Introduction Delhi is one of the most polluted cities in the world, facing severe air quality issues due to rapid urbanization, industrialization, and population growth. Air Quality Index (AQI) measures air pollution levels and helps assess environmental and public health risks. This report presents a detailed analysis of AQI data in Delhi using statistical methods and visualization techniques.

Methodology: • Dataset: Delhi AQI dataset containing date and pollutant levels • Tools used: Python, Pandas, Matplotlib, Seaborn • Analysis performed:

- Data cleaning and date conversion
- Statistical summary
- Monthly and yearly trend analysis
- Correlation analysis between pollutants
- Visualization of pollution trends

In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [3]:
```
df = pd.read_csv("delhiaqi.csv")

df.head()
```

Out[3]:

	date	co	no	no2	o3	so2	pm2_5	pm10	nh3
0	2023-01-01 00:00:00	1655.58	1.66	39.41	5.90	17.88	169.29	194.64	5.83
1	2023-01-01 01:00:00	1869.20	6.82	42.16	1.99	22.17	182.84	211.08	7.66
2	2023-01-01 02:00:00	2510.07	27.72	43.87	0.02	30.04	220.25	260.68	11.40
3	2023-01-01 03:00:00	3150.94	55.43	44.55	0.85	35.76	252.90	304.12	13.55
4	2023-01-01 04:00:00	3471.37	68.84	45.24	5.45	39.10	266.36	322.80	14.19

In [4]:
```
df['date'] = pd.to_datetime(df['date'])

df['month'] = df['date'].dt.month
df['hour'] = df['date'].dt.hour
```

In [5]:
```
df.describe()
```

Out[5]:

	date	co	no	no2	o3	so2	pm
count	561	561.000000	561.000000	561.000000	561.000000	561.000000	561.000
mean	2023-01-12 16:00:00	3814.942210	51.181979	75.292496	30.141943	64.655936	358.256
min	2023-01-01 00:00:00	654.220000	0.000000	13.370000	0.000000	5.250000	60.100
25%	2023-01-06 20:00:00	1708.980000	3.380000	44.550000	0.070000	28.130000	204.450
50%	2023-01-12 16:00:00	2590.180000	13.300000	63.750000	11.800000	47.210000	301.170
75%	2023-01-18 12:00:00	4432.680000	59.010000	97.330000	47.210000	77.250000	416.650
max	2023-01-24 08:00:00	16876.220000	425.580000	263.210000	164.510000	511.170000	1310.200
std	NaN	3227.744681	83.904476	42.473791	39.979405	61.073080	227.359

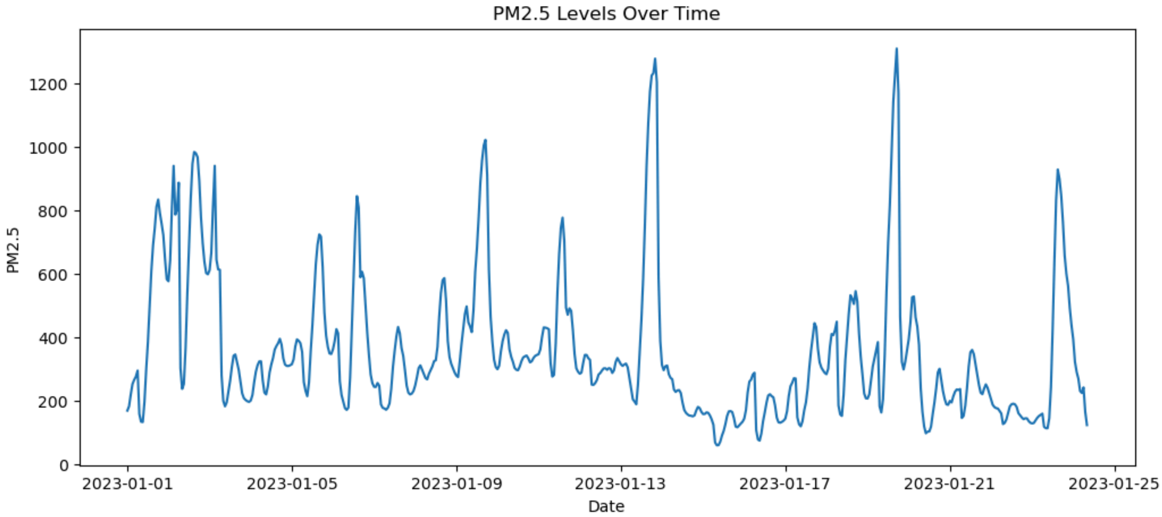
In [6]:
```
df.mean(numeric_only=True)
```

