

```
In [40]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
import warnings
warnings.filterwarnings("ignore")
import pandas.util.testing as tm
```

```
In [41]: data=pd.read_csv(r"C:\Users\IT\Documents\airQ\cpcb_dly_aq_tamil_nadu-2014.csv")
data.head()
```

Out[41]:

	Stn Code	Sampling Date	State	City/Town/Village/Area	Location of Monitoring Station	Agency	Type of Location	SO2	NO2	RSPM/PM10	PM 2.5
0	38	01-02-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	11.0	17.0	55.0	NaN
1	38	01-07-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	17.0	45.0	NaN
2	38	21-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	12.0	18.0	50.0	NaN
3	38	23-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	15.0	16.0	46.0	NaN
4	38	28-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	14.0	42.0	NaN

```
In [42]: data.tail()
```

Out [42]:

	Stn Code	Sampling Date	State	City/Town/Village/Area	Location of Monitoring Station	Agency	Type of Location	SO2	NO2	RSPM/PM10	PM 2.5
2874	773	12-03-14	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	15.0	18.0	102.0	NaN
2875	773	12-10-14	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	12.0	14.0	91.0	NaN
2876	773	17-12-14	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	19.0	22.0	100.0	NaN
2877	773	24-12-14	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	15.0	17.0	95.0	NaN
2878	773	31-12-14	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	14.0	16.0	94.0	NaN

```
In [6]: data.shape
```

```
Out[6]: (2879, 11)
```

```
In [7]: data.describe()
```

Out[7]:

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

```
In [8]: data.info()
```

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

```
In [9]: data.isnull().sum()
```

Out[9]:

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

```
In [10]: dataCleaned=data.dropna()
dataCleaned.isnull().sum()
```

Out[10]:

Stn Code	0.0
Sampling Date	0.0
State	0.0
City/Town/Village/Area	0.0
Location of Monitoring Station	0.0
Agency	0.0
Type of Location	0.0
SO2	0.0
NO2	0.0
RSPM/PM10	0.0
PM 2.5	0.0
dtype: float64	

```
In [14]: print(data.isnull().sum())
print(data.info())
```

```
Stn Code                                0
Sampling Date                          0
State                                  0
City/Town/Village/Area                 0
Location of Monitoring Station          0
Agency                                0
Type of Location                       0
SO2                                    11
NO2                                    13
RSPM/PM10                              4
PM 2.5                                2879
dtype: int64
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2879 entries, 0 to 2878
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Stn Code                             2879 non-null  int64
1   Sampling Date                        2879 non-null  object
2   State                               2879 non-null  object
3   City/Town/Village/Area               2879 non-null  object
4   Location of Monitoring Station        2879 non-null  object
5   Agency                               2879 non-null  object
6   Type of Location                     2879 non-null  object
7   SO2                                  2868 non-null  float64
8   NO2                                  2866 non-null  float64
9   RSPM/PM10                           2875 non-null  float64
10  PM 2.5                               0 non-null     float64
dtypes: float64(4), int64(1), object(6)
memory usage: 247.5+ KB
None
```

```
In [23]: data.columns
```

```
Out[23]: Index(['Stn Code', 'Sampling Date', 'State', 'City/Town/Village/Area',
        'Location of Monitoring Station', 'Agency', 'Type of Location', 'SO2',
        'NO2', 'RSPM/PM10', 'PM 2.5'],
        dtype='object')
```

```
In [24]: data.dtypes
```

Out[24]:

Stn Code	int64
Sampling Date	object
State	object
City/Town/Village/Area	object
Location of Monitoring Station	object
Agency	object
Type of Location	object
SO2	float64
NO2	float64
RSPM/PM10	float64
PM 2.5	float64
dtype: object	

```
In [50]: l=[]
for i in range(len(data.columns)):
    f=data.columns[i]
    count=0
    for j in range(len(data[f])):
        if data[f][j]==-200:
            count+=1
    l.append((f,count))
print("values from each column that needs to be replaced with avg \n ",l)

values from each column that needs to be replaced with avg
1
```

```
In [53]: num=data._get_numeric_data()
num[num<=0]=0
data
```

In [53]:

```
num=data._get_numeric_data()
num[num<0]=0
data
```

Out [53]:

	Stn Code	Sampling Date	State	City/Town/Village/Area	Location of Monitoring Station	Agency	Type of Location	SO2	NO2	RSPM/PM10	PM 2.5	
	0	38	01-02-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	11.0	17.0	55.0	NaN
	1	38	01-07-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	17.0	45.0	NaN
	2	38	21-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	12.0	18.0	50.0	NaN
	3	38	23-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	15.0	16.0	46.0	NaN
	4	38	28-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	14.0	42.0	NaN
	
	2874	773	12-03-14	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	15.0	18.0	102.0	NaN
	2875	773	12-10-14	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	12.0	14.0	91.0	NaN
	2876	773	17-12-14	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	19.0	22.0	100.0	NaN
	2877	773	24-12-14	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	15.0	17.0	95.0	NaN
	2878	773	31-12-14	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	14.0	16.0	94.0	NaN

2879 rows × 11 columns

```
In [107.. np.random.seed(10)
data=pd.read_csv(r"C:\Users\IT\Documents\airQ\cpcb_dly_aq_tamil_nadu-2014.csv")
print(data)
```

tn [107...

```
np.random.seed(10)
data=pd.read_csv(r"C:\Users\IT\Documents\airQ\cpcb_dly_aq_tamil_nadu-2014.csv")
print(data)
```

	Stn Code	Sampling Date	State	City/Town/Village/Area	\
0	38	01-02-14	Tamil Nadu	Chennai	
1	38	01-07-14	Tamil Nadu	Chennai	
2	38	21-01-14	Tamil Nadu	Chennai	
3	38	23-01-14	Tamil Nadu	Chennai	
4	38	28-01-14	Tamil Nadu	Chennai	
...	
2874	773	12-03-14	Tamil Nadu	Trichy	
2875	773	12-10-14	Tamil Nadu	Trichy	
2876	773	17-12-14	Tamil Nadu	Trichy	
2877	773	24-12-14	Tamil Nadu	Trichy	
2878	773	31-12-14	Tamil Nadu	Trichy	
	Location of Monitoring Station				\
0	Kathivakkam, Municipal	Kalyana Mandapam,	Chennai		
1	Kathivakkam, Municipal	Kalyana Mandapam,	Chennai		
2	Kathivakkam, Municipal	Kalyana Mandapam,	Chennai		
3	Kathivakkam, Municipal	Kalyana Mandapam,	Chennai		
4	Kathivakkam, Municipal	Kalyana Mandapam,	Chennai		
...				...	
2874		Central Bus Stand,	Trichy		
2875		Central Bus Stand,	Trichy		
2876		Central Bus Stand,	Trichy		
2877		Central Bus Stand,	Trichy		
2878		Central Bus Stand,	Trichy		
	Agency				\
0	Tamilnadu State Pollution Control Board				
1	Tamilnadu State Pollution Control Board				

0																						
1																						
2																						
3																						
4																						
...																						
2874																						
2875																						
2876																						
2877																						
2878																						

0																						
1																						
2																						
3																						
4																						
...																						
2874																						
2875																						
2876																						
2877																						
2878																						

[2879 rows x 11 columns]

```
In [ ]:
```

```
In [ ]:
```