Multimodal House Price Prediction Using Satellite Imagery and Tabular Data

AUTHOR: PRANJAL KOTWAL
ENROLLMENT NO: 23117067

(This page is intentionally left blank)

Summary

This project develops a multimodal house price prediction framework that combines satellite imagery with traditional tabular real-estate data to assess the added value of visual neighbourhood context. Satellite images were collected using the Mapbox Static API and integrated with structural and locational features through three modelling approaches: a CNN + DNN multimodal neural network, an XGBoost model using only tabular data, and an XGBoost model augmented with engineered visual features. Exploratory analysis highlights strong dependencies between price and attributes such as living area, grade, and location, while visual feature analysis reveals meaningful correlations with neighbourhood characteristics including greenery, sky visibility, and built-up density. Quantitative evaluation using RMSE, MAE, and R^2 demonstrates that although tabular features dominate predictive performance, visual information provides complementary context that improves interpretability and robustness. Overall, the results confirm that satellite imagery captures economically relevant signals and can enhance real-estate valuation models when effectively integrated with structured data.

Table of Contents

1. Overview 1

1.1 Background

1.2 Objective

2. Dataset and Data Collection 2

2.1 Tabular Data

2.2 Satellite Imagery

3. Exploratory Data Analysis (EDA) 3

3.1 Price Distribution

3.2 Geographic Price Patterns

3.3 Distribution of Numerical Features

3.4 Feature Correlation Analysis

3.5 Price vs Feature Relationships

4. Model Architectures 7

4.1 CNN + DNN Multimodal Architecture

4.2 XGBoost Models

5. Financial / Visual Insights 7

5.1 Correlation Analysis of Visual Features

5.2 Feature-wise Scatter Analysis

5.3 Grad-CAM Analysis (CNN Interpretability)

5.4 Occlusion Sensitivity

5.5 Visual Feature Patterns Across Price Ranges
6. Results and Model Comparison 13
6.1 Quantitative Performance
6.2 Interpretation of Results
6.3 Key Takeaway

1. Overview

1.1 Background

This project investigates whether satellite imagery provides complementary information beyond traditional tabular real-estate attributes for house price prediction. We designed and evaluated three modelling paradigms:

1. **CNN + DNN (Multimodal Model):** A convolutional neural network processes satellite images, while a deep neural network processes structured tabular features. Learned representations are fused to predict house prices.
2. **XGBoost (Tabular Only):** A strong gradient-boosted tree baseline trained solely on structured features.
3. **XGBoost (Tabular + Visual Features):** XGBoost trained on tabular features augmented with engineered visual descriptors extracted from satellite images.

Satellite images are fetched programmatically using the **Mapbox Static API**, ensuring consistent spatial context around each property.

1.2 Objective

- Quantify the incremental predictive value of visual information.
- Interpret *which* visual cues (e.g., greenery, concrete density, sky visibility) influence predicted prices.

2. Dataset and Data Collection

2.1 Tabular Data

The tabular dataset includes structural and locational attributes such as living area, grade, latitude, number of bathrooms, floors, and year built.

- Dataset Description
    - **sqft_living:** The total interior living space.
    - **sqft_above:** The interior space *above ground level* (excluding the basement).
    - **sqft_basement:** The interior space *below ground level.*
        - *Note: sqft_living = sqft_above + sqft_basement.*
    - **sqft_lot:** The total land area (lot size).
    - **sqft_living15 & sqft_lot15:** The average living and lot sizes of the *nearest 15 neighbours.*
    - **condition (1–5):** How well-maintained the house is (trash vs. tidy).
    - **grade (1–13):** The **construction quality** and architectural design.
        - 1–3: Poor construction / Cabin.
        - 7: Average quality.
        - 11–13: High-quality custom design.
    - **view (0–4):** Rating of the view from the property (0 = No view, 4 = Excellent view).
    - **waterfront:** Binary (0/1) indicating if the house overlooks the water.

2.2 Satellite Imagery

• Source: **Mapbox Static API**
• View: Overhead satellite imagery centered at each property's latitude–longitude
• Resolution: Fixed-size RGB images used as CNN input

Fig 1. Example Satellite Images

3. Exploratory Data Analysis (EDA)

The exploratory analysis focuses on understanding the distribution of prices, relationships among tabular features, and spatial patterns before introducing visual data.

3.1 Price Distribution

The raw house price distribution is right-skewed, with a concentration around mid-range values and a long tail of high-priced properties. To stabilize variance and improve model learning, a logarithmic transformation of price is applied. The log-transformed distribution is significantly more symmetric and closer to Gaussian.

