**Exercise: Predicting House Prices**

1. **Problem Statement:** You have been tasked with developing a linear regression model to predict house prices based on various features. The dataset provided includes information such as the size of the house, number of bedrooms, location, and other relevant factors.
2. **Dataset Description:**
   * **Features:**
     + Size of the house (in square feet)
     + Number of bedrooms
     + Number of bathrooms
     + Location (as categorical or numerical feature)
     + Age of the house (in years)
     + Distance to the nearest school (in miles)
   * **Target Variable:**
     + House price (in dollars)
3. **Tasks:** a. Load and explore the dataset. Perform data preprocessing steps such as handling missing values, encoding categorical variables, and scaling numeric features if necessary.

b. Split the dataset into training and testing sets (e.g., 70% training, 30% testing).

c. Implement a linear regression model using a suitable programming language or library (e.g., Python with scikit-learn).

d. Train the model on the training data and evaluate its performance using appropriate metrics (e.g., Mean Absolute Error, R-squared).

e. Interpret the coefficients of the linear regression model to understand the impact of each feature on the house price.

f. Utilize the trained model to make predictions on the test set and assess its predictive accuracy.

1. **Additional Considerations:**
   * Explore the distribution of house prices and other relevant features through visualizations like histograms, scatter plots, or correlation matrices.
   * Discuss any assumptions made during the analysis and potential challenges in predicting house prices accurately using linear regression.
2. **Deliverables:**
   * Python code implementing the above tasks.
   * A report summarizing the findings, including model performance metrics, interpretations of coefficients, insights into feature importance, and any recommendations for improving the model.