Satellite Imagery-Based Property Valuation

Overview

This project focuses on predicting residential property prices by combining structured housing attributes with satellite imagery in a multimodal regression framework. Traditional tabular features capture intrinsic property characteristics such as size, quality, and neighborhood statistics, while satellite images provide visual context about the surrounding environment, including green cover, road density, and proximity to water bodies.

To leverage both data modalities, we design a model with two parallel branches. A convolutional neural network (CNN) is used to extract high-dimensional visual embeddings from satellite images, while a multilayer perceptron (MLP) processes tabular features. These representations are fused and passed to a regression head to predict property prices. The objective is to evaluate whether environmental cues from satellite imagery can improve prediction performance beyond tabular data alone.

Dataset Description

The project uses a hybrid dataset composed of structured tabular data and satellite imagery. The base tabular dataset is derived from a real estate housing dataset containing detailed property attributes and geographic coordinates.

Tabular Data

The tabular dataset includes the following key features:

- `price`: Target variable representing the property sale price.
- `bedrooms`, `bathrooms`: Number of bedrooms and bathrooms.
- `sqft_living`: Total interior living area.
- `sqft_above`, `sqft_basement`: Above-ground and below-ground living areas.
- `sqft_lot`: Total land area of the property.
- `sqft_living15`, `sqft_lot15`: Average living and lot sizes of the nearest 15 neighboring properties, capturing neighborhood density.
- `condition`: Overall condition of the property (1–5).
- `grade`: Construction quality and architectural design (1–13).
- `view`: Quality of the property view (0–4).
- `waterfront`: Binary indicator of whether the property has a waterfront view.
- `lat`, `long`: Geographic coordinates of the property.

Satellite Imagery

Satellite images are programmatically fetched using the latitude and longitude coordinates for each property. Each image represents a localized region around the property and captures visual environmental context such as green spaces, road networks, and nearby water bodies. These images serve as the visual input to the CNN component of the multimodal model.

