```
import pandas as pd
import numpy as np
import numpy.random as rd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

```
df=pd.read_csv('airquality1.txt',index_col=0)
df
```

	Ozone	Solar.R	Wind	Temp	Month	Day
1	41.0	190.0	7.4	67	5	1
2	36.0	118.0	8.0	72	5	2
3	12.0	149.0	12.6	74	5	3
4	18.0	313.0	11.5	62	5	4
5	NaN	NaN	14.3	56	5	5
...	...	...	...	...	...	...
149	30.0	193.0	6.9	70	9	26
150	NaN	145.0	13.2	77	9	27
151	14.0	191.0	14.3	75	9	28
152	18.0	131.0	8.0	76	9	29
153	20.0	223.0	11.5	68	9	30

153 rows × 6 columns

```
df.isnull().sum()
```

```
Ozone      37
Solar.R     7
Wind        0
Temp        0
Month       0
Day         0
dtype: int64
```

```
mean=df.mean()
mean
```

```
Ozone       42.129310
Solar.R    185.931507
Wind         9.957516
Temp        77.882353
Month        6.993464
Day         15.803922
dtype: float64
```

```
mean_Ozone=df['Ozone'].mean()
mean_Ozone
```

```
42.12931034482759
```

```
df2=df['Ozone'].replace(np.nan,mean_Ozone)
df2
```

```
1      41.00000
2      36.00000
3      12.00000
4      18.00000
5      42.12931
         ...
149    30.00000
150    42.12931
151    14.00000
152    18.00000
153    20.00000
Name: Ozone, Length: 153, dtype: float64
```

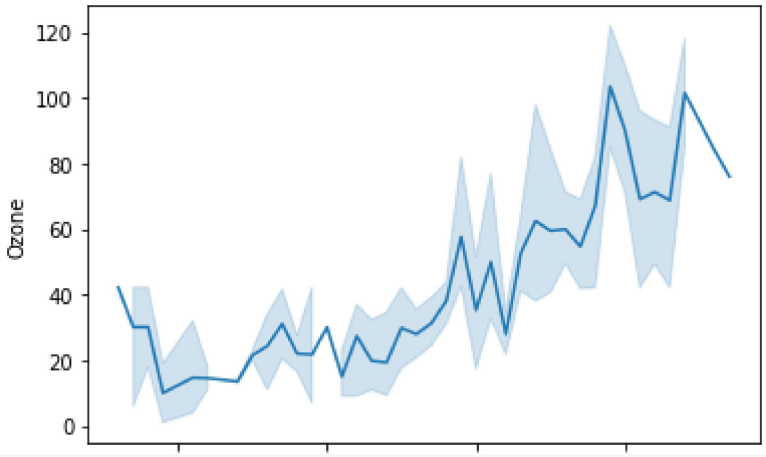
```
df2=df.replace(np.nan,mean)
df2
```

	Ozone	Solar.R	Wind	Temp	Month	Day
1	41.00000	190.000000	7.4	67	5	1
2	36.00000	118.000000	8.0	72	5	2
3	12.00000	149.000000	12.6	74	5	3
4	18.00000	313.000000	11.5	62	5	4
5	42.12931	185.931507	14.3	56	5	5
...	...	...	...	...	...	...
149	30.00000	193.000000	6.9	70	9	26
150	42.12931	145.000000	13.2	77	9	27
151	14.00000	191.000000	14.3	75	9	28
152	18.00000	131.000000	8.0	76	9	29
153	20.00000	223.000000	11.5	68	9	30

153 rows × 6 columns

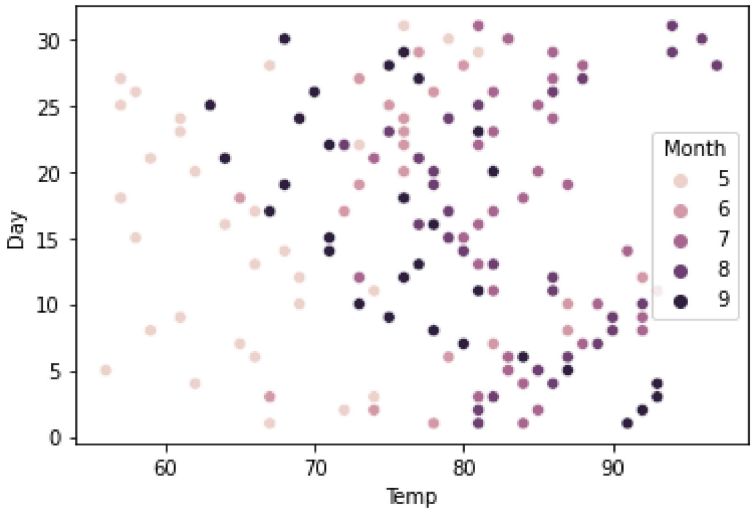
```
sns.lineplot(data=df2,x='Temp',y='Ozone')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6b86c07d90>
```




