# 

**INT217: INTRODUCTION TO DATA MANAGEMENT**

**PROJECT REPORT**  (Project Semester January-April 2025)

**Air Quality Analysis in India**

Submitted by

M MOHAN SAI PRABHAS  
  
 Registration No: - 12311722  
  
 Programme and Section: K23GS  
 Course Code: INT 217  
  
  
  
 Under the Guidance of

**Baljinder Kaur (32130)**

**Discipline of CSE/IT  
  
 Lovely School of computer science and engineering**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that M Mohan Sai Prabhas bearing Registration no. 12311722 has completed INT 217 project titled, **“Air Quality Analysis in India”** 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.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of computer science and engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 12/04/2025

**DECLARATION**

I, M Mohan Sai Prabhas, student of Computer science and Engineering under CSE/IT 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. 12311722 Name: M Prabhas

**Acknowledgement**  
  
I would like to express my heartfelt gratitude to my mentor and guide, Ms. Baljinder Kaur, for her constant encouragement, guidance, and insightful feedback throughout the duration of this project. Her support played a crucial role in helping me build confidence and stay focused while exploring and designing the dashboard.

I am thankful to Lovely Professional University for offering an academic environment that fosters creativity, critical thinking, and technical excellence. The infrastructure and mentorship provided were instrumental in shaping this project.

I would also like to extend my appreciation to [data.gov.in](https://data.gov.in), the Government of India’s open data platform, for making real-time, publicly accessible datasets available. The availability of structured and reliable data significantly contributed to the success and accuracy of this project.

Lastly, I wish to thank my peers, friends, and family members for their unwavering support and motivation throughout this journey**.**

**Table of Contents**

1. Introduction
2. Source of Dataset
3. Data Processing and cleaning

4Analysis on Dataset (for all the objectives)

4.1. Introduction  
4.2. General Description  
4.3. Specific Requirements, Functions, and Formulas  
4.4. Analysis Results  
4.5. Visualization

5.Conclusion

6.Future Scope

7.References

1. **Introduction**Air pollution has become one of the most pressing environmental challenges in recent years, particularly in developing countries like India. With rapid urbanization, industrial expansion, and an increasing number of vehicles, the quality of air has been deteriorating at an alarming rate. Exposure to polluted air not only affects public health but also contributes to climate change and environmental degradation.  
   This project aims to analyze real-time air quality data collected across various Indian cities and states to uncover key insights and patterns. Using Microsoft Excel, a user-friendly and widely used tool, an interactive dashboard was developed to visually represent the distribution and intensity of pollution across different locations.  
   The primary focus of this project is to help users quickly identify the most polluted regions, dominant pollutant types, and comparative pollution levels across states and cities. The dashboard provides dynamic filtering options through slicers and dropdowns, along with semi-circle charts and clear visual storytelling to make the analysis intuitive and accessible.  
   By organizing, cleaning, and analyzing the dataset sourced from data.gov.in, this project showcases how real-world data can be transformed into a powerful analytical tool for awareness, understanding, and decision-making related to environmental conditions**.**
2. **Source Of Dataset**

The dataset used in this project was obtained from [data.gov.in](https://data.gov.in) the official open data platform of the Government of India. The platform provides access to a wide range of publicly available datasets from various government departments and organizations, supporting transparency and data-driven decision-making.

For this project, the dataset selected relates to real-time air quality measurements recorded across different cities and states in India. It includes essential fields such as:

* Country, State, and City
* Station name
* Pollutant type
* Pollutant minimum, maximum, and average values
* Geographical coordinates (longitude and latitude)
* Last update timestamp

This dataset served as the foundation for building the interactive dashboard in Microsoft Excel. It enabled region-wise and pollutant-wise analysis of air quality, allowing users to explore pollution trends and identify areas with critical pollution levels.

