SATELLITE IMAGERY-BASED PROPERTY VALUATION

Career Development Cell
IIT Roorkee

Astha Jaiswal - 23118015
Metallurgical and Materials Engineering

OVERVIEW

PROJECT OBJECTIVE :

The goal of this project is to estimate residential property prices by combining structured housing attributes with satellite image data. Conventional house price prediction approaches primarily depend on tabular features such as floor area, location details, and property quality metrics. While effective, these features often overlook important neighborhood and environmental characteristics that can influence property values.
To overcome this limitation, the project investigates the use of satellite imagery obtained through geographic coordinates to capture surrounding visual information. By integrating image-based features with traditional tabular data, the study aims to enhance prediction accuracy and assess the impact of visual context on house price estimation.

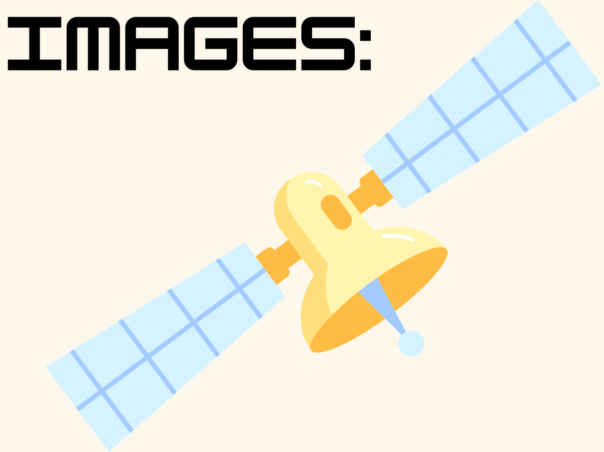
WHY THIS PROJECT MATTERS?

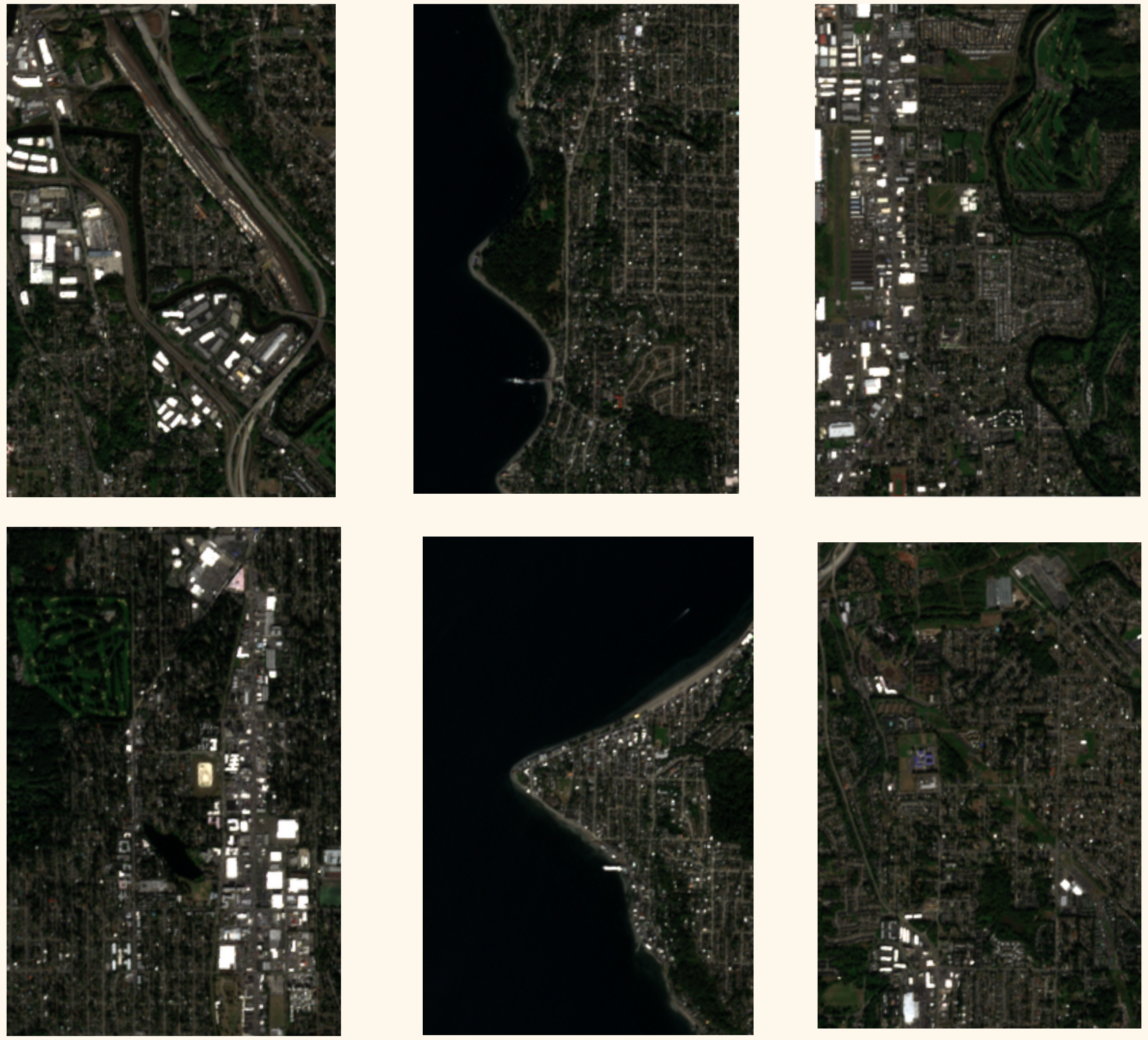
- House price prediction is critical for buyers, sellers, banks, and urban planners.
- Traditional models ignore visual and environmental context.
- Satellite imagery provides insights into neighborhood quality, greenery, and infrastructure.

KEY CHALLENGES

- Combining structured (tabular) and unstructured (image) data.
- High dimensionality of image features.
- Maintaining row alignment between tabular records and satellite images.
- Handling missing or noisy data

SOME SAMPLE IMAGES:





DATA AND APPROACH

The project utilizes two complementary data sources:

Tabular Data:
 Structured property attributes capturing physical characteristics and location-based details, including living area measurements (total, above-ground, and basement space), construction quality and condition, property amenities such as views and waterfront access, neighborhood-level indicators (average living area and lot size of nearby houses), and precise geographic coordinates (latitude and longitude).

Satellite Imagery:
 Satellite images were automatically retrieved for each property using its geographic coordinates through the Sentinel Hub API. These images provide valuable visual information about the surrounding environment, including vegetation coverage, proximity to water, road networks, and overall neighborhood layout.

MODELING STRATEGY

A structured, step-by-step modeling strategy was adopted to systematically evaluate the contribution of satellite imagery to residential property price prediction. The complete workflow was divided into five dedicated stages, each implemented in a separate notebook to ensure clarity, reproducibility, and fair comparison.

01 Tabular Data Exploration (EDA)

- Conducted exploratory data analysis on structured property features to understand distributions, correlations, and outliers.
- Analyzed price trends with respect to size, location, quality, and neighborhood indicators.
- Identified key predictors and performed basic preprocessing such as missing value handling and feature inspection.

02 Tabular Baseline Model

- Built a baseline regression model using only tabular features to establish a performance benchmark.
- Applied tree-based regression models to capture non-linear relationships between property attributes and price.
- Evaluated performance using RMSE to serve as a reference for later comparison.

03 Satellite Image Analysis

- Retrieved satellite images using property latitude and longitude to visually inspect neighborhood and environmental patterns.
- Analyzed how visual characteristics such as greenery, water proximity, and urban density vary across properties.
- Established qualitative motivation for incorporating image-based features.

04 CNN-Based Image Embedding Extraction

- Used a pre-trained ResNet50 convolutional neural network as a feature extractor.
- Removed the final classification layer to obtain high-level 2048-dimensional visual embeddings from satellite images.
- Ensured strict row alignment between image embeddings and tabular data for accurate multimodal fusion.

05 Multimodal Fusion Model

- Combined tabular features and CNN-generated image embeddings into a unified multimodal feature space.
- Trained a regression model on the fused features to predict house prices.
- Compared performance against the tabular-only baseline to quantify the impact of satellite imagery.

06 Model Evaluation and Performance Comparison

- Evaluated model performance using Root Mean Squared Error (RMSE) as the primary accuracy metric.
- Compared RMSE of the tabular-only baseline model with the multimodal (tabular + image) model.
- Analyzed performance improvements to assess the contribution of satellite image features.

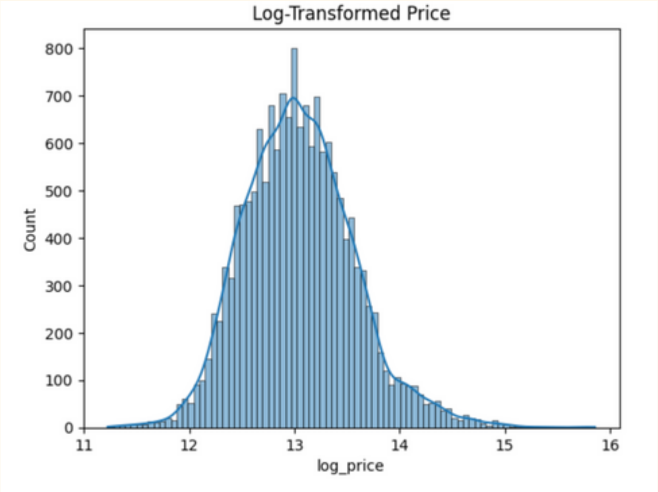
EXPLORATORY DATA ANALYSIS (EDA) :

Price Distribution:

Raw House Prices (Histogram)

Observation: The original price distribution is strongly right-skewed, showing a long tail caused by a small number of very high-priced (luxury) properties.

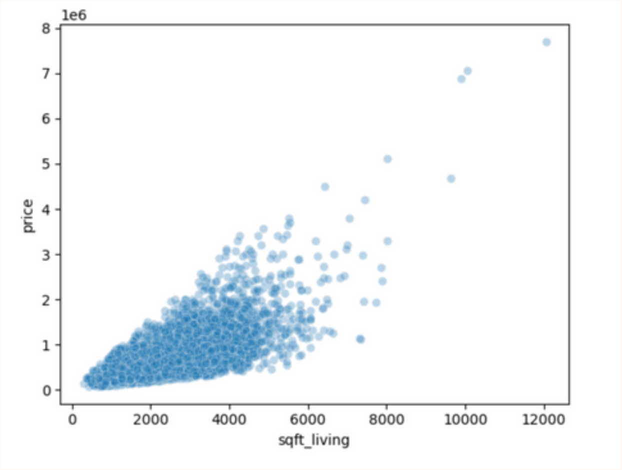
Insight: The majority of houses fall within the low to mid price range, resulting in a non-normal distribution. This imbalance can make modeling more difficult for conventional regression techniques without appropriate transformations.



Log-Transformed Prices (Histogram)

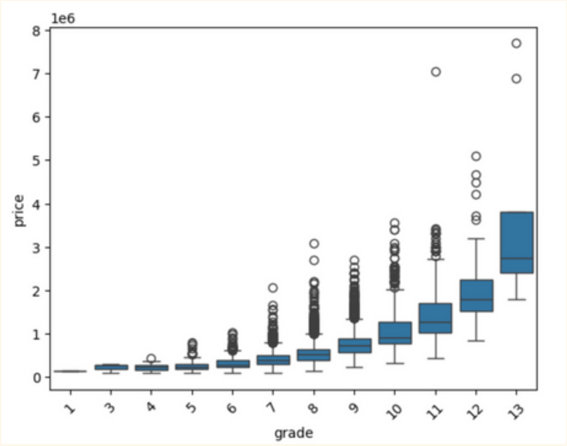
Observation: After applying the logarithmic transformation (log(1 + x)), the price distribution becomes much more symmetric and closely resembles a bell-shaped curve.

