Analysing the Air Quality Index (AQI) in Delhi: Pollutants, Seasonal Variations, and Geographical Impacts

Introduction:

Delhi, India's capital, consistently ranks among the most polluted cities globally. Rapid urbanization, high vehicular emissions, industrial activity, and geographical peculiarities exacerbate air quality issues, impacting millions. This study investigates the AQI trends in Delhi, identifies key pollutants, and analyzes seasonal and geographical factors affecting pollution levels to propose effective interventions.

Objectives:

- To analyze AQI trends in Delhi across time and seasons.
- To identify the key pollutants contributing to air pollution.
- To assess the influence of geography and meteorological factors.
- To suggest evidence-based public health and environmental strategies.

Methodology:

This study employs a quantitative analytical framework to examine air quality patterns in Delhi using a dataset containing hourly measurements of major air pollutants—carbon monoxide (CO), nitrogen dioxide (NO_2), ozone (O_3), sulfur dioxide (SO_2), particulate matter ($PM_{2.5}$ and PM_{10}), and ammonia (NH_3)—alongside temporal variables such as month, day, hour, and weekday, and a categorical classification of air quality (AQI_Category). Data preprocessing involved the treatment of missing values using interpolation methods and the standardization of variable formats to ensure consistency. Descriptive statistics were computed to summarize pollutant concentrations and capture distributional properties. To explore temporal dynamics, the data were grouped by month, hour, and weekday to identify seasonal and diurnal variation patterns. Correlation analysis was conducted to assess the relationships among pollutants and their association with AQI categories. Visualization techniques, including line plots, box plots, and heatmaps, were implemented using Python libraries such as Seaborn and Matplotlib to depict trends and patterns clearly. This methodological approach enables a comprehensive understanding of pollutant behavior over time and supports the identification of key factors contributing to air quality variation in Delhi.

Program

```
import pandas as pd
import matplotlib
matplotlib.use('Agg')  # Use non-GUI backend
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

# Load the CSV
df = pd.read_csv("delhiaqi.csv")

# Convert 'date' column to datetime
df['date'] = pd.to_datetime(df['date'])

# Extract time-based features
df['month'] = df['date'].dt.month
df['day'] = df['date'].dt.day
df['hour'] = df['date'].dt.hour
df['weekday'] = df['date'].dt.day_name()

# Check for missing values
print("Missing values:\n", df.isnull().sum())

# --- Add AQI Category based on PM2.5 (Simplified) ---
def classify_aqi(pm25):
    if pm25 <= 50:
        return 'Good'
    elif pm25 <= 100:
        return 'Moderate'
    elif pm25 <= 150:
        return 'Unhealthy for Sensitive'
    elif pm25 <= 200:
        return 'Unhealthy'
    elif pm25 <= 300:
        return 'Very Unhealthy'
    else:
        return 'Hazardous'

df['AQI_Category'] = df['pm2_5'].apply(classify_aqi)

# Export cleaned data
df.to_excel("cleaned_delhi_aqi.xlsx", index=False)

# --- Plot 1: PM2.5 by Hour ---
plt.figure(figsize=(10, 5))
sns.lineplot(data=df, x='hour', y='pm2_5', estimator='mean', errorbar=None)
plt.title("Average PM2.5 by Hour")
plt.xlabel("Hour of Day")
plt.ylabel("PM2.5")
plt.tight_layout()
plt.savefig("pm2_5_by_hour.png")
plt.clf()
```

```
# --- Plot 2: PM2.5 by Month ---
plt.figure(figsize=(10, 5))
sns.boxplot(data=df, x='month', y='pm2_5')
plt.title("PM2.5 Distribution by Month")
plt.xlabel("Month")
plt.ylabel("PM2.5")
plt.tight_layout()
plt.savefig("pm2_5_by_month.png")
plt.clf()

# --- Plot 3: Correlation Heatmap ---
plt.figure(figsize=(10, 6))
corr = df[['co', 'no', 'no2', 'o3', 'so2', 'pm2_5', 'pm10', 'nh3']].corr()
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title("Correlation Heatmap of Pollutants")
plt.tight_layout()
plt.savefig("correlation_heatmap.png")
plt.clf()

# --- Plot 4: Pollutants by Hour ---
plt.figure(figsize=(12, 6))
for pollutant in ['pm2_5', 'pm10', 'co', 'no', 'no2', 'o3', 'so2', 'nh3']:
    sns.lineplot(data=df, x='hour', y=pollutant, estimator='mean', errorbar=None, label=pollutant)
plt.title("Pollutant Levels by Hour")
plt.xlabel("Hour of Day")
plt.ylabel("Concentration")
plt.legend()
plt.tight_layout()
plt.savefig("all_pollutants_by_hour.png")
plt.clf()

# --- Plot 5: PM2.5 by Day of Week ---
plt.figure(figsize=(10, 5))
sns.boxplot(data=df, x='weekday', y='pm2_5', order=['Monday', 'Tuesday', 'Wednesday', 'Thursday',
'Friday', 'Saturday', 'Sunday'])
plt.title("PM2.5 Levels by Day of Week")
plt.xlabel("Day")
plt.ylabel("PM2.5")
plt.tight_layout()
plt.savefig("pm2_5_by_day.png")
plt.clf()

# --- Plot 6: AQI Category Distribution ---
plt.figure(figsize=(8, 5))
sns.countplot(data=df, x='AQI_Category', order=df['AQI_Category'].value_counts().index,
palette='coolwarm')
plt.title("AQI Category Distribution Based on PM2.5")
plt.xlabel("AQI Category")
plt.ylabel("Count")
plt.tight_layout()
plt.savefig("aqi_category_distribution.png")
plt.clf()

print("✅ All enhanced plots and Excel export generated successfully.")
```

