32. Scenario: You work as a data scientist for a real estate company. The company has collected

data on various houses, including features such as the size of the house, number of bedrooms,

location, and other relevant attributes. The marketing team wants to build a predictive model to

estimate the price of houses based on their features. They believe that linear regression modeling

can be an effective approach for this task.

Question:Your task is write a Python program to perform bivariate analysis and build a linear

regression model to predict house prices based on a selected feature (e.g., house size) from the

dataset. Additionally, you need to evaluate the model's performance to ensure its accuracy and

reliability.

Code:

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

# Load the house price dataset

file\_path =r"C:\Users\vara prasad\Downloads\house\_price\_data.csv"

df = pd.read\_csv(file\_path)

# Perform bivariate analysis (House Size vs. Price)

plt.figure(figsize=(10, 6))

sns.scatterplot(data=df, x='House\_Size\_sqft', y='Price', hue='Location', alpha=0.7)

plt.title('House Size vs. Price')

plt.xlabel('House Size (sqft)')

plt.ylabel('Price (USD)')

plt.show()

# Prepare data for linear regression

X = df[['House\_Size\_sqft']]

y = df['Price']

# Split data into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Train the linear regression model

model = LinearRegression()

model.fit(X\_train, y\_train)

# Make predictions

y\_pred = model.predict(X\_test)

# Evaluate the model

mse = mean\_squared\_error(y\_test, y\_pred)

r2 = r2\_score(y\_test, y\_pred)

print(f"Mean Squared Error (MSE): {mse:.2f}")

print(f"R-Squared (R2): {r2:.2f}")

# Plot the regression line

plt.figure(figsize=(10, 6))

sns.scatterplot(x=X\_test['House\_Size\_sqft'], y=y\_test, alpha=0.7, label='Actual Prices')

plt.plot(X\_test['House\_Size\_sqft'], y\_pred, color='red', linewidth=2, label='Regression Line')

plt.title('House Size vs. Price (Regression Line)')

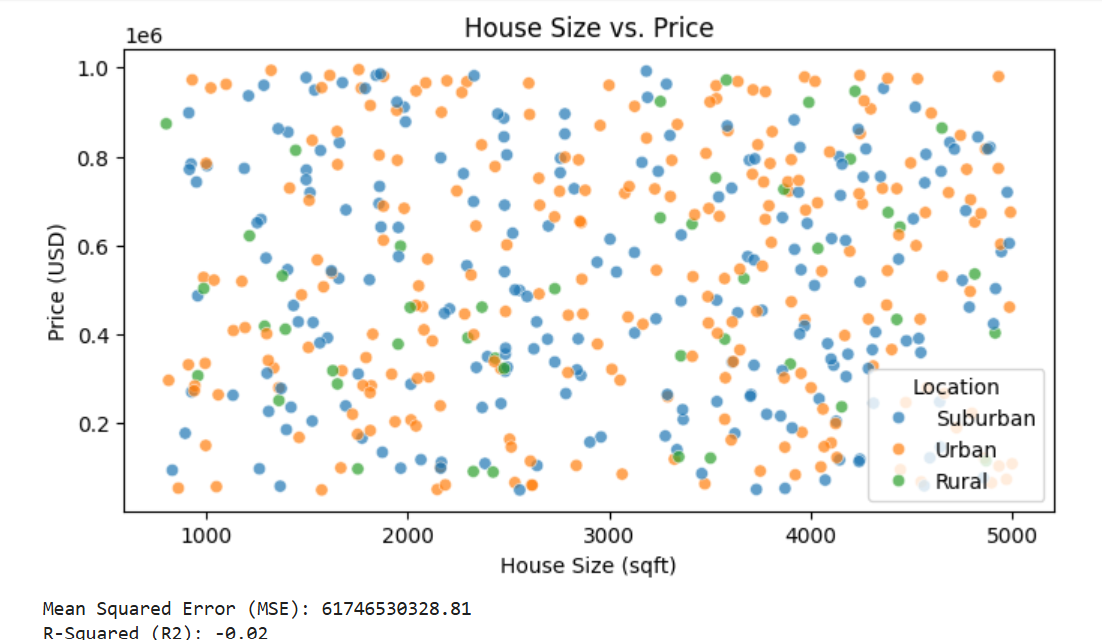
plt.xlabel('House Size (sqft)')