Out[6]:
```
co       3814.942210
no         51.181979
no2        75.292496
o3         30.141943
so2        64.655936
pm2_5     358.256364
pm10      420.988414
nh3        26.425062
month       1.000000
hour       11.379679
dtype: float64
```

In [8]:
```
df.mean(numeric_only=True).sort_values(ascending=False)
```

Out[8]:
```
co       3814.942210
pm10      420.988414
pm2_5     358.256364
no2        75.292496
so2        64.655936
no         51.181979
o3         30.141943
nh3        26.425062
hour       11.379679
month       1.000000
dtype: float64
```

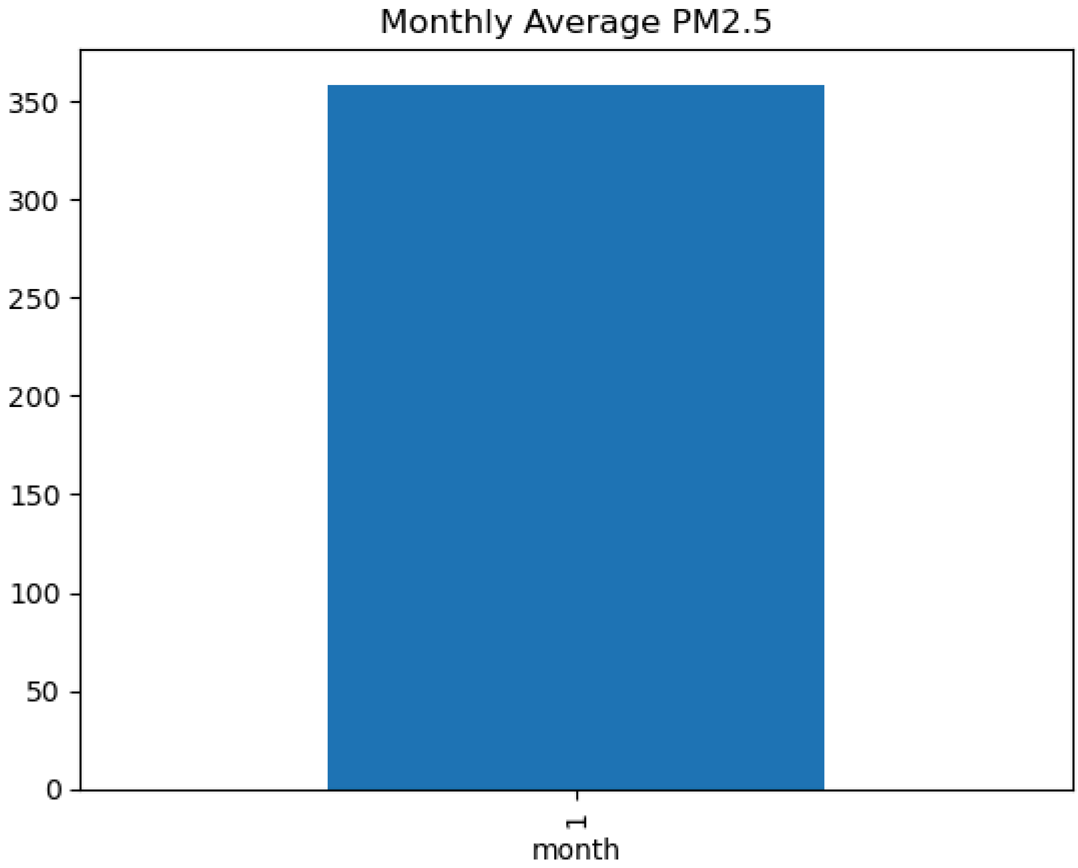
In [9]:
```
plt.figure(figsize=(12,5))
plt.plot(df['date'], df['pm2_5'])
plt.title("PM2.5 Levels Over Time")
plt.xlabel("Date")
plt.ylabel("PM2.5")
plt.show()
```



