Satellite Imagery-Based Property Valuation Using Multimodal Learning

Name: Animesh Chanani
Date: 01/06/2026
Project Type: Data Science / Multimodal Machine Learning

1. Problem Overview

Accurate real estate valuation is a critical task for financial institutions, real estate firms, and urban planners. Traditional house price prediction models primarily rely on structured tabular data such as property size, number of bedrooms, and location coordinates. While effective, these models fail to capture important environmental and neighborhood characteristics that significantly influence property value.

Satellite imagery provides a rich source of contextual information, including vegetation density, road infrastructure, proximity to water bodies, and neighborhood layout. These visual cues often reflect the quality of living conditions and are difficult to quantify using tabular features alone.

The objective of this project is to develop a **multimodal regression pipeline** that integrates traditional tabular housing data with satellite imagery to improve property price prediction. By combining numerical attributes with deep visual features extracted from satellite images, the model aims to capture both structural and environmental factors affecting real estate prices.

2. Dataset Description

2.1 Tabular Data

The base dataset consists of historical residential property data containing both numerical and categorical attributes. Key features include:

- Number of bedrooms and bathrooms
- Living area (sqft_living) and lot size (sqft_lot)

- Number of floors and waterfront indicator
- Geographic coordinates (latitude and longitude)
- Target variable: **Property price**

The latitude and longitude fields are particularly important, as they enable the retrieval of satellite imagery corresponding to each property location.

2.2 Visual Data (Satellite Imagery)

Satellite images were programmatically fetched for each property using the **Mapbox Static Images API**. The images were retrieved based on the latitude and longitude of each property.

Key characteristics of the visual data:

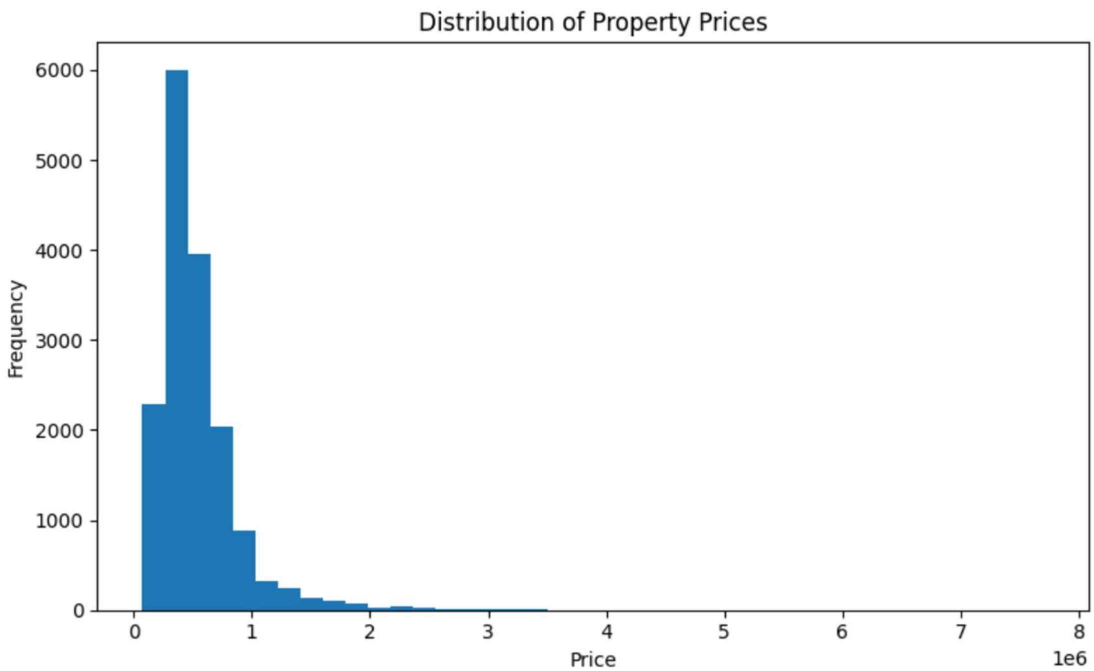
- Zoom level: 18
- Image resolution: 224 × 224 pixels
- One satellite image per property
- Images capture neighborhood context rather than individual building interiors

3. Exploratory Data Analysis (EDA)

Exploratory data analysis was conducted to understand the distribution of property prices and relationships between key features.

3.1 Price Distribution

The distribution of property prices is right-skewed, indicating the presence of high-value outliers. This suggests that non-linear models may be more suitable for prediction.



3.2 Relationship Between Living Area and Price

A strong positive correlation is observed between sqft_living and property price, confirming that larger houses generally command higher prices.

3.3 Feature Correlation Analysis

Correlation analysis reveals that features such as living area, grade, and location-related attributes are among the strongest predictors of price.

