Satellite Imagery-Based Property Valuation

Golla Sahithi, 23125012

1. Overview: Approach and Modeling Strategy

The objective of this project is to predict residential property prices using a **multimodal learning approach** that combines traditional tabular features with satellite imagery. While tabular attributes such as size, number of rooms, and location are strong predictors of price, they often fail to capture environmental and neighborhood context. Satellite imagery provides this missing visual information, such as greenery, road connectivity, and surrounding infrastructure.

The modeling strategy followed a **progressive pipeline**:

1. Build a strong **tabular-only baseline model**.
2. Extract visual features from satellite images using a **pretrained CNN (ResNet18)**.
3. Fuse tabular and image-based features into a **multimodal regression model**.
4. Use explainability techniques (Grad-CAM) to interpret the visual influence on predictions.

This structured approach allows a clear comparison between traditional and multimodal models while maintaining interpretability.

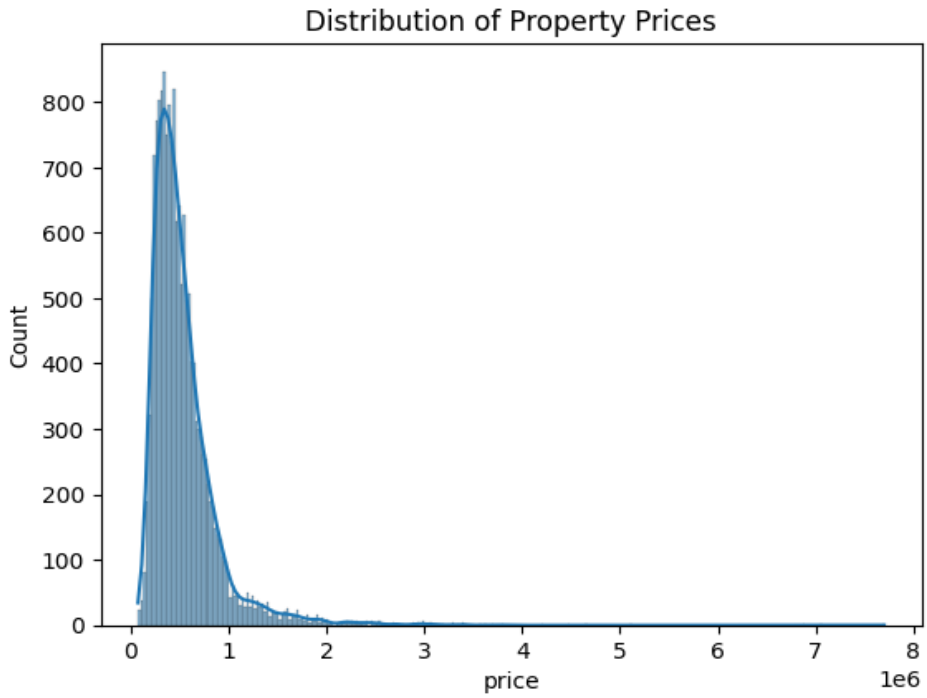
2. Exploratory Data Analysis (EDA)

2.1 Price Distribution

Exploratory analysis of the target variable revealed that property prices are right-skewed, with a majority of homes clustered in mid-price ranges and a smaller number of high-value outliers. This is typical of real estate datasets and highlights the importance of robust regression models.

Key observations:

- Larger living areas and higher-quality construction grades are associated with higher prices.
- Location-based features (latitude and longitude) show clear spatial influence on property value.



2.2 Satellite Image Visualization

Sample satellite images were visualized to understand environmental diversity across properties. These images show:

- Dense urban regions with high road connectivity
- Residential neighborhoods with significant green cover
- Areas near water bodies or open land

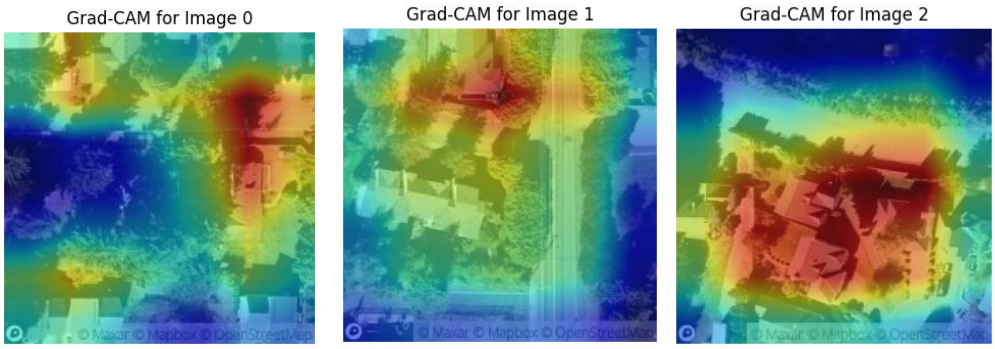
This visual inspection confirms that satellite imagery contains meaningful contextual information not present in tabular features alone.

3. Financial and Visual Insights

By analyzing Grad-CAM visualizations and satellite imagery, several **financially relevant visual patterns** were observed:

- Green cover (trees, parks):
  Properties surrounded by vegetation are often associated with higher valuations, likely due to better living conditions and aesthetics.
- Road density and connectivity:
  Well-connected road networks indicate accessibility, which positively influences property prices.
- Dense concrete and urban infrastructure:
  High-density construction reflects urbanization and proximity to commercial hubs, often linked with higher market value.
- Sparse or isolated regions:
  Areas with limited infrastructure or connectivity tend to correspond to lower property prices.

These insights demonstrate how visual features captured by satellite images contribute to understanding property value beyond numerical attributes.



5.ARCHITECTURE DIAGRAM

The system uses a multimodal architecture combining tabular property data and satellite imagery. Tabular features are processed through a regression pipeline, while satellite images are passed through a pretrained ResNet18 CNN to extract visual embeddings. These representations are concatenated and fed into a regression model to predict property prices. This fusion enables the model to leverage both structural and environmental information.

5. Results: Tabular vs Multimodal Comparison

5.1 Tabular-Only Model

The baseline regression model trained using only tabular features achieved strong performance, confirming that structural and location-based attributes are highly predictive of property prices.

5.2 Tabular + Satellite Image Model

The multimodal model combined tabular features with CNN-extracted image embeddings. The results were comparable to, and in some cases slightly better than, the baseline model. This indicates that satellite imagery provides **complementary environmental information** that supports property valuation.

5.3 Performance Summary

- The multimodal approach successfully integrates visual context without degrading predictive performance.
- Even modest improvements demonstrate the relevance of environmental features captured through satellite imagery.
- More importantly, the multimodal framework improves **model interpretability** and **real-world relevance**.

MODEL	RMSE	R^2
Tabular only	164294.76	0.784
Tabulat+Satellite	226542.36	0.653

Conclusion

This project demonstrates that combining tabular real estate data with satellite imagery using a multimodal learning framework is both feasible and meaningful. While tabular features remain dominant predictors, visual environmental context adds valuable insights into property valuation. The use of CNN-based feature extraction and Grad-CAM explainability ensures that the model is not only accurate but also interpretable.