Satellite Imagery-Based Property Valuation

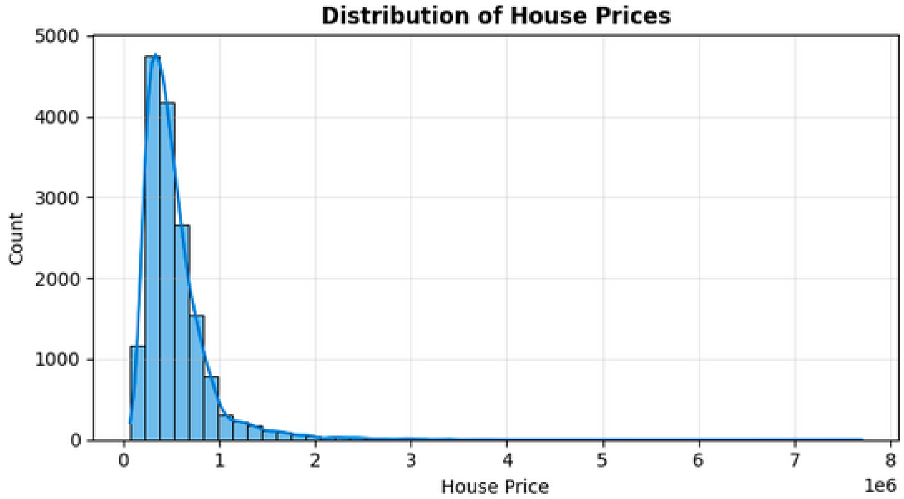
CDC X Yhills OPEN PROJECTS 2025-2026 - Data Science Track

Swastik Panda

23321030

Exploratory Data Analysis

Price Distribution

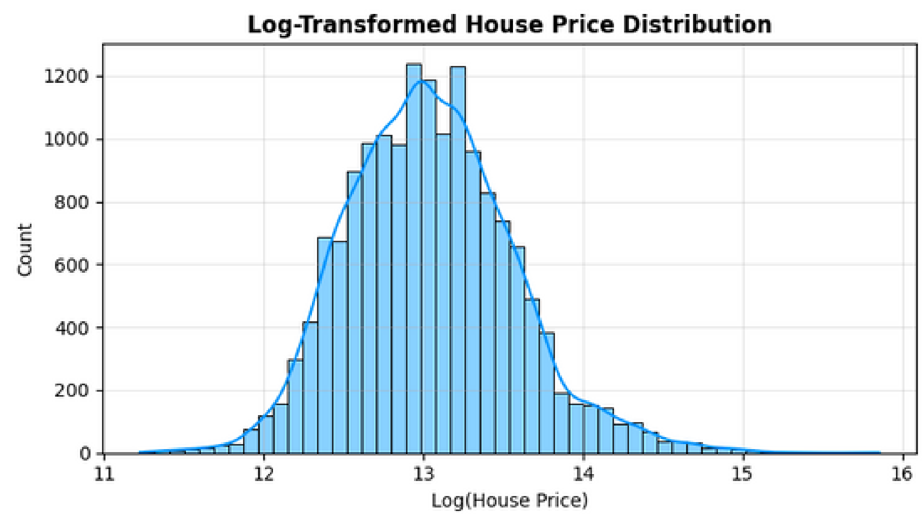


Raw House Prices (Histogram) Observation: The distribution of raw prices is heavily right-skewed, with a long tail representing high-value luxury properties.

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.

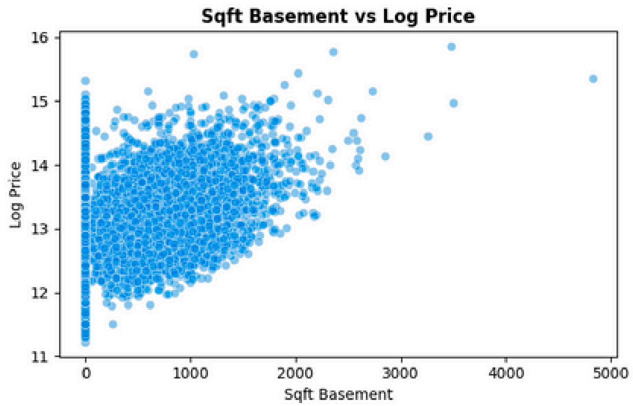
Insight: This transformation stabilizes the variance and reduces the impact of extreme outliers, ensuring better convergence and performance for predictive models.



