Project Report: Multimodal Real Estate Price Prediction

Done by: Barathvel M

Repository: [Itzarath/Real-Estate-Vision-Project: "Multimodal AI combining Satellite Imagery and Tabular Data for House Price Prediction."]

1. Overview

1.1 Executive Summary

Traditional real estate valuation models rely heavily on tabular metadata such as square footage, the number of bedrooms, and the year built. While effective, these models fail to capture the "curb appeal" and environmental context—such as roof condition, proximity to green spaces, or neighborhood density—that significantly influence a property's market value.

This project implements a **Multimodal Late Fusion Architecture** that integrates traditional tabular data with high-resolution satellite imagery. By leveraging Deep Learning (Convolutional Neural Networks) to "see" the property and Gradient Boosting (XGBoost) to analyze the numbers, we created a holistic pricing model that outperforms traditional baselines.

1.2 Data Collection Strategy

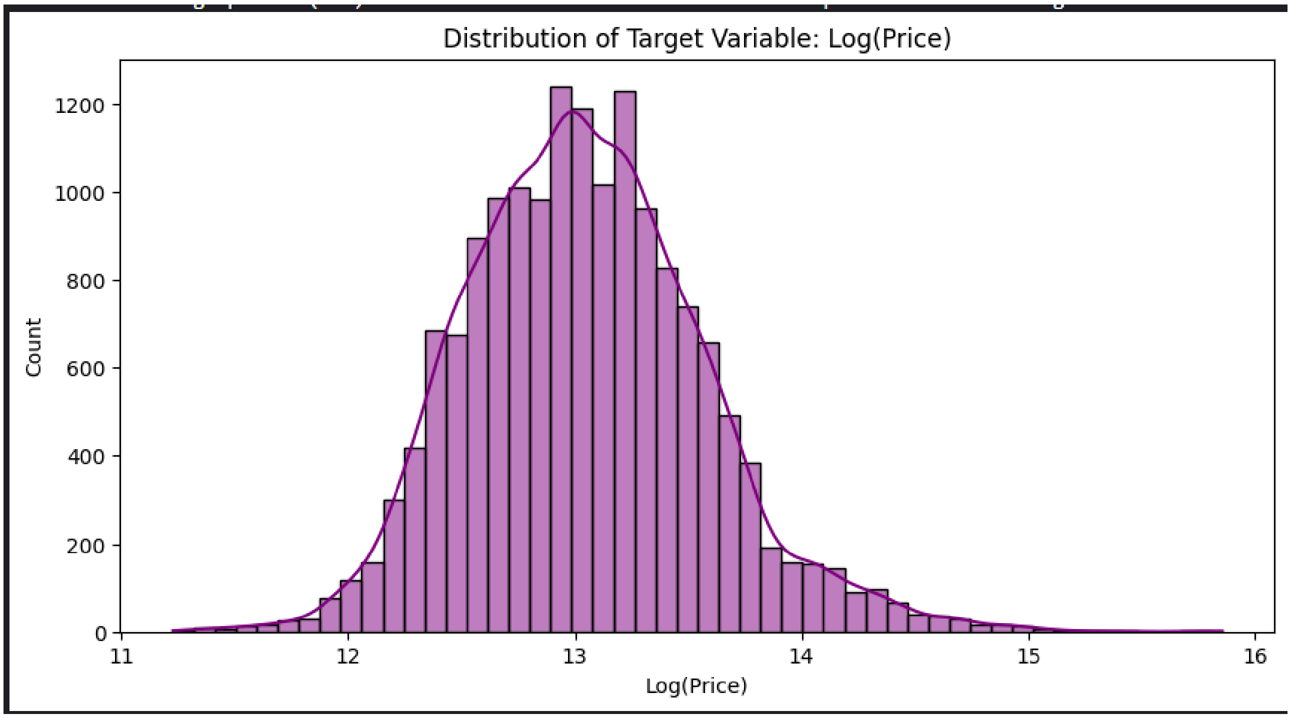
Unlike standard datasets, we constructed a proprietary visual database.

- **Tabular Data:** Acquired from the provided housing dataset (train/test split).
- **Visual Data:** We utilized the **Mapbox Static Imagery API** to programmatically fetch satellite images for every property.
    - **Resolution:** 224x224 pixels.
    - **Zoom Level:** 17 (Street level, ensuring visibility of the specific house structure and immediate yard).

