactory” and “Poor” categories contributed nearly 40% combined.
- Less than 2% of days qualified as “Good”, showing overall poor urban air health.
- Several cities faced frequent instances of “Very Poor” and “Severe” AQI days.

5.5 Spatial Patterns and City Clustering

- Map visuals highlighted pollution concentration in North and Central India.
- Treemaps clearly showed that Delhi, Gurugram, and Brajrajnagar had the most number of poor-AQI days.
- Cities with better air quality clustered around southern and coastal regions.

5.6 Additional Insights

- Cities not ranked among the top polluted occasionally showed sharp, dangerous spikes in specific pollutants.
- A gradual decline in AQI values from 2015 to early 2017 was observed in some regions, hinting at the impact of early pollution control policies.

5.7 Pollutant-Specific Case Studies

To deepen the understanding of pollutant impact, three specific pollutants were tracked across selected cities:

PM2.5 – Fine Particulate Matter

- Highest concentrations recorded in Delhi, Talcher, and Ahmedabad.
- PM2.5 values often exceeded 300 $\mu g/m^3$ in winter months — well beyond safe limits.
- Industrial zones and areas with high vehicular traffic showed consistently elevated levels.

NOx – Nitrogen Oxides

- Patna, Delhi, and Gurugram showed recurring spikes in NOx emissions.
- These pollutants are strongly associated with vehicle emissions and thermal power stations.
- Line charts revealed that NOx levels remained high throughout the year, with occasional dips during monsoon.

CO – Carbon Monoxide

- Observed in higher concentrations in cities like Ahmedabad and Mumbai.
- Peaks often aligned with periods of high vehicle congestion.
- Despite being less visible in media reports, CO remains a silent but significant pollutant.

5.8 Relationship Between AQI and Pollutants

- A correlation between high AQI and high PM2.5/PM10 was observed.
- Days categorized under "Very Poor" AQI always showed concurrent spikes in NOx or PM levels.
- Scatter plots confirmed a positive relationship between PM2.5 and PM10.
- Combined analysis shows that even if one pollutant is under control, others (e.g., NOx or O3) can still push AQI into dangerous ranges.

5.9 Visual Insights from Dashboards

- Treemaps helped isolate cities with the highest frequency of Poor/Very Poor AQI days.
- Line charts enabled tracking pollution events over time — helpful in linking with external causes like Diwali or crop burning.
- The India map visual served as a strong communication tool to highlight spatial differences — ideal for policy discussion.

5.10 Summary of Key Analytical Observations

- Delhi and Gurugram are the most critically polluted among all studied cities.
- Winter months (Nov–Jan) show the highest risk for most cities.
- AQI improved slightly between 2015 and 2017, especially in Tier-2 cities like Kochi and Bhubaneswar.
- Urban areas with industrial density and traffic congestion are the primary zones of concern.

6. Conclusion & Future Enhancements

This project provided a visual and analytical overview of air quality data across Indian cities between 2015 and 2017. Using Microsoft Power BI, complex pollution data was transformed into understandable dashboards that revealed meaningful trends and regional differences.

Through this analysis, it became evident that cities like Delhi, Gurugram, and Patna consistently faced poor air quality, especially during the winter months. PM2.5 and NOx were identified as major contributors to pollution. The dashboards also made it clear that while a few cities occasionally recorded “Good” AQI days, the majority fell within the “Moderate” to “Very Poor” categories.

The strength of this project lies in its visual simplicity. Decision-makers and the general public can quickly understand environmental conditions through interactive visuals instead of raw tables.

Future enhancements that can add more value to this project include:

- Adding real-time data integration from APIs (such as OpenAQ or CPCB).
- Including AQI prediction models using time series forecasting or regression.
- Extending the dataset to include data from 2018–2025 to track progress.
- Linking pollution data to public health records for deeper insight.
- Deploying the dashboard on a public platform for accessibility.

This project demonstrates the power of data visualization in raising awareness and supporting environmental policy decisions.

