```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
from sklearn import metrics
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
from sklearn.tree import DecisionTreeRegressor
import xgboost as xgb
from sklearn.cluster import KMeans

air=pd.read_csv('/content/Air quality-analysis-2014.csv')

air
```

```
      Stn Code Sampling Date        State City/Town/Village/Area  \
0           38      01-02-14  Tamil Nadu                 Chennai
1           38      01-07-14  Tamil Nadu                 Chennai
2           38      21-01-14  Tamil Nadu                 Chennai
3           38      23-01-14  Tamil Nadu                 Chennai
4           38      28-01-14  Tamil Nadu                 Chennai
...        ...           ...         ...                     ...
2874       773      12-03-14  Tamil Nadu                  Trichy
2875       773      12-10-14  Tamil Nadu                  Trichy
2876       773      17-12-14  Tamil Nadu                  Trichy
2877       773      24-12-14  Tamil Nadu                  Trichy
2878       773      31-12-14  Tamil Nadu                  Trichy

                      Location of Monitoring Station  \
0     Kathivakkam, Municipal Kalyana Mandapam, Chennai
1     Kathivakkam, Municipal Kalyana Mandapam, Chennai
2     Kathivakkam, Municipal Kalyana Mandapam, Chennai
3     Kathivakkam, Municipal Kalyana Mandapam, Chennai
4     Kathivakkam, Municipal Kalyana Mandapam, Chennai
...                                                ...
2874                        Central Bus Stand, Trichy
2875                        Central Bus Stand, Trichy
2876                        Central Bus Stand, Trichy
2877                        Central Bus Stand, Trichy
2878                        Central Bus Stand, Trichy

                                     Agency  \
0     Tamilnadu State Pollution Control Board
1     Tamilnadu State Pollution Control Board
2     Tamilnadu State Pollution Control Board
```

```
3       Tamilnadu State Pollution Control Board
4       Tamilnadu State Pollution Control Board
...                                             ...
2874  Tamilnadu State Pollution Control Board
2875  Tamilnadu State Pollution Control Board
2876  Tamilnadu State Pollution Control Board
2877  Tamilnadu State Pollution Control Board
2878  Tamilnadu State Pollution Control Board

                          Type of Location   SO2   NO2  RSPM/PM10  PM
2.5
0                          Industrial Area  11.0  17.0      55.0
NaN
1                          Industrial Area  13.0  17.0      45.0
NaN
2                          Industrial Area  12.0  18.0      50.0
NaN
3                          Industrial Area  15.0  16.0      46.0
NaN
4                          Industrial Area  13.0  14.0      42.0
NaN
...                                    ...   ...   ...       ...     ..
.
2874  Residential, Rural and other Areas  15.0  18.0     102.0
NaN
2875  Residential, Rural and other Areas  12.0  14.0      91.0
NaN
2876  Residential, Rural and other Areas  19.0  22.0     100.0
NaN
2877  Residential, Rural and other Areas  15.0  17.0      95.0
NaN
2878  Residential, Rural and other Areas  14.0  16.0      94.0
NaN

[2879 rows x 11 columns]

air.describe()

           Stn Code          SO2          NO2    RSPM/PM10  PM 2.5
count  2879.000000  2868.000000  2866.000000  2875.000000     0.0
mean    475.750261    11.503138    22.136776    62.494261     NaN
std     277.675577     5.051702     7.128694    31.368745     NaN
min      38.000000     2.000000     5.000000    12.000000     NaN
25%     238.000000     8.000000    17.000000    41.000000     NaN
50%     366.000000    12.000000    22.000000    55.000000     NaN
75%     764.000000    15.000000    25.000000    78.000000     NaN
max     773.000000    49.000000    71.000000   269.000000     NaN

air.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2879 entries, 0 to 2878
Data columns (total 11 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   Stn Code                     2879 non-null   int64
 1   Sampling Date                2879 non-null   object
 2   State                        2879 non-null   object
 3   City/Town/Village/Area       2879 non-null   object
 4   Location of Monitoring Station  2879 non-null  object
 5   Agency                       2879 non-null   object
 6   Type of Location             2879 non-null   object
 7   SO2                          2868 non-null   float64
 8   NO2                          2866 non-null   float64
 9   RSPM/PM10                    2875 non-null   float64
 10  PM 2.5                       0 non-null      float64
dtypes: float64(4), int64(1), object(6)
memory usage: 247.5+ KB

air.isnull().sum()

Stn Code                         0
Sampling Date                    0
State                            0
City/Town/Village/Area           0
Location of Monitoring Station   0
Agency                           0
Type of Location                 0
SO2                             11
NO2                             13
RSPM/PM10                        4
PM 2.5                        2879
dtype: int64

air_fillna = air

air_fillna.fillna(air_fillna.mean(), inplace=True)
# count the number of NaN values in each column
print(air_fillna.isnull().sum())

air_fillna

Stn Code                         0
Sampling Date                    0
State                            0
City/Town/Village/Area           0
Location of Monitoring Station   0
Agency                           0
Type of Location                 0
SO2                              0
```

```
NO2                                               0
RSPM/PM10                                         0
PM 2.5                                         2879
dtype: int64

<ipython-input-10-644c425b2295>:1: FutureWarning: The default value of
numeric_only in DataFrame.mean is deprecated. In a future version, it
will default to False. In addition, specifying 'numeric_only=None' is
deprecated. Select only valid columns or specify the value of
numeric_only to silence this warning.
  air_fillna.fillna(air_fillna.mean(), inplace=True)

      Stn Code Sampling Date        State City/Town/Village/Area  \
0           38      01-02-14  Tamil Nadu                 Chennai
1           38      01-07-14  Tamil Nadu                 Chennai
2           38      21-01-14  Tamil Nadu                 Chennai
3           38      23-01-14  Tamil Nadu                 Chennai
4           38      28-01-14  Tamil Nadu                 Chennai
...        ...           ...         ...                     ...
2874       773      12-03-14  Tamil Nadu                  Trichy
2875       773      12-10-14  Tamil Nadu                  Trichy
2876       773      17-12-14  Tamil Nadu                  Trichy
2877       773      24-12-14  Tamil Nadu                  Trichy
2878       773      31-12-14  Tamil Nadu                  Trichy

                          Location of Monitoring Station  \
0     Kathivakkam, Municipal Kalyana Mandapam, Chennai
1     Kathivakkam, Municipal Kalyana Mandapam, Chennai
2     Kathivakkam, Municipal Kalyana Mandapam, Chennai
3     Kathivakkam, Municipal Kalyana Mandapam, Chennai
4     Kathivakkam, Municipal Kalyana Mandapam, Chennai
...                                                ...
2874                         Central Bus Stand, Trichy
2875                         Central Bus Stand, Trichy
2876                         Central Bus Stand, Trichy
2877                         Central Bus Stand, Trichy
2878                         Central Bus Stand, Trichy

                                       Agency  \
0       Tamilnadu State Pollution Control Board
1       Tamilnadu State Pollution Control Board
2       Tamilnadu State Pollution Control Board
3       Tamilnadu State Pollution Control Board
4       Tamilnadu State Pollution Control Board
...                                        ...
2874    Tamilnadu State Pollution Control Board
2875    Tamilnadu State Pollution Control Board
2876    Tamilnadu State Pollution Control Board
2877    Tamilnadu State Pollution Control Board
2878    Tamilnadu State Pollution Control Board
```

```
                        Type of Location    SO2    NO2    RSPM/PM10    PM
2.5
0                       Industrial Area    11.0   17.0       55.0
NaN
1                       Industrial Area    13.0   17.0       45.0
NaN
2                       Industrial Area    12.0   18.0       50.0
NaN
3                       Industrial Area    15.0   16.0       46.0
NaN
4                       Industrial Area    13.0   14.0       42.0
NaN
...                                          ...    ...    ...         ...       ..
.
2874   Residential, Rural and other Areas  15.0   18.0      102.0
NaN
2875   Residential, Rural and other Areas  12.0   14.0       91.0
NaN
2876   Residential, Rural and other Areas  19.0   22.0      100.0
NaN
2877   Residential, Rural and other Areas  15.0   17.0       95.0
NaN
2878   Residential, Rural and other Areas  14.0   16.0       94.0
NaN

[2879 rows x 11 columns]

le=LabelEncoder()
air['State']=le.fit_transform(air['State'])
air

      Stn Code  Sampling Date  State  City/Town/Village/Area  \
0           38      01-02-14       0                 Chennai
1           38      01-07-14       0                 Chennai
2           38      21-01-14       0                 Chennai
3           38      23-01-14       0                 Chennai
4           38      28-01-14       0                 Chennai
...        ...           ...     ...                     ...
2874       773      12-03-14       0                  Trichy
2875       773      12-10-14       0                  Trichy
2876       773      17-12-14       0                  Trichy
2877       773      24-12-14       0                  Trichy
2878       773      31-12-14       0                  Trichy

                     Location of Monitoring Station  \
0      Kathivakkam, Municipal Kalyana Mandapam, Chennai
1      Kathivakkam, Municipal Kalyana Mandapam, Chennai
2      Kathivakkam, Municipal Kalyana Mandapam, Chennai
3      Kathivakkam, Municipal Kalyana Mandapam, Chennai
```

```
4       Kathivakkam, Municipal Kalyana Mandapam, Chennai
...                                                   ...
2874                          Central Bus Stand, Trichy
2875                          Central Bus Stand, Trichy
2876                          Central Bus Stand, Trichy
2877                          Central Bus Stand, Trichy
2878                          Central Bus Stand, Trichy

                                        Agency  \
0       Tamilnadu State Pollution Control Board
1       Tamilnadu State Pollution Control Board
2       Tamilnadu State Pollution Control Board
3       Tamilnadu State Pollution Control Board
4       Tamilnadu State Pollution Control Board
...                                         ...
2874    Tamilnadu State Pollution Control Board
2875    Tamilnadu State Pollution Control Board
2876    Tamilnadu State Pollution Control Board
2877    Tamilnadu State Pollution Control Board
2878    Tamilnadu State Pollution Control Board

                          Type of Location   SO2   NO2  RSPM/PM10  PM
2.5
0                            Industrial Area  11.0  17.0       55.0
NaN
1                            Industrial Area  13.0  17.0       45.0
NaN
2                            Industrial Area  12.0  18.0       50.0
NaN
3                            Industrial Area  15.0  16.0       46.0
NaN
4                            Industrial Area  13.0  14.0       42.0
NaN
...                                      ...   ...   ...        ...    ..
.
2874  Residential, Rural and other Areas  15.0  18.0      102.0
NaN
2875  Residential, Rural and other Areas  12.0  14.0       91.0
NaN
2876  Residential, Rural and other Areas  19.0  22.0      100.0
NaN
2877  Residential, Rural and other Areas  15.0  17.0       95.0
NaN
2878  Residential, Rural and other Areas  14.0  16.0       94.0
NaN

[2879 rows x 11 columns]
```

```
le=LabelEncoder()
air['Stn Code']=le.fit_transform(air['Stn Code'])
air
```

