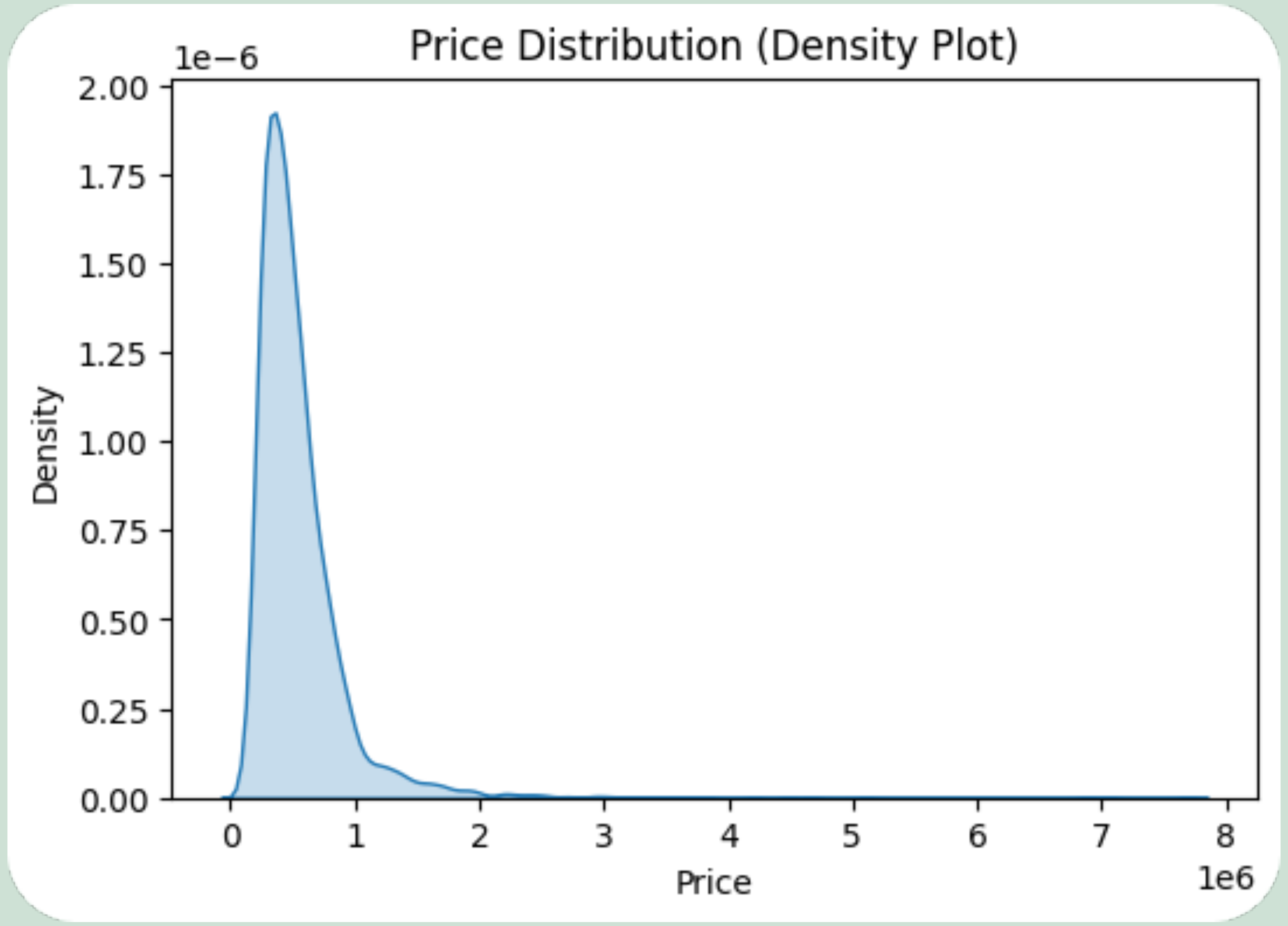

Career Development Cell
YHills
Learning Beyond Expectations
OPEN PROJECT
ABHISHEK 22321003

