Satellite Imagery–Enhanced Property Valuation

CDC × Yhills Open Projects (2025–26)

Project: Multimodal regression for residential property price prediction

Author : SACHIN MEENA

Enroll No : 23116085

Branch : ECE(3y)

GITHUB : [Satellite-Imagery-Based-Property-Valuation_final_12/ at main · Sachin121222/Satellite-Imagery-Based-Property-Valuation_final_12]

1. Introduction

1.1 Conventional Property Valuation Methods

Residential property valuation is a widely studied problem in real estate analytics and financial modelling. Most existing automated valuation models (AVMs) rely heavily on structured tabular data. These datasets typically include variables such as total living area, lot size, number of bedrooms and bathrooms, construction quality, year of construction, and coarse geographic indicators like latitude, longitude, or zip codes.

Such features primarily describe the internal and structural attributes of a property and have historically served as the foundation for predictive pricing systems.

1.2 Shortcomings of Structure-Only Data

Although tabular features provide valuable information, they fail to adequately represent the surrounding environment of a property. In real-world markets, two houses with similar internal specifications may differ significantly in price due to neighborhood-level characteristics.

Factors such as:

- access to water bodies,
- availability of green and open spaces,
- quality of road connectivity,
- surrounding urban density,

- proximity to infrastructure and amenities,

play a crucial role in shaping buyer perception and market value. Unfortunately, these aspects are either weakly approximated or completely missing in standard tabular datasets.

1.3 Role of Satellite Imagery in Valuation

Satellite imagery offers a practical way to capture neighborhood and environmental context at scale. High-resolution satellite images encode rich visual patterns related to land use, vegetation cover, water proximity, road networks, and overall urban planning.

These visual signals closely mirror how humans evaluate neighborhood desirability, making satellite imagery a meaningful complementary data source for residential price prediction.

1.4 Rationale for a Multimodal Learning Framework

The motivation behind this project is to investigate whether integrating satellite imagery with structured housing data can lead to more context-aware and realistic valuation models.

A multimodal framework enables:

- structured features to represent intrinsic property attributes, and

- image-based features to encode external neighborhood information.

Rather than replacing traditional valuation techniques, this approach seeks to enhance existing models by incorporating visual cues that are difficult to quantify numerically.

2. Problem Statement and Objectives

2.1 Problem Statement

Standard housing datasets lack explicit environmental and neighbourhood context, which limits their ability to fully explain property price variation.

This project aims to address this gap by developing a multimodal machine learning pipeline that fuses structured tabular data with satellite imagery. The central challenge lies in extracting informative visual representations and effectively combining them with tabular features in a unified predictive model.

2.2 Project Objectives

The key objectives of this study are:

- To establish a strong baseline property valuation model using only tabular housing attributes.
- To programmatically collect satellite images based on geographic coordinates associated with each property.

- To extract high-level visual features from satellite imagery using a pretrained Convolutional Neural Network (CNN).
- To train and evaluate image-only and multimodal regression models.
- To compare tabular-only, image-only, and multimodal approaches using standard regression metrics.
- To apply model explainability techniques (Grad-CAM) to interpret the influence of satellite imagery on predictions.

2.3 Scope and Expected Contributions

The scope of this project is restricted to residential property valuation using historical transaction data and publicly accessible satellite imagery.

The expected contributions include:

- insights into the effectiveness of multimodal learning for real estate applications,
- empirical evidence on how environmental context influences housing prices,
- and a reproducible end-to-end pipeline demonstrating the challenges of integrating visual data into traditional machine learning workflows.

3. Dataset Description

3.1 Data Source

The primary dataset consists of publicly available residential housing transaction records. Each entry contains structured property attributes along with latitude and longitude coordinates, enabling the retrieval of satellite imagery corresponding to property locations.

