**GOVERNMENT COLLEGE OF ENGINEERING ERODE**

****

B.E Electronics and Communication Engineering

PREDICTING HOUSE PRICES USING MACHINE LEARNING

**Name of the Students: University Register no:**

**Team Leader:**

M.Priyadharshini 731121106038

**Team Members:**

M.Sugi 731121106048

K.V.Hema 731121106017

Under the mentor of

**Dr.M.Sathyakala**

**Department of Information Technology(IT)**

**Department of Electronics and Communication Engineering**

Government College of Engineering

Near Vasavi College(PO),Erode,TamilNadu-638316,

Affiliated to Anna University ,Chennai.

**INTRODUCTION:**

* Predicting house prices using machine learning involves developing a model or algorithm that uses historical data and various features of a property to estimate its market value or selling price.
* Predicting house prices using machine learning involves training a model to recognize patterns in data and make accurate predictions. It needs a dataset that includes features like location, number of rooms, area, etc., along with corresponding house prices. Then, by using various machine learning algorithms like linear regression, decision trees, or neural networks to train the model. The model learns from the data and can make predictions on new, unseen data. It's important to evaluate the model's performance using metrics like mean squared error or R-squared to ensure its accuracy.

**PROJECT OVERVIEW:**

The overall, predicting house prices using machine learning involves a comprehensive data science pipeline, from data collection and preprocessing to model development, evaluation, deployment and ongoing maintenance.

**OBJECTIVE:**

* To develop a machine learning model that accurately predicts house prices based on various features.
* It is a helping tool for real estate professionals and individuals looking to buy or sell properties.

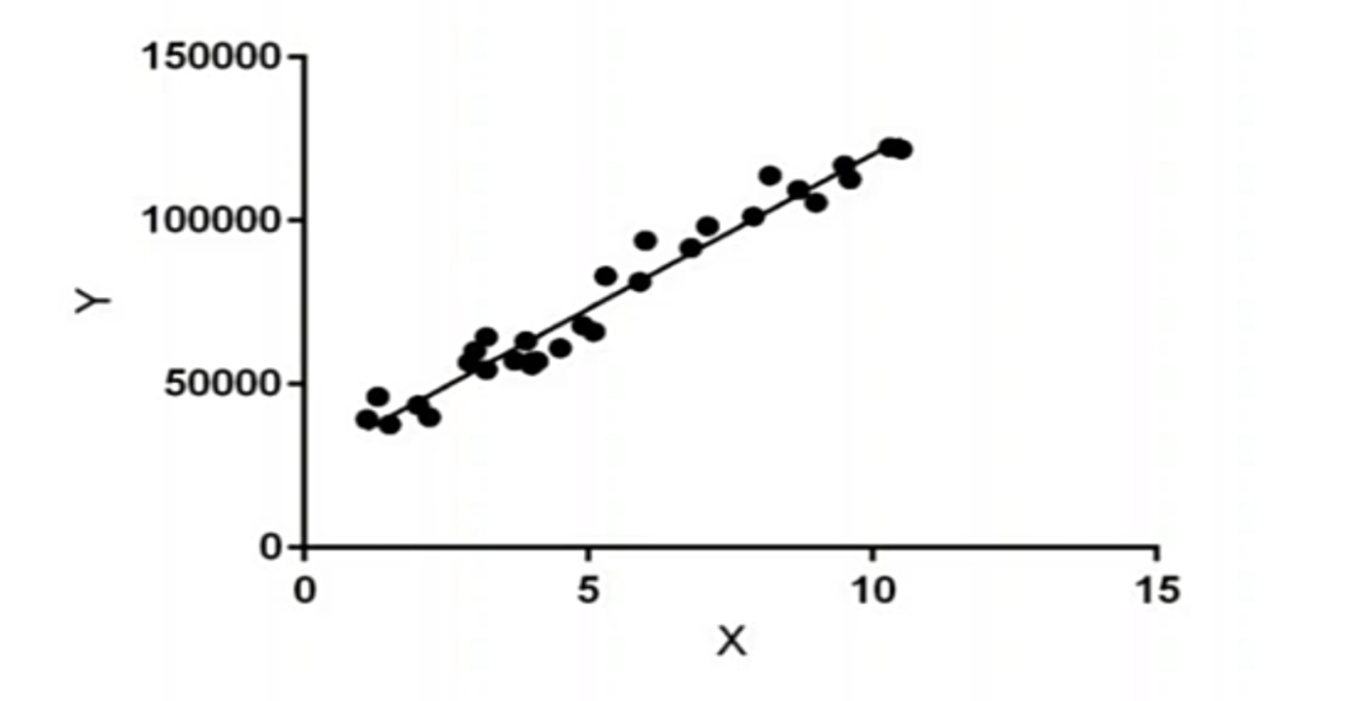
**DATASET:**

<https://www.kaggle.com/datasets/vedavyasv/usa-housing>

* ‘Avg.Area Income’- Avg. The income of the householder of the city house is located.
* ‘Avg.Area House Age’- Avg.Age of houses in the same city.
* ‘Avg.Area Number of Rooms’- Avg.Number of rooms for houses in the same city.
* ‘Avg.Area Number of Bedrooms’- Avg.Number of bedrooms for houses in the same city.
* ‘Area Population’- Population of the city.
* ‘Price’- Price that the house sold at.
* ‘Address’- Address of the houses.