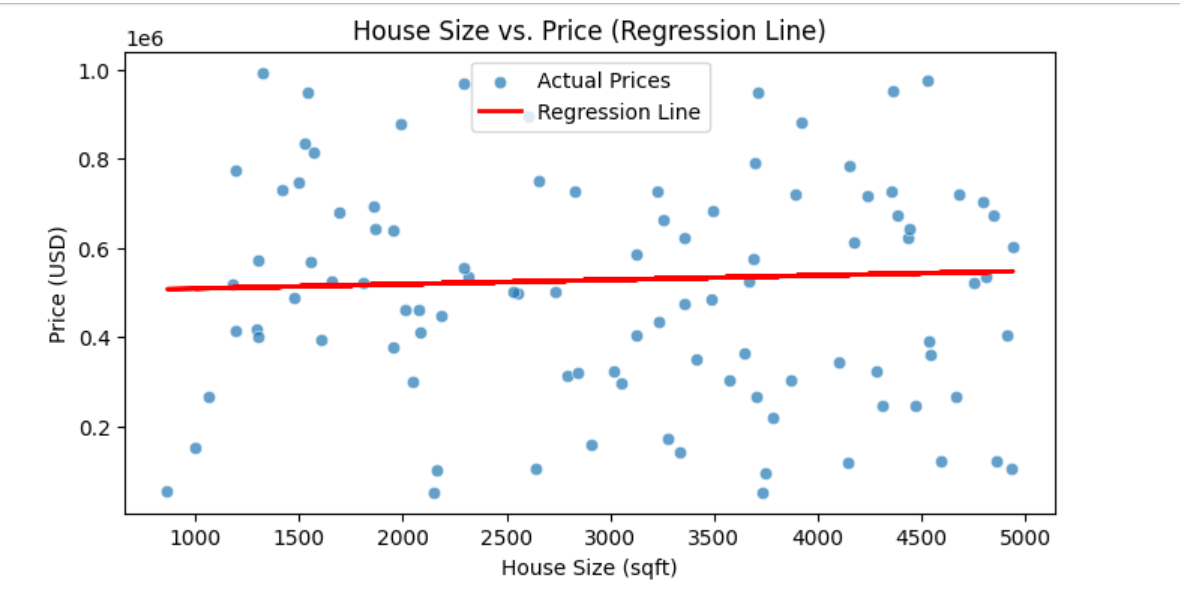
plt.ylabel('Price (USD)')

plt.legend()

plt.show()

output;