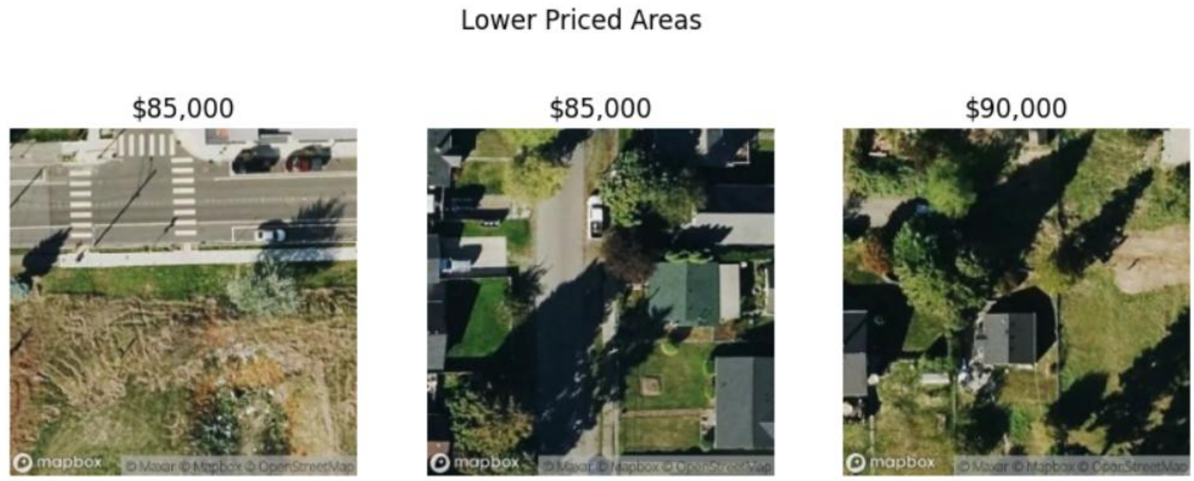
3.2 Structured Tabular Features

Feature Category	**Representative Attributes**
Property Size	sqft_living, sqft_lot
Quality Indicators	condition, grade
Location	lat, long
Neighborhood Proxies	sqft_living15, sqft_lot15

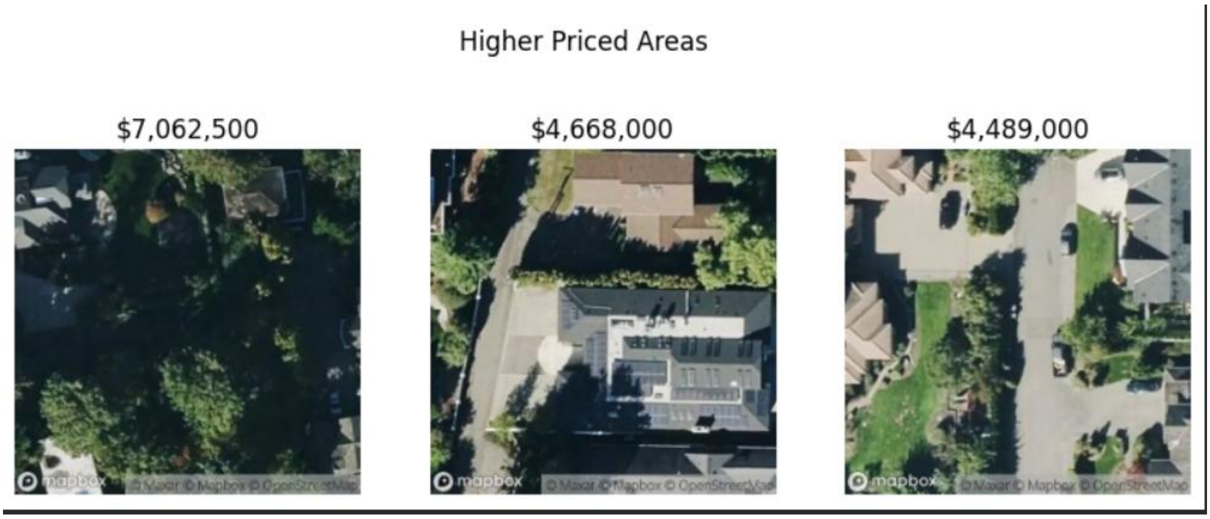
3.3 Satellite Image Collection

Satellite images were collected programmatically using property coordinates to capture neighborhood-level visual context. Due to API usage limits and computational constraints, imagery was obtained for a stratified subset of properties spanning the entire price range.

