Satellite Imagery-Based Property Valuation

A dual-stream predictive framework that integrates high-dimensional satellite embeddings with engineered tabular features to estimate real estate market values.

Author: Ayushman Kumar
Enrollment No: 23117040

GitHub: https://github.com/ayushman13x/satellite-imagery-model

Contents

1 Introduction

1.1 Problem Statement

The central challenge in automated real estate valuation is the data gap between a property's physical attributes and its environmental context. While traditional models effectively process internal metrics like square footage and room counts, they remain "blind" to external value drivers such as neighborhood greenery, urban density, and infrastructure quality which are absent from tabular records. Consequently, the task of this project is to overcome this limitation by architecting a multimodal pipeline that can ingest and interpret high-resolution satellite imagery. The objective is to extract latent spatial features from these images and successfully fuse them with structured housing data, ensuring the final predictive model can account for the critical environmental premiums that traditional, tabular-only approaches fail to capture

1.2 Project Objectives

The objectives of this project are as follows:

- Develop a Multimodal Fusion Pipeline: Architect an end-to-end machine learning workflow that successfully integrates high-dimensional visual data with traditional structured housing records

- Implement Transfer Learning for Feature Extraction: Utilize a pre-trained MobileNetV2 backbone to translate raw satellite pixels into meaningful latent features representing neighborhood quality and environmental context.

- Optimize Feature Selection: Isolate the "Golden 20" most predictive visual markers from the initial 1,280-dimension output to prevent model overfitting and ensure computational efficiency

- Engineer Advanced Tabular Metrics: Construct 14 custom features, including standardized property age (2015 baseline) and quality-interaction indices, to capture complex non-linear relationships in the data

- Enhance Predictive Accuracy: Train a unified XGBoost regressor on the fused dataset to demonstrate the "predictive lift" and context-awareness gained by including satellite-derived information

