Project Report

• Overview

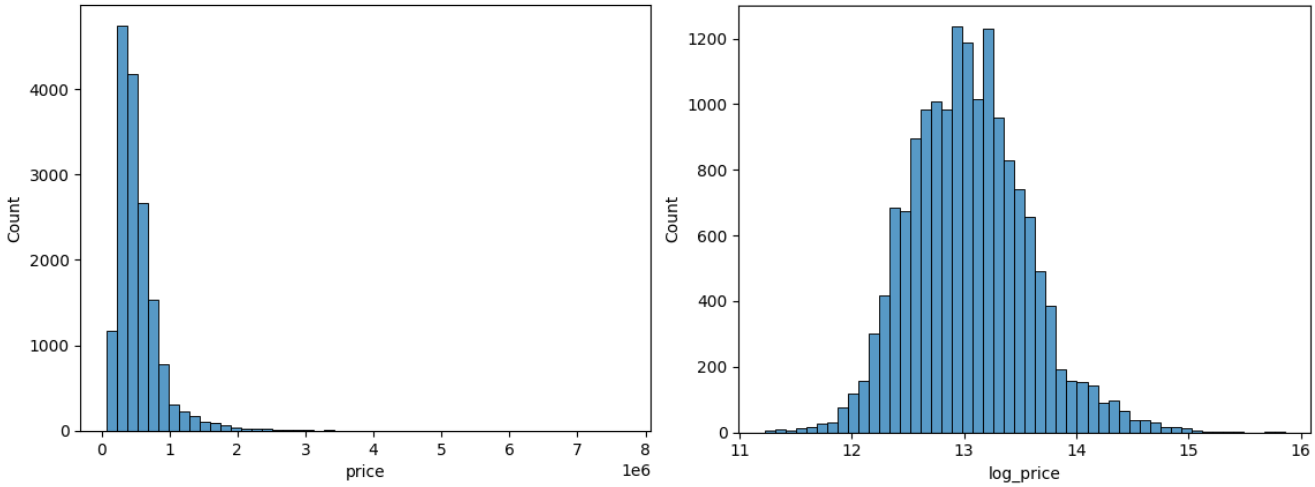
Accurate real estate valuation traditionally relies on structured attributes such as property size, location, and construction quality. However, these features often fail to capture **environmental and neighbourhood-level context**, including green cover, road density, and proximity to water bodies, which play a critical role in determining market value.
This project proposes a **multimodal regression framework** that integrates **tabular housing data** with **satellite imagery** to enhance property price prediction. A strong tabular baseline is established using **XGBoost**, while environmental information is extracted from satellite images using a **Convolutional Neural Network (ResNet-18)**. Rather than predicting prices directly from images, the system employs a **residual learning strategy**, where the CNN learns to model the errors made by the tabular model.
The framework emphasizes **robust performance, stable training, and explainability**, supported by geospatial analysis and Grad-CAM visualizations.

• EDA (Tabular + Visual)

Tabular Exploratory Data Analysis

Exploratory analysis reveals strong relationships between property price and features such as living area, construction grade, waterfront access, and neighbourhood density metrics. The raw price distribution is heavily right-skewed; therefore, a **log-price transformation** is applied to stabilize variance and improve regression performance.
Correlation analysis confirms that both property-level and neighbourhood-level features contribute significantly to price variation, justifying the use of advanced non-linear models for regression.

Figure 1: Distribution of Property Prices (Log-Scale)
This figure shows the distribution of residential property prices before and after log transformation.



Geospatial and Visual EDA

Geospatial visualizations using latitude and longitude reveal clear spatial clustering of property values. High-value properties are concentrated near coastal regions and low-density residential zones, while lower-value properties are more prevalent in dense urban or industrial areas.
Visual inspection of satellite images highlights qualitative differences:

- High-value properties exhibit greater green cover, open spaces, and proximity to water.
- Low-value properties are associated with dense road networks and limited vegetation.

