Experiment No- 1.3

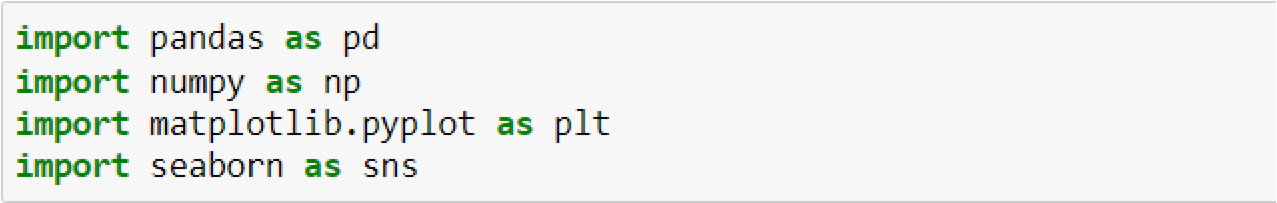
Student Name: Yash Kumar **UID: 20BCS9256**
Branch: CSE **Section/Group: 616-B**
Semester: 5th **Date of Performance: 03-09-22**
Subject Name: Machine Learning **Subject Code: 20CSP -317**

1. **Aim/Overview of the practical:** Implement linear regression.

2. Task to be done/ Which logistics used: To perform linear regression on housing.csv dataset.

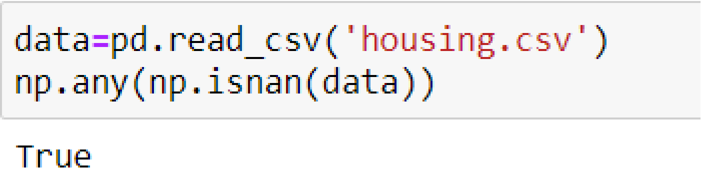
3. Steps of experiment/Code:

1. **Importing all the important libraries required for data analysis.**



```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

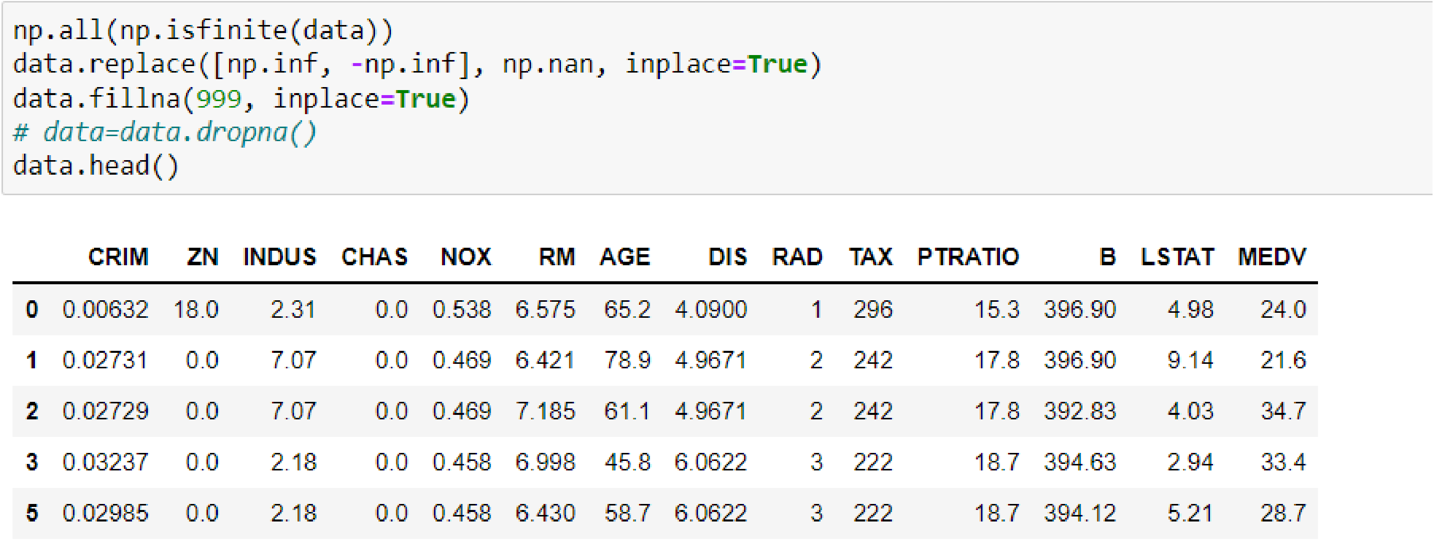
2. **Now importing the dataset using ‘read_csv’ function.**



