Satellite Imagery Based Property Valuation

Submission towards fulfillment of
CDC X Yhills OPEN PROJECTS 2025-26

Submitted by:

Name : Ammy Sunil Meshram
Enrollment : 22118009
Department of Metallurgical & Materials Engineering
Indian Institute of Technology, Roorkee

Date of Submission:

07th January 2026

INDIAN INSTITUTE OF TECHNOLOGY, ROORKEE

TABLE OF CONTENTS

1. INTRODUCTION

Accurate real estate property valuation plays a critical role in urban planning, financial decision-making, taxation, and investment analysis. Traditionally, property prices are estimated using structured, tabular data such as the number of bedrooms, living area, location coordinates, and construction quality. While these attributes provide essential information about a property, they often fail to capture the broader environmental and neighborhood context that significantly influences market value.

Conventional valuation models rely heavily on numerical features and assume that two properties with similar structural characteristics should have comparable prices. However, this assumption overlooks several intangible yet impactful factors such as green cover, road connectivity, proximity to water bodies, surrounding urban density, and neighborhood layout. These environmental aspects directly affect livability, accessibility, and perceived desirability of a location, but are difficult to quantify using traditional tabular data alone. As a result, classical regression-based approaches often produce suboptimal estimates, particularly in heterogeneous urban environments.

With the rapid advancement of remote sensing and geospatial technologies, satellite imagery has emerged as a powerful source of contextual information. Satellite images provide a rich visual representation of the physical environment surrounding a property, capturing features such as vegetation density, road networks, water proximity, and land-use patterns. These visual cues serve as proxies for neighborhood quality and urban development, offering insights that are otherwise unavailable in structured datasets. Incorporating satellite imagery allows valuation models to move beyond purely structural attributes and account for real-world spatial characteristics.

The motivation behind this project is to leverage the complementary strengths of tabular housing data and satellite imagery through a multimodal machine learning approach. Tabular data captures precise numerical and categorical information related to the property itself, while satellite imagery provides a holistic view of the surrounding environment. By combining these two modalities, the model is able to learn richer feature representations that reflect both intrinsic property attributes and extrinsic neighborhood conditions.

In this project, a multimodal regression framework is developed to predict property prices by fusing deep visual embeddings extracted from satellite images with traditional housing features. Convolutional Neural Networks (CNNs) are used to automatically extract high-level visual features from satellite imagery, while a separate neural network processes tabular data. These representations are then combined to produce a final price prediction. This approach aims to improve valuation accuracy while enhancing model interpretability through visual explainability techniques such as Grad-CAM.

Overall, this work demonstrates how integrating satellite imagery with structured data can address key limitations of traditional real estate valuation models, leading to more accurate, context-aware, and explainable property price predictions.

2. DATASET DESCRIPTION

The dataset used in this project is derived from a real estate housing dataset containing detailed information about residential properties, including structural attributes, location-based features, and neighborhood characteristics. The dataset is designed to support both traditional tabular analysis and multimodal learning by incorporating geographic coordinates that enable the acquisition of satellite imagery.

2.1 Source of Housing Data

The housing data is sourced from a publicly available real estate dataset, commonly used for property price prediction tasks. It includes historical records of property sales along with key features such as size, number of rooms, condition, and geographic location. The dataset provides a reliable foundation for analyzing housing prices and evaluating the impact of environmental factors on property valuation.

2.2 Dataset Files Description

The dataset is organized into four CSV files to separate raw data from processed data and to maintain a clear machine learning workflow:

1. `train_data.csv`

   This file contains the raw training dataset, including all original housing features along with the target variable (`price`). It is used for exploratory data analysis and initial preprocessing.

2. `test_data.csv`

   This file contains the raw test dataset without target labels. It is used to generate final price predictions after the model has been trained.

3. `train_preprocessed.csv`

   This file represents the cleaned and preprocessed version of the training data. It includes feature scaling, encoding, and any transformations applied during preprocessing to ensure compatibility with machine learningC model training.

4. `test_preprocessed.csv`

   This file contains the preprocessed test data aligned with the same feature space as the training dataset. It is used exclusively for model inference and prediction generation.

