Pollution Chronicles: A Tale of 8 Indian Cities and 8 Hidden Culprits

A Descriptive and Inferential Study of Air Quality

By
Vivekananda Mallik

Supervised by
Mr. Anjan Samanta
Assistant professor
Department of Applied Statistics
Maulana Abul Kalam Azad University of Technology

January 9, 2024

Contents

1 Abstract

Pollution Chronicles: A Tale of 8 Indian Cities and 8 Hidden Culprits,this project aims to determine the air quality of 8 cities across India including Delhi, Kolkata from 1^{st} January 2020 to August 2023 by analyzing daily emission of PM2.5, PM10, NO2, NH3, NO, NOx, SO2, CO. The most dangerous pollutants in each city is determined by Mann-Whitney U test. It has been found that PM10 and PM2.5 are the most dangerous pollutants for each city. Using the same test it is also found that effect of PM2.5 and PM10 are high in Delhi. Cities like Visakhapatnam is highly affected by the pollutants like NO2, NOx, NH3 because of the presence of various heavy industries. Ambala is affected by the emission of NH3 because of ammonia-based fertilizers and agricultural practices in the region. It has been found that in Winter the pollution level of different pollutants increase in almost every city and in Summer they tend to decrease. Using unsupervised machine learning techniques Cities are grouped together according to the pollution pattern.

2 Introduction

"Pollution Chronicles: A Tale of 8 Indian Cities and 8 Hidden Culprits" embarks on a journey that explores the multi-faceted dimensions of pollution in one of the world's most populous nations. From the bustling metropolises of Delhi and Kolkata to the serene landscapes of Jaipur, we delve into the stark contrasts and shared challenges faced by 8 diverse Indian cities. Along the way, we uncover the hidden culprits responsible for the deterioration of air quality.

Air pollution is a major subgroup of environmental pollution which poses a serious threat to the ecosystem. The risk of global sustainability can be reduced by controlling anthropogenic activities responsible for the emission of air pollutants in the environment. India accounts for having one of the most polluted capitals and cities within the globe. Over 120000 people died in India in 2020 as a result of air pollution and related problems [8]. An estimated 11000 and 54000 avoidable deaths in Hyderabad and Delhi, respectively, in 2020, have been attributed to air pollution [7]. Not only it has impacted the health but also impacted the economy heavily. It's economic cost is estimated to exceed $150 billion. Epidemiological studies of air pollution and children's lung function reveals that: (1) living in areas of high air pollution is associated with lower lung function, (2) chronic exposure to elevated level of air pollution is associated with lower rates of lung function growth, (3) improvement in air quality leads to improvement in lung function level and/or growth rate, and (4) children who spend a significant amount of time outdoors in polluted environments or those with poor nutrition may be more strongly affected by air pollution [6]. The most significant reduction was observed for nitrogen dioxide (NO2) (3–79 percent) and carbon monoxide (CO) (2–61 percent), pollutants which are mainly related to traffic emissions. Ozone (O3) showed a mixed trend with increasing levels at some cities which may be attributed to lower titration of O3 by NO. Maximum reduction observed in PM10 and PM2.5 was 58 and 57 percent [2], respectively during the lockdown period in 2020 as compared to the previous year. Air quality of the cities also improved in 2020. During the lockdown period in 2020, AQI of only 15 percent of cities was in the 'Unhealthy' category (151–200) while in 2019, 56 percent of cities were in the 'Unhealthy' category. In Ghaziabad and Patiala all the pollutants showed significant reduction after lockdown implementation except O3. Diurnal patterns of PM10, PM2.5, CO and NO2 showed lower values during lockdown period in 2020 with less distinct bimodal patterns as compared to 2019. [3]

In this project we will tour the 8 cities across India, revealing their pollution pattern and similarities. The project is going to answer following questions such as Which gas is affecting most on country's capital? , Which cities have similar type of pollution?, Which season of the year is most polluted? and many more. The project is based on the daily data from 1st January 2020 to August 2023 and the data has been collected from the website of Central Pollution Control Board, a Government of India website. The techniques that are applied here from Statistical inference to Unsupervised machine learning. Based on the nature of the data, non-parametric statistical techniques such as Kruskall-Wallis test to Mann-Whitney U test are applied for the evaluation. Hierarchical clustering is used to evaluate the similarities among the cities. The techniques that are mentioned above have helped us to determine the pollutants having more impact on a city.

3 Literature review

[9] In this study, we characterize the impacts of COVID-19 on air pollution using NO2 and Aerosol Optical Depth (AOD) from TROPOMI and MODIS satellite datasets for 41 cities in India. Specifically, our results suggested a 13 NO2 reduction during the lockdown (March 25–May 3rd, 2020) compared to the pre-lockdown (January 1st–March 24th, 2020) period. Also, a 19 reduction in NO2 was observed during the 2020-lockdown as compared to the same period during 2019. The top cities where NO2 reduction occurred were New Delhi (61.74 percent), Delhi (60.37 percent), Bangalore (48.25 percent), Ahmedabad (46.20 percent), Nagpur (46.13 percent), Gandhinagar (45.64) and Mumbai (43.08 percent) with less reduction in coastal cities. The temporal analysis revealed a progressive decrease in NO2 for all seven cities during the 2020 lockdown period. Results also suggested spatial differences, i.e., as the distance from the city center increased, the NO2 levels decreased exponentially.

[5] The analysis reveals that in many of the Indian States emission sources that are outside of their immediate jurisdictions make the dominating contributions to (population-weighted) ambient pollution levels of PM2.5. Consequently, most of the States cannot achieve significant improvements in their air quality and population exposure on their own without emission reductions in the surrounding regions, and any cost-effective strategy requires regionally coordinated approaches. Advanced technical emission control measures could provide NAAQS-compliant air quality for 60 percent of the Indian population. However, if combined with national sustainable development strategies, an additional 25 percent population will be provided with clean air, which appears to be a significant co-benefit on air quality (totaling 85 percent).

[4] 13 of the worst 20 cities in terms of air quality lie in India, as highlighted by a recent WHO report. The Indian capital, New Delhi leads the list with the worst air quality. In recent times, the share of eco-health problems prevalent in various Indian cities has risen significantly resulting in a spike of respiratory and cardiac diseases. Ground measurements of SO2, NO2, PM2.5 and PM10 levels are significantly above National Ambient Air Quality Standard (NAAQS) for most cities and for New Delhi, it has been found that at current rate and policies PM2.5 will not reach the recommended NAAQS values even by 2030.

[1] The average reduction in the concentration of NO, NO2, NOX, PM2.5, and O3 between 01 March and 12 May 2020 was found to be 63 percent, 48 percent, 48 percent, 18 percent, and 23percent respectively when compared to the same period in 2019. Similarly, the comparative analysis of pollutant concentrations between pre-lockdown (01–23 March 2020) and lockdown (24 March–12 May 2020) periods has shown a huge reduction in the ambient concentration of air pollutants, 47.3 percent (NO), 49 percent (NO2), 49 percent (NOX), 10 percent (SO2), 37.7 percent (PM2.5), and 15.6 percent (O3), resulting in improved air quality over Bangalore during the COVID-19 lockdown period. It is shown that the strict lockdown resulted in a significant reduction in the pollution levels. Such lockdowns may be useful as emergency intervention strategies to control air pollution in megacities when ambient air quality deteriorates dangerously.

4 Objective

As it was said in the introduction that 8 cities data have been collected each having 8 pollutants. The objective of this study are as follows:

1. Find worst pollutants of each city.

2. Find the highly affected city by the emission of each pollutant .

3. Find the season of a year when the emission increases or decreases.

4. To find which cities have same pollution pattern.

5 Methodology

5.1 Data acquisition

Data has been collected from the website of Central Pollution Control Board https://cpcb.nic.in/. Data set contains 8 different cities named Kolkata, Howrah, Delhi, Asansol, Brajarajnagar, Jaipur, Ambala, Visakhapattanam. Each city have 8 differnet pollutants named PM2.5 (ug/$\mathbf{m^3}$), PM10 (ug/$\mathbf{m^3}$), NO (ug/$\mathbf{m^3}$), NO2 (ug/$\mathbf{m^3}$), NOx (ppb), NH3 (ug/$\mathbf{m^3}$), SO2 (ug/$\mathbf{m^3}$), CO (ug/$\mathbf{m^3}$). Particulate Matter 2.5, It refers to tiny, inhalable particles or aerosols in the air that have a diameter of 2.5 micrometers (µm) or smaller. Particulate Matter 10, It refers to airborne particles or aerosols with a diameter of 10 micrometers (µm) or smaller. Data is collected during the period from 1^{st} January 2020 to 21^{st} August 2023.

PM2.5	**PM10**	**NO**	**NO2**	**NOx**	**NH3**	**SO2**	**CO**	**Date**
114.95	235.47	56.93	47.88	105.36	44.98	12.31	1.6	01-01-2020
125.01	193.7	39.3	37.61	77.27	40.46	9.47	1.26	02-01-2020
52.55	83.7	16.18	21.65	38.16	34.91	5.01	0.47	03-01-2020
42.87	73.82	57.26	23.45	81.04	34.23	3.55	1.13	04-01-2020
86.68	122.62	78.24	22.77	101.32	35.75	3.25	1.19	05-01-2020

Table 1: This is the data for the city Kolkata

5.2 Data preprocessing

It is a crucial step in the data analysis pipeline that involves cleaning, transforming, and organizing raw data into a format suitable for analysis. The goal of data preprocessing is to ensure that the data is accurate, consistent, and ready for modeling or analysis.

5.2.1 MICE

Multiple Imputation by Chained Equations (MICE) is a statistical method used for handling missing data in a data set. The key idea behind MICE is to impute missing values by iteratively modeling each variable with missing data conditional on the observed values of the other variables. The process is repeated multiple times to create several imputed data sets, and the final analysis is performed separately on each imputed data set. Initialize the missing values with some initial guesses (Say mean) (Iteration 0). For each column that contains missing values, replace the null values with that column's corresponding mean. Now start from left side columns and then gradually moves from left to right columns. Make the extreme left column's computed null values again null. That column become our response column. Taking other columns as predictors we use any model (Linear regression, Random Forest etc), to predict the null values of that column. Repeat the step for each column.

Now our all columns values are predicted and denote the dataset as Iteration 1. The next step is find difference of Iteration 1 and Iteration 0. It is obvious that only the missing values that are imputed will have some difference, others values are zero. Our aim is to make those differences also zero. What we do next is that: Taking Iteration 1 data set as primary data set, we repeat the above procedure again and again until the difference become zero or near to zero for all the values. Generally we iterate it 30 to 40 times to get the final data set.

5.2.2 Data integration

MICE is applied separately for all the cities. Now we have combined the data of 8 cities as a single data frame that will be necessary for implementing clustering techniques.