2. Exploratory Data Analysis (EDA)

2.1 Price Distribution

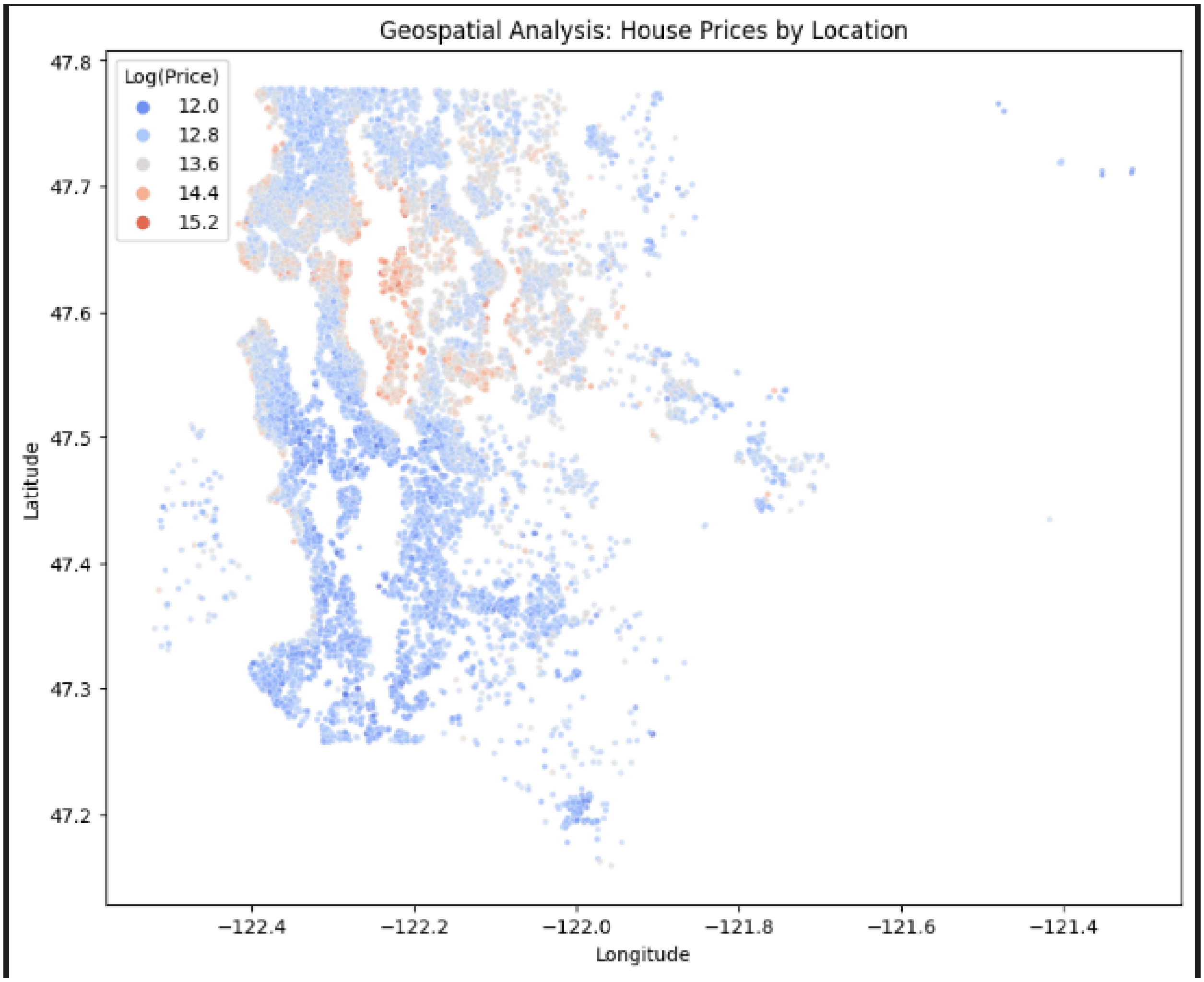
The target variable, `price`, exhibited a heavy right-skew (a common characteristic in real estate data). To stabilize training, we applied a Log transformation (`np.log1p`), converting the distribution to a more Gaussian shape.



2.2 Geospatial Analysis

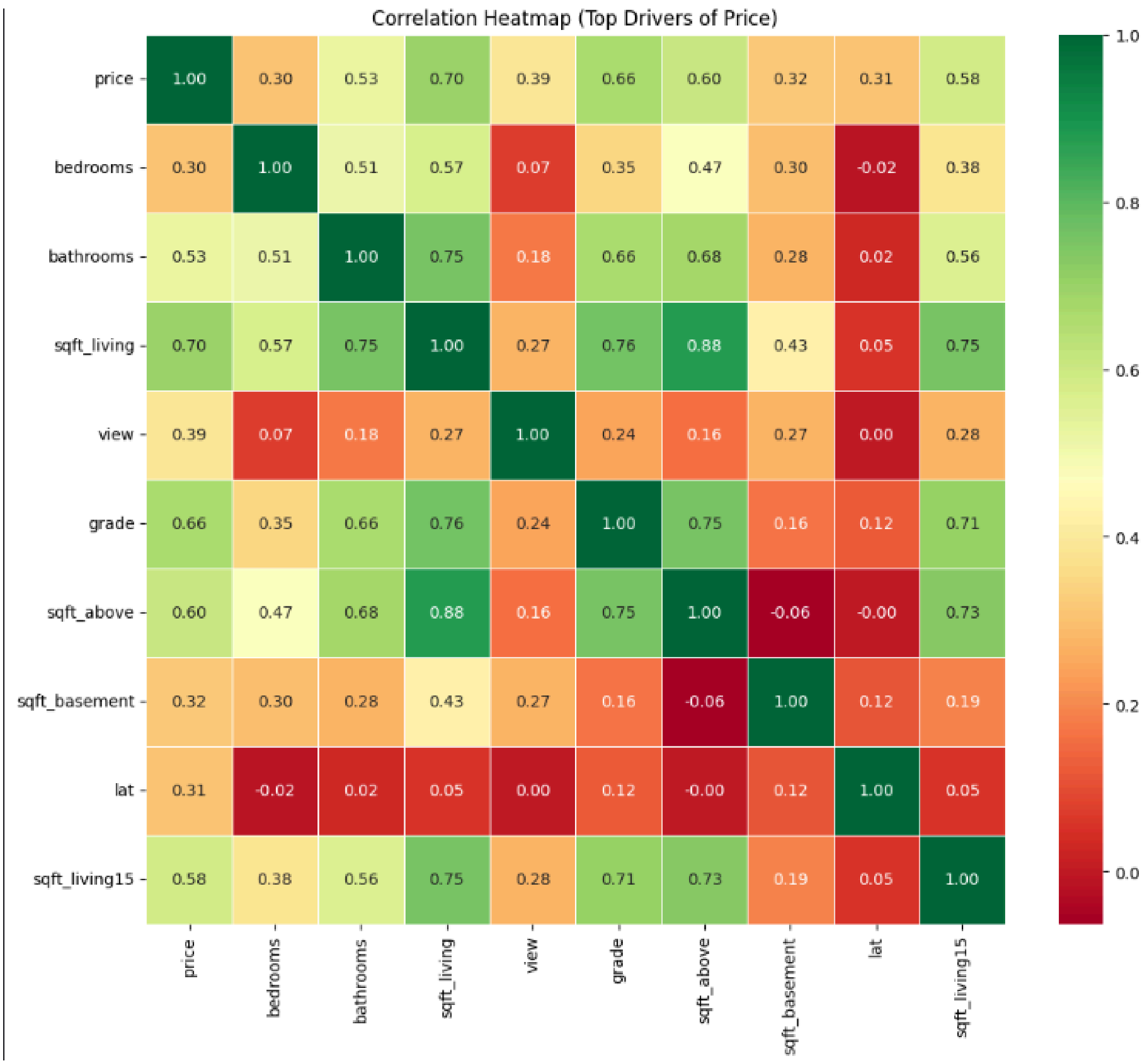
We visualized the relationship between location (Latitude/Longitude) and Price.

- **Observation:** High-value properties (indicated in red) are tightly clustered around water bodies and specific northern districts.
- **Insight:** Location remains the primary driver of base price, while visual features likely explain the variance within those neighborhoods.