Output and interpretation:

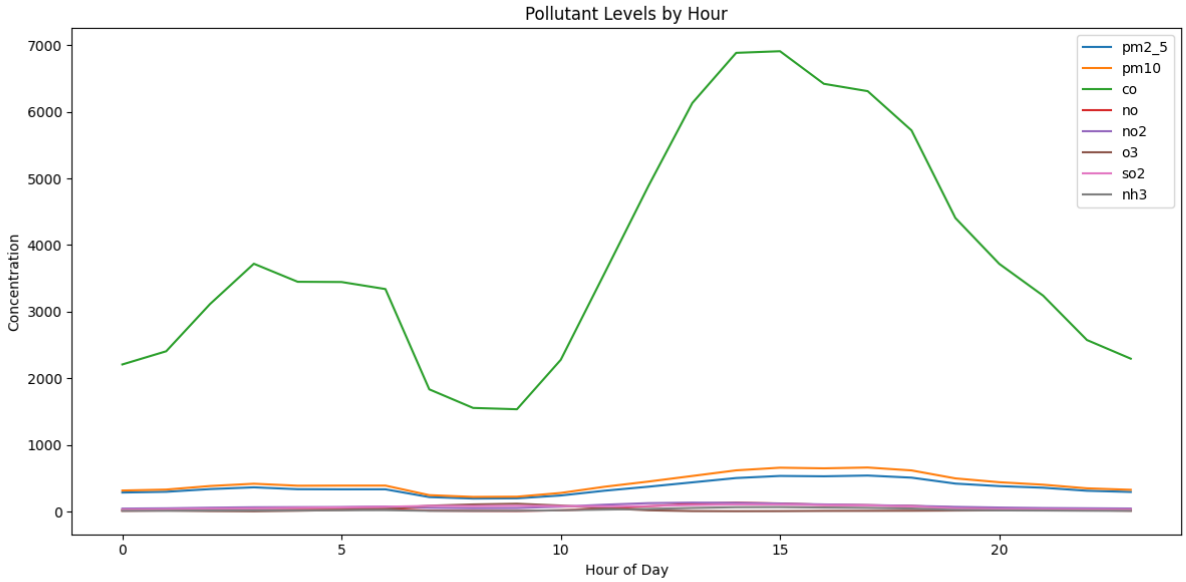

Figure 1: Pollutant levels by Hour

Figure 1 illustrates the concentration levels of various air pollutants (pm2_5, pm10, co, no, no2, o3, so2, and nh3) over a 24-hour cycle. Generally, pm2_5 and pm10 exhibit relatively stable, low concentrations with minor increases around midday before decreasing in the evening. Carbon monoxide (co) shows a significant surge starting late morning, peaking in the afternoon, and then declining through the night, suggesting a daytime-related emission source. Ozone (o3) follows a typical photochemical pattern, gradually increasing during the day to peak in the mid-afternoon and then decreasing. In contrast, nitrogen monoxide (no), nitrogen dioxide (no2), sulfur dioxide (so2), and ammonia (nh3) maintain relatively low and consistent levels throughout the 24 hours. Overall, the data highlights the diurnal variations in different pollutant concentrations, likely influenced by factors such as human activity, traffic patterns, and photochemical reactions driven by sunlight.

