Zillow Predictive Power: House Price Prediction Model

# 1. Introduction

The objective of this project is to develop a machine learning model that accurately predicts house prices based on historical housing data from Zillow. The model leverages features such as location, square footage, number of bedrooms, and market indicators to offer actionable insights for real estate investors, buyers, and sellers.

# 2. Dataset Description

The dataset used in this project is sourced from Zillow Economic Data, containing information on housing prices and relevant property and market features. Key variables include location details, square footage, number of bedrooms, median sale prices, and more.

# 3. Data Preprocessing

Initial data exploration was performed to understand the structure and content of the dataset. Missing values were identified and handled by imputing the mean for numerical columns. The dataset was further cleaned by removing or treating outliers, and feature engineering was applied where appropriate.

# 4. Exploratory Data Analysis (EDA)

EDA involved visualizing distributions, identifying patterns and correlations, and understanding relationships among variables. Plots were generated to show the impact of square footage, number of bedrooms, and location on housing prices.

# 5. Model Development

Two models were developed and compared:  
- Linear Regression: Used as a baseline model. Provided interpretable coefficients but had limitations with non-linear data.  
- Random Forest Regressor: A robust ensemble model that handled non-linearity and feature interactions effectively. Hyperparameters such as number of estimators and tree depth were tuned for better performance.

# 6. Evaluation Metrics

The performance of the models was evaluated using the following metrics:  
- Mean Absolute Error (MAE)  
- Mean Squared Error (MSE)  
- Root Mean Squared Error (RMSE)  
  
The Random Forest model demonstrated superior predictive power over the Linear Regression model across all metrics.

# 7. Results and Insights

The final model provided accurate predictions and highlighted key influencing features such as property size, location, and number of bedrooms. The model offers valuable decision-support capabilities for stakeholders in the real estate domain.

# 8. Conclusion

This project successfully developed a predictive model for house prices using Zillow data. Through data preprocessing, exploratory analysis, and machine learning modeling, the project delivers a practical tool for understanding and anticipating real estate price movements.