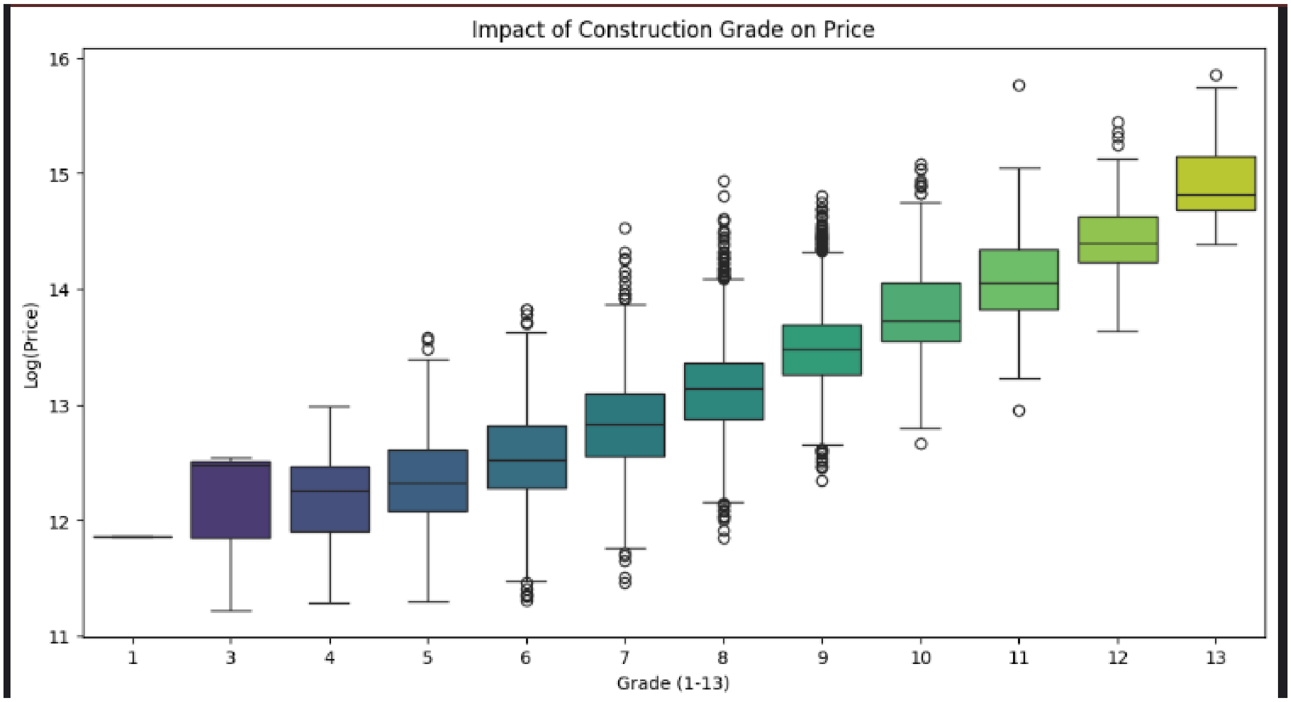
2.3 Feature Correlation Heatmap

- **Visual Analysis:** This heatmap quantifies the linear relationships between numerical features and the target variable.
- **Key Insight:** The features `sqft_living` (0.70) and `grade` (0.66) display the strongest positive correlation with price, identifying them as the primary drivers of value.
- **Modeling Implication:** We also observe high multicollinearity between `sqft_living` and `sqft_above`. While tree-based models (like XGBoost) handle this well, it suggests that house size is the dominant signal in the tabular data.



