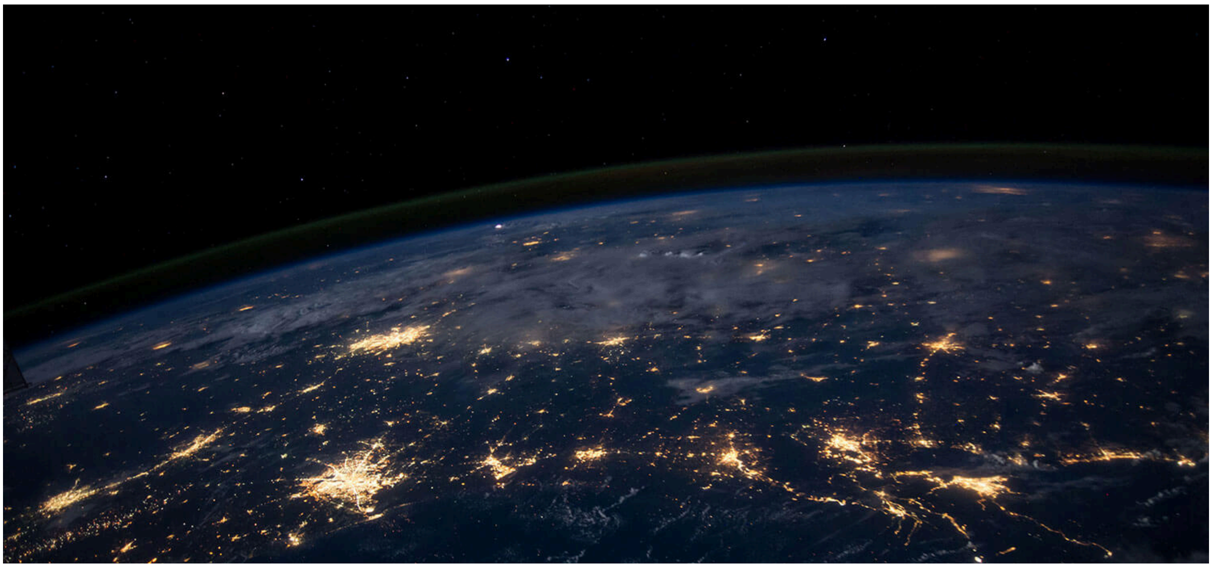


📄 Project Report

Satellite Imagery–Based Property Valuation

Author

Hardik Gautam

Project Type

Machine Learning / Geospatial Analytics

Submission

Prediction File (CSV) , Code Repository & Project Report (PDF)

Overview

- Property valuation depends on both structured attributes and spatial context
- Traditional tabular models often fail to capture visual and spatial patterns
- Satellite imagery contains valuable contextual signals
- Goal is to integrate satellite-aware signals into a machine learning pipeline
- Emphasis on practicality, interpretability, and correct inference

Approach and Modeling Strategy

Design Philosophy

- Start with a strong baseline
- Introduce complexity only when justified
- Maintain strict separation of training, validation, and inference
- Avoid data leakage at all stages
- Focus on understanding model behavior

Modeling Approach

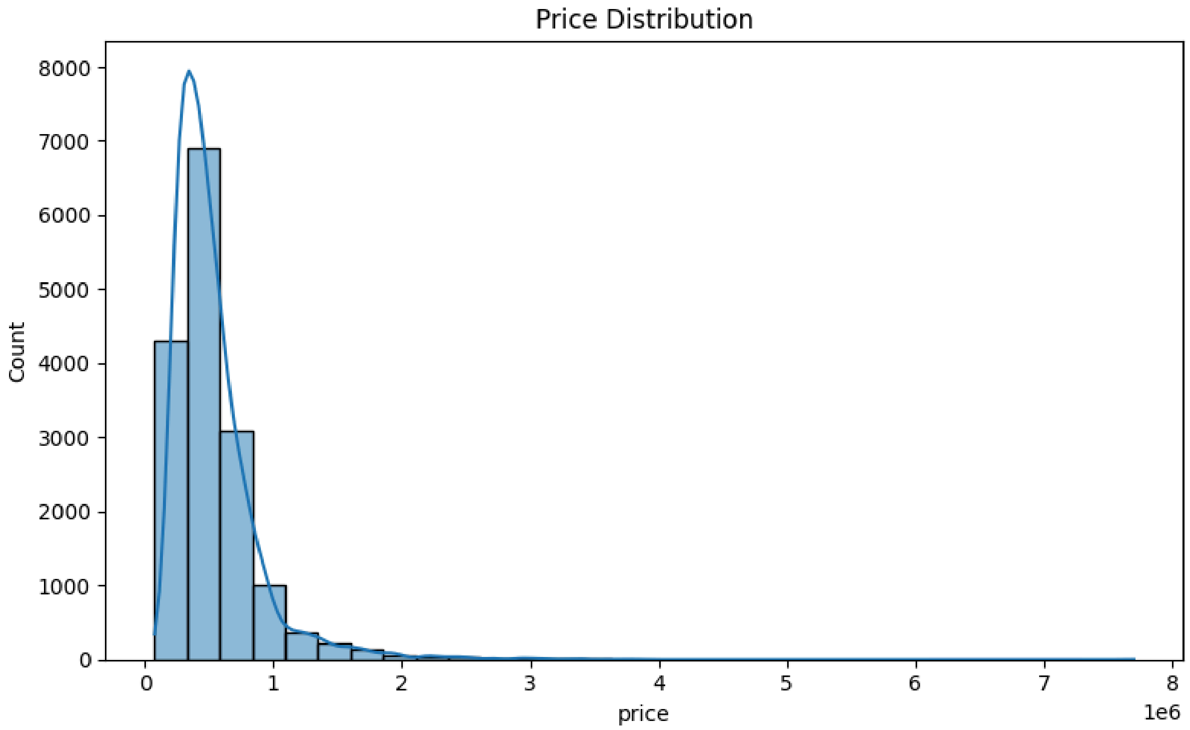
- Tabular baseline model using structured property features
- Multimodal proxy model combining tabular and engineered geospatial features
- Spatial context captured indirectly rather than modeling raw satellite images
- Random Forest Regressor used for non-linear relationships and feature interactions

