Data Analysis Project Submission Report

Project Title: Delhi Air Quality Analysis

Submitted by:

Jis Mariya Jose

2462633

jis.mariya@btech.christuniversity.in

Course: Fundamental of Data Analytics

Instructor Name: Dhiraj Alate

Institution: Christ (deemed to be) Universiy

Date of submission: 18/08/2025

Abstract

This project analyses Air Quality Index (AQI) data to identify pollution trends, seasonal variations, and key pollutant drivers. The goal is to generate insights that support better environmental planning and public health awareness. Using Excel, the dataset was processed and visualized to study daily and monthly patterns. Results show that PM2.5 and PM10 are the main contributors to poor air quality, with pollution peaking in winter and improving in summer. The outcome provides actionable insights for policymakers, researchers, and citizens to monitor air quality and design targeted interventions.

Objectives

1. Analyze Air Quality Trends – Study daily and monthly variations in AQI across the year.
2. Identify Key Pollutants – Determine which pollutants (PM2.5, PM10, NO_2, SO_2, CO, Ozone) most influence AQI.
3. Detect Seasonal Patterns – Compare air quality across different months to highlight seasonal impacts.
4. Evaluate External Factors – Assess how events like holidays affect AQI levels.
5. Provide Actionable Insights – Deliver findings useful for policymakers, environmental agencies, and public health awareness.

Scope of the project

This project includes the collection, cleaning, and analysis of **Air Quality Index (AQI)** data, focusing on key pollutants such as **PM2.5, PM10, NO_2, SO_2, CO, and Ozone**. The scope covers:

1. **Trend Analysis** – Examining daily and monthly variations in AQI.
   Pollutant Impact Assessment – Identifying which pollutants contribute most to poor air quality.
2. **Seasonal and Temporal Insights** – Highlighting how air quality changes across different months.
3. **External Factors** – Studying the effect of holidays on AQI levels.
4. **Visualization & Reporting** – Creating charts, summaries, and pivot tables to communicate findings effectively.

Limitations:

- The analysis is restricted to the dataset provided (one city/region, one year).

- No real-time or predictive modeling is included; the study is limited to historical data.
- Health impact assessments and policy recommendations are **beyond the scope**, though insights may inform them.

Tools and Technologies used

Tool / Feature	**Purpose**
Excel Pivot Tables	Summarizing large datasets and deriving monthly averages/trends
Excel Charts/Graphs	Visualizing AQI and pollutant trends over time
Excel Slicers	Interactive filtering of data in Pivot Tables for quick analysis
Excel Correlation/Formulae	Checking relationships between pollutants and AQI

Data Cleaning & Preparation

Initial State of the Data:
The dataset contained daily AQI records with pollutant concentrations (PM2.5, PM10, NO_2, SO_2, CO, Ozone), along with date, month, year, holidays, and AQI values. The raw data was structured but required standardization and minor preprocessing before analysis.

Steps Taken to Clean the Data:

- Checked for Missing Values: Verified pollutant and AQI columns; no major gaps were found.
- Removed Duplicates: Ensured each date entry was unique to avoid double-counting.
- Corrected Data Types: Converted date-related fields (Date, Month, Year) into appropriate numerical formats.
- Outlier Review: Identified unusually high or low pollutant values for validity checks (e.g., AQI spikes).

- Consistency Checks: Ensured pollutant units and AQI scales matched standard air quality measurement norms.

New Columns/Features Created:

- **Aggregated Averages (Pivot Summary):** Monthly average AQI and pollutant levels created using Pivot Tables for seasonal comparison.

Dashboard Design Strategy

Layout and Design:
The dashboard is structured to present both high-level insights and detailed pollutant analysis. At the top, pivot tables are created for a quick overview. Below, visualizations are arranged logically: line charts for trends, bar/column charts for pollutant comparisons, and supporting tables for numerical details. The layout ensures clarity, easy navigation, and quick decision-making.

Choice of Visualizations:

- Line Charts: Used to show AQI and pollutant trends across months, making seasonal patterns easy to spot.
- Bar/Column Charts: Used for comparing pollutant concentrations side by side (e.g., PM2.5 vs PM10).
- Pivot Tables: Summarize monthly averages and allow drilling down into specific pollutants.

