import pandas as pd  
import numpy as np  
  
from sklearn.model\_selection import train\_test\_split  
from sklearn import metrics  
from sklearn.linear\_model import LinearRegression

data = pd.read\_csv("../data/Housing.csv")  
  
columns = ['mainroad', 'guestroom', 'basement', 'hotwaterheating', 'airconditioning', 'prefarea']  
  
for col in columns:  
 data[col].replace(['yes', 'no'], [1, 0], inplace=True)  
  
data['furnishingstatus'].replace(['furnished', 'semi-furnished', 'unfurnished'], [2, 1, 0], inplace=True)

y = data.price  
x = data.drop(['price'], axis=1)

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size=0.3, random\_state=42)

lm = LinearRegression()  
  
lm.fit(x\_train, y\_train)

LinearRegression()

predictions = lm.predict(x\_test)

mae = metrics.mean\_absolute\_error(y\_test, predictions)  
mse = metrics.mean\_squared\_error(y\_test, predictions)  
rmse = np.sqrt(mse)  
  
print('MAE:', mae)  
print('MSE:', mse)  
print('RMSE:', rmse)

MAE: 925543.5483156563  
MSE: 1535047758428.0498  
RMSE: 1238970.4429194627

new\_house\_features = [[2500, 5, 4, 2, 1, 1, 1, 0, 0, 1, 1, 2]]  
new\_house\_price = lm.predict(new\_house\_features)  
  
print(f"Estimated price for a new house with the given features: ${new\_house\_price[0]:.2f}")

Estimated price for a new house with the given features: $8457936.18

/opt/homebrew/lib/python3.11/site-packages/sklearn/base.py:464: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names  
 warnings.warn(