Lower priced area images



Higher priced area images



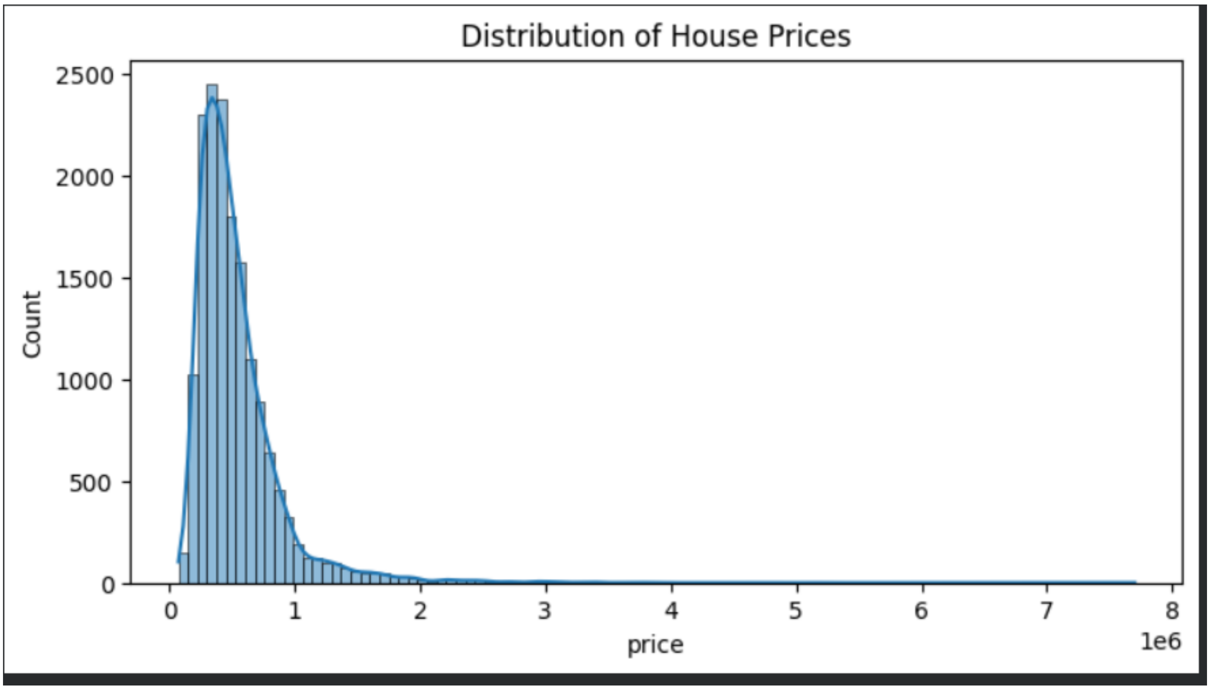
3.4 Data Preprocessing

- Tabular features were cleaned, standardized, and the target variable (price) was log-transformed to mitigate skewness.
- Satellite images were resized and normalized before being passed through a pretrained CNN.
- The CNN produced fixed-length image embeddings, which enabled efficient integration of visual information into regression models.

4. Exploratory Data Analysis

4.1 Distribution of Property Prices

Property prices exhibit a strong right-skew, with a small fraction of extremely expensive properties. To stabilize variance and improve model learning, the target variable was log-transformed prior to training.