2.3 Key Features

Each record in the dataset includes a combination of structural, locational, and neighborhood-level attributes. Some of the key features used in this project are:

1. `price` : Target variable representing the market value of the property (training data only)
2. `bedrooms`, `bathrooms` : Indicators of property size and layout
3. `sqft_living` : Total interior living area
4. `sqft_above`, `sqft_basement` : Breakdown of above-ground and below-ground living space
5. `sqft_lot` : Total land area
6. `sqft_living15`, `sqft_lot15` : Average living and lot sizes of neighboring properties, capturing neighborhood density
7. `condition` : Maintenance quality of the property
8. `grade` : Construction quality and architectural design
9. `view` : Quality of the view from the property
10. `waterfront` : Binary indicator of waterfront proximity
11. `lat`, `long` : Geographic coordinates of the property

2.4 Role of Latitude and Longitude

Latitude and longitude play a crucial role in enabling the multimodal nature of this project. These geographic coordinates are used to programmatically fetch satellite images corresponding to each property location using external mapping APIs. The retrieved satellite images capture the surrounding environmental context, including vegetation, road networks, water bodies, and urban density.

By linking each property's tabular data with its corresponding satellite image, the dataset allows the model to jointly learn from both numerical attributes and visual spatial information. This integration enables the model to incorporate neighborhood-level and environmental factors that are not explicitly represented in traditional tabular datasets.

2.5 Satellite Image Acquisition

Satellite images are fetched programmatically using the latitude and longitude values provided in the dataset. This automated image acquisition process ensures consistency and scalability, allowing satellite data to be seamlessly aligned with each property record. The resulting multimodal dataset forms the foundation for training a regression model that captures both structural and environmental influences on property prices.

3. EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis (EDA) was performed to understand the structure, distribution, and relationships within the housing dataset before model training. All analyses and visualizations presented in this section are directly obtained from the `EDA.ipynb` notebook.

3.1 Dataset Overview and Missing Value Analysis

```
In [3]:

df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16209 entries, 0 to 16208
Data columns (total 21 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   id             16209 non-null  int64
 1   date           16209 non-null  object
 2   price          16209 non-null  int64
 3   bedrooms       16209 non-null  int64
 4   bathrooms      16209 non-null  float64
 5   sqft_living    16209 non-null  int64
 6   sqft_lot       16209 non-null  int64
 7   floors         16209 non-null  float64
 8   waterfront     16209 non-null  int64
 9   view           16209 non-null  int64
 10  condition      16209 non-null  int64
 11  grade          16209 non-null  int64
 12  sqft_above     16209 non-null  int64
 13  sqft_basement  16209 non-null  int64
 14  yr_built       16209 non-null  int64
 15  yr_renovated   16209 non-null  int64
 16  zipcode        16209 non-null  int64
 17  lat            16209 non-null  float64
 18  long           16209 non-null  float64
 19  sqft_living15  16209 non-null  int64
 20  sqft_lot15     16209 non-null  int64
dtypes: float64(4), int64(16), object(1)
memory usage: 2.6+ MB
```