## Understanding Linear Regression:

Linear regression is a fundamental supervised learning algorithm in machine learning. It aims to establish a linear relationship between a dependent variable (target) and one or more independent variables (features). In the context of house price prediction, the dependent variable will be the house price, and the independent variables can be factors like the size of the house, number of bedrooms, location, etc.

****

**House Price Prediction with Linear Regression Involves Following Steps:**

* **Dataset Collection**: Gather historical house price data and corresponding features from platforms like Zillow or Kaggle.
* **Data Preprocessing**: Clean the data, handle missing values, and perform feature engineering, such as converting categorical variables to numerical representations.
* **Splitting the Dataset**: Divide the dataset into training and testing sets for model building and evaluation.
* **Building the Model**: Create a linear regression model to learn the relationships between features and house prices.
* **Model Evaluation**: Assess the model’s performance on the testing set using metrics like MSE or RMSE.
* **Fine-tuning the Model**: Adjust hyperparameters or try different algorithms to improve the model’s accuracy.
* **Deployment and Prediction**: Deploy the robust model into a real-world application for predicting house prices based on user inputs.

**PYTHON PROGRAM:**

**Import Libraries:**

 import pandas as pd

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 dataset from CSV**

df = pd.read\_csv('house\_data.csv')

**# Exploratory Data Analysis (EDA)**

print(df.head())

**# Summary statistics of the dataset**

print(df.describe())

**# Check for missing values**

print(df.isnull().sum())

**# Correlation matrix to understand feature relationships**

correlation\_matrix = df.corr()

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm')

plt.title("Correlation Matrix")

plt.show()

**# Preprocessing:**

Selecting features and target variable

X = df[['bedrooms', 'bathrooms', 'sqft\_living', 'sqft\_lot', 'floors', 'waterfront', 'view', 'condition']]

y = df['price']

**# Splitting the dataset 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)

**# Building the Linear Regression Model**

model = LinearRegression()

**# Fitting the model on the training data**

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

**# Model Evaluation**

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

**# Mean Squared Error and R-squared for model evaluation**

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

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

print("Mean Squared Error:", mse)

print("R-squared:", r2)

**# Predictions and Visualization**

plt.scatter(y\_test, y\_pred)

plt.xlabel("Actual Prices")

plt.ylabel("Predicted Prices")

plt.title("Actual Prices vs. Predicted Prices")

plt.show()

**# creating a residual plot to check the model's performance**

residuals = y\_test - y\_pred

plt.scatter(y\_test, residuals)

plt.axhline(y=0, color='red', linestyle='--')

plt.xlabel("Actual Prices")

plt.ylabel("Residuals")

plt.title("Residual Plot")

plt.show()

