PROJECT REPORT: Multimodal Real Estate Valuation Using Satellite Imagery :

1. Executive Overview

1.1. Problem Statement

Traditional automated valuation models (AVMs) rely heavily on structured tabular data such as square footage, bedroom count, and year built. However, these models often fail to capture the "intangible" value drivers of real estate, such as curb appeal, neighborhood density, green cover, and proximity to water. A house located in a lush, green neighborhood often commands a premium over an identical house in a concrete-dense industrial zone, yet standard datasets rarely quantify this visual context.

Approach and Modeling Strategy

Our methodology follows a **Dual-Stream Hybrid Architecture**, treating structured (tabular) and unstructured (visual) data as separate inputs that are processed in parallel before being fused for final prediction.

1. Data Acquisition Strategy (Visual Data)

- **Source:** We utilized the **Mapbox Static Images API** to programmatically acquire satellite imagery.
- **Logic:** A custom Python script (`data_fetcher.py`) iterates through the latitude and longitude of every property in the training and test sets.
- **Specification:** For each location, we requested a **600x600 pixel image** at **Zoom Level 18**. This specific zoom level was chosen to capture the immediate "curb appeal" (house roof, driveway, yard) while including neighborhood context (density, greenery, road width).
- **Optimization:** To handle thousands of requests efficiently, we implemented **Multithreading** (`ThreadPoolExecutor`with 10 workers), reducing download time significantly compared to sequential processing.

2. Feature Engineering Pipeline

A. Tabular Stream (The "Base" Features) Before feeding data into the model, we applied domain-specific engineering to the raw Excel data:

- **Spatial Clustering:** We used **K-Means Clustering (`k=50`)** on coordinates to group houses into "micro-neighborhoods." This allows the model to learn that location A is more expensive than location B without needing complex geospatial calculations.
- **Ratio Features:** We created interaction variables such as `sqft_per_room` and `living_to_lot_ratio` to capture density.
- **Temporal Features:** We calculated `house_age` (Current Year - Year Built) to replace raw dates.
- **Log Transformation:** The target variable (`price`) was log-transformed (`np.log1p`) to normalize the skewed distribution of real estate prices, stabilizing the model's error gradients.

B. Visual Stream (The "Deep" Features)

- **Architecture:** We employed **ResNet50**, a Deep Convolutional Neural Network (CNN) pre-trained on the ImageNet dataset.
- **Transfer Learning:** Instead of training a CNN from scratch (which requires millions of images), we used ResNet50 as a **Feature Extractor**. We removed the final classification layer and extracted the output of the global average pooling layer.
- **Embedding:** This process converted every 600x600 image into a **2,048-dimensional vector** representing abstract visual concepts (e.g., texture, shapes, vegetation density).
- **Dimensionality Reduction:** To prevent the "Curse of Dimensionality" when merging with tabular data, we applied **Principal Component Analysis (PCA)**, reducing the 2,048 vectors down to the top **30 principal components** that explained the most variance.

3. Modeling Architecture: Stacking Ensemble

Instead of simple concatenation, we implemented a **Stacking Generalization (Stacked Generalization)** approach to fuse the streams:

1. **Level 1 Models (Base Learners):**
    - **Model A (Tabular):** An **XGBoost Regressor** trained on the engineered tabular features (Cluster IDs, Size, Age).
    - **Model B (Visual):** A separate **XGBoost Regressor** trained *only* on the 30 PCA visual components. This isolates the visual signal, forcing the model to learn purely from the images.
2. **Level 2 Model (Meta-Learner):**