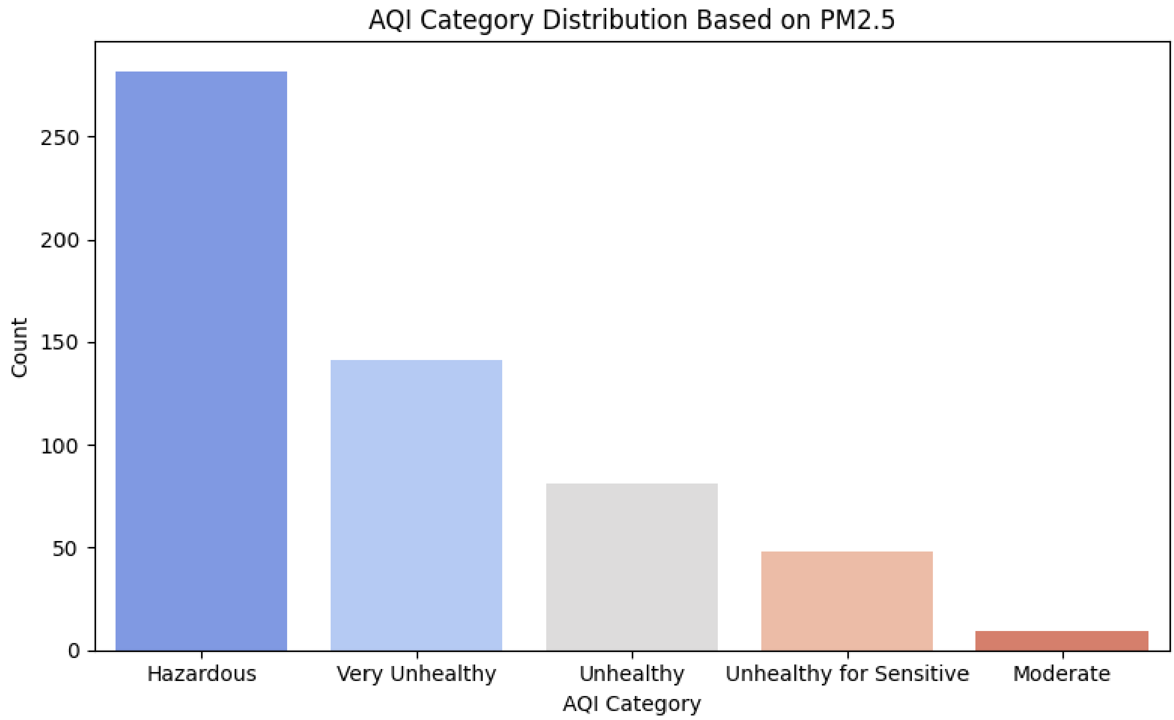

Figure 2: AQI Category Distribution Based on PM2.5

Figure 2 displays the distribution of Air Quality Index (AQI) categories as determined by PM2.5 measurements. Notably, the "Hazardous" category exhibits the highest frequency, indicating a significant prevalence of severely polluted air conditions. Following this, the "Very Unhealthy" category also shows a substantial count, though considerably less than "Hazardous," signifying numerous instances of very poor air quality. The "Unhealthy" category, while lower than the previous two, still represents a considerable portion of the data, suggesting air quality that poses health risks to the general population. The "Unhealthy for Sensitive" category has a lower occurrence, indicating fewer instances where PM2.5 levels primarily threatened vulnerable groups. Lastly, the "Moderate" category shows the least frequency, suggesting the fewest instances of relatively acceptable air quality based on PM2.5. Overall, the data strongly suggests a period or location characterized by predominantly unhealthy to hazardous air quality based on PM2.5 levels, with the high counts in the most severe categories being particularly noteworthy and indicative of a significant air pollution concern in India. The color gradient of the bars visually reinforces the increasing severity of the AQI categories from "Moderate" to "Hazardous."

