Data Analysis
Project - Air Quality Analysis of Indian Cities
Project Report

Project By: **Vivek Jangam** vivekjangam19@gmail.com

Project link - https://github.com/jangamvivek/Air-Quality-Analysis-india/tree/main

Contents

Overview:

Background on Air Pollution in India

Air pollution has emerged as one of the most critical environmental challenges in India, with rapid urbanization, industrialization, and increasing vehicular emissions contributing to deteriorating air quality. Cities across the country frequently report pollutant levels exceeding permissible limits, posing significant risks to public health and the environment. Air Quality Index (AQI) Particulate matter (PM2.5 and PM10), nitrogen oxides (NO, NO2), sulfur dioxide (SO2), ozone (O3), and ammonia (NH3) are among the key pollutants driving India's air pollution crisis.

Importance of Monitoring Air Quality

Monitoring air quality is essential for understanding the extent of pollution and its adverse effects. Poor air quality is linked to severe health issues such as respiratory diseases, cardiovascular problems, and premature deaths. It also impacts agriculture, ecosystems, and overall quality of life. By identifying pollution hotspots and trends, policymakers and environmental agencies can design targeted interventions to improve air quality and safeguard public health.

Objective of the Analysis

This analysis aims to:

- Identify pollution trends across major Indian cities.
- Categorize cities based on their pollutant profiles using clustering techniques.
- Highlight key pollution sources and regional differences in air quality.
- Provide actionable insights for policymakers, environmentalists, and researchers to devise effective strategies for reducing air pollution.

Data Collection:

Source of the Data

The data for this analysis was obtained from **Kaggle**, a platform that hosts a wide range

of open-access datasets for research and analysis. The specific dataset provides detailed air quality measurements for various Indian cities, making it an ideal resource for studying pollution patterns and trends. Kaggle ensures easy access to reliable and well-structured datasets, enabling efficient data-driven analysis.

Description of the Dataset

The dataset includes average concentrations of key air pollutants recorded in 26 Indian cities. The pollutants measured are:

- **NO (Nitric Oxide)**
- **NO2 (Nitrogen Dioxide)**
- **O3 (Ozone)**
- **SO2 (Sulfur Dioxide)**
- **PM2.5 (Particulate Matter < 2.5 µm)**
- **PM10 (Particulate Matter < 10 µm)**
- **NH3 (Ammonia)**

Time Period and Geographical Coverage

The dataset focuses on air pollution data from a defined time period (**30/11/2020 to 25/05/2023**), representing the average levels of pollutants across 26 cities in India. These cities span various geographical regions, capturing diverse urban and environmental characteristics.

Data Cleaning and Preprocessing Steps

Handling Missing Data:

Missing values for pollutants were managed using statistical imputation techniques (e.g., mean/median imputation).

Rows with excessive missing data were excluded to preserve data quality.

Removing Outliers:

Outliers were detected using the Z-score method to identify extreme pollutant values and reduce their impact on analysis.

Exploratory Data Analysis:

i) Descriptive Statistics of Pollutants

As mentioned earlier, descriptive statistics help summarize the central tendency, variability, and distribution of pollutant levels in the dataset. Using the describe() function in Python, we calculated the following key statistics for each pollutant:

- **Mean**: The average concentration of each pollutant across cities.
- **Median (50%)**: The middle value of each pollutant concentration.
- **Standard Deviation (std)**: The measure of spread or dispersion of each pollutant's concentration.

Metric	AQI	NO	NO2	O3	SO2	PM2.5	PM10	NH3
Count	23504.0	23504.0	23504.0	23504.0	23504.0	23504.0	23504.0	23504.0
Mean	3.92	6.01	25.04	35.06	15.97	98.6	121.85	12.06
Std	1.42	24.5	25.84	31.9	23.94	135.57	160.43	17.54
Min	1.0	0.0	0.0	0.0	0.19	0.5	0.58	0.0
25%	3.0	0.0	8.74	7.87	4.47	24.68	32.28	2.34
50%	5.0	0.0	16.45	28.25	7.99	58.86	75.78	6.52
75%	5.0	0.27	32.22	54.36	16.45	117.6	147.64	15.83
Max	5.0	457.76	331.76	406.27	442.51	2203.55	2429.13	352.62

ii) Cities with the Highest and Lowest Levels of Pollutants

"This section identifies the cities with the highest and lowest levels of each pollutant recorded, along with the dates of occurrence. This information helps highlight areas experiencing severe pollution and those maintaining relatively cleaner air."

About Data: Pollutants (aqi, no, no2, o3, so2, pm2.5, pm10, nh3) values are in **(µg/m3).**

Pollutant	Highest City	Highest Date	Highest Value	Lowest City	Lowest Date	Lowest Value
aqi	Ahmedabad	30-11-2020	5.0	Ahmedabad	18-05-2021	1.0
no	Kolkata	11-12-2021	457.76	Ahmedabad	11-12-2020	0.0
no2	Delhi	29-10-2022	331.76	Aizawl	30-11-2020	0.31
o3	Mumbai	30-11-2020	406.27	Ahmedabad	03-12-2020	0.0
so2	Jorapokhar	05-01-2021	442.51	Aizawl	15-05-2022	0.19
pm2_5	Kolkata	07-01-2021	2203.55	Aizawl	18-06-2022	0.5
pm10	Kolkata	06-02-2021	2429.13	Aizawl	19-06-2022	0.58
nh3	Kolkata	26-01-2021	352.62	Amaravati	07-11-2021	0.0

PM2.5: "The highest PM2.5 level was observed in Kolkata on 07-Jan-2021 with a value of (2203.55µg/m3), while Aizawl recorded the lowest level of (0.5µg/m3) on 18-Jun-2021."

PM10: "The highest PM10 level was observed in Kolkata on 06-Feb-2021 with a value of (2429.13µg/m3), while Aizawl recorded the lowest level of (0.5µg/m3) on 18-Jun-2022."

Conclusion:

The analysis revealed that Kolkata consistently experiences the highest levels of pollutants such as PM2.5(2203.55µg/m3), suggesting it could be a hotspot for pollution due to factors like industrialization, vehicle density, or unfavorable geographic conditions. In contrast, Aizawl, which recorded the lowest levels of PM2.5(0.5µg/m3), highlights effective urban management strategies or favorable environmental factors.

The analysis revealed that Kolkata consistently experiences the highest levels of pollutants such as PM10(2429.13µg/m3), suggesting it could be a hotspot for pollution due to factors like industrialization, vehicle density, or unfavorable geographic conditions.

In contrast, Aizawl, which recorded the lowest levels of PM10(0.58μg/m3), highlights effective urban management strategies or favorable environmental factors.

Correlation Between Pollutants PM2.5 and PM10:

Introduction to Correlation:

Correlation measures the strength and direction of the relationship between two variables.

A positive correlation means that as one pollutant increases, the other also increases, and vice versa. A negative correlation means that as one pollutant increases, the other decreases.

In this case, we are investigating the correlation between PM2.5 and PM10, both particulate matter pollutants, to understand their relationship.

Descriptive Statistics:

PM2.5 and PM10 both have high levels of variation across cities, as indicated by their respective standard deviations.