	Stn Code	Sampling Date	State	City/Town/Village/Area
0	0	01-02-14	0	Chennai
1	0	01-07-14	0	Chennai
2	0	21-01-14	0	Chennai
3	0	23-01-14	0	Chennai
4	0	28-01-14	0	Chennai
...	...	...	...	...
2874	29	12-03-14	0	Trichy
2875	29	12-10-14	0	Trichy
2876	29	17-12-14	0	Trichy
2877	29	24-12-14	0	Trichy
2878	29	31-12-14	0	Trichy

	Location of Monitoring Station
0	Kathivakkam, Municipal Kalyana Mandapam, Chennai
1	Kathivakkam, Municipal Kalyana Mandapam, Chennai
2	Kathivakkam, Municipal Kalyana Mandapam, Chennai
3	Kathivakkam, Municipal Kalyana Mandapam, Chennai
4	Kathivakkam, Municipal Kalyana Mandapam, Chennai
...	...
2874	Central Bus Stand, Trichy
2875	Central Bus Stand, Trichy
2876	Central Bus Stand, Trichy
2877	Central Bus Stand, Trichy
2878	Central Bus Stand, Trichy

	Agency
0	Tamilnadu State Pollution Control Board
1	Tamilnadu State Pollution Control Board
2	Tamilnadu State Pollution Control Board
3	Tamilnadu State Pollution Control Board
4	Tamilnadu State Pollution Control Board
...	...
2874	Tamilnadu State Pollution Control Board
2875	Tamilnadu State Pollution Control Board
2876	Tamilnadu State Pollution Control Board
2877	Tamilnadu State Pollution Control Board
2878	Tamilnadu State Pollution Control Board

	Type of Location	SO2	NO2	RSPM/PM10	PM 2.5
0	Industrial Area	11.0	17.0	55.0	NaN
1	Industrial Area	13.0	17.0	45.0	NaN
2	Industrial Area	12.0	18.0	50.0	

```
NaN
3                        Industrial Area  15.0  16.0      46.0
NaN
4                        Industrial Area  13.0  14.0      42.0
NaN
...                                      ...   ...   ...       ...      ..
.
2874  Residential, Rural and other Areas  15.0  18.0     102.0
NaN
2875  Residential, Rural and other Areas  12.0  14.0      91.0
NaN
2876  Residential, Rural and other Areas  19.0  22.0     100.0
NaN
2877  Residential, Rural and other Areas  15.0  17.0      95.0
NaN
2878  Residential, Rural and other Areas  14.0  16.0      94.0
NaN

[2879 rows x 11 columns]

le=LabelEncoder()
air['SO2']=le.fit_transform(air['SO2'])
air

      Stn Code Sampling Date  State City/Town/Village/Area  \
0            0      01-02-14      0                Chennai
1            0      01-07-14      0                Chennai
2            0      21-01-14      0                Chennai
3            0      23-01-14      0                Chennai
4            0      28-01-14      0                Chennai
...        ...           ...    ...                    ...
2874        29      12-03-14      0                 Trichy
2875        29      12-10-14      0                 Trichy
2876        29      17-12-14      0                 Trichy
2877        29      24-12-14      0                 Trichy
2878        29      31-12-14      0                 Trichy

                        Location of Monitoring Station  \
0     Kathivakkam, Municipal Kalyana Mandapam, Chennai
1     Kathivakkam, Municipal Kalyana Mandapam, Chennai
2     Kathivakkam, Municipal Kalyana Mandapam, Chennai
3     Kathivakkam, Municipal Kalyana Mandapam, Chennai
4     Kathivakkam, Municipal Kalyana Mandapam, Chennai
...                                                ...
2874                        Central Bus Stand, Trichy
2875                        Central Bus Stand, Trichy
2876                        Central Bus Stand, Trichy
2877                        Central Bus Stand, Trichy
2878                        Central Bus Stand, Trichy
```

```
                                              Agency  \
0       Tamilnadu State Pollution Control Board
1       Tamilnadu State Pollution Control Board
2       Tamilnadu State Pollution Control Board
3       Tamilnadu State Pollution Control Board
4       Tamilnadu State Pollution Control Board
...                                        ...
2874    Tamilnadu State Pollution Control Board
2875    Tamilnadu State Pollution Control Board
2876    Tamilnadu State Pollution Control Board
2877    Tamilnadu State Pollution Control Board
2878    Tamilnadu State Pollution Control Board

                         Type of Location  SO2   NO2  RSPM/PM10  PM 2.5

0                          Industrial Area    9  17.0       55.0     NaN

1                          Industrial Area   12  17.0       45.0     NaN

2                          Industrial Area   11  18.0       50.0     NaN

3                          Industrial Area   14  16.0       46.0     NaN

4                          Industrial Area   12  14.0       42.0     NaN

...                                    ...  ...   ...        ...     ...

2874  Residential, Rural and other Areas   14  18.0      102.0     NaN

2875  Residential, Rural and other Areas   11  14.0       91.0     NaN

2876  Residential, Rural and other Areas   18  22.0      100.0     NaN

2877  Residential, Rural and other Areas   14  17.0       95.0     NaN

2878  Residential, Rural and other Areas   13  16.0       94.0     NaN


[2879 rows x 11 columns]

le=LabelEncoder()
air['Agency']=le.fit_transform(air['Agency'])
air

      Stn Code Sampling Date  State City/Town/Village/Area  \
0            0      01-02-14      0                 Chennai
1            0      01-07-14      0                 Chennai
2            0      21-01-14      0                 Chennai
3            0      23-01-14      0                 Chennai
4            0      28-01-14      0                 Chennai
...        ...           ...    ...                    ...
```

```
2874    29    12-03-14    0                                    Trichy
2875    29    12-10-14    0                                    Trichy
2876    29    17-12-14    0                                    Trichy
2877    29    24-12-14    0                                    Trichy
2878    29    31-12-14    0                                    Trichy

                         Location of Monitoring Station    Agency  \
0      Kathivakkam, Municipal Kalyana Mandapam, Chennai         1
1      Kathivakkam, Municipal Kalyana Mandapam, Chennai         1
2      Kathivakkam, Municipal Kalyana Mandapam, Chennai         1
3      Kathivakkam, Municipal Kalyana Mandapam, Chennai         1
4      Kathivakkam, Municipal Kalyana Mandapam, Chennai         1
...                                               ...        ...
2874                      Central Bus Stand, Trichy         1
2875                      Central Bus Stand, Trichy         1
2876                      Central Bus Stand, Trichy         1
2877                      Central Bus Stand, Trichy         1
2878                      Central Bus Stand, Trichy         1

                         Type of Location  SO2   NO2   RSPM/PM10   PM 2.5

0                           Industrial Area    9  17.0       55.0      NaN

1                           Industrial Area   12  17.0       45.0      NaN

2                           Industrial Area   11  18.0       50.0      NaN

3                           Industrial Area   14  16.0       46.0      NaN

4                           Industrial Area   12  14.0       42.0      NaN

...                                      ...  ...   ...        ...      ...

2874  Residential, Rural and other Areas   14  18.0      102.0      NaN

2875  Residential, Rural and other Areas   11  14.0       91.0      NaN

2876  Residential, Rural and other Areas   18  22.0      100.0      NaN

2877  Residential, Rural and other Areas   14  17.0       95.0      NaN

2878  Residential, Rural and other Areas   13  16.0       94.0      NaN


[2879 rows x 11 columns]

le=LabelEncoder()
air['RSPM/PM10']=le.fit_transform(air['RSPM/PM10'])
air
```

```
      Stn Code Sampling Date  State City/Town/Village/Area  \
0            0      01-02-14      0                Chennai
1            0      01-07-14      0                Chennai
2            0      21-01-14      0                Chennai
3            0      23-01-14      0                Chennai
4            0      28-01-14      0                Chennai
...        ...           ...    ...                    ...
2874        29      12-03-14      0                 Trichy
2875        29      12-10-14      0                 Trichy
2876        29      17-12-14      0                 Trichy
2877        29      24-12-14      0                 Trichy
2878        29      31-12-14      0                 Trichy

                     Location of Monitoring Station  Agency  \
0     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
1     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
2     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
3     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
4     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
...                                               ...     ...
2874                      Central Bus Stand, Trichy       1
2875                      Central Bus Stand, Trichy       1
2876                      Central Bus Stand, Trichy       1
2877                      Central Bus Stand, Trichy       1
2878                      Central Bus Stand, Trichy       1

                    Type of Location  SO2   NO2  RSPM/PM10  PM 2.5

0                      Industrial Area    9  17.0         43     NaN

1                      Industrial Area   12  17.0         33     NaN

2                      Industrial Area   11  18.0         38     NaN

3                      Industrial Area   14  16.0         34     NaN

4                      Industrial Area   12  14.0         30     NaN

...                                ...  ...   ...        ...     ...

2874  Residential, Rural and other Areas   14  18.0         91     NaN

2875  Residential, Rural and other Areas   11  14.0         80     NaN

2876  Residential, Rural and other Areas   18  22.0         89     NaN

2877  Residential, Rural and other Areas   14  17.0         84     NaN

2878  Residential, Rural and other Areas   13  16.0         83     NaN
```

```
[2879 rows x 11 columns]

air['Sampling Date'] =air['Sampling Date'].str.replace('-', ' ')
air
```

