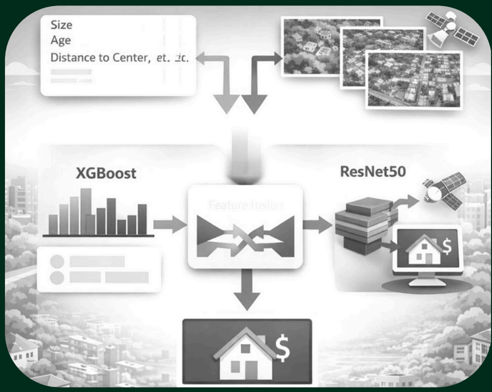
SATELLITE-IMAGERY-BASED PROPERTY VALUATION



PREPARED BY	Enrollment No.
Mangesh Rakhewar	24117072

Introduction

Problem statement

Conventional house price prediction models rely primarily on structured tabular data and often fail to capture neighborhood-level visual and environmental factors that influence property value. Satellite imagery provides rich contextual information about surroundings, but integrating unstructured visual data with numerical features remains a technical challenge. This project formulates property valuation as a multimodal regression problem, combining tabular housing data and satellite imagery to evaluate whether visual context can enhance predictive performance beyond traditional models.

Project Overview

Accurate house price prediction depends not only on internal property attributes but also on external neighborhood and environmental factors that are difficult to quantify using structured data alone. This project proposes an end-to-end multimodal house price prediction system that integrates tabular housing data with satellite imagery to capture both property-level and neighborhood-level information.

A strong tabular-only baseline is first established using XGBoost to model structured features such as property size, age, location, and engineered spatial distances. To evaluate the additional contribution of visual context, deep visual features extracted from satellite images are fused with tabular features in a multimodal learning pipeline. Performance of the multimodal model is then compared against the tabular baseline to quantify the satellite imagery importance.

The project also demonstrates a realistic deployment scenario by performing live satellite image inference at prediction time, ensuring alignment between training and real-world usage.

Objectives

1. **To develop a robust tabular baseline model** for house price prediction using engineered spatial, temporal, and categorical features, establishing a strong reference for multimodal comparison.
2. **To programmatically acquire satellite imagery** using geographic coordinates through a web-based API in order to capture neighborhood-level environmental context unavailable in traditional tabular datasets.
3. **To extract high-level visual representations** from satellite images using a pretrained convolutional neural network, enabling structured learning from unstructured visual data.
4. **To design and implement a multimodal regression pipeline** that integrates tabular features and satellite image embeddings through feature-level fusion for property value prediction.
5. **To ensure computational efficiency and scalability** by applying dimensionality reduction techniques and selective image sampling while preserving meaningful visual information.
6. **To analyze the limitations of multimodal learning in real-estate valuation**, identifying factors such as image resolution, feature dominance, and noise that affect performance gains.

Data description

Tabular Housing Dataset

- Data source-tabular_Dataset.csv
- Key attributes-['bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot', 'floors',

  'waterfront', 'view', 'condition', 'grade', 'sqft_above', 'zipcode'
  'sqft_basement', 'lat', 'long', 'sqft_living15', 'sqft_lot15',]-original
  ['log_dist_city', 'log_dist_water', 'log_dist_tech',
  'location_cluster', 'is_renovated']-Derived
- Target variable- price

Satellite Imagery Dataset

- Image Source- **Mapbox Static Images API**, which provides high-resolution satellite imagery based on geographic coordinates (latitude and longitude), enabling consistent and automated acquisition of neighborhood-level visual data.
- Resolution- 512 × 512 pixels ,Zoom Level-18

This provides a balanced representation of neighborhood-scale spatial context. This configuration captures surrounding infrastructure, road networks, vegetation, and building density within the vicinity of each property, while avoiding excessive focus on individual structures.

- Coverage area- approx. **0.06 km²** per image at this resolution and zoom level

Data Challenges

- Missing values-there are no missing values in tabular dataset
- skewness- high skewness in the target variable (house price)

House prices typically exhibit a **right-skewed distribution**, where a small number of high-value properties significantly exceed the majority of observations. This skewness introduces challenges for regression models, as extreme values can dominate the loss function and bias the model toward expensive properties. models trained on highly skewed targets often produce unstable predictions, poor generalization for mid-range properties, and inflated error metrics.

To mitigate these effects, a logarithmic transformation of the target variable is applied, which compresses extreme values, stabilizes variance, and enables the model to learn more balanced relationships across the price spectrum.

- High Cardinality of Zipcode Feature-Zipcode is a high-cardinality categorical feature that represents location-specific socioeconomic and infrastructural characteristics influencing house prices. Using raw zipcodes as categorical labels is not directly compatible with most regression models and can lead to dimensional explosion when one-hot encoding is applied.

To address this, zipcode encoding is required to transform location identifiers into meaningful numerical representations that preserve their relationship with the target variable. Target encoding is used in this project to capture the average price behavior associated with each zipcode while maintaining model efficiency.