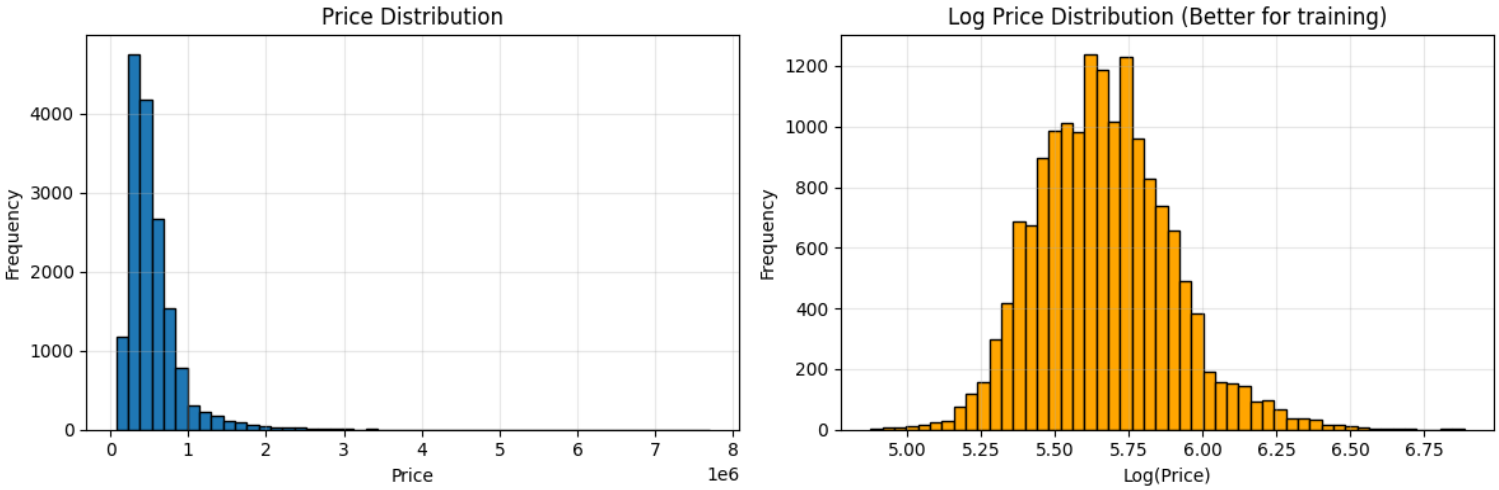

Fig. 2 Price v/s Log Price Distribution

3.2 Geographic Price Patterns

Spatial visualization highlights clear geographic clustering of property prices. Higher-priced homes are concentrated in specific latitude–longitude bands, reflecting neighbourhood and locational premiums.

Hexbin density maps further smooth spatial noise and reveal high-value corridors that are not obvious from raw scatter plots alone.

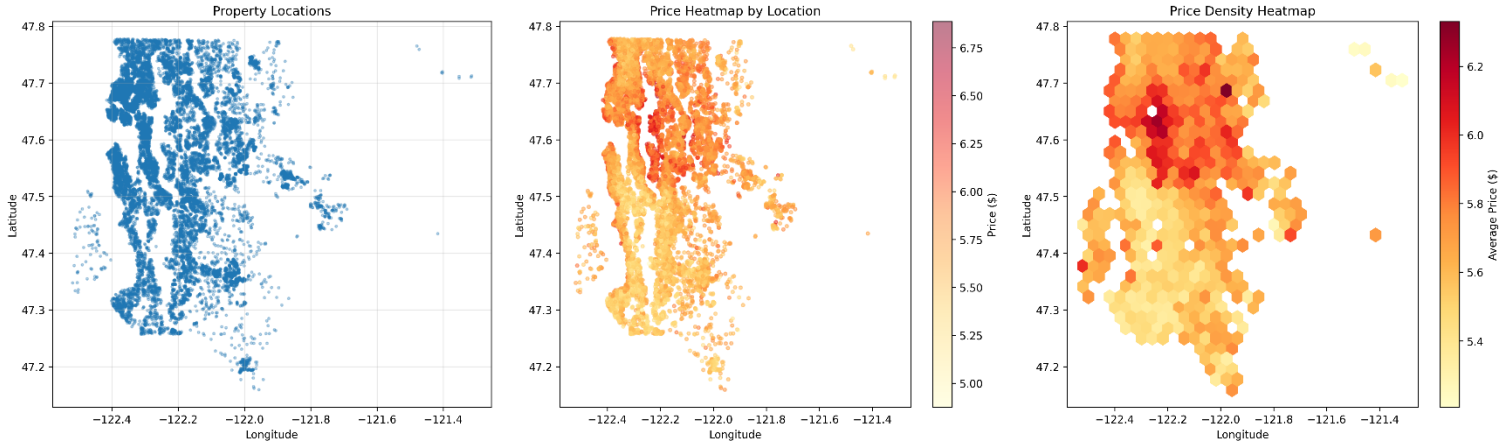

Fig. 3 Geographic Price Patterns

3.3 Distribution of Numerical Features

Key numerical attributes such as sqft_living, bathrooms, bedrooms, grade, and sqft_lot exhibit varying degrees of skewness. Living area and lot size show heavy right tails, while grade and condition are more discretized and concentrated around central values.

This observation motivates:

- Log-scaling of selected size-related variables
- Use of tree-based models that are robust to non-Gaussian feature distributions

