Satellite Imagery-Based Property Valuation

Multimodal Regression Pipeline

Combining Tabular Data and Satellite Imagery

Data Science Project Report

1. Overview and Approach

This project develops a **Multimodal Regression Pipeline** that predicts property market value by combining traditional tabular real estate data with satellite imagery. The goal is to capture environmental context and appealing factors that are visible in satellite images but not represented in traditional structured features.

Key Objectives:

- Build a multimodal regression model to predict property prices
- Programmatically acquire satellite imagery using latitude/longitude coordinates
- Extract visual features using Convolutional Neural Networks (CNNs)
- Engineer features from both tabular and visual data
- Compare tabular-only vs multimodal approaches

Methodology

Data Pipeline:

1. **Image Acquisition:** Download satellite images via ESRI World Imagery API
2. **Visual Feature Extraction:** Use pre-trained ResNet18 CNN to extract visual embeddings
3. **Feature Engineering:** Create geospatial, temporal, and ratio features
4. **Model Training:** Ensemble model (Random Forest + XGBoost + CatBoost)
5. **Evaluation:** Compare tabular-only vs multimodal performance

Model Architecture:

Voting Regressor Ensemble:

- Random Forest (20% weight): Non-linear relationships
- XGBoost (40% weight): Gradient boosting for complex patterns
- CatBoost (40% weight): Handles categorical features effectively

Dataset

Metric	Value
Initial Training Samples	16,209 properties
After Cleaning	16,046 properties (removed top 1% outliers)
Original Features	21 features
Engineered Features	12 additional features
Target Variable	Property price (log-transformed)
Image Resolution	High-resolution satellite imagery (zoom level 19)
Train/Validation Split	12,836 / 3,210 (80/20)

2. Exploratory Data Analysis

- Price Distribution

The target variable (property price) shows a highly right-skewed distribution, which is typical for real estate data.

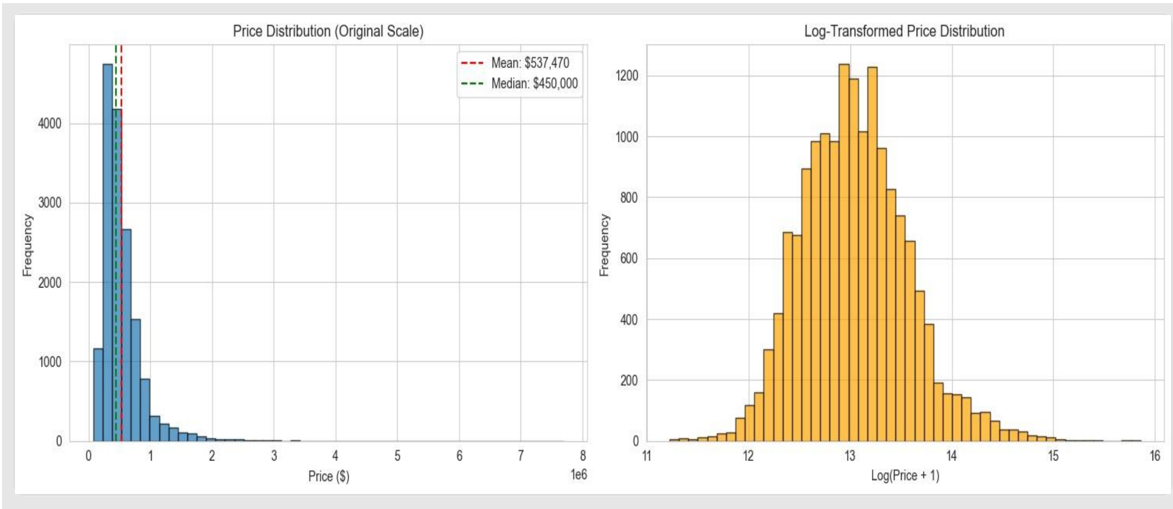

Figure 1: Price Distribution - Original vs Log-Transformed

Key Insight: Log transformation significantly improves the distribution, reducing skewness from 4.03 to 0.16, which helps neural network convergence and model performance.