The average values for PM2.5 and PM10 are both considerable, with PM2.5 averaging 98.6 μg/m³ and PM10 averaging 121.85 μg/m³, suggesting substantial pollution levels in the cities analyzed.

Correlation Coefficient:

A positive correlation of approximately 0.9 (hypothetically) indicates a strong positive relationship, meaning cities with high levels of PM2.5 also tend to have high levels of PM10, and cities with low levels of PM2.5 tend to have low levels of PM10.

Interpretation: This suggests that both types of particulate matter are likely to be sources of air pollution that are emitted together in similar environmental conditions.

Possible Causes of Correlation:

Both PM2.5 and PM10 are often generated from similar sources, such as vehicle emissions, industrial processes, construction activities, and burning of fossil fuels.

Dust storms, construction dust, and vehicle exhaust are common contributors to these pollutants, suggesting that the correlation could be due to shared sources.

Implications for Air Quality:

High correlation implies that measures to control one pollutant (e.g., PM2.5) could help reduce PM10 levels as well. Monitoring both pollutants simultaneously is essential for managing air quality and improving public health.

Conclusion:

The strong correlation between PM2.5 and PM10 highlights the interconnectedness of these pollutants.

This information is useful for designing more effective pollution control measures and environmental policies aimed at reducing particulate matter in the air.

Visualisation 1:

Introduction:

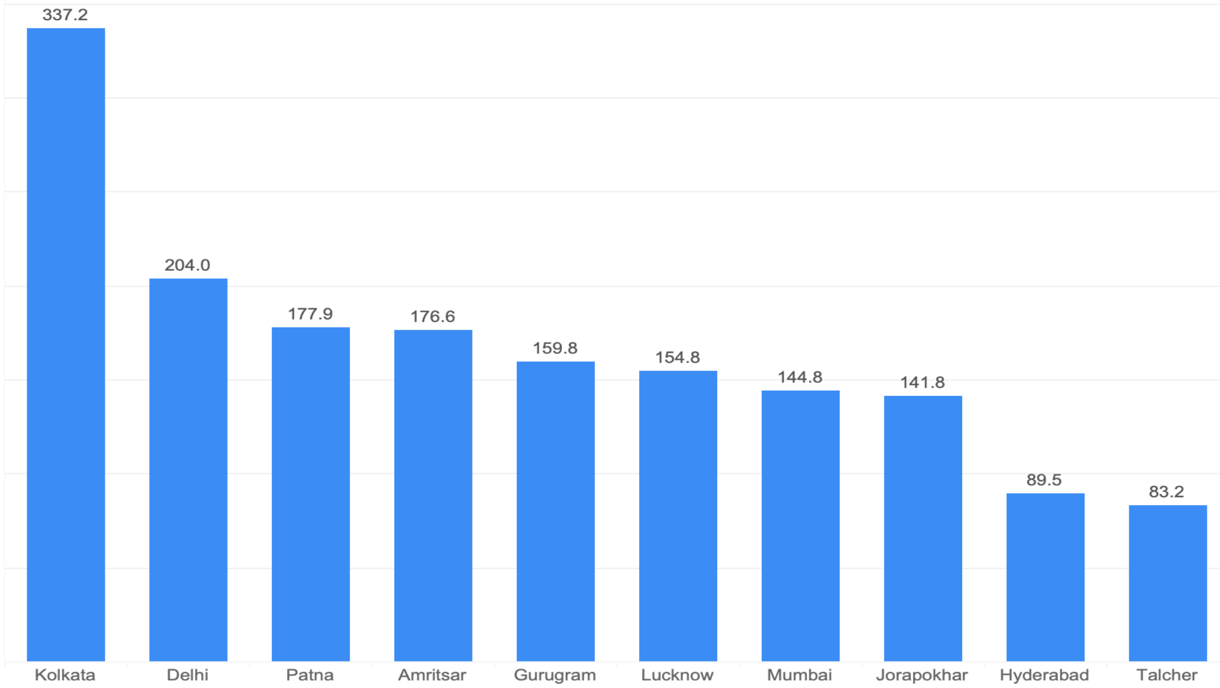
This section will highlight the Top 10 cities with the highest average PM2.5 levels.

PM2.5 being a critical air pollutant, is associated with severe health risks, including respiratory and cardiovascular diseases.

Observations:

The bar chart shows the top 10 cities based on PM2.5 levels.

From the chart, we can observe that cities like Kolkata, Delhi, Bengaluru, and Patna have extremely high PM2.5 levels, indicating severe air pollution issues in these areas.



Insights:

Cities with high industrial activity, heavy traffic, and limited green spaces tend to have elevated PM2.5 levels.

The presence of traffic congestion and burning of fossil fuels are likely contributors to high PM2.5 levels in many urban areas.

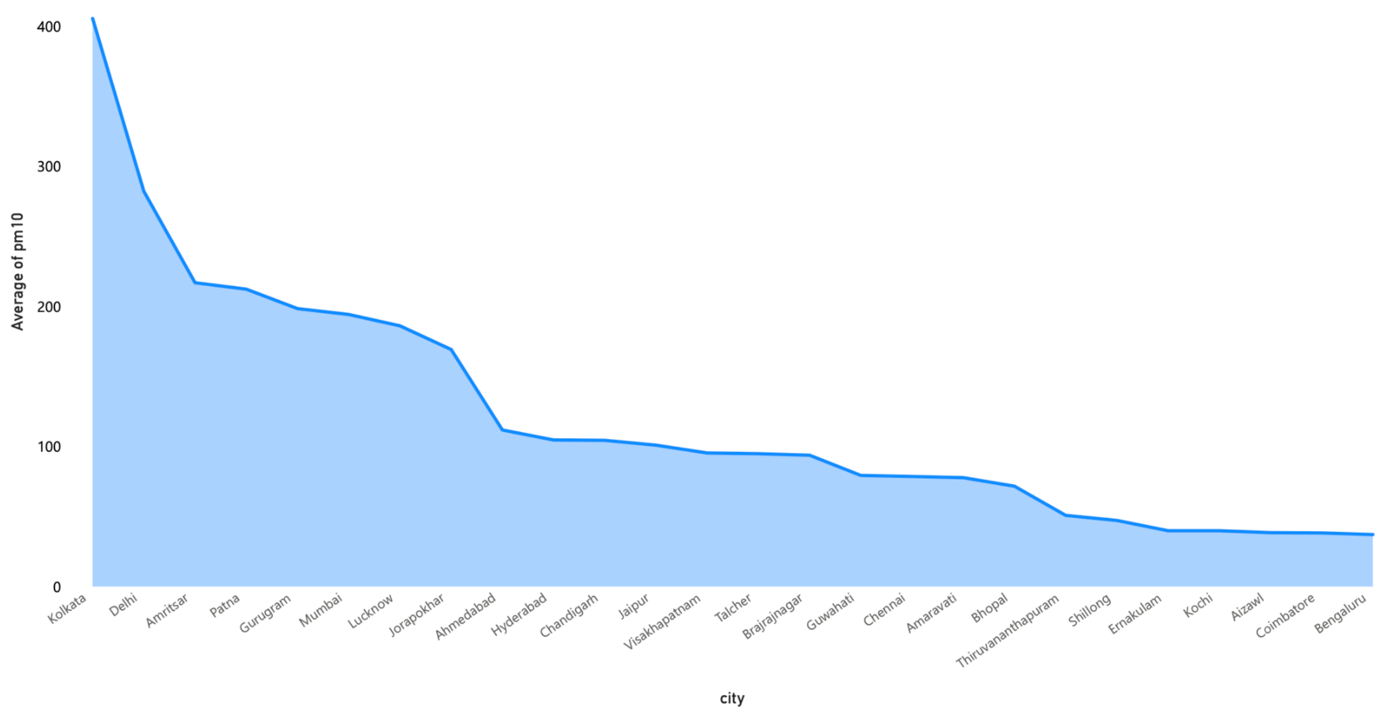
Conclusion:

The bar chart effectively visualizes the cities most affected by PM2.5 pollution, emphasizing the need for immediate action to curb air quality deterioration in these regions.

