House Price Prediction and Data Visualization

This Python script performs house price prediction using linear regression and includes data visualization of the Boston Housing Dataset. The script uses scikit-learn for dataset loading and modeling, seaborn for data visualization, and other libraries for data manipulation and analysis.

Dataset Used

The code utilizes the **Boston Housing Dataset** from scikit-learn. This dataset contains various features related to housing, such as crime rate, number of rooms, and more. The target variable is the **median value of owner-occupied homes** in thousands of dollars.

Code Structure

The code can be divided into the following key sections:

1. **Data Loading and Preprocessing**:
    a. Load the Boston Housing Dataset using scikit-learn.
    b. Handle missing values by imputing with the mean.
    c. Perform one-hot encoding for categorical columns.
2. **Data Visualization**:
    a. Visualize the distribution of selected categorical columns using countplots.
    b. Create box plots to compare selected numerical columns with house prices.
3. **Data Splitting**:
    a. Split the dataset into training and testing sets for model training and evaluation.
4. **Model Building**:
    a. Create a linear regression model using scikit-learn.

b. Fit the model to the training data.

5. **Model Evaluation**:
    a. Make predictions on the test data.
    b. Calculate Mean Squared Error (MSE) and R-squared (R2) as performance metrics.
6. **Visualization of Predictions**:
    a. Create a scatterplot to visualize actual vs. predicted house prices.

Usage

1. Clone this repository or download the Python script (e.g., **house_price_prediction.py**).
2. Ensure that you have the required libraries installed (e.g., scikit-learn, seaborn).
3. Run the script using Python:

python house_price_prediction.py

4. The script will load the dataset, perform data visualization, train a linear regression model, evaluate its performance, and display a scatterplot of actual vs. predicted house prices.
5. **Dependencies**

- scikit-learn
- seaborn

License

This project is licensed under the MIT License - see the LICENSE file for details.

Acknowledgments

- scikit-learn - Machine learning library for Python.
- seaborn - Data visualization library for Python.
- scikit-learn Boston Housing Dataset - Boston Housing Dataset in scikit-learn.