Dataset:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| HouseID | House\_Size\_sqft | Bedrooms | Bathrooms | Location | Lot\_Size\_acres | Year\_Built | Price |
| 1 | 1660 | 1 | 4 | Suburban | 2.32 | 1958 | 525593.1 |
| 2 | 4572 | 3 | 4 | Suburban | 1.4 | 1999 | 804789.8 |
| 3 | 3892 | 4 | 2 | Urban | 2.12 | 1955 | 721732.2 |
| 4 | 1266 | 2 | 4 | Suburban | 1.35 | 2009 | 97714.71 |
| 5 | 4244 | 4 | 4 | Suburban | 1.6 | 1993 | 119256.7 |
| 6 | 3971 | 4 | 2 | Urban | 0.31 | 1960 | 432729.6 |
| 7 | 3719 | 5 | 3 | Suburban | 1.91 | 2015 | 330526 |
| 8 | 930 | 2 | 4 | Suburban | 0.41 | 1996 | 270765.1 |
| 9 | 2485 | 4 | 4 | Suburban | 2.08 | 2022 | 316954.3 |
| 10 | 1569 | 4 | 4 | Suburban | 1.98 | 1952 | 813308.6 |
| 11 | 3191 | 2 | 1 | Suburban | 1.8 | 1979 | 932766.7 |
| 12 | 3233 | 2 | 2 | Suburban | 0.19 | 1970 | 434847.6 |
| 13 | 1984 | 4 | 1 | Suburban | 0.83 | 1977 | 910805.5 |
| 14 | 4185 | 2 | 2 | Suburban | 0.73 | 2011 | 355420.9 |
| 15 | 4917 | 4 | 3 | Suburban | 0.96 | 1982 | 502615.1 |
| 16 | 3704 | 4 | 3 | Suburban | 0.31 | 1984 | 264727.6 |
| 17 | 1274 | 5 | 4 | Suburban | 2.35 | 2001 | 658452.2 |
| 18 | 1882 | 1 | 2 | Urban | 1.43 | 1983 | 980032.1 |
| 19 | 3358 | 4 | 3 | Suburban | 0.83 | 1990 | 623318.4 |
| 20 | 2847 | 3 | 2 | Suburban | 1.05 | 2000 | 389923.4 |
| 21 | 3547 | 1 | 3 | Urban | 1.17 | 1996 | 665426.6 |
| 22 | 1775 | 1 | 1 | Suburban | 1.54 | 2007 | 166774.6 |
| 23 | 2606 | 1 | 3 | Urban | 1.34 | 2006 | 894226.1 |
| 24 | 989 | 5 | 1 | Urban | 2.31 | 1972 | 527929.8 |
| 25 | 3534 | 4 | 3 | Suburban | 1.29 | 1997 | 476882.3 |
| 26 | 3805 | 5 | 4 | Urban | 2.48 | 1971 | 606571.6 |
| 27 | 2699 | 4 | 1 | Suburban | 2.14 | 1966 | 643544.7 |
| 28 | 2067 | 5 | 3 | Suburban | 0.6 | 2018 | 118187 |
| 29 | 2328 | 5 | 4 | Suburban | 2.33 | 1960 | 698486.4 |
| 30 | 4002 | 3 | 1 | Urban | 0.38 | 1964 | 279835.1 |
| 31 | 4356 | 5 | 4 | Urban | 2.06 | 1984 | 728255 |
| 32 | 4690 | 2 | 2 | Suburban | 1.01 | 1982 | 831408.1 |
| 33 | 1446 | 3 | 2 | Rural | 2.21 | 1965 | 813760.6 |
| 34 | 3688 | 5 | 1 | Suburban | 2.18 | 1983 | 574875.9 |
| 35 | 3235 | 1 | 3 | Urban | 2.03 | 2015 | 544161.4 |
| 36 | 1400 | 2 | 1 | Suburban | 2 | 1989 | 185732.2 |
| 37 | 3163 | 2 | 4 | Suburban | 0.83 | 1977 | 786578.8 |
| 38 | 2861 | 2 | 3 | Suburban | 0.29 | 1996 | 307838.9 |
| 39 | 1041 | 3 | 2 | Urban | 1.07 | 1988 | 521860.7 |
| 40 | 2841 | 5 | 3 | Suburban | 0.52 | 2019 | 320060.4 |
| 41 | 3624 | 5 | 2 | Suburban | 1.77 | 2022 | 177136.9 |
| 42 | 3412 | 1 | 3 | Rural | 0.93 | 1969 | 648079.8 |
| 43 | 2163 | 1 | 3 | Urban | 2.44 | 1950 | 101615.4 |
| 44 | 2278 | 2 | 3 | Suburban | 1.64 | 2021 | 761213 |
| 45 | 3356 | 1 | 4 | Rural | 2.07 | 2012 | 351707.5 |
| 46 | 1575 | 3 | 4 | Urban | 0.42 | 2016 | 50127.96 |
