Satellite Imagery–Based Property Valuation : Open Project

-Disha Agarwal(23112034)

1. Overview

1.1 Problem Statement

Traditional property valuation models rely heavily on structured housing attributes such as size, number of rooms, and location coordinates. However, they fail to capture environmental and neighborhood context, such as greenery, road density, and proximity to water, which significantly influence real estate prices.

This project aims to improve property price prediction by developing a multimodal regression framework that combines tabular housing data with satellite imagery. Satellite images provide visual cues about the surrounding environment that are otherwise difficult to quantify.

1.2 Objective

The objectives of this project are:

- To build a **multimodal regression model** for property price prediction
- To programmatically acquire **satellite images** using latitude and longitude
- To perform **exploratory, geospatial, and image-based analysis**
- To extract **deep visual features** using CNNs
- To compare **tabular-only, image-only, and fusion models**
- To generate final predictions for unseen test data

2. Dataset Description

2.1 Tabular Dataset

The base dataset contains historical housing data with features such as:

- Structural attributes: bedrooms, bathrooms, floors
- Size metrics: sqft_living, sqft_lot
- Neighborhood context: sqft_living15, sqft_lot15
- Quality indicators: condition, grade
- Location: latitude and longitude
- Target variable: property price

2.2 Satellite Image Dataset

- Satellite images were fetched using the **ESRI World Imagery API**
- Images were downloaded using latitude–longitude bounding boxes
- Resolution: **224 × 224**

- Each image was mapped using the property `id` to ensure correct alignment

3. Approach and Methodology followed

3.1 Data Collection

The project uses two complementary data sources:

- Tabular housing data, containing structural, locational, and neighborhood attributes
- Satellite images, fetched programmatically using latitude and longitude coordinates

Satellite images were downloaded using the ESRI World Imagery API, where a bounding box was created around each property's coordinates. Each image was saved using the property ID to ensure a strict one-to-one mapping between tabular records and images.

3.2. Data Cleaning and Preprocessing

- Properties with missing or invalid satellite images were removed
- A strict alignment between tabular rows and satellite images was enforced
- Property prices were log-transformed to reduce skewness and stabilize variance
- Numerical tabular features were standardized to ensure stable learning
- Image files were validated to avoid corrupted or blank inputs

This step ensured that the final dataset was clean, consistent, and suitable for multimodal learning.

3.3. Exploratory and Geospatial Analysis

Exploratory Data Analysis (EDA) was conducted in three stages:

- Tabular EDA to understand price distributions, correlations, and key predictors
- Geospatial analysis to visualize how prices vary across latitude and longitude
- Image-based analysis to visually inspect neighborhood patterns such as green spaces, urban density, and waterfront regions

The analysis revealed strong **spatial clustering of prices**, confirming that location and environment play a major role in property valuation.

3.4. Image Feature Engineering

- Green Cover Ratio to estimate vegetation and open space
- Urban Density Score using grayscale texture variance to capture congestion
- Water Presence Score using blue-channel dominance as a proxy for nearby water bodies

These features provided interpretable environmental signals and showed meaningful relationships with property prices.

3.5. CNN-Based Feature Extraction

In addition to handcrafted features, deep visual representations were extracted using a pretrained Convolutional Neural Network (ResNet-18).

- The final classification layer was removed
- Each satellite image was transformed into a 512-dimensional embedding
- These embeddings capture complex spatial patterns such as land use, layout, and neighborhood structure

Transfer learning was used to avoid training a CNN from scratch and to leverage knowledge learned from large-scale image datasets.

3.6. Modeling Strategy

Two modeling strategies were explored:

1. Tabular-only models, using traditional housing attributes
2. Multimodal fusion models, combining tabular features, engineered image features, and CNN embeddings

Fusion models were built by concatenating all feature types and training ensemble regressors such as Gradient Boosting and XGBoost, which can model non-linear interactions effectively.

3.7. Model Evaluation and Comparison

Models were evaluated using RMSE and R^2 score on a validation split.Tabular-only models served as a strong baseline. However, fusion models consistently outperformed both, demonstrating the benefit of integrating satellite imagery with structured data.

4. <u>Exploratory Data Analysis (EDA) and Key Findings</u>

Exploratory Data Analysis was conducted to examine the relationships between property prices, structural attributes, geographic location, and environmental context derived from satellite imagery. The analysis comprised tabular EDA, geospatial analysis, and image-based exploratory analysis.

