**California Housing Price Prediction using Linear Regression**

**1. Introduction**

Predicting housing prices is a crucial task in real estate analytics, helping buyers, sellers, and policymakers make informed decisions. This project utilizes the **California Housing dataset** to build a **Linear Regression** model that predicts house values based on various features such as population, income, and house characteristics.

**2. Dataset and Preprocessing**

We used the **California Housing dataset** from sklearn.datasets, which contains **20,640 samples** with **8 numerical features**, including:

* **MedInc**: Median income in block group
* **HouseAge**: Median house age in block group
* **AveRooms**: Average number of rooms per household
* **AveBedrms**: Average number of bedrooms per household
* **Population**: Block group population
* **AveOccup**: Average household occupancy
* **Latitude**: Geographic coordinate
* **Longitude**: Geographic coordinate

The target variable (House\_Value) represents the **median house value (in $100,000s)**.

The dataset was split into **80% training data** and **20% test data** using train\_test\_split().

**3. Model Training**

We implemented **Linear Regression** using LinearRegression() and trained it on the **training dataset**.

**4. Model Evaluation**

After training, the model was tested on the unseen **test dataset**. The performance was measured using the **R² score**, which evaluates how well the model explains the variance in house prices.

* **R² Score:** The model achieved an **accuracy of 0.61**, meaning it explains **61% of the variance** in house prices.