AQI Data Analysis Dashboard

Lovely Professional University

Capstone Project Report

A Project Report Submitted in partial fulfilment of the requirements for the Award of the degree of

"Master of Computer Applications (MCA)"

By

Shriyansh Shukla

323102774

Centre for Distance and Online Education

LOVELY PROFESSIONAL UNIVERSITY PHAGWARA, PUNJAB

Year

2025

Link: https://github.com/ShriyanshShukla/AQI-Data-Analysis-Dashboard

Declaration

To whom-so-ever it may concern

I, **Shriyansh Shukla, 323102774**, hereby declare that the work done by me on "**AQI Data Analysis Dashboard**", is a record of original work for the partial fulfilment of the requirements for the award of the degree, **Master of Computer Applications (MCA)**.

Name: Shriyansh Shukla

Registration No: 323102774

Dated: 25-05-2025

CHAPTER 1

INTRODUCTION

1.1 Background and Context

Air pollution has become an enormous issue for the environment and for Indian health in the past few decades. With rapidly growing cities, increased factories, increased cars, and wood and fossil fuel burning, the air quality in most Indian cities has significantly deteriorated. Air pollution affects many areas-not only the environment-but human health and lives of millions. Most studies have shown that air pollution is related to respiratory illness, heart disease, premature death, as well as harmful effects in children and the elderly. In response, the Government of India and numerous scientific communities have taken steps to monitor, assess, and control air quality in the country.

The Central Pollution Control Board made AQI system to provide the public with a simple process for understanding air pollution. This index shows the level of pollutants and makes it very simple and categorize them from 'Good' to 'Severe' so the normal public can understand the quality of the air and take the appropriate actions. This AQI is calculated from seven key pollutants (PM2.5, PM10, NO2, SO2, CO2 O3 and NH3). Although these levels of pollutants are monitored at different monitoring stations, the AQI provides unambiguous health-oriented insight into the data. The present project initiated with the aim of developing a multi-page dashboard that demonstrates historical pollution trends of Indian cities and states between 2015 and 2020. The project integrates environmental science, data analysis, and visualization to offer useful insights through an easy-to-use interactive dashboard.

1.2 Objective of the Project

The main aim of this project is to provide a multifaceted dashboard that visualizes existing air quality data from all over India from the years 2015- 2020, and builds intuitions around pollution in various Indian cities and states using Power BI. The dashboard provides multiple views of the data, visualizing the patterns, trends, and distribution of air pollutants over time, but more than that the dashboard aims to create awareness, improve understanding, and support decision making by turning complex environmental data into informative visualizations.

A primary aim of the Project is to convert all the raw and unstructured data to structured, clean, and analyzable information. To achieve these aims the project utilizes Microsoft Excel Power Query to pull, match, and clean the datasets provided from multiple credible online sources. Specifically, this project uses a dataset that includes daily observations of air quality from 2015 - 2020 Additionally, the dataset includes air quality observations from multiple monitoring stations across India with air quality measurements for the major pollutants and a calculated AQI value and associated AQI category. The Project provides air quality data merged with appropriate geographical information so that spatial findings can be derived.

The object of this project is to analyze the cities and regions with the most, and least, pollution over time, compare pollution by state, and analyze levels of pollutants by type. For example, viewing pollution levels by year or season can provide a better understanding of the reasons behind pollution levels or measure the impact of policies over time. Additionally, this project hopes to develop a framework that can be reused or replicated in future similar projects. Although this version of the project is focused strictly on historical data analysis, the dashboard structure in this version project could be developed in the future to incorporate real-time sensors or predictive analytics features.

1.3 Scope and Applicability

Scope: The boundaries of this project are focused on a review of India's ambient air quality in the six-year time span from 2015 to 2020. It specifically examines just the seven pollutants in AQI methods used in India. Any additional gaseous compounds or fractions of particulates are not included in this work. Geographically, the project includes urban and peri-urban monitoring stations collated at the levels of cities and states. The analytical component will produce five report pages that will all focus on a different set long-term trend, ranking of cities, the categorical AQI, individual pollutants, and a map of average pollution levels. Each of the pages will include filters that allow users to focus the content.

Applicability: The dashboard is useful in many real-world situations, for government agencies and policymakers at the environmental level, the dashboard identifies pollution hotspots and identifies temporal trends that could be used to create place-based regulation or environmental policy. For example, if a city is often listed in the top most polluted cities then based on the dashboard data the government could implement city-specific interventions maybe impose vehicle bans, conduct industrial audits of pollution sources or create public awareness campaigns. For researchers and data analysts the dashboard is an organized and rich source of historical air quality data for correlation studies, environmental forecasting, and academic analysis. The data is available for six years and provides for opportunities to explore how significant events such as festivals or policy changes influenced pollution levels. Educators and students can also view the dashboard as a teaching or learning tool. The dashboard provides real-life, pragmatic opportunities for educators and students to learn about the concept of environmental science, public health, and data analysis - all of which are critical topics in interdisciplinary education today.

1.4 Significance and Relevance

This project is important because it distills a lot of raw air-quality data into simple and usable information. A lot of people cannot process all the detail presented in tables or extensive spreadsheets, so they might miss important patterns Thats where this dashboard comes and provides charts and maps to make it straightforward for anyone to understand the changes in air pollution trends.

Benefits for Everyday People: For general public just being aware of the air quality in their area can help them make good decisions and act accordingly to their health. For example, the dashboard may show that air quality in Delhi is poor in the winter months and if a family knows this so they can make the decision to use air purifiers in their home or wear masks when outdoors.

Support for Policy Makers: You are a government official or environmental agency and you want a quicker way to identify worse areas. This is one way to allow for filtering by city, month, or AQI category to view pollution hotspots within cities. If you see the city is reappearing in the Poor or Severe AQI, you can examine why the areas that people can drive in or to check if factories follow the current air rules/regulations.

Professional and Technical Skills: Whether you're learning about data analytics or data science, it's a good project for your portfolio. It demonstrates that you've completed the entire project cycle finding a data set, cleaning a dataset, and developing a working dashboard. Employers in a variety of fields business, government, health, etc. look for these skills. When you finish this project, you have proven that you have the knowledge for industry-standard tools and can produce actionable useful results on time.

CHAPTER 2

REVIEW OF LITERATURE

2.1 AQI Research Overview

The Air Quality Index is a vital instrument to rank air pollution levels and communicate them to a wider audience. The CPCB launched the National Air Quality Index (NAQI) in 2014 giving us a systematic method of measuring the air quality of regions across India. Since then, there has been a considerable body of work using AQI data to investigate the time and nature of air pollution.

