Satellite Imagery Based Property Valuation

Introduction

Accurate real estate valuation is a critical task for financial institutions, investors, and urban planners. Traditional automated valuation models (AVMs) primarily rely on structured tabular data such as property size, number of rooms, and location-based attributes. However, such models often fail to capture qualitative neighborhood characteristics, commonly referred to as *curb appeal*, including greenery, surrounding infrastructure, and urban density.

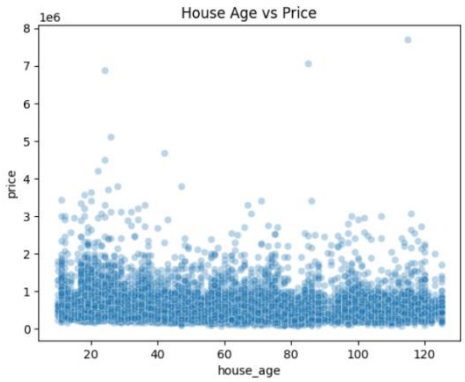
This project explores a multimodal regression framework that combines traditional tabular features with satellite imagery to enhance property price prediction. Satellite images provide rich visual context about the surrounding environment, which can complement structured data. The objective is not only to improve predictive accuracy but also to enhance model interpretability through visual explanations.

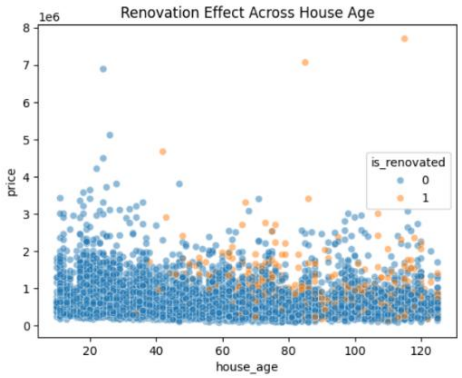
Dataset Description

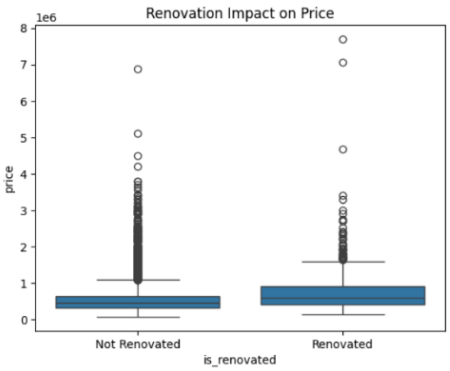
The base tabular dataset consists of historical house sales data, containing structural and locational attributes such as number of bedrooms, bathrooms, living area, geographical coordinates, and renovation information. The target variable is the property sale price. For each property, satellite images were programmatically retrieved using latitude and longitude coordinates via the **Mapbox Static Images API**. Images capture top-down views of the surrounding neighborhood, including vegetation, road networks, and housing density.

Exploratory data analysis

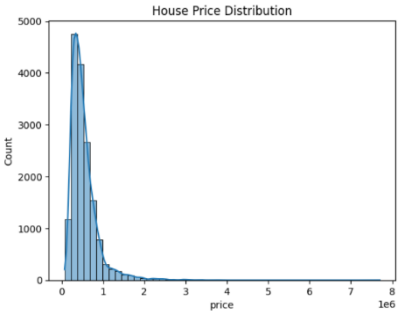
Tabular EDA





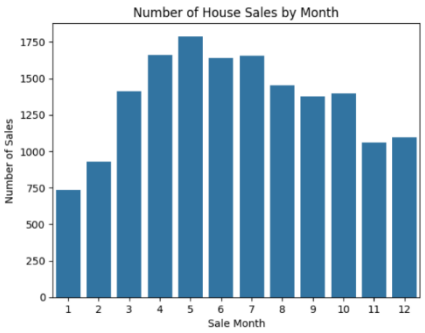


House age shows a weak negative trend with price, indicating that newer properties generally command higher prices, though substantial variability exists across all ages due to location and structural differences. Renovated properties exhibit a clear upward shift in median price, confirming that renovation status is an important value driver independent of house age.