	Stn Code	Sampling Date	State	City/Town/Village/Area
0	0	01 02 14	0	Chennai
1	0	01 07 14	0	Chennai
2	0	21 01 14	0	Chennai
3	0	23 01 14	0	Chennai
4	0	28 01 14	0	Chennai
...	...	...	...	...
2874	29	12 03 14	0	Trichy
2875	29	12 10 14	0	Trichy
2876	29	17 12 14	0	Trichy
2877	29	24 12 14	0	Trichy
2878	29	31 12 14	0	Trichy

	Location of Monitoring Station	Agency
0	Kathivakkam, Municipal Kalyana Mandapam, Chennai	1
1	Kathivakkam, Municipal Kalyana Mandapam, Chennai	1
2	Kathivakkam, Municipal Kalyana Mandapam, Chennai	1
3	Kathivakkam, Municipal Kalyana Mandapam, Chennai	1
4	Kathivakkam, Municipal Kalyana Mandapam, Chennai	1
...	...	...
2874	Central Bus Stand, Trichy	1
2875	Central Bus Stand, Trichy	1
2876	Central Bus Stand, Trichy	1
2877	Central Bus Stand, Trichy	1
2878	Central Bus Stand, Trichy	1

	Type of Location	SO2	NO2	RSPM/PM10	PM 2.5
0	Industrial Area	9	17.0	43	NaN
1	Industrial Area	12	17.0	33	NaN
2	Industrial Area	11	18.0	38	NaN
3	Industrial Area	14	16.0	34	NaN
4	Industrial Area	12	14.0	30	NaN
...	...	...	...	...	...
2874	Residential, Rural and other Areas	14	18.0	91	NaN
2875	Residential, Rural and other Areas	11	14.0	80	NaN
2876	Residential, Rural and other Areas	18	22.0	89	NaN

```
2877  Residential, Rural and other Areas   14  17.0         84      NaN

2878  Residential, Rural and other Areas   13  16.0         83      NaN


[2879 rows x 11 columns]
air

      Stn Code Sampling Date  State City/Town/Village/Area  \
0            0     01 02 14      0                  Chennai
1            0     01 07 14      0                  Chennai
2            0     21 01 14      0                  Chennai
3            0     23 01 14      0                  Chennai
4            0     28 01 14      0                  Chennai
...        ...          ...    ...                      ...
2874        29     12 03 14      0                   Trichy
2875        29     12 10 14      0                   Trichy
2876        29     17 12 14      0                   Trichy
2877        29     24 12 14      0                   Trichy
2878        29     31 12 14      0                   Trichy

                         Location of Monitoring Station  Agency  \
0     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
1     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
2     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
3     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
4     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
...                                               ...     ...
2874                          Central Bus Stand, Trichy       1
2875                          Central Bus Stand, Trichy       1
2876                          Central Bus Stand, Trichy       1
2877                          Central Bus Stand, Trichy       1
2878                          Central Bus Stand, Trichy       1

                        Type of Location  SO2  NO2  RSPM/PM10  PM 2.5

0                          Industrial Area    9  17.0         43      NaN

1                          Industrial Area   12  17.0         33      NaN

2                          Industrial Area   11  18.0         38      NaN

3                          Industrial Area   14  16.0         34      NaN

4                          Industrial Area   12  14.0         30      NaN

