Property Price Prediction: From Data Cleaning to Model Evaluation

For my data science project, I worked on predicting house prices using a dataset containing features such as the number of bedrooms, bathrooms, location, size, and other details about each house. My goal was to build a machine learning model that could estimate a house's selling price based on its features.

Step 1: Data Cleaning
1. Cleaned column names to remove spaces and special characters.
2. Handled missing values: filled numeric columns with median, categorical columns with mode or 'no'. 3. Fixed inconsistent text and ensured correct numeric data types.

Step 2: Feature Engineering
- Created house_age from year built.
- Converted Yes/No columns to binary 1/0.
- One-hot encoded multi-category columns like location and furnishing.

Step 3: Splitting the Dataset
- Used 80% of the data for training and 20% for testing with train_test_split.

Step 4: Building and Training the Model
- Applied Linear Regression.
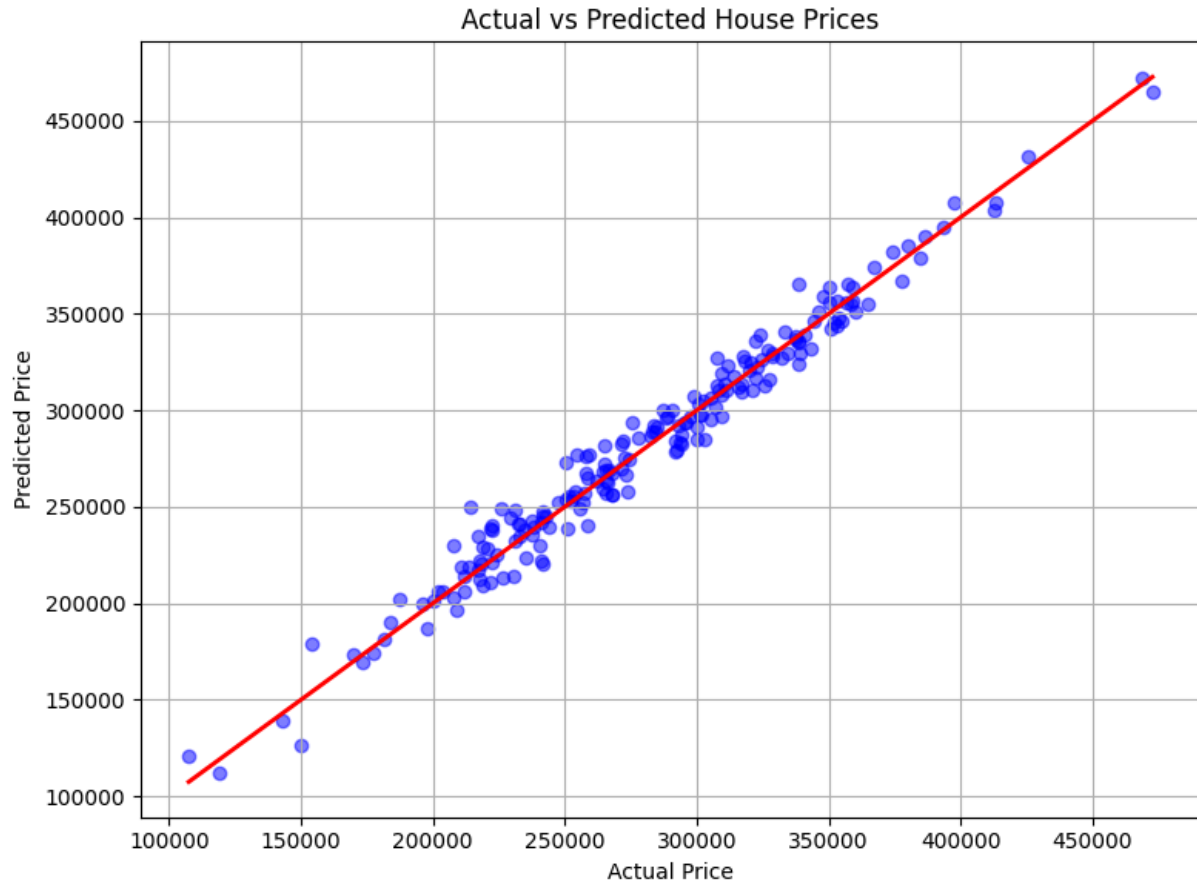- Trained the model using the training dataset.

Step 5: Model Evaluation
We measured the model performance using:
- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- R^2 Score
Below is the summary table:

Metric	Value
Mean Absolute Error (MAE)	7,718.49
Mean Squared Error (MSE)	97,746,453.88
R^2 Score	0.9739

Actual vs Predicted House Prices

Actual vs Predicted House Prices

Step 6: Conclusion
The Linear Regression model achieved an R² score of 0.9739, showing strong predictive power. The charts indicate that predictions closely match actual prices, with relatively small errors.