**Frequently Asked Questions (FAQs)**

**1. What is the main goal of this project?**

The goal of this project is to analyze and predict house prices in King County, USA, using a dataset containing various features of homes. The project involves data wrangling, exploratory data analysis, and building predictive models to understand how different features affect house prices and to forecast future prices.

**2. What steps did you take in the data preprocessing phase?**

In the data preprocessing phase, I performed the following steps:

* **Dropped unnecessary columns:** Removed columns like id and Unnamed: 0 that did not contribute to the analysis.
* **Handled missing values:** Replaced missing values in the bedrooms and bathrooms columns with the mean of their respective columns.
* **Converted data types:** Ensured that all columns had appropriate data types for analysis and modeling.

**3. How did you explore the data to understand the features better?**

I conducted exploratory data analysis (EDA) by:

* **Describing the dataset:** Used statistical summaries to understand the distribution and central tendencies of numerical features.
* **Visualizing correlations:** Created scatter plots and correlation matrices to explore relationships between different features and the target variable (price).
* **Boxplots:** Used boxplots to compare the distribution of house prices based on whether the property had a waterfront view.

**4. What predictive models did you use and how did you evaluate their performance?**

I used several predictive models, including:

* **Linear Regression:** Fit a linear regression model using features such as sqft\_living to predict house prices.
* **Ridge Regression:** Applied Ridge Regression to handle multicollinearity and regularize the model.
* **Polynomial Features:** Used polynomial features to capture non-linear relationships and improve model performance.

I evaluated the performance of these models using the R^2 score, which measures the proportion of variance in the target variable that is predictable from the features.

**5. What were the key findings from the analysis?**

Key findings from the analysis include:

* **Important Features:** sqft\_living, grade, and sqft\_above were found to be strongly correlated with house prices.
* **Impact of Waterfront View:** Properties with a waterfront view generally had higher prices compared to those without.
* **Model Performance:** The R^2 scores indicated that linear models performed reasonably well, but incorporating polynomial features and regularization improved the model’s ability to predict house prices.

**6. What improvements or future work could be done?**

Future improvements could include:

* **Feature Engineering:** Exploring additional features or creating new features to enhance model performance.
* **Advanced Models:** Experimenting with more complex models such as ensemble methods or neural networks.
* **Geospatial Analysis:** Incorporating geospatial analysis to better understand location-based factors affecting house prices.