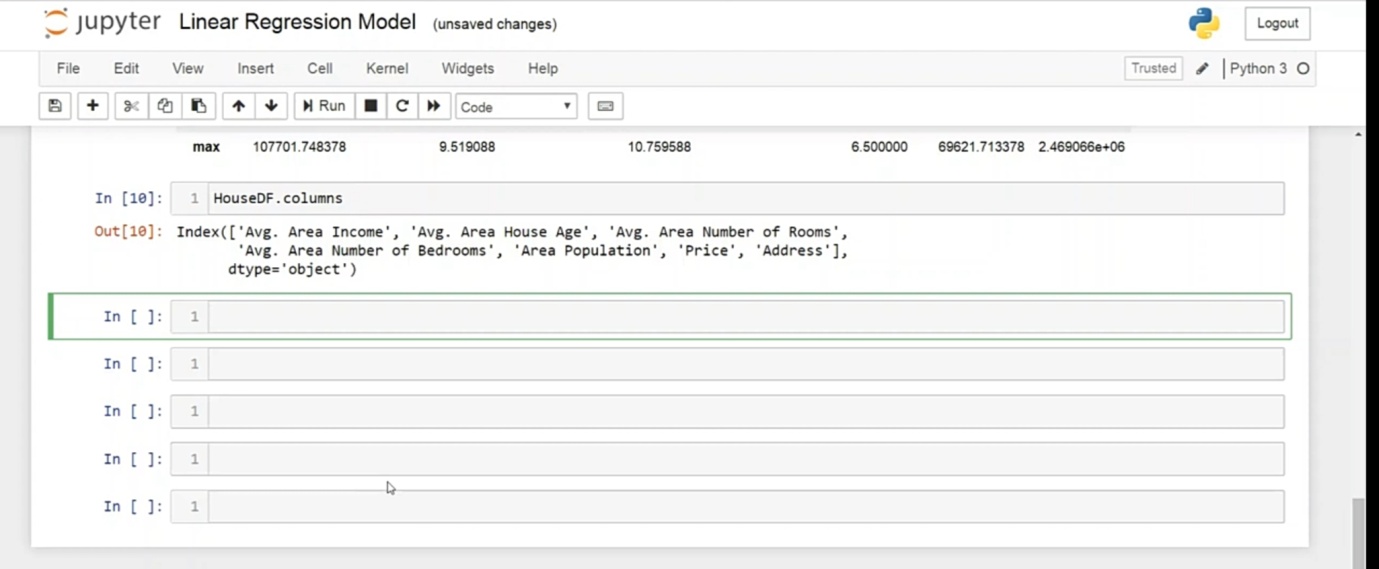
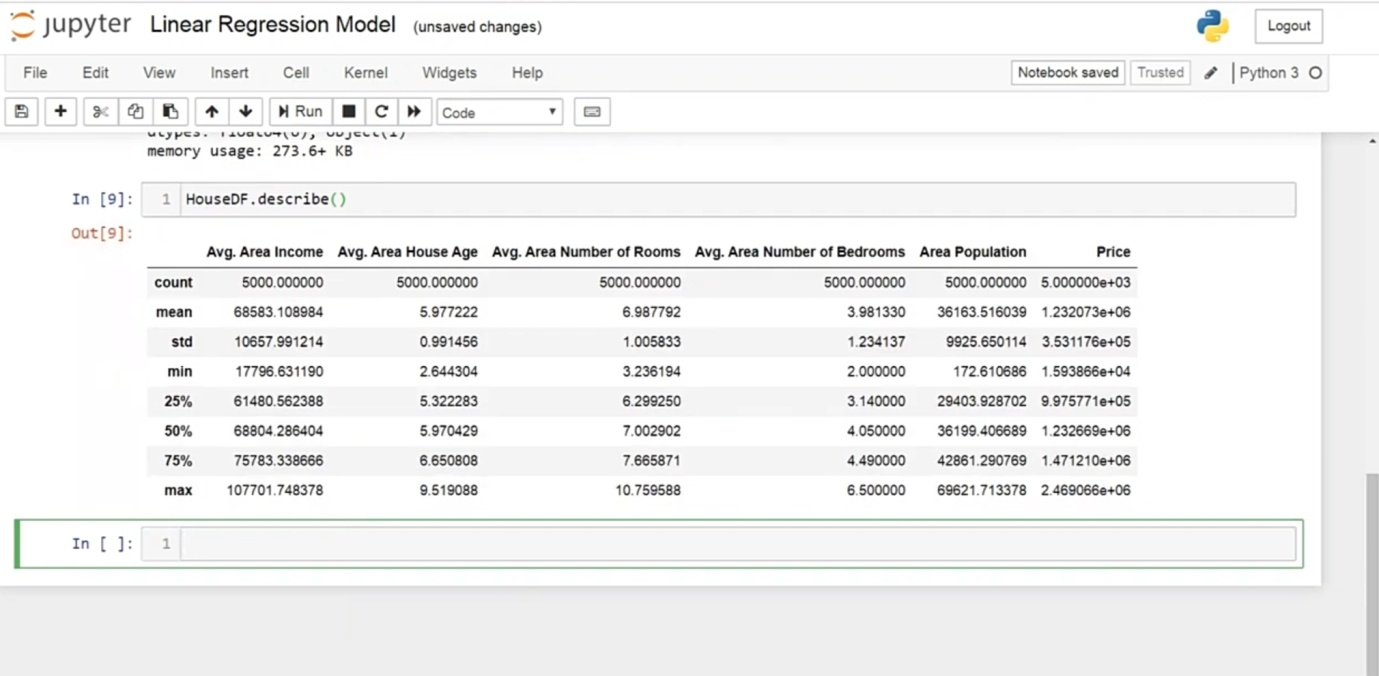
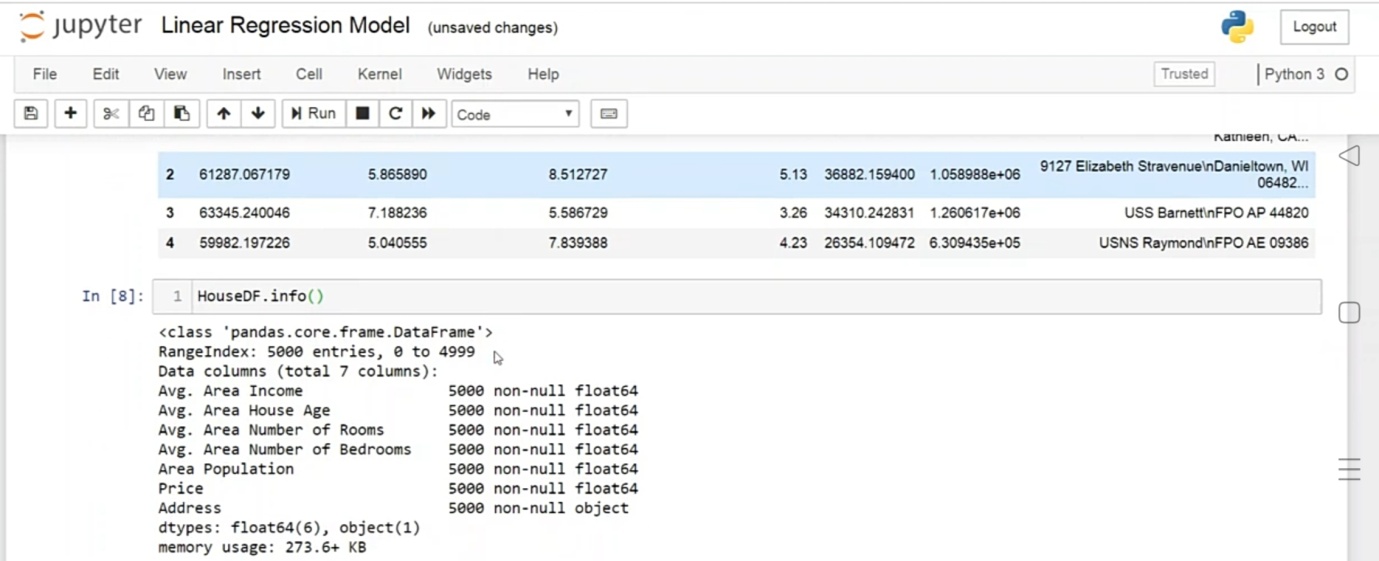
