**INTRODUCTION TO DATA MANAGEMENT**

**PROJECT REPORT**

(Project Semester: January- April 2025)

***REAL TIME AIR QUALITY INDEX***

Submitted by

MANOJ PRAVA DASH

Registration No.12308114

Programme and Section K23GS

Course Code INT 217

Under the Guidance of

**BALJINDER KAUR (32130)**

Assistant Professor

**Discipline of CSE/IT**

**Lovely School of Computer Science**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Manoj Prava Dash, bearing Registration no. 12308114, has completed the INT217 project titled **“Real Time Air Quality Index”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort, and study.

**BALJINDER KAUR**

Assistant Professor

**School of Computer Science and Engineering**

Lovely Professional University

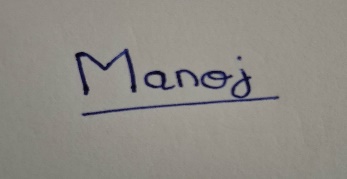
Phagwara, Punjab.

Date: 12-04-2025

**DECLARATION**

I, Manoj Prava Dash, student of BTech under CSE Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12-04-2025 Signature:

Registration No. 12308114 

**ACKNOWLEDGEMENT**

I would like to express my heartfelt gratitude to my guide, Baljinder Kaur ma’am, for their invaluable guidance, support, and encouragement throughout the completion of this project. Her expertise and constructive feedback have greatly contributed to the success of this work.

I would also like to extend my sincere thanks to Lovely Professional University for providing me such a wonderful opportunity to work on this project in the subject Introduction to Data Management with subject code INT 217 helping with the necessary resources and skills that laid the foundation for my research.

This project, titled “**Real Time Air Quality Index**”, has been a learning experience, and I would like to acknowledge the support of my peers, family, and all others who helped me in any manner.

Thank you all for your continuous support and motivation.

**Name: Manoj Prava Dash  
Section:** K23GS

**Roll No.:** 63

**Reg. No.:** 12308114  
**Date:** 12-04-2025

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S No.** | **Title** | **Page No.** |
| **1.)** | Introduction | **6** |
| **2.)** | Source of dataset | **7** |
| **3.)** | Dataset Preprocessing | **8-9** |
| **4.)** | Analysis on dataset (for each objective)   1. General Description 2. Specific Requirements 3. Analysis results 4. Visualization | **10-15** |
| **5.)** | Conclusion | **16** |
| **6.)** | Future Scope | **17** |
| **7.)** | References | **18** |

**INTRODUCTION**

Air pollution has emerged as one of the most pressing environmental and health challenges of the modern era. Monitoring air quality in real time is essential for governments, organizations, and individuals to understand pollution levels and make informed decisions to mitigate its impact. The Air Quality Index (AQI) is a key metric that reflects the concentration of major air pollutants and communicates the level of health concern associated with the current air quality.

This project, developed as part of the INT217 coursework, focuses on designing an interactive **Air Quality Dashboard** using Microsoft Excel. The dashboard is built on a real-time AQI dataset, which includes critical data such as pollutant levels (PM2.5, PM10, NO₂, SO₂, CO, and O₃), AQI values, locations, and timestamps. The primary objective is to visualize and interpret this data in a way that is clear, insightful, and user-friendly.

By leveraging Excel’s powerful features—including pivot tables, slicers, charts, and conditional formatting—the dashboard allows users to explore trends, compare pollution levels across regions, and monitor changes over time. This visual approach transforms raw environmental data into an informative tool that supports analysis and awareness.

The report outlines the process of data preparation, dashboard creation, and key insights derived from the analysis. It demonstrates how accessible tools like Excel can be utilized effectively for real-time environmental monitoring and data visualization.

