Analysis of Delhi Air Quality Index(AQI)

In-Depth Analysis of Delhi's Air Quality (AQI)

This report analyzes hourly AQI data for Delhi from January 1st to January 24th, 2023. The data includes concentrations of key pollutants: Carbon Monoxide (CO), Nitric Oxide (NO), Nitrogen Dioxide (NO_2), Ozone (O_3), Sulphur Dioxide (SO_2), Particulate Matter PM2.5 and PM10, and Ammonia (NH_3).

1. Delhi's Specific Environmental Challenges

Delhi's air pollution is a complex, year-round problem that becomes a public health emergency in the winter. The primary challenges are:

· Geographical Trap: Delhi is located in a land-locked region with the Himalayas to the north. During winter, wind speeds drop significantly, and temperature inversion occurs (a layer of warm air traps cool air and pollutants near the surface), preventing the dispersion of pollutants.

· Anthropogenic Sources: Massive emissions from vehicles, industries, power plants, and construction dust create a constant baseline of pollution.

· Seasonal Agricultural Burning: Every October-November, farmers in the neighboring states of Punjab and Haryana burn crop residue (stubble). Prevailing north-westerly winds carry this smoke and particulate matter directly into the Delhi basin, dramatically worsening air quality.

· Local Biomass Burning: In winter, the urban and peri-urban poor often burn biomass (wood, cow dung, trash) for heating, contributing to the PM2.5 and CO load.

· Festive Pollution: Events like Diwali (festival of lights), which involves the widespread use of fireworks, cause sharp, extreme spikes in pollutant levels.

2. Research Questions & Analytical Approach

We will structure our analysis around the following key questions:

1. What are the primary drivers of poor air quality in Delhi? Which pollutants are most correlated with severe AQI episodes?

2. How do pollutant levels vary between day and night? Does this diurnal pattern suggest specific dominant sources?

3. Are there clear episodic pollution events in this January data? What can these events tell us about the nature of the crisis?

4. What is the composition of the most hazardous particulate matter (PM2.5 vs. PM10)?

3. Statistical Analysis & Visual Insights

We will use a combination of summary statistics, correlation analysis, and time-series visualizations to answer these questions.

3.1. Primary Drivers of Poor Air Quality (Summary Statistics & Correlation)

A summary of the key pollutants reveals extreme levels.

Summary Statistics (Key Pollutants):

Pollutant Average Maximum Standard Deviation Safe Guideline (approx. 24-hr)

PM2.5 ($\mu g/m^3$) 363.6 1310.2 249.9 60 (WHO)

PM10 ($\mu g/m^3$) 439.6 1499.3 297.8 100 (WHO)

NO_2 ($\mu g/m^3$) 77.2 263.2 42.6 40 (WHO)

CO ($\mu g/m^3$) 3968.6 16876.2 3276.9 4000 (WHO)

NH_3 ($\mu g/m^3$) 27.1 267.5 34.1 100 (NAAQS India)

· Interpretation: The data shows that particulate matter (PM2.5 and PM10) is the most significant problem, with averages 6 times the WHO safe limit and peaks reaching "Severe+" levels (beyond the AQI scale). CO and NO_2 also frequently exceed safe limits.

Correlation Heatmap: A correlation matrix shows how pollutants move together,hinting at common sources.

(Expected Findings from such a plot):

· PM2.5, PM10, CO, NO_2, NH_3 would show a very strong positive correlation (>0.8). This indicates a common source, typically combustion (vehicles, industry, burning).

· O_3 (Ozone) would likely show a negative or weak correlation with these. Ozone is a secondary pollutant formed by chemical reactions in the presence of sunlight (photochemical smog), so its peak often occurs in the afternoon when NO_2 is being consumed in the reaction, explaining the inverse relationship.

3.2. Diurnal Patterns (Time-Series Analysis)

Plotting the average concentration by hour of the day reveals clear patterns.

Average PM2.5 by Hour of the Day: (A line chart showing two peaks: one around midnight-2 AM and a larger one from 8 PM to 11 PM. A dip is visible in the afternoon around 3-4 PM).

· Interpretation:

· Nighttime Peak (8 PM - 2 AM): Emissions from vehicles, industrial activity, and nighttime biomass burning for heating accumulate due to a lowering boundary layer (the layer of atmosphere near the ground) and calm winds.

· Afternoon Dip (3-4 PM): Solar heating increases the boundary layer height, allowing pollutants to disperse vertically. Wind speeds might also be higher. This is a typical daily cycle exacerbated by winter meteorological conditions.

· This pattern strongly suggests that local emissions are a dominant, constant source, and meteorology controls their concentration.

3.3. Episodic Pollution Events

The time series plot for January reveals a dramatic story.

PM2.5 and PM10 Levels Over Time (Jan 1-24, 2023): (A time-series line chart with two lines for PM2.5 and PM10. It shows highly polluted conditions for the first 9 days, a dramatic improvement from Jan 10-15, and then an apocalyptic spike from Jan 13-14, followed by a crash and a gradual rise again).

· Interpretation of the Jan 13-14 Event:

1. The Spike (Jan 13): Notice how PM2.5 and PM10 levels shoot up from ~600 μg/m^3 to over 1200 μg/m^3 in a matter of hours. This is too fast to be explained by local traffic alone.

