**House Price Prediction\_Phase4**

**(Development Part-II)**

**Importing and loading data from dataset:**

You must import a number of libraries and modules to aid with data processing, model creation, and evaluation while creating an artificial intelligence system that predicts home prices.

1. Prepare Your Dataset
2. Dataset Format
3. Load the Dataset (using pandas library)

**Preprocessing the dataset:**

Preparing data for a forecast of home prices to guarantee that the data is in the proper format and quality for training a neural network or other machine learning model, artificial intelligence (AI) comprises a number of crucial processes. The standard preprocessing steps are as follows:

**Data collection**:

Compile pertinent information about homes and their attributes, such as square footage, the number of bedrooms and bathrooms, location, facilities, and past sales prices.

**Cleaning Data:**

Managing Missing Values: Recognise, add, or delete any missing values. Imputation using the feature's mean, median, or mode is a common technique.

Finding and managing outliers that may distort the model's predictions is important. Outliers can be changed or eliminated.

**Transformation of Data:**

**Feature engineering**: Construct additional features as needed, such as a house's age, price per square foot, or total area.

**Normalise the Target Variable**: To make training easier, normalise the house prices if they have a large range.

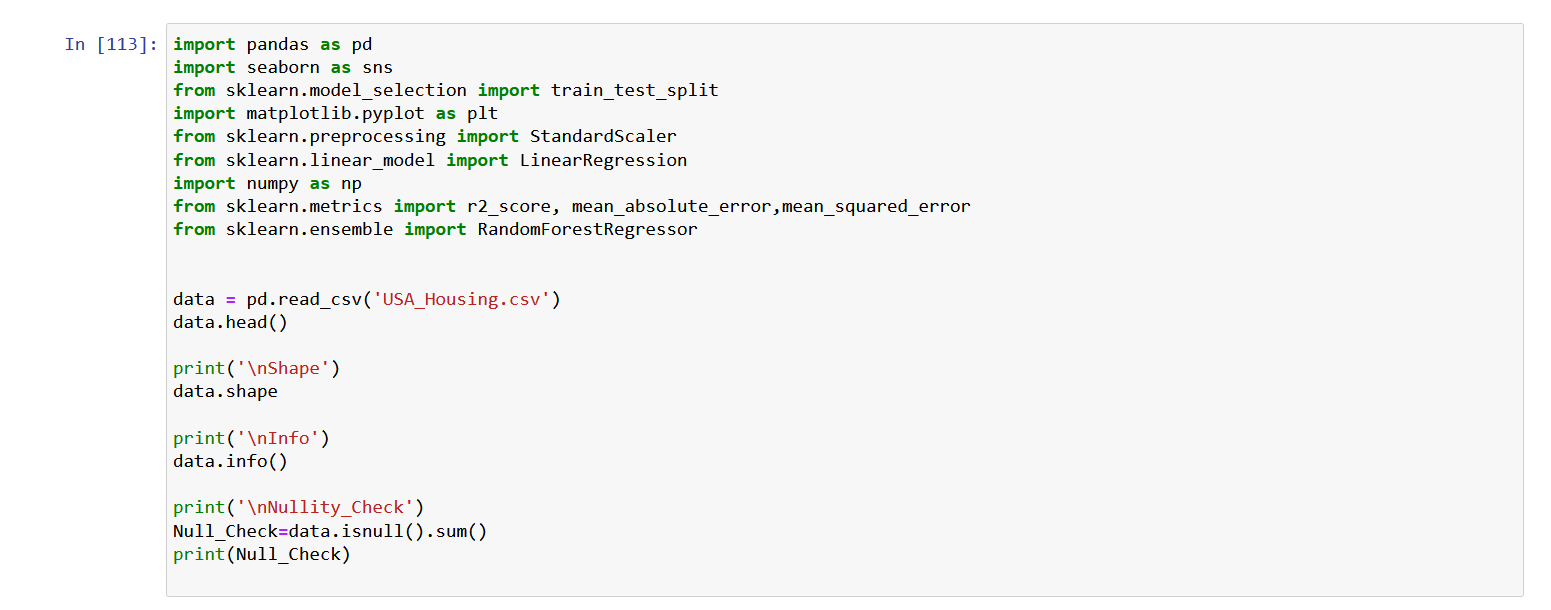
**Data Visualization:**

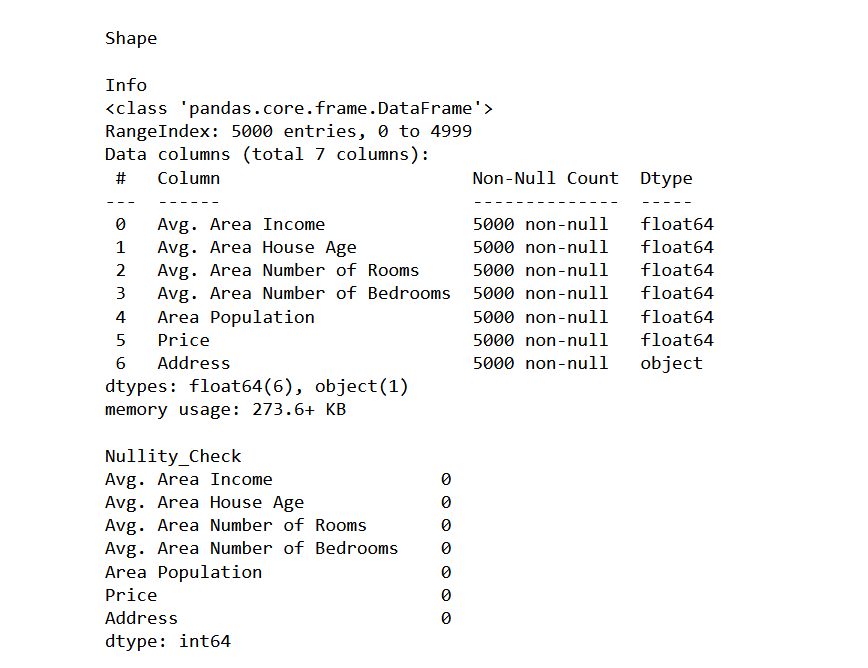
Explore the data through plots and graphs to better understand the relationships between features and the target variable.

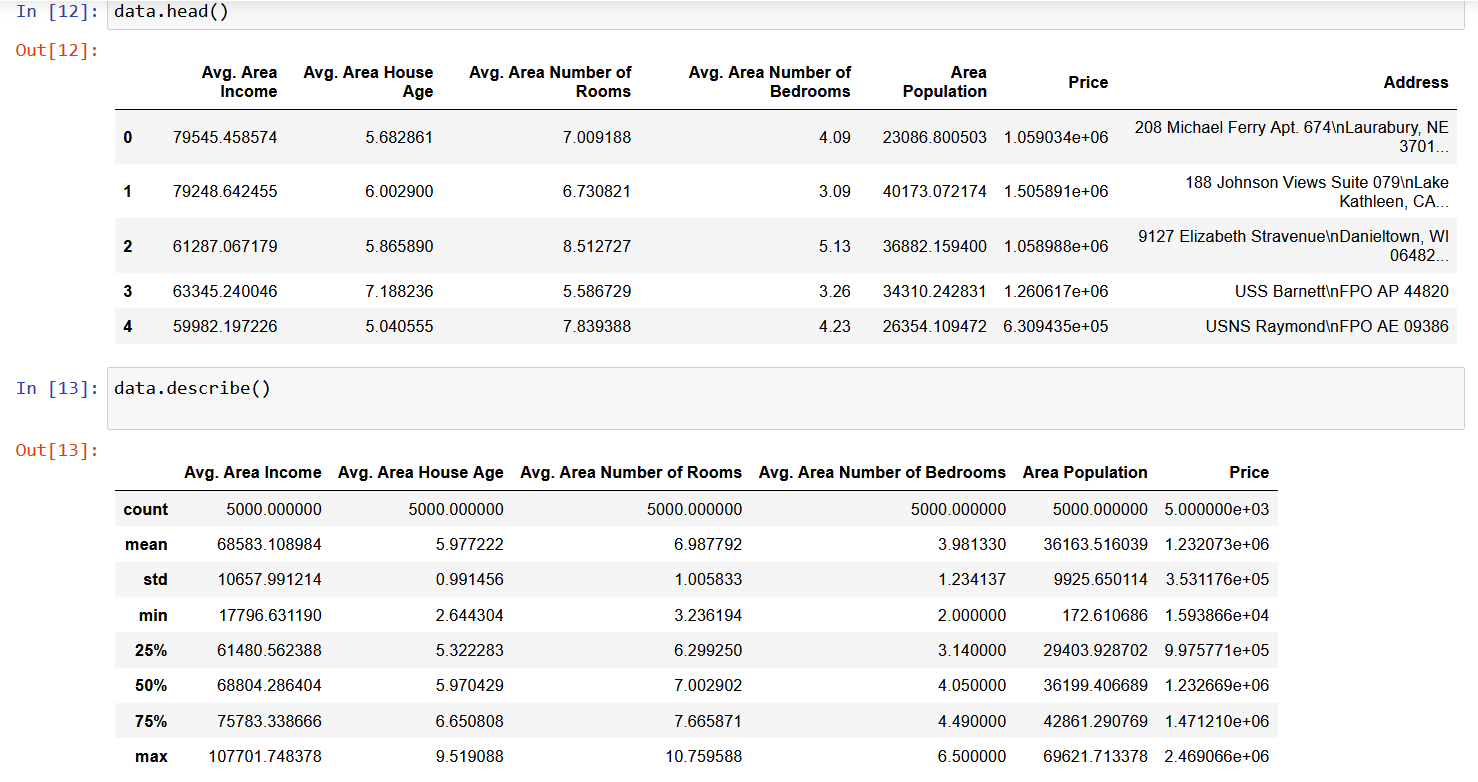
**Feature scaling**:

To give numerical characteristics a consistent scale, normalise or standardize them. For algorithms that are sensitive to feature sizes, such as neural networks, this is particularly crucial.

