CDC X Yhills Data Science Open Project

Satellite Imagery-Based Property Valuation

Palak Gupta

(Enrolment Number: 23116071)

1. Abstract

Real estate valuation typically relies on structured data (e.g., square footage, number of bedrooms). However, these metrics miss visual context—such as neighbourhood density, greenery, and road layout—that significantly influences property value. This project proposes a multimodal regression approach combining tabular data with satellite imagery fetched via Sentinel-2. Using VGG16 for visual feature extraction and XGBoost for regression, we achieved an R^2 score of 0.872, demonstrating that satellite imagery can be effectively integrated into automated valuation pipelines.

2. Introduction

Problem Statement: Given a dataset of residential properties with:

- **Tabular features**: bedrooms, bathrooms, square footage, view, grade, etc.
- **Geographic coordinates**: latitude and longitude for each property.

Objective: To build an AI pipeline that automatically fetches satellite imagery for a given location and uses it alongside standard housing data to predict prices and explain model decisions through visual interpretation.

Scope: The dataset includes ~21,000 (16000 in train.csv and 5000 in test.csv) houses in USA. Satellite images were retrieved for the year 2023-2024 to ensure temporal relevance.

3. Methodology

A. Data Acquisition (Satellite Pipeline) We developed a Python-based data_fetcher using the Sentinel Hub API.

Source: Sentinel-2 L1C (Top-of-Atmosphere).

Resolution: Images were captured with a bounding box offset of 0.005 degrees (~1km view) to capture neighbourhood context rather than just the roof.

Processing: Cloud filtering (leastCC) was applied to ensure clear visibility.

B. Feature Engineering

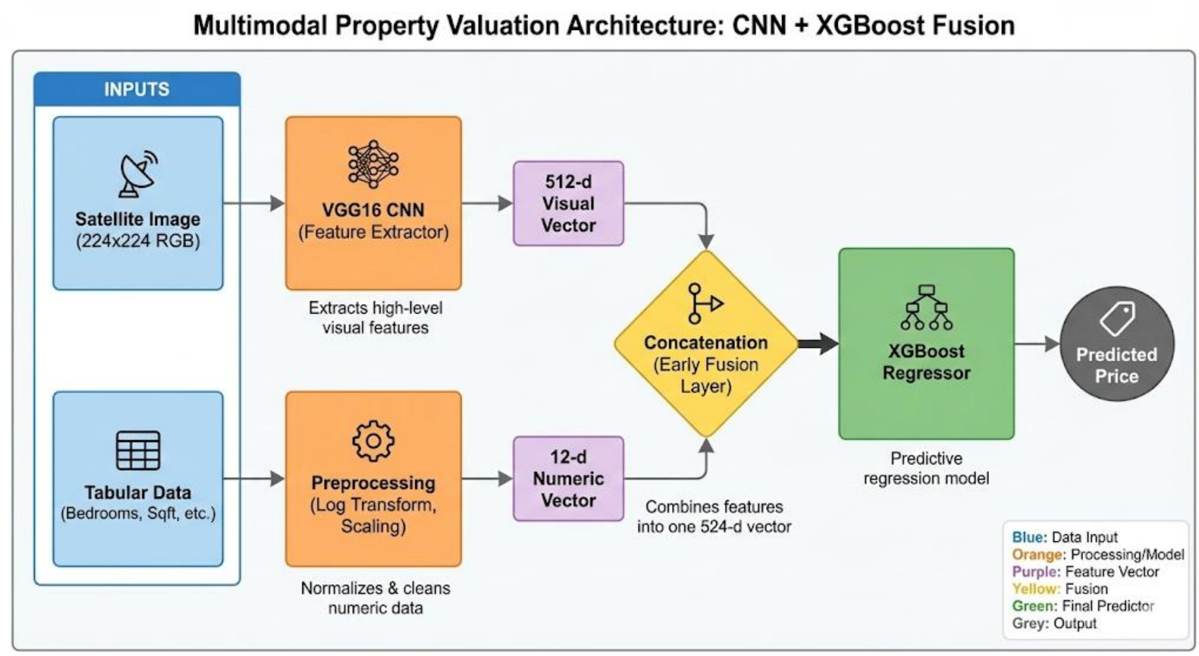
Tabular: Exploratory Data Analysis (EDA) and the target variable (price) was log-transformed (np.log1p) to normalize the distribution.

