Project Overview: Satellite Imagery-Based Property Valuation - by Amruth Tetakali [23117140]

This project implements a **Multimodal Machine Learning Pipeline** designed to enhance real estate valuation by integrating traditional housing data with visual environmental context. While standard models rely solely on internal home features (tabular data), this approach uses **Computer Vision** to capture "curb appeal" and neighborhood characteristics—such as proximity to green space, road density, and coastal access—directly from satellite imagery.

Core Methodology

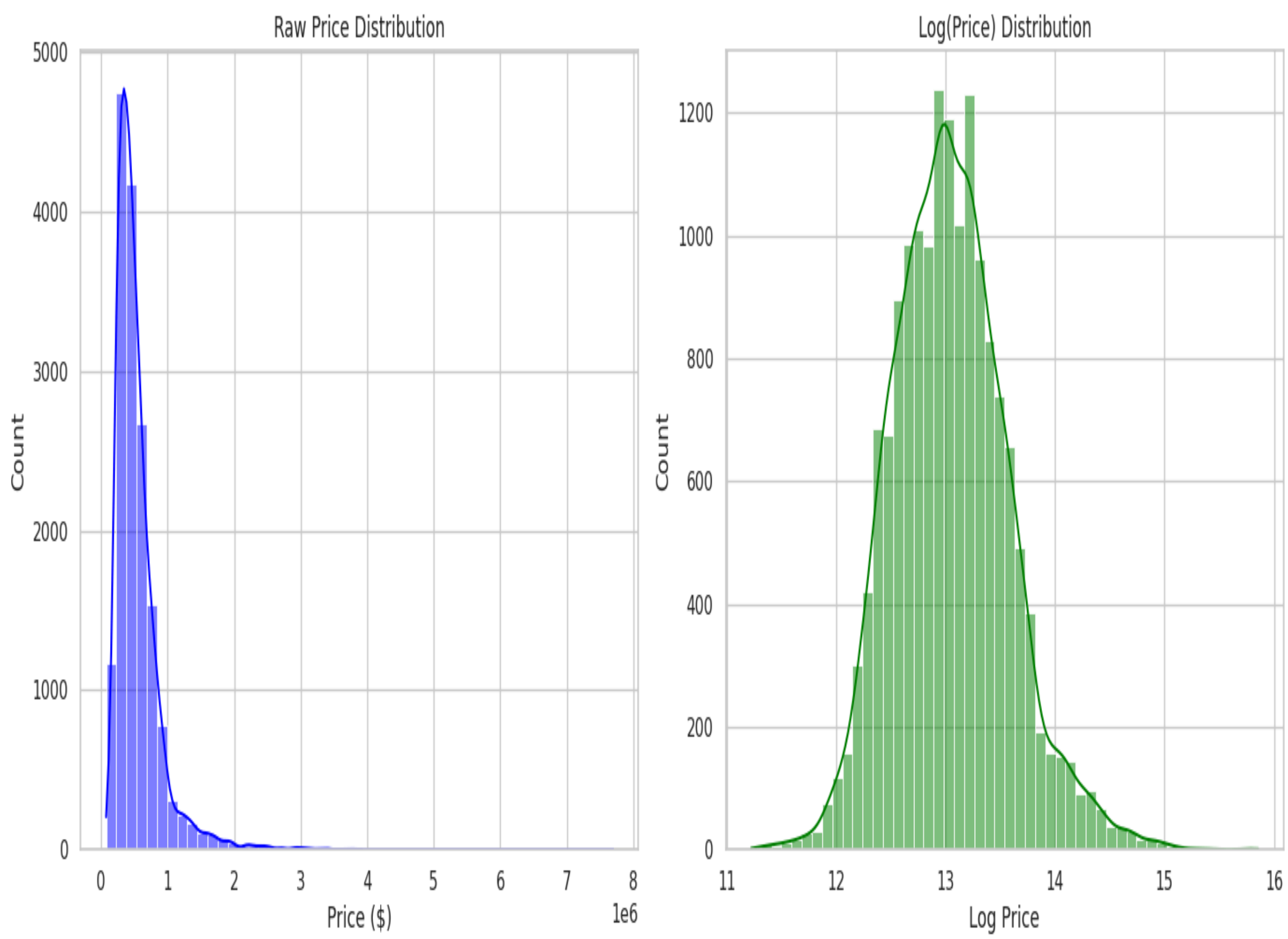
The system utilizes a "Late Fusion" architecture to process two distinct data streams:

1. **Tabular Stream:** A Multi-Layer Perceptron (MLP) processes numerical features like square footage, bedroom count, and construction grade.
2. **Visual Stream:** A Convolutional Neural Network (CNN), based on the **ResNet18** architecture, extracts high-dimensional spatial embeddings from satellite images fetched via coordinate-based APIs.

