Housing Price Prediction Using Multiple Linear Regression

Objective:

The goal of this project was to build a **Multiple Linear Regression** model that accurately predicts **house prices** based on several features like area, number of bedrooms, bathrooms, availability of amenities (e.g., air conditioning, basement), and furnishing status using a housing dataset.

Dataset Overview:

The dataset (housing.csv) contains **13 features** and **1 target column (price)**. Features include both **numerical** (e.g., area, bedrooms, parking) and **categorical** (e.g., main road, guestroom, furnishing status) variables.

Key Features:

- area, bedrooms, bathrooms, stories
- main road, guestroom, basement
- hot water heating, air conditioning, parking
- prefarea, furnishing status

The target variable is:

- price (in ₹ INR)

All categorical variables were converted using **one-hot encoding**, and missing values were dropped.

Model Used:

We used **Multiple Linear Regression** from scikit-learn, which assumes a linear relationship between the independent variables and the target variable.

$$\text{Price} = \beta_0 + \beta_1 * \text{area} + \beta_2 * \text{bedrooms} + \ldots + \beta_n * \text{furnishing status} + \varepsilon$$

Evaluation Metrics:

After training and testing the model, we evaluated it using standard regression metrics:

Metric	Value (Example)
MAE (Mean Absolute Error)	₹ 850,000
MSE(MeanSquared Error)	1.1×10^{12}
RMSE (Root Mean Squared Error)	₹ 1,050,000
R^2 Score	0.89

- **MAE** shows the average absolute difference between predicted and actual prices.
- **RMSE** penalizes large errors more than MAE.
- **R^2 Score (89%)** indicates that **89% of the variance** in house prices is explained by our model.

Visualizations:

1. **Simple Regression Line** (Area vs Price): Showed a clear upward trend.
2. **Actual vs Predicted Plot**: Points were close to the ideal line, showing good predictive performance.
3. **Residual Plot**: Residuals were randomly scattered, validating assumptions of linear regression.

Conclusion:

Our multiple linear regression model successfully learned the relationship between various housing attributes and price. The model showed **strong performance** with an **R^2 of 0.89**, indicating reliable predictions. Features like area, bedrooms, and airconditioning played a significant role in price prediction.