Overview Approach and Modeling Strategy

This project builds a property price prediction system using both tabular housing data and satellite images. The idea is that tabular data captures structural and location details of a house, while satellite images capture the surrounding environment such as roads, development density, and nearby infrastructure. The goal is to check whether combining these two sources improves valuation accuracy.

The approach is kept simple and controlled. First, a strong baseline model is built using only tabular data with an XGBoost regressor. Then, satellite images for each property are processed using a pretrained ResNet50 model to extract visual features. These image features are combined with tabular features using a late fusion approach. To reduce noise from high dimensional image features, PCA is applied before combining them with tabular data. The focus throughout the project is on fair comparison, clear evaluation, and interpretability rather than over optimization.

Exploratory Data Analysis EDA

The price variable is right skewed, so a logarithmic transformation is applied during modeling. Strong nonlinear relationships are observed between price and features such as living area, construction grade, and location coordinates. These features clearly drive most of the variation in house prices.

Geographic analysis shows clear spatial clustering of prices. Houses located near water, urban centers, or dense neighborhoods tend to have higher prices, while houses in sparsely developed areas are cheaper. Satellite image samples further highlight these differences. Lower priced houses are often located in areas with fewer roads and less development, while higher priced houses are surrounded by dense road networks, developed neighborhoods, and urban infrastructure. This analysis motivates the use of satellite imagery as an additional data source.

Financial and Visual Insights

Analysis of satellite images and Grad CAM visualizations shows that the CNN focuses on large scale environmental patterns rather than individual buildings. High value properties are associated with dense road networks, structured urban layouts, developed surroundings, and infrastructure rich regions. These areas usually appear visually complex and highly connected.

In contrast, areas with open land, sparse development, or uniform textures show weaker visual activation and are generally linked to lower property prices. This suggests that dense built environments and concrete heavy regions are strong visual indicators of higher value. However, many of these visual signals are already indirectly captured by tabular features such as latitude, longitude, neighborhood statistics, and waterfront indicators. As a result, satellite imagery provides overlapping information rather than completely new signals in this dataset.

Architecture Diagram

The system follows a late fusion architecture. Tabular data and satellite images are processed separately and combined only at the final regression stage. Satellite images are resized and normalized, then passed through a pretrained ResNet50 model to extract image embeddings. These embeddings are optionally reduced using PCA. Tabular features are taken directly from the dataset. Both feature sets are concatenated and passed to a regression model to predict the final house price. This design keeps the system modular, interpretable, and easy to debug.

Results Tabular vs Multimodal Comparison

The tabular only model performs very well and explains almost 90 percent of the variation in house prices. When raw high dimensional image embeddings are added, model performance drops due to noise and redundancy. After applying PCA to reduce image feature dimensions, multimodal performance improves and becomes close to the tabular baseline but does not clearly surpass it.

This shows that satellite imagery does capture meaningful environmental information, but its added value is limited when detailed tabular features already encode location and neighborhood effects. The results highlight that multimodal learning works best when different data sources provide complementary information rather than overlapping signals.