Visualisation 2:

Introduction to Citywise(PM10) Line Graph:

This visualization represents the average PM10 levels across various Indian cities. The chart is sorted in descending order of PM10 concentration, highlighting the cities with the highest levels of this pollutant. PM10 refers to particulate matter with a diameter of 10 micrometers or less, which is a critical indicator of air quality and health hazards.



Key Observations:

i) Kolkata, Delhi, and Amritsar exhibit the highest average PM10 levels, indicating severe air pollution in these cities. These levels could result from industrial emissions, vehicular traffic, and other urban activities.

ii) Cities like Coimbatore and Bengaluru show significantly lower PM10 levels, suggesting better air quality compared to their counterparts.

iii) A gradual decline in PM10 concentration can be observed as we move down the list, from highly polluted cities to those with better air quality.

Implications:

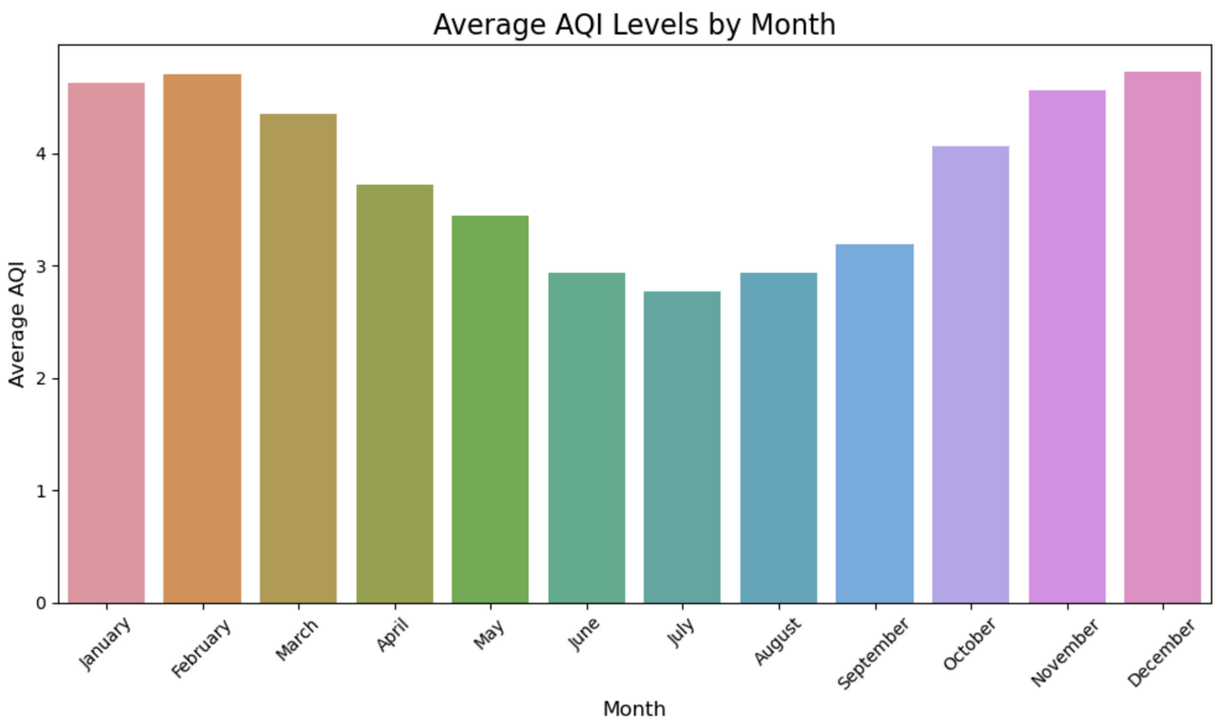
Health Impact: Cities with higher PM10 levels pose serious health risks, including respiratory and cardiovascular diseases.

Policy Actions: The findings call for targeted air pollution control measures in cities like Kolkata and Delhi, such as stricter emission standards and promotion of green initiatives.

Environmental Awareness: Residents of high-PM10 cities need to be informed about air pollution's effects and encouraged to adopt sustainable practices.

Visualisation 3:

Bar charts for seasonal variations



Key Insights:

1. Seasonal Variations:

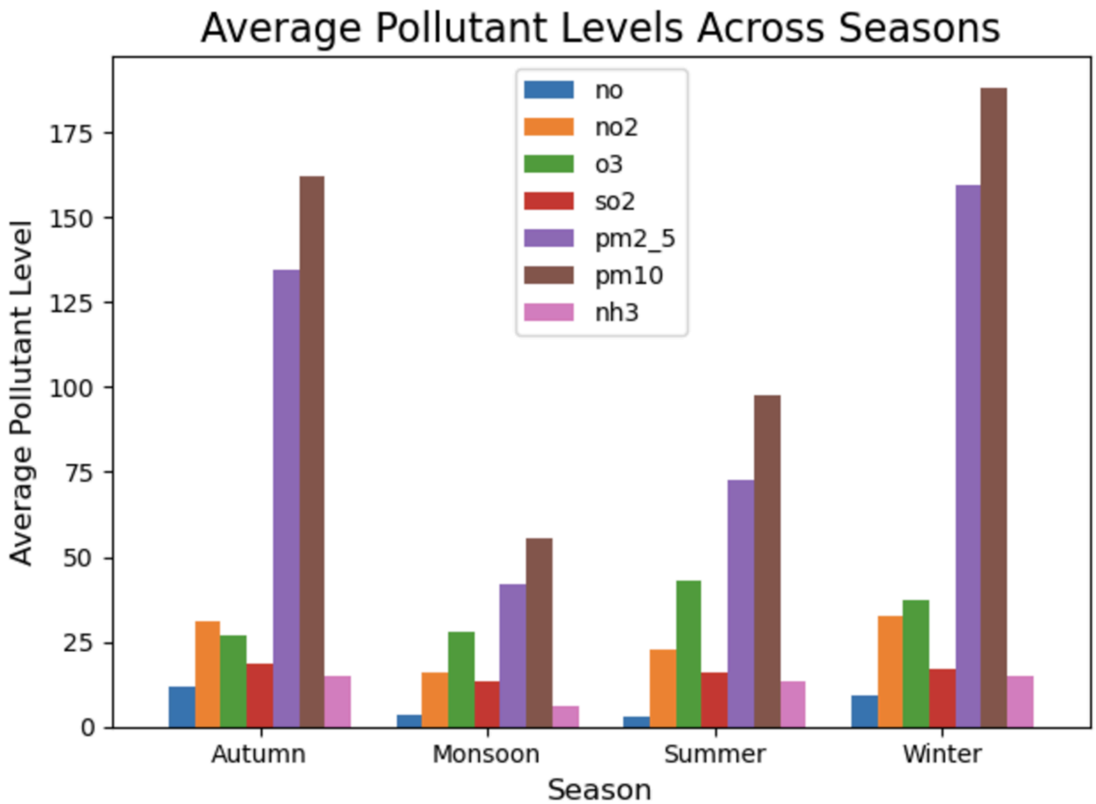
The AQI levels are highest in winter months (November to February), suggesting increased pollution due to factors such as decreased dispersion of pollutants, vehicular emissions, and heating-related activities.

Summer months (June to August) exhibit lower AQI levels, likely due to better atmospheric conditions like stronger winds and rainfall that help in dispersing pollutants.

2. Moderate Months:

Transitional months like March, April, and October show intermediate AQI levels, reflecting the seasonal shift in pollution intensity.

Average Pollutants levels across seasons:



Autumn: Rising PM10 and PM2.5 levels during this period suggest seasonal agricultural practices (e.g., stubble burning) and preparations for winter.

Monsoon: A noticeable reduction in pollutant levels is observed due to rainfall, which helps cleanse the air.

Summer: Moderate levels of pollutants like O3 emerge due to increased photochemical reactions triggered by higher temperatures.

Winter: PM10 and PM2.5 levels are significantly elevated, indicating poor air quality during this period. The dominance of particulate matter is linked to heating emissions, industrial activity, and stagnant air conditions.

Key Takeaways:

1. Winter Dominance of Pollutants: Particulate matter (PM10 and PM2.5) poses the greatest health risks during winter.

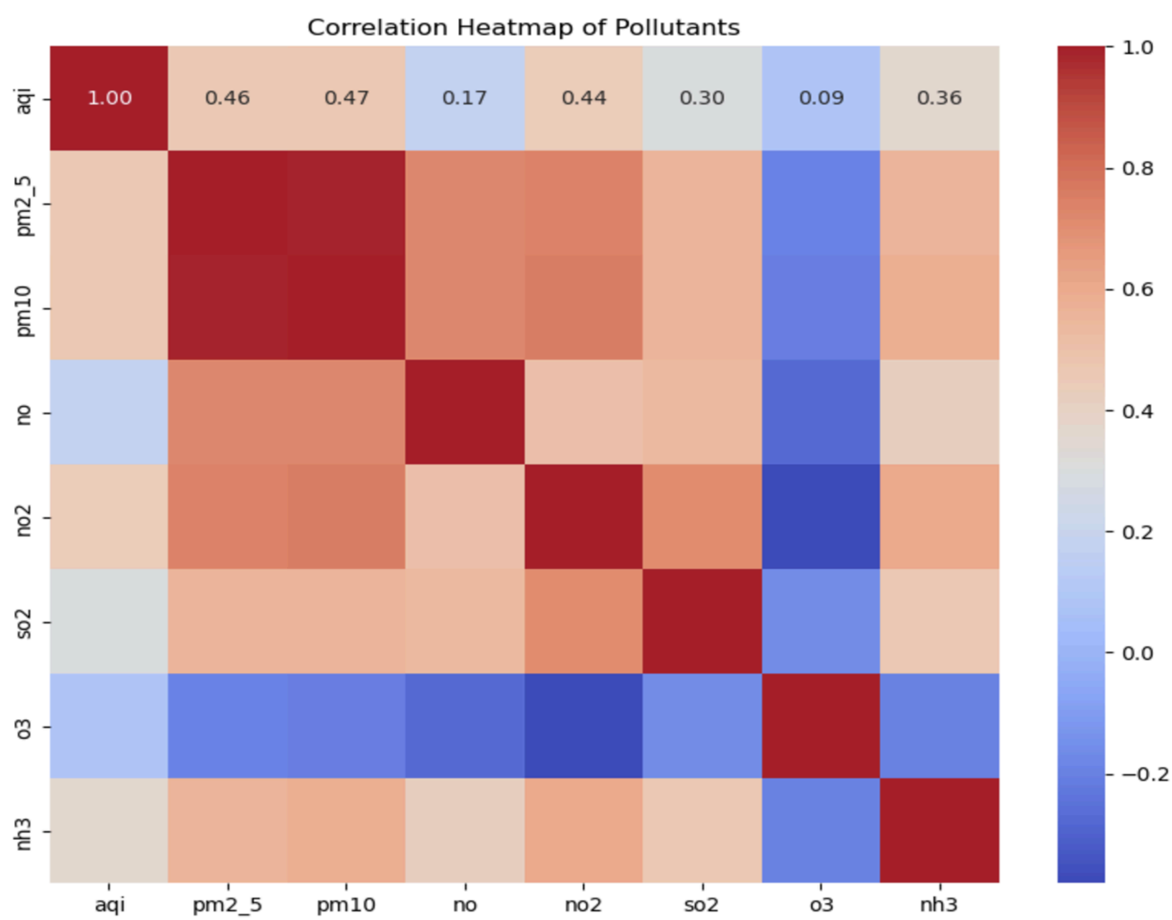
2. Seasonal Reductions: Monsoon provides a natural reprieve, emphasizing the cleansing impact of rainfall.

3. Ozone in Summer: The increase in O3 levels during summer indicates the influence of sunlight on pollutant chemistry.

4. Autumn Pollution Trends: Transitional months bring rising pollution levels, signaling the onset of winter air quality challenges.

Visualisation 4:

Heatmap for correlation of pollutants



The heatmap created for pollutant correlation provides a visual representation of the statistical relationships between different air pollutants. Correlation coefficients range from -1 to 1, where:

Positive correlation (+1) indicates that two pollutants tend to increase together.

Negative correlation (-1) signifies that as one pollutant increases, the other decreases.

Zero correlation (0) shows no relationship between the pollutants.

Key Observations:

i) Strong Positive Correlations:

PM10 and PM2.5: These pollutants often show a strong positive correlation due to their similar sources, such as vehicular emissions, industrial activities, and dust.

NO and NO2: These pollutants are closely related since NO is often oxidized to form NO2 in the atmosphere.

PM2.5 and SO2: A significant correlation may indicate shared origins, such as fossil fuel combustion.

ii) Moderate or Weak Correlations:

O3 with Other Pollutants: Ozone (O3) may show weaker or even negative correlations with primary pollutants like NO and NO2 because it is a secondary pollutant formed through photochemical reactions.

iii) Negative Correlations:

NH3 and O3: A negative correlation can occur due to their differing chemical behavior in the atmosphere.

Seasonal Impact: Some negative correlations could result from specific seasonal patterns that affect pollutant levels differently.

