PROJECT REPORT

Multimodal Property Price Prediction Using Tabular Data and Satellite Imagery

Submitted by

Harsh Rishi

Enrollment Number: **24113050**

Academic Year 2025–2026

1 Overview: Multimodal House Price Prediction

Accurate residential property valuation traditionally relies on structured tabular attributes such as property size, age, location, and quality indicators. However, these features often fail to capture neighborhood-level context, including surrounding infrastructure, greenery, road density, and urban layout.

This project builds a **multimodal regression pipeline** that combines tabular housing data with satellite imagery-derived features to predict residential property prices. The primary objective is to quantify whether incorporating satellite imagery provides measurable improvements over strong tabular-only baselines.

2 Modeling Strategy

The modeling strategy follows a modular multimodal pipeline, progressing from single-modality baselines to complex fusion models.

2.1 Data Preparation and Baseline Establishment

- Tabular preprocessing with cleaning and log-transformations.
- Baseline tabular-only and image-only regression models.

2.2 Vision Feature Engineering

- Satellite image acquisition using NAIP imagery.
- Feature extraction using a pretrained ResNet-50.
- PCA-based dimensionality reduction.

2.3 Multimodal Fusion Techniques

- Early fusion using concatenated features.
- Intermediate fusion with modality-specific representations.

2.4 Evaluation and Explainability

- Evaluation using RMSE and R^2 on log-transformed prices.
- Interpretability via Grad-CAM visualizations.

3 Dataset Description

3.1 Tabular Data

The tabular dataset contains historical residential property sales information with a diverse set of structured attributes, including:

- **Structural attributes:** Property-level characteristics such as *sqft_living*, *grade*, and *yr_built*.
- **Temporal attributes:** Time-related features such as the year of sale (*sale_year*).
- **Geospatial attributes:** Latitude and longitude coordinates representing property location.
- **Derived features:** Engineered variables including property age at the time of sale.

The target variable is the residential property price. To stabilize variance and reduce skewness, the target is log-transformed and represented as *price_log*.

3.2 Satellite Imagery Data

Satellite images are collected for each property using its corresponding latitude and longitude coordinates. These images capture neighborhood-level context that is not directly available in tabular data.

Image acquisition details:

- **Resolution:** 512×512 pixels
- **Zoom level:** 18
- **Image format:** GeoTIFF
- **Coverage:** Neighborhood-scale spatial context surrounding each property

Each satellite image is aligned with a unique property identifier to ensure a one-to-one mapping between tabular records and visual inputs.

4 Data Preprocessing

4.1 Data Cleaning

The following data cleaning steps are applied to ensure data quality and consistency:

- Removal of duplicate property records.
- Identification and handling of invalid or extreme values.
- Correction of inconsistencies in bedroom and bathroom counts.
- Derivation of property age features using construction year and sale year.

4.2 Target Transformation

Residential property prices exhibit strong right-skewness and heteroscedasticity, which can negatively impact regression model performance. To address this, the target variable is transformed as follows:

$$price\_log = \log(price)$$

All models are trained and evaluated using the log-transformed price, which improves numerical stability and enables more interpretable error metrics such as RMSE and R^2.

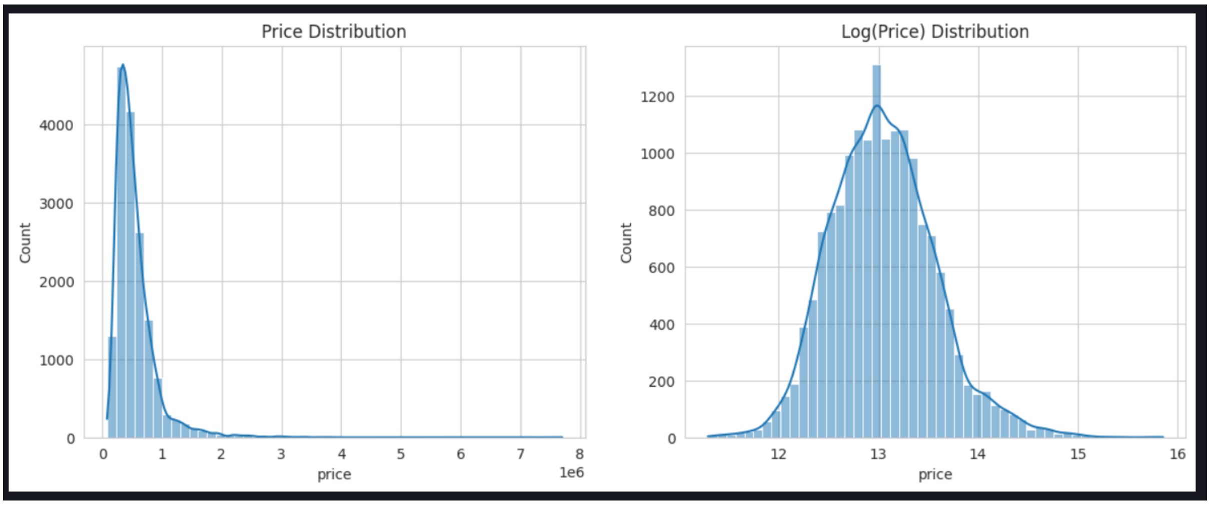

Figure 1: Sample Satellite Imagery for Residential Properties

5 Feature Engineering

In this stage, structural, temporal, and spatial features are engineered to enhance the predictive performance of the house price regression models.

5.1 Temporal Features

The age of each property at the time of sale is computed as a key temporal feature. To ensure accuracy, the renovation year is prioritized over the original construction year whenever a renovation has occurred. This allows the model to better reflect the effective age of the property.

5.2 Structural Features

Structural feature engineering focuses on capturing relationships between existing physical attributes. In particular, the ratio of living area to total lot area is computed. This feature helps the model understand how much of the available land is occupied by usable living space, which can serve as a proxy for urban density or property luxury.

5.3 Categorical Encoding

Although zip codes are represented numerically in the raw dataset, they function as categorical identifiers rather than continuous variables. Therefore, zip codes are explicitly treated as categorical features during preprocessing to prevent the model from learning invalid ordinal relationships.

5.4 Spatial Interaction Features

To capture localized geographic effects, spatial interaction features are engineered by combining latitude and longitude coordinates. These interactions enable the model to learn fine-grained spatial patterns that influence property prices beyond simple distance-based effects.

```
# Extract sale year
df['sale_year'] = df['date'].dt.year
# Effective construction year (renovation > built)
df['effective_year'] =df['yr_built']
df.loc[df['yr_renovated'] > 0, 'effective_year'] = df['yr_renovated']


# House age
df['house_age'] = df['sale_year'] - df['effective_year']
df.loc[df['house_age'] < 0, 'house_age'] = 0

# Living to lot ratio
df["living_to_lot_ratio"] = df["sqft_living"] / df["sqft_lot"]

# Zipcode as categorical
df["zipcode"] = df["zipcode"].astype(str)

# Latitude□Longitude interaction
df["lat_long_interaction"] = df["lat"] * df["long"]
```

Figure 2: Sample Satellite Imagery for Residential Properties

6 Exploratory Data Analysis (EDA)

6.1 Relationship Between Waterfront Status, Living Area, and Price

This analysis explores how waterfront status interacts with living area to influence residential property prices. Boxplots and scatter plots are used to examine both individual and interaction effects.

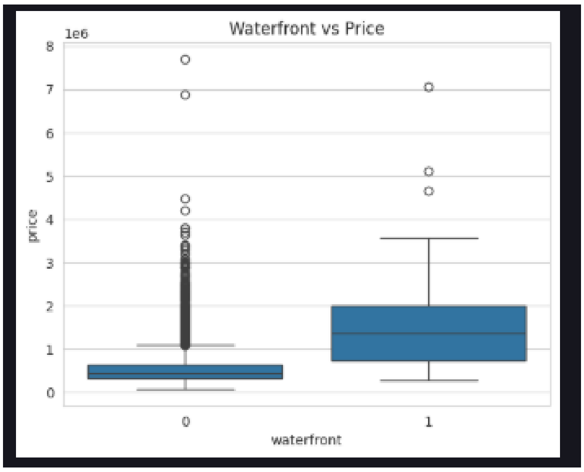

Figure 3: Waterfront vs Price Distribution

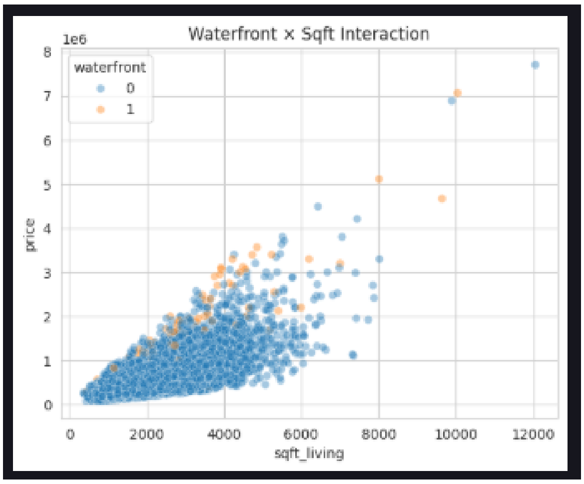

Figure 4: Waterfront and Living Area Interaction

Key Insights

1. **Waterfront properties command significantly higher prices.** The boxplot reveals a clear upward shift in price distribution for properties with waterfront access compared to non-waterfront properties. Both the median and upper quartile prices are substantially higher, with a greater concentration of high-end outliers.

2. **Living area exhibits a strong positive relationship with price.** Scatter plots indicate a strong positive correlation between *sqft_living* and property price for both waterfront and non-waterfront homes, confirming living area as a dominant pricing factor.

3. **Waterfront status acts as a price multiplier rather than a fixed premium.** For equivalent living areas, waterfront properties consistently achieve higher prices than non-waterfront properties. Moreover, the price differential increases with larger house sizes, indicating disproportionate value gains for larger waterfront homes.

6.2 Feature Correlation Analysis

This analysis examines the pairwise correlations between numerical features and the target variable using a correlation heatmap. The visualization helps identify strong predictors of property price, detect potential multicollinearity among features, and validate the effectiveness of engineered variables prior to model training.

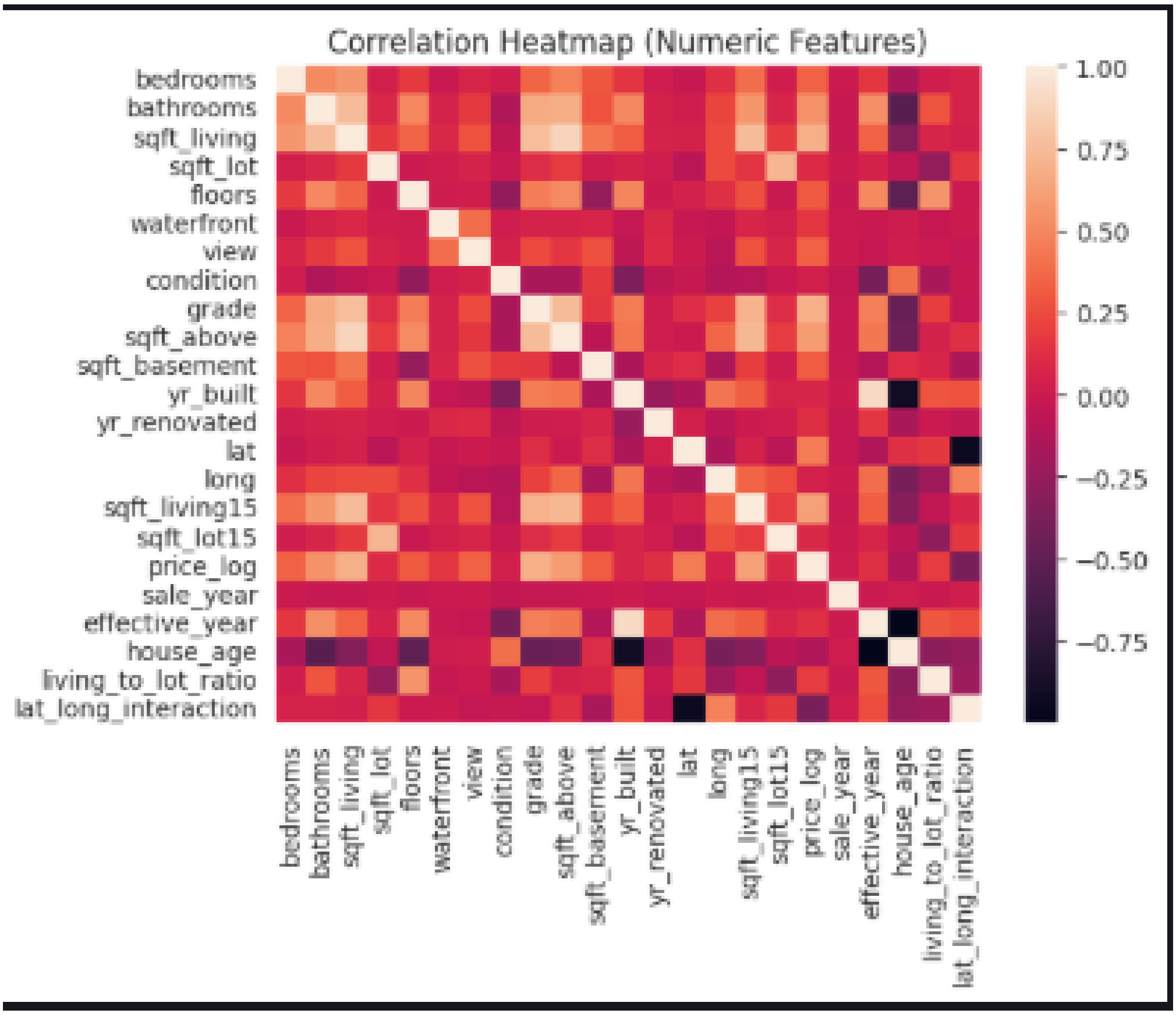

Figure 5: Correlation heatmap of numerical features and log-transformed property price

6.3 Spatial Distribution of Properties (Latitude–Longitude)

This analysis visualizes the geographic distribution of properties using latitude and longitude coordinates to understand spatial clustering and regional structure within the dataset.

Key Insights

1. **Strong geographic clustering.** Properties are not uniformly distributed across space; instead, they form dense clusters, indicating concentrated urban and suburban regions rather than a dispersed rural layout.

Figure 6: Spatial distribution of residential properties based on latitude and longitude

2. **Clear regional boundary.** The observed latitude–longitude range forms a tight spatial band, suggesting that all properties belong to a single metropolitan region, similar to a large city cluster, rather than multiple geographically distinct cities.

3. **Visible influence of water bodies.** Curved clusters and vertical gaps in the spatial distribution align with coastlines and water bodies. This spatial structure reinforces why waterfront proximity emerges as a strong price driver in earlier exploratory analyses.

6.4 Price Heatmap vs. Location

This analysis visualizes the spatial distribution of property prices by overlaying price intensity on geographic coordinates. The resulting heatmap highlights neighborhood-level price patterns that are not immediately evident from tabular features alone.

Key Insights

1. **High-value properties exhibit strong spatial clustering.** Expensive homes are not randomly distributed across the region; instead, they form distinct geographic hotspots, indicating localized pockets of high demand.

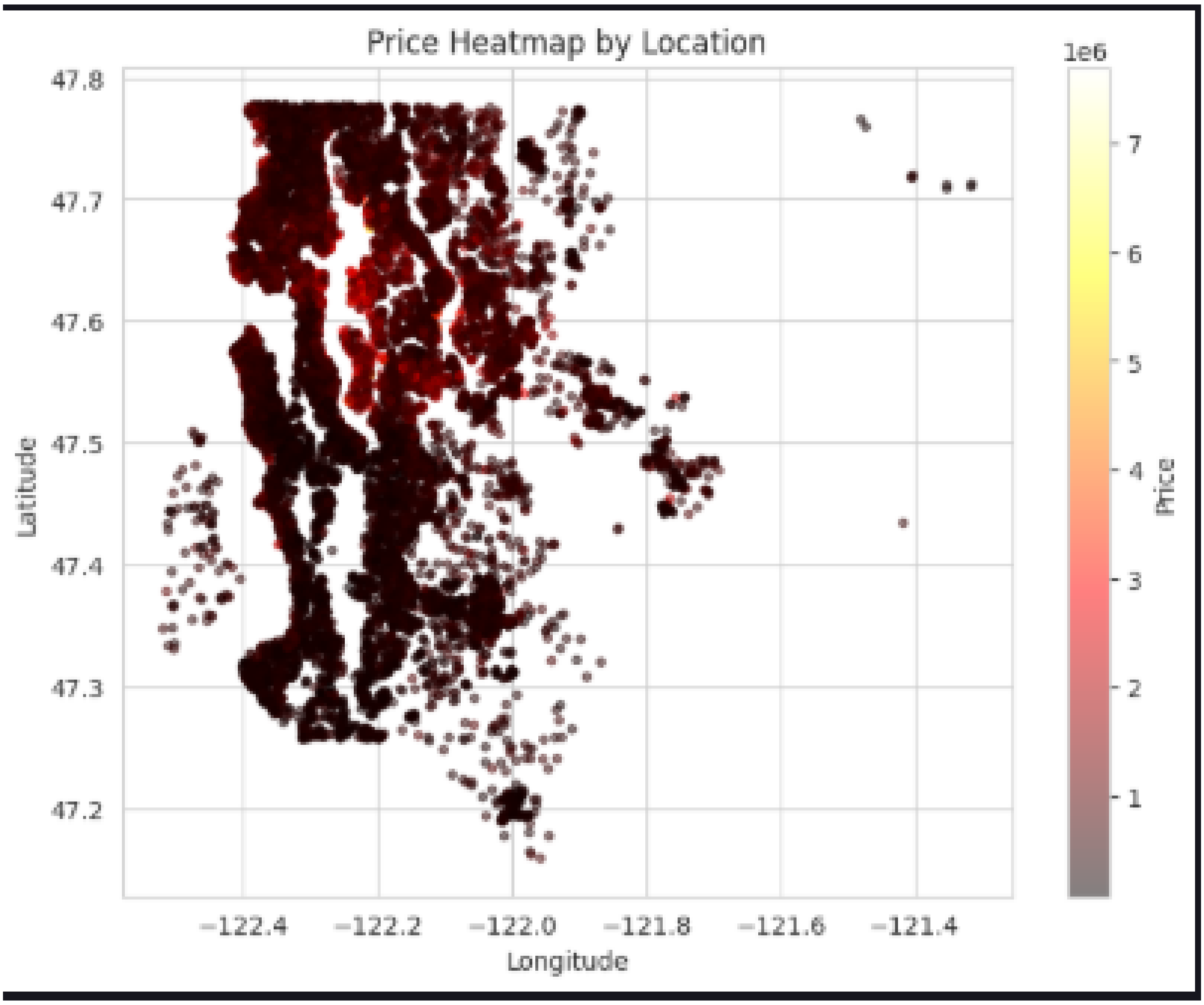

Figure 7: Spatial heatmap of residential property prices across geographic locations

2. **Premium prices align with waterfront and central urban areas.** Regions with the highest price intensity coincide with coastal or water-adjacent zones as well as well-developed urban centers, reinforcing the economic value of location and accessibility.

3. **Sharp price gradients occur across short distances.** Property prices change significantly over relatively small geographic shifts, highlighting strong neighborhood-level effects that extend beyond individual property characteristics.

6.5 Price vs. Distance to City Center

This analysis examines the relationship between residential property prices and their distance from the city center to quantify the effect of urban centrality on valuation.

Key Insights

1. **Strong negative relationship between price and distance.** Property prices generally decrease as the distance from the city center increases, indicating a clear location premium associated with central areas.

