PROJECT REPORT

SATELLITE IMAGERY-BASED PROPERTY VALUATION

Bharat Goyal

24116022

OVERVIEW

This project implements a **Multimodal Machine Learning Pipeline** designed to predict residential property prices by fusing structured tabular data with unstructured satellite imagery. The core hypothesis is that visual environmental features (e.g., neighborhood density, greenery, road patterns) captured in satellite images contain predictive signals distinct from standard housing metrics.

Modeling Strategy:

Data Sourcing: Numerical data was sourced from historical sales records. Visual data was programmatically acquired via the **Mapbox Static API**, fetching high-resolution satellite imagery (600x600, Zoom 18) for every property based on its GPS coordinates.

Feature Engineering:

Tabular: Standard preprocessing (Scaling) and creation of derived features like house_age and living_to_lot_ratio.

Visual: A **ResNet18** deep convolutional neural network (pretrained on ImageNet) was used as a **Feature Extractor**. The classification head was removed to extract 512-dimensional latent visual embeddings for each property.

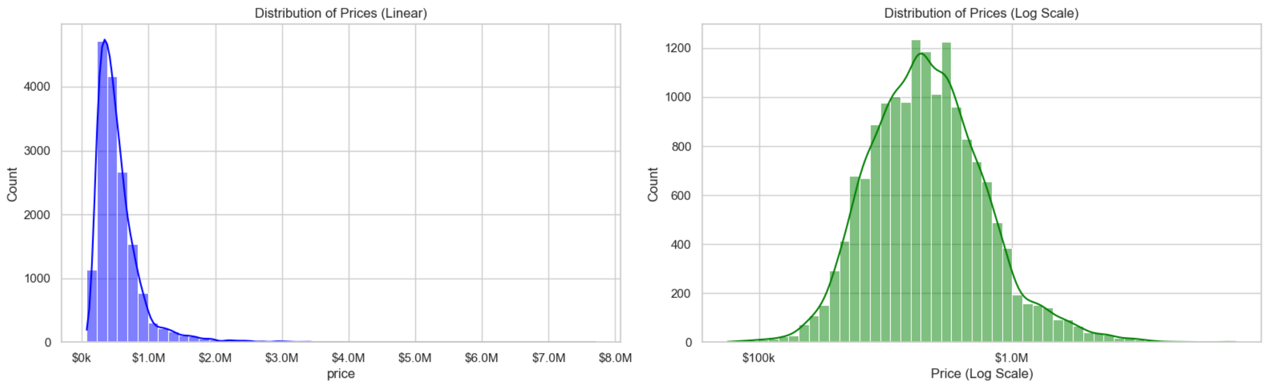
Dimensionality Reduction: Principal Component Analysis (PCA) was applied to compress the 512 visual features down to **20 principal components** to reduce noise and dimensionality before fusion.

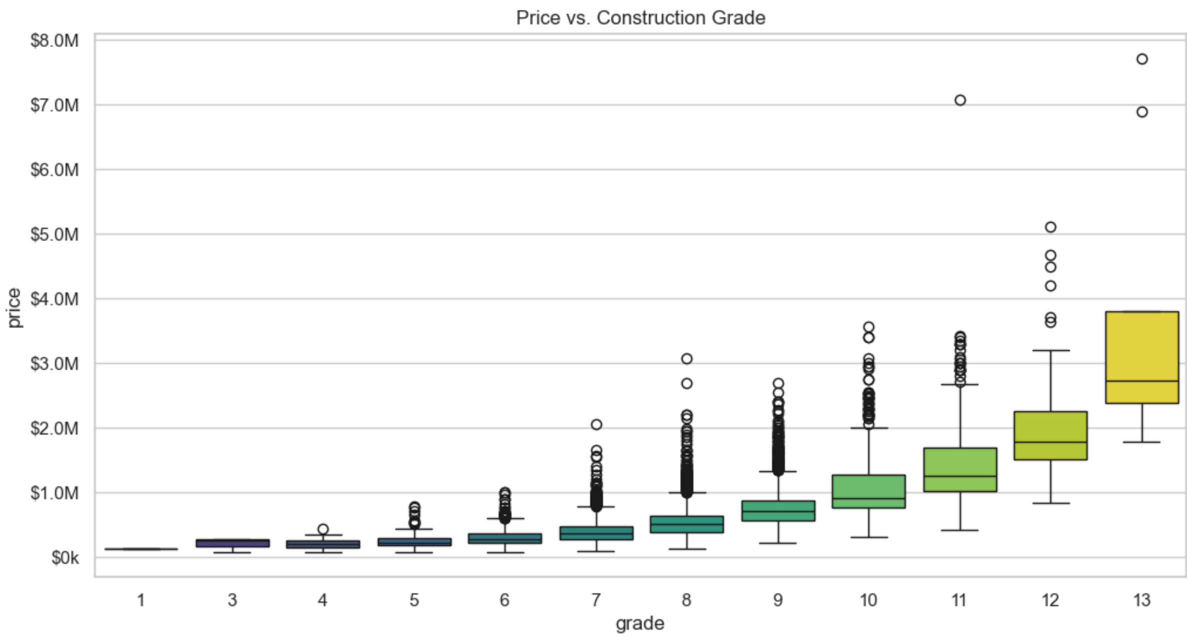
Fusion & Prediction: The compressed visual embeddings were concatenated with scaled tabular features. An **XGBoost Regressor** was trained on this hybrid dataset to predict the log-transformed price.

