**Data Analysis Project Submission Repor**t

PROJECT TITLE: AIR QUALITY DATASET FOR TEMPORAL POLLUTION ANALYSIS (DELHI REGION)

* **Submitted By:** 
  + **Student Name: ATHUL EMMANUEL STEPHEN**
  + **Roll Number: 2461207**
  + **College Email ID: athul.emmanuel@btech.christuniversity.in**
* **Course:** e.g., Fundamental Data Analysis
* **Instructor Name:**
* **Institution:** Christ University
* **Date of Submission:** 18/08/2025

### **2. ABSTRACT**

This project focuses on analysing air quality data to understand pollution trends and their impact on the overall Air Quality Index (AQI). The primary goal is to identify the major pollutants contributing to poor air quality, study seasonal variations, and compare AQI levels during holidays with those on regular days. Microsoft Excel was used as the core tool, leveraging its PivotTables, charts, and data visualization capabilities to transform raw environmental data into meaningful insights. The analysis highlights strong correlations between AQI and pollutants like PM2.5 and PM10, as well as seasonal patterns where winter months experience the worst air quality. The outcome is a structured, interactive dashboard and an analytical question–solutions sheet, both of which provide valuable insights for environmental monitoring, policymaking, and public awareness. This makes the project useful for stakeholders seeking to track, understand, and improve air quality management.

**3. OBJECTIVES**

The primary objectives of this project are:

1. **Clean and prepare the raw dataset** for analysis to ensure accuracy and consistency in results.
2. **Formulate and answer five key analytical questions** based on the dataset to extract meaningful insights.
3. **Create a comprehensive, user-friendly dashboard** in Microsoft Excel that highlights critical air quality metrics.
4. **Leverage appropriate charts, PivotTables, and graphs** to visualize trends, correlations, and seasonal variations effectively.
5. **Summarize the findings and their real-world implications** in a clear and concise manner, with a focus on supporting decision-making for environmental monitoring and policy improvement.

**4. Scope of the Project**

The scope of this project is limited to analyzing and visualizing the provided air quality dataset using Microsoft Excel. The project focuses on data cleaning, preparation, exploratory analysis, and visualization, ensuring that the dataset is structured and suitable for meaningful insights. The work does not involve programming languages such as Python or R, nor does it apply advanced statistical or machine learning models. Instead, the entire analysis is performed within a single Excel file, making use of PivotTables, charts, and dashboards to communicate findings effectively. The analysis is restricted to the given dataset and does not incorporate external data sources. This defined scope ensures the project remains concise, practical, and aligned to deliver a clear, interactive dashboard along with a question–answer analysis sheet for decision-making.

### **Tools & Technologies Used**

| **Tool/Technology** | **Purpose** |
| --- | --- |
| **Microsoft Excel** | **Data cleaning, manipulation, analysis, and dashboard creation** |
| **PivotTables** | **Summarizing, grouping, and comparing key metrics efficiently** |
| **Charts & Graphs** | **Visualizing AQI trends, pollutant contributions, and seasonal variations** |

**Data Cleaning & Preparation**

The dataset initially contained 1461 rows and 12 columns, representing daily air quality measurements with attributes such as pollutant concentrations (PM2.5, PM10, NO2, SO2, CO, Ozone), AQI values, date, month, year, and holiday counts. The data was generally well-structured, but a few checks and preparations were necessary before analysis.

Steps Taken:

* Checked for missing values: Verified that there were no null entries in key fields like AQI and pollutant concentrations.
* Checked for duplicates: Ensured there were no duplicate records for the same date.
* Validated data types: Confirmed that numerical values (e.g., AQI, PM2.5) were stored correctly as numbers, while Date, Month, and Year columns were consistent.
* Created calculated fields: Derived fields such as *Holiday vs Non-Holiday indicator* (based on Holidays\_Count) and combined *Year-Month* for time-series analysis.

These cleaning and preparation steps ensured that the dataset was reliable, structured, and ready for building PivotTables and dashboards in Excel.

**Dashboard Design Strategy**

The dashboard was designed to be intuitive, interactive, and visually clear, enabling quick insights into air quality trends and pollutant behavior. The layout is divided into four main sections:

1. Key Metrics (KPI Cards): At the top, summary cards display the Average AQI, Highest AQI, and Lowest AQI values to provide an at-a-glance understanding of overall air quality.
2. Trend Analysis: A line chart was used to show the AQI trend over time (by Year and Month) since it effectively captures seasonal fluctuations and long-term patterns.
3. Pollutant Contribution: A bar chart compares average levels of pollutants (PM2.5, PM10, NO2, SO2, CO, and Ozone) to highlight which pollutants dominate air quality variation.
4. Comparisons & Seasonal View: A clustered column chart shows monthly AQI averages across years, while a holiday vs non-holiday comparison chart demonstrates the effect of reduced activity days.

Interactive Elements:

* Slicers/Filters were added for Year, Month, and Holidays\_Count, allowing users to dynamically filter the dashboard and focus on specific time periods or conditions.