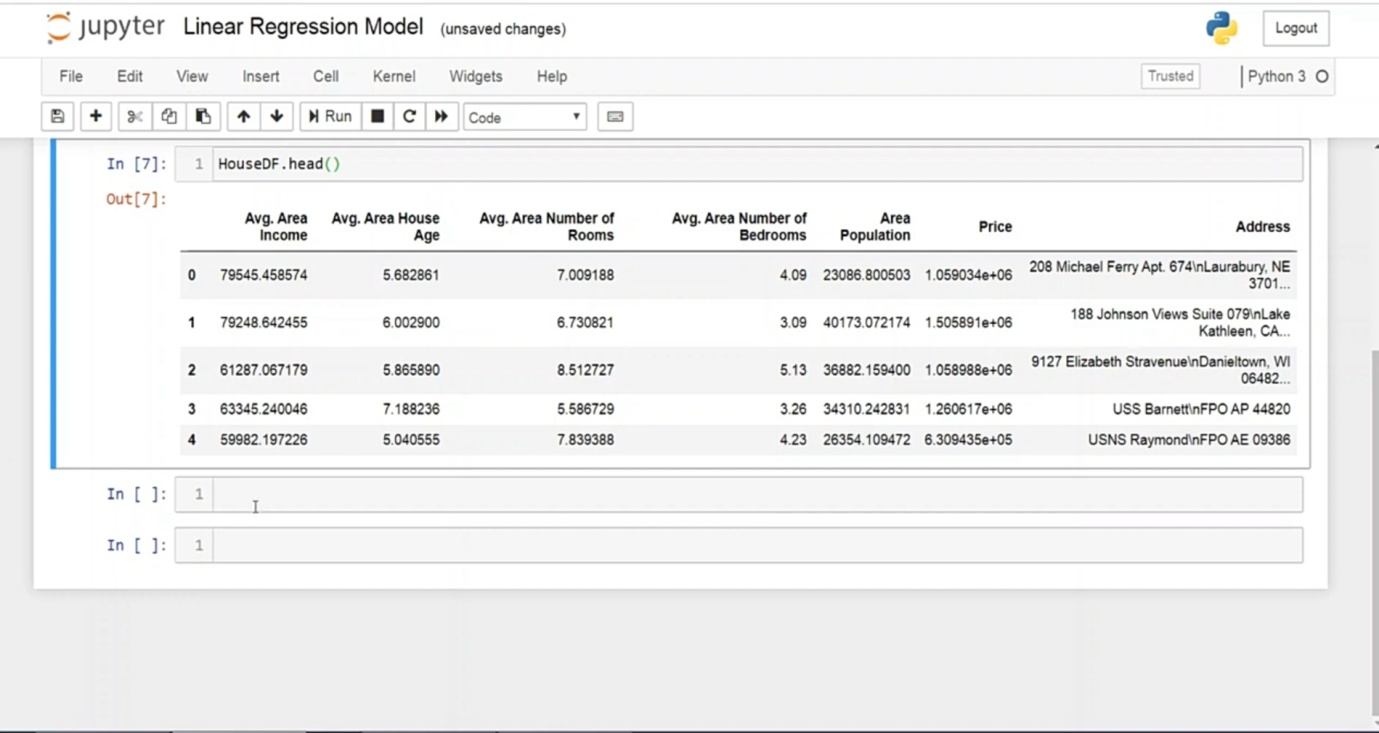
**# Using the trained model to make predictions on new data and visualize the results**

