Satellite Imagery–Based Property Valuation

CDC × Yhills Open Projects (2025–26)
Data Science Problem Statement

"A multimodal machine learning approach to predict residential property prices by combining structured housing data with satellite imagery."

Author: Rahul Chauhan

Domain: Data Science

Submission: CDC × Yhills Open Projects

Year: 2025–26

GitHub Repository:
https://github.com/Rahul2512Chauhan/Satellite-Imagery-Based-Property-Valuation

Introduction & Motivation

Traditional Property Valuation :

Property valuation is a core problem in real estate analytics and is traditionally addressed using structured, tabular data. Common valuation models rely on features such as property size, number of bedrooms and bathrooms, construction quality, year of build, and basic location indicators like zip codes or geographic coordinates. These attributes describe the internal characteristics of a house and form the foundation of most automated valuation systems.

Limitations of Tabular-Only Models :

While structured features capture important property-level details, they often fail to represent the surrounding environment. Two properties with similar internal characteristics can differ significantly in market value due to neighborhood-level factors. Elements such as proximity to water bodies, availability of green spaces, road connectivity, urban density, and nearby infrastructure strongly influence buyer perception and pricing, yet these factors are either weakly represented or entirely absent in tabular datasets.

Why Satellite Imagery Matters :

Satellite imagery provides a direct way to capture environmental and neighborhood context. High-resolution satellite images encode visual patterns related to land use, vegetation, water proximity, road networks, and overall urban layout. These visual cues align closely with human intuition about desirability and "curb appeal," making satellite imagery a valuable complementary data source for property valuation tasks.

Motivation for a Multimodal Approach :

The central motivation of this project is to explore whether combining structured housing data with satellite imagery can lead to more context-aware valuation models. A multimodal learning approach allows tabular features to represent intrinsic property attributes, while satellite images contribute external neighborhood information. Rather than replacing traditional features, this approach aims to enhance existing valuation frameworks by integrating visual context that is difficult to quantify through structured data alone.

Problem Statement & Objectives

Problem Statement :

Traditional housing data lacks the **environmental context** (neighborhood quality, green space, etc.) critical for accurate valuation.

This project aims to improve price predictions by developing a **multimodal pipeline** that fuses **tabular property data** with **satellite imagery**. The core challenge is effectively extracting visual features and integrating these diverse data types into a single, high-performance model.

Project Objectives :

The objectives of this project are as follows:

- To build a robust baseline property valuation model using only structured tabular housing data.
- To programmatically acquire satellite images using geographic coordinates (latitude and longitude) associated with each property.
- To extract high-level visual features from satellite images using a pretrained Convolutional Neural Network (CNN).
- To develop and evaluate image-only and multimodal regression models for property price prediction.
- To compare the performance of tabular-only, image-only, and multimodal models using standard regression metrics.
- To apply explainability techniques (Grad-CAM) to understand which visual regions in satellite images influence model predictions.

Scope and Expected Outcomes :

The scope of this project is limited to residential property valuation using historical housing transaction data and publicly available satellite imagery. The expected outcomes include a deeper understanding of the strengths and limitations of multimodal learning in real estate analytics, insights into how environmental context affects property prices, and a reproducible end-to-end pipeline that demonstrates the practical challenges of integrating visual data into traditional machine learning workflows.

Dataset Description

Data Source :

The primary dataset used in this project is a publicly available residential housing dataset containing historical property sale records. The dataset includes structured information describing property characteristics along with geographic coordinates (latitude and longitude) for each property. These coordinates enable the programmatic retrieval of satellite imagery corresponding to each location.

Tabular Data :

Category	Example Features
Size	sqft_living, sqft_lot
Quality	condition, grade
Location	lat, long
Neighborhood	sqft_living15, sqft_lot15

Satellite Imagery Data :

Satellite images were programmatically fetched using geographic coordinates to capture neighborhood-level context. Due to API and computational constraints, images were collected for a stratified subset of properties covering the full price distribution.

Data Preparation :

Tabular features were cleaned, log-transformed (price), and standardized prior to modeling. Satellite images were resized and normalized, then converted into fixed-length feature embeddings using a pretrained convolutional neural network. These embeddings enable efficient integration of visual information into downstream regression models.