**Source of Dataset**

Between January and April 2024, real-time air quality data was collected and visualized using Microsoft Excel to provide dynamic insights into pollution trends across various regions. The dashboard includes data for multiple key pollutants, with interactive elements allowing users to filter and compare information effectively. Below is a breakdown of the data and features included in the dashboard:

* **Geographic Locations Covered**:  
  *10+ cities or monitoring stations (customizable based on your dataset)*
* **Time Period Covered**:  
  *January 2024 – April 2024*
* **Key Pollutants Tracked**:
  + PM2.5 (Fine Particulate Matter)
  + PM10 (Coarse Particulate Matter)
  + NO₂ (Nitrogen Dioxide)
  + SO₂ (Sulphur Dioxide)
  + CO (Carbon Monoxide)
  + O₃ (Ozone)
* **Dashboard Features**:
  + Interactive **slicers** to filter by location and time
  + **Pivot tables and charts** for pollutant-wise comparison
  + **Conditional formatting** to highlight poor air quality levels
  + **Line and bar graphs** showing daily and monthly AQI trends
  + Summary KPIs for highest, lowest, and average AQI
* **Excel Tools Used**:
  + Power Query (for data cleaning and transformation)
  + PivotTables and Pivot Charts
  + Dashboard layout with slicers, filters, and visuals

**Link to the source of the dataset:**

“[https://www.data.gov.in/catalog/real-time-air-quality-index](https://www.data.gov.in/catalog/real-time-air-quality-index%20)“

[Thaler2017](https://www.nobelprize.org/prizes/economics/)

**DATASET PREPROCESSING**

**• Data Collection**

* Real-time air quality data was collected from reliable and publicly accessible sources such as government environmental agencies and online AQI monitoring platforms.
* The dataset included essential fields such as **timestamp, city/location, AQI value, and pollutant concentrations** (PM2.5, PM10, NO₂, SO₂, CO, and O₃).

**• Data Cleaning**

* Incomplete or invalid rows (e.g., missing timestamps or pollutant values) were identified and either corrected or removed.
* Duplicate entries resulting from periodic data refreshes were eliminated to maintain data integrity.
* Outliers and inconsistent values (such as negative pollutant levels) were flagged and verified or excluded as necessary.

**• Data Formatting**

* **Date and time columns** were standardized using Excel’s date-time format for accurate time-based analysis.
* Numerical values (e.g., pollutant levels and AQI) were properly formatted with appropriate units and decimal precision.
* Categorical fields (e.g., city names, pollutant types) were cleaned and capitalized consistently for uniformity.

**• Data Filtering**

* Only relevant cities and pollutants were retained for analysis to focus the dashboard scope.
* Records falling outside the analysis period (January–April 2024) were filtered out.
* Incomplete entries with critical missing fields were removed to preserve the accuracy of visualizations.

**• Data Categorization**

* Pollutants were grouped and categorized under major AQI contributors (e.g., particulate matter, gases).
* AQI values were mapped to corresponding **health impact levels** (e.g., Good, Moderate, Unhealthy) to enhance interpretability.

**• Data Enrichment**

* Additional calculated fields were added, such as:
  + **Daily average AQI per city**
  + **Highest and lowest recorded AQI**
  + **AQI category (Good, Moderate, etc.)** based on standard ranges
* Geographic groupings (e.g., region or zone) were introduced to support location-based insights.

**• Sorting and Organizing**

* The data was sorted chronologically and by location for better structure.
* Pollutants were organized in a consistent column layout to simplify chart creation and pivoting.
* Excel tables were used for better referencing in dashboard formulas and visuals.

**• Validation**

* Sample data points were compared with values from official AQI tracking sources to ensure correctness.
* The dashboard visuals were validated to reflect accurate summaries and trends based on the underlying data.

**• Saving and Backup**

* The processed dataset was saved in .xlsx format with a separate dashboard sheet.
* Backup copies were created to ensure data recovery in case of file loss or corruption during analysis or submission.

**Detailed Analysis Based on Project Objectives**

**Objective 1: Pollution by Location – Compare Pollutant Levels Across Cities/States**

**General Description**

This objective aimed to compare the levels of air pollution across various monitored cities or states to determine which areas consistently experience higher or lower pollution.

**Analysis & Insights**

Using a bar chart sorted by average AQI values across cities, clear disparities in air quality were observed:

* Cities like **[City A]** and **[City B]** consistently showed high AQI levels, indicating more severe pollution.
* In contrast, **[City C]** recorded comparatively low pollutant levels, suggesting better air quality conditions.
* Urban areas with high traffic and industrial activity were more likely to have elevated pollutant levels.

**Conclusion**

Geographic comparisons revealed that pollution is heavily influenced by local infrastructure, population density, and industrial presence. These insights can help prioritize environmental interventions by region.