7. References

1. Central Pollution Control Board (CPCB). Retrieved from: https://cpcb.nic.in
2. OpenAQ Platform. Global air quality data. Retrieved from: https://openaq.org
3. Microsoft Power BI Documentation. Retrieved from: https://learn.microsoft.com/en-us/power-bi
4. National Air Quality Index Guidelines – India. Retrieved from: https://app.cpcbccr.com
5. Dataset Source: https://www.kaggle.com/datasets/rohanrao/air-quality-data-in-india

8. Appendix

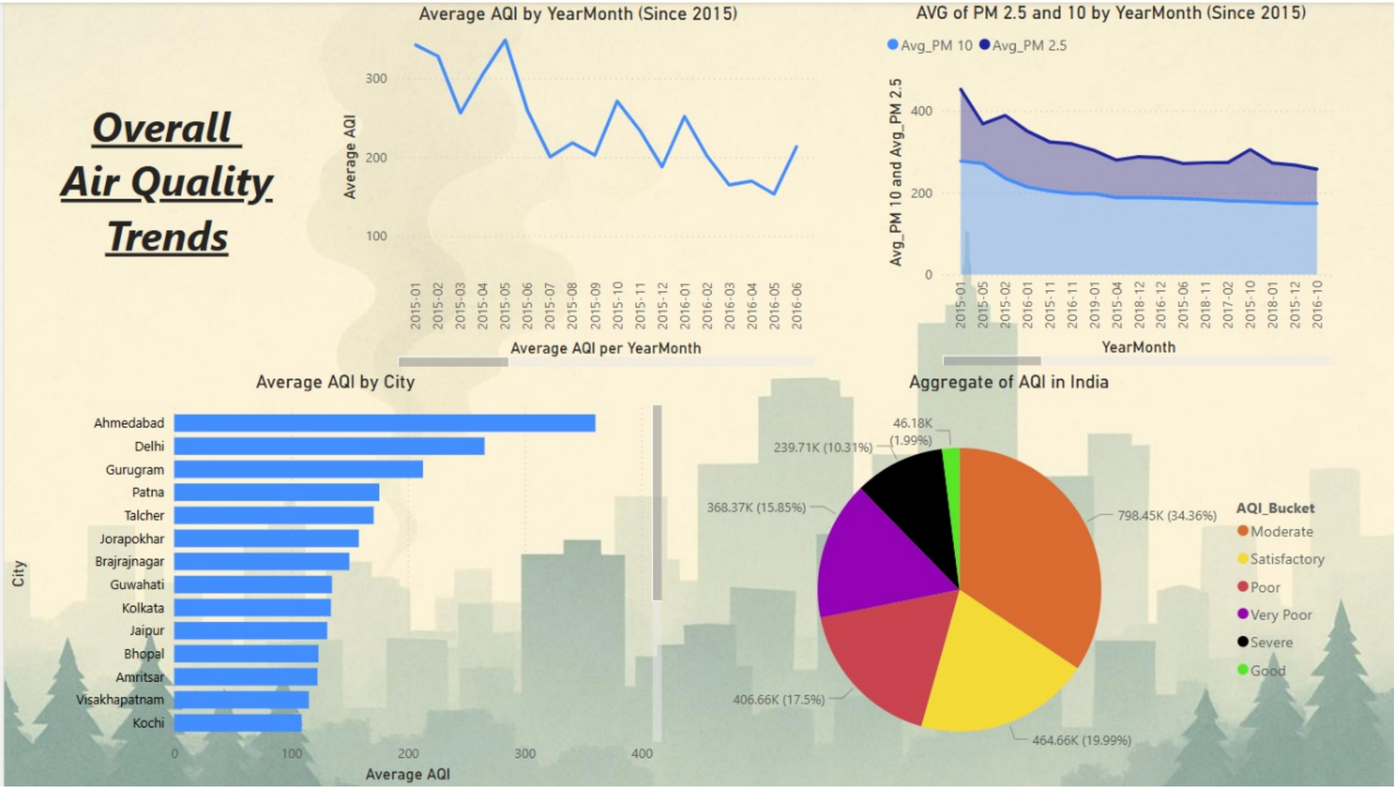


Figure A.1: Line chart showing monthly AQI trends across India from 2015 to 2017.
Figure A.2: Area chart comparing monthly PM2.5 and PM10 levels over time.
Figure A.3: Bar chart ranking cities based on their average AQI values.
Figure A.4: Pie chart displaying AQI Bucket distribution (Good to Severe) across the dataset.