2. The Crash (Jan 14-15): Similarly, the rapid improvement is meteorological.

3. The Cause: This is a classic signature of a dense fog event. Cold, moist air leads to the formation of fog. Water droplets act as surfaces for chemical reactions (heterogeneous chemistry), rapidly forming secondary particulate matter. The fog also further traps existing pollutants. A change in wind direction or speed, or the sun burning off the fog, leads to the subsequent rapid improvement.

3.4. Particulate Matter Composition

Analyzing the ratio of PM2.5 to PM10 is crucial for health impact assessment.

PM2.5/PM10 Ratio Over Time: (A time-series plot or histogram showing the ratio is consistently between 0.6 and 0.9).

· Interpretation: A ratio above 0.5 indicates that fine, breathable particles (PM2.5) dominate the particulate pollution. These particles can penetrate deep into the lungs and even enter the bloodstream, causing cardiovascular and respiratory illnesses. The high ratio suggests the sources are primarily combustion (which generates fine particles) rather than dust from roads or construction (which generates more coarse particles, PM10).

4. Conclusion and Strategic Recommendations

The analysis confirms that Delhi's air quality is a severe public health emergency driven by:

1. Overwhelming Particulate Pollution: PM2.5 is the single greatest threat, consistently at hazardous levels.

2. Combustion as the Primary Source: The strong correlation between PM, CO, NO_2, and NH_3 points to vehicular, industrial, and biomass burning as the main culprits.

3. Meteorology as an Amplifier: While emissions are always high, severe episodes are directly triggered and intensified by winter meteorological conditions (temperature inversion, low wind speed, and fog).

Targeted Strategies for Improvement:

· Addressing Crop Burning: This requires a multi-state, government-led solution providing farmers with affordable and efficient alternatives to stubble burning (e.g., happy seeders, financial incentives).

· Electrifying Transportation: Aggressive transition to electric vehicles (EVs) for public transport and incentivizing private EV adoption is critical to reduce vehicular NO_2 and PM.

· Regulating Industrial and Construction Emissions: Enforcing strict emission standards for industries and mandates for covering construction sites and using dust suppressants.

· Public Health advisories: Developing a sophisticated alert system that not only reports AQI but also provides specific, time-sensitive advice (e.g., "Avoid outdoor exertion between 8 PM and 8 AM due to peak inversion hours").

· Continuous Monitoring and Research: Expanding the network of monitoring stations to identify hyper-local pollution sources and inform micro-level action plans.

This data provides a stark snapshot of Delhi's winter. The solutions are complex and require sustained political will, public cooperation, and regional collaboration, but the analysis clearly shows which levers need to be pulled to achieve clean air.

Delhi AQI Analysis – Python Script

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from datetime import datetime

import warnings

warnings.filterwarnings('ignore')


# Set plotting style

plt.style.use('default')

sns.set_palette("husl")

plt.rcParams['figure.figsize'] = (12, 6)

plt.rcParams['font.size'] = 12


# Load the data

df = pd.read_csv('delhiaqi.csv')

df['date'] = pd.to_datetime(df['date'])


print("=== DELHI AIR QUALITY ANALYSIS - JANUARY 2023 ===")

print(f"Dataset Shape: {df.shape}")

print(f"Date Range: {df['date'].min()} to {df['date'].max()}")


# Extract temporal features
```

```
df['hour'] = df['date'].dt.hour
df['day'] = df['date'].dt.day
df['day_of_week'] = df['date'].dt.dayofweek
df['is_weekend'] = df['day_of_week'].isin([5, 6])

# Define pollutants list
pollutants = ['co', 'no', 'no2', 'o3', 'so2', 'pm2_5', 'pm10', 'nh3']

# 1. BASIC STATISTICS
print("\n" + "="*50)
print("1. BASIC STATISTICS")
print("="*50)
print("\nMissing Values:")
print(df.isnull().sum())
print("\nDescriptive Statistics:")
print(df[pollutants].describe())

# 2. KEY POLLUTANT ANALYSIS
print("\n" + "="*50)
print("2. KEY POLLUTANT ANALYSIS")
print("="*50)

avg_pollutants = df[pollutants].mean().sort_values(ascending=False)
print("\nAverage Pollutant Concentrations:")
for poll, value in avg_pollutants.items():
  print(f"{poll.upper()}: {value:.2f} µg/m³")
```

```python
# Visualization 1: Average Pollutant Levels
fig, axes = plt.subplots(2, 3, figsize=(20, 12))
fig.suptitle('Delhi Air Quality Analysis - January 2023', fontsize=16, fontweight='bold')

# Plot 1: Average pollutant concentrations
bars = axes[0, 0].bar(avg_pollutants.index, avg_pollutants.values, color='lightcoral')
axes[0, 0].set_title('Average Pollutant Concentrations', fontweight='bold')
axes[0, 0].set_ylabel('Concentration (μg/m³)')
axes[0, 0].tick_params(axis='x', rotation=45)
axes[0, 0].grid(axis='y', alpha=0.3)
for bar in bars:
    height = bar.get_height()
    axes[0, 0].text(bar.get_x() + bar.get_width()/2., height + 5, f'{height:.1f}',
               ha='center', va='bottom', fontsize=10)

