```
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
                    01-02-14 Tamil Nadu
            38
                                                         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
                    12-03-14 Tamil Nadu
2874
           773
                                                          Trichy
2875
           773
                    12-10-14 Tamil Nadu
                                                          Trichy
2876
           773
                    17-12-14 Tamil Nadu
                                                          Trichy
           773
                    24-12-14 Tamil Nadu
2877
                                                          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
                             Central Bus Stand, Trichy
2875
                             Central Bus Stand, Trichy
2876
2877
                             Central Bus Stand, Trichy
2878
                             Central Bus Stand, Trichy
                                       Agency \
      Tamilnadu State Pollution Control Board
1
      Tamilnadu State Pollution Control Board
2
      Tamilnadu State Pollution Control Board
```

3 4	Tamilnadu Sta Tamilnadu Sta						
2874 2875 2876 2877 2878	Tamilnadu Sta Tamilnadu Sta Tamilnadu Sta Tamilnadu Sta Tamilnadu Sta	te Pollution te Pollution te Pollution	Control E Control E Control E	Board Board Board			
		Type of	Location	S02	NO2 RSF	PM/PM10 P	M
2.5 0		Indust	rial Area	11.0	17.0	55.0	
NaN							
1 NaN		Indust	rial Area	13.0	17.0	45.0	
2		Indust	rial Area	12.0	18.0	50.0	
NaN 3		Indust	rial Area	15.0	16.0	46.0	
NaN		Taduat	niol Anoo	13.0	14.0	42.0	
4 NaN		Indust	rial Area	13.0	14.0	42.0	
2874 NaN	Residential,	Rural and ot	her Areas	15.0	18.0	102.0	
2875	Residential,	Rural and ot	her Areas	12.0	14.0	91.0	
NaN 2876	Residential,	Rural and ot	her Areas	19.0	22.0	100.0	
NaN 2877	Residential,	Rural and ot	her Areas	15.0	17.0	95.0	
NaN							
2878 NaN	Residential,	Rurat and ot	ner Areas	14.0	16.0	94.0	
[2879	rows x 11 col	umns1					
	escribe()						
ari .u	``	503		NOO	DCDM /DM1/	DM 2 F	
count mean	Stn Code 2879.000000 475.750261	S02 2868.000000 11.503138	2866.000		RSPM/PM10 2875.000000 62.494261	0.0	
std min	277.675577 38.000000	5.051702 2.000000			31.368745 12.000000		
25% 50%	238.000000	8.00000	17.000	0000	41.00000) NaN	
75%	366.000000 764.000000	12.000000 15.000000	25.000	0000	55.000000 78.000000) NaN	
max	773.000000	49.000000	71.000	0000	269.000000) NaN	
air.i	nfo()						

```
<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
                                      2879 non-null
     State
                                                       object
 3
     City/Town/Village/Area
                                      2879 non-null
                                                       object
4
     Location of Monitoring Station 2879 non-null
                                                       object
 5
                                      2879 non-null
     Agency
                                                       object
 6
     Type of Location
                                      2879 non-null
                                                       object
 7
     S02
                                      2868 non-null
                                                       float64
 8
     N02
                                      2866 non-null
                                                       float64
 9
     RSPM/PM10
                                      2875 non-null
                                                       float64
 10
     PM 2.5
                                                       float64
                                      0 non-null
dtypes: float64(4), int64(1), object(6)
memory usage: 247.5+ KB
air.isnull().sum()
Stn Code
                                      0
Sampling Date
                                      0
                                      0
State
City/Town/Village/Area
                                      0
                                      0
Location of Monitoring Station
                                      0
Agency
Type of Location
                                      0
S02
                                     11
N02
                                     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
                                      0
City/Town/Village/Area
                                      0
Location of Monitoring Station
                                      0
Agency
Type of Location
                                      0
                                      0
S02
```

```
N<sub>0</sub>2
                                      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
            38
1
                    01-07-14
                              Tamil Nadu
                                                          Chennai
2
            38
                    21-01-14
                               Tamil Nadu
                                                          Chennai
3
            38
                    23-01-14
                              Tamil Nadu
                                                          Chennai
4
                              Tamil Nadu
            38
                    28-01-14
                                                          Chennai
2874
           773
                    12-03-14
                               Tamil Nadu
                                                           Trichy
2875
           773
                    12-10-14
                              Tamil Nadu
                                                           Trichy
           773
                    17-12-14
2876
                              Tamil Nadu
                                                           Trichy
                    24-12-14
                              Tamil Nadu
2877
           773
                                                           Trichy
2878
           773
                    31-12-14
                              Tamil Nadu
                                                           Trichy
                        Location of Monitoring Station \
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
1
2
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
3
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
4
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
                              Central Bus Stand, Trichy
2874
2875
                              Central Bus Stand, Trichy
                              Central Bus Stand, Trichy
2876
                              Central Bus Stand, Trichy
2877
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
. . .
      Tamilnadu State Pollution Control Board
2874
      Tamilnadu State Pollution Control Board
2875
2876
      Tamilnadu State Pollution Control Board
2877
      Tamilnadu State Pollution Control Board
2878
      Tamilnadu State Pollution Control Board
```

```
Type of Location
                                            S02
                                                   N02
                                                        RSPM/PM10
                                                                   PM
2.5
                          Industrial Area
                                           11.0
                                                  17.0
                                                             55.0
0
NaN
                          Industrial Area
                                           13.0
                                                  17.0
                                                             45.0
NaN
                          Industrial Area
                                                             50.0
2
                                          12.0
                                                  18.0
NaN
                          Industrial Area
                                                             46.0
3
                                          15.0
                                                  16.0
NaN
                          Industrial Area
                                          13.0
                                                             42.0
4
                                                  14.0
NaN
2874
      Residential, Rural and other Areas
                                           15.0
                                                  18.0
                                                            102.0
NaN
      Residential, Rural and other Areas
2875
                                           12.0
                                                  14.0
                                                             91.0
NaN
      Residential, Rural and other Areas 19.0
                                                  22.0
                                                            100.0
2876
NaN
2877
      Residential, Rural and other Areas 15.0
                                                  17.0
                                                             95.0
NaN
      Residential, Rural and other Areas 14.0
                                                             94.0
2878
                                                  16.0
NaN
[2879 rows \times 11 columns]
le=LabelEncoder()
air['State']=le.fit transform(air['State'])
air
      Stn Code Sampling Date
                               State City/Town/Village/Area \
                                                     Chennai
0
            38
                     01-02-14
                                   0
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
                                                      Trichy
                                   0
2875
           773
                     12-10-14
                                   0
                                                      Trichy
2876
           773
                     17-12-14
                                   0
                                                      Trichy
           773
                     24-12-14
                                   0
2877
                                                      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, Munic	ipal Kalyana Manda	•		
2074		Control Duc (Trick:		
2874 2875		Central Bus S Central Bus S			
2876		Central Bus S			
2877		Central Bus S			
2878		Central Bus S			
2070		centrat bas s	Jeana, Trieny		
		Age	ency \		
0	Tamilnadu State Po		_		
1	Tamilnadu State Po	llution Control Bo	pard		
2	Tamilnadu State Po	llution Control Bo	pard		
3	Tamilnadu State Po				
4	Tamilnadu State Po	llution Control Bo	pard		
<u> </u>			• • •		
2874	Tamilnadu State Po				
2875	Tamilnadu State Po				
	Tamilnadu State Po				
	Tamilnadu State Po				
2878	Tamilnadu State Po	ctution control Bo	bard		
		Type of Location	S02 N02	RSPM/PM10	PM
2.5		Type of Location	302 1102	1131117111110	• • •
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				40.0	
3		Industrial Area	15.0 16.0	46.0	
NaN		Tuduatudal Ausa	12 0 14 0	42.0	
4 NaN		Industrial Area	13.0 14.0	42.0	
2874	Residential, Rural	and other Areas	15.0 18.0	102.0	
NaN	nesidentiat, narat	and benef meds	1510 1010	10210	
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					
[2070	rows x 11 columns]				
[20/9	IOWS V II COCUIIIIS]				