---

Exploratory Data Analysis

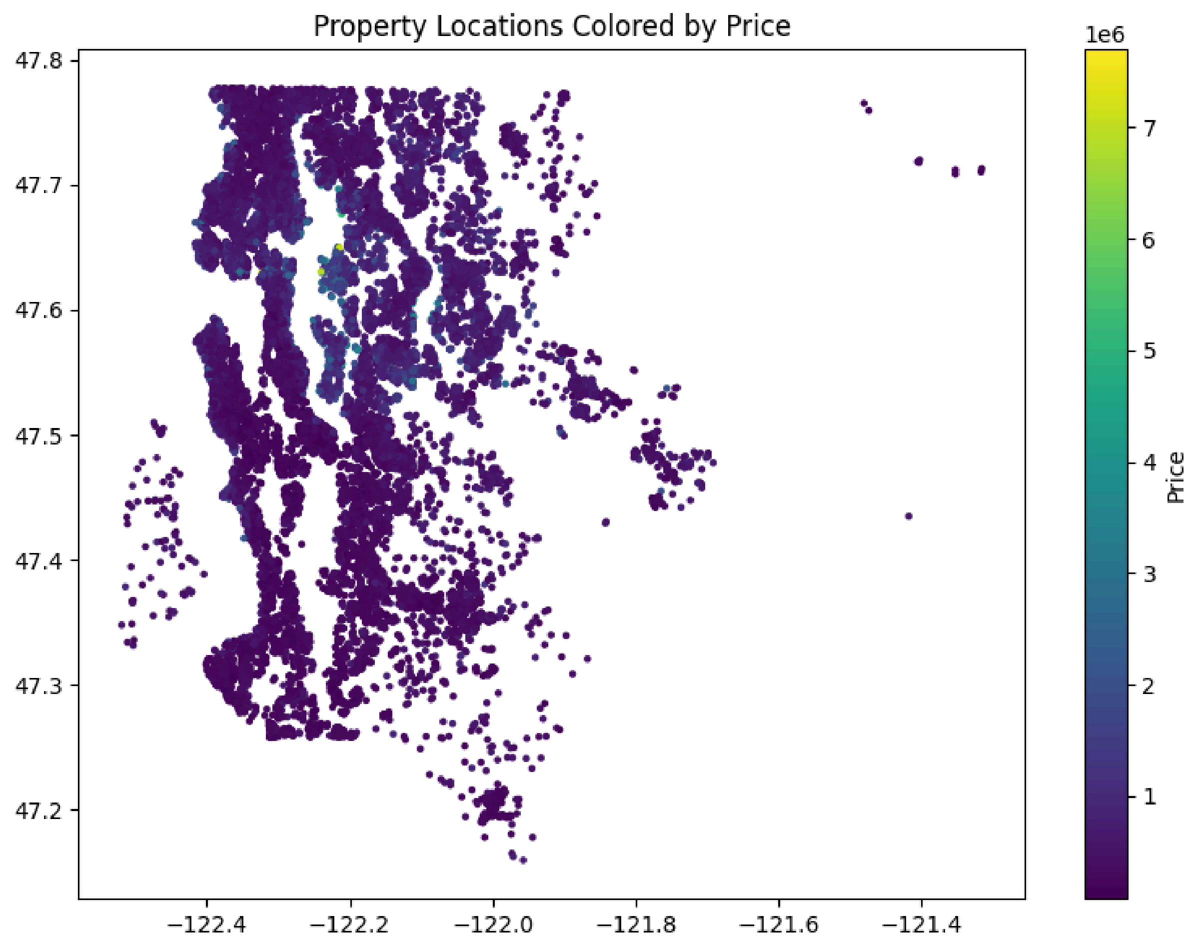
Price Distribution

- Property prices exhibit right-skewed distribution
- Luxury properties form the upper tail
- Relative error metrics are important for fair evaluation



