1. Introduction

This project focuses on predicting house prices using structured housing attributes along with geographical information. In addition to traditional tabular features, satellite images are fetched using Google Maps Static API to capture visual context such as neighborhood layout and infrastructure.

2. Dataset Description

The dataset consists of a training CSV file containing house attributes and prices, and a test CSV file with similar features excluding the target variable. Key features include number of bedrooms, bathrooms, square footage, condition, grade, and latitude-longitude coordinates.

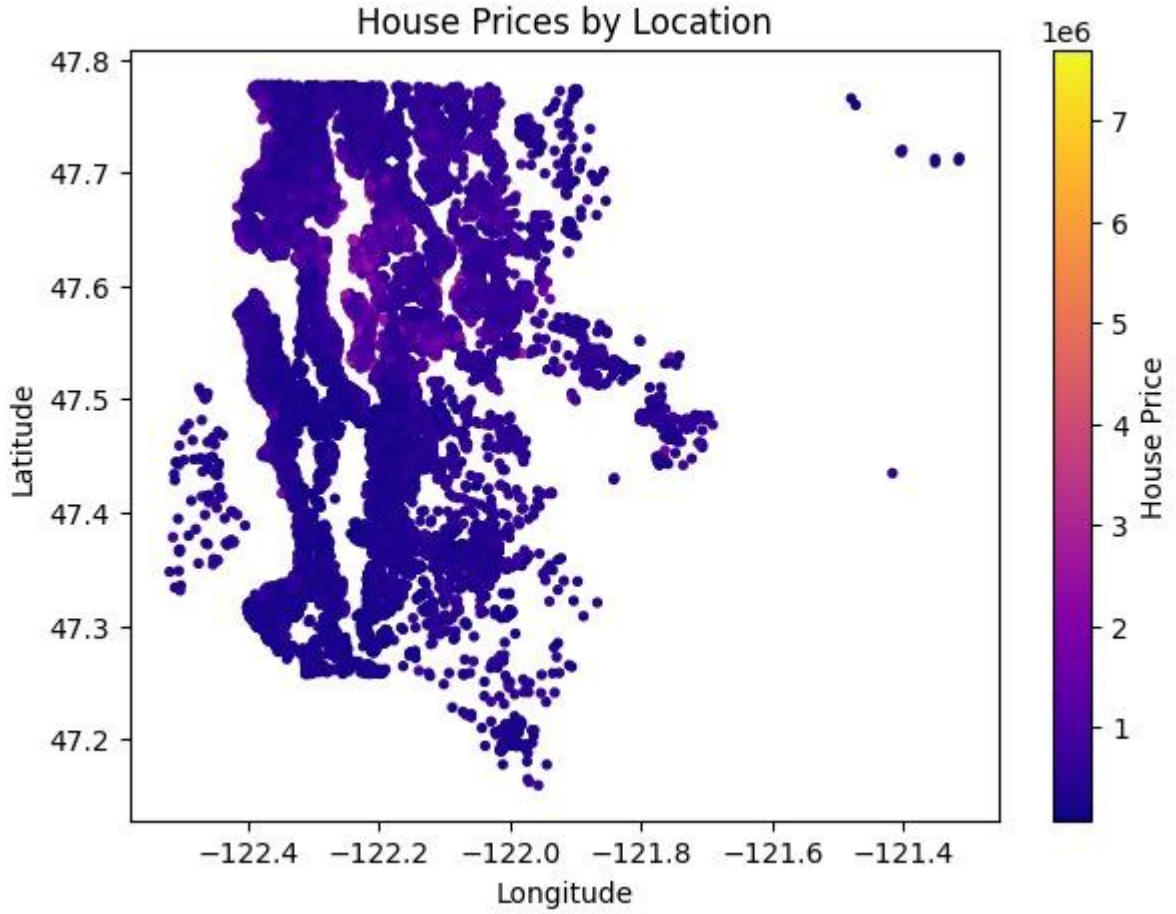
3. Satellite Image Collection

Satellite images are fetched using the Google Maps Static API. Each image corresponds to a house location defined by its latitude and longitude. Images are stored locally and can be used for future CNN-based feature extraction.