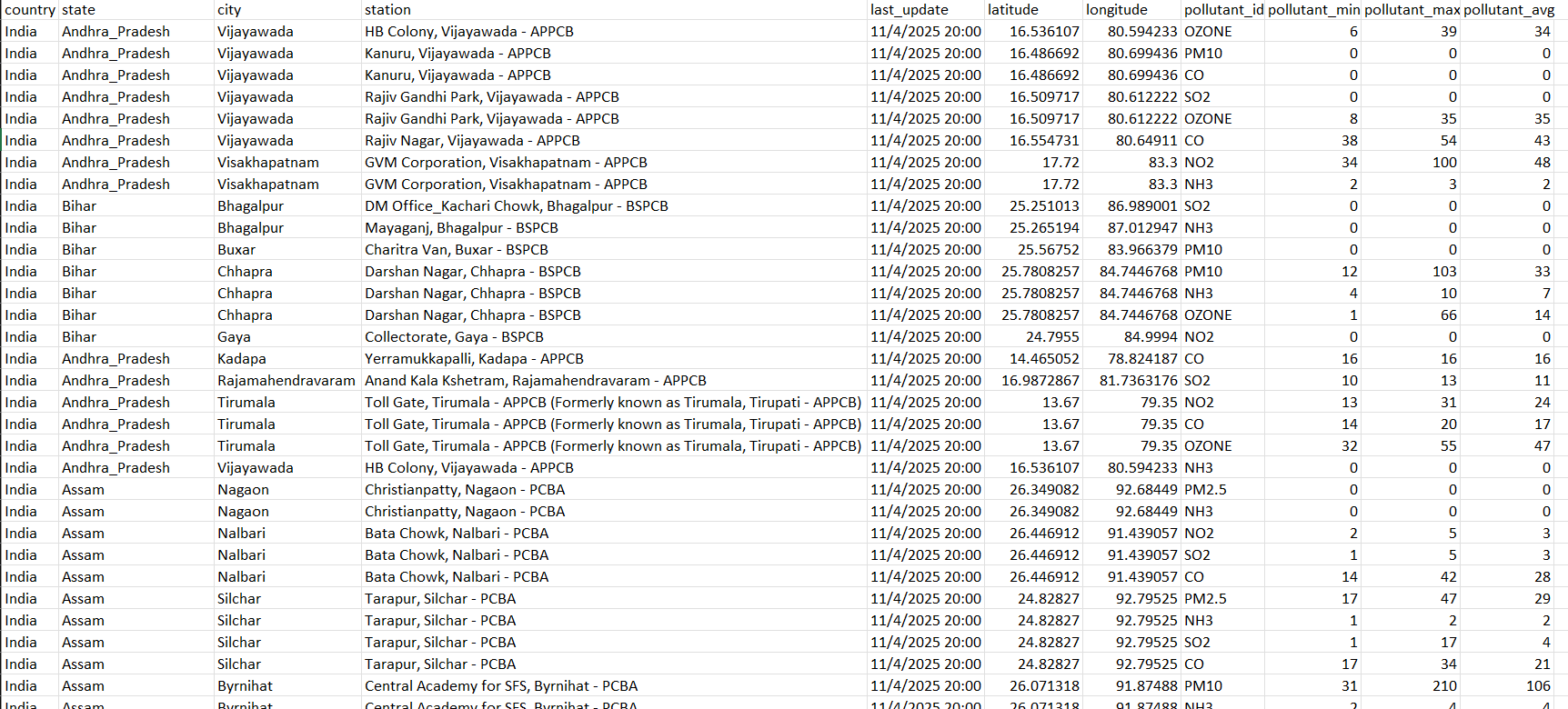
The data was downloaded in CSV format, then imported into Excel for cleaning, transformation, and visualization.

1. **Data Processing and cleaning**

Data cleaning is a crucial step in the data analysis process, ensuring that the dataset is accurate, complete, and suitable for analysis. In this project, the following data cleaning tasks were performed:

1. Conversion to Table Format: The first step was to convert the dataset into a structured table format using the Excel shortcut Ctrl + T. This table format provides several benefits, such as automatic handling of headers, the ability to sort and filter data easily, and better management of large datasets. It also enabled the use of Excel's built-in features such as data validation and formulas, which are vital for the cleaning and analysis process.
2. Handling Missing Values: After converting the data into a table, we checked for any missing values in the dataset. Missing values can skew the results of data analysis, so it’s essential to address them. In this dataset, missing values were represented as NA. To ensure consistency in the dataset, all instances of NA were replaced with 0. This step was carefully executed to prevent any incomplete records from affecting the overall analysis. Replacing NA with 0 ensures that the dataset remains complete and usable for calculations and visualizations.
3. Removal of Blank Rows: Next, we examined the dataset for any blank rows that might have been inadvertently included. Blank rows in a dataset can introduce errors or disrupt the flow of analysis. In this case, after thoroughly reviewing the dataset, it was confirmed that there were no blank rows present. This ensured that the dataset was continuous and free from any extraneous spaces that could have interfered with data manipulation.
4. Detection and Removal of Duplicate Entries: To maintain the integrity of the dataset, we performed a check to identify any duplicate records. Duplicate entries in the dataset can lead to biased analysis and incorrect results. After conducting this check, it was confirmed that no duplicates were found in the dataset. This step was vital in ensuring that each record represented a unique data point, which is crucial for producing reliable and accurate analysis results.

