Residential Real Estate Price Prediction — Model Validation

Real Estate Investment Trust • Investment Analytics Project

Overview

- This stage validates the selected predictive and benchmark models to ensure their reliability, robustness, and suitability for real-world deployment.
- Model diagnostics were conducted to verify residual behavior, stability, and generalization without altering the underlying data.

Objective

- Model predictions are unbiased and stable.
- Linear model assumptions are sufficiently satisfied for interpretation.
- The selected predictive model generalizes well to unseen data.
- No further feature removal or outlier treatment is required.

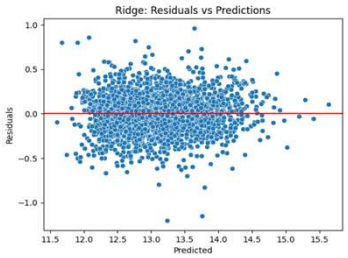
Results

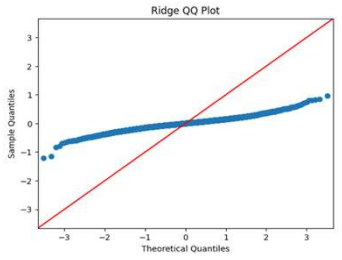
- Residuals are centered around zero, indicating no systematic bias.
- Residual variance is stable across predicted values.
- Log-transformation effectively stabilizes variance and normalizes error distribution.
- Train–test performance indicates stable generalization with no evidence of harmful overfitting.
- Multicollinearity is present among size-related features but is mitigated through Ridge regularization.
- Removing redundant features (e.g., year built vs. house age) preserves model performance while improving interpretability.

Model Generalization Check (Train vs Test RMSE)

Model	Train RMSE	Test RMSE
Ridge Regression	0.1805	0.1739
XGBoost	0.1062	0.1580Q

- **Ridge:** Very stable performance; minimal train–test gap indicates strong generalization and low overfitting risk.
- **XGB:** Expected train–test gap for boosted models; test performance remains strong, indicating controlled complexity.





Next Steps

- Interpreting Ridge regression coefficients to explain key price drivers.
- Translating model outputs into business-oriented insights.
- Presenting findings to support pricing strategy and investment decisions.