# Satellite Imagery-Based Property Valuation

## Overview

In this project, our goal was to predict house prices using both tabular data and satellite images.

Normally, house price prediction is done only using tabular data such as:

- bedrooms
- bathrooms
- living area
- location

But in this project, we wanted to go one step further and check whether satellite images of the surrounding area (trees, roads, water, buildings) can help improve price prediction.

To do this, I built:

- a tabular **machine learning** model
- a **deep learning CNN model** using satellite images
- a **fusion model** that combines both.

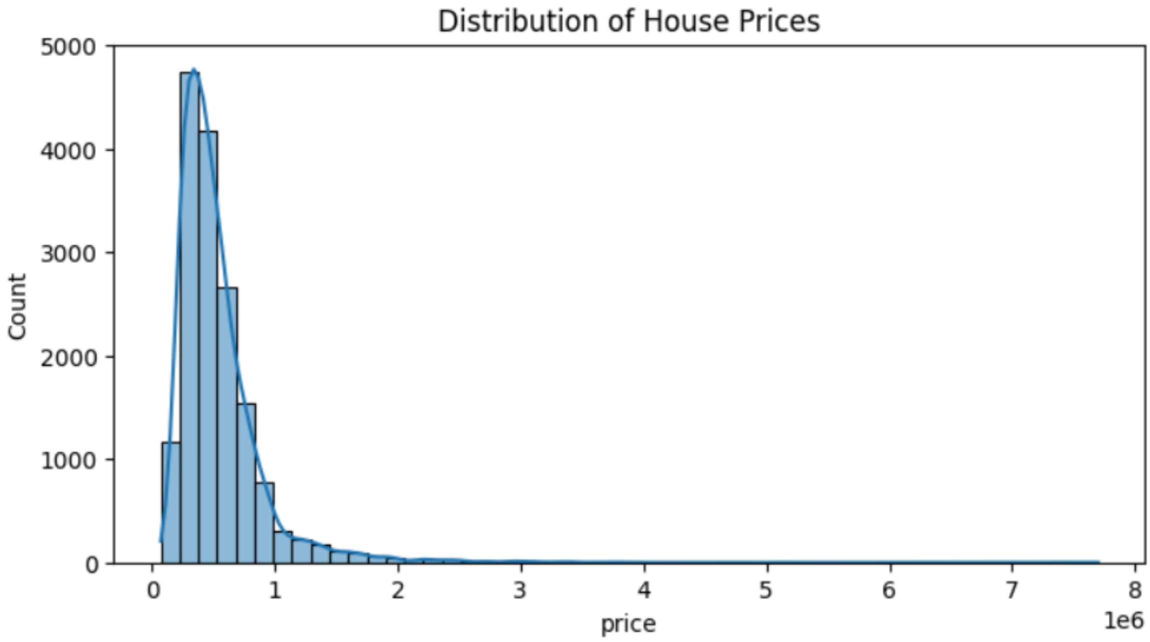
Finally, I compared all models and selected the best one based on performance.

## Exploratory Data Analysis (EDA)

### Price Distribution

First, I analyzed the distribution of house prices. I observed that:

- Most houses are priced in the middle range
- Very expensive houses are fewer
- The price distribution is right-skewed