- **Temporal and Spatial Analysis:** Studies have shown that seasonal variation affects air quality. It's often the case that pollution levels hit a peak in winter months because of held air conditions, temperature inversions and increased burning of biomass, a significant contributor to India's pollution bust.
- **Health Impact Studies:** Some studies have demonstrated that poor AQI is associated with a number of health conditions particularly respiratory and cardiovascular problems. These studies highlight that poor air would necessitate continuous AQI monitoring assuming people can derive some worst-case health risks.
- **Policy Formulation:** AQI data are an important resource in informing governments and policies to help mitigate pollution. For example, during high days, authorities might place restrictions on particular. Despite the years progress it continues to form a location, health, and activity risk-based figure with further development by researchers, stakeholders can only delivered greater preparedness to stakeholders, air pollution plans and data can continue to eliminate environmental challenges from decision making.

2.2 Historical Air Pollution in India

Air pollution has troubled India for many years and understanding the historical context within air pollution is essential to grasp the extent of the problem and the effort that is needed to be made to address it.

Air Pollution Recognition: Air Pollution became a widely known problem in India during the late 20th century. After understanding that polluted air could have bad health and many harmful effects, the Government of India adopted the Air (Prevention and Control of Pollution) Act in 1981 which aimed to prevent and control air pollution through the establishment of pollution control boards in a central and state governmental.

Rising Pollution Levels and Health Impacts: As ever, the only constant is change, air pollution legislation failed to control air pollution levels, which has been an increasing trend. Between 1990 and 2019, fine particulate matter (PM2.5), each particle representing natural and anthropogenic combustion, has severely increased by 115% in the hierarchy of pollution exposure in India.

Monitoring and Data Collection Efforts: To improve data assess air pollution in India, the country expanded its capacity for air quality monitoring. As of 2019, the Central Pollution Control Board and State Pollution Control Boards operated more than 750 manual monitoring stations across the country. The stations collected data on major from 2015-2019 providing useful trend information regarding pollution levels.

Government Initiatives and Programs: In response to the growing crisis, the Indian government initiated the launching of the National Clean Air Program (NCAP) in 2019 to address the air quality problems. NCAPs goal was to reduce PM pollution in the country by 20-30% by 2024 but This target was later revised to aim for a 40% reduction or meeting national standards by 2025-26.

2.3 AQI and Pollutant Basics

The AQI is a system that is originally set to measure the quality of air at a specific location, in a simpler style. It takes the complex scientific data about air pollution and condenses it to a single number, a color and a description so that the general public can understand how polluted the air is.

Key pollutants used to calculate AQI

1. Particulate Matter (PM10): They are the particles of 10 microns or less in size which is very little that's why these can enter into the lung and cause various respiratory impacts.
2. Particulate Matter (PM2.5): They are even smaller particles (2.5 microns or less) compare to PM10 that why they can enter the blood stream which will impact the lungs and heart.
3. Nitrogen Dioxide (NO2): It is a product of car emissions, industry and more, which can cause irritation while breathing and impact negatively to lung.
4. Sulphur Dioxide (SO2): This comes from when we burn the fossil fuels which can impact health and make existing heart disease even worse which is very bad.
5. Carbon Monoxide (CO): It is a colorless and odorless gas which comes from incomplete combustion so it can impact the transport of oxygen in the body.
6. Ozone (O3): It is beneficial in the upper atmosphere but at ground level causes chest coughing, throat pain and irritation.
7. Ammonia (NH3): It comes from farm activities and practices which produces eye and respiratory tract irritation.
8. Lead (Pb): It is a heavy metal related to industry and burns in products such as gasoline and paints. Lead is a well-known toxic substance specifically to the nervous system particularly in children.

AQI Categories and Health Implications

The AQI has six categories which correspond to increasing levels of health concern:

1. Good (0–50): This level of air quality is considered very normal level of pollution which have very little or no risk.
2. Satisfactory (51–100): This level of air quality is considered acceptable, but there can be some impact on a very small number of sensitive persons.
3. Moderate (101–200): This level of air quality is acceptable, however there will be a moderate health issue for a very small number of people who are very sensitive to air pollution.
4. Poor (201–300): Everyone may begin to experience health effects for members of sensitive groups may experience more serious health effects.
5. Very Poor (301–400): It's a health alert everyone may experience more serious health effects.
6. Severe (401–500): Health warnings of emergency conditions, the entire population is more likely to be affected.

Method of Calculation For the AQI:

1. Determine the concentration of each pollutant over a specified averaging time.
2. Convert the concentration of each pollutant to a sub-index using breakpoints.
3. Calculate the AQI by taking the highest sub-index among all pollutants.
4. This method ensures that the overall AQI is determined by the pollutant with the highest health risk at any point in time and that it provides a conservative indication of air quality.

2.4 Studies on AQI Impact on Health

Breathing and Heart Attack Prevention: Studies that have looked at higher levels of air pollutants (PM2.5, PM10) highlight the main exposures relating to respiratory and cardiovascular health. A short-term can exacerbate asthma and bronchitis (similar to other respiratory illness), while long-term exposure adversely impacts lung function as well led to decreased cardiac reserve.

Brain & Thinking: A large body of research has demonstrated that neurological health is also affected by air pollution in multiple studies across the world. It can enter the spinal cord and blood-brain barrier, and cognitive implications as well as neurodegenerative diseases like Alzheimer's and Parkinson's appears certainly.

High Risk Group: People have differential susceptibility of air pollution in terms of their overall health impact. Children Adolescent, elderly people and women who are pregnant as well people with chronic medical conditions. There are advisories from the National Centre for Disease Control (NCDC) in India which spell out the additional stakes for this populations.

Costs of Damage: The public health effects of air pollution are not limited to individual health status; they represent a burden of a nation. But the World Bank reported that 1.67 million deaths in India in 2019 were linked to air pollution-related health issues (17.8% of total deaths). The cost estimates above are large economic impositions that translates to a cumulative loss of ~$28.8 billion due premature death and $8 billion through morbidity.

2.5 International AQI Comparisons

India's National Air Quality Index (NAQI): It was launched in 2014 and took focus on mainly eight pollutants which are PM2.5, PM10, NO2. SO2, CO, O3, NH3, and Pb. The NAQI scales from 0 to 500 and has six levels which are Good, Satisfactory, Moderate, Poor, Very Poor, and Severe. This overall AQI value is calculated using the maximum sub-index from the eight pollutants in the location which they being measured.

United States Environmental Protection Agency (EPA): The U.S AQI is calculated by the EPA and they use a scale from 0-500 which is very same as over NAQI but they mainly focus on five pollutants like PM2.5, PM10, O3, CO, and SO2. In 2011 the U.S. set strict standards for the PM2.5 levels when making calculation for AQI. Let's take an example of PM2.5 concentration a 35.4 $\mu g/m^3$ in the USA will be a AQI value of 100.

