Multimodal Real Estate Price Prediction Report

Objective
The objective of this project is to predict residential property prices using a multimodal machine learning framework that combines structured tabular data with satellite imagery. The aim is to show how visual context improves valuation accuracy over traditional tabular-only models.

Dataset
The dataset contains housing records with numerical attributes such as bedrooms, bathrooms, living area, lot size, floors, condition, and geographic coordinates. Satellite images were downloaded using latitude and longitude to capture neighborhood-level information such as land use and urban density.

Tabular-Only Baseline
A baseline regression model was trained using only tabular features after preprocessing and normalization. While this approach captures intrinsic property characteristics, it cannot model environmental context, limiting prediction performance.

Multimodal Model
For the multimodal model, satellite images were passed through a pretrained ResNet-18 network to extract 512-dimensional embeddings. These embeddings were concatenated with tabular features and used as input to a regression model, enabling joint reasoning over numerical and visual information.

Training & Evaluation
The combined feature vectors were trained using Ridge Regression with an 80-20 train-validation split. Performance was evaluated using Root Mean Squared Error (RMSE) and R^2 score.

Results
The multimodal approach achieved strong predictive performance on the validation set:

Validation RMSE: 65,335.97
Validation R^2: 0.9695

This represents a significant improvement over the tabular-only baseline, demonstrating the value of incorporating satellite imagery.

Conclusion
The results confirm that multimodal learning enhances real estate price prediction by integrating spatial and environmental context. This approach is scalable and can be extended with higher-resolution imagery, end-to-end neural fusion, and temporal data.