...                                  ...  ...  ...        ...      ...

2874  Residential, Rural and other Areas   14  18.0         91      NaN
```

```
2875   Residential, Rural and other Areas     11   14.0          80          NaN

2876   Residential, Rural and other Areas     18   22.0          89          NaN

2877   Residential, Rural and other Areas     14   17.0          84          NaN

2878   Residential, Rural and other Areas     13   16.0          83          NaN


[2879 rows x 11 columns]
```

```
air.columns
```

```
Index(['Stn Code', 'Sampling Date', 'State', 'City/Town/Village/Area',
       'Location of Monitoring Station', 'Agency', 'Type of Location',
'SO2',
       'NO2', 'RSPM/PM10', 'PM 2.5'],
      dtype='object')
```

```python
corr = air.corr()
plt.figure(figsize=(8,8))
sns.heatmap(corr, cmap='viridis', annot=True)
```

```
<ipython-input-31-5aecfe2bd3f6>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it
will default to False. Select only valid columns or specify the value
of numeric_only to silence this warning.
  corr = air.corr()
```
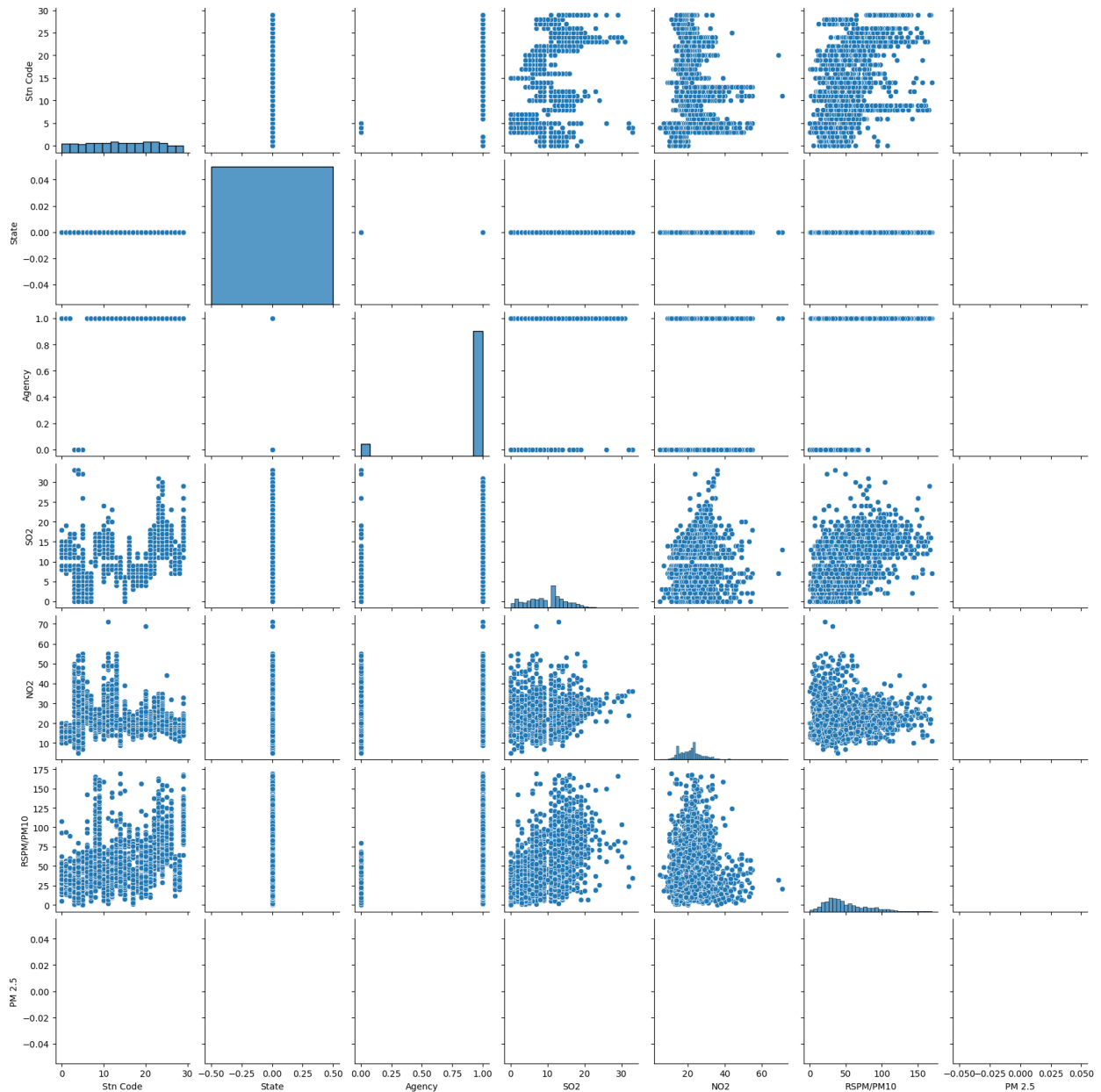
```
<Axes: >
```
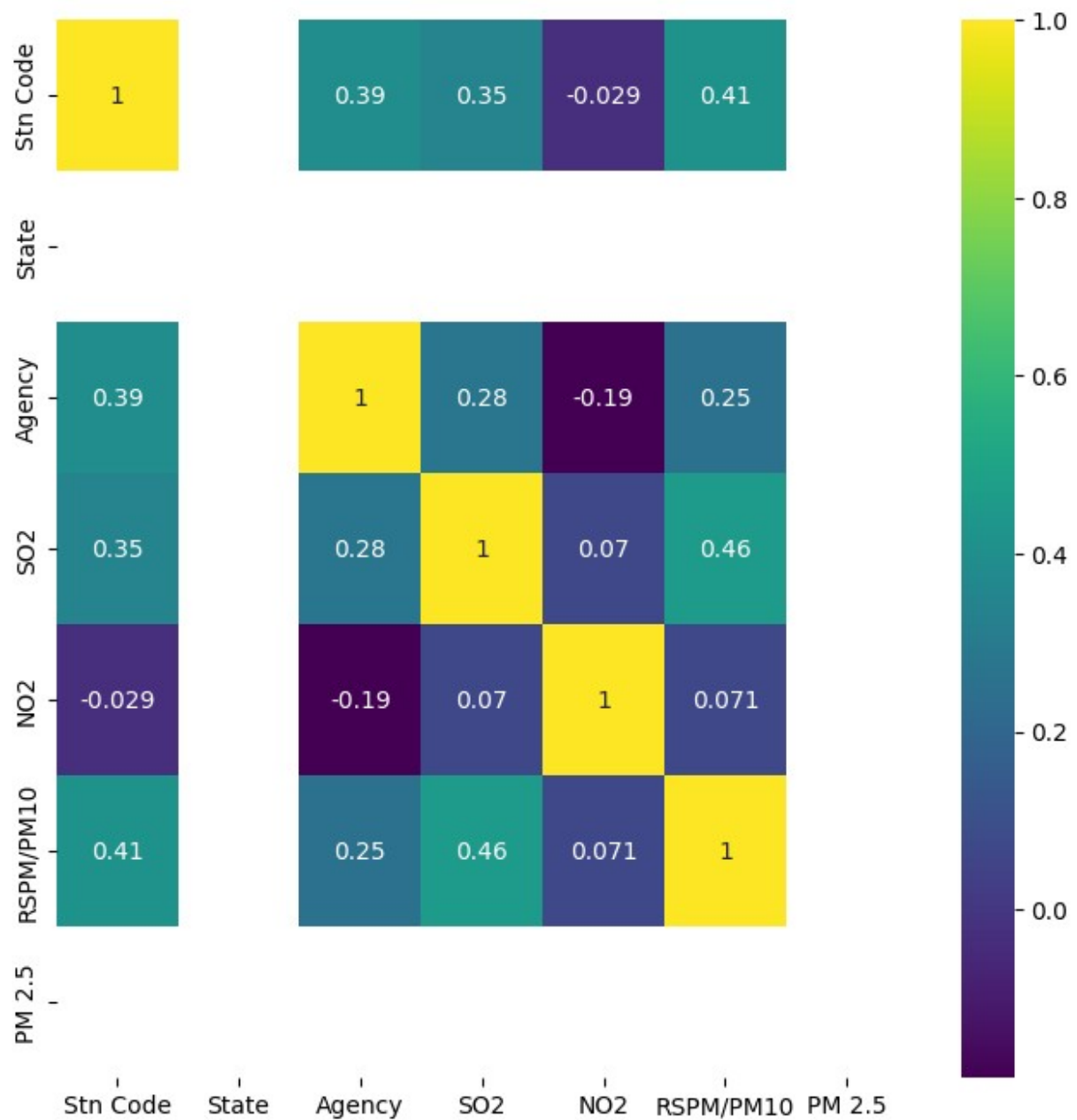
```
sns.pairplot(air)
```

```
<seaborn.axisgrid.PairGrid at 0x7e6e89283b50>
```

```
sns.regplot( y="Agency",x="Type of Location",  data=air)

-------------------------------------------------------------------------
-----
UFuncTypeError                           Traceback (most recent call
last)
<ipython-input-37-f1cb546c44ae> in <cell line: 1>()
----> 1 sns.regplot( y="Agency",x="Type of Location",  data=air)

/usr/local/lib/python3.10/dist-packages/seaborn/regression.py in
regplot(data, x, y, x_estimator, x_bins, x_ci, scatter, fit_reg, ci,
n_boot, units, seed, order, logistic, lowess, robust, logx, x_partial,
```

```
y_partial, truncate, dropna, x_jitter, y_jitter, label, color, marker,
scatter_kws, line_kws, ax)
    757        scatter_kws["marker"] = marker
    758        line_kws = {} if line_kws is None else copy.copy(line_kws)
--> 759        plotter.plot(ax, scatter_kws, line_kws)
    760        return ax
    761

/usr/local/lib/python3.10/dist-packages/seaborn/regression.py in
plot(self, ax, scatter_kws, line_kws)
    366
    367            if self.fit_reg:
--> 368                self.lineplot(ax, line_kws)
    369
    370            # Label the axes

/usr/local/lib/python3.10/dist-packages/seaborn/regression.py in
lineplot(self, ax, kws)
    411            """Draw the model."""
    412            # Fit the regression model
--> 413            grid, yhat, err_bands = self.fit_regression(ax)
    414            edges = grid[0], grid[-1]
    415

/usr/local/lib/python3.10/dist-packages/seaborn/regression.py in
fit_regression(self, ax, x_range, grid)
    197                    else:
    198                        x_min, x_max = ax.get_xlim()
--> 199                grid = np.linspace(x_min, x_max, 100)
    200            ci = self.ci
    201

/usr/local/lib/python3.10/dist-packages/numpy/core/overrides.py in
linspace(*args, **kwargs)

/usr/local/lib/python3.10/dist-packages/numpy/core/function_base.py in
linspace(start, stop, num, endpoint, retstep, dtype, axis)
    125        # Convert float/complex array scalars to float, gh-3504