Air Quality Index of China: Its AQI level has six main pollutants which are PM2.5, PM10, SO2, NO2, CO, and O3. The score that determines the level is from 0 – 500. As we can see they are similar to India's and the US ratings, but there are differences with the thresholds for some pollutants.

European Union Air Quality Index (EAQI): The European Union focus on the AQI which make calculation based on five pollutants (PM2.5, PM10, NO2, O3, and SO2). As mentioned before, the AQI systems in India, U.S., and China use composite scores summarized from information on several pollutants, whereas the Europe provides a separate index for each pollutant in addition to an overall score.

CHAPTER 3

IMPLEMENTATION OF PROJECT

3.1 Data Acquisition

The first stage of the project was to find a dataset that contained and accurately reflected the air quality numbers of India, which took a long time to do, however I finally found one on the well-known analytics site Kaggle. The dataset is called Air Quality Data in India, and you can find it at the following location:

Link: https://www.kaggle.com/rohanrao/air-quality-data-in-india

Reason for Choosing: With the choice of this Kaggle dataset, the data collection was simplified without the hassle of having to manually scrape data from multiple government portals or APIs. The CSV structured was well and broad so a lot of time was saved and I could focus more on cleaning and visualization. The popularity of the dataset and its documentation, further infused confidence into me for its reliability.

Files in the Dataset:

There are two main files in the dataset:

1. station_day.csv – This file is air quality data for each day and has columns including date of the report, station ID, and values for many pollutants such as PM2.5, PM10, NO2, SO2, O3, CO, NH3, etc. This file also has columns for AQI value and AQI bucket (for example, "Good", "Moderate", etc.).

2. station_names.csv – This file gives each station ID with the city and state where that station is located as this is important information to have when understanding the readings in relation to their location.

3.2 Data Cleaning and Transformation

Once I received the CSV files first, I used Power Query in Microsoft Excel to clean and transform the data so it can be used for analysis.

Importing

I imported the two files station_day.csv and station_names.csv into Power Query. The files were imported to Power Query with all columns and rows exactly the same way they were saved (unprocessed). Once the tables were opened in Power Query, I selected "Use First Row as Headers" to make the top row of the table use as the name of the columns (see Figure 3.1). This way stationid, date, and each pollutant name were properly classified.

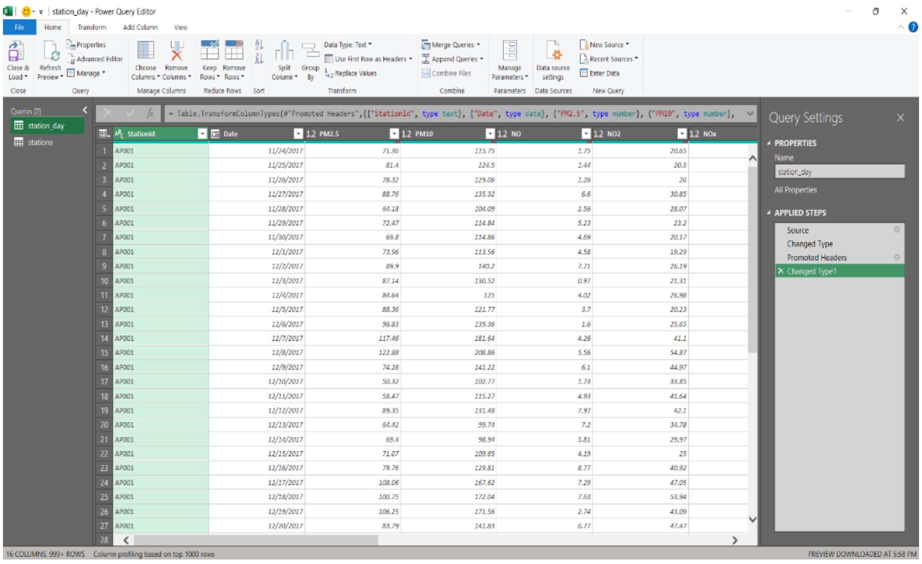


Figure 3.1: Importing AQI datasets and using first row as headers in Power Query.

Merging

Next thing in the procedure is that I needed to associate city and state information to each pollution reading. I did this by using the Merge Queries with a Left Outer Join based on the stationid column in station_day.csv and that same key in station_names.csv. After merging I expanded to only show city and state (Figure 3.2) and linked every daily measurement with a geographic location.

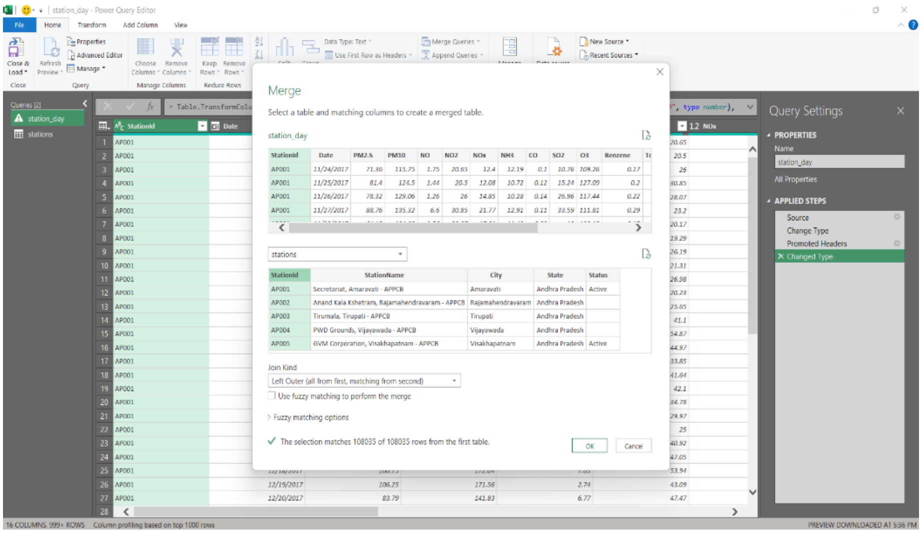


Figure 3.2: Merging datasets to attach city and state information using Left Outer Join.

Removing Columns

The merged table had more columns than were necessary for our AQI study and since the AQI formula adopted in India uses only seven pollutants, I simply deleted the following fields in a single action: stationid, no, nox, benzene, toluene, and xylene, and while I was at it, I also arranged the remaining columns in a rational order where date, the seven pollutants, AQI, AQI_bucket, city, and state are arranged beside one another (Figure 3.3).