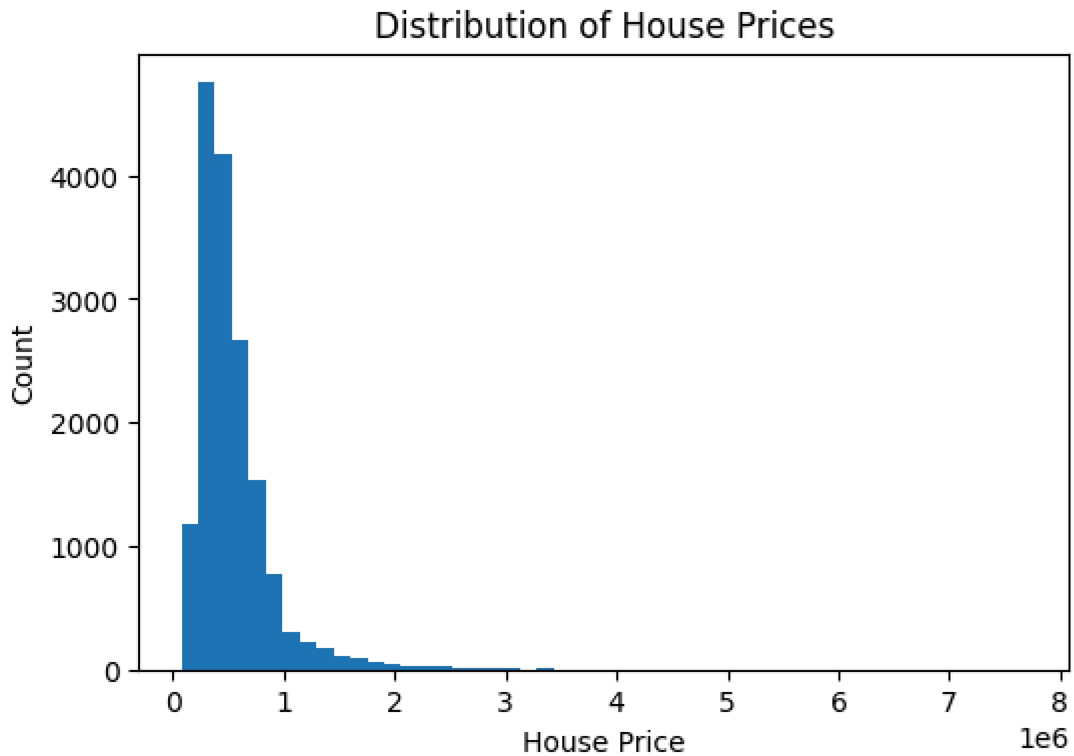
Data Splits

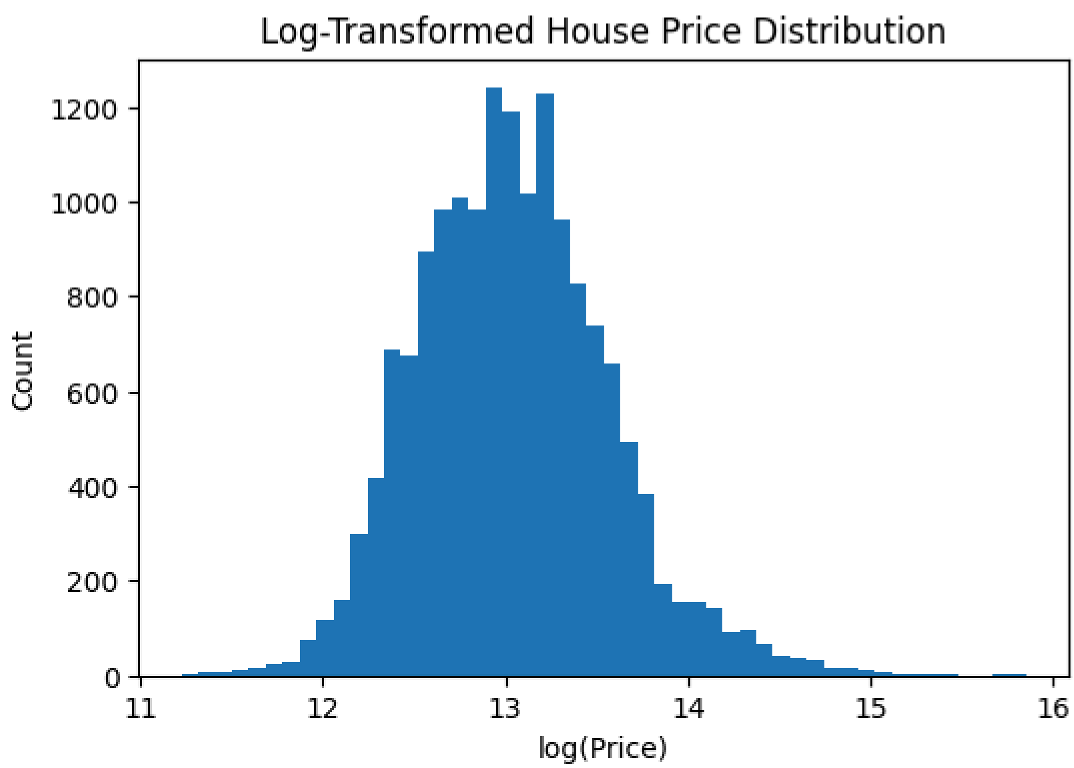
The dataset is split into training, validation, and test sets. Satellite images are available for a subset of properties due to API and computational constraints, and only samples with corresponding images are used for multimodal Training.

Exploratory and Geospatial Analysis

Exploratory data analysis was performed to understand the distribution of house prices, identify influential features, and examine the spatial characteristics of property values. Both tabular and geospatial perspectives were analyzed to motivate the use of satellite imagery in the modeling pipeline.