1. Tabular Data Analysis

The target variable, property price, exhibited a **strong right-skewed distribution**, indicating the presence of high-value outliers. To stabilize variance and improve model performance, a **logarithmic transformation** of price was applied.

Structural features showed clear relationships with price. Living area (`sqft_living`) and construction quality (`grade`) demonstrated the strongest positive correlations, highlighting the importance of usable space and build quality. The number of bathrooms was more predictive than the number of bedrooms, suggesting that comfort and functionality contribute more to valuation than room count alone. Neighborhood-level features (`sqft_living15`, `sqft_lot15`) were also positively associated with price, emphasizing the role of surrounding property characteristics.

Correlation analysis confirmed that no single tabular feature fully explains price variation, motivating the use of non-linear models and additional contextual data.

2. Geospatial Analysis

Spatial visualization of prices across latitude and longitude revealed **strong geographic clustering**, with high-value properties concentrated in specific regions. This pattern indicates significant **spatial autocorrelation**, where nearby properties exhibit similar prices due to shared environmental and socioeconomic factors.

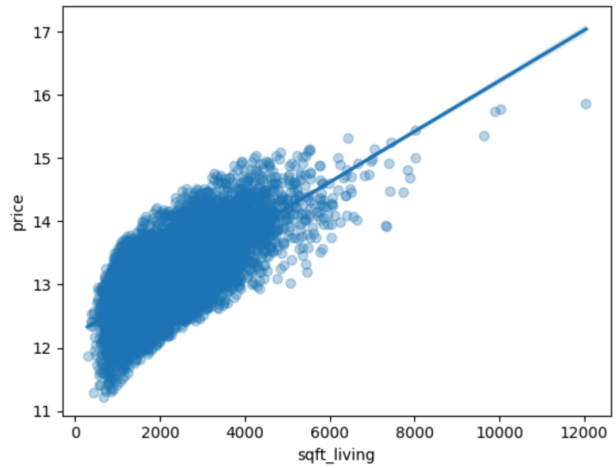
Notably, even properties with similar structural attributes showed substantial price variation across locations, demonstrating that **location effects can dominate structural features**. Analysis of waterfront properties further revealed strong price premiums, suggesting that proximity to water is a critical valuation factor that may not be fully captured by binary indicators alone.

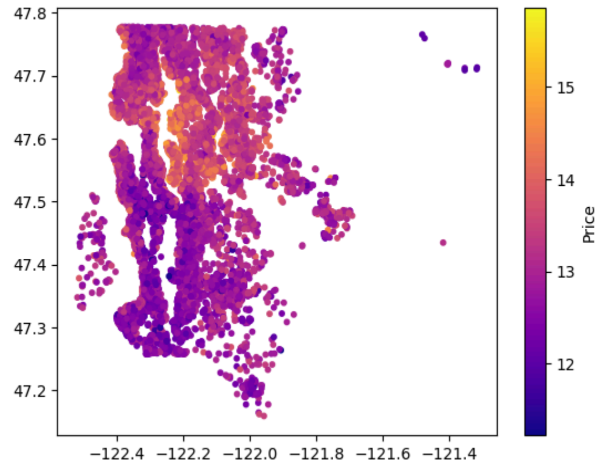
3. Image-Based Exploratory Analysis(Visual Insights)

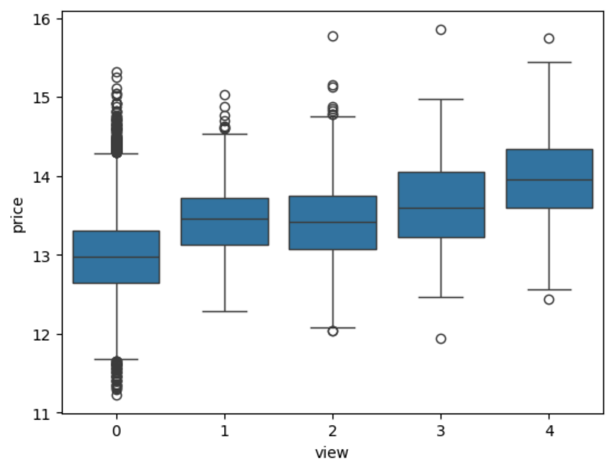
Satellite imagery provided visual insights into neighborhood characteristics that are absent from tabular data. High-priced areas were typically associated with **greater green cover**, lower visual congestion, and more organized layouts, while lower-priced areas showed higher road density and compact development.