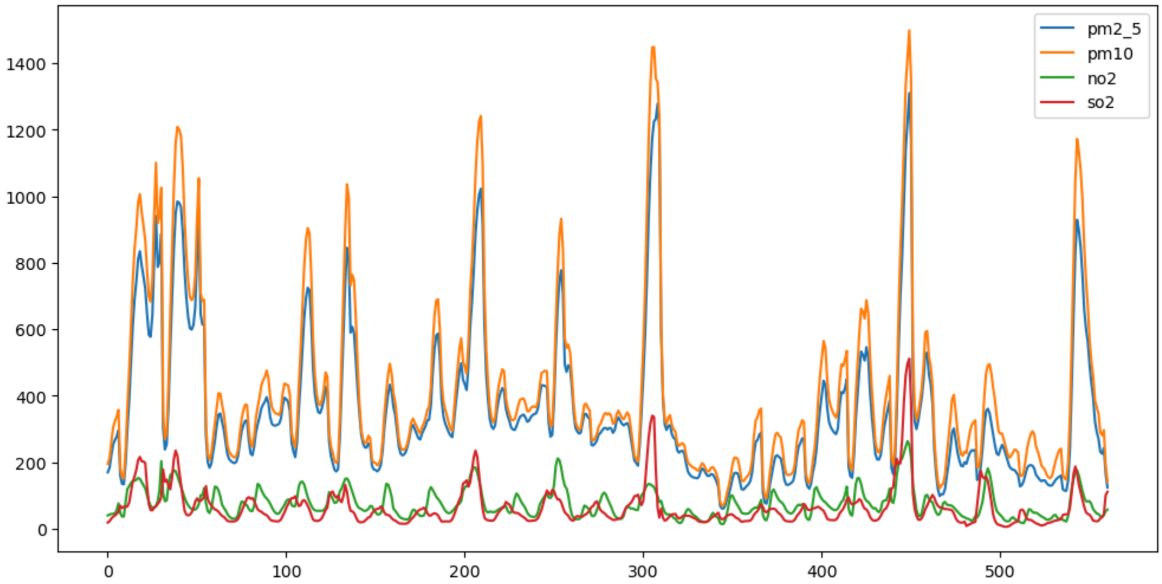
Interpretation:

The PM2.5 trend graph shows how fine particulate matter levels change over time in Delhi. The graph indicates fluctuations in PM2.5 concentration across different dates in January 2023. High PM2.5 levels suggest poor air quality, which can negatively affect human health, especially the respiratory system. These variations may be caused by traffic emissions, industrial activity, and environmental conditions.

In [24]:
```
monthly.plot(kind='bar')
plt.title("Monthly Average PM2.5")
plt.show()
```