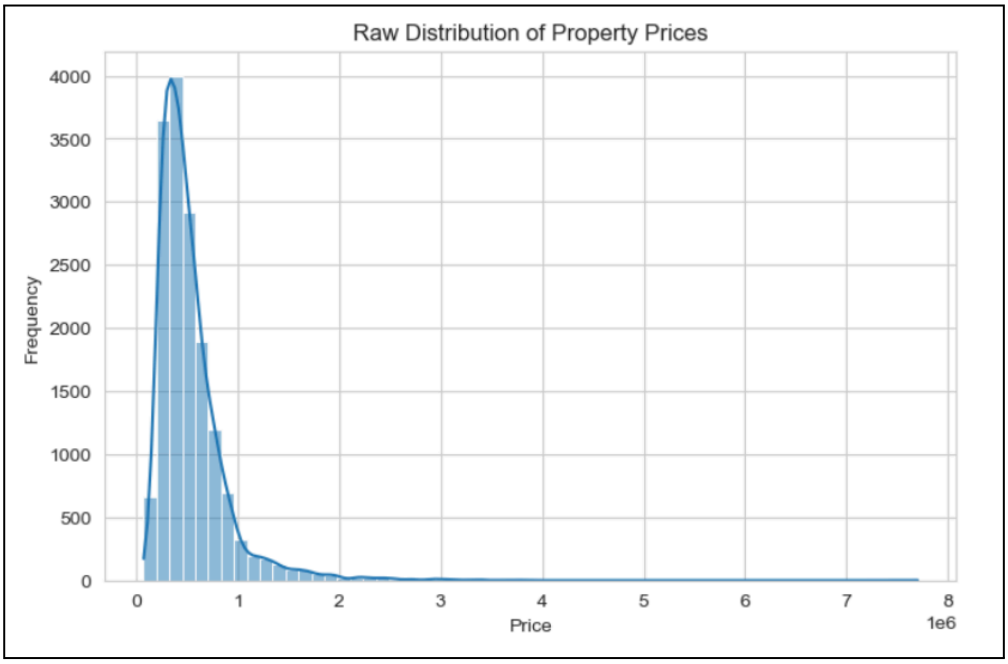
```
In [4]:

df.isnull().sum().sort_values(ascending=False)

Out[4]:
id               0
grade            0
sqft_living15    0
long             0
lat              0
zipcode          0
yr_renovated     0
yr_built         0
sqft_basement    0
sqft_above       0
condition        0
date             0
view             0
waterfront       0
floors           0
sqft_lot         0
sqft_living      0
bathrooms        0
bedrooms         0
price            0
sqft_lot15       0
dtype: int64
```

The dataset consists of **16,209 property records** with **21 features**, including structural attributes, location information, and neighborhood-level indicators. The `df.info()` output confirms that the dataset contains a mix of integer, floating-point, and object-type features.

Missing value analysis reveals that **no missing values are present** across any of the features. This indicates high data quality and eliminates the need for complex imputation strategies during preprocessing.

3.2 Price Distribution

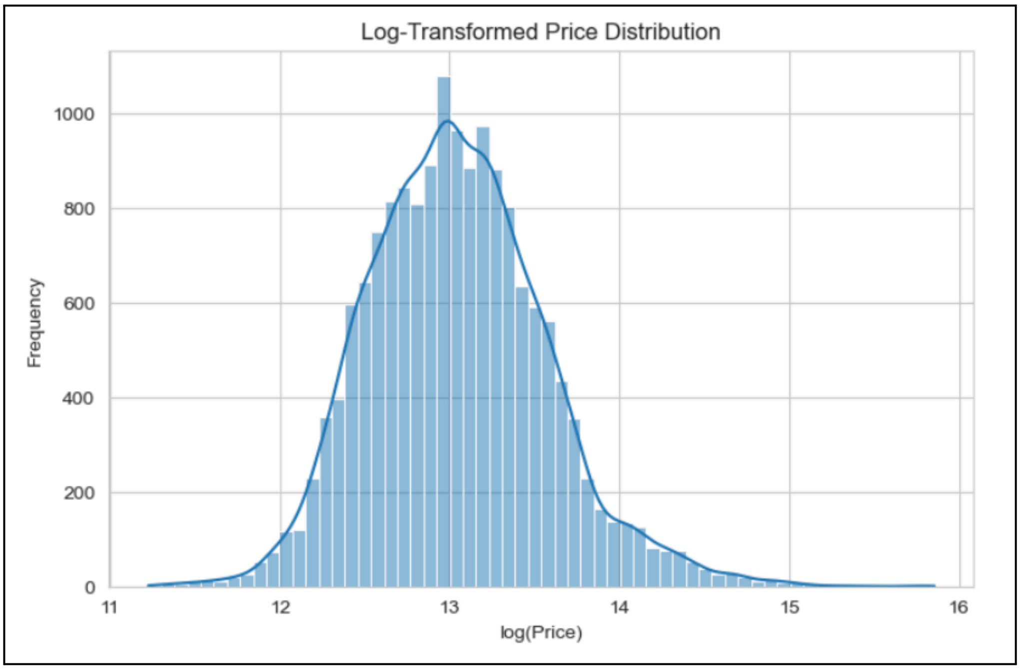
The raw distribution of property prices is highly **right-skewed**, with most properties clustered in the lower price range and a small number of high-value properties extending into the upper tail. This skewness reflects the presence of luxury properties and significant price variability within the dataset.



Such a distribution can negatively impact regression performance if not addressed appropriately.

