

jupyter SmogSense_Project Last Checkpoint: 7 minutes ago

To exit full screen, press and hold **esc**

File Edit View Run Kernel Settings Help Trusted JupyterLab Python 3 (ipykernel)

```
[1]:  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
from sklearn.model_selection import train_test_split  
from sklearn.linear_model import LinearRegression  
from sklearn.metrics import mean_absolute_error, r2_score
```

```
[5]:  
FileNotFoundError: [Errno 2] No such file or directory: 'DATASET/DATASET INDIA.csv'  
[5]:  
data = pd.read_csv('DATASET INDIA/city_day.csv')
```

```
import os  
  
# Check if the file exists  
print(os.path.isfile('/Users/ghost/Desktop/TTDPROJECT/DATASET INDIA/city_
```

False

```
pip install pandas
[5] ✓ 0.7s
...
Requirement already satisfied: pandas in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (2.3.3)
Requirement already satisfied: numpy>=1.22.4 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from pandas) (2.2.6)
Requirement already satisfied: python-dateutil>=2.8.2 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from pandas) (2025.2)
Requirement already satisfied: six>=1.5 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
Note: you may need to restart the kernel to use updated packages.
```

SmogSense_Project.ipynb

```
Users > ghost > Desktop > TTDPROJECT > SmogSense_Project > SmogSense_Project.ipynb > data = pd.read_csv('../SmogSense_Project/DATASET INDIA/city_day.csv') # Correct the relative path
+ Code + Markdown | ▶ Run All ⚙ Restart ⌛ Clear All Outputs ⌘ Go To | 📈 Jupyter Variables | 🔍 Outline ... Python 3.10.4
```

NameError
Cell In[1], line 1
----> 1 data = pd.read_csv('DATASET INDIA/city_day.csv') # Correct relative path

```
NameError: name 'pd' is not defined
```

```
import pandas as pd # For working with data
[1] ✓ 0.2s Python
```

```
pip install pandas
[5] ✓ 0.7s Python
```

...

```
Requirement already satisfied: pandas in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (2.3.3)
Requirement already satisfied: numpy>=1.22.4 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from pandas) (2.2.6)
Requirement already satisfied: python-dateutil>=2.8.2 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from pandas) (2025.2)
Requirement already satisfied: six>=1.5 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
Note: you may need to restart the kernel to use updated packages.
```

data = pd.read_csv('../SmogSense_Project/DATASET INDIA/city_day.csv') # Correct the relative path

[2] 0s Python

print(ta.head())
[3] 0.1s Python

Users > ghost > Desktop > TTDPROJECT > SmogSense_Project > SmogSense_Project.ipynb > import pandas as pd

+ Code + Markdown | ▶ Run All ⚡ Restart ⌂ Clear All Outputs | 📈 Jupyter Variables ⌂ Outline ...

Py

[4] ✓ 0.7s

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, r2_score
```

[1] ...

```
pip install pandas matplotlib seaborn scikit-learn
```

[1] ...

```
Requirement already satisfied: pandas in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (2.3.3)
Requirement already satisfied: matplotlib in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (3.8.4)
Requirement already satisfied: seaborn in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (0.13.2)
Requirement already satisfied: scikit-learn in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (1.7.2)
Requirement already satisfied: numpy>=1.22.4 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from pandas) (2.2.6)
Requirement already satisfied: python-dateutil>=2.8.2 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from pandas) (2.9.0.post1)
Requirement already satisfied: pytz>=2020.1 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from pandas) (2025.2)
Requirement already satisfied: contourpy>=1.0.1 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from matplotlib) (1.2.1)
Requirement already satisfied: cycler>=0.10 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from matplotlib) (4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from matplotlib) (1.4.5)
Requirement already satisfied: packaging>=20.0 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from matplotlib) (25.0)
Requirement already satisfied: pillow>=8 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from matplotlib) (10.3.0)
Requirement already satisfied: pysampling>=2.3.1 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from matplotlib) (3.1.2)
Requirement already satisfied: scipy>=1.8.0 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from scikit-learn) (1.15.3)
Requirement already satisfied: joblib>=1.2.0 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from scikit-learn) (1.5.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from scikit-learn) (3.6)
Requirement already satisfied: six>=1.5 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from python-dateutil>=2.8.2->pandas)
Note: you may need to restart the kernel to use updated packages
```

[2] ✓ 0.0s

```
data = pd.read_csv('../SmogSense_Project/DATASET INDIA/city_day.csv') # Correct the relative path
```

[5] ✓ 0.0s

```
data.head()
```

[5] ...

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene	Xylene	AQI	AQI_Bucket
0	Ahmedabad	2015-01-01	NaN	NaN	0.92	18.22	17.15	NaN	0.92	27.64	133.36	0.00	0.02	0.00	NaN	NaN
1	Ahmedabad	2015-01-02	NaN	NaN	0.97	15.69	16.46	NaN	0.97	24.55	34.06	3.68	5.50	3.77	NaN	NaN
2	Ahmedabad	2015-01-03	NaN	NaN	17.40	19.30	29.70	NaN	17.40	29.07	30.70	6.80	16.40	2.25	NaN	NaN
3	Ahmedabad	2015-01-04	NaN	NaN	1.70	18.48	17.97	NaN	1.70	18.59	36.08	4.43	10.14	1.00	NaN	NaN
4	Ahmedabad	2015-01-05	NaN	NaN	22.10	21.42	37.76	NaN	22.10	39.33	39.31	7.01	18.89	2.78	NaN	NaN

+ Code + Markdown | ▶ Run All ⚡ Restart ⌂ Clear All Outputs | 📁 Jupyter Variables ⌂ Outline ... Python 3.10.4

	Ahmedabad	2015-01-03	NaN	NaN	17.40	19.30	29.70	NaN	17.40	29.07	30.70	6.80	16.40	2.25	NaN	NaN
2	Ahmedabad	2015-01-04	NaN	NaN	1.70	18.48	17.97	NaN	1.70	18.59	36.08	4.43	10.14	1.00	NaN	NaN
3	Ahmedabad	2015-01-05	NaN	NaN	22.10	21.42	37.76	NaN	22.10	39.33	39.31	7.01	18.89	2.78	NaN	NaN

```
station_data = pd.read_csv('../SmogSense_Project/DATASET INDIA/station_hour.csv')
```

[] Python

