House Price Prediction – Complete Explanation

1. Data Creation

You manually created a dataset containing Area, Bedrooms, Bathrooms, Age, Location, and Price.

It was converted into a DataFrame using pandas.

2. Exploratory Data Analysis (EDA)

- df.info() shows datatypes and confirms no missing values.

- df.describe() provides statistical summary.

- df.isnull() confirms no missing data.

Pairplot and Heatmap:

- Shows relationships between variables.

- Area has strong positive correlation with Price.

3. One-Hot Encoding

Location column was converted into dummy variables using get_dummies with drop_first=True to avoid the dummy variable trap.

4. Train-Test Split

Data was split into 80% training and 20% testing using train_test_split.

5. Model Training

A Linear Regression model was used. The model learned relationships between features and Price.

6. Model Evaluation

Results:

- MAE ~ 2.5 lakh

- RMSE ~ 3.1 lakh

- R^2 ~ 0.985 (excellent performance)

7. Coefficients

Area increases price.

Age decreases price.

Chennai, Delhi, Mumbai dummies show relative price differences compared to Bangalore (base).

8. Prediction

A new house with: Area=2500, Bedrooms=3, Bathrooms=2, Age=15, Location=Delhi was predicted to cost about ■60.31 lakh.

9. Saving Model

Model saved using joblib.dump.

This completes the full ML pipeline.