5.3 Proposed methodology

5.3.1 Kruskal-Wallis Test

It is a non-parametric statistical hypothesis testing techniques. It is the non-parametric version of ANOVA. When the data sets do not satisfy the assumptions of ANOVA such as i. The data set comes from normal distribution ii.Variance are equal for all the data sets, then it is needed to use Kruskal-Wallis test for comparing more than two data sets.

Testing problem:

H_0 : There are no significant difference between the data sets

H_1 : There are significant differences between at least two of the groups

Test statistic:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1) \quad \text{where} \quad H \sim \chi^2(k-1)$$

R_i^2 : Sum of ranks in the i^{th} sample $\quad$ n_i : Size of the i^{th} sample

N : Total sample size $\quad$ k : Number of samples

Level of significance: $\alpha = 0.05$
Based on the critical value reject or accept null hypothesis.

5.3.2 Mann Whitney U test

Mann Whitney U test is a non parametric hypothesis testing techniques alternative to two sample t test. In this case the testing problem is as followed:
Testing problem:

H_0 : There are no significant difference between two independent data sets

H_1 : There are significant differences between two data sets

Test statistic:

$$U_1 = n_1.n_2 + \frac{n_1(n_1+1)}{2} - \sum_{i=1}^{n_1} R_i$$

$$U_2 = n_1.n_2 + \frac{n_2(n_2+1)}{2} - \sum_{j=1}^{n_2} R_j$$

$U = \min(U_1, U_2)$: This is the test statistic for the test.

$$\sum_{i=1}^{n_1} R_i : \text{Sum of ranks in the 1}^{\text{st}} \text{ sample} \quad \sum_{j=1}^{n_2} R_j : \text{Sum of ranks in the 2}^{\text{nd}} \text{ sample}$$

n_1 : Size of the 1$^{\text{st}}$ sample $\qquad$ n_2 : Size of the 2$^{\text{nd}}$ sample

$$E(U) = \frac{n_1 \cdot n_2}{2}$$

$$\text{std. dev} = \sqrt{\frac{n_a \cdot n_b \cdot (n_a + n_b + 1)}{12}}$$

$$Z = \frac{U - E(U)}{\text{std. dev}} = \frac{U - \frac{n_a \cdot n_b}{2}}{\sqrt{\frac{n_a \cdot n_b \cdot (n_a + n_b + 1)}{12}}} \quad Z \sim N(0, 1)$$

Level of significance: $\alpha = 0.05$

Compare the obtained Z value and the critical Z value to determine whether to retain or reject the null hypothesis.

5.3.3 Mann-Kendall test

The Mann-Kendall test, also known as the Mann-Kendall trend test, is a non-parametric statistical test used to detect trends in time series data to access whether a particular data series exhibits a monotonic trend (either increasing or decreasing) over time. This test is especially valuable when dealing with data that may not adhere to normal distribution assumptions and contains tied values or outliers.
Testing problem:

H_0 : There is no significant trend in data set
H_1 : There is significant trend in the data (either increasing or decreasing)

Test statistic: The Mann-Kendall test statistic, denoted as 'S,' quantifies the strength and direction of any trend present in the data. It is computed as follows:

$$S = \sum_{i<j} \text{sign}(x_j - x_i)$$

where x_i and x_j represent the values of the data series at time points i and j, respectively.
The variance of the Mann-Kendall test statistic can be calculated to assess the significance of the trend. It is computed using the following formula:

$$\text{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{t_i}(t_i(t_i - 1)(2t_i + 5))}{18}$$

t: is the number of tied groups in the data series. t_i : is the number of data points in each tied group. n: Number of data points.

Kendall's Tau (τ): Kendall's Tau is a related measure used to quantify the strength of correlation in ranked data. It can be calculated from the Mann-Kendall test statistic and is often used to describe the direction and magnitude of the trend.

$$\tau = \frac{2S}{n(n-1)}$$

and next we calculate

$$Z = \frac{\tau - 1}{\mathrm{Var}(\tau)}$$

The critical value of Z is obtained from the standard normal distribution based on the desired level of significance (usually 0.05).

The primary test statistic 'Z' is used to assess the significance of the trend. The sign and magnitude of 'S' help determine whether there is a significant trend in the data. The sign of 'S' (positive or negative) indicates the direction of the trend (increasing or decreasing).

Kendall's Tau is a measure of correlation and concordance. It quantifies the strength and direction of the relationship between data points and provides additional information about the trend. If τ is positive, it indicates a positive (increasing) association, and if it is negative, it suggests a negative (decreasing) association. The magnitude of τ represents the strength of the association. While τ is informative, it is often used in conjunction with 'S' to provide a more complete picture of the trend.

5.3.4 Hierarchical clustering

It is one type of clustering technique in unsupervised machine learning. There are two types of hierarchical clustering one is i. Agglomerative clustering and ii. Divisive clustering. Agglomerative clustering is used most of the time. Here, how it's work: consider six cities.

In Agglomerative clustering each city is considered as cluster.Suppose Kolkata and Howrah are closer to each other, they are merged together and consider as cluster 1. Store the result in a figure where Kolkata and Howrah are store together as cluster 1. Next Amaravati and Delhi are merged together and considered as cluster 2, stored it in the Figure. Now Asansol is nearer to the cluster 1. So another cluster is made considering cluster 1 and Asansol, named it cluster 3. Same process has been done for cluster 2 and Hyderabad, named it cluster 4. At the end cluster 3 and cluster 4 are formed as one cluster. As it is seen above that clusters are merged and store in a tree shape pattern. This tree shape pattern is known as Dendrogram. Dendrogram provides the information that, in which order, the clusters are merged.

6 Result and analysis

For PM2.5 AQI is calculated and based on AQI value on each day, the categories are assigned. There are 6 categories of AQI, Good (0-50), Satisfactory (50-100), Moderate (100-200), Poor (200-300), Very poor (300-400), Severe (400-500). The reason for showing the graphs only for pollutant PM2.5 is that it is the most harmful pollutants, so it will be more interested to know about it's emission.