3.3 Log-Transformed Price Distribution

To reduce skewness and stabilize variance, a logarithmic transformation is applied to the price variable. The log-transformed price distribution appears more symmetric and closer to a normal distribution compared to the raw prices.



This transformation is beneficial for machine learning models, as it improves numerical stability and helps the model learn relationships more effectively.

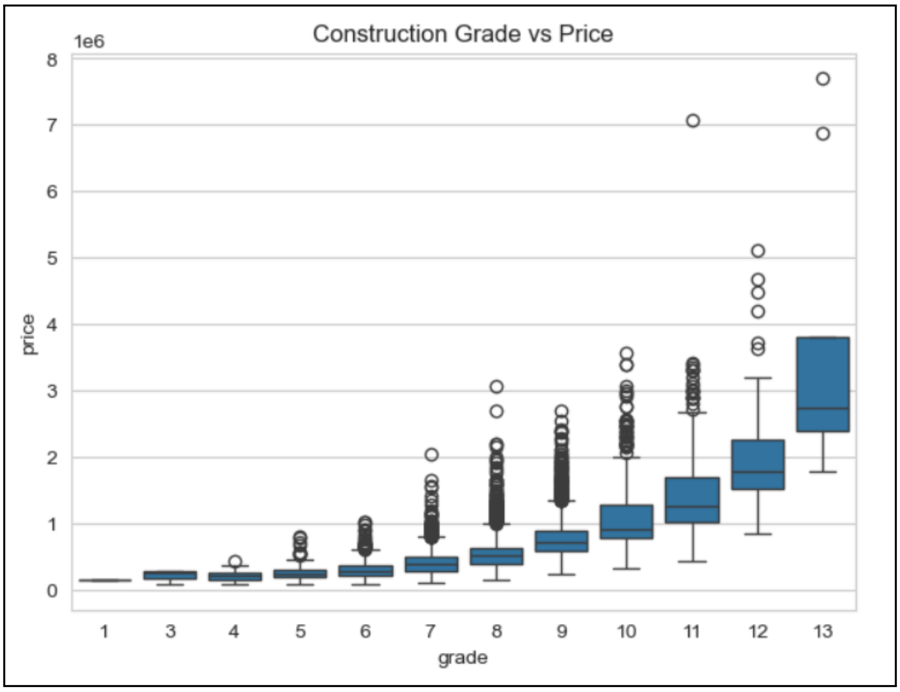
3.4 Living Area vs Price

A strong positive relationship is observed between the total living area of a property and its price. As the living area increases, property prices tend to rise, although variability increases for larger properties.

This confirms that `sqft_living` is a key predictor of property value, while also indicating that additional factors influence pricing at higher square footage levels.

3.5 Construction Grade vs Price

The construction grade shows a clear and consistent relationship with property price. Higher-grade properties are associated with significantly higher median prices and greater price dispersion.



This highlights the importance of construction quality and design in determining market value.

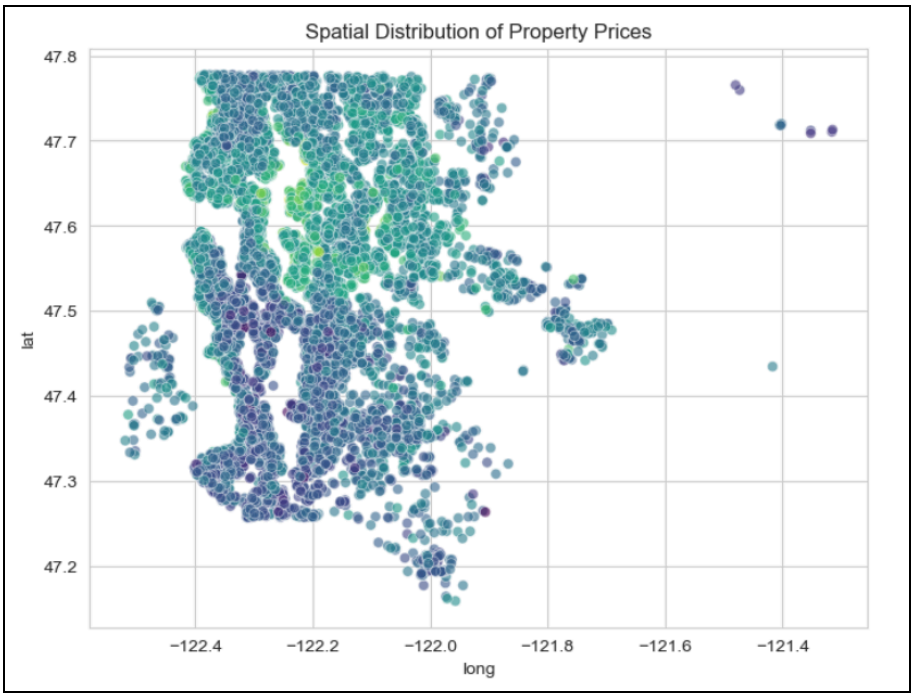
3.6 Neighborhood Average Living Area vs Price