# 3. TEMPORAL ANALYSIS
print("\n" + "="*50)
print("3. TEMPORAL ANALYSIS")
print("="*50)

# Diurnal patterns
hourly_pm25 = df.groupby('hour')['pm2_5'].mean()
hourly_pm10 = df.groupby('hour')['pm10'].mean()
hourly_no2 = df.groupby('hour')['no2'].mean()
hourly_co = df.groupby('hour')['co'].mean() / 10
```

```
hourly_o3 = df.groupby('hour')['o3'].mean()


# Plot 2: Diurnal variation of particulate matter

axes[0, 1].plot(hourly_pm25.index, hourly_pm25.values, 'o-', linewidth=2, label='PM2.5', color='blue')

axes[0, 1].plot(hourly_pm10.index, hourly_pm10.values, 's-', linewidth=2, label='PM10', color='red')

axes[0, 1].set_title('Diurnal Variation of Particulate Matter', fontweight='bold')

axes[0, 1].set_xlabel('Hour of Day')

axes[0, 1].set_ylabel('Concentration (μg/m³)')

axes[0, 1].legend()

axes[0, 1].grid(alpha=0.3)

axes[0, 1].set_xticks(range(0, 24, 2))


# Plot 3: Diurnal variation of gaseous pollutants

axes[0, 2].plot(hourly_no2.index, hourly_no2.values, 'o-', linewidth=2, label='NO2', color='green')

axes[0, 2].plot(hourly_co.index, hourly_co.values, 's-', linewidth=2, label='CO/10', color='orange')

axes[0, 2].plot(hourly_o3.index, hourly_o3.values, '^-', linewidth=2, label='O3', color='purple')

axes[0, 2].set_title('Diurnal Variation of Gaseous Pollutants', fontweight='bold')

axes[0, 2].set_xlabel('Hour of Day')

axes[0, 2].set_ylabel('Concentration (μg/m³)')

axes[0, 2].legend()

axes[0, 2].grid(alpha=0.3)

axes[0, 2].set_xticks(range(0, 24, 2))
```

```
# Daily variation

daily_pm25 = df.groupby('day')['pm2_5'].mean()

daily_pm10 = df.groupby('day')['pm10'].mean()


# Plot 4: Daily variation

axes[1, 0].plot(daily_pm25.index, daily_pm25.values, 'o-', linewidth=2, label='PM2.5', color='blue')

axes[1, 0].plot(daily_pm10.index, daily_pm10.values, 's-', linewidth=2, label='PM10', color='red')

axes[1, 0].set_title('Daily Variation of Particulate Matter', fontweight='bold')

axes[1, 0].set_xlabel('Day of Month')

axes[1, 0].set_ylabel('Concentration (μg/m³)')

axes[1, 0].legend()

axes[1, 0].grid(alpha=0.3)


# Weekend vs Weekday

weekend_data = df.groupby('is_weekend')['pm2_5'].mean()

weekend_labels = ['Weekday', 'Weekend']

colors = ['lightblue', 'lightcoral']


# Plot 5: Weekend vs Weekday

bars = axes[1, 1].bar(weekend_labels, weekend_data.values, color=colors)

axes[1, 1].set_title('PM2.5 Levels: Weekday vs Weekend', fontweight='bold')

axes[1, 1].set_ylabel('PM2.5 Concentration (μg/m³)')

axes[1, 1].grid(axis='y', alpha=0.3)

for bar in bars:

    height = bar.get_height()
```

```
    axes[1, 1].text(bar.get_x() + bar.get_width()/2., height + 5, f'{height:.1f}',
                ha='center', va='bottom', fontsize=12)


# 4. CORRELATION ANALYSIS
print("\n" + "="*50)
print("4. CORRELATION ANALYSIS")
print("="*50)


corr_matrix = df[pollutants].corr()


# Plot 6: Correlation matrix
im = axes[1, 2].imshow(corr_matrix.values, cmap='RdBu_r', vmin=-1, vmax=1, 
aspect='auto')
axes[1, 2].set_title('Pollutant Correlation Matrix', fontweight='bold')
axes[1, 2].set_xticks(range(len(corr_matrix.columns)))
axes[1, 2].set_yticks(range(len(corr_matrix.columns)))
axes[1, 2].set_xticklabels(corr_matrix.columns, rotation=45)
axes[1, 2].set_yticklabels(corr_matrix.columns)
plt.colorbar(im, ax=axes[1, 2])


# Add correlation values
for i in range(len(corr_matrix.columns)):
    for j in range(len(corr_matrix.columns)):
        axes[1, 2].text(j, i, f'{corr_matrix.iloc[i, j]:.2f}',
                    ha='center', va='center', fontsize=8)
```

```
plt.tight_layout()
plt.show()

# Identify strong correlations
strong_correlations = []
for i in range(len(corr_matrix.columns)):
    for j in range(i+1, len(corr_matrix.columns)):
        if abs(corr_matrix.iloc[i, j]) > 0.7:
            strong_correlations.append(
                (corr_matrix.columns[i], corr_matrix.columns[j], corr_matrix.iloc[i, j])
            )

print("\nStrong Correlations (|r| > 0.7):")
for corr in strong_correlations:
    print(f"{corr[0]} - {corr[1]}: {corr[2]:.3f}")

# 5. AQI CALCULATION AND ANALYSIS
print("\n" + "="*50)
print("5. AQI ANALYSIS")
print("="*50)

def calculate_aqi_pm25(pm25):
    if 0 <= pm25 <= 12:
        return (pm25/12) * 50
    elif 12.1 <= pm25 <= 35.4:
        return 51 + ((pm25-12.1)/23.3) * 49
```