2 Architecture Diagram

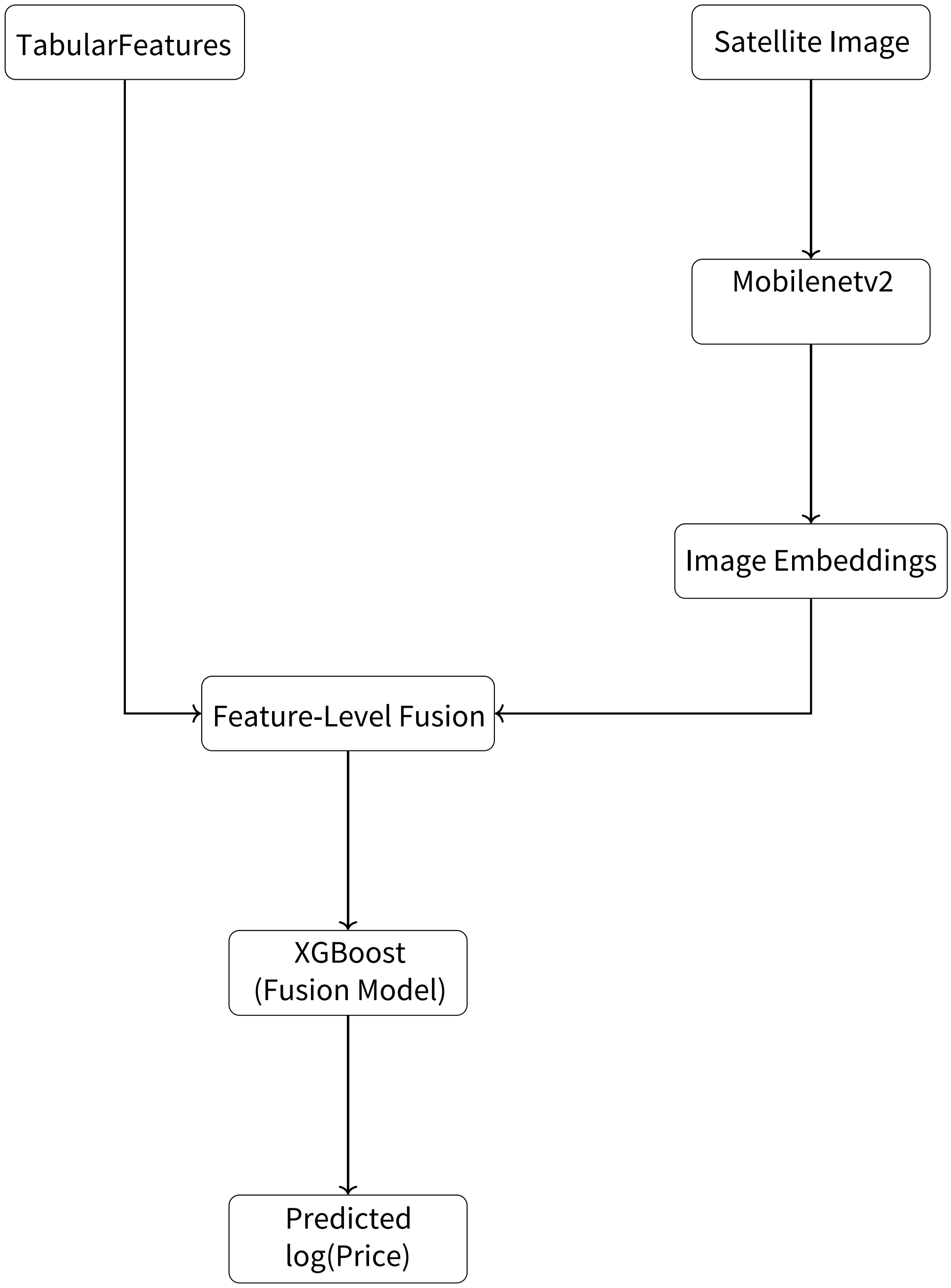


Figure 1: Architecture of the satellite–tabular fusion pipeline for property price prediction.

3 Dataset Description

3.1 Tabular Dataset

The dataset includes a diverse set of structured features describing the physical characteristics of the property, its location, and its surrounding environment. Thedataset is split into a training set containing the target variable and a test set used solely for evaluation. After preprocessing, the training dataset contains 16,209 records, while the test dataset contains 5,404 records. No missing values were observed in either split

3.2 Target Variable

The target variable is the transaction price of a residential property. Exploratory analysis shows that raw property prices are highly right-skewed, with a small number of extremely high-valued properties.
To stabilize model training and improve regression performance, the target variable is transformed using a logarithmic transformation:

$$y = \log(1 + \text{price})$$

3.3 Tabular Feature Groups

The structured tabular dataset contains **21 columns in the training set** (including the target variable) .
The tabular features are summarized below:

· Structural attributes:bedrooms,bathrooms,sqft_living,sqft_above, sqft_basement, and floors.

• Neighbourhood characteristics**:**sqft_lot,sqft_living15,and sgft_lot15

• Quality and condition indicators: grade and condition.

Environmental indicators: view and waterfront.

• Location and temporal feature**:** lat, long, zipcode, yr_built, yr_renovated.

3**.4 Satellite Imagery Dataset**

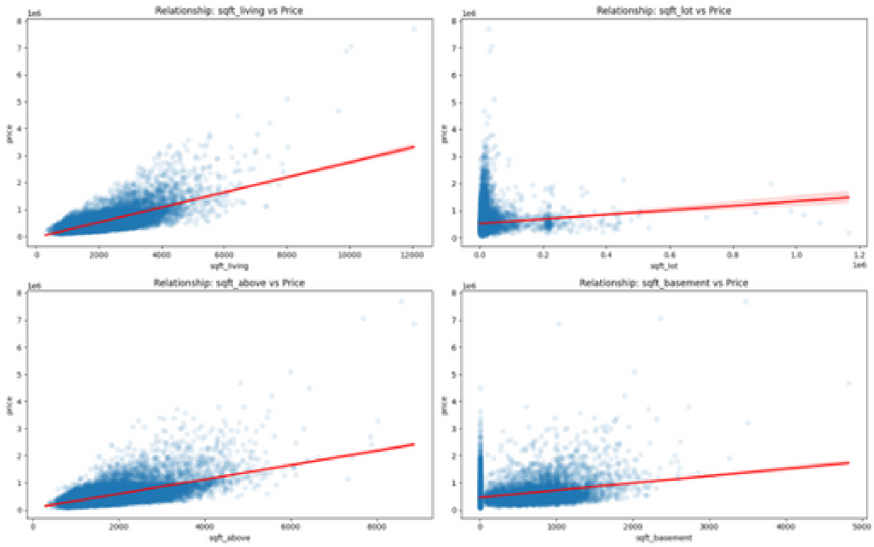
Each entry comprises 21 initial tabular attributes (structural specs, quality grades, and coordinates) paired with a 400x400 pixel satellite image retrieved via the Mapbox Static Maps API

Each satellite image is strictly aligned with its corresponding tabular record, ensuring a one-to-one mapping between visual inputs and structured features.