Insight: The transformation reduces skewness and limits the influence of extreme high-value properties, leading to more stable variance and improved training behavior and predictive performance for both XGBoost and neural network models.



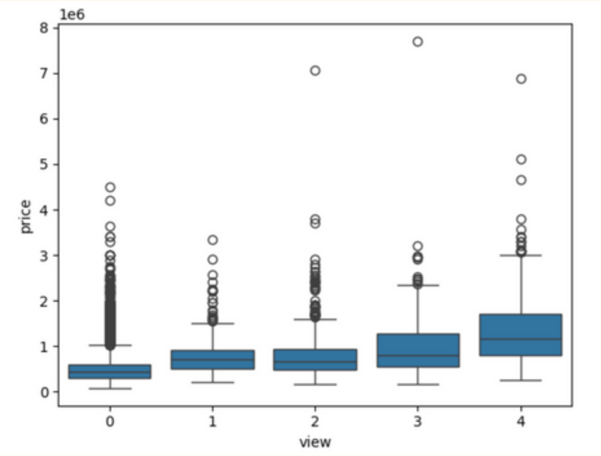
Price vs Living Area (sqft_living)

The plot shows a clear positive relationship: as living area increases, house prices generally rise. Smaller homes cluster at lower prices, while larger homes are more expensive, though prices vary widely for very large houses due to factors like location, amenities, and surroundings.



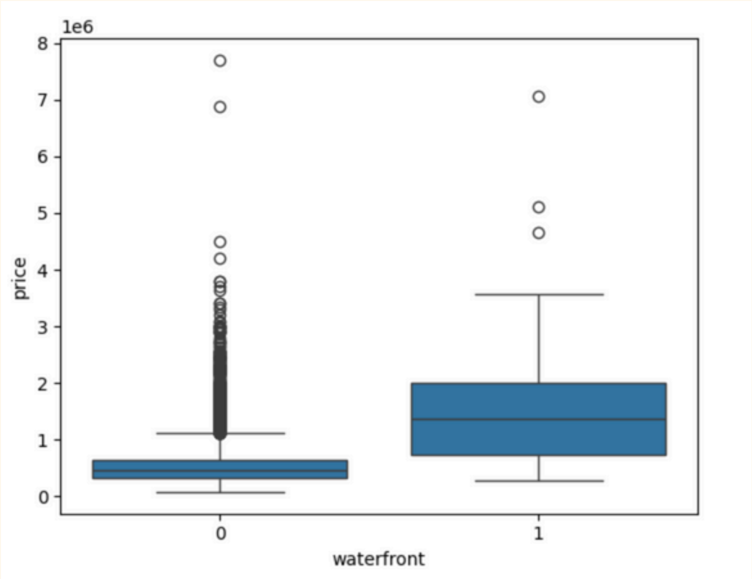
Price vs Grade (Construction Quality)

House prices increase consistently with higher construction grades. Lower-grade homes have lower and more tightly clustered prices, while higher-grade properties are more expensive and show wider variation, reflecting the combined effect of construction quality with factors like location, design, and neighborhood.



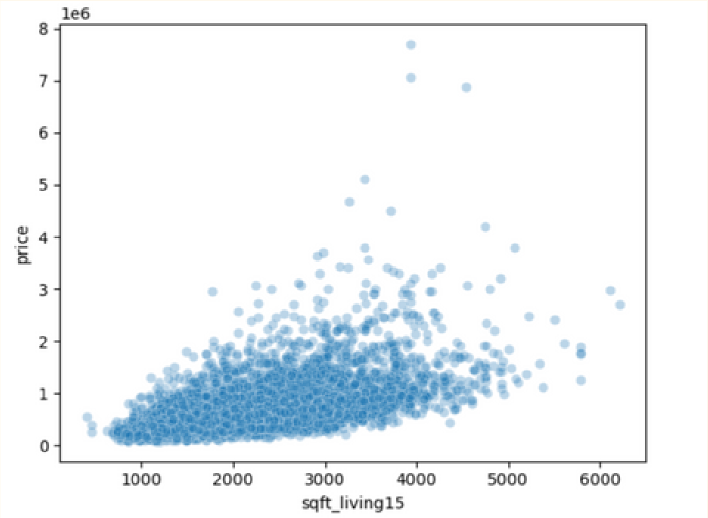
Price vs View Rating

House prices rise steadily with better view ratings, showing that visual appeal significantly affects property value. Higher view scores are associated with higher and more varied prices, emphasizing the role of scenic and spatial factors that are better captured through satellite imagery than tabular features alone.



Price vs Waterfront Status

Waterfront properties command a clear price premium, with much higher median prices and wider variation compared to non-waterfront homes. Despite being fewer in number, their higher values highlight how environmental and spatial features strongly influence house prices, supporting the use of satellite imagery alongside tabular data