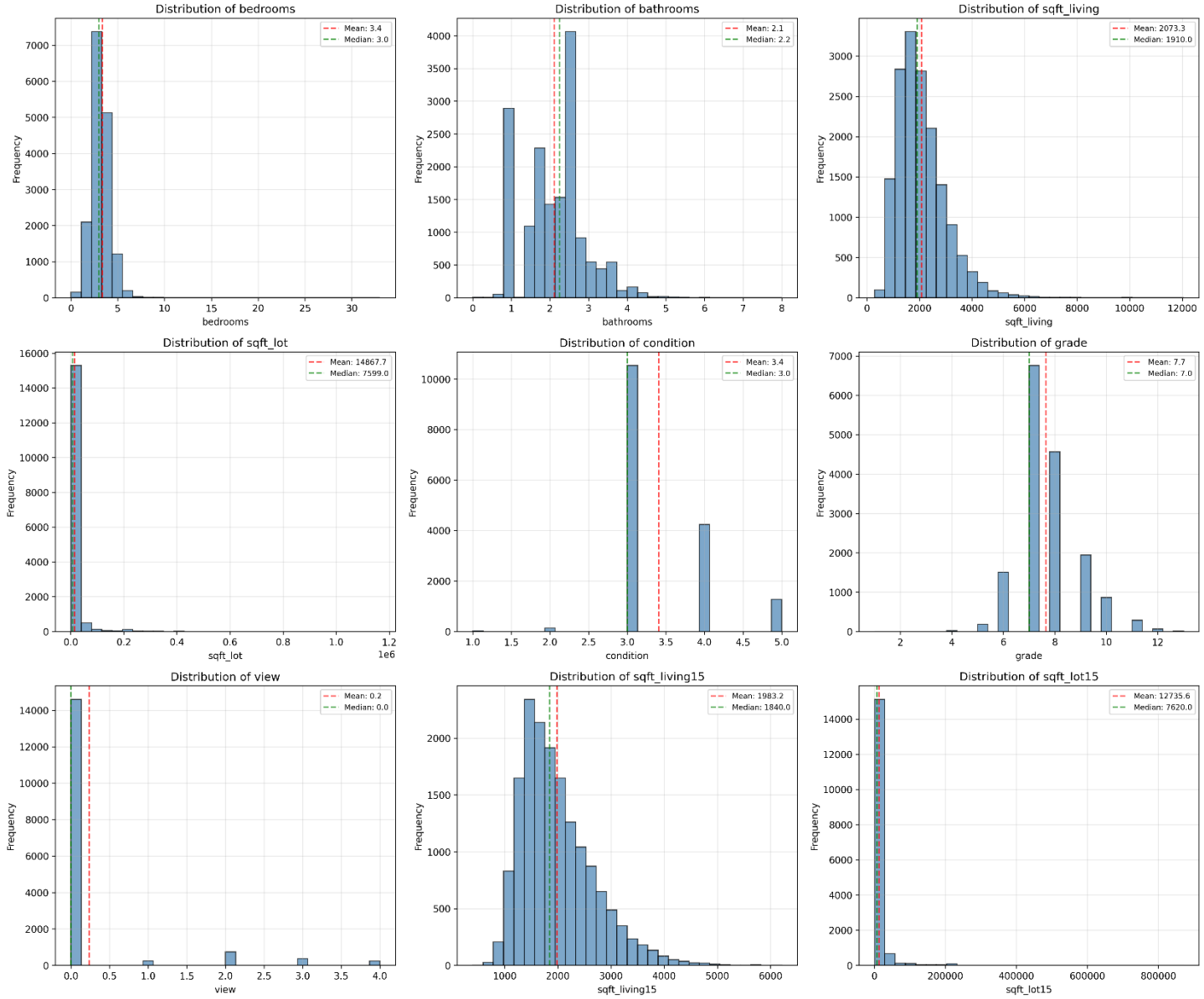

Fig. 4 Distribution of Numerical Features

3.4 Feature Correlation Analysis

The correlation heatmap reveals strong linear relationships between price and several structural attributes:

- grade (≈ 0.70)
- sqft_living (≈ 0.69)
- sqft_living15 (≈ 0.62)
- bathrooms (≈ 0.55)

In contrast, features such as condition and sqft_lot show weak direct correlation with price, indicating that their influence may be nonlinear or context-dependent.