Feature Analysis - Area

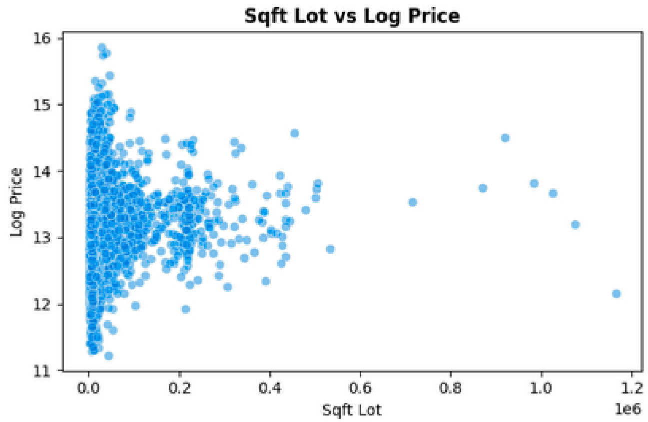
There is a strong positive correlation between living area and the log-transformed house price, though the variance in price increases as square footage grows.

The scatter plot shows a strong 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.

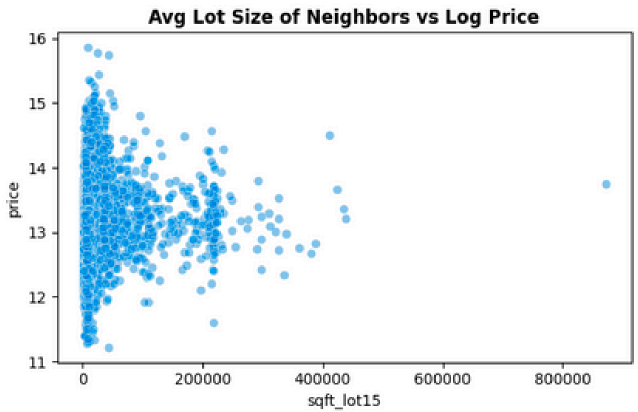


The scatter plot shows a moderate positive correlation between the square footage of the basement and the log price, suggesting that while a basement adds value, it is a less significant predictor than above-ground space.

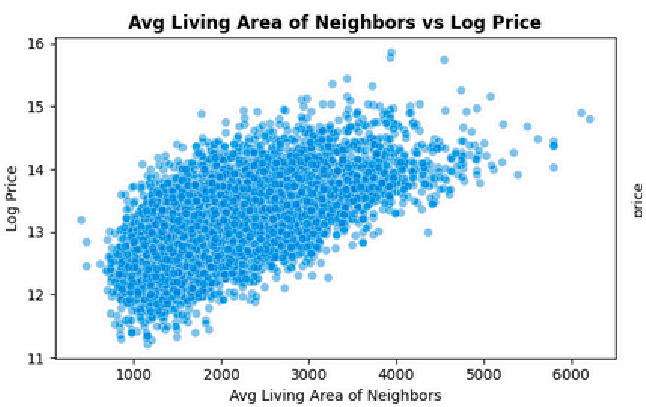
Feature Analysis - Area



The scatter plot shows a weak correlation between lot size and log price, indicating that property value is not very dependent on lot square footage than on interior living space.