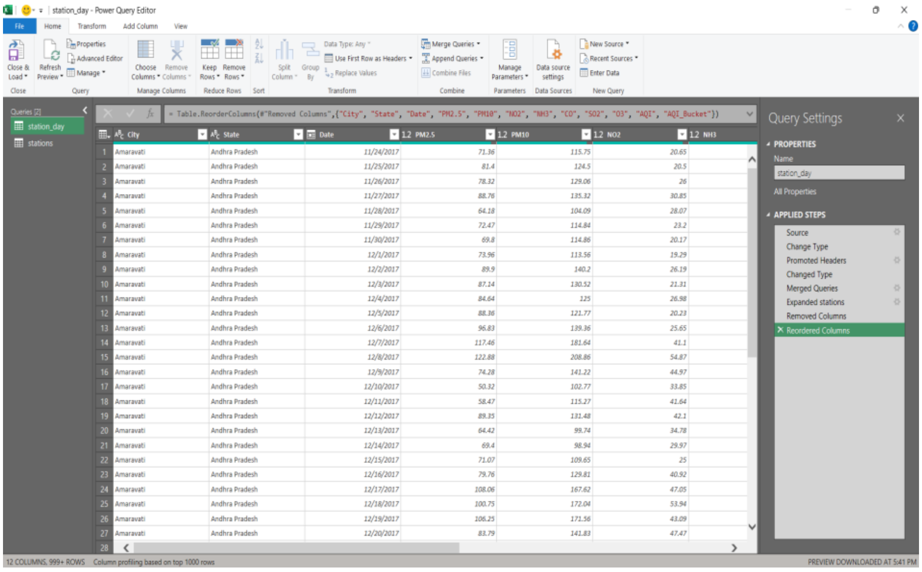


Figure 3.3: Cleaning the dataset by removing unused columns and organizing relevant fields.

Filtering and Formatting

Once I eliminated the non-relevant columns and reordered the relevant columns, a key follow-up component was to filter and format the data set to be ready for analysis in preparation for visualizing the results in Power BI. These steps were essential validity checks to ensure the data is cleaned and is both consistent, as complete, as possible, and in the right format so it can be displayed in Power BI.

Null Values: The raw data set had many rows where some values were missing for pollutants and AQI, So I filtered out rows that was missing many important columns.

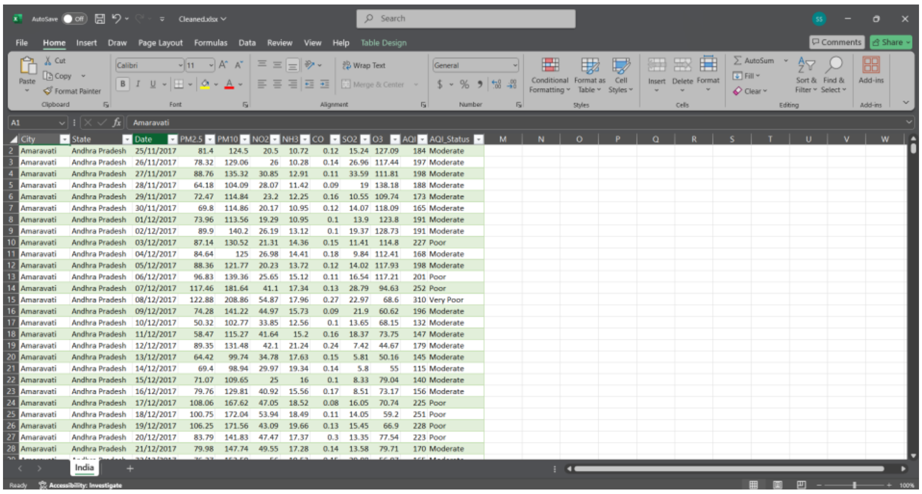
Validating Data Types: Each column was assigned the correct type for the most effective use in Power BI visuals. Numeric areas such as pollutant, and the AQI column were set to Decimal Number. The date column was assigned a Date type to allow for time-based visuals (like line graphs). Categorical areas such as city, state, and AQI_bucket were assigned Text type to allow for suitable grouping and filtering when creating dashboards.

3.3 Final Dataset Structure

The cleaned data that is undergone the necessary data clean-up and transformation steps successfully in Power Query is loaded and viewed in Excel and it was now meaningfully organized and ready for use in Power BI for visualization. A screenshot of the cleaned dataset available in Excel is provided within this section to demonstrate the generated layout of the cleaned dataset and how clearly organized information can be displayed.

This dataset which consists of 12 significant important columns which will be important for AQI review defined by Indian standards like date, PM2.5, PM10, NO2, SO2, CO, O3, NH3, AQI, AQI_bucket, city, and state and each column provides valuable information about pollutant levels and air quality in different regions over time.

This was organized by keeping only the valuable information columns for the analysis and by removing columns with extraneous information. Examples of redundant pollutants removed such as no, nox, benzene, toluene, and xylene allowed for a cleaner dataset as these pollutants were not included in the Indian AQI assessment calculations for air quality.



City	State	Date	PM2.5	PM10	NO2	NH3	CO	SO2	O3	AQI	AQI_Status
Amaravati	Andhra Pradesh	25/11/2017	81.4	124.5	20.5	10.72	0.12	15.24	127.09	184	Moderate
Amaravati	Andhra Pradesh	26/11/2017	78.32	129.06	26	10.28	0.14	26.96	117.44	197	Moderate
Amaravati	Andhra Pradesh	27/11/2017	88.76	135.32	30.85	12.91	0.11	33.59	111.81	198	Moderate
Amaravati	Andhra Pradesh	28/11/2017	64.18	104.09	28.07	11.42	0.09	19	138.18	188	Moderate
Amaravati	Andhra Pradesh	29/11/2017	72.47	114.84	23.2	12.25	0.16	10.55	109.74	173	Moderate
Amaravati	Andhra Pradesh	30/11/2017	69.8	114.86	20.17	10.95	0.12	14.07	118.09	165	Moderate
Amaravati	Andhra Pradesh	01/12/2017	73.96	113.56	19.29	10.95	0.1	13.9	123.8	191	Moderate
Amaravati	Andhra Pradesh	02/12/2017	89.9	140.2	26.19	13.12	0.1	19.37	128.73	191	Moderate
Amaravati	Andhra Pradesh	03/12/2017	87.14	130.52	21.31	14.36	0.15	11.41	114.8	227	Poor
Amaravati	Andhra Pradesh	04/12/2017	84.64	125	26.98	14.41	0.18	9.84	112.41	168	Moderate
Amaravati	Andhra Pradesh	05/12/2017	88.36	121.77	20.23	13.72	0.12	14.02	117.93	198	Moderate
Amaravati	Andhra Pradesh	06/12/2017	96.83	139.36	25.65	15.12	0.11	16.54	117.21	201	Poor
Amaravati	Andhra Pradesh	07/12/2017	117.46	181.64	41.1	17.34	0.13	28.79	94.63	252	Poor
Amaravati	Andhra Pradesh	08/12/2017	122.88	208.86	54.87	17.96	0.27	22.97	68.6	310	Very Poor
Amaravati	Andhra Pradesh	09/12/2017	74.28	141.22	44.97	15.73	0.09	21.9	60.62	196	Moderate
Amaravati	Andhra Pradesh	10/12/2017	50.32	102.77	33.85	12.56	0.1	13.65	68.15	132	Moderate
Amaravati	Andhra Pradesh	11/12/2017	58.47	115.27	41.64	15.2	0.16	18.37	73.75	147	Moderate
Amaravati	Andhra Pradesh	12/12/2017	89.35	131.48	42.1	21.24	0.24	7.42	44.67	179	Moderate
Amaravati	Andhra Pradesh	13/12/2017	64.42	99.74	34.78	17.63	0.15	5.81	50.16	145	Moderate
Amaravati	Andhra Pradesh	14/12/2017	69.4	98.94	29.97	19.34	0.14	5.8	55	115	Moderate
Amaravati	Andhra Pradesh	15/12/2017	71.07	109.65	25	16	0.1	8.33	79.04	140	Moderate
Amaravati	Andhra Pradesh	16/12/2017	79.76	129.81	40.92	15.56	0.17	8.51	73.17	156	Moderate
Amaravati	Andhra Pradesh	17/12/2017	108.06	167.62	47.05	18.52	0.08	16.05	70.74	225	Poor
Amaravati	Andhra Pradesh	18/12/2017	100.75	172.04	53.94	18.49	0.11	14.05	59.2	251	Poor
Amaravati	Andhra Pradesh	19/12/2017	106.25	171.56	43.09	19.66	0.13	15.45	66.9	228	Poor
Amaravati	Andhra Pradesh	20/12/2017	83.79	141.83	47.47	17.37	0.3	13.35	77.54	223	Poor
Amaravati	Andhra Pradesh	21/12/2017	79.98	147.74	49.55	17.28	0.14	13.58	79.71	170	Moderate

Figure 3.4: Final cleaned dataset

3.4 Visualization

The next step was to import it into Power BI Desktop to visualize it. The following section discusses each step in connecting, checking, and preparing the data inside Power BI.

Importing: I opened Power BI Desktop and clicked Get Data from the Home tab. Then I navigated to the folder for my cleaned Excel workbook and selected the worksheet that contained the cleaned-up dataset. Power BI places a preview of the table into a new window, showing all 12 columns and a few sample rows.

Loading and Verifying Data: After clicking Load, Power BI imported close to 85,000 rows of data. I could see each of the columns listed under the dataset name in the Fields pane. I clicked on the Data view icon to look through the entire table and make sure there are no unintended blank values or extra fields. Up to this point, this data was in exactly the same form as it was in Excel and every row was included and every column correctly labeled.

Visualizations: Before I started charting, I lastly checked the position of each column in the data section this will allow me easy access later on and more specifically, speed up the report design in the future. I also thought I would explore the sample data and try different types of charts while I designed the report and I ended up developing a multi-page interactive dashboard that clearly and effectively displayed the findings.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Overview of Dashboard Structure

The Power BI dashboard for this project is designed to be consistent and easy to navigate on all five pages. All pages will have "AQI Data Analysis" labeled at the top and underneath that will be three slicers Date, AQI Bucket, and City which will hold the same position on all pages. These three slicers will allow the users to filter the data based off of Time range, Contact category (Good, Moderate, Poor) or specific to City.

- **Page 1** serves as the introduction to the overall topic of the dashboard, which will show the time series line chart of average AQI from 2015 to 2020. Below that will be two bar charts, ranking the seven most polluted cities and a bar chart ranking the 7 least polluted cities.
- **Page 2** will focus on overall distributions and city performance. This page will include a line chart of average AQI across, pie chart with days in each bucket, and a second pie chart showing average AQI value in each bucket category.
- **Page 3** focuses on individual pollutants which displays the overall average concentrations of PM2.5, PM10, O3, and NH3 in four card visuals and down the cards we have four bar charts who ranks the top ten cities for each of those pollutants.
- **Page 4** discusses two more gases NO2 and SO2 and two bar charts shows the top ten cities for each of those gases. A third bar chart on this page compares the average levels of all seven pollutants by state.
- **Page 5** ends with a choropleth map of India, which show each state by its average AQI. From this remapping view, users can quickly see where air quality is better or worse.

4.2 Pollution Over Time and Across Cities

First Page of the dashboard presents both a time series line chart and two bar charts to depict the progression of air quality in India from 2015 - 2020 and also shows which cities performed worst or best when it comes to pollution levels.

Trend Over Time

The main line chart indicates the daily average AQI from all monitoring stations over the 6-year period.

1. Overall Trend: From 2015 through 2020 the average AQI shows a slight upward trend which indicates that air quality on average was getting worse during this time.
2. Seasonal Peaks: There are clear spikes every winter (November through February) which is a result of cooler temperatures and low wind speeds resulting in pollutants getting trapped near the ground. These peaks often also accord with crop-burning in northern states, as well as the increased burning of biomass in homes for heating purposes.
3. Summer Lows: AQI values tend to be at their lowest during April through June, when winds are stronger and the monsoons arrive, which helps disperse particulate matter.

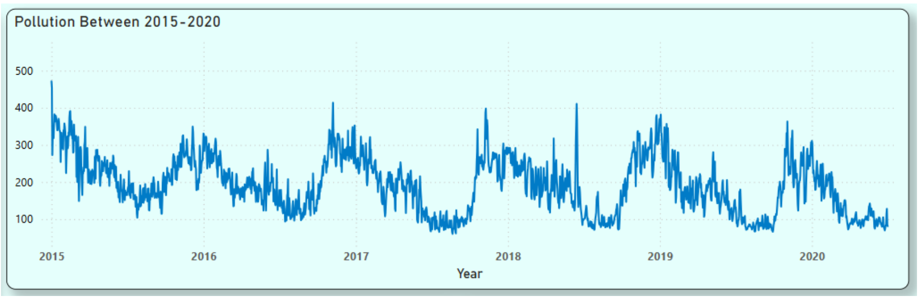


Figure 4.1: Line chart showing daily average AQI, 2015–2020

City Comparisons

Below the timeseries graph of two bar graphs show the top seven most polluted cities and the top seven least polluted cities based on their average AQI during the same time.

1. Most Polluted Cities: The most polluted includes industrial hubs and dense metropolitan centers. Cities of Delhi, Ahmedabad, and Lucknow are usually leading most polluted cities with an average AQI above 200 ("Poor" to "Very Poor") established a composite AQI consisting of heavy vehicle traffic, factory emissions, and relatively limited geography for dispersion of emissions.
2. Least Polluted Cities: Coastal or hill stations like Aizawl, Shillong, and Thiruvananthapuram find their place among the cleanest, having an average AQI below 100 ("Good" to "Satisfactory"). These AQI values benefit from geographic advantages whether they be the sea breezes or altitude benefiting from less concentration.

