SHADOWFOX DATASCIENCE INTERNSHIP:

Name: Srithesh

Email: sritheshp@gmail.com

Intermediate-level:

Air Quality Analysis of Delhi

Introduction:

Delhi is one of the most polluted cities in India. Air pollution mainly comes from vehicles, factories, construction work, and burning of waste and crops. Poor air quality causes many health problems such as breathing issues and heart diseases. This project studies air pollution data of Delhi to understand pollution patterns and problem areas.

1. Objective of the Study

The main goals of this project are:
- To study air pollution levels in Delhi
- To find the most harmful pollutants
- To understand monthly and seasonal pollution changes
- To provide useful insights for improving air quality

2. Methodology

Python was used to analyze the data. First, the data was cleaned by removing missing values and converting dates into the correct format. Then, charts and graphs were created to study pollution trends, monthly averages, seasonal patterns, and relationships between pollutants.

3. Data Analysis and Findings

- PM2.5 and PM10 are the most harmful pollutants in Delhi
- Pollution levels increase during winter months
- Pollution decreases during the monsoon season
- PM2.5 and PM10 are strongly related to overall air pollution

4. Seasonal Analysis

Winter has the worst air quality in Delhi. This happens because cold air traps pollution near the ground and wind speed is low. Burning of crop waste and heavy traffic also increase pollution. During the monsoon season, rain helps clean the air, so pollution levels are lower.

8. Conclusion

This project shows that air pollution in Delhi is a serious issue. High levels of PM2.5 and PM10 are dangerous for health. Seasonal changes, especially winter, play a major role in pollution levels.

9. Recommendations

- Use public transport and electric vehicles more often
- Reduce pollution from factories and construction work
- Stop burning waste and crop residue
- Warn people during high pollution days so they can stay safe

Delhi Air Quality Index (AQI) Analysis Report

This project analyses Delhi's Air Quality Index (AQI) using pollution data to understand trends, peak pollution hours, and relationships between pollutants.

