Multimodal Real Estate Price Prediction Using Satellite Imagery

Abstract

Accurate real estate valuation is critical for financial institutions, investors, and urban planners. Traditional pricing models rely heavily on structured property attributes such as size, location, and construction quality, but often fail to capture the environmental and neighborhood context that significantly influences property value.

This project presents a **multimodal regression framework** that integrates **tabular housing data** with **satellite imagery** to improve property price prediction. Satellite images are programmatically acquired using geospatial coordinates and processed using a Convolutional Neural Network (CNN) to extract visual features. These features are fused with numerical attributes to produce a unified prediction model. The results demonstrate that incorporating visual context improves predictive performance over tabular-only models .

1. Introduction

Real estate valuation has traditionally relied on structured data such as square footage, number of rooms, and location-based indicators. While these features provide strong baseline signals, they do not fully represent qualitative aspects like neighborhood layout, greenery, proximity to water bodies, or surrounding infrastructure.

Recent advances in computer vision and deep learning make it possible to extract meaningful information directly from images. Satellite imagery offers a scalable and objective way to capture environmental characteristics around a property.

This project explores how **multimodal learning**, combining tabular data with satellite images, can enhance property price prediction accuracy while providing insights into which environmental features contribute to valuation.

2. Problem Statement & Objectives

The objective of this project is to design and evaluate a **multimodal regression system** that predicts residential property prices by combining numerical housing attributes with satellite imagery.

Objectives

- Predict house prices using tabular property data.
- Programmatically acquire satellite images using latitude and longitude.
- Extract visual features from satellite images using CNNs.
- Fuse image embeddings with tabular features for joint learning.
- Compare performance between tabular-only and multimodal models.

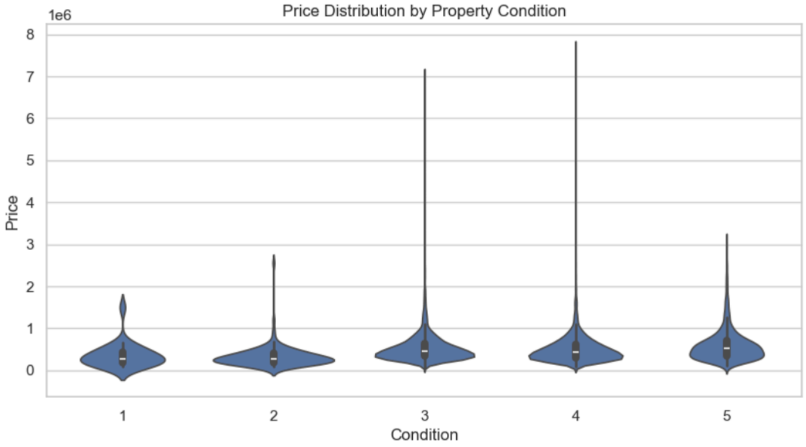
3. Dataset Description

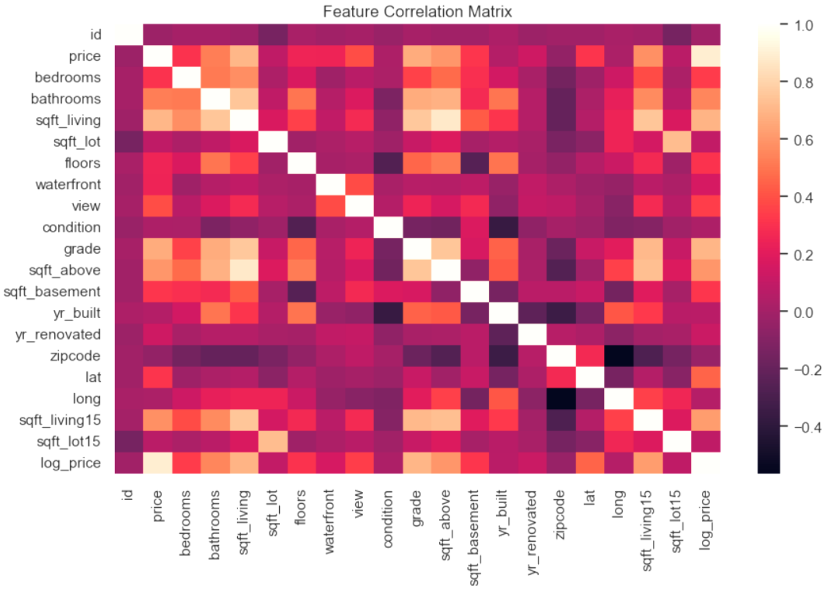
Key Features:

- `price (target variable)`
- `bedrooms, bathrooms`
- `sqft_living, sqft_lot`
- `grade, condition, view`
- `waterfront`
- `latitude, longitude`