This is also a living report as when a user selects a filter for actual year the lists made for that year's city standing. This interactive report allows for further exploration for example how the lockdowns in 2020 impacted the city standings, seeing how relatively polluted cities temporarily improved their rankings.

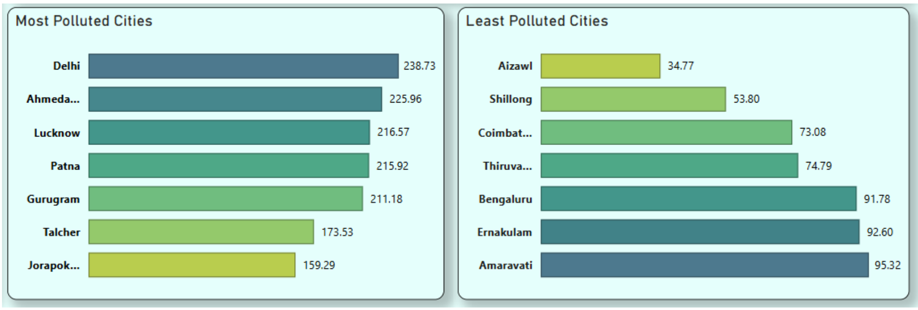


Figure 4.2: Bar charts of the seven most and least polluted cities based on average AQI

4.3 AQI Bucket and City Insights

The second page of the dashboard provides two complementary perspectives on air quality distribution across individual pollution categories (buckets) and cities.

City-Wise AQI Ranking

On this page the line chart plots the average AQI for each city as sorted in descending order. The default setting displays all cities long-term averages from 2015 to 2020. The overall descending trend line allows one to visualize the distance between cities with the worst pollution to cities with moderate or good air quality. For ease of use, simply hover over any point to reveal that city's average AQI.

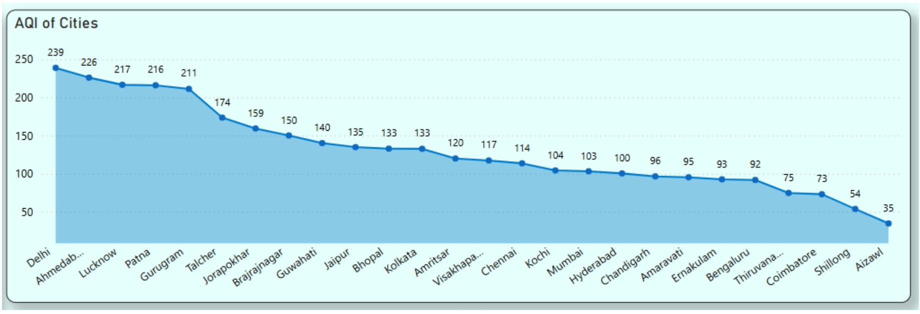


Figure 4.3: Line chart of cities by descending average AQI

AQI Bucket Distribution

Beneath the line chart is the first pie chart that shows the distribution of days within AQI buckets (Good, Satisfactory, Moderate, Poor, Very Poor, Severe). Each color-coded slice corresponds to the proportion of total records in its category. For instance, the largest slice is usually "Moderate" which indicates that most days are neither very clean nor severely polluted. Only the labels for Severe and Good are likely to be smaller slices.

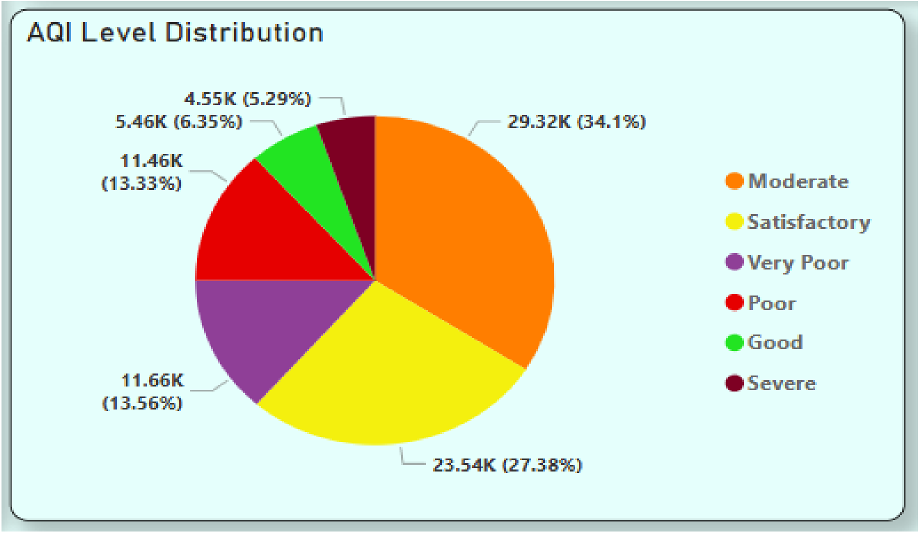


Figure 4.4: Pie chart showing percentage of days in each AQI bucket

Average AQI by Bucket

The second pie chart included on this page depicts the average AQI value within each bucket. It is important to note that the size of each slice refers to the mean AQI for that bucket and not the number of days. This expresses the severity of each bucket in quantitative terms. Even though the "Moderate" bucket may contain many days, it has a significantly lower AQI value than the "Severe" bucket, which is the largest slice in this pie chart.

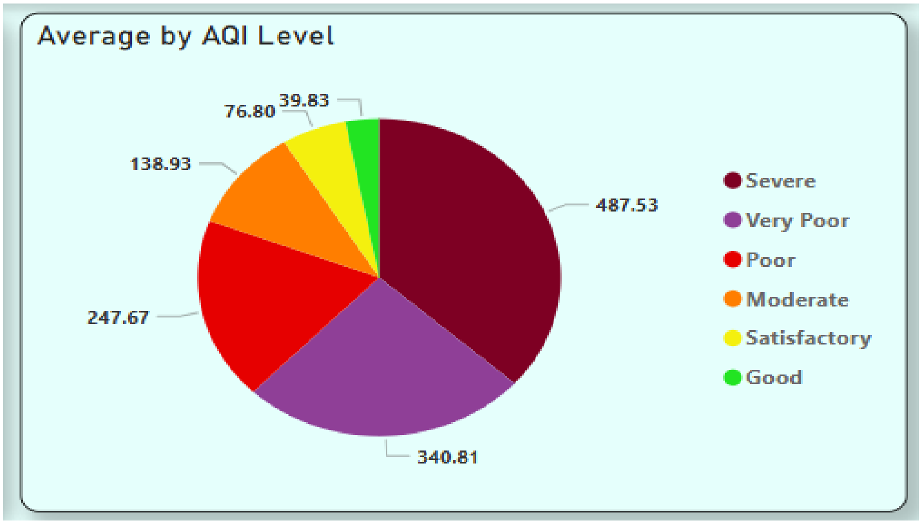


Figure 4.5: Pie chart of average AQI value per bucket

4.4 Pollutant-Wise City Comparison

Third and Fourth Page of the dashboard focus on comparing specific pollutants across cities.