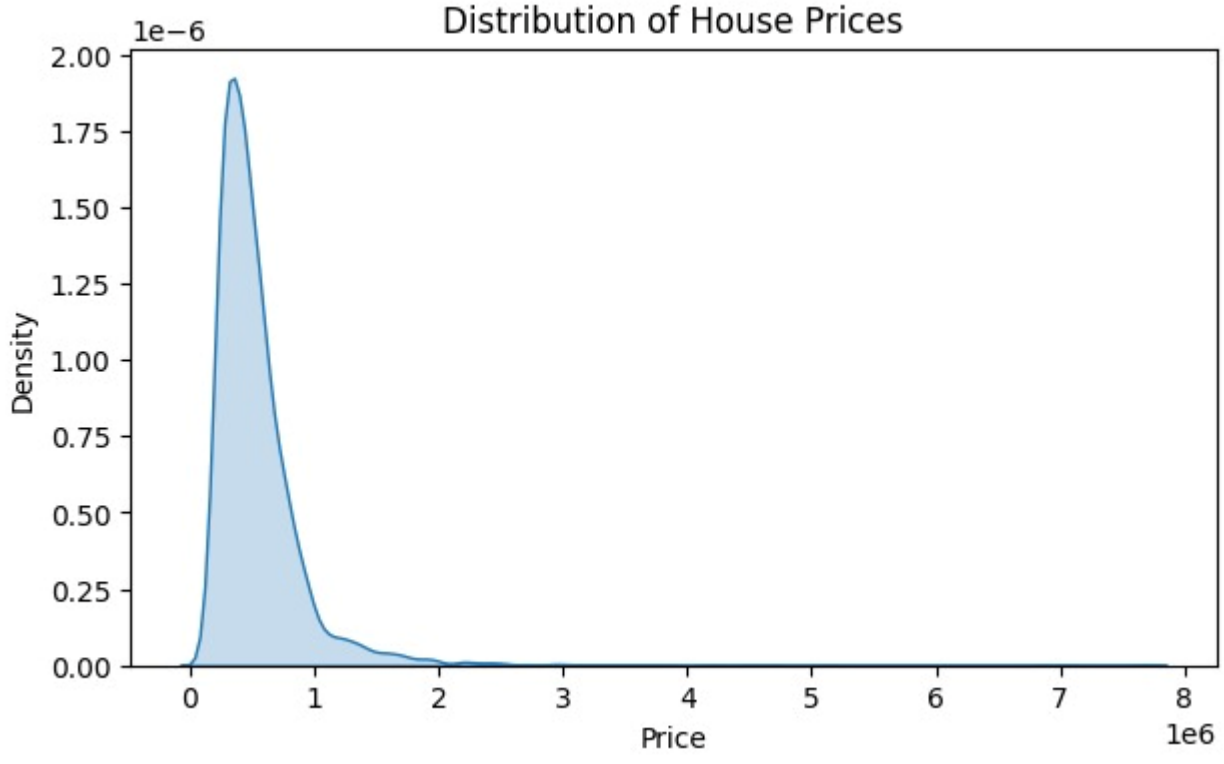
4. Data Preprocessing

Numerical features are selected for modeling. Missing values are handled using median imputation. Feature scaling is applied using standardization, and the dataset is split into training and validation subsets.

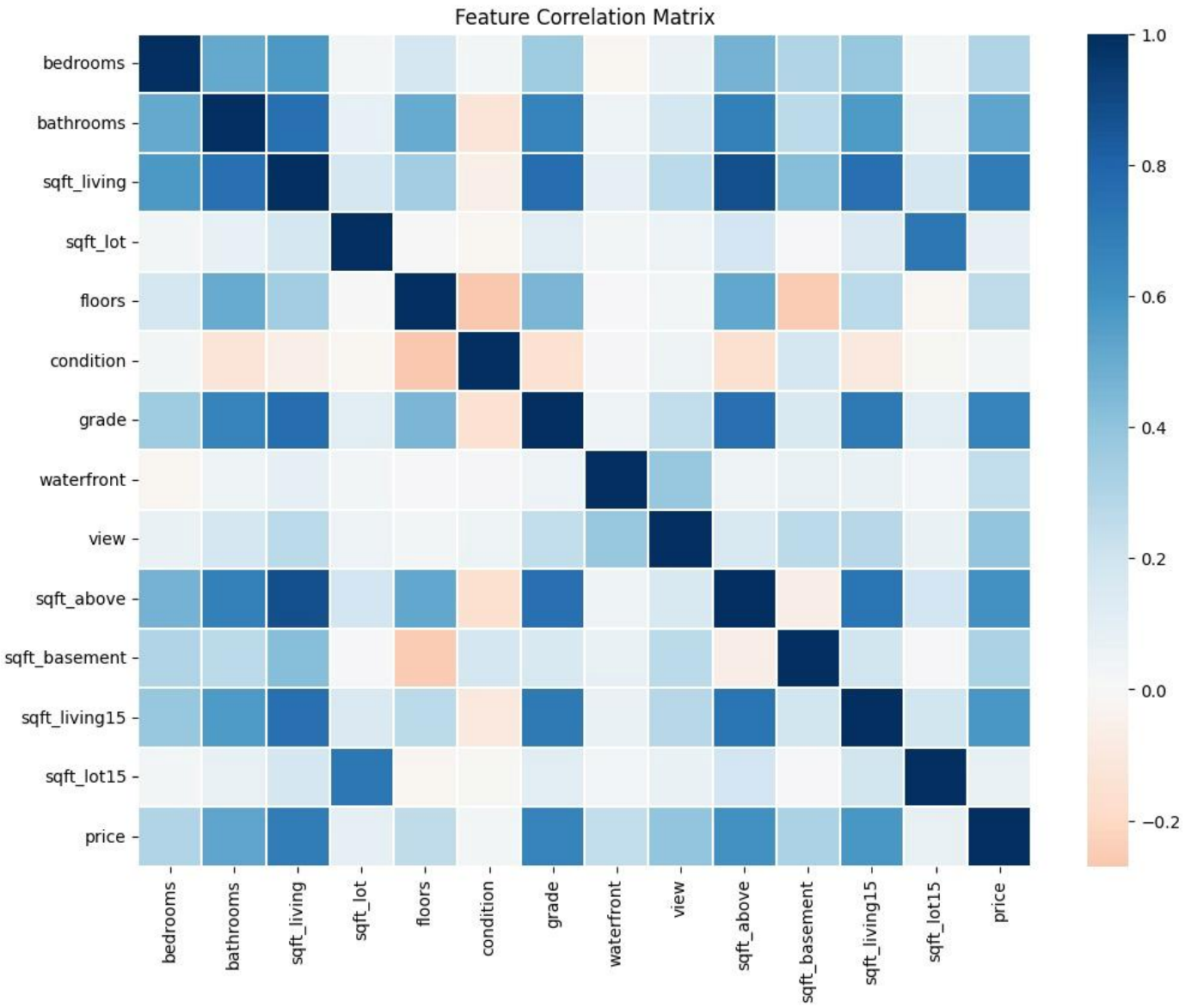
House prices by location



Distribution of house prices



Feature Correlation Matrix



5. Model Architecture

Two models are used for price prediction: a Neural Network implemented in PyTorch and an XGBoost Regressor. The neural network consists of fully connected layers with ReLU activation, while XGBoost is used as a strong baseline for tabular data.

6. Evaluation Metrics

Model performance is evaluated using Root Mean Squared Error (RMSE) and R-squared (R^2) score, which measure prediction accuracy and explained variance respectively.

7. Results

XGBoost outperformed the neural network in terms of RMSE and R^2 score, demonstrating its effectiveness on structured housing data.

8. Conclusion

This project demonstrates an end-to-end machine learning pipeline for house price prediction. The inclusion of satellite imagery provides a foundation for future multimodal learning approaches.

9. Future Scope

Future work may include convolutional neural networks for image feature extraction, multimodal learning, hyperparameter optimization, and deployment as a web application.