| 47 | 4814 | 5 | 1 | Rural | 2.17 | 1953 | 535572.7 |
| 48 | 834 | 2 | 1 | Suburban | 2.31 | 2007 | 94509.31 |
| 49 | 3952 | 1 | 2 | Urban | 1.27 | 1974 | 312361.1 |
| 50 | 2755 | 3 | 4 | Urban | 1.56 | 1996 | 721627.7 |
| 51 | 2385 | 3 | 2 | Suburban | 1.94 | 1987 | 109555.1 |
| 52 | 4743 | 1 | 1 | Urban | 0.52 | 1964 | 847371.6 |
| 53 | 3873 | 5 | 2 | Suburban | 1.31 | 2012 | 53628.93 |
| 54 | 1821 | 1 | 1 | Urban | 1.06 | 1970 | 284482.7 |
| 55 | 4261 | 2 | 4 | Suburban | 0.45 | 2006 | 753858.9 |
| 56 | 3413 | 1 | 3 | Urban | 0.98 | 2002 | 350456.6 |
| 57 | 4643 | 3 | 4 | Suburban | 0.26 | 1995 | 146797.8 |
| 58 | 2300 | 1 | 2 | Rural | 0.16 | 1963 | 392222.2 |
| 59 | 961 | 5 | 2 | Rural | 0.42 | 1966 | 306873.6 |
| 60 | 2781 | 4 | 4 | Suburban | 2.41 | 1988 | 850576.3 |
| 61 | 1795 | 1 | 4 | Urban | 1.42 | 1958 | 347680.5 |
| 62 | 4142 | 5 | 1 | Suburban | 2.42 | 1955 | 799485.9 |
| 63 | 4598 | 5 | 2 | Urban | 1.14 | 1972 | 897278.5 |
| 64 | 2075 | 3 | 2 | Urban | 0.85 | 2008 | 462126.6 |
| 65 | 1816 | 5 | 4 | Urban | 1.31 | 1995 | 914446 |
| 66 | 1137 | 5 | 3 | Urban | 1.15 | 1987 | 408452.2 |
| 67 | 1678 | 5 | 2 | Suburban | 0.35 | 1963 | 965874.5 |
| 68 | 1876 | 5 | 3 | Suburban | 1.64 | 2021 | 134825.1 |
| 69 | 4793 | 2 | 4 | Urban | 0.62 | 2011 | 702668.3 |
| 70 | 1179 | 2 | 4 | Urban | 1.59 | 1984 | 519123.3 |
| 71 | 1292 | 3 | 1 | Rural | 1.66 | 1968 | 418266.4 |
| 72 | 2862 | 1 | 2 | Urban | 0.46 | 1979 | 651076.3 |
| 73 | 4684 | 5 | 3 | Urban | 0.25 | 1974 | 718664.2 |
| 74 | 864 | 1 | 1 | Urban | 1.97 | 1989 | 54145.11 |
| 75 | 3368 | 1 | 2 | Suburban | 1.2 | 1963 | 208605.4 |
| 76 | 2827 | 3 | 1 | Suburban | 0.24 | 1975 | 727393.8 |
| 77 | 3495 | 5 | 4 | Urban | 2.49 | 1971 | 683066.1 |
| 78 | 2295 | 5 | 2 | Urban | 0.24 | 1997 | 967745.1 |
| 79 | 1191 | 4 | 1 | Suburban | 1.77 | 2000 | 772985.9 |
| 80 | 4361 | 1 | 1 | Suburban | 2.46 | 1952 | 953234.1 |
| 81 | 3078 | 1 | 3 | Urban | 0.67 | 1991 | 717413.5 |
| 82 | 3899 | 2 | 1 | Rural | 0.44 | 1968 | 333149.6 |
| 83 | 1000 | 4 | 2 | Urban | 0.39 | 2022 | 150096.7 |
| 84 | 3904 | 2 | 2 | Urban | 0.83 | 1977 | 792732.7 |
| 85 | 3254 | 2 | 3 | Rural | 0.34 | 1994 | 661941.8 |
| 86 | 4445 | 2 | 2 | Urban | 1.76 | 2020 | 95776.71 |
| 87 | 1604 | 3 | 4 | Suburban | 0.25 | 2012 | 392047.9 |
| 88 | 3531 | 3 | 2 | Urban | 1.32 | 2022 | 958959.8 |
| 89 | 3573 | 2 | 2 | Urban | 2.49 | 1968 | 525381.1 |
| 90 | 2370 | 4 | 1 | Rural | 2.05 | 1996 | 461092.7 |
| 91 | 3490 | 1 | 1 | Urban | 1.58 | 1963 | 484814.6 |
| 92 | 4640 | 4 | 2 | Suburban | 0.84 | 1987 | 248438.6 |
| 93 | 1828 | 5 | 3 | Urban | 1.6 | 2002 | 400271.8 |
| 94 | 1302 | 3 | 1 | Urban | 1.36 | 2010 | 401322.6 |
| 95 | 1670 | 1 | 3 | Urban | 1.12 | 1985 | 99737.27 |
| 96 | 1006 | 1 | 2 | Suburban | 0.41 | 1989 | 779189.3 |