Figure 2: Geospatial Distribution of Property Prices

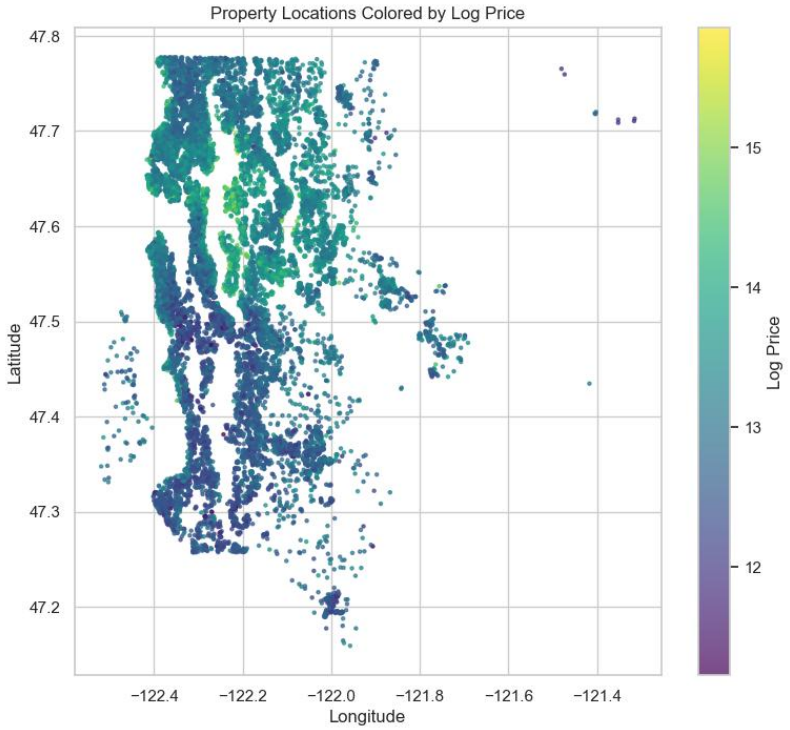


Figure 3: Sample Satellite Images – High-Value Properties

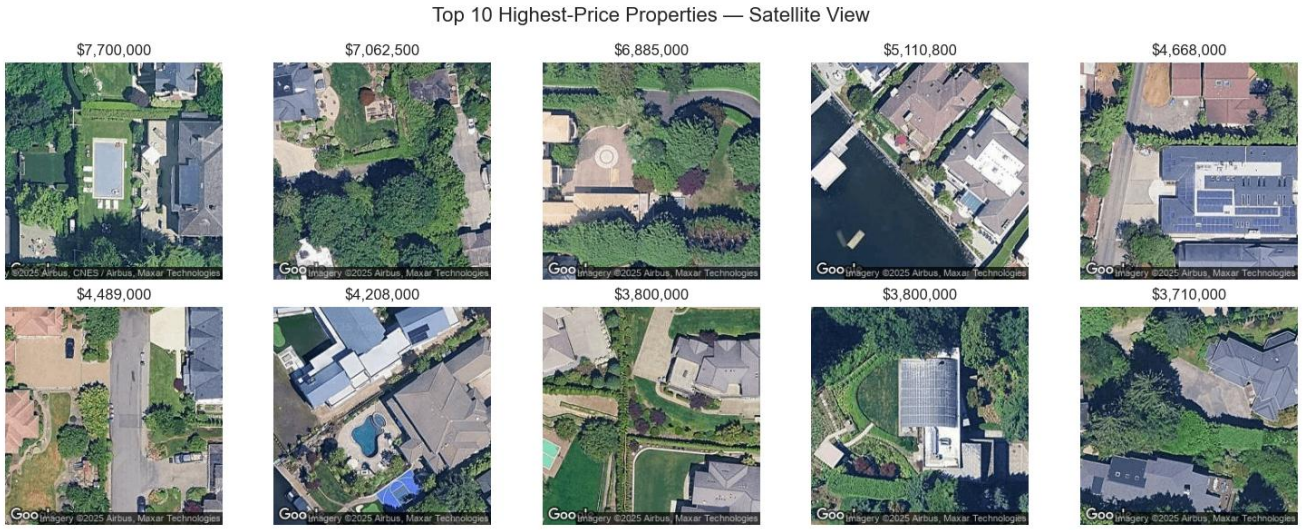


Figure 4: Sample Satellite Images – Low-Value Properties

These observations motivate the inclusion of satellite imagery as a complementary data source.