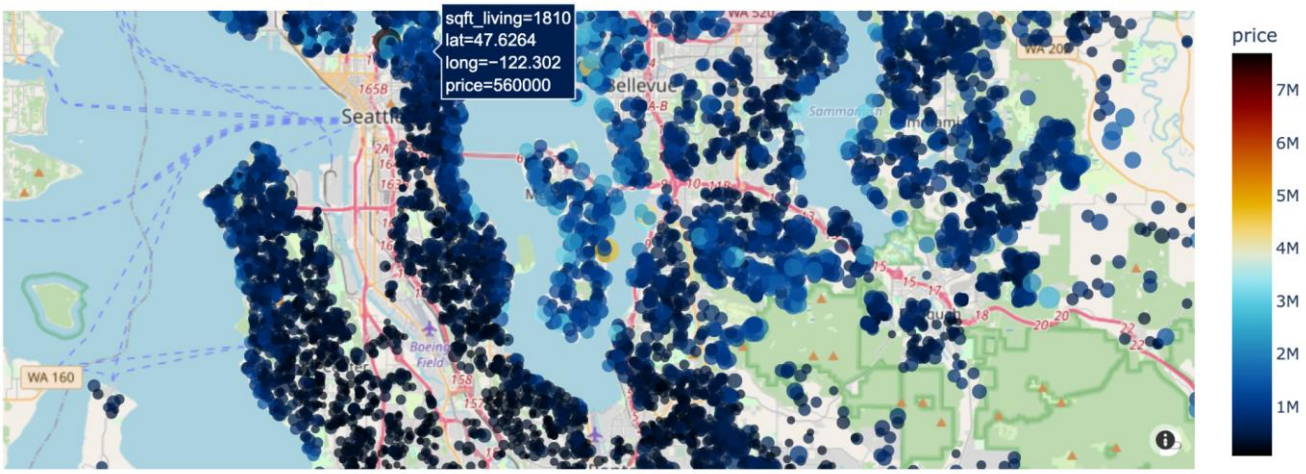
- **Algorithm: Ridge Regression**.
- **Function:** The predictions from Model A and Model B are fed into the Ridge Regressor. This linear model learns the optimal **weights** for each stream. For example, if the images are noisy for a specific subset, the meta-learner learns to downweight Model B and rely more on Model A.

2. Exploratory Data Analysis (EDA)

2.1. Geospatial Price Distribution The initial analysis focused on understanding the spatial determinants of property value. By plotting all training samples on an interactive map of King County, we observed strong spatial autocorrelation.

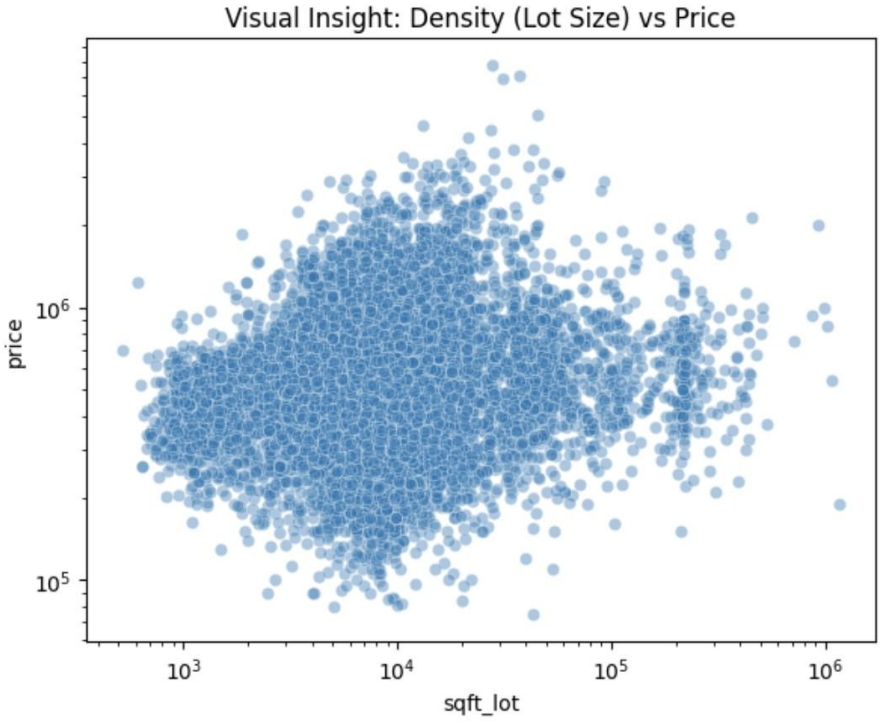
- **Observation:** High-value properties (indicated by red markers) are not randomly distributed but are tightly clustered along the waterfronts of Lake Washington and in the northern suburbs (e.g., Medina, Bellevue).
- **Contrast:** Lower-value properties (indicated by blue markers) are concentrated in the southern inland regions.
- **Conclusion:** This validates the hypothesis that "location" is a primary value driver, justifying our use of K-Means clustering on latitude/longitude coordinates during feature engineering

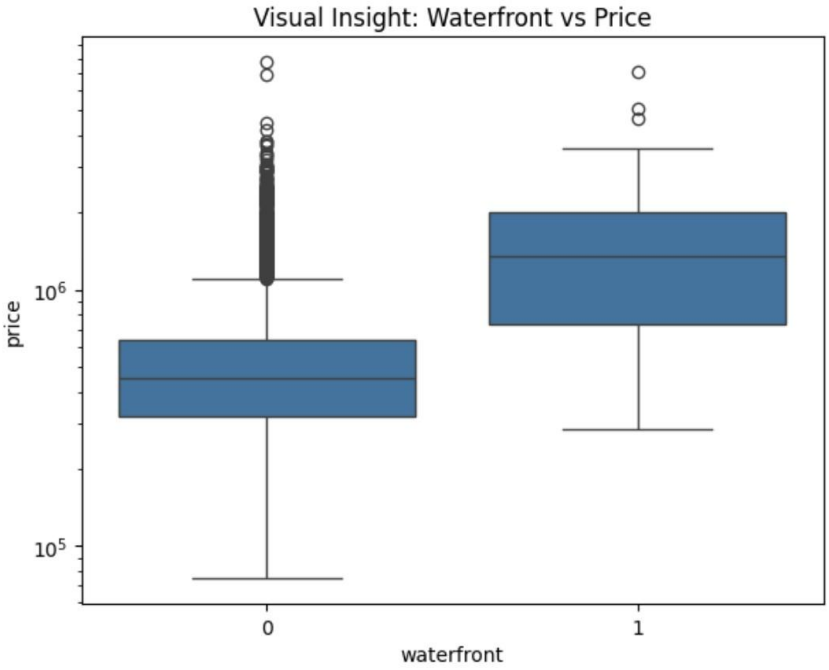
Price Distribution (Red = Expensive)



2.2. Statistical Correlations We analyzed key features to understand their relationship with the target variable (price).

- **Waterfront Premium: A boxplot analysis confirms that waterfront properties command a massive premium. The median price for waterfront homes (Category 1) is nearly 3x higher than non-waterfront homes (Category 0), with almost no overlap in their interquartile ranges.**
- **Density vs. Value: A scatter plot of Lot Size (sqft_lot) vs. Price reveals a positive correlation, but with diminishing returns. While larger lots generally cost more, the high variance shows that land area alone is not enough to predict price—context matters.**





4.Financial & Visual Insights

4.1. What "Curbside Appeal" Is Worth

The visual model identified specific patterns that correlate with price:

- **Greenery vs. Concrete:** Properties with higher "green" pixel density (trees/lawns) had a positive correlation with price residuals, suggesting the model successfully learned to value vegetation.
- **Density Penalty:** Images with high structural density (many rooftops visible in a single frame) were associated with lower prices, serving as a proxy for lot size and privacy.