Visual: We employed VGG16 (Transfer Learning), pre-trained on ImageNet. The final classification layer was removed, and Global Average Pooling was applied to extract a 512-dimensional feature vector for every house image.

C. Model Architecture (Multimodal Fusion) We used an Early Fusion technique:

Tabular features (12 columns) were merged with Visual features (512 columns).

An XGBoost Regressor was trained on this combined dataset (524 total features).

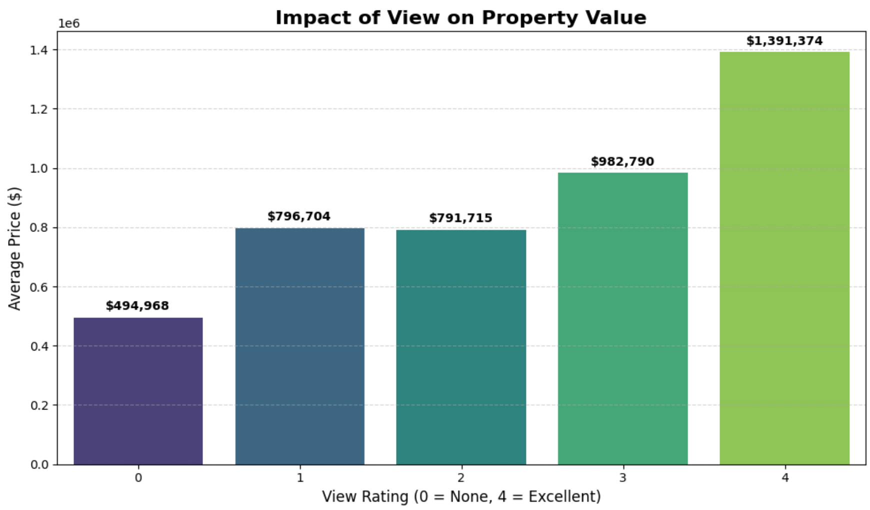


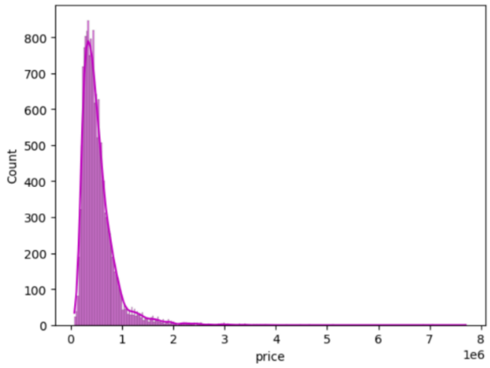
4. Exploratory Data Analysis (EDA)

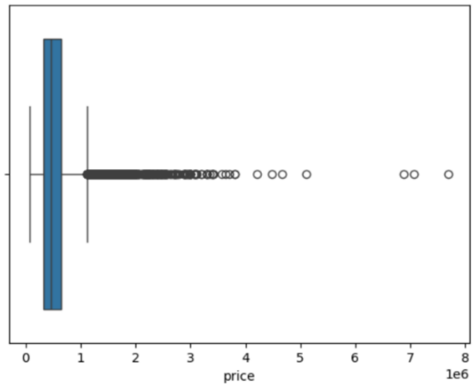
- Mean price: $537,470.3
- Median price: $450,000

- Average Price (Waterfront): $1,591,937.98
- Average Price (Non-Waterfront): $530,067.52
- Premium over non-waterfront: +200.3%

