House Price Prediction Test

Problem Description

Build a complete machine learning pipeline using the provided housing data to estimate home sale prices based on various property characteristics. Your primary objective is to develop a robust prediction model. **It is recommended that the final model use linear regression, although using a different model will not be rewarded or punished.**

Your solution will be evaluated on the following areas:

- **Data Handling and Exploration:** Import, clean, and prepare the dataset for analysis. Perform exploratory data analysis to identify trends and patterns.
- **Feature Engineering:** Create and select meaningful features from the dataset.
- **Modeling:** Develop your prediction model, applying best practices in model development.
- **Evaluation:** Assess the performance of your model using appropriate metrics.

Your approach should demonstrate technical proficiency while reflecting a thoughtful application of sound modeling practices throughout the entire pipeline. **Every assumption and modification made during the process must be clearly outlined.**

As output, a jupyter notebook-based project is preferred, but raw python is acceptable as well. Ensure all relevant outputs are included, the code is runnable, and any required packages are present in a requirements file.

You have 24 hours to complete this assessment. You can request a one-time 24 hours extension if needed. Please submit your project in a compressed folder to talent@openhouse.ai. If you cannot submit it due to size limitations please upload to a private file sharing system like Google Drive and share with the file available to anyone with the link. Please do not provide a publicly available or private github link as sharing this amongst our team has caused problems in the past.