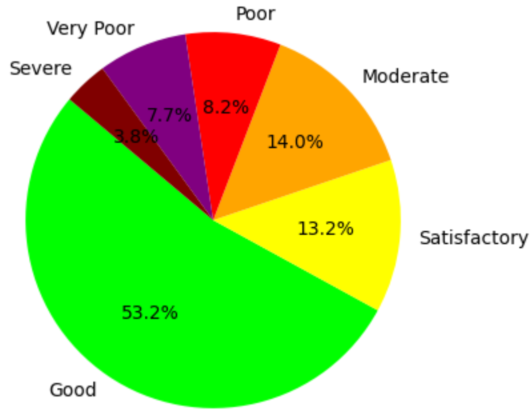


Figure 1:Kolkata's PM2.5 in 2020

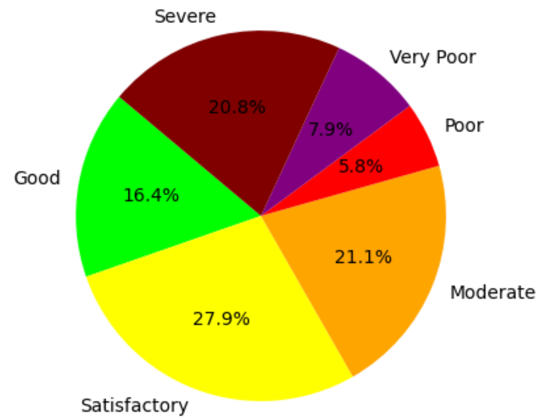


Figure 2:Delhi's PM2.5 in 2020

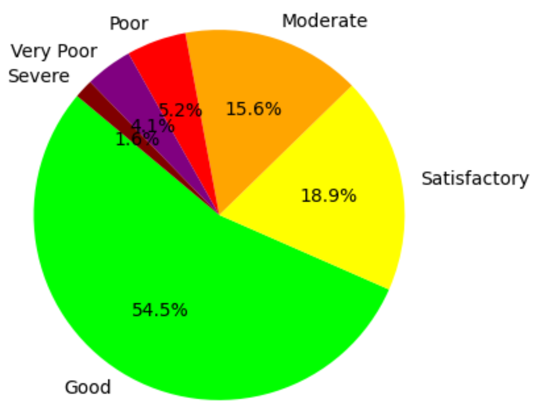


Figure 3:Howrah's PM2.5 in 2020

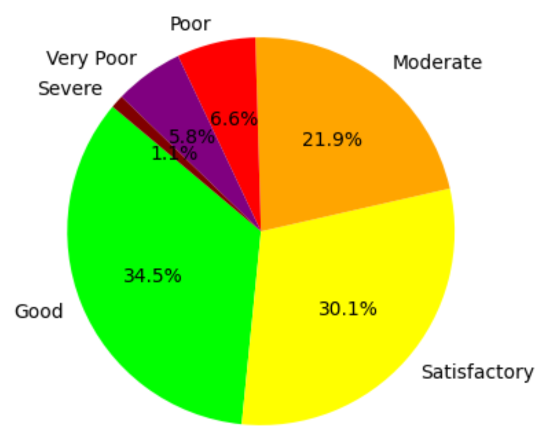


Figure 4:Asansol's PM2.5 in 2020

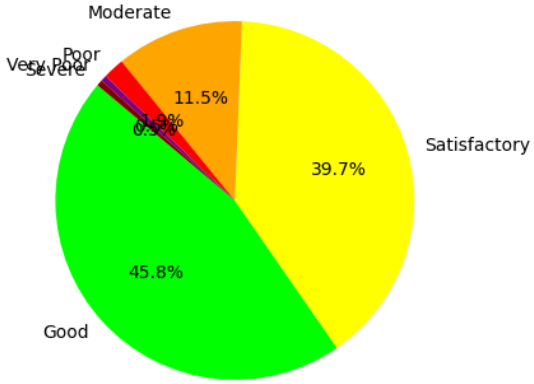


Figure 5:Jaipur's PM2.5 in 2020

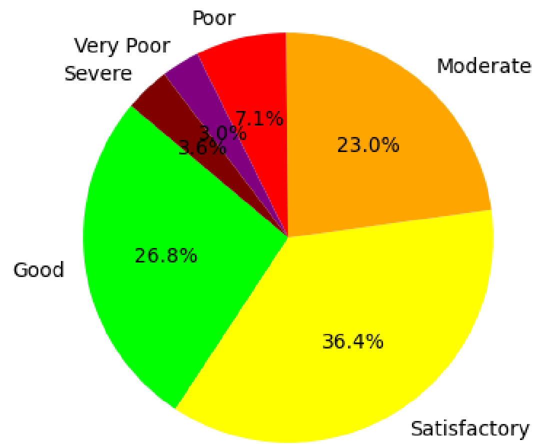


Figure 6:Ambala's PM2.5 in 2020

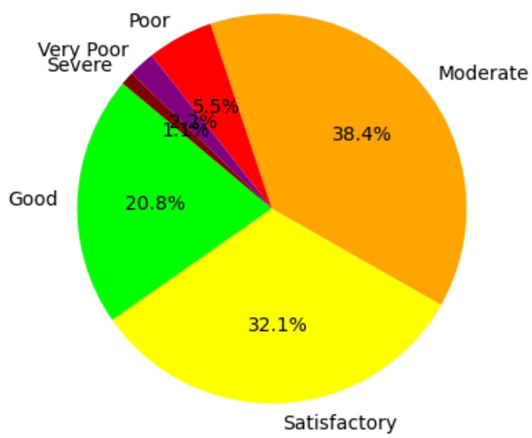


Figure 7:Brajarajnagar's PM2.5 in 2020

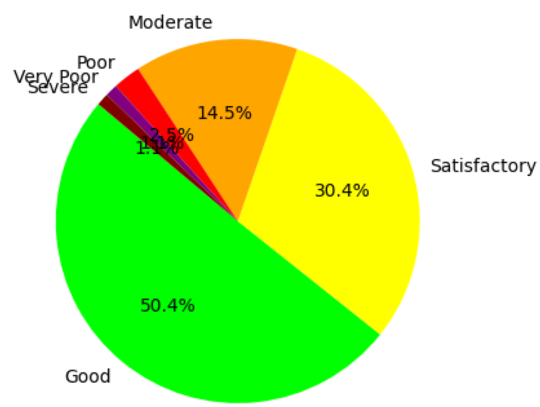


Figure 8:Visakhapattanam's PM2.5 in 2020

Pie diagrams show that in 2020 Kolkata, Howrah, Visakhapattanam were least polluted than the other cities. Delhi had the worst PM2.5 emission. 32 percent time of the year PM2.5 emission of Delhi were between poor to severe condition.

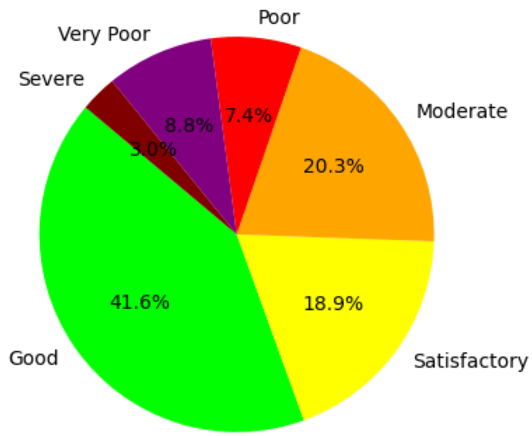


Figure 1:Kolkata's PM2.5 in 2021

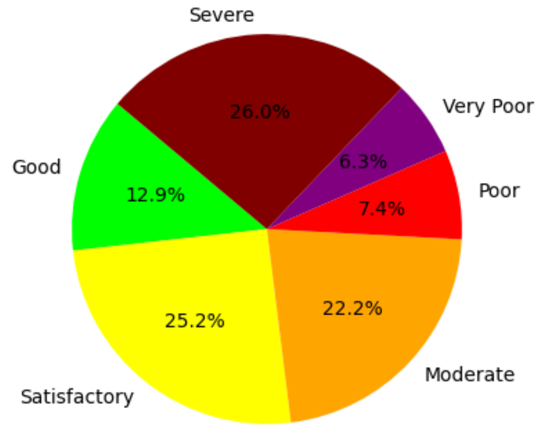


Figure 2:Delhi's PM2.5 in 2021

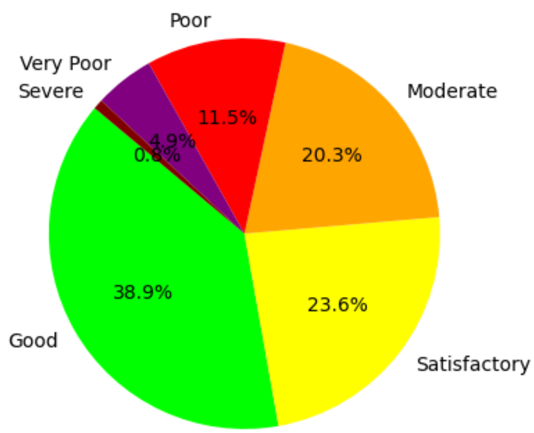


Figure 3:Howrah's PM2.5 in 2021

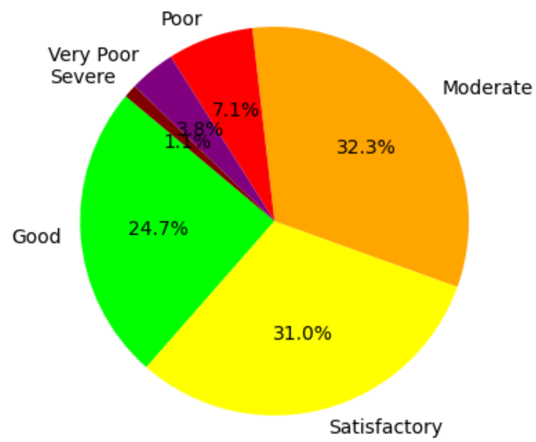


Figure 4:Asansol's PM2.5 in 2021

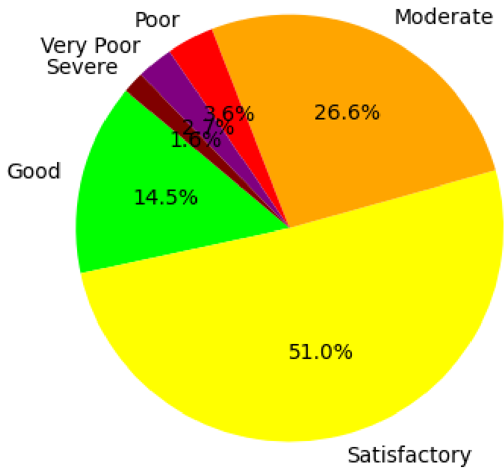


Figure 5:Jaipur's PM2.5 in 2021

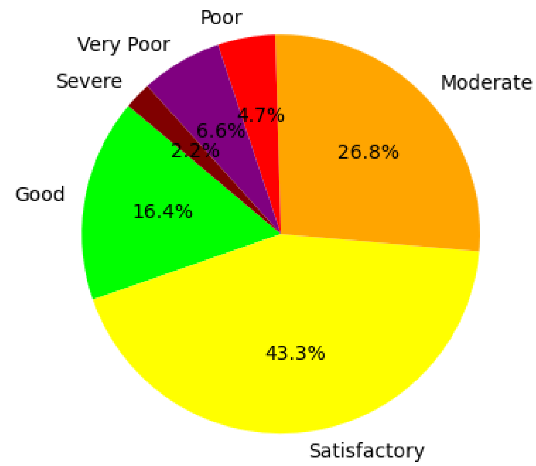


Figure 6:Ambala's PM2.5 in 2021

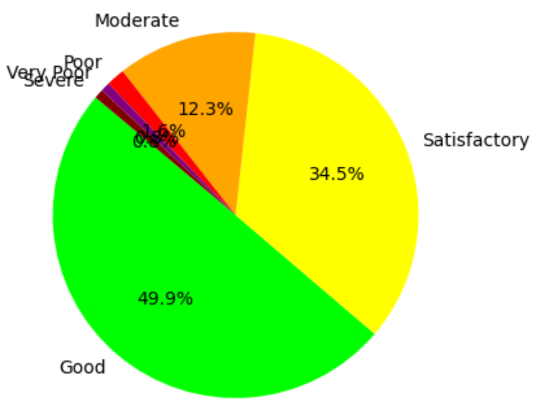


Figure 7:Brajarajnagar's PM2.5 in 2021

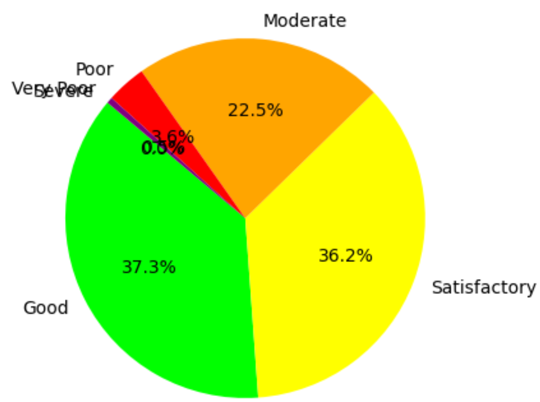


Figure 8:Visakhapattanam's PM2.5 in 2021

In 2021 Brajarajnagar among the least polluted cities in terms of PM2.5 emission. This year Delhi's PM2.5 emission got even worse than 2020.