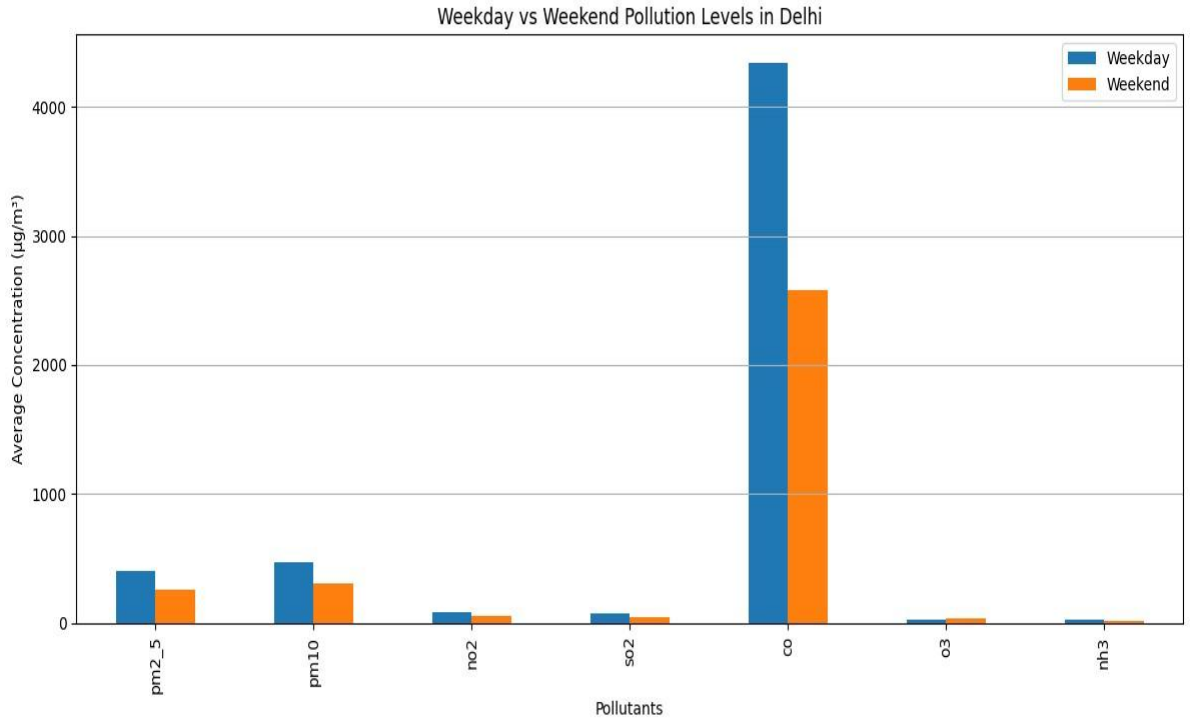
Weekday vs Weekend Pollution Analysis

Step 1: Pollution data is divided into weekdays and weekends.

Step 2: Average pollutant levels are calculated for both groups.

Step 3: A bar graph compares weekday and weekend pollution.

Step 4: Higher weekday pollution shows the effect of traffic and industries.



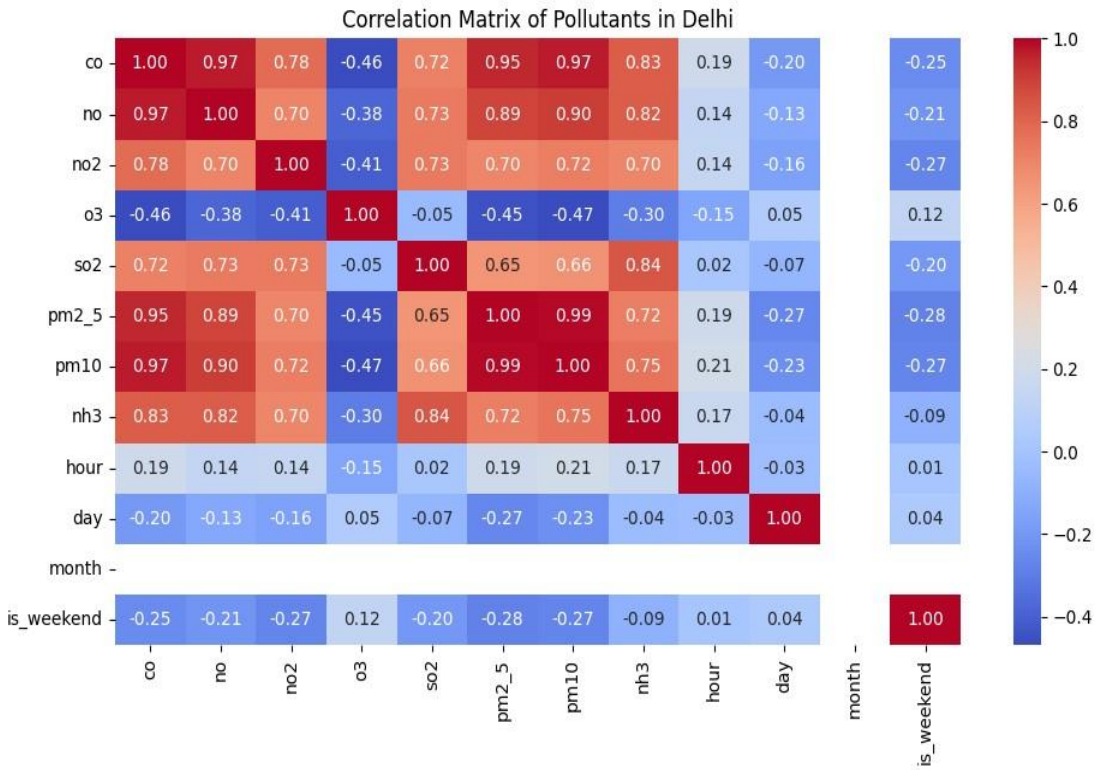
Correlation Analysis of Pollutants

Step 1: Correlation values between pollutants are calculated.