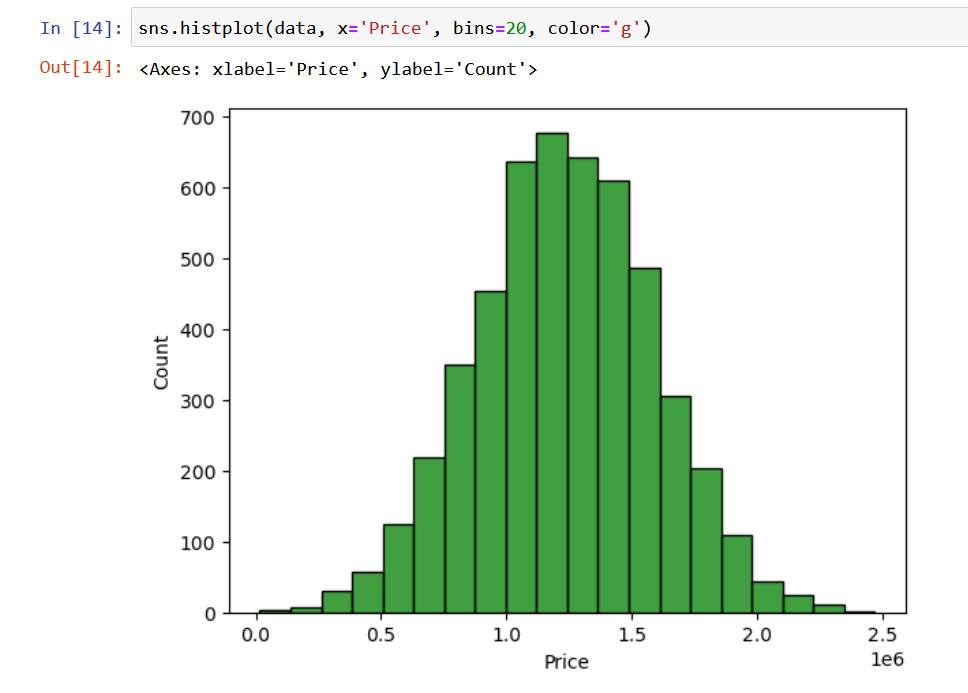
**Model-Specific Preprocessing**:

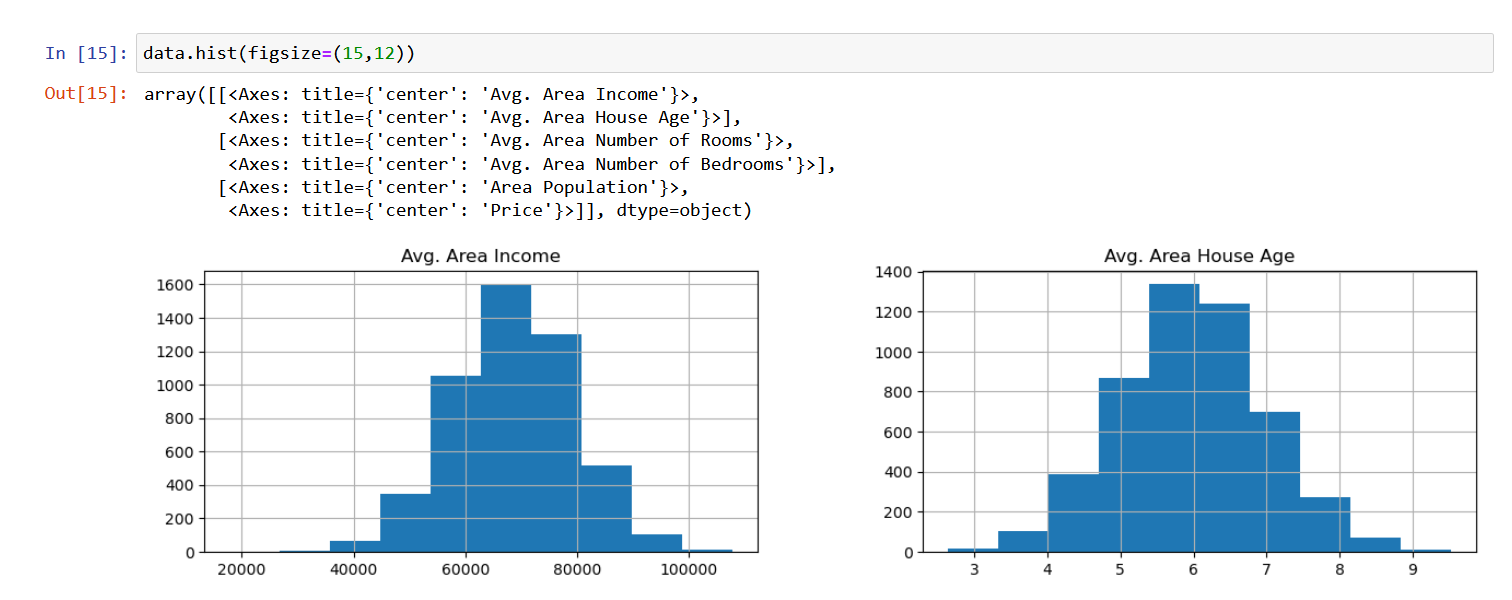
Neural networks are one type of AI model that may have particular needs. It can be necessary, for instance, to normalise input characteristics between 0 and 1, or to carry out further changes depending on the model.

