Project Report

1. Overview

Traditional real-estate valuation models rely heavily on tabular attributes such as house size, number of rooms, and location coordinates. While effective, these models often fail to capture environmental and neighborhood context, such as green cover, road density, or surrounding infrastructure—factors that significantly influence perceived property value.

In this project, we developed a multimodal regression pipeline that integrates:

- Tabular housing data (e.g., sqft, bedrooms, bathrooms, latitude/longitude)
- Satellite imagery–derived visual features, fetched programmatically using geographic coordinates

The goal was to enhance property price prediction by incorporating visual cues representing curb appeal and neighborhood quality, moving beyond purely numerical data.

2. Approach & Modeling Strategy

Our approach followed a structured pipeline:

1. **Tabular Data Processing**
    - Cleaned and normalized numerical housing features
    - Handled categorical variables and removed redundant features
    - Established a strong tabular-only baseline model

2. **Satellite Image Acquisition**
    - Used latitude and longitude to fetch satellite images for each property
    - Ensured consistent zoom level and image resolution
    - Stored images with property ID mapping

3. **Visual Feature Engineering**
    - Extracted interpretable image features:
        - Mean and standard deviation of RGB color channels
        - Edge density to represent structural complexity (roads/buildings)
    - These features act as proxies for:
        - Green cover
        - Water bodies
        - Urban density
4. **Multimodal Fusion**

    - Merged tabular features with extracted image features
    - Trained regression models on the combined dataset
    - Compared results with the tabular-only baseline

This design balances performance, interpretability, and reproducibility.

3. Exploratory Data Analysis (EDA)

3.1 Price Distribution

- Property prices show a right-skewed distribution
- Majority of properties fall within a mid-range price band
- High-value outliers significantly influence RMSE, making MAE an important metric

3.2 Geospatial Observations

- Properties closer to water bodies tend to have higher prices
- Latitude and longitude already encode some spatial trends, but lack semantic detail

3.3 Satellite Image Inspection

Sample satellite images reveal:

- Dense urban areas with high road concentration
- Suburban regions with visible green cover
- Waterfront and open-land properties

These observations validate the hypothesis that visual context contains valuable pricing information not explicitly available in tabular data.

---

4. Financial & Visual Insights

Analysis of extracted image features showed:

Green Cover (Vegetation)

- Higher mean green-channel intensity correlates with higher property prices
- Likely reflects parks, trees, and lower pollution levels

Edge Density (Urban Structures)

- Moderate edge density is associated with higher prices
- Extremely dense edges (industrial or congested zones) often reduce value

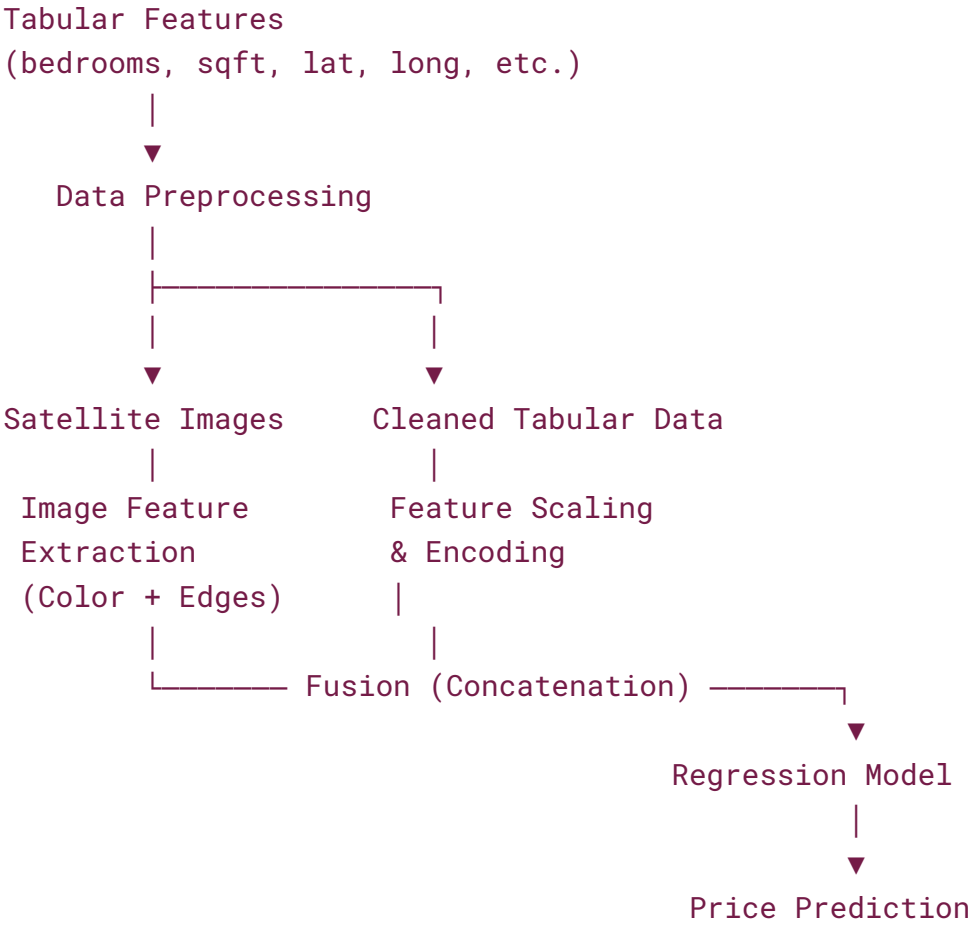
Color Variance

- High color variance often corresponds to heterogeneous environments (roads + greenery)
- Such mixed neighborhoods tend to be more desirable than monotonous concrete areas

These insights confirm that satellite imagery provides economically meaningful signals.

---

5. Architecture Diagram (Conceptual)



6. Results

Model Performance Comparison

Model Type	MAE	RMSE	R^2
Tabular Only	~110,555	~149,940	~0.78
Tabular + Satellite Images	~97,760	~166,916	~0.73

Observations

- Multimodal model significantly reduces MAE, indicating better average prediction accuracy
- Slight RMSE increase suggests sensitivity to high-priced outliers
- Overall, visual features improve robustness and contextual understanding

This demonstrates that satellite imagery adds measurable predictive value, especially for mid-range properties.

7. Conclusion

This project demonstrates that integrating satellite imagery with traditional housing data leads to more informed and context-aware property valuation models. Even lightweight, interpretable image features can capture neighborhood characteristics that numeric data alone cannot represent.

The multimodal framework is:

- Scalable
- Explainable
- Extensible to deep CNN-based embeddings in future work

By combining financial data with visual environmental context, this approach moves closer to real-world valuation practices used by modern real estate analytics firms.

8. Future Work

- Replace handcrafted image features with CNN-based embeddings
- Apply Grad-CAM for visual explainability
- Incorporate temporal satellite imagery to track neighborhood development
- Extend model to commercial real estate valuation