```
#sns.lineplot(data=df, x="Temp", y="Day", hue="Month")
sns.scatterplot(data=df2,x="Temp",y="Day",hue="Month")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6b86aee590>
```




```
X=df2[['Temp']]
y=df2[['Ozone']]
X,y
```

```
(     Temp
1      67
2      72
3      74
4      62
5      56
..    ...
149    70
150    77
151    75
152    76
153    68

[153 rows x 1 columns],         Ozone
1    41.00000
2    36.00000
3    12.00000
4    18.00000
```

```
5    42.12931
..        ...
149  30.00000
150  42.12931
151  14.00000
152  18.00000
153  20.00000

[153 rows x 1 columns])
```

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=0)
```

```
X_train
```

	Temp
127	93
103	86
93	81
23	61
67	83
...	...
10	69
104	86
68	88
118	86
48	72

122 rows × 1 columns

```
X_test
```

	Temp
27	57
136	77
64	81
106	80
25	57
8	59
45	80
102	92
113	77
55	76
87	82
95	82
84	82
140	67
105	82
41	87
34	67
112	78
70	92
120	97
57	78
9	61
137	71
81	85

```
y_train
```

	Ozone
127	91.00000
103	42.12931
93	39.00000
23	4.00000
67	40.00000

```
y_test
```

	Ozone
27	42.12931
136	28.00000
64	32.00000
106	65.00000
25	42.12931
8	19.00000

```
from sklearn.linear_model import LinearRegression
regressor=LinearRegression()
regressor.fit(X_train,y_train)
```

```
LinearRegression()
```