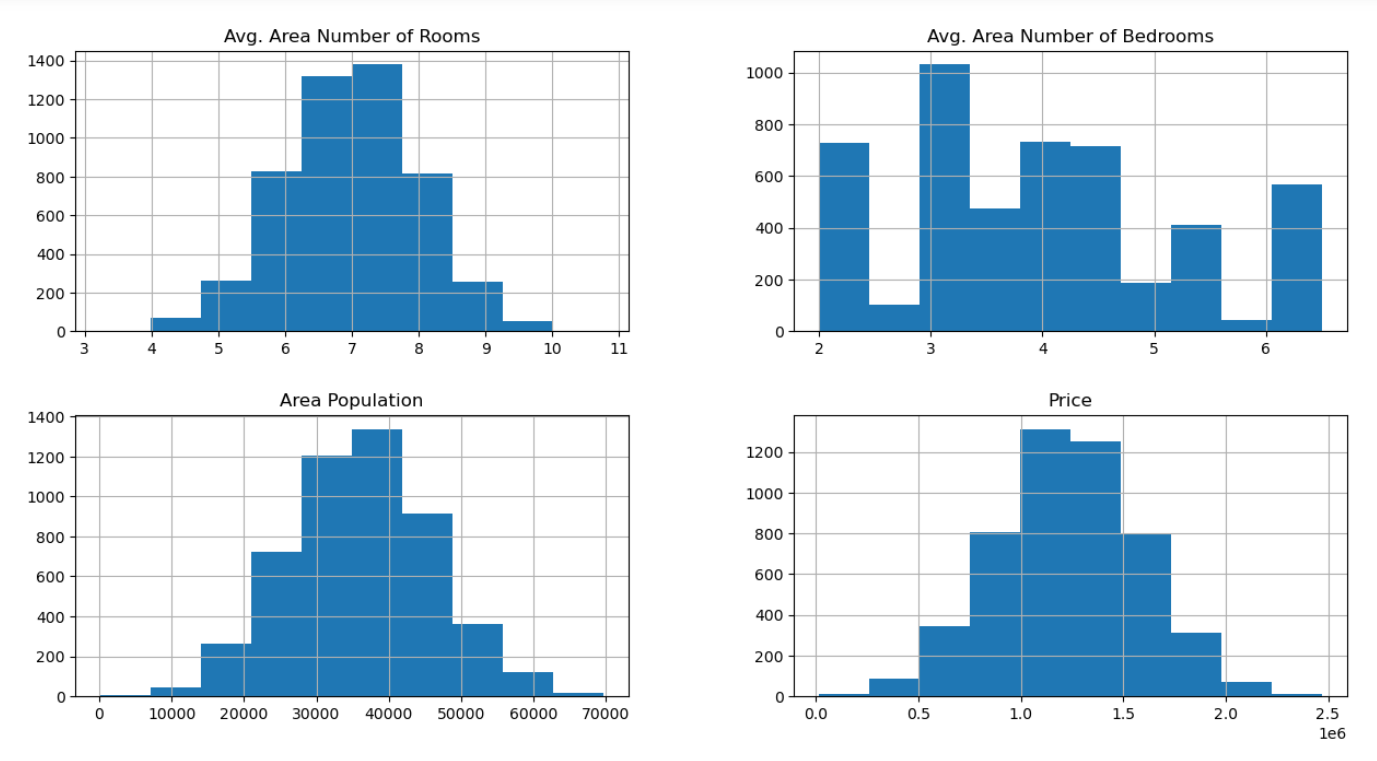
**Output:**

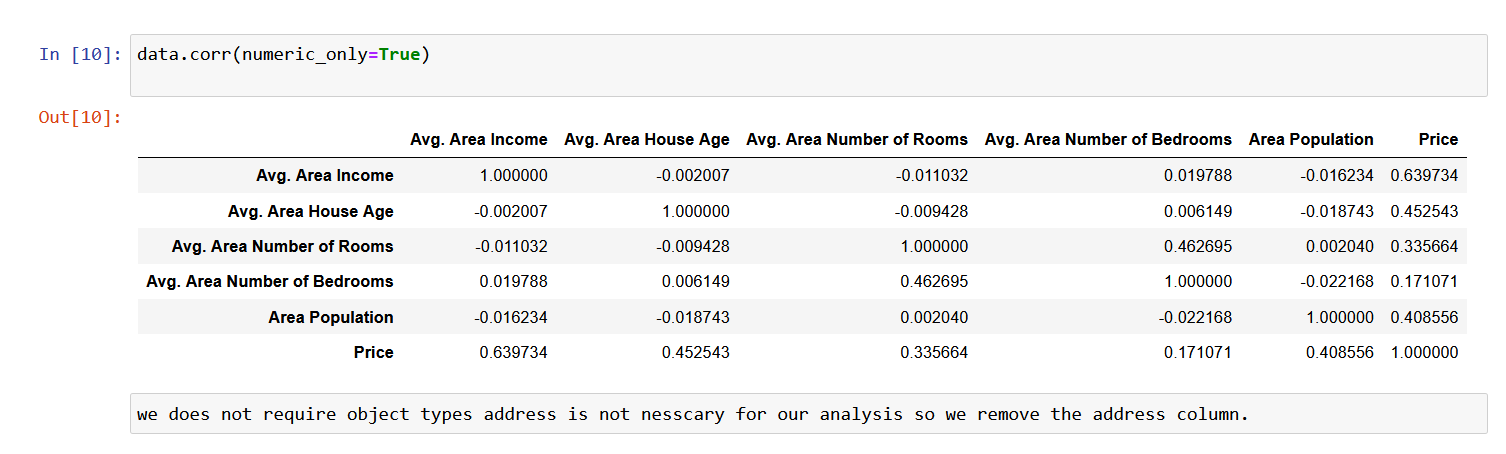


**There is no null values in the cells. So, it does need to drop or fill the null values**.