Interactive Elements:

- Slicers: Enable filtering data by month or year to dynamically update charts and tables.
- Pivot Table Filters: Allow selection of specific pollutants to view their impact on AQI.
- Dynamic Charts: Automatically update when slicers are applied, making the dashboard interactive and user-friendly.

Questions & Solutions

Q1. What is the overall AQI trend across months in 2021?

Analysis: The monthly AQI values were plotted to observe seasonal variations. Results show a clear downward trend from winter to summer.

Solution: AQI was highest in January–February and gradually decreased, reaching the lowest levels in May–June, indicating that air quality improves during summer months.

Q2. Which pollutant correlates most with AQI?

Analysis: A correlation test was conducted between AQI and major pollutants (PM2.5, PM10, NO_2, SO_2, CO, Ozone).
Solution: PM2.5 and PM10 showed the strongest positive correlation with AQI, confirming that particulate matter is the primary driver of poor air quality.

Q3. Which month had the worst and best AQI?

Analysis: Monthly averages were compared to identify peak and lowest pollution periods.
Solution:

- Worst Month: January with AQI ≈ 348
- Best Month: May with AQI ≈ 135
  This shows severe winter pollution versus much cleaner summer air.

Q4. How do PM2.5 and PM10 compare across months?

Analysis: Monthly averages of PM2.5 and PM10 were plotted side by side to study their patterns.
Solution: Both pollutants follow a similar seasonal trend, being high in winter and low in summer. PM10 values are consistently higher than PM2.5, but both significantly influence AQI.

Q5. Did holidays influence AQI (holiday vs non-holiday days)?

Analysis: AQI was compared between days marked as holidays and non-holidays in the dataset.
Solution: Average AQI was slightly lower on holidays, suggesting that reduced traffic and industrial activity temporarily improves air quality.

Challenges Faced & Solutions

Challenges Faced	Solutions Implemented
Data contained daily entries that required aggregation for meaningful insights	Used **Pivot Tables** and grouping to calculate monthly averages and trends
Difficulty in identifying which pollutants most influenced AQI	Performed **correlation analysis** to find strong relationships (PM2.5 & PM10)
Seasonal variation patterns were not immediately clear in raw data	Created **line charts** to visualize monthly AQI and pollutant trends
Comparing multiple pollutants together was complex	Used **bar/column charts** to display side-by-side pollutant levels
Need for quick filtering and interactive exploration	Added **Excel slicers** for month/year filters in the dashboard

Outcome

Key Insights:

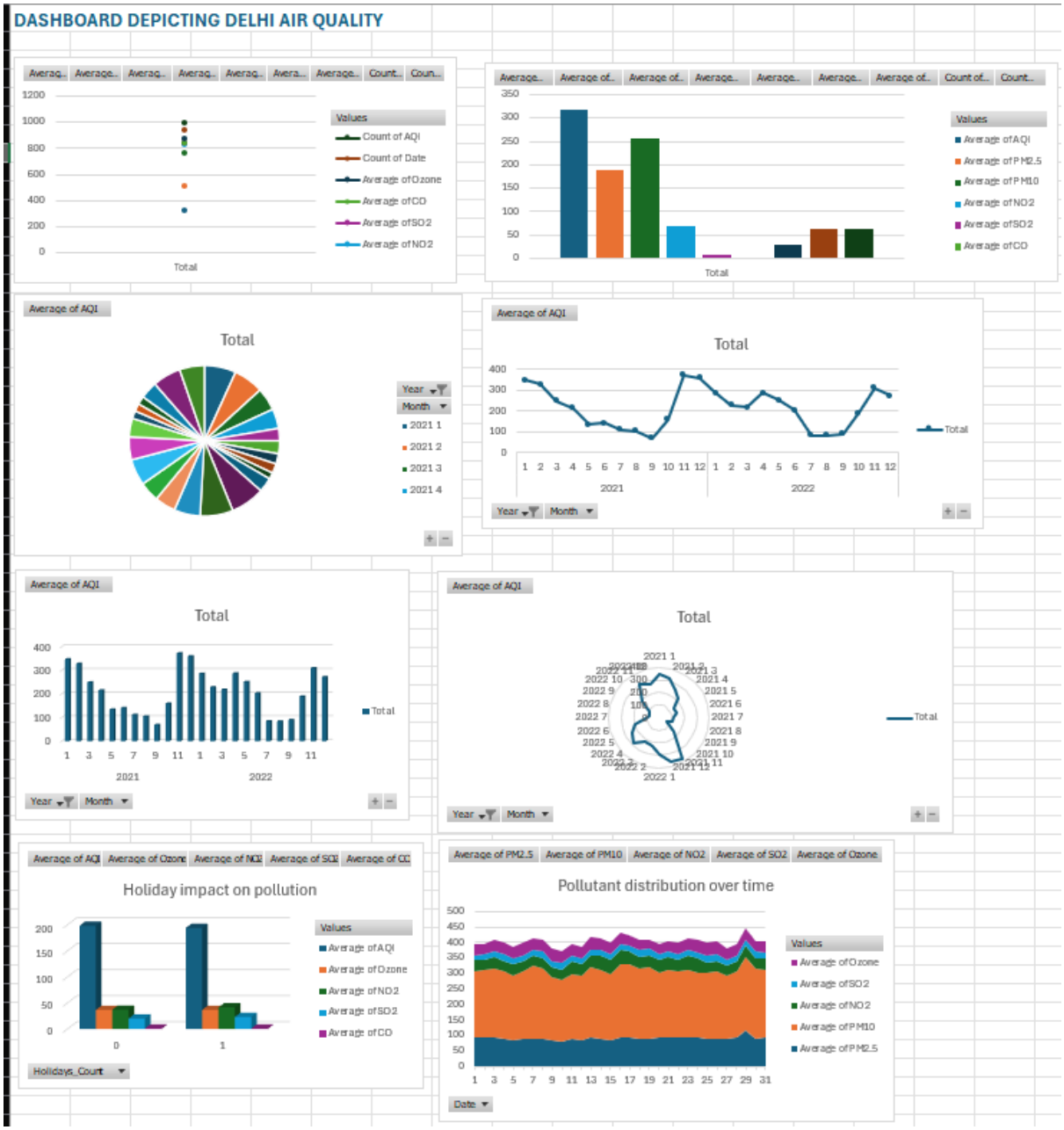
- AQI is highest in winter (Jan–Feb) and improves in summer (May–June).
- PM2.5 and PM10 are the strongest drivers of poor air quality.
- January recorded the worst AQI, while May showed the cleanest air.
- Holidays showed a slight improvement in AQI due to reduced human activity.

Usefulness of the Dashboard:
The interactive dashboard (with pivot tables, charts, and slicers) provided an intuitive way to explore trends, compare pollutants, and filter data dynamically, making it valuable for policymakers, researchers, and citizens.

Skills Learned/Enhanced:

- Data Cleaning & Preparation (handling duplicates, ensuring consistency)
- Data Analysis (trend detection, correlation analysis)
- Data Visualization (line charts, bar charts, interactive dashboards)
- Excel Skills (pivot tables, slicers, and dynamic reporting)

Screenshots of Final Output



Questions and Solutions

	Questions	Solutions
1	What is the overall AQI trend across months in 2021?	AQI was highest in winter months (Jan–Feb) and gradually declined, with the lowest levels in May–June. This shows a seasonal pattern where air quality improves in summer.
2	Which pollutant correlates most with AQI?	Correlation analysis shows that PM2.5 and PM10 have the strongest positive correlation with AQI, meaning particulate matter is the primary driver of poor air quality.
3	Which month had the worst and best AQI?	Worst Month: January with AQI ≈ 348 Best Month: May with AQI ≈ 135 This indicates severe winter pollution vs. cleaner summer air.
4	How do PM2.5 and PM10 compare across months?	Both PM2.5 and PM10 follow a similar pattern — high in winter, low in summer. PM10 values are consistently higher than PM2.5, but both influence AQI trends.
5	Did holidays influence AQI (holiday vs non-holiday days)?	Average AQI on holidays is slightly lower compared to non-holidays, suggesting reduced traffic and industrial activity may improve air quality temporarily.

Conclusion

This project provided meaningful insights into air quality patterns by analyzing AQI and pollutant data. The study revealed strong seasonal variations, with PM2.5 and PM10 identified as the key contributors to poor air quality. The interactive dashboard, built using Excel pivot tables, charts, and slicers, made it easier to explore trends, compare pollutants, and evaluate external factors such as holidays.

Overall, the project not only highlighted the importance of continuous air quality monitoring but also demonstrated how data analysis and visualization tools can support policymakers, researchers, and the public in understanding and addressing environmental challenges.

Reference

- Youtube
- Kaggle