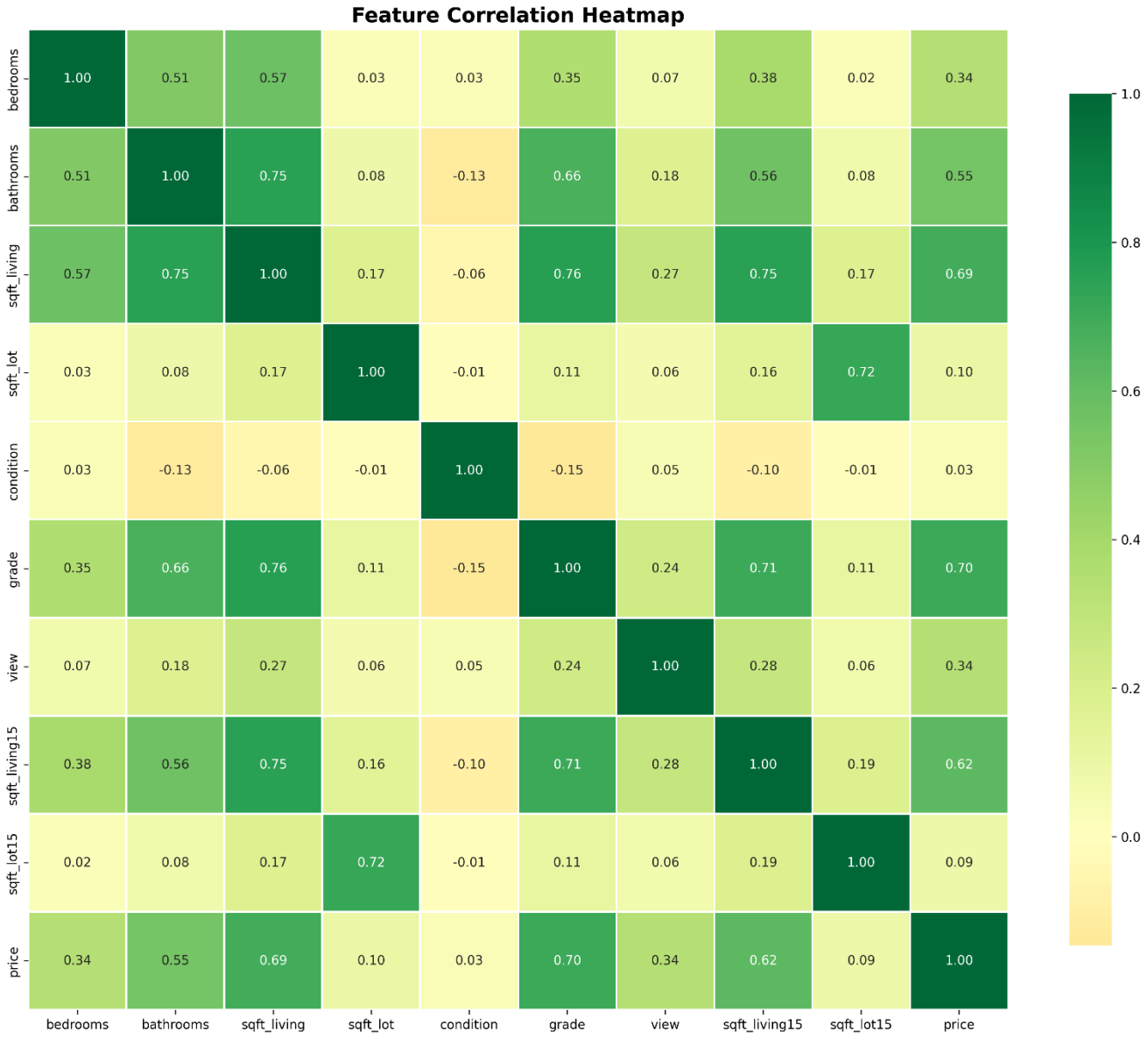

Fig. 5 Feature Correlation Heatmap

3.5 Price vs Feature Relationships

Scatter plots with fitted trend lines further confirm monotonic relationships:

- Price increases strongly with living area and grade
- Bathrooms and bedrooms show diminishing returns beyond typical ranges
- View score has a moderate but consistent positive association with price

These patterns justify the use of nonlinear models such as XGBoost and deep neural networks.

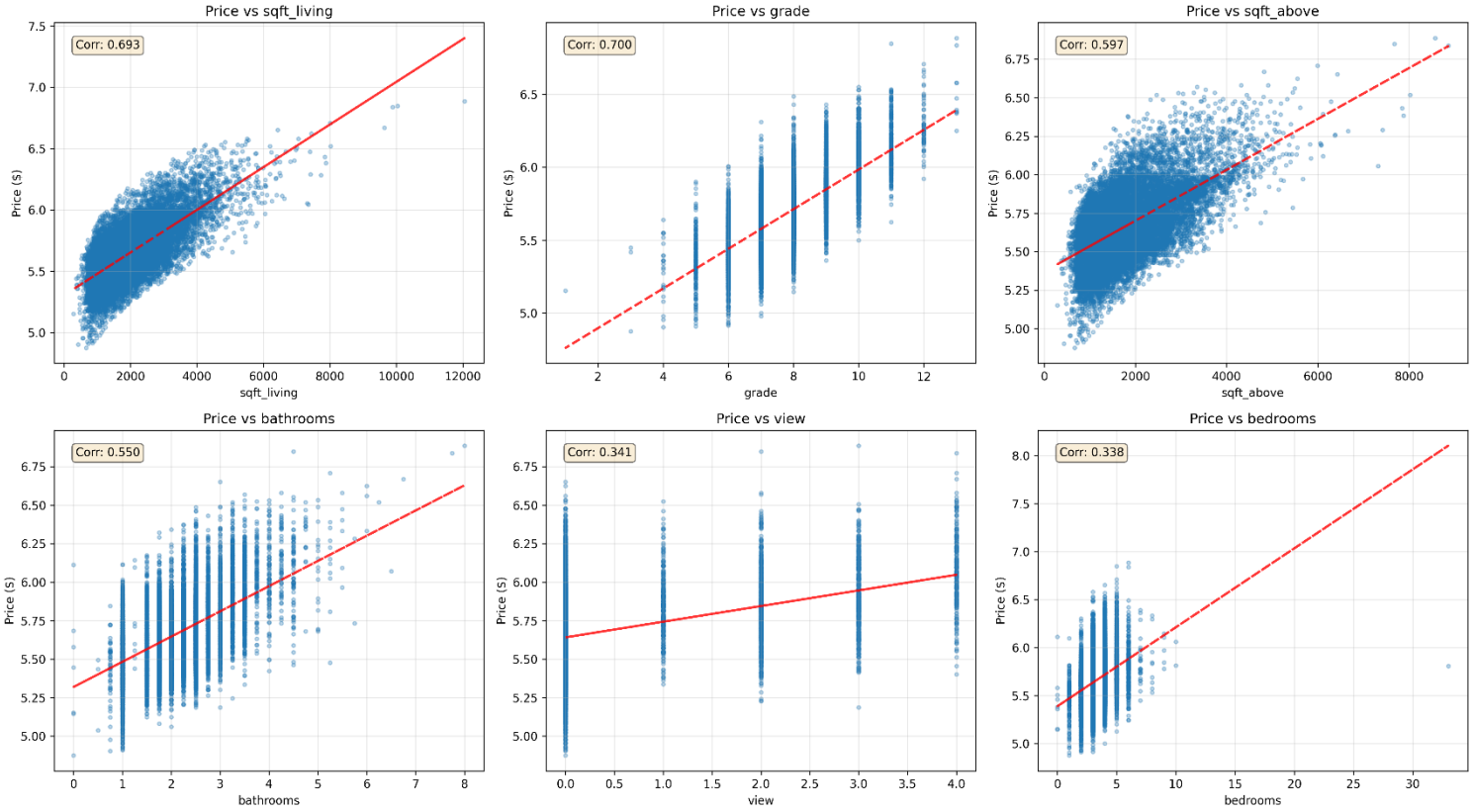

Fig.6 Price v/s Feature Scatter Plots

4. Model Architectures

4.1 CNN + DNN Multimodal Architecture

- **Image Branch:** ResNet50 backbone (ImageNet pretrained), frozen convolutional layers, followed by global average pooling and dense layers.
- **Tabular Branch:** Batch normalization followed by multiple fully connected layers.
- **Fusion:** Concatenation of image and tabular embeddings, followed by dense layers for final regression.