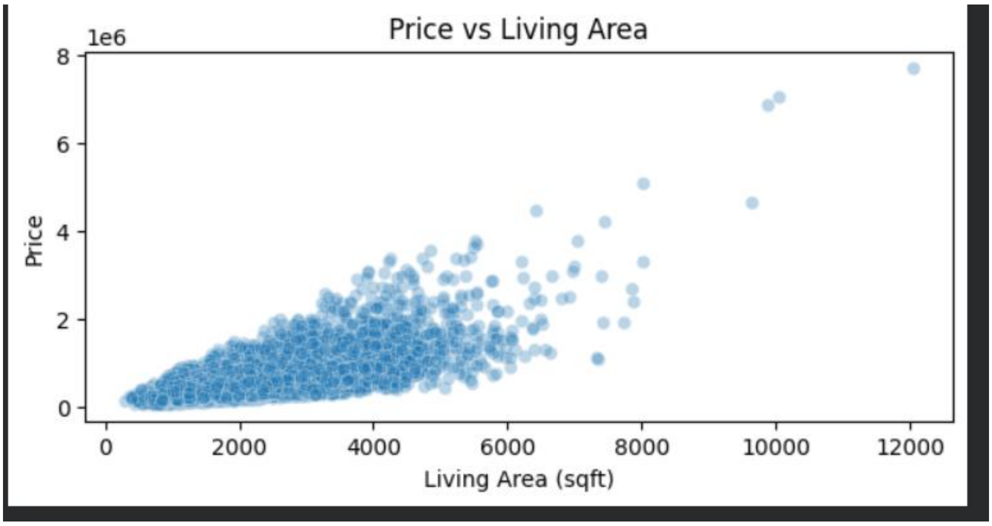


4.2 Key Feature Relationships

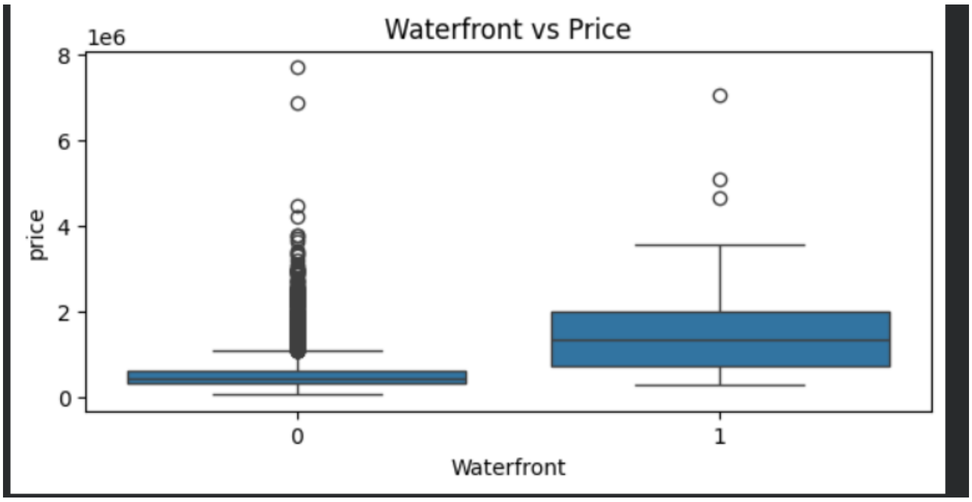
Several structured variables demonstrate strong associations with property price:

- **Living area (sqft_living)** shows a clear positive relationship with price.
- **Waterfront properties** command a significant premium.
- **View quality** correlates positively with higher valuations.

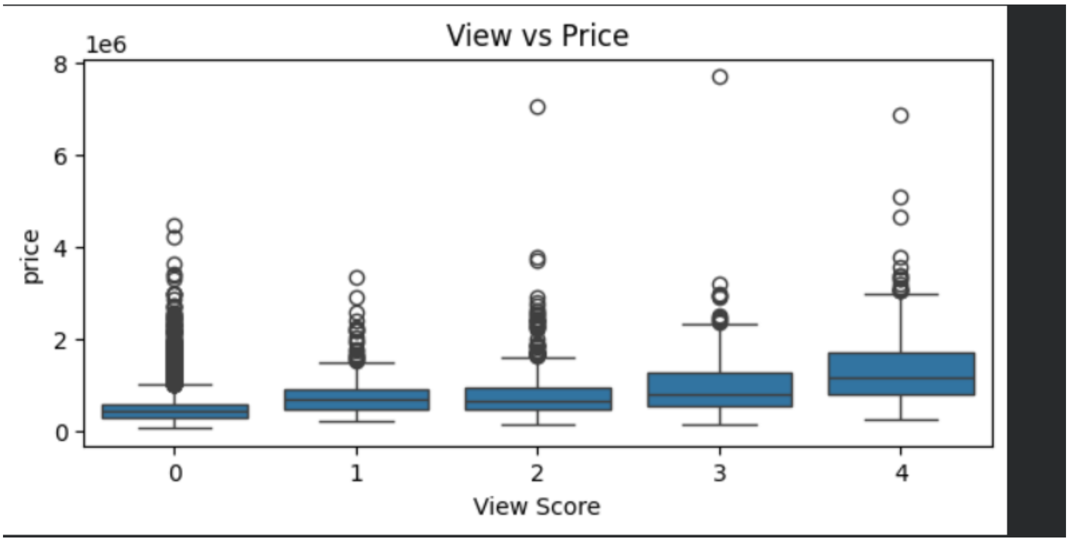
Relationship Between Living Area and Property Price