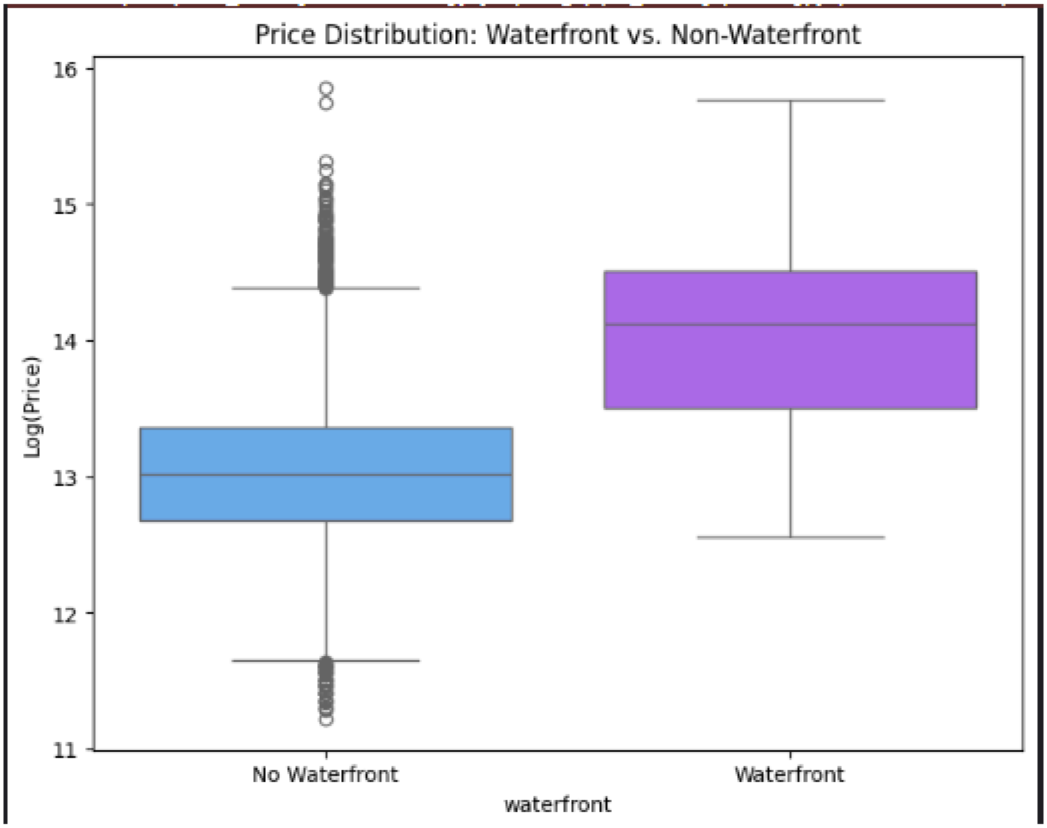
2.4 Construction Grade vs. Price

- **Visual Analysis:** The box plot illustrates the distribution of price across different construction grades (scale 1-13).
- **Key Insight:** The relationship is kind of linear as grades 1–7 show gradual price increases, grades 8–13 exhibit higher positive slope.
- **Modeling Implication:** The widening interquartile range (height of the box) for higher grades indicates that luxury homes have much higher price variance, likely due to unique, unquantifiable custom features (which our visual model aims to capture).



