📊 Linear Regression – Actual vs Predicted Analysis with Metrics

# 📈 Model Evaluation Metrics

These metrics summarize the performance of our Linear Regression model trained on the California Housing dataset.  
The target variable is the median house value (in $100,000 units).

- 📉 \*\*Mean Absolute Error (MAE): 0.5332\*\*   
 On average, the predictions are off by ~$53,320. MAE is easy to understand and good for measuring real-world prediction error.  
  
- 📉 \*\*Mean Squared Error (MSE): 0.5559\*\*   
 Squared error; penalizes larger errors more. Useful during optimization, though not in original units.  
  
- 📏 \*\*Root Mean Squared Error (RMSE): 0.7456\*\*   
 Square root of MSE, bringing the error back into original units. Our model is, on average, off by ~$74,560.  
  
- 📊 \*\*R² Score: 0.5758\*\*   
 The model explains about 57.58% of the variation in housing prices. This is a moderate result and indicates room for improvement.

# 📉 Actual vs Predicted House Prices – Plot Interpretation

This plot compares the actual house prices (x-axis) with the predicted prices (y-axis).   
Each dot represents a house, and the red dashed line is the ideal line (perfect predictions).  
The closer the dots are to the red line, the better the model is performing.

## ✅ 1. Decent Alignment (1.5 to 4 range)

In the range of actual values from about 1.5 to 4 (i.e., $150K to $400K), the predictions generally align well with the red line.  
This means the model is doing a fairly good job for mid-range house prices.

## ❌ 2. Clipping at 5

There’s a sharp vertical line at 5 (i.e., $500K), showing that actual house prices were capped at this value.  
But the model still tries to predict beyond 5, which leads to overpredictions and scattered dots above the red line.  
This results in \*\*error spikes\*\* at the upper end.

## ⚠️ 3. Spread Below 1.5

For lower house prices (below $150K), the model sometimes underpredicts.  
Dots tend to fall under the red line, indicating the model didn't fully capture the pricing patterns in that lower range.

# 📌 Summary Insight

- The model captures the overall trend, but struggles at the extremes — especially for very high or low house values.  
- This is expected from a basic Linear Regression model that assumes a linear relationship.  
- Next steps could include:  
 - Applying \*\*regularization\*\* (like Lasso or Ridge Regression)  
 - Trying \*\*non-linear models\*\* (Decision Trees, Random Forests)

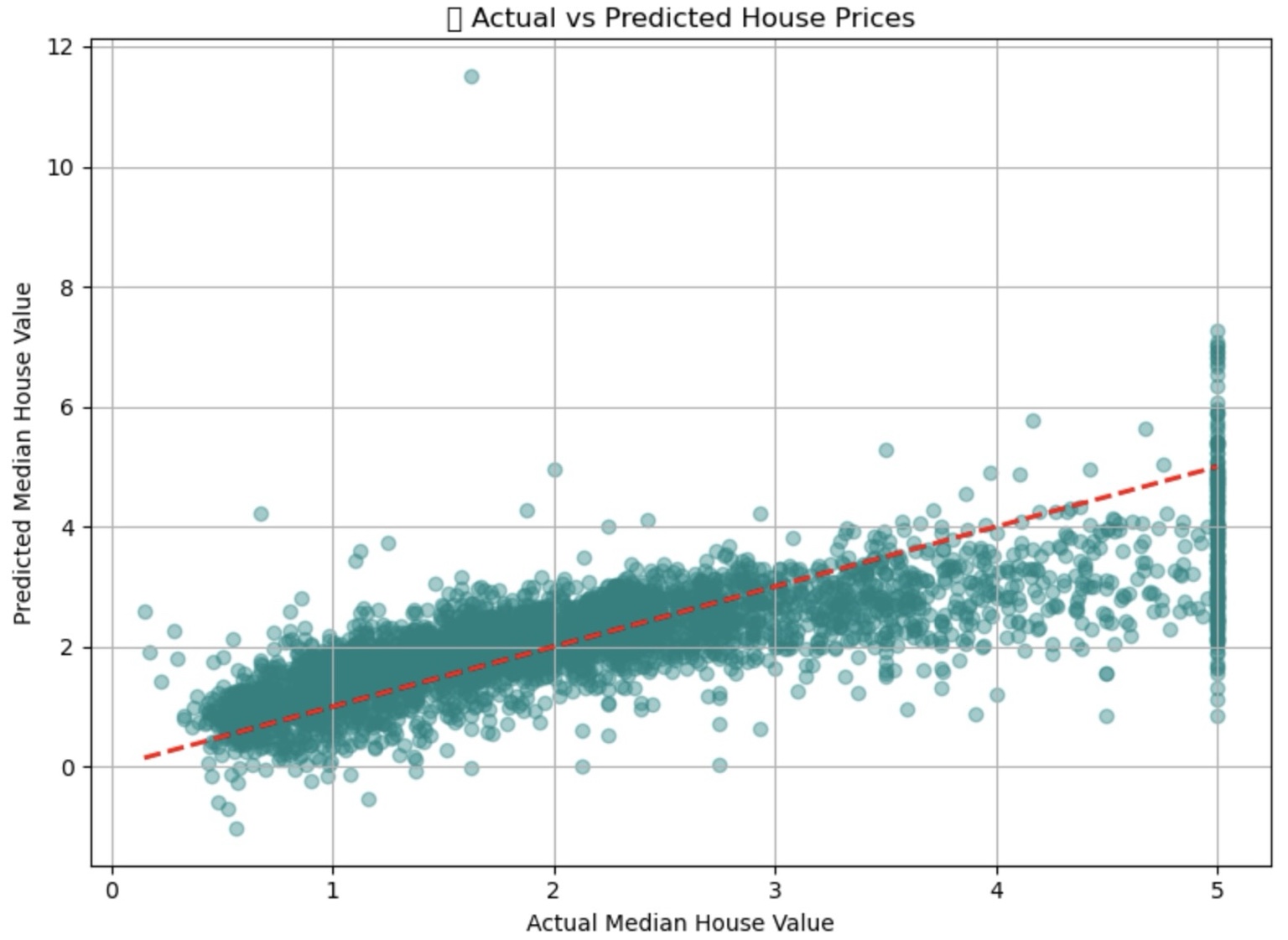


Figure: Actual vs Predicted House Prices using Linear Regression on California Housing Data.