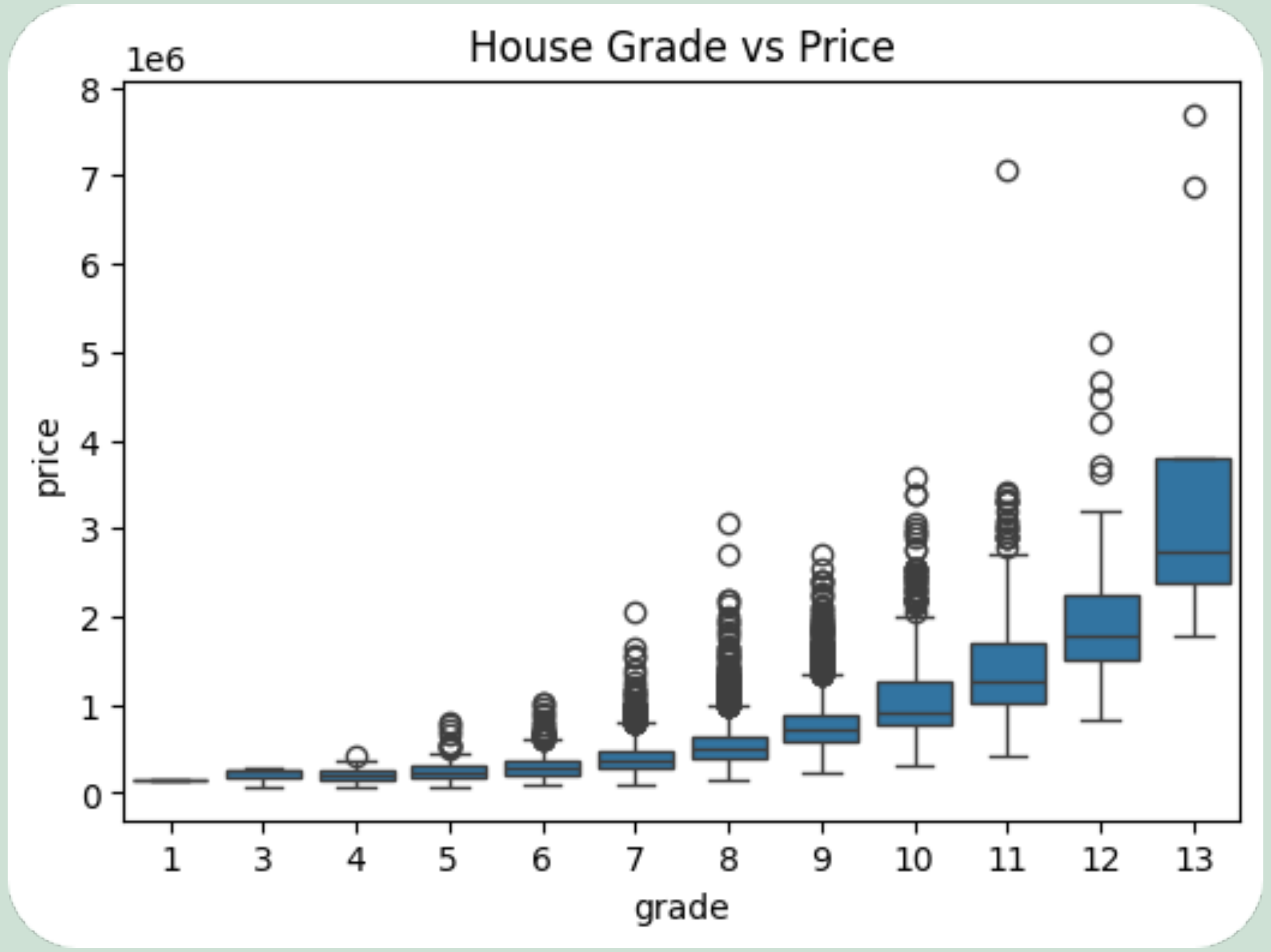
Project Overview

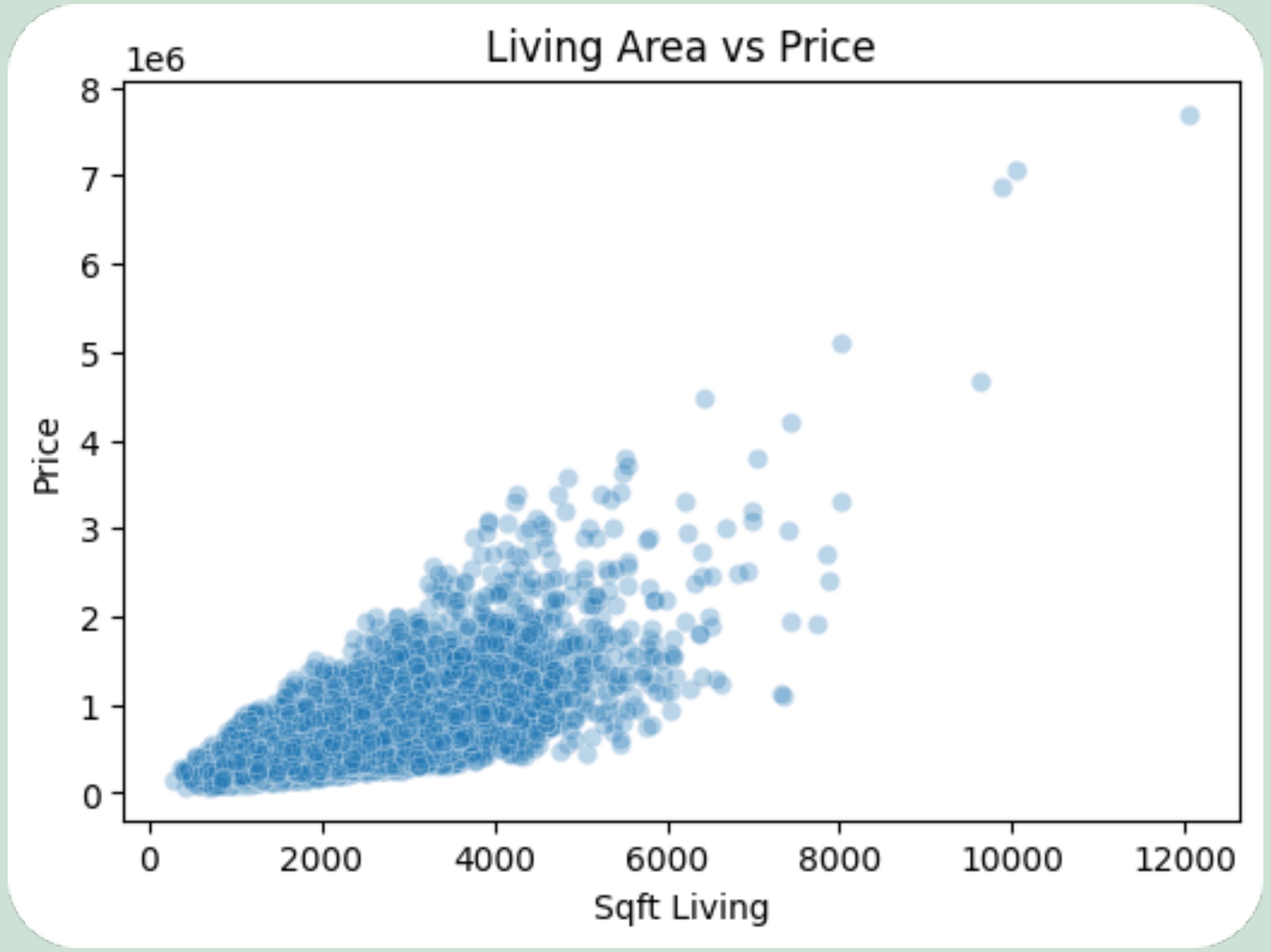
The objective of this project is to predict house prices using a multimodal machine learning approach.Instead of relying only on traditional tabular data (such as property attributes), this project also incorporates satellite imagery to capture spatial and environmental information around properties.