```
le=LabelEncoder()
air['Stn Code']=le.fit transform(air['Stn Code'])
air
      Stn Code Sampling Date
                               State City/Town/Village/Area \
0
                     01-02-14
                                                      Chennai
             0
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
            29
                     17-12-14
2876
                                    0
                                                       Trichy
2877
            29
                     24-12-14
                                    0
                                                       Trichy
2878
            29
                     31-12-14
                                    0
                                                       Trichy
                         Location of Monitoring Station \
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
0
1
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
2
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
3
4
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
. . .
2874
                              Central Bus Stand, Trichy
2875
                              Central Bus Stand, Trichy
                              Central Bus Stand, Trichy
2876
                              Central Bus Stand, Trichy
2877
                              Central Bus Stand, Trichy
2878
                                         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
      Tamilnadu State Pollution Control Board
2875
2876
      Tamilnadu State Pollution Control Board
2877
      Tamilnadu State Pollution Control Board
      Tamilnadu State Pollution Control Board
2878
                         Type of Location
                                                   N<sub>0</sub>2
                                                         RSPM/PM10
                                                                     PM
                                             S02
2.5
                          Industrial Area
                                            11.0
                                                   17.0
                                                              55.0
NaN
                          Industrial Area
                                            13.0
                                                   17.0
                                                              45.0
1
NaN
2
                          Industrial Area 12.0
                                                   18.0
                                                              50.0
```

```
NaN
                          Industrial Area 15.0
                                                             46.0
3
                                                  16.0
NaN
4
                          Industrial Area
                                          13.0
                                                  14.0
                                                             42.0
NaN
2874
      Residential, Rural and other Areas
                                           15.0
                                                            102.0
                                                  18.0
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
      Residential, Rural and other Areas 15.0
                                                             95.0
2877
                                                  17.0
NaN
2878
      Residential, Rural and other Areas 14.0
                                                             94.0
                                                 16.0
NaN
[2879 rows x 11 columns]
le=LabelEncoder()
air['S02']=le.fit transform(air['S02'])
air
      Stn Code Sampling Date
                               State City/Town/Village/Area \
                    01-02-14
0
             0
                                   0
                                                     Chennai
1
             0
                                   0
                    01-07-14
                                                     Chennai
2
             0
                    21-01-14
                                   0
                                                     Chennai
3
             0
                    23-01-14
                                   0
                                                     Chennai
4
             0
                    28-01-14
                                   0
                                                     Chennai
            29
                    12-03-14
                                   0
                                                      Trichy
2874
2875
            29
                    12-10-14
                                   0
                                                      Trichy
2876
            29
                    17-12-14
                                   0
                                                      Trichy
                    24-12-14
2877
            29
                                   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
                              Central Bus Stand, Trichy
2874
                              Central Bus Stand, Trichy
2875
                              Central Bus Stand, Trichy
2876
                              Central Bus Stand, Trichy
2877
2878
                              Central Bus Stand, Trichy
```