By performing these essential data cleaning steps, we ensured that the dataset was accurate, complete, and ready for the next stages of analysis. The cleaned dataset now provides a reliable foundation for further exploration, visualization, and analysis.  
After the cleaning the dataset:



**4. Data Analysis (For all Objectives)**

**Objective 1: To Identify the Average of pollutant max of each pollutant India**

* 1. **Introduction**

The aim is to pinpoint the cities and states in India that experience the highest pollution levels. Identifying these locations helps in targeting pollution control measures effectively.

* 1. **General Description**

The dataset contains real-time pollutant values collected from various cities and states across India. By grouping this data based on location, a clearer picture of the worst-affected areas was achieved.

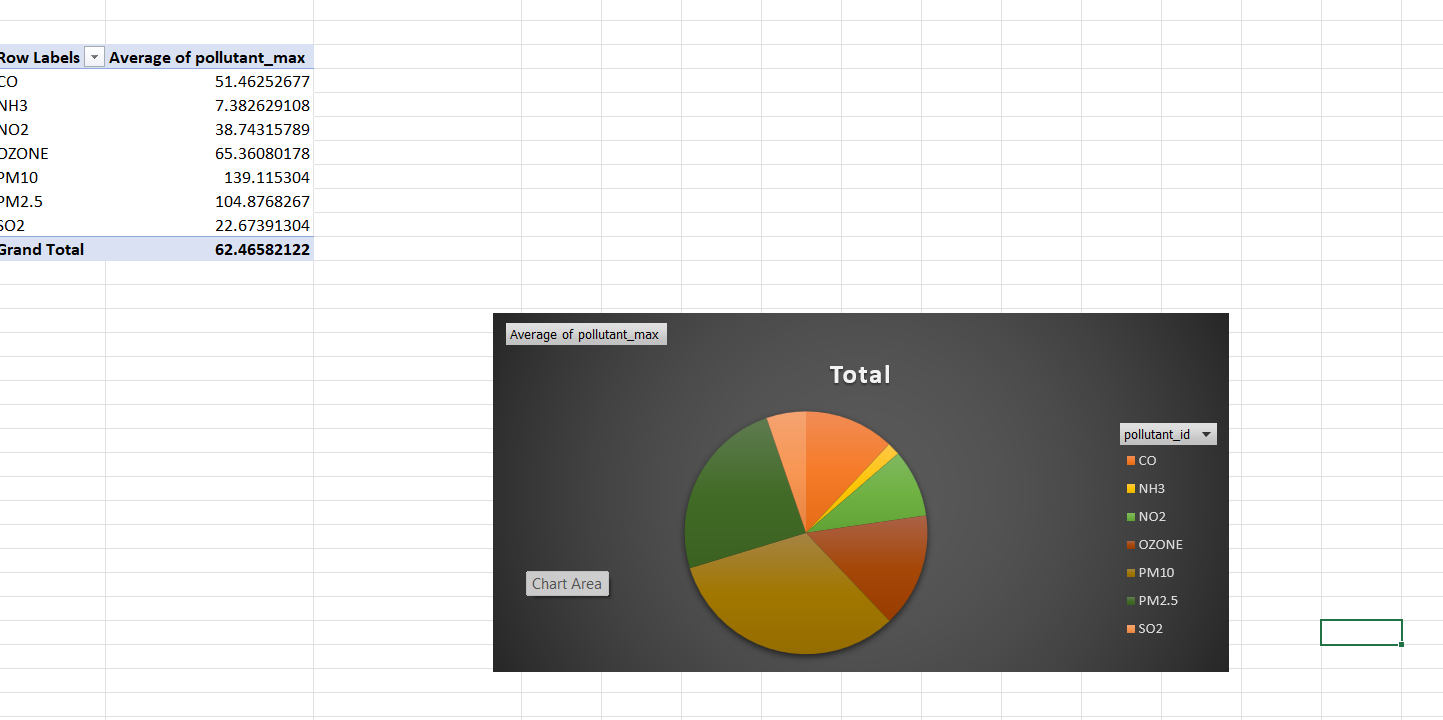
* 1. **Specific Requirements, Functions, and Formulas**

1. Used **Pivot Table** to summarize data by City and State. States in filter, city in rows and the pollutant avg on the value.
2. Applied value field settings – avg for the pollutant id.
3. Added **pollutant max** to Values section to get total values.
4. Applied **sorting** in descending order to highlight highest pollution levels
5. Inserted a **pie chart** from the Pivot Table for clear visualization.
6. Used chart formatting options to highlight the most affected locations.
7. Used slicer as well for the state and the pollutant type Here, for this objective only state slicer is used.
   1. Analysis Results

The chart revealed pollutant id with consistently high pollution values. These results highlight the geographical regions in India that face more severe air quality challenges.