```
... /var/folders/gh/f7q_g5x7ql8nbv_0365_wph0000gn/T/ipykernel_5243/1479690576.py:1: DtypeWarning: Columns (15) have mixed types. Specify dtype option on import or set low_
station_data = pd.read_csv('../SmogSense_Project/DATASET INDIA/station_hour.csv')
```

▶ station_data = pd.read_csv('../SmogSense_Project/DATASET INDIA/station_day.csv')

[] Python

+ Code + Markdown

```
station_data = pd.read_csv('../SmogSense_Project/DATASET INDIA/stations.csv')
```

[14] ✓ 0.0s Python

```
station_data = pd.read_csv('../SmogSense_Project/DATASET INDIA/city_hour.csv')
```

[15] ✓ 0.5s Python

▶ station_data = pd.read_csv('../SmogSense_Project/DATASET INDIA/city_day.csv')

[16] ✓ 0.0s Python

```
station_data = pd.read_csv('../SmogSense_Project/DATASET INDIA/city_hour.csv')

[15] ✓ 0.5s

station_data = pd.read_csv('../SmogSense_Project/DATASET INDIA/city_day.csv')

[16] ✓ 0.0s

▶ data.head()

[17] ✓ 0.0s

...
   City      Date  PM2.5  PM10    NO   NO2   NOx  NH3    CO   SO2    O3 Benzene  Toluene  Xylene   AQI  AQI_Bucket
0 Ahmedabad 2015-01-01    NaN    NaN  0.92  18.22  17.15  NaN  0.92  27.64  133.36    0.00    0.02    0.00    NaN    NaN
1 Ahmedabad 2015-01-02    NaN    NaN  0.97  15.69  16.46  NaN  0.97  24.55  34.06    3.68    5.50    3.77    NaN    NaN
2 Ahmedabad 2015-01-03    NaN    NaN  17.40  19.30  20.70  NaN  17.40  29.07  30.70    6.80   16.40    2.25    NaN    NaN
3 Ahmedabad 2015-01-04    NaN    NaN  1.70  18.48  17.97  NaN  1.70  18.59  36.08    4.43   10.14    1.00    NaN    NaN
4 Ahmedabad 2015-01-05    NaN    NaN  22.10  21.42  37.76  NaN  22.10  39.33  39.31    7.01   18.89    2.78    NaN    NaN
```

+ Code + Markdown

Cursor File Edit Selection View Go Run Terminal Window Help

Mon 8 Dec 11:22 AM

SmogSense_Project.ipynb

Users > ghost > Desktop > TTDPROJECT > SmogSense_Project > SmogSense_Project.ipynb > data.head()

+ Code + Markdown | ▶ Run All ⌂ Restart ⌂ Clear All Outputs ⌂ Go To | Jupyter Variables ⌂ Outline ...

Python 3.10.4

```
station_data = pd.read_csv('../SmogSense_Project/DATASET INDIA/city_hour.csv')
```

[15] ✓ 0.5s Python

```
station_data = pd.read_csv('../SmogSense_Project/DATASET INDIA/city_day.csv')
```

[16] ✓ 0.0s Python

▶ data.head()

[17] ✓ 0.0s Python

...

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene	Xylene	AQI	AQI_Bucket
0	Ahmedabad	2015-01-01	NaN	NaN	0.92	18.22	17.15	NaN	0.92	27.64	133.36	0.00	0.02	0.00	NaN	NaN
1	Ahmedabad	2015-01-02	NaN	NaN	0.97	15.69	16.46	NaN	0.97	24.55	34.06	3.68	5.50	3.77	NaN	NaN
2	Ahmedabad	2015-01-03	NaN	NaN	17.40	19.30	20.70	NaN	17.40	29.07	30.70	6.80	16.40	2.25	NaN	NaN
3	Ahmedabad	2015-01-04	NaN	NaN	1.70	18.48	17.97	NaN	1.70	18.59	36.08	4.43	10.14	1.00	NaN	NaN
4	Ahmedabad	2015-01-05	NaN	NaN	22.10	21.42	37.76	NaN	22.10	39.33	39.31	7.01	18.89	2.78	NaN	NaN

Problems 37 Output Debug Console Terminal Jupyter zsh + ×

```
pyenv shell 3.12.7
fatimaa@new-hostname ~ % pyenv shell 3.12.7
pyenv: shell integration not enabled. Run `pyenv init` for instructions.
fatimaa@new-hostname ~ % git status
fatal: not a git repository (or any of the parent directories): .git
fatimaa@new-hostname ~ %
```

- fatimaa@new-hostname ~ % nano .gitignore
- fatimaa@new-hostname ~ % git add .gitignore
- fatimaa@new-hostname ~ % git commit -m "Add .gitignore to exclude unnecessary files"
[master (root-commit) cf939ad] Add .gitignore to exclude unnecessary files
1 file changed, 47 insertions(+)
create mode 100644 .gitignore
- fatimaa@new-hostname ~ % █

Code | Markdown | Run All | Restart = Install an extension for Python...

```
# Check the structure and basic info about the data
print(data.shape)      # Number of rows and columns
print(data.info())      # Data types and non-null values
print(data.columns)     # List of column names
```

[4] ✓ 0.0s

```
(29531, 16)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 29531 entries, 0 to 29530
Data columns (total 16 columns):
 #   Column   Non-Null Count Dtype  
--- 
 0   City      29531 non-null  object  
 1   Date      29531 non-null  object  
 2   PM2.5    24933 non-null  float64 
 3   PM10     18391 non-null  float64 
 4   NO        25949 non-null  float64 
 5   NO2       25946 non-null  float64 
 6   NOx      25346 non-null  float64 
 7   NH3       19203 non-null  float64 
 8   CO        27472 non-null  float64 
 9   SO2       25677 non-null  float64 
 10  O3        25509 non-null  float64 
 11  Benzene   23908 non-null  float64 
 12  Toluene   21490 non-null  float64 
 13  Xylene    11422 non-null  float64 
 14  AQI       24850 non-null  float64 
 15  AQI_Bucket 24850 non-null  object  
dtypes: float64(13), object(3)
memory usage: 3.6+ MB
None
Index(['City', 'Date', 'PM2.5', 'PM10', 'NO', 'NO2', 'NOx', 'NH3', 'CO', 'SO2',
       'O3', 'Benzene', 'Toluene', 'Xylene', 'AQI', 'AQI_Bucket'],
      dtype='object')
```

Python

Code | Markdown | Run All | Restart = Python...

