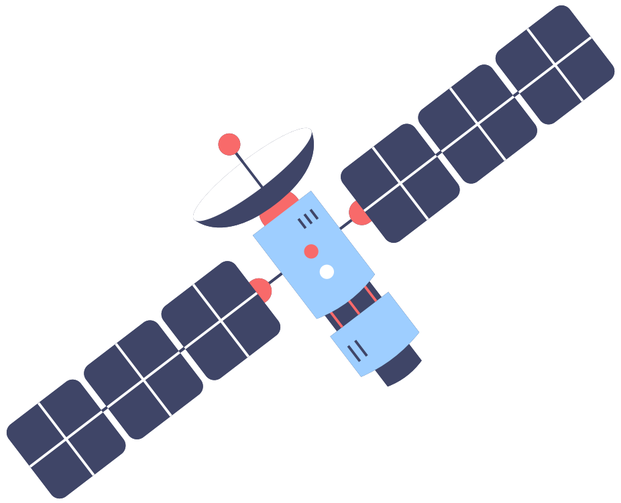
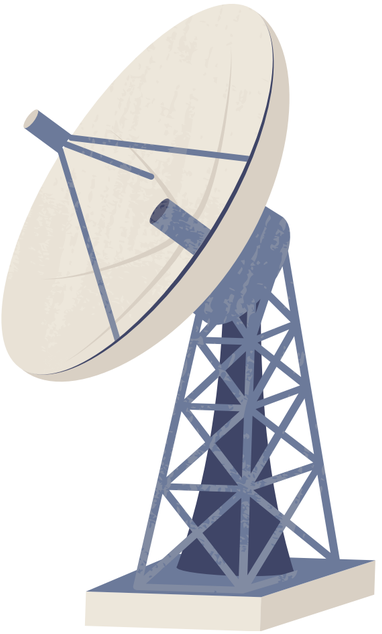


Satellite Imagery-Based Property Valuation

Career Development Cell
IIT Roorkee

Awani Soni 23322008
BS-MS Economics



Overview

Project Objective:

The objective of this project is to predict residential house prices using a combination of structured property data and satellite imagery. Traditional house price prediction models rely heavily on tabular features such as property size, location, and quality indicators. However, these features may not fully capture environmental and neighborhood-level information that influences property value.

To address this, this project explores whether satellite images, extracted using geographic coordinates, can provide additional visual context and improve price prediction performance when combined with tabular data.

Some Sample Images

Data and Approach:

The dataset consists of two primary data sources:

- **Tabular Data:** A comprehensive set of structural and environmental attributes including living area dimensions (sqft_living, above, and basement), construction quality and maintenance (grade and condition), property-specific amenities (view and waterfront status), and localized neighborhood metrics (sqft_living15/lot15) paired with precise geospatial coordinates (latitude and longitude).

- **Satellite Imagery:** Satellite images were programmatically fetched using latitude and longitude coordinates for each property via the Sentinel Hub API, capturing surrounding environmental context such as greenery, proximity to water bodies, road density, and overall urban structure.

Modeling Strategy

A structured, step-by-step modeling strategy was adopted to systematically evaluate the contribution of satellite imagery to property price prediction.

1 Baseline Tabular Modeling

The modeling process began with a tabular-only baseline, using the key features explicitly mentioned in the project guidelines: Bedrooms, Bathrooms, Sqft_living, Latitude, Longitude.
Multiple regression models were evaluated, and XGBoost emerged as the strongest baseline model based on RMSE and R^2 scores. This established a strong reference point against which multimodal approaches could be compared.

2 Multimodal Modeling with Image Embeddings

To incorporate visual information, satellite images were processed using a pre-trained ResNet50 convolutional neural network, where:

- The CNN was used as a fixed feature extractor.
- High-dimensional image embeddings were generated for each property.
- These embeddings were concatenated with the 5 key tabular features.

The resulting hybrid dataset (tabular + image embeddings) was used to train an XGBoost regression model. This experiment assessed whether satellite imagery adds predictive value beyond core numerical features.

3 Expanded Tabular Feature Modeling

After the dataset description was expanded, additional tabular features related to construction quality, neighborhood density, and visual context were incorporated: Grade, Sqft_living15, View, Waterfront.
Using correlation analysis and exploratory data analysis, a refined set of 9 informative tabular features was selected. A new feature-enriched tabular baseline was trained using XGBoost, resulting in improved predictive performance.

4 Multimodal Modeling with Expanded Features

Image embeddings extracted from ResNet50 were then combined with the expanded tabular feature set to form a second hybrid dataset. The same XGBoost framework was applied to ensure a fair comparison between:

- Tabular-only models
- Tabular + image-based models

This step helped evaluate whether satellite imagery still adds value when strong domain-specific tabular features are already present.

5 End-to-End Multimodal Neural Network

In addition to tree-based models, an end-to-end multimodal deep learning architecture was implemented:

- A pre-trained ResNet50 backbone processed satellite images.
- A Multi-Layer Perceptron (MLP) processed tabular features.
- Both feature streams were combined using late fusion.
- A joint regression head produced final price predictions.

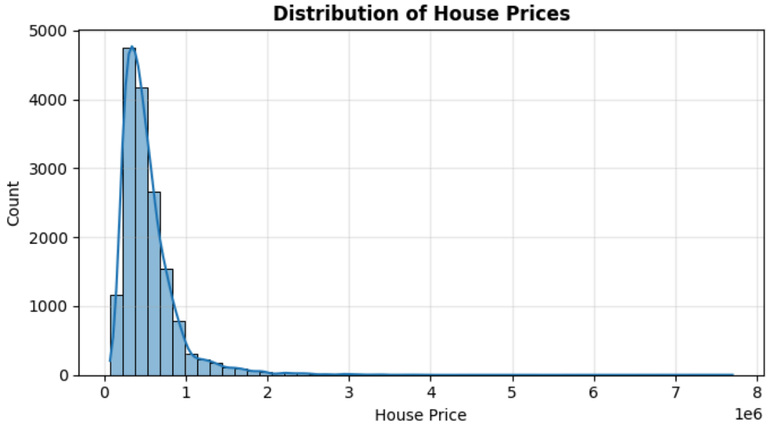
While this neural approach did not outperform XGBoost on tabular data, it provided a suitable framework for visual explainability.

6 Model Optimization and Explainability

- Optuna was used for systematic hyperparameter optimization of the XGBoost models.
- SHAP was applied to interpret feature importance in tabular models.
- Grad-CAM was used to visualize spatial regions of satellite images influencing the neural network's predictions.

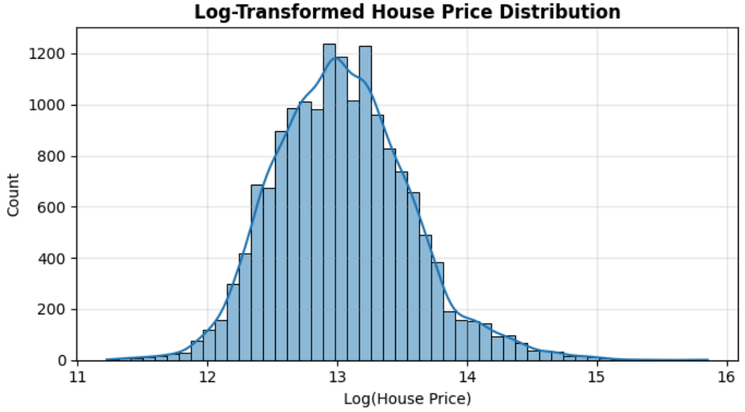
Exploratory Data Analysis

Price Distribution:



Raw House Prices (Histogram)

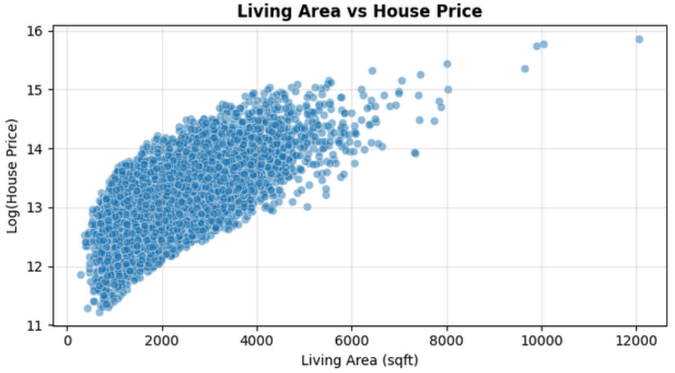
- **Observation:** The distribution of raw prices is heavily right-skewed, with a long tail representing high-value luxury properties.
- **Insight:** Most properties are concentrated in the lower to mid-price range, making the data non-normal and potentially challenging for standard regression models.



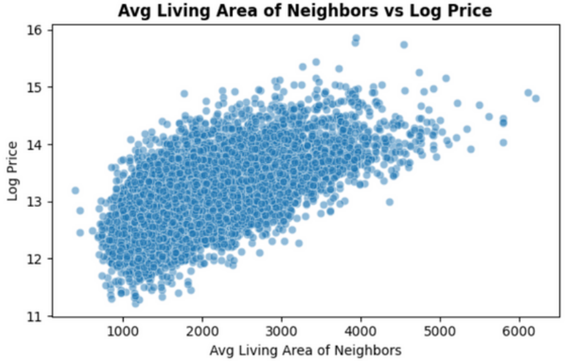
Log-Transformed Prices (Histogram)

- **Observation:** Applying a logarithmic transformation (log(1+x)) successfully normalized the distribution, resulting in a more symmetric, bell-shaped curve.
- **Insight:** This transformation stabilizes the variance and reduces the impact of extreme outliers, ensuring better convergence and performance for both the XGBoost and Neural Network models.

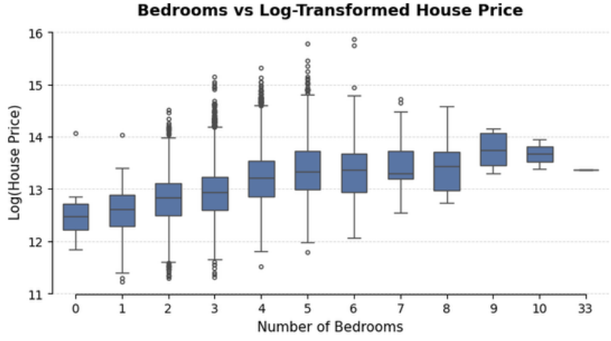
Relationship with key tabular features:



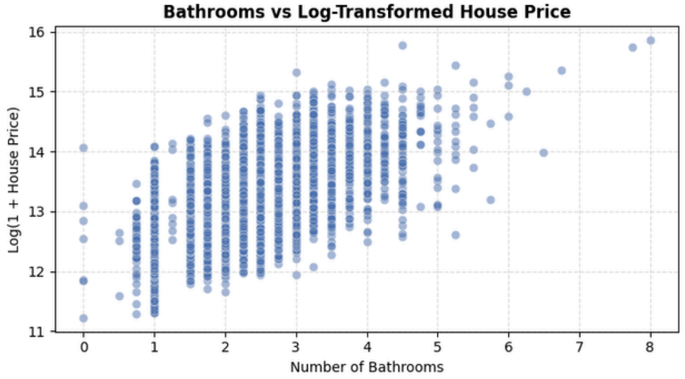
There is a strong positive correlation between living area and the log-transformed house price, though the variance in price increases as square footage grows.



The scatter plot shows a moderate positive correlation between the average living area of neighboring houses and the log price, indicating that property values tend to be higher in neighborhoods with larger homes.



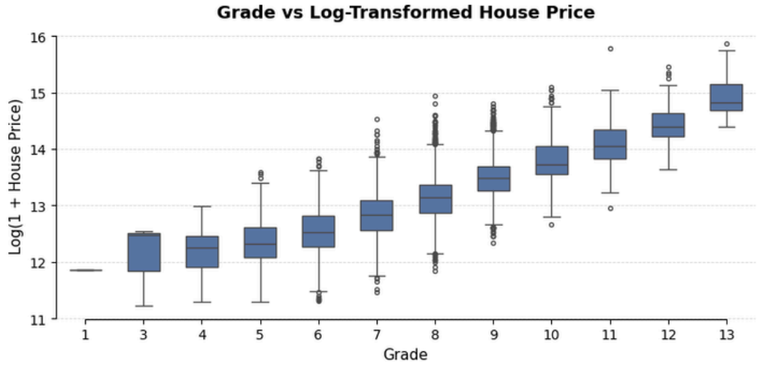
The box plot reveals that median house prices increase steadily with the number of bedrooms up to approximately five, after which the relationship plateaus and price variance increases significantly.



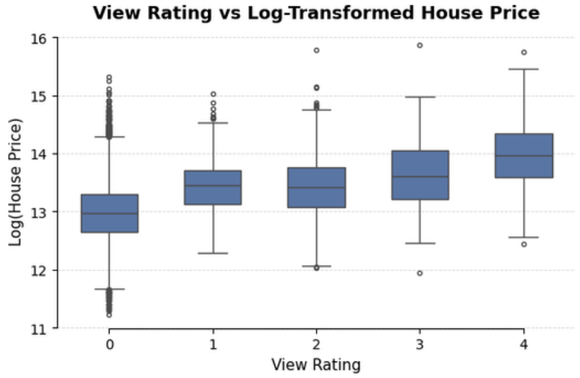
There is a clear upward trend between the number of bathrooms and house price, with the most significant price gains occurring as properties move from single to multi-bathroom configurations.

Exploratory Data Analysis: Continued

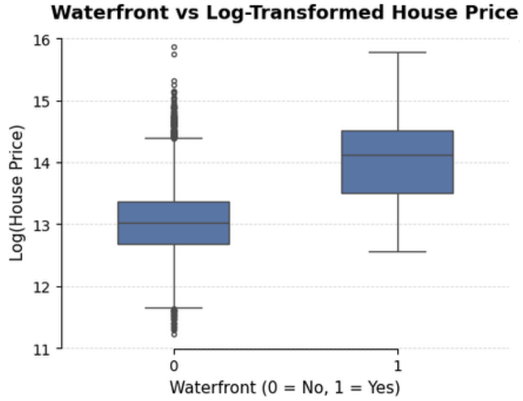
Relationship with key tabular features:



House prices exhibit a strong, consistent upward trend as the construction grade increases, with the highest-rated properties (grades 11–13) commanding a significant price premium.