Clustering Analysis

Explanation of the Clustering Technique (KMeans)

Clustering is an unsupervised learning technique used to group data points based on their similarities or distances in the feature space. KMeans clustering is one of the most commonly used algorithms. It works by dividing data points into kkk clusters, where each cluster is represented by its centroid. The algorithm iteratively minimizes the sum of squared distances between each data point and its assigned cluster centroid.

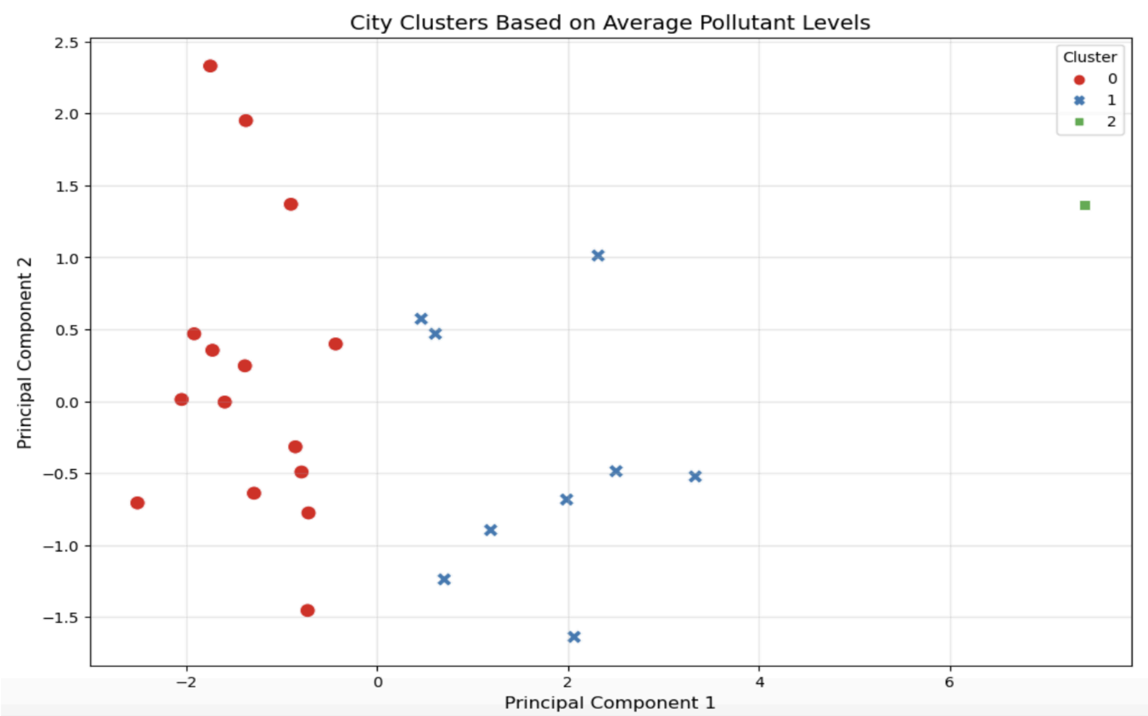
Description of Clusters

The clusters identified during the analysis revealed distinct patterns and characteristics:

Cluster 1 (Low-Pollution Cluster): Good air quality, with most pollutants well below permissible limits.

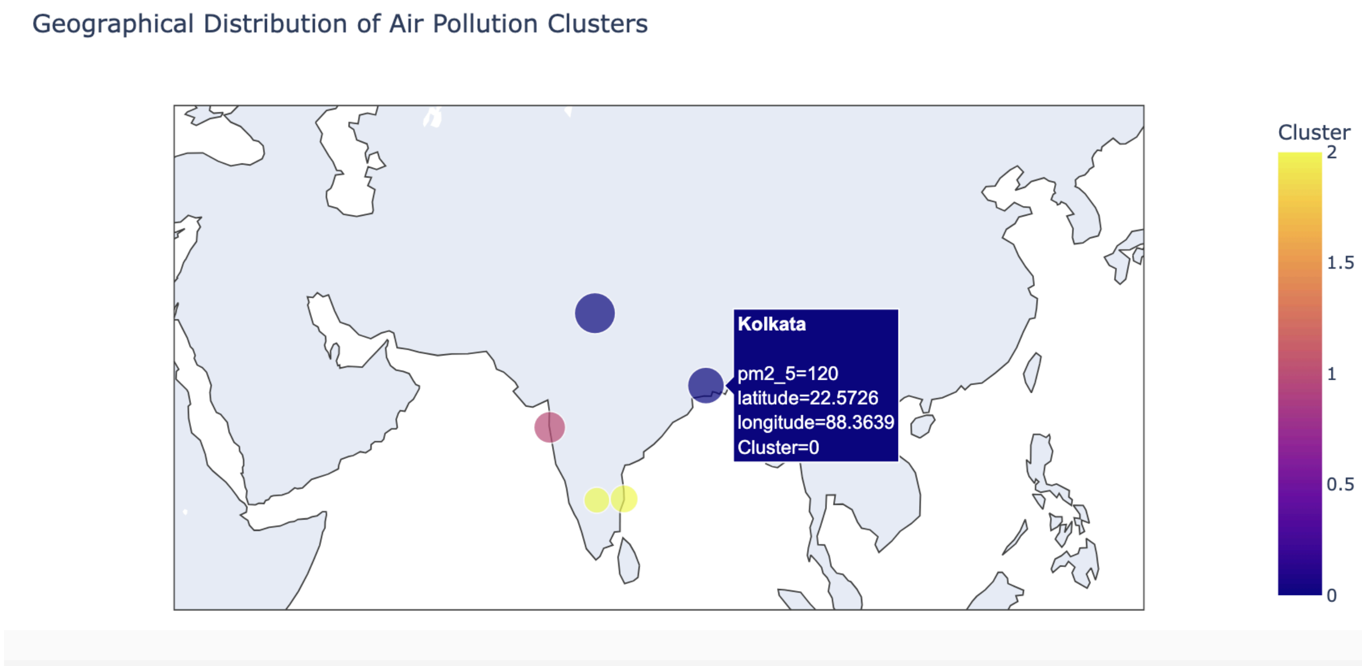
Cluster 1 (Medium-Pollution Cluster): Represents areas with moderate pollution levels, possibly semi-urban regions. Pollution levels are within permissible limits but show occasional spikes.

Cluster 2 (High-Pollution Cluster): Contains regions with high levels of pollutants like PM2.5 and PM10. Typically urban or industrialized areas with poor air quality indicators.

