

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
```

```
df=pd.read_csv('BostonHousing.csv')
```

```
df
```

| | crim | zn | indus | chas | nox | rm | age | dis | rad | tax |
|-----|---------|------|-------|------|-------|-------|------|--------|-----|-----|
| 0 | 0.00632 | 18.0 | 2.31 | 0 | 0.538 | 6.575 | 65.2 | 4.0900 | 1 | 296 |
| 1 | 0.02731 | 0.0 | 7.07 | 0 | 0.469 | 6.421 | 78.9 | 4.9671 | 2 | 242 |
| 2 | 0.02729 | 0.0 | 7.07 | 0 | 0.469 | 7.185 | 61.1 | 4.9671 | 2 | 242 |
| 3 | 0.03237 | 0.0 | 2.18 | 0 | 0.458 | 6.998 | 45.8 | 6.0622 | 3 | 222 |
| 4 | 0.06905 | 0.0 | 2.18 | 0 | 0.458 | 7.147 | 54.2 | 6.0622 | 3 | 222 |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 501 | 0.06263 | 0.0 | 11.93 | 0 | 0.573 | 6.593 | 69.1 | 2.4786 | 1 | 273 |
| 502 | 0.04527 | 0.0 | 11.93 | 0 | 0.573 | 6.120 | 76.7 | 2.2875 | 1 | 273 |
| 503 | 0.06076 | 0.0 | 11.93 | 0 | 0.573 | 6.976 | 91.0 | 2.1675 | 1 | 273 |
| 504 | 0.10959 | 0.0 | 11.93 | 0 | 0.573 | 6.794 | 89.3 | 2.3889 | 1 | 273 |
| 505 | 0.04741 | 0.0 | 11.93 | 0 | 0.573 | 6.030 | 80.8 | 2.5050 | 1 | 273 |

| | ptratio | b | lstat | medv |
|-----|---------|--------|-------|------|
| 0 | 15.3 | 396.90 | 4.98 | 24.0 |
| 1 | 17.8 | 396.90 | 9.14 | 21.6 |
| 2 | 17.8 | 392.83 | 4.03 | 34.7 |
| 3 | 18.7 | 394.63 | 2.94 | 33.4 |
| 4 | 18.7 | 396.90 | 5.33 | 36.2 |
| .. | ... | ... | ... | ... |
| 501 | 21.0 | 391.99 | 9.67 | 22.4 |
| 502 | 21.0 | 396.90 | 9.08 | 20.6 |
| 503 | 21.0 | 396.90 | 5.64 | 23.9 |
| 504 | 21.0 | 393.45 | 6.48 | 22.0 |
| 505 | 21.0 | 396.90 | 7.88 | 11.9 |

```
[506 rows x 14 columns]
```

```
df.columns
```

```
Index(['crim', 'zn', 'indus', 'chas', 'nox', 'rm', 'age', 'dis',  
      'rad', 'tax',  
      'ptratio', 'b', 'lstat', 'medv'],  
      dtype='object')
```

```
x=df[['crim', 'zn', 'indus', 'chas', 'nox', 'rm', 'age', 'dis',  
      'rad', 'tax', 'ptratio', 'b', 'lstat', 'medv']]
```

```
y=df['medv']
```

```
y
```

```
0      24.0  
1      21.6  
2      34.7  
3      33.4  
4      36.2
```

```
...  
501     22.4  
502     20.6  
503     23.9  
504     22.0  
505     11.9
```

```
Name: medv, Length: 506, dtype: float64
```

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25,rand  
om_state=42)
```

```
model=LinearRegression()  
model.fit(x_train,y_train)
```

```
LinearRegression()
```

```
y_predict=model.predict(x_test)  
y_predict
```

```
array([23.6, 32.4, 13.6, 22.8, 16.1, 20. , 17.8, 14. , 19.6, 16.8,  
21.5,  
      18.9,  7. , 21.2, 18.5, 29.8, 18.8, 10.2, 50. , 14.1, 25.2,  
29.1,  
      12.7, 22.4, 14.2, 13.8, 20.3, 14.9, 21.7, 18.3, 23.1, 23.8, 15.  
,  
      20.8, 19.1, 19.4, 34.7, 19.5, 24.4, 23.4, 19.7, 28.2, 50. ,  
17.4,  
      22.6, 15.1, 13.1, 24.2, 19.9, 24. , 18.9, 35.4, 15.2, 26.5,  
43.5,  
      21.2, 18.4, 28.5, 23.9, 18.5, 25. , 35.4, 31.5, 20.2, 24.1, 20.  
,  
      13.1, 24.8, 30.8, 12.7, 20. , 23.7, 10.8, 20.6, 20.8,  5. ,  
20.1,  
      48.5, 10.9,  7. , 20.9, 17.2, 20.9,  9.7, 19.4, 29. , 16.4, 25.]
```

```
,  
    25. , 17.1, 23.2, 10.4, 19.6, 17.2, 27.5, 23. , 50. , 17.9,  
9.6,  
    17.2, 22.5, 21.4, 12. , 19.9, 19.4, 13.4, 18.2, 24.6, 21.1,  
24.7,  
    8.7, 27.5, 20.7, 36.2, 31.6, 11.7, 39.8, 13.9, 21.8, 23.7,  
17.6,  
    24.4, 8.8, 19.2, 25.3, 20.4, 23.1])
```

```
model.score(x_train,y_train)
```

```
1.0
```

```
model.score(x_test,y_test)
```

```
1.0
```

```
np.sqrt(mean_squared_error(y_test,y_predict))
```

```
2.4126462142329238e-14
```