By: RIDDHI SIDANA
23322023

SATELLITE IMAGERY-BASED PROPERTY VALUATION

~A Multimodal Regression Approach

OVERVIEW

Accurate property valuation is a critical task in real estate analytics, traditionally driven by structured attributes such as house

This project addresses this limitation by developing a multimodal regression framework that integrates tabular housing data with satellite imagery–derived embeddings extracted using a pretrained Convolutional Neural Network, allowing the model to account for both numerical and visual drivers of property market values.

Each property was mapped to a unique satellite image using its latitude and longitude, stored in structured directories for training and testing.

Additionally, visual explainability technique such as Grad-CAM is used to understand which environmental features influence the model's predictions, increasing transparency and trustworthiness in real estate decision-making.

DATA DESCRIPTION

Tabular Dataset

Each row in the base dataset corresponds to a residential property, with the target variable being property price. The dataset includes a range of structural, locational, and neighbourhood-related features, such as:

- **Target:** Price
- Structural attributes: bedrooms, bathrooms, sqft_living, sqft_above, sqft_basement, floors
- Lot and neighbourhood characteristics: sqft_lot, sqft_living15, sqft_lot15
- Quality and condition indicators: grade (construction and design quality), condition (maintenance level)
- Environmental indicators: view (view quality), waterfront (binary indicator)
- Location information: latitude, longitude, zipcode
- Temporal features: yr_built, yr_renovated
- **Additional engineered features:** house age (derived to improve model expressiveness)

Train shape: (16209, 21) , Test shape : (5404, 20)

Null Values : 0

Satellite Imagery Dataset

- To capture environmental context, the **Google Maps Static API** was used to download satellite images corresponding to each property location, with a fixed zoom level of **19** and image resolution of **400 x 400** to ensure consistency across samples.
- The slight discrepancy between tabular records and available images arises because duplicate property entries were removed by retaining the most recent record based on transaction date.

Train images: 16110 , Test images: 5396 = no of. unique properties

METHODOLOGY

1

DATA COLLECTION AND PREPROCESSING

Structured housing data were cleaned, outliers were removed, followed by feature selection, basic preprocessing, and **log transformation of prices** to reduce skewness.

2

SATELLITE IMAGE ACQUISITION

Satellite images were programmatically downloaded via the **Google Maps Static API** to capture neighbourhood-level environmental context such as green cover, road layout, and surrounding density.

3

EXPLORATORY AND GEOSPATIAL ANALYSIS

Exploratory data analysis was conducted to study price distributions, structural feature relationships, spatial patterns, and preliminary associations between visual environmental factors and property prices.

4

VISUAL FEATURE ENGINEERING USING CNN

Satellite images were processed using a pretrained **ResNet50** convolutional neural network with frozen weights and global average pooling to extract high-dimensional image embeddings representing environmental characteristics.

5

IMAGE-ONLY AND TABULAR BASELINE MODELING

Separate regression models(Random Forest, Ridge, XgBoost) were trained using tabular features alone and image embeddings alone to quantify the individual predictive contribution of each data modality.

6

MULTIMODAL FEATURE FUSION

Dimensionality reduction **(PCA)** was applied to image embeddings, and an early fusion strategy was used to concatenate tabular features with visual embeddings into a unified feature representation.

7

MULTIMODAL REGRESSION MODELING AND EVALUATION

A gradient-boosted decision tree **(XGBoost)** regressor was trained on the fused feature set to predict property prices, leveraging both structural and visual information. Model performance was evaluated using RMSE and R^2, and results were compared across tabular-only, image-only, and multimodal fusion models to assess the benefit of incorporating satellite imagery.

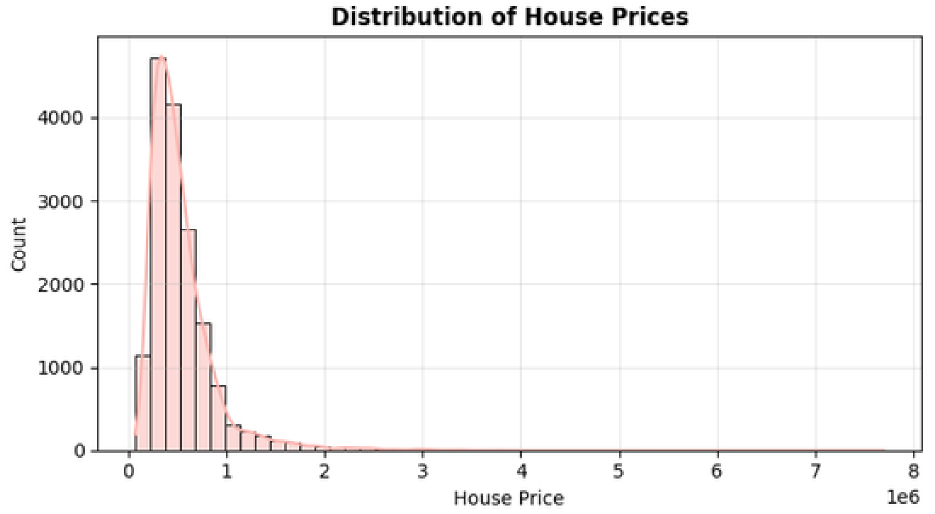
8

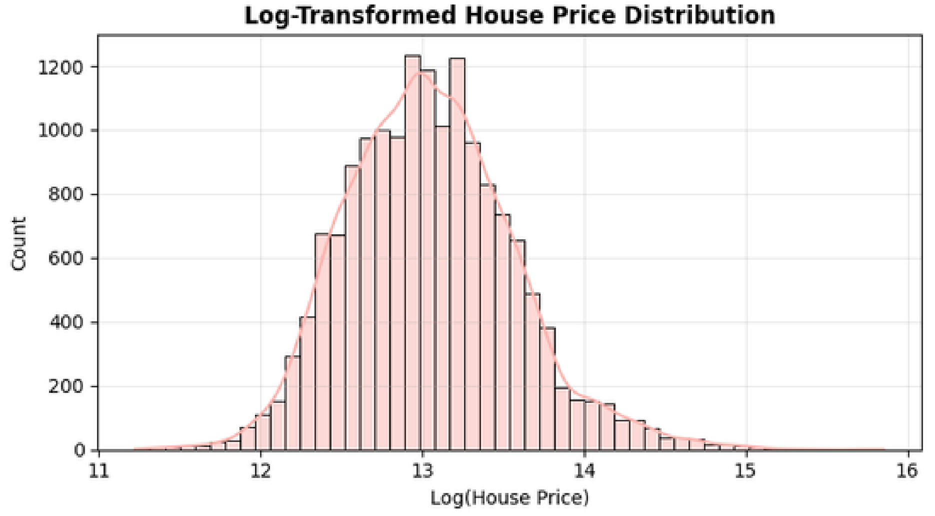
MODEL EXPLAINABILITY

To enhance interpretability, **Grad-CAM** was applied to the CNN feature extractor to visualise the satellite image regions most influential in learning visual representations related to property value and **SHAP** was applied to interpret feature importance in tabular models

EDA & INSIGHTS

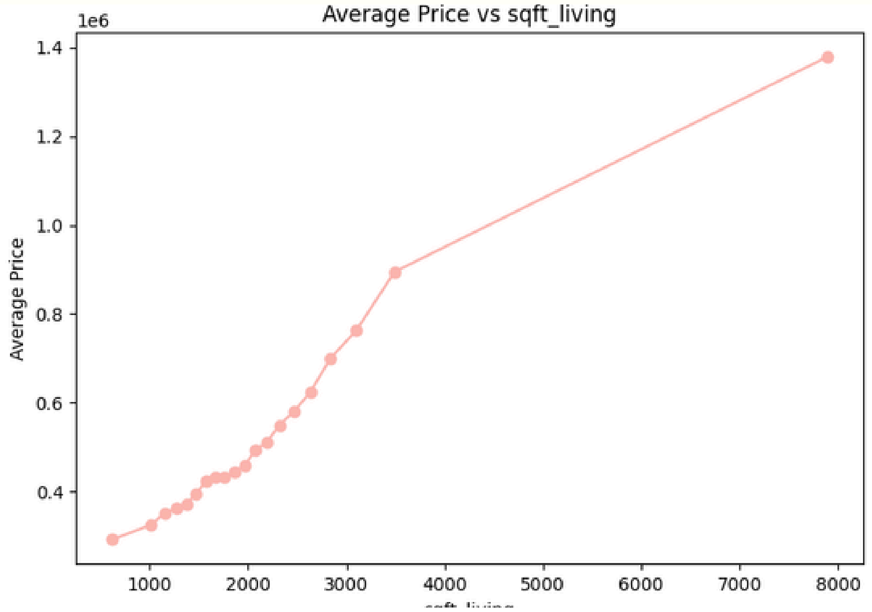
Target variable (price)

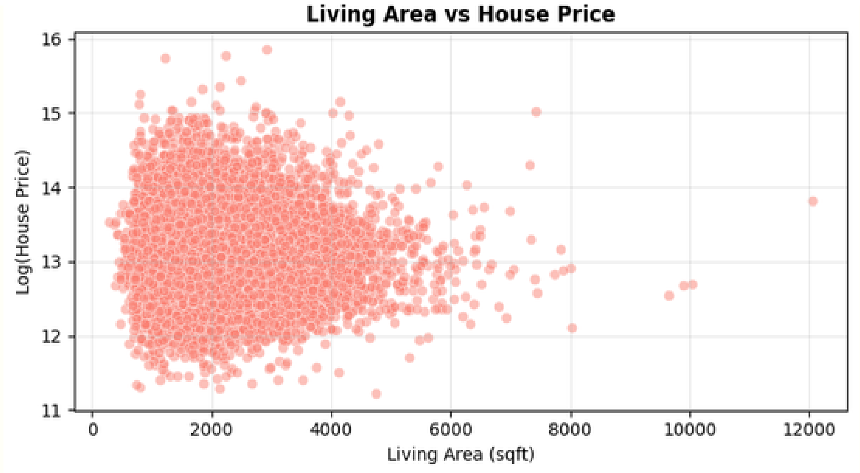




- Property prices exhibit a right-skewed distribution, with a small number of high-valued properties.
- A log transformation (log(1+x)) is therefore applied to stabilize variance and improve model learning.

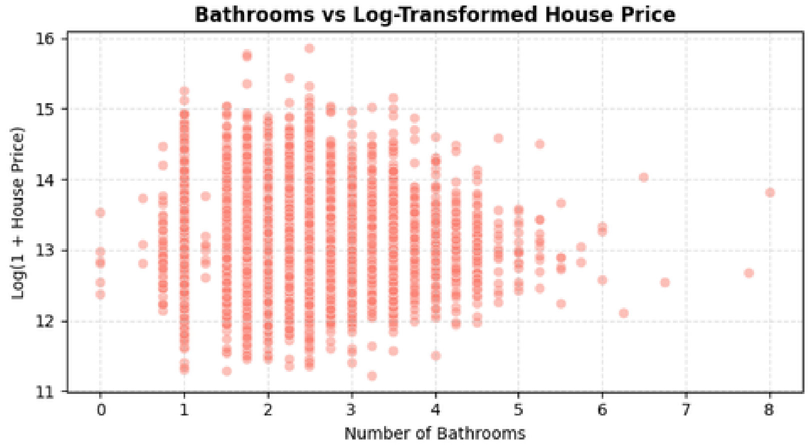
Sqft_living vs Price





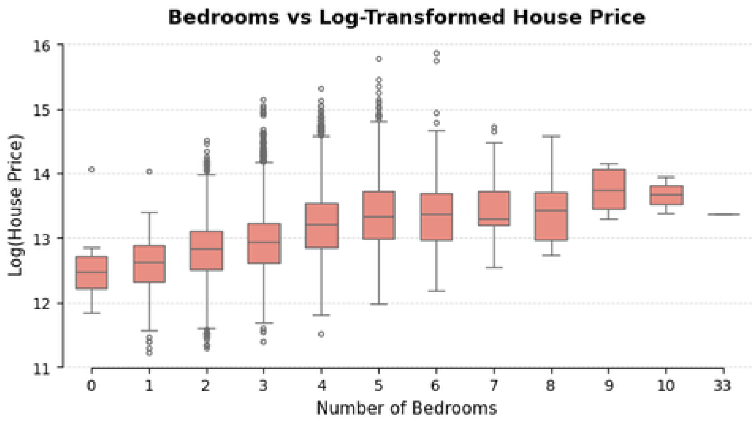
- Living area shows a strong positive relationship with price, confirming that usable interior space is one of the most influential drivers of property valuation.

Bathroom vs Price



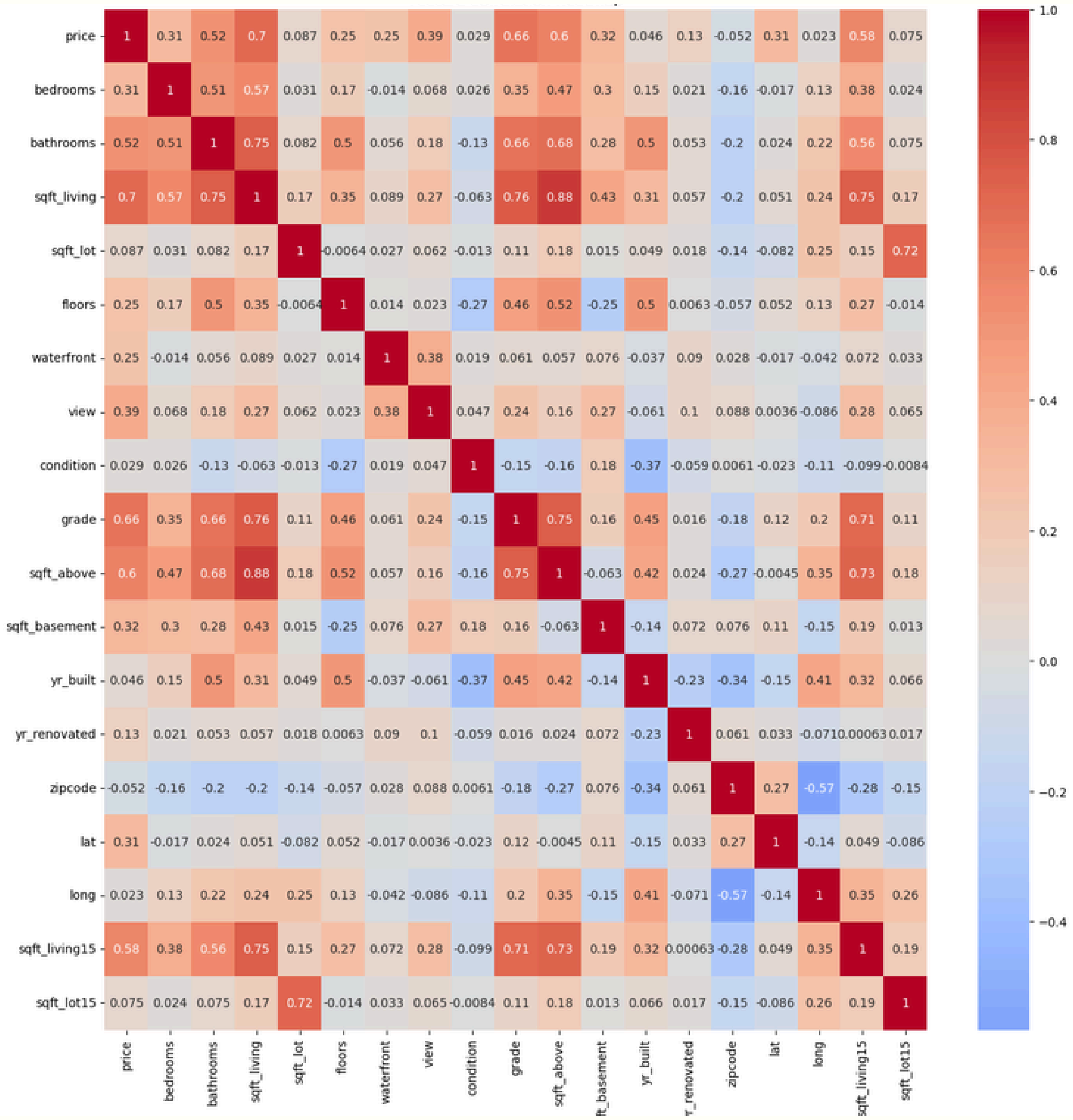
- Houses with more bathrooms tend to command higher prices, reflecting improved livability and functionality, especially for larger households.

Bedroom vs Price



- Property price generally increases with the number of bedrooms, although marginal gains diminish beyond a certain point, indicating that size alone does not fully determine value

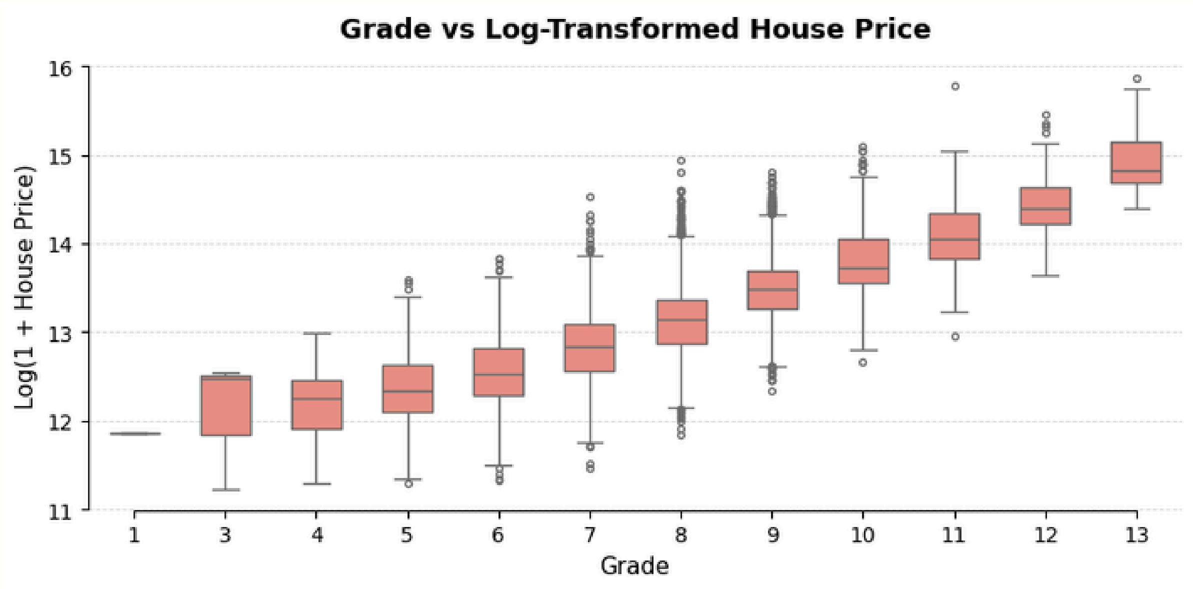
Correlation Heatmap



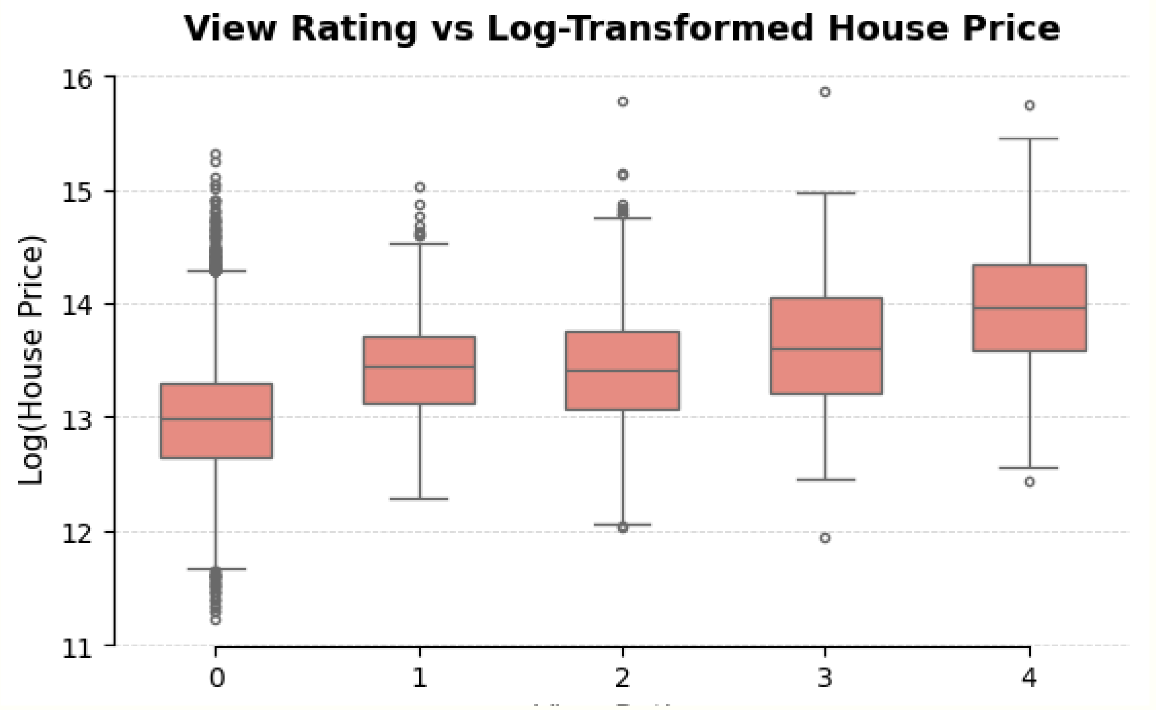
- The correlation heatmap shows that property **price is strongly associated with structural features** such as living area (sqft_living), sqft_above construction quality (grade), number of bathrooms, and above-ground area, while lot size and renovation year exhibit comparatively weaker relationships.
- High intercorrelation among size-related variables (e.g., sqft_living, sqft_above, and sqft_living15) indicates potential **multicollinearity**, justifying the use of tree-based models that can handle correlated features effectively.

EDA & INSIGHTS

Continued



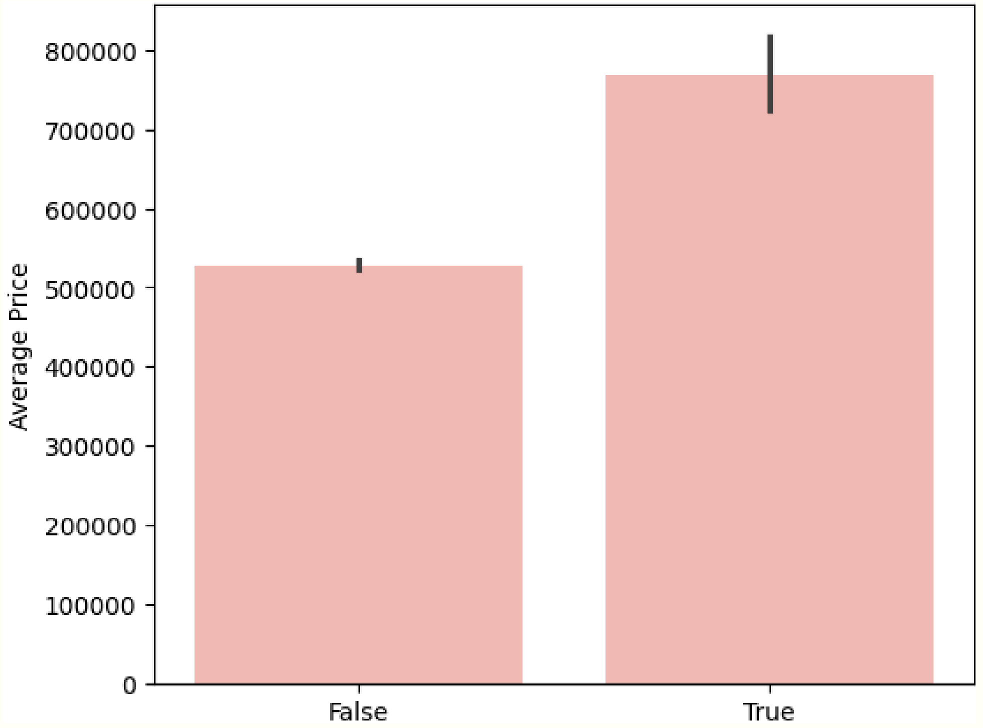
- Construction grade exhibits a clear monotonic relationship with price, where higher-quality architectural design and materials significantly increase property value.

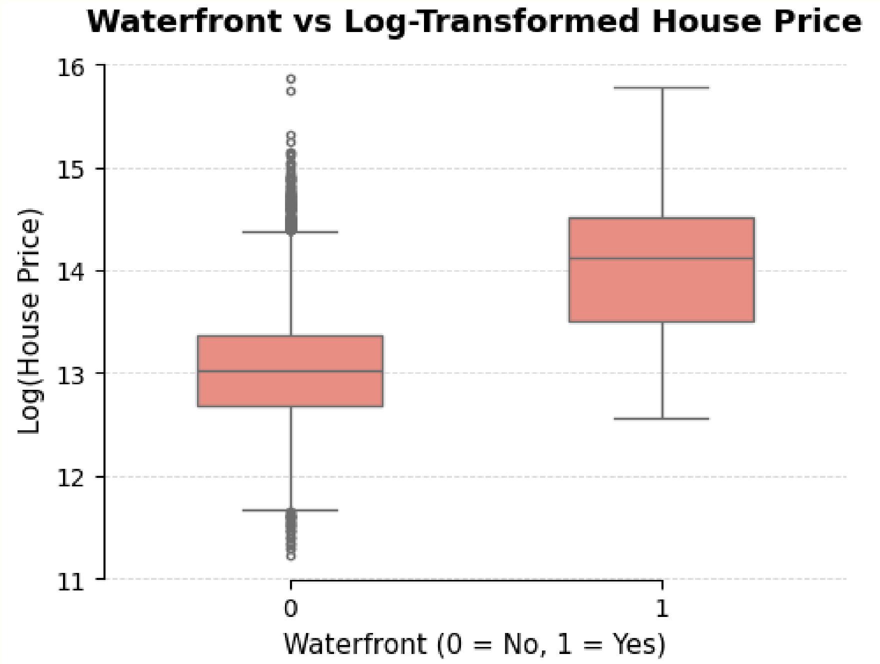


- Properties with higher view ratings consistently exhibit higher prices, highlighting the premium associated with scenic surroundings.

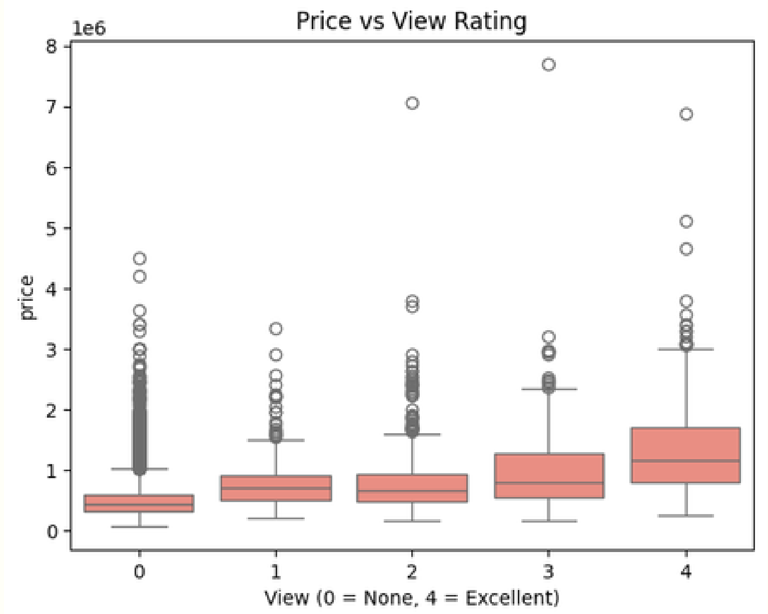
Impact Of Renovation on House Prices

Renovated properties exhibit significantly higher average prices compared to non-renovated homes, indicating that recent upgrades materially enhance property value.

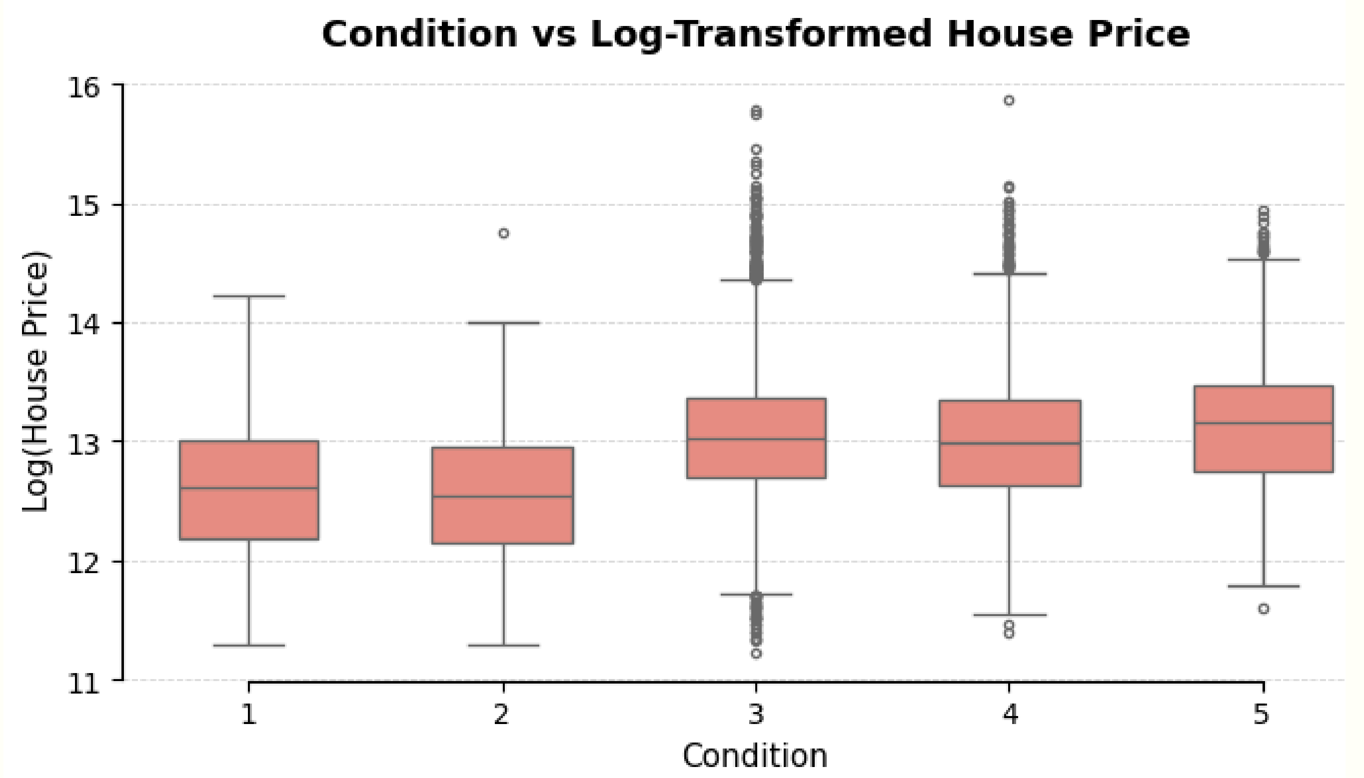




- Waterfront properties command a substantial price premium relative to non-waterfront homes, even after log transformation, highlighting the strong valuation impact of proximity to water.



- Property prices increase consistently with higher view ratings, demonstrating a monotonic relationship between scenic quality and valuation.
- Homes with premium views (ratings 3–4) show higher median prices and greater upside potential, reflecting strong buyer preference for aesthetic and visual amenities.



- Property prices increase with better maintenance condition, but the price uplift is incremental rather than steep, indicating that condition primarily acts as a hygiene factor rather than a dominant value driver.
- Once a baseline condition is met, buyers place relatively greater emphasis on structural size, location, and environmental amenities, consistent with diminishing marginal returns to maintenance quality.

GEOSPATIAL ANALYSIS

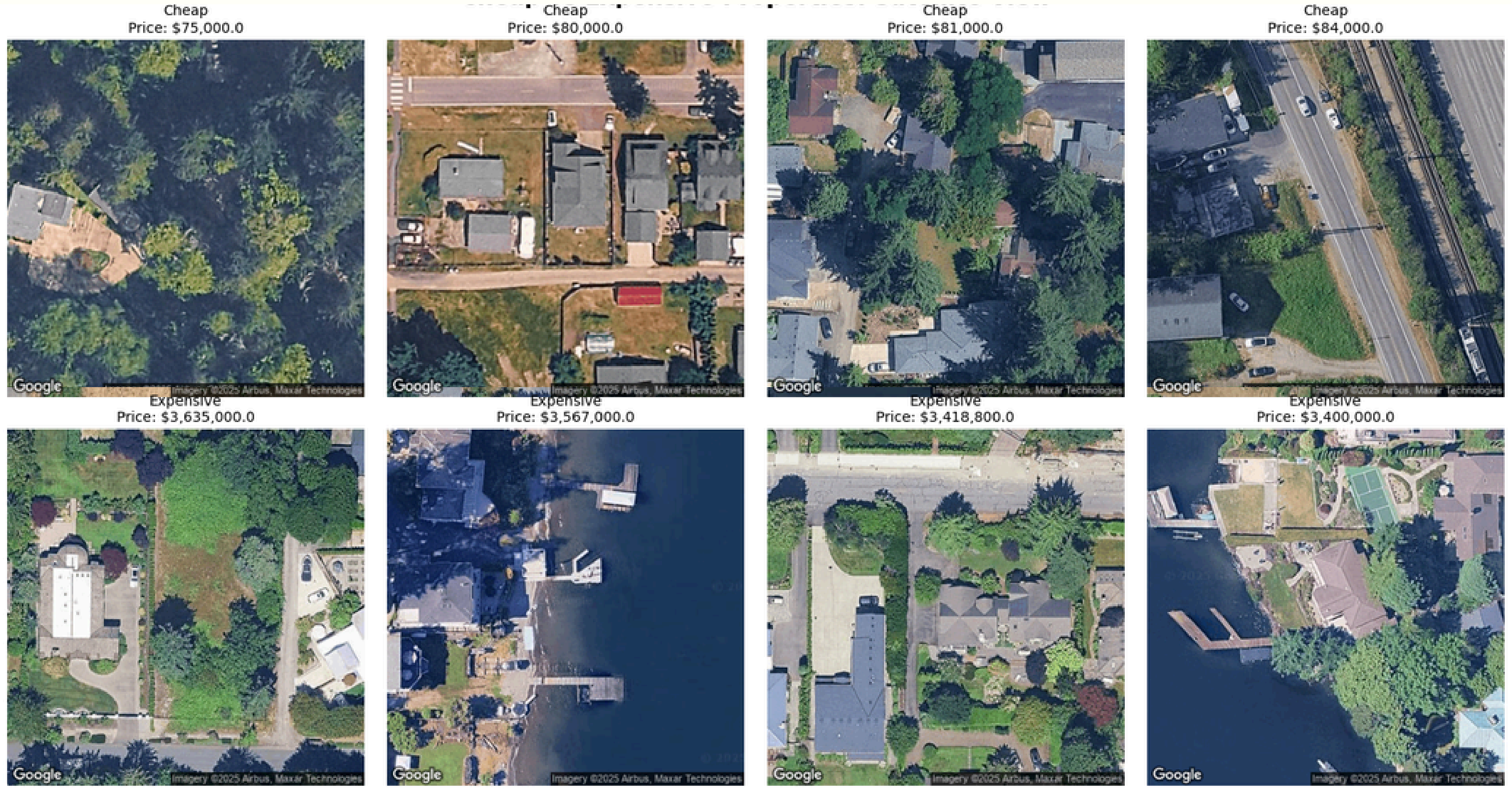
Cheap vs Expensive Properties: Satellite View

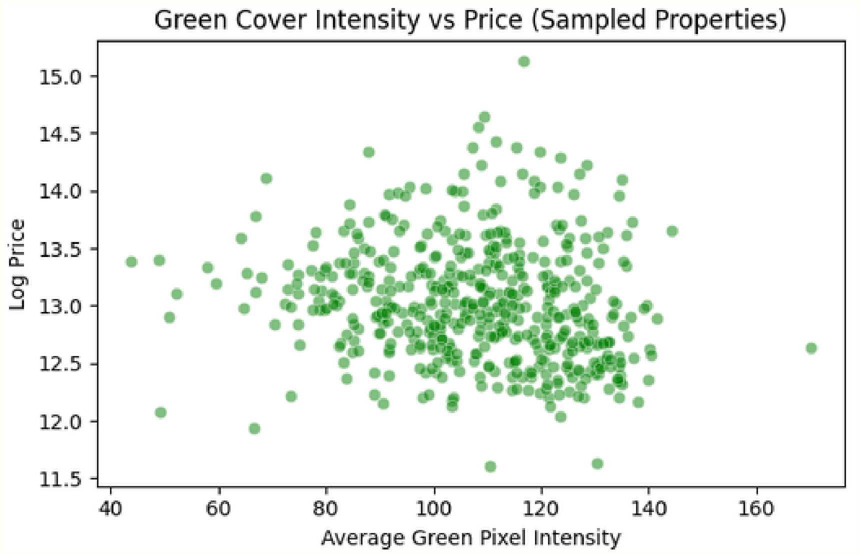
Lower-Priced Properties:

- Located in denser neighbourhoods with closely packed buildings and greater proximity to major roads or infrastructure.
- Smaller plot sizes with limited open space.
- Lower green cover and fewer landscaped areas.

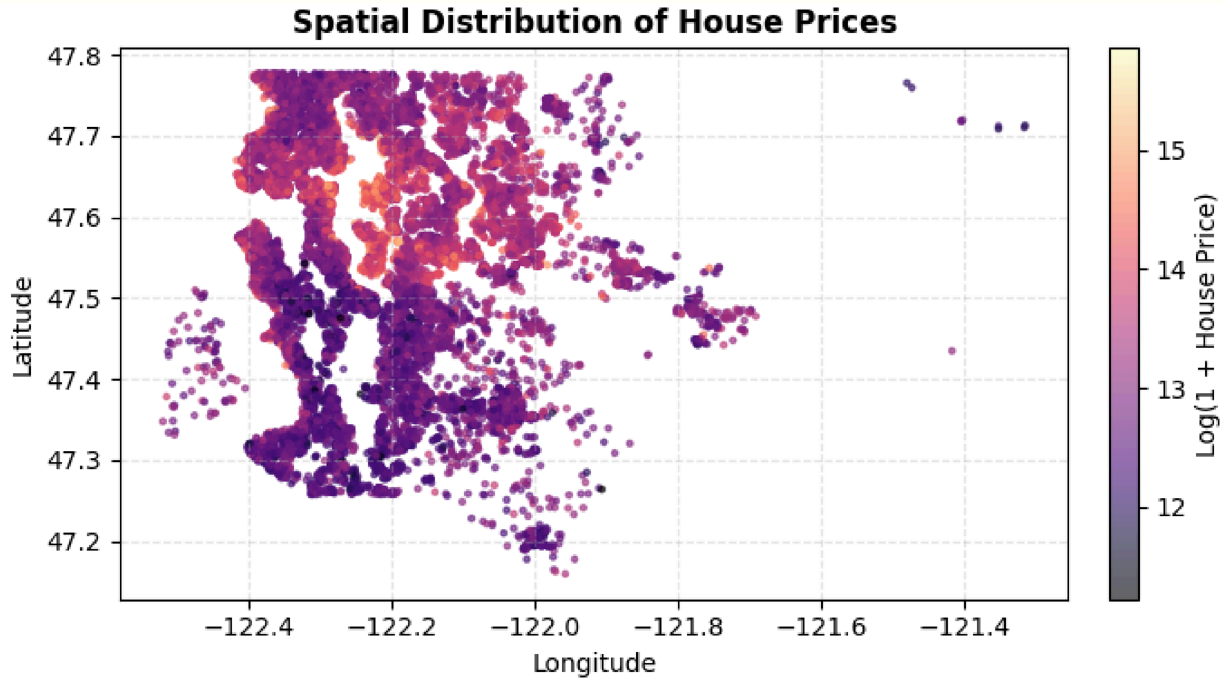
Higher-Priced Properties:

- Larger plots with well-defined open and private spaces and organised neighbourhood layouts with low congestion.
- Higher surrounding greenery and tree cover.
- Presence of premium features such as waterfront access or private amenities.