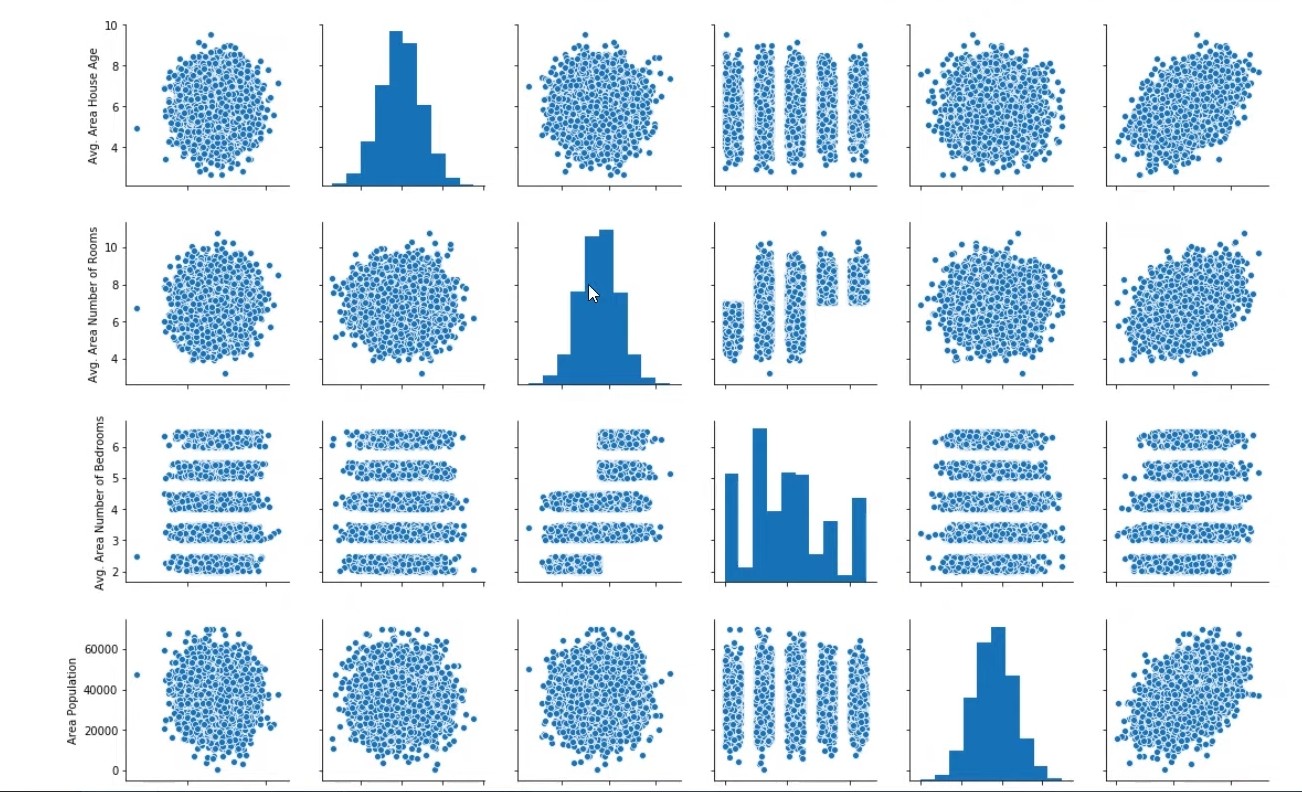
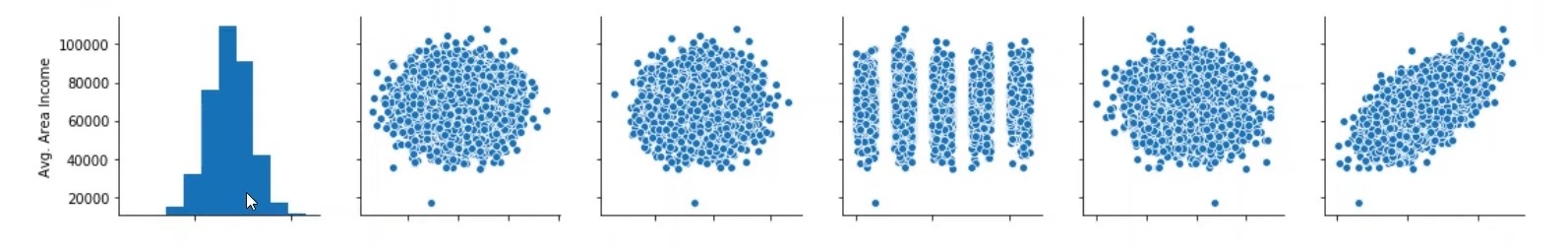
new\_data = [[3, 2, 1500, 4000, 1, 0, 0, 3]]