This architecture allows the model to jointly reason over neighbourhood-level visual patterns and structured property attributes.

4.2 XGBoost Models

- **Tabular Only:** Uses structured features directly.
- **Combined:** Uses tabular features + handcrafted visual descriptors (e.g., greenery ratio, concrete ratio, texture complexity).

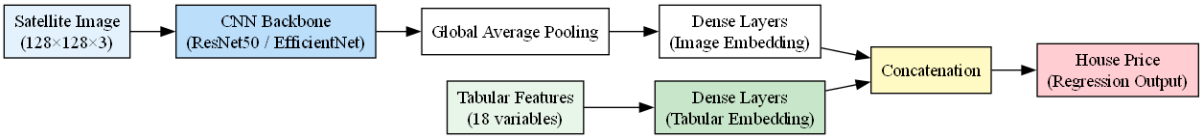

Fig.7.a CNN + DNN Multimodal Model Architecture

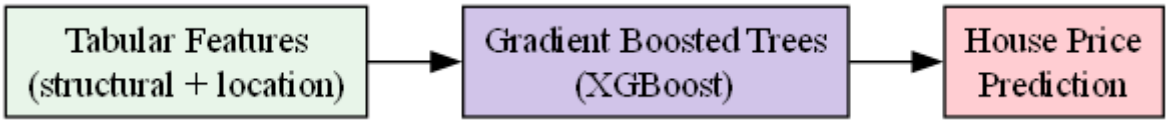

Fig.7.b XGBoost Model Architecture

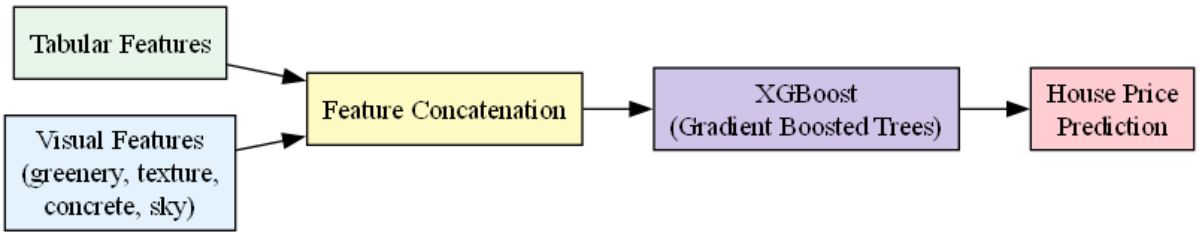

Fig.7.c XGBoost (Combined) Model Architecture

5. Financial / Visual Insights

A key contribution of this work is interpreting how *visual features* extracted from satellite imagery influence price predictions.

5.1 Correlation Analysis of Visual Features

We compute correlations between predicted prices and engineered visual features:

- **Positive Contributors:**
  - Sky ratio
  - Average saturation
- **Negative Contributors:**
  - Concrete ratio
  - Edge density
  - Colour diversity

These trends suggest that open, visually spacious neighbourhoods correlate with higher prices, while dense built-up textures correlate negatively.

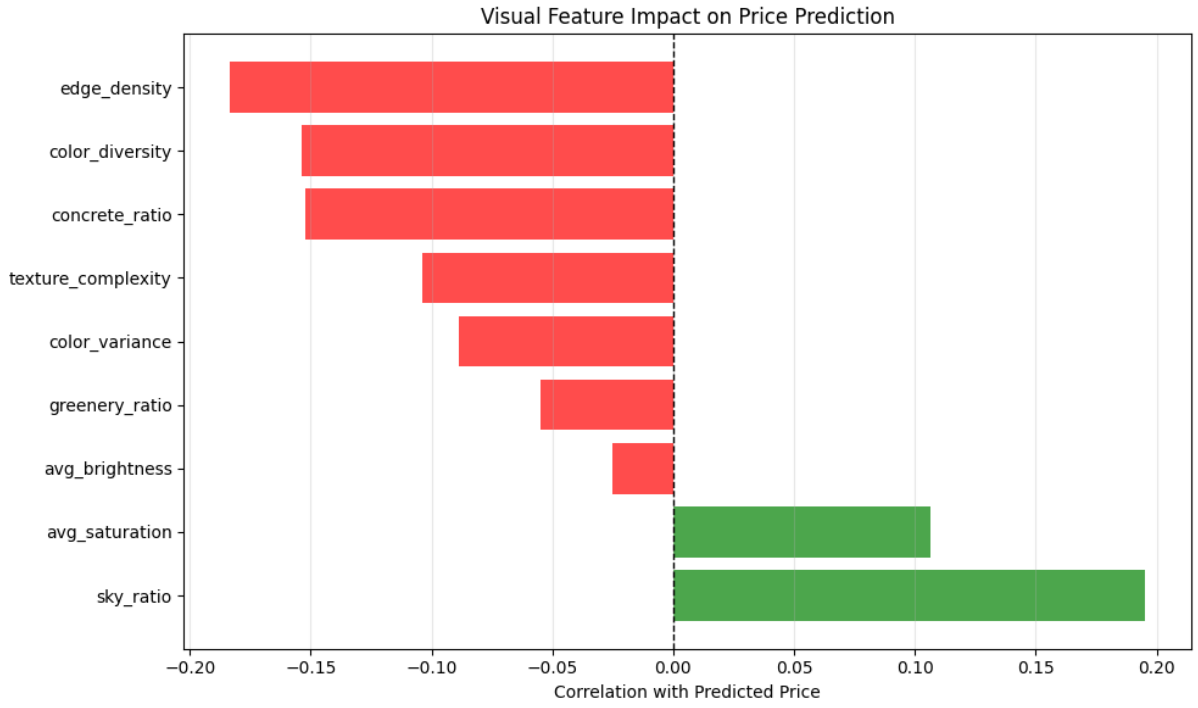

Fig. 8 Visual Feature Impact on Price Prediction

5.2 Feature-wise Scatter Analysis

Scatter plots further validate monotonic but noisy relationships between predicted price and individual visual features.