This design ensures that both high-level insights and detailed comparisons are accessible in one unified view.

**Questions & Solutions**

**● Question 1: What is the trend of AQI across different months?**

* **Analysis:** I grouped the data by *Year-Month* and calculated the average AQI for each period. This showed how AQI levels fluctuated month to month over the years.
* **Solution:** The monthly trend chart indicates that AQI levels peaked in **winter months (Nov–Jan)** and were comparatively lower during **monsoon months (Jun–Aug)**. This seasonal trend suggests weather conditions strongly influence air pollution.

**● Question 2: Which pollutant has the strongest impact on AQI?**

* **Analysis:** I computed the correlation between AQI and each pollutant (PM2.5, PM10, NO2, SO2, CO, Ozone). The closer the correlation is to **+1**, the stronger the relationship.
* **Solution:** The analysis shows **PM2.5** and **PM10** have the **strongest positive correlations** with AQI. This indicates particulate matter is the primary driver of poor air quality, and controlling these pollutants would have the largest effect on AQI reduction.

**● Question 3: Does the number of holidays in a month affect air quality?**

* **Analysis:** I compared the average AQI values across months with different numbers of holidays. If industrial and traffic activity reduces during holidays, we should see lower AQI.
* **Solution:** The results show that months with **more holidays generally had slightly lower AQI** compared to months with fewer holidays. This supports the idea that reduced human activity during holidays contributes to better air quality.

**● Question 4: Which year recorded the worst air quality?**

* **Analysis:** I calculated the average AQI for each year in the dataset to identify long-term patterns.
* **Solution:** The analysis reveals that **2021 recorded the highest average AQI** compared to subsequent years. This highlights a peak in pollution during that period, possibly linked to post-pandemic industrial and traffic activity increases.

**● Question 5: On which day of the week is AQI generally the highest?**

* **Analysis:** I grouped data by *day of the week* and calculated average AQI for each. This helps detect weekly cycles linked to traffic and industrial schedules.
* **Solution:** The results indicate that **Fridays and Mondays typically have the highest AQI levels**, while Sundays tend to have the lowest. This pattern reflects higher weekday activity and lower emissions during weekends.

**Challenges Faced & Solutions**

| **Challenge** | **Solution** |
| --- | --- |
| **Challenge 1: Large dataset with multiple pollutants and time variables** | **Solution:** Cleaned and structured the data by creating a proper **date column** (combining Date, Month, Year) and adding derived fields like *Day of the Week* and *Year-Month*. This made time-series analysis and trend visualization easier. |
| **Challenge 2: Identifying the most impactful pollutant on AQI** | **Solution:** Performed a **correlation analysis** between AQI and pollutants (PM2.5, PM10, NO2, SO2, CO, Ozone) to objectively determine which pollutants contributed most. |
| **Challenge 3: Choosing suitable chart types for insights** | **Solution:** Experimented with different chart types—**line chart** for monthly AQI trends, **bar chart** for yearly averages and pollutant correlations, and **column charts** for holidays and weekday patterns—to ensure clear data storytelling. |
| **Challenge 4: Displaying insights concisely for decision-making** | **Solution:** Added **KPI cards** (Average AQI, Highest AQI with date, Lowest AQI with date) on the dashboard for quick reference, making the analysis executive-friendly. |
| **Challenge 5: Ensuring results were interpretable** | **Solution:** Supplemented visualizations with a **Questions & Solutions sheet**, explaining each finding step-by-step (Question → Analysis → Solution) to provide context and actionable insights. |

**Outcome**

● Key Insights Gained

* Air Quality Index (AQI) shows a seasonal trend, with the worst levels observed in winter months (Nov–Jan) and relatively better air quality during the monsoon period (Jun–Aug).
* PM2.5 and PM10 were found to have the strongest correlation with AQI, making them the most critical pollutants to control.
* Holiday periods tend to have slightly lower AQI values, highlighting the impact of reduced human and industrial activity on pollution.
* Yearly analysis revealed that 2021 recorded the highest average AQI, while later years showed slight improvements.
* Weekly analysis showed Fridays and Mondays had the highest AQI, while Sundays recorded the lowest, reflecting human activity cycles.

● Usefulness of the Dashboard

* The dashboard provides a visual, interactive summary of air quality trends over time, pollutants’ impact, and external factors such as holidays and weekdays.
* Decision-makers and researchers can quickly identify problem periods, pollutants of concern, and behavioral patterns linked to AQI.
* KPI cards highlight extreme values, making it easier to monitor air quality performance at a glance.

● Reflection on Skills Learned & Enhanced

* Improved skills in data cleaning and transformation (merging date fields, deriving time-based variables).
* Strengthened ability to select appropriate visualizations for specific insights (line charts for trends, bar charts for comparisons, KPI cards for summaries).
* Gained hands-on experience in Excel dashboard creation, correlation analysis, and storytelling with data.
* Enhanced ability to translate raw data into actionable insights, an essential skill for academic and professional projects.

**Screenshots of Final Output**

