House Prices Dataset Summary

This dataset contains information about real estate properties sold in a specific region, including various features that affect the price of the homes.

The dataset consists of 18 columns, representing different attributes of the properties such as:

1. Price: The sale price of the house.
2. Bedrooms: The number of bedrooms in the house.
3. Bathrooms: The number of bathrooms.
4. sqft_living: The living area of the house (in square feet).
5. sqft_lot: The total land area of the property (in square feet).
6. Floors: The number of floors in the house.
7. Waterfront: Whether the property is located near a waterfront.
8. View: The view quality of the property (scale from 0 to 4).
9. Condition: The condition of the house (scale from 1 to 5).
10. sqft_above: The area of the house excluding the basement (in square feet).

Machine Learning Problem

The problem at hand is to predict the price of a house based on the given features. This is a regression problem, where the goal is to use the features

(such as the number of bedrooms, square footage, etc.) to predict the sale price of the house.

Methodology Diagram (Steps in the Project):

1. Data Loading: Load the dataset using pandas.
2. Data Cleaning: Handle missing or incorrect data.
3. Data Exploration and Analysis: Use exploratory data analysis (EDA) tools to understand the

relationship between the features and the house price.

4. Build Machine Learning Model: Use regression algorithms such as Linear Regression or more advanced models like Random Forest.

5. Model Evaluation and Improvement: Evaluate model performance using metrics such as MSE or RMSE and fine-tune the model for better accuracy.

6. Price Prediction: Use the trained model to predict house prices based on input data.