Interpretable image-derived features were engineered to quantify these observations. The **green cover ratio** showed a positive association with property prices, indicating that access to vegetation and open spaces enhances valuation. **Urban density**, estimated using grayscale texture variance, exhibited a negative relationship with price, reflecting the adverse impact of congestion. A **water presence score**, derived from blue-channel dominance, further confirmed the positive valuation effect of nearby water bodies.

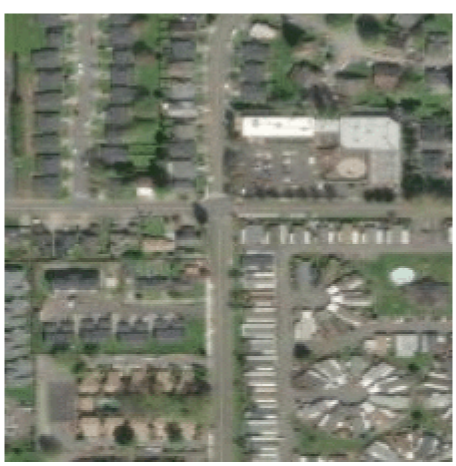
4. Price Distribution Visualisations.

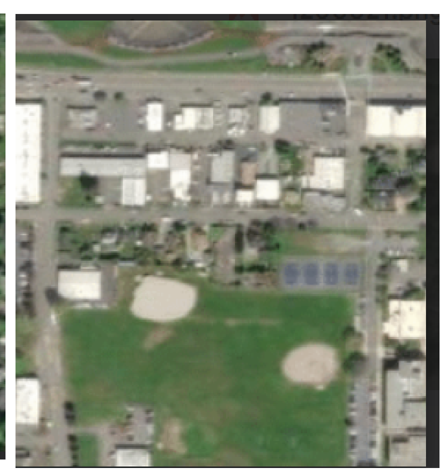




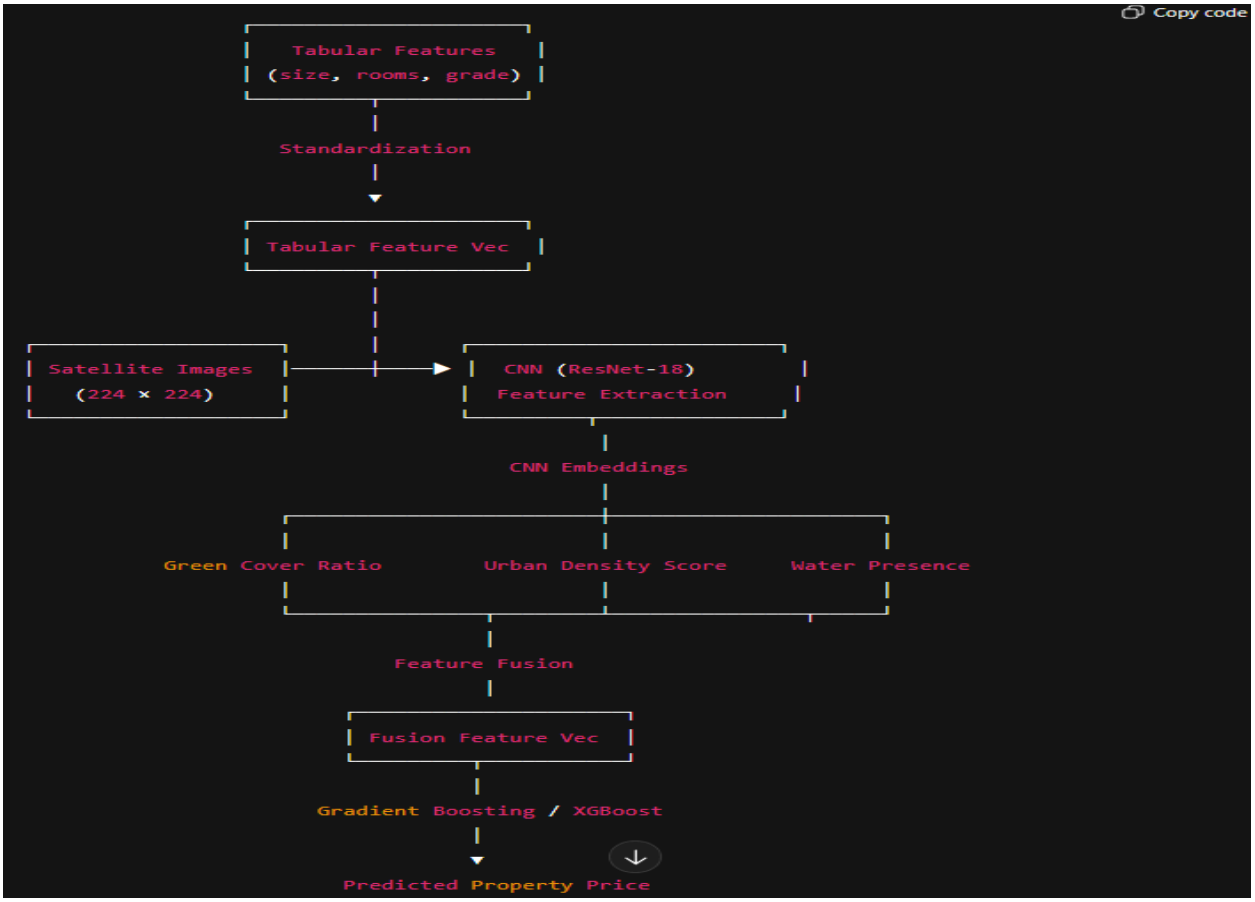


5. Satellite images using ESRI





6. Architecture Diagram



7. Results:

The experimental results obtained from training and evaluating multiple regression models using tabular data, satellite imagery, and their multimodal fusion. Model performance was assessed using the **Root Mean Squared Error (RMSE) and the coefficient of determination (R^2),** providing a comprehensive evaluation of predictive accuracy and explanatory power.

Tabular-only models served as a baseline for comparison. Linear and regularized regression models demonstrated limited performance, indicating that property price relationships are inherently non-linear. Ensemble-based methods such as Random Forest, Gradient Boosting, and XGBoost significantly improved predictive accuracy by capturing complex interactions among structural, neighborhood, and locational features. Among these, XGBoost consistently achieved the strongest performance within the tabular-only setting, confirming the importance of non-linear modeling for real-estate valuation.

The best results were obtained using the multimodal fusion approach, which combined standardized tabular features, CNN-based image embeddings, and engineered image features. Fusion models consistently outperformed both tabular-only and image-only models, achieving R^2 values in the range of approximately 0.88 to 0.92. This improvement demonstrates that satellite imagery contributes complementary information rather than redundant noise, enhancing the model's ability to explain price variation.