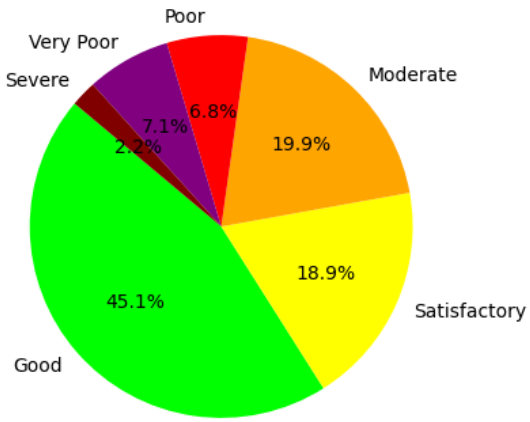


Figure 1:Kolkata's PM2.5 in 2022

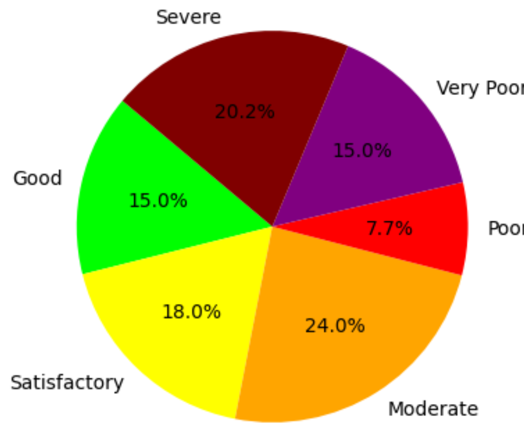


Figure 2:Delhi's PM2.5 in 2022

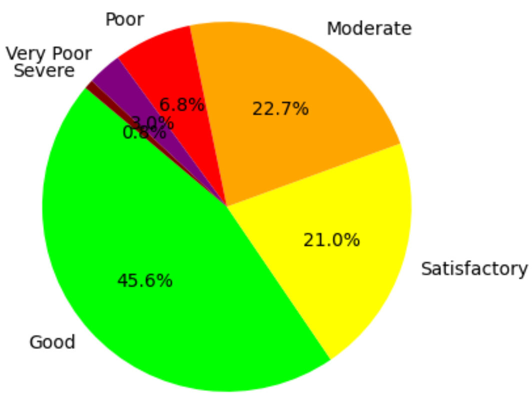


Figure 3:Howrah's PM2.5 in 2022

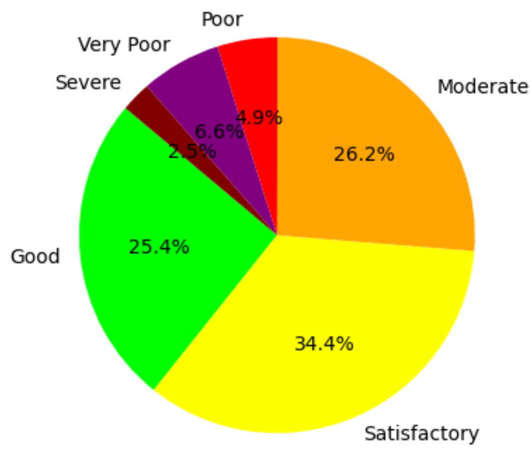


Figure 4:Asansol's PM2.5 in 2022

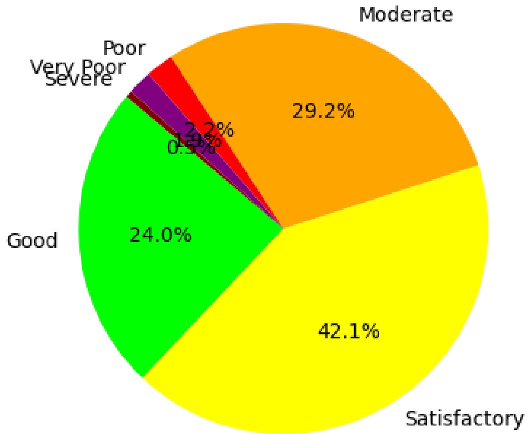


Figure 5:Jaipur's PM2.5 in 2022

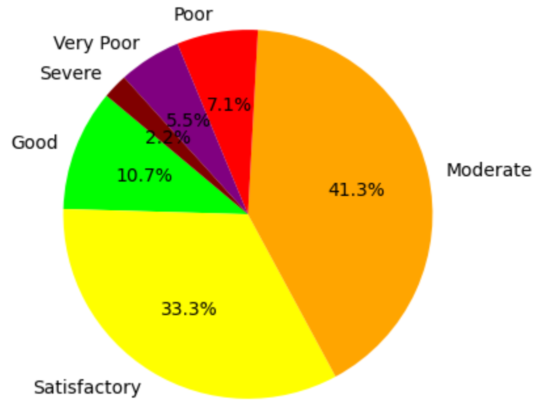


Figure 6:Ambala's PM2.5 in 2022

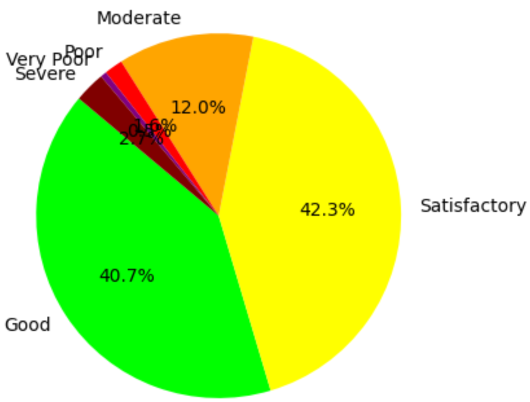


Figure 7:Brajarajnagar's PM2.5 in 2022

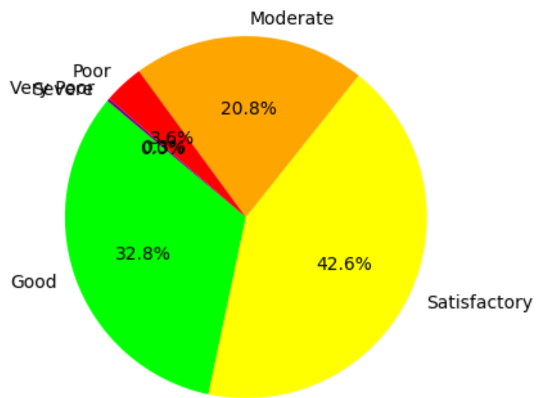


Figure 8:Visakhapattanam's PM2.5 in 2022

PM2.5 emission of Delhi were between poor to severe condition 42.9 percent time of the year. In case of Kolkata it was 16.1 percent. Visakhapattanam was lowest with 3.6 percent.