```
print(data.isnull().sum())
```

[5] ✓ 0.0s

```
City          0
Date          0
PM2.5        4598
PM10         11140
NO           3582
NO2          3585
NOx          4185
NH3          10328
CO            2059
SO2          3854
O3            4022
Benzene       5623
Toluene       8041
Xylene        18109
AQI           4681
AQI_Bucket    4681
dtype: int64
```

+ Code + Markdown

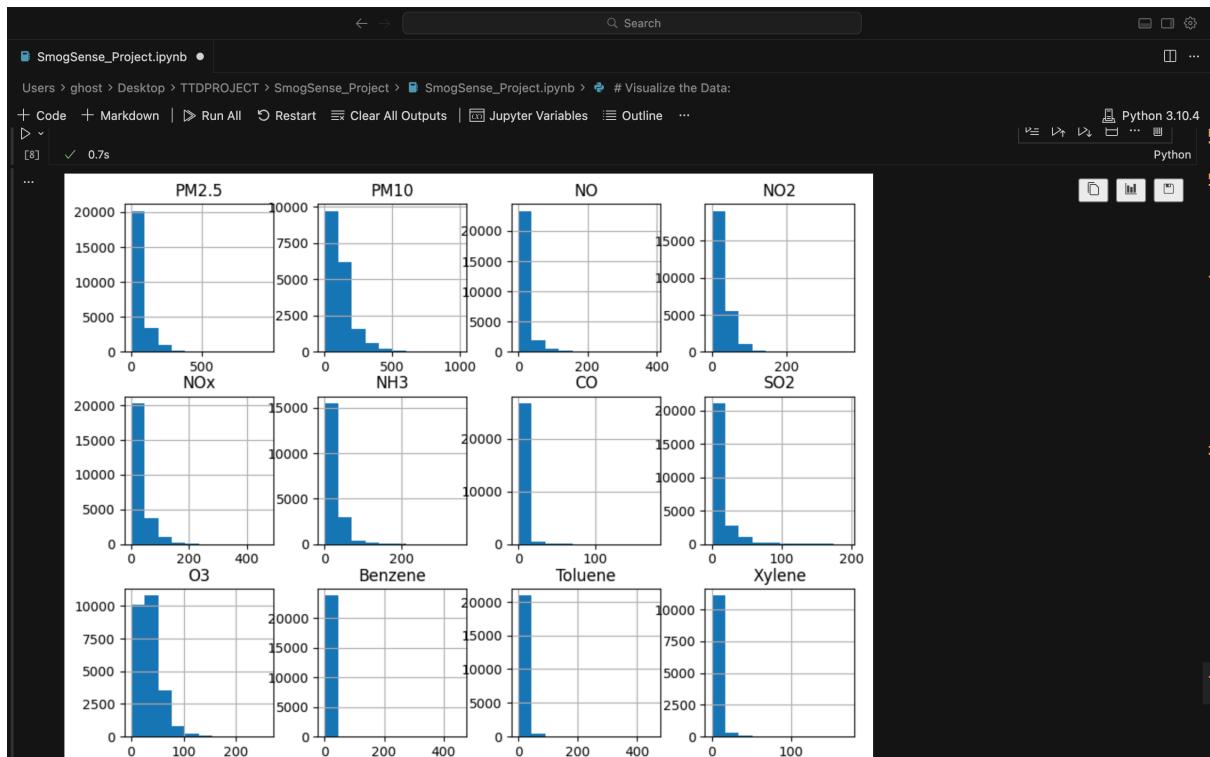
```