Step 2: Heatmap colors show strength of relationships.

Step 3: PM2.5 and PM10 are strongly related.

Step 4: Ozone shows different behavior from other pollutants.



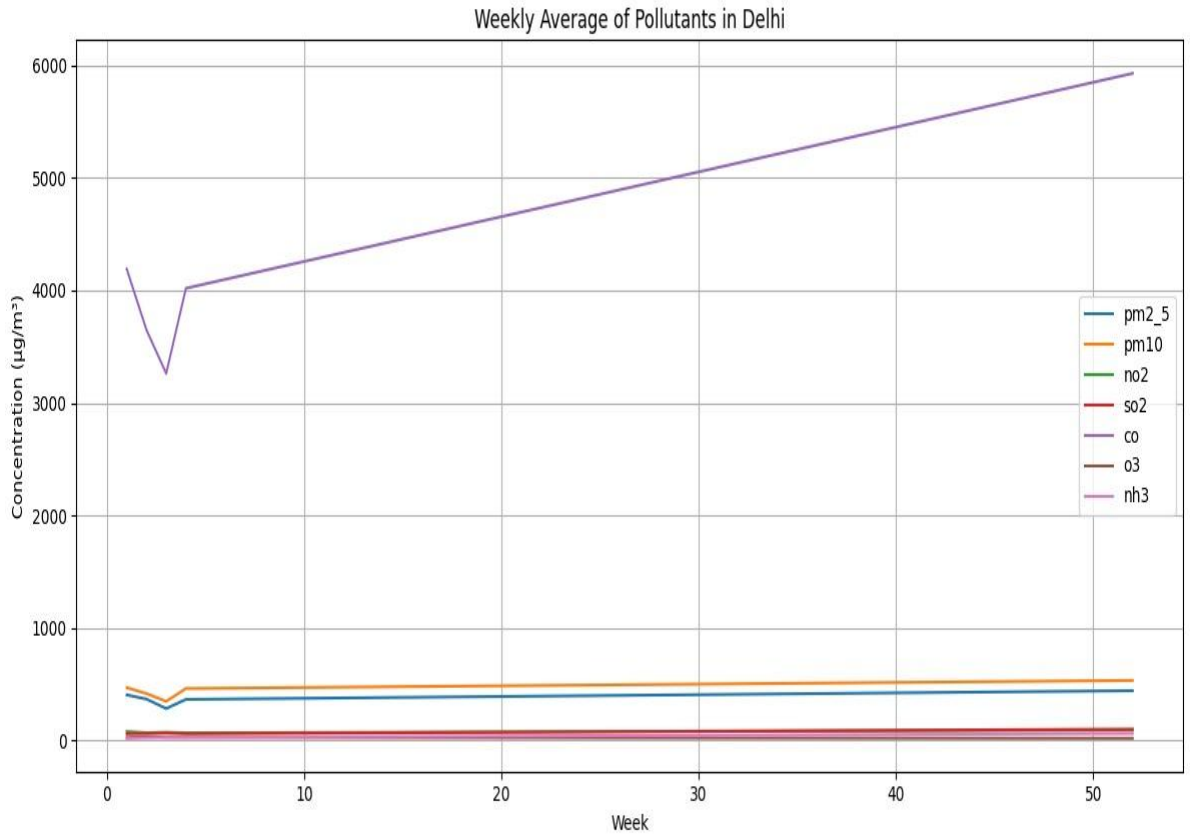
Weekly Trend of Pollutants

Step 1: Data is grouped by week.

Step 2: Weekly average of pollutants is calculated.

Step 3: Line graph shows pollution trend over time.

Step 4: Increasing trend shows worsening air quality.