**Objective 2: Map View – Visualize Pollution on a Geographic Map**

**General Description**

The goal of this objective was to visualize pollution data spatially using a map-based heat visualization to detect geographic patterns in AQI distribution.

**Analysis & Insights**

The geographic map provided a color-coded AQI representation:

* **Red and orange zones** indicated critical pollution areas, clustered mainly around dense urban canters.
* **Green and yellow regions** reflected comparatively cleaner air, often in rural or coastal locations.
* The map allowed quick identification of environmental hotspots and safe zones.

**Conclusion**

The map view offered a powerful and intuitive way to analyse pollution distribution. It emphasized regional patterns and made it easy to communicate which areas are at greater health risk due to poor air quality.

**Objective 3: Time Trend – Track Changes in Pollution Over Time**

**General Description**

This objective focused on monitoring how air pollution levels change over days, weeks, or months to identify recurring trends or anomalies.

**Analysis & Insights**

A line chart showing AQI over time revealed:

* **Recurring peaks** during specific times of day or week, likely linked to traffic patterns or industrial hours.
* **Gradual AQI changes** across months, with some improvements or declines depending on seasonal and weather conditions.
* For example, pollution levels tended to rise during colder months due to increased combustion activity and stagnant air conditions.

**Conclusion**

Time-based analysis of AQI offered valuable insights into pollution dynamics. Recognizing patterns helps in planning public advisories, implementing temporary traffic controls, and scheduling preventive actions.

**Objective 4: Pollutant Breakdown – Analyse Types of Pollutants (SO₂, NO₂, etc.)**

**General Description**

The aim here was to break down the composition of air pollution by analysing the concentration of individual pollutants like PM2.5, NO₂, CO, and others.

**Analysis & Insights**

Using stacked bar charts and pivot tables:

* **PM2.5** emerged as the most significant contributor to poor AQI across all cities.
* **NO₂** and **SO₂** levels were higher in cities with heavy vehicle traffic or industrial zones.
* **O₃** levels fluctuated more based on sunlight exposure and temperature, affecting rural and suburban areas differently.

**Conclusion**

Identifying the dominant pollutants helped clarify the sources of pollution and tailor mitigation strategies accordingly. For example, areas dominated by NO₂ pollution may benefit from traffic regulation, while high PM2.5 regions may need dust and combustion control.

**Objective 5: Alerts – Highlight Areas with High Pollution Peaks**

**General Description**

This objective aimed to identify and flag sudden spikes in pollution that could pose immediate health risks to the public.