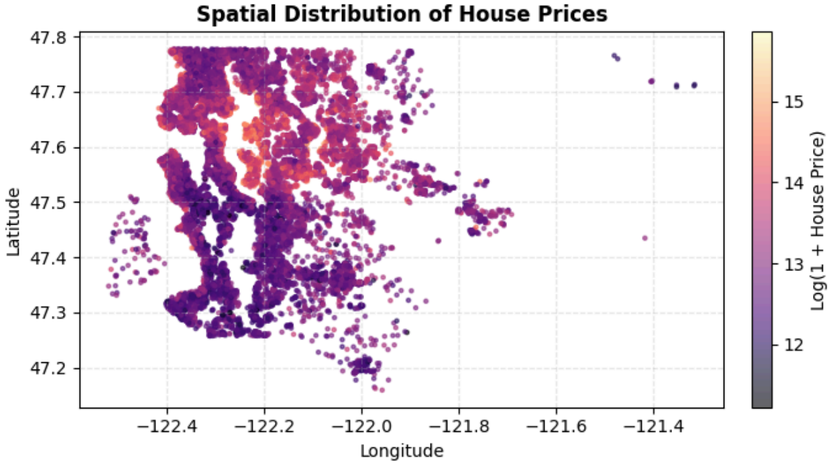


Properties with higher view ratings consistently command a price premium, with a significant jump in median value observed for homes with a top-tier rating of 4.

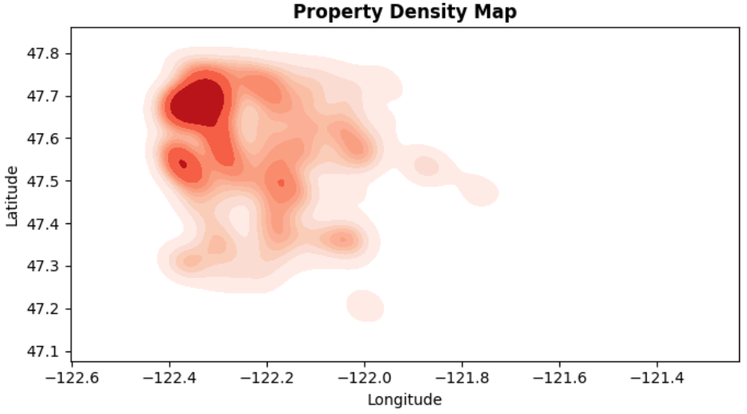


Waterfront properties command a substantial price premium, with median values significantly higher than non-waterfront homes despite a much smaller inventory of such listings.

Spatial analysis:



The spatial scatter plot maps properties by latitude and longitude, with a color gradient representing log-transformed house prices. We observe strong geographic clustering, where high-value properties are concentrated in specific coastal or central zones. This confirms the presence of spatial autocorrelation—the principle that a property's value is heavily dictated by its immediate neighbors.



The Kernel Density Estimation (KDE) plot visualizes the concentration of property listings. The data is not uniform; instead, it shows "hotspots" in dense urban and suburban centers. Because these high-density areas often overlap with high-price clusters, it suggests that urban amenities and infrastructure are significant drivers of value. Furthermore, the model has the most training examples in these dense regions, leading to higher localized predictive confidence.

The combined analysis demonstrates that house prices are influenced not only by structural attributes but also by the surrounding geographic and visual environment. Since neighboring properties share similar physical and environmental characteristics, satellite imagery can capture valuable contextual information such as land use patterns, proximity to water bodies, road networks, and urban density. This strongly motivates the inclusion of satellite imagery as an additional input to enhance predictive performance.

Exploratory Data Analysis: Continued

Sample Satellite Images :

To understand the visual context surrounding properties, satellite images were retrieved using the latitude and longitude of each house through the SentinelHub API. These images capture neighborhood-level characteristics such as road networks, greenery, urban density, and proximity to natural features.


$ 612,000


$ 392,000


$ 399,888


$ 385,000


$ 235,000


$ 390,000


$ 485,000


$ 1,695,000


$ 650,000


$ 275,000


$ 450,000


$ 625,000

Random Sample Images:

- Images show diverse neighborhood patterns:
  - Dense urban grids
  - Green/forested regions
  - Road networks
  - Mixed residential layouts
- Prices vary even when house structure is unseen.
- Indicates surrounding context matters, not just the building.

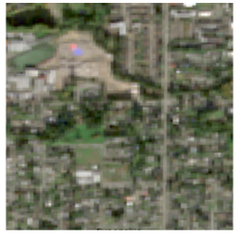
Key Idea:

Different visual surroundings correspond to different price levels.


$ 75,000


$ 80,000


$ 81,000


$ 82,000


$ 84,000


$ 85,000

Cheap Properties:

- Less organized layouts
- Fewer visible amenities
- Sparse or irregular road structure
- Less proximity to water or open spaces

Expensive Properties:

- Proximity to water bodies
- Planned neighborhoods
- Dense but organized housing
- More greenery and open space

Overall, these visualizations validate the inclusion of satellite imagery as an additional data modality. They demonstrate that geographic and visual context contains meaningful information that is not fully captured by tabular features alone and can enhance house price prediction.


$ 7,700,000


$ 7,062,500


$ 6,885,000


$ 5,110,800


$ 4,668,000


$ 4,489,000

Financial & Visual Insights

Environmental Quality as a Financial Signal:

Satellite imagery captures neighborhood-level environmental attributes that influence buyer willingness to pay but are difficult to quantify directly. Visual patterns such as **greenery, open spaces, proximity to water bodies,** and **planned residential layouts** are consistently associated with **higher property prices**.
These visual signals align closely with key **tabular features** used in the baseline models. Variables such as **sqft_living** and **grade** capture **internal size** and **construction quality**, while **sqft_living15** reflects **neighborhood affluence and density**. Features like **view** and **waterfront** explicitly encode access to **natural amenities**, and **latitude–longitude** act as **spatial proxies** for locational desirability.
Satellite imagery complements these features by capturing **broader environmental context**—such as **green cover, road structure, and spatial organization**—which reinforces numerical indicators of neighborhood quality.

Key Financial Insight:

Environmental quality operates as a neighborhood-level asset that increases willingness to pay. Satellite imagery captures this latent value at scale and economically contextualizes strong tabular signals.

Visual Indicator	Economic Driver	Impact on Valuation
High Green Cover	Perceived Prestige & Privacy	Positive (+)
Proximity to Water	Natural Amenity / Scarcity	High Positive (++)
Dense Concrete Grid	Urban Congestion / High Density	Neutral to Negative (-)
Organized Layouts	Planned Infrastructure	Positive (+)

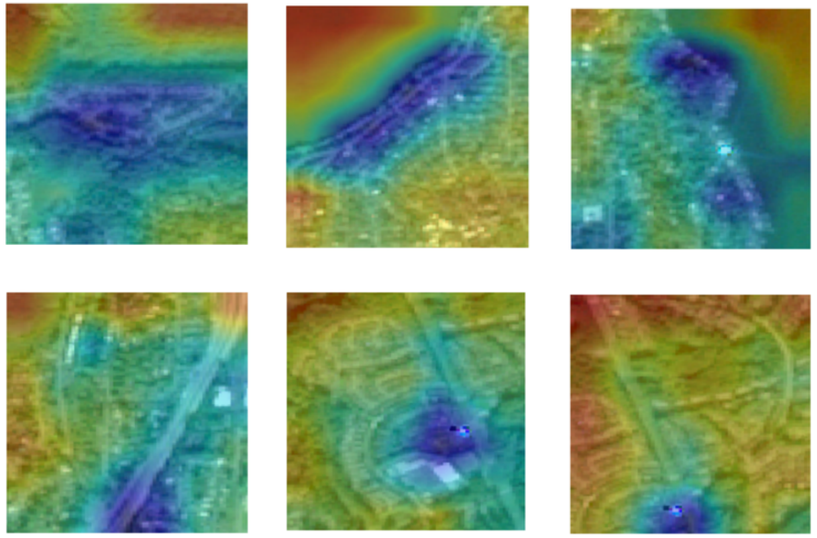
Model Interpretability Using Grad-CAM:

To interpret the visual information learned by the multimodal neural network, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to the ResNet50 image backbone. Grad-CAM enables visualization of the spatial regions in satellite imagery that contribute most strongly to the predicted property price.

- **Technical Implementation:**
  Activation maps were extracted from the final convolutional layer of ResNet50 (conv5_block3_out). By computing the gradients of the predicted price with respect to these feature maps, importance heatmaps were generated and overlaid on the original satellite images, highlighting the regions that most influenced the model's predictions.

- **Key Observation: Neighborhood Context:**
  As shown in the Grad-CAM visualizations, the model's attention is rarely concentrated solely on individual building rooftops. Instead, high-activation regions consistently correspond to surrounding greenery and open spaces, road structure and neighborhood layout, and proximity to water bodies or other natural amenities. This indicates that the CNN prioritizes broader environmental and spatial context rather than isolated structural details of the house.

- **Economic Interpretation:**
  This behavior closely mirrors real-world property appraisal practices, where neighborhood quality and environmental appeal play a central role in determining value. By focusing on contextual features, the model effectively learns a visual proxy for curb appeal, a qualitative factor that is difficult to capture using tabular variables alone.



Although multimodal neural networks were explored, the final predictions were generated using the tabular XGBoost model due to its consistently better performance. Visual models were retained for explainability and analysis of environmental context using Grad-CAM.

Financial & Visual Insights: Continued

Why Tabular Models Remain Dominant?

- Despite the inclusion of satellite imagery, tabular models consistently achieved superior predictive performance. This outcome is primarily driven by the strong explanatory power of engineered tabular features and their overlap with the information present in visual data.
- Several tabular variables already encode key drivers of property value. Features such as sqft_living and grade capture internal size and construction quality, while sqft_living15 reflects neighborhood density and affluence. Variables such as view and waterfront explicitly represent access to scenic and natural amenities, which are also visually observable in satellite images. Additionally, latitude and longitude act as strong spatial proxies, implicitly capturing neighborhood desirability, urban proximity, and infrastructure quality.
- As a result, much of the environmental and neighborhood context learned by the convolutional neural network through satellite imagery is already reflected in the tabular feature space. This leads to diminishing marginal returns when visual features are added to an already information-rich tabular model.
- Furthermore, gradient-boosted decision tree models such as XGBoost are particularly effective at exploiting structured data, capturing non-linear relationships and feature interactions with high efficiency. In contrast, satellite imagery introduces higher dimensionality and noise, which does not always translate into proportional performance gains when strong tabular signals are present.