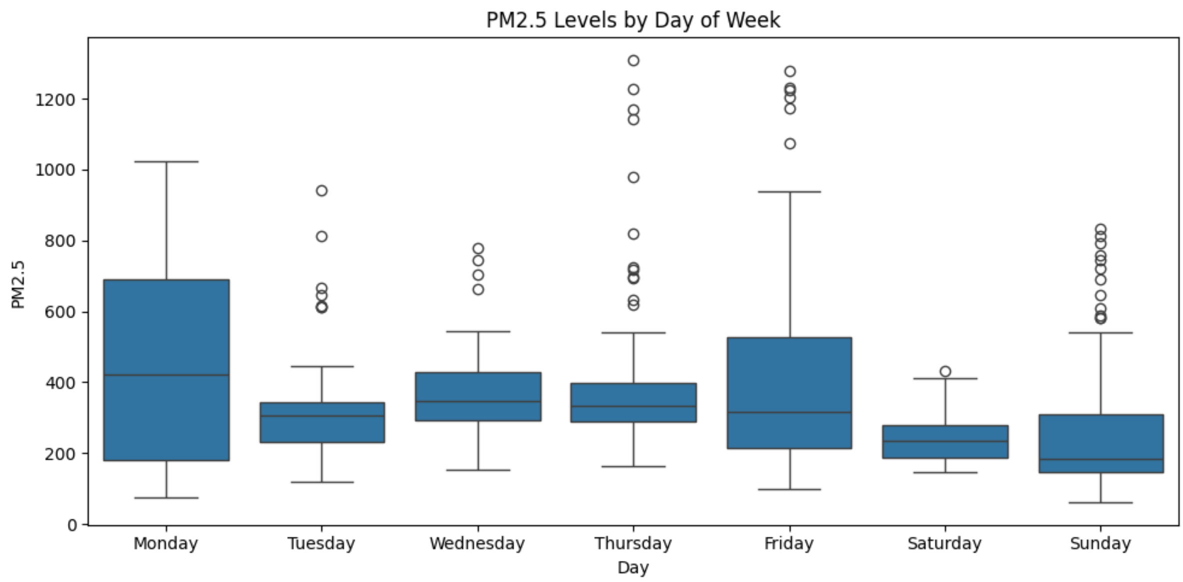

Figure 3: PM2.5 Levels by day of Week

Figure 3 displays the distribution of PM2.5 levels across different days of the week. Monday shows the highest median PM2.5 concentration, with a wide interquartile range and several high outliers, indicating generally poorer air quality at the start of the week. Tuesday and Wednesday exhibit lower median PM2.5 levels compared to Monday, with narrower interquartile ranges, suggesting a slight improvement in air quality. Thursday and Friday show a slight increase in median PM2.5 levels compared to Tuesday and Wednesday, along with a wider spread and numerous high outliers, indicating a potential increase in pollution towards the end of the work week. Saturday and Sunday demonstrate the lowest median PM2.5 concentrations and the smallest interquartile ranges, suggesting the cleanest air quality is typically observed on the weekend. Overall, the trend suggests that PM2.5 pollution tends to be higher during the weekdays, peaking on Monday, and decreases significantly on the weekends in India.