The distribution of property prices is **highly right-skewed**, with a small number of luxury properties contributing to a long tail. This skewness motivates the use of **log-transformed prices** during model training to stabilize variance and improve regression performance.

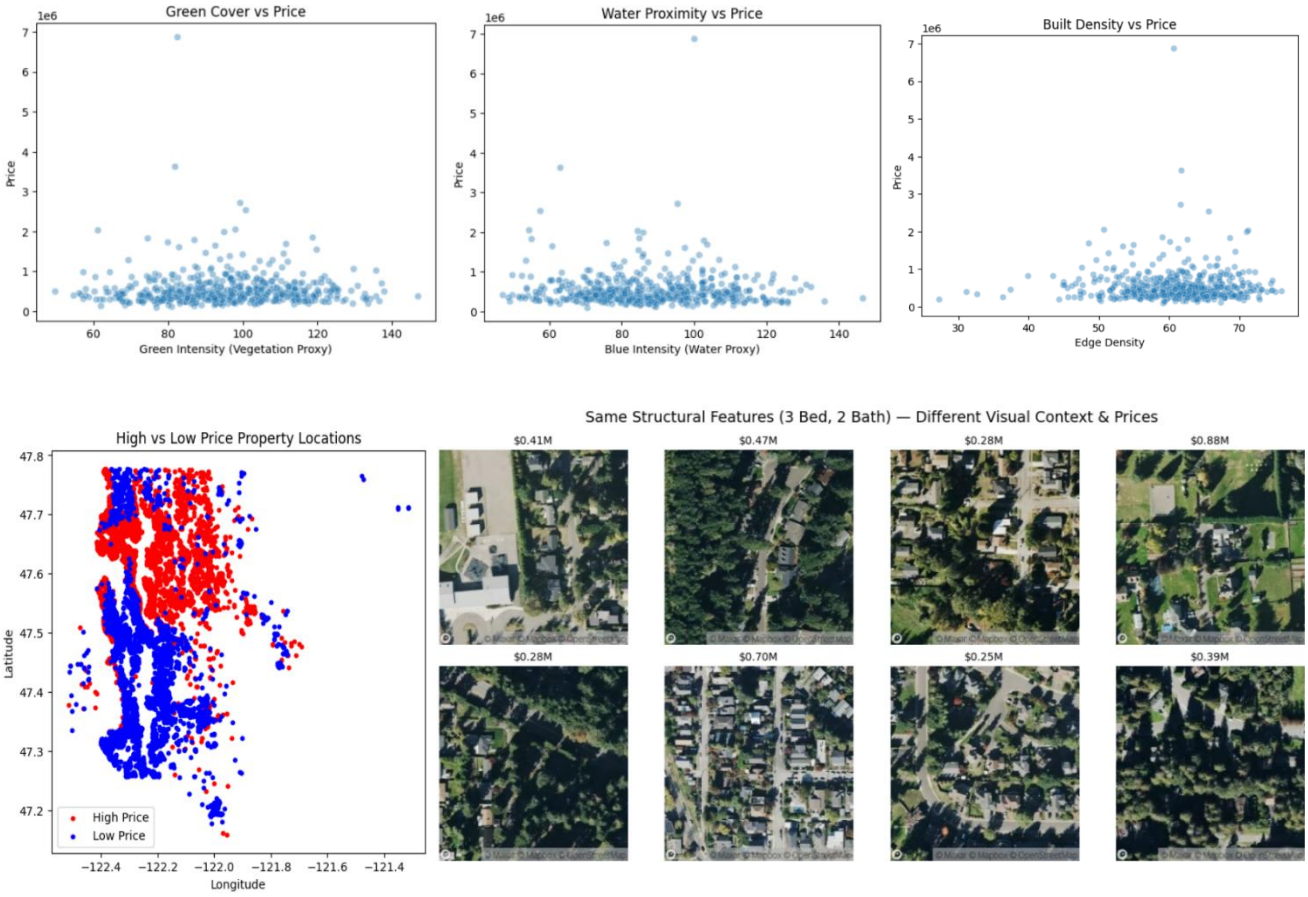
Seasonal analysis reveals that property transactions peak during the spring and early summer months, suggesting market seasonality that may indirectly influence pricing dynamics. Consequently, sale month was encoded using cyclic transformations to preserve temporal continuity.