Key Insight:
Tabular models outperform multimodal models because key environmental and locational signals captured by satellite imagery are already encoded in features such as grade, view, waterfront, neighborhood density, and geographic location. As a result, adding images provides limited incremental predictive benefit when strong structured features are available.

Practical Implications: Images as an Audit, Trust, and Robustness Layer

Although tabular models remain the primary drivers of price prediction accuracy, satellite imagery provides substantial practical value when used as a complementary layer rather than a replacement.

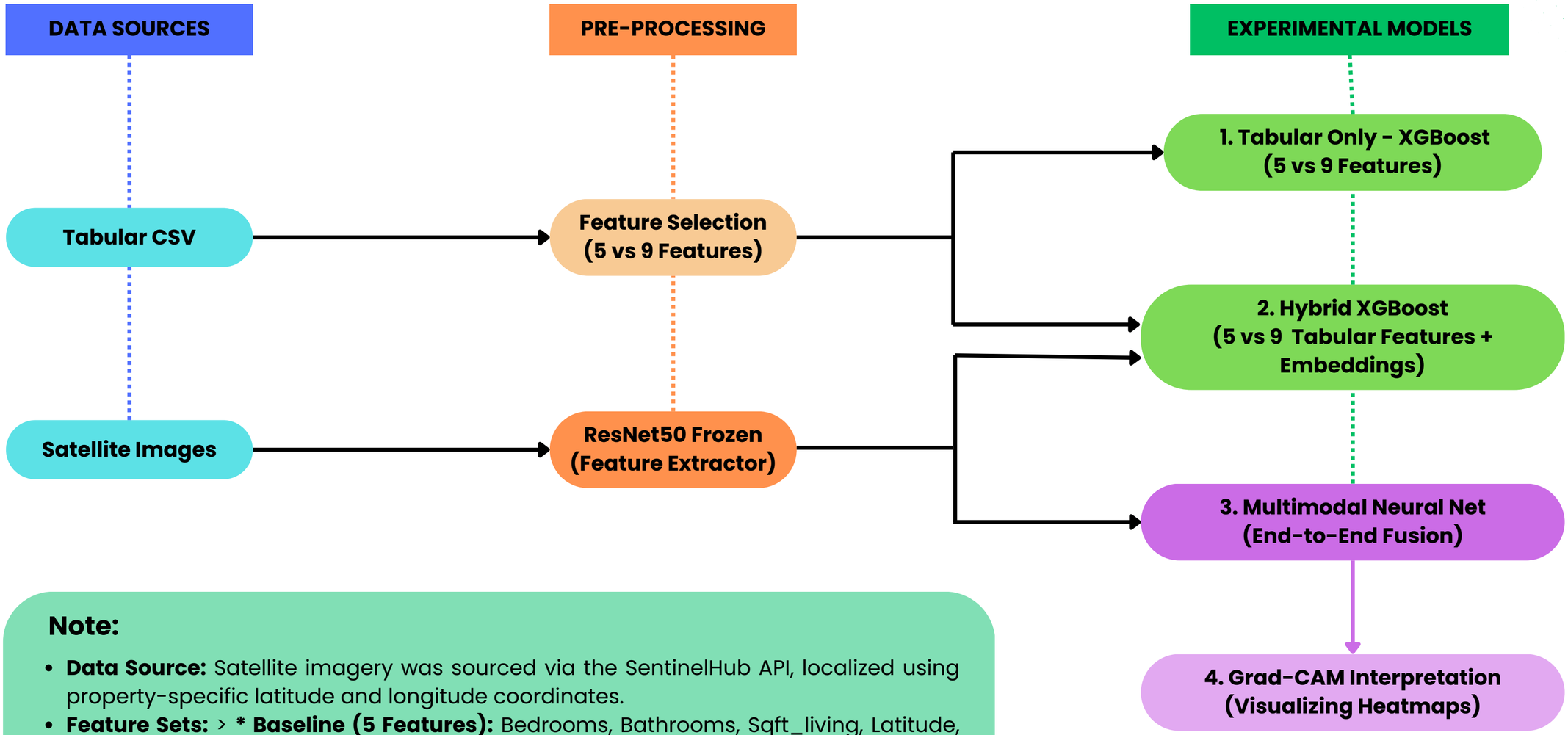
Audit Layer:
Satellite images enable manual and automated validation of tabular inputs. Visual inspection can reveal inconsistencies such as misreported views, incorrect neighborhood classification, or outdated location attributes that may not be evident from structured data alone.

Trust & Transparency:r:
Using Grad-CAM visualizations, stakeholders can observe which environmental factors influence model reasoning. By showing that the model attends to greenery, road structure, and neighborhood layout, the system becomes more interpretable and trustworthy for real estate professionals, policymakers, and financial institutions.