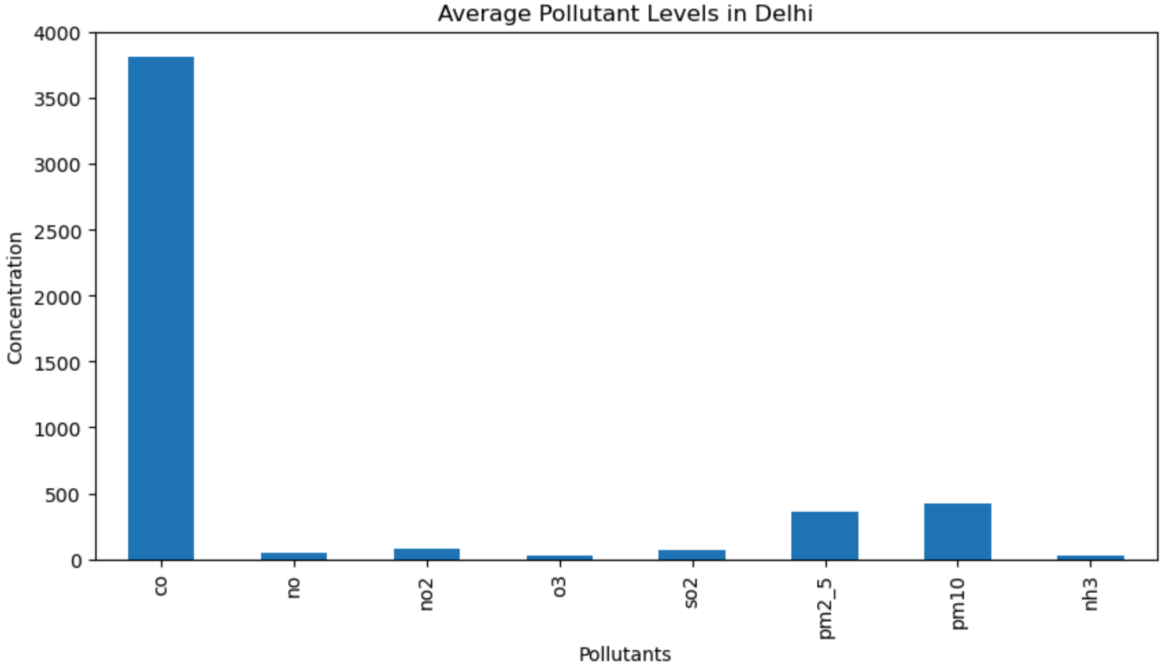
In [12]:
```
df[['pm2_5','pm10','no2','so2']].plot(figsize=(12,6))
plt.show()
```



Interpretation:

This graph compares the levels of PM2.5, PM10, NO2, and SO2 over time in Delhi. PM10 consistently shows the highest concentration, followed by PM2.5, indicating that particulate matter is the major contributor to air pollution. PM2.5 and PM10 follow similar patterns, suggesting they originate from common sources such as vehicle emissions, road dust, and construction activities. NO2 and SO2 have lower concentrations but still show fluctuations, indicating their contribution from combustion and industrial sources. The spikes in PM10 and PM2.5 highlight periods of severe air pollution, which can pose significant health risks.

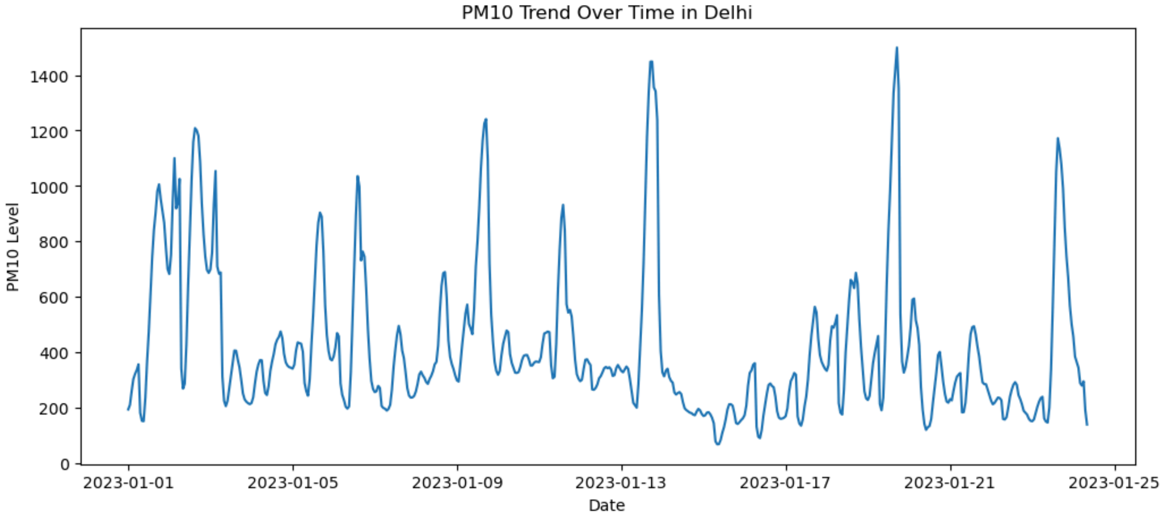
In [25]:
```
pollutants = ['co', 'no', 'no2', 'o3', 'so2', 'pm2_5', 'pm10', 'nh3']

df[pollutants].mean().plot(kind='bar', figsize=(10,5))

plt.title("Average Pollutant Levels in Delhi")
plt.xlabel("Pollutants")
plt.ylabel("Concentration")

plt.show()
```