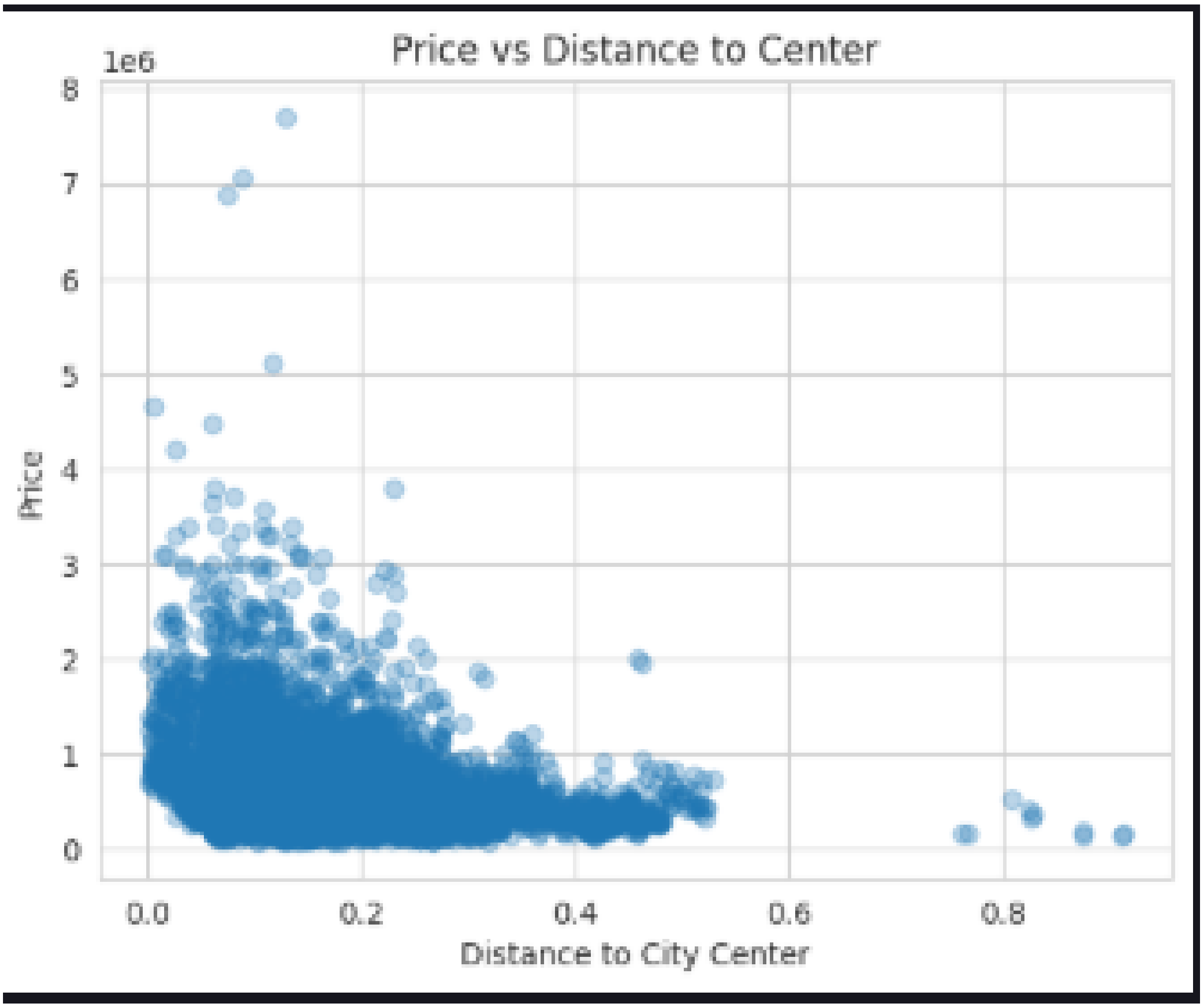

Figure 8: Relationship between property price and distance from the city center

2. **High-value properties are concentrated near the city center.** The most expensive homes are predominantly located at shorter distances from the center, reflecting strong centrality effects and higher demand for centrally located properties.

3. **Diminishing price sensitivity at larger distances.** Beyond a certain distance threshold, property prices tend to flatten and cluster at lower values, suggesting that proximity to the city center has the greatest influence within nearby neighborhoods.

6.6 Geographic Clusters vs. Location

This analysis investigates spatial clustering patterns to identify natural neighborhood groupings and localized submarkets within the study region.

Key Insights

1. **Distinct neighborhood segmentation.** Properties naturally separate into well-defined geographic clusters, indicating clear neighborhood or submarket boundaries within the metropolitan area.