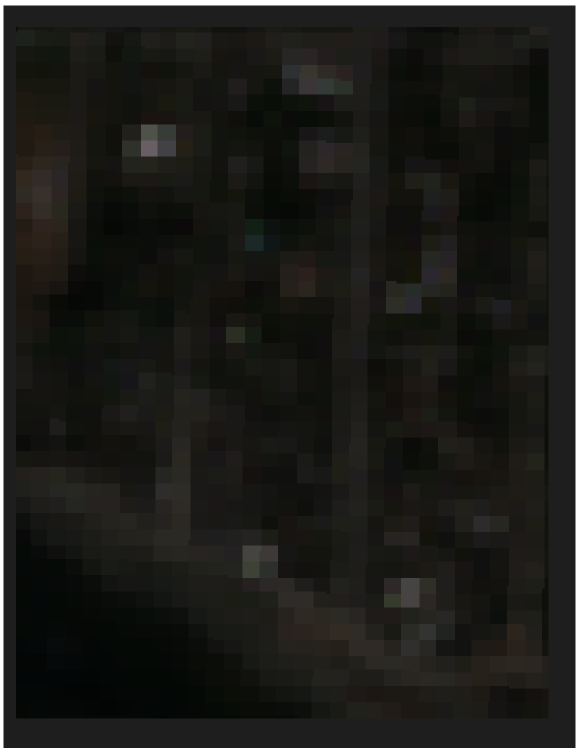
Price Distribution

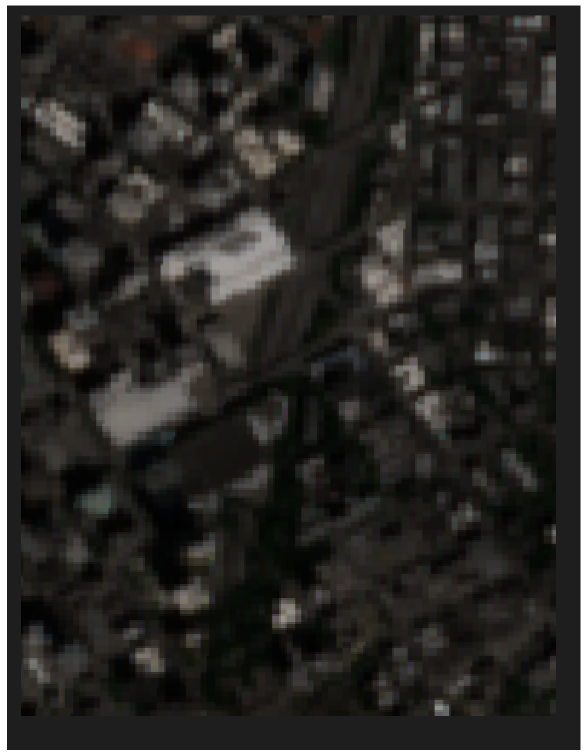




The raw distribution of house prices is highly right-skewed, with a small number of extremely high-priced properties. To stabilize variance and improve regression performance, a logarithmic transformation is applied to the target variable. The log-transformed price distribution is more symmetric and closer to a normal distribution, making it more suitable for predictive modeling.