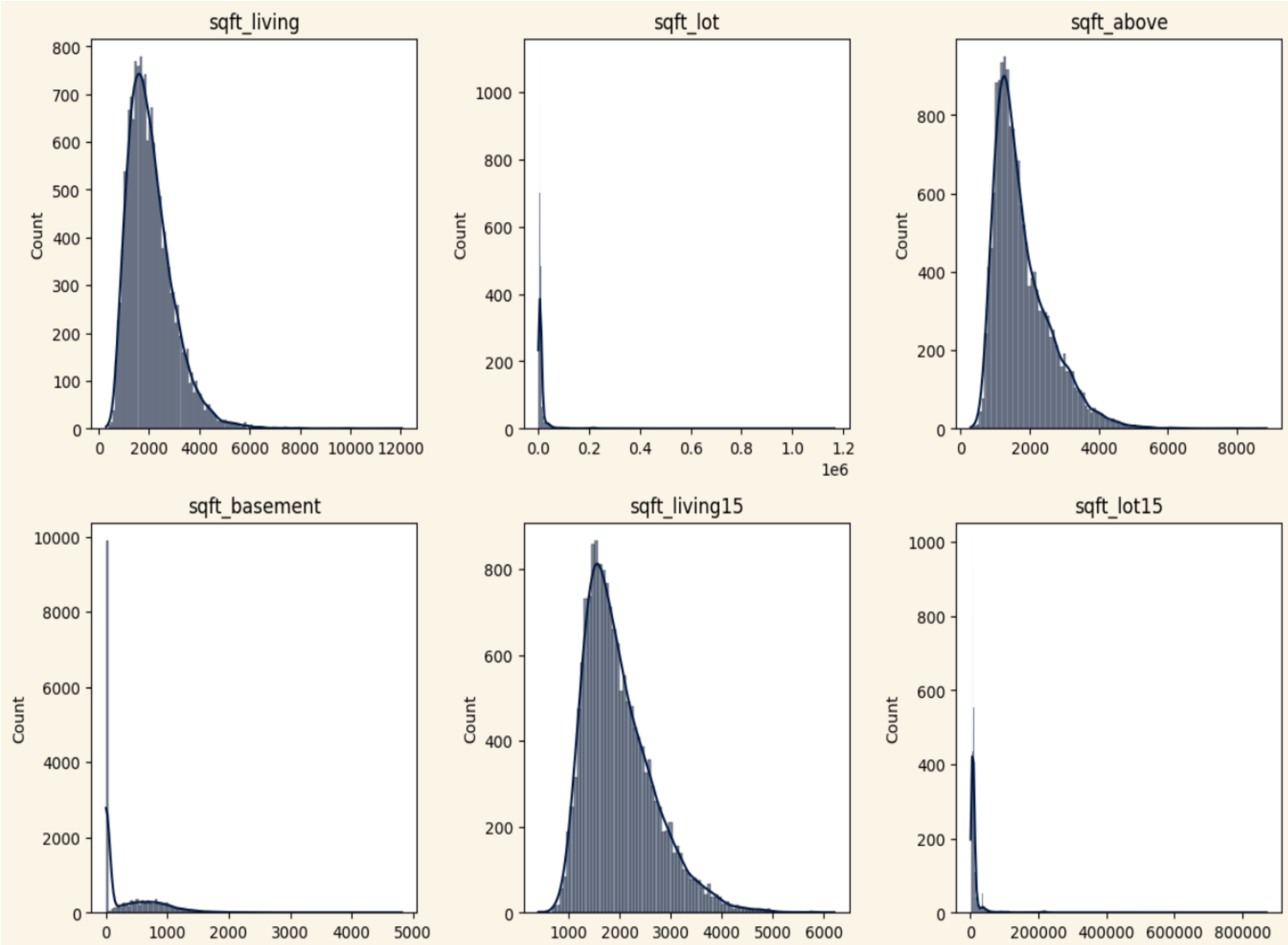

**Area-Based Features (sqft)**

Next, I analyzed area-related features such as:

- sqft_living
- sqft_lot
- sqft_above
- sqft_basement

From the plots, I observed that:

- Houses with larger living area usually have higher prices
- Basement area adds value, but less than living area
- Lot size varies a lot and is not always a strong price indicator




## House Prices by Location (Latitude–Longitude Plot)

To understand how location affects house prices, I plotted house prices on a latitude–longitude scatter plot.

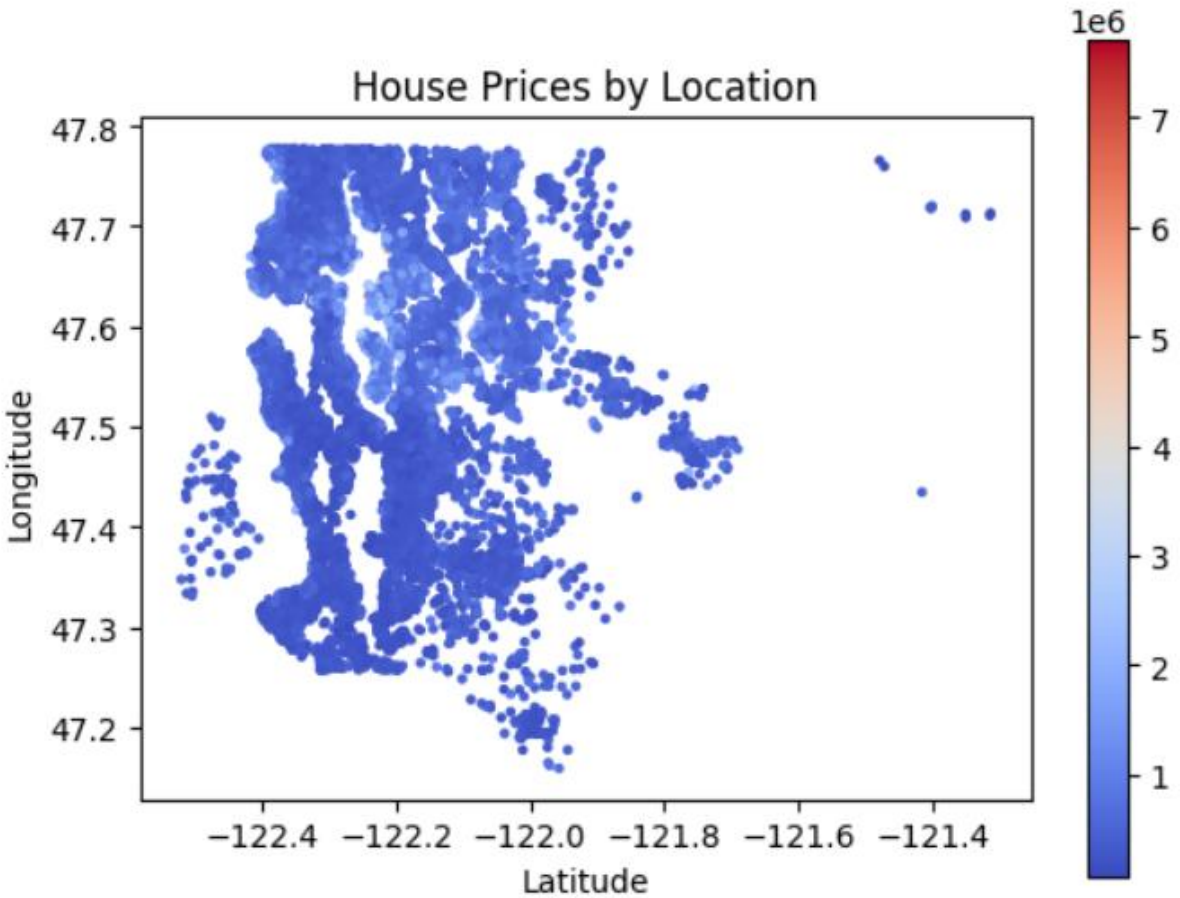
In this plot:

- Each point represents a house
- The x-axis shows latitude
- The y-axis shows longitude
- The color represents the house price

### What I observed

- Houses are not evenly distributed across locations

- Certain regions have a high concentration of houses, which indicates urban or residential clusters
- Areas with higher prices tend to appear grouped together, rather than being randomly spread
- This shows that location plays a very important role in determining house prices




**Correlation Heatmap**

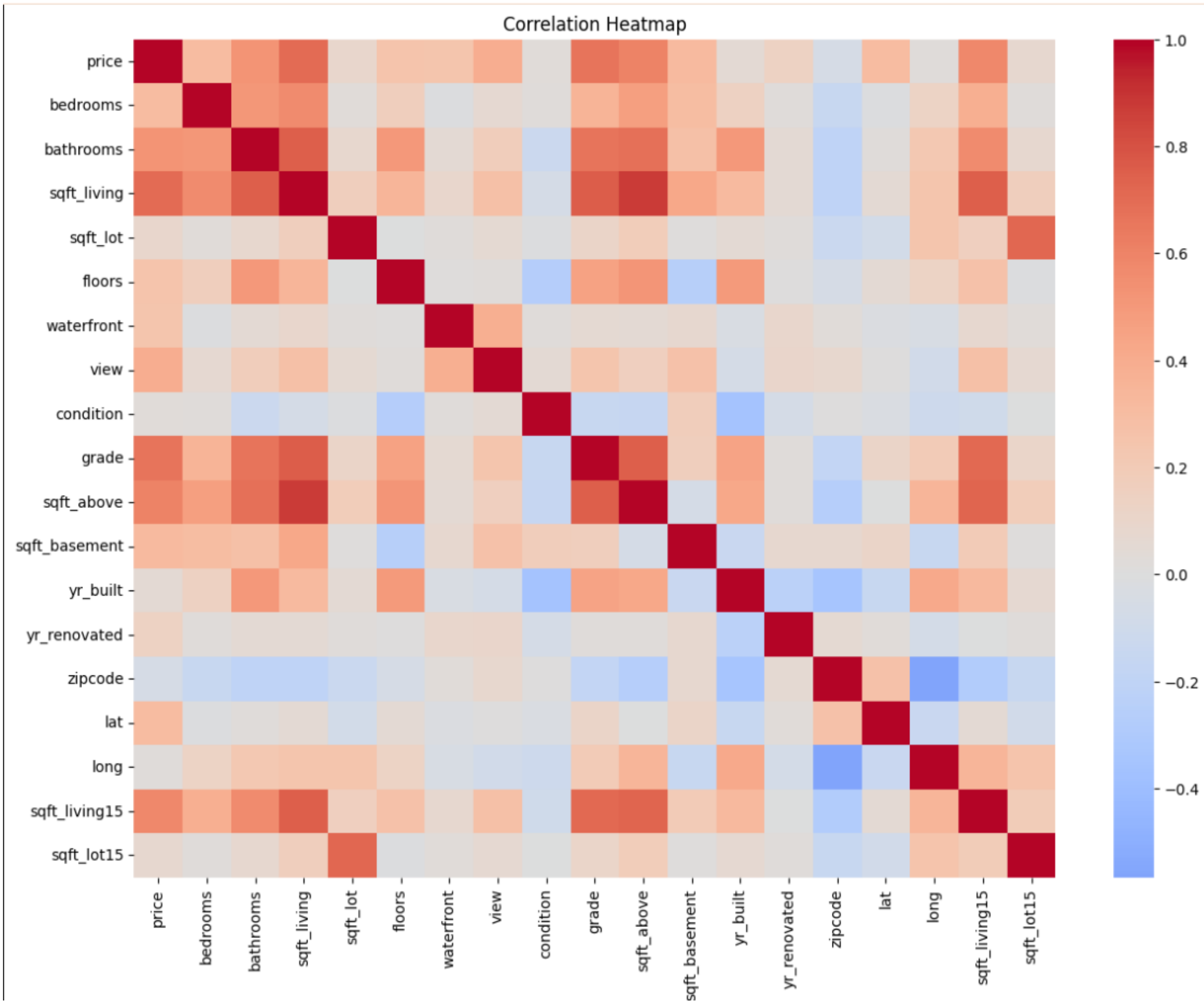
To understand how different features are related to house prices, I created a correlation heatmap.

From the heatmap, I observed that:

- sqft_living has a strong positive correlation with house price
- grade is also highly correlated with price

- bathrooms and sqft_above show moderate positive correlation
- Features like sqft_lot and zipcode have weaker correlation with price

This means that bigger houses with better construction quality usually cost more.