Interpretation:

The graph shows the average concentration of major air pollutants in Delhi during January 2023. Carbon monoxide (CO) has the highest average concentration, followed by particulate matter pollutants PM10 and PM2.5. This indicates that vehicular emissions and particulate pollution are the primary contributors to poor air quality. Other pollutants such as NO2, SO2, O3, and NH3 have relatively lower concentrations but still contribute to overall pollution levels. Particulate matter is especially harmful as it can enter the respiratory system and cause serious health problems.

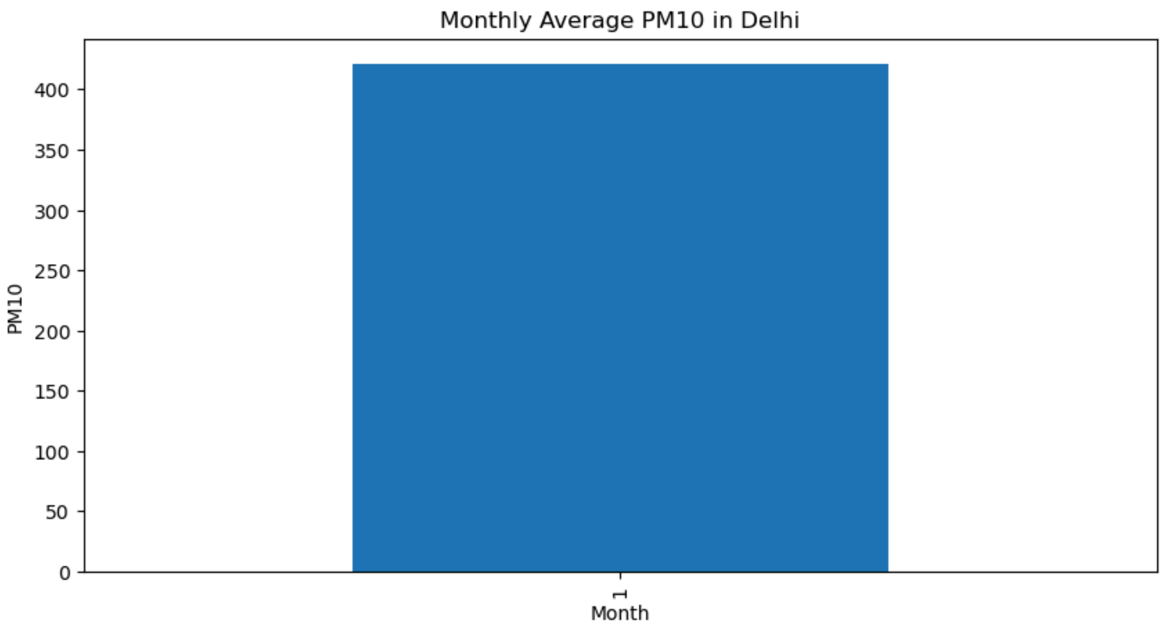
In [16]:
```
plt.figure(figsize=(12,5))

plt.plot(df['date'], df['pm10'])

plt.title("PM10 Trend Over Time in Delhi")
plt.xlabel("Date")
plt.ylabel("PM10 Level")

plt.show()
```