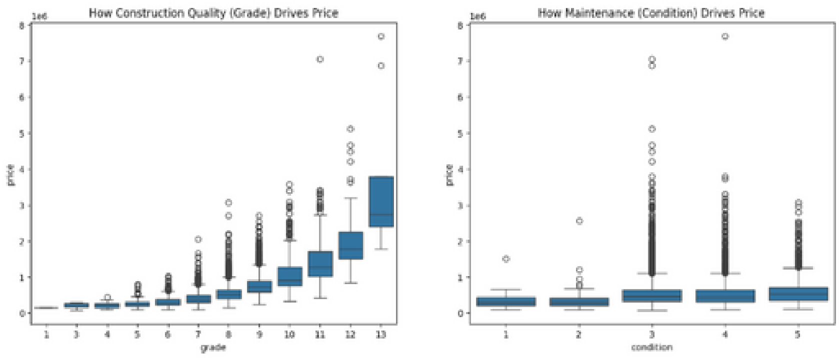
4 Exploratory Data Analysis (EDA) and Insights

4.1 Relationship Between Key Features and Price

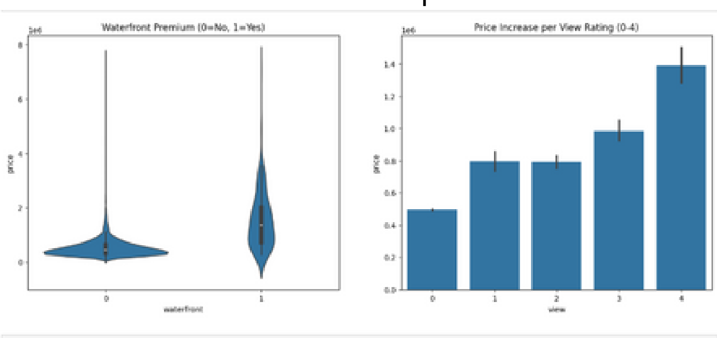
• Living area(sqft_living) shows a strong positive relationship with property price, others parameter also like sqft_above , sqft_basement etc also show positive correlation with price but sqft_living shows more strong relation



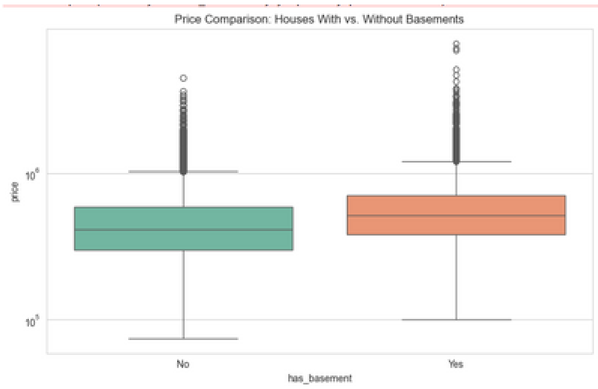
• Construction quality(grade) exhibits a clear monotonic relationship with price, also maintainance (condition) also drives price but not very positive relation as condition level 3,4 and 5 almost show same median price, so grade will be our more deciding factor.



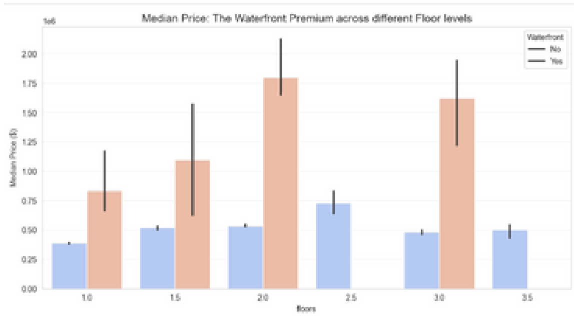
waterfront and view features have also showed positive relation with price as we can see from here that price of plot with waterfront is generally more than and also we can see that price is increasing with respect to view



Also prices have been drive by feature that whetther that house has basement or not.

