<u>Satellite Imagery-Based Property Valuation Report</u>

- **Sahaj Saxena**

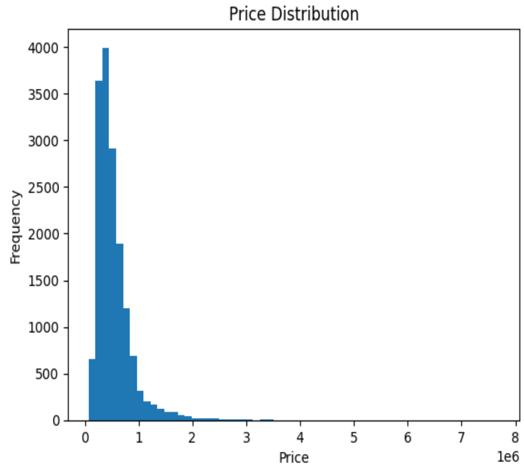
Overview

This project builds a multimodal property valuation system that fuses traditional housing attributes with environmental context captured from satellite imagery. A frozen ResNet-18 CNN is used to generate 512-dimensional visual embeddings, reduced to 50 dimensions using PCA, and concatenated with cleaned numeric features before training a final XGBoost regression model. The model is evaluated using RMSE and R^2, and interpretability is ensured using Grad-CAM overlays.

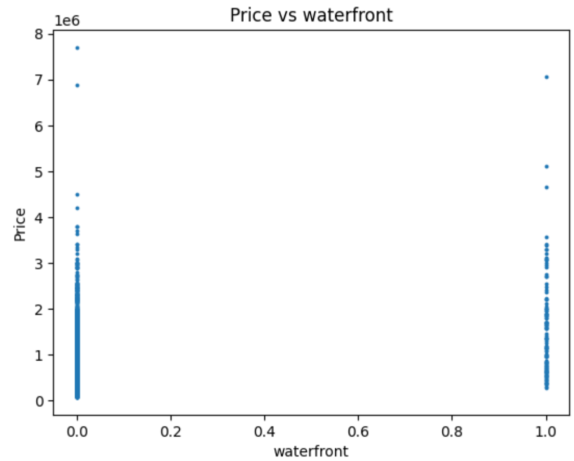
Exploratory Data Analysis (EDA)

Key insights from initial analysis:

- **Price distribution is highly right-skewed**, indicating the presence of premium outlier properties. Log scaling is later applied for visualization purposes in analysis but not used for final model training.
- **Waterfront properties show significantly higher prices**, validating that proximity to water strongly influences valuation.
- **Sqft_living and Grade show strong positive correlation with price**, while renovation year has low overall impact due to sparse non-zero values.
- A **correlation heatmap confirms multicollinearity** between size-based features (sqft_above, sqft_living15, sqft_lot15), justifying dimensionality reduction for image embeddings.
- Sample satellite images were fetched using the **Mapbox Static API** to capture neighborhood context including vegetation, road network, and water bodies.