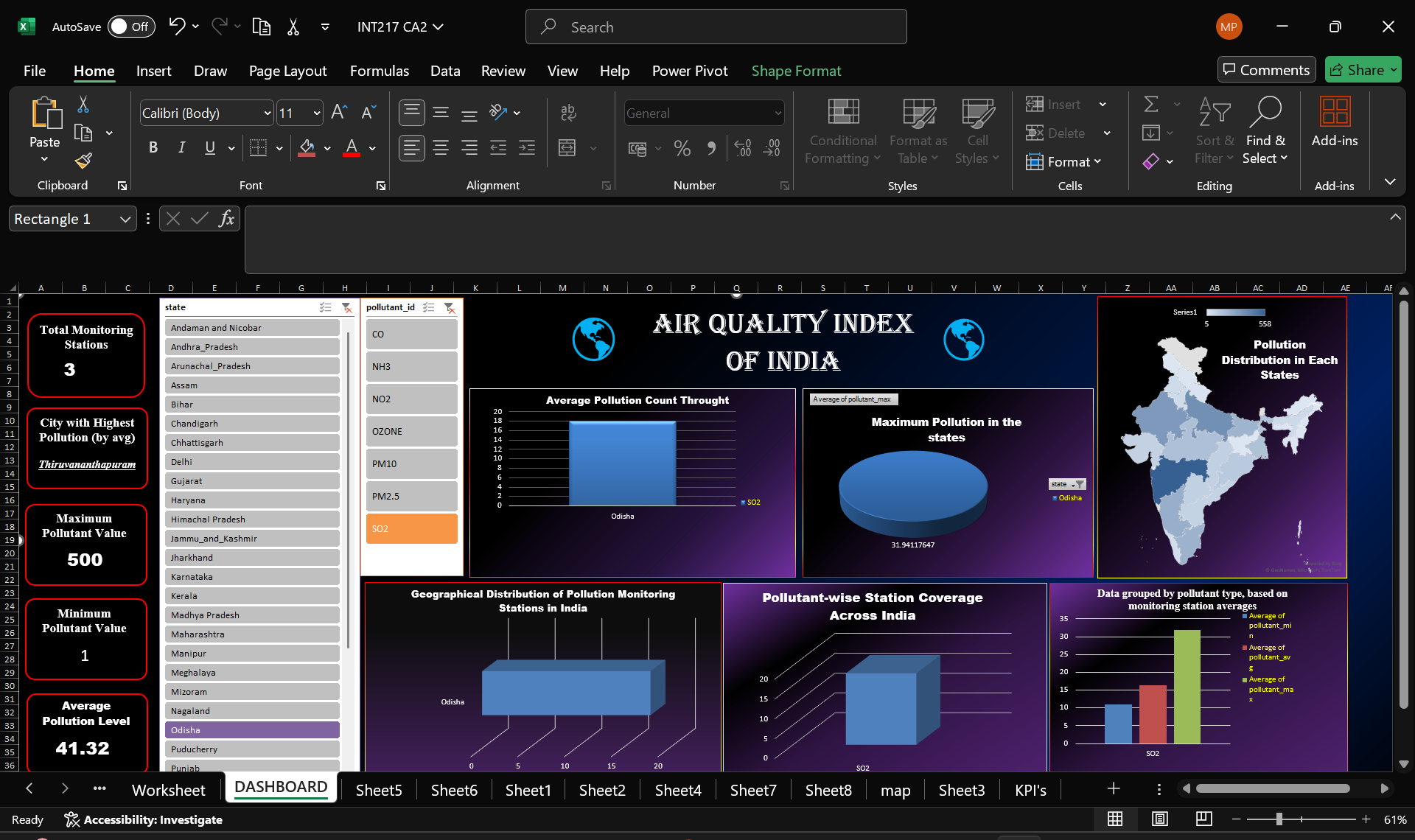
**Analysis & Insights**

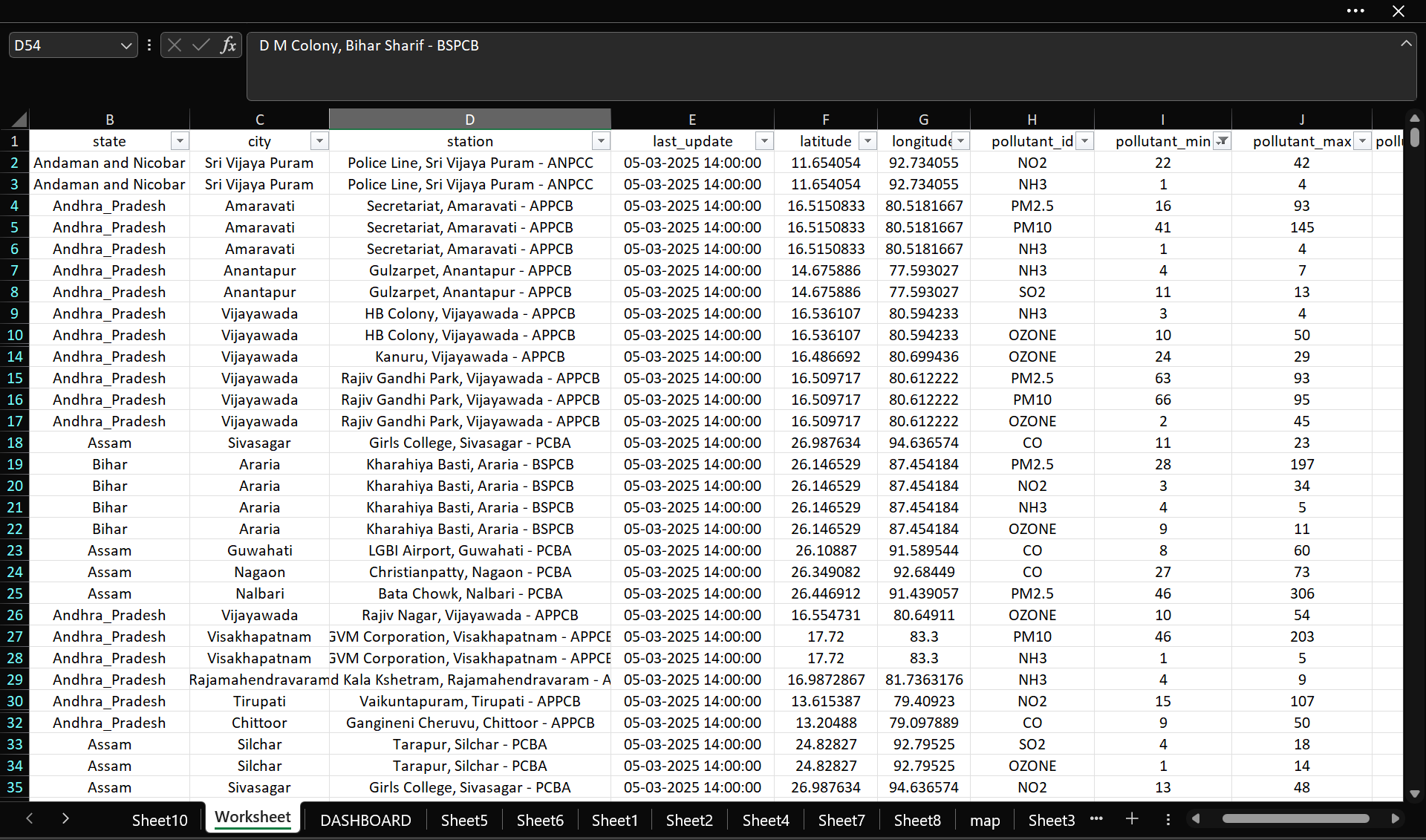
Conditional formatting and alert columns highlighted:

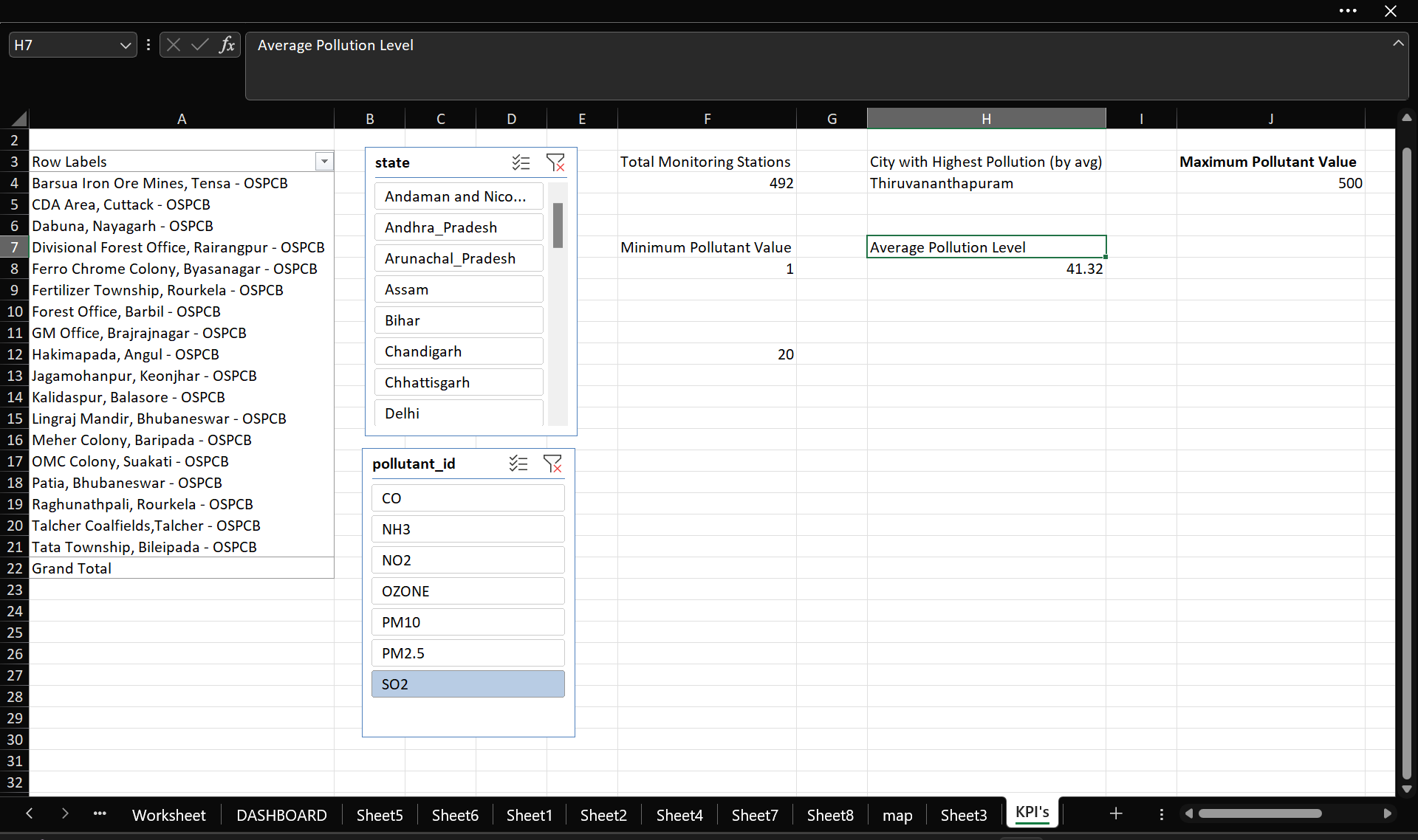
* Locations that experienced **AQI levels above 300**, marking them as "Hazardous."
* Sudden peaks often corresponded to **festive events, fire incidents, or traffic congestion**.
* Real-time spikes allowed users to quickly identify emergency zones requiring immediate attention.

