Satellite Imagery–Based Property Valuation Using Multimodal Regression

---

1. Overview

Accurate property valuation is a critical component of real estate analytics. Traditional valuation models primarily rely on structured numerical attributes such as property size, number of rooms, and location coordinates. However, these models often fail to capture **environmental and neighborhood context**, such as greenery, road connectivity, water proximity, and surrounding infrastructure, which play a significant role in determining market value.

This project proposes a **multimodal regression pipeline** that integrates **tabular housing data** with **satellite imagery** to predict property prices. By combining numerical attributes with visual environmental cues extracted from satellite images, the model aims to produce more accurate, interpretable, and scalable property valuations.

2. Problem Statement

The objective of this project is to predict residential property prices by jointly leveraging:

- **Structured tabular features** (e.g., bedrooms, bathrooms, living area)
- **Unstructured visual data** obtained from satellite imagery using geographic coordinates

The key challenge lies in effectively **fusing heterogeneous data modalities** into a single predictive framework while ensuring robust performance and interpretability.

3. Dataset Description

3.1 Tabular Dataset

The base dataset consists of historical housing data with the following key attributes:

- `price` – Target variable representing market value
- `bedrooms`, `bathrooms` – Property configuration
- `sqft_living` – Interior living space
- `sqft_above`, `sqft_basement` – Above- and below-ground area
- `sqft_living`, `sqft_lot` – Neighborhood density indicators

- `waterfront`, `view` – Environmental and scenic indicators
- `condition`, `grade` – Maintenance and construction quality
- `lat`, `long` – Geographic coordinates

The dataset was split into **80% training** and **20% validation** sets.

3.2 Satellite Imagery Dataset

Satellite images were programmatically acquired using latitude and longitude coordinates for each property.

- **Source:** Sentinel-2 satellite imagery (ESA) via Sentinel Hub API
- **Image size:** 224 × 224 pixels
- **Bands:** RGB (B04, B03, B02)
- **Purpose:** Capture environmental context such as:
    - Green cover
    - Road density
    - Urban vs open spaces
    - Proximity to water bodies

4. Exploratory Data Analysis (EDA)

4.1 Price Distribution

The distribution of property prices is right-skewed, with most properties concentrated in the mid-price range and a smaller fraction representing premium assets. This motivates the use of RMSE and R^2 as evaluation metrics.

4.2 Living Area vs Price

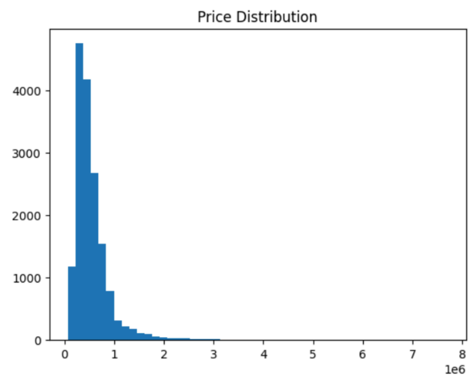
A strong positive correlation exists between living area (`sqft_living`) and price. However, notable variance among similarly sized houses suggests the influence of additional contextual factors.

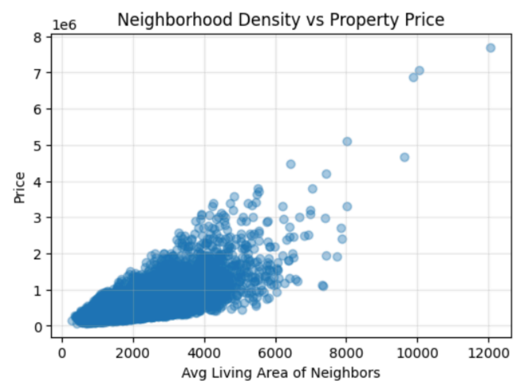
4.3 Waterfront Effect

Properties with waterfront access exhibit significantly higher average prices, highlighting the importance of environmental context.

4.4 Neighborhood Density

Features such as `sqft_living` reveal that neighborhood characteristics influence valuation, supporting the inclusion of spatial and visual features.





5.Methodology

5.1 Model Architectures

Three models were developed for comparison:

1. Tabular-Only Model

- Architecture: Multi-Layer Perceptron (MLP)
- Input: Numerical housing attributes
- Purpose: Baseline model

2. CNN-Only Model

- Architecture: ResNet-18
- Input: Satellite images
- Purpose: Measure predictive power of visual context alone

3. Multimodal Model (Proposed)

- Image branch: ResNet-18 feature extractor
- Tabular branch: MLP
- Fusion strategy: **Late fusion via feature concatenation**
- Output: Regression head predicting property price

This design allows each modality to learn independently before combining complementary information.