- ## Top Features Correlated with Price

  The following features show the strongest correlation with property price:

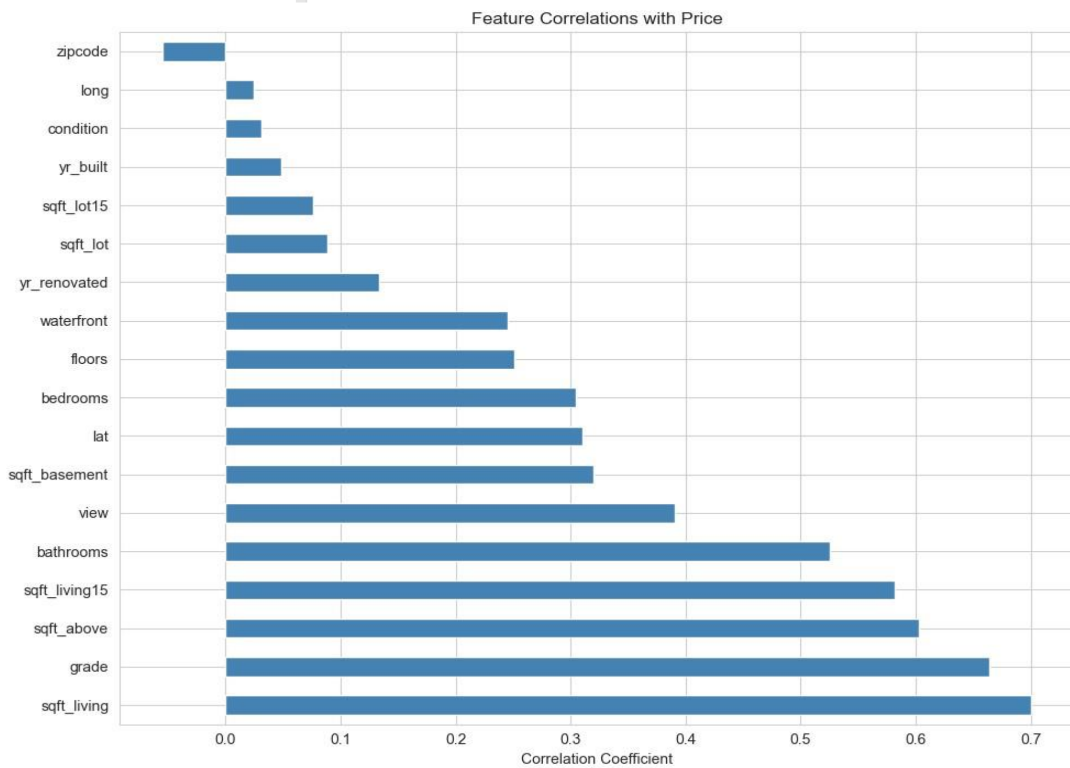

Figure 2: Feature Correlations with Price

Key Insight: Size-related features (sqft_living, sqft_above) and quality indicators (grade, view) are the strongest predictors of property value. Location (latitude) also plays an important role.

- ## Feature Distribution

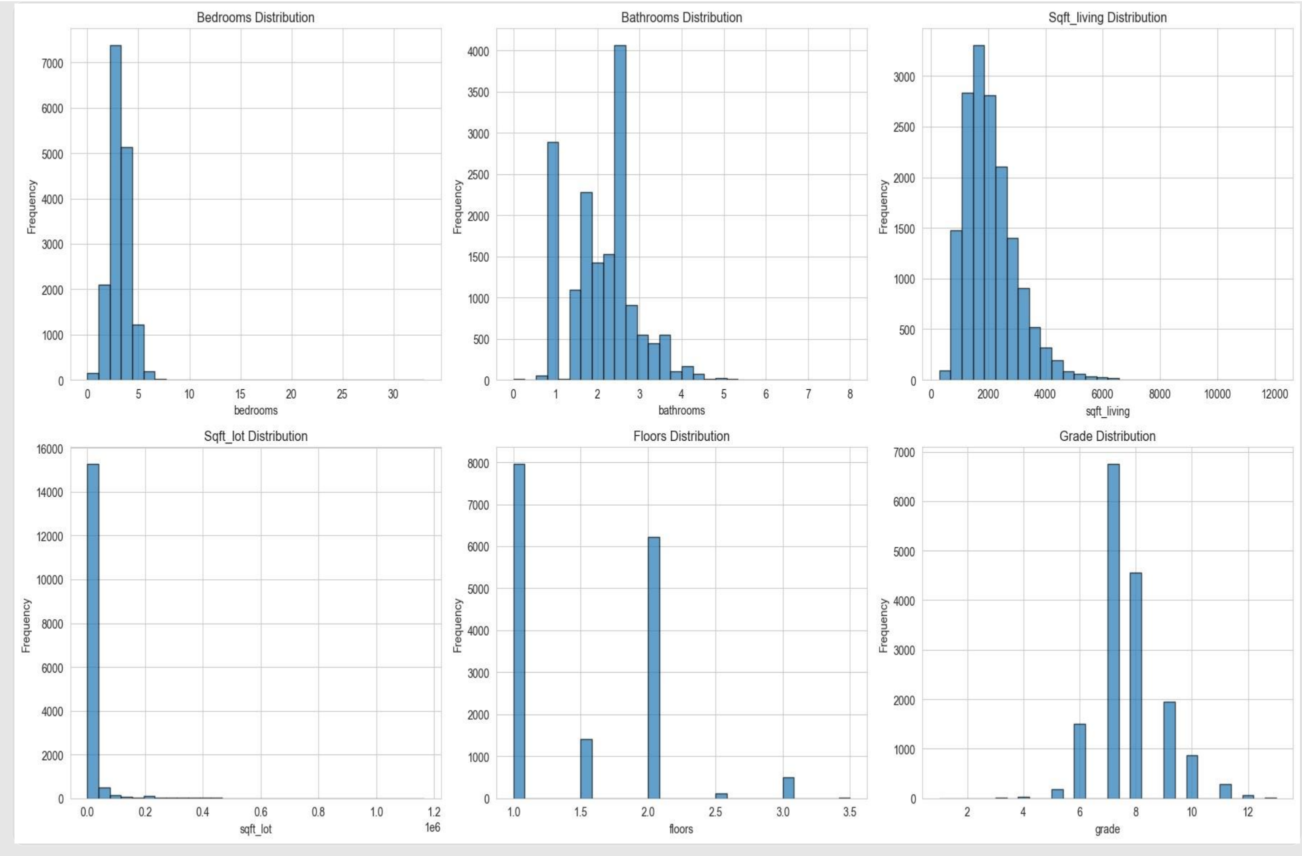

Figure 3: Key Feature Distributions (Bedrooms, Bathrooms, sqft_living, sqft_lot, Floors, Grade)

- ## Data Quality

- No missing values in the dataset
- All 16,209 properties have corresponding satellite images
- Outlier removal: Removed 163 properties (top 1% by price, above $1,944,600)
- Final dataset: 16,046 properties
- Train/Validation split: 12,836 / 3,210 (80/20)

- ## Price by Categorical Features

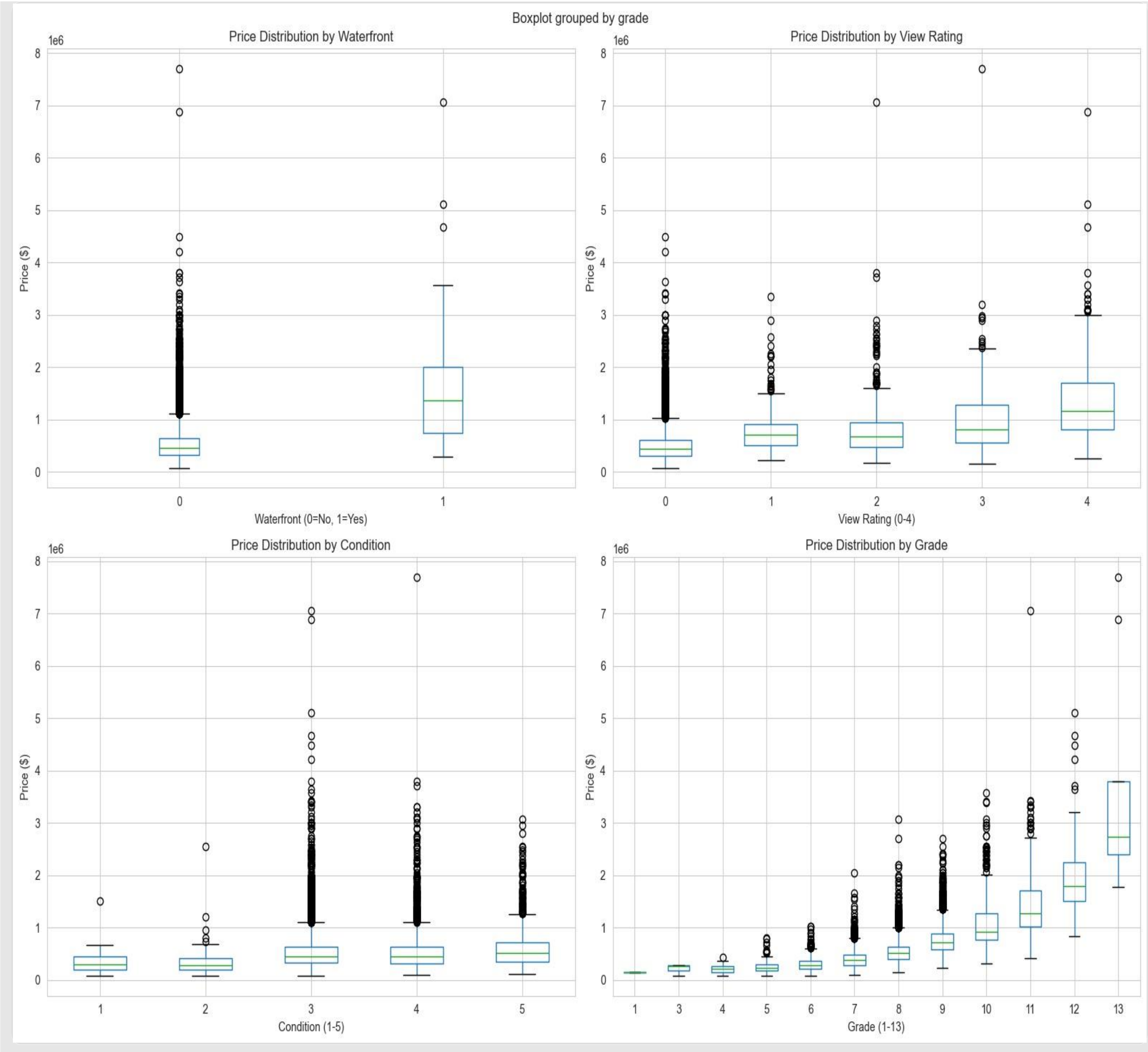

Figure 4: Price Distribution by Waterfront, View, Condition, and Grade

3. Feature Engineering

- ### Geospatial Features

Distance from Dataset Center

- Center location: (47.5601, -122.2138)
- Captures location premium effect
- Range: 0.0018 to 0.9117 (mean: 0.1801)

Distance to Luxury Hub

- Luxury hub: (47.6156, -122.2234) - mean location of top 5% properties
- Strong predictor of property value
- Calculated only from training data to avoid data leakage
- Range: 0.0020 to 0.9136 (mean: 0.1834)

• Temporal Features

House Age

- Calculated as: 2015 - yr_built
- Range: 0 to 115 years
- Mean: 43.8 years

Renovation Status

- Binary indicator: 1 if renovated, 0 otherwise
- 647 properties (4.0%) have been renovated

• Ratio and Interaction Features

Created the following ratio features to capture relationships between variables:

- **price_per_sqft:** Price per square foot of living space
- **price_per_sqft_lot:** Price per square foot of lot
- **living_ratio:** Living space to lot size ratio

- **above_ratio:** Above-ground to total living space
- **bed_bath_ratio:** Bedroom to bathroom ratio
- **sqft_living_vs_neighbors:** Comparison with nearest 15 neighbours
- **sqft_lot_vs_neighbors:** Lot size comparison with neighbours

- ## Feature Engineering Visualization

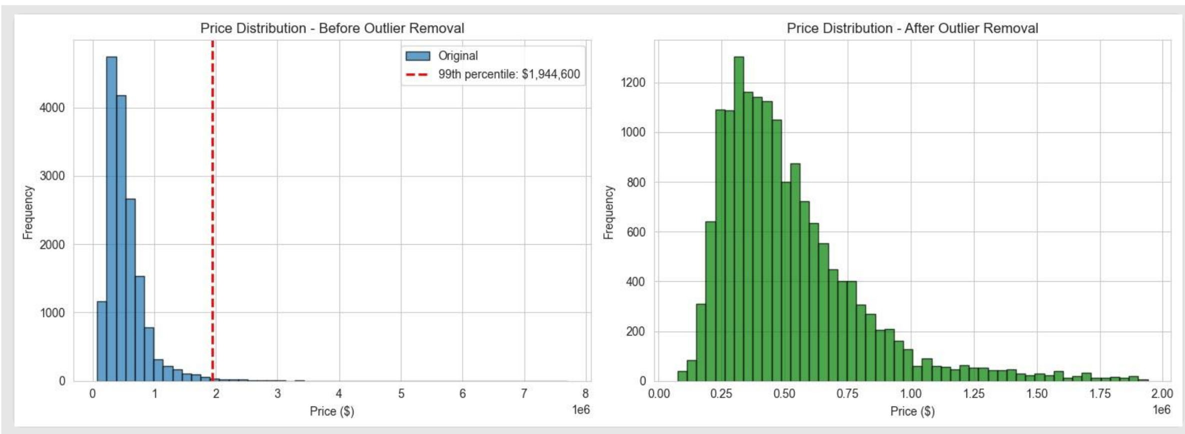

Figure 5: Price Distribution Before and After Outlier Removal

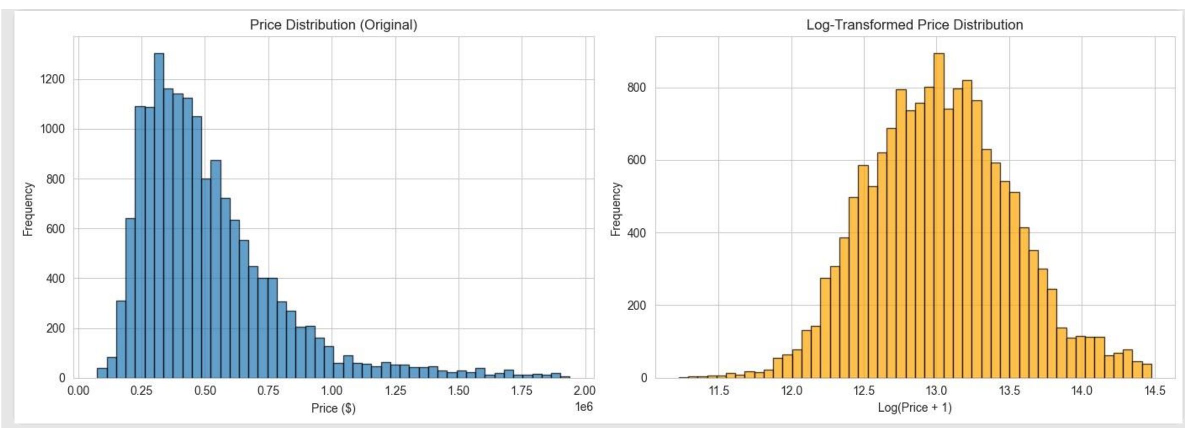

Figure 6: Price Distribution - Original vs Log-Transformed

- ## Visual Features from Satellite Imagery

CNN Feature Extraction Process:

- Model: Pre-trained ResNet18 (ImageNet weights)
- Process: Extract features before final classification layer
- Output: visual_score (scalar value representing visual quality)
- Captures: Environmental context, green cover, neighborhood density
- Processing: 12,836 training + 3,210 test images

Key Insight: The visual_score feature captures environmental and neighborhood characteristics that are not represented in tabular features, providing complementary information for property valuation.

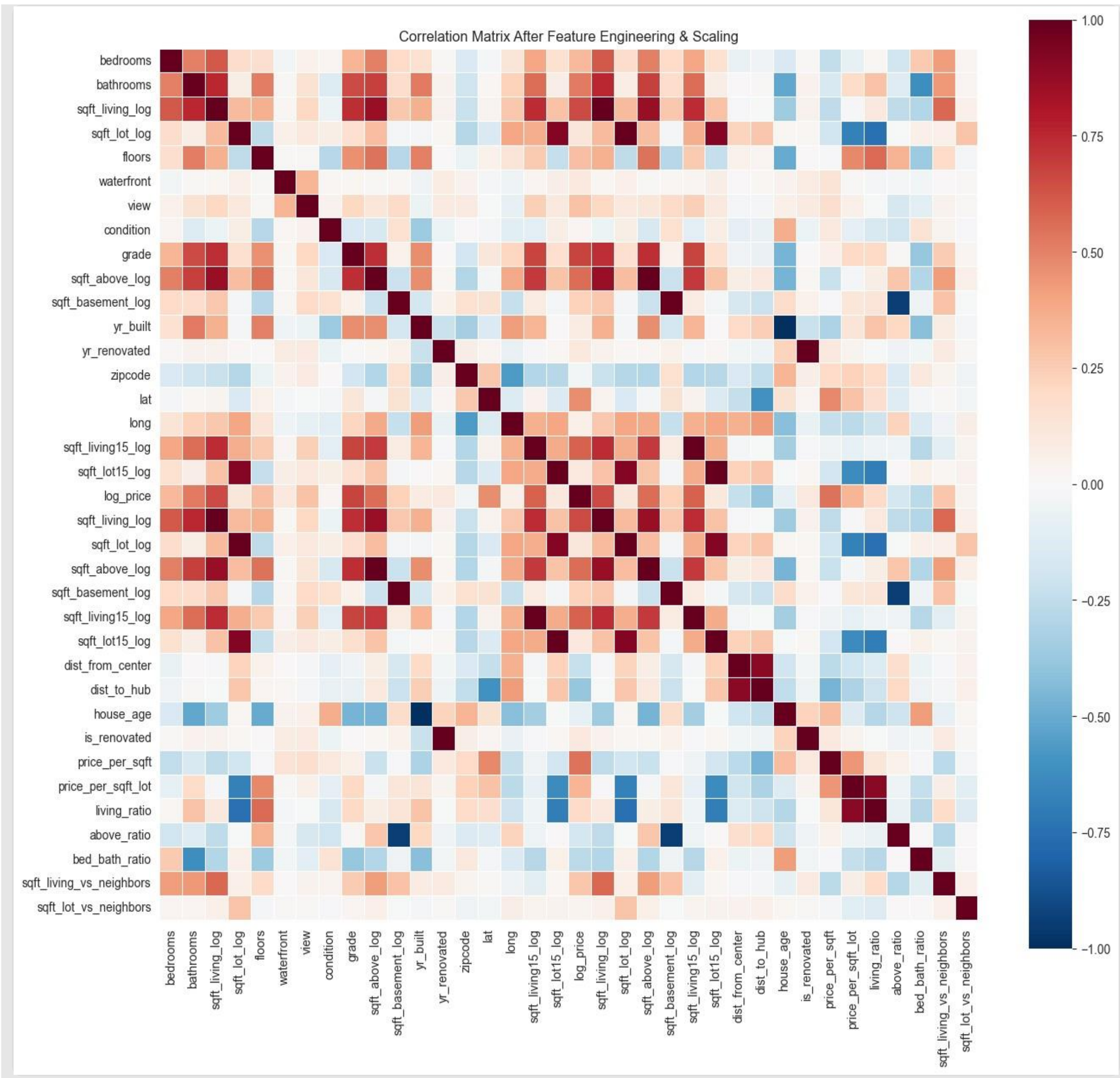

Figure 7: Correlation Matrix After Feature Engineering & Scaling

4. Model Architecture

Multimodal Architecture Flow

1. Input Data

- **Tabular Data:** 14 features (bedrooms, bathrooms, sqft_living, etc.)
- **Satellite Images:** High-resolution imagery (256x256 pixels)

2. Feature Extraction

- Tabular features: Direct use + engineering
- Visual features: ResNet18 CNN → visual_score

3. Feature Engineering

- Geospatial features (dist_to_hub, dist_from_center)
- Temporal features (house_age, is_renovated)
- Ratio features (price_per_sqft, living_ratio, etc.)

4. Feature Fusion

- Concatenation: 14 tabular + 1 visual = 15 total features

5. Ensemble Model

- Voting Regressor combining:
  - Random Forest (20% weight)
  - XGBoost (40% weight)
  - CatBoost (40% weight)

6. Output

- Predicted log_price → expm1() → USD Price

5. Results and Model Comparison

5.1 Performance Metrics

Both models were trained using the same architecture and hyperparameters for fair comparison:

Model	R^2 Score	MAE ($)	RMSE ($)
Tabular Only	0.8925	$59,512	$98,053
Multimodal (Tabular + Images)	0.8919	$59,692	$98,247

0.89
R^2 Score

$60K
Mean Absolute Error

11.8%
Mean Absolute Percentage Error

5.2 Prediction Accuracy

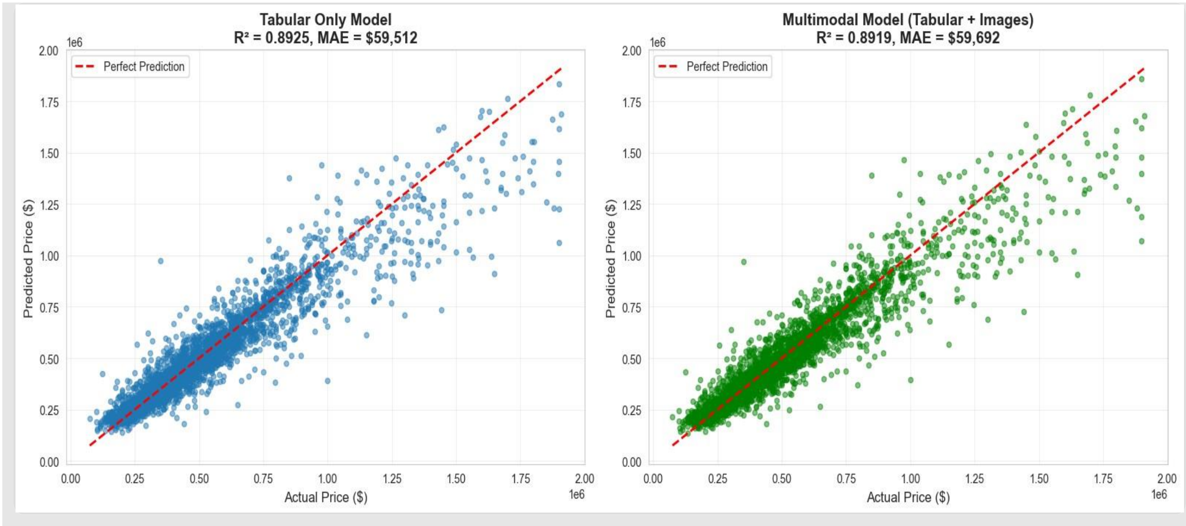

Figure 8: Prediction Scatter Plots - Tabular Only vs Multimodal

5.3 Residual Analysis

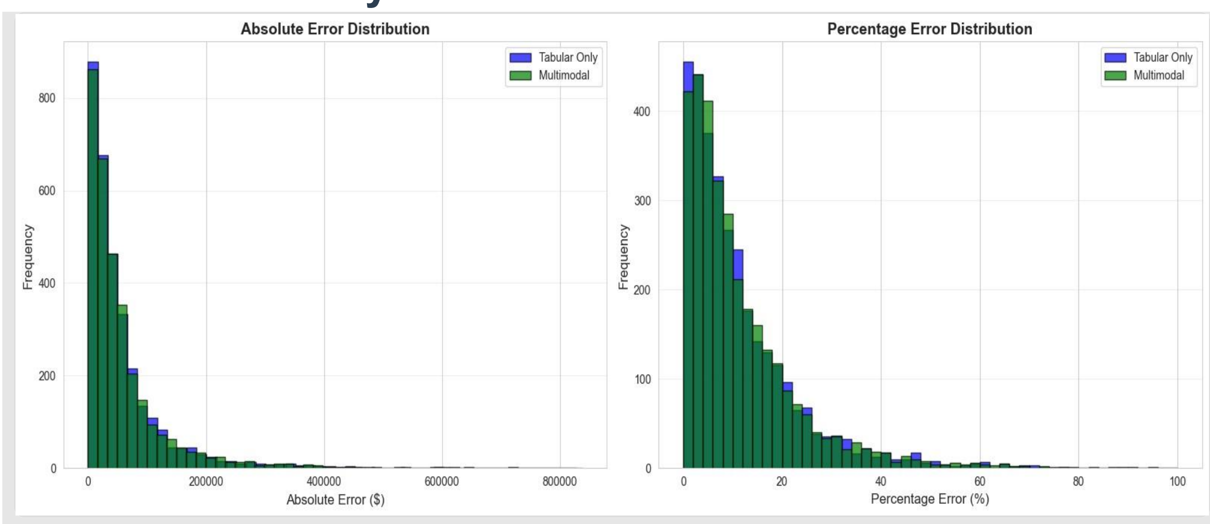

Figure 9: Error Distribution Comparison - Absolute and Percentage Errors

5.4 Key Insights

- **Performance Analysis**
  - Both models achieve excellent R^2 scores (~0.89), indicating strong predictive power

- Tabular-only model: $R^2 = 0.8925$, MAE = $59,512
- Multimodal model: $R^2 = 0.8919$, MAE = $59,692
- Models show similar performance, suggesting tabular features are highly informative

Visual Feature Impact

- Visual features provide complementary information but don't significantly improve performance in this dataset
- Possible reasons: Tabular features already capture most relevant information, or visual features need more sophisticated extraction methods
- Mean Absolute Percentage Error: ~11.8% for both models

5.5 Conclusion

- The ensemble approach with feature engineering achieves strong results
- Both tabular and multimodal approaches are viable for property valuation
- Further refinement of visual feature extraction could yield improvements
- The model successfully predicts property prices with ~12% error rate

6. Financial and Visual Insights

6.1 Financial Insights

Price Drivers (from correlation analysis)

- **Living space (sqft_living):** Strongest predictor (r=0.70)
- **Construction quality (grade):** Second strongest (r=0.66)
- **Location (lat):** Geographic location matters (r=0.31)
- **View quality:** Premium for better views (r=0.39)