```
data=pd.read_csv('housing.csv')
np.any(np.isnan(data))
```

```
True
```

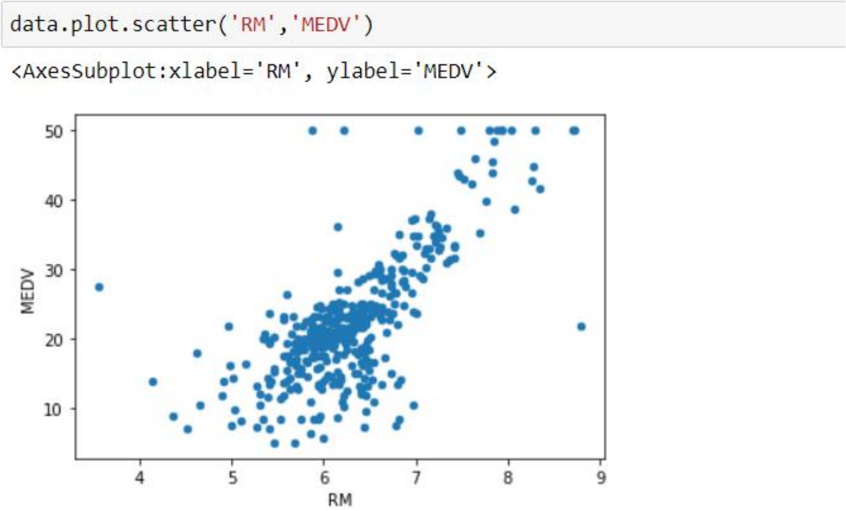
3. Replacing null values in the dataset.



```
np.all(np.isfinite(data))
data.replace([np.inf, -np.inf], np.nan, inplace=True)
data.fillna(999, inplace=True)
# data=data.dropna()
data.head()
```

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
5	0.02985	0.0	2.18	0.0	0.458	6.430	58.7	6.0622	3	222	18.7	394.12	5.21	28.7

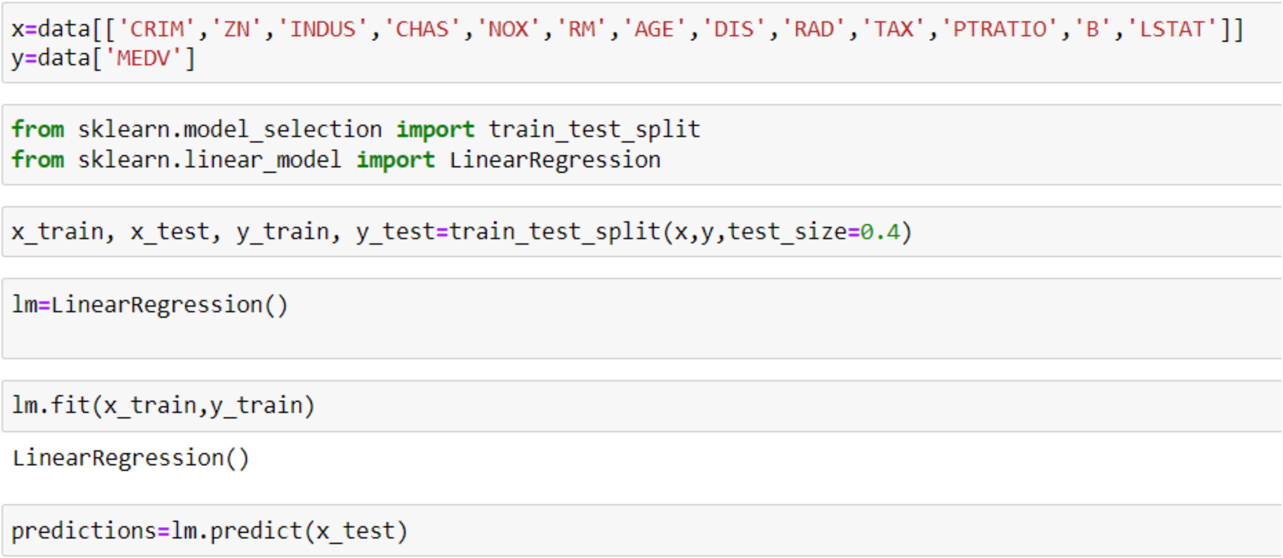
4. Scatter plot:



```
data.plot.scatter('RM','MEDV')
```

```
<AxesSubplot:xlabel='RM', ylabel='MEDV'>
```

5. Implementing linear regression using sklearn model:



```
x=data[['CRIM','ZN','INDUS','CHAS','NOX','RM','AGE','DIS','RAD','TAX','PTRATIO','B','LSTAT']]
y=data['MEDV']
```

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

```
x_train, x_test, y_train, y_test=train_test_split(x,y,test_size=0.4)
```

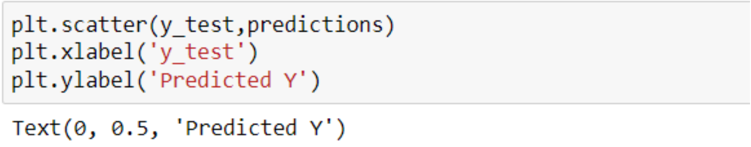
```
lm=LinearRegression()
```

```
lm.fit(x_train,y_train)
```

```
LinearRegression()
```

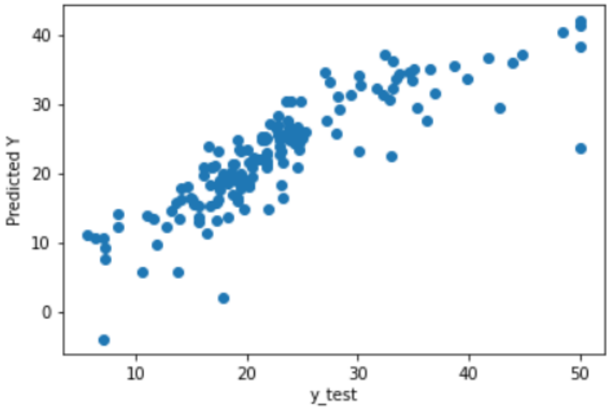
```
predictions=lm.predict(x_test)
```

6. Scatter plot representing predicted Y:

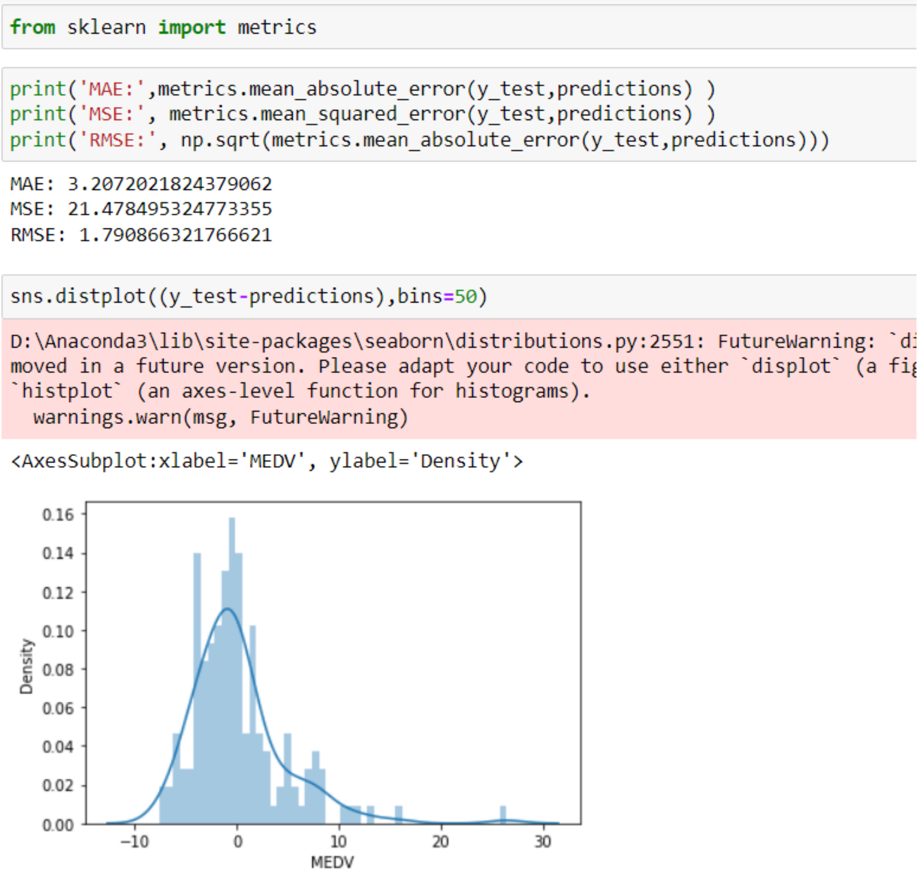


