Import dataset

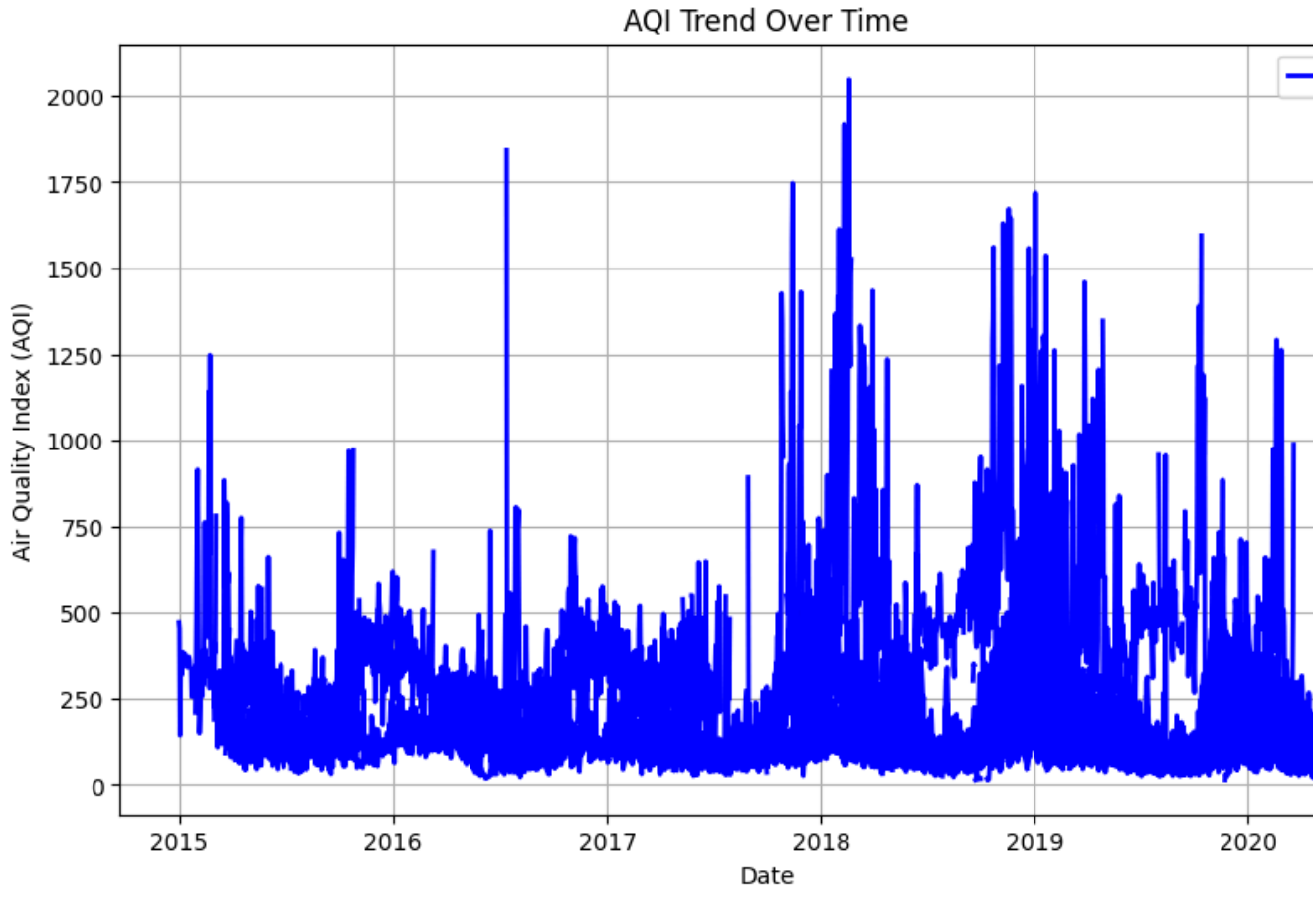
```
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
df = pd.read_csv("/content/city_day.csv")

# Convert 'Date' column to datetime for easier plotting
df['Date'] = pd.to_datetime(df['Date'])



# Explore the structure and content of the dataset
print(df.info())
print(df.describe())
print(df.head())
```

```
 13  Xylene      11422 non-null  float64
 14  AQI         24850 non-null  float64
 15  AQI_Bucket  24850 non-null  object
dtypes: datetime64[ns](1), float64(13), object(2)
memory usage: 3.6+ MB
None
                                Date         PM2.5          PM10  \
count                          29531  24933.000000  18391.000000
mean   2018-05-14 05:40:15.807118080     67.450578    118.127103
min              2015-01-01 00:00:00      0.040000      0.010000
25%              2017-04-16 00:00:00     28.820000     56.255000
50%              2018-08-05 00:00:00     48.570000     95.680000
75%              2019-09-03 00:00:00     80.590000    149.745000
max              2020-07-01 00:00:00    949.990000   1000.000000
std                              NaN     64.661449     90.605110

                 NO           NO2           NOx           NH3            CO  \
count  25949.000000  25946.000000  25346.000000  19203.000000  27472.000000
mean      17.574730     28.560659     32.309123     23.483476      2.248598
min        0.020000      0.010000      0.000000      0.010000      0.000000
25%        5.630000     11.750000     12.820000      8.580000      0.510000
50%        9.890000     21.690000     23.520000     15.850000      0.890000
75%       19.950000     37.620000     40.127500     30.020000      1.450000
max      390.680000    362.210000    467.630000    352.890000    175.810000
std       22.785846     24.474746     31.646011     25.684275      6.962884

                SO2            O3       Benzene       Toluene        Xylene  \
count  25677.000000  25509.000000  23908.000000  21490.000000  11422.000000
mean      14.531977     34.491430      3.280840      8.700972      3.070128
```

```
2  Ahmedabad 2015-01-03    NaN   NaN  17.40  19.30  29.70  NaN  17.40  29.07
3  Ahmedabad 2015-01-04    NaN   NaN   1.70  18.48  17.97  NaN   1.70  18.59
4  Ahmedabad 2015-01-05    NaN   NaN  22.10  21.42  37.76  NaN  22.10  39.33

       O3  Benzene  Toluene  Xylene  AQI AQI_Bucket
0  133.36     0.00     0.02    0.00  NaN        NaN
1   34.06     3.68     5.50    3.77  NaN        NaN
2   30.70     6.80    16.40    2.25  NaN        NaN
3   36.08     4.43    10.14    1.00  NaN        NaN
4   39.31     7.01    18.89    2.78  NaN        NaN
```

