KEY INSIGHTS FROM RESIDUAL ANALYSIS & DIAGNOSTICS

After fitting the Multiple Linear Regression model, I checked the residuals to see how well the model was learning the pattern in the data. The residual plot showed that most of the points were scattered around zero, which means the model was not making any major systematic errors. However, there were a few predictions where the residuals were larger, which suggests the model struggled slightly with some extreme or very high-priced houses. This is normal in real-estate datasets because expensive properties usually behave differently compared to normal houses. The residual distribution looked almost symmetric, which indicates that the model errors follow a reasonably normal pattern. This is a good sign because it means our linear regression assumptions are mostly being satisfied. Overall, the diagnostics showed that the model is performing consistently and not overfitting too much.

HOW REGULARIZATION IMPROVED MODEL STABILITY

Regularization techniques like Ridge, Lasso, and ElasticNet helped the model become more stable by controlling the effect of highly correlated features and reducing unnecessary complexity. **Ridge Regression** reduced the size of large coefficients, which helped handle multicollinearity and made the model more generalizable. **Lasso Regression** actually removed some unimportant features by shrinking their coefficients to zero. This helped simplify the model. **ElasticNet** combined the strengths of both Ridge and Lasso. It balanced feature selection and coefficient shrinkage, giving a much more reliable model on unseen data. Overall, regularization prevented the model from memorizing the training data and made the predictions smoother and more stable, especially for houses with unusual feature combinations.

RATIONALE FOR FINAL CHAMPION MODEL SELECTION

To pick the best model, I compared all models — baselines, SLR, MLR, Ridge, Lasso, and ElasticNet — using RMSE on the test dataset. The Champion Model was chosen based on lowest RMSE, meaning it made the smallest prediction errors on unseen test data. The winning model consistently performed well across:

- Test RMSE
- MAE
- R^2 sco
- Cross-validation results

and showed no major issues in the residual analysis. This means the final selected model does not just perform well on the training data, but also generalizes nicely to new houses it has never seen before. Overall, the Champion Model was selected because it provided the best balance between accuracy, stability, and generalization, which is exactly what we need in a real-world regression solution.