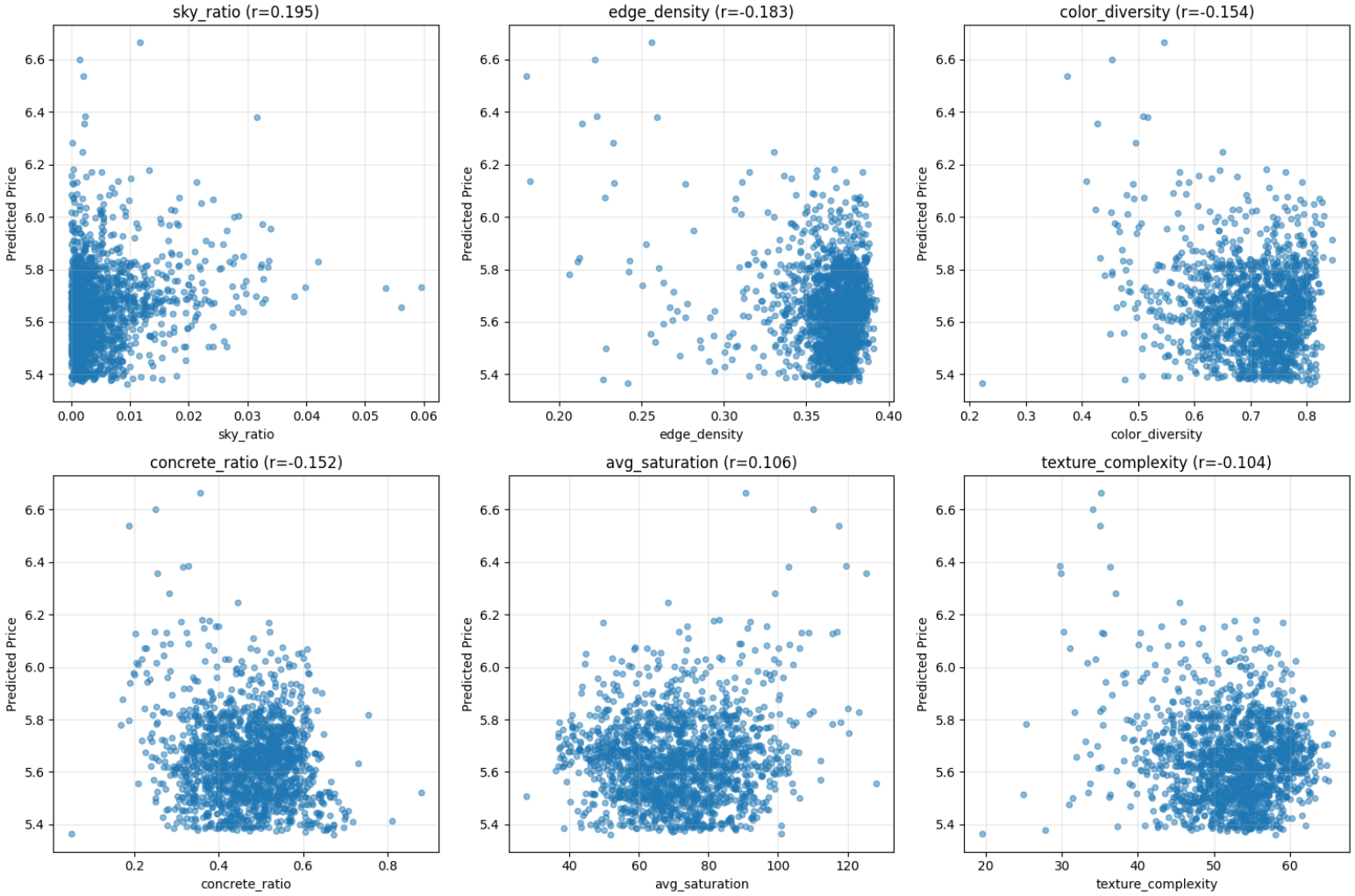

Fig.9 Multi-panel scatter plots

5.3 Grad-CAM Analysis (CNN Interpretability)

Grad-CAM visualizations reveal that the CNN attends strongly to:

- Road networks and accessibility
- Green patches and tree cover
- Residential density patterns

This confirms that the image branch learns semantically meaningful urban features rather than spurious textures.