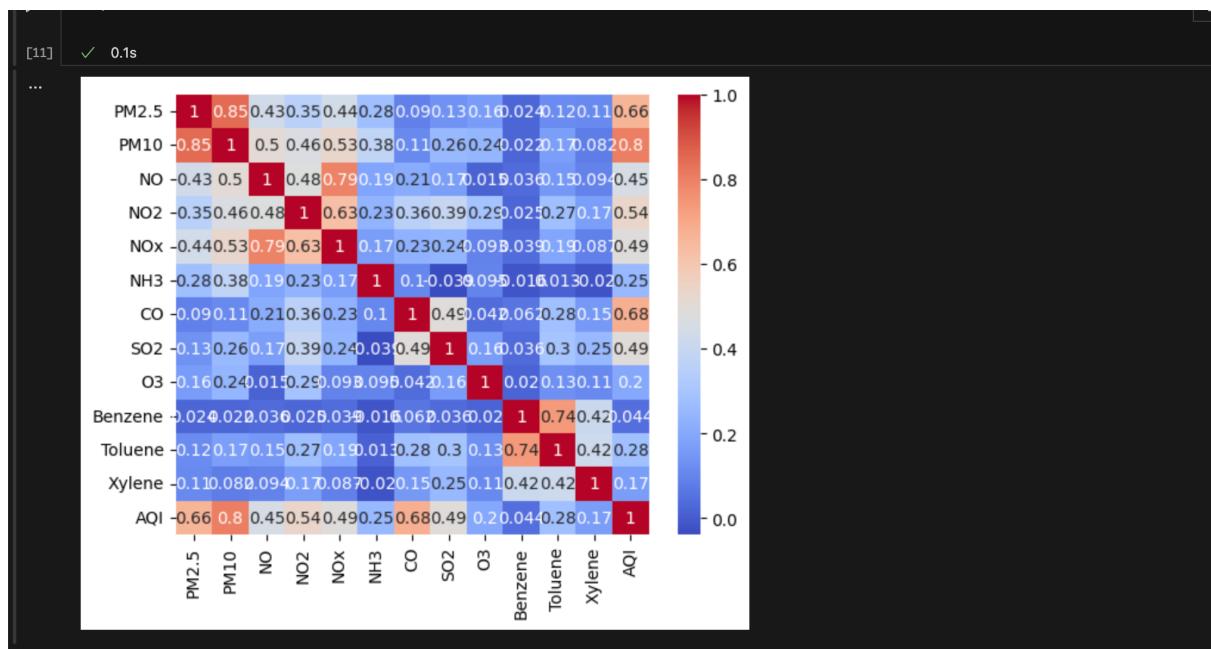
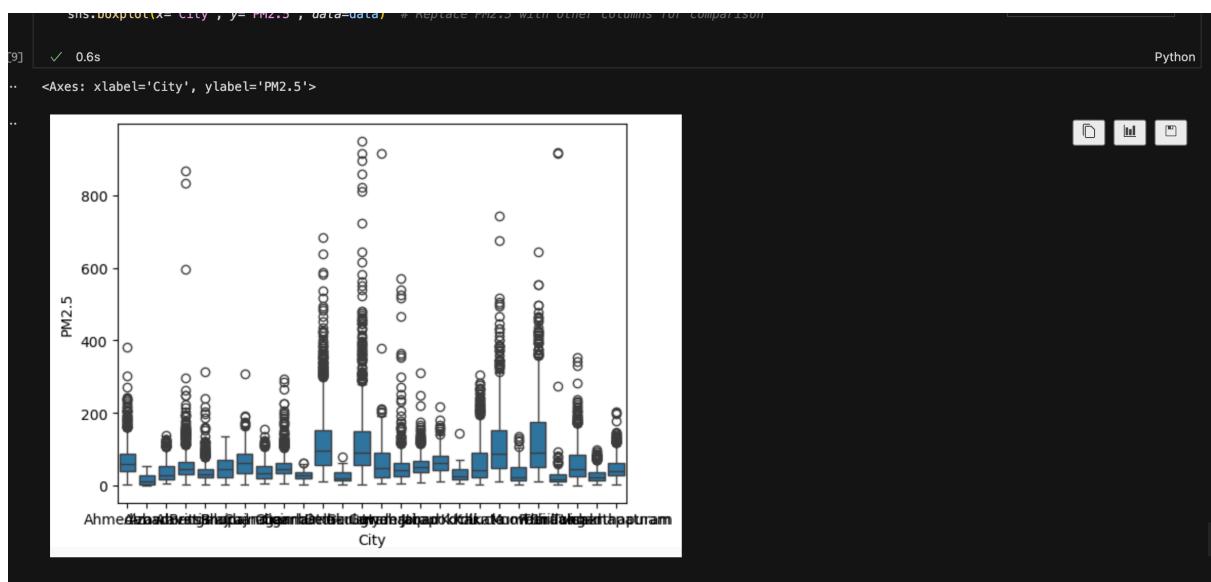
[6] ✓ 0.0s
...          PM2.5        PM10         NO        NO2        NOx \
count  24933.000000  18391.000000  25949.000000  25946.000000  25346.000000
mean   67.450578   118.127103   17.574730   28.560659   32.309123
std    64.661449   90.605110   22.785846   24.474746   31.646011
min    0.040000   0.010000   0.020000   0.010000   0.000000
25%   28.820000   56.255000   5.630000   11.750000   12.820000
50%   48.570000   95.680000   9.890000   21.690000   23.520000
75%   80.590000  149.745000  19.950000  37.620000  40.127500
max   949.990000  1000.000000  390.680000  362.210000  467.630000