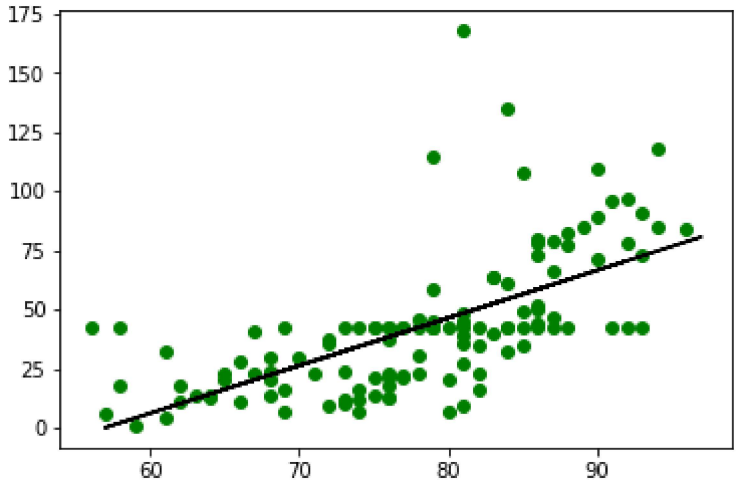
```
y_predict=regressor.predict(X_test)
y_predict
```

```
array([[7.05931271e-02],
       [4.04125964e+01],
       [4.84809970e+01],
       [4.64638969e+01],
       [7.05931271e-02],
       [4.10479345e+00],
       [4.64638969e+01],
       [7.06690988e+01],
       [4.04125964e+01],
       [3.83954962e+01],
       [5.04980972e+01],
       [5.04980972e+01],
       [5.04980972e+01],
       [2.02415948e+01],
       [5.04980972e+01],
       [6.05835980e+01],
       [2.02415948e+01],
       [4.24296965e+01],
       [7.06690988e+01],
       [8.07545996e+01],
       [4.24296965e+01],
       [8.13899378e+00],
       [2.83099954e+01],
       [5.65493977e+01],
       [5.04980972e+01],
       [6.46177983e+01],
       [2.83099954e+01],
       [5.25151974e+01],
       [4.84809970e+01],
       [4.04125964e+01],
       [1.82244946e+01]])
```

```
plt.scatter(X_train, y_train,color='g')
```

```
plt.plot(X_test, y_predict,color='k')
```

```
[<matplotlib.lines.Line2D at 0x7f6b869d0810>]
```




```
from sklearn.metrics import mean_squared_error
MSE=mean_squared_error(y_test,y_predict)
MSE
```

```
483.6166845156324
```

```
from sklearn.metrics import mean_squared_error
RMSE=mean_squared_error(y_test,y_predict,squared=False)
RMSE
```

```
21.991286558899468
```

```
import sklearn.metrics as met
r2score=met.r2_score(y_test,y_predict)
r2score
```

```
0.18150617685884907
```

```
X2=df2[["Temp","Solar.R","Wind"]]
y2=df2[["Ozone"]]
X2
```

	Temp	Solar.R	Wind
1	67	190.000000	7.4
2	72	118.000000	8.0
3	74	149.000000	12.6
4	62	313.000000	11.5

```
y2
```

	Ozone
1	41.00000
2	36.00000
3	12.00000
4	18.00000
5	42.12931
...	...
149	30.00000
150	42.12931
151	14.00000
152	18.00000
153	20.00000

153 rows × 1 columns

```
X2.shape
```

```
(153, 3)
```

```
X2_train,X2_test,y2_train,y2_test=train_test_split(X2,y2,test_size=0.2,random_state=0)
```

```
y2.shape
```

```
(153, 1)
```

```
X2_train
```

	Temp	Solar.R	Wind
127	93	189.0	4.6
103	86	137.0	11.5
93	81	83.0	6.9
23	61	25.0	9.7
67	83	314.0	10.9
...	...	...	...
10	69	194.0	8.6
104	86	192.0	11.5
68	88	276.0	5.1

```
X2_test
```

	Temp	Solar.R	Wind
27	57	185.931507	8.0
136	77	238.000000	6.3
64	81	236.000000	9.2
106	80	157.000000	9.7
25	57	66.000000	16.6
8	59	99.000000	13.8
45	80	332.000000	13.8
102	92	222.000000	8.6
113	77	259.000000	15.5
55	76	250.000000	6.3

```
y2_train
```

	Ozone
127	91.00000
103	42.12931
93	39.00000
23	4.00000
67	40.00000
...	...
10	42.12931
104	44.00000
68	77.00000
118	73.00000
48	37.00000

122 rows × 1 columns

```
y2_test
```

	Ozone
27	42.12931
136	28.00000
64	32.00000
106	65.00000
25	42.12931
8	19.00000
45	42.12931
102	42.12931
113	21.00000
55	42.12931
87	20.00000
95	16.00000
84	42.12931
140	18.00000
105	28.00000
41	39.00000
34	42.12931
112	44.00000
70	97.00000
120	76.00000
57	42.12931
9	8.00000
137	9.00000
81	63.00000
38	29.00000
99	122.00000

```
MSE2=mean_squared_error(y2_test,y2_predict)
MSE2
```

```
483.6166845156324
```

52 42.12931

```
RMSE2=mean_squared_error(y2_test,y2_predict,squared=False)
RMSE2
```

```
21.991286558899468
```

```
r2score2=met.r2_score(y_test,y_predict)
r2score2
```

```
0.18150617685884907
```

```

```