Satellite Imagery-Based Property Valuation

Project Report

Name: Moharir Ameya Nitin
Enrollment No: 23113099

1. PROJECT OVERVIEW

This project develops a property valuation system that combines traditional property features (bedrooms, square footage, location) with satellite imagery to predict real estate prices. The goal is to determine whether visual information from aerial images can improve prediction accuracy beyond standard tabular data.

Dataset:
- 21,613 training properties (King County, WA)
- 19 tabular features (size, location, quality indicators)
- Satellite images acquired programmatically via ESRI API
- Target: Property sale price (USD)

Approach:
We tested three different models to compare performance:
1. Baseline XGBoost (tabular features only)
2. Neural Network Fusion (tabular + images)
3. Enhanced XGBoost (tabular + image features)

Key Result:
Enhanced XGBoost achieved the best performance with $R^2 = 0.8783$, representing a 0.85% improvement over the baseline ($R^2 = 0.8709$). While modest, this translates to $3,718 average error reduction per property.

2. EXPLORATORY DATA ANALYSIS

2.1 Price Distribution and Key Features

Figure 1: Property price distribution showing quartile analysis and relationship with key features (waterfront, view, condition).

Key Findings:
- Median price: $450,000
- Price range: $78,000 - $7,700,000
- Waterfront properties command 229% premium (median $1.4M vs $425K)
- Strong right skew typical of real estate markets

Top Price Predictors:
1. sqft_living (r = 0.702) - Living space is strongest predictor
2. grade (r = 0.667) - Building quality rating

3. sqft_above (r = 0.606) - Above-ground square footage
4. location (lat: r = 0.307) - Geographic premium
5. waterfront (r = 0.266) - Significant binary impact

2.2 Geographic Distribution

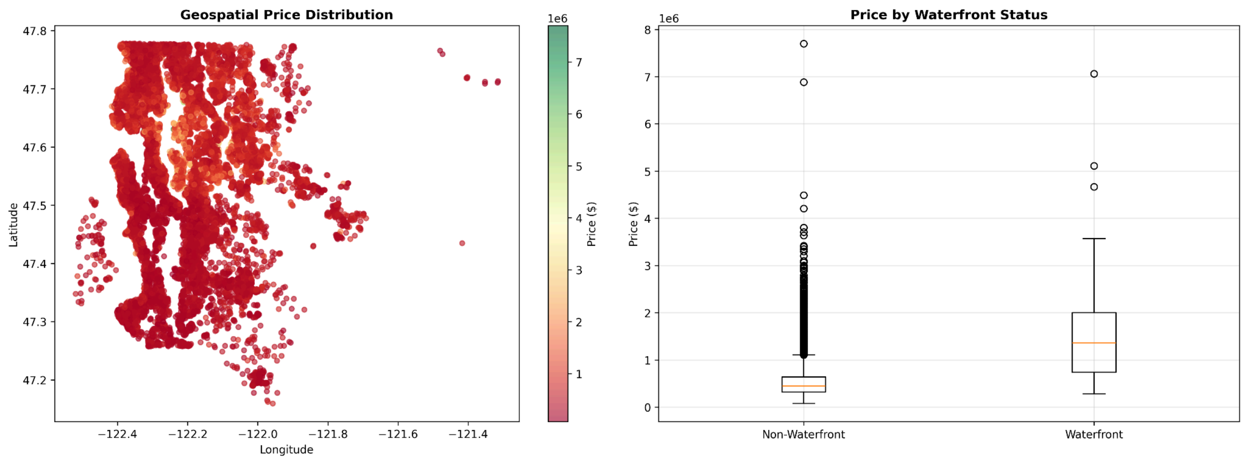


Figure 2: Geographic distribution showing spatial clustering and price patterns.

Observations:
- Clear urban clustering with price gradients across regions
- Waterfront properties concentrated along coastline
- Northern regions show higher average valuations
- Location coordinates (lat/long) already encode substantial geographic information