Exploratory Data Analysis (EDA)

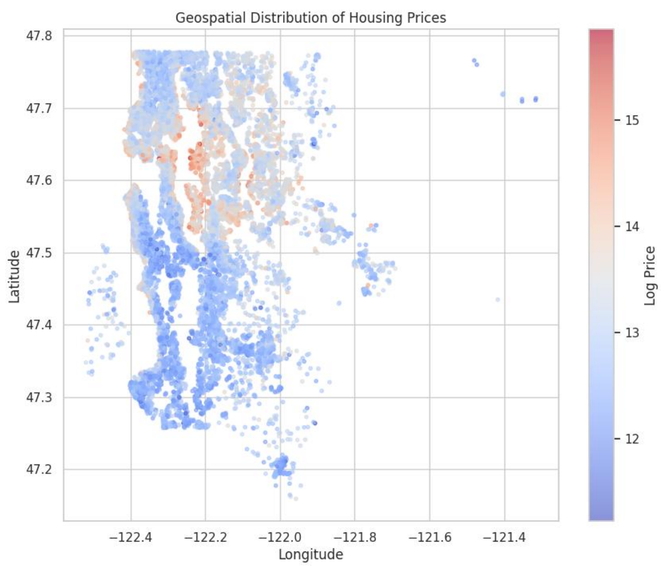
Data analysis was performed in preprocessing.ipynb to understand the underlying distributions and spatial relationships.

Price Distribution: The raw property prices exhibited a significant right-skew (long tail of luxury homes). A **Log Transformation (np.log1p)** was applied, successfully normalizing the target variable for better regression stability.

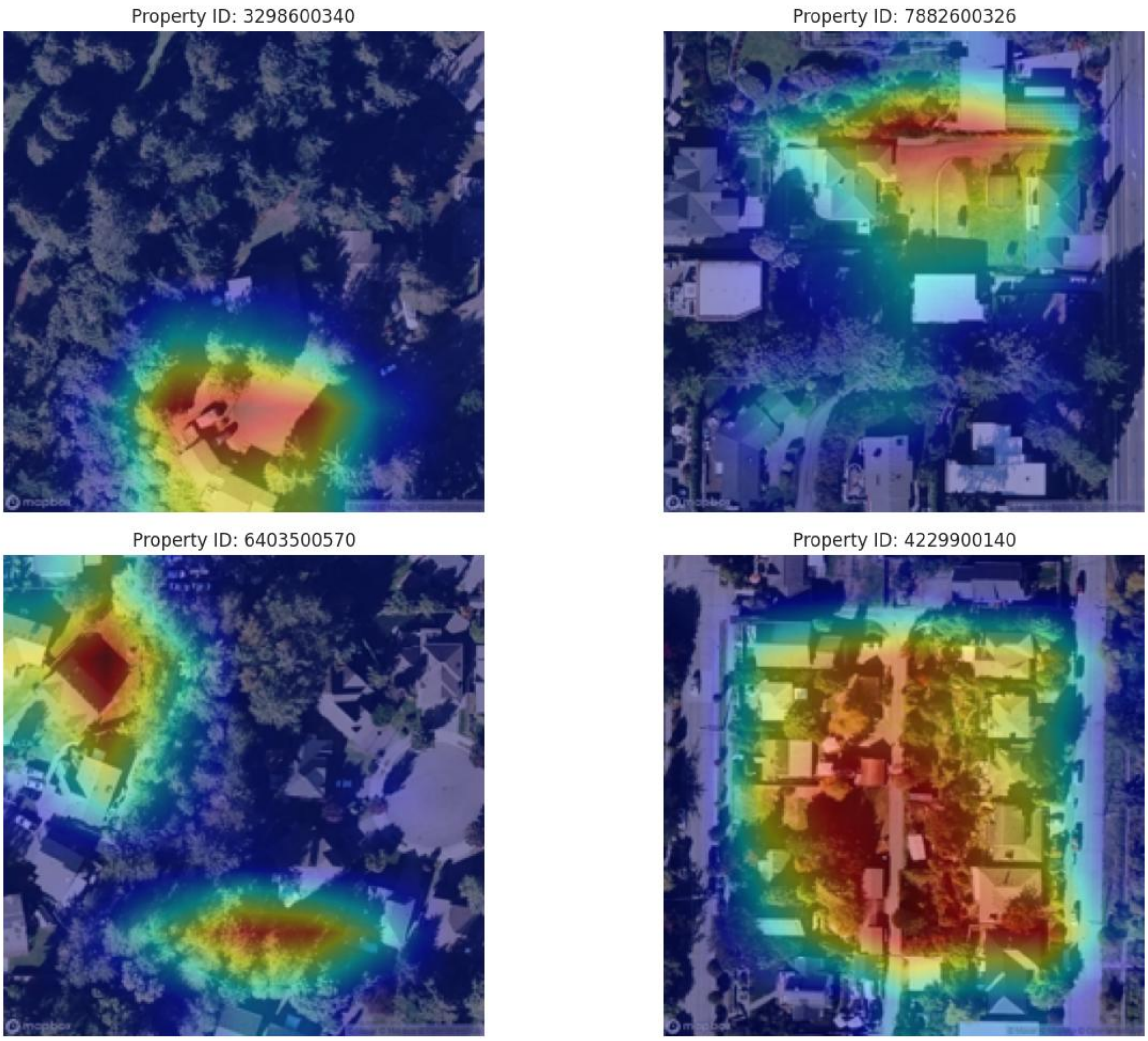




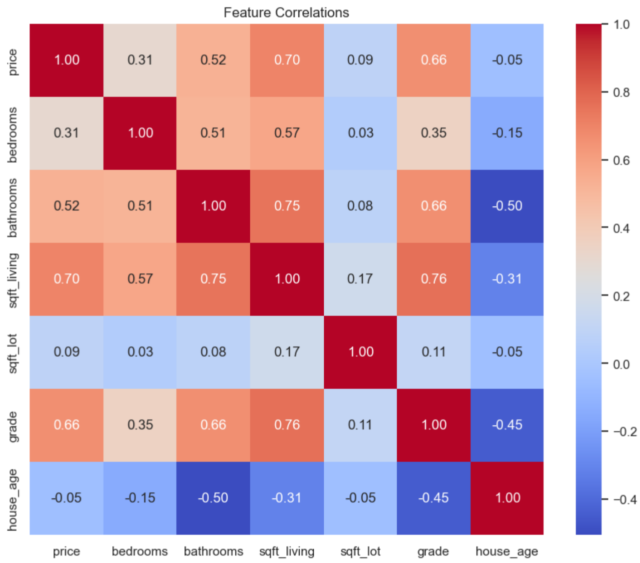
Geospatial Analysis: A scatter plot of Latitude vs. Longitude colored by Price revealed clear spatial clustering. High-value properties are concentrated in specific waterfront and northern zones, confirming that location is a dominant predictor.