Exploratory Data Analysis (EDA)

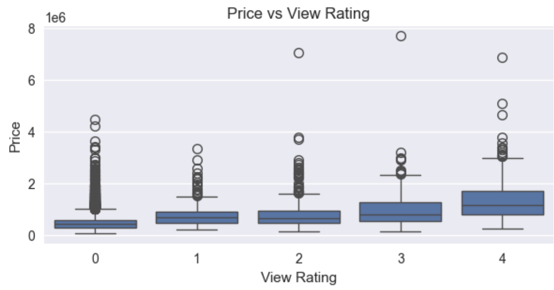
Price Distribution :

The distribution of property prices is highly right-skewed, with a small number of very expensive properties. To stabilize variance and improve regression performance, the target variable (price) was log-transformed for all modeling experiments.

Relationship Between Key Features and Price :

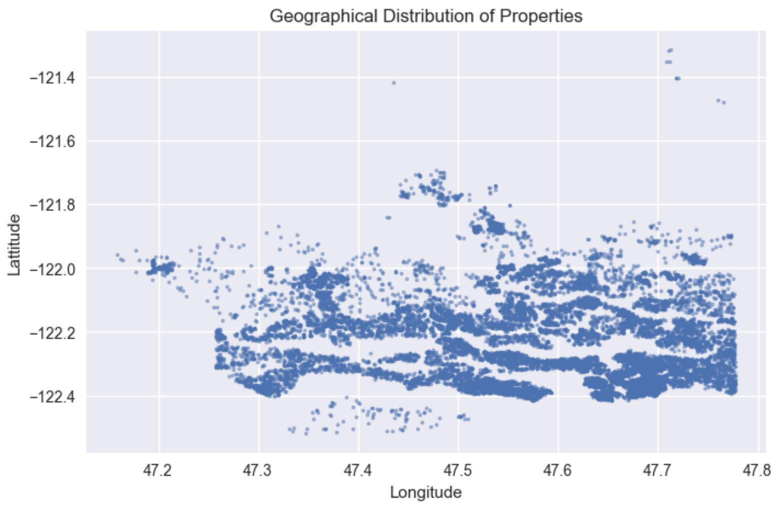
Several structured features show strong relationships with property price:

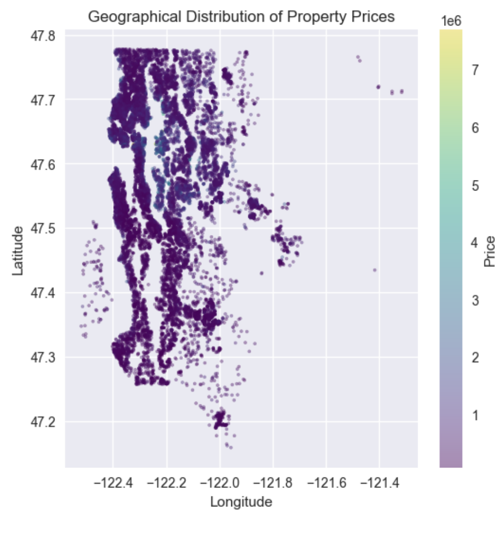
- **sqft_living:** Larger living areas are strongly associated with higher prices.
- **waterfront:** Waterfront properties command a significant price premium.
- **view:** Higher view ratings correspond to higher property values.



Geographic Distribution of Properties :

Mapping property locations reveals clear geographic clustering. High-priced properties tend to be concentrated near water bodies and premium residential zones, while lower-priced properties are more uniformly distributed inland.





Modelling Approach

Tabular Baseline :

A baseline model was first built using structured tabular features to establish reference performance. The target variable (price) was log-transformed to reduce skewness, and numerical features were standardized. A gradient boosting regressor was used due to its strong performance on structured data and ability to model non-linear relationships.

Image Feature Extraction :

Satellite images were converted into numerical representations using a pretrained ResNet-18 model. The final classification layer was removed to extract fixed-length (512-dimensional) image embeddings. All CNN weights were frozen during feature extraction to reduce overfitting and computational cost.

Image-Only Model :

An image-only regression model was trained using the extracted embeddings to evaluate whether satellite imagery contains standalone predictive signal. A shallow multi-layer perceptron (MLP) was used for this purpose.