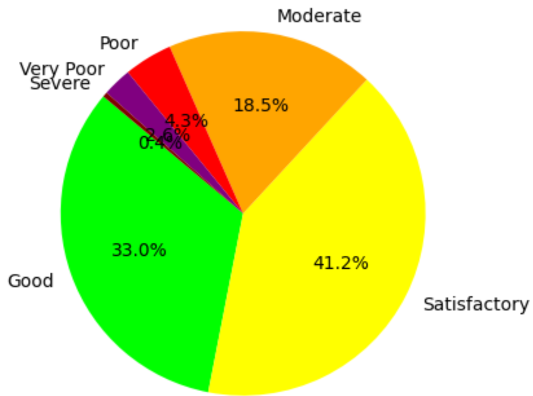


Figure 1:Kolkata's PM2.5 in 2023

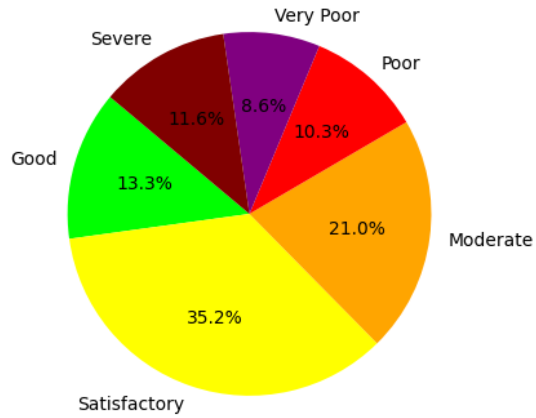


Figure 2:Delhi's PM2.5 in 2023

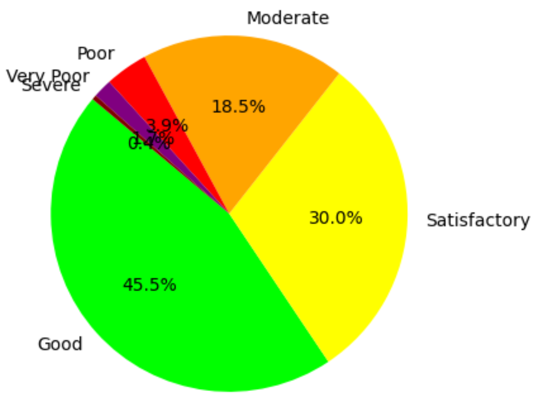


Figure 3:Howrah's PM2.5 in 2023

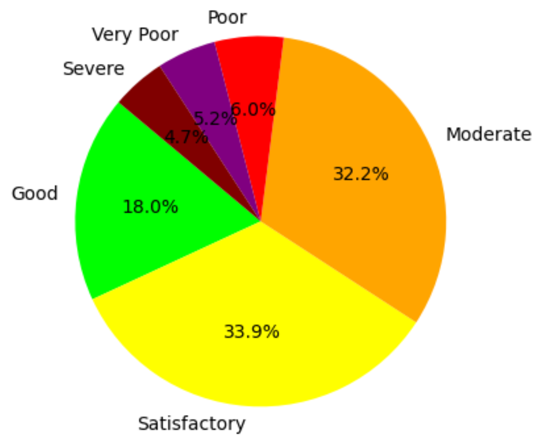


Figure 4:Asansol's PM2.5 in 2023

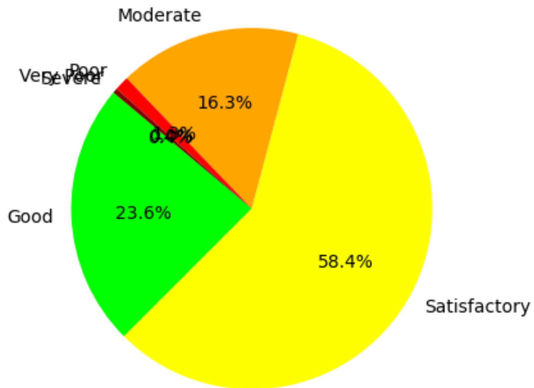


Figure 5:Jaipur's PM2.5 in 2023

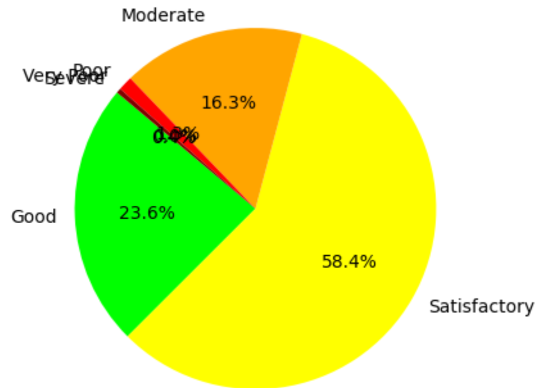


Figure 6:Ambala's PM2.5 in 2023

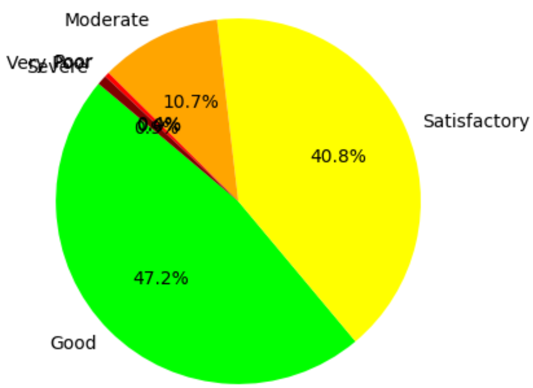


Figure 7:Brajarajnagar's PM2.5 in 2023

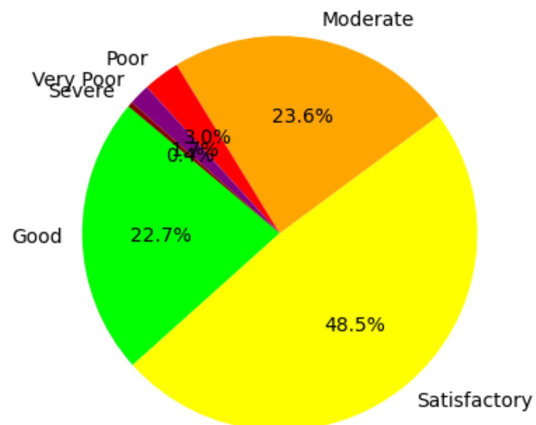
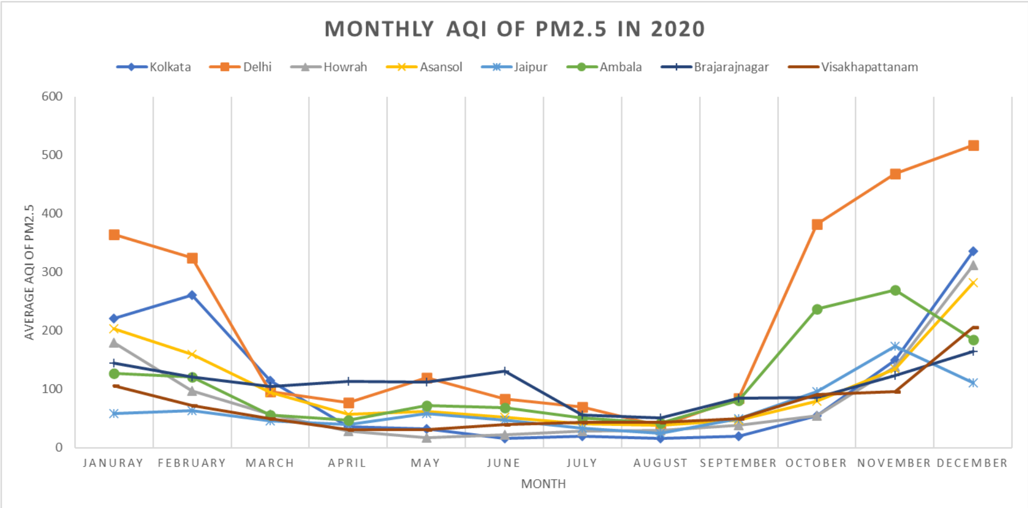
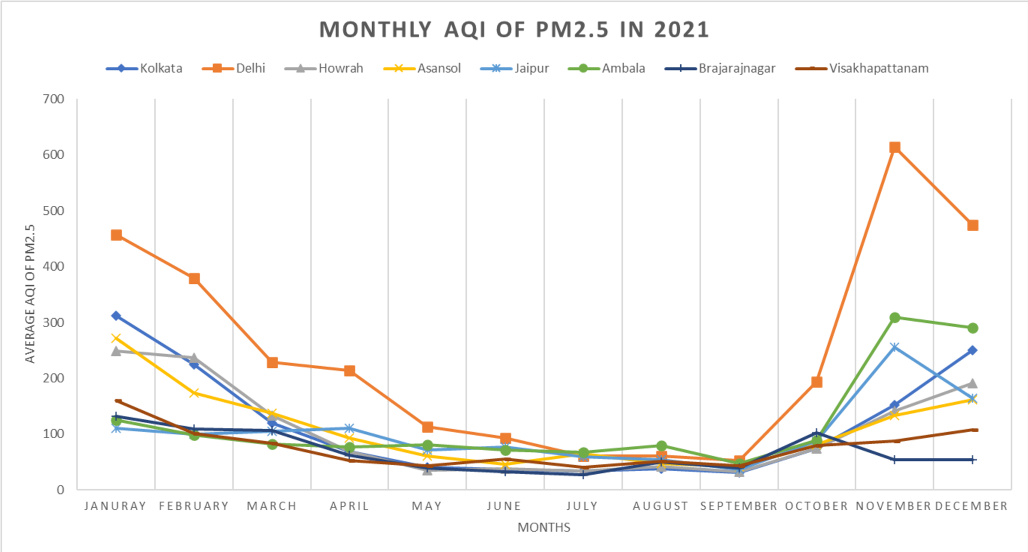
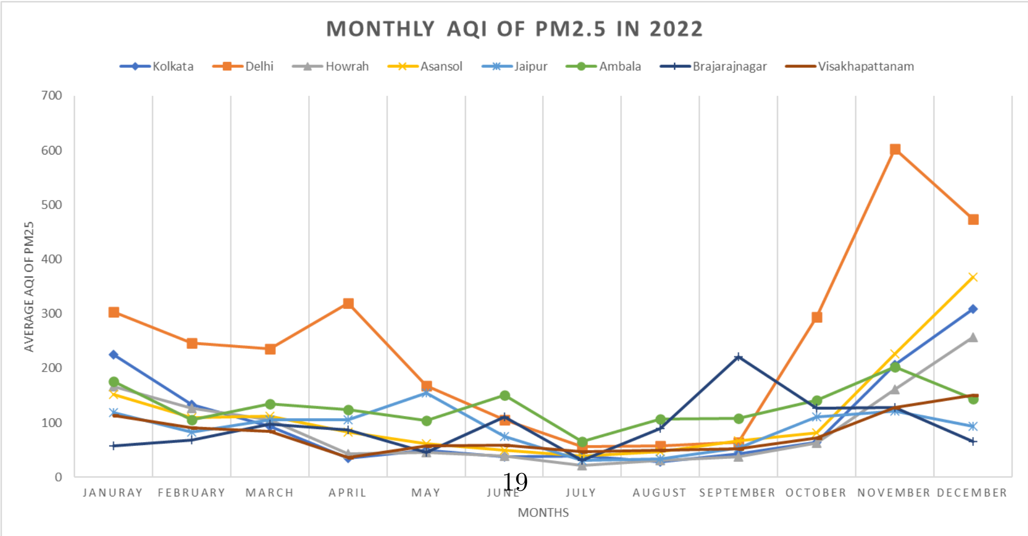


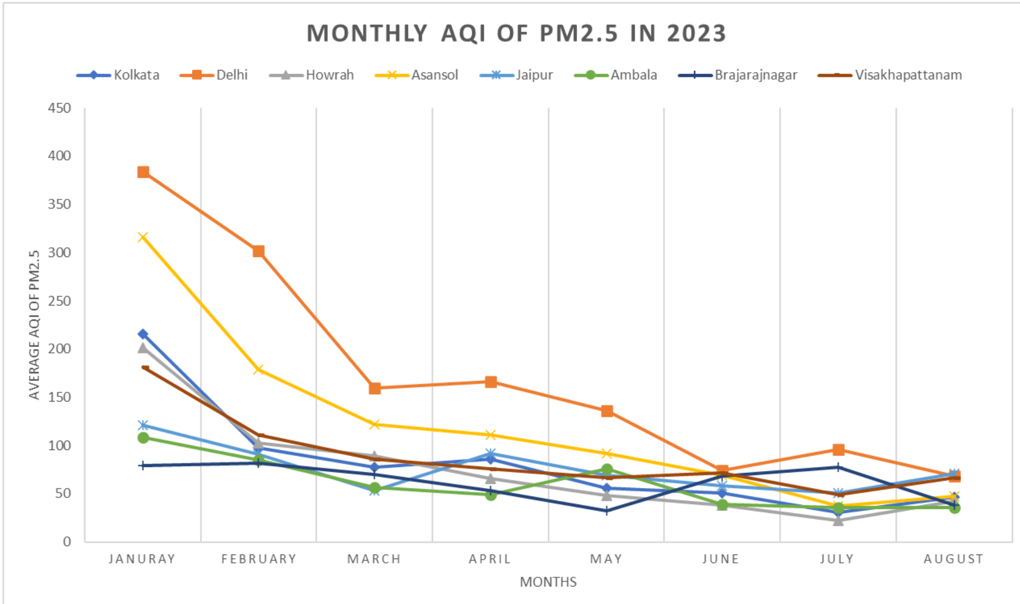
Figure 8:Visakhapattanam's PM2.5 in 2023

It remains more or less same for each city in compare to previous year.


MONTHLY AQI OF PM2.5 IN 2020
Kolkata
Delhi
Howrah
Asansol
Jaipur
Ambala
Brajarajnagar
Visakhapattanam
AVERAGE AQI OF PM2.5
0
100
200
300
400
500
600
JANURAY
FEBRUARY
MARCH
APRIL
MAY
JUNE
JULY
AUGUST
SEPTEMBER
OCTOBER
NOVEMBER
DECEMBER
MONTH


MONTHLY AQI OF PM2.5 IN 2021
Kolkata
Delhi
Howrah
Asansol
Jaipur
Ambala
Brajarajnagar
Visakhapattanam
AVERAGE AQI OF PM2.5
0
100
200
300
400
500
600
700
JANURAY
FEBRUARY
MARCH
APRIL
MAY
JUNE
JULY
AUGUST
SEPTEMBER
OCTOBER
NOVEMBER
DECEMBER
MONTHS


MONTHLY AQI OF PM2.5 IN 2022
Kolkata
Delhi
Howrah
Asansol
Jaipur
Ambala
Brajarajnagar
Visakhapattanam
AVERAGE AQI OF PM25
0
100
200
300
400
500
600
700
JANURAY
FEBRUARY
MARCH
APRIL
MAY
JUNE
JULY
AUGUST
SEPTEMBER
OCTOBER
NOVEMBER
DECEMBER
MONTHS



The above multiple line plots of monthly average PM2.5 emission give the idea that Delhi exhibits higher PM2.5 pollution in each year. The CPCB's annual average permissible limits for PM 2.5 is 40 ug/m3. But it is evident from the graphs that most of the cities are far above the safest limit.