Robustness to Data Limitations:
In scenarios where tabular data is incomplete, noisy, or unavailable—such as emerging markets or newly developed regions—satellite imagery can act as a fallback signal, capturing neighborhood quality and spatial context at scale.

Key Takeaway:
Satellite imagery should be viewed as a supporting intelligence layer that enhances model reliability, interpretability, and robustness, while final valuation decisions remain driven by strong tabular predictors.

Architecture Diagram



Note:

- **Data Source:** Satellite imagery was sourced via the SentinelHub API, localized using property-specific latitude and longitude coordinates.
- **Feature Sets:** > * **Baseline (5 Features):** Bedrooms, Bathrooms, Sqft_living, Latitude, and Longitude.
    - **Extended (9 Features):** Baseline + Grade, Sqft_living15 (neighbor area), View Rating, and Waterfront status.

Architecture Diagram – System Explanation

This section explains the end-to-end modeling pipeline used for property price prediction, as illustrated in the architecture diagram.

1. Data Sources:

- **Tabular Data (CSV):** The primary dataset consists of structured property attributes such as living area, bedrooms, bathrooms, and grade, etc.
- **Satellite Imagery:** Images sourced via the SentinelHub API using property-specific latitude and longitude. These images provide visual context regarding neighborhood density, greenery, and infrastructure.

2. Pre-Processing & Feature Engineering:

2.1 Model-Specific Tabular Processing

A critical distinction was made in how data was fed into different model types to respect their mathematical requirements:

- **XGBoost Path:** Features were used in their raw format. Since XGBoost is a tree-based ensemble, it is invariant to feature scaling, preserving the natural interpretability of the data (e.g., actual square footage).
- **Neural Network Path:** Tabular features underwent Z-score Standardization (using StandardScaler). This ensures all inputs have a mean of 0 and a standard deviation of 1, preventing high-magnitude features from dominating the gradient updates.

2.2 Image Pre-Processing

Satellite images were resized to 224 x 224 pixels and normalized using ImageNet-standard means and deviations to align with the pre-trained ResNet50 weights used for feature extraction.

3. Experimental Modeling Strategy:

The project followed an incremental "Experimental Ladder" to isolate the value of each data modality.

Experiment	Model Type	Tabular Features	Visual Data	Purpose
Stage 1	XGBoost	5 Core	None	Establish a baseline for price prediction using primary structural data.
Stage 2	Hybrid XGBoost	5 Core	ResNet Embeddings	Evaluate if visual environmental context (greenery, density) improves basic structural models.
Stage 3	Extended XGBoost	9 Full	None	Optimize the tabular model using locational proxies (Lat/Long) and neighborhood stats.
Stage 4	Hybrid XGBoost	9 Full	ResNet Embeddings	Determine if satellite imagery provides "marginal gains" over high-quality locational data.
Stage 5	Neural Network	9 Full	ResNet (Late Fusion)	Create an end-to-end multimodal architecture for Grad-CAM interpretability.

4. Final Model Optimization (Optuna):

To push the performance of the Final Extended Tabular Model (Stage 3) to its limit, Optuna was employed for automated hyperparameter tuning.

- **Bayesian Optimization:** Unlike a random search, Optuna uses past trials to "narrow in" on the best parameters, such as learning_rate, max_depth, and subsample.
- **Peak Performance:** This optimization was key to ensuring the 9-feature tabular model reached its maximum predictive accuracy.

5. Explainability (SHAP & Grad-CAM):

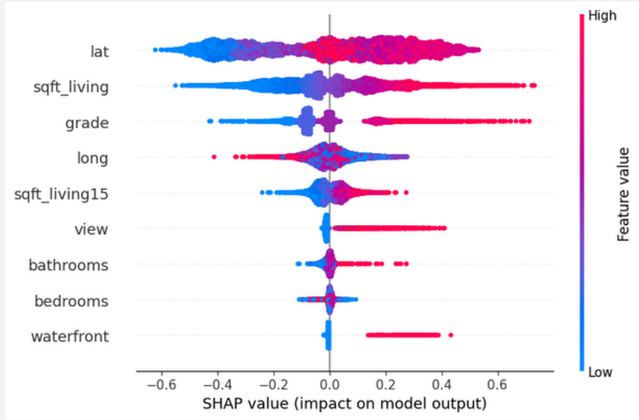
The architecture was designed to be a **Glass Box**, providing insights into how the models **think**:

- **SHAP (Tabular):** Used on the XGBoost models to rank which features (like lat or grade) had the highest impact on individual price predictions.
- **Grad-CAM (Visual):** Applied to the last convolutional layer of the Neural Network to produce heatmaps. This visually identifies which parts of a satellite image—such as green spaces or road layouts—influenced the model's valuation.

6. Overall Pipeline Summary:

The pipeline balances Predictive Accuracy (achieved through the Optimized 9-feature XGBoost) with Spatial Interpretability (achieved through the Multimodal Neural Network). While structured data proved to be the strongest predictor, the integration of satellite imagery validated the economic impact of environmental quality on property valuation.

SHAP Feature Importance Analysis



The SHAP summary plot illustrates how individual features influence the model's price predictions. Features are ranked by overall impact, with horizontal positioning indicating whether a feature increases or decreases the predicted value.

Key Findings:

- **Location Dominance:** Latitude (lat) is the most influential predictor, highlighting that geographic position is the primary driver of property value.
- **Structural Drivers:** House size (sqft_living) and construction quality (grade) show a direct positive correlation with price—larger, higher-grade homes consistently command higher valuations.
- **Spatial Context:** Longitude (long) and neighboring house sizes (sqft_living15) further refine the model by capturing regional trends and neighborhood affluence.
- **Amenity Premiums:** Premium features like a good view or waterfront status provide significant positive price spikes, even if they are less frequent in the dataset.
- **Diminishing Returns:** Basic utility counts, such as bedrooms and bathrooms, have a relatively minor impact compared to the overarching factors of location, size, and quality.

The analysis confirms that property value is primarily determined by where the house is located and how large it is, while environmental and neighborhood factors provide the final marginal value adjustments.

Results

Since ground-truth prices are unavailable for the test set, all model performance is evaluated on a validation split derived from the training data. The original dataset of approximately 16,209 samples was partitioned into 80% training and 20% validation.
All reported metrics are computed on the validation set, using log-transformed house prices as the prediction target. Model performance is assessed using Root Mean Squared Error (RMSE) and R^2 score.

Quantitative Performance Comparison

The following table summarizes the validation performance of all evaluated models:

Model Type	Features Used	RMSE ↓	R-Squared ↑
XGBoost (Baseline)	5 Tabular Features	0.2004	0.8545
XGBoost (Hybrid)	5 Tabular Features + Image Embeddings	0.1967	0.8598
XGBoost (Baseline)	9 Tabular Features	0.1755	0.8884
XGBoost (Hybrid)	9 Tabular Features + Image Embeddings	0.1794	0.8833
End to End Neural Network	9 Tabular + Images	0.4024	0.4132

(Lower RMSE and higher R^2 indicate better performance.)

Impact of Satellite Imagery

Case 1: Limited Tabular Information (5 Features)

When only a restricted set of 5 tabular features was used, incorporating satellite image embeddings led to a modest but consistent improvement in model performance. The hybrid model achieved a lower RMSE and higher R^2 compared to the tabular-only baseline.

Interpretation:

In settings with limited structural information, satellite imagery provides complementary environmental context—such as greenery, road density, and neighborhood layout—which helps improve predictive accuracy.

Case 2: Rich Tabular Information (9 Features)

When the feature set was expanded to 9 informative tabular variables, the tabular-only XGBoost model already achieved strong performance. In this case, adding image embeddings did not improve accuracy and resulted in a slight degradation in RMSE and R^2.

Interpretation:

Key tabular features such as property size, grade, location (latitude and longitude), and waterfront access already capture much of the information present in satellite images. As a result, the visual signal becomes largely redundant, limiting its marginal contribution.

Neural Network Performance

The end-to-end multimodal neural network underperformed relative to all tree-based models, exhibiting substantially higher RMSE and lower R^2 on the validation set.

Several factors contribute to this outcome:

- **Limited dataset size for deep multimodal learning:**
  Deep neural networks require large-scale data to effectively learn joint representations from images and structured features. The available dataset, while sufficient for tree-based models, is relatively small for end-to-end visual–tabular fusion.
- **Frozen CNN backbone restricting task-specific adaptation:**
  The ResNet50 image encoder was kept frozen to prevent overfitting, which limited the model's ability to adapt visual features specifically to property valuation rather than generic object recognition.
- **Strong inductive bias of tree-based models for tabular data:**
  Gradient-boosted trees (XGBoost) are inherently well-suited for structured economic variables, capturing non-linear interactions and threshold effects more efficiently than neural networks.
- **Redundancy between tabular and visual signals:**
  Key pricing information such as location, neighborhood quality, and amenities is already explicitly encoded in tabular features (e.g., latitude, longitude, view, waterfront), reducing the marginal contribution of satellite imagery in an end-to-end setup.
- **Higher sensitivity of neural networks to architectural and hyperparameter choices:**
  Neural models are more sensitive to learning rates, fusion strategies, and network depth, making stable optimization more challenging in comparison to tree-based methods.

Despite weaker predictive performance, the multimodal neural network remains valuable from an interpretability and analytical perspective. Grad-CAM visualizations demonstrate that the model learns meaningful spatial patterns related to neighborhood structure, greenery, and environmental context. This confirms that satellite imagery captures economically relevant signals, even when these signals do not translate into superior predictive accuracy.

Key Takeaways

- Satellite imagery adds value when tabular information is sparse.
- With strong tabular features, tree-based models outperform multimodal approaches.
- XGBoost consistently delivers the best accuracy for structured housing data.
- Multimodal neural networks are better suited for explainability than raw performance in this setting.

THANK YOU