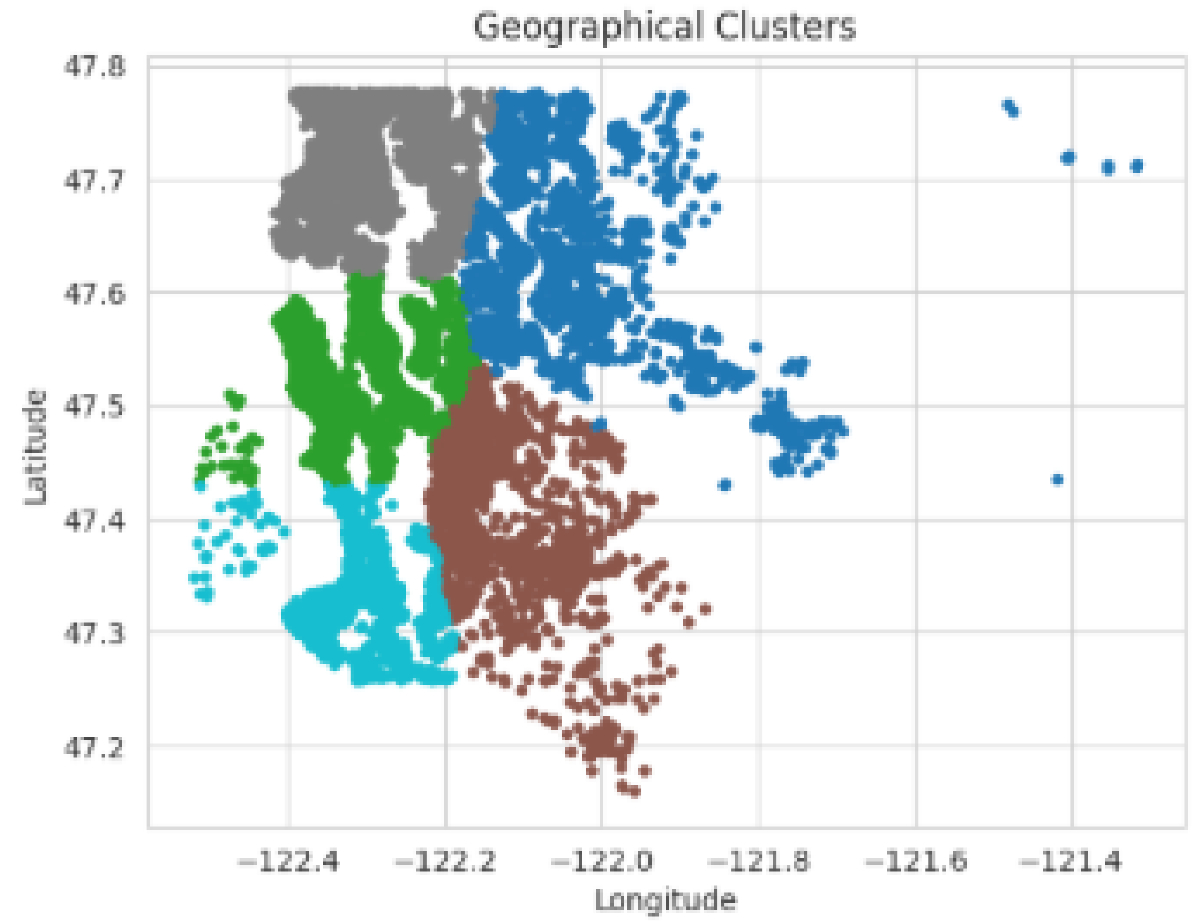

Figure 9: Geographic clustering of properties based on spatial location

2. **High intra-cluster similarity.** Properties within the same cluster are spatially proximate and are likely to share similar characteristics, such as accessibility, nearby amenities, and comparable price behavior.

3. **Strong inter-cluster differences.** Different clusters occupy visibly distinct regions, suggesting the presence of location-based pricing regimes rather than a single homogeneous housing market.

7 Image Feature Extraction

Satellite images are resized to 224×224 pixels to match the input requirements of a pretrained CNN.

A pretrained CNN (ResNet-based) is used as a fixed feature extractor, producing dense embeddings for each image. No end-to-end fine-tuning is performed.

To reduce dimensionality and noise:

- PCA is applied to image embeddings
- Embeddings are reduced to a compact representation

These embeddings act as high-level visual descriptors of neighborhood context.

8 Evaluation & Explainability

Grad-CAM is used to visualize which regions of satellite imagery most influence the model's predictions. Grad-CAM is applied to the pretrained CNN used for image feature extraction, providing qualitative insights into the visual patterns captured by the model.

The activation maps indicate that the model consistently focuses on:

- Dense urban regions
- Road networks and built-up areas
- Surrounding infrastructure patterns

Areas with higher greenery or lower construction density show comparatively weaker activation.

This suggests that the model associates urban development intensity and infrastructure density with higher predicted property prices.