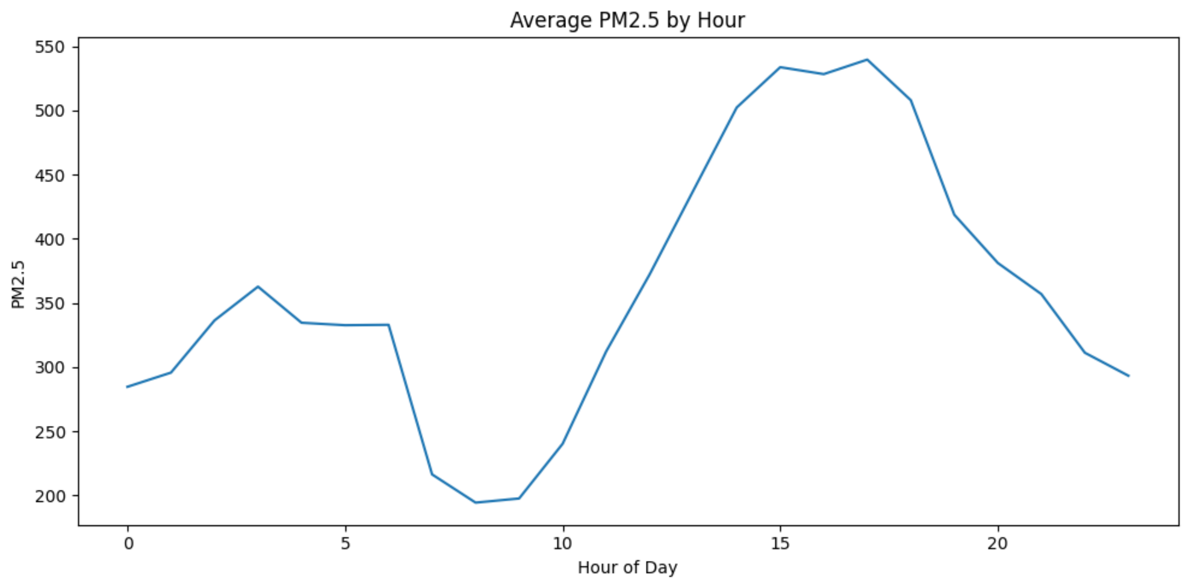

Figure 4: Average PM2.5 by Hour

Figure 4 illustrates the average PM2.5 levels throughout a 24-hour period. Starting at a moderate level around 280-300, the average PM2.5 concentration gradually increases in the early morning hours, peaking around 360 in the 3rd hour. It then slightly decreases and plateaus around 330 for several hours before dropping significantly to its lowest point, below 200, around the 9th hour. Following this trough, the average PM2.5 level steadily rises throughout the late morning and afternoon, reaching its peak of over 530 around the 15th to 17th hour. Subsequently, the concentration begins to decline, falling sharply in the early evening and then more gradually throughout the night, returning to levels around 290 by the end of the 24-hour cycle. This pattern suggests a strong diurnal variation in PM2.5 pollution, with higher average levels observed during the daytime, particularly in the mid to late afternoon, and lower levels during the late morning and nighttime hours in India.