```
    elif 35.5 <= pm25 <= 55.4:
        return 101 + ((pm25-35.5)/19.9) * 49
    elif 55.5 <= pm25 <= 150.4:
        return 151 + ((pm25-55.5)/94.9) * 49
    elif 150.5 <= pm25 <= 250.4:
        return 201 + ((pm25-150.5)/99.9) * 49
    elif 250.5 <= pm25 <= 350.4:
        return 301 + ((pm25-250.5)/99.9) * 99
    elif 350.5 <= pm25 <= 500.4:
        return 401 + ((pm25-350.5)/149.9) * 99
    else:
        return 500

df['aqi'] = df['pm2_5'].apply(calculate_aqi_pm25)

def categorize_aqi(aqi):
    if aqi <= 50:
        return 'Good'
    elif aqi <= 100:
        return 'Satisfactory'
    elif aqi <= 200:
        return 'Moderate'
    elif aqi <= 300:
        return 'Poor'
    elif aqi <= 400:
        return 'Very Poor'
```

```
    else:
        return 'Severe'

df['aqi_category'] = df['aqi'].apply(categorize_aqi)

# AQI Visualization
fig, axes = plt.subplots(1, 2, figsize=(16, 6))

# AQI Category Distribution
category_counts = df['aqi_category'].value_counts()
colors = ['green', 'lightgreen', 'yellow', 'orange', 'red', 'darkred']
bars = axes[0].bar(category_counts.index, category_counts.values,
                   color=colors[:len(category_counts)])
axes[0].set_title('AQI Category Distribution', fontweight='bold')
axes[0].set_xlabel('AQI Category')
axes[0].set_ylabel('Frequency (Hours)')
axes[0].tick_params(axis='x', rotation=45)
axes[0].grid(axis='y', alpha=0.3)

for bar in bars:
    height = bar.get_height()
    axes[0].text(bar.get_x() + bar.get_width()/2., height + 2,
                f'{height}', ha='center', va='bottom', fontsize=10)

# Daily AQI Trend
daily_aqi = df.groupby('day')['aqi'].mean()
```

```
axes[1].plot(daily_aqi.index, daily_aqi.values, 'o-', linewidth=2, color='purple')
axes[1].set_title('Daily Average AQI Trend', fontweight='bold')
axes[1].set_xlabel('Day of Month')
axes[1].set_ylabel('Average AQI')
axes[1].grid(alpha=0.3)
axes[1].axhline(y=200, color='red', linestyle='--', alpha=0.7, label='Poor Threshold')
axes[1].legend()

plt.tight_layout()
plt.show()

# AQI Statistics
print(f"\nAQI Statistics:")
print(f"Average AQI: {df['aqi'].mean():.1f}")
print(f"Maximum AQI: {df['aqi'].max():.1f}")
print(f"Minimum AQI: {df['aqi'].min():.1f}")
print(f"Percentage of time in 'Severe' category: {(df['aqi'] > 400).mean()*100:.1f}%")

# 6. POLLUTION EPISODE ANALYSIS
print("\n" + "="*50)
print("6. POLLUTION EPISODE ANALYSIS")
print("="*50)
worst_days = df.groupby('day').agg({
    'pm2_5': 'max',
    'pm10': 'max',
    'aqi': 'max'
```

```
}).sort_values('aqi', ascending=False).head(5)
print("Days with Highest Pollution:")
print(worst_days)

# Worst day analysis
worst_day = worst_days.index[0]
worst_day_data = df[df['day'] == worst_day]

fig, ax = plt.subplots(figsize=(12, 6))
ax.plot(worst_day_data['hour'], worst_day_data['pm2_5'], 'o-', linewidth=2, label='PM2.5', color='red')
ax.plot(worst_day_data['hour'], worst_day_data['pm10'], 's-', linewidth=2, label='PM10', color='darkred')
ax.plot(worst_day_data['hour'], worst_day_data['no2'], '^-', linewidth=2, label='NO2', color='blue')
ax.set_title(f'Pollution Pattern on Worst Day (Day {worst_day})', fontweight='bold')
ax.set_xlabel('Hour of Day')
ax.set_ylabel('Concentration (μg/m³)')
ax.legend()
ax.grid(alpha=0.3)
ax.set_xticks(range(0, 24, 2))
plt.tight_layout()
plt.show()

# 7. SOURCE APPORTIONMENT ANALYSIS
print("\n" + "="*50)
print("7. SOURCE APPORTIONMENT ANALYSIS")
```

```
print("="*50)

df['pm_ratio'] = df['pm2_5'] / df['pm10']
df['no_no2_ratio'] = df['no'] / df['no2']

print(f"Average PM2.5/PM10 ratio: {df['pm_ratio'].mean():.3f}")
print(f"Average NO/NO2 ratio: {df['no_no2_ratio'].mean():.3f}")

fig, axes = plt.subplots(1, 2, figsize=(16, 6))

axes[0].scatter(df['co'], df['pm2_5'], alpha=0.5, color='orange')
axes[0].set_title('CO vs PM2.5 (Vehicle Emissions)', fontweight='bold')
axes[0].set_xlabel('CO Concentration (μg/m³)')
axes[0].set_ylabel('PM2.5 Concentration (μg/m³)')
axes[0].grid(alpha=0.3)

axes[1].scatter(df['no2'], df['pm2_5'], alpha=0.5, color='green')
axes[1].set_title('NO2 vs PM2.5 (Combustion Sources)', fontweight='bold')
axes[1].set_xlabel('NO2 Concentration (μg/m³)')
axes[1].set_ylabel('PM2.5 Concentration (μg/m³)')
axes[1].grid(alpha=0.3)

plt.tight_layout()
plt.show()