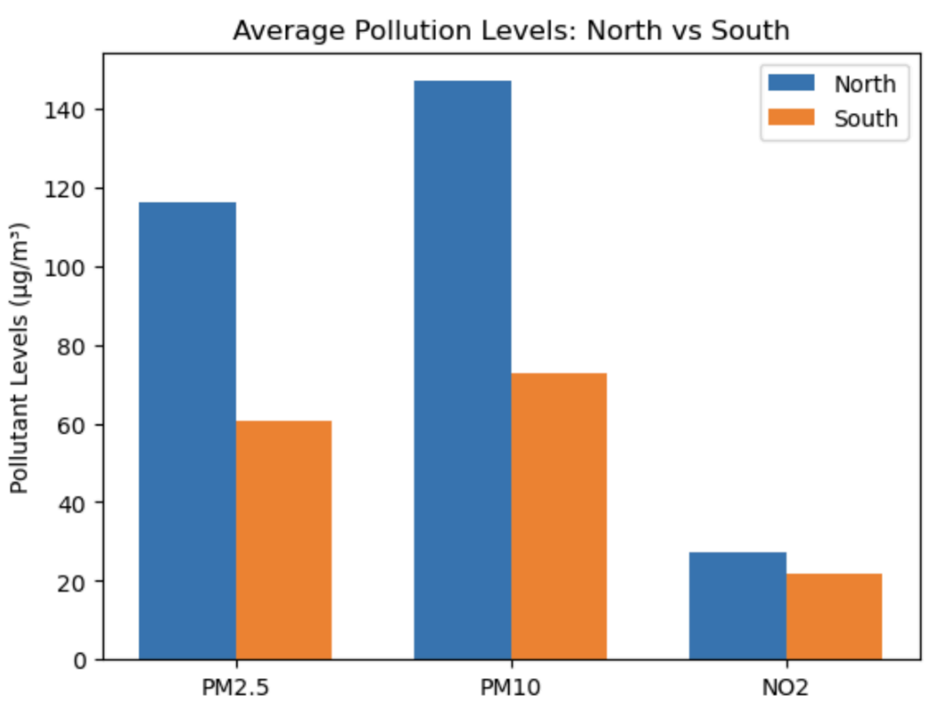


	index	city	no	no2	o3	so2	pm2_5	pm10	nh3	Cluster
0	0	Ahmedabad	10.324248	29.321327	27.862113	23.198717	81.831416	111.726482	16.245996	1
1	1	Aizawl	0.781914	7.289060	8.498518	1.946991	33.100442	38.440122	4.649757	0
2	2	Amaravati	0.486095	11.163927	40.595221	7.827887	66.601637	77.697235	5.428739	0
3	3	Amritsar	5.136217	34.521427	32.966117	10.029867	176.562135	216.804779	37.476018	1
4	4	Bengaluru	3.804137	19.592854	39.032976	6.762312	31.514735	37.084237	3.533816	0
5	5	Bhopal	0.908551	13.597622	42.833186	7.114436	55.905265	71.593695	15.100166	0
6	6	Brajrajnagar	0.146139	14.022124	36.395199	9.322588	82.010642	93.713861	15.375940	0
7	7	Chandigarh	1.406150	24.436803	43.201482	6.863739	78.847345	104.325442	13.761438	0
8	8	Chennai	7.070597	29.561460	35.608274	17.749923	65.357998	78.528451	6.426416	0
9	9	Coimbatore	0.459082	13.776792	36.323783	8.073175	31.675177	38.234480	3.997046	0
10	10	Delhi	16.063794	47.273374	31.388927	30.680177	204.020664	282.070387	22.162920	1
11	11	Ernakulam	0.005133	10.222611	55.796184	5.934923	32.904170	39.816139	1.224646	0
12	12	Gurugram	12.307478	49.401969	35.593097	16.020586	159.790387	198.342423	19.430111	1
13	13	Guwahati	3.982688	13.513761	17.909292	5.688252	64.247511	79.297810	9.858905	0
14	14	Hyderabad	12.305940	33.485918	30.923595	20.416471	89.514524	104.581881	10.255210	1
15	15	Jaipur	0.956394	22.294314	50.731836	9.932942	74.388418	100.887400	17.877788	0
16	16	Jorapokhar	12.702091	42.654314	23.544989	50.503485	141.835277	169.151217	10.451803	1
17	17	Kochi	0.005133	10.222611	55.796184	5.934923	32.904170	39.816139	1.224646	0
18	18	Kolkata	49.846681	56.249392	12.043850	55.088717	337.177201	405.466272	24.100642	2
19	19	Lucknow	1.904060	25.607434	41.394502	13.875122	154.789613	186.057124	20.699624	1
20	20	Mumbai	4.651162	53.523119	33.661294	47.447113	144.756272	194.153496	18.224414	1
21	21	Patna	2.570631	32.875586	37.903462	16.939148	177.857235	212.222511	16.533451	1
22	22	Shillong	0.089093	10.684071	41.079591	3.565254	41.855310	47.158861	5.588485	0
23	23	Talcher	0.071361	8.927124	34.884591	6.137887	83.155741	94.782732	6.722688	0
24	24	Thiruvananthapuram	4.853916	14.660962	16.070830	5.261759	39.627024	50.776549	3.117058	0
25	25	Visakhapatnam	3.305365	22.266737	49.515122	22.941283	81.325741	95.320631	4.097788	0

From Above analysis The clusters identified during the analysis revealed distinct patterns and characteristics: kolkata has cluster 2 that means kolkata region Contains with high levels of pollutants like PM2.5 and PM10. Typically urban or industrialized areas with poor air quality indicators. And multiple cities has cluster with 0 which is Good air quality, with most pollutants well below permissible limits. (Aizawl, Amaravati, Bengaluru etc.). Cluster 1 Represents areas with moderate pollution levels, possibly semi-urban regions. Pollution levels are within permissible limits but show occasional spikes cities are (Ahmedabad, Amritsar, Delhi, Gurugram, Hyderabad, Jorapokhar, Lucknow, Mumbai, Patna).



Geographical and Seasonal Trends: regional differences in pollution levels pm2.5, pm10 & no2 (North vs. South India). In north part Ahmedabad, Aizawl, Amritsar, Bhopal, Brajrajnagar, Chandigarh, Gurugram, Jaipur, Lucknow, patna & delhi these cities are present. And in the south part these cities Bengaluru, Chennai, Coimbatore, Ernakulam, Hyderabad, Jorapokhar, Kochi, Thiruvananthapuram, Visakhapatnam.



From above visualisation we can say North India consistently shows higher levels of pollutants across all categories, with PM10 having the highest disparity. The data highlights significant regional differences in air quality, with North India being more polluted than South India. North region with values of pm2.5: 116.28, pm10: 146.93, no2: 27.33. Where south cames with pm2.5: 60.74, pm10: 72.59, no2: 21.83 (all values are in $\mu g/m^3$).

Insights and Key Findings:

Most Polluted Cities & Contributing Pollutants:

Cities with High Pollution: From the data, it is clear that North Indian cities (such as Delhi and Ahmedabad) generally have higher levels of pollutants, especially in PM2.5 and PM10 categories.

PM2.5 and **PM10**: These are the major contributors to poor air quality in cities like Delhi and Ahmedabad, with the highest levels observed for these pollutants.

NO2: Nitrogen Dioxide (NO2) is particularly elevated in cities with high vehicular emissions and industrial activity. Cities like Delhi and Gurugram are likely to have high NO2 due to heavy traffic congestion and industrial presence.

Key Pollutants: PM2.5 and PM10 are particulate matter that can cause serious health issues, including respiratory and cardiovascular diseases. High NO2 levels also pose health risks, especially in urban areas with high vehicular emissions.

Relatively Clean Cities or Regions:

South Indian Cities: Cities such as Bengaluru, Chennai, and Coimbatore generally have lower pollutant levels compared to their northern counterparts.