* 1. Visualization

A pie chart was created to show the total pollutant id having how much of average pollutant.



**Objective 2: To Analysis Top 15 cities which is having max pollutants**

1. **Introduction**

This analysis top 15 cities which is haing max pollutants.

1. **General Description**

The dataset contains pollutant readings (PM2.5, NO₂, CO, etc.) across different cities and states. Averages were calculated to measure typical pollution levels and compare pollutants.

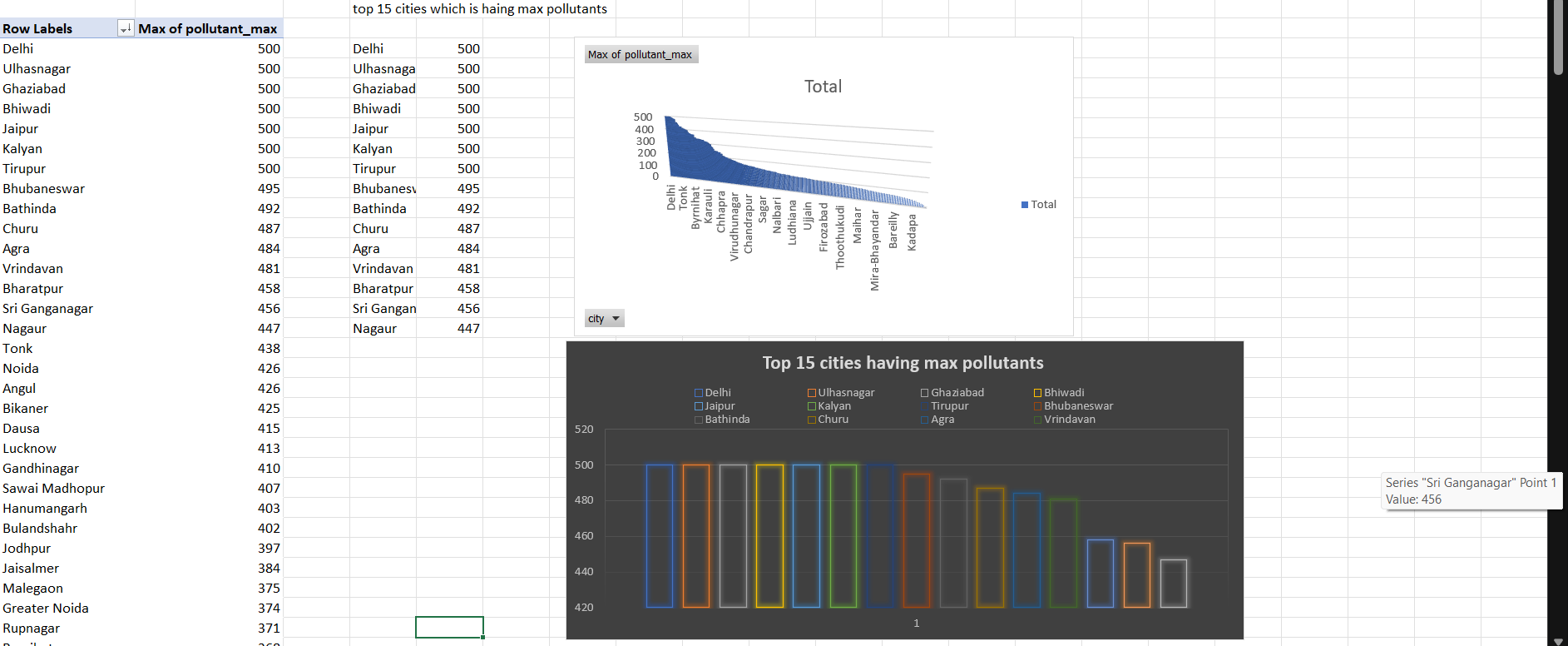
1. **Specific Requirements, Functions, and Formulas**
2. Created a **Pivot Table** with Pollutant Types in Row Labels. State in rows, city filtered and the pollutant average in value.
3. Set the Values field to **"Average"** for each pollutant.
4. Applied **Value Field Settings → Average** to get mean readings.
5. Created a **BOX Chart** from this Pivot Table to show pollutant averages.
6. Used different line markers and colors to represent each pollutant type.
7. **Analysis Results**

The line chart showed which pollutant types have higher average values, helping to understand which pollutants contribute more to poor air quality.

