# 📊 Linear Regression Report – Housing Dataset

This document compares the original Mini Guide steps with the actual code implementation for the Housing dataset and highlights extra enhancements for deeper insights.

## 📋 Comparison Table

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Mini Guide Description | Implemented Code | Extra Feature |
| 1 | Import and preprocess the dataset | Used pandas to load, removed NaNs, label encoded categorical features | Handled infinite values; encoded ocean\_proximity |
| 2 | Split data into train-test sets | Used train\_test\_split from sklearn for both simple and multiple models | Stratified splitting not needed as it's regression |
| 3 | Fit a Linear Regression model using sklearn.linear\_model | Used LinearRegression() for both simple and multiple models | Fitted separately for simple (1 feature) and multiple (all features) |
| 4 | Evaluate model using MAE, MSE, R² | Used sklearn.metrics: MAE, MSE, R² | Added RMSE for better error interpretation |
| 5 | Plot regression line and interpret coefficients | Plotted simple regression line using matplotlib, displayed coefficients | Added residual distribution plots using Seaborn |

## 🧾 Summary

The implemented Linear Regression task not only followed all the suggested steps from the mini guide but also introduced useful enhancements like RMSE evaluation, residual plotting, and encoding categorical data. This makes the dataset fully ready for both linear modeling and further enhancements like regularization or cross-validation