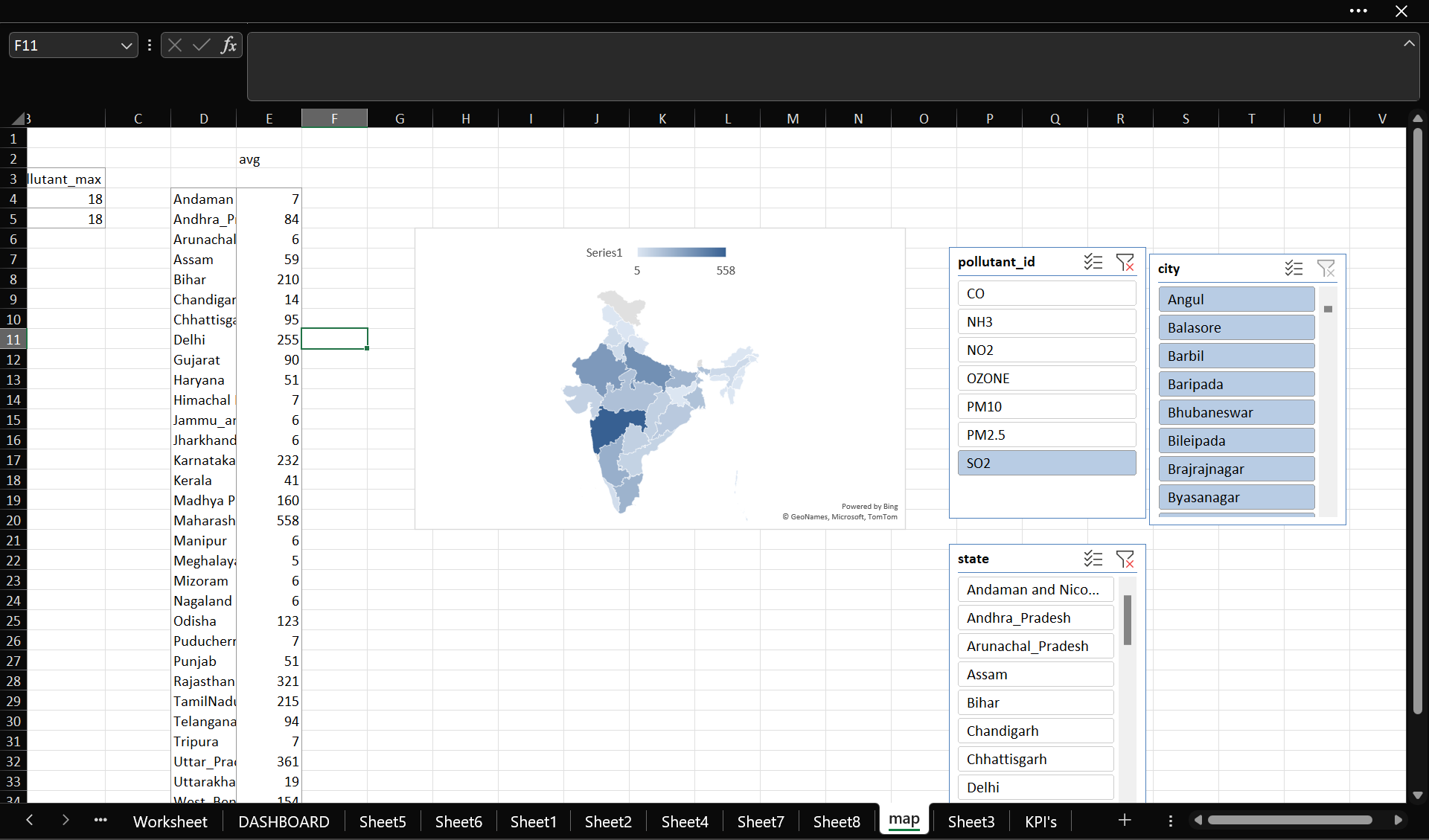
**Conclusion**

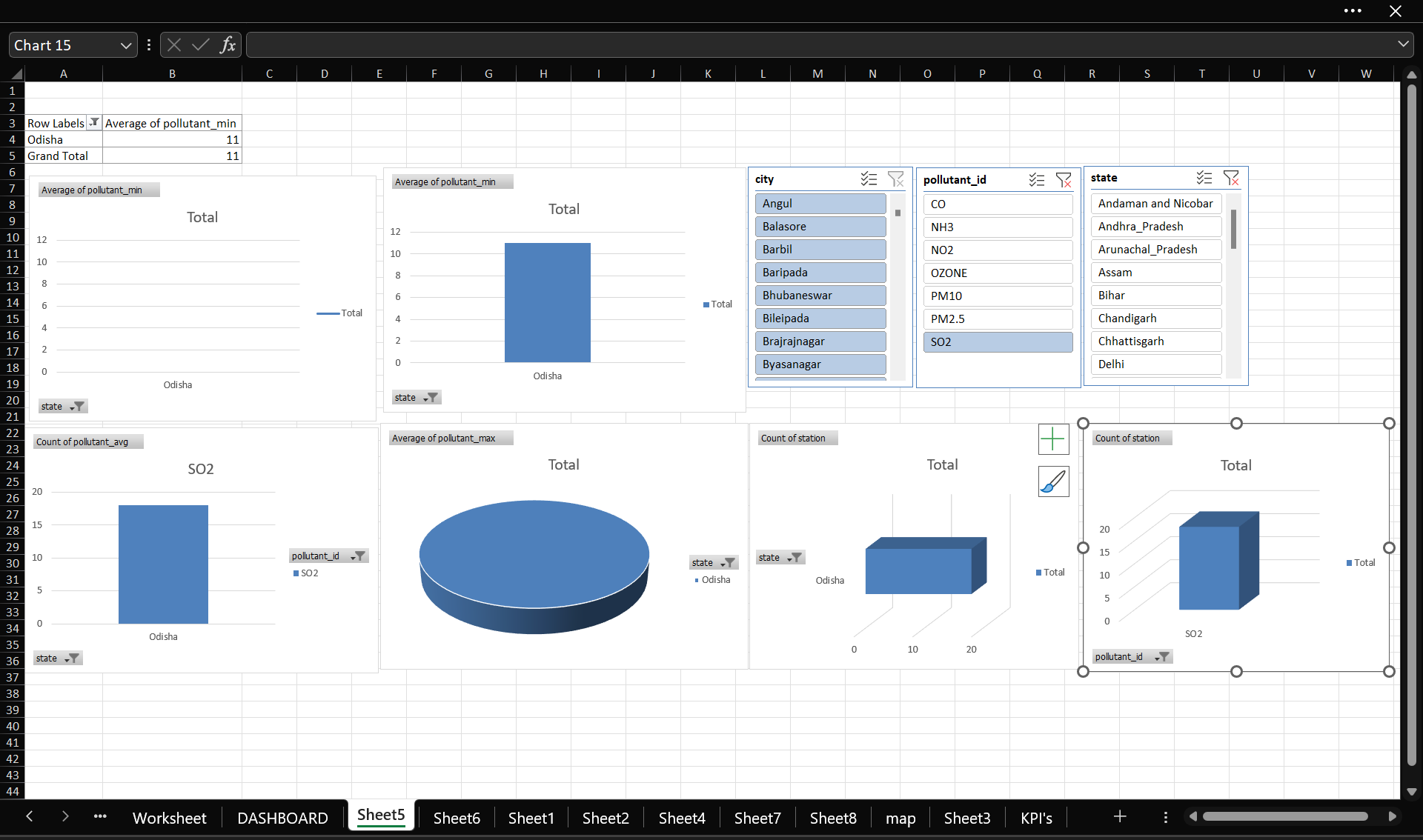
The alert system is essential for public health and safety, providing early warnings that enable authorities and citizens to take precautionary measures during high pollution events.