1. **Visualization**

A line chart was developed to display the average levels of various pollutants across all data points, showing the severity and differences between pollutant types.

**Objective 3: To Compare the Count of max and min of Different Pollutant Id**



1. **Introduction**

This analysis aimed to determine which pollutant id is having count of max and min.

1. **General Description**

The dataset records various pollutants such as PM2.5, CO, SO₂, NO₂, etc. Identifying which of these pollutants appear most often helps to understand the major contributors to air pollution.

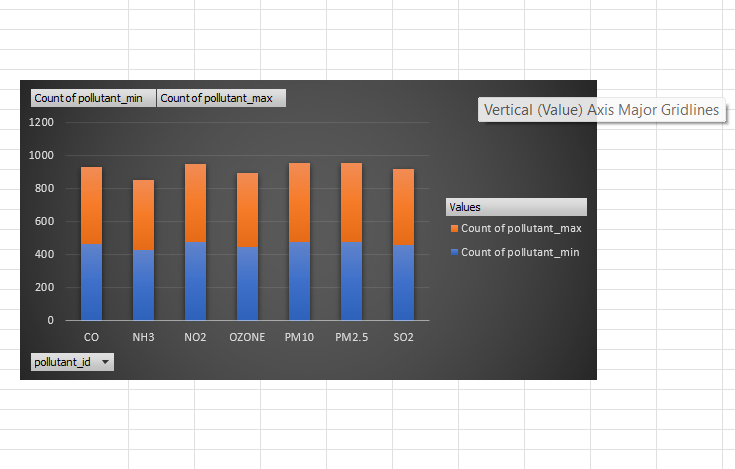
1. **Specific Requirements, Functions, and Formulas**
2. Built a **Pivot Table** using Pollutant Types in Row Labels.
3. Set the Values to **"Count"** to calculate how frequently each pollutant appears.
4. Sorted the data in descending order to highlight the most common pollutants.
5. Inserted a **Bar Chart** to display the proportion of each pollutant type.
6. **Converted values to percentage** within the chart for better understanding of distribution.
7. Enabled **data labels** to show the exact percentage on each donut slice.
8. Applied **colour formatting** for each pollutant type to enhance clarity and visual distinction.
9. And also the slicer for pollutant type.
10. **Analysis Results**

The result showed a few pollutant is having max and min Indian air quality. The most pollutant type in India.

1. **Visualization**

A bar chart was used to show the percentage distribution of each pollutant type. It visually represented the dominance of specific pollutants.

621



**Objective 4: To Analyze for each state what was the count of pollutant id**

**Introduction**

This analysis focused on count of pollution id and min levels vary across different states at a given time.

* 1. **General Description**

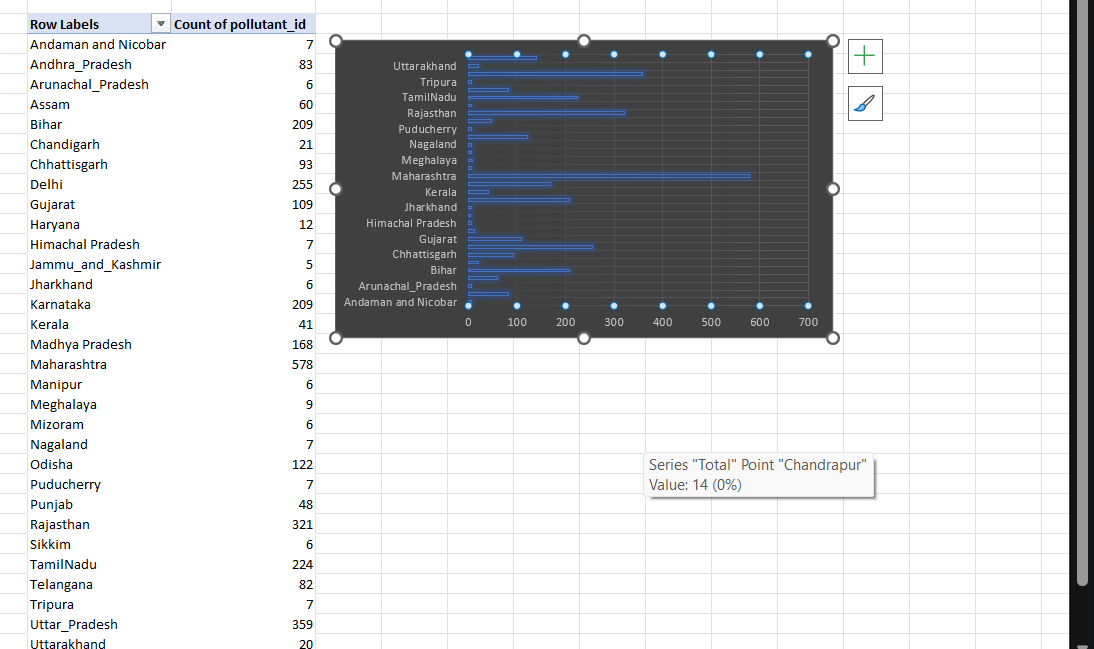
Pollution levels fluctuate across locations based on geography, traffic, industrial presence, and other factors. This analysis helps to identify such location-based variations.