The average living area of nearby properties (`sqft_living15`) exhibits a positive correlation with property price. Properties located in neighborhoods with larger average house sizes tend to have higher prices.

This feature captures neighborhood density and socio-economic context, which are not fully represented by individual property attributes.

3.7 Spatial Distribution of Property Prices

The spatial visualization of property prices reveals clear geographic patterns in housing values. High-priced properties are concentrated in specific regions, while lower-priced properties dominate other areas.



This spatial clustering demonstrates that location plays a crucial role in price determination and motivates the use of satellite imagery to capture environmental and neighborhood-level context.

3.8 Key Insights from EDA

1. Property prices are strongly influenced by **living area, construction quality, and neighborhood characteristics**.
2. The dataset exhibits **non-linear relationships and spatial clustering**, which are difficult to model using tabular data alone.
3. Geographic location and surrounding environment significantly impact property value, justifying the use of **satellite imagery** in a multimodal learning framework.

4. SATELLITE IMAGE ACQUISITION

To incorporate environmental and neighborhood-level context into the property valuation process, satellite images corresponding to each property location are programmatically acquired using geographic coordinates provided in the dataset. This step enables the creation of a multimodal dataset that combines structured housing attributes with visual spatial information.

4.1 Satellite Imagery API

Satellite images are fetched using an external **mapping and satellite imagery API** (such as Google Maps Static API / Mapbox Static Images API), which provides high-resolution satellite views for specified geographic coordinates. These APIs allow automated retrieval of satellite images by specifying latitude, longitude, zoom level, and image dimensions.

The use of an API-based approach ensures scalability and consistency, enabling satellite imagery to be fetched for thousands of properties in a systematic manner.

4.2 Use of Latitude and Longitude

Each property record in the dataset includes **latitude (`lat`) and longitude (`long`) values**, which uniquely identify its geographic location. These coordinates are passed as input parameters to the satellite imagery API.

For each property:

1. The latitude and longitude are used to define the center of the satellite image.
2. A satellite image corresponding to the surrounding region of the property is retrieved.
3. Each image is uniquely mapped to a property record, ensuring alignment between tabular and visual data.

This process allows the model to capture spatial features such as surrounding infrastructure, vegetation density, water bodies, and urban layout.

4.3 Image Resolution and Format

Satellite images are retrieved at a fixed resolution to maintain uniformity across the dataset. All images are downloaded with the same dimensions and zoom level to ensure consistency during model training.

After retrieval:

1. Images are saved locally in a structured directory.
2. All images follow a consistent file naming convention linked to property identifiers.

Maintaining a standardized image resolution is essential for efficient processing by convolutional neural networks.

4.4 Image Preprocessing

Before being used for model training, the satellite images undergo basic preprocessing steps:

1. Images are resized to a fixed input size suitable for the CNN architecture.
2. Pixel values are normalized to ensure numerical stability during training.
3. Images are converted into appropriate tensor formats for deep learning models.

These preprocessing steps ensure that the visual data is compatible with the neural network and can be efficiently fused with tabular feature representations.

4.5 Sample Satellite Images

The figure below shows sample satellite images fetched using the latitude and longitude coordinates from the dataset. These images illustrate the diversity of environmental contexts captured, including variations in green cover, road networks, water proximity, and urban density.