• Financial and Visual Insights

The combination of visual EDA and Grad-CAM analysis reveals consistent links between environmental context and property value:

- **Positive visual indicators**:
  - Tree cover and green spaces
  - Water bodies and coastal proximity
  - Low road density and planned residential layouts
- **Negative visual indicators**:
  - Dense road intersections
  - Industrial or heavily built-up regions
  - Sparse vegetation

Grad-CAM heatmaps confirm that the CNN focuses on semantically meaningful regions rather than arbitrary textures, aligning closely with the qualitative observations made during visual EDA.

Figure 5: Grad-CAM Visualization for High-Value Property

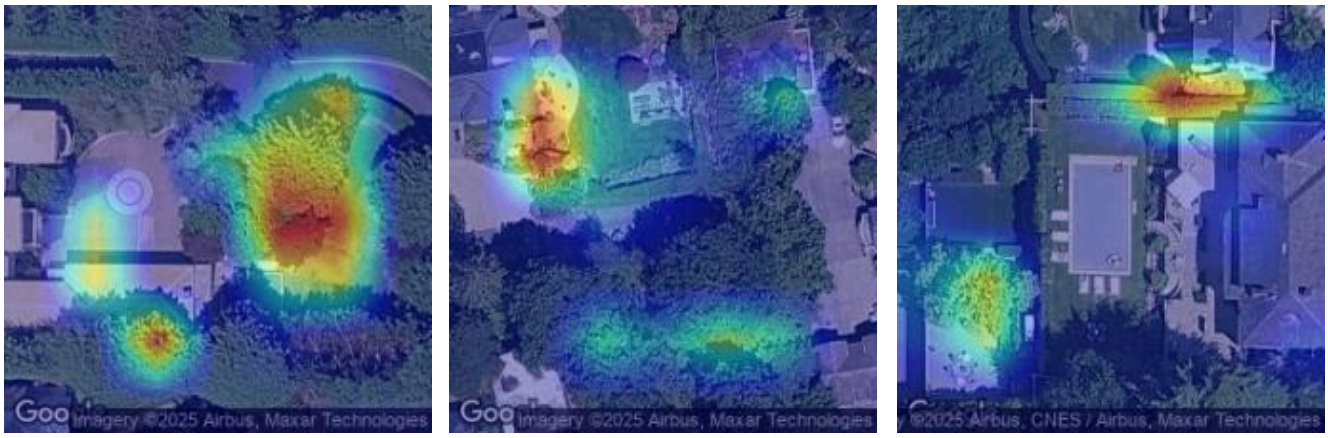
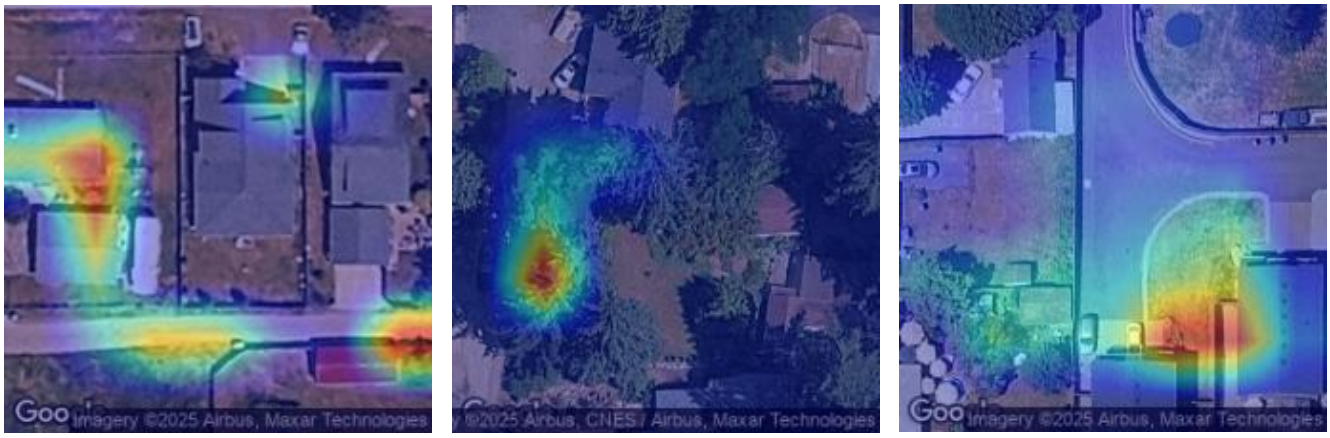


