Satellite Imagery–Based Property Valuation Using Multimodal Regression

---

Name: Jay Jain
Enrolment Number: 23112045
Year: 3rd
Branch: Chemical Engineering

1. Overview: Approach and Modeling Strategy

The objective of this project is to improve traditional real estate price prediction models by incorporating **satellite imagery** alongside structured property attributes. Conventional valuation models rely heavily on tabular data such as square footage, number of rooms, construction quality, and location coordinates. While these features are strong predictors of price, they do not explicitly represent **environmental and neighborhood context**, such as surrounding green cover, road connectivity, or urban density.

To address this limitation, this project proposes a **multimodal regression framework** that combines:

- **Tabular property features**, and
- **Satellite images** obtained using latitude and longitude information.

The modeling strategy follows a **comparative approach**:

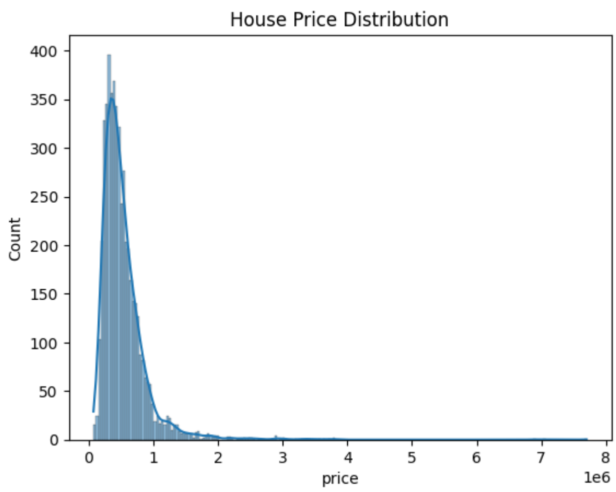
1. First, a **strong tabular-only baseline** is established using gradient boosting models to capture the maximum predictive power from structured data.
2. Next, a **multimodal deep learning model** is developed that fuses satellite image embeddings (extracted using a CNN) with tabular feature embeddings.
3. The two approaches are evaluated and compared to understand:
    - Whether satellite imagery improves prediction accuracy.
    - How satellite imagery contributes to **model interpretability**, even if accuracy gains are limited.

This approach ensures both **performance benchmarking** and **insight generation**, which are critical in real-world valuation systems.

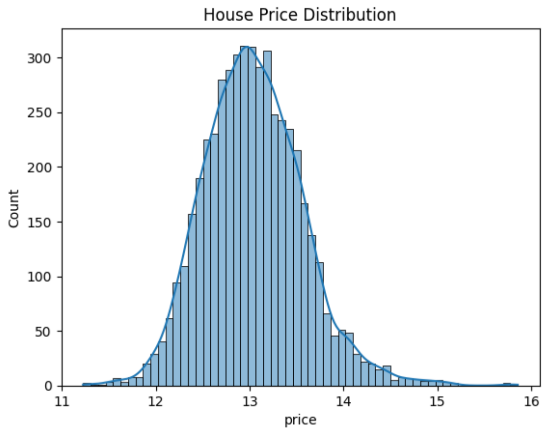
---

2. Exploratory Data Analysis (EDA)

2.1 Price Distribution



Exploratory analysis of the target variable (price) reveals a **strong right-skewed distribution**, with a small number of very high-priced properties. Such skewness can negatively impact regression models by increasing variance and biasing predictions toward extreme values.



To address this, a **logarithmic transformation** is applied:

y = log({price} + 1)

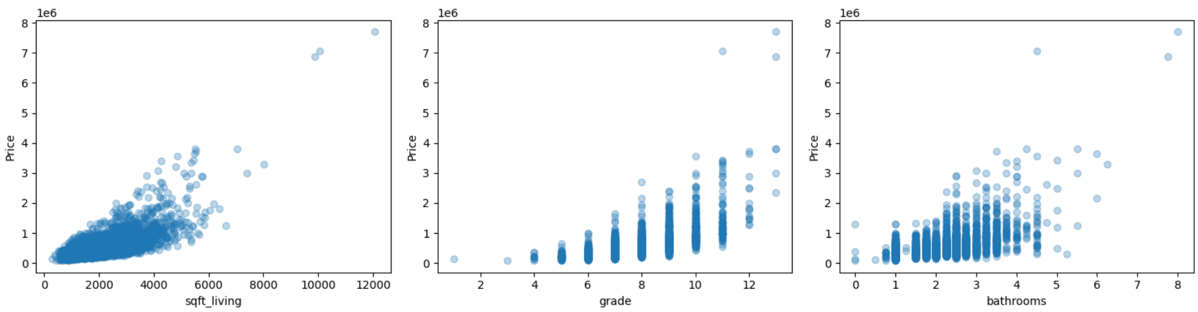
The log-transformed price distribution is significantly more symmetric, improving numerical stability and enabling more effective learning for both tabular and multimodal models.

2.2 Feature–Price Relationships

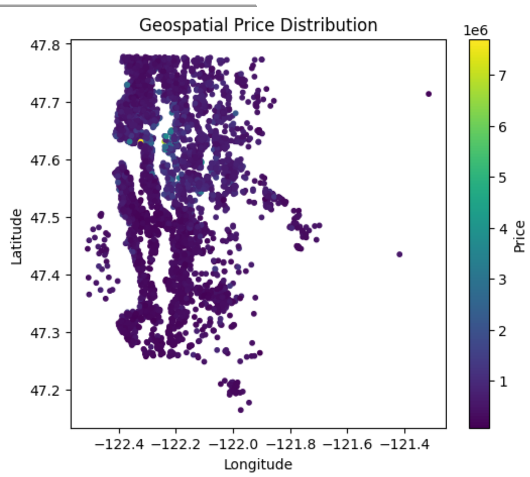
Analysis of individual features against price shows strong positive relationships with:

- **Living area (sqft_living)**
- **Construction quality (grade)**
- **Number of bathrooms**
- **Average neighborhood living area (sqft_living15)**

These relationships confirm that structural and neighborhood-level features are primary drivers of property value.



2.3 Spatial Analysis



A spatial visualization of latitude and longitude colored by price reveals **clear geographic clustering** of high-value properties. Certain regions consistently exhibit higher prices, highlighting the importance of location and neighborhood context.

This spatial clustering motivates the inclusion of satellite imagery, as images can capture **visual**

characteristics of neighborhoods—such as green spaces, road layouts, and development density—that are not explicitly represented in tabular form.

2.4 Sample Satellite Images

Sample satellite images show significant visual variation across properties:

- Suburban regions with visible greenery and open layouts.
- Dense urban areas dominated by roads and concrete structures.
- Mixed residential zones combining built-up areas and vegetation.

These variations suggest that satellite imagery may encode meaningful contextual information relevant to property valuation.



---

3. Financial and Visual Insights

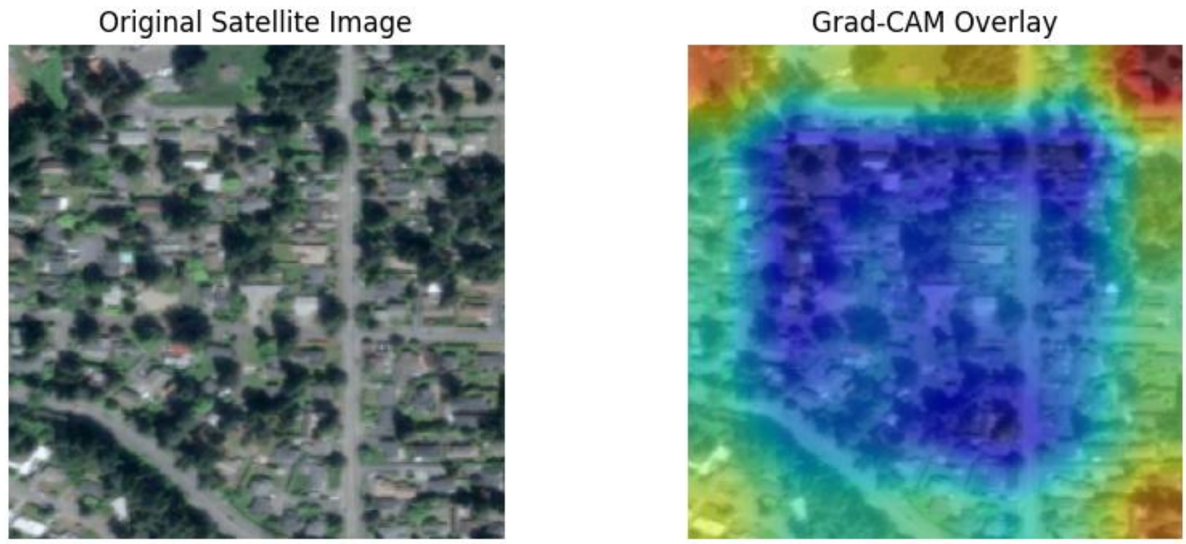
Although satellite imagery does not significantly outperform tabular models in raw predictive accuracy, it provides **valuable financial and visual insights** when analyzed using explainability techniques.

Using **Grad-CAM (Gradient-weighted Class Activation Mapping)** on the CNN image branch, the following patterns are observed:

- **Green Cover (Trees and Open Spaces):**
  Properties surrounded by visible vegetation and open spaces tend to receive higher predicted valuations. This aligns with known economic principles linking greenery to improved livability, air quality, and residential desirability.
- **Road Connectivity and Infrastructure:**
  In dense urban areas, the model focuses on road networks and infrastructure. Well-connected regions are associated with better accessibility and higher economic activity, positively influencing property prices.

- **Urban Density and Layout:**
  High-value neighborhoods often show organized layouts, wider roads, and consistent development patterns. Conversely, irregular layouts and highly congested areas correspond to lower predicted values.

These insights demonstrate that satellite imagery contributes **contextual understanding** rather than purely numerical gains, making the multimodal model useful for interpretability and qualitative analysis.



---

4. Architecture Diagram (Model Design)

The multimodal architecture integrates image and tabular data through a late-fusion strategy.

Architecture Description:

1. **Image Branch (CNN):**
   - Satellite images are passed through a pretrained ResNet-18.
   - The CNN extracts high-level visual embeddings representing neighborhood context.
   - CNN weights are frozen to prevent overfitting due to limited image data.
2. **Tabular Branch (MLP):**
   - Structured features are processed using a multilayer perceptron.
   - Features include structural, neighborhood, and geospatial attributes.
3. **Fusion Layer:**
   - Image embeddings and tabular embeddings are concatenated.

    o The combined representation is passed to a regression head.

4. **Output:**
    o The model predicts log-transformed property prices.

```
Satellite Image ──▶ CNN ──▶ Image Embedding      ┐
                                                 ├──▶ Fusion ──▶ Regressor ──▶ Price
Tabular Features ──▶ MLP ──▶ Tabular Embedding ┘
```

This design allows the model to combine numerical property attributes with visual neighborhood context.

---

5. Results and Model Comparison

All models are evaluated in **real price space** by converting predictions back from log-scale. This ensures that RMSE values are interpretable in currency units.

5.1 Performance Comparison

Model	Data Used	R^2	RMSE (Price)
Tabular Gradient Boosting	Tabular Only	~0.89	~132,000
Multimodal CNN + MLP	Tabular + Satellite Images	~0.63	Higher

5.2 Interpretation of Results

- The **tabular-only model** achieves very strong performance, indicating that structured features capture most of the variance in property prices.
- The **multimodal model** does not outperform the tabular baseline in terms of RMSE or R^2.
- However, the multimodal approach provides **interpretability benefits** by visually identifying environmental and neighborhood features influencing valuation.

This outcome reflects a realistic scenario in applied machine learning, where additional data modalities enhance understanding rather than always improving numerical accuracy.

---

6. Conclusion

This project demonstrates the successful development of a **multimodal regression pipeline** for real estate valuation. Key conclusions include:

- Structured tabular features remain the dominant predictors of property prices.
- Satellite imagery provides limited marginal improvement in predictive accuracy.
- Visual explainability using Grad-CAM reveals meaningful environmental and infrastructural factors influencing valuation.
- Multimodal learning is valuable for **contextual insights and decision support**, even when accuracy gains are modest.

Overall, the project highlights the importance of combining strong baselines with interpretability-driven models in real-world financial applications.