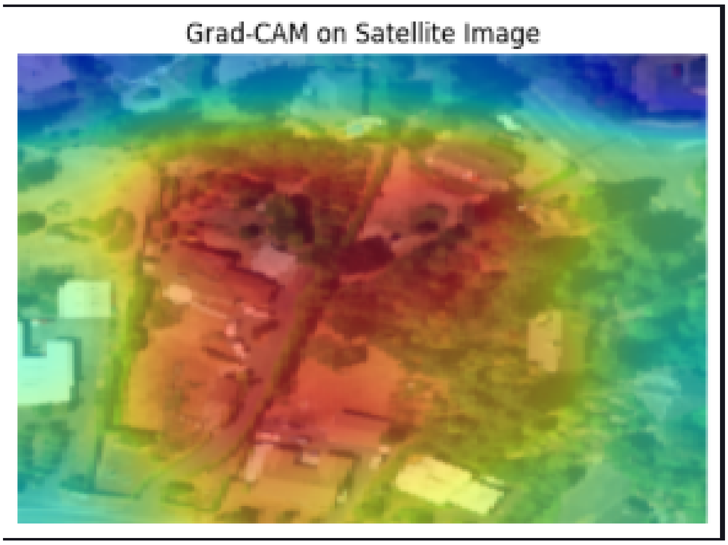

Figure 10: Geographic clustering of properties based on spatial location

9 Financial & Visual Insights

1. **Green Cover vs. Built Density:** Properties surrounded by visible green spaces, lower road density, and open land tend to receive slightly higher predicted values compared to areas dominated by dense concrete structures. This suggests that environmental quality and perceived livability contribute positively to valuation.

2. **Urban Structure and Connectivity:** Regular road layouts, organized housing clusters, and well-connected neighborhoods correlate with higher predicted prices. In contrast, irregular or congested built patterns are associated with lower valuations.

3. **Water and Open-Space Proximity:** Visual indicators of nearby water bodies or open spaces introduce localized premiums, even when such information is not explicitly present in the tabular features.

10 Architecture Diagram

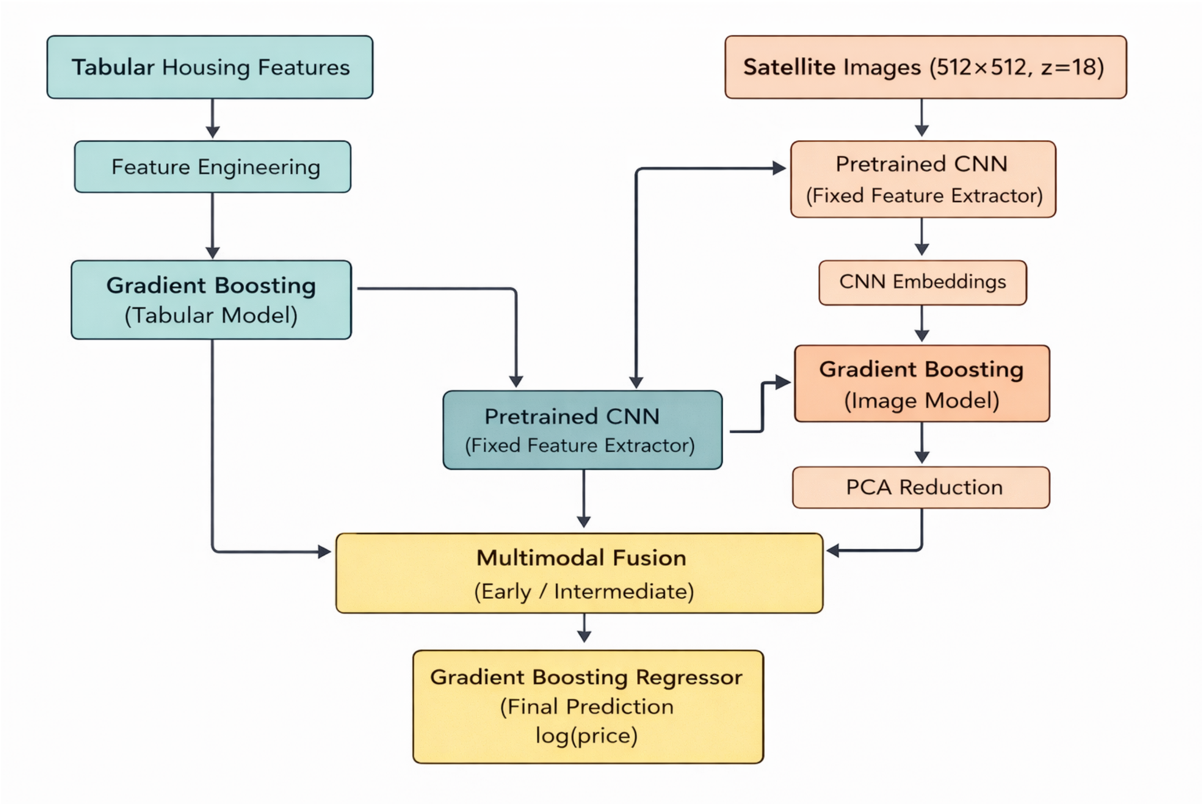

Figure 11: Geographic clustering of properties based on spatial location

11 Results: Tabular Data Only vs. Tabular + Satellite Images

11.1 Model Comparison Setup

1 Tabular-Only Model (Baseline)

Model Used: Gradient Boosting Regressor (GBR)

Input Features

- Structured housing attributes only
- Numeric features (after scaling)
- Target: *price_log*

Purpose

- Establish a strong baseline
- Measure how well traditional tabular data alone predicts prices

2 Early Fusion Model (Tabular + Image)

Model Used: Gradient Boosting Regressor

Fusion Strategy

- CNN image embeddings concatenated directly with tabular features
- Single combined feature vector → GBR

Input Features

- Scaled tabular features
- Satellite image embeddings (pre-extracted)
- Target: *price*

Purpose

- Test whether adding raw visual information improves prediction

3 Intermediate Fusion Model (Tabular + Image)

Model Used: Gradient Boosting Regressor

Fusion Strategy

- Learn compact representations of:
  - Tabular features
  - Image embeddings
- Combine learned representations → GBR

Input Features

- Transformed tabular representation
- Transformed image representation
- Target: *price*

Purpose

- Allow each modality to learn independently before combining

11.2 Quantitative Performance Comparison

The models were evaluated using Root Mean Squared Error (RMSE), where lower values indicate better performance, and the R^2 score, where higher values indicate stronger explanatory power. The comparison highlights the impact of different data modalities and fusion strategies on prediction accuracy.