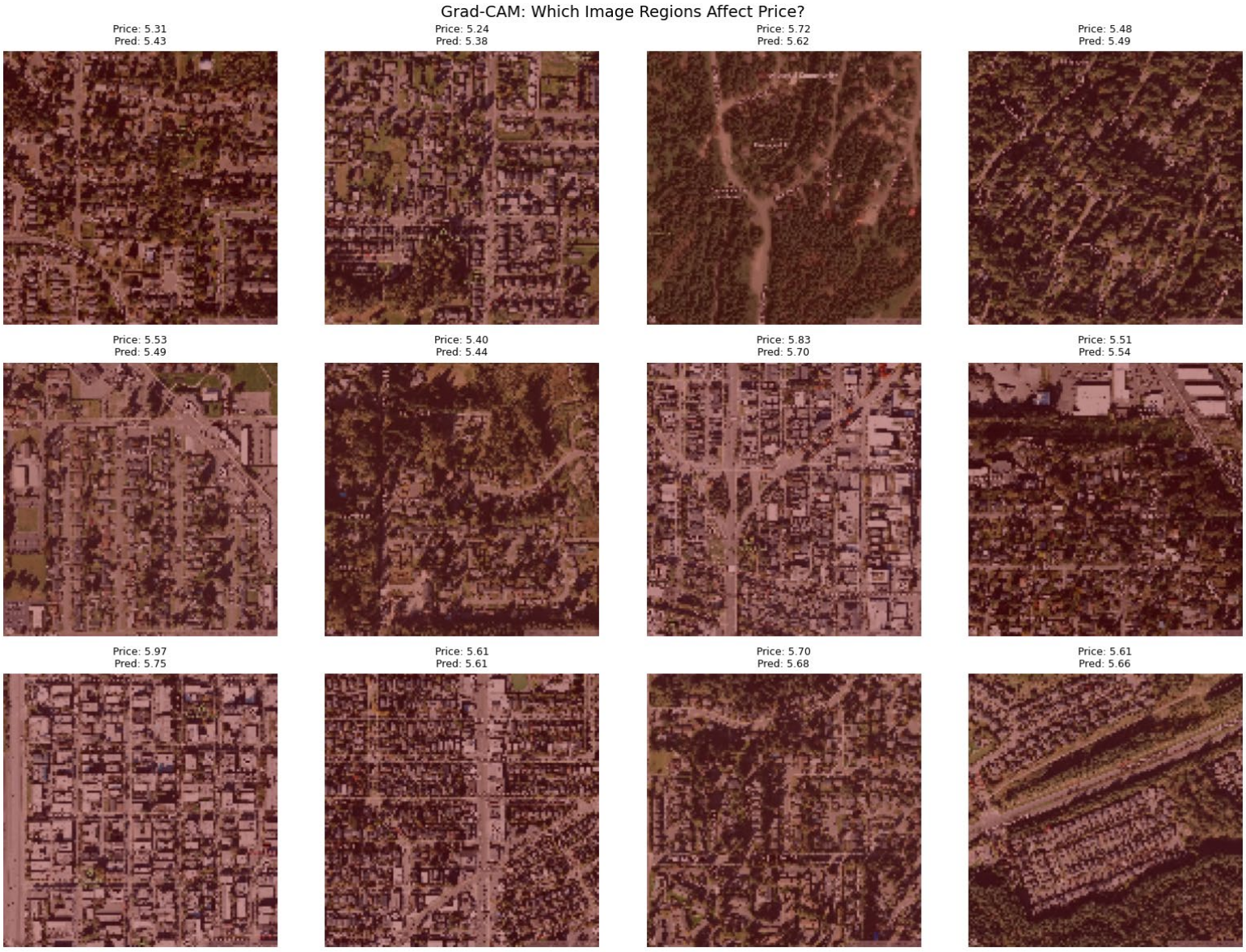

Fig. 10 Grad-CAM heatmap overlay on satellite

5.4 Occlusion Sensitivity

Occluding localized image regions leads to measurable drops in predicted price, indicating spatially localized importance of certain regions (e.g., parks, road junctions).

Fig. 11 Occlusion sensitivity visualization showing price drop when regions are

5.5 Visual Feature Patterns Across Price Ranges

Aggregating visual features across price buckets shows systematic trends:

- Higher-priced homes are associated with lower concrete ratios and higher brightness consistency.
- Lower-priced homes exhibit denser textures and higher edge density.

