In [1]:

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

In [2]:

```python
df = pd.read_csv("C:/Users/ameya/OneDrive/Desktop/DSBDAL/housing.csv")
```

In [3]:

```python
df.head()
```

Out[3]:

	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	b	lstat	m
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	2
1	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	2
2	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	3
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	3
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	3

In [4]:

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

Out[4]:

```
crim       0
zn         0
indus      0
chas       0
nox        0
rm         0
age        0
dis        0
rad        0
tax        0
ptratio    0
b          0
lstat      0
medv       0
dtype: int64
```

In [5]:

```python
print("The shape of the data is: ")
df.shape
```

The shape of the data is:

Out[5]:

(506, 14)

In [6]:

```python
x = df.iloc[:,:-1].values
y = df.iloc[:,-1].values
```

In [10]:

```python
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)
```

In [11]:

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

Out[11]:

LinearRegression()

In [18]:

```python
y_pred = regressor.predict(x_test)
y_pred
```

Out[18]:

```
array([24.88963777, 23.72141085, 29.36499868, 12.12238621, 21.44382254,
       19.2834443 , 20.49647539, 21.36099298, 18.8967118 , 19.9280658 ,
        5.12703513, 16.3867396 , 17.07776485,  5.59375659, 39.99636726,
       32.49654668, 22.45798809, 36.85192327, 30.86401089, 23.15140009,
       24.77495789, 24.67187756, 20.59543752, 30.35369168, 22.41940736,
       10.23266565, 17.64816865, 18.27419652, 35.53362541, 20.96084724,
       18.30413012, 17.79262072, 19.96561663, 24.06127231, 29.10204874,
       19.27774123, 11.15536648, 24.57560579, 17.5862644 , 15.49454112,
       26.20577527, 20.86304693, 22.31460516, 15.60710156, 23.00363104,
       25.17247952, 20.11459464, 22.90256276, 10.0380507 , 24.28515123,
       20.94127711, 17.35258791, 24.52235405, 29.95143046, 13.42695877,
       21.72673066, 20.7897053 , 15.49668805, 13.98982601, 22.18377874,
       17.73047814, 21.58869165, 32.90522136, 31.11235671, 17.73252635,
       32.76358681, 18.7124637 , 19.78693475, 19.02958927, 22.89825374,
       22.96041622, 24.02555703, 30.72859326, 28.83142691, 25.89957059,
        5.23251817, 36.72183202, 23.77267249, 27.26856352, 19.29492159,
       28.62304496, 19.17978838, 18.97185995, 37.82397662, 39.22012647,
       23.71261106, 24.93076217, 15.88545417, 26.09845751, 16.68819641,
       15.83515991, 13.10775597, 24.71583588, 31.25165267, 22.16640989,
       20.25087212,  0.59025319, 25.44217132, 15.57178328, 17.93719475,
       25.30588844, 22.3732326 ])
```

In [14]:

```python
from sklearn.metrics import r2_score
r2_score(y_pred,y_test)
```

Out[14]:

0.35236530087887474

In [17]:

```python
y_pred = pd.DataFrame(y_pred)
y_pred
```

Out[17]:

	0
0	24.889638
1	23.721411
2	29.364999
3	12.122386
4	21.443823
...	...
97	25.442171
98	15.571783
99	17.937195
100	25.305888
101	22.373233

102 rows × 1 columns

In [ ]: