Satellite Imagery–Based Property Valuation Using Residual Multimodal Learning

Submitted By : Milan Bambhaniya (23112026)

1 Problem Statement

Property price prediction is usually done using tabular housing data such as area, number of rooms, location, and construction quality. These features explain most of the property value, but they ignore the surrounding environment. As a result, houses with similar structural features can still have different market prices due to differences in neighborhood, greenery, road density, or nearby water bodies.

The goal of this project is to predict residential property prices by using both tabular housing data and satellite images, and to generate final price predictions for an unseen test dataset.

2 Overview of the Approach

To solve this problem, a multimodal machine learning system is developed using two types of data:

- Tabular housing features that describe the structure and location of a property
- Satellite images that capture environmental and neighborhood context

Satellite images are obtained using the ESRI imagery service, and deep learning is used to extract meaningful visual features from these images. The final output of the system is a CSV file containing predicted prices for the test set.

3 Modeling Strategy

The core strategy of this project is to treat tabular data as the main driver of property value and use satellite imagery only as a supporting signal. A tabular model is first trained to predict the base price of a property. This model captures most of the price variation.

Instead of directly combining tabular and image features, a residual learning approach is used. The difference between the true price and the tabular model prediction is learned using satellite image features. This allows visual information to correct the remaining errors without overpowering the strong structural signals.

This strategy results in a stable, interpretable, and realistic multimodal valuation model.

4 Data Description

The dataset consists of residential property records along with their geographic coordinates. The task is to predict property prices for an unseen test dataset using information learned from the training data.

The training dataset contains approximately 16,200 records and includes the target variable price. Each training sample is associated with a corresponding satellite image downloaded using the ESRI imagery service.

The test dataset contains approximately 5,400 records with the same tabular features as the training data but without price labels. Satellite images for the test data are obtained using the same extraction logic to maintain consistency between training and inference.

Both training and test datasets follow the same feature structure, enabling a unified preprocessing and modeling pipeline.

5 Preprocessing and Feature Engineering

5.1 Tabular Data: Feature Engineering

5.1.1 Baseline Handling

- No significant missing values were observed
- No imputation or outlier removal was performed
- Non-informative columns (id, date) were removed
- Same feature set maintained for both train and test data

5.1.2 New Feature From Exixting Feature

Feature Name	Description	Why It Was Created
house_age	Current year − year built	Captures depreciation and age-related value changes
is_renovated	Binary indicator of renovation	Renovated houses often command higher prices
total_sqft	Above-ground + basement area	Represents total usable living space
room_density	Living area ÷ (bedrooms + 1)	Proxy for spaciousness and layout quality

5.2 Satellite Images: Preprocessing and Feature Design

5.2.1 Standardization

- All images resized to **224 × 224 pixels**
- Ensures consistent input to the CNN

5.2.2 Deep Visual Features

- Pretrained **ResNet50** used as a feature extractor
- Outputs high-level visual representations of surroundings
- No fine-tuning performed to avoid overfitting

5.2.3 Dimensionality Reduction

- CNN features compressed using **PCA**
- PCA fitted **only on training data**
- Same transformation applied to test data

5.2.4 Interpretable Visual Signals

To complement CNN features, simple visual indicators were computed:

Visual Feature	Captures
Green ratio	Vegetation and parks
Blue ratio	Water bodies
Edge density	Roads and building concentration
Brightness	Overall urban intensity

These signals provide **human-interpretable context** that supports the residual learning framework.

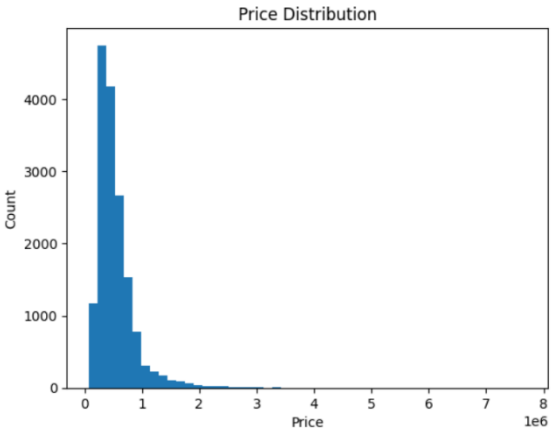
6 Exploratory Data Analysis (Structural Signals)

This section analyzes how property prices relate to key structural and location-based features. The objective is to understand whether tabular features capture meaningful price patterns and justify their role as the primary predictors in the model.