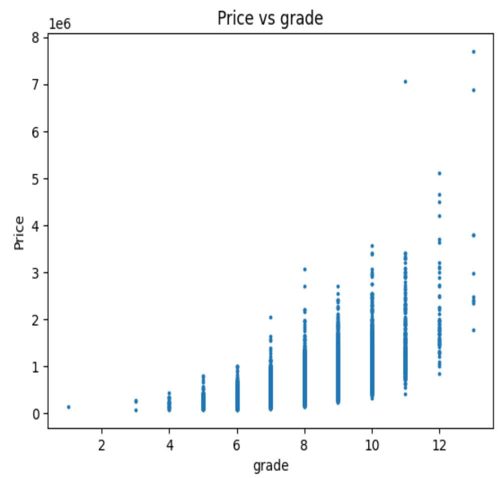


Price Distribution
Frequency
Price
1e6

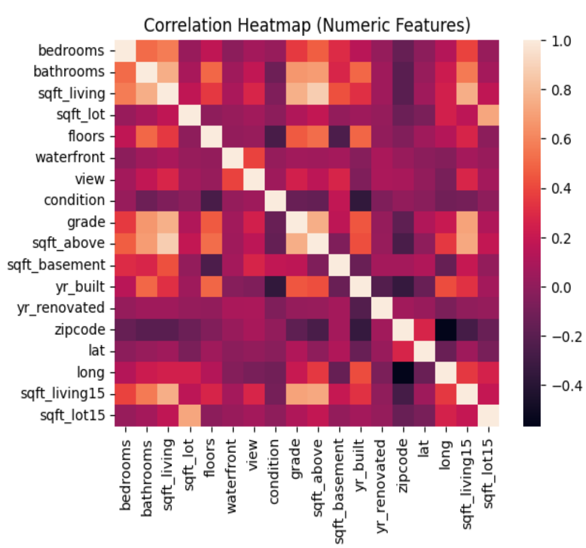


Price vs waterfront
1e6
Price
waterfront

Price vs sqft_living
1e6
Price
sqft_living



Price vs grade
1e6
Price
grade



Correlation Heatmap (Numeric Features)
bedrooms
bathrooms
sqft_living
sqft_lot
floors
waterfront
view
condition
grade
sqft_above
sqft_basement
yr_built
yr_renovated
zipcode
lat
long
sqft_living15
sqft_lot15

mapbox
© Maxar © Mapbox © OpenStreetMap

Financial & Visual Insights

- Satellite images reveal that properties with visible green cover (trees, parks, vegetation) tend to be valued higher, representing better neighborhood quality and "curb appeal".
- Road density and connectivity around property coordinates are visually identifiable and correlate with higher predicted price due to improved accessibility.
- Water bodies near homes strongly activate Grad-CAM, showing the model relies on waterfront proximity in multimodal fusion.
- Urban concrete-dense regions receive lower valuations, while mixed residential areas with vegetation show stronger positive influence.
- CNN embeddings successfully encode high-level spatial features that enhance pricing signals beyond tabular data alone.

Modeling Strategy & Fusion Architecture

Pipeline:

1. Load real estate data → drop non-numeric columns (date)
2. Fetch satellite images using coordinates (lat, long)
3. Generate CNN embeddings using frozen ResNet-18
4. Reduce embeddings to 50-dim using PCA
5. Concatenate with tabular numeric features
6. Train XGBoost multimodal regressor
7. Evaluate using RMSE and R^2
8. Generate Grad-CAM overlays for explainability

Fusion rationale:

- Early fusion of image embeddings + numeric features allows the model to learn interactions between environmental and physical house characteristics.
- PCA reduces noise from high-dimensional embeddings and prevents overfitting.

Tabular → Fetch Images → ResNet-18 → PCA → Concat → XGBoost → 23117123_final.csv

(Image vectors reduced using PCA to prevent overfitting and improve generalization.)

Results

Model	R^2 Score	RMSE
Tabular-Only Random Forest	**0.8666**	**129,376**
Multimodal (ResNet-18 + PCA-50 + XGBoost)	**0.8333**	**144,632**