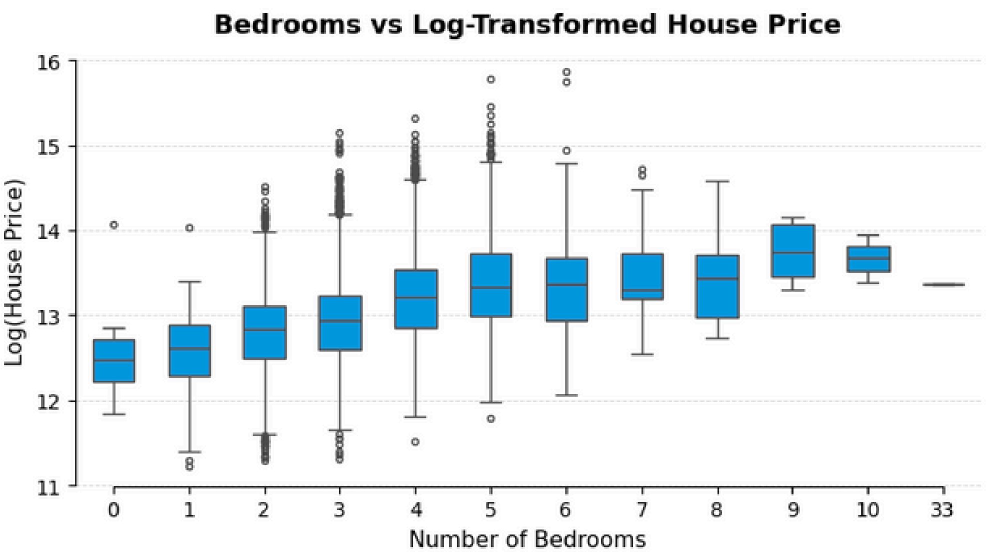
The scatter plot shows a weak correlation between the average lot size of the nearest fifteen neighbors and the log price, which was to be expected after the preceding graph

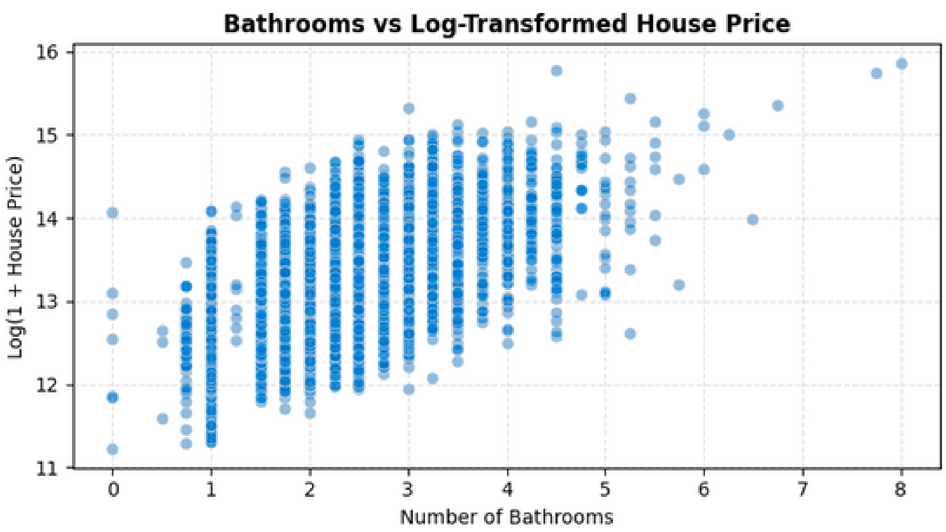


The scatter plot shows a moderate positive correlation between the average living area of neighbors and the log price, indicating that property values tend to be higher in neighborhoods with larger homes