6.1 Price Distribution

The price distribution shows a strong right-skewed pattern. Most properties fall in the lower to mid-price range, while a small number of high-value properties extend the tail of the distribution.

Insight:
This skewed but continuous distribution supports the use of regression models and indicates that extreme values exist but do not dominate the dataset.



6.2 Living Area vs Price (sqft_living)

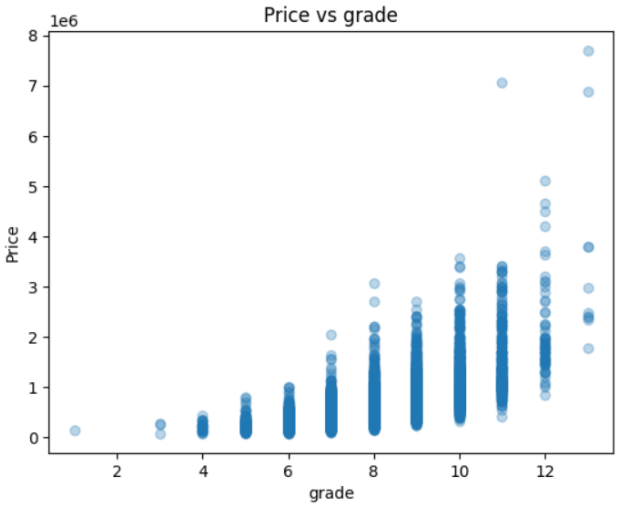
A clear positive relationship is observed between living area and property price. As the living area increases, prices generally rise, though the spread becomes wider for larger homes.

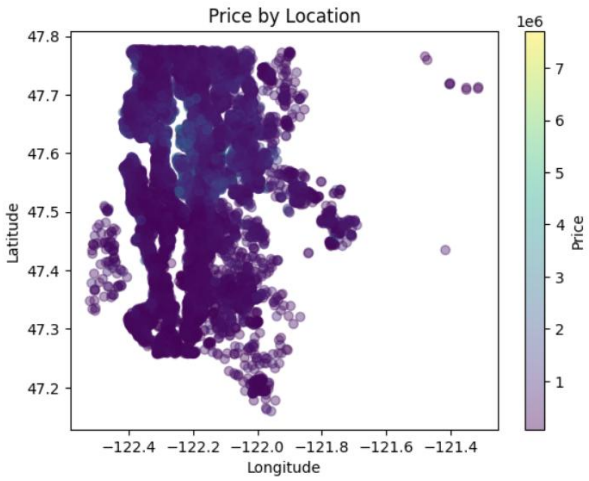
Insight:
Living area is a strong predictor of price, but the increasing variance at higher square footage suggests that size alone does not fully explain price differences. This motivates the use of additional features and non-linear models.

6.3 Construction Grade vs Price

Property price increases consistently with higher construction grades. Each increase in grade corresponds to a noticeable upward shift in price levels.

Insight:
Construction quality is a highly informative feature and explains large price jumps. This confirms that engineered and categorical structural features play a critical role in valuation.





6.4 Geographic Distribution of Prices

The location plot shows strong spatial clustering of prices. Certain geographic regions consistently exhibit higher prices, while others remain in lower price bands despite having similar structural features.

Insight:
Location plays a significant role in property valuation, but its effect is complex and non-linear. Latitude and longitude capture broad spatial trends but fail to describe fine-grained neighborhood characteristics.

6.5 EDA Summary and Its Role in Model Design

Exploratory data analysis helped clarify how different factors influence property prices and directly shaped the modeling approach. Structural features such as living area, construction grade, and total space showed strong and consistent relationships with price, confirming that tabular data should act as the primary source of prediction. At the same time, EDA revealed non-linear behavior and increasing variance for larger or higher-grade properties, indicating the need for flexible, non-linear models. Geographic analysis showed clear spatial price clusters but also highlighted the limitations of latitude and longitude in representing neighborhood quality. Most importantly, EDA demonstrated that properties with similar structural characteristics can still have different prices, pointing to missing contextual information beyond tabular data.

Key takeaways from EDA and how they were used:

- Strong tabular feature–price relationships → **Tabular model chosen as the base predictor**
- Non-linear and heteroscedastic patterns → **Tree-based models used instead of linear models**
- Spatial clustering with location limitations → **Motivation to add contextual information**
- Price variation among structurally similar houses → **Satellite imagery introduced**
- Limited but complementary visual signal → **Residual learning used instead of direct feature fusion**

These insights led to a modeling strategy where tabular data captures the core property value, and satellite imagery is used only to correct remaining prediction errors in a controlled and stable manner.