Line Plot for Overall AQI Trend Over Time

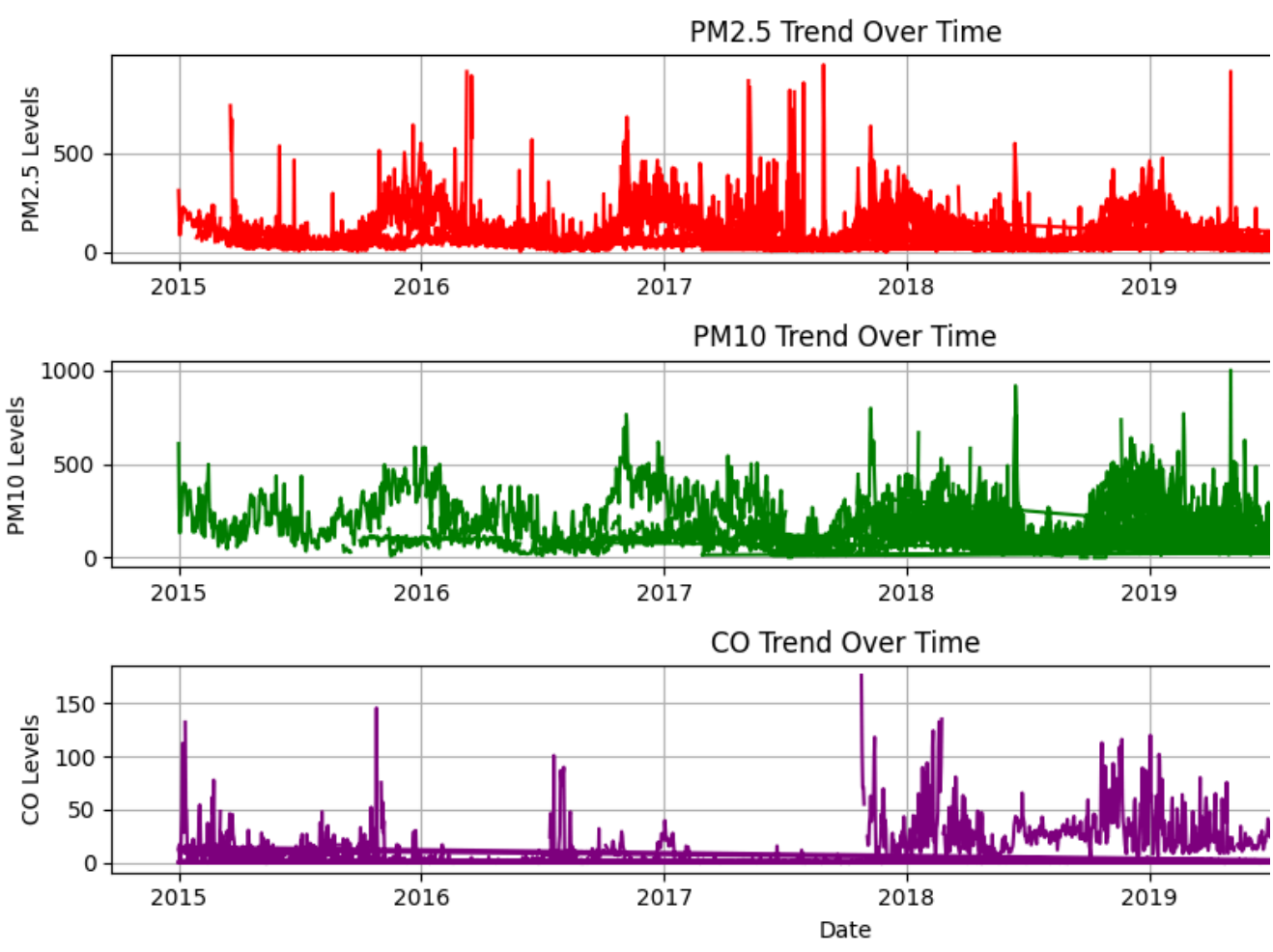
```
plt.figure(figsize=(10,6))
plt.plot(df['Date'], df['AQI'], label='AQI', color='blue', linewidth=2)
plt.title('AQI Trend Over Time')
plt.xlabel('Date')
plt.ylabel('Air Quality Index (AQI)')
plt.legend()
plt.grid(True)
plt.show()
```