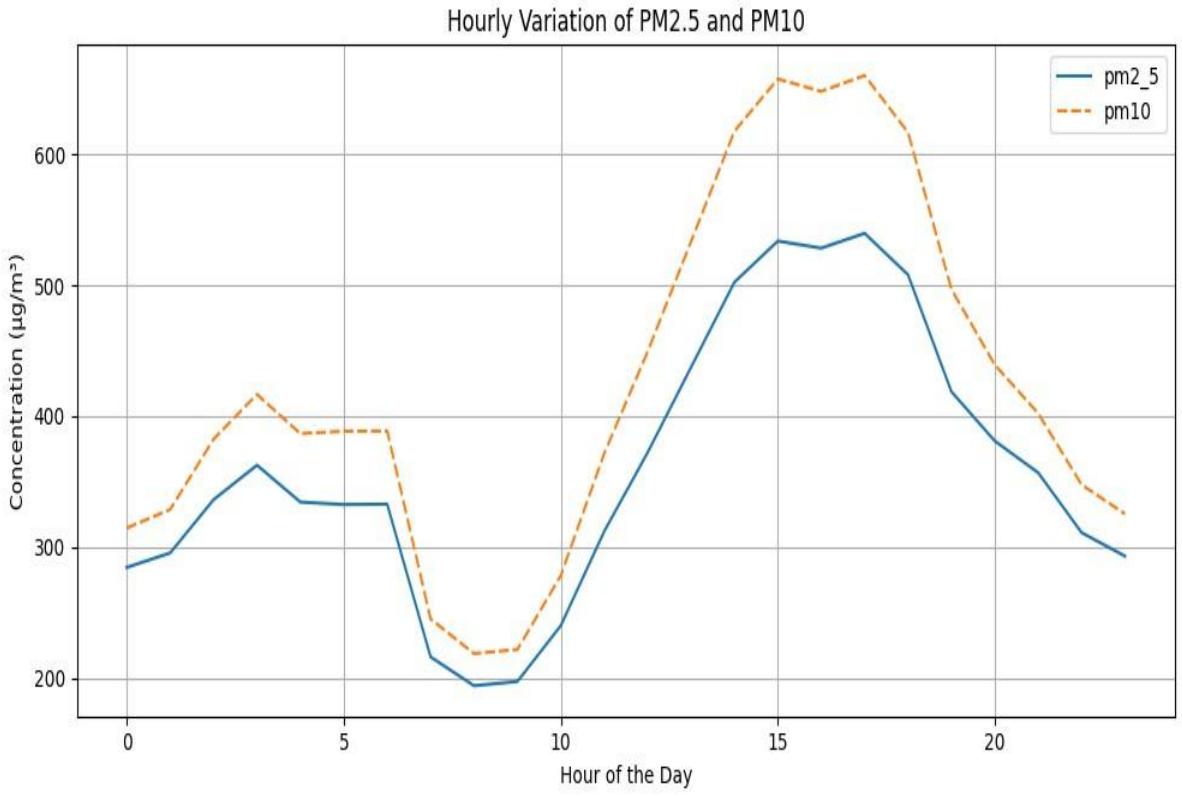
Hourly Variation of PM2.5 and PM10

Step 1: Pollution data is grouped by hour.

Step 2: Hourly averages of PM2.5 and PM10 are calculated.

Step 3: Line graph shows hourly pollution changes.

Step 4: Highest pollution occurs in afternoon and evening.



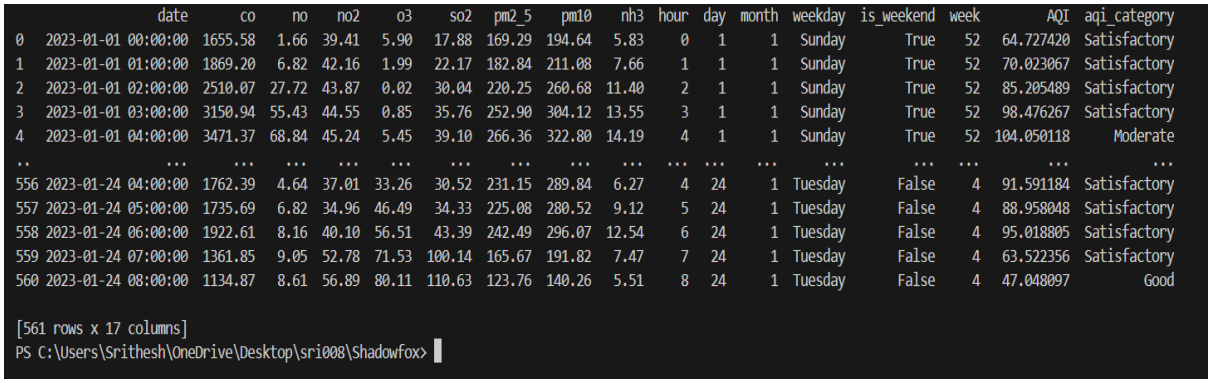
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv("delhiaqi.csv")
df.head()