Feature Analysis - Bedrooms & Bathrooms

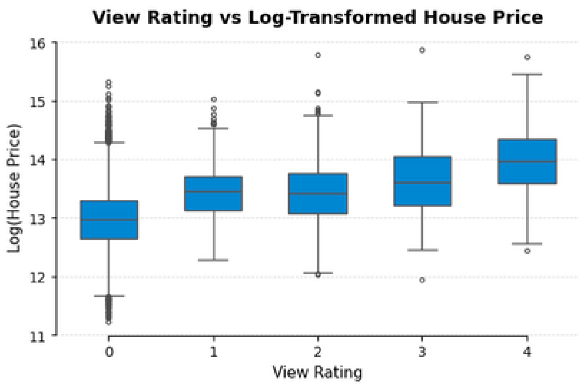
The box plot reveals that median house prices increase steadily with the number of bedrooms up to approximately five, after which the relationship plateaus and price variance increases significantly



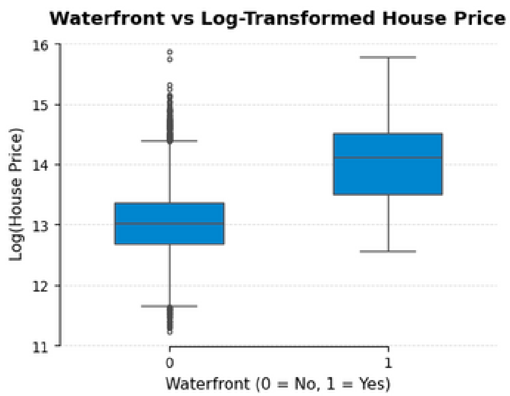


There is a clear upward trend between the number of bathrooms and house price, with the most significant price gains occurring as properties move from single to multi-bathroom configurations.

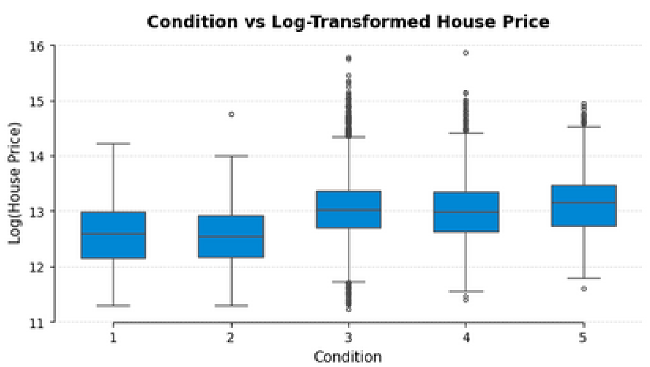
Feature Analysis - View Rating + Waterfront



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.

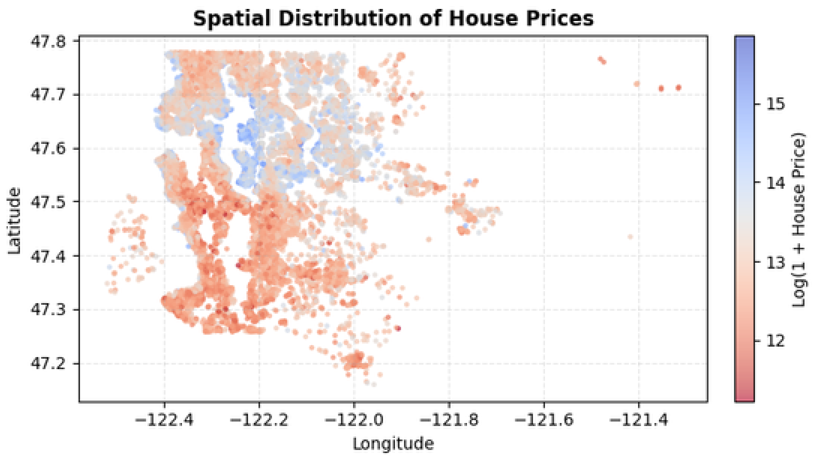


Waterfront properties command a substantial price premium, with median values significantly higher than non-waterfront homes despite a much smaller inventory of such listings