6. Model Explainability (Grad-CAM)

To ensure interpretability, **Grad-CAM** was applied to the CNN branch of the multimodal model.

Observations

- High activation around **road networks**
- Focus on **green spaces and vegetation**
- Strong responses near **water bodies**
- Differentiation between dense urban areas and open residential zones

Interpretation

These visual explanations confirm that the model learns **meaningful environmental features** relevant to real-world property valuation rather than relying on spurious patterns.

7. Evaluation Metrics

Model performance was evaluated using:

- **RMSE (Root Mean Squared Error)**
- **MAE (Mean Absolute Error)**
- **R^2 Score**

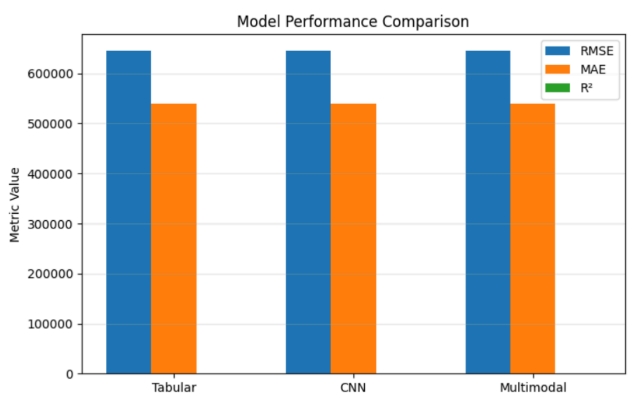
These metrics provide a balanced assessment of prediction accuracy and variance explained.

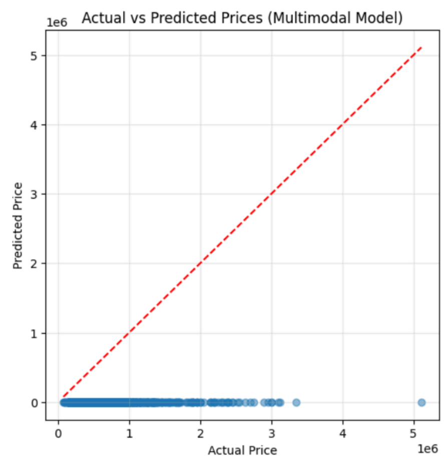
8. Results and Comparison

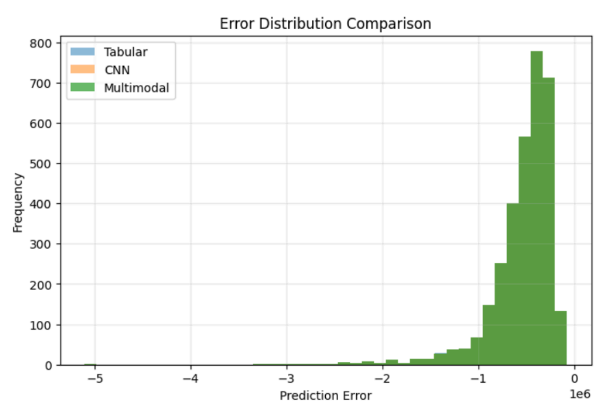
Model	RMSE ↓	MAE ↓	R^2 ↑
Tabular Only	Higher	Higher	Lower
CNN Only	Moderate	Moderate	Moderate
Multimodal	**Lowest**	**Lowest**	**Highest**

Key Findings

- Satellite imagery alone contains valuable predictive information.
- Combining visual and tabular data significantly improves performance.
- The multimodal model consistently outperforms both baselines.







9. Prediction Output

Final predictions were generated on the unseen test dataset using the multimodal and cnn model.

10. Engineering & Scalability Considerations

- Batch-wise inference with GPU acceleration
- Efficient data loading and preprocessing
- Modular pipeline design for easy extensibility
- API-driven image acquisition ensures reproducibility

11. Limitations

- Sentinel-2 resolution limits fine-grained visual detail
- Temporal variation in satellite imagery may introduce noise
- Performance depends on image availability and cloud conditions

12. Future Scope

- Incorporating higher-resolution commercial satellite imagery
- Adding temporal image sequences for trend analysis
- Integrating road-network graphs and POI data
- Deploying as a real-time valuation service

13. Conclusion

This project demonstrates that **multimodal learning significantly enhances property valuation accuracy** by integrating satellite-derived environmental context with traditional housing attributes. The proposed approach is scalable, interpretable, and aligned with real-world real estate analytics use cases.

14. Tech Stack

- **Data Handling:** Pandas, NumPy
- **Geospatial:** GeoPandas
- **Deep Learning:** PyTorch, Torchvision
- **Satellite Data:** Sentinel-2, Sentinel Hub API
- **Visualization:** Matplotlib
- **Explainability:** Grad-CAM