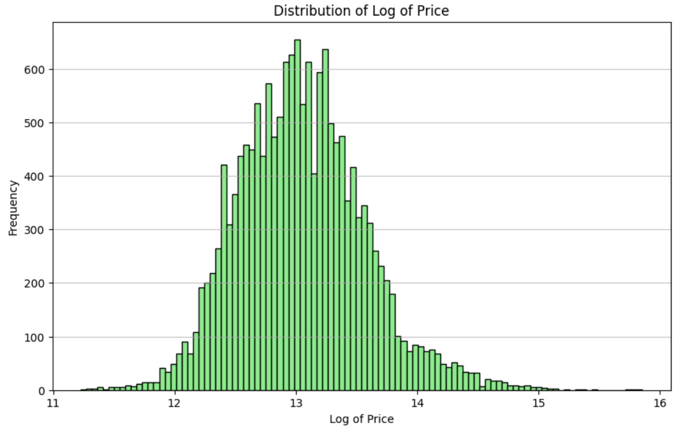
- Outliers and Extreme Property Values- The dataset includes properties with unusually high prices, sizes, or land areas. These outliers can disproportionately influence model training and distort loss minimization, particularly in regression tasks.
- Data Leakage Risks-Certain attributes (e.g., yr_renovation or sale_yr) may indirectly encode future information if not handled carefully, zipcode also majorly contribute to price data posing a risk of data leakage and artificially inflated performance.

Exploratory Data Analysis(EDA)

Price Distribution Analysis

skewness in price distribution

House prices typically exhibit a **highly right-skewed distribution**, where a small number of high-value properties are significantly more expensive than the majority of observations. This skewness poses challenges for regression models, as extreme values can disproportionately influence the learning process and lead to unstable predictions.



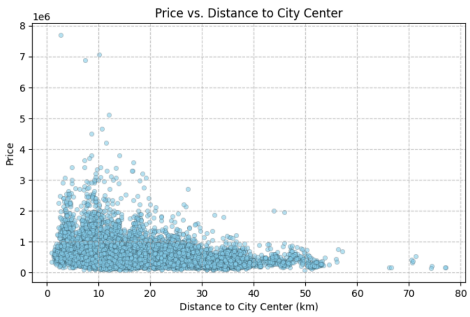
Log transformation

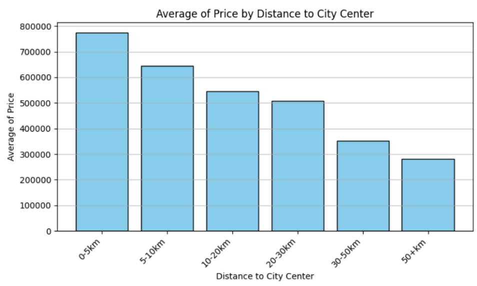
Applying a logarithmic transformation to the price variable helps **compress the scale of extreme values**, thereby reducing skewness and stabilizing variance across the dataset. This transformation results in a more symmetric distribution, allowing the model to learn relationships that are more consistent across different price ranges.

From a modeling perspective, log transformation improves numerical stability, reduces the impact of outliers on loss minimization, and enables the model to focus on **relative price differences** rather than absolute magnitudes.

Distance-Based Price Variation

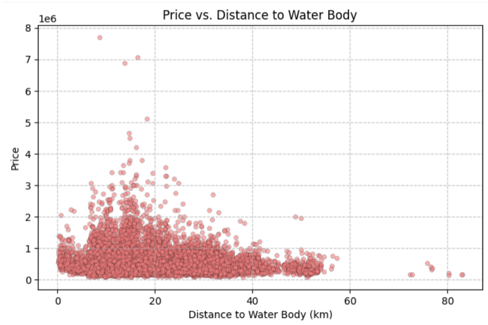
To understand the influence of spatial proximity on property valuation, house prices are analyzed against distance from three key geographic reference points: the city center, nearby water bodies, and major technology hubs. The scatter plots reveal a consistent and interpretable spatial trend across all three dimensions.

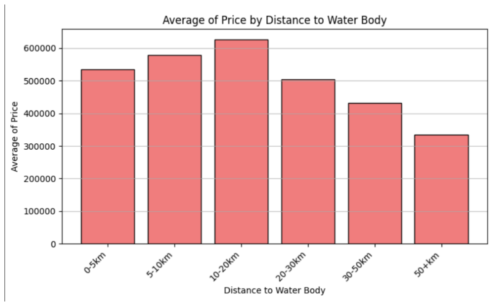




The ***Price vs. Distance to City Center*** plot shows a strong inverse relationship between property price and distance from the urban core. Properties located within close proximity to the city center exhibit the highest price values, including the majority of premium and outlier-priced houses. As distance increases, both the average price and the upper price range decline steadily.

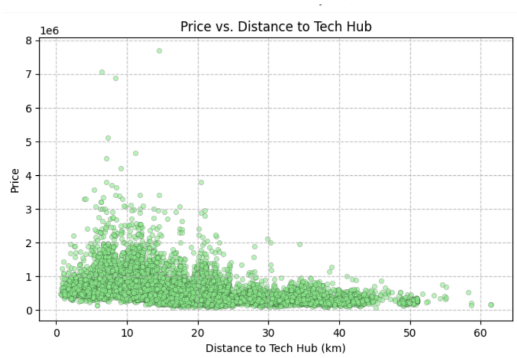
This pattern reflects the economic importance of central urban areas, where accessibility to employment, transportation, amenities, and services is highest. Beyond a certain distance threshold, high-priced properties become increasingly rare, indicating diminishing location premium.

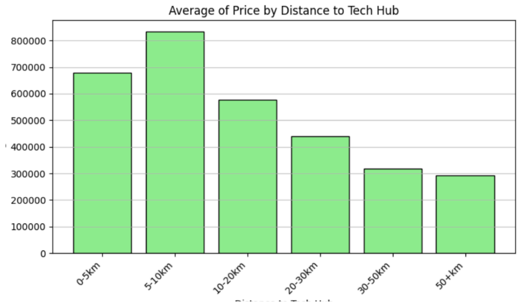




The ***Price vs. Distance to Water Body*** plot highlights a pronounced price premium for properties located near water features. Houses situated closer to water bodies show a higher concentration of expensive properties, while prices generally decline as distance increases.

This trend indicates that proximity to water contributes positively to property valuation due to aesthetic appeal, recreational opportunities, and perceived environmental quality.





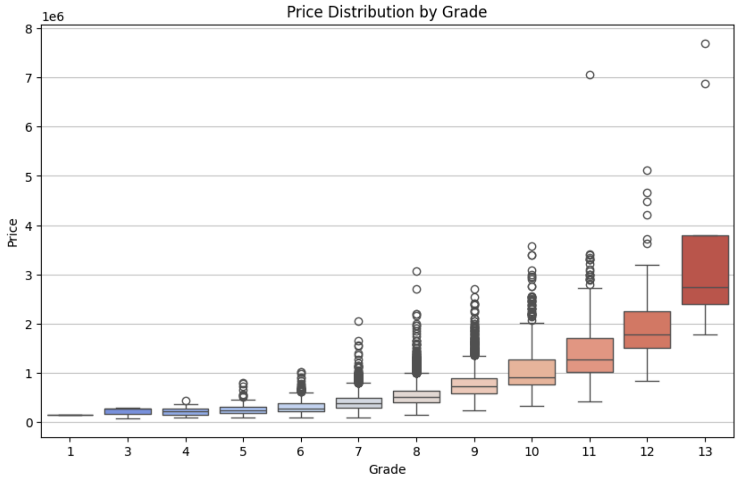
The ***Price vs. Distance to Tech Hub*** plot demonstrates a clear negative relationship between house price and distance from major employment centers. Properties closer to technology hubs tend to have higher prices, reflecting strong demand driven by reduced commuting time and access to high-income employment opportunities.

As distance from tech hubs increases, both the median price and the frequency of high-priced properties decline noticeably. This trend suggests that employment proximity is a significant economic driver of housing demand and reinforces the importance of incorporating distance-to-tech-hub features into valuation models.

Impact of Property Grade on House Price

The box plot illustrates a strong positive relationship between property grade and house price. As the grade increases, both the median price and the overall price distribution shift upward, indicating that higher-grade properties consistently command higher market values.

Lower grades exhibit tightly clustered price distributions with limited upper ranges, reflecting standardized construction quality and lower valuation. In contrast, higher-grade properties show greater price dispersion and significantly higher upper outliers, suggesting that premium construction quality, design, and materials substantially influence property value.

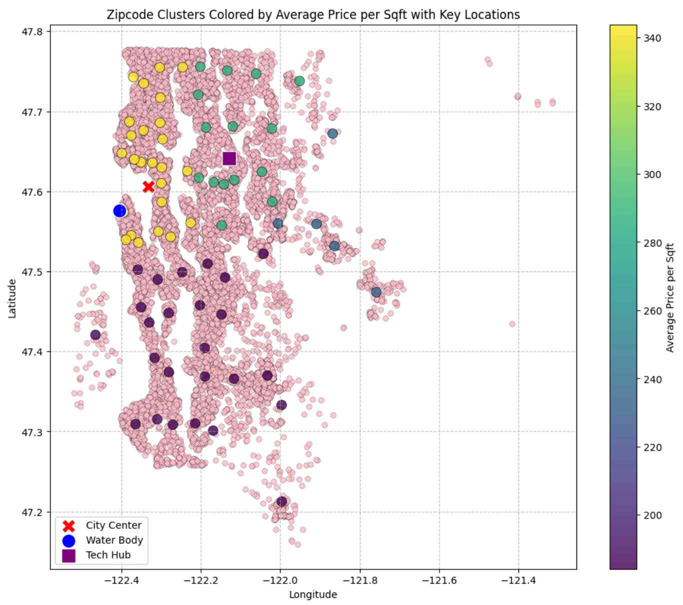


Overall, this analysis confirms property grade as a key determinant of house price and justifies its inclusion as an important predictive feature in the model.

Zipcode-Level Spatial Price Insights

The zipcode-level spatial visualization reveals significant regional variation in average house prices, indicating strong location dependency in property valuation.

Zipcodes located closer to the city center exhibit consistently higher prices, reflecting the premium associated with urban accessibility and infrastructure. Similarly, regions near water bodies show elevated price levels, highlighting the environmental and lifestyle benefits of waterfront proximity. Proximity to major technology hubs also corresponds to higher prices, driven by employment accessibility and reduced commuting costs. In contrast, zipcodes farther from these key locations display lower and more uniform pricing, indicating diminished location-based premiums.



These observations confirm that zipcode encodes critical geographic and socioeconomic information, justifying its inclusion as an encoded feature in the predictive model.

Impact of Property Condition on House Price

Property condition is a critical determinant of house prices and is primarily influenced by factors such as **property age** and **renovation status**. To analyze this effect, the relationship between log-transformed house prices and property age is examined alongside a comparative analysis of renovated and non-renovated properties.

Effect of Property Age

The ***log(Price) vs. Property Age*** scatter plot reveals a weak but consistent negative relationship between property age and house price. As properties age, their average log price shows a gradual downward trend, indicating depreciation over time due to structural wear, outdated design, and increased maintenance requirements.

However, the wide dispersion of prices across all age groups suggests that age alone does not fully determine property value. Older properties can still command high prices depending on factors such as location, neighborhood quality, and structural upgrades.

Slope of regression line: `-0.0014`

Effect of Renovation Status

The box plot comparing log price distributions of **renovated and non-renovated properties** shows a clear upward shift in prices for renovated houses. Renovated properties exhibit a higher median price and a higher upper price range, indicating a positive renovation premium.

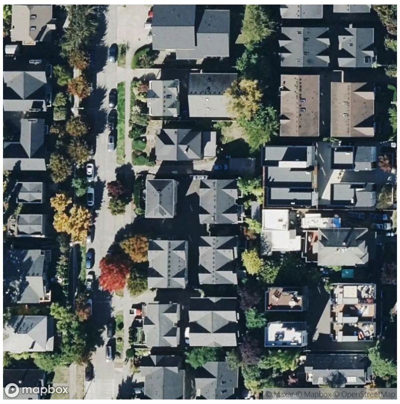
This shift suggests that renovation mitigates the depreciation effect of aging by improving functional quality, aesthetics, and market appeal.

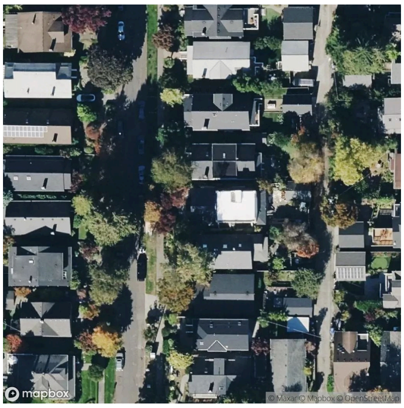
Visual Data Exploration

Satellite imagery is explored **to qualitatively assess neighborhood-level characteristics** that influence property valuation but are not explicitly represented in tabular data. Representative images corresponding to different neighborhood typologies are analyzed to identify recurring spatial and environmental patterns.

Dense Urban Area

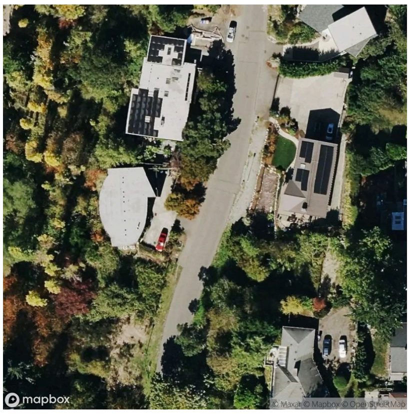
The satellite image representing dense urban regions shows compact building arrangements with minimal spacing between structures, high road density, and limited visible green spaces. Such regions are typically associated with high population density, strong infrastructure connectivity, and close proximity to commercial and employment centers. These characteristics often correspond to higher land values but may also introduce congestion and limited open space. The visual compactness and infrastructure concentration observed in these images provide contextual cues related to urban accessibility and economic activity.

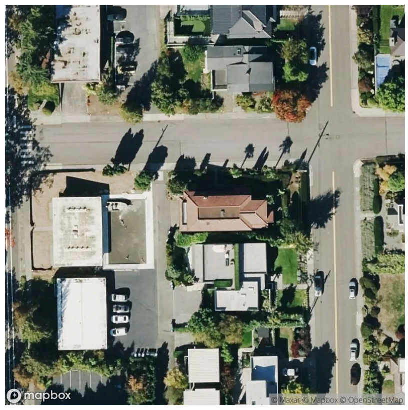




Suburban Residential Area

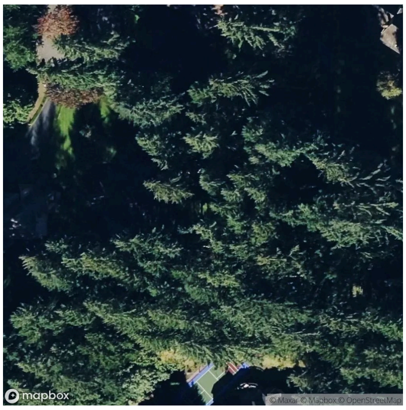
Suburban residential images exhibit moderately spaced housing units, organized street layouts, and a balance between built structures and open areas. These neighborhoods often represent family-oriented residential zones with relatively lower congestion compared to urban cores.The visual patterns suggest improved living space and reduced density, which can influence price differently depending on proximity to urban centers and employment hubs.





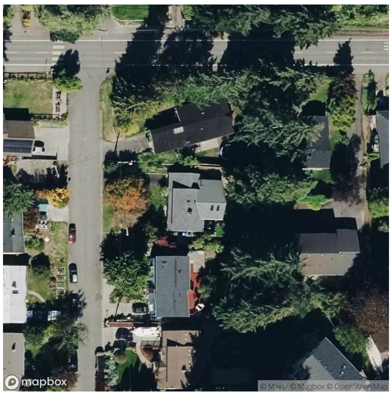
Green-Rich Neighborhood

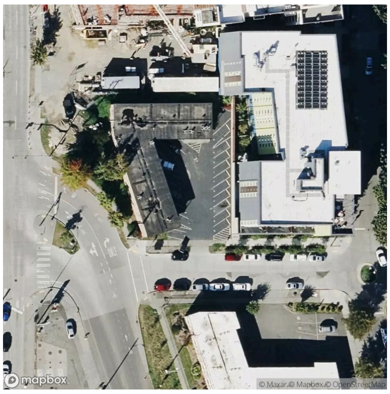
Green-rich neighborhoods are characterized by abundant tree cover, open spaces, and visible vegetation surrounding residential properties. These visual cues are indicative of environmental quality, recreational access, and improved livability. Such areas often command a price premium due to aesthetic appeal and perceived health benefits. The presence of extensive greenery is difficult to quantify using tabular data alone, making satellite imagery particularly valuable for capturing this information.

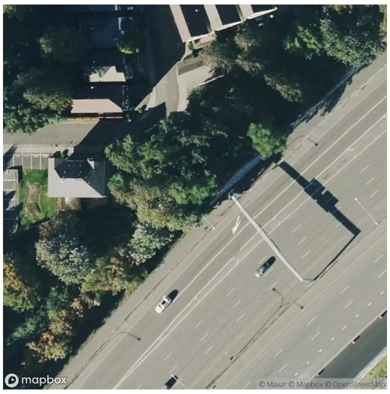


Road-Dense / Infrastructure-Heavy Area

Images dominated by extensive road networks, intersections, and transportation infrastructure reflect regions with high connectivity and accessibility. While improved road infrastructure facilitates commuting and economic activity, excessive road density may also introduce noise and congestion. The visual representation of infrastructure intensity provides important contextual information regarding accessibility and urban development patterns.







etc.

Observed Neighborhood Patterns

Building Density

Satellite images show clear variation in housing density across regions. Dense urban areas exhibit compact building layouts with minimal spacing, while suburban neighborhoods display more spaced residential structures, indicating lower congestion and larger plot sizes.

Road Network Structure

Urban neighborhoods are characterized by dense, well-connected road networks, often arranged in grid-like patterns. In contrast, suburban and peripheral areas show sparser road structures with fewer intersections, reflecting lower traffic intensity and reduced connectivity.

Green Cover

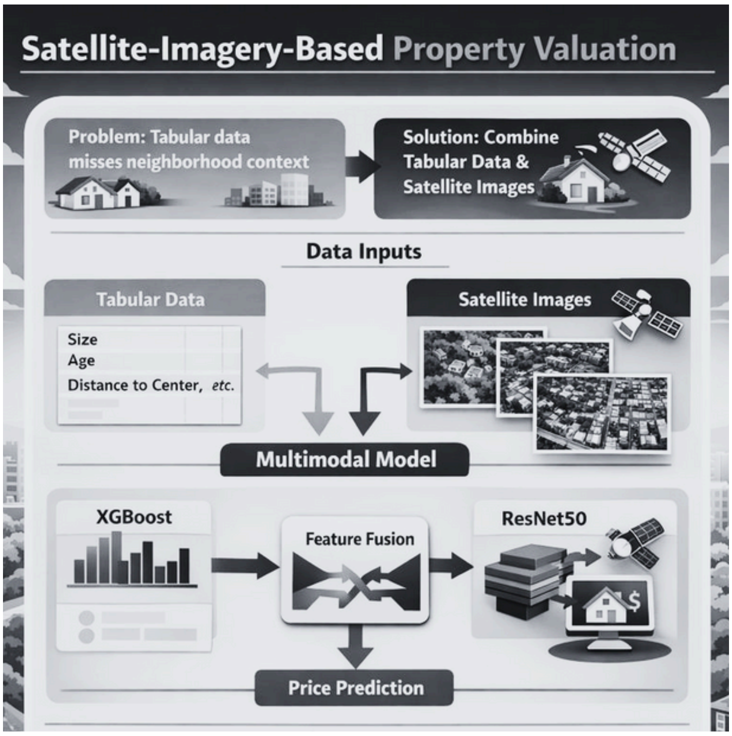
Significant differences in green cover are observed across neighborhoods. Tree-rich areas with visible vegetation indicate higher environmental quality, whereas concrete-dominant regions reflect highly developed urban zones with limited open spaces.

Environmental Features

Certain neighborhoods exhibit proximity to water bodies and open spaces, contributing to scenic value and improved livability. Areas lacking such features appear more infrastructure-heavy and uniform in visual structure.

Methodology

The methodology outlines the end-to-end pipeline developed for multimodal house price prediction, detailing the preprocessing of tabular data, acquisition and processing of satellite imagery, feature fusion strategy, and model architecture. The overall approach integrates structured and unstructured data to capture both property-level and neighborhood-level information.



Tabular Data Preprocessing

Tabular data preprocessing is a critical step in ensuring that raw housing data is transformed into a structured, informative, and model-ready format. This process involves data cleaning, feature engineering, spatial transformation, and encoding techniques designed to capture economic, temporal, and geographic factors influencing house prices.

- Date and Target Variable Processing

The transaction date is converted into a datetime format to extract temporal information relevant to market trends. From this, the **sale year** is derived to capture time-dependent price variations.

House prices typically exhibit strong right skewness; therefore, a **logarithmic transformation of the price** is applied. This transformation stabilizes variance, reduces the influence of extreme values, and improves regression performance by enabling the model to learn relative price differences more effectively.

- Property Condition Feature Engineering

To capture the physical condition of a property, **property age** is computed as the difference between the sale year and the year of construction. Property age reflects depreciation effects due to structural wear and outdated design.

Additionally, a binary **renovation indicator** is created to distinguish renovated and non-renovated properties. Renovation status helps model the positive price impact of structural upgrades and mitigates the depreciation effect associated with aging properties.

- Geospatial Distance Feature Engineering

Raw latitude and longitude values do not explicitly encode spatial relationships. To address this limitation, **Haversine distance calculations** are used to compute distances from each property to key geographic anchors:

- City center
- Nearest water body
- Major technology hub

These distances are then log-transformed to reduce skewness and capture non-linear decay effects, where proximity has a stronger influence at shorter distances.

- Location Clustering

To further capture neighborhood-level spatial structure, geographic coordinates are clustered using **KMeans clustering**. Each property is assigned to a location cluster representing spatially similar regions. This approach allows the model to learn neighborhood-specific patterns without relying solely on raw coordinates.

- Categorical Encoding

Zipcode is a high-cardinality categorical variable that encodes socioeconomic and geographic information. Direct usage of raw zipcodes is unsuitable for regression models. Therefore, **target encoding** is applied, transforming zipcodes into numerical values based on their relationship with the target variable. This approach preserves location-based price signals while avoiding dimensional explosion associated with one-hot encoding.

- Interaction Feature Construction

An interaction feature combining **waterfront presence and living area** is introduced to model the amplified effect of larger properties located near water bodies. This feature captures non-linear interactions where waterfront benefits scale with property size.

- Feature Selection

A comprehensive set of features is selected, including property attributes, spatial features, temporal variables, and engineered indicators. This ensures that both intrinsic property characteristics and extrinsic location-based factors are represented in the model.

Satellite Image Acquisition

Satellite image acquisition is performed to obtain neighborhood-level visual context for each property, which cannot be captured through tabular data alone.

- API-Based Image Fetching

Satellite images are downloaded using the **Mapbox Static Images API**, which provides satellite imagery based on latitude and longitude coordinates. For each property, a request is constructed using its geographic coordinates to retrieve a satellite image centered on the property location. This automated approach enables scalable and reproducible image acquisition for the entire dataset.

Each image is fetched using a fixed zoom level and resolution to maintain uniform spatial context across all properties. The images are stored locally using a structured naming convention that includes the property identifier and geographic coordinates.

- Design Choices: Zoom Level and Resolution

All images are retrieved at **zoom level 18** with a **512 × 512 pixel resolution**. This configuration is selected to capture neighborhood-scale spatial information, including surrounding buildings, road networks, green cover, and nearby infrastructure, rather than focusing on individual property details.

- Error Handling and Robustness

The download process includes exception handling to manage network issues, timeouts, or API failures. Failed downloads are tracked separately, ensuring transparency in data availability and preventing silent data loss. Successfully downloaded images are logged, enabling verification of dataset completeness.

- Image–Property Mapping

A summary file is generated to map each property identifier to its corresponding satellite image path. This mapping ensures reliable alignment between tabular records and visual data during subsequent feature extraction and multimodal modeling stages.

Image Feature Extraction

- CNN Architecture

A pretrained **ResNet50** convolutional neural network is employed to extract visual features from satellite images. The network's final classification layer is removed, enabling it to function as a feature extractor rather than a classifier.

- Pretrained Model Usage

The ResNet50 model is initialized with ImageNet-pretrained weights and used in inference mode. Standard image preprocessing steps, including resizing and normalization, are applied to ensure compatibility with the pretrained network.

- Feature Representation

Each satellite image is transformed into a **512-dimensional feature vector** that encodes high-level neighborhood characteristics such as infrastructure layout, road patterns, and green cover. These embeddings serve as numerical representations of visual context.

- Sampling Strategy

To reduce computational overhead, image features are extracted from a **random 20% subset** of the available images. This approach enables efficient training while preserving sufficient visual diversity for multimodal learning.

Multimodal Feature Fusion

- Feature Alignment

Tabular records and satellite image features are aligned using a common property identifier to ensure correct correspondence between structured attributes and visual embeddings.

- Tabular Feature Preparation

A selected set of engineered tabular features capturing property characteristics, spatial distances, location clusters, and renovation status is used. The high-cardinality zipcode feature is transformed using **target encoding**, preserving location-based price signals while avoiding dimensional explosion.

All tabular features are scaled using **RobustScaler** to reduce sensitivity to outliers.

- Image Feature Processing

Satellite image embeddings extracted from ResNet50 are independently scaled using RobustScaler. Due to the high dimensionality and redundancy of CNN embeddings, **Principal Component Analysis (PCA)** is applied to reduce image features from **512 to 64 dimensions**, retaining dominant visual patterns while suppressing noise.

- Fusion Strategy

Feature-level fusion is employed by **concatenating scaled tabular features with PCA-reduced image embeddings**. This unified representation enables the regression model to jointly learn interactions between structured property attributes and neighborhood-level visual context.

Baseline Model Training (Tabular Data Only)

A baseline regression model is trained using only tabular housing features to establish a strong reference for evaluating the contribution of satellite imagery. The target variable is log-transformed house price, which improves numerical stability and reduces the influence of extreme values during training. The dataset is split into training and validation sets using an 80–20 split to enable unbiased evaluation.

- Learning Algorithm

An **XGBoost Regressor** is employed as the baseline model due to its strong performance on structured data, ability to model non-linear relationships, and robustness to feature interactions. Feature scaling is applied prior to training to standardize input distributions and improve optimization stability.

- Evaluation Metrics

1. **R^2 score**, to measure explained variance
2. **Root Mean Squared Error (RMSE)**, computed in the original price scale for interpretability

Multimodal regression pipeline

- Feature-Level Multimodal Integration

After independent preprocessing of tabular and image features, multimodal learning is achieved through feature-level fusion. Scaled tabular features are concatenated with PCA-reduced image embeddings to form a unified multimodal feature vector. This combined representation allows the model to jointly learn relationships between structured property attributes and neighborhood-level visual context.

- Hyperparameter Optimization

To obtain an optimal regression model, **Optuna** is used for automated hyperparameter tuning of the XGBoost regressor. The optimization objective minimizes the **root mean squared error (RMSE)** on the validation set. Key hyperparameters tuned include the number of trees, learning rate, tree depth, subsampling ratios, and regularization terms.

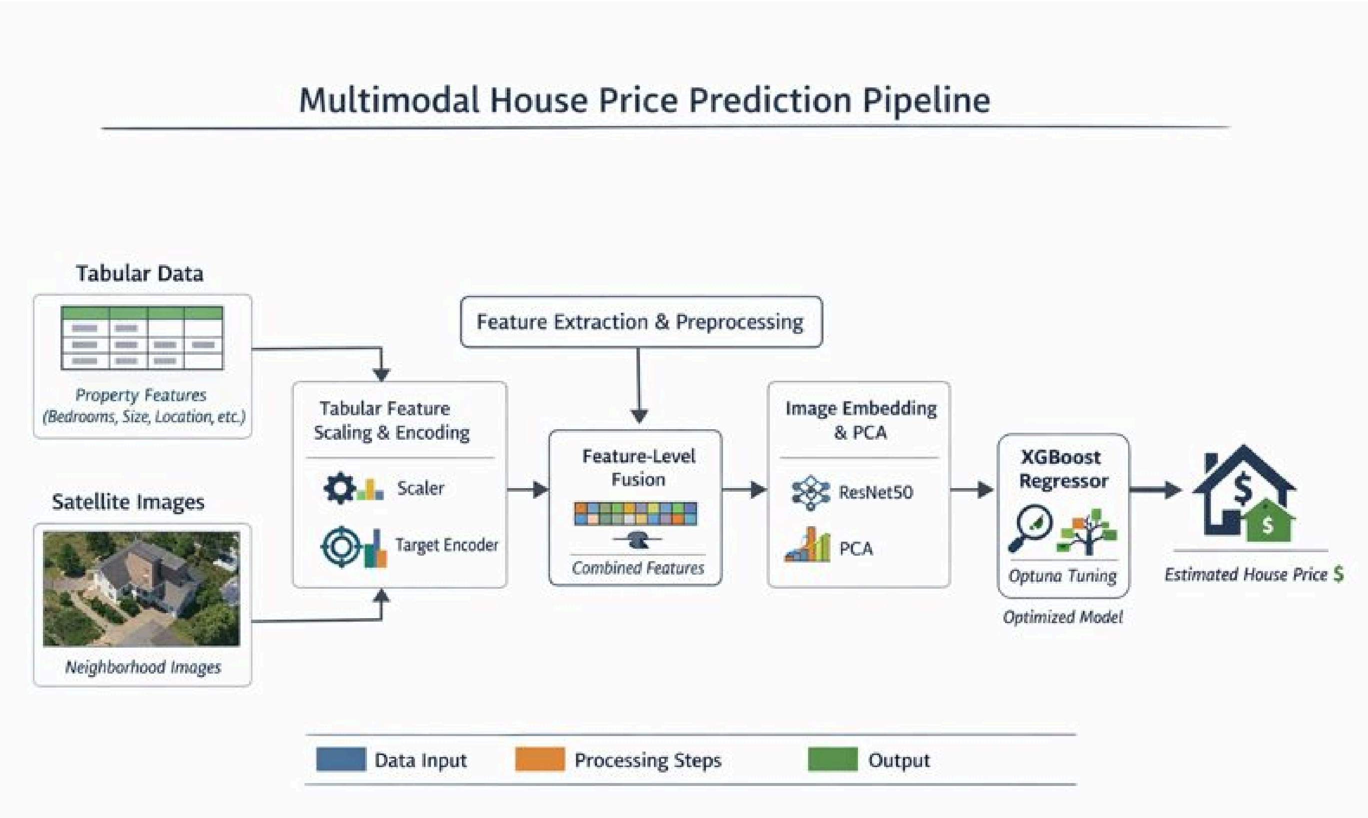
- Model Training

Using the best hyperparameters identified by Optuna, a final **XGBoost regression model** is trained on the multimodal feature set. The model operates on log-transformed house prices to stabilize learning and is capable of capturing non-linear interactions between tabular and image-derived features.

- Pipeline Construction

All preprocessing components—including tabular scalers, image scalers, PCA transformation, zipcode encoder, and the trained regression model—are packaged into a single pipeline object. This pipeline is serialized and saved, enabling consistent preprocessing and prediction during inference and deployment.

Architecture diagram



Architecture of the multimodal regression pipeline integrating tabular features and satellite image embeddings for house price prediction.

Experimental Setup

This section describes the experimental configuration adopted to ensure fair evaluation, reproducibility, and reliable comparison between tabular and multimodal models.

Training–Validation Split

- Dataset Partitioning

The dataset is partitioned into **training and validation sets** using an **80–20 split**. The training set is used for model learning and hyperparameter optimization, while the validation set is reserved for unbiased performance evaluation.

A fixed random seed is applied during splitting to ensure experiment reproducibility and consistent comparisons across different modeling approaches.

- Sampling Strategy

For the multimodal setup, only samples with both **tabular records and corresponding satellite image features** are retained to maintain data alignment. Additionally, image feature extraction is performed on a randomly selected subset of satellite images to balance computational feasibility and representative visual diversity.

Evaluation Metrics

- Root Mean Squared Error (RMSE)

RMSE measures the average magnitude of prediction errors in the **original price scale**, making it directly interpretable in monetary terms. Lower RMSE values indicate better predictive accuracy and reduced deviation between predicted and actual house prices.

- R^2 Score

The R^2 score quantifies the proportion of variance in house prices explained by the model. Higher R^2 values indicate stronger explanatory power and better overall model fit.

Results and Discussion

This section presents the experimental results of both the tabular-only and multimodal models and discusses their performance characteristics, comparative behavior.

Baseline Model Performance

The tabular-only model achieves the following results:

- **R^2 score:** 0.898
- **RMSE:** $113,044

The strong performance of the tabular baseline highlights that **property attributes and engineered spatial features capture the majority of price-determining factors**. Features related to location, property size, grade, and proximity to key geographic anchors play a dominant role in valuation.

This high baseline performance establishes a **challenging benchmark** for multimodal learning. Any additional improvement obtained by incorporating satellite imagery must therefore provide complementary information beyond already strong tabular signals.

Multimodal Model Performance

The multimodal model, which combines tabular features with satellite image embeddings, achieves the following results:

- **R^2 score:** 0.883
- **RMSE:** $182,014

The multimodal model incorporates neighborhood-level visual context alongside structured property attributes. While satellite imagery contributes additional contextual information related to surrounding infrastructure, green cover, and spatial layout, its overall impact on predictive performance is limited. The results indicate that visual features provide complementary signals but do not significantly enhance accuracy beyond the strong tabular baseline.

The observed performance suggests that the predictive capacity of structured features dominates in this dataset, leaving limited scope for improvement through visual data. Nonetheless, the multimodal framework demonstrates the feasibility of integrating heterogeneous data sources and provides qualitative interpretability through visual feature analysis.

Explainability and Insights

This section focuses on interpreting model behavior and understanding the contribution of both tabular and visual features to price prediction.

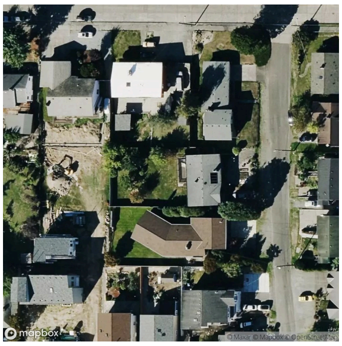
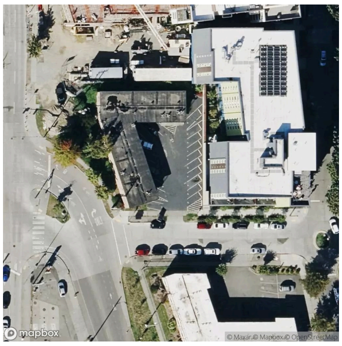
Feature Importance (Tabular Features)

Analysis of feature importance from the tabular model highlights property size, grade, location-based features, and proximity to key geographic anchors as dominant predictors. These results align with real estate valuation principles and confirm that engineered spatial and condition-related features play a central role in the model's decisions.

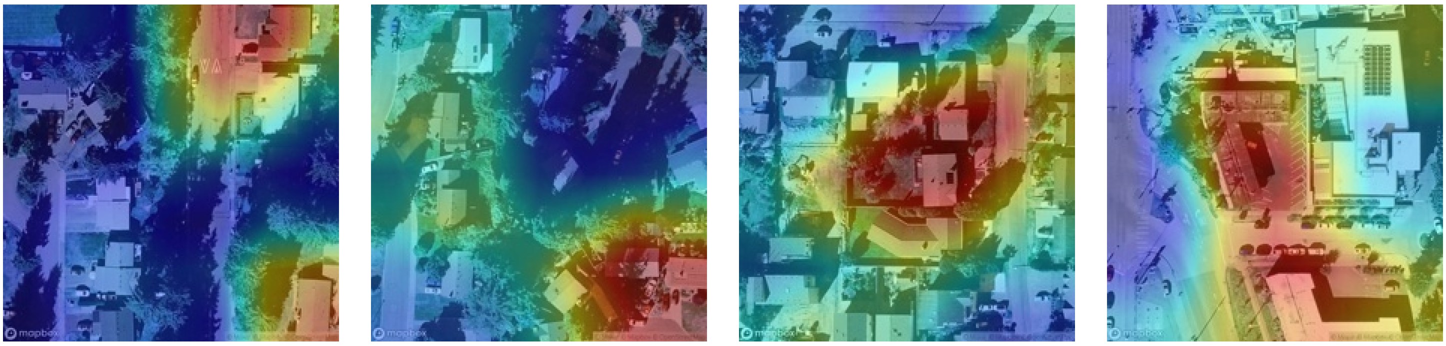
Visual Feature Interpretation

Gradient-weighted Class Activation Mapping (Grad-CAM) is used to interpret the visual features learned from satellite images by highlighting regions that contribute most strongly to the extracted image representations. By generating heatmaps over satellite imagery, Grad-CAM provides qualitative insight into which neighborhood-level visual cues influence the model, such as road networks, building density, green cover, and proximity to water bodies.



Grad-CAM visualizations applied to the CNN feature extractor provide qualitative insights into which regions of satellite images influence learned representations. The highlighted areas often correspond to road density, building concentration, green cover, and proximity to water bodies. These visual cues support the hypothesis that satellite imagery captures neighborhood-level context relevant to property valuation, even if its quantitative impact is limited.



Overall insight

The results indicate that while satellite imagery provides meaningful contextual signals, structured tabular data remains the dominant source of predictive power in this setting. Multimodal learning offers complementary insights but requires higher-resolution imagery, improved fusion strategies, or task-specific visual training to achieve stronger gains.

Conclusion

This project presented an end-to-end multimodal house price prediction framework that integrates structured tabular data with satellite imagery to capture both property-specific and neighborhood-level information. A strong tabular baseline was established using engineered spatial, temporal, and condition-based features, demonstrating that traditional structured attributes explain the majority of price variation.

Satellite images were incorporated through pretrained convolutional neural networks to extract neighborhood-level visual features. Although the quantitative improvement from multimodal learning was limited, qualitative analysis and Grad-CAM visualizations confirmed that the model learned meaningful visual cues related to infrastructure density, green cover, and environmental context. This highlights the complementary role of visual data in refining property valuation.

The experimental results emphasize that while multimodal approaches are feasible and interpretable, their effectiveness depends on image quality, feature fusion strategies, and task-specific visual representations. Overall, the project demonstrates a reproducible and scalable framework for multimodal real estate valuation and provides insights into both the strengths and limitations of integrating satellite imagery with traditional pricing models.