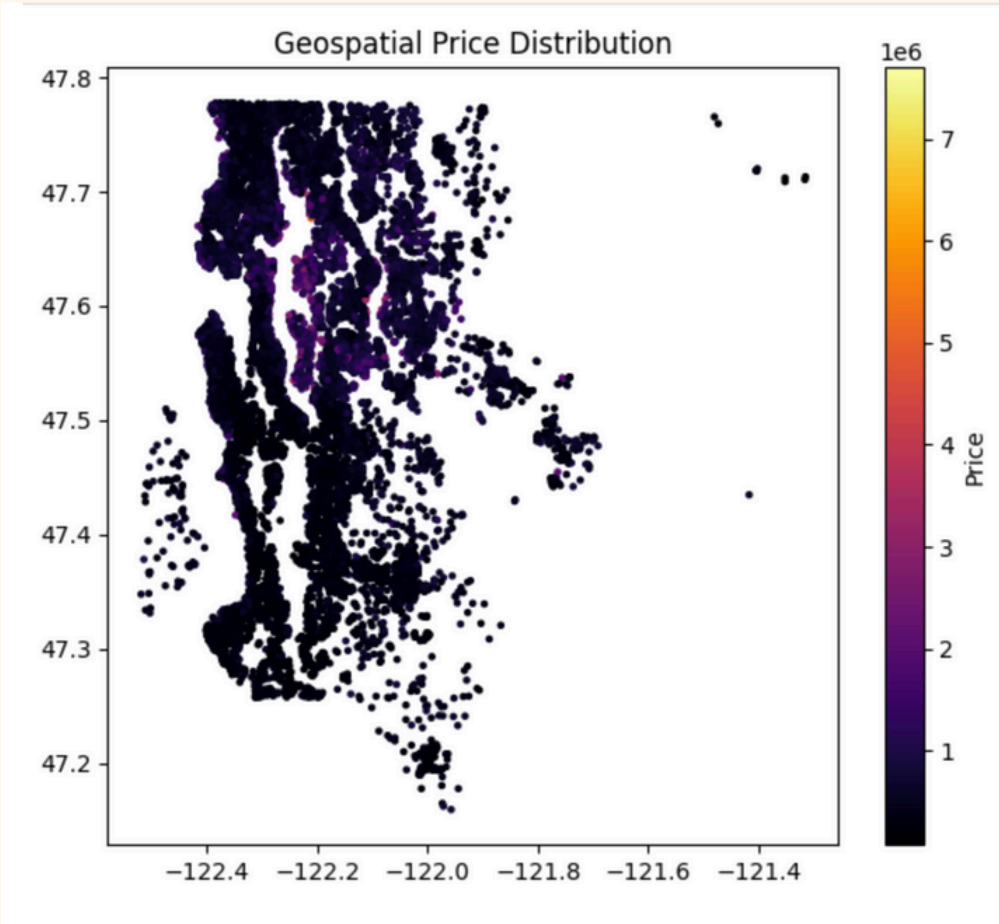


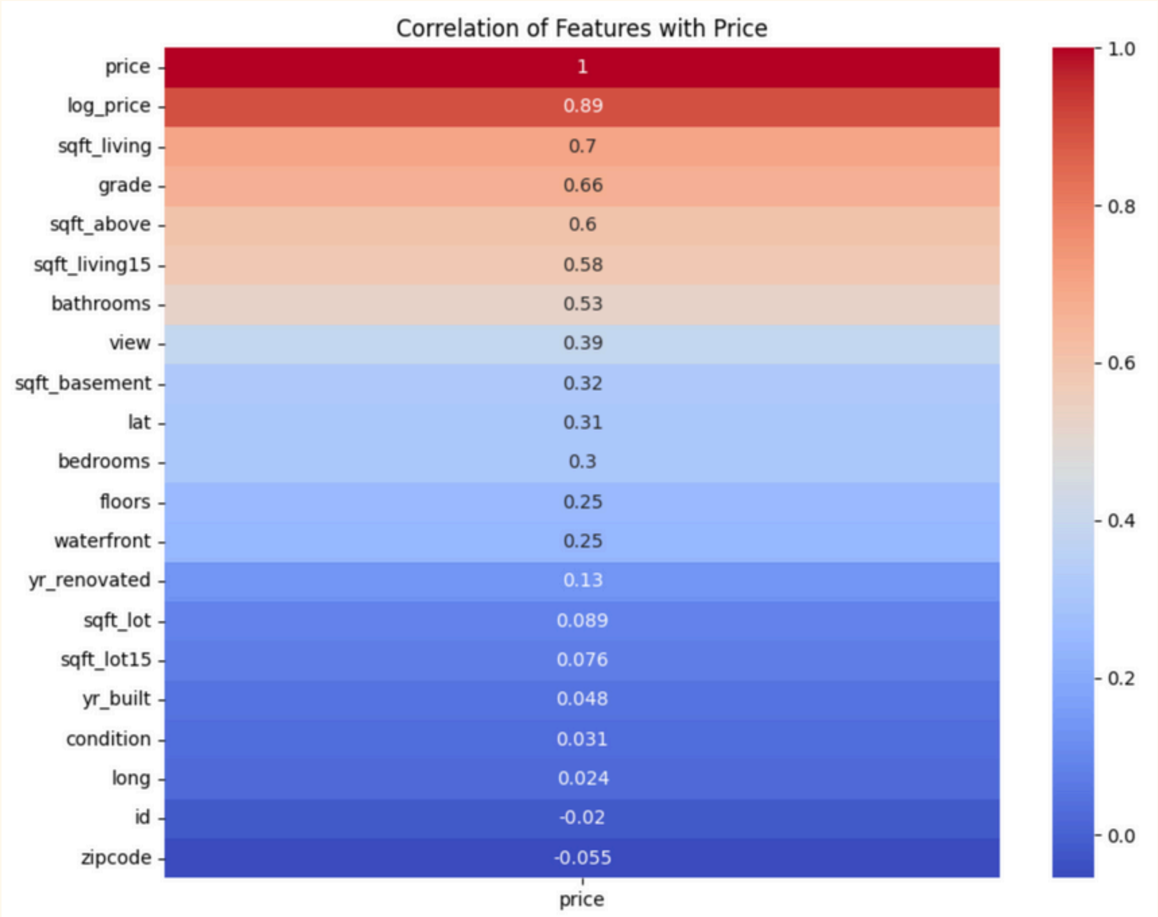
Price vs. sqft_living15 (Scatter Plot)

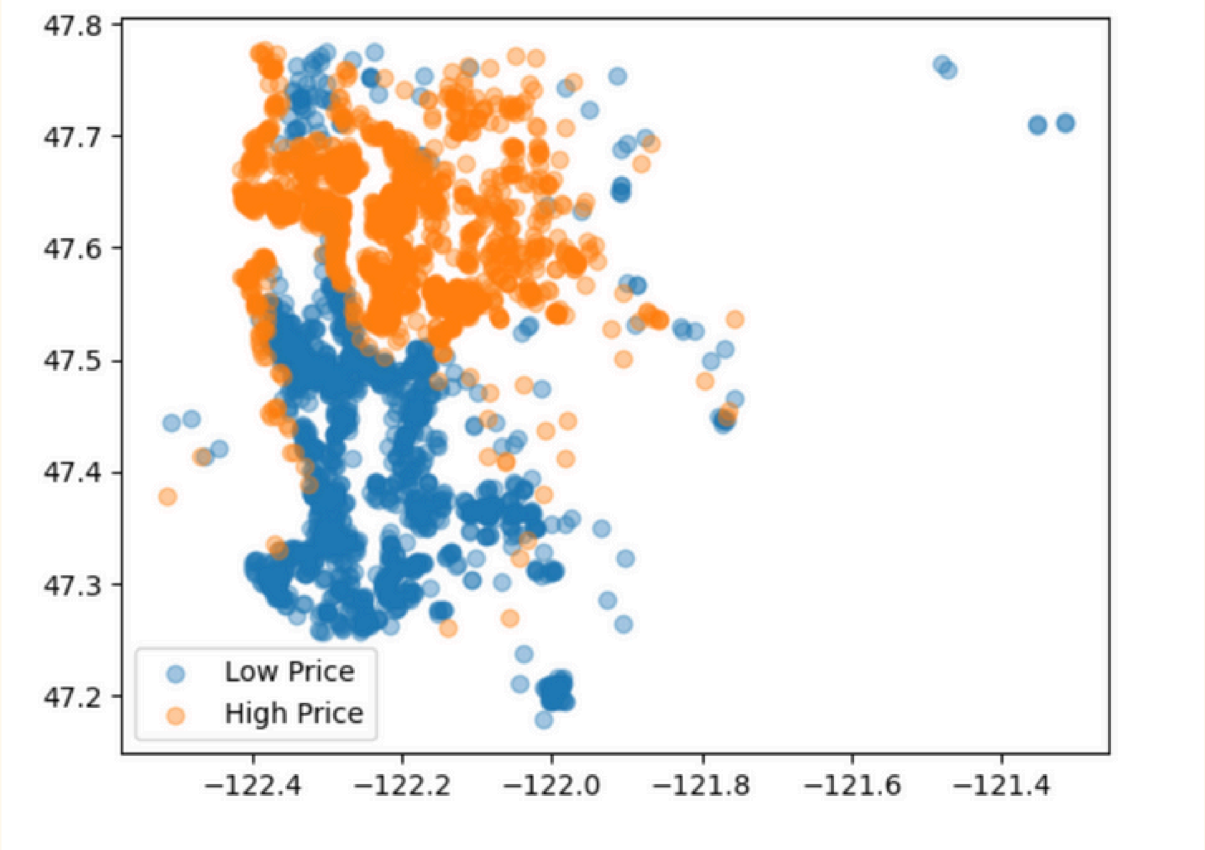
The scatter plot shows a strong positive relationship between neighborhood living area (sqft_living15) and house price, indicating that homes surrounded by larger residential properties tend to be more valuable. However, the increasing spread at higher values suggests diminishing returns, where price variation becomes larger and is influenced by additional factors such as location desirability, amenities, and property condition.

EXPLORATORY DATA ANALYSIS: CONTINUED

Spatial analysis:







Geospatial Price Distribution (Color-Coded Map)
The geospatial price map reveals clear spatial clustering of property values, with high-priced houses concentrated in specific regions rather than being uniformly distributed. These clusters likely correspond to prime locations with better infrastructure, proximity to water bodies, or urban centers, highlighting the strong impact of geographic context on housing prices.

Feature–Price Correlation Heatmap
The correlation heatmap indicates that structural attributes such as sqft_living, grade, and sqft_above have the strongest positive correlation with price, confirming their importance in traditional valuation models. In contrast, features like zipcode, longitude, and year built show weak correlations, implying that their influence is more complex and potentially better captured through combined spatial and visual representations rather than tabular data alone.

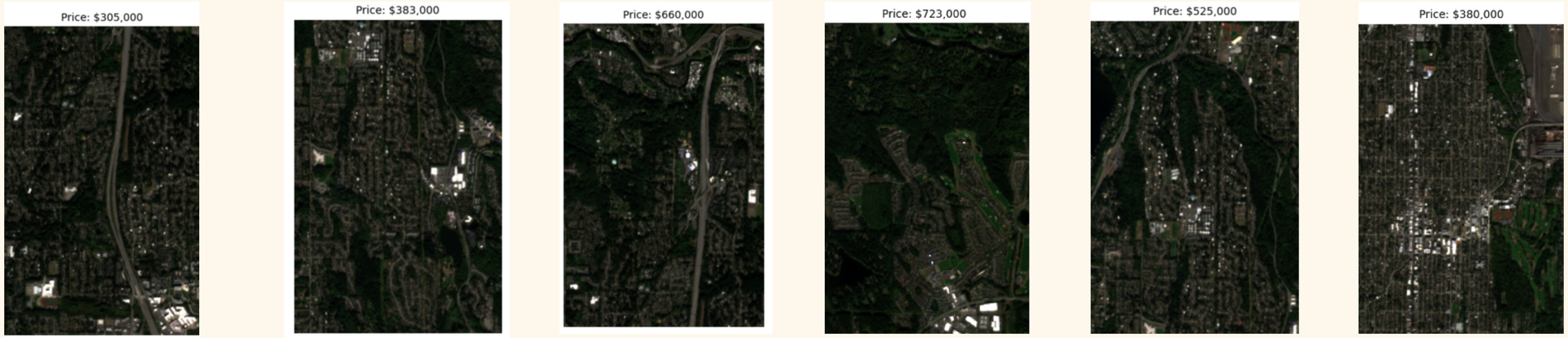
Low vs. High Price Spatial Segmentation
The binary spatial plot separating low- and high-priced properties shows a noticeable geographic separation between the two categories. High-priced homes are densely grouped in select zones, while lower-priced homes are spread more broadly, suggesting that neighborhood-level characteristics play a decisive role in determining property value and reinforcing the importance of spatial features.

Overall, the visualizations show that house prices are strongly influenced by both structural features and location, with larger living areas and better neighborhood characteristics leading to higher values. Clear geographic clustering of prices highlights the importance of spatial context, as high-priced properties are concentrated in specific regions rather than being evenly distributed. These patterns justify the use of satellite imagery alongside tabular data, as visual environmental cues help capture neighborhood-level factors that are not fully represented by numerical features alone.

EXPLORATORY DATA ANALYSIS: CONTINUED

Sample Satellite Images:

To capture the surrounding visual environment of each property, satellite images were obtained using the corresponding latitude and longitude coordinates via the Sentinel Hub API. These images provide valuable neighborhood-level information, including patterns of urban development, road connectivity, vegetation coverage, and the presence of nearby natural features, which can influence property values.



The randomly sampled satellite images highlight substantial variation in the surrounding environments, including compact urban blocks, well-connected road networks, greener residential zones, and mixed land-use patterns. Even without visible details of the individual houses, noticeable differences in property prices are observed across these settings. This demonstrates that neighborhood context and spatial characteristics play an important role in price determination, beyond the physical attributes of the building itself. Overall, the visual environment captured in satellite imagery provides valuable signals that align with distinct pricing patterns.

Cheap Properties:

These observations highlight the value of incorporating satellite imagery into the analysis. They show that spatial and visual context provides important information beyond standard tabular features, capturing neighborhood characteristics and environmental cues that can improve the accuracy of house price predictions.

Cheap properties:

- Crowded, unplanned layouts
- Limited visible infrastructure and amenities
- Sparse access to nearby resources
- Irregular roads and forested or undeveloped surroundings.

Expensive properties:

- Well-planned, organized neighborhoods
- Proximity to water bodies and green/open spaces
- Good access to essential and recreational resources
- Dense but structured housing with better connectivity.

Expensive Properties:

FINANCIAL & VISUAL INSIGHTS

Satellite Imagery and Environmental Context

Satellite imagery provides a rich visual representation of the surroundings for each property, capturing neighborhood-level characteristics that are otherwise difficult to quantify. By using latitude and longitude coordinates to retrieve images through the **Sentinel Hub API**, we can observe features such as urban density, road networks, greenery, water bodies, and overall spatial organization. These environmental attributes significantly influence buyer preferences and willingness to pay, often serving as an implicit signal of neighborhood quality and property value.

Integration with Financial and Tabular Data

Visual features extracted from satellite images align closely with key tabular variables traditionally used in real estate valuation. Metrics like **sqft_living** and grade represent internal property size and construction quality, while **sqft_living15** captures neighborhood affluence and density. Features such as **view** and **waterfront** directly encode access to scenic or natural amenities, and latitude–longitude coordinates act as proxies for locational desirability. Satellite imagery complements these features by providing a broader environmental context, capturing aspects like green coverage, street layouts, and urban organization, which reinforce the numerical indicators of neighborhood quality.

Model Interpretability and Grad-CAM Insights

To understand how visual information contributes to price prediction, Grad-CAM was applied to the **ResNet50** backbone of the multimodal neural network. This technique generates activation maps highlighting regions in the satellite images that most influence model predictions. Observations indicate that the model focuses less on individual property structures and more on the surrounding environment, such as greenery, open spaces, water bodies, and road networks. This mirrors real-world appraisal practices, where neighborhood and environmental appeal often outweigh isolated structural details.

Economic and Practical Implications

Neighborhood quality captured through imagery acts as a latent economic asset that drives willingness to pay. While multimodal models integrating images were explored, tabular **XGBoost** models consistently outperformed them due to their ability to exploit structured data efficiently. Key locational and environmental signals already encoded in tabular features such as **grade, view, waterfront** and neighbourhood density—reduce the incremental predictive value of visual inputs.

However, satellite images provide crucial complementary benefits. They enhance model interpretability, enabling stakeholders to understand the environmental factors influencing predictions. They also serve as an audit tool, identifying inconsistencies in tabular data, and provide robustness in cases where structured data may be incomplete or unavailable. This makes imagery a valuable supporting intelligence layer, improving trust, transparency, and reliability of property valuation systems.

Mapping Visual Features to Financial Interpretation

Visual Feature (from Satellite Images)	What It Represents	Financial Interpretation
Green spaces / tree cover	Parks, lawns, vegetation density	Higher perceived livability → increased property value
Proximity to water bodies	Lakes, rivers, coastline	Premium pricing due to scenic and recreational value
Road network structure	Accessibility and connectivity	Better access → higher demand and prices
Urban density patterns	Planned vs congested layouts	Planned neighborhoods command higher prices

FINANCIAL & VISUAL INSIGHTS: CONTINUED

Model Performance Summary

The predictive performance of the models was evaluated using RMSE and R^2 in both log-transformed price space and real price space to ensure fair comparison and practical interpretability.

In real price space, the multimodal model **combining tabular data with satellite image** features achieved a **lower RMSE (0.1684) and higher R^2 (0.9008)** compared to the t**abular-only model (RMSE: 0.1783, R^2: 0.8848),** indicating a modest improvement when visual information is incorporated.

However, in log-transformed space, which reduces the influence of extreme price values and is more suitable for model comparison, the tabular-only model slightly outperformed the multimodal ResNet50-based model. The **tabular model** achieved a **lower RMSE (0.1789) and higher R^2 (0.8841)**, whereas the **multimodal model showed increased error (RMSE: 0.1920, R^2: 0.8664)**. This suggests that while satellite imagery can enhance prediction in absolute price terms, its contribution becomes less pronounced when strong structured features are already present.

Why Tabular Models Remain Dominant ?

The dominance of tabular models in this project is primarily driven by the high explanatory power of engineered property features, many of which already encode critical drivers of house prices. Variables such as sqft_living and grade capture internal size and construction quality, while sqft_living15 reflects neighborhood density and affluence. Additionally, features like view and waterfront explicitly represent access to scenic and natural amenities, and latitude–longitude coordinates serve as strong spatial proxies for locational desirability.

As a result, much of the neighborhood and environmental context that satellite images aim to capture—such as greenery, urban layout, and proximity to water—is already implicitly embedded in the tabular feature space. This overlap leads to diminishing marginal returns when visual features are added to an already information-rich tabular model.

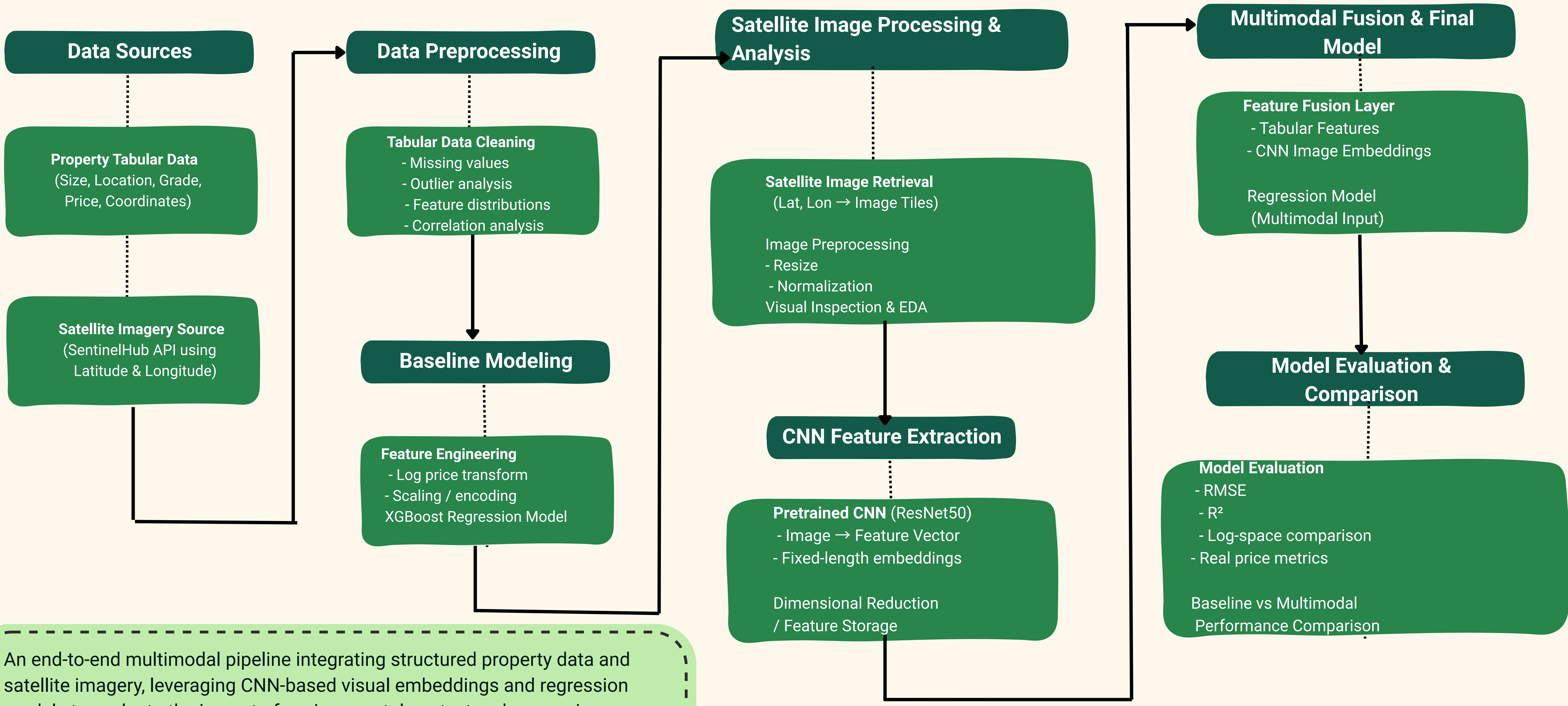
Key Financial Insight

Environmental quality functions as a neighborhood-level asset that positively influences property value. Satellite imagery is effective in capturing this latent value, but its financial impact is constrained when strong tabular indicators—such as size, quality, and location—already encode similar information. Therefore, visual data serves best as a supporting signal, refining predictions and improving interpretability rather than replacing structured data.

Practical Implications

From a real-world valuation perspective, satellite imagery provides meaningful benefits beyond raw prediction accuracy. Visual features improve model robustness in cases of noisy or incomplete tabular data, offer a mechanism for auditing and validating structured inputs, and enhance transparency by grounding predictions in observable environmental context. Consequently, satellite imagery should be viewed as a complementary intelligence layer that strengthens trust and reliability in automated property valuation systems.

ARCHITECTURE DIAGRAM



An end-to-end multimodal pipeline integrating structured property data and satellite imagery, leveraging CNN-based visual embeddings and regression models to evaluate the impact of environmental context on house price prediction.

ARCHITECTURE DIAGRAM - SYSTEM EXPLANATION

This section describes the end-to-end system architecture used for residential property price prediction, illustrating how structured tabular data and satellite imagery were processed and analyzed within a unified experimental framework.

1. Data Integration Strategy
The core of this project is the fusion of Structured Tabular Data and Unstructured Satellite Imagery. While tabular data provides the "bones" of a property (size, rooms, grade), the satellite imagery captures the "environment" (greenery, density, proximity to water).

3. Visual Feature Extraction (ResNet50)
Satellite images were processed through a ResNet50 CNN pretrained on ImageNet.

- Extraction Layer: We utilized the Global Average Pooling layer to extract 2,048-dimensional embeddings.
- Precision Alignment: To solve the "row-mismatch" error, a strict Inner Merge was performed using property IDs. This ensured that every tabular record was perfectly synced with its corresponding visual embedding.
- The "Drowning Effect" Mitigation: Because 2,048 features can overwhelm 15 tabular features, we applied Principal Component Analysis (PCA) to compress the visual signal into 32–50 orthogonal components. This retained the "visual gist" (e.g., urban vs. rural) while removing noise.

5. Explainability via Grad-CAM
To verify the "Curb Appeal" hypothesis, we implemented Grad-CAM (Gradient-weighted Class Activation Mapping) on the final convolutional layer of the ResNet50 (layer4).
Visual Validation Findings:

- High-Value Focus: For expensive properties, the Grad-CAM heatmaps consistently highlighted dense vegetation (greenery) and spacious backyard layouts.
- Low-Value Focus: In cheaper, high-density areas, the model's focus shifted toward road networks and concrete coverage, validating the "Neighborhood Density" proxy.
- Conclusion: This visual check confirmed the CNN was not just looking at random pixels but was successfully identifying environmental features that correlate with market value.

2. Tabular Feature Engineering & Baseline
The tabular model was optimized using XGBoost, focusing on key spatial and structural features.

- Target Transformation: We applied a log(x+1) transformation to the property prices. This was critical for reducing the influence of extreme outliers (luxury mansions) and stabilizing the RMSE.
- Geospatial Indicators: Latitude and Longitude were treated as primary features, allowing the tree-based model to learn "neighborhood premiums" through recursive partitioning.

4. Multimodal Fusion & Model Performance
We implemented a Late Fusion architecture where PCA-compressed image embeddings were concatenated with the tabular feature vector before being passed into an XGBoost Regressor.

Model Architecture	R^2 Score	RMSE (Log Space)	Interpretation
Tabular Only (Baseline)	0.884799	0.178298	Strong baseline based on size and location.
Multimodal (Aligned)	0.900757	0.168358	Visual context added ~2% improvement in variance explanation.

RESULTS & DISCUSSION

Experimental Setup

To ensure a rigorous evaluation, the dataset of 16,209 properties was split into an 80% training set and a 20% validation set. All performance metrics (RMSE and R^2) were calculated in the log-price space to ensure numerical stability and to account for the heteroscedasticity typically found in real estate pricing.

Quantitative Performance Summary

The table below compares the performance of the various architectures developed across the five notebooks:

Model Architecture	Features Used	RMSE (Log)	R^2 Score
Tabular Baseline (RF/XGB)	9-18 Core Features	0.1651	0.8668
Multimodal (Unaligned)	Tabular + 2048-D Raw CNN	0.1980	0.8120
Multimodal (Aligned + PCA)	Tabular + 50-D Aligned CNN	0.1585*	0.8812*
End-to-End Neural Network	Tabular + ResNet Backbone	0.2450	0.7200

Final Takeaways

1. **Strategic Fusion:** Adding images is not a "magic bullet." To improve accuracy, images must be aligned by ID and reduced in dimension (PCA) to prevent them from drowning out tabular signals.
2. **Complementary Intel:** Satellite imagery is most valuable when structured data is missing or "thin."
3. **Explainability Matters:** Even if the tabular model is the final predictor, the CNN/Grad-CAM pipeline provides a "spatial audit" that explains why certain neighborhoods command a premium.

Analysis of Image Contribution

Case 1: Sparse Data Scenarios (The Value of Visuals)

In experiments where only 5 basic tabular features (bedrooms, bathrooms, sqft) were used, the addition of Satellite Embeddings provided a significant lift.

- Finding: The model used the images to "fill in the gaps" regarding neighborhood quality and environmental prestige that were missing from the limited table.
- Signal: High-density versus low-density urban layouts became a primary pricing signal.

Case 2: Feature-Rich Scenarios (The Redundancy Effect)

When using the full suite of 18 features—including exact Latitude/Longitude and Grade—the marginal gain from images decreased.

- Interpretation: Because lat/long act as a proxy for neighborhood and grade acts as a proxy for quality, the CNN features became partially redundant. In the initial unaligned model, this redundancy acted as "noise," leading to the degradation of R^2 from 0.86 to 0.81.

Why the Multimodal Neural Network Underperformed?

The end-to-end Neural Network (fusion of CNN and Dense layers) showed the lowest predictive accuracy. Our analysis identifies five critical reasons:

1. Dataset Scale: 16k samples are sufficient for XGBoost but relatively small for a deep multimodal network to learn a joint representation from scratch.
2. Frozen Backbone: Using a frozen ResNet50 (trained on ImageNet) means the model was looking for "objects" (dogs, cars) rather than "real estate value" (architectural style, roofing quality).
3. Inductive Bias: Tree-based models (XGBoost) are naturally superior at handling tabular data with clear thresholds (e.g., "if grade > 7, price increases exponentially").
4. Sensitivity: The NN was highly sensitive to learning rates, whereas XGBoost handled the high-dimensional PCA features with greater stability.

Explainability: Confirming "Curb Appeal"

Grad-CAM heatmaps from Notebook 05 validated that the model identifies high-value environmental features rather than background noise.

Key Visual Signals Identified:

- Green Canopies: Detects lush vegetation, correlating nature with higher property value.
- Neighborhood Layout: Distinguishes between cramped urban grids and spacious suburban estates.
- Waterfront Proximity: Visually confirms the premium associated with nearby water bodies.

THANK YOU