Overview of the key pollutants

At the top of page 3 and 4 there are card visuals provide the national averages for the four pollutants PM2.5, PM10, O3, and NH3 using the full time-period of 2015–2020.

PM2.5 (Particulate Matter): The exposure from these smallest particles that can travel deep into the lungs.

PM10 (Particulate Matter): The exposure this larger size particulate matter that can be in dust or in anything more particulate matter shaped like that in construction work.

O3 (Ozone): This pollutant can cause the airway to swell and aggravate the symptoms of asthma.

NH3 (Ammonia): This comes from agricultural practices and garbage management have an overall impact.

NO2 (Nitrogen Dioxide): This comes when combustion activities from vehicles and power plants produce the pollutant that can aggravate the airway.

SO2 (Sulphur Dioxide): This comes from coal and industry affecting levels of pollution and respiratory ailments and acid rain.

Figure 4.6: Card visuals of average levels pollutants

Top Cities by Pollutant

The ten cities with the highest average of each key pollutant over the 2015-2020 period are listed below. These charts show which urban areas are most affected by fine particulates, gases, or other pollutants. The comparison of these top-ten lists is side by side which showcase struggle of cities with specific pollutants. This information helps targeting specific pollutants.

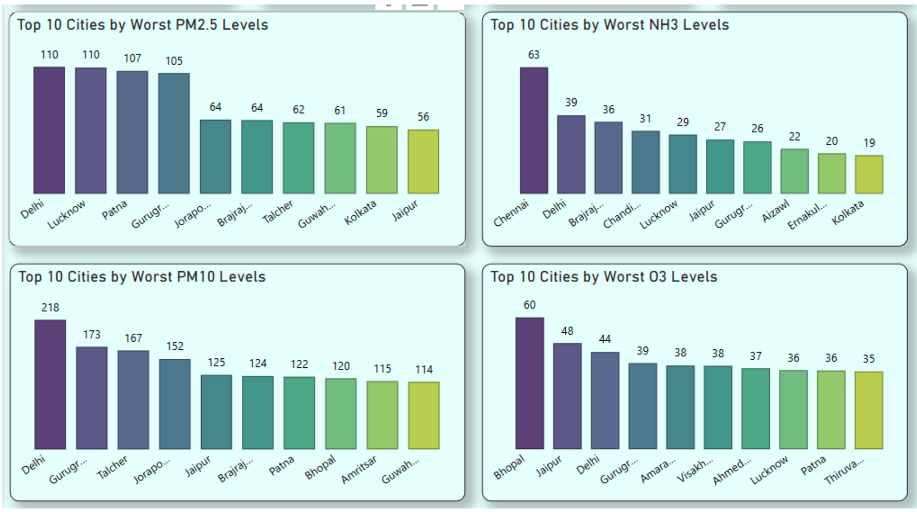


Figure 4.7: Bar charts of top ten cities by PM2.5, PM10, O3, and NH3

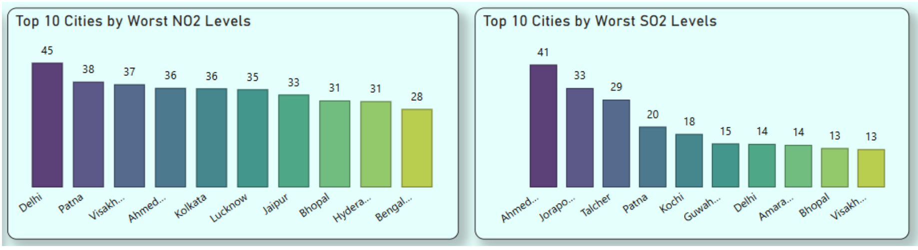


Figure 4.8: Bar charts of top ten cities by NO_2 and SO_2

4.5 State-Level Pollution Patterns

This bar chart shows the average concentrations for six key pollutants which are PM2.5, PM10, NO2, SO2, O3, and NH3 for each Indian state for the period 2015–2020. Data for each state coincides together with one bar for each pollutant which allows for easy comparisons of pollutants between the states.

Finding Heavy Polluters: As expected, states with higher readings of PM2.5 and PM10 such as Uttar Pradesh and Delhi had the highest bars in these categories. These high pollutant readings are likely associated with industrial activity, traffic, and agricultural field burning and states like Jharkhand and Gujarat had higher SO2 readings because of their heavy industry.

Gas Levels: Bihar and Delhi had quite high average levels of NO2 readings due to dense traffic and industrial activity. Tamil Nadu and Haryana had higher average levels of NH3 due to agricultural production intensity and fertilizer application.

Cleaner Areas: States along the coast such as Kerala and Karnataka tended to have lower readings of most pollutants. Mizoram and Meghalaya also show lower levels of pollutants in part due to their Eastern Himalayan geographical landscape with hilly areas and regions removed from industrial developments and population densities especially for particulates.

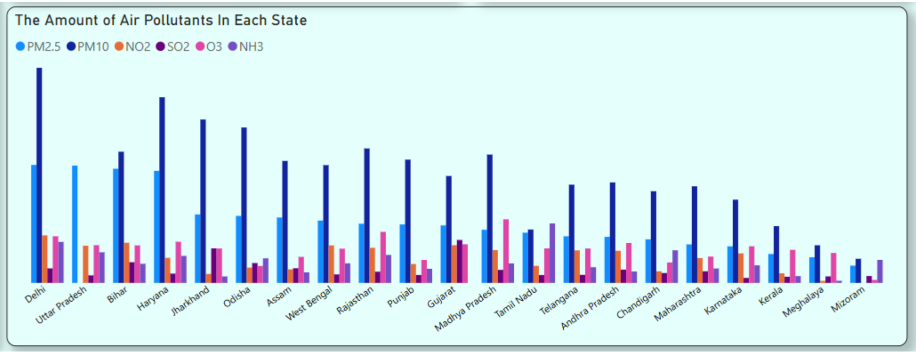


Figure 4.9: Average concentrations of PM2.5, PM10, NO2, SO2, O3, and NH3 by state

This Last Page displays a choropleth map of India showing the average AQI in each state over the six-year time period where the darker shades indicating worse air quality and lighter shades showing regions with better air quality.