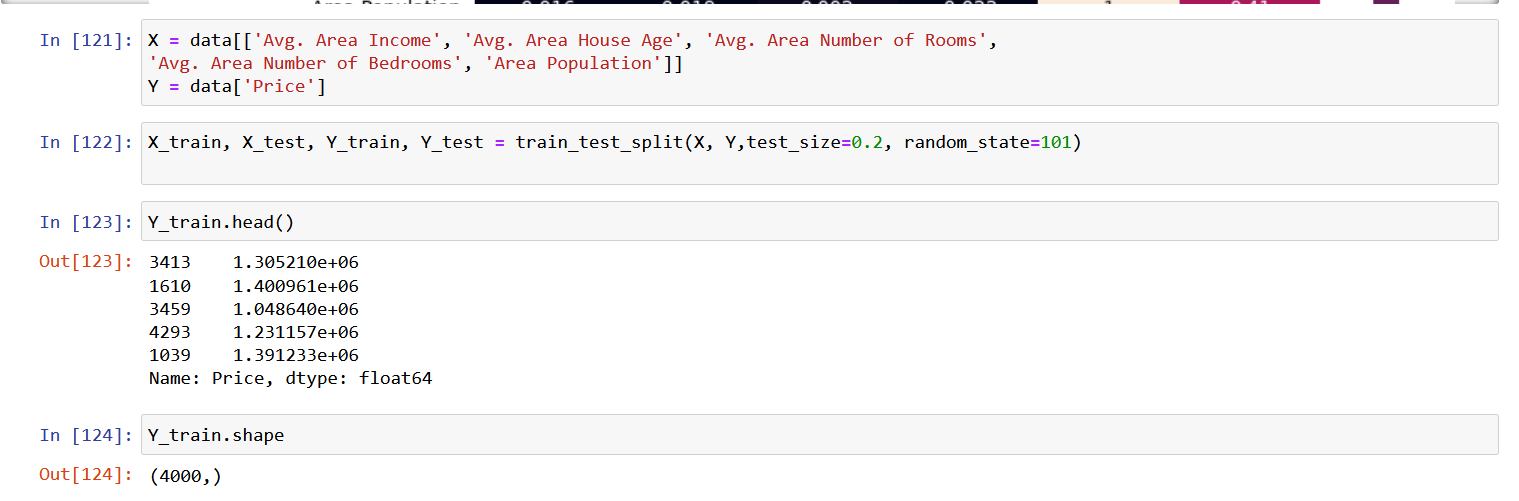




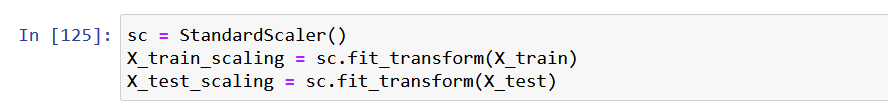




**Dividing dataset & Train, Test, split the data**

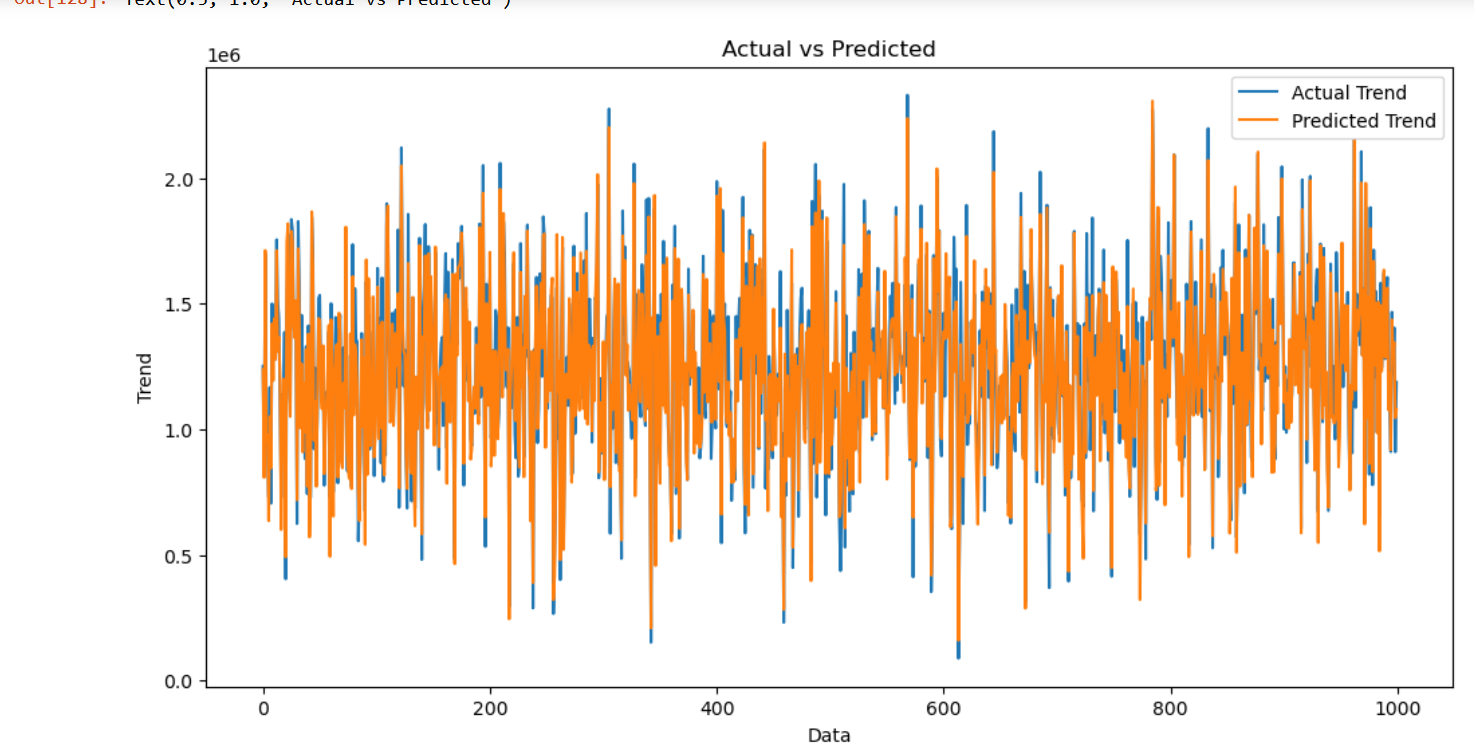