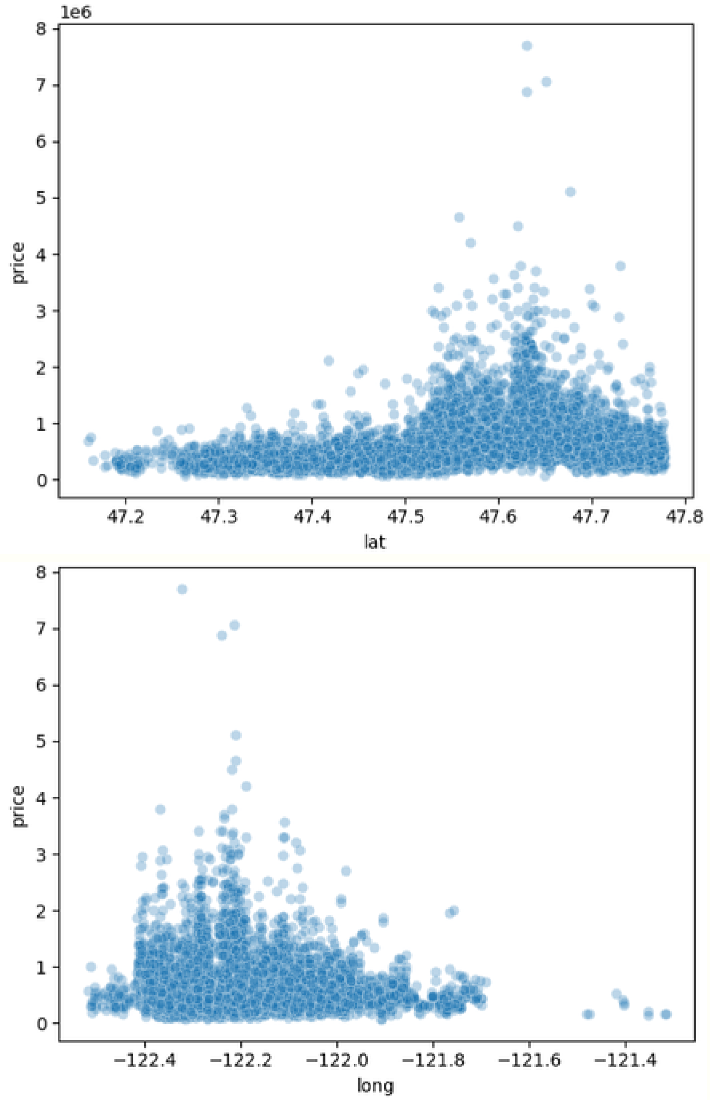
- A positive but non-linear relationship is observed between green cover intensity and property prices, indicating that environmental greenery enhances value when combined with favorable location and neighborhood characteristics.

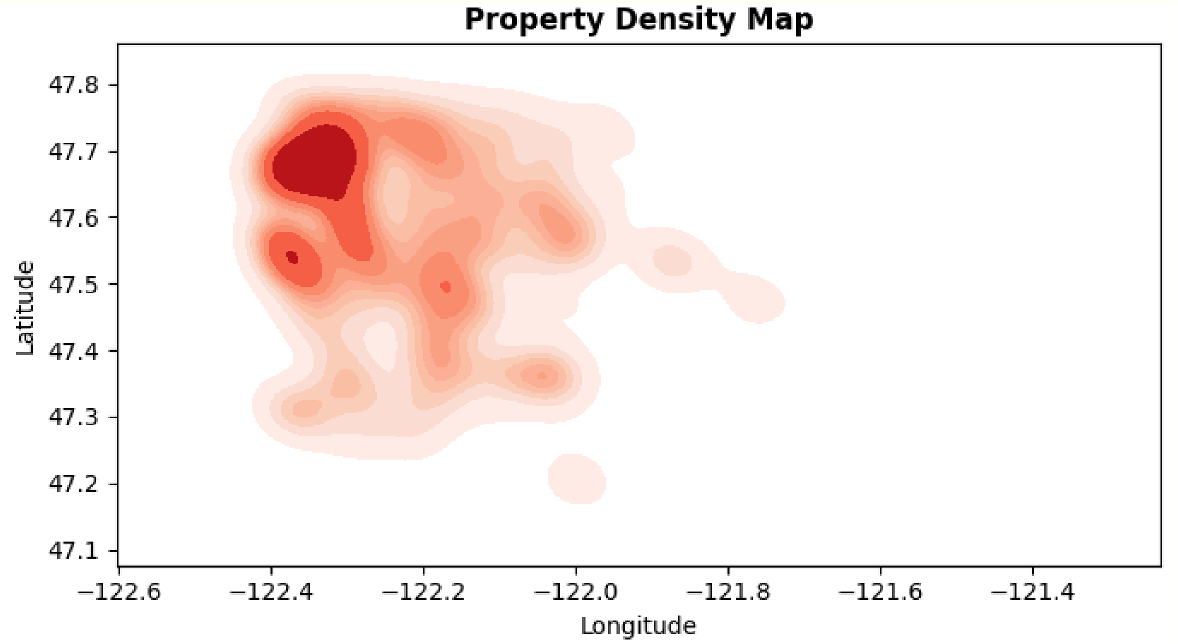


- The spatial scatter plot maps properties by latitude and longitude, with a colour gradient representing log-transformed house prices.
- House prices exhibit pronounced spatial clustering, with high-value properties concentrated in specific geographic pockets rather than being uniformly distributed, highlighting the strong influence of localised neighbourhood effects.

Latitude–Longitude and Price Relationship

- High-priced properties tend to cluster within narrow latitude–longitude bands corresponding to established urban and high-amenity neighbourhoods, while properties outside these bands show greater price dispersion and generally lower valuations.





- The Kernel Density Estimation (KDE) plot visualises the concentration of property listings.
- The data is not uniform; instead, it shows hotspots in dense urban and suburban centres. Because these high-density areas often overlap with high-price clusters, it suggests that urban amenities and infrastructure are significant drivers of value.

MODELING STRATEGY AND PERFORMANCE COMPARISON

Model performance was evaluated using RMSE and R^2 metrics and compared across tabular-only, image-only, and multimodal fusion approaches.

01 BASELINE MODELING (TABULAR ONLY)

- A total of **14** carefully selected tabular features capturing structural attributes, quality indicators, and neighborhood characteristics were used to train the tabular-only regression model.
- Gradient Boosted Decision Trees (XGBoost) were employed due to their ability to model non-linear relationships and handle correlated features effectively.
- Among all tabular baselines, **XGBoost** achieved the highest performance, indicating that structured housing attributes contain strong predictive signals for property valuation.
- **The tabular-only model achieved the highest standalone performance,** reflecting the strong predictive power of structured housing attributes.

02 BASELINE MODELING (IMAGE-ONLY)

- Satellite images were processed using a **pretrained ResNet50 CNN,** initialised with ImageNet weights and configured without the classification head (include_top=False) and with **global average pooling.**
- The CNN weights were kept frozen to leverage pretrained visual representations while avoiding overfitting.
- This architecture produced **2048-dimensional** image embeddings representing environmental and neighborhood-level context.
- After aligning these embeddings with tabular records, an image-only XGBoost regressor was trained, achieving an R^2 of approximately 0.46, confirming that satellite imagery alone captures meaningful but incomplete valuation information.

03 MULTIMODAL FUSION APPROACH

- To combine both modalities, a multimodal early-fusion approach was adopted.
- **16** final tabular features were concatenated with reduced-dimensional image embeddings, where **PCA** was applied to the image feature space to improve numerical stability and reduce redundancy.
- The fused feature representation was then modelled using XGBoost.

- **The multimodal fusion model outperformed the image-only approach and achieved performance comparable to the tabular-only baseline, with nearly identical R^2 values.**

MODEL EXPLAINABILITY

Model Type	Data Modality	Features Used	Model	RMSE (log-price)	R^2
Tabular-Only	Structured housing data	14 tabular features	XGBoost	**0.163**	**0.901**
Image-Only	Satellite imagery	ResNet50 embeddings (2048-D)	XGBoost	0.367	0.464
Multimodal Fusion	Tabular + Satellite imagery	16 tabular features + PCA-reduced image embeddings	XGBoost	0.169	0.894