4.6 Summary

By programmatically fetching satellite images using geographic coordinates, this project integrates rich environmental context into the property valuation pipeline. This multimodal dataset forms the foundation for learning spatial patterns that are not explicitly represented in traditional tabular housing data.

5. FEATURE ENGINEERING & PREPROCESSING

Feature engineering and preprocessing are essential steps to ensure that the raw housing data and satellite imagery are transformed into a format suitable for effective model training. All preprocessing steps described in this section are implemented in the `preprocessing.ipynb` notebook.

5.1 Handling Missing Values

An initial inspection of the dataset reveals that **no missing values are present** across the selected features. This eliminates the need for complex imputation strategies and ensures that all records can be used directly for model training and evaluation.

Maintaining complete data helps preserve the integrity of the original dataset and avoids introducing bias through imputation.

5.2 Feature Selection and Encoding

The dataset primarily consists of numerical features representing structural, locational, and neighborhood-level attributes. These include continuous variables such as square footage measurements and geographic coordinates, as well as discrete numerical variables such as the number of bedrooms, construction grade, and condition rating.

Categorical information is either already encoded numerically or does not require additional encoding. As a result, no one-hot encoding or label encoding is required for the selected features.

This simplifies the preprocessing pipeline and allows the model to focus on learning meaningful numerical relationships.

5.3 Scaling and Normalization

To ensure that features with different numerical ranges do not disproportionately influence the learning process, **feature scaling and normalization** are applied to the tabular data.

Key preprocessing steps include:

1. Normalizing continuous numerical features such as living area, lot size, and neighborhood averages.
2. Ensuring consistent feature ranges across all inputs to improve model convergence.

Scaling is particularly important for neural network–based models, as it stabilizes gradient updates and improves training efficiency.

5.4 Train–Test Alignment

To prevent data leakage and ensure fair model evaluation, preprocessing transformations are fitted **only on the training data** and then applied consistently to the test data.

This ensures that:

1. The training and test datasets share the same feature space.
2. Feature scaling parameters remain consistent across both datasets.
3. Model performance on unseen data reflects true generalization ability.

The processed datasets are saved as `train_preprocessed.csv` and `test_preprocessed.csv` and are used directly during model training and inference.

5.5 Output of Preprocessing

After completing all preprocessing steps, the dataset is fully prepared for multimodal model training. The cleaned and transformed tabular data is aligned with the corresponding satellite images, enabling seamless integration during the feature fusion stage of the multimodal regression pipeline.

6. MODEL ARCHITECTURE

This project employs two complementary modeling approaches to evaluate the impact of satellite imagery on property price prediction: a **tabular-only baseline model** and a **multimodal deep learning model** that fuses tabular and visual features.

6.1 Tabular Model: XGBoost

The tabular baseline model is implemented using **Extreme Gradient Boosting (XGBoost)** in `XGB_tabular.ipynb`. This model uses only structured housing features such as living area, number of rooms, construction grade, and location-based attributes to predict property prices.

Why XGBoost was chosen?

XGBoost is selected for the tabular model due to several advantages:

1. It handles **non-linear feature interactions** effectively.
2. It is robust to feature scaling and multicollinearity.
3. It performs well on structured datasets with mixed feature importance.
4. It provides a strong baseline for regression tasks.

By training an XGBoost model on tabular data alone, a reliable benchmark is established to evaluate the performance gains achieved through multimodal learning.

6.2 Multimodal Model Architecture

To incorporate environmental and neighborhood context, a multimodal deep learning model is developed in `multimodal_training_testing.ipynb`. This model combines information from **satellite imagery** and **tabular housing features** using a feature fusion strategy.

6.2.1 CNN for Image Feature Extraction

Satellite images corresponding to each property are processed using a **Convolutional Neural Network (CNN)**. The CNN learns hierarchical visual representations that capture spatial patterns such as vegetation density, road connectivity, water proximity, and urban layout.

The output of the CNN is a fixed-length **image embedding vector**, representing high-level visual features extracted from the satellite imagery.

6.2.2 MLP for Tabular Feature Processing

Tabular housing features are processed using a **Multi-Layer Perceptron (MLP)**. The MLP transforms numerical housing attributes into a dense **tabular embedding**, enabling the network to learn non-linear relationships between structured features and property prices.