## Financial/Visual Insights

To understand what visual features influence house prices, I used Grad-CAM on the CNN model.

### Green Areas

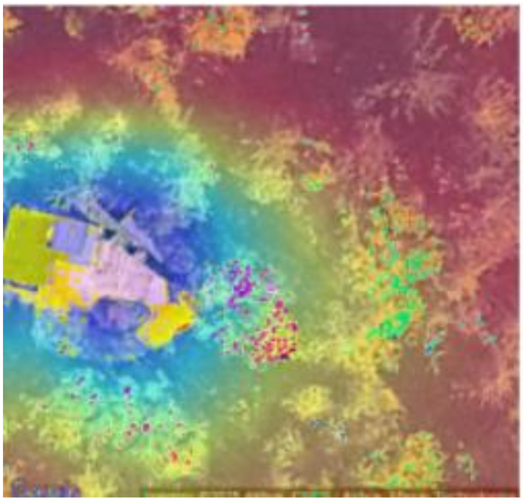
- Houses surrounded by trees and open spaces are usually more expensive
- The CNN focused more on green regions for high-priced houses.

**Roads and Connectivity**

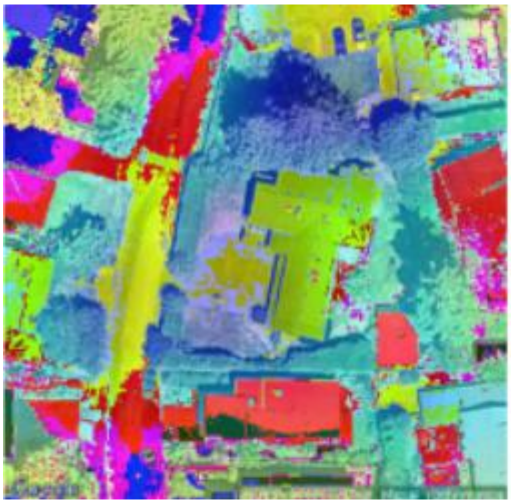
- Good road connectivity increases house value
- Very dense road networks sometimes reduce value due to noise and traffic

**Dense Concrete Areas**

- Very crowded building regions generally reduce price
- Less open space usually means lower value



**Grad-CAM visualization for a low-priced house**



**Grad-CAM visualization for a high-priced house**

## Architecture Diagram

### Tabular Model (XGBoost)

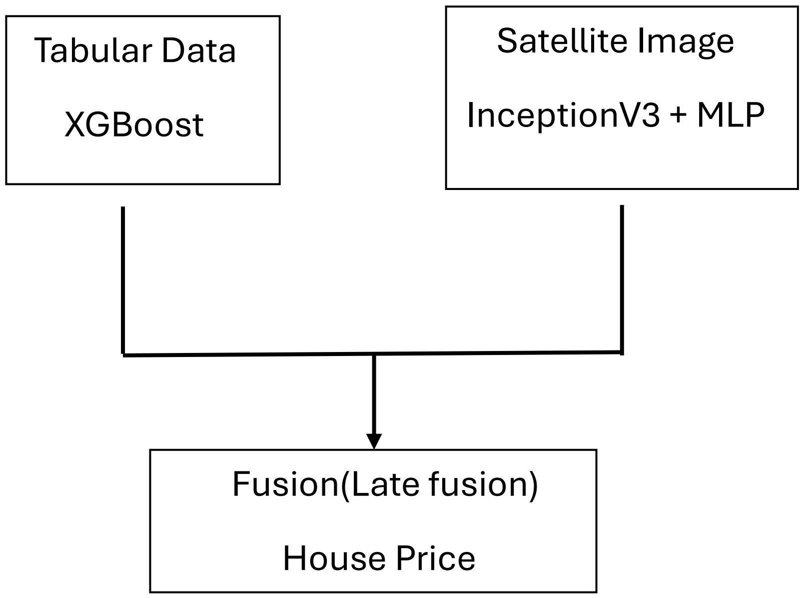
- Uses numerical features like area, rooms, and location
- This model performed very well and captured most price patterns