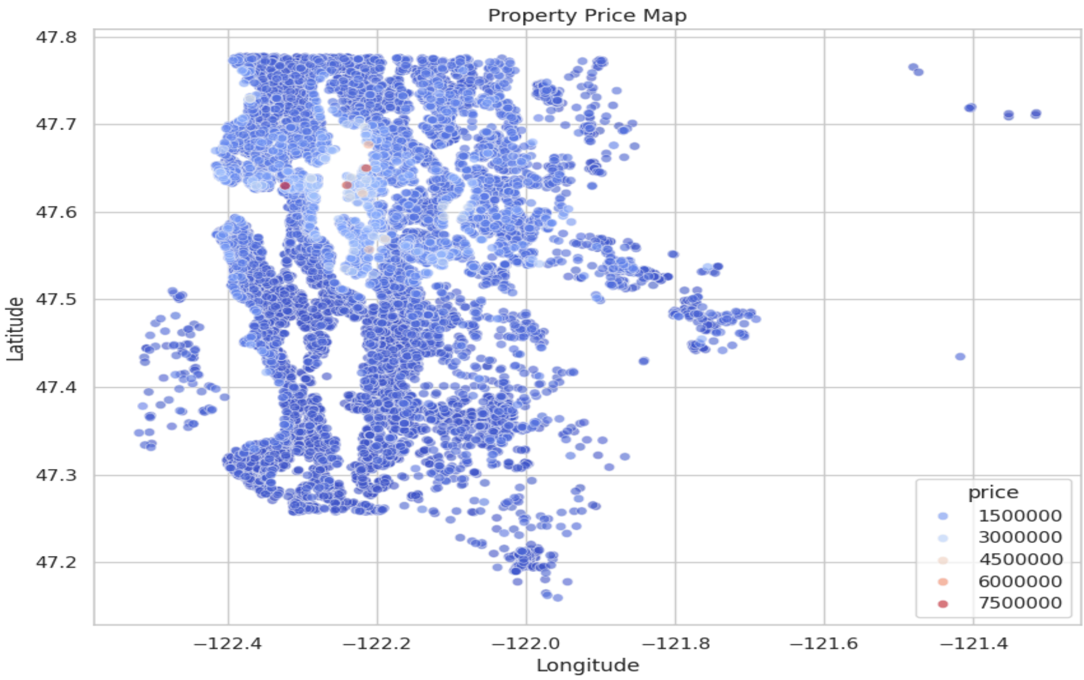
2. EXPLORATORY DATA ANALYSIS (EDA)

Key Insights from the Data:

Histogram of Price Distribution -



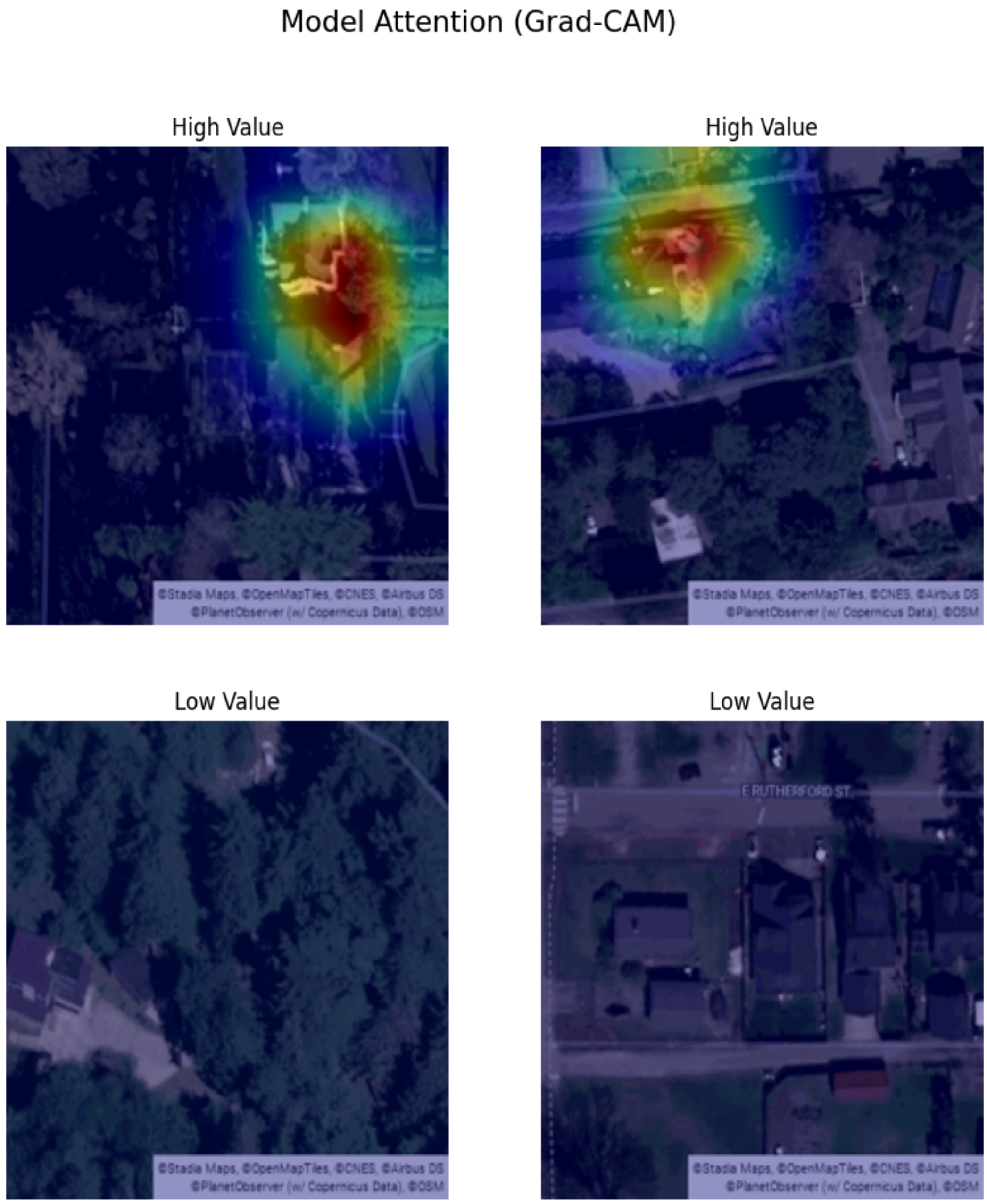
Map of Expensive Neighborhoods -



Sample Satellite Images -

FINANCIAL & VISUAL FORENSICS

To understand what the "Black Box" Neural Network is seeing, we applied Grad-CAM (Gradient-weighted Class Activation Mapping). This technique highlights the pixels that most influenced the price prediction.

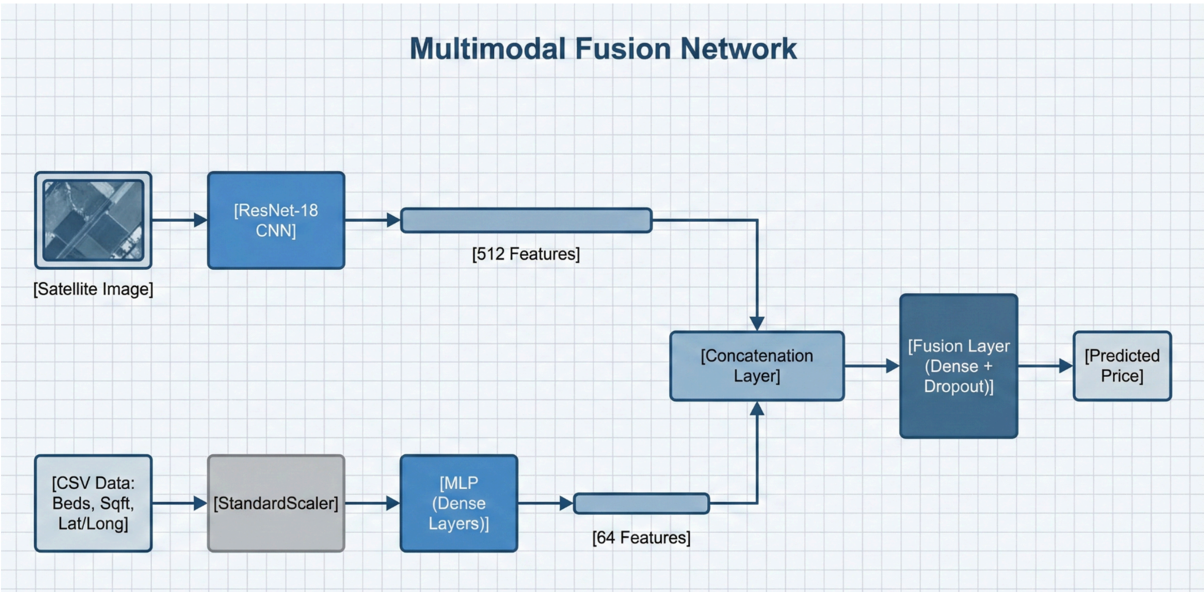


- High Value Images (Top Row): The model's attention (red hotspots) strongly focuses on the main structure of the house, indicating that building size and footprint are key drivers.

- Low Value Images (Bottom Row): The attention is more diffuse or focuses on the lot boundaries/surrounding pavement, suggesting a lack of distinct high-value structural features.)

Findings:
1. "Structure Over Surroundings": For high-value homes, the model learned to "look at" the house itself, validating that visual house quality (size, complexity) matters.
2. "Contextual Clues": For lower-value properties, the model often attended to the surrounding context (roads, density), implying that location density is a stronger signal when the house structure itself is modest.
3. Validation: This confirms the visual branch is contributing meaningful signals beyond just random noise, actively identifying property footprints.

ARCHITECTURE COMPOSITION



The model independently processes visual satellite context through a ResNet-18 CNN and numerical property data through a dedicated MLP.

Individual embeddings (512 visual and 64 tabular features) are merged into a unified vector to represent the property's total market value.

The fusion layer maps the combined features directly to a price prediction, allowing the model to learn how visual "curb appeal" influences physical house specs.

RESULTS & PERFORMANCE

We compared a robust Gradient Boosting Regressor (XGBoost) trained on tabular data against our Multimodal Neural Network.

Metric: R2 Score

1. Tabular Baseline (XGBoost): 0.886
2. Multimodal Network (Ours): 0.7458