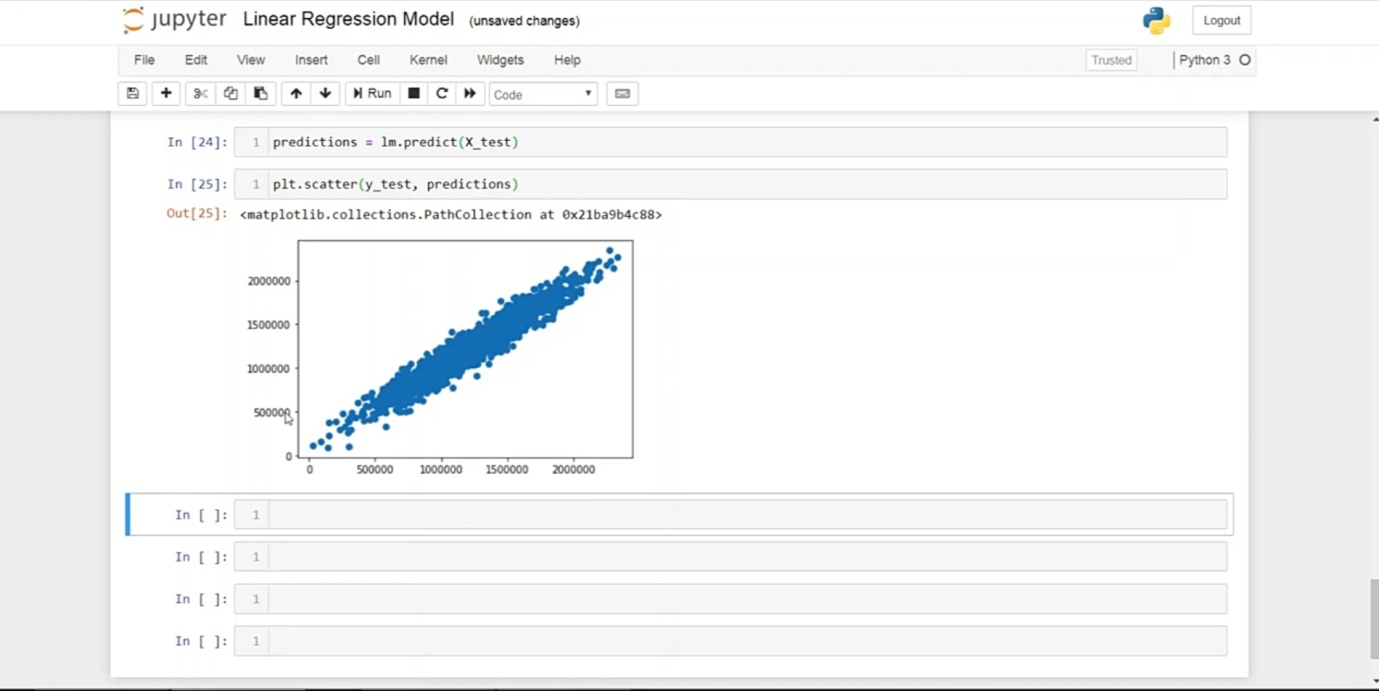
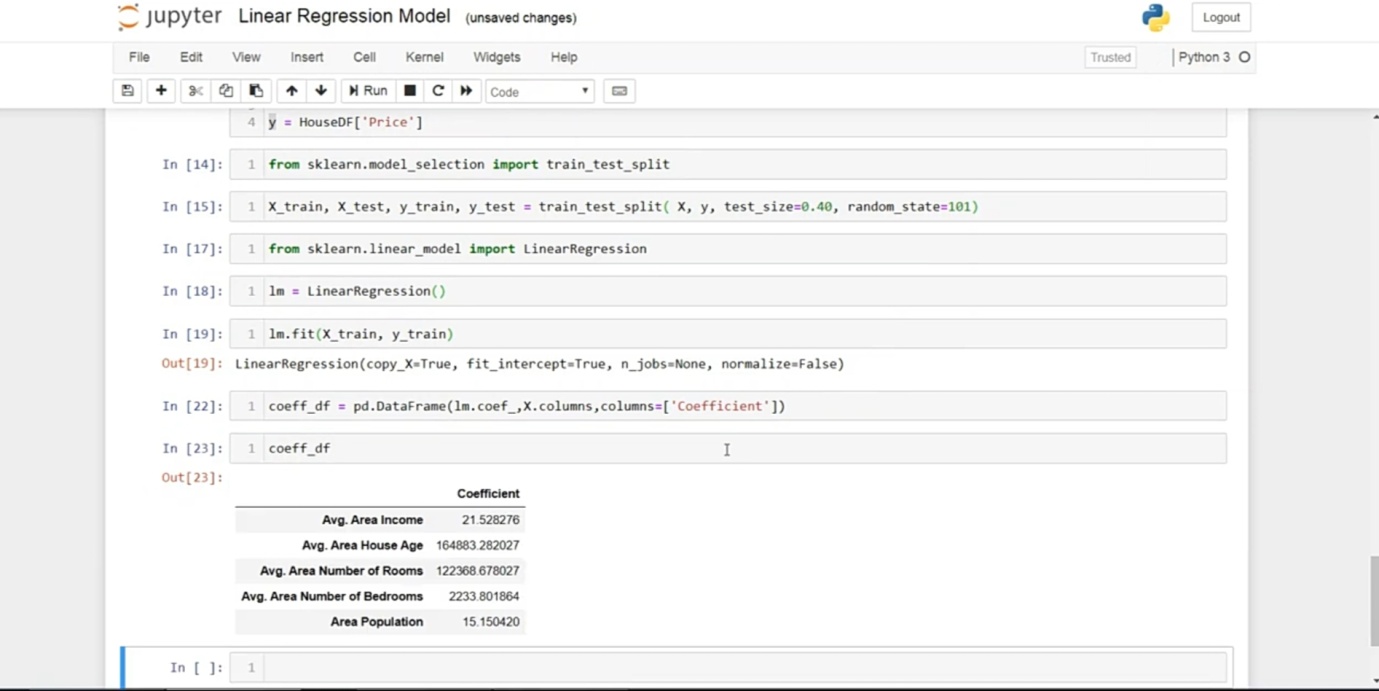
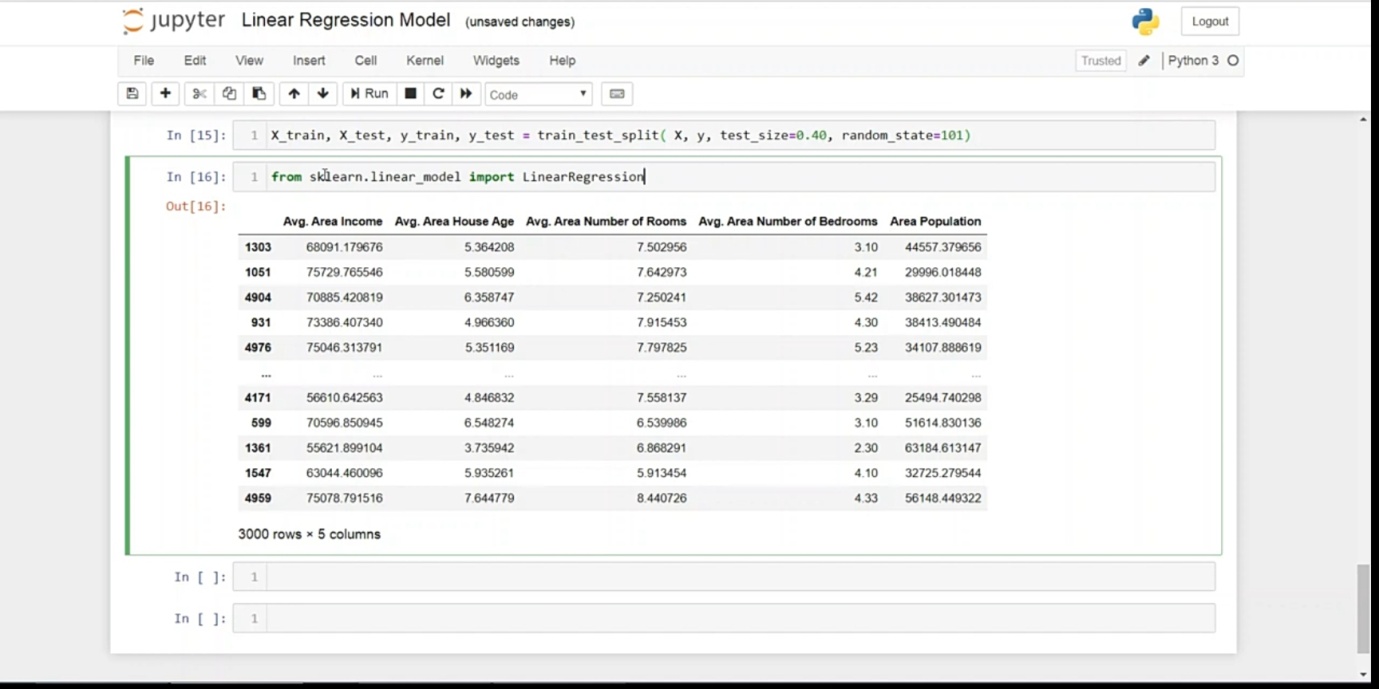
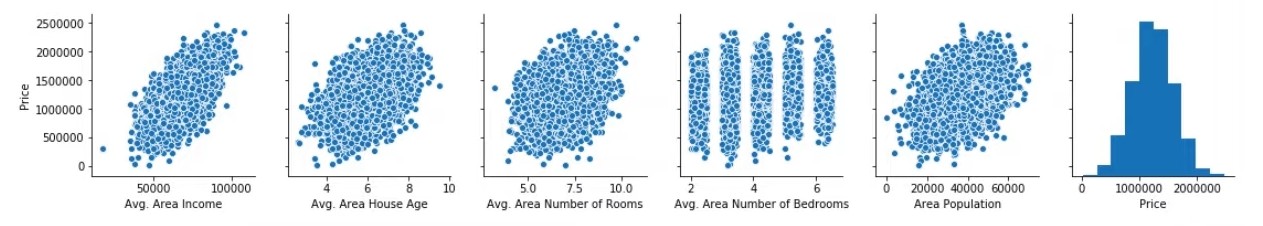
predicted\_price = model.predict(new\_data)

