devl4b

November 13, 2024

[33]:

**import pandas as pd**

**from sklearn.model\_selection import** train\_test\_split

**from sklearn.linear\_model import** LinearRegression

**from sklearn.metrics import** mean\_absolute\_error,r2\_score,mean\_squared\_error

[34]:

data=pd.DataFrame(pd.read\_csv("/content/archive (15).zip")) data

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| [34]: |  | price |  | area | bedrooms | bathrooms | stories | mainroad | guestroom | basement | \ |
|  | 0 | 13300000 |  | 7420 | 4 | 2 | 3 | yes | no | no |  |
|  | 1 | 12250000 |  | 8960 | 4 | 4 | 4 | yes | no | no |  |
|  | 2 | 12250000 |  | 9960 | 3 | 2 | 2 | yes | no | yes |  |
|  | 3 | 12215000 |  | 7500 | 4 | 2 | 2 | yes | no | yes |  |
|  | 4 | 11410000 |  | 7420 | 4 | 1 | 2 | yes | yes | yes |  |
|  | .. | … | … |  | … | … … | … | … | … |  |  |
| 540 | | 1820000 | 3000 | | 2 | 1 | 1 | yes | no | yes | |
| 541 | | 1767150 | 2400 | | 3 | 1 | 1 | no | no | no | |
| 542 | | 1750000 | 3620 | | 2 | 1 | 1 | yes | no | no | |
| 543 | | 1750000 | 2910 | | 3 | 1 | 1 | no | no | no | |
| 544 | | 1750000 | 3850 | | 3 | 1 | 2 | yes | no | no | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | hotwaterheating | airconditioning | parking |  | prefarea | furnishingstatus |
| 0 | no | yes | 2 |  | yes | furnished |
| 1 | no | yes | 3 |  | no | furnished |
| 2 | no | no | 2 |  | yes | semi-furnished |
| 3 | no | yes | 3 |  | yes | furnished |
| 4 | no | yes | 2 |  | no | furnished |
| .. | … | … | … | … |  | … |
| 540 | no | no | 2 |  | no | unfurnished |
| 541 | no | no | 0 |  | no | semi-furnished |
| 542 | no | no | 0 |  | no | unfurnished |
| 543 | no | no | 0 |  | no | furnished |
| 544 | no | no | 0 |  | no | unfurnished |

[545 rows x 13 columns]

[35]:

**from sklearn.preprocessing import** LabelEncoder label\_encoder=LabelEncoder() data["mainroad"]=label\_encoder.fit\_transform(data["mainroad"]) data["guestroom"]=label\_encoder.fit\_transform(data["guestroom"]) data["basement"]=label\_encoder.fit\_transform(data["basement"])

data["hotwaterheating"]=label\_encoder.fit\_transform(data["hotwaterheating"]) data["airconditioning"]=label\_encoder.fit\_transform(data["airconditioning"]) data["prefarea"]=label\_encoder.fit\_transform(data["prefarea"]) data["furnishingstatus"]=label\_encoder.fit\_transform(data["furnishingstatus"])

[66]:

q1=data['area'].quantile(0.25) q3=data['area'].quantile(0.75) iqr=q3-q1

low\_b=q1-1.5\*iqr upp\_b=q3+1.5\*iqr

data=data.loc[(data['area']>low\_b) & (data['area']<upp\_b)] data

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| [66]: |  | price |  | area | bedrooms | bathrooms | stories | mainroad | guestroom | \ |
|  | 0 | 13300000 |  | 7420 | 4 | 2 | 3 | 1 | 0 |  |
|  | 1 | 12250000 |  | 8960 | 4 | 4 | 4 | 1 | 0 |  |
|  | 2 | 12250000 |  | 9960 | 3 | 2 | 2 | 1 | 0 |  |
|  | 3 | 12215000 |  | 7500 | 4 | 2 | 2 | 1 | 0 |  |
|  | 4 | 11410000 |  | 7420 | 4 | 1 | 2 | 1 | 1 |  |
|  | .. | … | … |  | … | … … | … | … |  |  |
| 540 | | 1820000 | 3000 | | 2 | 1 | 1 | 1 | 0 | |
| 541 | | 1767150 | 2400 | | 3 | 1 | 1 | 0 | 0 | |
| 542 | | 1750000 | 3620 | | 2 | 1 | 1 | 1 | 0 | |
| 543 | | 1750000 | 2910 | | 3 | 1 | 1 | 0 | 0 | |
| 544 | | 1750000 | 3850 | | 3 | 1 | 2 | 1 | 0 | |

basement hotwaterheating airconditioning parking prefarea \ 0 0 0 1 2 1

1 0 0 1 3 0

2 1 0 0 2 1

3 1 0 1 3 1

4 1 0 1 2 0

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| .. … | … | … |  | … | … |
| 540 1 |  | 0 | 0 |  | 2 0 |
| 541 0 |  | 0 | 0 |  | 0 0 |
| 542 0 |  | 0 | 0 |  | 0 0 |
| 543 0 |  | 0 | 0 |  | 0 0 |
| 544 0 |  | 0 | 0 |  | 0 0 |

furnishingstatus

0 0

1 0

|  |  |  |
| --- | --- | --- |
| 2 |  | 1 |
| 3 |  | 0 |
| 4 |  | 0 |
| .. | … |  |
| 540 |  | 2 |
| 541 |  | 1 |
| 542 |  | 2 |
| 543 |  | 0 |
| 544 |  | 2 |

[529 rows x 13 columns]

[67]:

x=data[["area"]].values y=data["price"].values

[68]:

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.2,random\_state=0)

[69]:

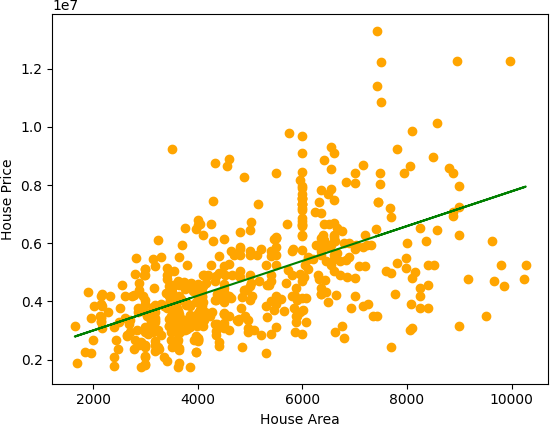
model=LinearRegression() model.fit(x\_train,y\_train) y\_pred=model.predict(x\_test) mae=mean\_absolute\_error(y\_test,y\_pred) mse=mean\_squared\_error(y\_test,y\_pred) r2=r2\_score(y\_test,y\_pred) print("MAE:",mae,"MSE:",mse,"R2:",r2)

MAE: 1136248.8927377404 MSE: 2270218857637.408 R2: 0.2686574976577394

[70]:

**import matplotlib.pyplot as plt** plt.scatter(x,y,color="orange") plt.plot(x,model.predict(x),color="green") plt.xlabel("House Area")

plt.ylabel("House Price") plt.show()



[64]:

new\_data=[[3456]] predicted\_price=model.predict(new\_data) print(predicted\_price)

[3879738.36970015]

[ ]: