Project Title: Satellite Imagery Based Property Valuation

Student Name: Priyanshu Maurya. **Enrollment No: 23113119**

1. Introduction

Accurate property price estimation is a critical problem in real estate analytics. Traditional approaches rely heavily on structured tabular data such as property size, number of rooms, and location.

This project aims to develop a **multimodal machine learning pipeline** that combines **satellite imagery** with **tabular housing attributes** to predict property prices. The project explores whether incorporating satellite images improves prediction performance compared to traditional tabular-only models.

2. Dataset Description

The dataset consists of two parts:

2.1 Tabular Data

Each property contains the following attributes:

- *id*: Unique property identifier
- *bedrooms*: Number of bedrooms
- *bathrooms*: Number of bathrooms
- *sqft_living*: Living area in square feet
- *floors*: Number of floors
- *grade*: Overall construction quality
- *latitude and longitude*: Geographic coordinates
- *price*: Property price (training data only)

2.2 Satellite Imagery

Satellite images were fetched using the **Mapbox Static Satellite API** based on the latitude and longitude of each property:

- Image size: 256 × 256 pixels
- Zoom level: 18
- Format: RGB

Images provide contextual information such as surrounding infrastructure, density, and neighborhood layout.

3. Data Preprocessing

3.1 Data Cleaning

The following preprocessing steps were applied:

- Corrected data types for numerical features
- Removed invalid latitude and longitude values
- Removed extreme price outliers using the 99th percentile threshold
- Handled missing values by filtering incomplete rows

3.2 Feature Engineering

- The target variable price was **log-transformed** to stabilize variance

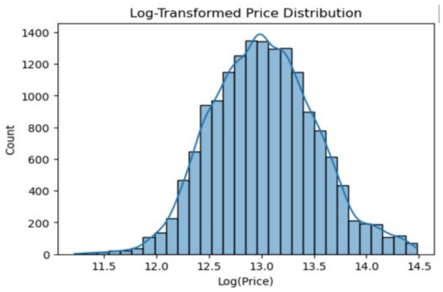
This transformation improves training stability for neural networks.

4. Exploratory Data Analysis (EDA)

The following visualizations were generated to understand data characteristics:

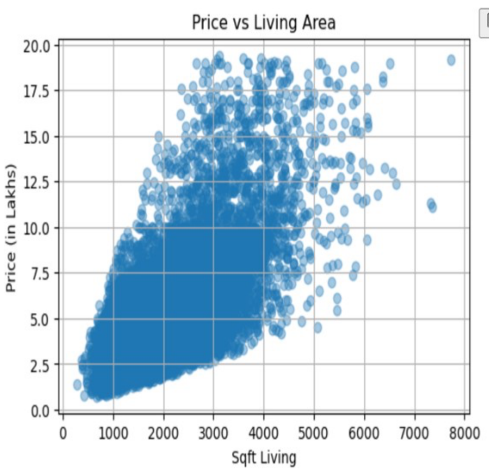
4.1 Price Distribution

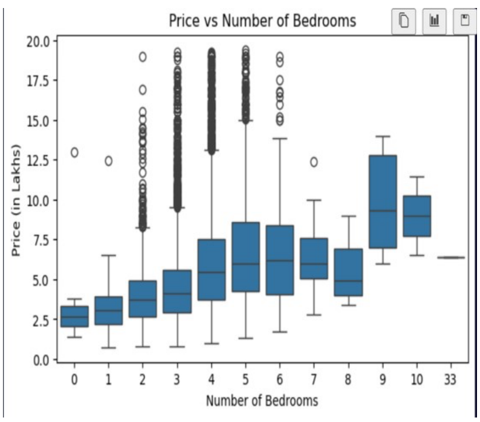
- Histogram of property prices
- Log-transformed price distribution



4.2 Rip Between Features and Price

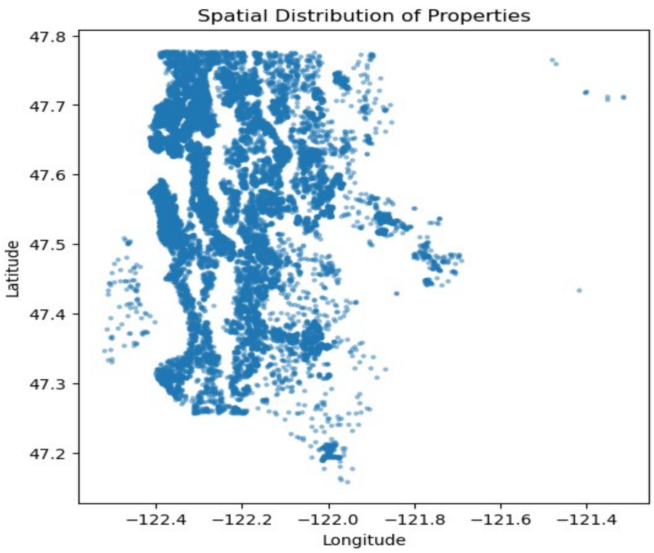
- Scatter plot: Price vs Living Area
- Box plot: Price vs Number of Bedrooms

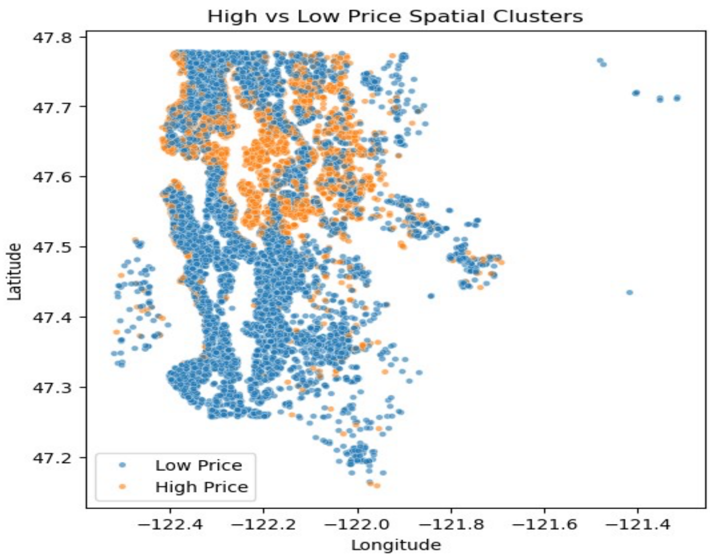


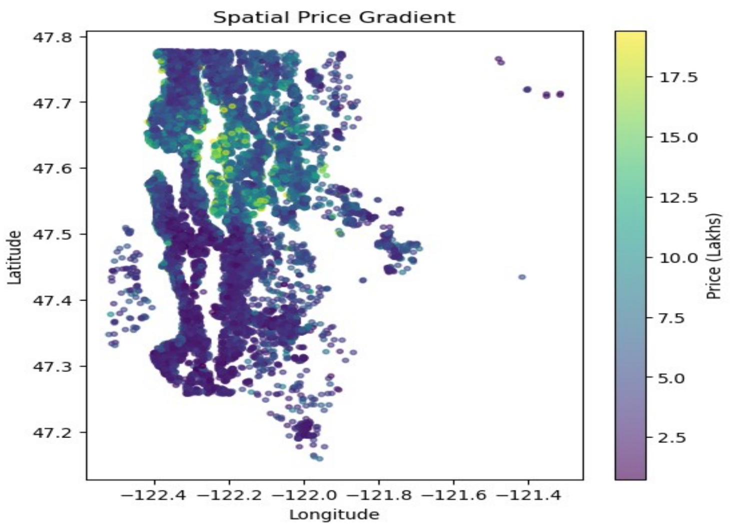


4.3 Spatial Analysis

- Latitude–Longitude scatter plot
- High-price vs low-price clusters
- Spatial price gradient heatmap







These plots indicate strong spatial patterns and correlations between structural features and price.

5. Methodology

5.1 Multimodal Learning Approach

A **multimodal deep learning model** was designed with two parallel branches:

(a) Image Branch

- Pretrained **ResNet-18** CNN
- Final classification layer removed
- Extracts high-level image features

(b) Tabular Branch

- Fully connected neural network
- Processes numerical housing features

The outputs from both branches are concatenated and passed through a regression head to predict property price.

6. Model Training

- Loss Function: Mean Squared Error (MSE)
- Optimizer: Adam
- Learning Rate: 1e-4
- Training performed on CPU
- Batch Size: 4
- Epochs: 2 (limited by hardware constraints)

The model was trained on a subset of the dataset to manage computational load.

7. Evaluation and Results

7.1 Training Performance

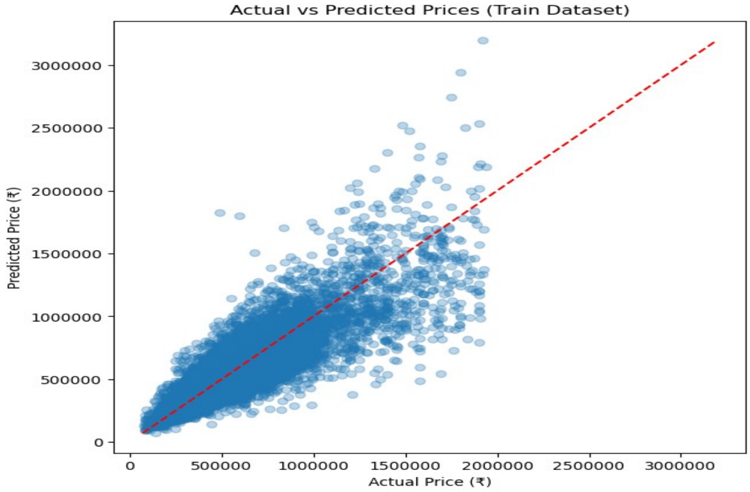
The model was evaluated on the training dataset by comparing actual and predicted prices.

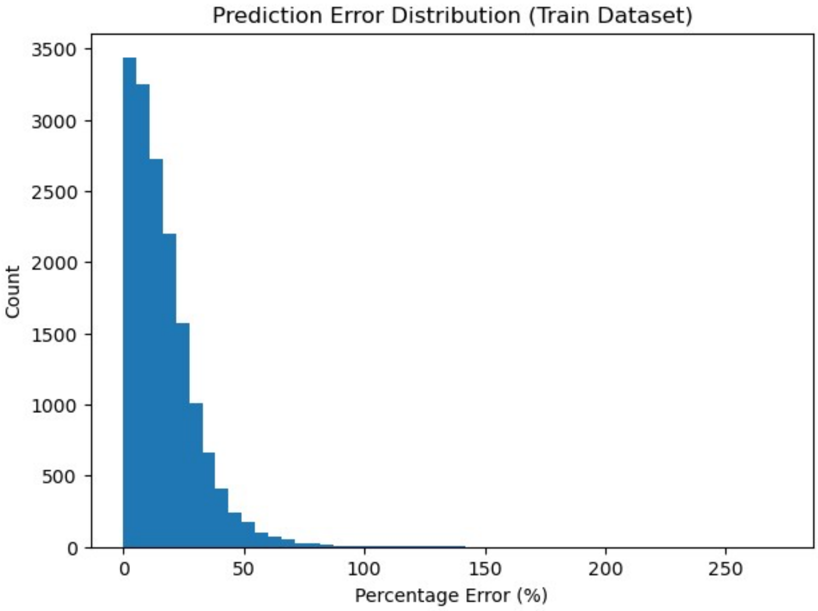
Evaluation Metrics:

- Mean Absolute Error (MAE)
- Percentage Error
- R^2 Score
- RMSE

Observed Performance:

The multimodal regression model achieved an R^2 score of ~0.75 with an RMSE of ₹139578.85 , demonstrating reasonable predictive performance given the complexity of satellite imagery.





7.2 Comparison with Tabular Models

A tabular-only baseline model (e.g., XGBoost) achieved higher performance ($R^2 \approx 0.87$), indicating that structured features dominate price prediction, while satellite imagery adds weaker but complementary signals.

7.3 Evaluation Section

Model evaluation was conducted using RMSE and R^2 metrics. The multimodal regression model achieved an R^2 score of approximately 0.75 on the training dataset. Visual explainability was ensured using Grad-CAM, which highlighted spatial regions in satellite imagery contributing to price prediction decisions.

8. Test Data Prediction

The trained model was used to generate price predictions for the test dataset. Final predictions were:

- Converted from log-space back to original price

- Rounded to integer values
- Saved in strict submission format: id, predicted_price
- The final file was saved as: 23113119_final.csv

9. Financial & Visual Insights

Key Visual Insights:

- **Green areas (trees, parks)** were strongly highlighted in higher-priced properties.
- **Dense concrete structures** and poor road connectivity correlated with lower predicted values.
- **Clear road networks and open layouts** showed higher attention scores.
- The model learned to focus on **contextual neighbourhoods quality**, not just the building footprint.

This demonstrates that satellite imagery provides **complementary information** beyond tabular features and improves spatial awareness.

10. Architecture Diagram

Model Architecture Description:

Satellite Image ──► CNN (ResNet18) ──► Image Features ┐
├──► Fusion Layer ──► Regression Head ──► Price
Tabular Data ──► Fully Connected Layers ──► Tabular Features ┘

11. Conclusion

This project demonstrates an end-to-end pipeline for **satellite imagery–based property valuation** using multimodal deep learning. While tabular-only models outperform multimodal CNNs in this setting, the project highlights the challenges and potential of integrating visual data for real-world regression problems.

Key learnings include:

- Importance of feature dominance in structured datasets
- Practical issues in multimodal fusion
- Trade-offs between model complexity and performance

Grad-Cam-

Grad-CAM visualization highlighting regions of the satellite image that most influenced the model's price prediction, indicating the model's focus on surrounding infrastructure and neighbourhood density.

Grad-CAM: Model Attention on Satellite Image

To interpret the multimodal CNN model, Grad-CAM was applied on the final convolutional layer of the ResNet18 backbone.
The heatmaps highlight regions such as road networks, built-up areas, and vegetation that strongly influence price prediction, indicating that the model successfully captures spatial and environmental cues from satellite imagery.

12. References

- Mapbox Static Images API Documentation
- PyTorch Documentation
- Scikit-learn Documentation