The idea behind this approach is that surrounding infrastructure, greenery, road connectivity, and urban development, visible from satellite images- can significantly influence property prices. By combining tabular data and satellite image features, we aim to build a more informative and robust predictive model.

Exploratory Data Analysis (EDA)







This plot shows the distribution of house prices in the dataset. The distribution is right-skewed, indicating that most properties are priced in the lower to mid-range, while a small number of luxury houses have very high prices. This skewness justifies the use of robust models and explains why extreme values (outliers) need careful handling.

The box plot illustrates a strong positive relationship between house grade and price. As the grade increases, both the median price and price variability increase significantly. Higher grades correspond to better construction quality and design, which strongly influences house valuation. This confirms that grade is a highly informative feature for price prediction.

This scatter plot shows a clear upward trend between living area (square feet) and house price. Larger houses generally have higher prices, although some spread exists due to factors like location and house quality. The pattern indicates that living area is one of the most influential numerical predictors in the dataset.

Data Preprocessing

Tabular Data

The following steps were applied:

Removal of non-predictive identifiers (id)

- Conversion of date features into numerical components (sale_year, sale_month)
- Removal of raw date strings to avoid model errors
- Feature engineering (e.g., living_to_lot_ratio)
- Log transformation of target variable (price) to reduce skewness

Image Data

Satellite images are processed using a ResNet-based CNN as a feature extractor:

- Images resized to 224 × 224
- Normalized and converted to tensors
- Passed through CNN with final classification layer removed
- Output is a 2048-dimensional image embedding per property

These embeddings capture high-level visual patterns relevant to property valuation.

Dimensionality Reduction using PCA

Image embeddings are high-dimensional and may contain redundant information. To address this, Principal Component Analysis (PCA) is applied:

- Image features are first standardized
- PCA reduces dimensions from 2048 → 128
- Most visual variance is preserved while reducing noise

Benefits of PCA:

- Improves computational efficiency
- Reduces overfitting
- Enhances multimodal model performance

Multimodal Feature Fusion

After preprocessing:

- Tabular features represent property attributes
- PCA-reduced image features represent neighborhood context

These two feature sets are concatenated horizontally to form a multimodal feature vector for each property.

This fused representation allows the model to learn from both structured and visual data simultaneously.

Key Observation(Overall Project)