Plot Individual Pollutant Trends Over Time

```
plt.figure(figsize=(10,6))

# Plot for PM2.5
plt.subplot(3, 1, 1)
plt.plot(df['Date'], df['PM2.5'], color='red', label='PM2.5')
```

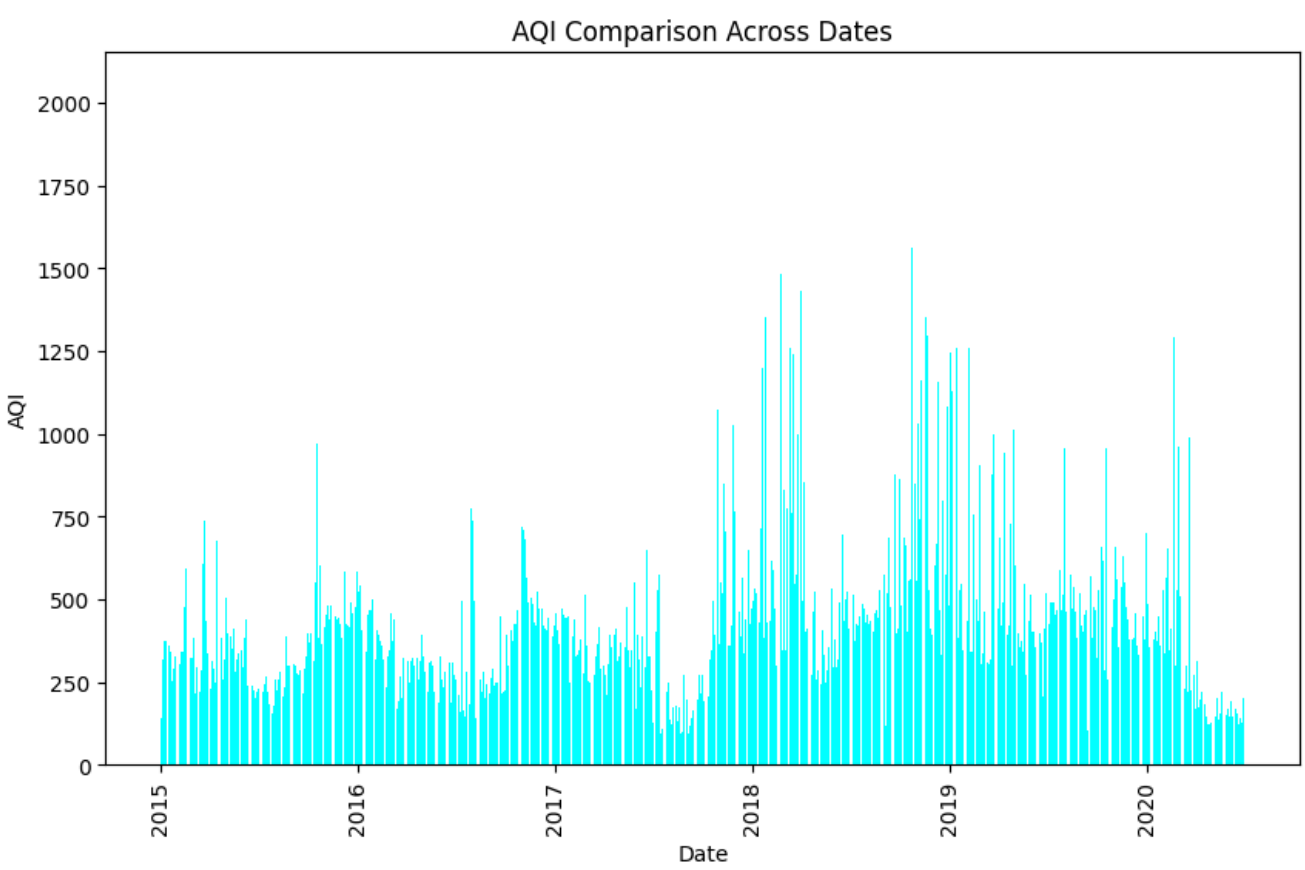
```
plt.title('PM2.5 Trend Over Time')
plt.ylabel('PM2.5 Levels')
plt.grid(True)

# Plot for PM10
plt.subplot(3, 1, 2)
plt.plot(df['Date'], df['PM10'], color='green', label='PM10')
plt.title('PM10 Trend Over Time')
plt.ylabel('PM10 Levels')
plt.grid(True)

# Plot for CO
plt.subplot(3, 1, 3)
plt.plot(df['Date'], df['CO'], color='purple', label='CO')
plt.title('CO Trend Over Time')
plt.ylabel('CO Levels')
plt.xlabel('Date')
plt.grid(True)

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



Bar Plot for AQI Comparison Across Dates