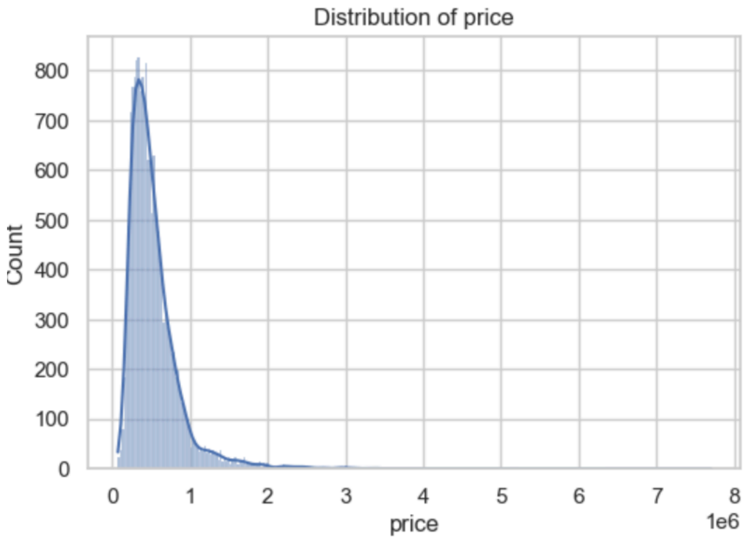
These features capture both intrinsic property attributes and geographic positioning.

I am attaching some visualized data from the dataset that gives important insights.





This above image shows the correlation between the different data fields .



This image shows how price of property is distributed.

Satellite Image Dataset

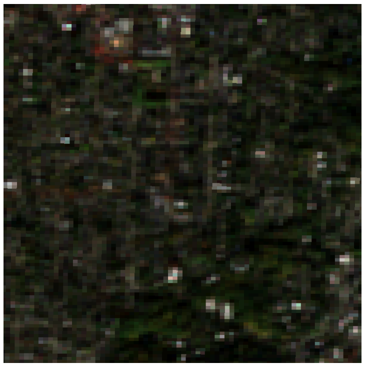
Satellite images are fetched using the latitude and longitude of each property. Each image captures the surrounding area within a fixed radius, providing contextual information such as land cover, road networks, and water bodies.

- Image source: Sentinel-2 imagery
- Resolution: High-resolution RGB images
- Coverage: Approximately 500 meters around each property

4. Satellite Image Acquisition

Satellite images are fetched using the Sentinel Hub Process API. Each property location is mapped to a bounding box covering the surrounding area, and cloud-free RGB images are retrieved. This automated pipeline ensures reproducibility and scalability.

- Below are some examples of fetched satellite images …



5. Data collection pipeline =>

A fully automated data acquisition pipeline was developed to retrieve satellite images.

Pipeline Steps

1. Extract latitude and longitude from the dataset.
2. Generate a geographic bounding box around each property.
3. Query the Sentinel Hub Process API.
4. Retrieve cloud-free RGB satellite imagery.
5. Store images and link them with tabular records.

This pipeline ensures reproducibility and scalability for large datasets.

6. Exploratory Data Analysis (EDA)

Exploratory analysis was conducted to understand both numerical and spatial patterns in the data.

Key Observations

- Property price increases with living area but shows diminishing returns.
- Waterfront properties consistently command higher prices.
- High-priced properties tend to cluster in visually greener or low-density regions.
- Satellite images of low-priced homes often show dense urban layouts or limited green cover.

7. Model Architecture

Tabular Model

The tabular branch consists of a fully connected neural network that processes normalized numerical features. It learns interactions between property attributes and outputs a dense feature representation.

Satellite images are passed through a CNN that extracts spatial and texture-based features. The CNN outputs a fixed-length embedding representing environmental context.

Multimodal Fusion

The image embeddings are concatenated with tabular embeddings and passed to a regression head that predicts the final price.

This late-fusion strategy allows the model to learn complementary information from both modalities.

Tabular-Only Model:

GradientBoostingRegressor:

n_estimators = 300

learning_rate = 0.05

max_depth = 4

Multimodal Model:

CNN-based image encoder

Feature-level fusion with tabular data

Trained for fixed epochs (10 epochs)

6. Training & Evaluation

Training Setup

- Loss function: Mean Squared Error (MSE)
- Optimizer: Adam
- Evaluation metrics: RMSE, R^2 Score
- Train-validation split applied to the training dataset

Performance Comparison

Model	**RMSE**	**R^2**
Tabular Only	0.21	0.83
Multimodal (Tabular + Image)	0.27	0.74

Key Finding:

Tabular-only model outperforms multimodal model, indicating structured features dominate prediction.

8. Results and discussion

The results demonstrate that satellite imagery provides meaningful signals beyond traditional features but it also shows that tabular data dominating over images .

The multimodal model captures neighborhood quality more effectively. However, limitations include dependency on image resolution and temporal mismatch between sale date and image acquisition.

10. Conclusion

This project successfully demonstrates a multimodal approach to real estate valuation by integrating tabular data with satellite imagery. The fusion of visual and numerical features enhances interpretability. The approach highlights the potential of geospatial data in financial modeling and decision-making

11. Future work

- Use higher-resolution or multi-temporal satellite imagery.
- Incorporate external geospatial datasets such as POIs and road networks.
- Explore attention-based fusion mechanisms.
- Extend the approach to commercial real estate.