```
Agency
      Tamilnadu State Pollution Control Board
0
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
      Tamilnadu State Pollution Control Board
2875
2876
      Tamilnadu State Pollution Control Board
      Tamilnadu State Pollution Control Board
2877
2878 Tamilnadu State Pollution Control Board
                        Type of Location
                                           S02
                                                 N02
                                                      RSPM/PM10
                                                                  PM 2.5
                         Industrial Area
                                             9
                                                17.0
                                                            55.0
                                                                     NaN
                         Industrial Area
                                            12
                                               17.0
                                                            45.0
                                                                     NaN
                          Industrial Area
                                            11
                                                18.0
                                                            50.0
                                                                     NaN
3
                          Industrial Area
                                            14
                                                16.0
                                                            46.0
                                                                     NaN
                          Industrial Area
                                            12 14.0
                                                            42.0
                                                                     NaN
      Residential, Rural and other Areas
                                                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
                                                                     NaN
                                            14 17.0
                                                            95.0
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
                    01-02-14
                                                     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
```

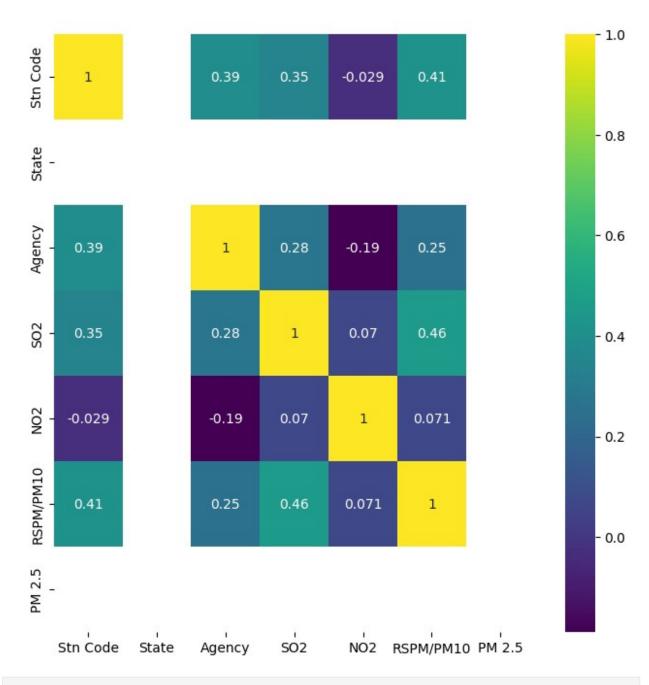
```
2874
            29
                    12-03-14
                                   0
                                                     Trichy
2875
            29
                    12-10-14
                                                     Trichy
                                   0
2876
            29
                    17-12-14
                                   0
                                                     Trichy
                    24-12-14
                                                     Trichy
2877
            29
                                   0
2878
            29
                    31-12-14
                                   0
                                                     Trichy
                        Location of Monitoring Station Agency \
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
                                                              1
1
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
                                                              1
2
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
                                                              1
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
3
                                                              1
4
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
                                                              1
                             Central Bus Stand, Trichy
2874
                                                              1
                             Central Bus Stand, Trichy
                                                              1
2875
                             Central Bus Stand, Trichy
2876
                                                              1
                             Central Bus Stand, Trichy
                                                              1
2877
                             Central Bus Stand, Trichy
2878
                                                              1
                        Type of Location SO2
                                                      RSPM/PM10
                                                                 PM 2.5
                                                 N02
                         Industrial Area
                                           9
                                               17.0
                                                           55.0
                                                                    NaN
1
                         Industrial Area
                                            12
                                                17.0
                                                           45.0
                                                                    NaN
                         Industrial Area
2
                                            11
                                                18.0
                                                           50.0
                                                                    NaN
                         Industrial Area
                                            14 16.0
                                                           46.0
                                                                    NaN
                         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
      Residential, Rural and other Areas
2877
                                            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
```

0 1 2 3 4 2874 2875 2876 2877 2878	Stn Code Sam	pling Date 01-02-14 01-07-14 21-01-14 23-01-14 28-01-14 12-03-14 12-10-14 17-12-14 24-12-14 31-12-14	State 0 0 0 0 0 0 0 0 0 0 0 0	City/	Town/\	Che Che Che Tr Tr Tr	Area \ ennai ennai ennai ennai ennai ennai ennai ennai chy richy richy richy richy	
0 1 2 3 4 2874 2875 2876 2877	Kathivakkam, Kathivakkam, Kathivakkam, Kathivakkam, Kathivakkam,	Municipal Municipal Municipal Municipal	Kalyana Kalyana Kalyana Kalyana Central Central	Mand Mand Mand Mand Mand Bus Bus Bus	dapam, dapam, dapam, dapam, dapam, Stand, Stand,	Chennai Chennai Chennai Chennai		
2878		Indu	Central of Loca ustrial	Bus ntion Area	Stand, SO2	Trichy NO2 F 17.0	1 RSPM/PM10 43	PM 2.5 NaN
1 2 3 4		Indu Indu	ustrial ustrial ustrial ustrial	Area Area	12 11 14 12	17.0 18.0 16.0 14.0	33 38 34 30	NaN NaN NaN NaN
2874 2875	Residential,	Rural and	other A	 ireas	12 14 11	18.0 14.0	91 80	 NaN NaN
2876 2877 2878	Residential, Residential, Residential,	Rural and	other A	reas	18 14 13	22.0 17.0 16.0	89 84 83	NaN NaN NaN