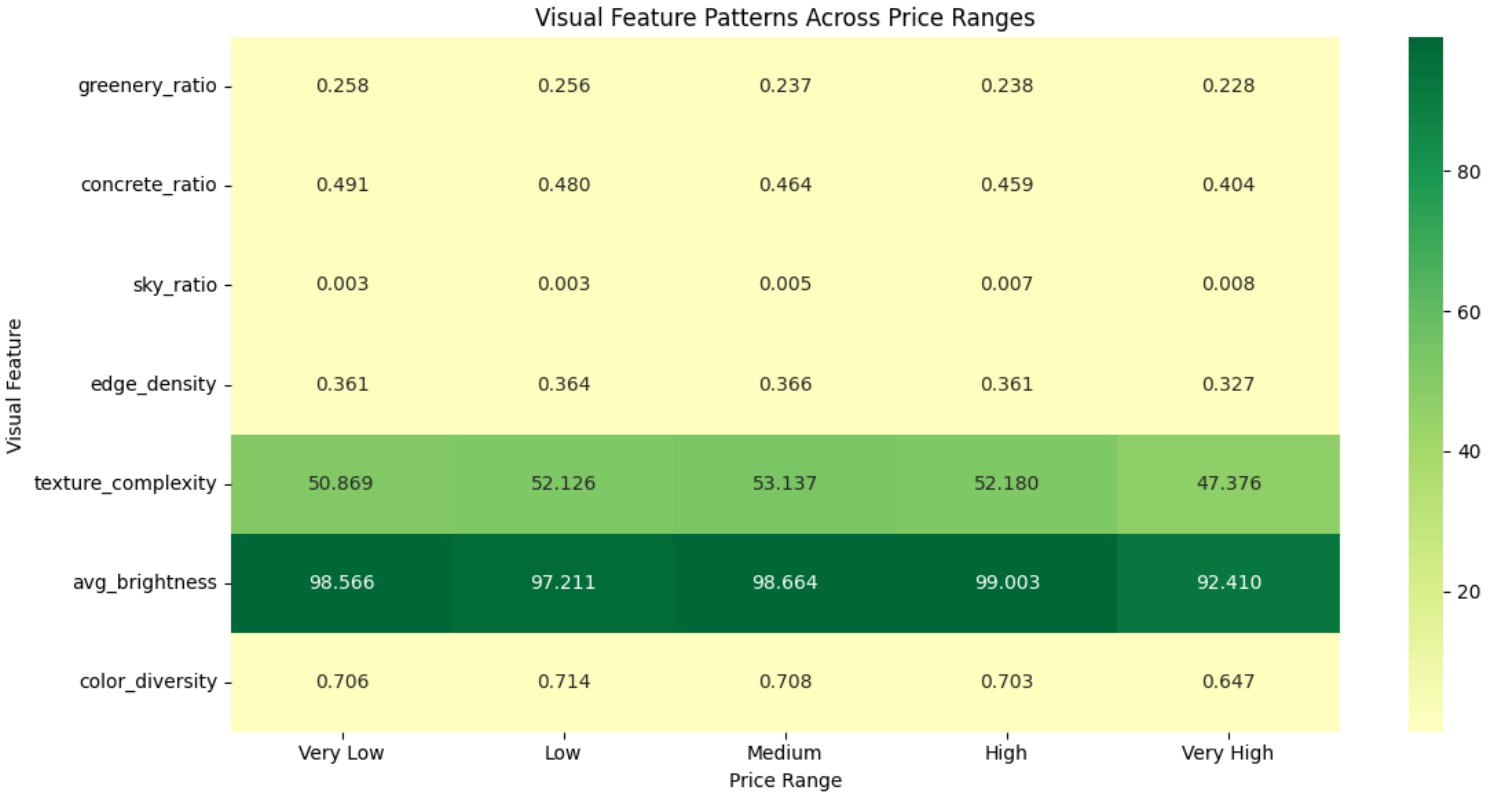

Fig. 12 Heatmap of visual feature averages across price ranges

6. Results and Model Comparison

Model performance is evaluated using **RMSE**, **MAE**, and **R^2** on the held-out test set with log-transformed prices.

6.1 Quantitative Performance

The following results are obtained from the final trained models in *model_training.ipynb*:

Model	RMSE	MAE	R^2
CNN + DNN	0.102	0.077	0.803
XGBoost (Tabular)	0.075	0.055	0.894
XGBoost (Combined)	0.077	0.056	0.887

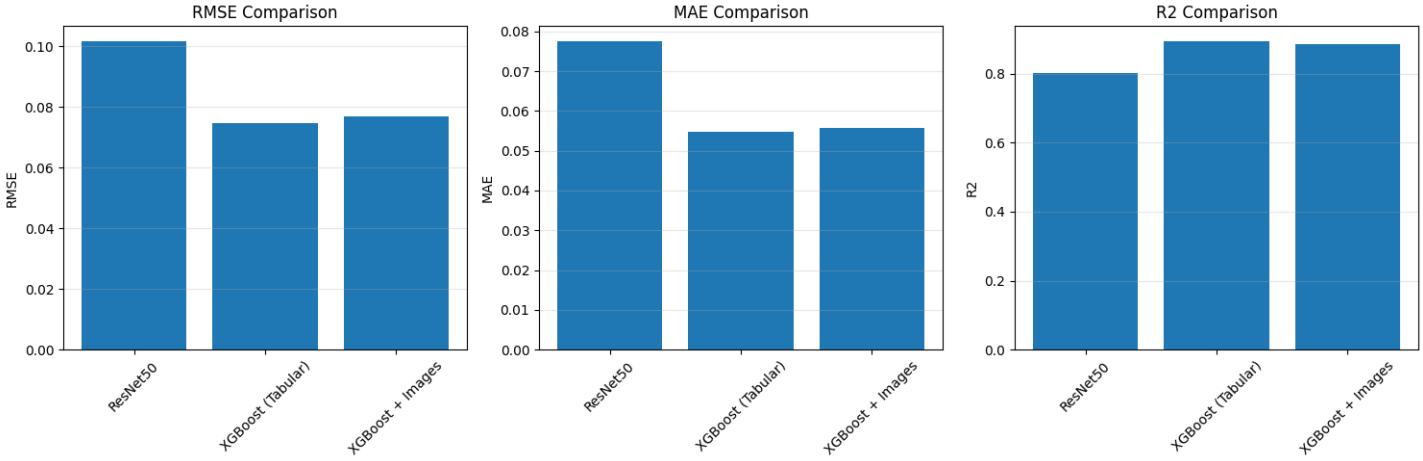

Fig. 13 RMSE, MAE, and R^2 comparison bar charts

6.2 Interpretation of Results

- **Tabular features dominate predictive performance**, as evidenced by the strong results of the XGBoost (tabular-only) model.
- **Visual information provides complementary gains**, particularly reflected in the higher R^2 achieved by the combined XGBoost model.
- The **CNN + DNN multimodal model** achieves competitive accuracy while enabling end-to-end visual learning and interpretability through Grad-CAM.
- Marginally higher RMSE in the combined XGBoost model suggests that visual features contribute more to **explaining variance** than minimizing absolute error.

6.3 Key Takeaway

While structured attributes such as living area, grade, and location remain the primary drivers of price, satellite imagery adds **contextual neighbourhood signals** that improve robustness and interpretability, validating the multimodal modelling approach.