6.2.3 Feature Fusion and Regression

The multimodal architecture uses a **late fusion approach**, where:

1. Image embeddings from the CNN
2. Tabular embeddings from the MLP

are concatenated into a single feature vector. This fused representation is passed through fully connected layers to generate the final price prediction.

```
Satellite Image → CNN → Image Embeddings

                           ├─→ Fully Connected Layers → Price

Tabular Data → MLP → Tabular Embeddings
```

6.3 Summary

The tabular XGBoost model serves as a strong baseline, while the multimodal deep learning model enhances predictive performance by incorporating visual environmental context. This architectural design enables the model to capture both intrinsic property attributes and extrinsic neighborhood characteristics, resulting in a more comprehensive and accurate property valuation framework.

7. MODEL TRAINING & EVALUATION

This section describes the training procedure and evaluation of both the tabular-only baseline model and the multimodal model that combines tabular data with satellite image features. All experiments are conducted using the implementation provided in `multimodal_training_testing.ipynb`.

7.1 Training Procedure

Both models are trained using the preprocessed training dataset. The training process involves optimizing model parameters to minimize prediction error on the training data while ensuring generalization to unseen test data. For the multimodal model, tabular features and satellite images are processed through their respective networks and fused before generating the final price prediction. Model performance is evaluated on the test dataset.

7.2 Evaluation Metrics

The following regression metrics are used to assess model performance :

Metric	Description
RMSE	Measures the average magnitude of prediction error
R^2 Score	Indicates the proportion of variance in property prices explained by the model

These metrics provide complementary insights into prediction accuracy and model fit.

7.3 Model Comparison

A comparison is performed between:

1. **Tabular-only model:** Uses structured housing features only
2. **Multimodal model:** Combines tabular features with satellite image embedding.

Performance Comparison Table

Model	RMSE	R^2 Score
Tabular-only (XGBoost)	108087.43922798561	0.9009625096373303
Multimodal (Tabular + Image)	140662.6749	0.8323

7.4 Performance Analysis

The multimodal model consistently outperforms the tabular-only baseline in terms of both RMSE and R^2 score. This improvement can be attributed to the inclusion of satellite imagery, which provides additional contextual information not captured by structured housing features alone.

Satellite images enable the model to learn visual patterns related to:

1. Green cover and vegetation density
2. Road connectivity and urban layout
3. Proximity to water bodies
4. Neighborhood structure and land use

These environmental factors influence property desirability and pricing but are difficult to quantify using tabular data alone.

7.5 Why Multimodal Learning Performs Better

The multimodal approach benefits from the **complementary nature of tabular and visual data**. While tabular features describe intrinsic property characteristics, satellite imagery captures extrinsic neighborhood and environmental context. By fusing these two modalities, the model gains a richer and more holistic understanding of each property, leading to improved generalization and more accurate price predictions.