print("Predicted Price:", predicted\_price[0])

**OUTPUT:**

****

****

****

****

**ADVANTAGES:**

Predicting house prices using machine learning offers several

significant advantages:

**1.Accuracy:**

Machine learning models can process and analyze vast amounts of data, including various property and market factors. This result is more accurate house price predictions compared to traditional methods that rely on a limited set of variables.

**2.Complex Data Handling:**

Machine learning algorithms can handle complex, non-linear relationships in the data. They can recognize patterns and interactions among different features, allowing for a more comprehensive assessment of a property's value.

**3.Continuous Learning:**

Machine learning models can be continually updated with new data, enabling them to adapt to changing market conditions and trends.

**4.Efficiency:**

Automated valuation models powered by machine learning can evaluate properties rapidly. This efficiency is beneficial for both property appraisers and individuals looking to determine the value of a property quickly.

**5. Data Integration:**

Machine learning models can incorporate a wide range of data sources, including property characteristics, neighborhood information, economic indicators, and even social trends. This holistic approach provides a more complete picture of the factors influencing house prices.

**6.Reduced Bias:**

Machine learning can help reduce human bias in property valuation. It evaluates properties objectively based on data, which can lead to fairer and more consistent pricing.

**7.Market Insights:**

By analyzing historical data and current market conditions, machine learning can offer valuable insights into market trends, helping investors and developers make informed decisions.

**8.Risk Assessment:**

Machine learning can assess the risk associated with a property, which is crucial for mortgage lenders and investors. It helps identify potential issues or opportunities related to a property's value.

**9. Transparency:**

Machine learning models can provide clear and transparent explanations for their predictions, which is essential for building trust among stakeholders in the real estate market.

**10. Scalability:**

Machine learning models can be deployed at scale, making it possible to assess property values in large real estate portfolios, entire neighborhoods, or even across entire cities.

**11. Time and Cost Savings:**

Using machine learning for property valuation can save time and reduce costs associated with manual appraisals, making it an efficient and cost-effective solution for businesses.

**12. Customization:**

Machine learning models can be customized to cater to specific markets, types of properties, or regional variations, allowing for more tailored and precise predictions.

**CONCLUSION:**

Predicting house prices using machine learning is a transformative and promising approach that has the potential to revolutionize the real estate industry. By this exploration, the project determines remarkable capabilities of machine learning in providing more accurate data-driven, and nuanced predictions for property values.