

# House Price Prediction using Linear Regression

## 1. Introduction

This project focuses on building a House Price Prediction system using a Linear Regression model. The goal is to understand the complete machine learning workflow including data loading, exploratory data analysis (EDA), model training, evaluation, and result interpretation.

## 2. Dataset Description

The dataset used for this project is a House Price Prediction Dataset provided in CSV format. It contains multiple features related to house characteristics, which are used to predict the target variable: house price.

## 3. Exploratory Data Analysis (EDA)

EDA was performed to understand the data distribution and relationships between variables. Missing values were checked and handled. Correlation analysis was conducted using a heatmap to identify important features influencing house prices.

## **4. Model Building**

A Linear Regression model was chosen due to its simplicity and interpretability. The dataset was split into training and testing sets. The model was trained using the training data and later used to predict house prices on unseen test data.

## **5. Model Evaluation**

The performance of the model was evaluated using standard regression metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and  $R^2$  Score. These metrics helped assess prediction accuracy and overall model performance.

## **6. Results**

The Linear Regression model achieved reasonable accuracy and demonstrated the relationship between input features and house prices. Visualization of actual vs predicted values showed how closely the model predictions matched real prices.

## **7. Conclusion**

This project successfully demonstrated the end-to-end implementation of a machine learning regression model. It provided hands-on experience in data preprocessing, EDA, model training, and evaluation.

## **8. Future Improvement Ideas**

Future improvements may include feature scaling, trying advanced regression models such as Ridge or Lasso, hyperparameter tuning, and deploying the model using a web interface for real-time predictions.