```
[2879 rows x 11 columns]
air['Sampling Date'] =air['Sampling Date'].str.replace('-', ' ')
air
      Stn Code Sampling Date
                               State City/Town/Village/Area \
                                                     Chennai
                     01 02 14
1
                     01 07 14
             0
                                   0
                                                     Chennai
2
             0
                     21 01 14
                                   0
                                                     Chennai
3
             0
                     23 01 14
                                                     Chennai
                                   0
4
             0
                     28 01 14
                                   0
                                                     Chennai
                     12 03 14
2874
            29
                                   0
                                                      Trichy
            29
                     12 10 14
2875
                                   0
                                                      Trichy
            29
                     17 12 14
2876
                                   0
                                                      Trichy
2877
            29
                     24 12 14
                                   0
                                                      Trichy
            29
                     31 12 14
2878
                                                      Trichy
                         Location of Monitoring Station
                                                          Agency \
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
0
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
1
                                                                1
2
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
                                                                1
3
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
                                                                1
4
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
                                                                1
. . .
                              Central Bus Stand, Trichy
2874
                                                                1
                              Central Bus Stand, Trichy
2875
                                                                1
                              Central Bus Stand, Trichy
2876
                                                                1
2877
                              Central Bus Stand, Trichy
                                                                1
                              Central Bus Stand, Trichy
2878
                         Type of Location S02
                                                  N02
                                                       RSPM/PM10
                                                                   PM 2.5
                          Industrial Area
                                              9
                                                 17.0
                                                               43
                                                                      NaN
                          Industrial Area
                                             12
                                                 17.0
                                                               33
                                                                      NaN
                          Industrial Area
                                                                      NaN
                                             11
                                                 18.0
                                                               38
3
                          Industrial Area
                                             14 16.0
                                                               34
                                                                      NaN
                          Industrial Area
                                             12
                                                 14.0
                                                               30
                                                                      NaN
      Residential, Rural and other Areas
                                             14
                                                 18.0
                                                               91
                                                                      NaN
2875
      Residential, Rural and other Areas
                                             11
                                                 14.0
                                                               80
                                                                      NaN
      Residential, Rural and other Areas
2876
                                             18
                                                 22.0
                                                               89
                                                                      NaN
```

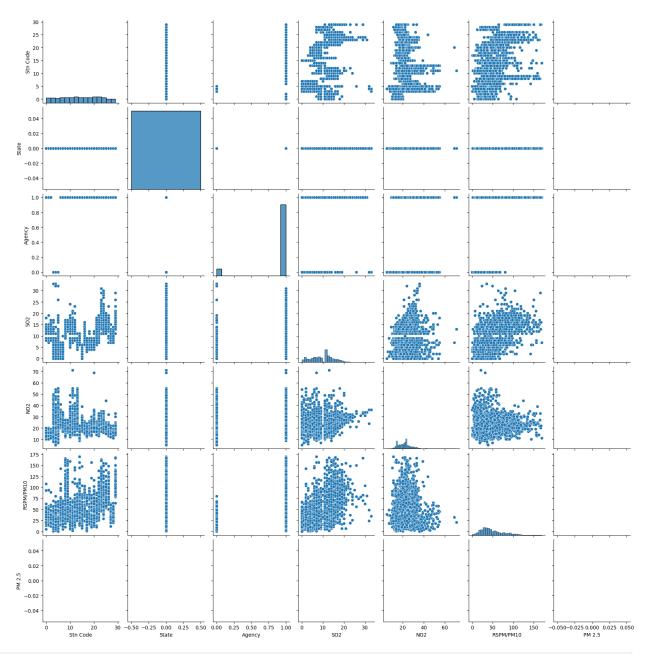
Residential, Rural and other Areas 14 17.0 84 NaN										
Stn Code Sampling Date State City/Town/Village/Area O	2877	Residential,	Rural and	other A	reas	14	17.0	84		NaN
Stn Code Sampling Date	2878	Residential,	Rural and	other A	reas	13	16.0	83		NaN
0	_	rows x 11 co	lumns]							
0 Kathivakkam, Municipal 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 5 Kathivakkam, Municipal Kalyana Mandapam, Chennai 1 6 Kathivakkam, Municipal Kalyana Mandapam, Chennai 1 7 Central Bus Stand, Trichy 1 7 Central Bus Stand, Tric	1 2 3 4 2874 2875 2876 2877	0 0 0 0 29 29 29	01 02 14 01 07 14 21 01 14 23 01 14 28 01 14 12 03 14 12 10 14 17 12 14 24 12 14	0 0 0 0 0 0	City/T	own/V	Ch Ch Ch Ch T T T	ennai ennai ennai ennai ennai richy richy richy richy		
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	1 2 3	Kathivakkam, Kathivakkam, Kathivakkam,	Municipal Municipal Municipal Municipal	Kalyana Kalyana Kalyana Kalyana	Manda Manda Manda Manda	pam, pam, pam, pam,	Chenna Chenna Chenna Chenna	i 1 i 1 i 1	\	
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 2875 2876 2877			Central Central Central	Bus S Bus S Bus S	tand, tand, tand,	Trich Trich Trich	y 1 y 1 y 1 y 1		
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			Туре	of Loca	tion	S02	N02	RSPM/PM10	PM	2.5
2 Industrial Area 11 18.0 38 NaN 3 Industrial Area 14 16.0 34 NaN 4 Industrial Area 12 14.0 30 NaN	0		Ind	ustrial	Area	9	17.0	43		NaN
3 Industrial Area 14 16.0 34 NaN 4 Industrial Area 12 14.0 30 NaN	1		Ind	ustrial	Area	12	17.0	33		NaN
4 Industrial Area 12 14.0 30 NaN	2		Ind	ustrial	Area	11	18.0	38		NaN
	3		Ind	ustrial	Area	14	16.0	34		NaN
	4		Ind	ustrial	Area	12	14.0	30		NaN
2874 Residential, Rural and other Areas 14 18.0 91 NaN										
	2874	Residential,	Rural and	other A	reas	14	18.0	91		NaN

```
2875
      Residential, Rural and other Areas
                                                                   NaN
                                           11 14.0
                                                            80
2876
      Residential, Rural and other Areas
                                           18 22.0
                                                            89
                                                                   NaN
      Residential, Rural and other Areas
2877
                                           14 17.0
                                                            84
                                                                   NaN
      Residential, Rural and other Areas
2878
                                           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',
'S02',
       'NO2', 'RSPM/PM10', 'PM 2.5'],
      dtype='object')
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>

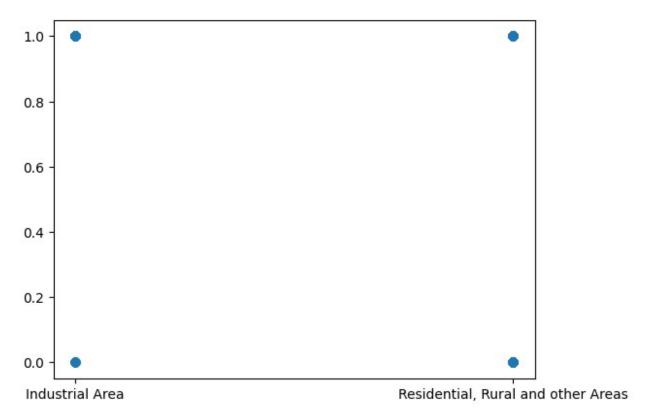


```
y partial, truncate, dropna, x jitter, y jitter, label, color, marker,
scatter kws, line kws, ax)
    757
            scatter kws["marker"] = marker
    758
            line kws = {} if line kws is None else copy.copy(line kws)
            plotter.plot(ax, scatter_kws, line kws)
--> 759
    760
            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 req:
--> 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)
                """Draw the model."""
    411
    412
                # Fit the regression model
--> 413
                grid, yhat, err bands = self.fit regression(ax)
                edges = qrid[0], qrid[-1]
    414
    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
            start = asanyarray(start) * 1.0
--> 127
    128
            stop = asanyarray(stop) * 1.0
    129
UFuncTypeError: ufunc 'multiply' did not contain a loop with signature
matching types (dtype('<U15'), dtype('float64')) -> None
```