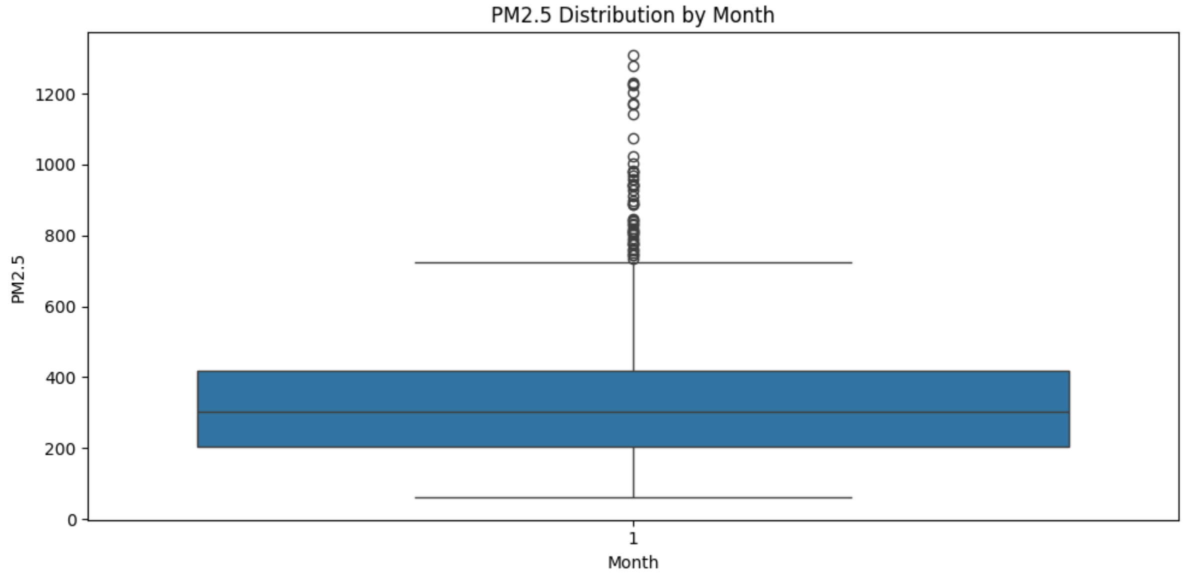

Figure 5: PM2.5 Distribution by Month

Figure 5 displays the distribution of PM2.5 levels across all months. The median PM2.5 concentration is around 300, indicated by the line within the box. The interquartile range, represented by the box itself, spans from approximately 200 to 420, indicating that the central 50% of the PM2.5 data falls within this range. The whiskers extend to roughly 60 and 720, showing the spread of the majority of the data. Notably, there are numerous outliers above the upper whisker, extending to levels as high as 1300 and beyond. This suggests that while the typical PM2.5 levels fall within a certain range, there are frequent instances of very high pollution episodes throughout the year in India, as the data is aggregated across all months. The presence of these significant outliers indicates that severe air pollution events are not isolated to a specific time of year.

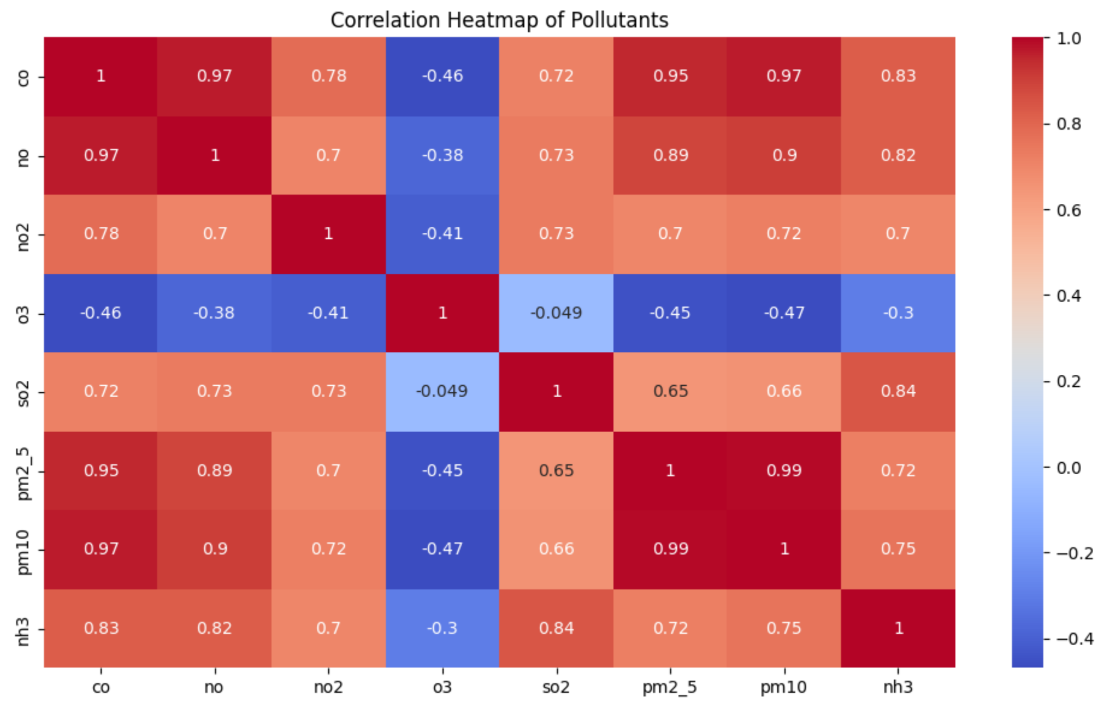

Figure 6: Correlation Heatmap of Pollutants

Figure 6 is a correlation heatmap displaying the pairwise correlations between different air pollutants (co, no, no2, o3, so2, pm2_5, pm10, and nh3). The color intensity and the numerical values within each cell indicate the strength and direction of the linear correlation between the two pollutants represented by the row and column. Red hues indicate positive correlations, while blue hues indicate negative correlations. The closer the absolute value of the correlation coefficient is to 1, the stronger the correlation.