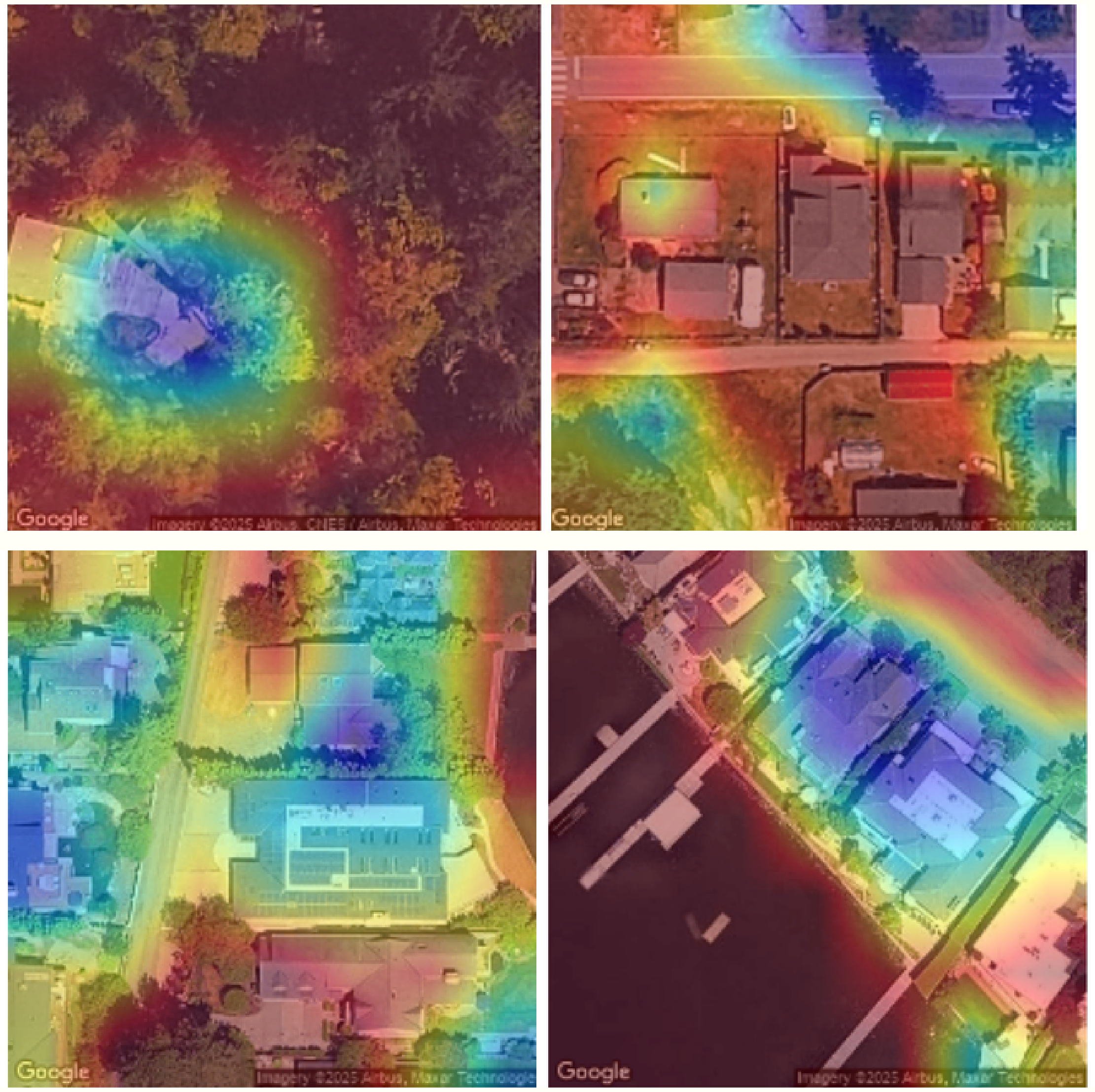
To improve transparency and trust in model predictions, explainability techniques were applied separately to the image-based and tabular-based components of the system. Since the multimodal fusion model relies on both CNN-extracted visual features and structured attributes, explainability was performed at the feature-extraction level for each modality rather than on the final regressor.

- For satellite imagery, **Grad-CAM (Gradient-weighted Class Activation Mapping)** was used to identify image regions that most strongly influenced the CNN's learned representations. For tabular data, **SHAP (SHapley Additive exPlanations)** was used to quantify the contribution of individual features to price predictions in the XGBoost model.

- Grad-CAM explains where the CNN is looking, not the final price prediction, thereby providing interpretability for visual feature learning.
- It was applied to the final convolutional layer (conv5_block3_out) to generate spatial heatmaps highlighting regions that contributed most to the model's output activations. The resulting heatmaps were overlaid on the original satellite images to visually interpret model attention.
- Grad-CAM analysis was conducted across low-priced, mid-priced, and high-priced properties to understand how environmental context varies across different valuation ranges.
- For low-priced properties, the CNN primarily attends to dense infrastructure and constrained layouts, reflecting lower environmental amenity and limited private space. Attention is concentrated around roads, dense built-up areas, and limited open spaces and minimal focus on greenery or landscaped regions.
- For high-priced properties, Grad-CAM consistently highlights environmental amenities such as greenery, open spaces, and waterfront proximity, validating the importance of visual context in premium valuation.

Low-Priced Properties:
Dense development and proximity to roads are associated with lower perceived neighborhood quality and reduced valuation.

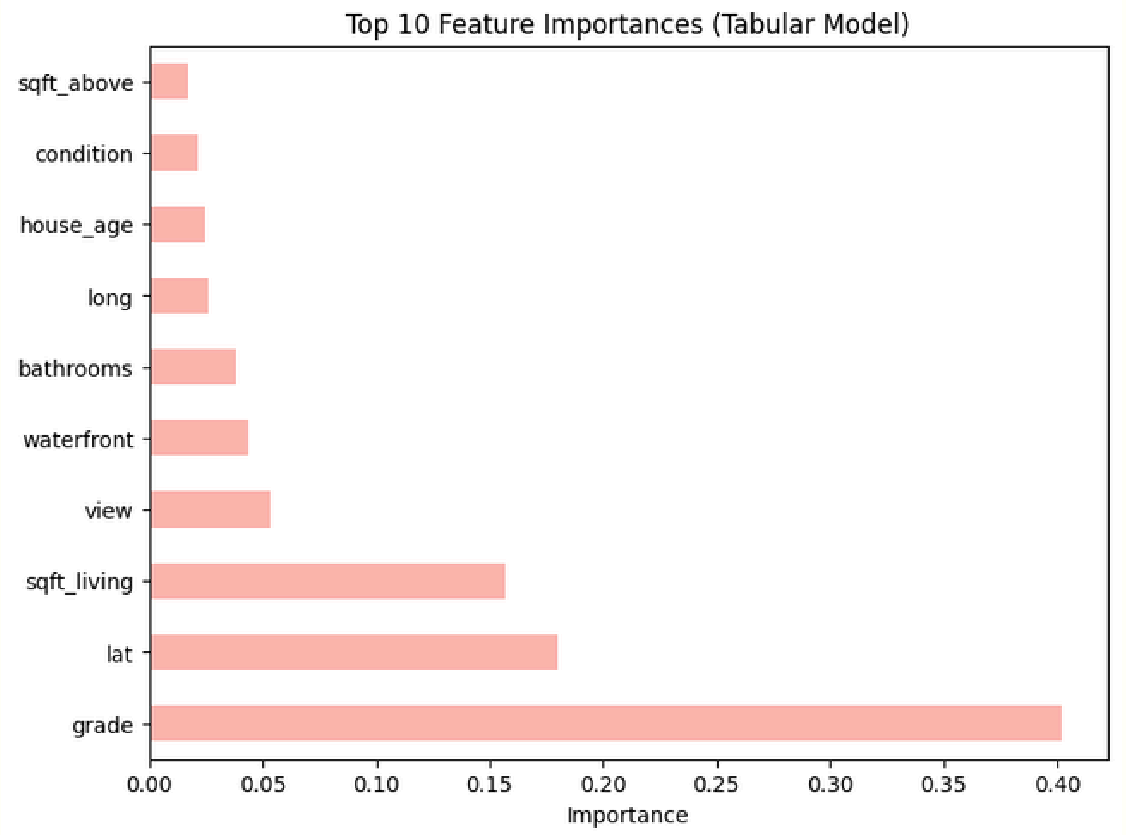
High-Priced Properties:
Environmental quality and exclusivity are key drivers of high-end property prices beyond structural attributes alone.