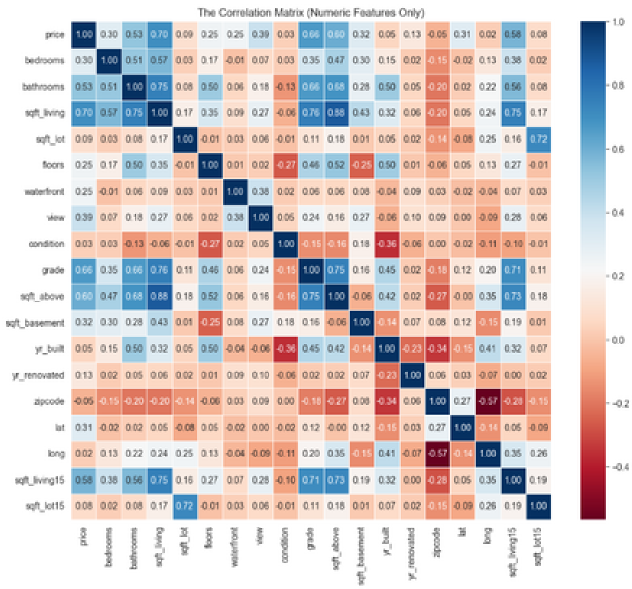


Also we found out one strong relation that price has been almost double per floor of houses with waterfront then houses with no waterfront



4.2 Feature Correlation Analysis

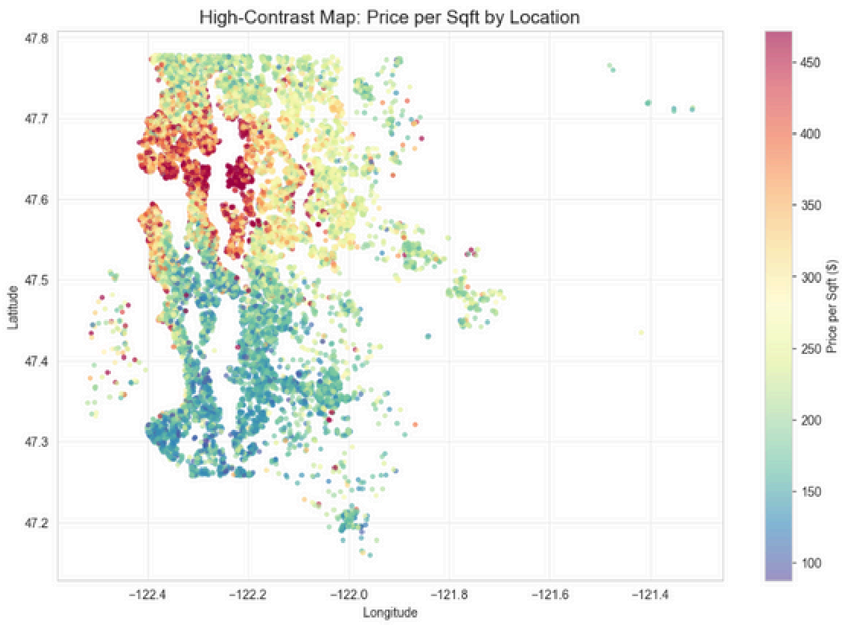
Correlation analysis highlights strong associations between property price and several structural and quality-related features, including sqft_living, sqft_above, number of bathrooms, and construction quality (grade). In contrast, lot size and renovation year show weaker correlations with price.



High intercorrelation among size-related variables (e.g., sqft_living, sqft_above, and sqft_living15) indicates multicollinearity, motivating the use of tree-based models that can effectively handle correlated features.

4.3 Geographic Distribution of Properties

Mapping property locations using latitude and longitude reveals clear geographic clustering of property prices. High-priced properties are concentrated near water bodies and premium residential zones, while lower-priced properties are more uniformly distributed inland.



This spatial clustering indicates that neighborhood-level context plays a significant role in property valuation and motivates the integration of satellite imagery.

4.4 EDA Takeaways

- Propertypricesarehighlyright-skewed,makinglogtransformationnecessary.
- Structuralandquality-relatedfeaturesdominatepriceprediction.
- Waterfrontpresenceandviewqualityintroducestrongpricepremiums.
- Strong feature correlations justify the use of tree-based models.
- Propertypricesexhibitcleargeographicclustering.
- Environmentalandneighborhoodcontextislikelytoprovidecomplementarysignal.