### Image Model (CNN – InceptionV3)

- Uses satellite images as input
- Learns visual patterns such as greenery and density

### Fusion Approach

- Outputs from both models were combined and evaluated

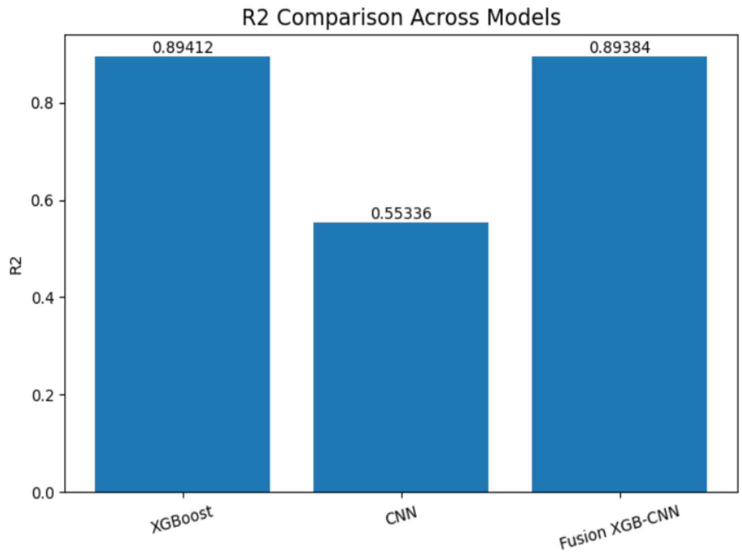


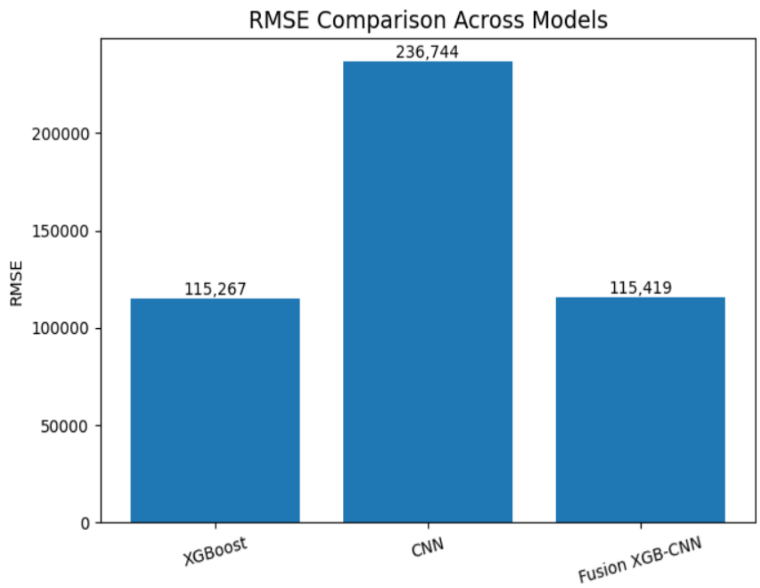

***Multimodal architecture combining tabular features using XGBoost and satellite images using an InceptionV3-based CNN. The outputs of both models are fused to predict house prices***

## Results

| **Model** | **R2 score** | **RMSE** |
|---|---|---|
| Tabular Data (XGBoost) | 0.8941 | 115,267 |

| | | |
|---|---|---|
| CNN (Images) | 0.5536 | 236,744 |
| Tabular + Images (Fusion) | 0.8938 | 115,419 |




## Observation

Even though satellite images contain useful visual information, the tabular features already explain most of the house price variation.

Adding satellite images increased model complexity but did not improve accuracy.

## Final Conclusion

- Tabular data is very strong for house price prediction
- Satellite images help in understanding the surroundings
- Fusion models should be used only when images add new information