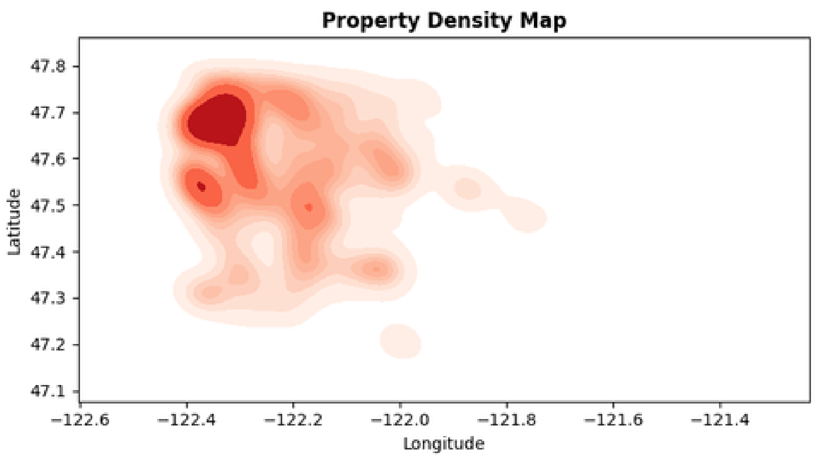


The box plot illustrates that house prices generally improve as the overall condition of the property increases from 1 to 5, with the most notable jump in median value occurring between properties in fair and good condition.

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.

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.

Visual Data

Satellite Image Acquisition

To facilitate the multi-modal analysis, satellite imagery for each property was downloaded at three distinct zoom levels i.e 16x, 17x and 18x to capture varying degrees of contextual detail. After evaluating the trade-offs between local structural features and neighborhood environment, zoom level 17 was selected as the primary input for the analysis.

To capture the complex spatial features of the property locations, I utilized ConvNeXt Tiny, a streamlined version of the modernized convolutional architecture that leverages the benefits of both Swin transformer and Resnet.

Despite its smaller parameter footprint, the 'Tiny' variant maintains the advanced design elements of the full model such as 7x7 depth-wise convolutions and inverted bottlenecks - making it exceptionally efficient for extracting high-dimensional environmental features from satellite tiles without prohibitive computational costs.

Feature Extraction Methods + Gradcam

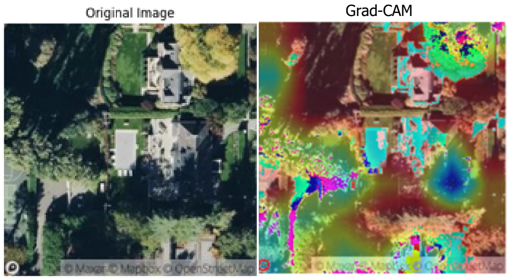
Approach 1: Baseline Global Feature Extraction
This method treats the pre-trained ConvNeXt Tiny purely as a static feature generator. By removing the classification head and applying Global Average Pooling, the entire image is summarized into a single 768-dimensional vector.

Goal: To capture the high-level semantic "gist" of the satellite image (e.g., general urban density) using generic features learned from ImageNet, without task-specific fine-tuning.

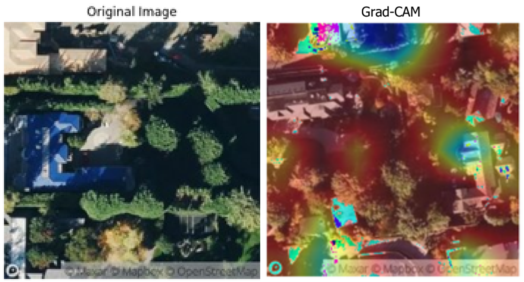
Reason of Rejection: While the extracted embeddings were archived for potential downstream tasks, the Grad-CAM visualizations revealed a critical lack of focus.
Although the model occasionally identified relevant environmental factors like road networks and vegetation cover, it frequently attended to distorted artifacts and high-frequency noise.
It also highlighted all or none of the image sections as important.

This indicates that with only high level features , the pre-trained weights remain optimized for generic object classification rather than identifying the specific visual cues that drive real estate value.