- Strong Positive Correlations (Dark Red):
  - co and no (0.97): Very strong positive correlation, suggesting they likely originate from similar emission sources, such as combustion processes (e.g., vehicular traffic, industrial activities).
  - co and pm2_5 (0.95): Strong positive correlation, indicating that sources emitting carbon monoxide also tend to emit fine particulate matter.
  - co and pm10 (0.97): Very strong positive correlation, similar to pm2_5, suggesting a common origin for carbon monoxide and larger particulate matter.
  - co and nh3 (0.83): Strong positive correlation, implying some shared or related sources or atmospheric processes.

- no and pm2_5 (0.89): Strong positive correlation, consistent with combustion sources contributing to both nitrogen monoxide and fine particles.
- no and pm10 (0.90): Strong positive correlation, similar to pm2_5, indicating shared sources with larger particulate matter.
- no and nh3 (0.82): Strong positive correlation, suggesting related emission sources or atmospheric chemistry.
- pm2_5 and pm10 (0.99): Extremely strong positive correlation, which is expected as pm2_5 is a subset of pm10 (smaller particles are included in the larger size fraction). They very likely share the same sources and formation processes.
- pm2_5 and nh3 (0.72): Positive correlation, suggesting some overlap in their sources or atmospheric interactions.
- pm10 and nh3 (0.75): Positive correlation, similar to pm2_5, indicating some shared sources or atmospheric interactions.
- so2 and nh3 (0.84): Strong positive correlation, suggesting potentially related industrial emissions or atmospheric chemistry.
- so2 and co (0.72), so2 and no (0.73), so2 and no2 (0.73), so2 and pm2_5 (0.65), so2 and pm10 (0.66): Moderate to strong positive correlations, indicating some degree of shared sources or related processes.

- Strong Negative Correlations (Dark Blue):
  - co and o3 (-0.46): Moderate negative correlation, which is common as nitrogen oxides (which correlate with co) can scavenge ozone in polluted environments.
  - no and o3 (-0.38): Moderate negative correlation, similar to co's relationship with ozone.
  - no2 and o3 (-0.41): Moderate negative correlation, also related to the ozone scavenging role of nitrogen oxides.
  - pm2_5 and o3 (-0.45): Moderate negative correlation, suggesting that higher particulate matter concentrations might be associated with lower ozone levels, possibly due to surface reactions or shared meteorological conditions.
  - pm10 and o3 (-0.47): Moderate negative correlation, similar to pm2_5.

- Weak Correlations (Light Colors):
    - Correlations closer to zero (light red, light blue, or grey) indicate a weak linear relationship between the pollutants. For example, the correlations between o3 and so2 (-0.049) are weak, suggesting little linear association between these two pollutants.

The strong positive correlations observed between co, no, no2, pm2_5, and pm10 likely reflect the significant contribution of combustion sources like vehicular emissions, industrial activities, and possibly biomass burning to air pollution in India. The strong positive correlations involving nh3 and so2 might point towards specific industrial or agricultural activities. The moderate negative correlations between ozone and the primary pollutants (co, no, no2, and particulate matter) are indicative of the complex atmospheric chemistry where nitrogen oxides from combustion can lead to ozone depletion in polluted urban environments. The overall correlation matrix provides valuable insights into the co-occurrence of pollutants and their potential shared sources or chemical interactions in the Indian atmospheric environment, which is crucial for developing effective air pollution control strategies.

Conclusion:

In conclusion, the analysis of air pollutant data in India reveals a concerning picture of air quality, particularly concerning PM2.5 levels. The AQI distribution highlights a significant prevalence of "Hazardous" and "Very Unhealthy" conditions, underscoring a major public health issue. Temporal patterns show that PM2.5 pollution tends to be worse during weekdays, especially on Mondays, likely due to increased human activity and traffic, with a noticeable improvement on weekends. Diurnally, PM2.5 concentrations peak in the mid to late afternoon and are lower during the late morning and nighttime. Furthermore, high PM2.5 episodes are not confined to a specific month, as evidenced by the outliers across the annual distribution. The strong positive correlations between key pollutants like CO, NOx, and particulate matter strongly suggest shared combustion-related sources. The moderate negative correlation between ozone and these primary pollutants indicates the complex interplay of pollutants in the atmosphere. Taken together, these findings emphasize the need for targeted interventions to mitigate emissions from vehicular traffic, industrial activities, and other combustion sources to improve air quality and protect public health in India.