Grad-CAM: Visual Explainability for Satellite Images

MODEL EXPLAINABILITY

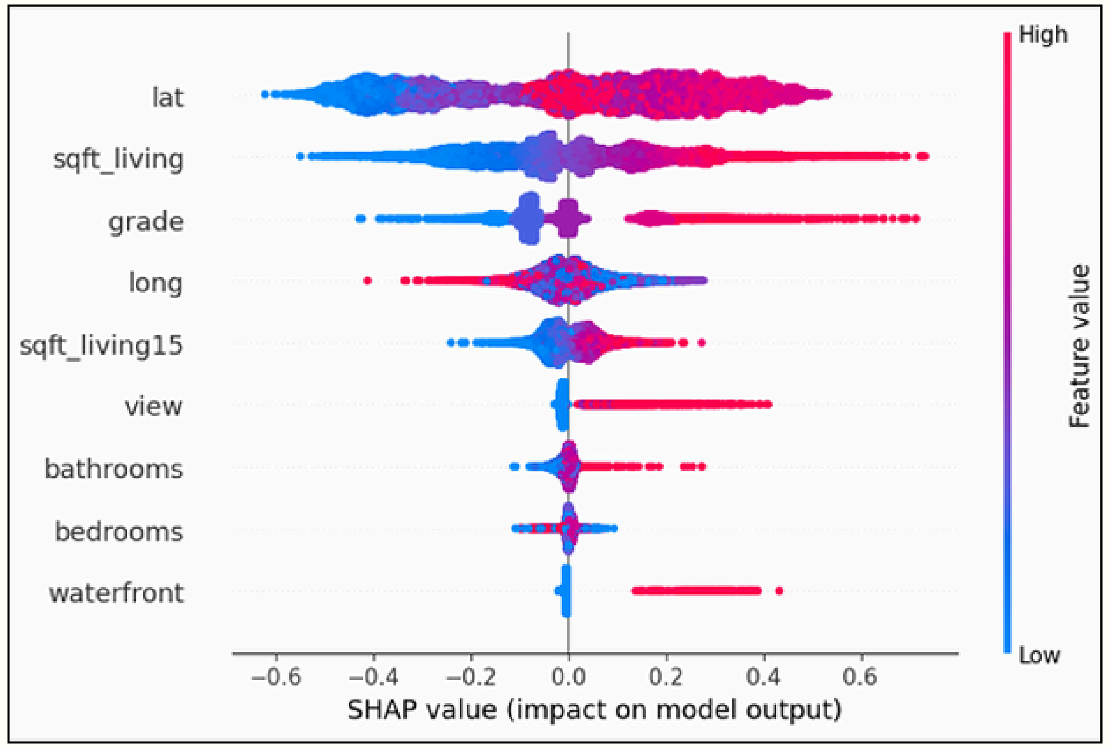
SHAP Explainability for Tabular Model



SHAP values were computed for the tabular-only XGBoost model to quantify the marginal contribution of each feature to the predicted log-price. This enabled both global feature importance ranking and directional interpretation of feature effects.

Key **Observations:**

- grade, sqft_living, and latitude emerge as the most influential predictors
- Higher values of living area, construction quality, and favorable location increase predicted prices
- Features such as waterfront, bathrooms, and neighborhood size contribute secondary but meaningful uplift.



KEY FINANCIAL & ECONOMIC INSIGHTS

- Environmental features such as green cover and water access act as long-term value stabilizers, depreciating more slowly than physical structures and offering greater price resilience for real estate investments.
- Waterfront access and high view ratings command durable price premiums, as these non-replicable amenities are directly capitalized into property values through strong buyer willingness to pay.
- Green surroundings enhance property value only in conjunction with favorable neighborhood and infrastructure conditions, highlighting the complementary nature of environmental and locational attributes in valuation.
- Proximity to dense road networks and infrastructure imposes negative externalities, where noise, congestion, and reduced privacy suppress residential value despite improved accessibility.
- High-end property markets exhibit significantly greater price dispersion, reflecting heterogeneous buyer preferences and necessitating individualized valuation approaches rather than uniform pricing heuristics.
- Multimodal fusion models closely align with **hedonic pricing theory**, capturing how buyers evaluate properties as bundles of structural, locational, and environmental attributes, even when aggregate predictive gains are modest.

CONCLUSION

End-to-End Multimodal Framework

This project demonstrates an end-to-end multimodal property valuation framework that integrates structured housing attributes with satellite-derived environmental context. Parallel pipelines process tabular data and satellite images, which are subsequently fused to model property prices in a unified framework.

Visual Feature Learning through CNNs

Satellite images acquired via the Google Maps Static API are processed using a pretrained and frozen ResNet50 model to extract high-level visual embeddings representing neighborhood characteristics such as green cover, layout, and proximity to water. These embeddings capture environmental signals not explicitly available in tabular data.

Multimodal Fusion and Predictive Performance

Tabular and visual features are combined using an early fusion strategy and modeled with an XGBoost regressor. While the tabular-only model achieves the highest standalone accuracy, the multimodal fusion model delivers comparable performance, demonstrating that satellite imagery provides complementary contextual information.

Model Explainability and Economic Validity

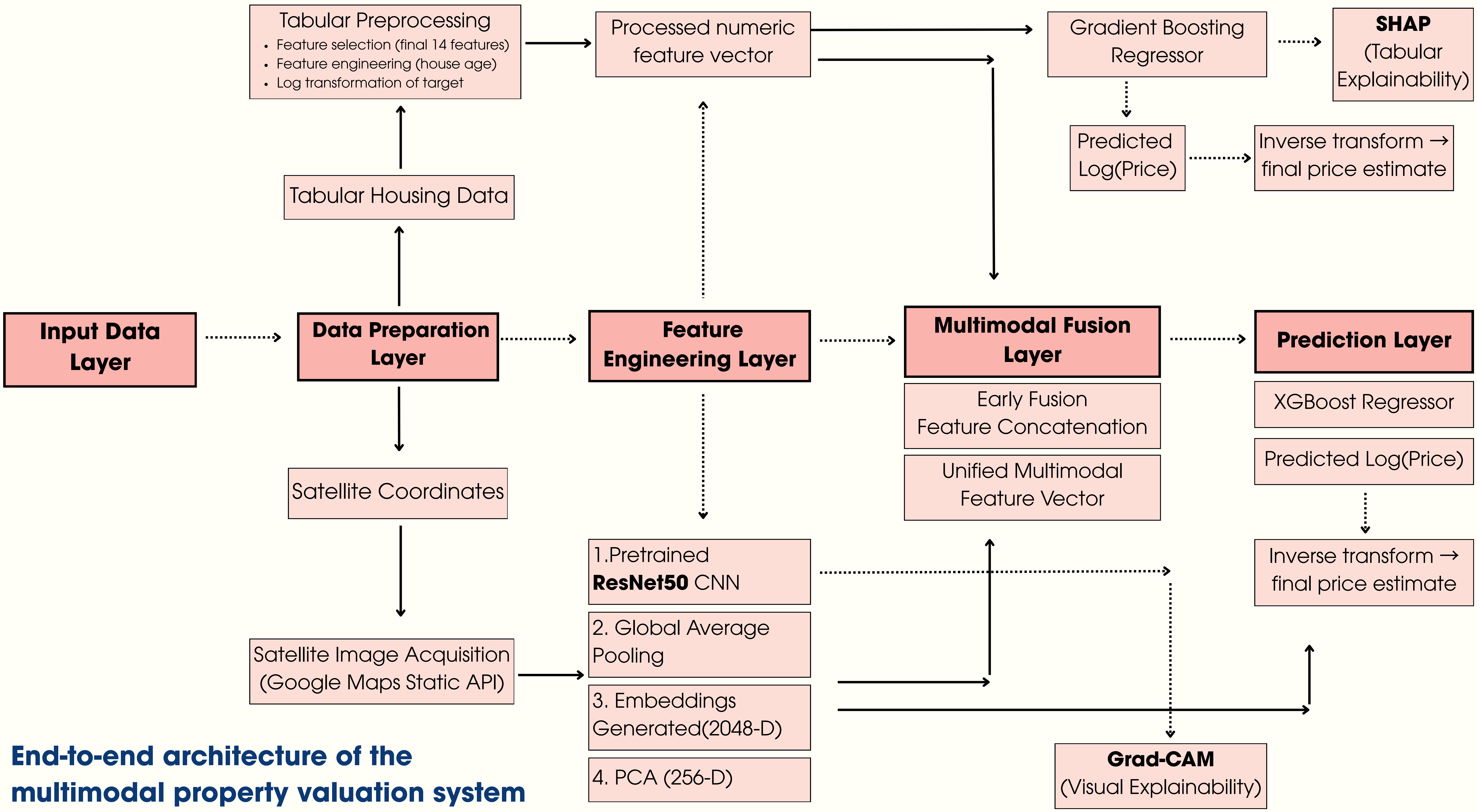
Explainability analysis using Grad-CAM and SHAP reveals that the model focuses on economically meaningful factors. Grad-CAM highlights environmentally salient regions in satellite images, while SHAP confirms the dominance of structural features such as construction grade and living area in pricing decisions.

Implications for Real Estate Valuation

The results indicate that effective property valuation requires both quantitative fundamentals and qualitative neighborhood context. The proposed framework aligns with hedonic pricing theory by treating properties as bundles of structural, locational, and environmental attributes.

FUTURE SCOPE:

Future extensions of this work could include fine-tuning the CNN on real estate–specific imagery, incorporating temporal satellite data to capture neighborhood evolution, and integrating higher-resolution or street-level imagery. Additionally, experimenting with late fusion architectures and end-to-end multimodal learning may further enhance predictive performance and interpretability.



End-to-end architecture of the multimodal property valuation system

THANK YOU