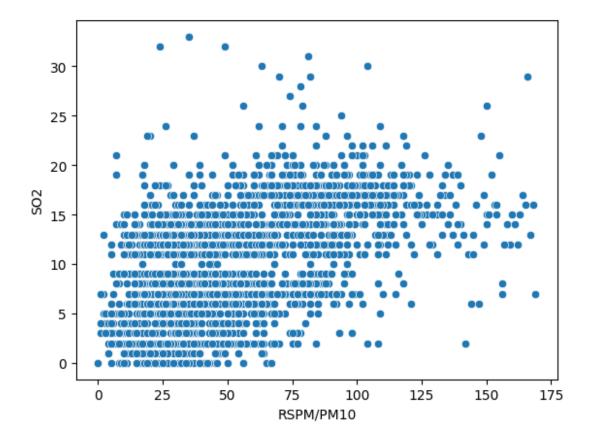


```
sns.regplot( y="Agency", x="Type of Location", data=air)
UFuncTypeError
                                          Traceback (most recent call
last)
<ipython-input-39-f1cb546c44ae> in <cell line: 1>()
----> 1 sns.regplot( y="Agency", x="Type of Location", data=air)
/usr/local/lib/python3.10/dist-packages/seaborn/regression.py 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 req:
--> 368
                    self.lineplot(ax, line kws)
    369
```

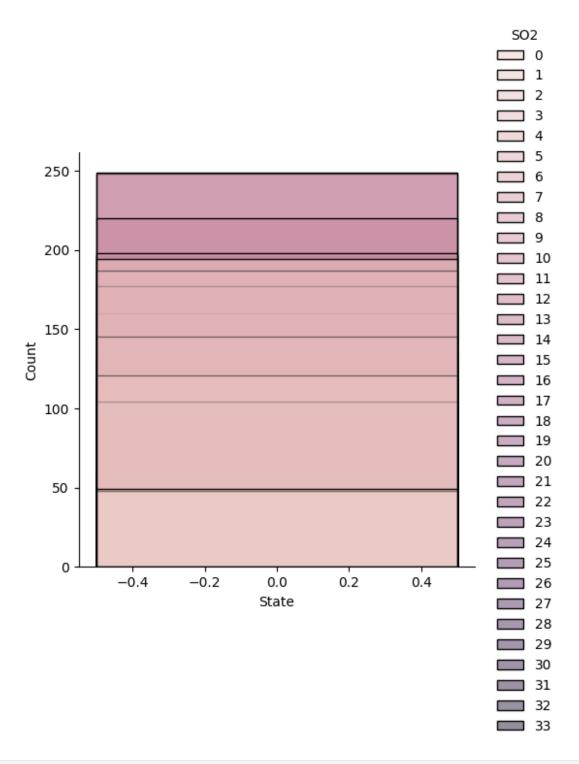
```
370
                # Label the axes
/usr/local/lib/python3.10/dist-packages/seaborn/regression.py in
lineplot(self, ax, kws)
                """Draw the model."""
    411
    412
                # Fit the regression model
--> 413
                grid, yhat, err_bands = self.fit_regression(ax)
                edges = grid[0], grid[-1]
    414
    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)
                ci = self.ci
    200
    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="S02",x="RSPM/PM10", data=air)
<Axes: xlabel='RSPM/PM10', ylabel='S02'>
```



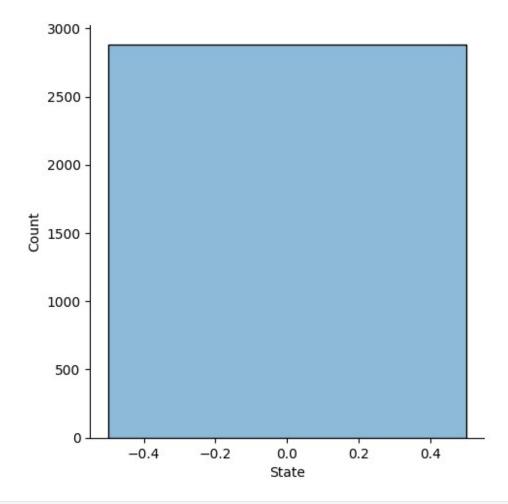
sns.displot(air, x="State", hue="S02", common_norm=False)
<seaborn.axisgrid.FacetGrid at 0x7e6e8838d240>



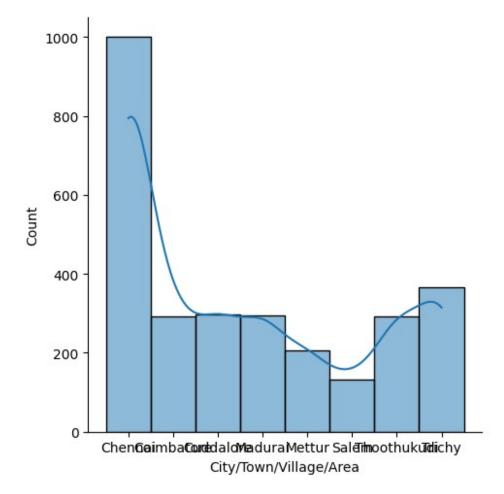
sns.scatterplot(air, x='Type of Location',y="State")
<Axes: xlabel='Type of Location', ylabel='State'>



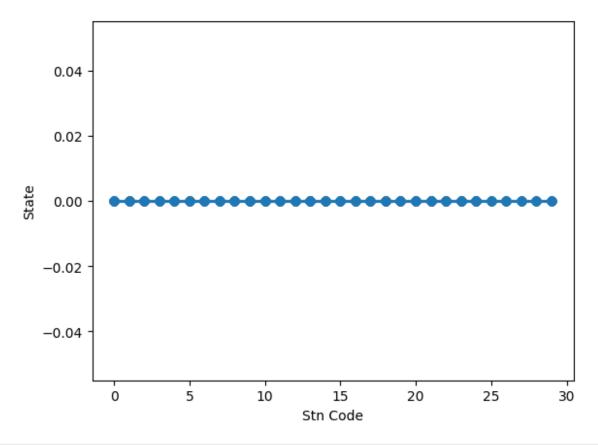
sns.displot(air, x="State",kde=True)
<seaborn.axisgrid.FacetGrid at 0x7e6e84973700>