2020	2021	2022	2023
PM 10	PM10	PM10	PM10
PM 2.5	NOX	PM2.5	PM2.5
NOx	PM2.5	NOx	NOx
NO2	NO2	NO2	NO2
NO	NH3	NH3	NH3
NH3	NO	NO	NO
SO2	SO2	SO2	SO2
CO	CO	CO	CO

Table 2: Kolkata

2020	2021	2022	2023
PM 10	PM10	PM10	PM10
PM 2.5	PM2.5	PM2.5	PM2.5
NO2=NH3	NO2	NO2	NO2=NH3
NOx	NOx	NOx=NH3	NOx
SO2	NH3=SO2	SO2	SO2
N0	NO	NO	NO
CO	CO	CO	CO

Table 3: Delhi

2020	2021	2022	2023
PM 10	PM10	PM10	PM10
PM2.5	PM2.5	PM2.5	PM2.5
NOx	NOx	NOx	NOx
NO2	NH3	NO2	NO2
NH3	NO2	NH3	NH3
N0	NO	SO2	SO2
SO2	SO2	NO	NO
CO	CO	CO	CO

Table 5: Asansol

2020	2021	2022	2023
PM 10	PM10	PM10	PM10
NOx	PM2.5=NOx	PM2.5	PM2.5=NOx
PM2.5=NO2	NO2	NOx	NO2
NH3=SO2	NH3	NO2	NO
NO	NO	NO	NH3
C0	SO2	NH3	SO2
	CO	SO2	CO
		CO	

Table 4: Howrah

2020	2021	2022	2023
PM 10	PM10	PM10	PM10
PM2.5=NOx=NH3	PM2.5	PM2.5=NOx	PM2.5=NOx=NH3
NO2	NOx=NH3	NO2=NH3	NO2
SO2	NO2	SO2	SO2
N0	SO2	NO	N0
C0	NO	CO	CO
	CO		

Table 6: Jaipur

2020	2021	2022	2023
PM 10	PM10	PM10	PM10
PM2.5	PM2.5	PM2.5	NH3
NH3	NH3	NH3	PM2.5
NO2	SO2	NO2	NOx
NOx	NOx	NOx	SO2
S02	NO2	SO2	NO2
NO	NO	NO	NO
CO	CO	CO	CO

Table 7: Ambala

2020	2021	2022	2023
PM 10	PM10	PM10	PM10
PM2.5	PM2.5	PM2.5	NOx
NOx=NH3	SO2	NO=NOx	PM2.5=NO
NO=NO2=SO2	NOx	SO2	NO2
CO	NO2=NH3	NO2	SO2
	NO	NH3	NH3
	CO	CO	CO

Table 8: Brajarajnagar

2020	2021	2022	2023
PM 10	PM10	PM10	PM10
PM2.5	PM2.5=NO2	PM2.5	PM2.5
NO2	NOx	NO2=NOx	NOx
NOx	NH3	NO	NO2
SO2	SO2	NH3	NO
NH3	NO	SO2	NH3
NO	CO	CO	SO2
CO			CO

Table 9: Visakhapattanam

There is a rule that says: 'Fit Model to Data not Data to Model'.The reason for saying that, first check the characteristics of the data then fit suitable model. Not fit any pre-determined model, that may lead to wrong result. In our case, the data does not satisfy the parametric test's assumptions. So we have turned to their non-parametric counterparts.

Let's choose a city. In that city there are 8 pollutants. Now construct a test named Kruskal-Wallis, and the H_0 : There is no significant difference between 8 pollutants Versus H_1: There are significant difference between at least two pollutants. After comparing critical value with test statistic we have reached the conclusion that "There is significant difference between at least two pollutants.

Now, let us remind our aim again, which is: find the pollutants which has most impact in each city. To accomplish that goal we need to consider another test named Mann-Whitney U test which is the alternative of parametric t test . In that case H_0 will be : There are no significant difference between two pollutants (for example PM2.5=PM10) versus H_1: There is significant difference between two pollutants (for example $PM2.5 \neq PM10$). There are 8 pollutants, for the test 2 pollutants are compared each time. So total 28 times the test is conducted. After obtaining the result there may arise two situation. First situation is H_0 is accepted, then we need not to do one sided test. But if second situation arises, that is H_1 is accepted, in that case we have to do one sided test, to obtain the result.

After conducting all the test for each city, and comparing the result, the result is shown in Table 2 to Table 9. It can be concluded that PM10 is affecting each city's air most and after that PM2.5. CO has lowest impact on each city's air pollution.

Now if we are interested to know which pollutant is affecting which cities most? It can be answered in the same way using same techniques. In that case, pollutant is fixed and comparing that pollutant's effect on each city.

Let's take a testing problem then it will be more clearer: H_0: There is no difference between kolkata's PM2.5 effect and Howrah's PM2.5 effect and the H_1: There is difference between kolkata's PM2.5 effect and Howrah's PM2.5 effect. Using Mann-Whitney U test and comparing p value with given level of significance we have reached the conclusion that There is no difference between Kolkata's PM2.5 and Howrah's PM2.5 effect. In that way we have identified the cities for each pollutants. And the result is given as followed:

PM2.5	PM10	NO2
Delhi	Delhi	Visakhapattnam
Asansol=Ambala	Asansol=Ambala	Jaipur
Jaipur	Visakhapattanam	Asansol
Kolkata=Howrah=Brajarajnagar=Visakhapattanam	Jaipur	Delhi
	Kolkata=Howrah	Kolkata=Howrah
	Brajarajnagar	Brajarajnagar
		Ambala

Table 10: PM2.5, PM10, NO2 effect on Cities

NOx	NO	NH3
Jaipur=Asansol	Brajarajnagar	Jaipur
Howrah	Visakhapattanam	Ambala
Kolkata=Visakhapattanam	Howrah=Asansol	Kolkata=Delhi=Asansol
Brajarajnagar	Kolkata	Visakhapattanam
Delhi	Delhi=Jaipur	Howrah=Brajarajnagar
Asansol	Ambala	

Table 11: NOx, NO, NH3 effect on Cities

SO2
Brajarajnagar
Visakhapattanam=Jaipur
Delhi
Howrah
Kolkata=Ambala
Asansol

Table 12: SO2 effect on Cities

The next objective is to check if the pollutants show any pattern during different seasons of a year. For that we have divided each year into Summer and Winter. Using Mann-Kendall test for trend the results are obtained.

City	2020	2021	2022	2023
Kolkata	Decreasing	Decreasing	Decreasing	Decreasing
Howrah	Decreasing	Decreasing	Decreasing	Decreasing
Asansol	Decreasing	Decreasing	Decreasing	Decreasing
Ambala	Increasing	No Trend	No Trend	No Trend
Jaipur	No Trend	Decreasing	No Trend	No Trend
Delhi	No Trend	Decreasing	Decreasing	Decreasing
Brajarajnagar	No Trend	Decreasing	Decreasing	Decreasing
Visakhapattanam	No Trend	Decreasing	No Trend	No Trend

Table 13: Summer PM2.5

City	2020	2021	2022	2023
Kolkata	Decreasing	Decreasing	Decreasing	Decreasing
Howrah	Decreasing	Decreasing	Decreasing	Decreasing
Asansol	Decreasing	Decreasing	Decreasing	Decreasing
Ambala	Increasing	No Trend	No Trend	Decreasing
Jaipur	No Trend	Decreasing	Increasing	Increasing
Delhi	No Trend	Decreasing	No Trend	Decreasing
Brajarajnagar	No Trend	Decreasing	Decreasing	Decreasing
Visakhapattanam	No Trend	Decreasing	Decreasing	No Trend

Table 14: Summer PM10

City	2020	2021	2022	2023
Kolkata	Decreasing	Decreasing	No Trend	Decreasing
Howrah	Decreasing	Decreasing	Decreasing	Decreasing
Asansol	Decreasing	Decreasing	Increasing	Decreasing
Ambala	Increasing	No Trend	Decreasing	No Trend
Jaipur	No Trend	Decreasing	Decreasing	Decreasing
Delhi	Decreasing	No Trend	Decreasing	No Trend
Brajarajnagar	No Trend	No Trend	Increasing	Decreasing
Visakhapattanam	No Trend	Decreasing	No Trend	Increasing

Table 15: Summer NO2

City	2020	2021	2022	2023
Kolkata	Decreasing	No Trend	Decreasing	Decreasing
Howrah	Decreasing	Decreasing	Decreasing	Decreasing
Asansol	Decreasing	Decreasing	No Trend	Decreasing
Ambala	Decreasing	Decreasing	Decreasing	Increasing
Jaipur	Increasing	No Trend	Decreasing	No Trend
Delhi	Decreasing	No Trend	Increasing	Increasing
Brajarajnagar	Decreasing	Decreasing	No Trend	No Trend
Visakhapattanam	No Trend	Decreasing	Increasing	Increasing

Table 16: Summer NH3

City	2020	2021	2022	2023
Kolkata	Decreasing	Decreasing	No Trend	No Trend
Howrah	No Trend	Decreasing	Decreasing	Decreasing
Asansol	Decreasing	Decreasing	Decreasing	Decreasing
Ambala	Decreasing	No Trend	No Trend	Increasing
Jaipur	Increasing	Decreasing	Decreasing	Increasing
Delhi	No Trend	No Trend	Decreasing	No Trend
Brajarajnagar	No Trend	Decreasing	Increasing	Decreasing
Visakhapattanam	No Trend	Decreasing	No Trend	Decreasing

Table 17: Summer NO

City	2020	2021	2022	2023
Kolkata	Decreasing	Decreasing	No Trend	Decreasing
Howrah	Decreasing	Decreasing	Decreasing	Decreasing
Asansol	Decreasing	Decreasing	Increasing	Decreasing
Ambala	No Trend	No Trend	Decreasing	Increasing
Jaipur	Increasing	Decreasing	Decreasing	No Trend
Delhi	Decreasing	Decreasing	Decreasing	No Trend
Brajarajnagar	No Trend	Decreasing	Increasing	Decreasing
Visakhapattanam	No Trend	Decreasing	No Trend	Decreasing

Table 18: Summer NOx

City	2020	2021	2022	2023
Kolkata	Decreasing	Decreasing	No Trend	Decreasing
Howrah	Decreasing	Decreasing	Decreasing	No Trend
Asansol	Decreasing	Decreasing	Decreasing	Decreasing
Ambala	Decreasing	No Trend	No Trend	No Trend
Jaipur	Decreasing	Decreasing	Decreasing	Decreasing
Delhi	Decreasing	Decreasing	Decreasing	Decreasing
Brajarajnagar	Increasing	No Trend	Decreasing	Increasing
Visakhapattanam	No Trend	No Trend	Decreasing	Decreasing

Table 19: Summer SO2

City	2020	2021	2022	2023
Kolkata	Decreasing	Decreasing	Decreasing	Decreasing
Howrah	Decreasing	Increasing	Increasing	Decreasing
Asansol	Decreasing	Decreasing	Decreasing	Decreasing
Ambala	No Trend	Decreasing	Decreasing	Increasing
Jaipur	No Trend	Decreasing	No Trend	No Trend
Delhi	Decreasing	No Trend	Decreasing	No Trend
Brajarajnagar	Increasing	Decreasing	No Trend	Decreasing
Visakhapattanam	No Trend	No Trend	Decreasing	Decreasing

Table 20: Summer CO

City	2020	2021	2022
Kolkata	Increasing	Increasing	Increasing
Howrah	Increasing	Increasing	Increasing
Delhi	Increasing	Increasing	Increasing
Ambala	Increasing	Increasing	No Trend
Jaipur	Increasing	Increasing	No Trend
Brajarajnagar	Increasing	No Trend	No Trend
Visakhapattanam	Increasing	Increasing	Increasing
Asansol	No Trend	Increasing	Increasing

Table 21: Winter PM2.5

City	2020	2021	2022
Kolkata	Increasing	Increasing	Increasing
Howrah	Increasing	Increasing	Increasing
Delhi	No Trend	Increasing	Increasing
Ambala	Decreasing	Increasing	No Trend
Jaipur	No Trend	Increasing	No Trend
Brajarajnagar	Increasing	Increasing	No Trend
Visakhapattanam	Increasing	No Trend	Increasing
Asansol	Increasing	Increasing	Increasing

