

Summary Report

Data Preprocessing

- Missing values handled using median (numerical) and most frequent (categorical) strategies.
- Categorical variables encoded using One-Hot Encoding.
- Numerical features scaled using StandardScaler.

Model Evaluation

- Linear Regression

- MAE: 979679.69
- RMSE: 1331071.42
- R^2 Score: 0.65

- Decision Tree

- MAE: 1257110.09
- RMSE: 1710758.39
- R^2 Score: 0.42

Findings

- **Linear Regression:** Performs relatively well with an R^2 score of 0.65, meaning it explains 65% of the variance in house prices. However, the RMSE is still quite high, indicating prediction errors.

- **Decision Tree:** Has a lower R^2 score (0.42), suggesting it overfits the training data and lacks generalization. It also has a higher MAE and RMSE than Linear Regression, making it less suitable for this dataset.

- Limitations:

- The models do not capture complex non-linear relationships effectively.
- Feature engineering and hyperparameter tuning (especially for Decision Trees) could improve performance.
- Additional data, feature selection, and trying ensemble methods (Random Forest, Gradient Boosting) might yield better results.