- **Northern Hotspots:** States like Delhi, Punjab and Uttar Pradesh are shown with the darkest colors because they have the highest average AQI values which means worst air quality.
- **Central and Eastern India:** States such as Jharkhand and Chhattisgarh are shown with dark colors from the pollution caused by industries' excessive emissions of particulate matter.
- **Southern and Northeastern India:** States in Southern and eastern parts of the country have the lightest colors, which represent that those states have extremely low AQI rankings, thus being less polluted overall.

This map helps a user understand regional air-quality patterns more quickly, and allows a user to click on any state to filter the rest of the dashboard pieces to that state. For example, if the user selects only Maharashtra, all the visuals will update to reflect that state only, which allows them to focus their analysis on Maharashtra.

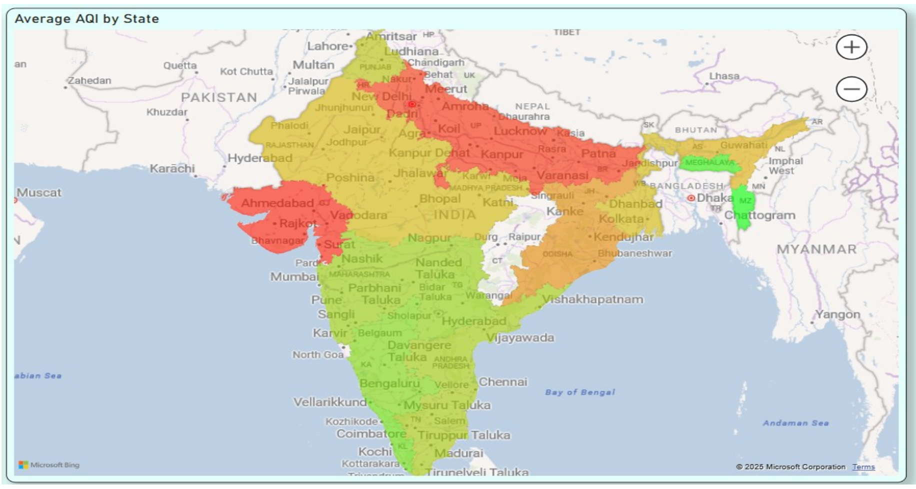


Figure 4.10: Choropleth map of average AQI by state

CHAPTER 5

Conclusion and Future Scope

5.1 Summary of Findings

Overall, average AQI values rose over the examined period from 2015 to 2020, with median seasonal peaks in the winter months when cooler temperatures and stagnant air led to traps of pollution at the surface.

City-level comparisons showed considerable class variation, for example some cities consistently had the highest pollution levels and tended to be classified as "Poor" or "Very Poor" on the AQI scale. While cities some cities improved air quality due to geography and a low concentration of polluting industries. Most records fell into the "Moderate" AQI category and while analyzing I found that individual pollutants like PM2.5 and PM10 were found in most quantity with their averages exceeding safe limits in many urban centers.

The findings show that although air pollution is a significant problem across India, the extent and proportion of pollution vary by season, city, and state. Pollution generally worsens in winter and reduces during summer. Also, there are cities and states that are more polluted than others, so targeted actions such as better traffic legislation, reduced pollution from factories, and more sustainable farming methods can make these actions more effective.

5.2 Future Scope

Real-Time Data Integration: At present, the dashboard only provides data as recent as 2020. By integrating real-time air-quality feeds (government APIs or IoT sensor networks), the dashboard could present live AQI readings. A live dashboard would allow users to make instantaneous decisions on whether or not to go outside, as well as do chores that day.

Predictive Modeling with Machine Learning: An easy step forward would be to provide some sort of predictive modeling. For instance, if the desktop dashboard uses linear regression or time-series models like ARIMA or Prophet the displayed AQI levels can be correlated to future (typically, short-range) AQI levels, with which future pollution trends can visually be illustrated. A linear regression or time-series prediction would provide citizens and policymakers with an expected AQI level for the upcoming week or month, and subsequently, engage in preventive actions.

Web and Mobile Accessibility: Developing the Power BI report as a web or mobile app would make it more usable and shareable. Embedding it in a web page or developing a mobile-friendly interface would do the trick.

Integration with Weather and Traffic Data: Connecting air-quality data to weather patterns (wind, rain, temperature) may uncover deeper patterns. For example, if there are drops in aqi with raining, then rain has a role in the cleanliness of the air.

Increased Geographic Scope: The current dashboard only captures cities and states in India if we Extend the dashboard to include cross-border regions (neighboring countries) would allow air quality comparison on a global scale.

5.3 Final Conclusion

This project has taken years of raw air-quality data and transformed it into one interactive dashboard that helps reveal overall pollution trends, better understand health risks, and inform action. This project was completed by cleaning and aggregating data in Excel Power Query and then creating visuals in Power BI to demonstrate how average pollution levels have generally changed over time, to show the specific dates where air quality was worse, and to show which areas have the most significant challenges. The final dashboard integrates line graphs, bar charts, pie charts, and maps, piggybacking on the user-friendly functionality of the software so any user can examine trends by date, category of pollution, and location in only a few clicks. In a way, the project showed how the simplicity of using free, no-code solutions, like Power Query and Power BI can enable people to take messy datasets from publication to clean and trustworthy insights, all without writing a single line of code. Aside from the tangible results, the entire experience yielded incredible opportunities in the area of skills learned, including finding and merging datasets, applying consistent cleaning steps, assigning appropriate data types, and building reports in a straightforward manner. The output is more than just a report or series of reports; it is a living document that can be updated, added to, and shared between teams and devices. It serves as a stable platform for real-time monitoring, and simple forecasting, and provides the possibility of including another data source as a future step. Most importantly for decision-makers, researchers, and community members, it gives them a straightforward way to track air quality, compare changes over time, and target actions where they will have the most impact. The takeaway is that this work gives an opportunity to see where air quality performance was in the past, and points in a better, more deliberate, and data-informed direction towards cleaner air and healthier lives.

Project Summary

Registration No: 323102774 Name of student: Shriyansh Shukla

Title of Capstone Project / Project Work:
AQI Data Analysis Dashbaard

Objectives of the Project:

1. I want to give policymakes an Interactive platform for AQI
2. Convert complex AQI data into clear Insights.
3. Enable quick comparison of pollution Levels.
4. Raise greater public awareness of Air Pollution.

Results and Findings:
In my dashboard we can see general rise and low in India AQI from 2015-2020, with clear winter spikes and significant regional differences.

Specific outcomes of the Project:
A documented and repeatable pipeline that produce a clean 12 column dataset and A five page dashboard with Time series trends.

Learnings from the Project:
I got skilled in end-to-end data analytics. I Learnd about cleaning and transforming data and making Interactive visuals.

Any Challenges/issues faced during the Project:
I faced inconsistent and missing data during cleaning. It was little hard to clean data.

Signature of the Student:

15