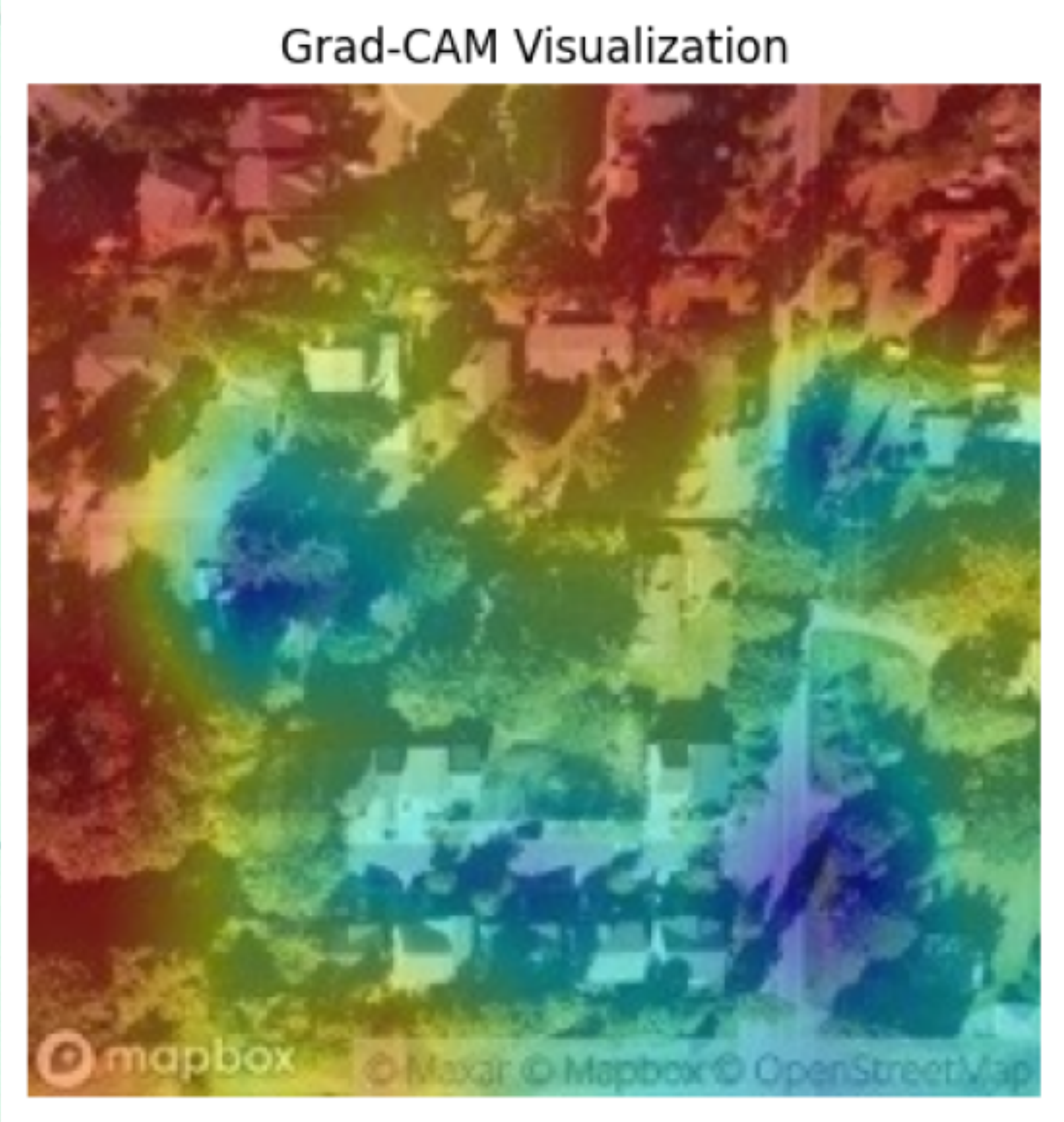
The tabular-only model achieved the best performance with an R^2 of approximately 0.87, indicating that structured property attributes (such as living area, grade, and other numerical features) are highly predictive of house prices.
The multimodal model, which combines tabular features with satellite image embeddings, achieved a comparatively lower R^2 (0.75). This suggests that while satellite imagery adds contextual information, it does not dominate prediction performance in the current setup.
Satellite images were processed using a CNN-based feature extractor (ResNet), and the extracted embeddings were merged with tabular data using the property id. Images were therefore used indirectly as numerical representations, not as raw visual inputs during model training.
Grad-CAM visualisation was used to interpret the image-based model behavior. The heatmap highlights regions such as:

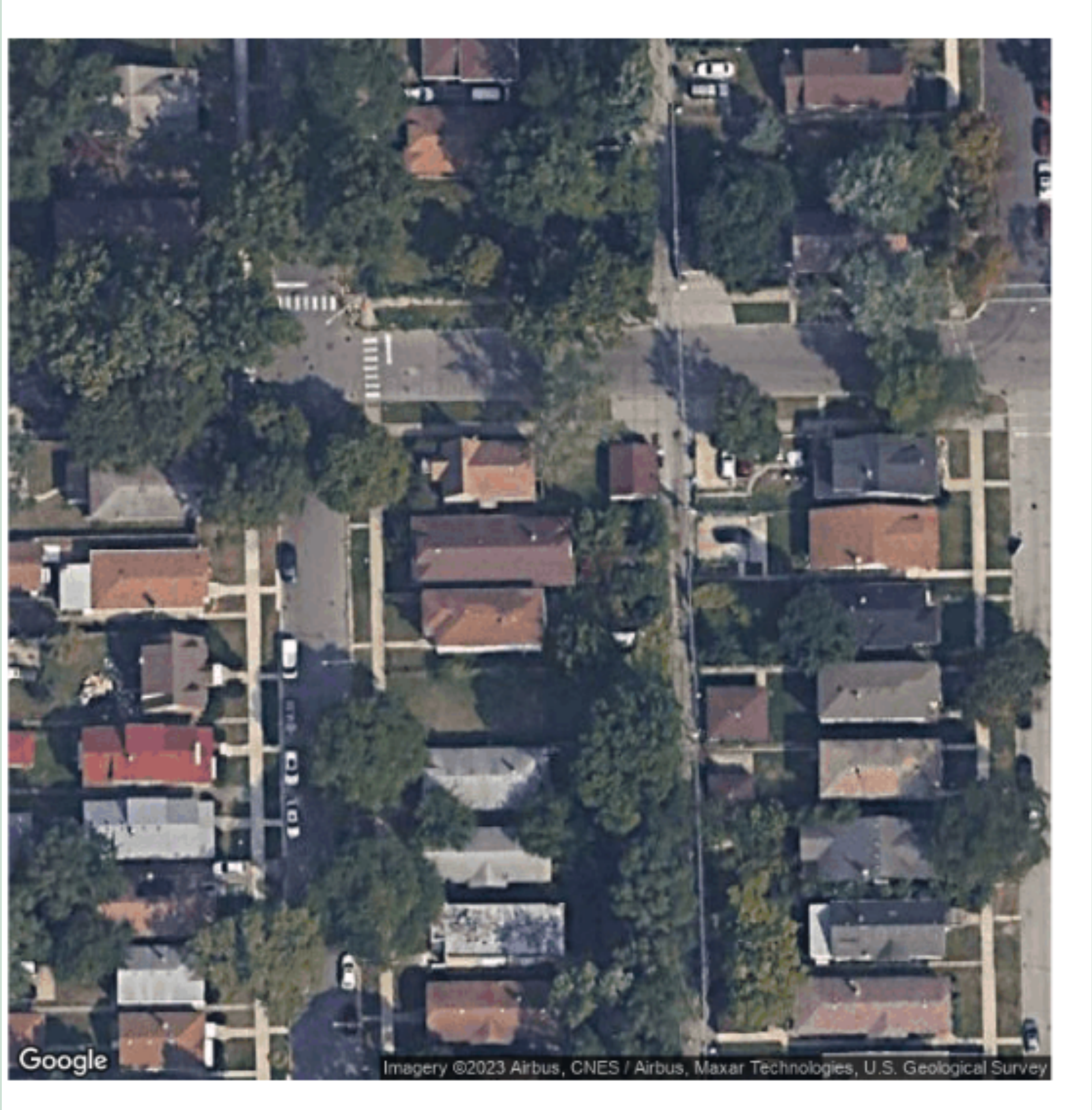
- Built-up areas
- Road networks
- Dense housing clusters

This indicates that the CNN focuses on urban density and infrastructure patterns, which are logically related to property valuation

Performance:	**R^2**	**RMSE**
Tabular :	**0.872**	**126461.20019878604**
Multimodal :	**0.758**	**174919.0024793155**

Key Observation(Overall Project)





Key Observation(Overall Project)

Despite meaningful spatial attention shown by Grad-CAM, the image features alone were not sufficient to outperform tabular data, likely due to:

- Limited image resolution
- Absence of temporal or neighborhood-level metadata
- Simple feature fusion strategy (concatenation)

Dimensionality reduction using PCA was necessary to handle high-dimensional image embeddings, improving model stability but potentially removing fine-grained visual information.