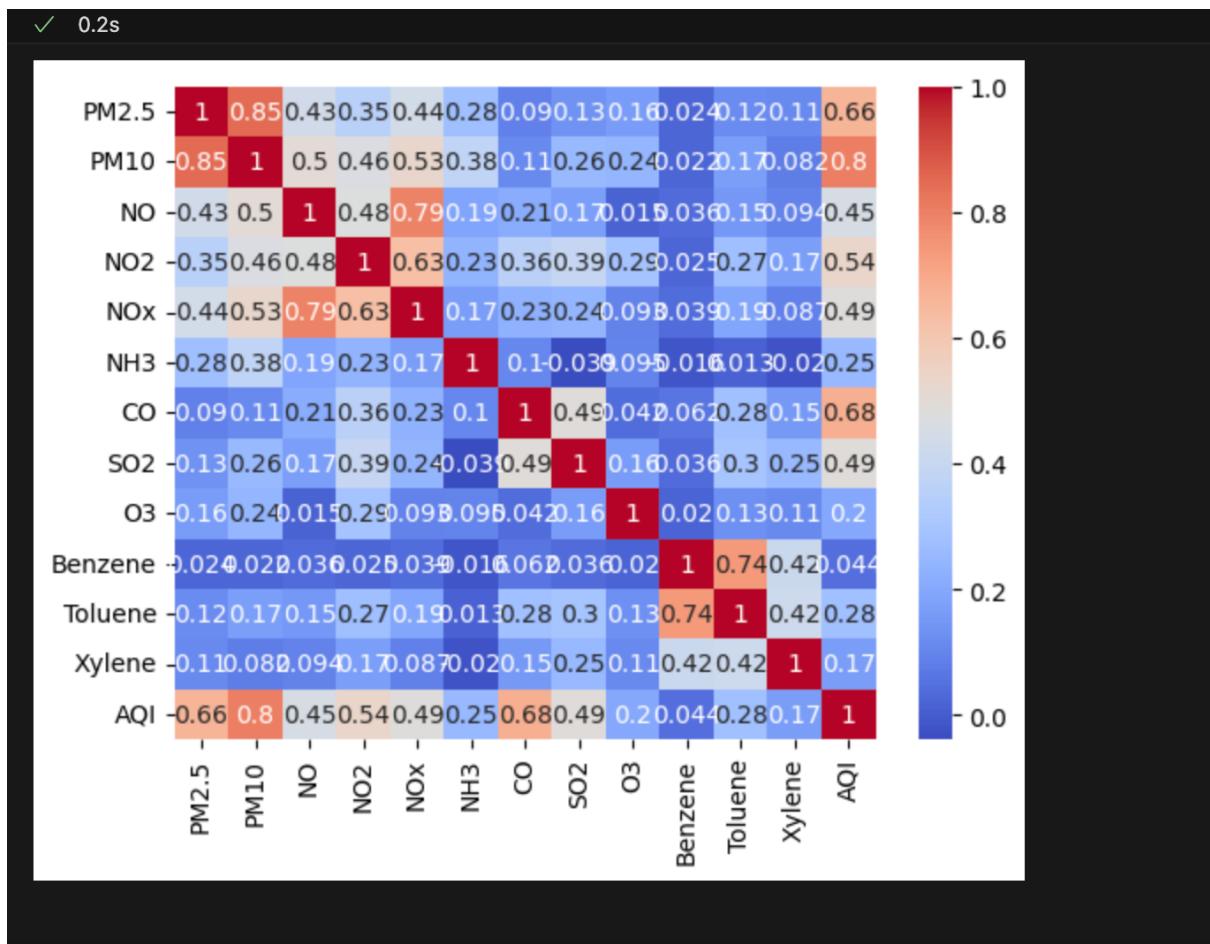
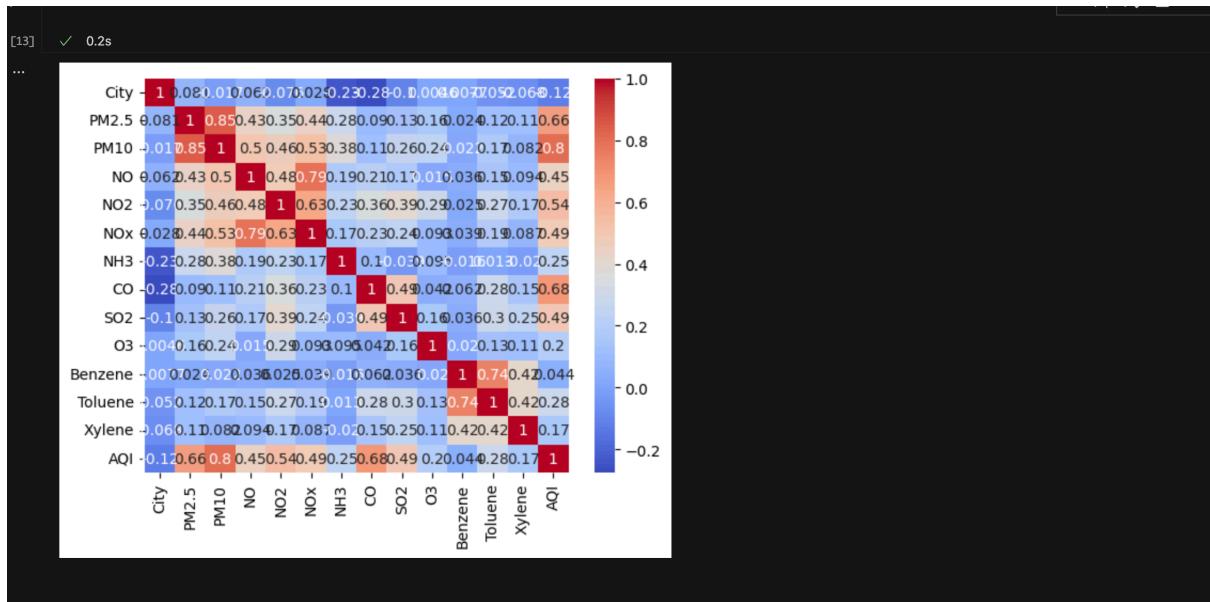
          NH3          CO         SO2         O3      Benzene \
count  19203.000000  27472.000000  25677.000000  25509.000000  23908.000000
mean   23.483476   2.248598   14.531977   34.491430   3.280840
std    25.684275   6.962884   18.133775   21.694928   15.811136
min    0.010000   0.000000   0.010000   0.010000   0.000000
25%   8.580000   0.510000   5.670000   18.860000   0.120000
50%   15.850000   0.890000   9.160000   30.840000   1.070000
75%   30.020000   1.450000  15.220000   45.570000   3.080000
max   352.890000  175.810000  193.860000  257.730000  455.030000

       Toluene        Xylene        AQI
count  21490.000000  11422.000000  24850.000000
mean   8.700972    3.070128   166.463581
std    19.969164   6.323247   140.696585
min    0.000000   0.000000   13.000000
25%   0.600000   0.140000   81.000000
50%   2.970000   0.980000  118.000000
75%   9.150000   3.350000  208.000000
max   454.850000  170.370000  2049.000000

```







```
▷ [18] print(data.dtypes)