print(df.info())
print(df.describe())
df.isnull().sum
df['date'] = pd.to_datetime(df['date'],format = "%d-%m-%Y %H:%M")
df['hour'] = df['date'].dt.hour
df['day'] = df['date'].dt.day
df['month'] = df['date'].dt.month
df['weekday'] = df['date'].dt.day_name()
df['is_weekend'] = df['weekday'].isin(['Saturday', 'Sunday'])

# Compute average pollutant levels for weekdays vs weekends
pollutants = ['pm2_5', 'pm10', 'no2', 'so2', 'co', 'o3', 'nh3']
weekday_vs_weekend_avg = df.groupby('is_weekend')[pollutants].mean().T
weekday_vs_weekend_avg.columns = ['Weekday', 'Weekend']

# Plot comparison
plt.figure(figsize=(10, 6))
weekday_vs_weekend_avg.plot(kind='bar', figsize=(12, 6))
plt.title("Weekday vs Weekend Pollution Levels in Delhi")
plt.ylabel("Average Concentration (µg/m³)")
plt.xlabel("Pollutants")
plt.grid(axis='y')
plt.tight_layout()
plt.savefig("Weekday vs Weekend Pollution")
plt.show()


correlation_matrix = df.drop(columns='date').corr(numeric_only=True)
print(correlation_matrix)
```

```python
#correlation heatmap
plt.figure(figsize=(10, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Matrix of Pollutants in Delhi")
plt.tight_layout()
plt.savefig("aqi_correlation_heatmap.png")
plt.show()


df['week'] = df['date'].dt.isocalendar().week
weekly_avg = df.groupby('week')[['pm2_5', 'pm10', 'no2', 'so2', 'co', 'o3', 'nh3']].mean()
print(weekly_avg)


plt.figure(figsize=(12, 6))
for pollutant in weekly_avg.columns:
    plt.plot(weekly_avg.index, weekly_avg[pollutant], label=pollutant)
plt.title("Weekly Average of Pollutants in Delhi")
plt.xlabel("Week")
plt.ylabel("Concentration (µg/m³)")
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.savefig("weekly_avg_pollutants.png")
plt.show()


hourly_avg = df.groupby('hour')[['pm2_5', 'pm10']].mean()
print(hourly_avg)

plt.figure(figsize=(10, 5))
sns.lineplot(data=hourly_avg)
plt.title("Hourly Variation of PM2.5 and PM10")
plt.xlabel("Hour of the Day")
plt.ylabel("Concentration (µg/m³)")
plt.grid(True)
plt.tight_layout()
plt.savefig("hourly_pm_variation.png")
plt.show()

```

```python
# Identify peak pollution hours for PM2.5 and PM10
peak_pm2_5_hour = hourly_avg['pm2_5'].idxmax()
peak_pm10_hour = hourly_avg['pm10'].idxmax()


print(f"Peak PM2.5 concentration hour: {peak_pm2_5_hour}:00")
print(f"Peak PM10 concentration hour: {peak_pm10_hour}:00")


df["AQI"]=(
    (df["pm2_5"]/df['pm2_5'].max())*0.6+(df["pm10"]/df['pm10'].max())*0.4
)*500