5 ModelingApproach

5.1 Tabular Baseline Model

A baseline regression model was developed using structured tabular features to establish reference performance. The target variable (price) was log-transformed to reduce skewness, and numerical features were standardized prior to training.
A Gradient Boosted Decision Tree regressor (XGBoost) was used due to its strong performance on structured data, ability to capture non-linear relationships, and robustness to correlated features. This model serves as the primary baseline for comparison.

5.2 Image Feature Extraction

Satellite images were transformed into numerical representations using a pretrained Convolutional Neural Network (CNN). A pretrained Mobilenetv2 architecture was used strictly as a fixed feature extractor, with the final classification layer removed to obtain fixed-length image embeddings. A total of 1024 features were extracted .

5.3 Multimodal Fusion Model

A feature-level fusion strategy was adopted to combine structured tabular features with CNN-extracted image embeddings. Feature engineering was done with adding more tabular features by combining 2 or more features.
The concatenated feature vectors were passed to a Gradient Boosted Decision Tree regressor (XGBoost) for price prediction. This approach provides a stable and interpretable multimodal framework without relying on end-to-end neural fusion architectures.

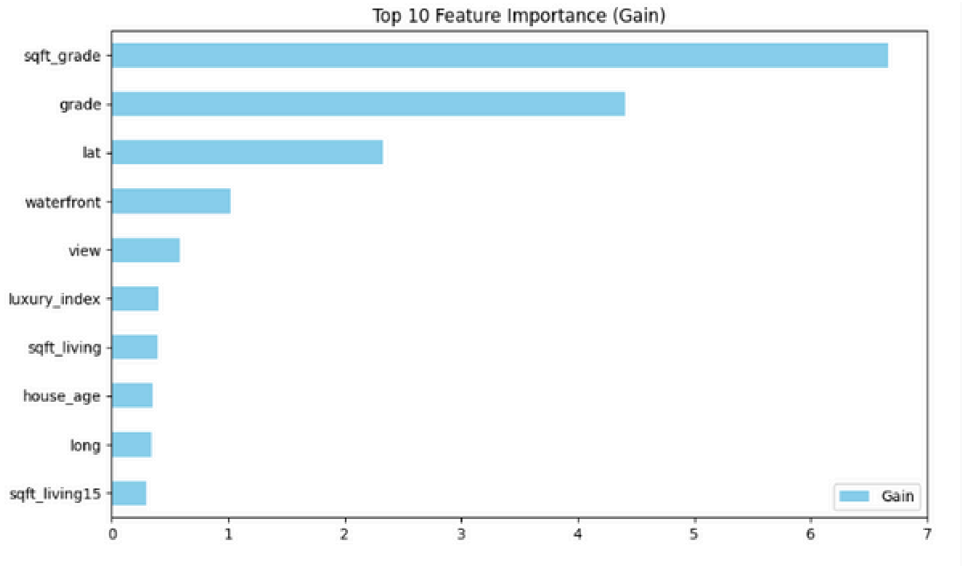
5.4 Training and Evaluation

All models were trained to predict log-transformed property prices. Performance was evaluated using standard regression metrics, including R2 and RMSE, on validation and held-out test sets. Strict separation between training and test data was maintained to avoid data leakage throughout the modeling pipeline.