Impact of Waterfront Presence on Property Prices

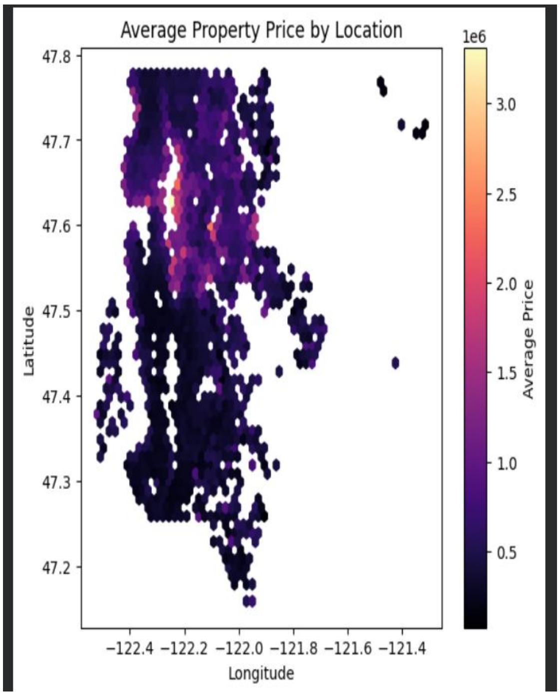


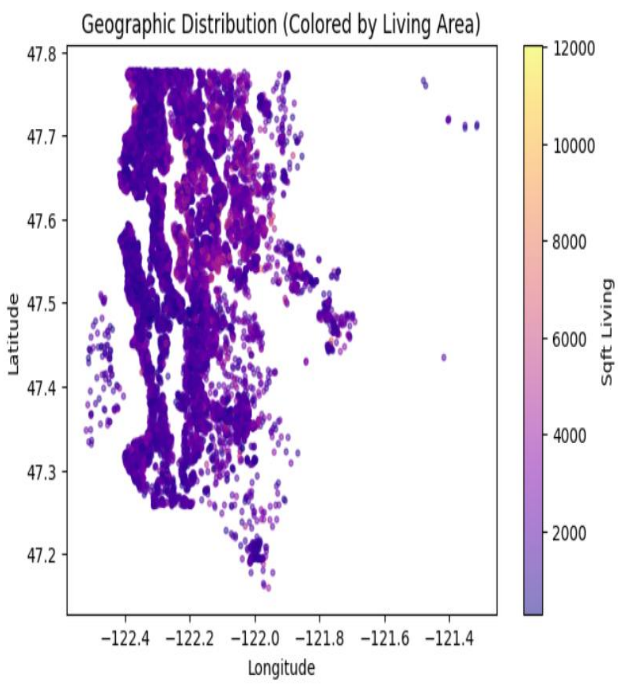
Effect of View Quality on Property Prices



4.3 Spatial Patterns

Geospatial analysis reveals clear clustering of high-value properties near water bodies and premium residential areas, whereas lower-priced properties are more evenly distributed inland.





5. Modeling Methodology

5.1 Tabular Baseline Model

A baseline regression model was trained using structured features only. Prices were modeled in log-space, and all numerical features were standardized.

A gradient boosting regressor was selected due to its strong performance on tabular data and ability to capture non-linear interactions.

5.2 Visual Feature Extraction

Satellite images were processed using a pretrained **ResNet-18** architecture. The final classification layer was removed, yielding 512-dimensional feature embeddings.

To reduce overfitting and computational cost, CNN weights were frozen during feature extraction.

5.3 Image-Only Regression Model

An image-only model was trained using the extracted embeddings to assess whether satellite imagery alone carries predictive information. A shallow multi-layer perceptron (MLP) served as the regression model.

5.4 Multimodal Fusion Strategy

An early-fusion approach was implemented, where standardized tabular features were concatenated with image embeddings and passed through a fully connected neural network.

```
Tabular Features —▶ Encoder ─┐
                             ├─▶ Fusion Network –▶ Price
Satellite Images –▶ CNN –▶ Embeddings ─┘
```

This strategy was chosen for its simplicity and interpretability.

5.5 Training Configuration

Neural network models were trained using mean squared error loss on log-transformed prices and optimized using the Adam optimizer. Validation-based early stopping was applied to prevent overfitting.

6. Model Explainability Using Grad-CAM

6.1 Motivation for Explainability

Deep learning models incorporating image data are often perceived as black boxes. In property valuation, it is critical to ensure that predictions rely on meaningful neighborhood features rather than spurious visual artifacts.

6.2 Grad-CAM Methodology

Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to the convolutional layers of the pretrained ResNet-18 model. Grad-CAM generates heatmaps highlighting image regions that most influence model predictions.