[21…

	Model	RMSE	R^2
0	Tabular Only	0.181751	0.883536
1	Image Only	0.532884	-0.001164
2	Early Fusion	0.181613	0.883712
3	Intermediate Fusion	0.206117	0.850215

+ Code + Markdown

Figure 12: Geographic clustering of properties based on spatial location

Interpretation & Insights

- **Tabular-Only Model** Indicates that structured housing attributes explain a large portion of price variance.
- **Image-Only Model** Negative R^2 indicates that the model performs worse than predicting the mean, showing that satellite imagery alone is insufficient for accurate valuation.
- **Early Fusion Model** Achieves the best overall performance. Even though the improvement over the tabular-only model is marginal, it demonstrates that adding image features provides incremental predictive value.
- **Intermediate Fusion Model** Performance drops slightly compared to early fusion, suggesting that the learned intermediate representations may have introduced information loss or over-compression.

11.3 Tabular Feature Importance Analysis

Tabular Feature Insights

The most influential tabular features include:

- Geographic location (latitude and spatial interactions)
- Property size and construction quality
- Neighborhood housing density indicators

These findings align with real estate domain knowledge, highlighting the importance of location and the built environment while reinforcing confidence in the model's predictions.

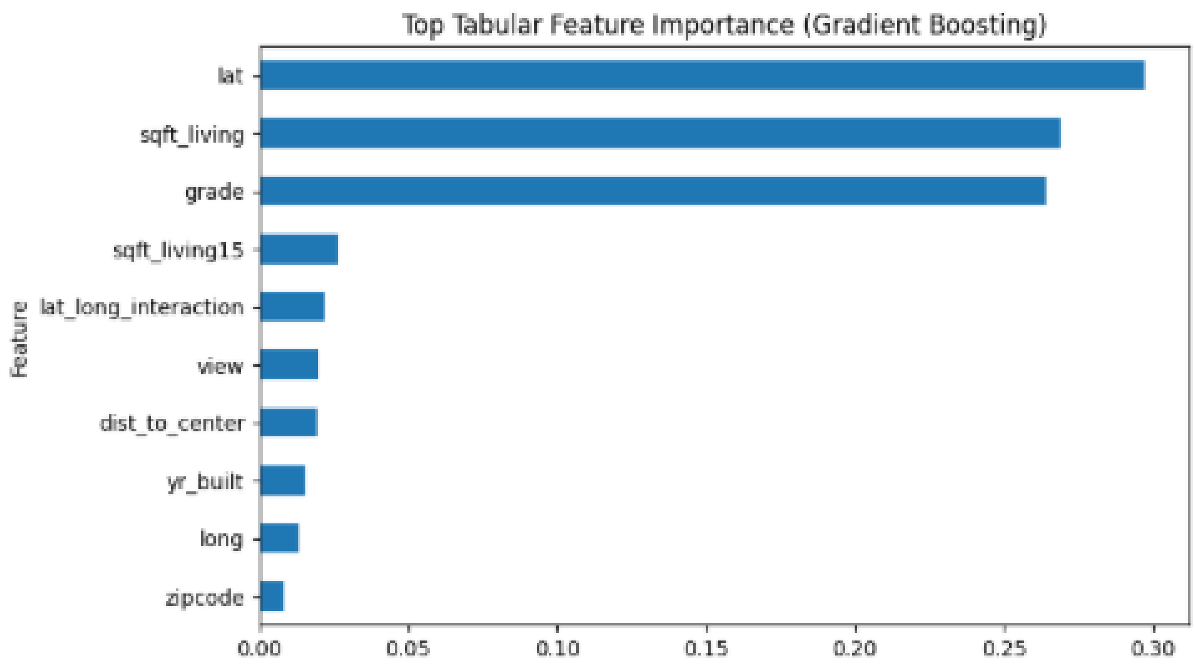

Figure 13: Top tabular features importance

12 Conclusion

This project explores a multimodal approach to residential property valuation by combining tabular housing data with satellite imagery. Strong tabular-only models confirm that location, property size, and construction quality are the primary drivers of price prediction.

The inclusion of satellite imagery provides complementary neighborhood-level context and results in modest but consistent performance improvements when fused with tabular features. Explainability analysis using Grad-CAM further indicates that the model focuses on meaningful urban and infrastructural patterns.

Overall, the findings demonstrate that while tabular data remains the dominant signal, visual features enhance predictive robustness, supporting the value of multimodal learning in real estate analytics.