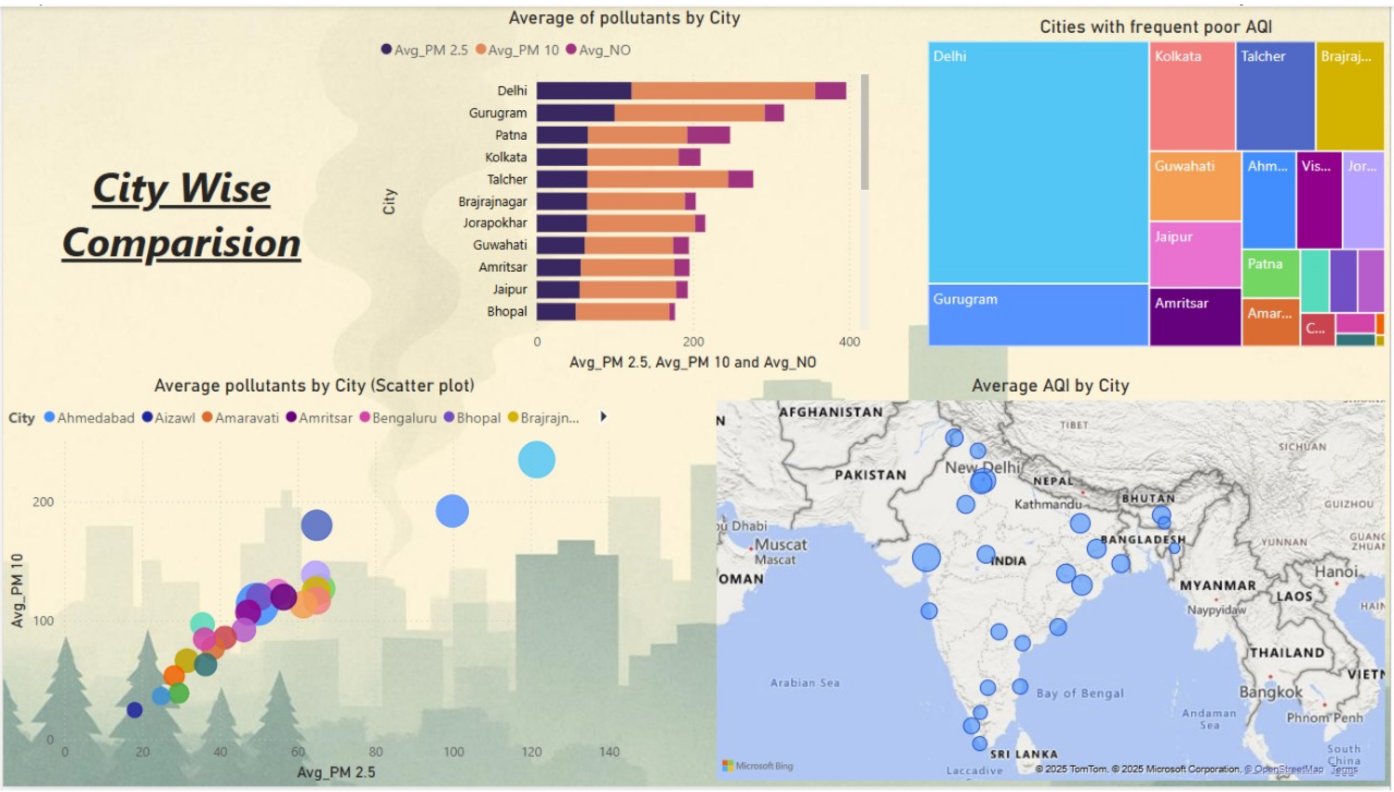


Figure A.5: Stacked bar chart comparing PM2.5, PM10, and NO values by city.
Figure A.6: Treemap showing cities with the highest frequency of “Poor” or worse AQI days.
Figure A.7: Scatter plot showing correlation between PM2.5 and PM10 across cities.
Figure A.8: Map visualizing average AQI values by geographic location in India.