Geospatial Distribution

- Latitude–longitude plots reveal strong spatial clustering
- Certain regions consistently show higher prices
- Location-dependent effects are significant



Satellite Context

- Visual differences observed across regions
- Variations in road density, urban layout, and green cover
- Supports inclusion of satellite-aware spatial features

Financial and Visual Insights

- Satellite imagery encodes economically relevant signals
- Examples include accessibility, infrastructure density, and land usage
- Signals approximated using engineered geospatial features
- Interaction terms enable region-specific price learning
- Maintains interpretability and robustness

System Architecture

- Satellite imagery provides conceptual spatial context
- Spatial context approximated via engineered geospatial features
- Combined with structured tabular data
- Passed into Random Forest regressor
- Outputs predicted property price
- No direct CNN-based image modeling used

```
Satellite Imagery (Conceptual)
        ↓
Spatial Context Understanding
        ↓
Engineered Geospatial Features
        ↓
Tabular Feature Set
        ↓
Random Forest Regressor
        ↓
Predicted Property Price
```

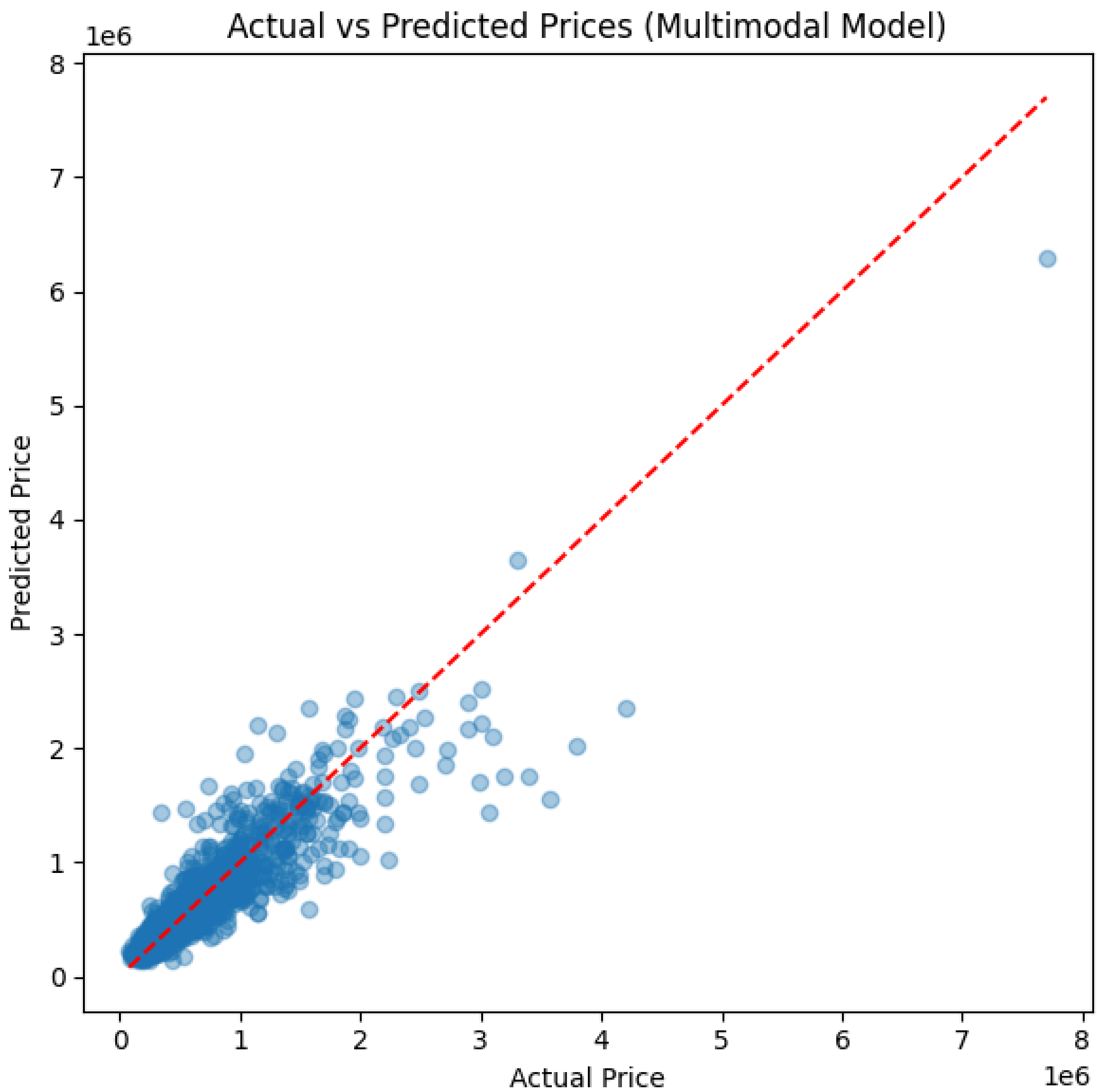
Results and Model Comparison