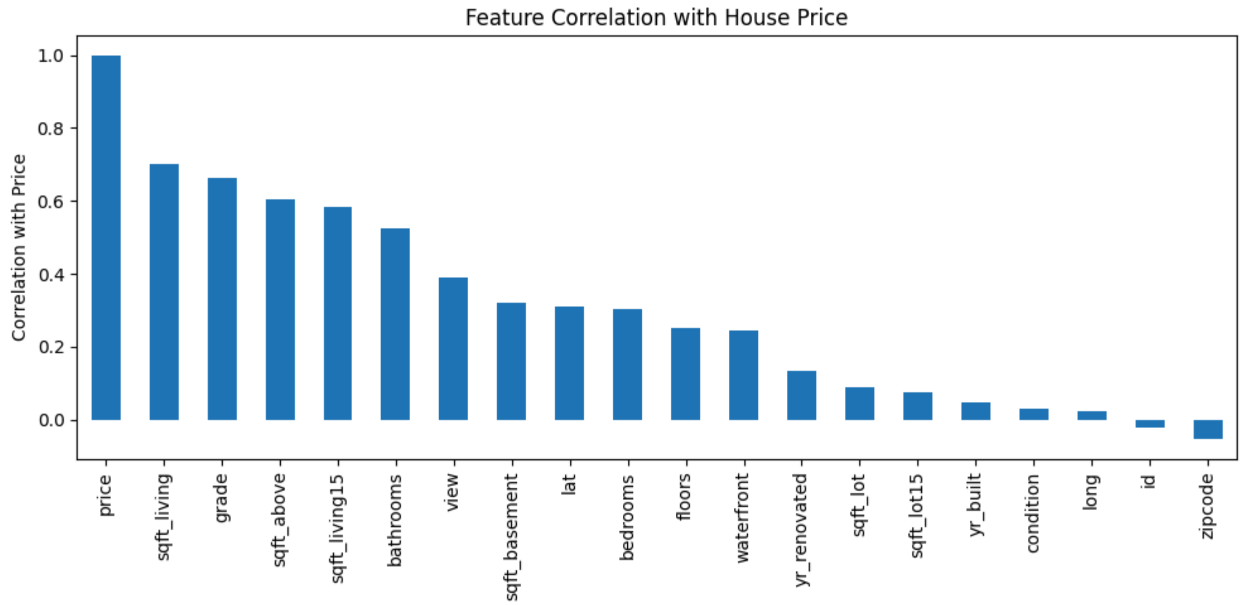
Sample Satellite Images





Sample satellite images illustrate the visual information available to the model. These images capture environmental characteristics such as surrounding green cover, road layouts, building density, and proximity to water bodies. Such visual cues are not explicitly represented in the tabular data and provide additional contextual information for property valuation.

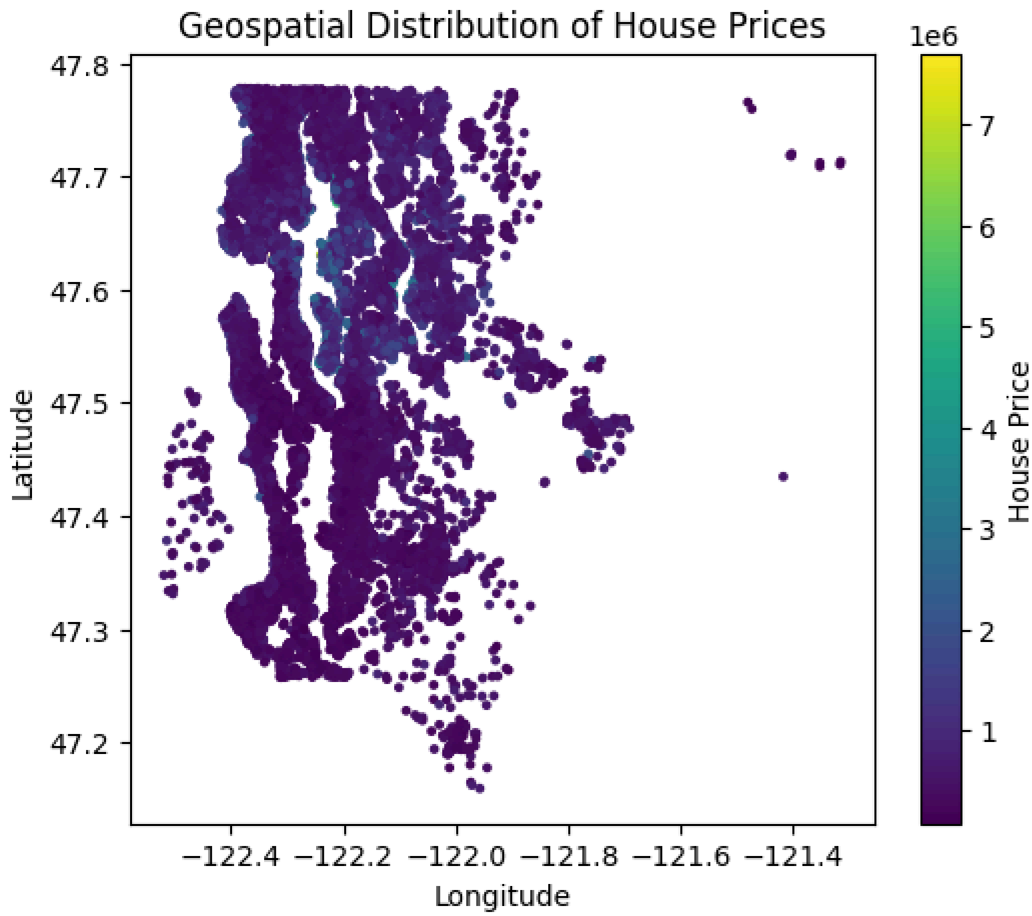
Feature Correlation Analysis



Correlation analysis indicates that structural features such as living area (`sqft_living`), construction quality (`grade`), and above-ground space (`sqft_above`) exhibit strong positive relationships with price. Environmental and neighborhood-related attributes such as `view` and `waterfront` also show positive correlations, while purely identifier-based features like `zipcode` display weak linear relationships.