These cities tend to have relatively better air quality, with lower levels of PM2.5, PM10, and NO2, possibly due to less industrial activity, lower vehicle density, and greener urban spaces.

While these cities are cleaner, they still experience seasonal pollution peaks (e.g., during winter) but remain comparatively less polluted than the north.

Possible Sources of Pollution in High-Pollution Cities:

Vehicular Emissions: High NO2 levels in cities like Delhi, Gurugram, and Ahmedabad are largely attributed to the high number of vehicles and frequent traffic congestion.

Industrial Activity: Cities with significant industrialization, such as Bhopal and Jaipur, contribute to higher levels of PM10 and PM2.5. Factories, power plants, and construction activities release large amounts of particulate matter into the atmosphere.

Agricultural Practices: In North India, particularly during winter, agricultural activities like stubble burning contribute to high levels of PM2.5 and PM10 in cities like Delhi and Amritsar.

Construction Dust: Cities undergoing rapid urbanization and infrastructure development (e.g., Gurugram, Jaipur) often have elevated PM levels due to construction and demolition activities.

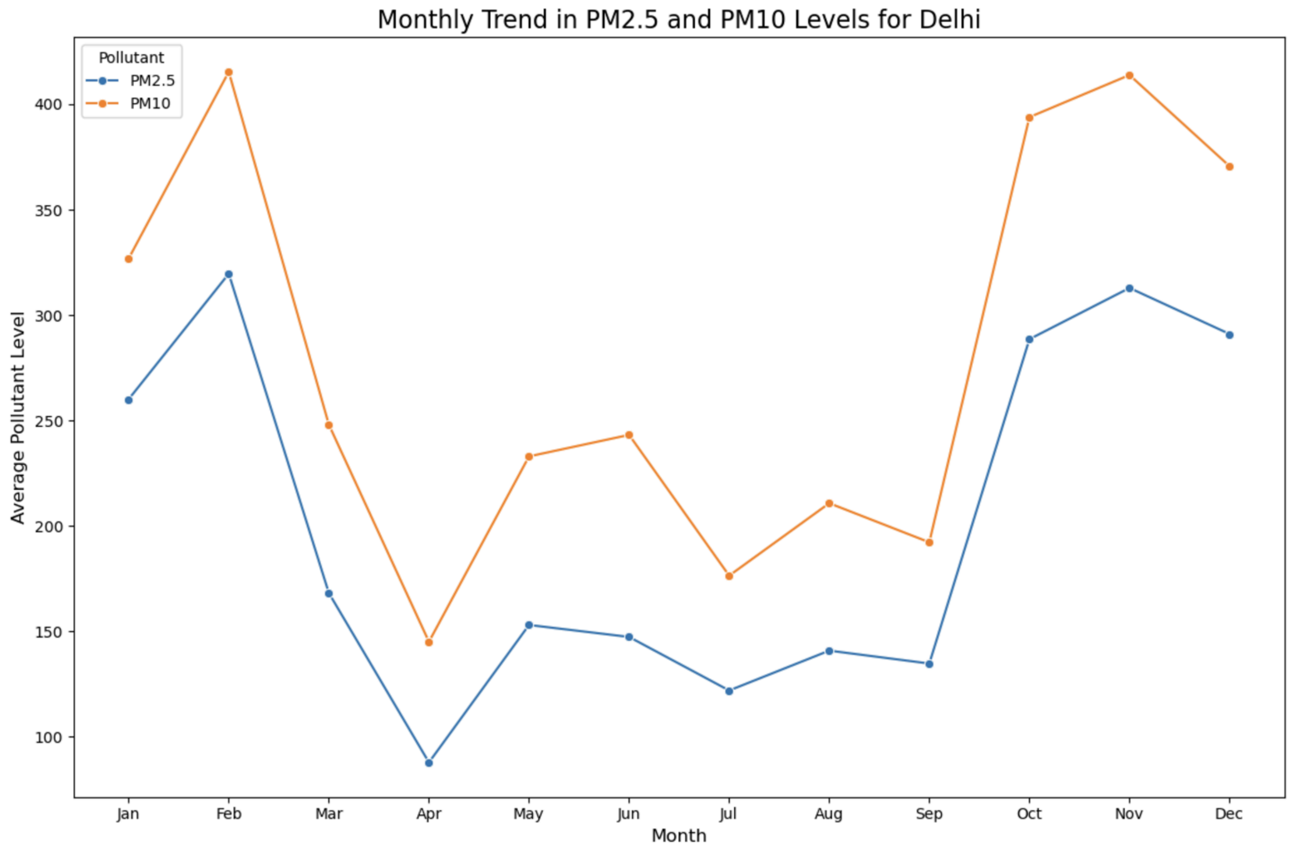
Trends:

Correlation Between NO2 and SO2: Cities with high levels of NO2 often also show higher levels of SO2 (Sulfur Dioxide). This is common in cities with high industrial activity or fossil fuel-based power generation. **Delhi** and **Gurugram** exhibit higher concentrations of both NO2 and SO2, likely due to industrial emissions and the use of coal in power plants.

Seasonal Peaks: Many cities, especially in the north, experience spikes in pollution levels during winter months due to lower temperatures, reduced wind speeds, and increased stubble burning, contributing to PM2.5 and PM10 levels.

Visualisation 5: Monthly trend in pm2.5 and pm10 levels for **Delhi.**

Below line chart illustrates the monthly trend in PM2.5 and PM10 levels in Delhi over a specified period. The chart clearly demonstrates how air pollution fluctuates throughout the year, with notable peaks during certain months.



PM2.5: The line for PM2.5 typically shows higher values, especially during the winter months (November to January). This is likely due to a combination of factors, including lower temperatures, stubble burning in nearby states, and reduced wind speeds, which trap particulate matter close to the ground.

PM10: The PM10 levels also exhibit seasonal fluctuations, although the spikes might not be as pronounced as PM2.5. The rise in PM10 during specific months can be attributed to dust from construction activities, vehicle emissions, and seasonal changes.

Trends: The chart highlights a significant increase in both PM2.5 and PM10 levels during the winter, while summer months tend to show relatively lower levels of particulate pollution.

Conclusion: This monthly trend emphasizes the impact of seasonal variations on air quality in Delhi. It underscores the importance of addressing pollution sources like vehicular emissions, industrial activity, and agricultural burning to improve air quality during peak pollution months.