def aqi_category(a):
    if a <=50:
        return "Good"
    elif a<=100:
        return "Satisfactory"
    elif a<=200:
        return "Moderate"
    elif a<=300:
        return "Poor"
    elif a<=400:
        return "Very poor"
df["aqi_category"]=df["AQI"].apply(aqi_category)

print(df)
```

```
RangeIndex: 561 entries, 0 to 560
Data columns (total 9 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   date    561 non-null    object
 1   co      561 non-null    float64
 2   no      561 non-null    float64
 3   no2     561 non-null    float64
 4   o3      561 non-null    float64
 5   so2     561 non-null    float64
 6   pm2_5   561 non-null    float64
 7   pm10    561 non-null    float64
 8   nh3     561 non-null    float64
dtypes: float64(8), object(1)
memory usage: 39.6+ KB
None
                co          no         no2          o3         so2       pm2_5        pm10         nh3
count   561.000000  561.000000  561.000000  561.000000  561.000000  561.000000  561.000000  561.000000
mean   3814.942210   51.181979   75.292496   30.141943   64.655936  358.256364  420.988414   26.425062
std    3227.744681   83.904476   42.473791   39.979405   61.073080  227.359117  271.287026   36.563094
min     654.220000    0.000000   13.370000    0.000000    5.250000   60.100000   69.080000    0.630000
25%    1708.980000    3.380000   44.550000    0.070000   28.130000  204.450000  240.900000    8.230000
50%    2590.180000   13.300000   63.750000   11.800000   47.210000  301.170000  340.900000   14.820000
75%    4432.680000   59.010000   97.330000   47.210000   77.250000  416.650000  482.570000   26.350000
max   16876.220000  425.580000  263.210000  164.510000  511.170000 1310.200000 1499.270000  267.510000
                co          no         no2          o3         so2       pm2_5        pm10         nh3        hour       day  month  is_weekend
co        1.000000    0.969740    0.776402   -0.463082    0.716831    0.953083    0.966801    0.826299    0.191138 -0.196306    NaN   -0.249787
no        0.969740    1.000000    0.702201   -0.377813    0.734503    0.888810    0.903339    0.823638    0.135450 -0.131474    NaN   -0.214414
no2       0.776402    0.702201    1.000000   -0.407177    0.734961    0.698696    0.720050    0.700254    0.139751 -0.162382    NaN   -0.272112
o3       -0.463082   -0.377813   -0.407177    1.000000   -0.049158   -0.450458   -0.468477   -0.299663   -0.147007  0.051575    NaN    0.123092
so2       0.716831    0.734503    0.734961   -0.049158    1.000000    0.648996    0.658325    0.843635    0.016675 -0.070568    NaN   -0.202575
pm2_5     0.953083    0.888810    0.698696   -0.450458    0.648996    1.000000    0.994088    0.720303    0.191102 -0.265522    NaN   -0.284301
pm10      0.966801    0.903339    0.720050   -0.468477    0.658325    0.994088    1.000000    0.754468    0.206265 -0.229061    NaN   -0.272881
nh3       0.826299    0.823638    0.700254   -0.299663    0.843635    0.720303    0.754468    1.000000    0.167693 -0.043720    NaN   -0.092112
hour      0.191138    0.135450    0.139751   -0.147007    0.016675    0.191102    0.206265    0.167693    1.000000 -0.030332    NaN    0.011338
day      -0.196306   -0.131474   -0.162382    0.051575   -0.070568   -0.265522   -0.229061   -0.043720   -0.030332  1.000000    NaN    0.036701
month          NaN         NaN         NaN         NaN         NaN         NaN         NaN         NaN         NaN       NaN    NaN         NaN
is_weekend -0.249787   -0.214414   -0.272112    0.123092   -0.202575   -0.284301   -0.272881   -0.092112    0.011338  0.036701    NaN    1.000000
           pm2_5        pm10        no2         so2          co         o3        nh3
week
1     406.534762  471.023631  80.098512   60.376667  4190.683333  27.353452  21.519286
2     369.237321  418.176310  70.252083   63.132440  3649.632381  35.886012  24.094167
3     285.133512  349.163631  74.195179   65.884226  3262.321369  28.827024  27.900060
4     366.518788  463.286364  69.022121   60.595455  4019.390606  28.226667  28.800303
```