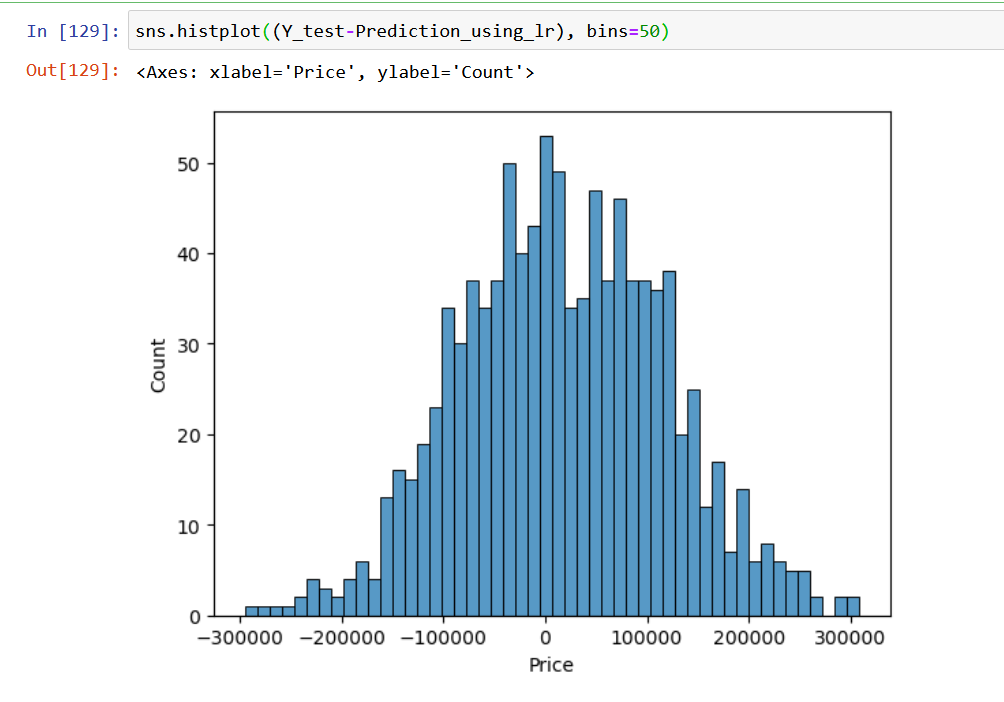
# Standardizing the data



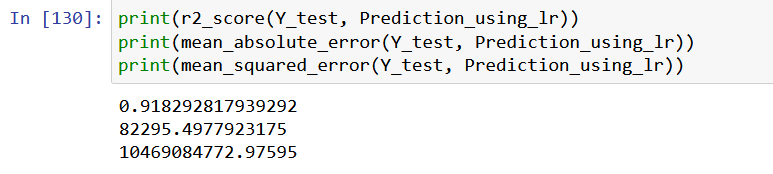
**Model building and Evaluation using Linear Regression algorithm**