Multimodal Fusion :

A simple early-fusion strategy was adopted, where standardized tabular features and image embeddings were concatenated and passed through a fully connected neural network for price prediction. This approach was chosen for its simplicity and interpretability.

```
Tabular Features ──▶ Tabular Encoder  ─┐
                                       ├─▶ Fusion MLP ─▶ Price
Satellite Image ─▶ CNN ─▶ Embedding  ──┘
```

Training Setup :

Neural network models were trained using mean squared error loss on log-transformed prices and optimized with Adam. A validation split and early stopping were used to control overfitting.

Model Explainability (Grad-CAM)

Motivation :

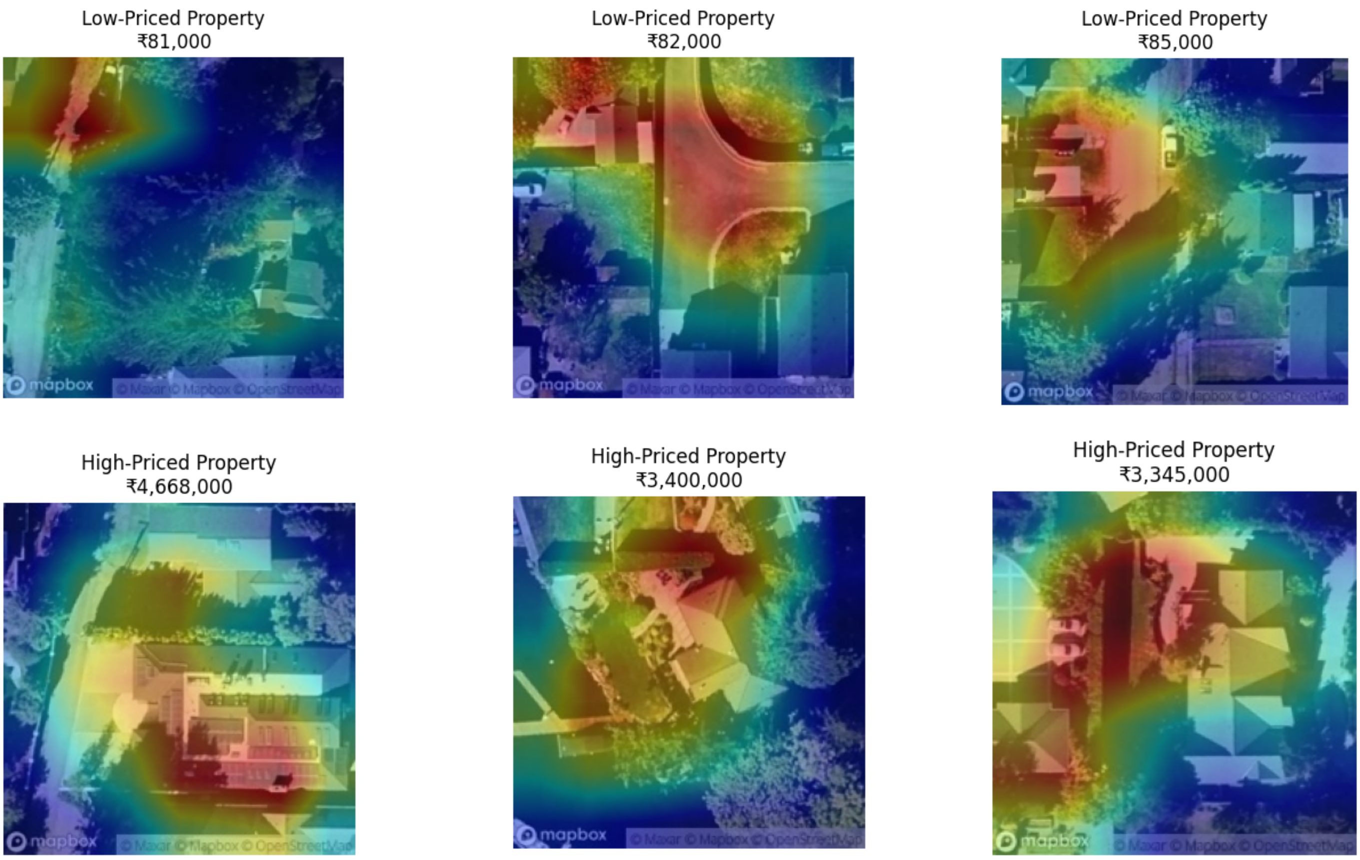
Deep learning models that use image data are often considered black boxes. In the context of property valuation, it is important to verify that the model focuses on meaningful neighborhood features rather than spurious patterns. Explainability helps validate whether satellite imagery contributes interpretable and realistic signals to the prediction process.