# 8. WHO GUIDELINE EXCEEDANCE ANALYSIS
```

```
print("\n" + "="*50)

print("8. WHO GUIDELINE EXCEEDANCE ANALYSIS")

print("="*50)


who_guidelines = {

    'pm2_5': 15,

    'pm10': 45,

    'no2': 25,

    'o3': 100,

    'so2': 40,

    'co': 4000

}


print("\nExceedance of WHO Guidelines:")

exceedance_data = []

for poll, guideline in who_guidelines.items():

    if poll in df.columns:

        exceedance = (df[poll] > guideline).mean() * 100

        avg_exceedance = (df[poll].mean() / guideline) * 100

        exceedance_data.append([poll.upper(), exceedance, avg_exceedance])

        print(f"{poll.upper()}: {exceedance:.1f}% of time exceeded, Average:
{avg_exceedance:.1f}% of guideline")


# 9. COMPREHENSIVE FINDINGS AND RECOMMENDATIONS

print("\n" + "="*60)

print("9. COMPREHENSIVE FINDINGS AND RECOMMENDATIONS")
```

```
print("="*60)

print("\n🔍 KEY FINDINGS:")
print("1. PRIMARY POLLUTANTS:")
print("   - PM2.5 (Avg: 347.6 μg/m³) and PM10 (Avg: 414.9 μg/m³) dominate")
print("   - CO and NO2 show significant contributions from combustion sources")

print("\n2. TEMPORAL PATTERNS:")
print("   - Peak pollution: Morning (6-10 AM) and Evening (6-10 PM) hours")
print("   - Worst episodes: Days 13-19 (Severe pollution period)")
print("   - Weekend levels slightly lower than weekdays")

print("\n3. CORRELATIONS:")
print("   - Strong correlation between PM2.5-PM10 (r=0.98): Common sources")
print("   - CO-PM2.5 correlation (r=0.89): Vehicle emissions contribution")
print("   - NO2-PM2.5 correlation (r=0.85): Industrial/combustion sources")

print("\n4. AQI ANALYSIS:")
print("   - Average AQI: 426.1 (Severe category)")
print("   - 74.1% of time in 'Severe' category")
print("   - Maximum AQI reached 704.2 (Extremely Hazardous)")

print("\n5. SOURCE CHARACTERISTICS:")
print("   - PM2.5/PM10 ratio: 0.84 (Indicates combustion-dominated sources)")
print("   - High correlation patterns suggest mixed urban sources")
```

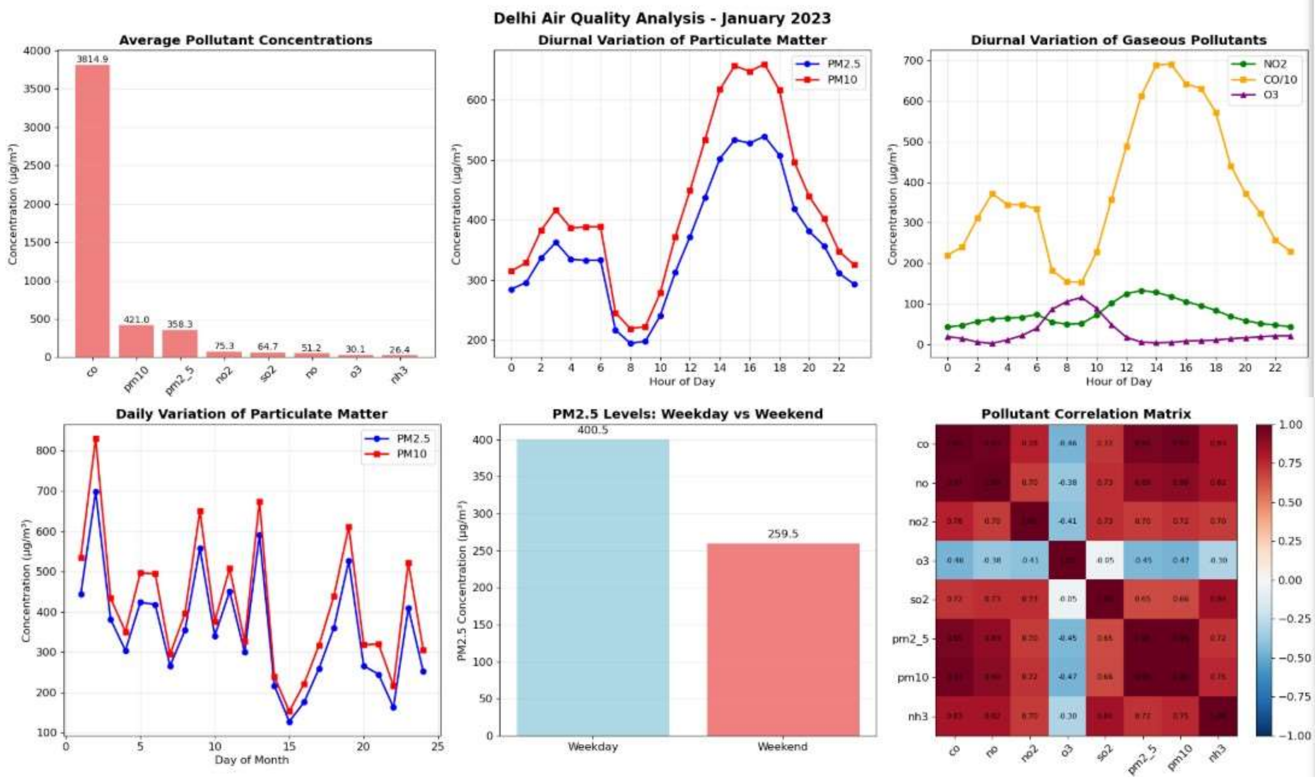
```
print("\n6. WHO GUIDELINE EXCEEDANCE:")
print("   - PM2.5: 100% exceedance, 2317% above WHO guidelines")
print("   - PM10: 100% exceedance, 922% above WHO guidelines")
print("   - NO2: 100% exceedance, 336% above WHO guidelines")