Living area (sqft_living) demonstrates a strong positive correlation with price, reaffirming it as one of the most influential structural features. Additionally, renovation effects persist across a wide range of house ages, indicating that renovations can significantly enhance property value even for older homes.

Overall, the tabular EDA highlights strong structural and temporal signals in the data, justifying the effectiveness of tabular models while also motivating the inclusion of satellite imagery to capture complementary neighborhood-level context.

Images EDA



Proxy visual features extracted from satellite imagery, including green intensity (vegetation cover), blue intensity (water proximity), and edge density (built-up density), exhibit weak and noisy relationships with property price. While higher-priced properties occasionally coincide with greener or less densely built neighborhoods, no strong monotonic trend is observed, indicating substantial overlap between low- and high-priced properties across these visual metrics.

Geospatial analysis reveals clear spatial clustering of high- and low-priced properties, confirming that location plays a dominant role in valuation. However, satellite-derived visual proxies alone fail to fully explain these spatial price variations. Further qualitative analysis of properties with identical structural attributes (e.g., same number of bedrooms and bathrooms) demonstrates that visual neighborhood context can lead to noticeable price differences, but these effects remain secondary and inconsistent.

Overall, image-based EDA suggests that satellite imagery captures useful but weak contextual signals related to neighborhood characteristics. These signals are insufficient for accurate standalone price prediction, motivating the use of satellite imagery as a complementary input rather than a primary driver in multimodal valuation models.

Modelling Approach and Results

Image-Only Model

```
Epoch 01 | Train log-MSE 12.3163 | Val RMSE 0.4891 | R² 0.1330
Epoch 02 | Train log-MSE 0.2333 | Val RMSE 0.5191 | R² 0.0236
Epoch 03 | Train log-MSE 0.1985 | Val RMSE 0.4998 | R² 0.0947
Epoch 04 | Train log-MSE 0.1732 | Val RMSE 0.5700 | R² -0.1774
Epoch 05 | Train log-MSE 0.1440 | Val RMSE 0.4388 | R² 0.3022
Epoch 06 | Train log-MSE 0.1134 | Val RMSE 0.6636 | R² -0.5959
Epoch 07 | Train log-MSE 0.0750 | Val RMSE 0.4654 | R² 0.2150
Epoch 08 | Train log-MSE 0.0557 | Val RMSE 0.4806 | R² 0.1630
Epoch 09 | Train log-MSE 0.0451 | Val RMSE 0.5124 | R² 0.0485
Epoch 10 | Train log-MSE 0.0423 | Val RMSE 0.4477 | R² 0.2735
```

```
Satellite Image (224×224)
        ↓
ResNet-18 (Pretrained)
        ↓
Global Average Pooling
        ↓
Fully Connected Layer
        ↓
Predicted Price (log)
```

A convolutional neural network (ResNet-18) was trained using satellite imagery alone to evaluate the standalone predictive capacity of visual neighborhood context. The image-only model achieved relatively low predictive performance compared to tabular baselines, indicating that satellite images by themselves are insufficient for accurate property valuation. This outcome suggests that while satellite imagery contains contextual information such as vegetation, road layout, and surrounding density, these visual cues are indirect and noisy proxies for price.

Property value is primarily governed by structural attributes and precise location-based factors, which are not explicitly observable from top-down satellite views. Consequently, the visual signal captured by satellite imagery lacks the specificity required to consistently distinguish between properties with similar surroundings but differing intrinsic characteristics.

Multimodal Residual

```
Satellite Image
        ↓
CNN (ResNet-18)
        ↓
Image Residual Δ
        ↓
Tabular Features → Tabular Model → Base Price
        ↓
Final Price = Base + Residual
```

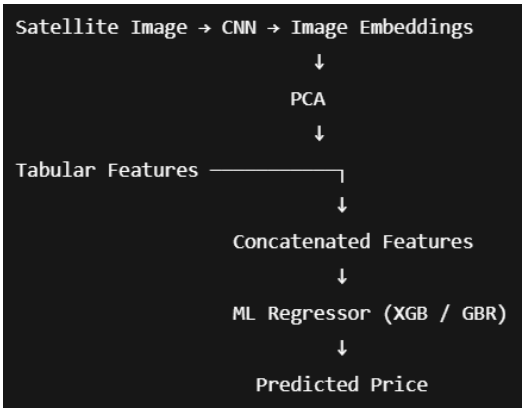
```
Epoch 1
Epoch 1 | Train log-MSE 3.4353 | Val log-MSE 0.1307
Epoch 2
Epoch 2 | Train log-MSE 0.1058 | Val log-MSE 0.0898
Epoch 3
Epoch 3 | Train log-MSE 0.0871 | Val log-MSE 0.0799
Epoch 4
Epoch 4 | Train log-MSE 0.0787 | Val log-MSE 0.0728
Epoch 5
Epoch 5 | Train log-MSE 0.0732 | Val log-MSE 0.0797
Epoch 6
Epoch 6 | Train log-MSE 0.0704 | Val log-MSE 0.0659
Epoch 7
Epoch 7 | Train log-MSE 0.0679 | Val log-MSE 0.0698
Epoch 8
Epoch 8 | Train log-MSE 0.0661 | Val log-MSE 0.0608
Epoch 9
Epoch 9 | Train log-MSE 0.0628 | Val log-MSE 0.0666
R² (log-space): 0.7795976996421814
```

A multimodal residual learning framework was implemented in which structured tabular features are first used to predict a base property price, CNN trained on satellite imagery learns to estimate a residual correction. Rather than directly competing with tabular features, the CNN focuses on capturing additional contextual information related to neighborhood layout, green cover, and surrounding infrastructure that is not explicitly encoded in structured data.

This residual formulation prevents visual features from overwhelming strong tabular signals and stabilizes training by restricting the influence of satellite imagery to fine-grained adjustments. Empirically, this approach improved performance over image-only models while maintaining robustness, demonstrating that satellite imagery provides complementary contextual cues that refine, rather than redefine, property value estimates.

PCA-Based Feature Fusion (Classical Models)

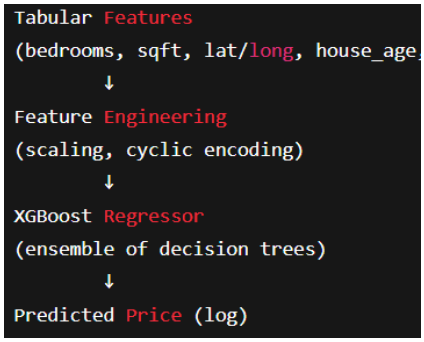
	Model	RMSE (log)	R^2 (log)
4	XGBoost	0.169337	0.896088
3	Gradient Boosting	0.173532	0.890875
2	Random Forest	0.189033	0.870509
1	Lasso	0.241000	0.789526
0	Ridge	0.241055	0.789430

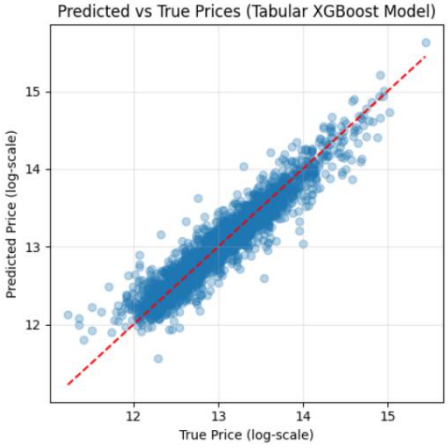


To further investigate whether visual information extracted by CNNs contributes measurable predictive value, high-dimensional image embeddings obtained from a pretrained ResNet-18model were standardized and reduced using Principal Component Analysis (PCA) and were then concatenated with scaled tabular features and used as input to multiple classical regression models.Ensemble-based models trained on the fused feature representation, particularly XGBoost and Gradient Boosting, achieve strong predictive performance, with XGBoost attaining an R^2 score of approximately 0.90 in log-price space

Overall, the limited improvement observed after embedding fusion reinforces the conclusion that satellite imagery contributes weak but complementary signals, insufficient to substantially enhance valuation accuracy beyond what is already captured by structured attributes.