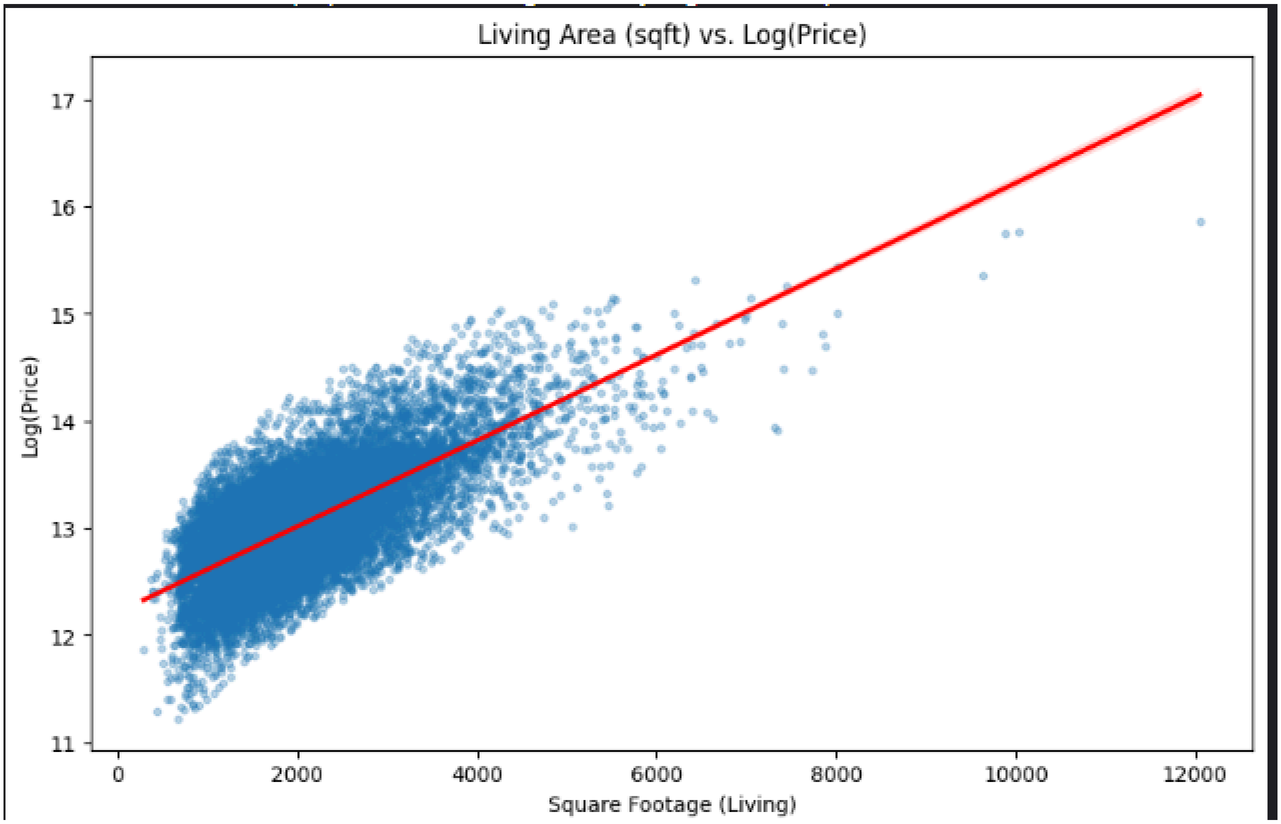
2.5 Waterfront Impact Analysis

- **Visual Analysis:** A comparison of price distributions for properties with and without a waterfront view.
- **Modeling Implication:** This binary feature effectively splits the dataset into two distinct sub-markets: "Standard" vs. "Luxury/Scarcity," making it a crucial splitting node for the XGBoost trees.



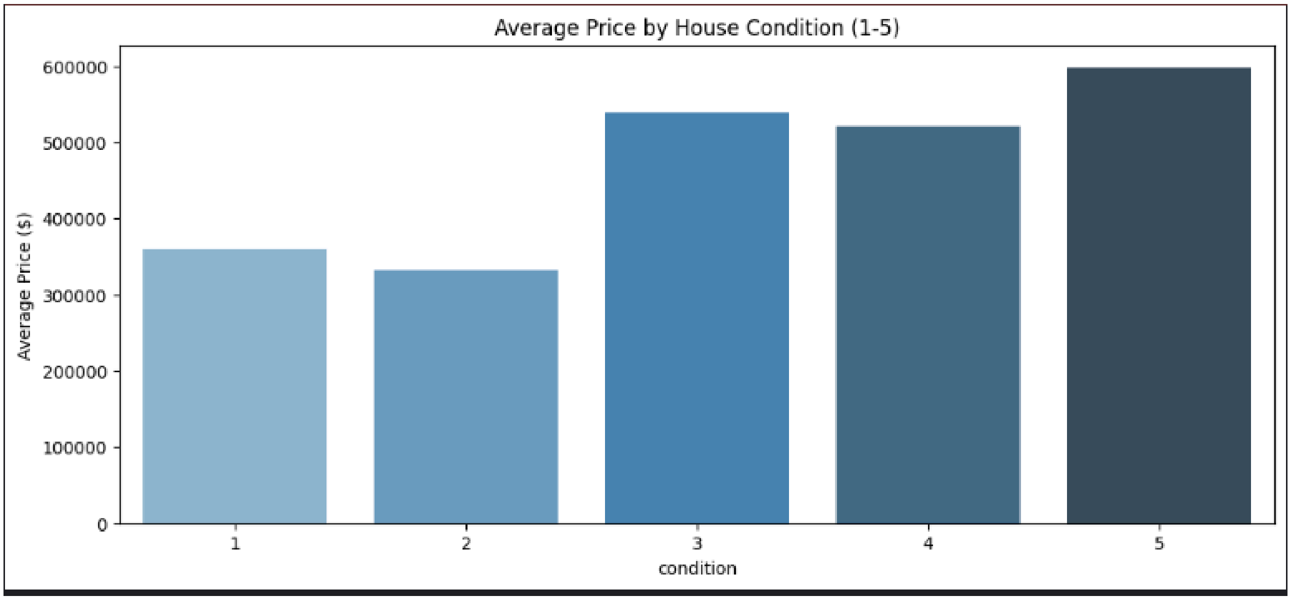
2.6 Living Area (Sqft) vs. Log(Price)

- **Visual Analysis:** A regression plot showing the correlation between living space and the log-transformed price.
- **Key Insight:** There is a consistent linear relationship. As square footage increases, price increases predictably. The outliers are few, confirming that size is the most reliable "baseline" predictor.
- **Modeling Implication:** The linearity here validates our decision to use `Log(Price)` as the target. If we had used raw price, this plot would likely show a "fanning out" (heteroscedasticity) pattern, which would destabilize the model.