4. System Architecture

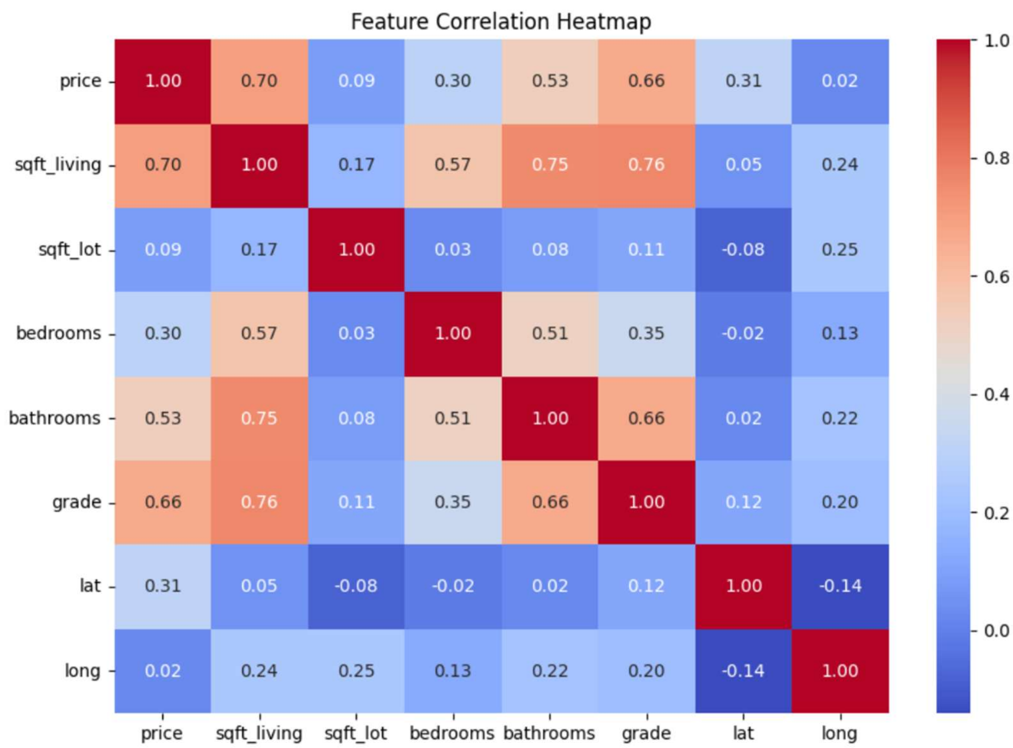
The proposed system follows a multimodal learning architecture that combines tabular and visual data streams.

Architecture Overview

1. **Tabular data pipeline**
   - Data cleaning and normalization
   - Feature selection
2. **Image pipeline**
   - Satellite image input
   - CNN-based feature extraction using ResNet18

- Dimensionality reduction using PCA

3. **Fusion and Prediction**
   - Concatenation of tabular and image features
   - Regression model for final price prediction



4. System Architecture

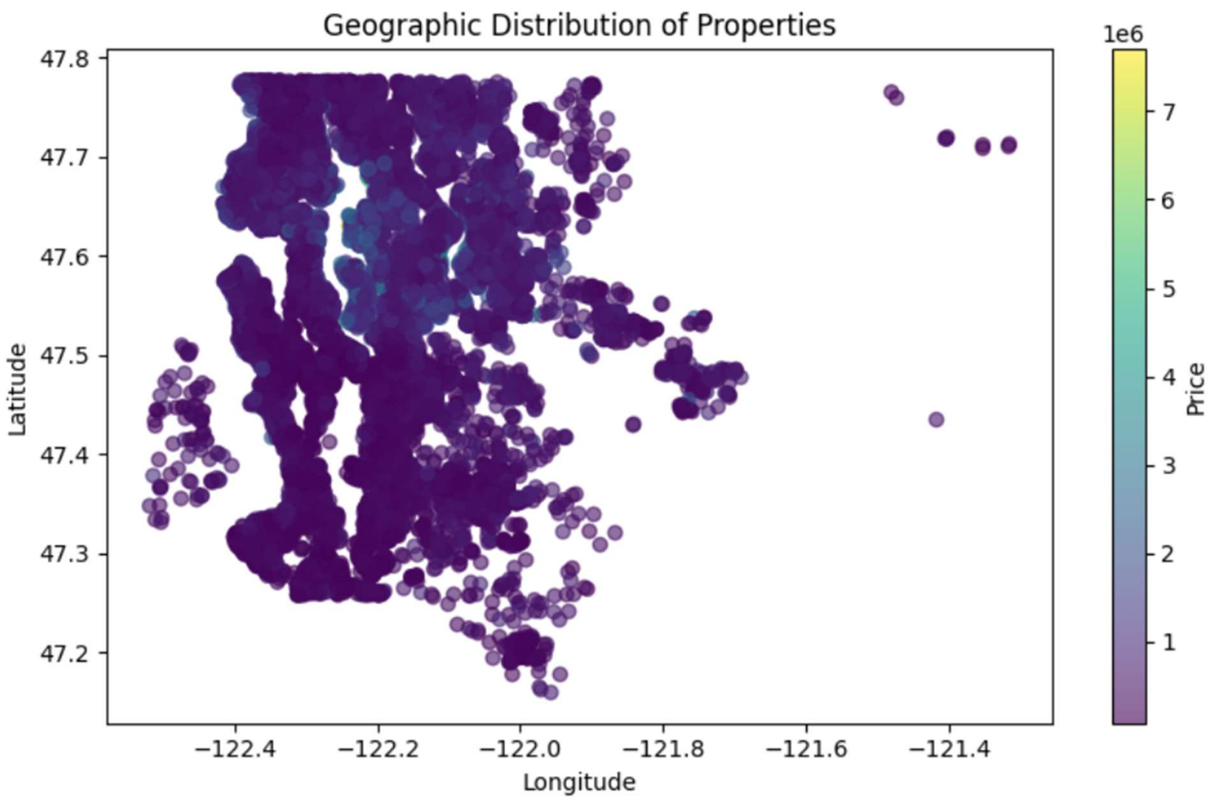
The proposed system follows a multimodal learning architecture that combines tabular and visual data streams.