```
plt.scatter(y_test,predictions)
plt.xlabel('y_test')
plt.ylabel('Predicted Y')
```

```
Text(0, 0.5, 'Predicted Y')
```



7. Calculating mean square error, mean absolute error, root mean square error:



```
from sklearn import metrics
```

```
print('MAE:',metrics.mean_absolute_error(y_test,predictions) )
print('MSE:', metrics.mean_squared_error(y_test,predictions) )
print('RMSE:', np.sqrt(metrics.mean_absolute_error(y_test,predictions)))
```

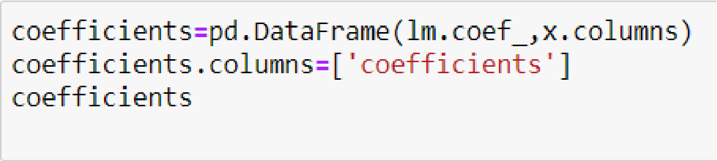
```
MAE: 3.2072021824379062
MSE: 21.478495324773355
RMSE: 1.790866321766621
```

```
sns.distplot((y_test-predictions),bins=50)
```

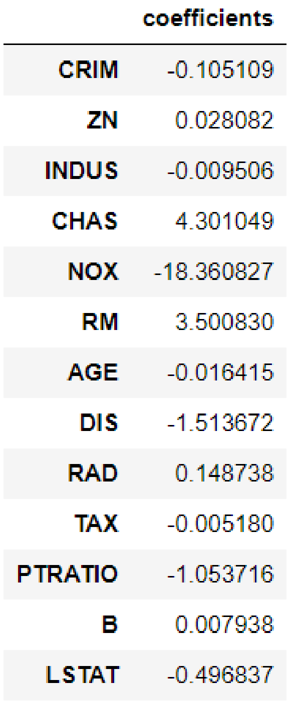
```
D:\Anaconda3\lib\site-packages\seaborn\distributions.py:2551: FutureWarning: `di
moved in a future version. Please adapt your code to use either `displot` (a fig
`histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

```
<AxesSubplot:xlabel='MEDV', ylabel='Density'>
```

8. **Finding coefficients:**



```
coefficients=pd.DataFrame(lm.coef_,x.columns)
coefficients.columns=['coefficients']
coefficients
```



	coefficients
CRIM	-0.105109
ZN	0.028082
INDUS	-0.009506
CHAS	4.301049
NOX	-18.360827
RM	3.500830
AGE	-0.016415
DIS	-1.513672
RAD	0.148738
TAX	-0.005180
PTRATIO	-1.053716
B	0.007938
LSTAT	-0.496837

Learning Outcomes (What I have learnt):

1. **I have learnt about the linear regression.**
2. **Analyzing the dataset.**
3. **Finding coefficients and analyzing regression through scatter plot.**
4. **Analyzing dataset statistically.**

Evaluation Grid:

	Parameters	Marks Obtained	Maximum Marks
1.	Student Performance (Conduct of experiment) objectives/Outcomes.		12
2.	Viva Voce		10
3.	Submission of Work Sheet (Record)		8
	Total		30