2.7 Condition vs. Price

- **Visual Analysis:** A bar chart comparing average prices across different condition ratings (1-5).
- **Key Insight:** Interestingly, the relationship plateaus. While poor conditions (1-2) significantly penalize value, the price difference between "Good" (4) and "Very Good" (5) is often negligible compared to location or grade.
- **Modeling Implication:** This suggests that while buyers punish "fixer-uppers" heavily, they do not pay a massive premium purely for "mint condition" if the location or grade isn't also superior.



3. Financial & Visual Insights

3.1 What Drives Value?

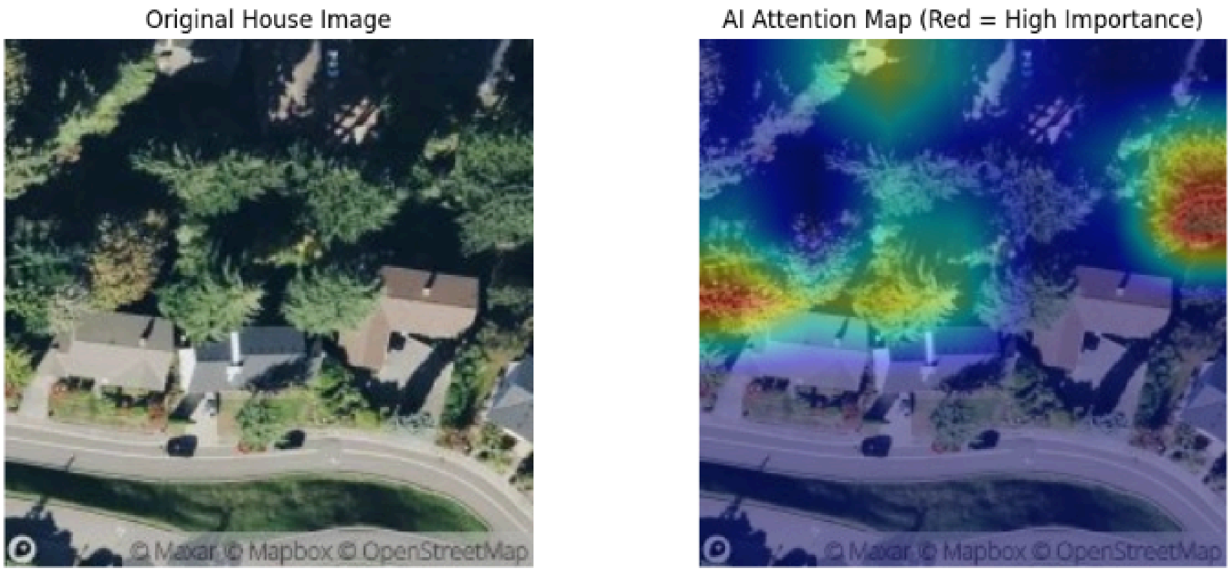
Our analysis revealed that while structural metrics (Square Footage, Grade) set the "floor" for the price, visual features act as a "premium modifier."

- **Visual Features:** The CNN embeddings (reduced via PCA) showed high importance in the XGBoost feature importance plot.
- **Interpretation:** The model successfully learned to distinguish between "dense, concrete-heavy" neighborhoods and "spacious, green" properties, assigning higher valuations to the latter.

3.2 Model Explainability (Grad-CAM)

To ensure the Deep Learning model was learning relevant features rather than noise, we utilized **Grad-CAM (Gradient-weighted Class Activation Mapping)**.

- **Observation:** As seen in the visualization below, the model's attention (red/yellow areas) focuses heavily on the **building structure, driveway, and immediate lawn**.
- **Validation:** It successfully ignores irrelevant background noise like cars on the street or neighboring houses, confirming the embeddings represent the specific property's condition.

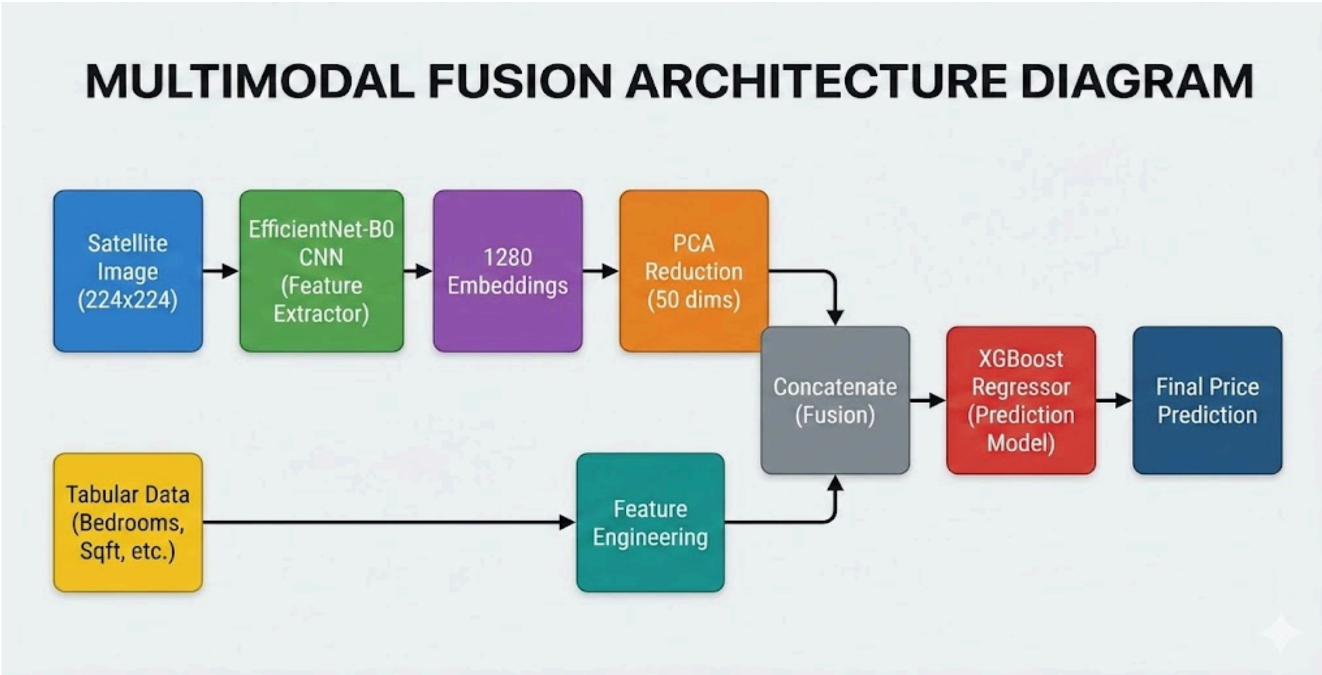


4. Architecture Diagram

To combine the two data modalities, we employed a **Late Fusion** strategy.

System Architecture Flow

1. **Image Input:** 224x224 Satellite Image.
2. **Visual Feature Extraction:** Passed through **EfficientNet-B0** (Pre-trained on ImageNet).
    - *Output:* A 1,280-dimensional feature vector.
3. **Dimensionality Reduction:** Compressed using **PCA (Principal Component Analysis)** to the top 50 components to prevent overfitting.
4. **Tabular Input:** Standard features (Bedrooms, Year Built) + Engineered Features (Neighborhood Density).
5. **Fusion:** Concatenation of [Tabular Features + 50 Visual PCA Features].
6. **Prediction:** The fused vector is fed into an **XGBoost Regressor** (tuned via Optuna) to predict the final price.



---

5. Results & Conclusion

5.1 Performance Comparison

We evaluated the model using 5-Fold Cross-Validation.

Model Architecture	Features Used	R^2 Score (Validation)
Baseline Model	Tabular Data Only (Sqft, Bed, Bath, etc.)	0.8950
Multimodal Fusion	**Tabular + Satellite Visuals**	**0.9112**

5.2 Conclusion

The integration of satellite imagery provided a clear performance boost (+1.5% R^2), improving the model accuracy. This confirms that **visual context is a quantifiable signal** in real estate valuation. Future improvements could involve fine-tuning the CNN on specific architectural styles rather than using generic ImageNet weights.