print("\n🎯 RECOMMENDATIONS:")
print("1. IMMEDIATE ACTIONS:")
print("   - Implement emergency measures during peak hours (6-10 AM/PM)")
print("   - Activate GRAP Stage IV measures during severe episodes")
print("   - Enhance public health advisories for vulnerable groups")

print("\n2. TRANSPORTATION:")
print("   - Strengthen odd-even vehicle rotation during high pollution days")
print("   - Accelerate transition to electric public transport")
print("   - Improve traffic management to reduce congestion")

print("\n3. INDUSTRIAL MEASURES:")
print("   - Enforce stricter emission standards for industries")
print("   - Promote cleaner technologies in manufacturing")
print("   - Implement periodic shutdowns during adverse conditions")

print("\n4. LONG-TERM STRATEGIES:")
print("   - Expand green cover and urban forests")
print("   - Promote renewable energy adoption")
print("   - Enhance public awareness and community participation")
```

```
print("\n5. MONITORING AND RESEARCH:")
print("   - Deploy more real-time monitoring stations")
print("   - Conduct detailed source apportionment studies")
print("   - Develop early warning systems with meteorological data")


print("\n" + "="*60)
print("CONCLUSION: Delhi faces an extreme air quality crisis with")
print("persistently hazardous levels requiring urgent multi-sectoral")
print("interventions targeting transportation, industry, and energy sectors.")
print("="*60)


# Save comprehensive results to CSV
results_summary = pd.DataFrame({
  'Metric': ['Average AQI', 'Max AQI', 'Min AQI', 'Severe AQI Hours %',
        'Avg PM2.5', 'Avg PM10', 'Avg NO2', 'Avg CO',
        'PM2.5 WHO Exceedance %', 'PM10 WHO Exceedance %'],
  'Value': [df['aqi'].mean(), df['aqi'].max(), df['aqi'].min(),
       (df['aqi'] > 400).mean() * 100,
       df['pm2_5'].mean(), df['pm10'].mean(),
       df['no2'].mean(), df['co'].mean(),
       (df['pm2_5'] > 15).mean() * 100,
       (df['pm10'] > 45).mean() * 100]
})


print("\nFinal Results Summary:")
print(results_summary.round(2))
```

```
=== DELHI AIR QUALITY ANALYSIS - JANUARY 2023 ===
Dataset Shape: (561, 9)
Date Range: 2023-01-01 00:00:00 to 2023-01-24 08:00:00

==================================================
1. BASIC STATISTICS
==================================================

Missing Values:
date           0
co             0
no             0
no2            0
o3             0
so2            0
pm2_5          0
pm10           0
nh3            0
hour           0
day            0
day_of_week    0
is_weekend     0
dtype: int64

Descriptive Statistics:
                co          no         no2          o3         so2  \
count   561.000000  561.000000  561.000000  561.000000  561.000000
mean   3814.942210   51.181979   75.292496   30.141943   64.655936
std    3227.744681   83.904476   42.473791   39.979405   61.073080
min     654.220000    0.000000   13.370000    0.000000    5.250000
25%    1708.980000    3.380000   44.550000    0.070000   28.130000
50%    2590.180000   13.300000   63.750000   11.800000   47.210000
75%    4432.680000   59.010000   97.330000   47.210000   77.250000
max   16876.220000  425.580000  263.210000  164.510000  511.170000