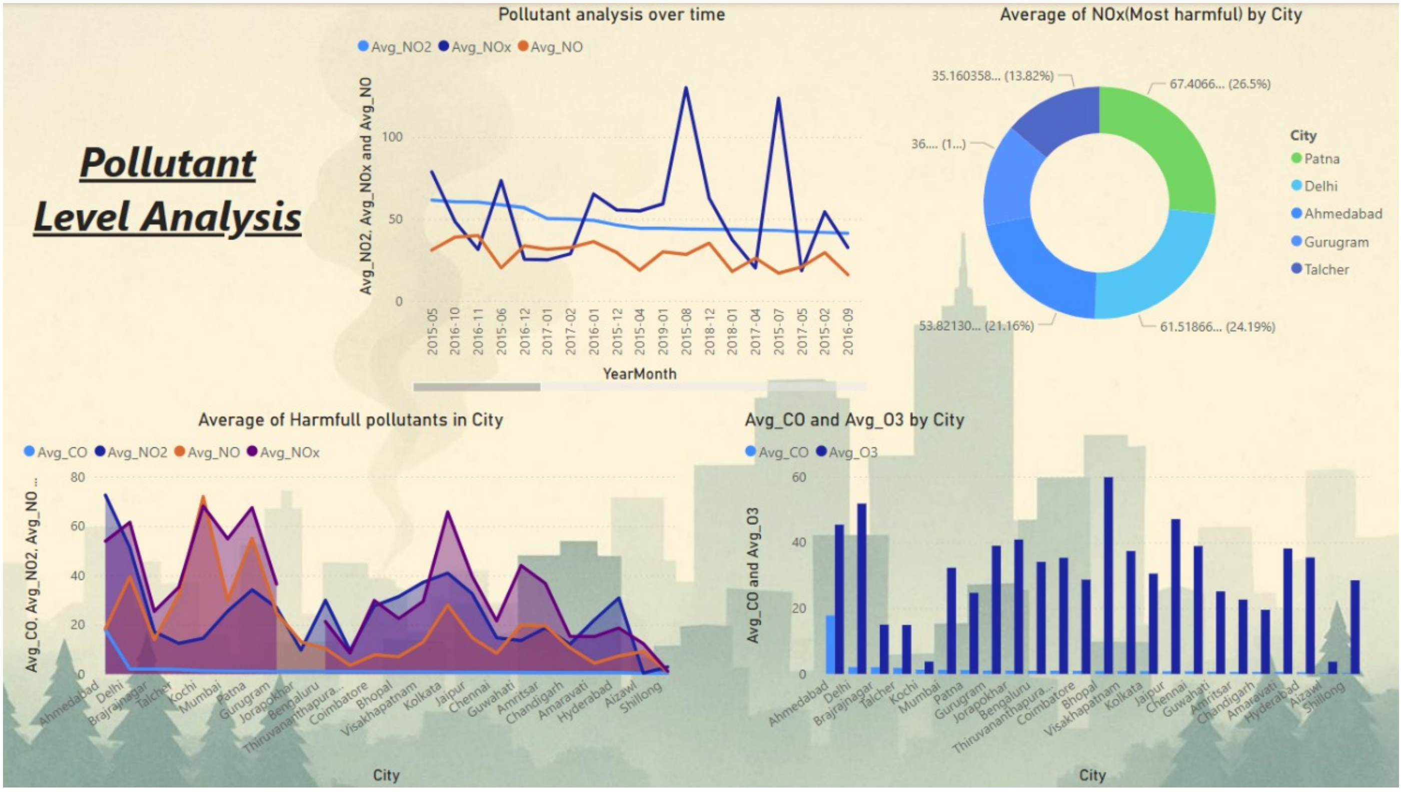


Figure A.9: Line chart showing monthly trends of NO, NO2, and NOx from 2015 to 2017.
Figure A.10: Donut chart showing average NOx levels by city, highlighting top five contributors.
Figure A.11: Stacked area chart comparing average harmful pollutant levels by city.
Figure A.12: Bar chart showing city-wise averages of CO and O3 levels.

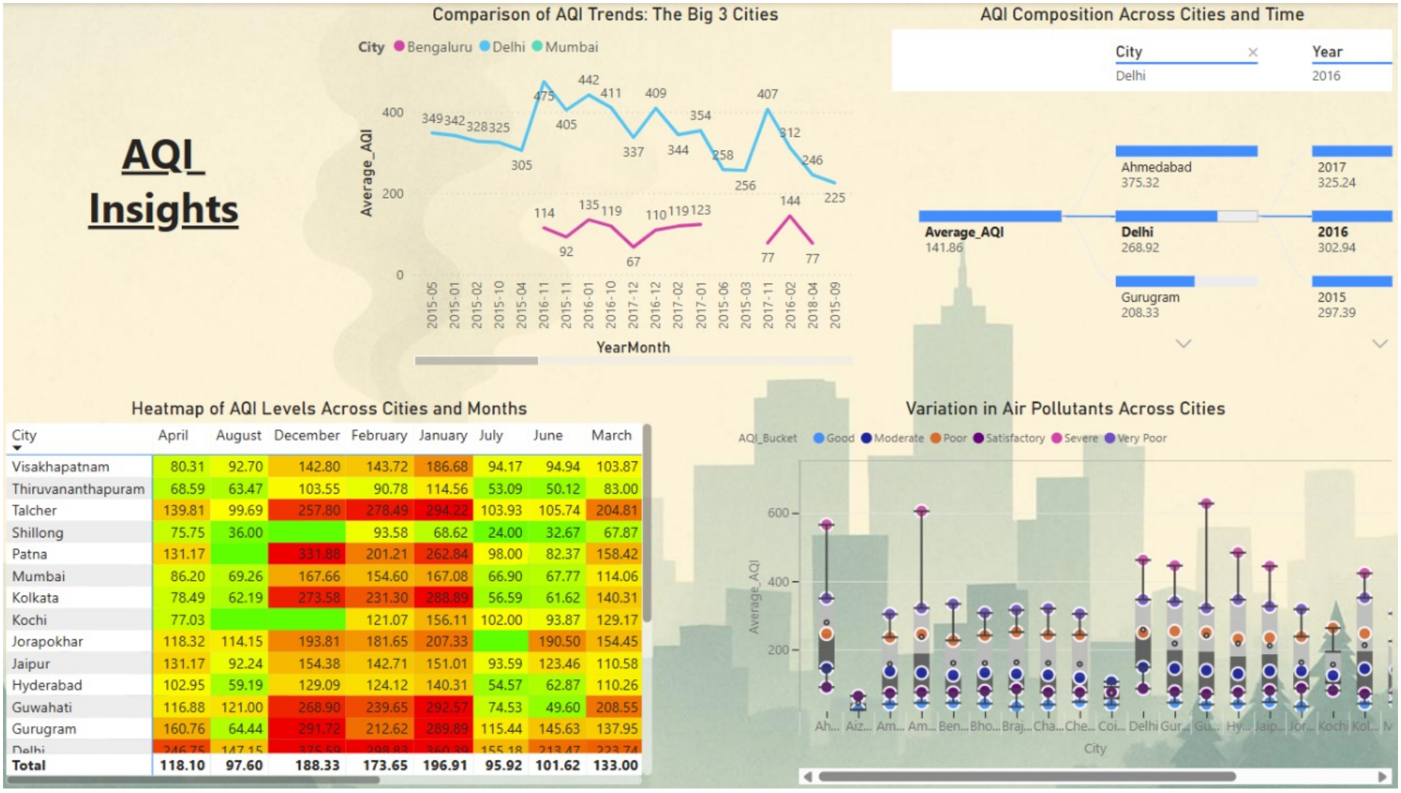


Figure A.13: Line chart comparing AQI trends in Delhi, Mumbai, and Bengaluru over time.
Figure A.14: Sankey diagram showing average AQI for Delhi, Ahmedabad, and Gurugram across years.
Figure A.15: Heatmap showing monthly AQI variation across Indian cities.
Figure A.16: Bubble bar chart showing AQI Bucket distribution across all cities.