XGBoost on tabular data





To establish a strong baseline and assess the predictive power of structured data, multiple tabular regression models were evaluated using engineered property-level features. These features capture key structural, locational, and temporal characteristics known to influence real estate prices, including living area, quality grade, geographic coordinates, renovation status, and seasonal effects.

Among all tabular approaches, XGBoost emerged as the most effective model, achieving the highest R^2 score and lowest RMSE in log-price space. XGBoost's gradient boosting framework, combined with regularization and subsampling, enables it to robustly model heterogeneous effects across features such as location, size, and condition while maintaining strong generalization.

```
XGBoost RMSE (log): 0.16324739075544945
XGBoost R² (log): 0.903426873986678
```

Final Model Selection and Conclusion

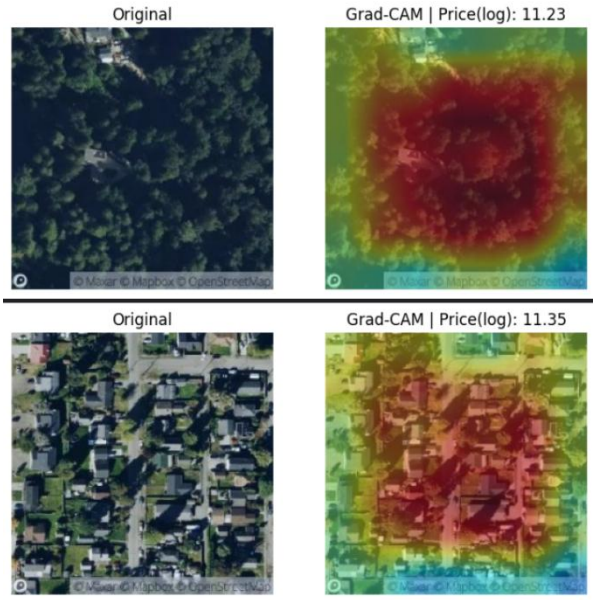
This project evaluated multiple modeling paradigms to assess the relative contributions of structured tabular data and satellite imagery for property valuation. **Image-only CNN** models showed limited predictive power, indicating that while satellite images capture neighborhood-level context, they lack direct indicators of critical value drivers such as interior quality, construction grade, and precise location effects.

To leverage visual information without degrading performance, multimodal strategies were explored. A **multimodal residual learning framework** allowed satellite imagery to provide contextual corrections to a strong tabular baseline, improving interpretability while maintaining model stability. Additionally, **PCA-compressed CNN embeddings fused with tabular features** achieved performance close to tabular-only models, reinforcing the conclusion that visual features act as auxiliary context rather than primary predictors.

Across all experiments, the **tabular XGBoost regressor** consistently delivered the highest accuracy by effectively modeling nonlinear relationships among structural, locational, and temporal attributes. Consequently, it was selected as the final model for test-set prediction. Overall, satellite imagery enhanced contextual understanding and explainability but remained complementary to structured data in driving valuation accuracy.

GradCAM

Gradient-weighted Class Activation Mapping (Grad-CAM) is used to interpret convolutional neural network (CNN) predictions by highlighting the regions of an input image that most strongly influence the model's output.



Grad-CAM visualizations reveal that the CNN primarily focuses on **broad neighborhood-level patterns** rather than precise property-specific features. The activations are diffuse and region-wide, indicating reliance on coarse visual context such as housing density, road layout, and surrounding greenery. This explains the limited standalone performance of the image-only model, as satellite images do not capture critical factors like interior quality, construction grade, or amenities. Overall, the analysis confirms that while satellite imagery provides useful contextual information, it lacks strong discriminative power on its own.

Overall, the project highlights the importance of structured data for reliable property valuation, while showing that satellite imagery, when used judiciously, can enhance interpretability and contextual understanding within multimodal frameworks.