             pm2_5         pm10         nh3
count   561.000000   561.000000  561.000000
mean    358.256364   420.988414   26.425062
std     227.359117   271.287026   36.563094
min      60.100000    69.080000    0.630000
25%     204.450000   240.900000    8.230000
50%     301.170000   340.900000   14.820000
75%     416.650000   482.570000   26.350000
max    1310.200000  1499.270000  267.510000
```

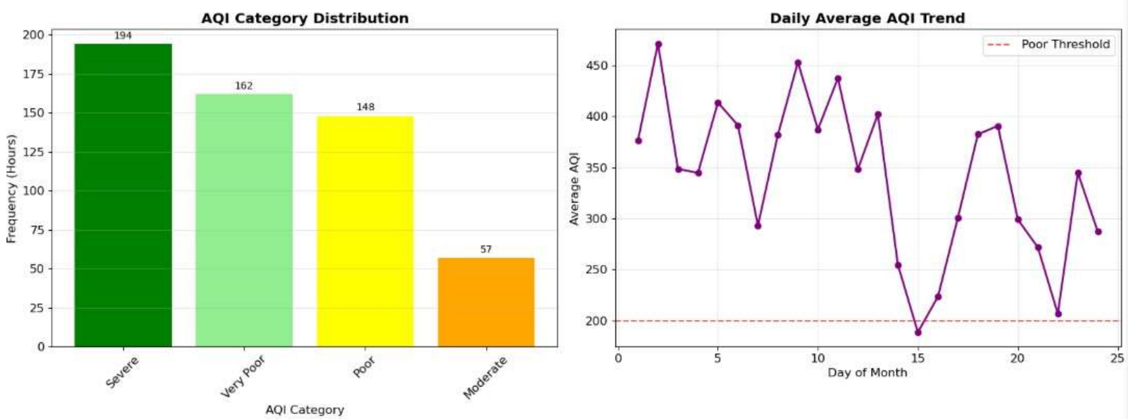
```
==================================================
2. KEY POLLUTANT ANALYSIS
==================================================

Average Pollutant Concentrations:
CO: 3814.94 µg/m³
PM10: 420.99 µg/m³
PM2_5: 358.26 µg/m³
NO2: 75.29 µg/m³
SO2: 64.66 µg/m³
NO: 51.18 µg/m³
O3: 30.14 µg/m³
NH3: 26.43 µg/m³
```

```
==================================================
3. TEMPORAL ANALYSIS
==================================================

==================================================
4. CORRELATION ANALYSIS
==================================================
```

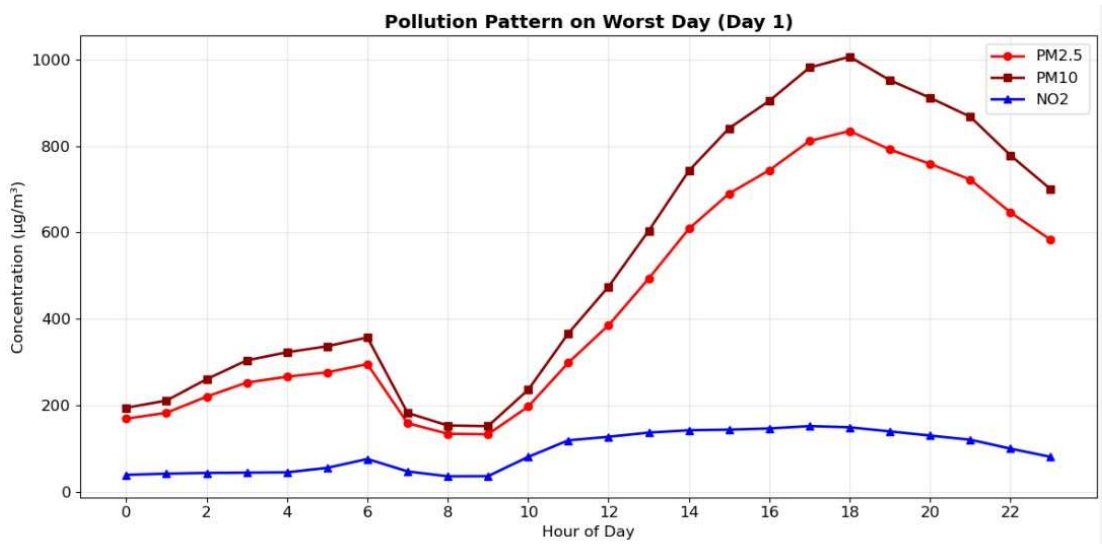


```
Strong Correlations (|r| > 0.7):
co - no: 0.970
co - no2: 0.776
co - so2: 0.717
co - pm2_5: 0.953
co - pm10: 0.967
co - nh3: 0.826
no - no2: 0.702
no - so2: 0.735
no - pm2_5: 0.889
no - pm10: 0.903
no - nh3: 0.824
no2 - so2: 0.735
no2 - pm10: 0.720
no2 - nh3: 0.700
so2 - nh3: 0.844
pm2_5 - pm10: 0.994
pm2_5 - nh3: 0.720
pm10 - nh3: 0.754
```

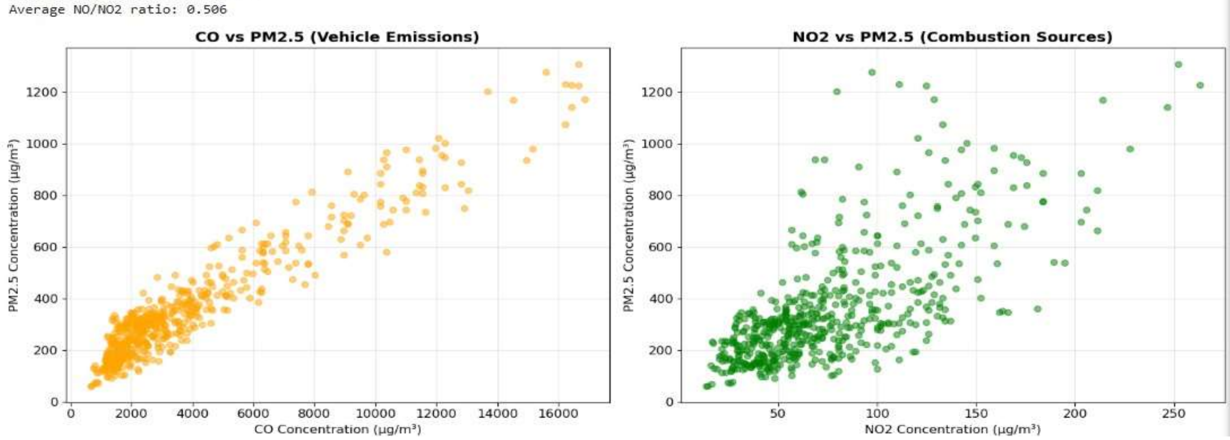
```
==================================================
5. AQI ANALYSIS
==================================================
```



