**Project Title: Predictive Analysis of Housing Prices in the USA**

Introduction

This project focuses on building a predictive model to analyze and predict housing prices in the United States using a Linear Regression algorithm. This document provides a detailed overview of the project, including the dataset, problem statement, methodology, and results.

Description

Import all the required packages.

import the dataset using pandas.

After importing the dataset we analyze the dataset with head, tail, describe methods.

After we will find the relationship among the columns in the dataset with the help of plots,

histogram plot is used to find the each and every price frequency.

box plot used to find the outlier

heatmap is used to find the correlation among all the columns in the dataset.

After analyzing the data we do preprocessing by scale down all the values to 0 to 1 except price column.

After we split the dataset into two part train, test.

After we build the LinearRegression model to predict the price of the house.

by using test data we calculate the accuracy of the model.

Dataset

The dataset used in this project is named "USA\_Housing.csv." It contains information about housing features and prices, with the following columns:

- 'Avg. Area Income': Average income of the residents in the area.

- 'Avg. Area House Age': Average age of houses in the area.

- 'Avg. Area Number of Rooms': Average number of rooms in houses in the area.

- 'Avg. Area Number of Bedrooms': Average number of bedrooms in houses in the area.

- 'Area Population': Population of the area.

- 'Price': Price of the house.

The goal is to predict the 'Price' of houses based on these features.

Problem Statement

The primary objective of this project is to create a predictive model that can estimate housing prices based on various property-related features. This model has practical applications in real estate, assisting buyers and sellers in making informed decisions about property values.

Methodology

Data Preprocessing

1. Imported necessary libraries: pandas, numpy, seaborn, matplotlib.pyplot, and scikit-learn modules for data analysis and machine learning.

2. Loaded the dataset from 'USA\_Housing.csv' using pd.read\_csv ().

3. Explored the dataset with:

- dataset.head () and dataset.tail () to display the first and last rows.

- dataset.info () to gather information about the dataset.

- dataset.describe () to provide summary statistics.

- Dataset.columns to list the dataset columns.

4. Visualized the dataset with the following plots:

- Histogram of 'Price' using sns.histplot () and sns.boxplot ().

- Hexbin plot between 'Avg. Area House Age' and 'Price' using sns.jointplot ().

- Scatter plot between 'Avg. Area Income' and 'Price' using sns.jointplot ().

- Pairplot and histograms for all dataset features using sns.pairplot () and dataset.hist ().

5. Calculated the correlation matrix and displayed it using a heatmap with sns.heatmap () to identify relationships between variables.

Data Splitting

1. Split the dataset into training and testing sets using train\_test\_split (). The test size was set to 20%.

Data Scaling

1. Standardized the data using StandardScaler () to make the features suitable for linear regression.

Model Building

1. Created a Linear Regression model using LinearRegression ().

2. Trained the model on the training data with model\_linear\_regression.fit (X\_train\_scal, Y\_train).

3. Made predictions on the test data using model\_linear\_regression.predict (X\_test\_scal).

4. Visualized the actual vs. predicted trends of housing prices using matplotlib.

Model Evaluation

1. Calculated evaluation metrics including:

- R-squared score (r2\_score)

- Mean Absolute Error (mean\_absolute\_error)

- Mean Squared Error (mean\_squared\_error)

2. Displayed the results of the model's performance.

Results

The Linear Regression model demonstrated a [r2\_score] R-squared value, indicating [interpretation of R-squared]. The Mean Absolute Error (MAE) and Mean Squared Error (MSE) were [MAE and MSE values], suggesting [interpretation of MAE and MSE]. The model successfully predicted housing prices based on the provided features.