**Estimating House Price — Variable Selection & Analysis**

**Objective**

The goal of this analysis is to identify the most important variables that can be used to estimate house prices effectively, based on exploratory data analysis and correlation analysis.

**Dataset**

The dataset used contains various features of houses, including:

* Structural attributes (sqft\_living, sqft\_above, bathrooms, bedrooms, floors)
* Quality metrics (grade, condition)
* Geographical attributes (zipcode, lat, long)
* View and waterfront indicators
* Renovation status and year built

**Analysis Steps**

**1. Data Cleaning**

* Checked for null values and handled missing data.
* Verified and corrected data types.

**2. Exploratory Data Analysis (EDA)**

* Visualized data distributions.
* Examined pairwise relationships between variables.
* Generated correlation matrix.

**3. Correlation Analysis**

* Computed Pearson correlation coefficients between features and house price.

**Key Insights**

**Variables with Strong Positive Correlation with House Price**

| **Variable** | **Correlation with Price** |
| --- | --- |
| sqft\_living | High |
| grade | High |
| sqft\_above | High |
| bathrooms | Moderate to High |
| view | Moderate |
| sqft\_living15 | Moderate |

**Interpretation:**

* Larger homes (sqft\_living, sqft\_above) tend to command higher prices.
* Quality indicators (grade) strongly influence price.
* More bathrooms and good views positively impact price.

**Variables with Weak or No Significant Correlation**

| **Variable** | **Correlation with Price** |
| --- | --- |
| zipcode | Low |
| yr\_renovated | Low |
| condition | Low |
| floors | Low |

**Interpretation:**

* Geographical grouping (zipcode) is too coarse to directly predict price.
* Condition and renovation year are less consistently related to price in this dataset.

**Recommended Features for Model**

For predicting house price, the following features are recommended:

* sqft\_living
* grade
* sqft\_above
* bathrooms
* view
* sqft\_living15
* Optionally: bedrooms, waterfront, latitude

**Visual Insights**

