**Home Price Prediction Using XGBoost**

**Introduction** This project focuses on predicting home prices using the XGBoost machine learning model. The dataset used for this prediction includes various factors such as the number of bedrooms, bathrooms, living space, lot size, floors, waterfront presence, view quality, condition, grade, year built, renovations, location, and nearby housing statistics.

**Dataset Overview** The dataset, "kc\_house\_data.csv," contains 21 columns, with the target variable being the home price. The key features considered for training the model include:

* **Bedrooms, Bathrooms, Sqft\_Living, Sqft\_Lot:** Structural attributes of the house.
* **Floors, Waterfront, View, Condition, Grade:** Quality and condition-related attributes.
* **Yr\_Built, Yr\_Renovated:** Historical construction and renovation data.
* **Zipcode, Latitude, Longitude:** Geographic location features.
* **Sqft\_Living15, Sqft\_Lot15:** Additional housing size data for nearby properties.

**Data Preprocessing** To ensure data quality and improve model performance, the following preprocessing steps were performed:

* Handling missing values by removing rows with NaN values.
* Standardizing numerical features using StandardScaler to normalize data.
* Splitting the dataset into training (80%) and testing (20%) subsets.

**Model Training and Evaluation** The XGBoost Regressor was chosen due to its efficiency in handling structured data. The model parameters were set as follows:

* Objective: reg:squarederror
* Number of estimators: 100
* Learning rate: 0.1
* Maximum depth: 5

After training the model, it was evaluated using the following metrics:

* **Mean Absolute Error (MAE):** Measures the average magnitude of errors.
* **Mean Squared Error (MSE):** Provides a quadratic loss function for evaluation.
* **R2 Score:** Indicates the proportion of variance explained by the model.

**Results** The model produced the following results on the test dataset:

* **Mean Absolute Error (MAE):** [Value to be filled after execution]
* **Mean Squared Error (MSE):** [Value to be filled after execution]
* **R2 Score:** [Value to be filled after execution]

**Conclusion** The XGBoost model successfully predicts home prices based on multiple influencing factors. The results indicate that the model performs well in explaining variance in house prices. Future improvements could include hyperparameter tuning, feature engineering, and using additional data sources for improved predictions.

**Future Work**

* Fine-tuning hyperparameters to optimize model performance.
* Exploring feature importance to refine input variables.
* Integrating external economic and market data for better accuracy.