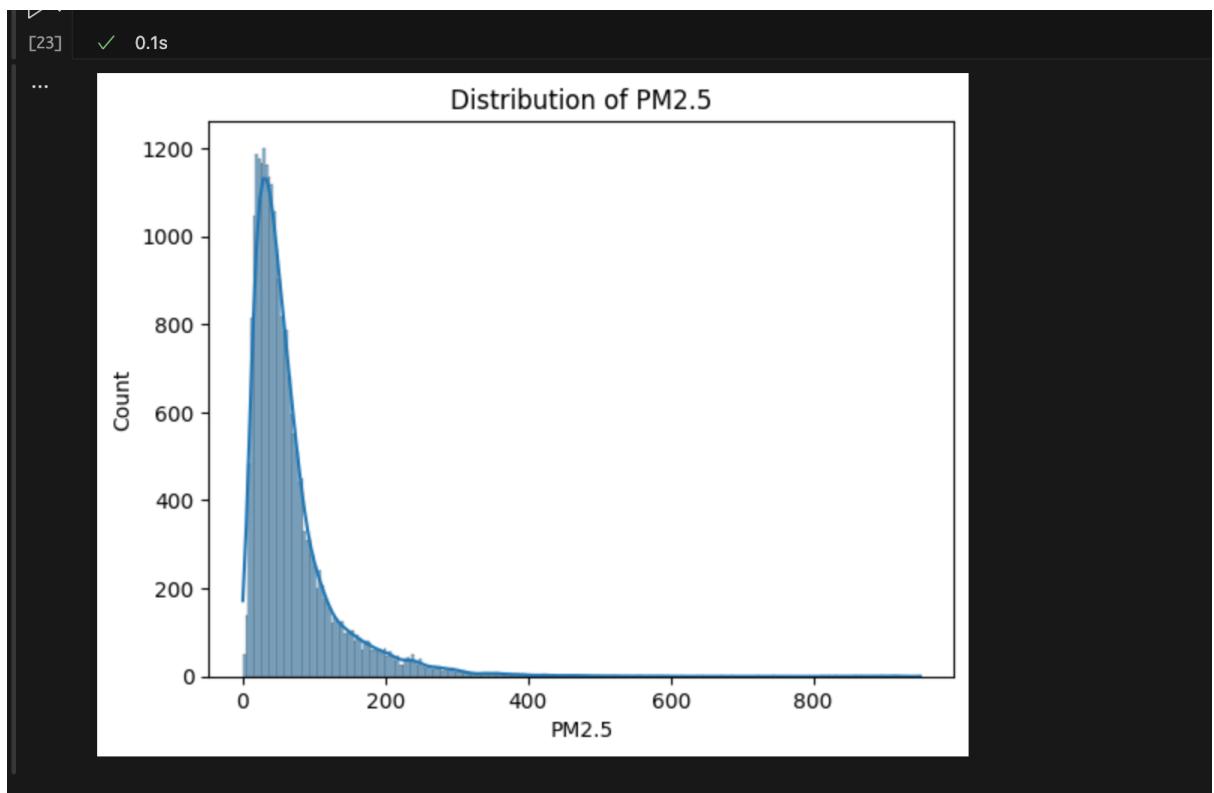
[18] ✓ 0.0s
...   City        int64
Date      object
PM2.5     float64
PM10      float64
NO        float64
NO2       float64
NOx       float64
NH3       float64
CO        float64
SO2       float64
O3        float64
Benzene    float64
Toluene    float64
Xylene     float64
AQI       float64
AQI_Bucket object
dtype: object
```

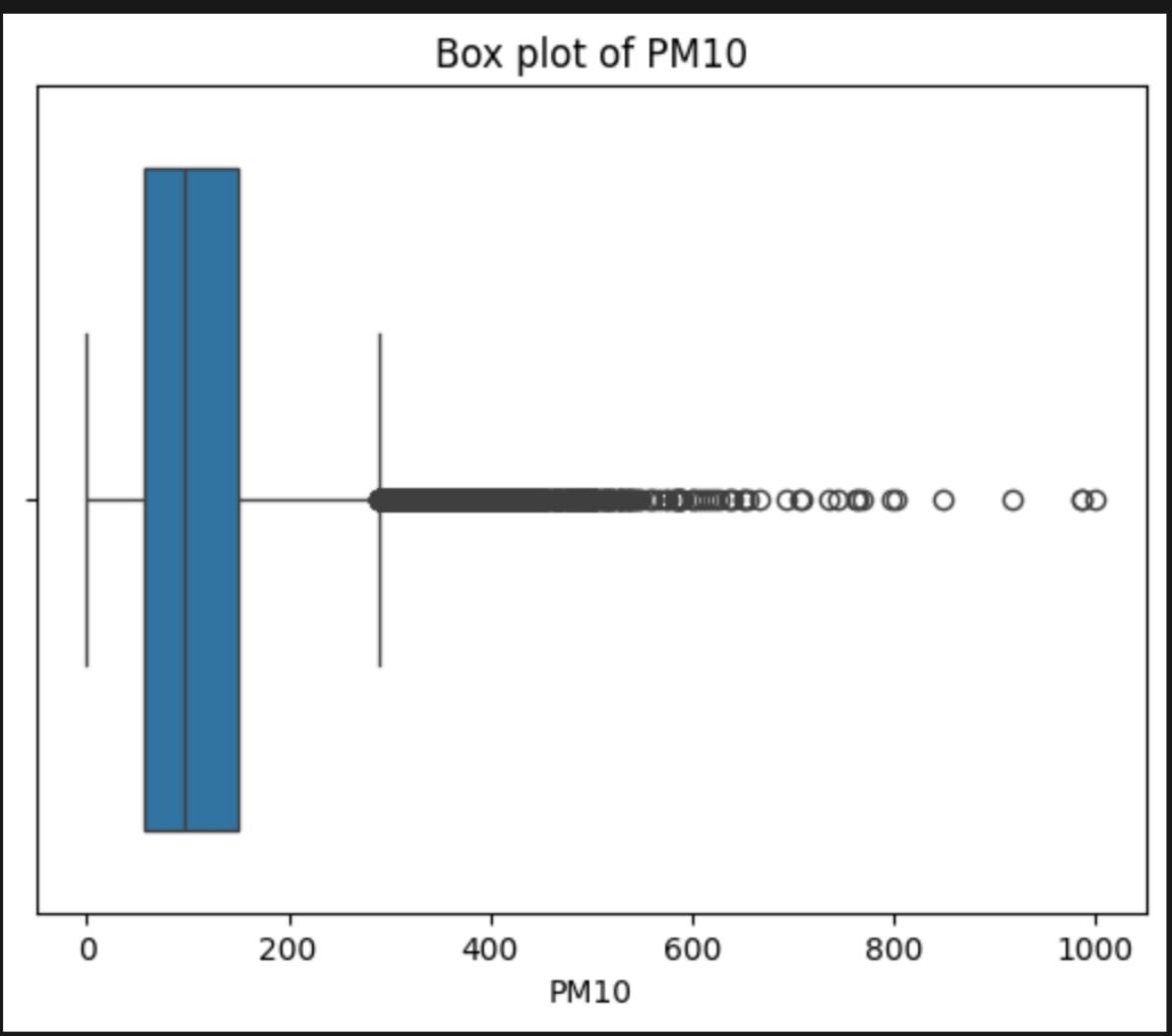
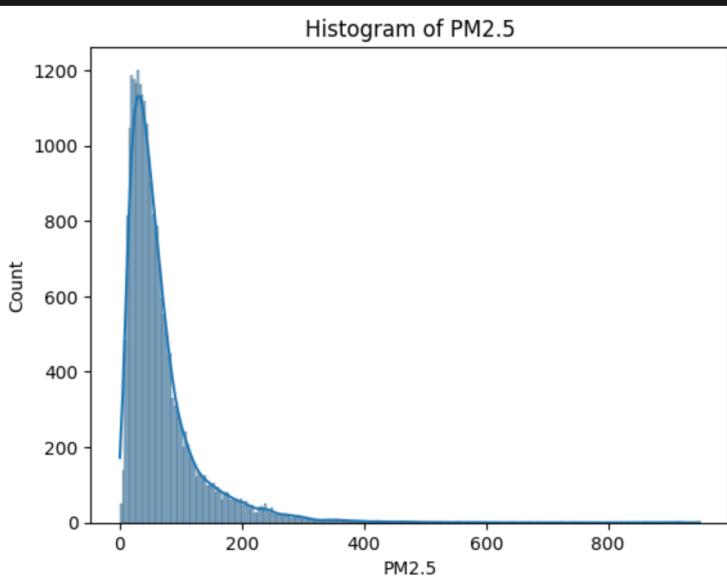
[+ Code](#) [+ Markdown](#)

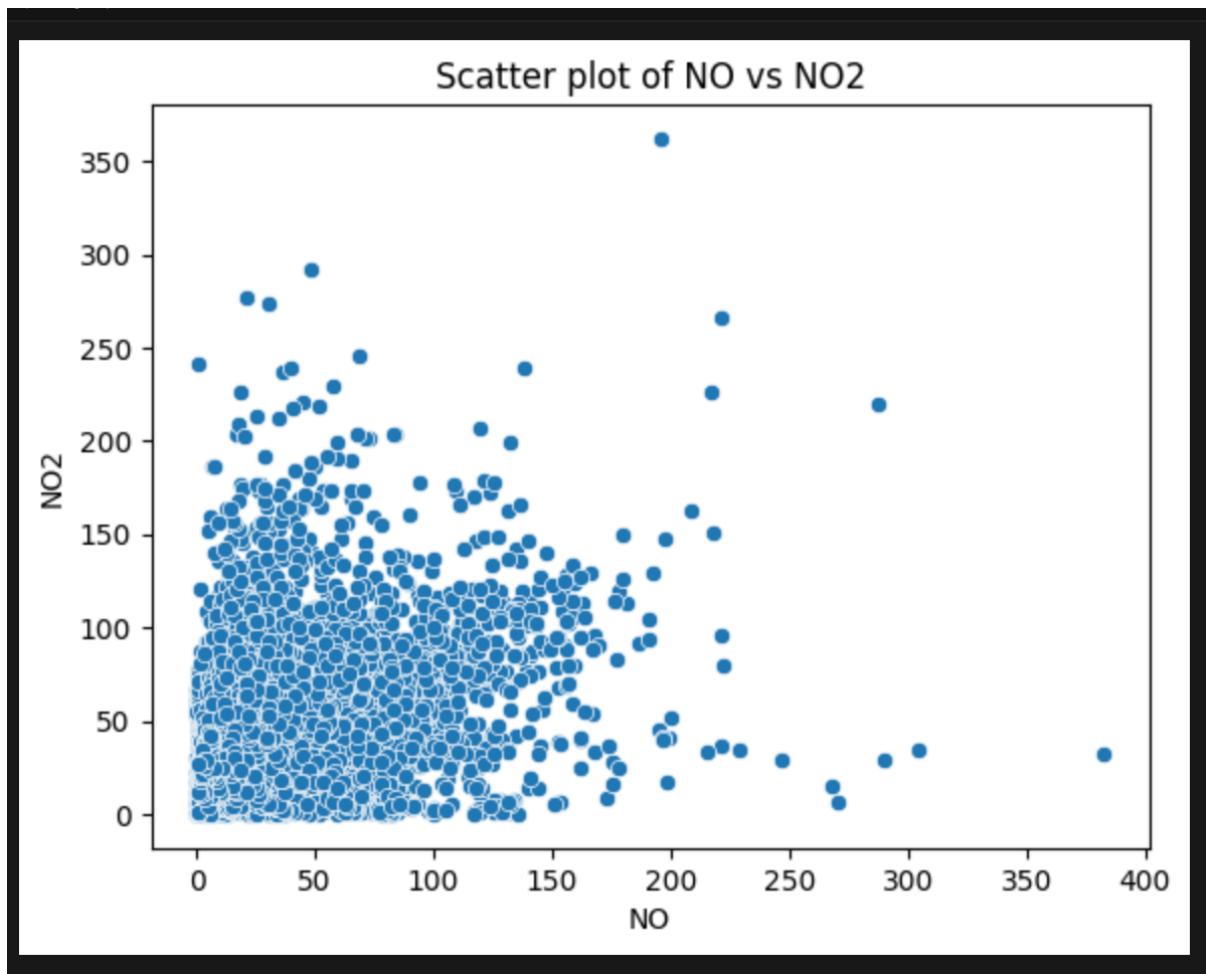
```
▷ [21] print(data_filled.isnull().sum())

