EXPERINMENT-32

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 matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
data = pd.read_csv("housing.csv")   # columns: area, bedrooms, age, location, price
# Example: house size (area) vs price
plt.scatter(data["area"], data["price"])
plt.xlabel("House Size (sq ft)")
plt.ylabel("House Price")
plt.title("Bivariate Analysis: Area vs Price")
plt.grid(True)
plt.show()
X = data[["area"]]   # selected feature
y = data["price"
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=0
)
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
print("Mean Squared Error (MSE):", mean_squared_error(y_test, y_pred))
print("R² Score:", r2_score(y_test, y_pred))
size = float(input("Enter house size (sq ft): "))
prediction = model.predict([[size]])[0]
print("\nPredicted House Price:", round(prediction, 2))
```

Output:

Bivariate Analysis: Area vs Price