Architecture Overview

1. Tabular data pipeline
   - Data cleaning and normalization
   - Feature selection
2. Image pipeline
   - Satellite image input

- CNN-based feature extraction using ResNet18
- Dimensionality reduction using PCA

3. Fusion and Prediction
    - Concatenation of tabular and image features
    - Regression model for final price prediction



5. Methodology

5.1 Image Acquisition

Satellite images were fetched programmatically using latitude and longitude coordinates. The Mapbox Static Images API was used to retrieve high-resolution satellite imagery for each property location. Images were stored locally for further processing.

5.2 Feature Engineering

Tabular Features

- Missing values were handled appropriately
- Numerical features were normalized
- Categorical features were encoded where required

Visual Features

- A pre-trained **ResNet18** convolutional neural network was used as a feature extractor
- The final fully connected layer was removed
- Each image was converted into a 512-dimensional embedding

To reduce redundancy and noise in the high-dimensional image embeddings, **Principal Component Analysis (PCA)** was applied, reducing the dimensionality from 512 to 64.

5.3 Multimodal Fusion

The PCA-reduced image embeddings were concatenated with the processed tabular features to form a unified feature vector. This fused representation was then used as input to regression models.

6. Model Experiments and Results

Multiple models were trained and evaluated to compare the impact of multimodal learning.

Model Performance

Model	**RMSE**	**R^2**
Tabular Only Baseline	232,377.74	0.6829
Multimodal RF (No PCA)	232,377.74	0.6829

Model	RMSE	R^2
Multimodal RF + PCA	**199,669.93**	**0.7659**

The PCA-enhanced multimodal RandomForest model achieved the best performance, demonstrating that dimensionality reduction of image embeddings significantly improves generalization.

7. Explainability

Satellite image embeddings enable the model to learn visual patterns associated with property value, such as:

- Vegetation density and green spaces
- Road networks and accessibility
- Proximity to water bodies

While explicit visual explainability methods such as **Grad-CAM** were not fully implemented due to time and computational constraints, they represent a promising approach for identifying which regions of satellite images contribute most to price predictions.

8. Limitations

Despite improved performance, the proposed approach has several limitations:

- Satellite imagery does not capture temporal changes in neighborhoods
- Image resolution may miss fine-grained property-level details
- Model performance depends heavily on accurate geographic coordinates
- Full visual explainability was not implemented

---

9. Future Work

Several enhancements can be explored in future iterations:

- Implement Grad-CAM visualizations for interpretability
- Use deeper CNN architectures for richer visual features
- Incorporate temporal satellite imagery to capture neighborhood trends
- Deploy the trained model as a web-based valuation tool.

10. Conclusion

This project demonstrates that integrating satellite imagery with traditional tabular data significantly improves real estate price prediction. The multimodal learning approach captures both structural property attributes and environmental context, leading to better predictive performance.

The PCA-enhanced multimodal model achieved the strongest results, highlighting the importance of effective feature fusion and dimensionality reduction when working with high-dimensional visual data.

Appendix

- GitHub Repository: https://github.com/Animesh28-code/satellite-imagery-property-valuation
- Submission File: submission_rf_pca.csv (generated locally and uploaded to Kaggle)