Assign4_7323

February 5, 2026

```python
[66]: # Dushyant Krishna Sharma
      import pandas as pd
      import numpy as np
      import matplotlib.pyplot as plt
      import seaborn as sns

      from sklearn.model_selection import train_test_split
      from sklearn.linear_model import LinearRegression
      from sklearn.metrics import mean_squared_error, r2_score
```

```python
[67]: # data description
      data = pd.read_csv("BostonHousing.csv")
      print(data.head())
```

```
      CRIM    ZN  INDUS  CHAS    NOX     RM   AGE     DIS  RAD  TAX  PTRATIO  \
0  0.00632  18.0   2.31   0.0  0.538  6.575  65.2  4.0900    1  296     15.3
1  0.02731   0.0   7.07   0.0  0.469  6.421  78.9  4.9671    2  242     17.8
2  0.02729   0.0   7.07   0.0  0.469  7.185  61.1  4.9671    2  242     17.8
3  0.03237   0.0   2.18   0.0  0.458  6.998  45.8  6.0622    3  222     18.7
4  0.06905   0.0   2.18   0.0  0.458  7.147  54.2  6.0622    3  222     18.7

        B  LSTAT  MEDV
0  396.90   4.98  24.0
1  396.90   9.14  21.6
2  392.83   4.03  34.7
3  394.63   2.94  33.4
4  396.90    NaN  36.2
```

```python
[68]: # null values
      print(data.isnull().sum())
```

```
CRIM      20
ZN        20
INDUS     20
CHAS      20
NOX        0
RM         0
AGE       20
```

```
DIS          0
RAD          0
TAX          0
PTRATIO      0
B            0
LSTAT       20
MEDV         0
dtype: int64
```

[69]:
```python
# removing null values
data = data.dropna()
```

[70]:
```python
X = data.drop("MEDV", axis=1)
y = data["MEDV"]
```

[71]:
```python
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)
```

[72]:
```python
model = LinearRegression()
model.fit(X_train, y_train)
```
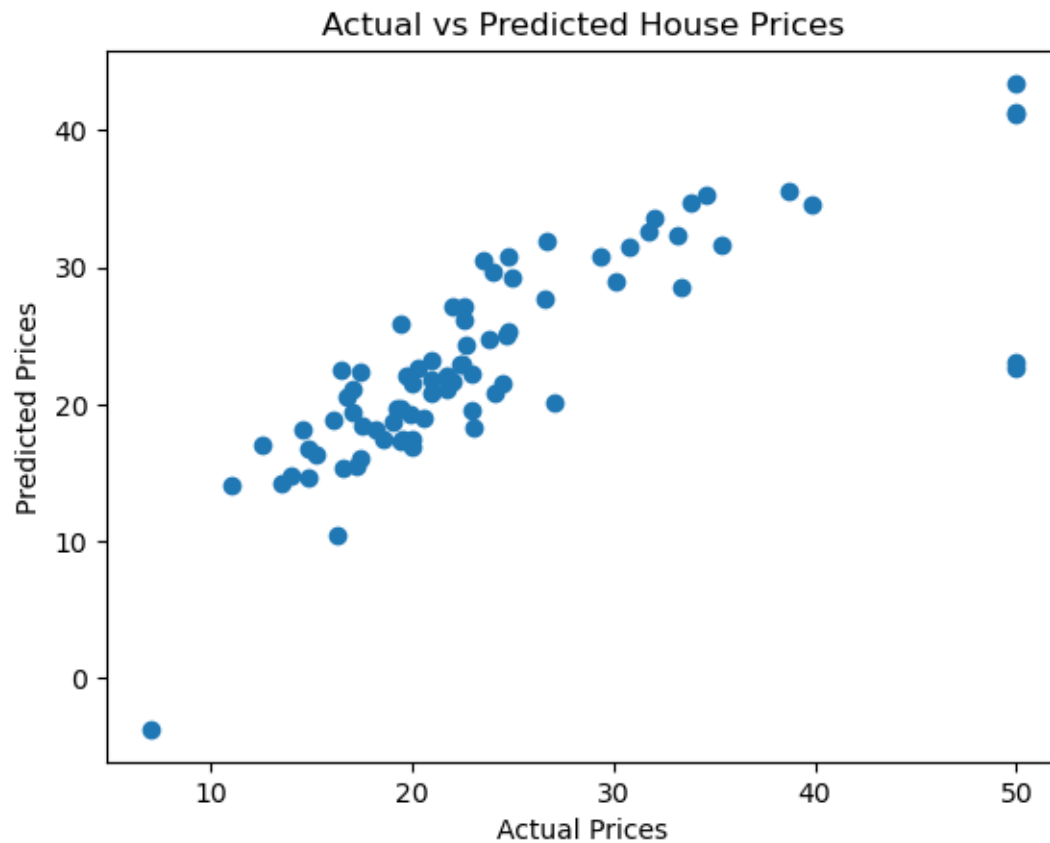
[72]: LinearRegression()