* 1. **Specific Requirements, Functions, and Formulas**
  2. Created a **Pivot Table** with City/State in rows and pollutant values in columns and pollutant types in filters.
  3. Displayed **real-time data** for pollutants across different locations.
  4. Used **Clustered bar Chart** to represent pollutant variation per location.
  5. Enabled data labels and axis titles for better interpretation.
  6. Made use of slicers or filters to refine comparison between states or cities.
  7. Inserted slicers for State and Pollutant Type to allow dynamic filtering and easier comparison across different regions and pollutant categories.
  8. Applied formatting for easy distinction between multiple bars and pollutant levels.
  9. **Analysis Results**

The column chart clearly highlighted how certain locations showed much higher pollutant levels compared to others at the same time, indicating location-based pollution trends.

* 1. **Visualization**

A chart was created to count the pollutant id of cities and states using real-time data. It allowed easy visual comparison of pollution across locations.



**Objective 5: To Examine the Relationship Between cities and count of Pollutant max**

1. **Introduction**

The objective was to find patterns or relationships between specific locations and the count pollutants most commonly present there.

1. **General Description**

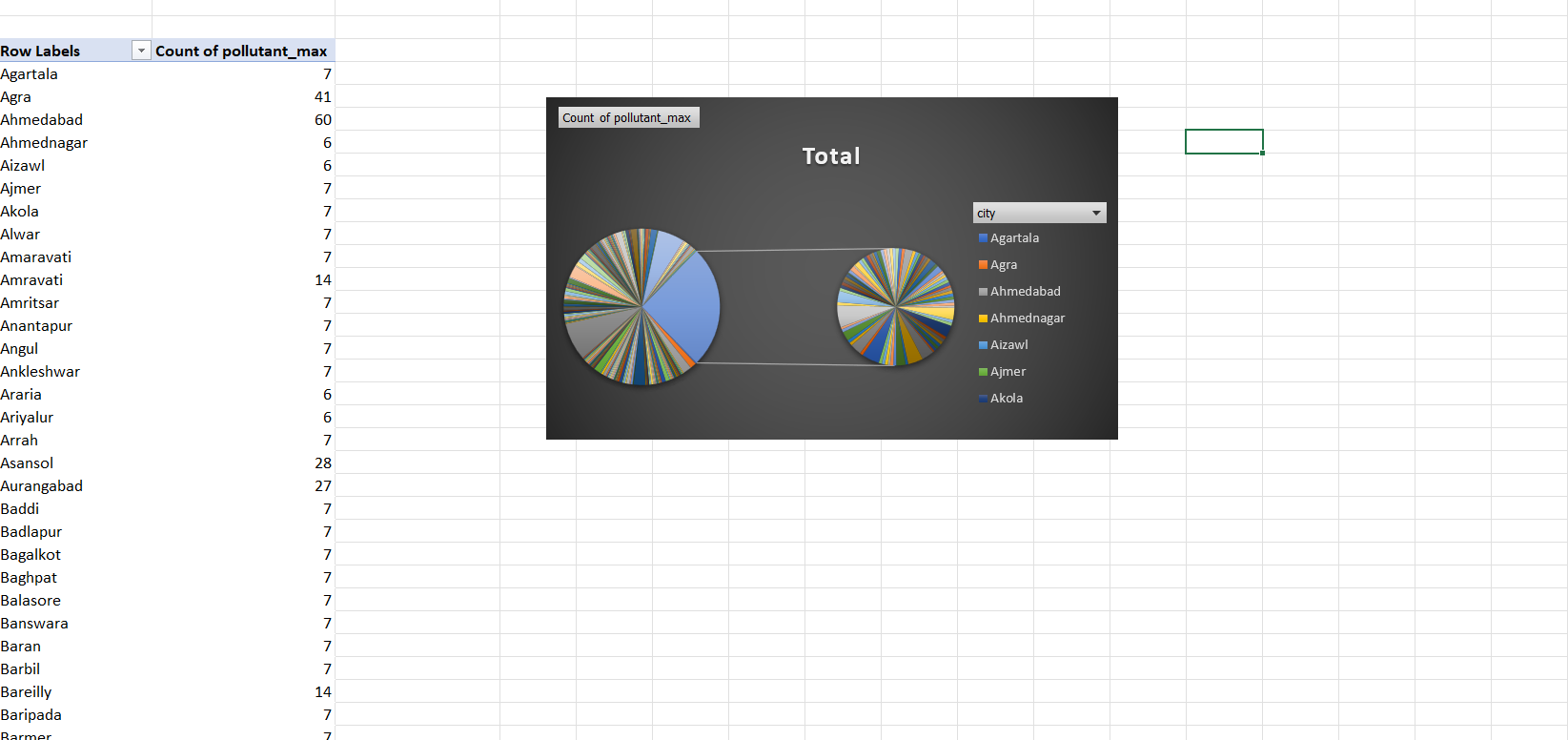
The dataset links locations with pollutant types. Analyzing this relationship helps to understand whether certain pollutants are more prevalent in specific regions.