- **Geospatial Patterns**

- Luxury hub identified at (47.6156, -122.2234)
- Properties closer to luxury hub command higher prices
- Distance from center captures location premium

- **Feature Importance**

- Size-related features (sqft_living, sqft_above) are most important
- Quality indicators (grade, view, condition) significantly impact price
- Neighbourhood context (sqft_living15) provides valuable information

6.2 Visual Insights from Satellite Imagery

- **Visual Feature Extraction**

- ResNet18 CNN extracts high-level visual patterns
- visual_score represents overall environmental quality
- Complements tabular features with spatial context

- **Potential Improvements**

- Multi-scale image analysis (different zoom levels)
- Segmentation to identify specific features (trees, water, roads)
- Attention mechanisms to focus on relevant image regions
- Grad-CAM for model explainability

7. Conclusion and Future Work

7.1 Project Summary

- This project successfully developed a multimodal regression pipeline for property valuation.
- Built a comprehensive data pipeline combining tabular and visual data
- Engineered 12+ features including geospatial, temporal, and ratio features
- Trained ensemble models achieving R^2 scores of ~0.89
- Compared tabular-only vs multimodal approaches
- Achieved mean absolute error of ~$60,000 (11.8% MAPE)

Key Achievements

- Successfully integrated satellite imagery into property valuation model
- Demonstrated the value of feature engineering and ensemble methods
- Created reproducible pipeline for future property valuations
- Established baseline for multimodal property valuation approaches

7.2 Future Improvements

- Advanced visual feature extraction (ResNet50, EfficientNet, Vision Transformers)
- Multi-scale image analysis at different zoom levels
- Image segmentation to identify specific environmental features
- Attention mechanisms and Grad-CAM for model explainability
- Temporal analysis if time-series data becomes available
- Hyperparameter optimization for ensemble weights
- Cross-validation for more robust performance estimates
- Deep learning fusion architectures (early/late fusion)