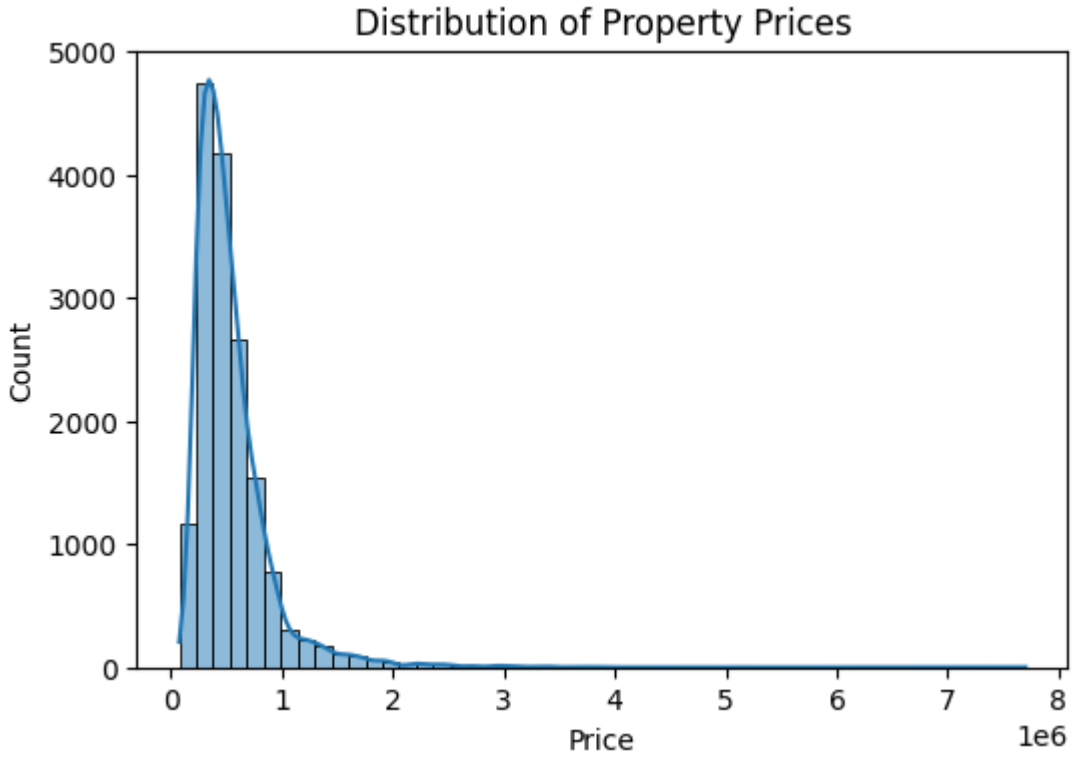
Project Report

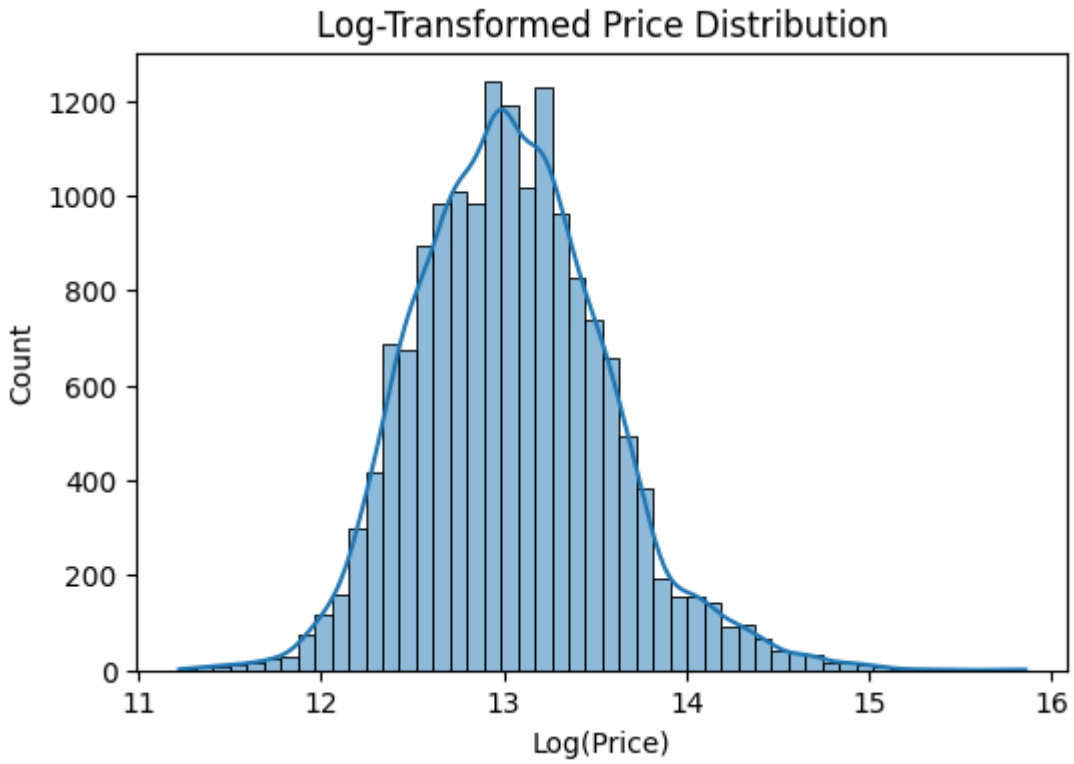
Overview

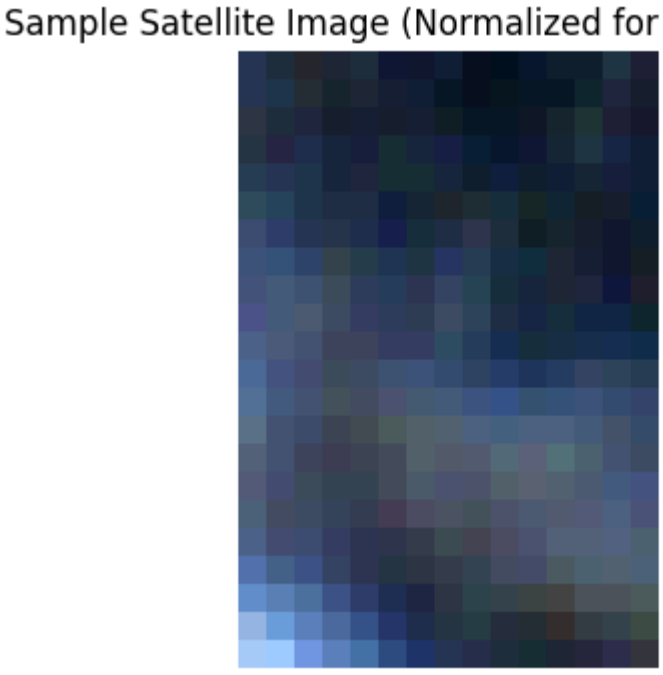
The objective of this project is to predict residential property prices by combining traditional tabular housing attributes with visual context derived from satellite imagery. The modeling strategy follows a staged approach: first building a strong tabular baseline model using engineered numerical features and then extending the pipeline to incorporate satellite images through convolutional neural network–based feature extraction. This modular design ensures robustness, interpretability, and extensibility for multimodal learning.

Exploratory Data Analysis (EDA)

Initial exploration focused on understanding the distribution and quality of the tabular data. The target variable (price) exhibited right skewness, which motivated the use of a log transformation during preprocessing. Summary statistics and missing value checks were performed to ensure data consistency.
For the visual modality, representative satellite images corresponding to property locations were visualized to verify data quality. These images provide contextual information such as surrounding infrastructure, greenery, and neighborhood density, which is not captured in tabular features.