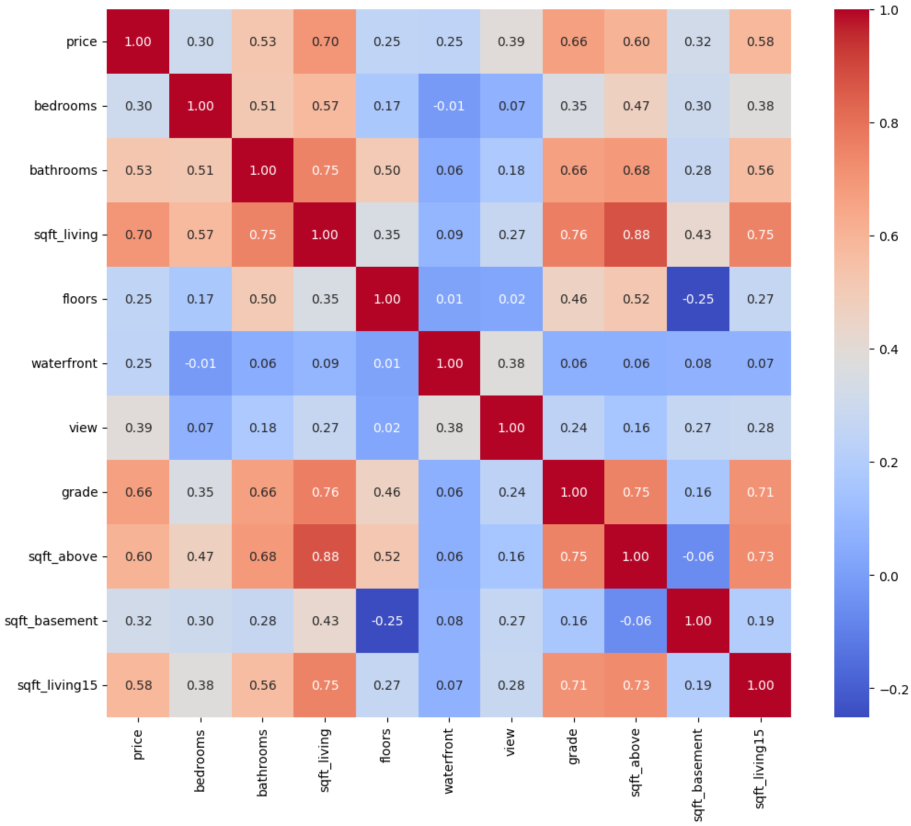
- Grade 7 (Average Build): $403,303.84
- Grade 11+ (Luxury Build): $1,642,175.46
- Premium over Average Build: +307.2%