[73]:
```python
y_pred = model.predict(X_test)
```

[74]:
```python
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)

print("Mean Squared Error:", mse)
print("Root Mean Squared Error:", rmse)
print("R^2 Score:", r2)
```
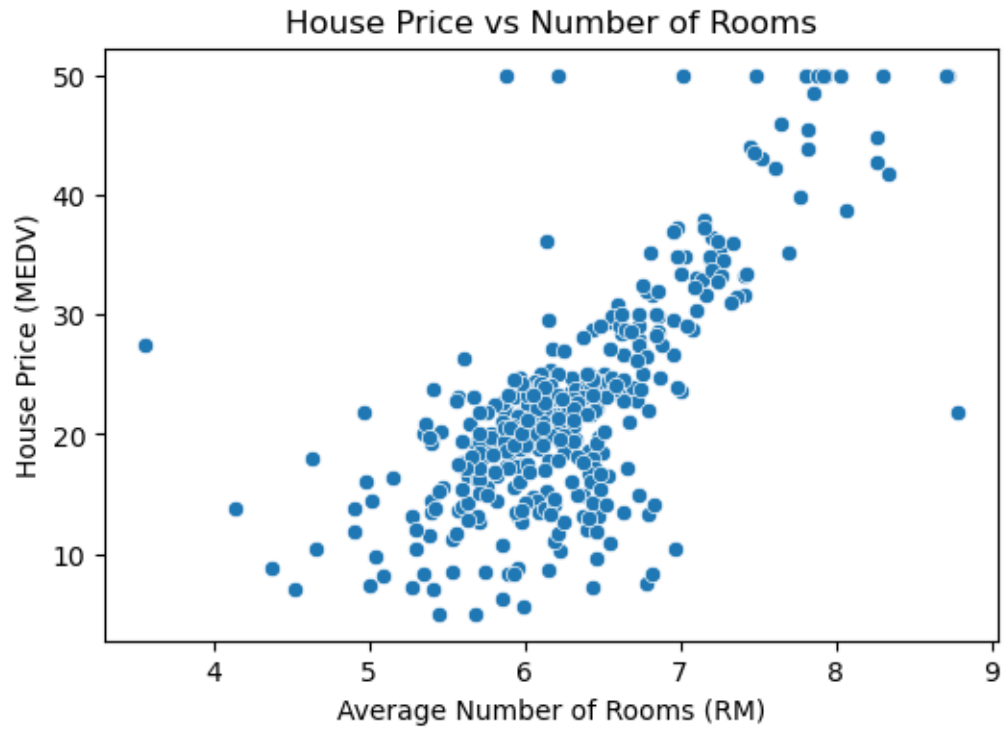
```
Mean Squared Error: 31.454047664950867
Root Mean Squared Error: 5.608390826694486
R^2 Score: 0.6270849941673191
```
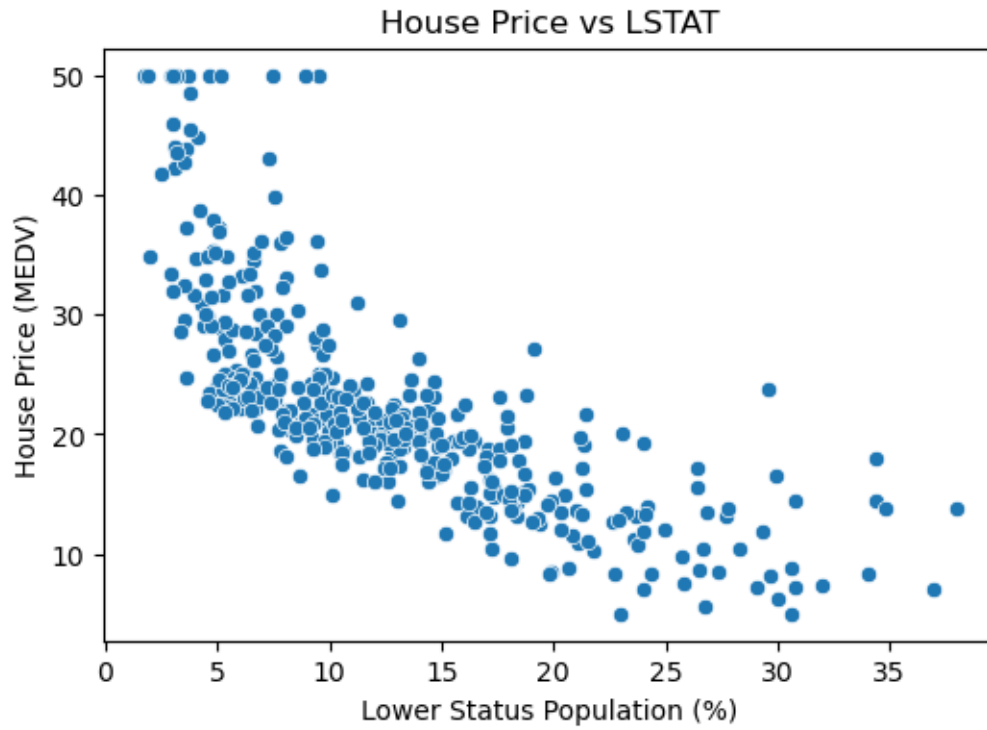
[75]:
```python
plt.scatter(y_test, y_pred)
plt.xlabel("Actual Prices")
plt.ylabel("Predicted Prices")
plt.title("Actual vs Predicted House Prices")
plt.show()
```

Actual vs Predicted House Prices


```
[76]: plt.figure(figsize=(6, 4))
      sns.scatterplot(x=data["RM"], y=data["MEDV"])
      plt.xlabel("Average Number of Rooms (RM)")
      plt.ylabel("House Price (MEDV)")
      plt.title("House Price vs Number of Rooms")
      plt.show()
```

House Price vs Number of Rooms

```
[77]: plt.figure(figsize=(6, 4))
      sns.scatterplot(x=data["LSTAT"], y=data["MEDV"])
      plt.xlabel("Lower Status Population (%)")
      plt.ylabel("House Price (MEDV)")
      plt.title("House Price vs LSTAT")
      plt.show()
```

House Price vs LSTAT


```
[78]: plt.figure(figsize=(6, 4))
      sns.scatterplot(x=data["CRIM"], y=data["MEDV"])
      plt.xscale("log")
      plt.xlabel("Crime Rate (log scale)")
      plt.ylabel("House Price (MEDV)")
      plt.title("House Price vs Crime Rate")
      plt.show()
```

House Price vs Crime Rate