7 Satellite Imagery: Challenges and Design Implications

Satellite imagery provides useful neighborhood context, but it also has several limitations that must be considered when designing a valuation model.

7.1 Limitations of Raw Satellite Images

- Satellite images contain **visual noise** such as shadows, lighting differences, and seasonal variation that are unrelated to property value.
- **Resolution and image quality** are inconsistent across locations, which reduces reliability.
- Images provide a **top-down view only**, missing interior quality and street-level factors.
- On their own, satellite images are **weak predictors** compared to structural tabular features.

7.2 Implications for Model Design

- End-to-end CNN training was avoided to reduce overfitting and noise sensitivity.
- A pretrained CNN was used only for **feature extraction**, not direct price prediction.
- Visual features were compressed and treated as **secondary information** rather than primary drivers.

Key outcome:
These limitations motivated a **restrained, correction-oriented use of visual data**, where satellite imagery refines tabular predictions instead of dominating the valuation process.

8 Visual Feature Engineering

Satellite images contain rich contextual information, but raw pixel values are difficult to interpret and often noisy. To make visual data useful for property valuation, image information was converted into structured features that balance expressive power with interpretability.

8.1 CNN Feature Extraction

A pretrained **ResNet50** model was used to extract high-level visual representations from satellite images. The network was not fine-tuned and was used only as a feature extractor to reduce overfitting and avoid learning noise specific to the dataset.

Each image was transformed into a **2048-dimensional feature vector**, capturing spatial patterns such as vegetation coverage, road layouts, and built-up regions. To reduce redundancy and stabilize learning, these features were compressed using **Principal Component Analysis (PCA)**. PCA was fitted only on the training data, and the top **64 components** were retained.

8.2 Interpretable Visual Features

While CNN features capture complex patterns, they remain difficult to interpret. To reduce black-box behavior and align visual information with real-world meaning, additional **interpretable visual features** were created.

These features were designed to explicitly represent neighborhood characteristics that influence property value but are not available in tabular data.

Feature	Description	Why It Was Created
green_ratio	Proportion of vegetation and green areas	Green surroundings are associated with higher livability and premium neighborhoods
blue_ratio	Presence of water bodies	Proximity to water often increases property desirability and price
edge_density	Density of roads and buildings	High density reflects congestion and heavy urbanization
brightness	Overall surface intensity	Differentiates dense concrete areas from balanced residential zones

Key Insight:
Interpretable features ground deep visual embeddings in economic intuition.

By combining compressed CNN features with interpretable visual signals, the model captures environmental context while remaining transparent and economically meaningful.

9 Residual Learning Framework and Architecture

This page describes the core modeling logic used to combine tabular data and satellite imagery in a controlled and interpretable way.

9.1 Base Tabular Model

The modeling process begins with a machine learning model trained only on **tabular housing features**. These features capture the structural and economic characteristics of a property, such as size, quality, and location coordinates.

This model predicts a **base price**, which represents the portion of property value that can be explained using tabular data alone. Since tabular features show strong relationships with price, this base model captures most of the overall price variation.

9.2 Residual Modeling with Images

After predicting the base price, the remaining error—known as the **residual**—represents information that the tabular model fails to explain. These residuals often arise due to neighborhood quality and environmental context, which are not fully captured by tabular data.

A second model is trained using **satellite image features** as input and the **residuals** as the target. Instead of predicting prices directly, this image-based model learns how much the base price should be corrected based on visual surroundings.

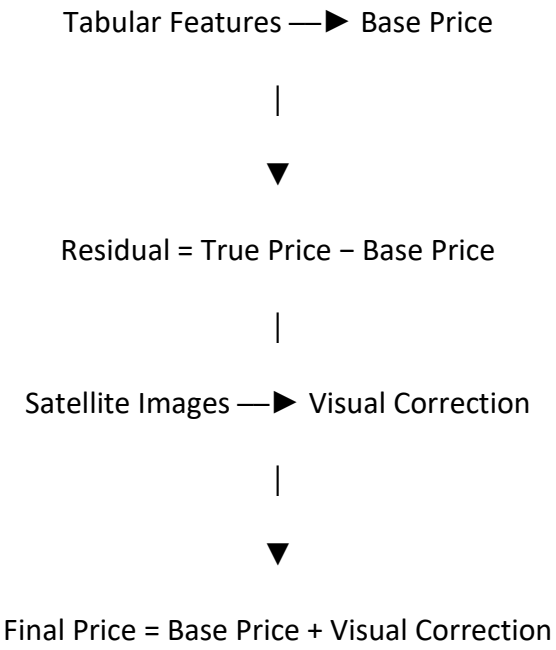
This design ensures that satellite imagery contributes only **incremental corrections**, rather than competing with strong structural signals.