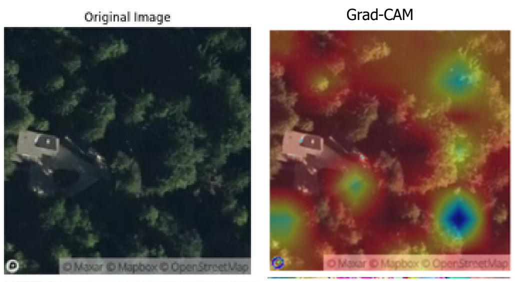
GRADCAM IMPORTANCE COLOUR MAP



Most expensive property



2nd Most expensive property



Least expensive property

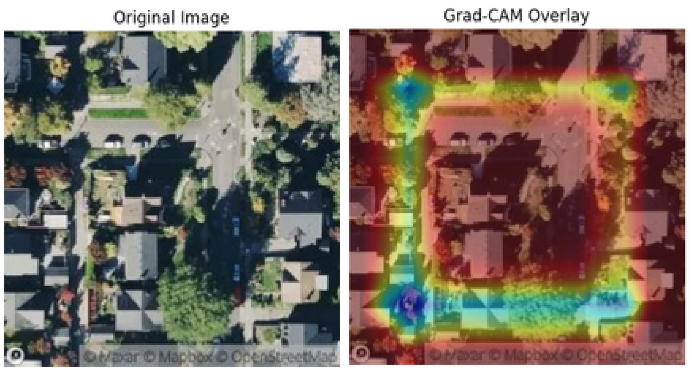
2nd Least expensive property

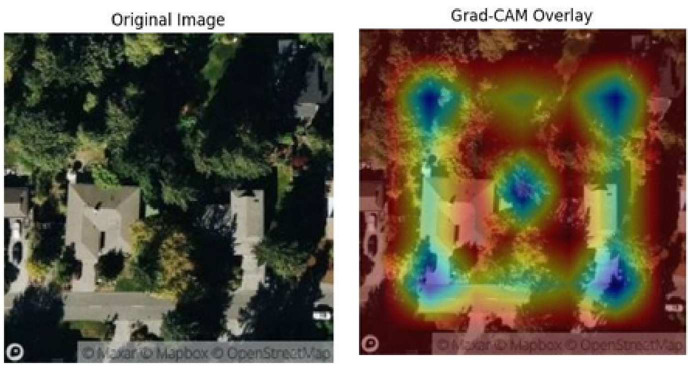
Feature Extraction Methods + Gradcam

Approach 2: End-to-End Visual Regression
Instead of using the network as a static feature generator, this approach appended a regression head directly to the ConvNeXt backbone. The model was trained end-to-end to predict the house price solely from the satellite image.

Goal: To perform supervised representation learning. By forcing the model to minimize price error, we aimed to calibrate the weights to identify specific "high-value" visual markers (e.g., swimming pools, manicured lawns) rather than generic objects.

Reason for Rejection: Analysis of the Grad-CAM overlays revealed that the end-to-end regression model failed to learn semantically meaningful features. Instead of attending to specific value drivers such as the primary residence, driveway, or backyard—the network highlighted everything around the house in an arbitrary pattern.





GRADCAM IMPORTANCE COLOUR MAP

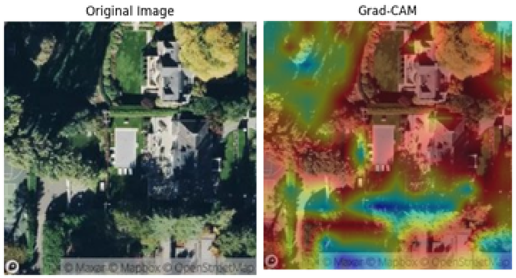
Feature Extraction Methods + Gradcam

Approach 3: Multi-Layer Feature Blending
This advanced strategy extracts feature maps from multiple depths of the network—specifically Stage 2 (mid-level features) and Stage 3 (deep semantic features). These maps are independently pooled, concatenated, and projected through a blending layer consisting of a Linear projection, Layer Normalization, and GELU activation.