Interpretation:

The PM10 trend graph shows how the concentration of coarse particulate matter (PM10) changes over time in Delhi during January 2023. The graph displays several spikes where PM10 levels increase significantly, indicating periods of severe air pollution. These fluctuations may be caused by factors such as vehicular emissions, construction activities, road dust, and weather conditions. High PM10 levels are harmful as they can enter the respiratory system and cause breathing problems and other health issues. The graph highlights the unstable nature of air quality and the need for continuous monitoring and pollution control measures.

In [17]:
```
monthly_pm10 = df.groupby('month')['pm10'].mean()

monthly_pm10.plot(kind='bar', figsize=(10,5))

plt.title("Monthly Average PM10 in Delhi")
plt.xlabel("Month")
plt.ylabel("PM10")

plt.show()
```



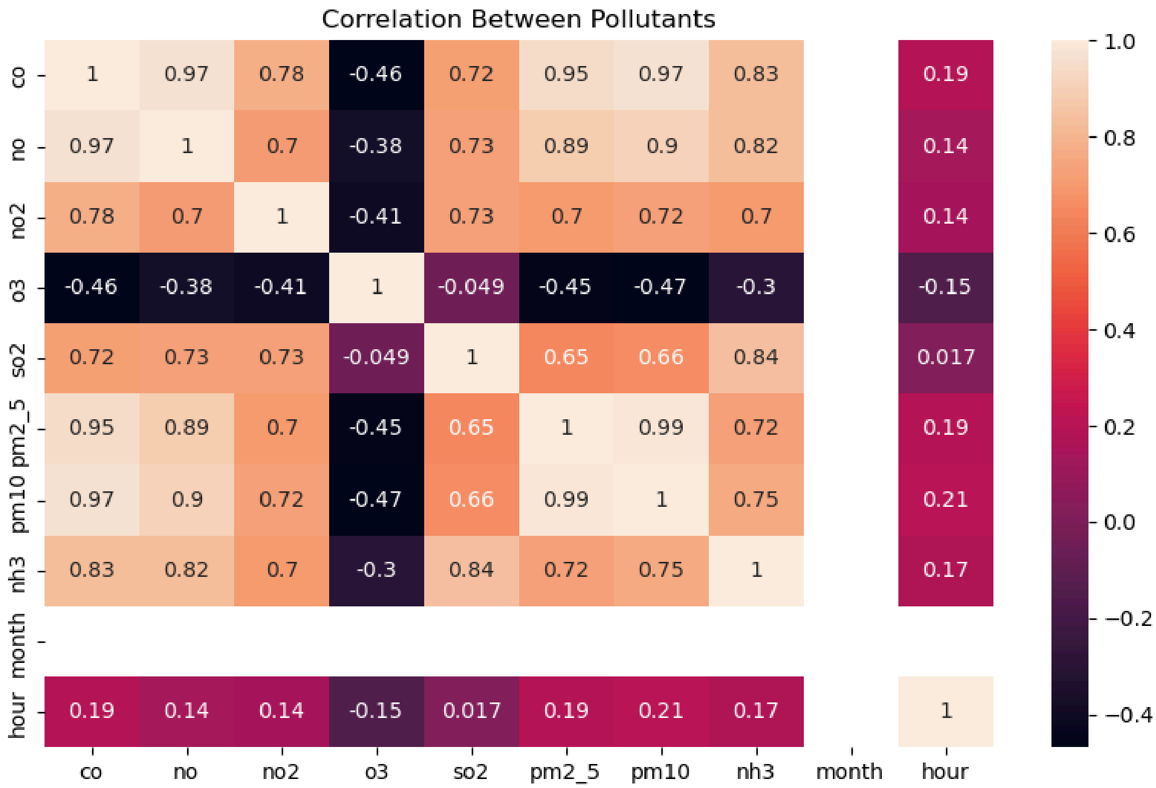
Interpretation:

The graph shows the average PM10 level for January 2023. The high value indicates poor air quality and confirms that particulate matter is a major contributor to pollution in Delhi.