    126        # and make sure one can use variables that have an
__array_interface__, gh-6634
--> 127        start = asanyarray(start) * 1.0
    128        stop  = asanyarray(stop)  * 1.0
    129

UFuncTypeError: ufunc 'multiply' did not contain a loop with signature
matching types (dtype('<U15'), dtype('float64')) -> None
```

```
sns.regplot( y="Agency",x="Type of Location",  data=air)

---------------------------------------------------------------------
-----
UFuncTypeError                             Traceback (most recent call
last)
<ipython-input-39-f1cb546c44ae> in <cell line: 1>()
----> 1 sns.regplot( y="Agency",x="Type of Location",  data=air)

/usr/local/lib/python3.10/dist-packages/seaborn/regression.py in
regplot(data, x, y, x_estimator, x_bins, x_ci, scatter, fit_reg, ci,
n_boot, units, seed, order, logistic, lowess, robust, logx, x_partial,
y_partial, truncate, dropna, x_jitter, y_jitter, label, color, marker,
scatter_kws, line_kws, ax)
    757       scatter_kws["marker"] = marker
    758       line_kws = {} if line_kws is None else copy.copy(line_kws)
--> 759       plotter.plot(ax, scatter_kws, line_kws)
    760       return ax
    761

/usr/local/lib/python3.10/dist-packages/seaborn/regression.py in
plot(self, ax, scatter_kws, line_kws)
    366
    367             if self.fit_reg:
--> 368                 self.lineplot(ax, line_kws)
    369
```

```
    370           # Label the axes

/usr/local/lib/python3.10/dist-packages/seaborn/regression.py in
lineplot(self, ax, kws)
    411           """Draw the model."""
    412           # Fit the regression model
--> 413           grid, yhat, err_bands = self.fit_regression(ax)
    414           edges = grid[0], grid[-1]
    415

/usr/local/lib/python3.10/dist-packages/seaborn/regression.py in
fit_regression(self, ax, x_range, grid)
    197                   else:
    198                       x_min, x_max = ax.get_xlim()
--> 199               grid = np.linspace(x_min, x_max, 100)
    200           ci = self.ci
    201

/usr/local/lib/python3.10/dist-packages/numpy/core/overrides.py in
linspace(*args, **kwargs)

/usr/local/lib/python3.10/dist-packages/numpy/core/function_base.py in
linspace(start, stop, num, endpoint, retstep, dtype, axis)
    125     # Convert float/complex array scalars to float, gh-3504
    126     # and make sure one can use variables that have an
__array_interface__, gh-6634
--> 127     start = asanyarray(start) * 1.0
    128     stop  = asanyarray(stop)  * 1.0
    129

UFuncTypeError: ufunc 'multiply' did not contain a loop with signature
matching types (dtype('<U15'), dtype('float64')) -> None
```

```
sns.scatterplot( y="SO2",x="RSPM/PM10",  data=air)
<Axes: xlabel='RSPM/PM10', ylabel='SO2'>
```
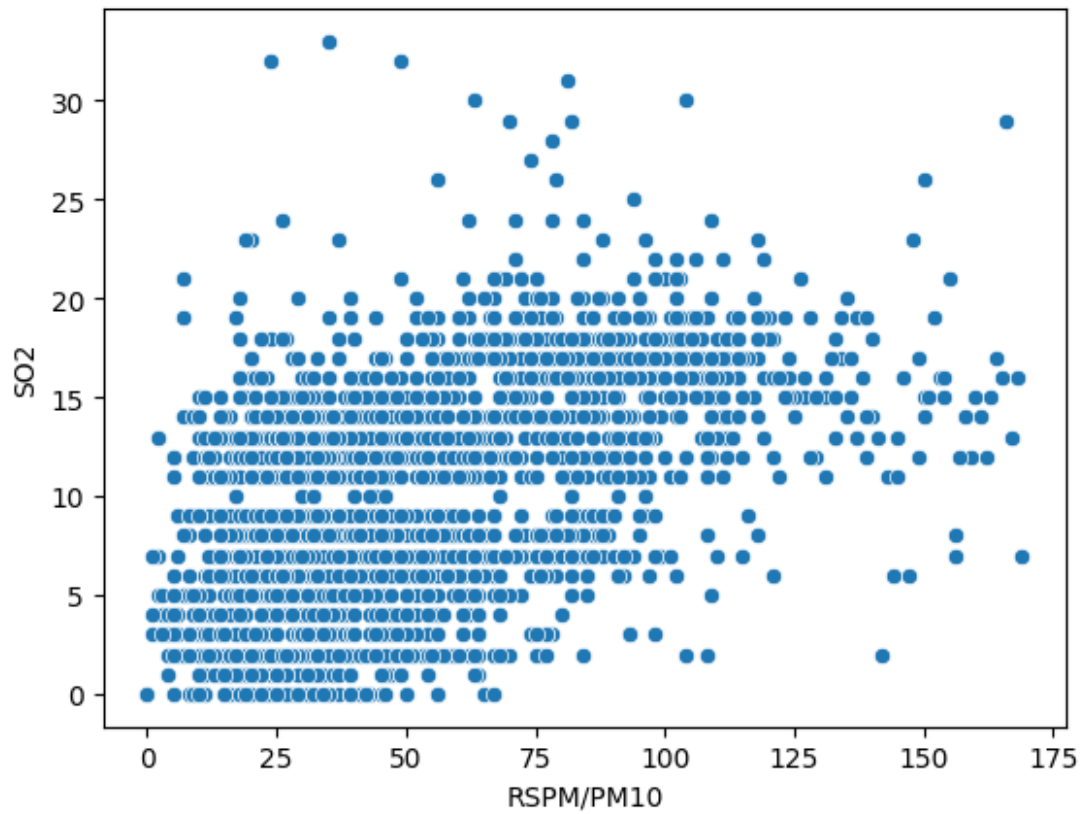
```
sns.displot(air, x="State", hue="SO2",  common_norm=False)
<seaborn.axisgrid.FacetGrid at 0x7e6e8838d240>
```
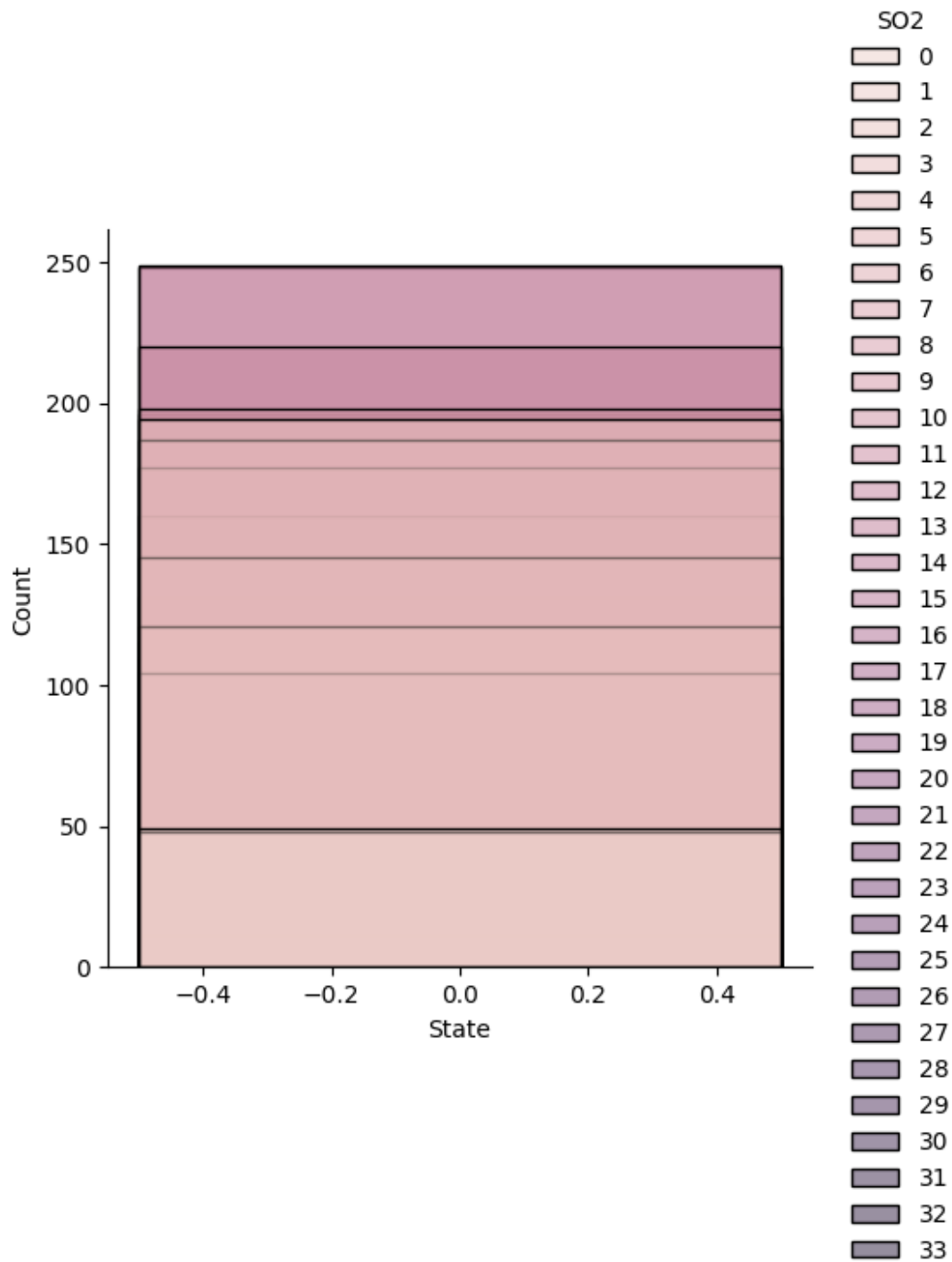
```
sns.scatterplot(air, x='Type of Location',y="State")
<Axes: xlabel='Type of Location', ylabel='State'>
```
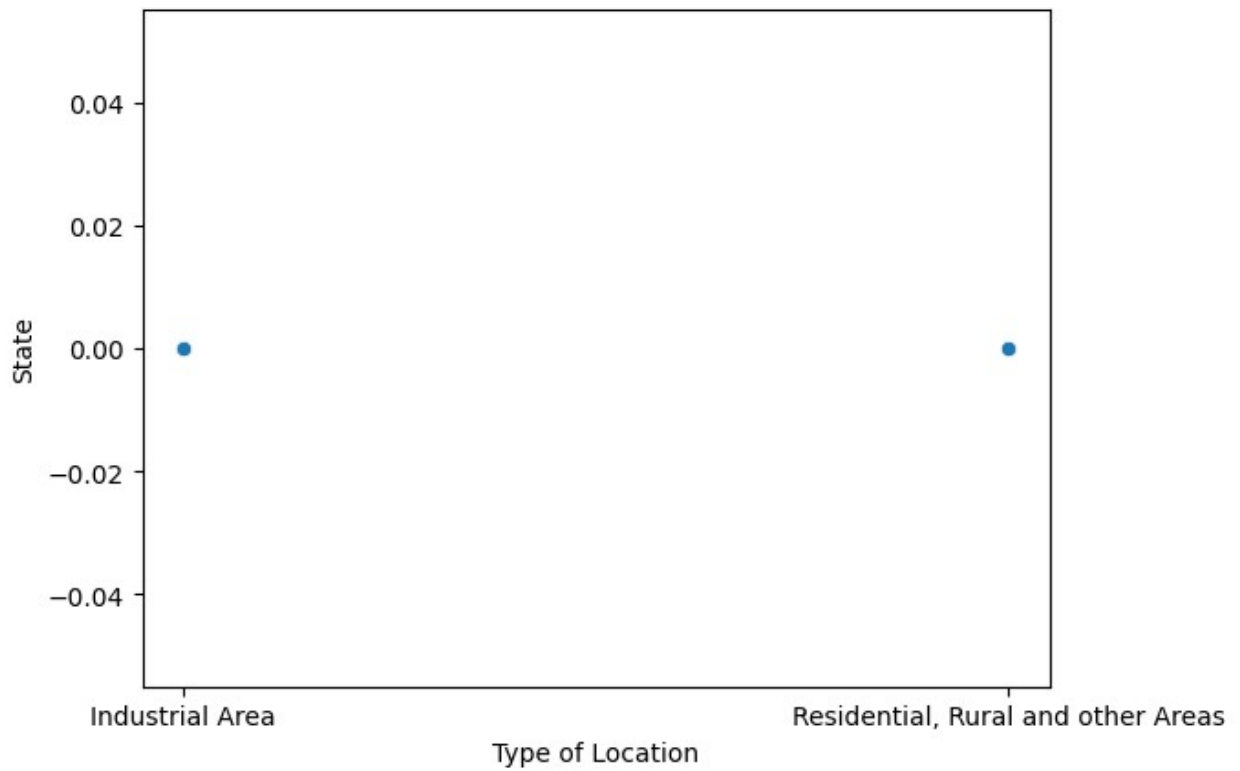
```
sns.displot(air, x="State",kde=True)
<seaborn.axisgrid.FacetGrid at 0x7e6e84973700>
```
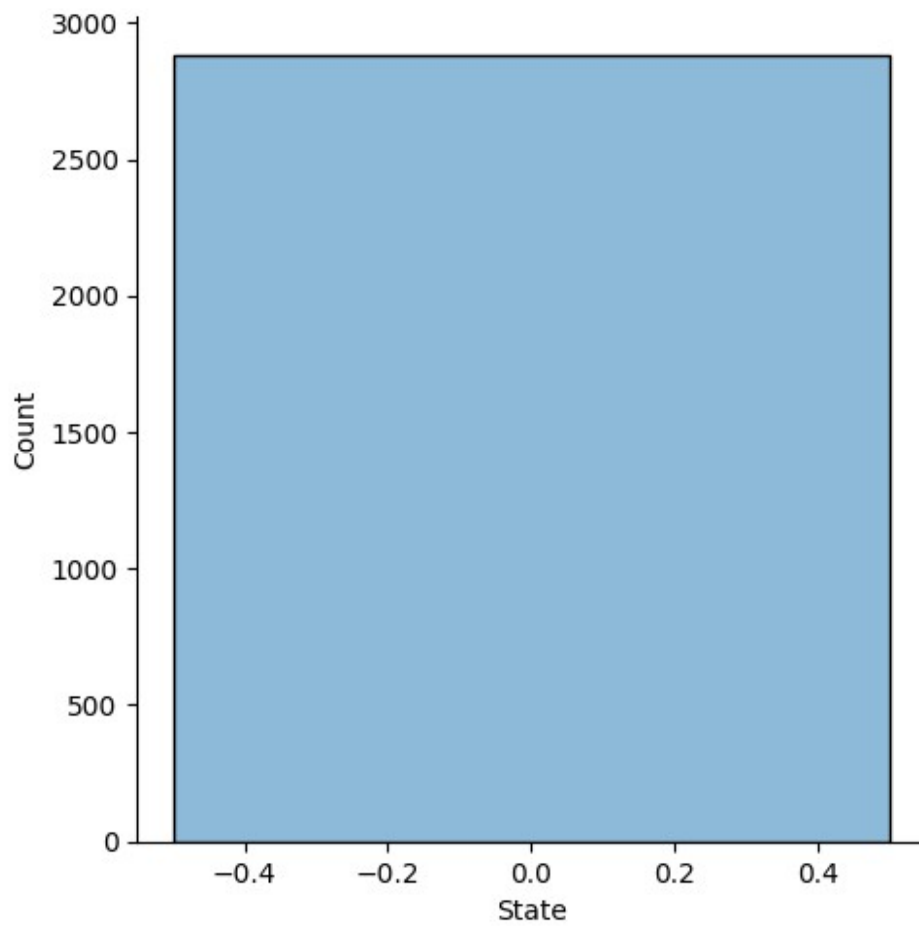
```
sns.displot(air, x="City/Town/Village/Area", kde=True)

<seaborn.axisgrid.FacetGrid at 0x7e6e88465b40>
```
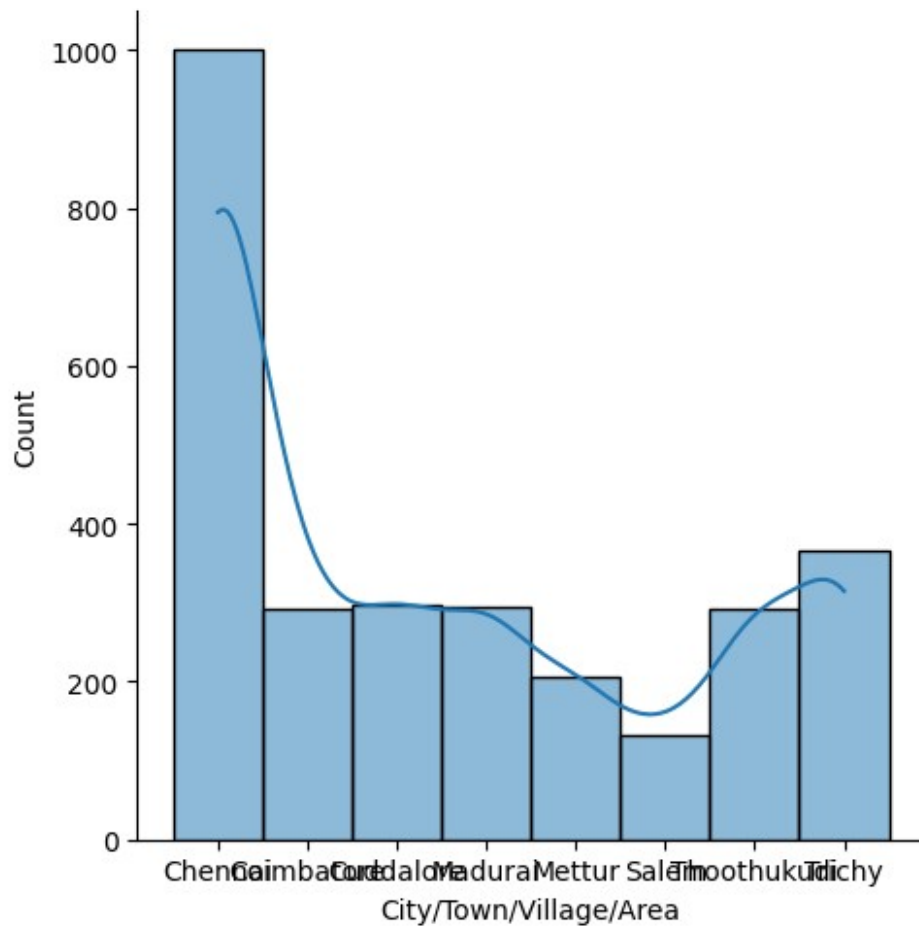
```
sns.regplot( y="State",x="Stn Code",  data=air)
```

```
<Axes: xlabel='Stn Code', ylabel='State'>
```
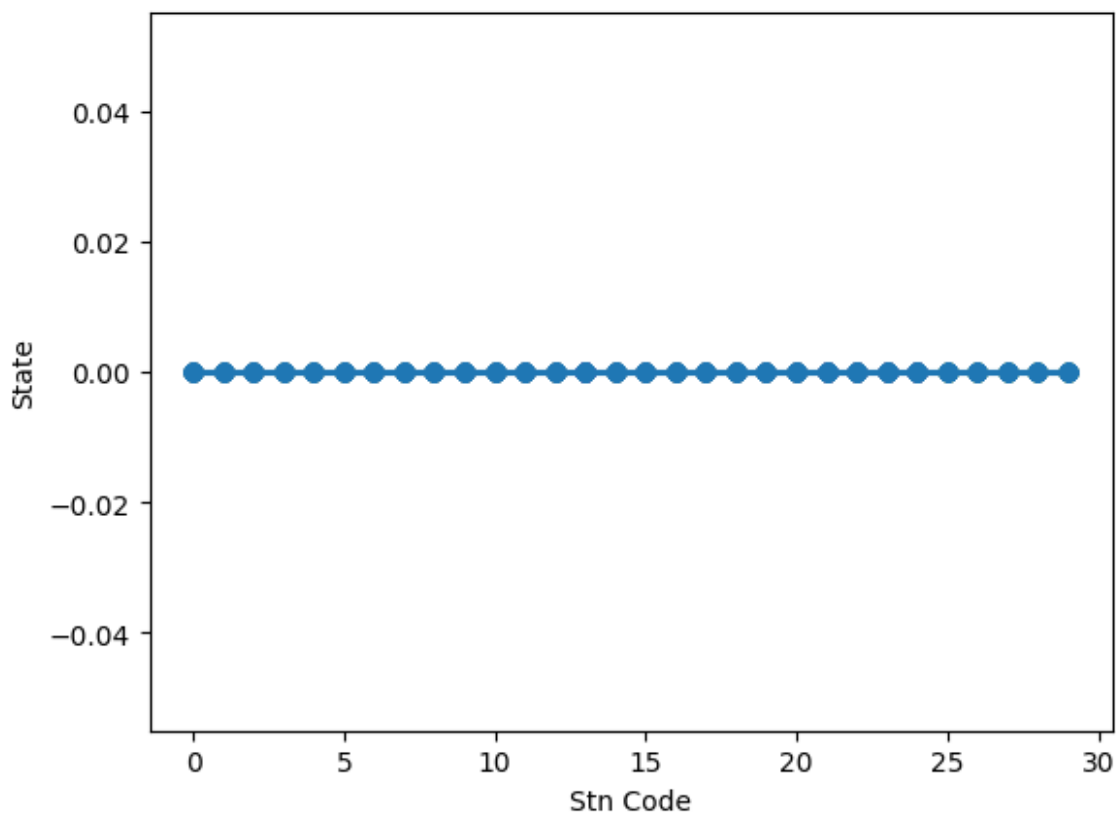
```
x=air[['Stn Code','Sampling Date',    'State',
      'City/Town/Village/Area',  'Location of Monitoring Station',
      'Agency',  'Type of Location',   'SO2',       'NO2',
      'RSPM/PM10',     'PM 2.5']]

x

      Stn Code Sampling Date   State City/Town/Village/Area  \
0            0       01 02 14       0                Chennai
1            0       01 07 14       0                Chennai
2            0       21 01 14       0                Chennai
3            0       23 01 14       0                Chennai
4            0       28 01 14       0                Chennai
...        ...            ...     ...                    ...
2874        29       12 03 14       0                  Trichy
2875        29       12 10 14       0                  Trichy
2876        29       17 12 14       0                  Trichy
2877        29       24 12 14       0                  Trichy
2878        29       31 12 14       0                  Trichy

                      Location of Monitoring Station   Agency  \
0      Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
1      Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
2      Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
3      Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
```

```
4       Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
...                                                   ...     ...
2874                            Central Bus Stand, Trichy       1
2875                            Central Bus Stand, Trichy       1
2876                            Central Bus Stand, Trichy       1
2877                            Central Bus Stand, Trichy       1
2878                            Central Bus Stand, Trichy       1

                            Type of Location  SO2  NO2  RSPM/PM10  PM 2.5

0                            Industrial Area    9  17.0         43     NaN

1                            Industrial Area   12  17.0         33     NaN

2                            Industrial Area   11  18.0         38     NaN

3                            Industrial Area   14  16.0         34     NaN

4                            Industrial Area   12  14.0         30     NaN

...                                      ...  ...   ...        ...     ...

2874  Residential, Rural and other Areas   14  18.0         91     NaN

2875  Residential, Rural and other Areas   11  14.0         80     NaN

2876  Residential, Rural and other Areas   18  22.0         89     NaN

2877  Residential, Rural and other Areas   14  17.0         84     NaN

2878  Residential, Rural and other Areas   13  16.0         83     NaN


[2879 rows x 11 columns]
```

```python
y=air[['RSPM/PM10']]
y
```

```
      RSPM/PM10
0            43
1            33
2            38
3            34
4            30
...         ...
2874         91
2875         80
2876         89
2877         84
2878         83
```

```
[2879 rows x 1 columns]

x_train, x_test, y_train, y_test = train_test_split(x,y,
random_state=42)

x_train
```

	Stn Code	Sampling Date	State	City/Town/Village/Area	\
1385	16	12 10 14	0	Cuddalore
1125	6	25 03 14	0	Coimbatore
482	22	25 09 14	0	Chennai
1765	11	18 10 14	0	Madurai
1000	15	01 04 14	0	Coimbatore
...	...	...	...	...
1638	10	07 10 14	0	Madurai
1095	15	12 06 14	0	Coimbatore
1130	6	04 11 14	0	Coimbatore
1294	16	01 08 14	0	Cuddalore
860	5	20 05 14	0	Chennai

	Location of Monitoring Station	Agency	\
1385	Eachangadu Villagae	1
1125	SIDCO Office, Coimbatore	1
482	Anna Nagar, Chennai	1
1765	Fenner (I) Ltd. Employees Assiciation Building...	1
1000	Poniarajapuram, On the top of DEL, Coimbatore	1
...	...	...
1638	Highway (Project -I) Building, Madurai	1
1095	Poniarajapuram, On the top of DEL, Coimbatore	1
1130	SIDCO Office, Coimbatore	1
1294	Eachangadu Villagae	1
860	Thiruvottiyur Municipal Office, Chennai	0

	Type of Location	SO2	NO2	RSPM/PM10	PM 2.5
1385	Residential, Rural and other Areas	9	21.0	56	NaN
1125	Industrial Area	2	26.0	142	NaN
482	Residential, Rural and other Areas	13	27.0	41	NaN
1765	Industrial Area	11	16.0	18	NaN
1000	Residential, Rural and other Areas	2	23.0	56	NaN
...	...	...	...	...	...
1638	Residential, Rural and other Areas	6	22.0	16	NaN
1095	Residential, Rural and other Areas	2	18.0	104	NaN

		Type of Location	SO2	NO2	RSPM/PM10	PM 2.5
1130		Industrial Area	3	23.0	64	NaN
1294	Residential, Rural and other Areas		7	23.0	79	NaN
860		Industrial Area	11	27.0	44	NaN

[2159 rows x 11 columns]

x_test

	Stn Code	Sampling Date	State	City/Town/Village/Area	\
471	22	25 08 14	0	Chennai	
1453	18	19 08 14	0	Cuddalore	
2377	14	09 02 14	0	Thoothukudi	
1601	10	17 02 14	0	Madurai	
1094	15	12 03 14	0	Coimbatore	
...	...	...	...	...	
1413	18	24 03 14	0	Cuddalore	
1090	15	19 11 14	0	Coimbatore	
1512	17	25 03 14	0	Cuddalore	
630	24	29 01 14	0	Chennai	
2507	9	17 12 14	0	Thoothukudi	

	Location of Monitoring Station	Agency	\
471	Anna Nagar, Chennai	1	
1453	District Environmental Engineer Office, Imperi...	1	
2377	AVM Jewellery Building, Tuticorin	1	
1601	Highway (Project -I) Building, Madurai	1	
1094	Poniarajapuram, On the top of DEL, Coimbatore	1	
...	...	...	
1413	District Environmental Engineer Office, Imperi...	1	
1090	Poniarajapuram, On the top of DEL, Coimbatore	1	
1512	SIPCOT Industrial Complex, Cuddalore	1	
630	Kilpauk, Chennai	1	
2507	Raja Agencies, Tuticorin	1	

	Type of Location	SO2	NO2	RSPM/PM10	PM 2.5
471	Residential, Rural and other Areas	9	21.0	33	NaN
1453	Residential, Rural and other Areas	4	15.0	30	NaN
2377	Residential, Rural and other Areas	7	11.0	69	NaN
1601	Residential, Rural and other Areas	15	23.0	42	NaN
1094	Residential, Rural and other Areas	2	23.0	75	NaN
...		...	...	...	...

1413	Residential, Rural and other Areas	9	25.0	88	NaN
1090	Residential, Rural and other Areas	2	23.0	57	NaN
1512	Industrial Area	7	21.0	75	NaN
630	Residential, Rural and other Areas	22	25.0	71	NaN
2507	Industrial Area	14	18.0	139	NaN

[720 rows x 11 columns]

y_train

```
      RSPM/PM10
1385         56
1125        142
482          41
1765         18
1000         56
...         ...
1638         16
1095        104
1130         64
1294         79
860          44
```

[2159 rows x 1 columns]

y_test

```
      RSPM/PM10
471          33
1453         30
2377         69
1601         42
1094         75
...         ...
1413         88
1090         57
1512         75
630          71
2507        139
```

[720 rows x 1 columns]

LR=LinearRegression()

PyCaret's Regression module (pycaret.regression) is a supervised machine learning module which is used for predicting continuous values / outcomes using various techniques and algorithms. Regression can be used for predicting values / outcomes such as sales, units sold, temperature or any number which is continuous.

PyCaret's regression module has over 25 algorithms and 10 plots to analyze the performance of models. Be it hyper-parameter tuning, ensembling or advanced techniques like stacking, PyCaret's regression module has it all.

```python
dataset=pd.read_csv('/content/Air quality-analysis-2014.csv')

data = dataset.sample(frac=0.9,
random_state=786).reset_index(drop=True)
data_unseen = dataset.drop(data.index).reset_index(drop=True)

print('Data for Modeling: ' + str(data.shape))
print('Unseen Data For Predictions: ' + str(data_unseen.shape))

Data for Modeling: (2591, 11)
Unseen Data For Predictions: (288, 11)

dataset_fillna = dataset

dataset_fillna.fillna(dataset_fillna.mean(), inplace=True)
# count the number of NaN values in each column
print(dataset_fillna.isnull().sum())

dataset_fillna
```

```
Stn Code                            0
Sampling Date                       0
State                               0
City/Town/Village/Area              0
Location of Monitoring Station      0
Agency                              0
Type of Location                    0
SO2                                 0
NO2                                 0
RSPM/PM10                           0
PM 2.5                           2879
dtype: int64
```

```
<ipython-input-66-b5b95eac7e1e>:1: FutureWarning: The default value of
numeric_only in DataFrame.mean is deprecated. In a future version, it
will default to False. In addition, specifying 'numeric_only=None' is
deprecated. Select only valid columns or specify the value of
numeric_only to silence this warning.
  dataset_fillna.fillna(dataset_fillna.mean(), inplace=True)

     Stn Code Sampling Date        State City/Town/Village/Area  \
0          38       01-02-14  Tamil Nadu                Chennai
```

```
1               38     01-07-14  Tamil Nadu                         Chennai
2               38     21-01-14  Tamil Nadu                         Chennai
3               38     23-01-14  Tamil Nadu                         Chennai
4               38     28-01-14  Tamil Nadu                         Chennai
...            ...          ...         ...                             ...
2874           773     12-03-14  Tamil Nadu                          Trichy
2875           773     12-10-14  Tamil Nadu                          Trichy
2876           773     17-12-14  Tamil Nadu                          Trichy
2877           773     24-12-14  Tamil Nadu                          Trichy
2878           773     31-12-14  Tamil Nadu                          Trichy

                             Location of Monitoring Station   \
0        Kathivakkam, Municipal Kalyana Mandapam, Chennai
1        Kathivakkam, Municipal Kalyana Mandapam, Chennai
2        Kathivakkam, Municipal Kalyana Mandapam, Chennai
3        Kathivakkam, Municipal Kalyana Mandapam, Chennai
4        Kathivakkam, Municipal Kalyana Mandapam, Chennai
...                                                    ...
2874                         Central Bus Stand, Trichy
2875                         Central Bus Stand, Trichy
2876                         Central Bus Stand, Trichy
2877                         Central Bus Stand, Trichy
2878                         Central Bus Stand, Trichy

                                                  Agency   \
0        Tamilnadu State Pollution Control Board
1        Tamilnadu State Pollution Control Board
2        Tamilnadu State Pollution Control Board
3        Tamilnadu State Pollution Control Board
4        Tamilnadu State Pollution Control Board
...                                           ...
2874     Tamilnadu State Pollution Control Board
2875     Tamilnadu State Pollution Control Board
2876     Tamilnadu State Pollution Control Board
2877     Tamilnadu State Pollution Control Board
2878     Tamilnadu State Pollution Control Board

                           Type of Location  SO2   NO2  RSPM/PM10  PM
2.5
0                           Industrial Area  11.0  17.0      55.0
NaN
1                           Industrial Area  13.0  17.0      45.0
NaN
2                           Industrial Area  12.0  18.0      50.0
NaN
3                           Industrial Area  15.0  16.0      46.0
NaN
4                           Industrial Area  13.0  14.0      42.0
NaN
...                                     ...   ...   ...       ...   ..
```

```
.
2874   Residential, Rural and other Areas  15.0  18.0      102.0
NaN
2875   Residential, Rural and other Areas  12.0  14.0       91.0
NaN
2876   Residential, Rural and other Areas  19.0  22.0      100.0
NaN
2877   Residential, Rural and other Areas  15.0  17.0       95.0
NaN
2878   Residential, Rural and other Areas  14.0  16.0       94.0
NaN

[2879 rows x 11 columns]

le=LabelEncoder()
dataset['State']=le.fit_transform(dataset['State'])
dataset

      Stn Code Sampling Date  State City/Town/Village/Area  \
0           38      01-02-14      0                 Chennai
1           38      01-07-14      0                 Chennai
2           38      21-01-14      0                 Chennai
3           38      23-01-14      0                 Chennai
4           38      28-01-14      0                 Chennai
...        ...           ...    ...                    ...
2874       773      12-03-14      0                  Trichy
2875       773      12-10-14      0                  Trichy
2876       773      17-12-14      0                  Trichy
2877       773      24-12-14      0                  Trichy
2878       773      31-12-14      0                  Trichy

                        Location of Monitoring Station  \
0     Kathivakkam, Municipal Kalyana Mandapam, Chennai
1     Kathivakkam, Municipal Kalyana Mandapam, Chennai
2     Kathivakkam, Municipal Kalyana Mandapam, Chennai
3     Kathivakkam, Municipal Kalyana Mandapam, Chennai
4     Kathivakkam, Municipal Kalyana Mandapam, Chennai
...                                                ...
2874                        Central Bus Stand, Trichy
2875                        Central Bus Stand, Trichy
2876                        Central Bus Stand, Trichy
2877                        Central Bus Stand, Trichy
2878                        Central Bus Stand, Trichy

                                   Agency  \
0     Tamilnadu State Pollution Control Board
1     Tamilnadu State Pollution Control Board
2     Tamilnadu State Pollution Control Board
3     Tamilnadu State Pollution Control Board
4     Tamilnadu State Pollution Control Board
```

```
...                                                ...
2874   Tamilnadu State Pollution Control Board
2875   Tamilnadu State Pollution Control Board
2876   Tamilnadu State Pollution Control Board
2877   Tamilnadu State Pollution Control Board
2878   Tamilnadu State Pollution Control Board

                              Type of Location   SO2   NO2   RSPM/PM10   PM
2.5
0                              Industrial Area  11.0  17.0       55.0
NaN
1                              Industrial Area  13.0  17.0       45.0
NaN
2                              Industrial Area  12.0  18.0       50.0
NaN
3                              Industrial Area  15.0  16.0       46.0
NaN
4                              Industrial Area  13.0  14.0       42.0
NaN
...                                        ...   ...   ...        ...      ..
.
2874   Residential, Rural and other Areas  15.0  18.0      102.0
NaN
2875   Residential, Rural and other Areas  12.0  14.0       91.0
NaN
2876   Residential, Rural and other Areas  19.0  22.0      100.0
NaN
2877   Residential, Rural and other Areas  15.0  17.0       95.0
NaN
2878   Residential, Rural and other Areas  14.0  16.0       94.0
NaN

[2879 rows x 11 columns]

le=LabelEncoder()
dataset['Stn Code']=le.fit_transform(dataset['Stn Code'])
dataset

      Stn Code Sampling Date  State City/Town/Village/Area  \
0            0      01-02-14      0                 Chennai
1            0      01-07-14      0                 Chennai
2            0      21-01-14      0                 Chennai
3            0      23-01-14      0                 Chennai
4            0      28-01-14      0                 Chennai
...        ...           ...    ...                     ...
2874        29      12-03-14      0                   Trichy
2875        29      12-10-14      0                   Trichy
2876        29      17-12-14      0                   Trichy
2877        29      24-12-14      0                   Trichy
2878        29      31-12-14      0                   Trichy
```

```
                          Location of Monitoring Station  \
0      Kathivakkam, Municipal Kalyana Mandapam, Chennai
1      Kathivakkam, Municipal Kalyana Mandapam, Chennai
2      Kathivakkam, Municipal Kalyana Mandapam, Chennai
3      Kathivakkam, Municipal Kalyana Mandapam, Chennai
4      Kathivakkam, Municipal Kalyana Mandapam, Chennai
...                                                 ...
2874                        Central Bus Stand, Trichy
2875                        Central Bus Stand, Trichy
2876                        Central Bus Stand, Trichy
2877                        Central Bus Stand, Trichy
2878                        Central Bus Stand, Trichy

                                           Agency  \
0       Tamilnadu State Pollution Control Board
1       Tamilnadu State Pollution Control Board
2       Tamilnadu State Pollution Control Board
3       Tamilnadu State Pollution Control Board
4       Tamilnadu State Pollution Control Board
...                                         ...
2874    Tamilnadu State Pollution Control Board
2875    Tamilnadu State Pollution Control Board
2876    Tamilnadu State Pollution Control Board
2877    Tamilnadu State Pollution Control Board
2878    Tamilnadu State Pollution Control Board

                          Type of Location   SO2   NO2   RSPM/PM10   PM
2.5
0                            Industrial Area  11.0  17.0       55.0
NaN
1                            Industrial Area  13.0  17.0       45.0
NaN
2                            Industrial Area  12.0  18.0       50.0
NaN
3                            Industrial Area  15.0  16.0       46.0
NaN
4                            Industrial Area  13.0  14.0       42.0
NaN
...                                      ...   ...   ...        ...   ..
.
2874   Residential, Rural and other Areas  15.0  18.0      102.0
NaN
2875   Residential, Rural and other Areas  12.0  14.0       91.0
NaN
2876   Residential, Rural and other Areas  19.0  22.0      100.0
NaN
2877   Residential, Rural and other Areas  15.0  17.0       95.0
NaN
2878   Residential, Rural and other Areas  14.0  16.0       94.0
```