Table 22: Winter PM10

City	2020	2021	2022
Kolkata	Increasing	Increasing	No Trend
Howrah	No Trend	Increasing	Increasing
Delhi	Decreasing	Increasing	Increasing
Ambala	Decreasing	Increasing	Increasing
Jaipur	No Trend	No Trend	Increasing
Brajarajnagar	No Trend	Increasing	Decreasing
Visakhapattanam	Decreasing	No Trend	No Trend
Asansol	Increasing	Increasing	Increasing

Table 23: Winter NO2

City	2020	2021	2022
Kolkata	Increasing	Increasing	No Trend
Howrah	Increasing	Decreasing	Increasing
Delhi	No Trend	No Trend	Increasing
Ambala	Increasing	Increasing	Decreasing
Jaipur	No Trend	No Trend	Increasing
Brajarajnagar	Increasing	Increasing	No Trend
Visakhapattanam	No Trend	Decreasing	Increasing
Asansol	Increasing	Increasing	Increasing

Table 24: Winter NH3

City	2020	2021	2022
Kolkata	Increasing	Increasing	Increasing
Howrah	No Trend	No Trend	Increasing
Delhi	No Trend	Increasing	No Trend
Ambala	No Trend	Increasing	Decreasing
Jaipur	Increasing	No Trend	Increasing
Brajarajnagar	Increasing	Increasing	Increasing
Visakhapattanam	Increasing	No Trend	Decreasing
Asansol	Increasing	Increasing	Increasing

Table 25: Winter NO

City	2020	2021	2022
Kolkata	Increasing	Increasing	No Trend
Howrah	No Trend	Increasing	Increasing
Delhi	No Trend	Increasing	Increasing
Ambala	No Trend	Increasing	No Trend
Jaipur	No Trend	No Trend	Increasing
Brajarajnagar	Increasing	Increasing	No Trend
Visakhapattanam	Increasing	No Trend	No Trend
Asansol	Increasing	Increasing	Increasing

Table 26: Winter NOx

City	2020	2021	2022
Kolkata	Increasing	Increasing	No Trend
Howrah	Increasing	Increasing	Increasing
Delhi	Decreasing	Increasing	No Trend
Ambala	No Trend	Increasing	Increasing
Jaipur	Increasing	No Trend	No Trend
Brajarajnagar	Increasing	Decreasing	No Trend
Visakhapattanam	Increasing	Increasing	No Trend
Asansol	Increasing	Increasing	Increasing

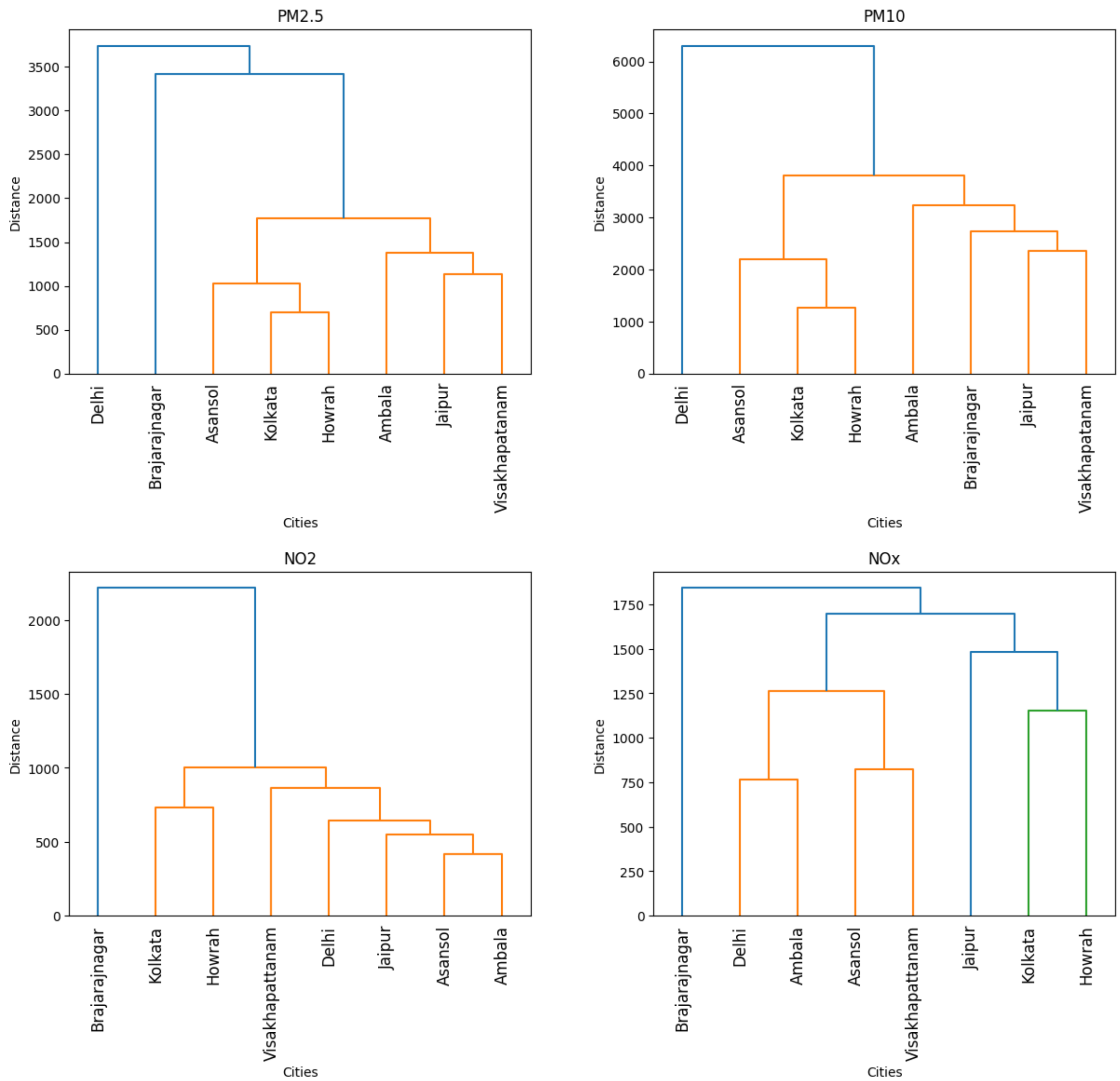
Table 27: Winter SO2

City	2020	2021	2022
Kolkata	Increasing	Increasing	Increasing
Howrah	Increasing	Increasing	Increasing
Delhi	Decreasing	Increasing	Increasing
Ambala	Increasing	Increasing	Increasing
Jaipur	Increasing	No Trend	No Trend
Brajarajnagar	Increasing	Increasing	No Trend
Visakhapattanam	Increasing	No Trend	Increasing
Asansol	Increasing	Increasing	Increasing

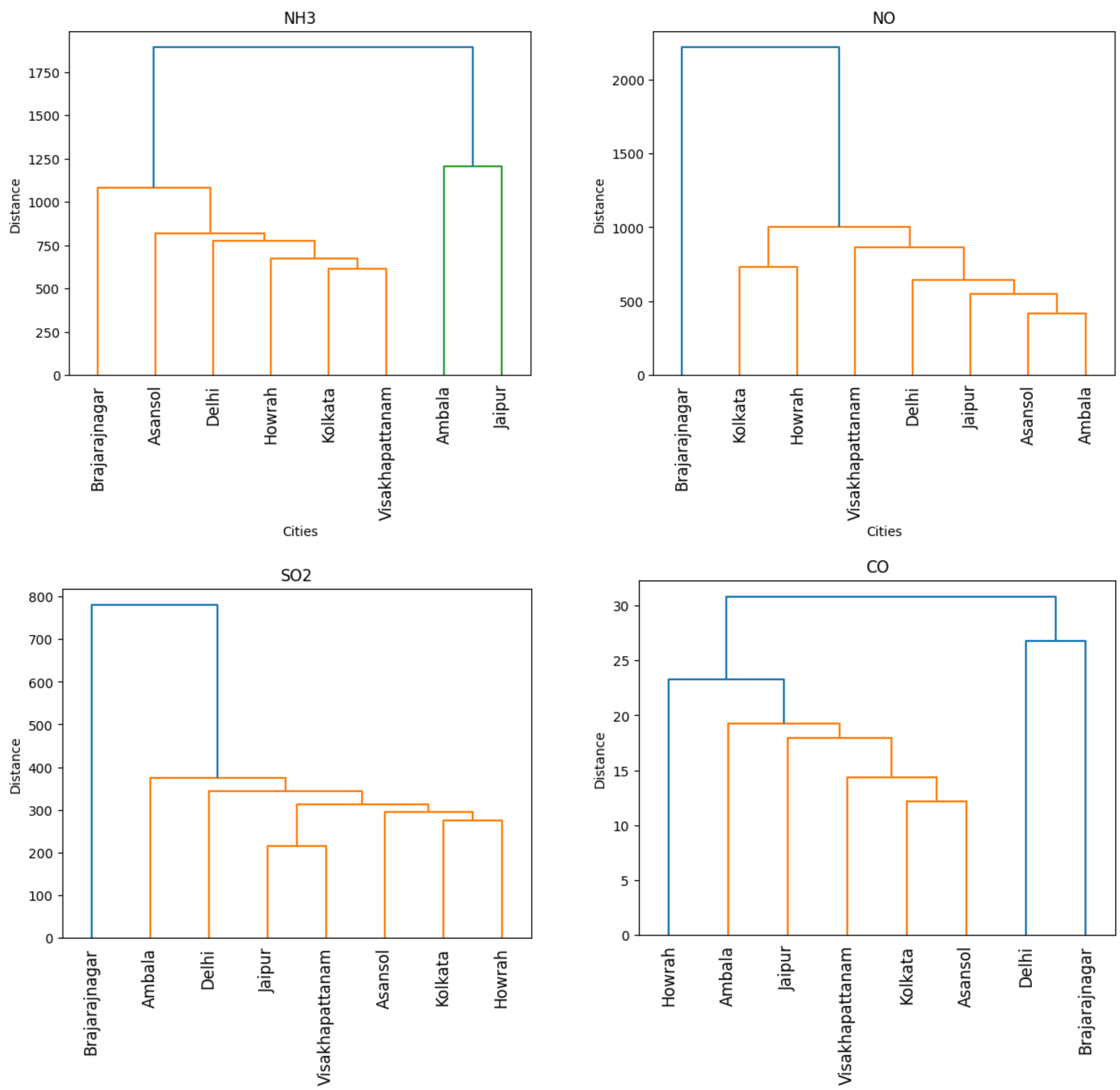
Table 28: Winter CO

Let's discuss our observation from the above tables [Table 13 : Table 28]. The seasonal variation in pollutant emissions (PM2.5, PM10, NO, NH3, NOx, NO2, SO2, CO) in eight Indian cities (Kolkata, Delhi, Howrah, Asansol, Visakhapattanam, Brajarajnagar, Jaipur, Ambala) from January 1, 2020, to August 21, 2023.In summer time, in almost every Indian cities the pollution level tends to decreases and in winter the pollution shows increasing trend.