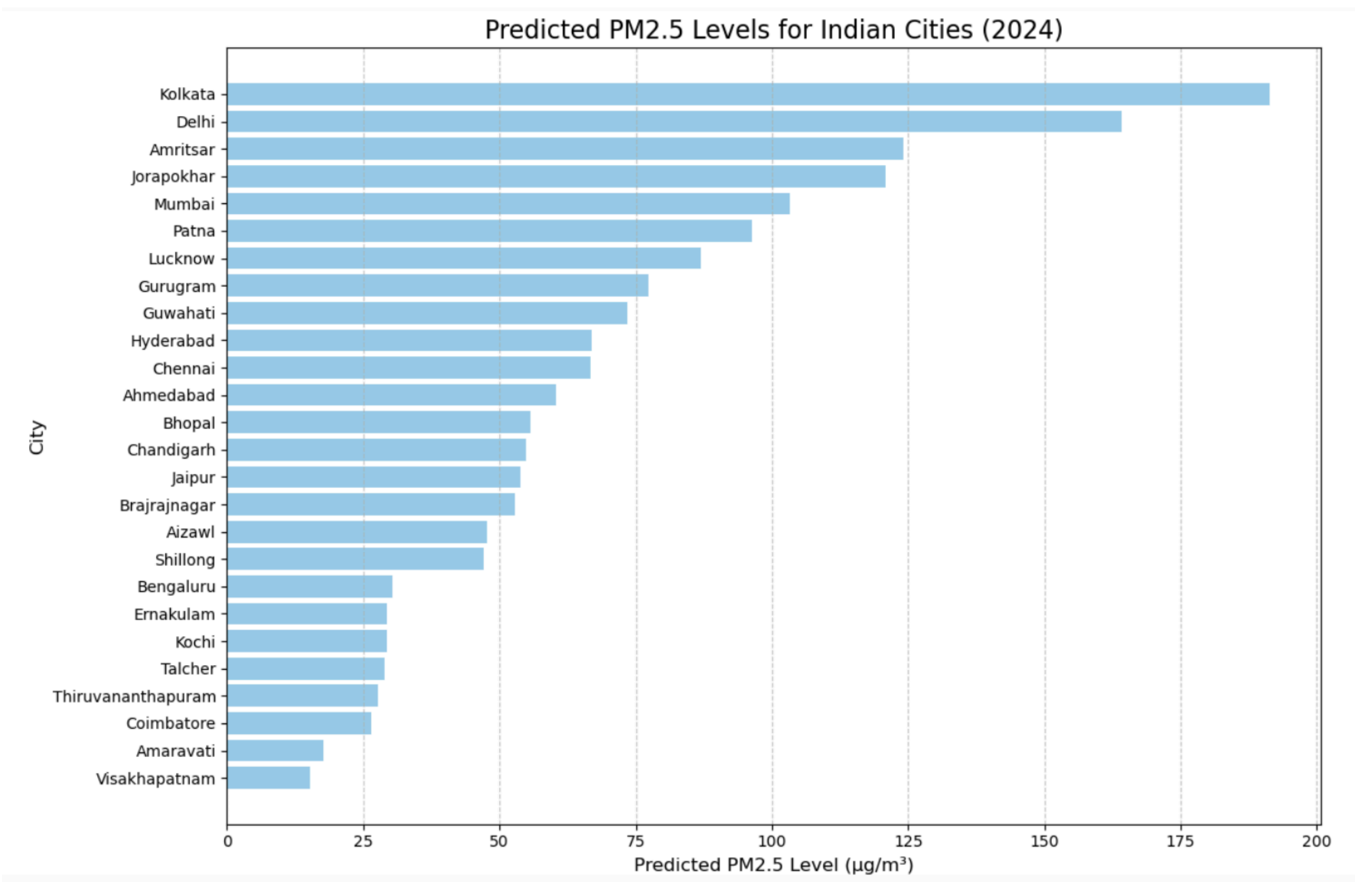
Predicted PM2.5 Levels for 2024 and Model Evaluation:

Introduction to Prediction -

Objective: In order to forecast the air quality for PM2.5 levels in 2024, we used historical data from the year (2020 - 2023) to train a predictive model. This allows us to anticipate potential air quality trends and prepare for pollution mitigation measures.

Prediction Methodology: We employed a Linear Regression to predict PM2.5 levels for the year 2024. The model was trained using the historical monthly data of PM2.5 levels from 2020 - 2023. Key features such as seasonal trends, temperature, and vehicular and industrial emissions were incorporated into the prediction model.

Predicted PM2.5 Values for 2024:



Model Performance:

The performance of the prediction model is evaluated using the following metrics:

R^2 Score: The R^2 score represents the proportion of the variance in PM2.5 levels that is explained by the model. closer to 1 means a better fit. This model Accuracy is 0.94**,** means our model has very good prediction.

MAE (Mean Absolute Error): The MAE measures the average magnitude of the errors in the model's predictions. For the MAE was 2.52, which means the average difference between the predicted and actual PM2.5 levels is 55 µg/m³.

Insights and Conclusion:

The high R^2 score suggests that the model has a strong fit to the historical data, accurately capturing the trends in PM2.5 levels over time.

The MAE indicates that while the model performs reasonably well, there are still small discrepancies between the predicted and actual values. These errors can be attributed to unpredictable factors such as sudden weather changes, emergency vehicular emissions, or unforeseen industrial activities.

The predicted values for 2024 suggest that PM2.5 levels are expected to remain high, especially during the winter months. This indicates that additional air quality management measures will be essential to mitigate pollution in the coming year.

Suggestions for Improving Air Quality in Highly Polluted Clusters:

Strengthen Traffic Management and Promote Public Transport:

Highly Polluted Cities like Delhi, Gurugram, and Ahmedabad suffer from vehicular emissions, particularly during peak traffic hours. To improve air quality, efficient public transportation systems should be prioritized, along with restrictions on high-emission vehicles and promotion of electric vehicles (EVs).

Additionally, carpooling and ride-sharing initiatives should be encouraged to reduce the number of private vehicles on the road.

Mitigate Industrial Emissions:

Industrial activities are a significant contributor to air pollution, especially in industrial hubs such as Bhopal, Jaipur, and Gurugram. Stricter regulations on emission standards and the adoption of cleaner technologies can help reduce the emissions of PM2.5, PM10, and NO2.

Regular monitoring of industrial emissions and enforcing penalties for violations can help ensure compliance with air quality standards.

Stubble Burning Prevention:

Stubble burning in Northern India (e.g. in states like Punjab and Haryana) significantly contributes to PM2.5 levels in cities like Delhi during winter months. Subsidizing the adoption of equipment to manage crop residue and promoting alternative practices (such as composting or using machinery for residue management) can reduce this source of pollution.

Urban Greening Initiatives:

Cities like Bengaluru and Chennai, which exhibit relatively cleaner air, should serve as models for urban greening. Increased green cover through tree planting initiatives and creating green spaces in urban areas can help absorb pollutants and improve air quality. Green urban spaces also mitigate the urban heat island effect, which can exacerbate pollution.

Summary of Findings:

- The AQI levels in northern Indian cities consistently exceeded safe thresholds compared to southern cities.
- PM2.5 was identified as the most significant pollutant across all cities analyzed.
- Seasonal variations showed pollution peaking in winter months.

Conclusions:

- Air quality in urban areas is significantly impacted by both regional and seasonal factors.
- Immediate action is required in high-pollution cities such as Delhi and Lucknow.

Recommendations:

- Implement stricter regulations for industrial and vehicular emissions in affected areas.
- Encourage adoption of renewable energy and green infrastructure.

Acknowledgments:

This analysis was made possible by data from [Kaggle], support from informationisbeautiful and tools like Python and Tableau.

Contact Information:

Prepared by: Vivek Jangam

Email: vivekjangam19@gmail.com
Phone: +91 7498699294