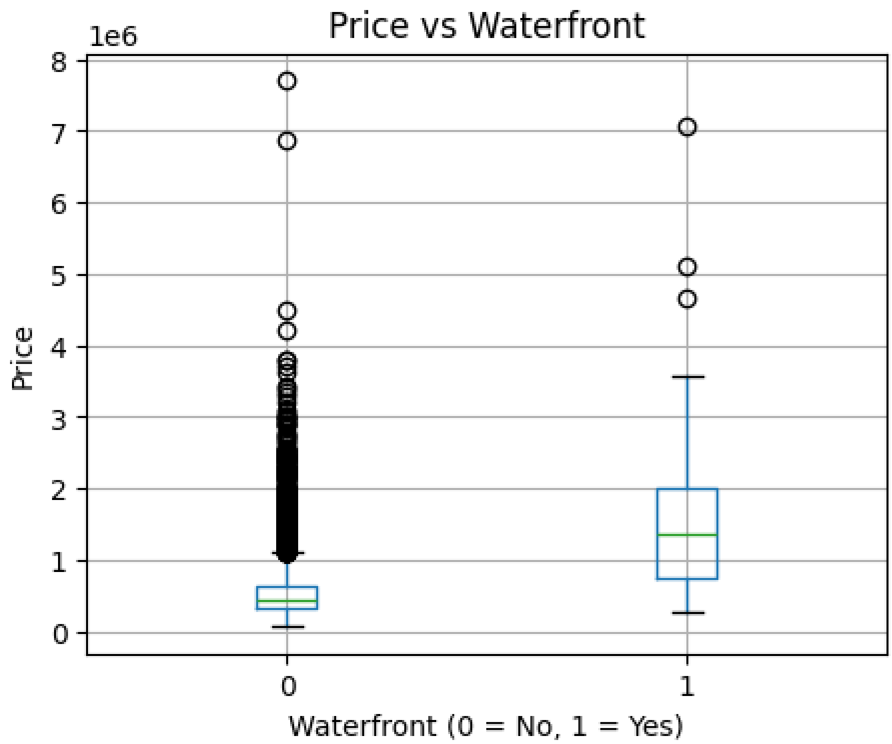
These observations suggest that while tabular features capture important structural information, they may not fully represent environmental and spatial context, motivating the incorporation of satellite imagery.

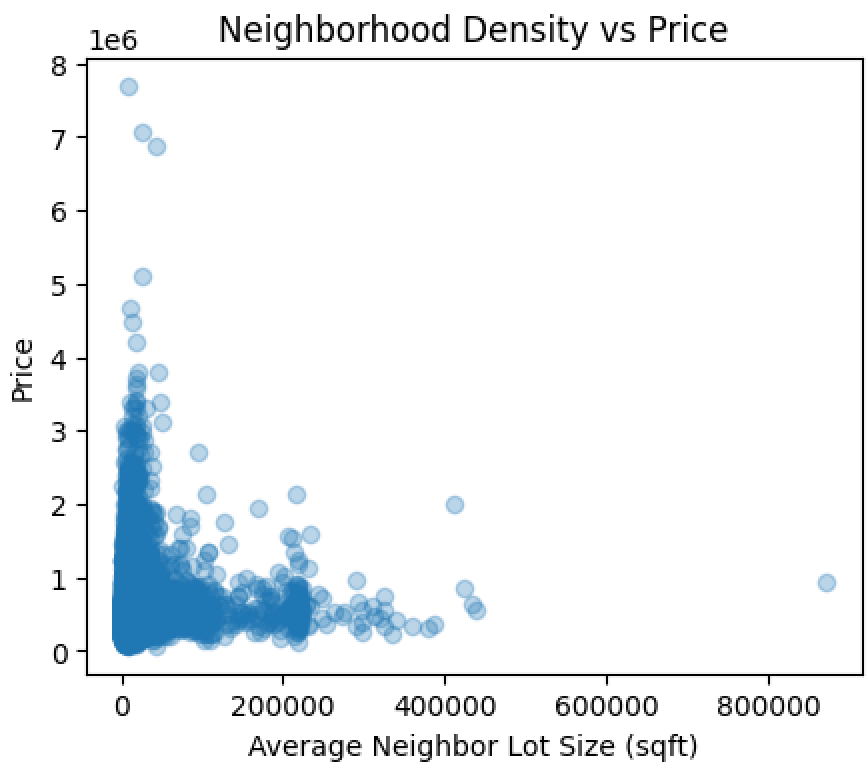
Geospatial Price Distribution



Geospatial visualization reveals strong spatial clustering of property prices, with higher-priced homes concentrated in specific geographic regions. This pattern highlights the importance of location and surrounding environment in real estate valuation and suggests that spatial context plays a critical role in price determination.