Grad-CAM Approach :

Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to the convolutional layers of the pretrained ResNet-18 model used for image feature extraction. Grad-CAM produces heatmaps that highlight image regions with the greatest influence on the model's output, which can be overlaid on the original satellite images for interpretation.

Observations :

Grad-CAM visualizations show consistent and interpretable patterns. High-value properties tend to activate regions corresponding to water bodies, green spaces, and well-organized residential layouts. In contrast, low-value properties often highlight dense urban structures or areas with limited greenery. These observations align with known real estate valuation principles and confirm that satellite imagery captures meaningful neighborhood context.



Results & Model Comparison

Evaluation Metrics :

Model performance was evaluated using **Root Mean Squared Error (RMSE)** and **R^2 score** on a held-out validation set. RMSE measures prediction error magnitude in the original price scale, while R^2 indicates the proportion of variance explained by the model. For neural network models, predictions were made on log-transformed prices and converted back to the original scale for evaluation.

Model Performance Comparison :

Three modeling approaches were evaluated:

- **Tabular-only model** using structured housing features
- **Image-only model** using satellite image embeddings
- **Multimodal fusion model** combining tabular and image features

Table : Model Performance Comparison :

Model	RMSE	R^2
Tabular Only	117699.03	0.8896
Image Only	517667.21	0.7985
Multimodal Fusion	426696.07	0.2938

Discussion of Results :

The tabular-only model achieved the strongest performance, showing that structured property attributes capture most of the predictive signal. The image-only model performed worse than the tabular baseline but better than random, indicating that satellite imagery contains useful information. The multimodal fusion model did not outperform the tabular model, suggesting limitations of simple early-fusion strategies.

Key Takeaway :

Tabular features remain the dominant driver of property price prediction. Satellite imagery captures meaningful neighborhood context, but naïve multimodal fusion does not guarantee performance improvement.

Limitations & Future Work

Limitations :

This project has several limitations. Satellite imagery was collected for a stratified subset of properties due to API and computational constraints, rather than the full dataset. The multimodal model used a simple early-fusion strategy, which may not fully capture complex interactions between tabular and visual features. Additionally, satellite images provide a static view of the environment and do not account for temporal changes such as recent development or seasonal effects.

Future Work :

Future improvements could include experimenting with more advanced fusion techniques such as attention-based or late-fusion architectures. Fine-tuning the CNN on satellite imagery rather than using frozen embeddings may also improve performance. Incorporating additional spatial features, temporal data, or higher-resolution imagery could further enhance model accuracy and robustness.

Conclusion

This project explored residential property valuation as a multimodal regression problem by combining structured housing data with satellite imagery. A strong tabular baseline model was first established, followed by image-based and multimodal models to assess the contribution of visual neighborhood context.

The results show that structured tabular features remain the primary drivers of property price prediction. Satellite imagery was found to contain meaningful neighborhood-level information, as demonstrated by image-only modeling and Grad-CAM explainability. However, simple early-fusion multimodal approaches did not outperform the tabular baseline, highlighting the challenges of effectively integrating heterogeneous data modalities.

Overall, this project demonstrates both the potential and the limitations of using satellite imagery for real estate valuation. While visual context provides valuable insights, careful fusion design and more advanced modeling strategies are required to fully realize its benefits in predictive performance.

Reproducibility & Code Availability

All code used in this project is available in a public GitHub repository and is organized to ensure reproducibility. The repository includes data preprocessing scripts, image ingestion utilities, model training notebooks, and evaluation workflows.

Satellite images are fetched programmatically using geographic coordinates and are intentionally excluded from version control due to API usage constraints. Detailed setup instructions, dependencies, and execution steps are provided in the repository README.

GitHub Repository:
https://github.com/Rahul2512Chauhan/Satellite-Imagery-Based-Property-Valuation