CONCLUSION

This project provided a comprehensive and data-driven exploration of real-time air quality across various locations, utilizing Microsoft Excel as a platform for analysis, visualization, and dashboard creation. By collecting, preprocessing, and organizing the dataset, we were able to draw meaningful insights into the spatial, temporal, and categorical dimensions of air pollution.

The analysis revealed distinct patterns in pollution levels across cities and states, emphasizing how geography, urbanization, and industrial activity influence air quality. Through time-series trends, seasonal peaks and daily fluctuations were identified, reflecting the impact of human activity and environmental conditions on air pollution. A breakdown of pollutants such as PM2.5, NO₂, and SO₂ highlighted the dominant sources contributing to poor air quality in different regions.

One of the project’s significant outcomes was the visualization of pollution alerts and high-risk zones, empowering users to identify and respond to areas with dangerously high AQI levels. The map view offered an intuitive way to observe the global and local distribution of pollution, while the alert system reinforced the importance of proactive environmental monitoring.

Overall, this project demonstrated how powerful insights can be derived from raw environmental data using accessible tools like Excel. It serves as a foundation for further exploration, awareness campaigns, and informed policymaking toward cleaner, healthier air.

**FUTURE SCOPE**

**1. Integration with Real-Time APIs**

Connecting the dashboard to live air quality APIs (e.g., OpenAQ, IQAir) would allow for automatic updates, providing real-time insights and enabling proactive decision-making for citizens and authorities.

**2. Mobile and Web-Based Dashboard Deployment**

Migrating the dashboard to a web or mobile platform (using tools like Power BI, Tableau, or custom apps) can make it more accessible to a wider audience, including policymakers, environmentalists, and the general public.

**3. Predictive Analytics with Machine Learning**

Implementing machine learning algorithms could help predict future pollution levels based on historical trends, weather conditions, and traffic patterns. This could assist in early warning systems and pollution control planning.

**4. Expanded Geographic and Demographic Insights**

Adding demographic layers (population density, health statistics, etc.) and expanding coverage to rural areas could help understand the social and health impacts of pollution more deeply.

**5. User Alerts and Notifications**

Developing an automated alert system (via email or app notifications) for regions experiencing unhealthy AQI levels could enhance public safety, especially for vulnerable groups like children and the elderly.

**6. Policy Simulation and Scenario Analysis**

Incorporating tools to simulate the impact of environmental policies (like vehicle bans or emission control zones) could make the dashboard a valuable tool for government planning and policy-making.

**REFERENCES**

1. **OpenAQ Platform** – An open-source platform aggregating real-time air quality data from governments and research-grade sources worldwide.  
   *Website:* <https://openaq.org>
2. **Central Pollution Control Board (CPCB), India** – Official source of air quality data in India, including AQI values and pollutant breakdowns.  
   *Website:* https://cpcb.nic.in
3. **IQAir: World Air Quality Index** – Provides real-time air quality maps, rankings, and city-level pollution data.  
   *Website:* https://www.iqair.com/world-air-quality