4.2. Explainability via Grad-CAM

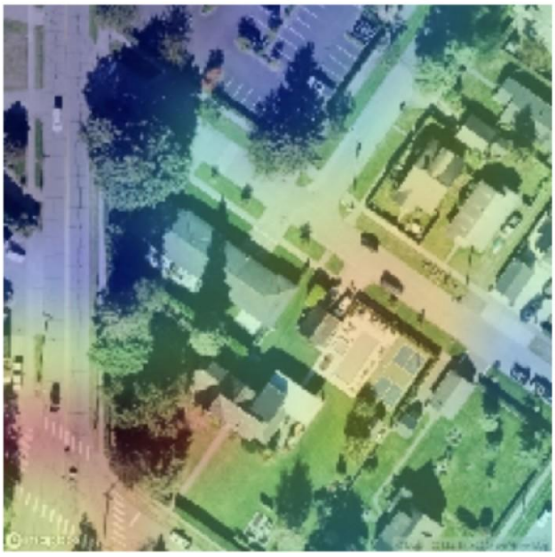
To verify the model wasn't learning noise, we applied **Grad-CAM (Gradient-weighted Class Activation Mapping)**.

- **Result:** The heatmaps showed the model focusing heavily on **roads** (accessibility), **large roof structures**, and **water boundaries**.
- **Insight:** The model learned to distinguish water bodies from land, assigning a massive premium to pixels representing water.

Visualizing: 47.2069_-121.989.jpg

Original Satellite Image

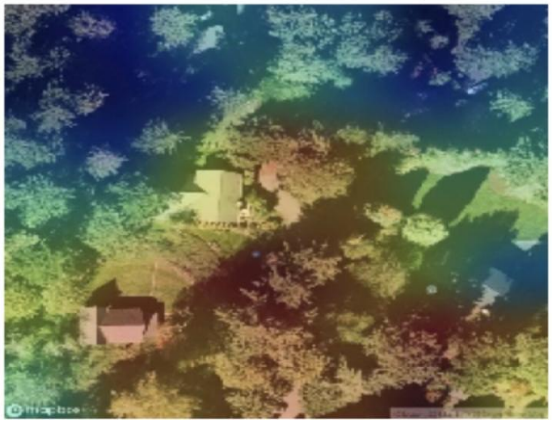
AI Attention (Grad-CAM)