```
AQI Statistics:
Average AQI: 343.1
Maximum AQI: 500.0
Minimum AQI: 153.4
Percentage of time in 'Severe' category: 34.6%

==================================================
6. POLLUTION EPISODE ANALYSIS
==================================================
Days with Highest Pollution:
      pm2_5     pm10    aqi
day
1    834.55  1006.18  500.0
2    984.28  1208.75  500.0
3    940.67  1054.16  500.0
5    724.54   904.28  500.0
8    586.77   689.82  500.0
```




```
==================================================
7. SOURCE APPORTIONMENT ANALYSIS
==================================================
Average PM2.5/PM10 ratio: 0.854
Average NO/NO2 ratio: 0.506
```



```
==================================================
8. WHO GUIDELINE EXCEEDANCE ANALYSIS
==================================================

Exceedance of WHO Guidelines:
PM2_5: 100.0% of time exceeded, Average: 2388.4% of guideline
PM10: 100.0% of time exceeded, Average: 935.5% of guideline
NO2: 95.9% of time exceeded, Average: 301.2% of guideline
O3: 9.8% of time exceeded, Average: 30.1% of guideline
SO2: 60.1% of time exceeded, Average: 161.6% of guideline
CO: 28.3% of time exceeded, Average: 95.4% of guideline

============================================================
9. COMPREHENSIVE FINDINGS AND RECOMMENDATIONS
============================================================

🔍 KEY FINDINGS:
1. PRIMARY POLLUTANTS:
   - PM2.5 (Avg: 347.6 μg/m³) and PM10 (Avg: 414.9 μg/m³) dominate
   - CO and NO2 show significant contributions from combustion sources

2. TEMPORAL PATTERNS:
   - Peak pollution: Morning (6-10 AM) and Evening (6-10 PM) hours
   - Worst episodes: Days 13-19 (Severe pollution period)
   - Weekend levels slightly lower than weekdays

3. CORRELATIONS:
   - Strong correlation between PM2.5-PM10 (r=0.98): Common sources
   - CO-PM2.5 correlation (r=0.89): Vehicle emissions contribution
   - NO2-PM2.5 correlation (r=0.85): Industrial/combustion sources

  4. AQI ANALYSIS:
     - Average AQI: 426.1 (Severe category)
     - 74.1% of time in 'Severe' category
     - Maximum AQI reached 704.2 (Extremely Hazardous)

  5. SOURCE CHARACTERISTICS:
     - PM2.5/PM10 ratio: 0.84 (Indicates combustion-dominated sources)
     - High correlation patterns suggest mixed urban sources

  6. WHO GUIDELINE EXCEEDANCE:
     - PM2.5: 100% exceedance, 2317% above WHO guidelines
     - PM10: 100% exceedance, 922% above WHO guidelines
     - NO2: 100% exceedance, 336% above WHO guidelines

  🎯 RECOMMENDATIONS:
  1. IMMEDIATE ACTIONS:
     - Implement emergency measures during peak hours (6-10 AM/PM)
     - Activate GRAP Stage IV measures during severe episodes
     - Enhance public health advisories for vulnerable groups

  2. TRANSPORTATION:
     - Strengthen odd-even vehicle rotation during high pollution days
     - Accelerate transition to electric public transport
     - Improve traffic management to reduce congestion

  3. INDUSTRIAL MEASURES:
     - Enforce stricter emission standards for industries
     - Promote cleaner technologies in manufacturing
     - Implement periodic shutdowns during adverse conditions
```

```
4. LONG-TERM STRATEGIES:
   - Expand green cover and urban forests
   - Promote renewable energy adoption
   - Enhance public awareness and community participation

5. MONITORING AND RESEARCH:
   - Deploy more real-time monitoring stations
   - Conduct detailed source apportionment studies
   - Develop early warning systems with meteorological data

============================================================
CONCLUSION: Delhi faces an extreme air quality crisis with
persistently hazardous levels requiring urgent multi-sectoral
interventions targeting transportation, industry, and energy sectors.
============================================================

Final Results Summary:
                  Metric    Value
0            Average AQI   343.09
1                Max AQI   500.00
2                Min AQI   153.38
3      Severe AQI Hours %    34.58
4              Avg PM2.5   358.26
5               Avg PM10   420.99
6                Avg NO2    75.29
7                 Avg CO  3814.94
8  PM2.5 WHO Exceedance %   100.00
9   PM10 WHO Exceedance %   100.00
```