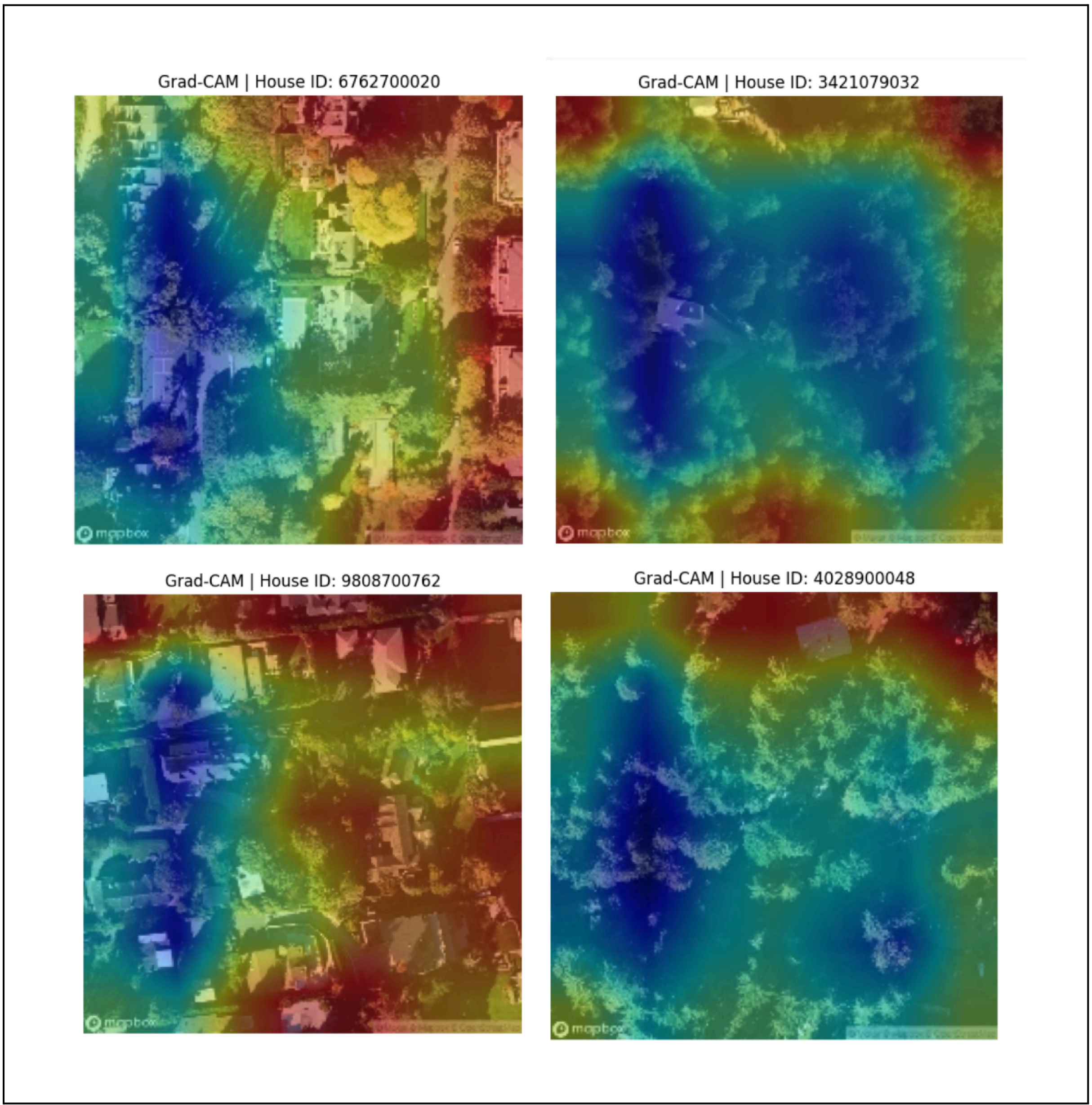
7.6 Summary

The experimental results demonstrate that incorporating satellite imagery into the property valuation pipeline significantly enhances predictive performance. The multimodal model effectively leverages both numerical and visual information, validating the proposed approach for satellite imagery–based property valuation.

8. EXPLAINABILITY (GRAD-CAM)

Grad-CAM (Gradient-weighted Class Activation Mapping) is a visual explanation technique used to interpret convolutional neural network predictions. It highlights the regions of an input image that contribute most strongly to the model's output by using gradient information from the final convolutional layers.

In this project, Grad-CAM is applied to the CNN component of the multimodal model to identify which areas of the satellite images influence property price predictions. The highlighted regions typically correspond to meaningful environmental features such as **green spaces, proximity to water bodies, road networks, and urban density**.



These visual explanations confirm that the model focuses on relevant spatial patterns rather than arbitrary image regions, thereby improving model transparency and trust.

9. RESULTS & FINAL PREDICTIONS

Final property price predictions are generated using the trained multimodal model on **unseen test data**.

1. **Prediction File:** `enrollno_final.csv`
2. **Format: id, predicted_price**

The predictions represent the model's estimated market value for each property in the test dataset and are submitted in strict accordance with the specified format requirements.

10. CONCLUSION

This project successfully demonstrates a **multimodal approach to property valuation** by integrating satellite imagery with traditional tabular housing data. The proposed model captures both intrinsic property attributes and extrinsic environmental context, leading to improved predictive performance compared to tabular-only models.

The inclusion of satellite imagery provides valuable spatial insights that are otherwise difficult to quantify, highlighting the importance of neighborhood and environmental factors in real estate valuation. Future work may explore higher-resolution imagery, additional geospatial features, or advanced fusion techniques to further enhance model performance and interpretability.