```
plt.figure(figsize=(10,6))
plt.bar(df['Date'], df['AQI'], color='cyan')
plt.title('AQI Comparison Across Dates')
plt.xlabel('Date')
plt.ylabel('AQI')
```

```
plt.xticks(rotation=90)
plt.show()
```

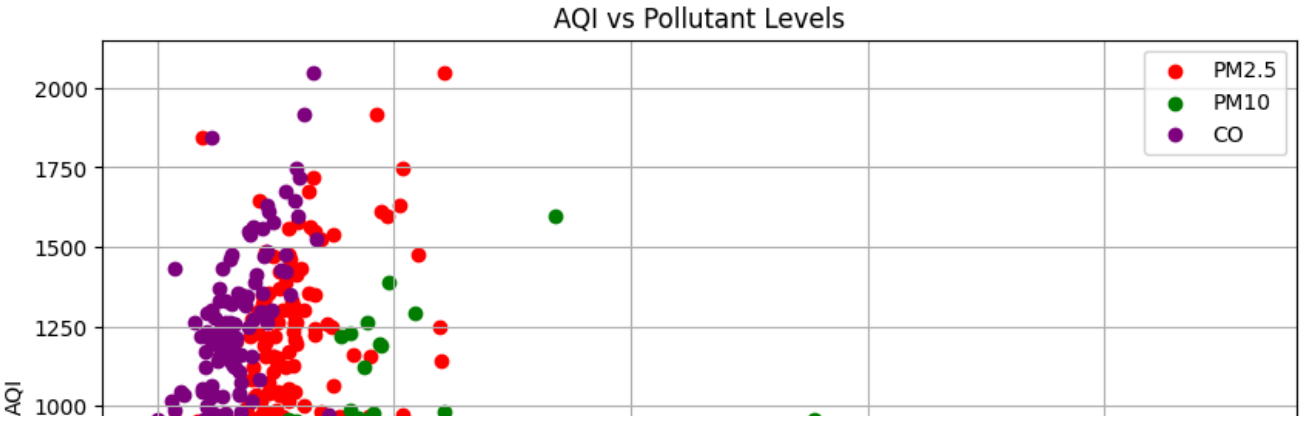


Box Plot for AQI Distribution

```
plt.figure(figsize=(8,6))
plt.boxplot(df['AQI'])
plt.title('AQI Distribution')
plt.ylabel('AQI')
plt.grid(True)
plt.show()
```

Scatter Plot for AQI vs. Pollutant Levels

```
plt.figure(figsize=(10,6))

# Scatter plot for AQI vs PM2.5
plt.scatter(df['PM2.5'], df['AQI'], color='red', label='PM2.5')
plt.scatter(df['PM10'], df['AQI'], color='green', label='PM10')
plt.scatter(df['CO'], df['AQI'], color='purple', label='CO')

plt.title('AQI vs Pollutant Levels')
plt.xlabel('Pollutant Levels')
plt.ylabel('AQI')
plt.legend()
plt.grid(True)
plt.show()
```


AQI vs Pollutant Levels
PM2.5
PM10
CO
AQI
2000
1750
1500
1250
1000