Environmental and Neighborhood Effects





Properties with waterfront access exhibit significantly higher prices compared to non-waterfront properties, confirming the premium associated with proximity to water bodies. Additionally, neighborhood density analysis shows that properties located in lower-density areas tend to command higher prices, reflecting the desirability of open and less congested environments.

Methodology

This section describes the modeling approach used for property price prediction. We first establish a tabular only baseline model and then extend it to a multimodal architecture that incorporates satellite imagery. This allows for a direct comparison between traditional structured-data modeling and a multimodal learning approach.

Tabular-Only Model

The tabular baseline model uses structured housing attributes such as property size, construction quality, neighborhood statistics, and geographic coordinates. All numerical features are standardized, and the target variable is log-transformed to reduce skewness.

A multilayer perceptron (MLP) is trained on these features to predict house prices. This model serves as a baseline to evaluate the benefit of incorporating satellite imagery.

Multimodal Model

The multimodal model consists of two parallel branches. The image branch uses a convolutional neural network (CNN) to extract visual embeddings from satellite images, capturing environmental features such as green spaces, road density, and nearby water bodies. The tabular branch uses an MLP to encode structured housing attributes.

The embeddings from both branches are concatenated and passed through a regression head to predict the log-transformed house price. This fusion allows the model to jointly learn from both numeric and visual information.

Architecture Diagram

```
Satellite Image (256×256)
        │
        ▼
   CNN Encoder
        │
        ▼
  Image Embedding
        │
        ├───────────────┐
                        ▼
Tabular Features → MLP Encoder → Tabular Embedding
                        │
                        ▼
                  Concatenation
                        │
                        ▼
                 Regression Head
                        │
                        ▼
              Predicted House Price
```

The proposed architecture consists of two parallel encoders: a CNN-based image encoder that processes satellite imagery and an MLP-based encoder that processes tabular housing features. The learned embeddings from both modalities are fused using concatenation and passed to a regression head to predict the final house price.

Financial & Visual Insights

This section analyzes how both tabular features and visual cues from satellite imagery influence house prices.

1. Insights from Tabular Features (Financial Factors)

From correlation analysis and exploratory data analysis, we observe:

- Living area (`sqft_living`) and construction quality (`grade`) show the strongest positive correlation with house price, indicating that larger and better-built homes are valued significantly higher.

- Bathrooms and above-ground area (`sqft_above`) also contribute strongly to price, reflecting buyer preference for usable living space.
- Properties with waterfront access exhibit a noticeably higher average price compared to non-waterfront properties, confirming the premium associated with water proximity.
- Neighborhood-level features such as average living area of nearby houses (`sqft_living15`) highlight the impact of neighborhood density and surrounding property quality on valuation.

These results suggest that both individual property characteristics and local neighborhood context play an important role in pricing.

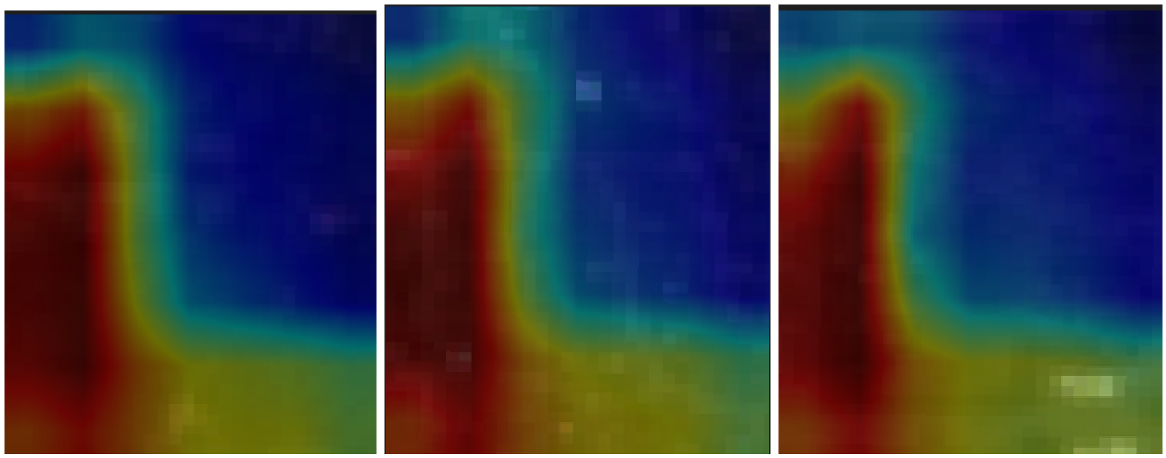
2. Insights from Satellite Imagery (Visual Factors)