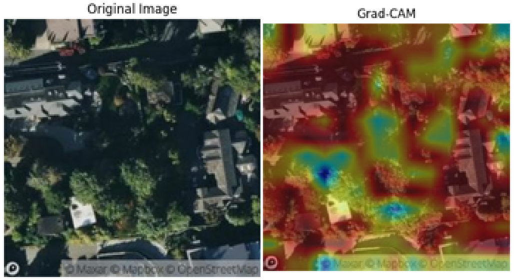
Goal: To capture a holistic view of the property. While deep layers understand broad semantic concepts like "neighborhood density," mid-level layers preserve finer details such as vegetation texture and road patterns. Blending them creates a richer, more robust representation.

Why it was selected: Unlike the regression or baseline model, the Grad-CAM visualizations for this approach demonstrated a stable and logical focus.
The model correctly highlighted key environmental value drivers such as vegetation density, sidewalks, and road proximity, confirming that it was learning to "see" the specific neighborhood characteristics that influence property prices rather than relying on background noise

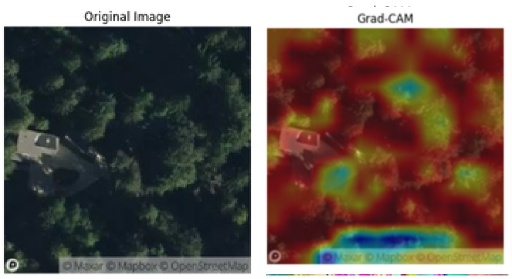
GRADCAM IMPORTANCE COLOUR MAP



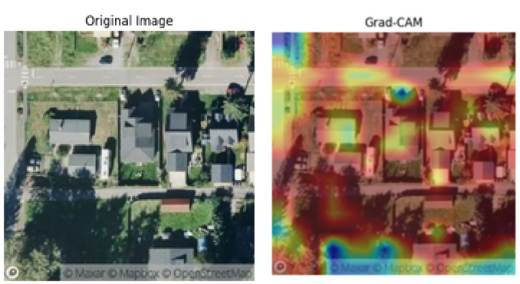
Most expensive property



2nd Most expensive property



Least expensive property



2nd Least expensive property

Models

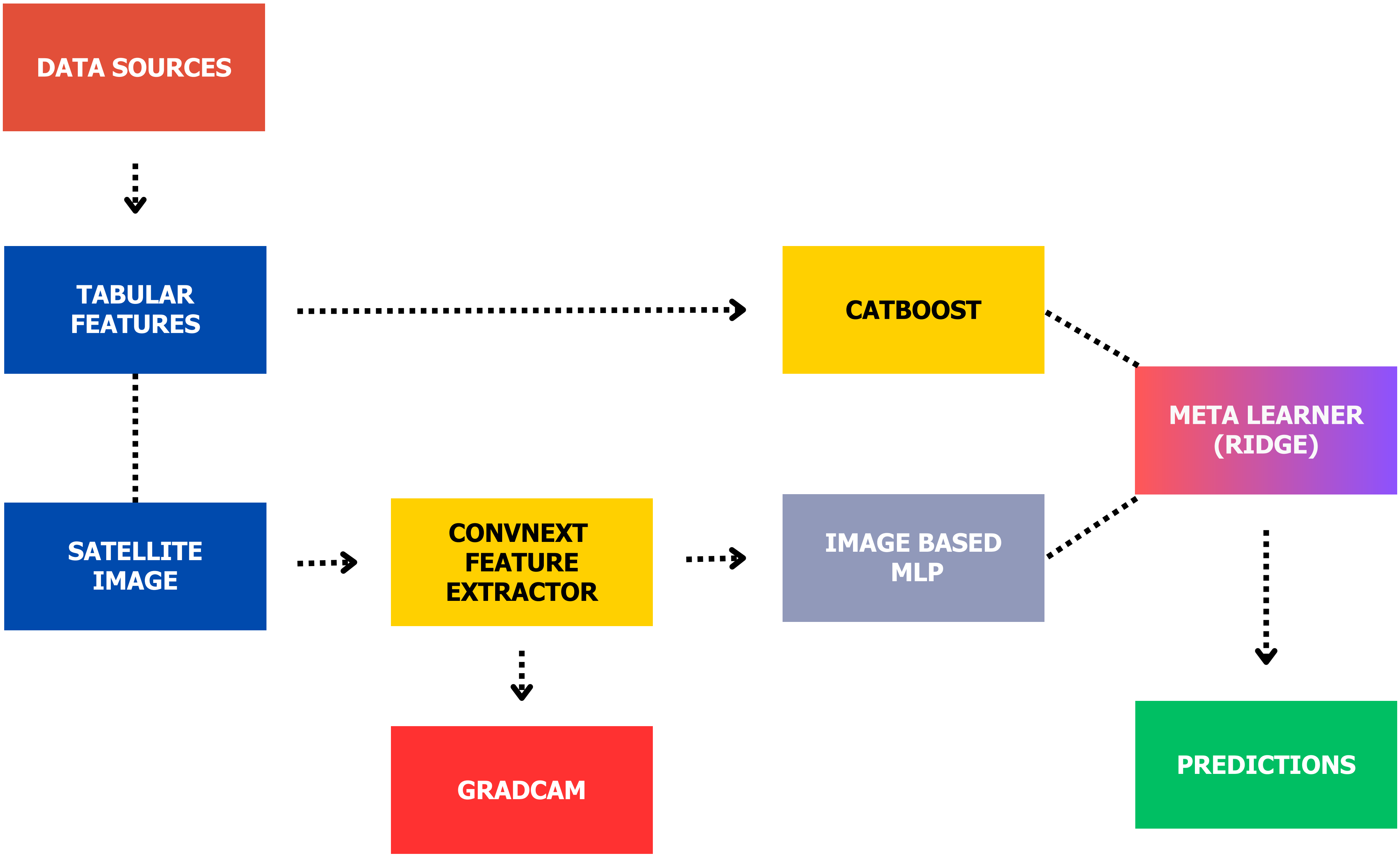
Tabular Only

Models	R Squared	log-RMSE
Linear Regression	**0.777**	0.248
Random Forest	**0.880**	0.181
LightGBM	**0.897**	-
Autogluon*	**0.913**	-
Catboost	**0.906**	0.160

We selected CatBoost as the final model for tabular data utilizing the complete multi-modal feature set. Although important features were identified, best results were found with the complete feature set.