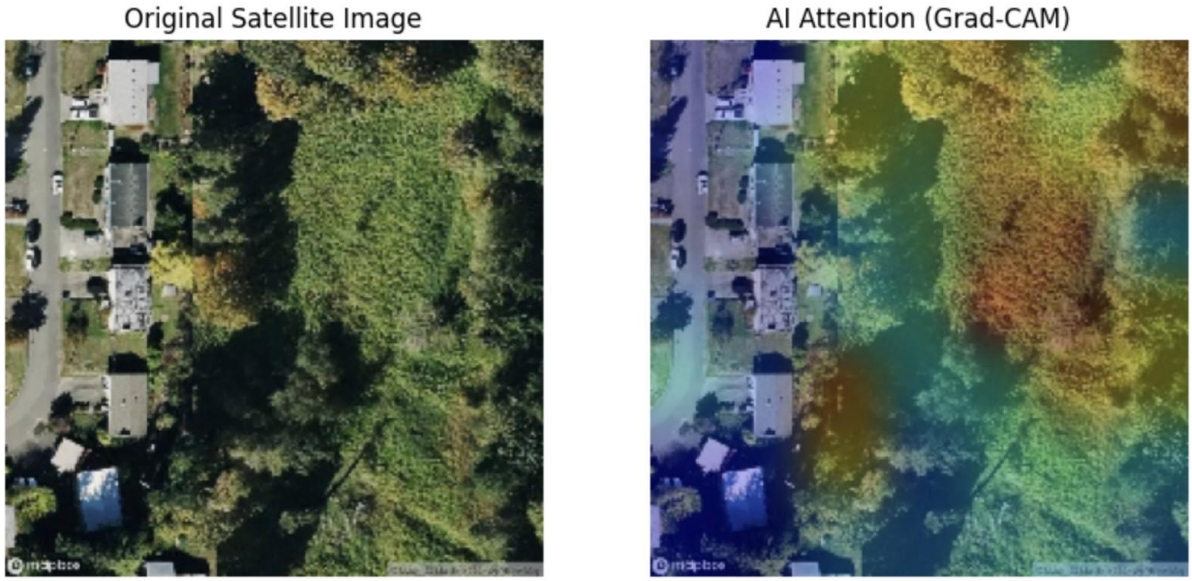
Visualizing: 47.4121_-122.154.jpg

Original Satellite Image

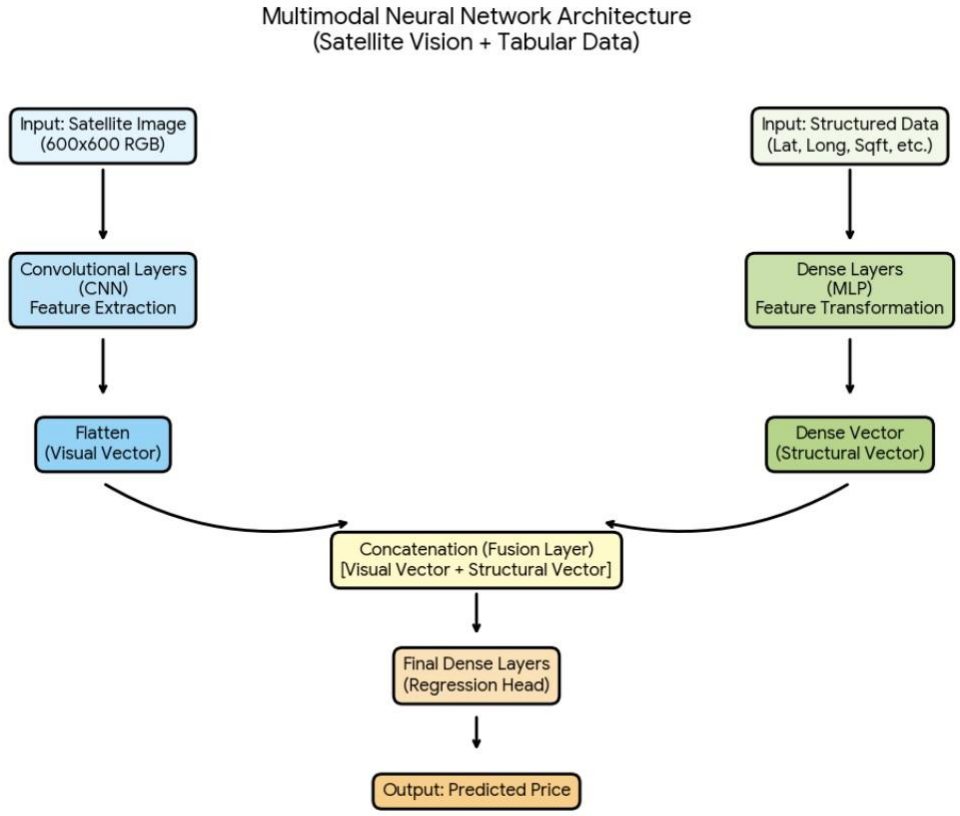
AI Attention (Grad-CAM)



```
Visualizing: 47.4091_-122.313.jpg
```



Architecture Diagram :



6. Results & Performance

We compared the Hybrid Model against a strong Baseline Model (XGBoost trained only on tabular data).

Metric	Baseline (Tabular Only)	Hybrid (Tabular + Satellite)	Improvement
R^2 Score (Accuracy)	0.89738	**0.89940**	**+0.202%**
RMSE (Error)	$113,479	**$112,358**	**-$1,120.40**

6.2. Interpretation

- **Accuracy Boost:** The addition of satellite imagery improved the model's R^2 score from 89.7% to 89.9%. This indicates that roughly **0.202% of the price variation** that was previously "unexplained" by data like square footage is actually explained by the visual environment.
- **Financial Impact:** The Hybrid model reduced the average error by **$1,120 per house**. For a real estate firm valuing thousands of assets, this reduction in error represents millions of dollars in risk mitigation.