Residential Real Estate Price Prediction — EDA Summary

Real Estate Investment Trust • Investment Analytics Project

Project Overview:

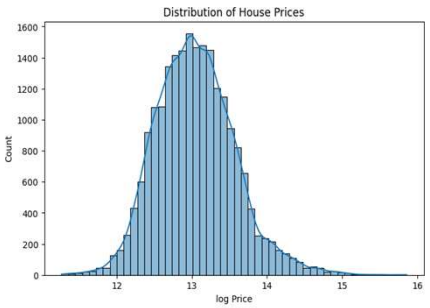
- This project focuses on analyzing residential property data to understand price behavior and key value drivers.
- Exploratory Data Analysis (EDA) was performed to validate data quality, study relationships, and prepare the dataset for modeling.

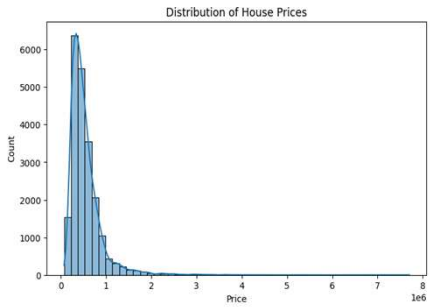
Key Insights

- House prices exhibit a strong right-skewed distribution, driven by a small number of high-value luxury properties.
- Applying a log transformation to price significantly reduced skewness and revealed clearer linear relationships with key predictors.
- Living area, construction quality (grade), and number of bathrooms show the strongest relationships with property prices.
- Waterfront properties consistently command a substantial price premium compared to non-waterfront homes.
- Size-related variables show strong overlap, indicating multicollinearity among square footage features.
- Location-related variables also display redundancy when represented in multiple forms.
- A small number of extreme observations were identified and flagged for further evaluation rather than removed.

Details

- Distribution analysis confirmed that raw prices are highly skewed; log-transformed prices are more stable and suitable for modeling.
- These extreme values represent luxury or unique properties rather than widespread data issues.
- Missing values were detected during EDA and addressed by removing the affected column, resulting in a clean dataset.
- Extreme values were flagged rather than removed to preserve data integrity and allow model-level evaluation.





Next Steps

- Create derived features to reduce multicollinearity and better capture property value (e.g., total area, ratios, and age-related features).
- Encode categorical and binary variables appropriately for modeling.
- Prepare a final, model-ready feature set aligned with modeling requirements.