| 97 | 2284 | 5 | 2 | Urban | 2.23 | 1979 | 445679.2 |
| 98 | 1663 | 5 | 4 | Suburban | 1.18 | 1958 | 831070.9 |
| 99 | 3590 | 2 | 1 | Urban | 0.57 | 2013 | 857830.6 |
| 100 | 1363 | 3 | 2 | Rural | 0.98 | 1982 | 251394.4 |
| 101 | 4991 | 3 | 4 | Urban | 1.09 | 2001 | 674485.8 |
| 102 | 2557 | 4 | 1 | Suburban | 2.09 | 1963 | 498667.7 |
| 103 | 2478 | 2 | 2 | Suburban | 1.86 | 2002 | 886148.9 |
| 104 | 4042 | 2 | 2 | Suburban | 1.95 | 2008 | 254953.7 |
| 105 | 1859 | 2 | 1 | Suburban | 0.13 | 1990 | 693922.5 |
| 106 | 2522 | 2 | 1 | Suburban | 1.1 | 2017 | 627364.5 |
| 107 | 4114 | 3 | 1 | Suburban | 1.26 | 2017 | 330536.4 |
| 108 | 3957 | 3 | 2 | Urban | 0.15 | 1957 | 179770.9 |
| 109 | 3425 | 2 | 2 | Urban | 0.72 | 1951 | 669057.8 |
| 110 | 3529 | 4 | 4 | Rural | 1.92 | 1990 | 751667.6 |
| 111 | 2397 | 1 | 3 | Suburban | 0.43 | 2004 | 349846 |
| 112 | 3860 | 1 | 1 | Suburban | 1.38 | 1965 | 662591 |
| 113 | 3493 | 4 | 2 | Urban | 0.62 | 2011 | 425374.4 |
| 114 | 4427 | 2 | 2 | Urban | 0.13 | 1997 | 727546.5 |
| 115 | 2163 | 3 | 1 | Urban | 0.68 | 1979 | 239254.5 |
| 116 | 2781 | 1 | 3 | Suburban | 2.44 | 1961 | 895704.3 |
| 117 | 2463 | 5 | 1 | Urban | 2.02 | 1984 | 323039.1 |
| 118 | 2329 | 5 | 3 | Urban | 2.4 | 1960 | 399396.7 |
| 119 | 2838 | 4 | 2 | Urban | 1.27 | 2013 | 105187.4 |
| 120 | 4102 | 2 | 3 | Urban | 0.36 | 1999 | 155936.6 |
| 121 | 3037 | 1 | 4 | Suburban | 1.42 | 1970 | 540068.8 |
| 122 | 2106 | 2 | 2 | Urban | 1.19 | 2018 | 304213 |
| 123 | 4829 | 1 | 4 | Suburban | 2.13 | 1997 | 843706.2 |
| 124 | 3475 | 4 | 1 | Urban | 0.34 | 1969 | 63958.72 |
| 125 | 2082 | 4 | 1 | Urban | 1.27 | 2016 | 410139.4 |
| 126 | 1509 | 5 | 1 | Urban | 0.46 | 1993 | 370478.3 |
| 127 | 4548 | 2 | 3 | Urban | 0.88 | 1982 | 68361.25 |
| 128 | 1463 | 5 | 2 | Urban | 1.87 | 2022 | 168150 |
| 129 | 2798 | 2 | 1 | Urban | 1.24 | 2010 | 442960.5 |
| 130 | 4245 | 5 | 1 | Suburban | 1 | 1958 | 518221.4 |
| 131 | 4543 | 3 | 2 | Urban | 1.05 | 2017 | 434075.3 |
| 132 | 2295 | 3 | 1 | Suburban | 1.2 | 1984 | 554390.8 |
| 133 | 4104 | 3 | 4 | Suburban | 1.98 | 2014 | 615375.1 |
| 134 | 4563 | 3 | 2 | Suburban | 2.24 | 1957 | 59427.63 |
| 135 | 2653 | 1 | 2 | Urban | 2.39 | 1994 | 490890.4 |
| 136 | 2091 | 3 | 2 | Urban | 1.99 | 1996 | 965324.3 |
| 137 | 4381 | 4 | 4 | Urban | 0.86 | 1989 | 543080.7 |
| 138 | 4257 | 4 | 2 | Urban | 1.75 | 2003 | 693665.2 |
| 139 | 2436 | 1 | 4 | Rural | 1.15 | 1995 | 346270.7 |
| 140 | 4496 | 3 | 2 | Urban | 0.71 | 1996 | 785291.8 |
| 141 | 3799 | 3 | 3 | Urban | 2.12 | 1964 | 784275.4 |
| 142 | 3952 | 3 | 1 | Suburban | 0.19 | 1968 | 545231.3 |
| 143 | 1498 | 5 | 3 | Suburban | 2.26 | 1975 | 977213.1 |
| 144 | 2960 | 2 | 1 | Suburban | 1.21 | 2001 | 169272.7 |
| 145 | 4897 | 5 | 4 | Urban | 1.63 | 1971 | 66100.28 |