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', 'S02',
      'RSPM/PM10', 'PM 2.5']]
Х
      Stn Code Sampling Date
                               State City/Town/Village/Area \
0
                    01 02 14
                                                     Chennai
             0
                                   0
1
             0
                     01 07 14
                                   0
                                                     Chennai
2
             0
                     21 01 14
                                   0
                                                     Chennai
3
             0
                     23 01 14
                                                     Chennai
                                   0
4
             0
                     28 01 14
                                   0
                                                     Chennai
2874
                     12 03 14
            29
                                   0
                                                      Trichy
2875
            29
                     12 10 14
                                   0
                                                      Trichy
                     17 12 14
2876
            29
                                   0
                                                      Trichy
2877
            29
                     24 12 14
                                   0
                                                      Trichy
            29
2878
                     31 12 14
                                                      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
                              Central Bus Stand, Trichy
2875
                                                               1
2876
                              Central Bus Stand, Trichy
                                                               1
                              Central Bus Stand, Trichy
2877
                                                               1
2878
                              Central Bus Stand, Trichy
                                                               1
                        Type of Location
                                                 N02
                                           S02
                                                      RSPM/PM10
                                                                  PM 2.5
                          Industrial Area
                                            9
                                                17.0
                                                              43
                                                                     NaN
                          Industrial Area
                                            12 17.0
                                                              33
                                                                     NaN
2
                          Industrial Area
                                            11
                                                18.0
                                                              38
                                                                     NaN
3
                          Industrial Area
                                            14 16.0
                                                              34
                                                                     NaN
                          Industrial Area
                                            12 14.0
                                                              30
                                                                     NaN
      Residential, Rural and other Areas
2874
                                            14
                                                18.0
                                                              91
                                                                     NaN
2875
      Residential, Rural and other Areas
                                                                     NaN
                                            11 14.0
                                                              80
2876
      Residential, Rural and other Areas
                                            18
                                                22.0
                                                              89
                                                                     NaN
      Residential, Rural and other Areas
2877
                                            14 17.0
                                                              84
                                                                     NaN
2878
      Residential, Rural and other Areas
                                            13 16.0
                                                              83
                                                                     NaN
[2879 rows x 11 columns]
y=air[['RSPM/PM10']]
У
      RSPM/PM10
             43
0
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
                     12 10 14
            16
                                   0
                                                   Cuddalore
1125
                     25 03 14
             6
                                   0
                                                  Coimbatore
482
            22
                     25 09 14
                                   0
                                                     Chennai
                    18 10 14
1765
            11
                                   0
                                                     Madurai
1000
            15
                    01 04 14
                                   0
                                                  Coimbatore
. . .
           . . .
1638
            10
                    07 10 14
                                   0
                                                     Madurai
            15
                    12 06 14
1095
                                   0
                                                  Coimbatore
                    04 11 14
1130
             6
                                   0
                                                  Coimbatore
1294
                    01 08 14
                                   0
                                                   Cuddalore
            16
                    20 05 14
860
                                                     Chennai
                          Location of Monitoring Station Agency \
                                     Eachangadu Villagae
1385
                                                                 1
1125
                                SIDCO Office, Coimbatore
                                                                1
                                     Anna Nagar, Chennai
                                                                 1
482
1765
      Fenner (I) Ltd. Employees Assiciation Building...
                                                                 1
1000
          Poniarajapuram, On the top of DEL, Coimbatore
                                                                 1
                 Highway (Project -I) Building, Madurai
1638
                                                                 1
          Poniarajapuram, On the top of DEL, Coimbatore
1095
                                                                 1
1130
                                SIDCO Office, Coimbatore
                                                                 1
1294
                                     Eachangadu Villagae
                                                                 1
                Thiruvottiyur Municipal Office, Chennai
860
                         Type of Location S02 N02
                                                       RSPM/PM10
                                                                  PM 2.5
1385
      Residential, Rural and other Areas
                                                 21.0
                                                              56
                                                                      NaN
1125
                          Industrial Area
                                           2
                                                 26.0
                                                             142
                                                                      NaN
482
      Residential, Rural and other Areas
                                            13
                                                 27.0
                                                              41
                                                                      NaN
1765
                                             11
                          Industrial Area
                                                 16.0
                                                              18
                                                                      NaN
1000 Residential, Rural and other Areas
                                            2 23.0
                                                              56
                                                                      NaN
      Residential, Rural and other Areas
                                                 22.0
                                                              16
                                                                      NaN
      Residential, Rural and other Areas
1095
                                                 18.0
                                                             104
                                                                      NaN
```

1130		Indu	ustrial	Area	3	23.0	64	NaN
1294	Residential, Rural	and	other	Areas	7	23.0	79	NaN
860		Indu	ustrial	Area	11	27.0	44	NaN
[2159	rows x 11 columns]							
x_tes	t							
471	Stn Code Sampling 22 25 0		State 0	_	Ōwn/\	/illage/ Cher		
1453	18 19 0	8 14	0			Cuddal	Lore	
2377 1601	14 09 0 10 17 0		0			Thoothuk Madu		
1094	15 12 0		0			Coimbat		
1413	 18 24 0	3 14				Cuddal	ore	
1090	15 19 1	1 14	0			Coimbat	tore	
1512 630	17 25 0 24 29 0		0			Cuddal Cher		
2507	9 17 1		0			Thoothuk		
		Loca	ation o	of Monit	oring	g Station	n Agency	\
471	District Facilianum			Anna N	lagar,	, Chennai	i 1	
1453 2377	District Environme AV					ιmperi Γuticorir		
1601	Highway	(Pro	oject -	I) Buil	.ding,	, Madurai	i 1	
1094	Poniarajapuram	, un	the to	ор от ре	L, C	oimbatore 		
1413 1090	District Environme							
1512	Poniarajapuram SIPCO			•		Cuddalore		
630 2507			Paia			, Chennai Γuticorir		
2307			_	_				
		Type	of Loc	ation	502	N02 R9	SPM/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	r	Areas	9	25.0	88	NaN
1090	Residential,	Rural	and othe	r	Areas	2	23.0	57	NaN
1512			Industri	.al	Area	7	21.0	75	NaN
630	Residential,	Rural	and othe	r	Areas	22	25.0	71	NaN
2507			Industri	.al	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.