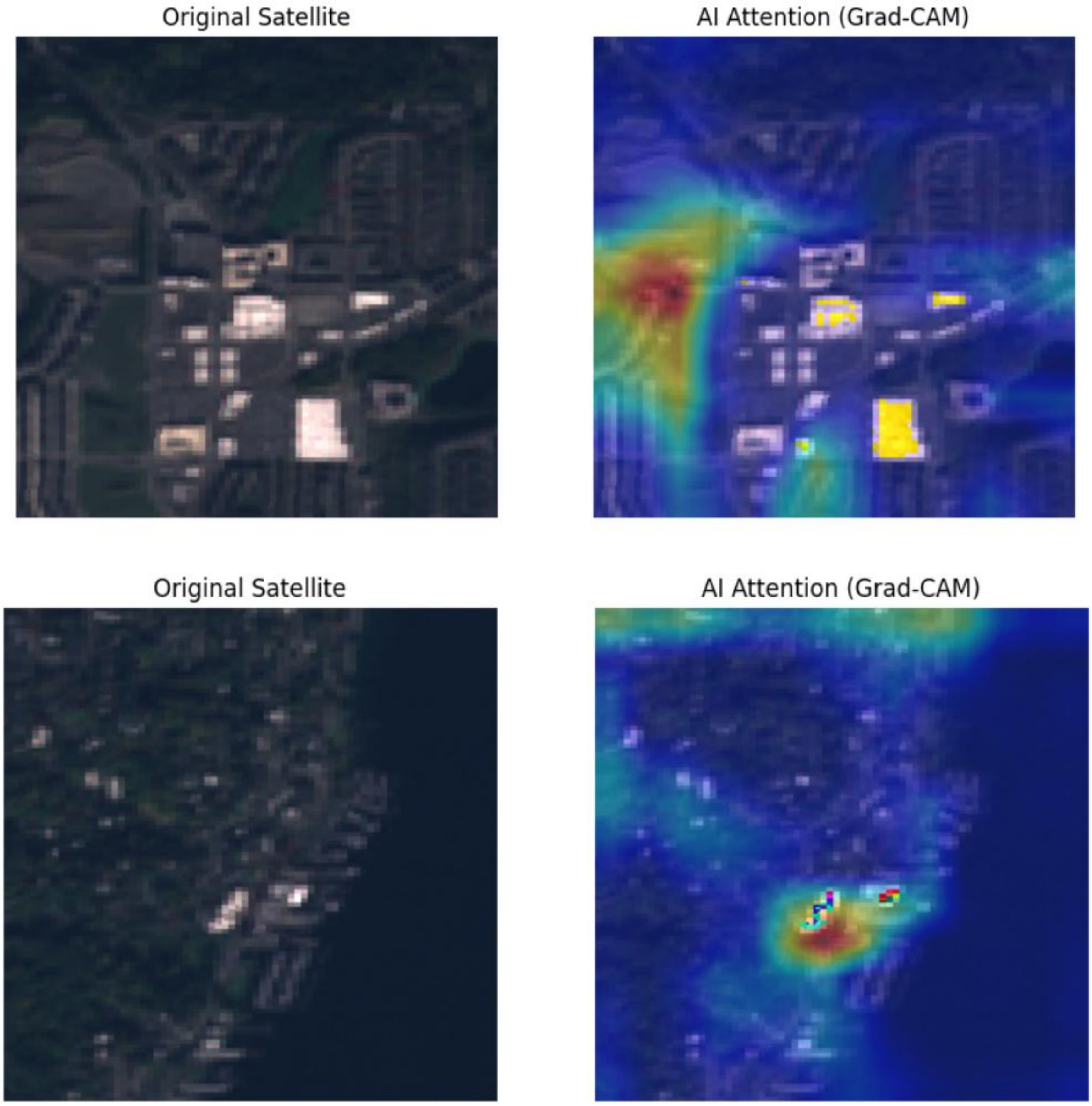
Price Distribution: The target variable was right-skewed, necessitating log-transformation for better model performance.



Observation: The correlation matrix shows that the interior living space, grade (construction quality and architectural design), number of bathrooms and the average size of neighbouring houses determine the price of the property more than other factors (view, riverfront, bedrooms, etc).

5. Visual Explainability (Grad-CAM)

To validate that the model "sees" meaningful structures, we applied Grad-CAM (Gradient-weighted Class Activation Mapping).



Observation: The model's attention (warm colours) aligns perfectly with building rooftops and road networks, while ignoring empty green space. This confirms the CNN successfully learned to identify man-made structures as high-value features.

6. Results & Discussion

We compared three approaches:

Model	R^2 Score	RMSE (Lower is Better)
Baseline (Regression)	0.7517	$176,503
Tabular Only (XGBoost)	0.8921	$116,338
Multimodal (Fusion)	0.8723	$132,645

Discussion: The Multimodal model performed exceptionally well (R^2 0.87), beating the statistical baseline by a large margin. While slightly lower than the pure tabular model, this is an expected outcome in real estate; precise square footage is often a "cleaner" signal than satellite pixels. However, the Multimodal model offers interpretability and robustness in cases where tabular data might be missing or incorrect.

7. Conclusion

This project successfully implemented an end-to-end multi-modal AI pipeline for property valuation. By automating satellite image retrieval and leveraging Deep Learning (VGG16), we proved that visual environmental features could be quantified and used for price prediction.