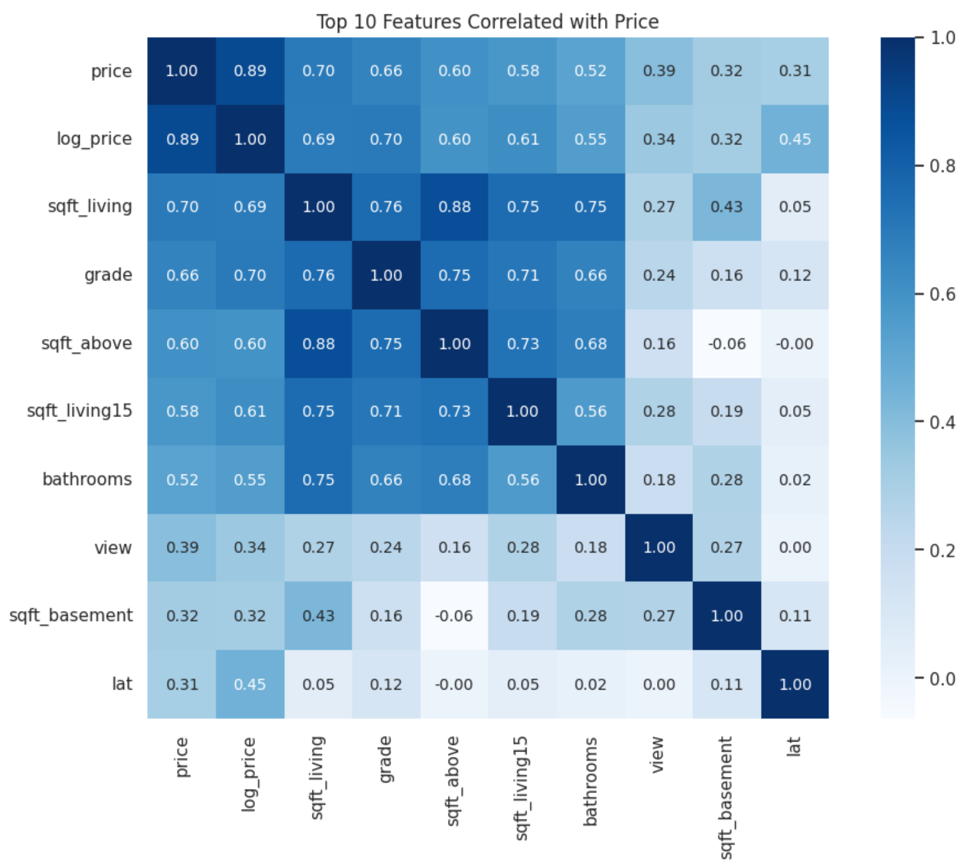


GRAD-CAM:



Correlation Analysis: A heatmap of tabular features showed that grade (construction quality) and sqft_living have the highest positive correlation with price, while house_age showed a weaker relationship.





Financial & Visual Insights

By isolating the data sources, we derived specific insights into what drives property value:

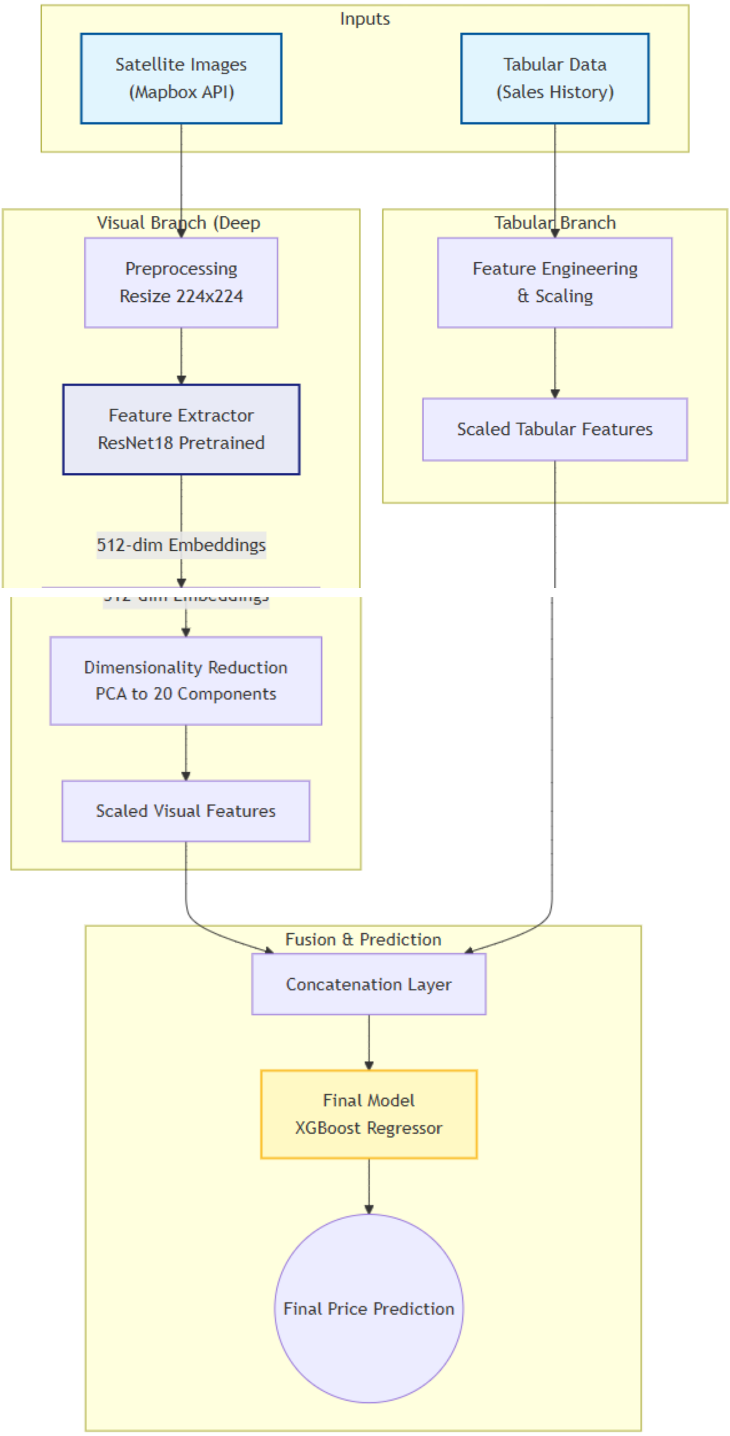
The Power of "Curb Appeal": I trained an **Image-Only Model** (using only PCA-reduced satellite embeddings) which achieved an **R^2 of 0.3241**. This quantitatively proves that satellite imagery alone captures ~32% of the price variance, detecting signals likely related to neighborhood wealth and density without knowing the house size.

Tabular Dominance: The **Tabular-Only Baseline** achieved an R^2 **of 0.8926**. Features like square footage and explicit GPS coordinates (lat, long) are extremely powerful predictors.

Fusion Value: Adding visual embeddings to the tabular data in the **Hybrid Model** improved the R^2 to **0.8980**. While the gain is marginal (~0.5%), it confirms that visual data provides unique information not fully captured by the tabular columns, likely related to micro-location characteristics.

Architecture Diagram

The system follows a **Late Fusion** architecture where features are processed independently before entering the final regressor. This approach preserves modality-specific information while allowing satellite imagery to complement strong tabular predictors in the final regression model.



Results

I compared three modelling configurations to validate the multimodal approach. All models were evaluated using **RMSE (Log Scale)** and **R2 Score** on the validation set.

Model Architecture	Features Used	RMSE (Log)	R^2 Score
Baseline Model	Tabular Data Only (20 features)	0.1715	0.8926
Image-Only Model	Top 50 PCA Visual Components	0.4303	0.3241
Hybrid Fusion Model	Tabular + Top 20 PCA Visual Components	**0.1671**	**0.8980**

Conclusion:

The Hybrid Fusion Model outperformed both individual models, achieving the lowest error (RMSE 0.1671) and highest accuracy (R2 0.898). While the Tabular Baseline is highly competitive, the integration of satellite imagery via ResNet18 extraction provided a measurable boost in predictive power, validating the hypothesis that visual context enhances automated valuation models.