1. **Specific Requirements, Functions, and Formulas**
2. Used a Pivot Table with Pollutant Types in columns and State in columns.
3. Values set to count to measure pollutant max presence by state.
4. Inserted a pie of pie Chart to display regional differences in pollutant type.
5. Chart formatting applied for clearer comparison between multiple pollutants.
6. The state slicer will show the bar chart according to state.
7. Enabled legends and gridlines to make interpretation more effective.
8. **Analysis Results**

The chart revealed that some states are more affected by certain pollutants than others, showing geographic patterns in pollutant presence. With the help of the slicer, it will show particular state pollutant types variation in bar chart format.

1. **Visualization**

A bar chart was created to showcase which pollutant types are dominant in specific states or cities, revealing regional patterns in pollution sources.  
  
  
**Objective 6: Interactive Dashboard Components and Real-Time Insights**



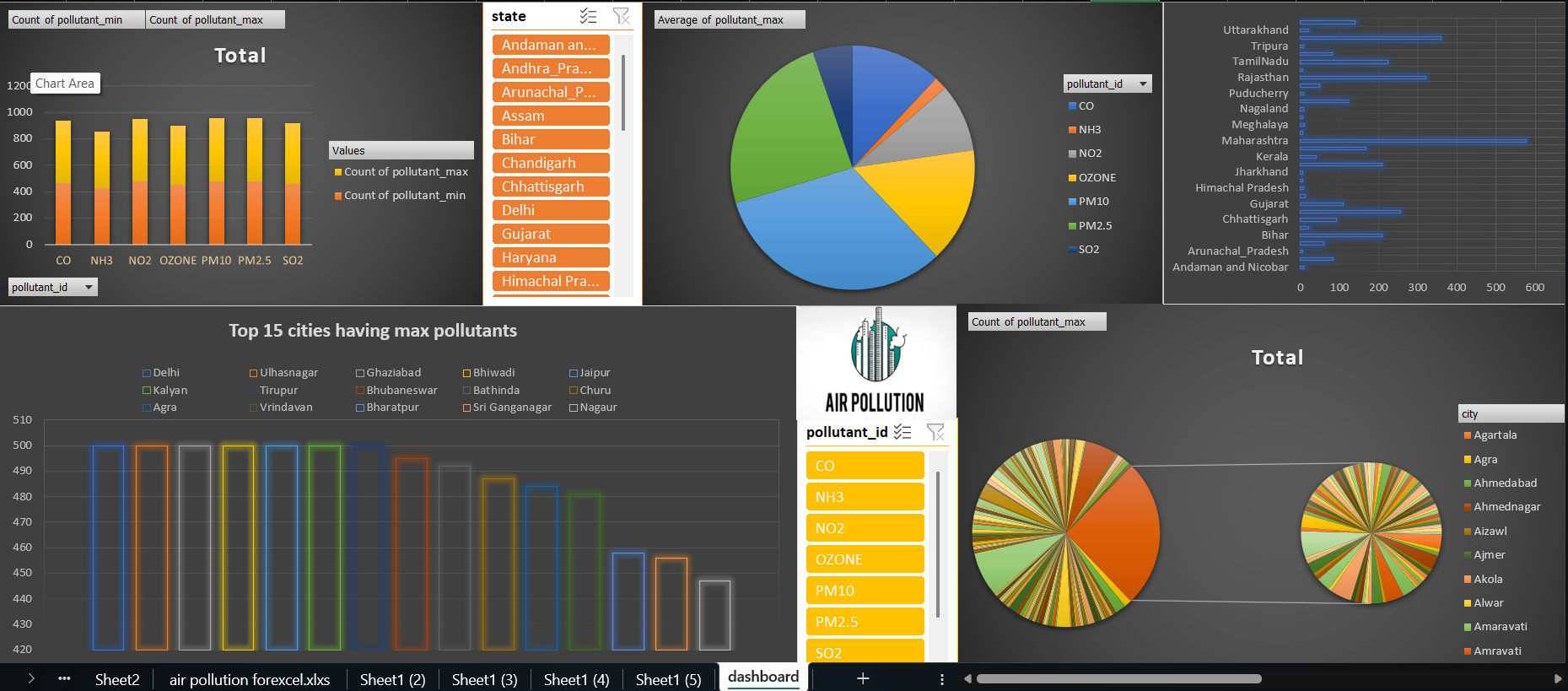
1. **General Description**

This section includes interactive features that allow users to explore the most polluted cities, the dominant pollutant type, and assess real-time pollution risk levels using dropdowns and dynamic visuals.

**Analysis Results**

1. Helps users instantly understand:
2. Which city is currently the most polluted.
3. What pollutant dominates across India.
4. Pollution risk level of any selected city.

**Dashboard:**



**5. Conclusion**This project offered a comprehensive analysis of air quality across India using Microsoft Excel's powerful data handling and visualization tools. By organizing and cleaning the dataset, applying advanced Excel functions, and designing clear, interactive dashboards, we were able to transform raw data into meaningful insights.

Each objective focused on a specific aspect of air quality—from identifying the most polluted cities and states to analyzing the most frequent and severe pollutant types. The use of Pivot Tables and charts like clustered bar charts, line graphs, and donut charts helped highlight patterns, fluctuations, and dominant pollutant sources with clarity. These visualizations made it easier to detect pollution trends across locations and over time.