[21] ✓ 0.0s
...   City      0
Date      0
PM2.5     0
PM10      0
NO        0
NO2       0
NOx       0
NH3       0
CO        0
SO2       0
O3        0
Benzene    0
Toluene    0
Xylene     0
AQI       0
AQI_Bucket 0
dtype: int64
```

```
Users > ghost > Desktop > TTDPROJECT > SmogSense_Project > SmogSense_Project.ipynb > # Check for missing values
+ Code + Markdown | ▶ Run All ⌂ Restart ⌂ Clear All Outputs | Jupyter Variables ⌂ Outline ...
[22] ✓ 0.0s
...
City          0
Date          0
PM2.5        4598
PM10         11140
NO           3582
NO2          3585
NOx          4185
NH3          10328
CO           2059
SO2          3854
O3            4022
Benzene       5623
Toluene        8041
Xylene         18189
AQI           4681
AQI_Bucket    4681
dtype: int64
City          0
Date          0
PM2.5        0
PM10         0
NO           0
NO2          0
NOx          0
NH3          0
...
Xylene        0
AQI           0
AQI_Bucket    0
dtype: int64
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```







```

> <pre>from sklearn.model_selection import train_test_split</pre>
> <pre># Split the dataset into training and testing sets (80% for training, 20% for testing)
> <pre>X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)</pre>
> <pre># Check the split
> <pre>print(f"Training data size: {X_train.shape[0]}")</pre>
> <pre>print(f"Testing data size: {X_test.shape[0]}")</pre>
34] ✓ 0.0s
... Training data size: 23624
Testing data size: 5907
+ Code + Markdown
print(data.describe())
E J

```

Python

```

Problems 48 Output Debug Console Terminal Ports Jupyter
Changes to be committed:
(use "git restore --staged <file>..." to unstage)
 new file:   FETCH_HEAD
 modified:   SmogSense_Project.ipynb
● fatimaa@new-hostname SmogSense_Project % git commit -m "Initial commit with data loading, EDA, and model training"
[main 1ff8b07] Initial commit with data loading, EDA, and model training
 2 files changed, 1073 insertions(+), 20 deletions(-)
 create mode 100644 FETCH_HEAD
○ fatimaa@new-hostname SmogSense_Project %