Its superior performance is largely attributed to its specialized handling of categorical variables, allowing it to effectively leverage critical property attributes like 'grade', 'condition', and 'view', etc.

*While the AutoGluon pipeline achieved a marginally higher score, it relies on automated stacking and complex ensembling (AutoML). To ensure a fair evaluation of our custom satellite embeddings and maintain architectural transparency, we prioritized a single, highly effective gradient boosting model over a "black-box" automated pipeline.

Architecture of the Multimodal Meta model



Tabular + Image

Models	R Squared	RMSE
Multimodal CatBoost (PCA Embeddings, 90% Variance)	**0.900**	**0.1116 (log)**
Multimodal MLP based Regressor	**0.659**	**206753.67**
Meta Learner (Catboost + MLP)	**0.895**	**-116471.45**

Conclusion

We observed that the CatBoost model trained solely on tabular data delivered the optimal performance, achieving an exceptionally high R^2 of 0.906. This indicates that the tabular features are already intrinsically rich and capture the primary variance in house prices, rendering the satellite imagery largely redundant.

This conclusion is quantitatively confirmed by our ensemble metamodel (combining CatBoost with an image-based MLP), which yielded coefficients of [0.99988947, 0.01005528].
This heavily skewed weighting reveals that the tabular model contributes nearly all the predictive power, confirming that the visual data is unnecessary for this specific task.