```
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
                                     0
Agency
Type of Location
                                     0
S02
                                     0
N02
                                     0
RSPM/PM10
                                     0
PM 2.5
                                  2879
dtype: int64
<ipython-input-66-b5b95eac7ele>: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 \
                    01-02-14 Tamil Nadu
0
            38
                                                         Chennai
```

```
1
            38
                     01-07-14
                               Tamil Nadu
                                                           Chennai
2
                               Tamil Nadu
            38
                     21-01-14
                                                           Chennai
3
            38
                     23-01-14
                               Tamil Nadu
                                                           Chennai
                     28-01-14
4
            38
                               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
                              Central Bus Stand, Trichy
2876
2877
                              Central Bus Stand, Trichy
                              Central Bus Stand, Trichy
2878
                                         Agency \
0
      Tamilnadu State Pollution Control Board
1
      Tamilnadu State Pollution Control Board
2
      Tamilnadu State Pollution Control Board
      Tamilnadu State Pollution Control Board
3
4
      Tamilnadu State Pollution Control Board
2874
      Tamilnadu State Pollution Control Board
      Tamilnadu State Pollution Control Board
2875
2876
      Tamilnadu State Pollution Control Board
2877
      Tamilnadu State Pollution Control Board
2878
      Tamilnadu State Pollution Control Board
                                                                     PM
                         Type of Location
                                             S02
                                                    N<sub>0</sub>2
                                                         RSPM/PM10
2.5
                          Industrial Area
                                            11.0
                                                   17.0
                                                              55.0
0
NaN
                          Industrial Area
                                            13.0
                                                   17.0
                                                              45.0
1
NaN
2
                          Industrial Area
                                                              50.0
                                            12.0
                                                   18.0
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
                                                            102.0
                                                  18.0
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
                                                     Chennai
                                   0
            38
1
                    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
. . .
           . . .
           773
                    12-03-14
2874
                                   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 \
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
0
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
1
2
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
3
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
4
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
                              Central Bus Stand, Trichy
2874
                              Central Bus Stand, Trichy
2875
                              Central Bus Stand, Trichy
2876
2877
                              Central Bus Stand, Trichy
                              Central Bus Stand, Trichy
2878
                                        Agency \
      Tamilnadu State Pollution Control Board
0
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 2875 2876 2877 2878	Tamilnadu S Tamilnadu S Tamilnadu S	tate Polluti tate Polluti tate Polluti tate Polluti tate Polluti	on Conti on Conti on Conti	rol Bo rol Bo rol Bo	oard oard oard			
		Type	of Locat	tion	S02	N02	RSPM/PM10	PM
2.5		. , , , ,						
0		Indu	strial A	Area	11.0	17.0	55.0	
NaN								
1		Indu	strial A	Area	13.0	17.0	45.0	
NaN 2		Tndu	strial A	۸ د ۵۵	12.0	18.0	50.0	
∠ NaN		Illuu	Stilat F	Area	12.0	10.0	30.0	
3		Indu	strial A	Area	15.0	16.0	46.0	
NaN								
4		Indu	strial A	Area	13.0	14.0	42.0	
NaN								
2874	Pacidential	, Rural and	other A	reac	15.0	18.0	102.0	
NaN	Nestuentiat	, Nurat and	other Ai	cas	13.0	10.0	102.0	
2875	Residential	, Rural and	other A	reas	12.0	14.0	91.0	
NaN		•						
2876	Residential	, Rural and	other A	reas	19.0	22.0	100.0	
NaN	Danidantial	Dunal and	. ۸ مرمط لحم		15 0	17 0	05.0	
2877 NaN	Residential	, Rural and	otner Ai	reas	15.0	17.0	95.0	
2878	Residential	, Rural and	other A	reas	14.0	16.0	94.0	
NaN		, marae ana	0 21101 711		2.10	10.0	30	
[2879	rows x 11 c	olumns]						
le=La	belEncoder()							
	et['Stn Code	']=le.fit tr	ansform	(data:	set['S	tn Cod	e'])	
datas		_						
	Stn Code Sar	mpling Data	State (Ci+v/	Town ///:	i 11 200	/Area \	
0	0	01-02-14	0	стсу/	IOWII/ V.	_	ennai	
1	Ö	01-07-14	0				ennai	
2	0	21-01-14	0				ennai	
3	Θ	23-01-14	0				ennai	
4	0	28-01-14	0			Ch	ennai	
 2071	20	12 02 14				_	richy	
2874 2875	29 29	12-03-14 12-10-14	0 0				richy richy	
2876	29	17-12-14	0				richy	
2877	29	24-12-14	0				richy	
2878	29	31-12-14	0				richý	

```
Location of Monitoring Station \
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
0
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
                              Central Bus Stand, Trichy
2875
                              Central Bus Stand, Trichy
2876
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
. . .
      Tamilnadu State Pollution Control Board
2874
      Tamilnadu State Pollution Control Board
2875
2876
      Tamilnadu State Pollution Control Board
      Tamilnadu State Pollution Control Board
2877
2878
      Tamilnadu State Pollution Control Board
                         Type of Location
                                                        RSPM/PM10
                                                                   PM
                                            S02
                                                   N<sub>0</sub>2
2.5
                          Industrial Area 11.0
                                                             55.0
0
                                                  17.0
NaN
                          Industrial Area
                                          13.0
                                                             45.0
1
                                                  17.0
NaN
2
                          Industrial Area
                                          12.0
                                                  18.0
                                                             50.0
NaN
                          Industrial Area
                                                             46.0
3
                                          15.0
                                                  16.0
NaN
                          Industrial Area
                                                             42.0
4
                                           13.0
                                                  14.0
NaN
2874
      Residential, Rural and other Areas
                                          15.0
                                                  18.0
                                                            102.0
NaN
      Residential, Rural and other Areas 12.0
2875
                                                             91.0
                                                  14.0
NaN
2876
      Residential, Rural and other Areas 19.0
                                                  22.0
                                                            100.0
NaN
      Residential, Rural and other Areas 15.0
2877
                                                  17.0
                                                             95.0
NaN
      Residential, Rural and other Areas 14.0
2878
                                                             94.0
                                                  16.0
```

```
NaN
[2879 rows x 11 columns]
le=LabelEncoder()
dataset['Agency']=le.fit transform(dataset['Agency'])
dataset
      Stn Code Sampling Date
                               State City/Town/Village/Area \
0
             0
                    01-02-14
                                                     Chennai
                                   0
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
                    12-03-14
                                                      Trichy
2874
            29
                                   0
                    12-10-14
2875
            29
                                   0
                                                      Trichy
            29
2876
                    17-12-14
                                   0
                                                      Trichy
2877
            29
                    24-12-14
                                   0
                                                      Trichy
                                   0
2878
            29
                    31-12-14
                                                      Trichy
                         Location of Monitoring Station Agency \
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
                                                               1
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
1
                                                               1
2
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
                                                               1
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
3
                                                                1
4
                                                               1
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
. . .
2874
                              Central Bus Stand, Trichy
                                                               1
2875
                              Central Bus Stand, Trichy