The above analysis that are already been done gives us the sense that which pollutants are dangerous ,which cities are mostly affected by it and which time the pollutants show any significant pattern. But now if we are interested to know which cities have similar pollution pattern, and we want to group them based on it. Then we need to use unsupervised machine learning techniques, in our case hierarchical clustering.



PM2.5
Distance
0
500
1000
1500
2000
2500
3000
3500
Delhi
Brajarajnagar
Asansol
Kolkata
Howrah
Ambala
Jaipur
Visakhapatanam
Cities
PM10
Distance
0
1000
2000
3000
4000
5000
6000
Delhi
Asansol
Kolkata
Howrah
Ambala
Brajarajnagar
Jaipur
Visakhapatanam
Cities
NO2
Distance
0
500
1000
1500
2000
Brajarajnagar
Kolkata
Howrah
Visakhapattanam
Delhi
Jaipur
Asansol
Ambala
Cities
NOx
Distance
0
250
500
750
1000
1250
1500
1750
Brajarajnagar
Delhi
Ambala
Asansol
Visakhapattanam
Jaipur
Kolkata
Howrah
Cities


NH3
Distance
1750
1500
1250
1000
750
500
250
0
Brajarajnagar
Asansol
Delhi
Howrah
Kolkata
Visakhapattanam
Ambala
Jaipur
Cities
NO
Distance
2000
1500
1000
500
0
Brajarajnagar
Kolkata
Howrah
Visakhapattanam
Delhi
Jaipur
Asansol
Ambala
Cities
SO2
Distance
800
700
600
500
400
300
200
100
0
Brajarajnagar
Ambala
Delhi
Jaipur
Visakhapattanam
Asansol
Kolkata
Howrah
Cities
CO
Distance
30
25
20
15
10
5
0
Howrah
Ambala
Jaipur
Visakhapattanam
Kolkata
Asansol
Delhi
Brajarajnagar
Cities

	PM2.5
Cluster 1	Delhi
Cluster 2	Brajarajnagar
Cluster 3	Asansol, Kolkata, Howrah, Ambala, Jaipur, Visakhapattanam

(a) Cluster formed by cities based on PM2.5

	PM10
Cluster 1	Delhi
Cluster 2	Asansol, Kolkata, Howrah, Ambala, Jaipur, Visakhapattanam, Brajarajnagar

(b) Cluster formed by cities based on PM10

	NO2
Cluster 1	Brajarajnagar
Cluster 2	Asansol, Kolkata, Howrah, Ambala, Jaipur, Visakhapattanam, Delhi

(c) Cluster formed by cities based on NO2

	NOx
Cluster 1	Brajarajnagar
Cluster 2	Asansol, Kolkata, Howrah, Ambala, Jaipur, Visakhapattanam, Delhi

(d) Cluster formed by cities based on NOx

	NH3
Cluster 1	Ambala, Jaipur
Cluster 2	Asansol, Kolkata, Howrah, Brajarajnagar, Visakhapattanam, Delhi

(e) Cluster formed by cities based on NH3

	NO
Cluster 1	Brajarajnagar
Cluster 2	Asansol, Kolkata, Howrah, Ambala, Jaipur, Visakhapattanam, Delhi

(f) Cluster formed by cities based on NO

	SO2
Cluster 1	Brajarajnagar
Cluster 2	Asansol, Kolkata, Howrah, Ambala, Jaipur, Visakhapattanam, Delhi

(g) Cluster formed by cities based on SO2

	CO
Cluster 1	Delhi,Brajarajnagar
Cluster 2	Asansol, Kolkata, Howrah, Ambala, Jaipur, Visakhapattanam

(h) Cluster formed by cities based on CO

In case of each pollutants, some clusters are formed, In the hierarchical clustering analysis, cities that are grouped together are considered to be more similar to each other in terms of their pollution patterns compared to the cities in other clusters.

If we compare the dendrogram of PM2.5 and PM10, it can be seen that Delhi stands out only as a cluster in both cases. Also in our statistical analysis it was found that there is the significant difference between Delhi's PM2.5 and others cities PM2.5 emission, same goes for PM10.

7 Conclusion

PM2.5 and PM10 have the highest effects in all cities is significant and suggests common factors contributing to the increased levels of particulate matter during certain seasons. There are several possible reasons for this phenomenon: Particulate matter emissions, especially PM2.5 and PM10, are often associated with combustion processes. In winter, the atmospheric conditions tend to be more stable, leading to lower dispersion of pollutants. This results in the accumulation of particulate matter in the air.During winter, the increased use of heating systems and burning of solid fuels, such as wood and coal, can significantly contribute to higher PM2.5 and PM10 emissions. This is particularly relevant in regions where heating is a common practice in cold weather. The demand for heating and increased vehicular emissions during cold weather leads to higher emissions of PM from sources like burning of biomass and fossil fuels. The incomplete combustion of fossil fuels in vehicles and heating systems releases higher levels of NOx and CO in the atmosphere.

In contrast, during the summer months, there is a decrease in PM2.5 and PM10 emissions due to increased wind speed and atmospheric mixing, which facilitates the dispersion of particulate matter and reduces their concentration. Reduced heating requirements and better fuel combustion efficiency contribute to lower emissions of particulate matter.

PM2.5 and PM10 are most dangerous among all pollutants. Delhi is highly affected by the emission of PM2.5 and PM10. Delhi's geographical location plays a significant role. It is situated in a landlocked region, which means that pollutants can become trapped within the city due to the absence of natural ventilation from the sea. This geographical confinement exacerbates pollution levels, especially during periods of stagnant air. Delhi is one of the most densely populated cities in the world. The large population contributes to higher emissions from vehicles, industries, and residential heating, all of which increase the concentration of pollutants in the air.

NO2, NOx, and NH3 have the highest impact on cities like Jaipur, Visakhapatnam, and Ambala during the period from January 1, 2020, to August 21, 2023, there are several possible reasons for this phenomenon. Jaipur is located in the state of Rajasthan, known for its significant agricultural activities. The use of ammonia-based fertilizers in agriculture can be a major source of NH3 emissions in the region. Visakhapatnam is an industrial hub with several heavy industries and manufacturing units. Industrial processes can be a primary source of NO2 and NOx emissions. The city has a significant port and maritime activities, including shipping and shipping-related industries. These activities can contribute to air pollution, especially NOx emissions from ship engines. Ambala, located in Haryana, is an agriculturally rich area. NH3 emissions can be elevated due to the use of ammonia-based fertilizers and agricultural practices in the region.

Now if we focus on the result of hierarchical clustering outcomes, we can see here Delhi exhibits different pollution pattern in case of PM2.5 and PM10. The statistical test, we have conducted gives us the information about the location parameters such as average. Dendrogram grouped cities based on the similarities in their pollution pattern. Kolkata and Howrah are grouped together in almost every dendrogram, because these two cities are situated nearly and they share almost same pollution patterns. In case of Brajarajnagar, Coal mining equipment, such as diesel-powered vehicles and machinery, can emit nitrogen oxides (NOx), including nitrogen dioxide (NO2), as a byproduct of combustion. Explosives used in mining operations can also release NOx and NO2 into the atmosphere. Ammonia (NH3) emissions can result from the use of ammonium nitrate-based explosives in mining, which can lead to the release of ammonia gas. Sulfur dioxide (SO2) emissions can occur during coal mining when the coal contains sulfur, which is common in many coal deposits. When sulfur-containing coal is exposed to air and water, it can undergo chemical reactions that produce SO2.

So in summary:

1. Delhi is the most polluted cities in terms of PM2.5 and PM10 emission
2. Brajarajnagar has many coal fields that's why it shows different pollution pattern in terms of the emission of NOx, NO2, NH3, SO2
3. In winter the pollution level is increased and in summer it tends to decrease.

References

[1] KC Gouda, Priya Singh, Nikhilasuma P, Mahendra Benke, Reshama Kumari, Geeta Agnihotri, Kiran M Hungund, and Ramesh V. Assessment of air pollution status during covid-19 lockdown (march–may 2020) over bangalore city in india. *Environmental Monitoring and Assessment*, 193(7):395, 2021.

[2] Sunil Gulia, SM Shiva Nagendra, Jo Barnes, and Mukesh Khare. Urban local air quality management framework for non-attainment areas in indian cities. *Science of the Total Environment*, 619:1308–1318, 2018.

[3] Sonal Kumari, Anita Lakhani, K Maharaj Kumari, et al. Covid-19 and air pollution in indian cities: World's most polluted cities. *Aerosol and Air Quality Research*, 20(12):2592–2603, 2020.

[4] Prakhar Misra and Wataru Takeuchi. Analysis of air quality in indian cities using remote sensing and economic growth parameters. In *Proceedings of the ACRS*. Citeseer, 2015.

[5] Pallav Purohit, Markus Amann, Gregor Kiesewetter, Peter Rafaj, Vaibhav Chaturvedi, Hem H Dholakia, Poonam Nagar Koti, Zbigniew Klimont, Jens Borken-Kleefeld, Adriana Gomez-Sanabria, et al. Mitigation pathways towards national ambient air quality standards in india. *Environment international*, 133:105147, 2019.

[6] Saidur Rahaman, Selim Jahangir, Ruishan Chen, Pankaj Kumar, and Swati Thakur. Covid-19's lockdown effect on air quality in indian cities using air quality zonal modeling. *Urban climate*, 36:100802, 2021.

[7] Rohit Sharma, Raghvendra Kumar, Devendra Kumar Sharma, Le Hoang Son, Ishaani Priyadarshini, Binh Thai Pham, Dieu Tien Bui, and Sakshi Rai. Inferring air pollution from air quality index by different geographical areas: case study in india. *Air Quality, Atmosphere & Health*, 12:1347–1357, 2019.

[8] AK Singh, HK Gupta, K Gupta, P Singh, VB Gupta, and RC Sharma. A comparative study of air pollution in indian cities. *Bulletin of environmental contamination and toxicology*, 78(5):411–416, 2007.

[9] Krishna Prasad Vadrevu, Aditya Eaturu, Sumalika Biswas, Kristofer Lasko, Saroj Sahu, JK Garg, and Chris Justice. Spatial and temporal variations of air pollution over 41 cities of india during the covid-19 lockdown period. *Scientific Reports*, 10(1):16574, 2020.