Satellite images provide additional spatial and environmental context that is not fully captured by tabular features alone.

Using Grad-CAM visualizations on the trained CNN, we observe that:

- The model tends to focus on coastal boundaries, water bodies, and open spaces, particularly for high-value properties.
- Areas with dense greenery or well-planned residential layouts receive stronger activations, suggesting a positive influence on predicted prices.
- In contrast, regions with high road density or industrial patterns show weaker activations, corresponding to relatively lower predicted prices.

These visual patterns indicate that satellite imagery helps the model capture environmental attractiveness and land-use characteristics, which complements traditional numerical features.



The Grad-CAM visualizations reveal that the convolutional neural network consistently attends to spatially localized regions within the satellite imagery when predicting house prices. Warmer regions (red and yellow) indicate areas that contributed most strongly to the model's output, while cooler regions (blue) represent less influential areas. Across multiple samples, the model focuses on coherent spatial zones rather than random pixels, suggesting that it learns meaningful neighborhood-level visual patterns. These regions likely encode environmental context such as built density, land-use structure, or surrounding infrastructure, which complements the structured tabular features. This demonstrates that satellite imagery provides additional predictive signals beyond traditional housing attributes and supports the effectiveness of the multimodal learning approach.

Results Comparison

To evaluate the contribution of satellite imagery, we compare the performance of two models: a tabular-only regression model and a multimodal model that integrates both tabular features and satellite image embeddings. Model performance is assessed using Root Mean Squared Error (RMSE) and R^2 score on the validation set.

Quantitative Results

Model	RMSE ↓	R^2 ↑
Tabular Only	0.64	0.67
Tabular + Satellite Images	0.25	0.77

The multimodal model achieves a substantial reduction in RMSE and a significant improvement in R^2 score compared to the tabular-only baseline. This indicates that incorporating satellite imagery enables the model to explain a larger fraction of price variance and produce more accurate predictions.

Analysis

The tabular-only model relies solely on structured attributes such as square footage, number of rooms, location coordinates, and neighborhood statistics. While these features capture many important factors, they fail to represent visual and spatial characteristics of the surrounding environment. By integrating CNN-derived image embeddings, the multimodal model gains access to additional contextual cues related to neighborhood layout and land-use patterns. The observed performance improvement confirms that satellite imagery provides complementary information beyond traditional tabular features

Incorporating satellite imagery via a CNN-based encoder significantly improves house price prediction accuracy over tabular features alone, validating the effectiveness of the multimodal regression approach.

Conclusion

In this project, we developed a multimodal regression pipeline for property price prediction by combining structured housing attributes with satellite imagery. Traditional tabular features such as square footage, location, and neighborhood statistics were complemented with visual context extracted from satellite images using a convolutional neural network (CNN). This approach allowed the model to capture both numerical and environmental factors that influence real estate valuation.

Exploratory and geospatial analysis revealed strong spatial patterns in housing prices, including higher values near waterfronts and in dense, well-developed neighborhoods. Log-transforming the target variable stabilized variance and improved model training. Visual analysis of satellite imagery further suggested that land-use patterns and surrounding infrastructure contribute meaningful signals beyond tabular data alone.

Quantitative evaluation demonstrated that the multimodal model significantly outperformed the tabular-only baseline, achieving lower RMSE and higher R^2 scores. This improvement confirms that satellite imagery provides complementary information that enhances predictive accuracy. Grad-CAM visualizations offered interpretability by highlighting image regions that influenced model predictions, supporting the model’s reliance on meaningful spatial features rather than noise.

Overall, the results validate the effectiveness of multimodal learning for property valuation tasks. Future work could extend this framework by incorporating higher-resolution imagery, temporal satellite data, or additional geospatial features to further improve performance and interpretability.