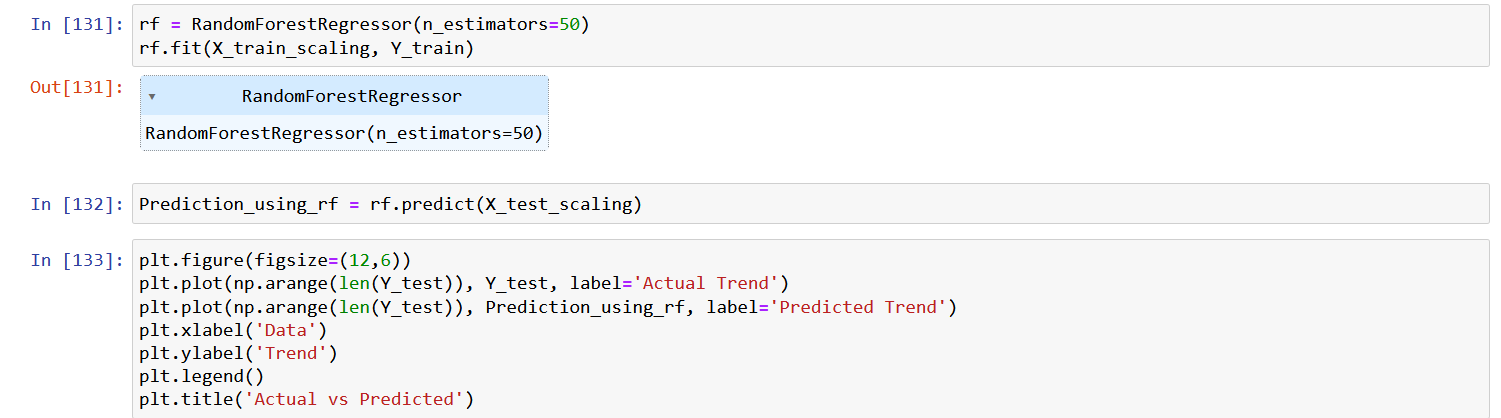


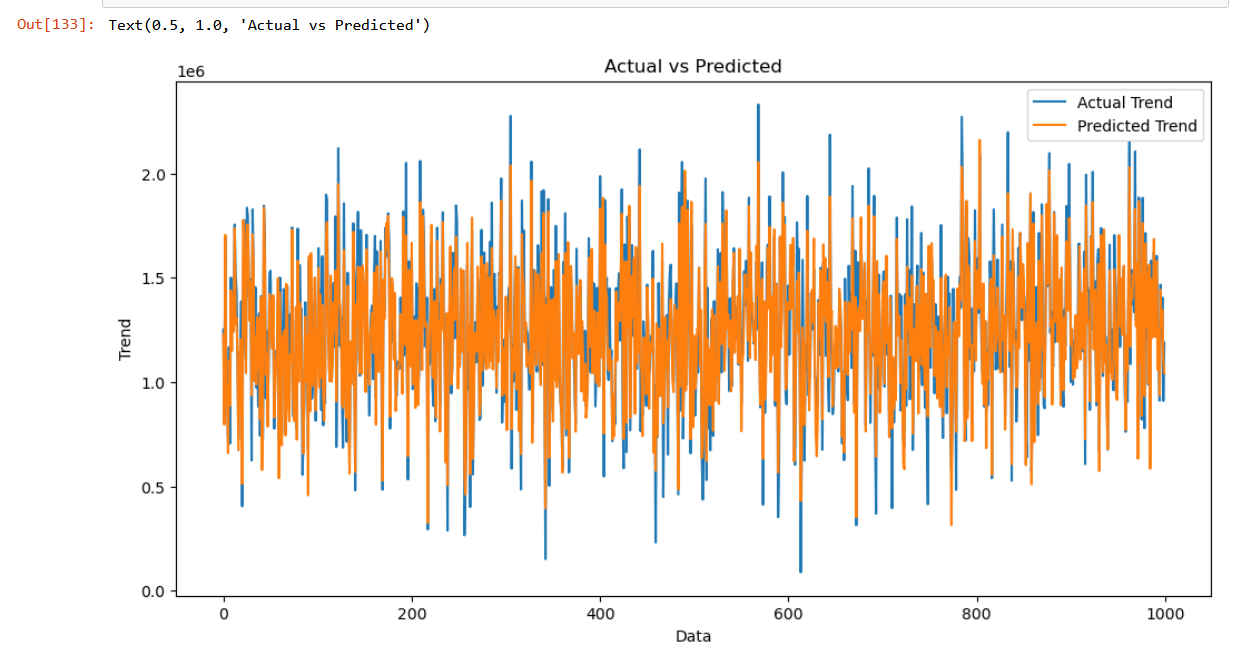


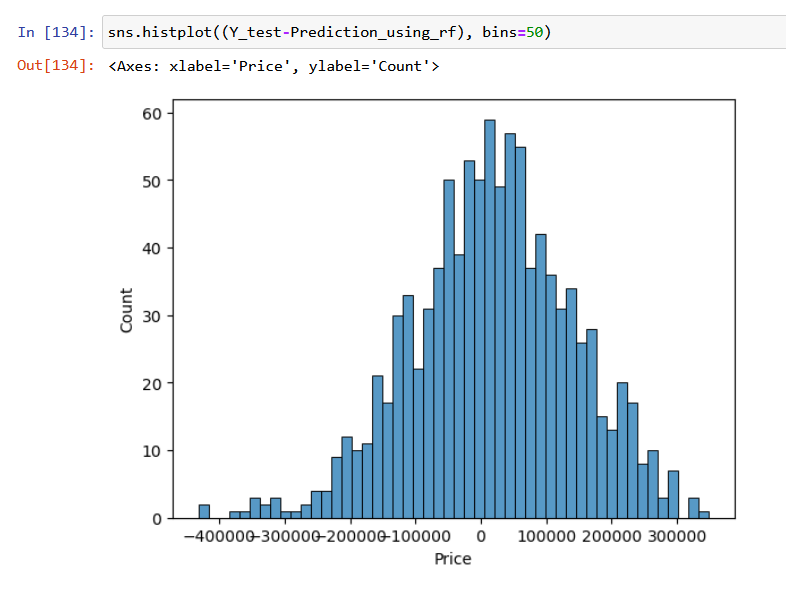
**Testing view in histogram graphical view**

**Our prediction Score value**

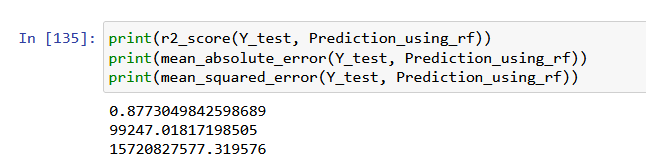
****

**Model building and Evaluation using Random Forest Regressor algorithm**





**Prediction Score using random forest regressor**



Both are nearly give the r2\_score, absolute error and squared error. And the both graphical view for actual vs predicted values are also nearly same.

**Dataset link:**

[**https://www.kaggle.com/datasets/vedavyasv/usa-housing**](https://www.kaggle.com/datasets/vedavyasv/usa-housing)

**Drive link:**

**https://drive.google.com/file/d/1Hztmh2WjORVqYtrsvRZsJ-p5bKsAcWgX/view?usp=sharing**

**Github link:**

**https://github.com/Pontamilselvan2004/House\_Price\_Prediction\_Phase3/blob/main/Preprocessing\_Data.ipyn**