CHANDIGARH UNIVERSITY
Discover. Learn. Empower.

NAAC GRADE A+
Accredited University

Experiment 1[c]

Student Name: Rahul Saxena

UID: 24MCI10204

Branch: MCA (AI-ML)

Section/Group: MAM - 3(B)

Semester: II

Subject Name: Machine Learning Lab (24CAP-672)

Aim: Write a program to implement Linear Regression on JupyterNotebook.

a. **Download different dataset from Kaggle or UCI ML repository.**
b. **Import all the necessary modules for the datasets.**
c. **Find out the best fit line along with the MSE, RMSE.**

Steps:

- Download datasets from Kaggle or UCI ML repository.
- Import necessary libraries like Pandas, NumPy, and Scikit-learn.
- Perform exploratory data analysis (EDA).
- Split the dataset into training and testing sets.
- Train a linear regression model.
- Evaluate the model using MSE and RMSE.

Source Code:

```
import pandas as pd
import numpy as np
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("USA_Housing.csv")
print("First 5 rows of the dataset:")
print(data.head())

print("\nMissing values in the dataset:")
print(data.isnull().sum())
X = data[['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',
       'Avg. Area Number of Bedrooms', 'Area Population']]
y = data['Price']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
```

```
y_pred = model.predict(X_test)

mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)

print("\nModel Evaluation Metrics:")
print(f"Mean Squared Error (MSE): {mse}")
print(f"Root Mean Squared Error (RMSE): {rmse}")
print(f"R-squared (R2 Score): {r2}")

plt.scatter(y_test, y_pred, color='blue', alpha=0.5)
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], color='red', linestyle='--')
plt.xlabel("Actual Prices")
plt.ylabel("Predicted Prices")
plt.title("Actual vs Predicted Prices")
plt.show()
```

Output:

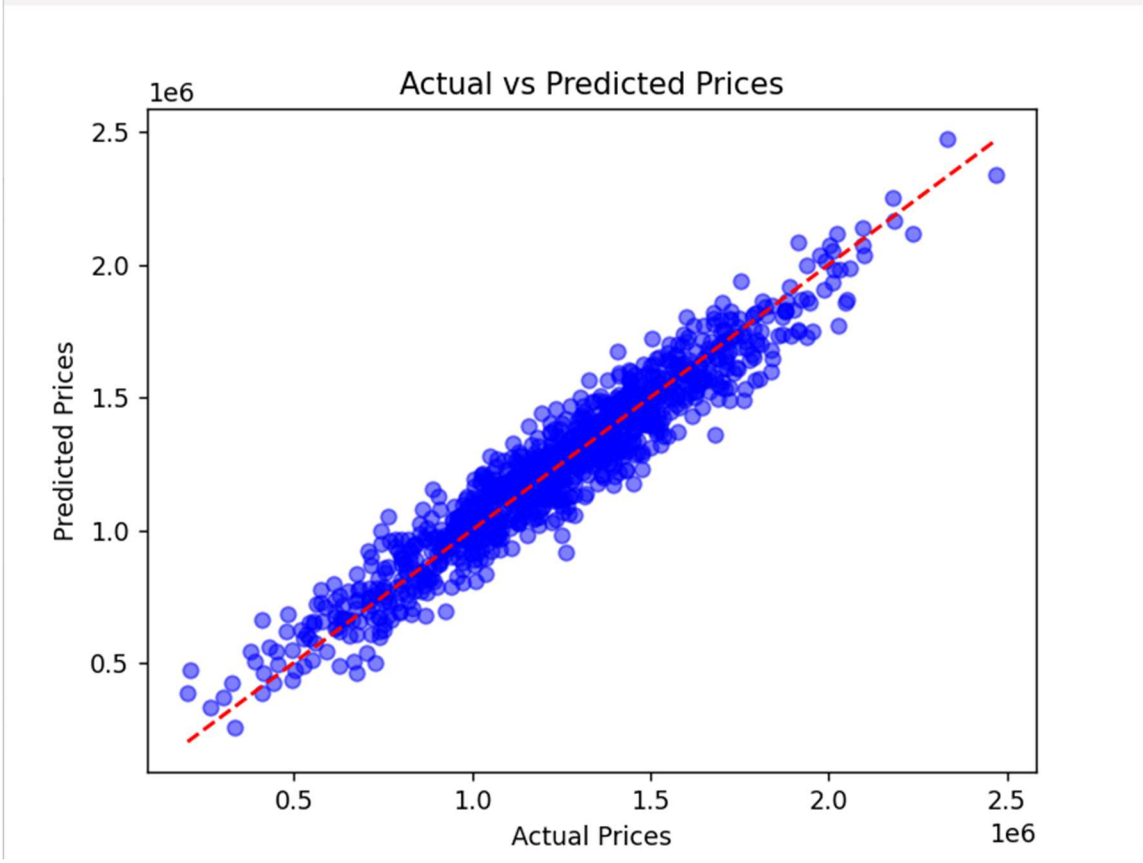
```
PS C:\Users\saxen\Downloads\New> & C:/Users/saxen/AppData/Local/Programs/Python/Python312/python.exe c:/Users/saxen/Downloads/New/index.py
First 5 rows of the dataset:
   Avg. Area Income  Avg. Area House Age  Avg. Area Number of Rooms  ...  Area Population         Price                                            Address
0      79545.458574             5.682861                   7.009188  ...     23086.800503  1.059034e+06  208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
1      79248.642455             6.002900                   6.730821  ...     40173.072174  1.505891e+06  188 Johnson Views Suite 079\nLake Kathleen, CA...
2      61287.067179             5.865890                   8.512727  ...     36882.159400  1.058988e+06  9127 Elizabeth Stravenue\nDanieltown, WI 06482...
3      63345.240046             7.188236                   5.586729  ...     34310.242831  1.260617e+06                            USS Barnett\nFPO AP 44820
4      59982.197226             5.040555                   7.839388  ...     26354.109472  6.309435e+05                            USNS Raymond\nFPO AE 09386

[5 rows x 7 columns]

Missing values in the dataset:
Avg. Area Income              0
Avg. Area House Age           0
Avg. Area Number of Rooms     0
Avg. Area Number of Bedrooms  0
Area Population               0
Price                         0
Address                       0
dtype: int64

Model Evaluation Metrics:
Mean Squared Error (MSE): 10089009300.892078
Root Mean Squared Error (RMSE): 100444.0605555753
R-squared (R2 Score): 0.9179971706834487
```

Graphical Representation of the Actual VS Predicted Prices



Learning Outcome:

- Learned how to perform linear regression on a real-world dataset.
- Understood how to evaluate the performance of a regression model.
- Visualized the relationship between actual and predicted prices.