9.3 Final Prediction Logic

The final property price is obtained by combining both components:

- The **base price** predicted from tabular features
- The **visual correction** predicted from satellite image features

This additive approach enforces a clear separation of responsibilities between data modalities and improves stability.



10 Model Training and Validation

10.1 Model Choices

Both the base tabular model and the image residual model were implemented using **XGBoost regression**. XGBoost was chosen due to its strong performance on structured data, ability to model non-linear relationships, and robustness to feature scaling. Using the same model family for both stages ensured consistency in learning behavior while allowing each model to focus on a different source of information.

10.2 Validation Strategy

Model performance was evaluated using a held-out validation split from the training data. The tabular model was trained first and evaluated independently to establish a strong baseline. Residuals from this model were then computed on the same split and used as targets for training the image residual model.

Final validation performance was measured by combining the base tabular predictions with the predicted visual corrections. This two-stage evaluation ensured that improvements from satellite imagery were assessed realistically and without leakage.

11 Results

11.1 Validation Performance

Model	R^2 (Validation)	RMSE (Validation)
Tabular Only	0.8901	117,445
Tabular + Satellite (Residual)	0.8913	116,810

```
[23]
✓ 0s
from sklearn.metrics import r2_score, mean_squared_error
import numpy as np

tab_val_pred = tab_model.predict(X_tab.loc[val_idx])

r2 = r2_score(y.loc[val_idx], tab_val_pred)
rmse = np.sqrt(mean_squared_error(y.loc[val_idx], tab_val_pred))

print("Tabular Validation R²:", r2)
print("Tabular Validation RMSE:", rmse)

Tabular Validation R²: 0.8900825381278992
Tabular Validation RMSE: 117445.25328850035
```

```
[27]
✓ 0s
from sklearn.metrics import r2_score, mean_squared_error
import numpy as np

residual_val_pred = img_model.predict(X_image[val_idx])
final_val_pred = tab_val_pred + residual_val_pred

r2_final = r2_score(y.loc[val_idx], final_val_pred)
rmse_final = np.sqrt(mean_squared_error(y.loc[val_idx], final_val_pred))

print("FINAL Multimodal Validation R²:", r2_final)
print("FINAL Multimodal Validation RMSE:", rmse_final)

FINAL Multimodal Validation R²: 0.8912675380706787
FINAL Multimodal Validation RMSE: 116810.4499434875
```

11.2 Result Interpretation

The multimodal model shows a **consistent improvement** over the tabular-only baseline. The increase in R^2, along with a reduction in RMSE, indicates that satellite imagery contributes additional information that helps refine price predictions.

Importantly, the improvement is **modest rather than dramatic**. This behavior is expected and desirable, as tabular features already explain most of the price variation. The role of satellite imagery is therefore not to replace structural valuation, but to correct remaining errors related to neighborhood and environmental context.

Key interpretation:
The modest gain confirms complementary signal integration rather than metric inflation.

These results validate the residual learning strategy, demonstrating that controlled use of visual information can improve model performance while preserving stability and interpretability.

12 Discussion and Conclusion

12.1 Discussion

This project demonstrates that multimodal learning for property valuation is most effective when each data source is used according to its true explanatory strength. Tabular features such as size, quality, and location coordinates capture the core determinants of property value. Satellite imagery, while informative, contains noise and limited standalone predictive power.

By using a residual learning framework, the model ensures that satellite imagery contributes only where tabular data falls short. This design improves robustness, avoids overfitting, and maintains interpretability. The small performance improvement observed is therefore a sign of a well-calibrated model rather than a limitation.

12.2 Limitations

- Satellite images provide only a top-down view and cannot capture interior quality.
- Image resolution and update frequency may vary across locations.
- The approach is better suited for residential properties than high-rise or commercial buildings.

12.3 Conclusion

In conclusion, this project presents a disciplined and realistic approach to multimodal property valuation. By treating satellite imagery as a corrective signal rather than a primary predictor, the model balances performance, interpretability, and stability. The resulting system acts as a **valuation lens that integrates structure and surroundings**, offering a practical framework for real-world real estate analytics.