Additionally, interactive elements such as slicers, dropdowns, and dynamic info boxes added depth to the analysis, allowing users to filter data in real time and draw location-specific conclusions. For example, users can now instantly identify which city is currently facing the highest pollution level, understand the dominant pollutant type, and assess the health risk level based on selected cities.

In conclusion, this report not only reflects technical skills in Excel-based data analysis but also demonstrates how data can drive awareness and action. The insights generated here can assist policymakers, researchers, and concerned citizens in making informed decisions toward improving air quality and promoting environmental health in India.  
**6. Future Scope**  
  
While this project successfully delivers valuable insights into India’s air quality using Excel, there are several ways it can be enhanced and expanded in future work:

1. Integration with Live Data Sources: By connecting to APIs or real-time data feeds, the dashboard can be made dynamic, updating pollution data automatically for up-to-date monitoring.
2. Advanced Predictive Analytics: Machine learning models can be integrated in Python or R to forecast pollution trends and predict risk levels based on historical data.
3. Health Impact Analysis: Future versions can include metrics that estimate potential health impacts (e.g., asthma rates, lung disease risks) based on pollutant exposure.
4. City-wise Policy Comparison: Add data on government regulations and policies to compare their effectiveness in improving air quality.
5. Mobile-Friendly Dashboard: Creating a Power BI or web-based version of the dashboard can make it accessible on mobile and web platforms for public use.
6. User Feedback Loop: Allowing users to report local pollution levels or anomalies can help validate and enrich the dataset.
7. **Reference  
   Microsoft Excel Documentation:**  
   Microsoft Support – *Formulas, Functions, Pivot Tables, and Charts in Excel*  
   <https://support.microsoft.com/excel>  
   *(Used for building pivot tables, charts, slicers, and data formatting.)*
8. **Government & Research Data:**  
   Central Pollution Control Board (CPCB) – *Air Quality Index and Pollutant Guidelines*  
   https://cpcb.nic.in  
   *(Referred for pollutant categories and AQI interpretation.)*
9. **General Air Quality Guidelines:**  
   World Health Organization (WHO) – *Air Quality Levels and Health Impact*  
   <https://www.who.int/health-topics/air-pollution>  
   *(Consulted for understanding pollution impact levels and classifications.)*

**Linkedin Screenshot:**

