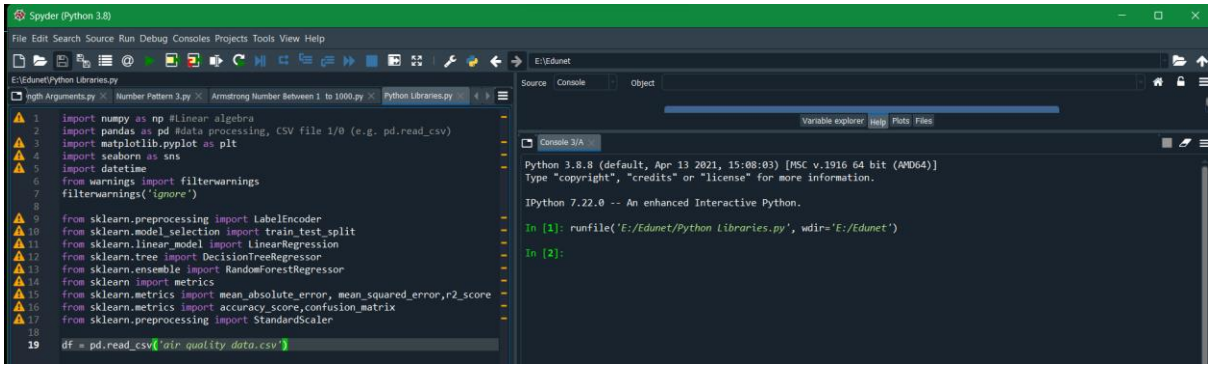
Air quality Prediction Model

Name :- Arshad Liyaquat Shaikh

AICTE Student ID: STU671a198cd1ec81729763724

AICTE Internship ID: 17283057006703da244fae9




```
In [2]: df.head()
Out[2]:
       City        Date  PM2.5  PM10  ...  Toluene  Xylene  AQI  AQI_Bucket
0  Ahmedabad  2015-01-01    NaN   NaN  ...     0.02    0.00  NaN         NaN
1  Ahmedabad  2015-01-02    NaN   NaN  ...     5.50    3.77  NaN         NaN
2  Ahmedabad  2015-01-03    NaN   NaN  ...    16.40    2.25  NaN         NaN
3  Ahmedabad  2015-01-04    NaN   NaN  ...    10.14    1.00  NaN         NaN
4  Ahmedabad  2015-01-05    NaN   NaN  ...    18.89    2.78  NaN         NaN

[5 rows x 16 columns]
```

```
In [3]: df.shape
Out[3]: (29531, 16)
```

```
In [4]: df.info()
<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
```

```
In [5]: df.isnull().sum()
Out[5]:
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
```

```
In [6]: df.duplicated().sum()
Out[6]: 0
```

```
In [7]: df1= df.dropna(subset=['AQI'],inplace=True)

In [8]: df.isnull().sum().sort_values(ascending=False)
Out[8]:
Xylene        15372
PM10           7086
NH3            6536
Toluene        5826
Benzene        3535
NOx            1857
O3              807
PM2.5           678
SO2             605
CO              445
NO2             391
NO              387
City              0
Date              0
AQI               0
AQI_Bucket        0
dtype: int64
```

```
In [9]: df.shape
Out[9]: (24850, 16)

In [10]: df.describe().T
Out[10]:
           count        mean         std  ...      50%       75%      max
PM2.5    24172.0   67.476613   63.075398  ...   48.785   80.9250   914.94
PM10     17764.0  118.454435   89.487976  ...   96.180  150.1825   917.08
NO       24463.0   17.622421   22.421138  ...    9.910   20.0300   390.68
NO2      24459.0   28.978391   24.627054  ...   22.100   38.2400   362.21
NOx      22993.0   32.289012   30.712855  ...   23.680   40.1700   378.24
NH3      18314.0   23.848366   25.875981  ...   16.310   30.3600   352.89
CO       24405.0    2.345267    7.075208  ...    0.930    1.4800   175.81
SO2      24245.0   14.362933   17.428693  ...    9.220   15.1400   186.08
O3       24043.0   34.912885   21.724525  ...   31.250   46.0800   257.73
Benzene  21315.0    3.458668   16.036020  ...    1.290    3.3400   455.03
Toluene  19024.0    9.525714   20.881085  ...    3.575   10.1800   454.85
Xylene    9478.0    3.588683    6.754324  ...    1.420    4.1200   170.37
AQI      24850.0  166.463581  140.696585  ...  118.000  208.0000  2049.00

[13 rows x 8 columns]
```

```
In [11]: null_values_percentage = (df.isnull().sum()/df.isnull().count()*100).sort_values(ascending=False)

In [12]: null_values_percentage
Out[12]:
Xylene        61.859155
PM10          28.515091
NH3           26.301811
Toluene       23.444668
Benzene       14.225352
NOx            7.472837
O3             3.247485
PM2.5          2.728370
SO2            2.434608
CO             1.790744
NO2            1.573441
NO             1.557344
City           0.000000
Date           0.000000
AQI            0.000000
AQI_Bucket     0.000000
dtype: float64
```