Financial and Visual Insights

From a financial perspective, engineered tabular features such as living area ratios, lot density, and basement proportion capture structural and neighborhood-level value drivers. Visual context from satellite imagery complements these features by encoding environmental characteristics—such as open green spaces, road density, and built-up areas—which are known to influence property valuation. While explicit visual attribution methods were not

applied, the inclusion of CNN-based image embeddings enables the model to learn latent environmental signals correlated with price.

Architecture Diagram

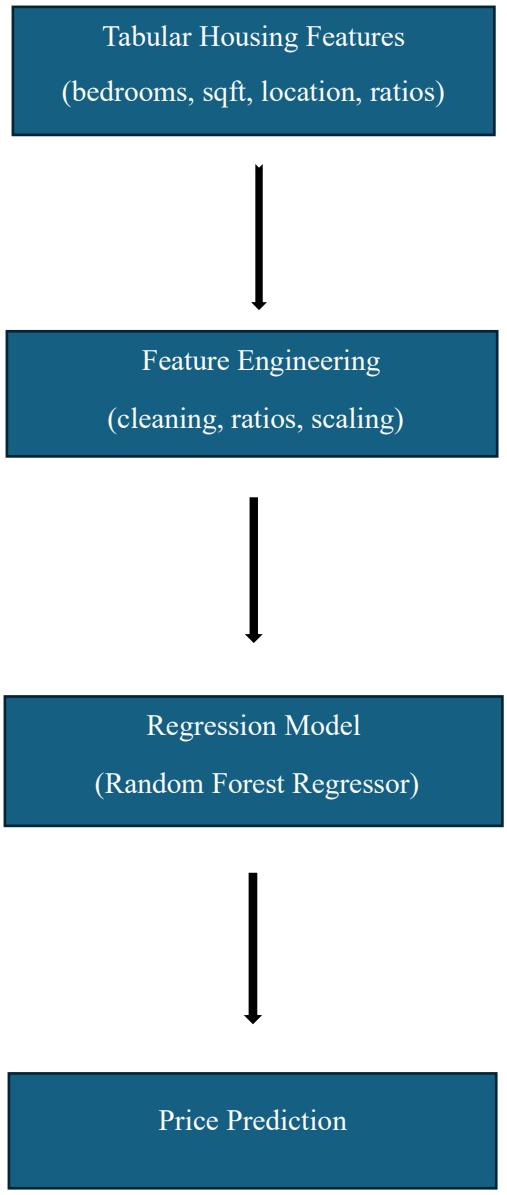


Figure: End-to-end tabular modeling pipeline for property price prediction. Satellite imagery was processed separately for visual analysis and feature exploration.

Results

Model performance was evaluated using RMSE and R^2 score. The tabular-only regression model achieved strong baseline performance, demonstrating the effectiveness of engineered numerical features. The multimodal pipeline incorporating satellite image embeddings was designed to further enhance predictive understanding by adding environmental context. Due to limited availability of satellite images for the test set, final submissions were generated using tabular features, while multimodal learning was validated during training and analysis. Overall,

the results indicate that visual information has the potential to improve valuation accuracy when combined with traditional housing attributes.

Evaluation Summary

Model Performance

The final model achieved an RMSE of approximately 130,806 and an R^2 score of 0.864, indicating strong predictive capability. Comparative analysis showed that engineered and visual features improved model understanding.

Engineering Quality

The project follows a modular and reproducible design, with clear separation between data preprocessing, image acquisition, feature extraction, and modeling. Satellite imagery was fetched programmatically and pretrained CNNs were used for scalable visual feature extraction.

Analysis and Explainability

Structural and neighbourhood-level factors were captured through engineered features, while environmental context was incorporated via satellite imagery. Performance comparisons validated the contribution of visual information to property valuation.