**House Price Prediction Documentation**

**Introduction**

This document provides an overview of the process involved in predicting house prices using machine learning techniques. The dataset used for this analysis is the "kc\_house\_data.csv" dataset.

**1. Data Loading and Exploration**

Importing necessary libraries: NumPy, Pandas, Matplotlib, Seaborn, and Scikit-learn modules.

Loading the dataset using Pandas read\_csv function and displaying the first few rows to understand the data structure.

Descriptive statistics are computed to get an overview of numerical features like mean, median, min, max, etc.

Visualizing the distribution of the number of bedrooms using a bar plot.

Exploring the geographical data with joint plots for latitude and longitude.

**2. Data Visualization**

Scatter plots are created to visualize relationships between house price and different features such as square feet living area, location (latitude and longitude), number of bedrooms, waterfront status, floors, condition, and zipcode.

Bar plots are used to show the distribution of floors in the dataset.

**3. Preprocessing and Modeling**

Linear Regression model: Preprocessing steps include converting date values to binary, dropping unnecessary columns, splitting the data into training and testing sets, and fitting the Linear Regression model to predict house prices.

Gradient Boosting Regressor model: Similar preprocessing steps are applied, and the Gradient Boosting Regressor model is used for prediction.

**4. Evaluation and Validation**

The performance of both models is evaluated using the R-squared score (coefficient of determination).

The training and test set deviance for the Gradient Boosting Regressor model is plotted to visualize the model's learning process and overfitting.

**5. Principal Component Analysis (PCA)**

PCA is applied to the scaled feature data to reduce dimensionality and identify important components that explain the variance in the dataset.

**Conclusion**

House price prediction is a crucial task in real estate analysis, and machine learning models like Linear Regression and Gradient Boosting Regressor can provide valuable insights and accurate predictions based on various features. Proper data preprocessing, model selection, and evaluation techniques are essential for building reliable predictive models.