                                                               1
                              Central Bus Stand, Trichy
                                                                1
2876
2877
                              Central Bus Stand, Trichy
                                                               1
                              Central Bus Stand, Trichy
2878
                         Type of Location
                                            S02
                                                   N02
                                                        RSPM/PM10
                                                                    PM
2.5
0
                          Industrial Area 11.0
                                                  17.0
                                                             55.0
NaN
                          Industrial Area 13.0
                                                             45.0
                                                  17.0
1
NaN
                                                             50.0
2
                          Industrial Area 12.0
                                                  18.0
NaN
                                                             46.0
3
                          Industrial Area 15.0
                                                  16.0
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
      Residential, Rural and other Areas 14.0
2878
                                                               94.0
                                                   16.0
NaN
[2879 rows x 11 columns]
le=LabelEncoder()
dataset['Type of Location']=le.fit transform(dataset['Type of
Location'l)
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
            29
                     12-10-14
2875
                                    0
                                                       Trichy
            29
                     17-12-14
                                    0
2876
                                                       Trichy
2877
            29
                     24-12-14
                                    0
                                                       Trichy
            29
2878
                     31-12-14
                                    0
                                                       Trichy
                         Location of Monitoring Station Agency \
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
0
                                                                 1
1
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
                                                                 1
2
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
                                                                 1
3
                                                                 1
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
4
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
                                                                 1
2874
                               Central Bus Stand, Trichy
                                                                 1
                               Central Bus Stand, Trichy
2875
                                                                 1
                               Central Bus Stand, Trichy
                                                                 1
2876
                               Central Bus Stand, Trichy
                                                                 1
2877
2878
                               Central Bus Stand, Trichy
                                                                 1
      Type of Location
                                      RSPM/PM10
                                                  PM 2.5
                          S02
                                 N02
0
                         11.0
                               17.0
                                           55.0
                                                     NaN
                      0
1
                      0
                         13.0
                                17.0
                                           45.0
                                                     NaN
2
                         12.0
                                18.0
                      0
                                           50.0
                                                     NaN
3
                      0
                         15.0
                                16.0
                                           46.0
                                                     NaN
4
                         13.0
                      0
                                14.0
                                           42.0
                                                     NaN
                          . . .
                                 . . .
                                                     . . .
                                             . . .
                         15.0
                                          102.0
2874
                                18.0
                                                     NaN
```

```
2875
                          12.0
                                14.0
                                            91.0
                                                      NaN
2876
                          19.0
                                22.0
                                           100.0
                                                      NaN
                       1
2877
                       1
                          15.0
                                17.0
                                            95.0
                                                      NaN
                          14.0
2878
                                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 \
                     01 02 14
                                                       Chennai
              0
                                     0
                     01 07 14
1
              0
                                     0
                                                       Chennai
2
              0
                                     0
                     21 01 14
                                                       Chennai
3
                     23 01 14
              0
                                     0
                                                       Chennai
4
              0
                     28 01 14
                                     0
                                                       Chennai
. . .
                     12 03 14
2874
             29
                                     0
                                                        Trichy
2875
             29
                     12 10 14
                                     0
                                                        Trichy
                     17 12 14
                                                        Trichy
             29
2876
                                     0
2877
             29
                     24 12 14
                                     0
                                                        Trichy
             29
                     31 12 14
2878
                                     0
                                                        Trichy
                          Location of Monitoring Station
                                                            Agency \
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
0
                                                                  1
      Kathivakkam, Municipal Kalyana Mandapam, Chennai
1
                                                                  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
                               Central Bus Stand, Trichy
2875
                                                                  1
                               Central Bus Stand, Trichy
                                                                  1
2876
2877
                               Central Bus Stand, Trichy
                                                                  1
2878
                               Central Bus Stand, Trichy
                                                                  1
      Type of Location
                           S02
                                 N02
                                       RSPM/PM10
                                                   PM 2.5
0
                          11.0
                                17.0
                                            55.0
                                                      NaN
                      0
1
                          13.0
                                17.0
                                            45.0
                                                      NaN
2
                                            50.0
                          12.0
                                18.0
                      0
                                                      NaN
3
                      0
                          15.0
                                16.0
                                            46.0
                                                      NaN
4
                          13.0
                      0
                                14.0
                                            42.0
                                                      NaN
2874
                      1
                          15.0
                                18.0
                                           102.0
                                                      NaN
                                14.0
2875
                      1
                          12.0
                                            91.0
                                                      NaN
2876
                      1
                          19.0
                                22.0
                                           100.0
                                                      NaN
2877
                          15.0
                                17.0
                                            95.0
                      1
                                                      NaN
2878
                          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().