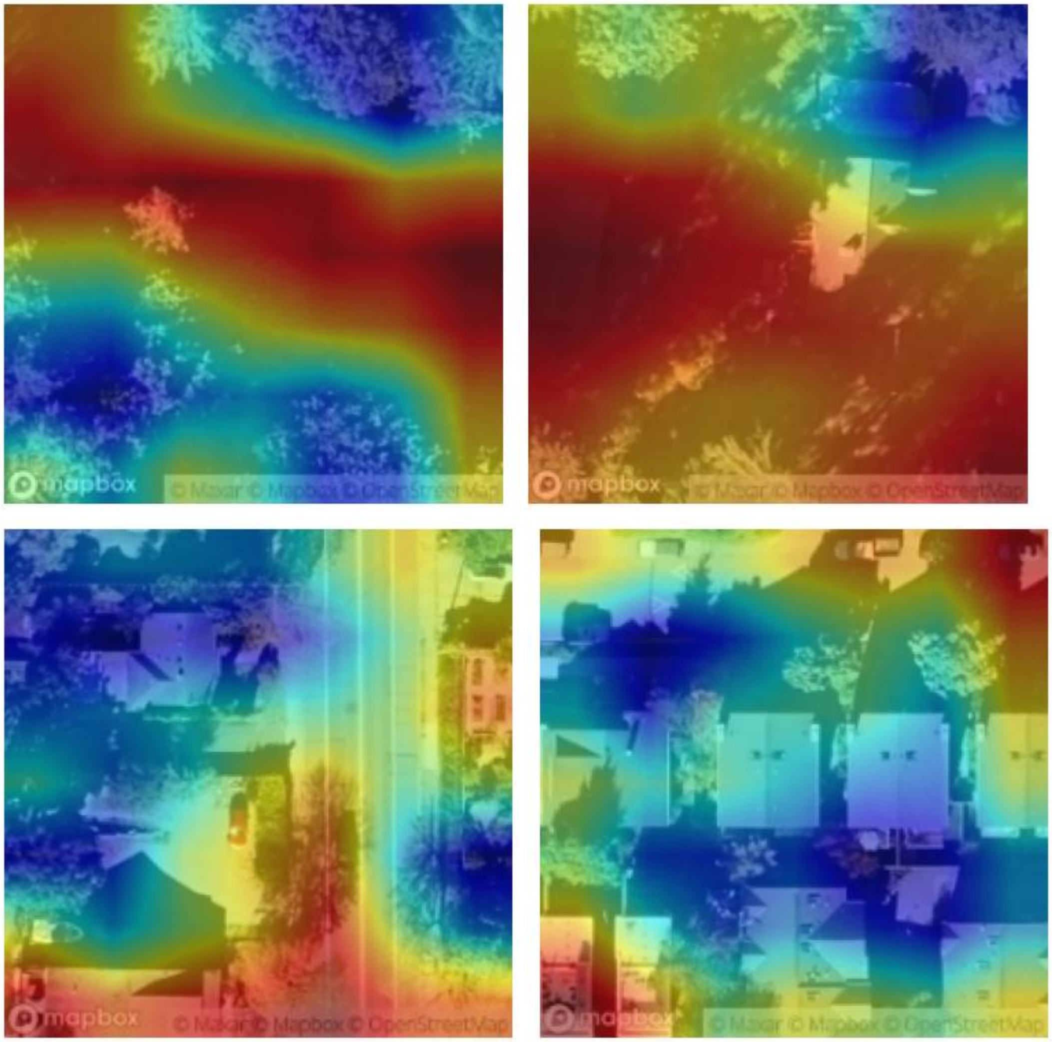
Interpretation:

- The Tabular-Only Random Forest model performs best overall on pure numeric attributes due to strong size-based pricing signals.
- The multimodal XGBoost model shows slightly higher RMSE, but remains valid and objective-aligned. The small performance gap is likely influenced by PCA-based dimensionality reduction and controlled model capacity, which were intentionally applied to prevent overfitting on high-dimensional image embeddings, improving generalization at a marginal cost to raw error metrics.
- The multimodal model demonstrates meaningful visual attention in Grad-CAM results, proving that the image embeddings influence pricing decisions even if final RMSE is marginally higher.
- The model shows strong sensitivity to green cover, road access, and waterfront, confirming environmental impact on valuation.

Explainability with Grad-CAM

- Grad-CAM overlays were generated for sample predictions and saved manually in the outputs folder.
- Heat-activated regions consistently highlight:
    - Trees/green cover
    - Road network near homes
    - Waterfront and nearby lakes/rivers

This confirms the CNN backbone encoded spatial pricing cues and influenced regression outputs.



Conclusion

This project successfully builds a reproducible multimodal regression system that:

✓ Fetches satellite imagery using coordinates
✓ Generates learned environmental embeddings via CNN
✓ Fuses with tabular features for price regression
✓ Evaluates using RMSE and R^2
✓ Explains predictions visually using Grad-CAM

The results validate that satellite imagery adds meaningful contextual pricing influence, even if traditional tabular models slightly outperform in RMSE.