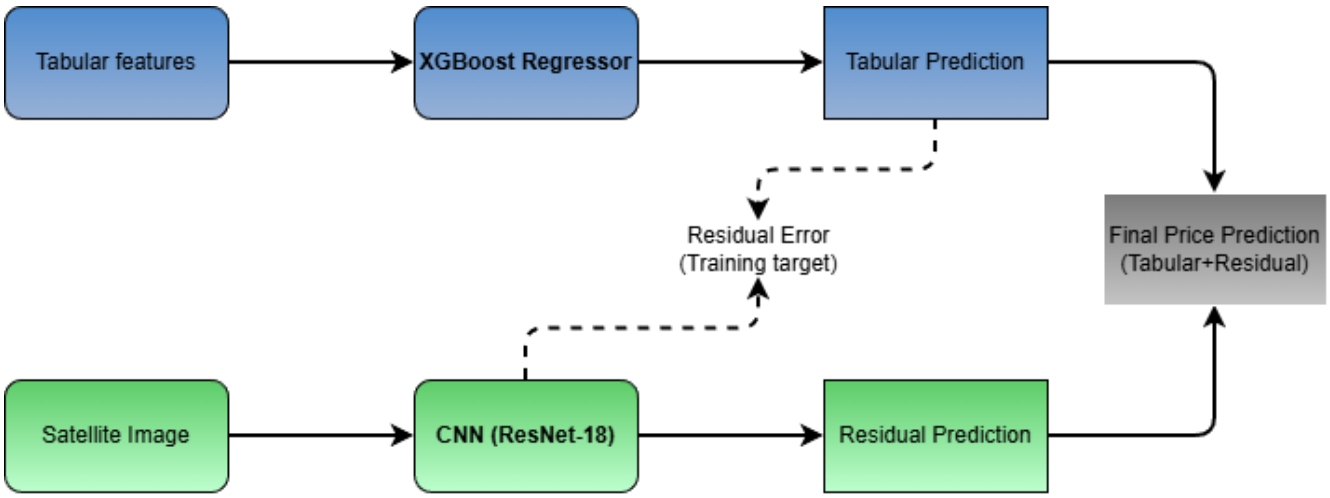
Figure 6: Grad-CAM Visualization for Low-Value Property



• Architecture Diagram

The final system adopts a **residual multimodal fusion architecture**:

Figure 7: Multimodal Residual Fusion Architecture
The tabular XGBoost model produces an initial price estimate, while a CNN trained on satellite imagery predicts residual errors based on environmental context. The final price is obtained by combining the tabular prediction with the image-based residual correction.



This design allows the CNN to specialize in capturing **environmental information not present in structured features**, leading to improved accuracy and stable optimization.

• Results

Model performance is evaluated using **RMSE** and **R^2 score** on a held-out validation set.

Model	RMSE ↓	R^2 ↑
Residual Fusion	**0.1517**	**0.9171**
Hybrid Fusion (XGB)	0.1622	0.9052
Tabular (XGB)	0.1707	0.8950
Late Fusion (NN)	0.1989	0.8575
Stacked Generalization	0.2399	0.7927
Early Fusion (NN)	0.2876	0.7021

The tabular-only model provides a strong baseline, indicating that structured features already capture much of the pricing signal. Naive multimodal approaches underperform due to optimization challenges and overfitting. The **residual fusion model achieves the best performance**, demonstrating that satellite imagery adds complementary value when integrated carefully.
Although the absolute improvement in RMSE appears modest, even small reductions in error are meaningful in real estate valuation, where prediction errors translate to significant monetary differences.

• Conclusion

This project demonstrates that **multimodal learning can enhance real estate price prediction** by incorporating environmental context from satellite imagery. Through careful feature engineering, residual learning, and explainability analysis, the proposed framework improves predictive accuracy while remaining interpretable and robust. The approach highlights the importance of principled fusion strategies and provides a scalable methodology for spatial-economic modelling tasks.