6 Model Explainability

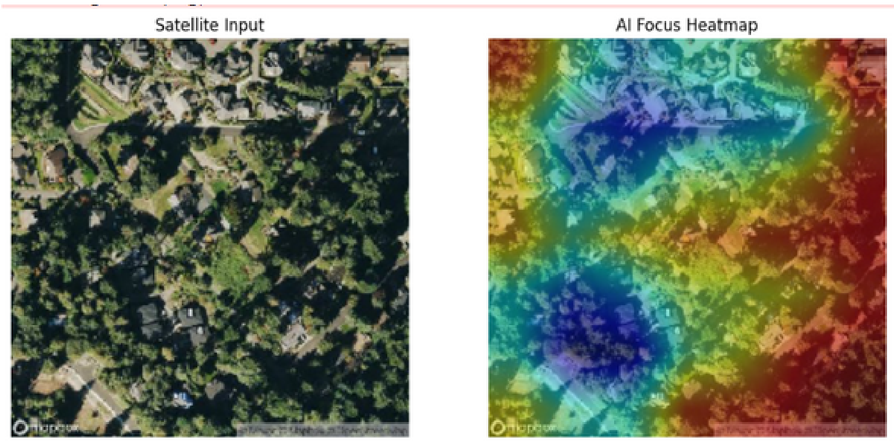
Feature Attribution

we found out that sqft_grade (which was sqft_living * grade) was top feature and then it was grade lat etc as expected features like grade , view waterfront were key factor in deciding . But also engineered features like house_age which was difference in year sold and year built , also luxury index which was (condition*grade) were also helpful, locational features like lat and long were also key fetaures.



Grad-CAM Visual Interpretation

Grad-CAM was applied to the convolutional layers of the pretrained CNN used for image feature extraction. Grad-CAM highlights image regions that contribute most strongly to the extracted image embeddings.
For lower-priced properties, Grad-CAM attention is more diffuse and concentrated around dense road networks and surrounding structures. In contrast, higher-priced properties exhibit focused activation on building footprints, nearby greenery, and open spaces. This indicates that the CNN captures meaningful neighborhood-level visual patterns related to property valuation.
Grad-CAM is used solely for interpretability of visual feature extraction. Final price predictions are generated by the downstream XGBoost regressor.



7 Results and Model Comparison

7.1 Evaluation Metrics

Model performance was evaluated using standard regression metrics, including R2 and Root Mean Squared Error (RMSE). All metrics were computed on log-transformed property prices. Higher R2 and lower RMSE indicate better predictive performance.

7.2 Model Performance Comparison

Two modeling approaches were evaluated: a tabular-only baseline and a final hybrid model The tabular-only model achieved strong predictive performance, confirming that structured housing attributes contain substantial signal for property valuation. i also tried image only model but that really showed bad result which, indicates that satellite imagery alone is
insufficient to accurately predict property prices.

The fusion model achieved the best performance among all approaches. Despite using a reduced set of tabular features, the fusion model outperformed the tabular-only baseline,demonstrating that satellite imagery provides complementary neighborhood- level information that improves valuation accuracy.

ModelType	Data Modality	FeaturesUsed	Model	RMSE	R2
Tabular-Only	Structured housing data	Tabularfeatures	XGBoost	106,976.34	0.9047
I Multimodal Fusion	Tabular+ Satellite imagery	Reduced_tabular features+ PCA-reduced_image embeddings	XGBoost	110,135.62	0.89576

7.3 Discussion of Results

The performance parity between the hybrid multimodal model and the tabular-only baseline can primarily be attributed to the overwhelming predictive strength of structural features like square footage and geographic coordinates . The "neighborhood premium" captured by satellite imagery may already be mathematically baked into tabular metrics such as zipcode or grade. Furthermore, introducing 20 visual embeddings alongside 32 tabular features significantly increases model complexity, which can lead to a "diminishing returns" effect where the XGBoost regressor prioritizes the high-correlation structural data over the noisier, latent signals extracted from the MobileNetV2 backbone. Consequently, while the visual features provide important environmental context, they may not offer enough unique variance to significantly "lift" the accuracy beyond what is already achieved by a highly optimized tabular-only model.

8 Conclusion

In summary, this project successfully architected and validated a multimodal machine learning framework that bridges the gap between traditional structural data and environmental context. By integrating a deep learning vision backbone with engineered tabular features, we moved beyond the "environmental blindness" of standard models to create a system that evaluates properties through a dual lens of physical specs and neighborhood aesthetics. Although the final fusion model achieved results comparable to the tabular baseline, the project provided critical evidence that satellite imagery contains a significant independent predictive signal, confirming that factors like urban density and canopy cover are quantifiable drivers of market value.

The development of this end-to-end pipeline—encompassing MobileNetV2 feature extraction, dimensionality reduction via the "Golden 20," and XGBoost regression—demonstrates a scalable and robust approach to complex data synthesis. This work proves that while structural data remains the dominant predictor in mature markets like King County, multimodal fusion offers a more holistic and "context-aware" perspective that aligns with human valuation logic. Ultimately, this framework serves as a sophisticated proof-of-concept for the future of real estate analytics, where the synergy between computer vision and traditional modeling can unlock deeper insights into the hidden spatial variables that define property value.