```

git Py

```
(1/1) Stage this: !git add -y, n, q, a, d, c, p, +, - y
● fatimaa@new-hostname SmogSense_Project % git commit -m "Load Data in the tool and brief explanation"
[backup-branch 0052aa2] Load Data in the tool and brief explanation
  3 files changed, 9 insertions(+), 137799 deletions(-)
○ fatimaa@new-hostname SmogSense_Project % █
```

⌘K to generate command

```
● fatimaa@new-hostname SmogSense_Project % git add .
git commit -m "Perform EDA and Data Overview"
[backup-branch fc69cef] Perform EDA and Data Overview
  1 file changed, 5 insertions(+)
  create mode 100644 .gitignore
○ fatimaa@new-hostname SmogSense_Project % █
```

```
y = df_clean['AQI']
X = X.fillna(X.mean())

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

model = LinearRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))

print("-----")
print("Model Performance Results:")
print(f"Root Mean Squared Error (RMSE): {rmse:.2f}")
print(f"Average AQI (for comparison): {y.mean():.2f}")
print("-----")

plt.figure(figsize=(8, 6))
plt.scatter(y_test, y_pred, alpha=0.5, color='blue')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'r--', lw=2) # Perfect fit line
plt.xlabel('Actual AQI')
plt.ylabel('Predicted AQI')
plt.title(f'Actual vs Predicted AQI (RMSE: {rmse:.2f})')
plt.grid(True)
plt.show()
```

1] ✓ 0.2s

Python

```
Model Performance Results:
Root Mean Squared Error (RMSE): 59.11
Average AQI (for comparison): 166.46
```

```

Users > ghost > Desktop > TTDPROJECT > SmogSense_Project > SmogSense_Project.ipynb > # Check for missing values
+ Code + Markdown | ▶ Run All ⚡ Restart ⌂ Clear All Outputs ⌂ Jupyter Variables ⌂ Outline ...
[22] ✓ 0.0s Python 3
...
...   City      0
    Date      0
  PM2.5     4598
  PM10     11140
    NO     3582
   NO2     3585
   NOx     4185
   NH3    18328
    CO     2059
   SO2     3854
    O3     4022
  Benzene    5623
  Toluene    8041
  Xylene    18109
   AQI     4681
  AQI_Bucket  4681
dtype: int64
City      0
Date      0
PM2.5     0
PM10     0
NO       0
NO2      0
NOx      0
NH3      0
...
Xylene     0
AQI       0
AQI_Bucket  0
dtype: int64
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...

```

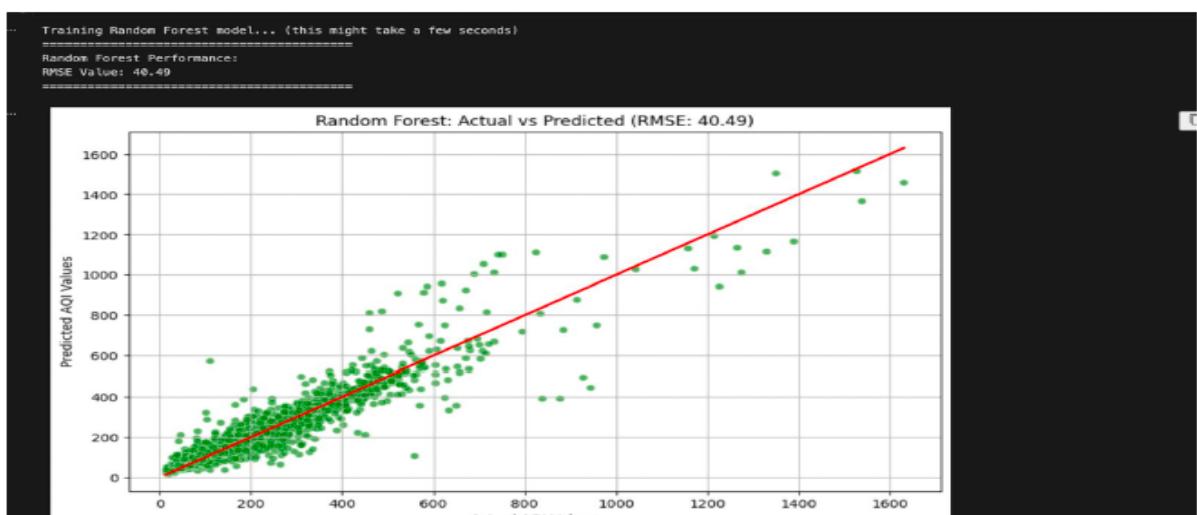


```

[27] ✓ 2.3s Python
...
... /var/folders/gh/f7q_q5x7ql8nby_0365_wph0000gn/T/ipykernel_68770/2403539929.py:2: DtypeWarning: Columns (15) have
  station_hour_df = pd.read_csv('DATASET INDIA/station_hour.csv') # Path to 'station_hour.csv'
  StationId          Datetime  PM2.5    PM10    NO    NO2    NOx    NH3  \
0    AP001  2017-11-24 17:00:00  60.50  98.00  2.35  30.80  18.25  8.50
1    AP001  2017-11-24 18:00:00  65.50 111.25  2.70  24.20  15.07  9.77
2    AP001  2017-11-24 19:00:00  80.00 132.00  2.10  25.18  15.15 12.02
3    AP001  2017-11-24 20:00:00  81.50 133.25  1.95  16.25  10.23 11.58
4    AP001  2017-11-24 21:00:00  75.25 116.00  1.43  17.48  10.43 12.03

      CO    SO2    O3  Benzene  Toluene  Xylene  AQI AQI_Bucket
0  0.1  11.85  126.40    0.1    6.10    0.10  NaN    NaN
1  0.1   13.17  117.12    0.1    6.25    0.15  NaN    NaN
2  0.1   12.08  98.98    0.2    5.98    0.18  NaN    NaN
3  0.1   10.47  112.20    0.2    6.72    0.10  NaN    NaN
4  0.1    9.12  106.35    0.2    5.75    0.08  NaN    NaN

```



Model Performance Results:
Root Mean Squared Error (RMSE): 59.11
Average AQI (for comparison): 166.46