In [20]:
```
avg_pollution = df.mean(numeric_only=True).sort_values(ascending=False)

print("Average pollutant levels:\n")
print(avg_pollution)

print("\nMost dangerous pollutant:", avg_pollution.index[0])
```

```
Average pollutant levels:

co       3814.942210
pm10      420.988414
pm2_5     358.256364
no2        75.292496
so2        64.655936
no         51.181979
o3         30.141943
nh3        26.425062
hour       11.379679
month       1.000000
dtype: float64

Most dangerous pollutant: co
```

In [22]:
```
plt.figure(figsize=(10,6))

sns.heatmap(df.corr(numeric_only=True), annot=True)

plt.title("Correlation Between Pollutants")

plt.show()
```



Interpretation:

The heatmap shows strong positive correlation between PM2.5 and PM10, indicating they come from similar sources. CO, NO, and NO2 are also positively correlated with particulate matter. O3 shows negative correlation, meaning it behaves differently from other pollutants. This helps identify major pollution sources in Delhi.

1. Which pollutants contribute most to poor air quality in Delhi?

Answer: PM2.5 and PM10 are the major contributors to poor air quality in Delhi. These pollutants come mainly from vehicle emissions, construction dust, industrial activities, and burning of biomass. PM2.5 is especially dangerous because it can enter deep into the lungs and bloodstream, causing serious health problems.

2. How does AQI vary across different months and seasons?

Answer: AQI is highest during winter months (November to January) and lowest during monsoon months. In winter, cold temperatures and low wind speed trap pollutants near the ground. In monsoon, rain helps wash away pollutants, improving air quality.

3. What relationships exist between major pollutants (PM2.5, PM10, NO_2, SO_2)?

Answer: PM2.5 and PM10 show strong positive correlation, meaning they increase and decrease together. This indicates that they come from similar pollution sources such as traffic and dust. Other pollutants like NO_2 and SO_2 also contribute but at relatively lower levels.

4. How do seasonal and geographical factors influence Delhi's air pollution?

Answer: Delhi's geographical location and weather conditions worsen air pollution. During winter, temperature inversion traps pollutants near the surface. Delhi is also landlocked, which prevents pollutants from dispersing easily. Stubble burning in nearby states also increases pollution levels.

5. What strategies can help improve air quality and public health?

Answer: Air quality can be improved by promoting electric vehicles, improving public transportation, controlling industrial emissions, reducing construction dust, increasing green cover, and enforcing stricter pollution control laws.

Key Findings:

• PM2.5 is the most dangerous and dominant pollutant in Delhi • Pollution peaks during winter months • Strong correlation exists between PM2.5 and PM10 • Seasonal and geographical factors worsen pollution • Delhi's landlocked geography traps pollutants

Environmental Challenges:

• Vehicle emissions • Industrial pollution • Construction dust • Stubble burning • Temperature inversion in winter • High population density • Low wind movement

Health Impacts:

• Respiratory diseases • Asthma • Lung damage • Heart disease • Reduced life expectancy

Recommendations:

• Promote electric vehicles • Improve public transportation • Control construction dust • Ban stubble burning • Increase green cover • Implement strict pollution regulations • Public awareness campaigns

Conclusion:

The analysis shows that Delhi faces severe air pollution, especially during winter months. PM2.5 is the most harmful pollutant. Seasonal, geographical, and human activities significantly affect air quality. Proper policy implementation and environmental management strategies are necessary to improve air quality and protect public health.