```
           pm2_5        pm10        no2         so2          co         o3        nh3
week
1     406.534762  471.023631  80.098512   60.376667  4190.683333  27.353452  21.519286
2     369.237321  418.176310  70.252083   63.132440  3649.632381  35.886012  24.094167
3     285.133512  349.163631  74.195179   65.884226  3262.321369  28.827024  27.900060
4     366.518788  463.286364  69.022121   60.595455  4019.390606  28.226667  28.800303
52    443.940000  535.040417  93.236250  102.260417  5929.152500  21.290833  63.490833
           pm2_5        pm10
hour
0     284.571250  314.638750
1     295.627083  328.781250
2     336.278750  382.432917
3     362.630417  416.650000
4     334.481667  386.673333
5     332.568750  388.529583
6     332.872500  388.692083
7     216.307917  245.275417
8     194.313750  218.862083
9     197.552609  221.964783
10    240.183478  278.340870
11    312.113913  371.545217
12    372.226087  449.208261
13    437.294348  533.203043
14    502.223043  617.469130
15    533.620870  657.307826
16    528.229565  647.846522
17    539.517826  659.962174
18    507.915652  616.495217
19    418.675217  496.733043
20    381.065652  439.151739
21    356.821739  402.109130
22    311.110870  347.679130
23    293.195652  325.403478
Peak PM2.5 concentration hour: 17:00
Peak PM10 concentration hour: 17:00
```



	date	co	no	no2	o3	so2	pm2_5	pm10	nh3	hour	day	month	weekday	is_weekend	week	AQI	aqi_category
0	2023-01-01 00:00:00	1655.58	1.66	39.41	5.90	17.88	169.29	194.64	5.83	0	1	1	Sunday	True	52	64.727420	Satisfactory
1	2023-01-01 01:00:00	1869.20	6.82	42.16	1.99	22.17	182.84	211.08	7.66	1	1	1	Sunday	True	52	70.023067	Satisfactory
2	2023-01-01 02:00:00	2510.07	27.72	43.87	0.02	30.04	220.25	260.68	11.40	2	1	1	Sunday	True	52	85.205489	Satisfactory
3	2023-01-01 03:00:00	3150.94	55.43	44.55	0.85	35.76	252.90	304.12	13.55	3	1	1	Sunday	True	52	98.476267	Satisfactory
4	2023-01-01 04:00:00	3471.37	68.84	45.24	5.45	39.10	266.36	322.80	14.19	4	1	1	Sunday	True	52	104.050118	Moderate
..	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
556	2023-01-24 04:00:00	1762.39	4.64	37.01	33.26	30.52	231.15	289.84	6.27	4	24	1	Tuesday	False	4	91.591184	Satisfactory
557	2023-01-24 05:00:00	1735.69	6.82	34.96	46.49	34.33	225.08	280.52	9.12	5	24	1	Tuesday	False	4	88.958048	Satisfactory
558	2023-01-24 06:00:00	1922.61	8.16	40.10	56.51	43.39	242.49	296.07	12.54	6	24	1	Tuesday	False	4	95.018805	Satisfactory
559	2023-01-24 07:00:00	1361.85	9.05	52.78	71.53	100.14	165.67	191.82	7.47	7	24	1	Tuesday	False	4	63.522356	Satisfactory
560	2023-01-24 08:00:00	1134.87	8.61	56.89	80.11	110.63	123.76	140.26	5.51	8	24	1	Tuesday	False	4	47.048097	Good

[561 rows x 17 columns]
PS C:\Users\Srithesh\OneDrive\Desktop\sri008\Shadowfox>

Conclusion

The analysis shows Delhi’s air pollution is higher on weekdays and peak hours. Controlling vehicle emissions and industrial pollution can improve AQI.