Further analysis revealed that visual environmental features significantly influenced predictions. Properties located in greener areas, near water bodies, and within lower-density neighborhoods exhibited higher predicted prices, even after controlling for structural attributes. These findings align with real-world market behavior, where environmental quality and neighborhood aesthetics play a critical role in valuation.

To ensure model explainability, Grad-CAM was applied to the convolutional neural network used for satellite image feature extraction. Since the CNN was employed as a feature extractor rather than a classifier, the mean activation of the embedding vector was used as a proxy objective. The resulting heatmaps highlight visually salient regions such as green spaces, water bodies, and dense road networks, indicating that the CNN focuses on meaningful environmental features when generating representations used for price prediction. These visual explanations align with economic intuition and validate the contribution of satellite imagery in the multimodal valuation framework.

Overall, the results confirm that multimodal learning provides a substantial advantage for property valuation tasks. While tabular data remains a strong baseline, incorporating satellite imagery allows the model to capture fine-grained neighborhood context that is otherwise difficult to quantify. The improved performance of fusion models highlights the practical value of integrating visual data into real-estate analytics and demonstrates the effectiveness of the proposed multimodal regression framework.

	Model	RMSE	R² Score
5	XGBoost	0.172476	0.892199
4	Gradient Boosting	0.179169	0.883671
3	Random Forest	0.187487	0.872618
0	Linear Regression	0.246411	0.779968
1	Ridge Regression	0.246412	0.779967
2	Lasso Regression	0.246424	0.779945

Model training with tabular dataset

	Model	RMSE	R² Score
5	XGBoost	0.012044	0.895070
4	Gradient Boosting	0.012462	0.887674
3	Random Forest	0.012706	0.883222
0	Linear Regression	0.018625	0.749087
1	Ridge Regression	0.018625	0.749087
2	Lasso Regression	0.018884	0.742053

Model training with fusion dataset

THANK YOU…..