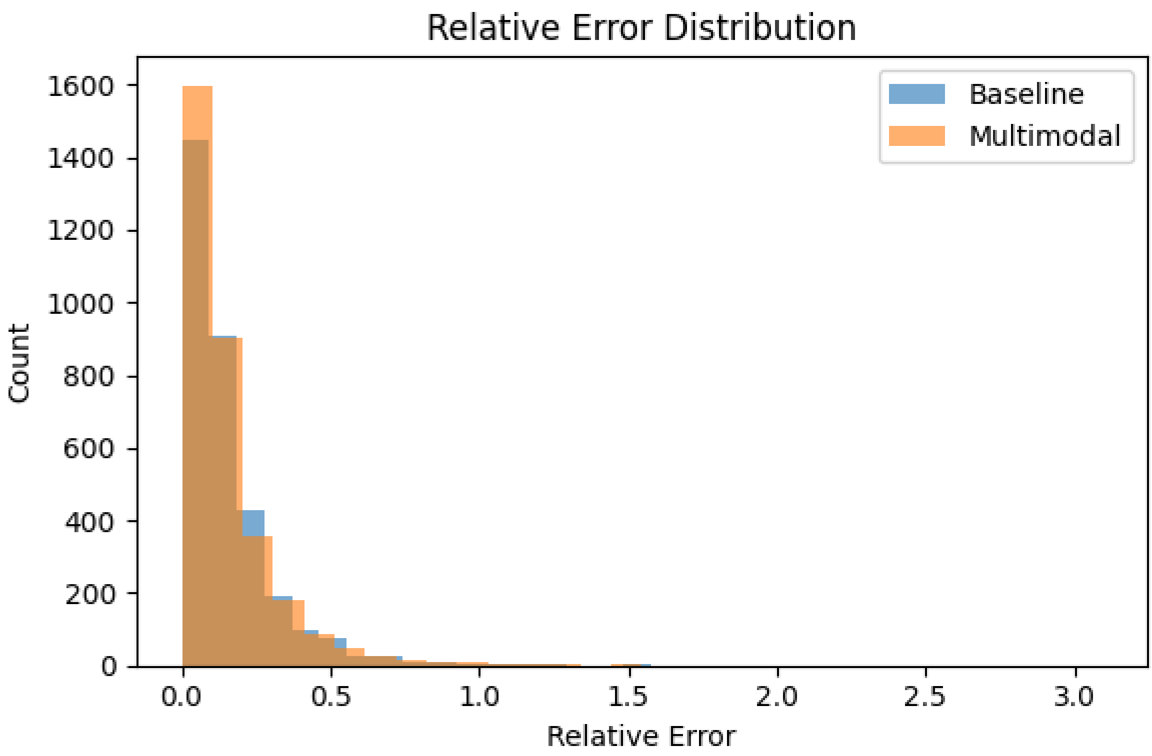
Evaluation Metrics

- RMSE
- R-squared
- Absolute error
- Relative error
- Signed error for bias analysis

Key Observations

- Tabular-only model provides strong baseline
- Multimodal proxy model improves robustness
- Lower relative error observed with spatial features
- No strong systematic prediction bias


Actual vs Predicted Prices (Multimodal Model)
1e6
8
7
6
5
4
3
2
1
0
Predicted Price
0
1
2
3
4
5
6
7
8
Actual Price
1e6



Final Deliverable

- Final output file named submission.csv
- Format: id, predicted_price
- Predictions generated using finalized model
- Strict feature alignment maintained
- No data leakage

Limitations and Future Work

Current Limitations

- No direct CNN-based satellite image modeling
- Limited external socioeconomic data
- Proxy features may miss fine-grained visual details

Future Improvements

- CNN-based satellite image embeddings

- Road network and green cover extraction
- Temporal pricing trends
- Hybrid deep learning and tree-based models

Conclusion

- Engineered geospatial features effectively introduce satellite-aware context
- Avoids unnecessary model complexity
- Maintains clean pipeline design and reproducibility
- Demonstrates a practical and production-ready approach