NaN

[2879 rows x 11 columns]

```python
le=LabelEncoder()
dataset['Agency']=le.fit_transform(dataset['Agency'])
dataset
```

	Stn Code	Sampling Date	State	City/Town/Village/Area
0	0	01-02-14	0	Chennai
1	0	01-07-14	0	Chennai
2	0	21-01-14	0	Chennai
3	0	23-01-14	0	Chennai
4	0	28-01-14	0	Chennai
...	...	...	...	...
2874	29	12-03-14	0	Trichy
2875	29	12-10-14	0	Trichy
2876	29	17-12-14	0	Trichy
2877	29	24-12-14	0	Trichy
2878	29	31-12-14	0	Trichy

	Location of Monitoring Station	Agency
0	Kathivakkam, Municipal Kalyana Mandapam, Chennai	1
1	Kathivakkam, Municipal Kalyana Mandapam, Chennai	1
2	Kathivakkam, Municipal Kalyana Mandapam, Chennai	1
3	Kathivakkam, Municipal Kalyana Mandapam, Chennai	1
4	Kathivakkam, Municipal Kalyana Mandapam, Chennai	1
...	...	...
2874	Central Bus Stand, Trichy	1
2875	Central Bus Stand, Trichy	1
2876	Central Bus Stand, Trichy	1
2877	Central Bus Stand, Trichy	1
2878	Central Bus Stand, Trichy	1

	Type of Location	SO2	NO2	RSPM/PM10	PM 2.5
0	Industrial Area	11.0	17.0	55.0	NaN
1	Industrial Area	13.0	17.0	45.0	NaN
2	Industrial Area	12.0	18.0	50.0	NaN
3	Industrial Area	15.0	16.0	46.0	NaN
4	Industrial Area	13.0	14.0	42.0	NaN
...	...	...	...	...	...
2874	Residential, Rural and other Areas	15.0	18.0	102.0	NaN

```
2875   Residential, Rural and other Areas  12.0  14.0       91.0
NaN
2876   Residential, Rural and other Areas  19.0  22.0      100.0
NaN
2877   Residential, Rural and other Areas  15.0  17.0       95.0
NaN
2878   Residential, Rural and other Areas  14.0  16.0       94.0
NaN

[2879 rows x 11 columns]

le=LabelEncoder()
dataset['Type of Location']=le.fit_transform(dataset['Type of
Location'])
dataset

      Stn Code Sampling Date  State City/Town/Village/Area  \
0            0      01-02-14      0                 Chennai
1            0      01-07-14      0                 Chennai
2            0      21-01-14      0                 Chennai
3            0      23-01-14      0                 Chennai
4            0      28-01-14      0                 Chennai
...        ...           ...    ...                     ...
2874        29      12-03-14      0                  Trichy
2875        29      12-10-14      0                  Trichy
2876        29      17-12-14      0                  Trichy
2877        29      24-12-14      0                  Trichy
2878        29      31-12-14      0                  Trichy

                       Location of Monitoring Station  Agency  \
0     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
1     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
2     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
3     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
4     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
...                                               ...     ...
2874                     Central Bus Stand, Trichy       1
2875                     Central Bus Stand, Trichy       1
2876                     Central Bus Stand, Trichy       1
2877                     Central Bus Stand, Trichy       1
2878                     Central Bus Stand, Trichy       1

      Type of Location   SO2   NO2  RSPM/PM10  PM 2.5
0                    0  11.0  17.0       55.0     NaN
1                    0  13.0  17.0       45.0     NaN
2                    0  12.0  18.0       50.0     NaN
3                    0  15.0  16.0       46.0     NaN
4                    0  13.0  14.0       42.0     NaN
...                ...   ...   ...        ...     ...
2874                 1  15.0  18.0      102.0     NaN
```

```
2875                       1  12.0  14.0     91.0     NaN
2876                       1  19.0  22.0    100.0     NaN
2877                       1  15.0  17.0     95.0     NaN
2878                       1  14.0  16.0     94.0     NaN

[2879 rows x 11 columns]

dataset['Sampling Date'] = dataset['Sampling Date'].str.replace('-', '
')
dataset
```

```
      Stn Code Sampling Date   State City/Town/Village/Area  \
0            0     01 02 14       0                  Chennai
1            0     01 07 14       0                  Chennai
2            0     21 01 14       0                  Chennai
3            0     23 01 14       0                  Chennai
4            0     28 01 14       0                  Chennai
...        ...          ...     ...                      ...
2874        29     12 03 14       0                   Trichy
2875        29     12 10 14       0                   Trichy
2876        29     17 12 14       0                   Trichy
2877        29     24 12 14       0                   Trichy
2878        29     31 12 14       0                   Trichy

                         Location of Monitoring Station  Agency  \
0     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
1     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
2     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
3     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
4     Kathivakkam, Municipal Kalyana Mandapam, Chennai       1
...                                              ...     ...
2874                       Central Bus Stand, Trichy       1
2875                       Central Bus Stand, Trichy       1
2876                       Central Bus Stand, Trichy       1
2877                       Central Bus Stand, Trichy       1
2878                       Central Bus Stand, Trichy       1

      Type of Location   SO2   NO2  RSPM/PM10  PM 2.5
0                    0  11.0  17.0       55.0     NaN
1                    0  13.0  17.0       45.0     NaN
2                    0  12.0  18.0       50.0     NaN
3                    0  15.0  16.0       46.0     NaN
4                    0  13.0  14.0       42.0     NaN
...                ...   ...   ...        ...     ...
2874                 1  15.0  18.0      102.0     NaN
2875                 1  12.0  14.0       91.0     NaN
2876                 1  19.0  22.0      100.0     NaN
2877                 1  15.0  17.0       95.0     NaN
2878                 1  14.0  16.0       94.0     NaN
```

```
[2879 rows x 11 columns]
```

The setup() function initializes the environment in pycaret and creates the transformation pipeline to prepare the data for modeling and deployment. setup() must be called before executing any other function in pycaret. It takes two mandatory parameters: a pandas dataframe and the name of the target column. All other parameters are optional and are used to customize the pre-processing pipeline (we will see them in later tutorials).

When setup() is executed, PyCaret's inference algorithm will automatically infer the data types for all features based on certain properties. The data type should be inferred correctly but this is not always the case. To account for this, PyCaret displays a table containing the features and their inferred data types after setup() is executed. If all of the data types are correctly identified enter can be pressed to continue or quit can be typed to end the expriment. Ensuring that the data types are correct is of fundamental importance in PyCaret as it automatically performs a few pre-processing tasks which are imperative to any machine learning experiment. These tasks are performed differently for each data type which means it is very important for them to be correctly configured.

In later tutorials we will learn how to overwrite PyCaret's infered data type using the numeric_features and categorical_features parameters in setup().