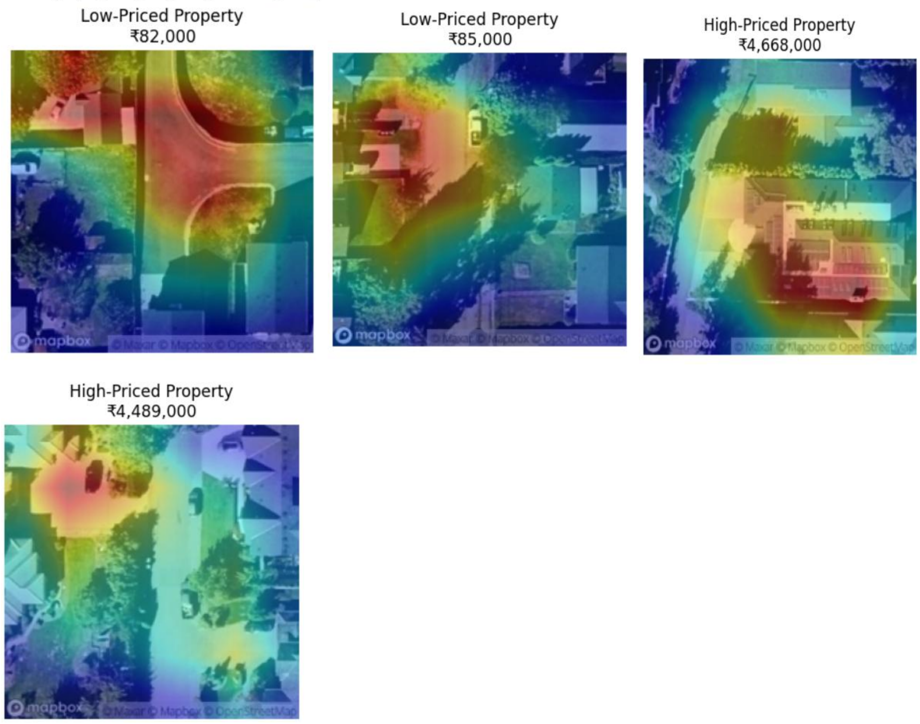
6.3 Visual Interpretation

Grad-CAM visualizations reveal consistent and interpretable patterns:

- High-value properties emphasize water bodies, green spaces, and organized residential layouts.

- Lower-value properties highlight dense urban areas with limited greenery.

These patterns align well with established real estate valuation principles.



7. Experimental Results

7.1 Evaluation Metrics

Models were evaluated using:

- Root Mean Squared Error (RMSE) on the original price scale
- R^2 score to measure explained variance

Predictions from neural networks were inverse-transformed before evaluation.

7.2 Model Performance Comparison

Model	RMSE	R^2
Tabular-Only	88521.20	0.9376
Image-Only	539792	-1.2756
Multimodal Fusion	489662.99	-0.6775

7.3 Discussion

The tabular-only model achieved the best performance, indicating that structured property attributes remain the dominant predictors of price. The image-only model performed worse than the baseline but better than random, confirming that satellite imagery contains meaningful information.

The multimodal early-fusion model failed to outperform the tabular baseline, highlighting the limitations of simple fusion strategies.

7.4 Key Insight

While satellite imagery captures valuable neighborhood context, naïve multimodal fusion does not automatically improve predictive performance.

8. Limitations and Future Directions

8.1 Limitations

- Satellite imagery was available only for a subset of properties.
- Early-fusion architecture may not capture complex cross-modal relationships.
- Satellite images provide a static snapshot and ignore temporal changes.

8.2 Future Work

- Explore attention-based and late-fusion architectures.
- Fine-tune CNN models on satellite imagery.
- Integrate temporal and higher-resolution spatial data.

9. Conclusion

This project examined residential property valuation through a multimodal learning lens by combining structured housing data with satellite imagery. While tabular features remain the strongest predictors, satellite imagery captures meaningful neighborhood-level signals, as evidenced by image-only modeling and Grad-CAM analysis.

The findings underscore both the promise and the challenges of multimodal learning in real estate analytics and emphasize the need for more sophisticated fusion strategies.

10. Reproducibility and Code Availability

All code is publicly available and structured to ensure reproducibility. Satellite imagery is fetched programmatically and excluded from version control due to API constraints. Detailed setup instructions and execution steps are provided in the project repository.