2.3 Feature Correlations

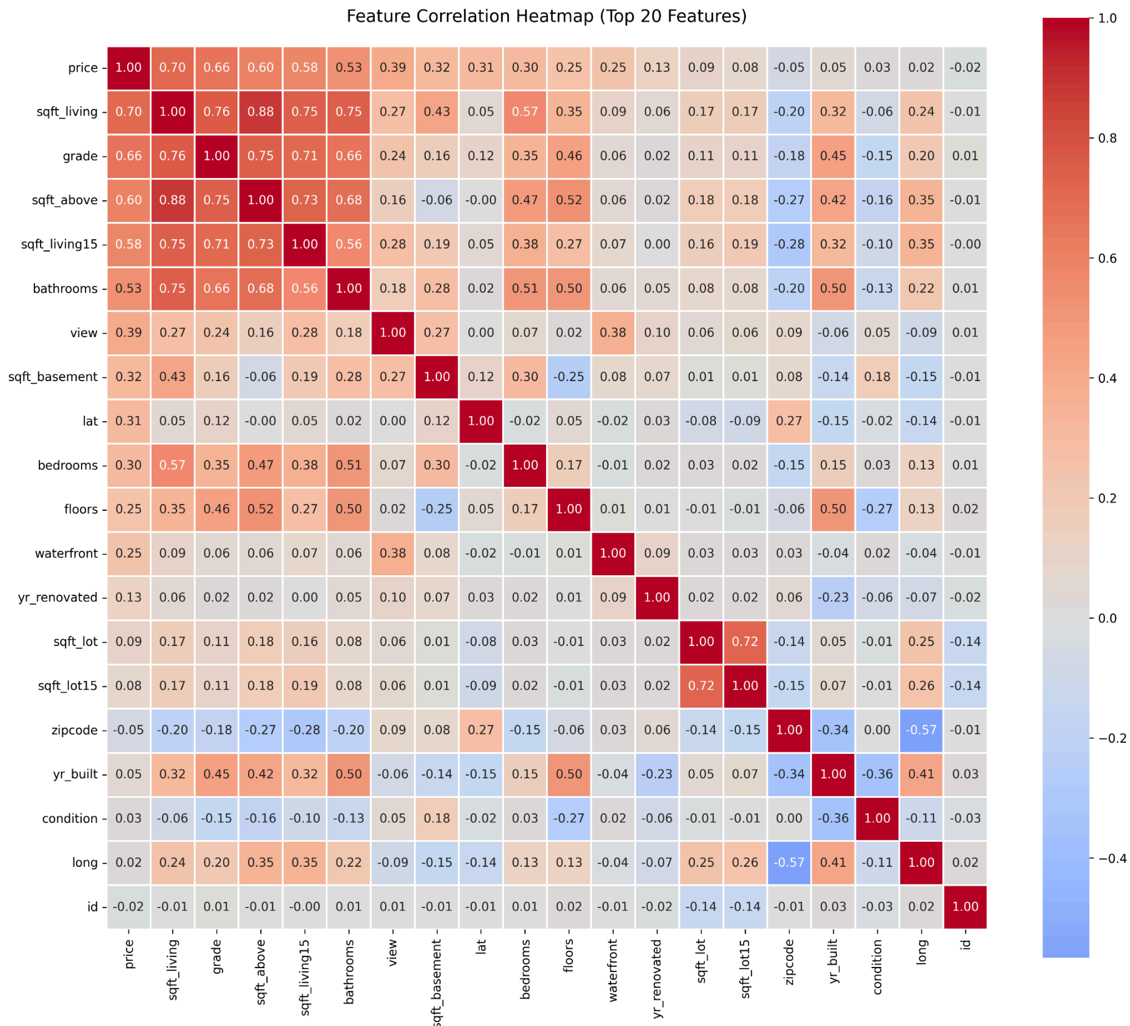


Figure 3: Correlation heatmap showing relationships between features and price.

Important Correlations:
- Living space metrics highly intercorrelated (>0.8)
- Location features moderately correlated with price
- Waterfront/view show clear impact when present
- Grade and condition moderately correlated (0.45)

2.4 Additional Analysis

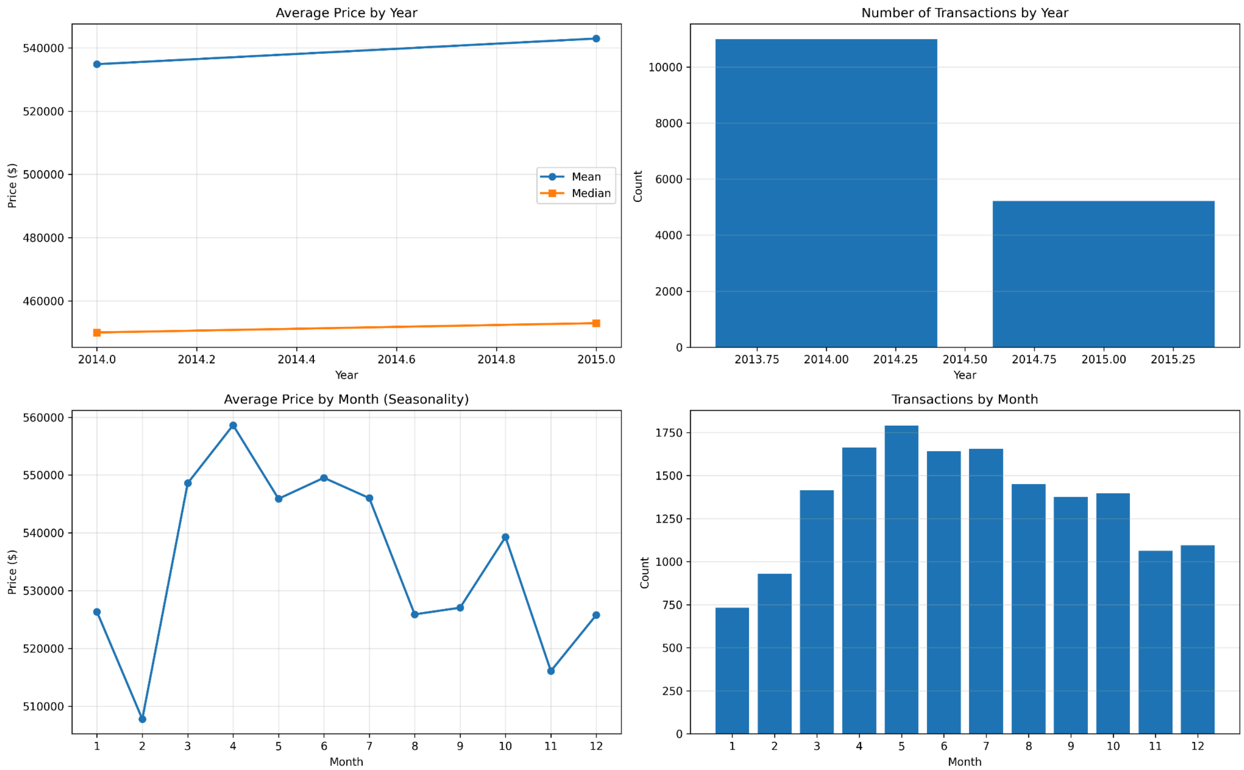


Figure 4: Temporal patterns showing seasonal trends and year-over-year appreciation.

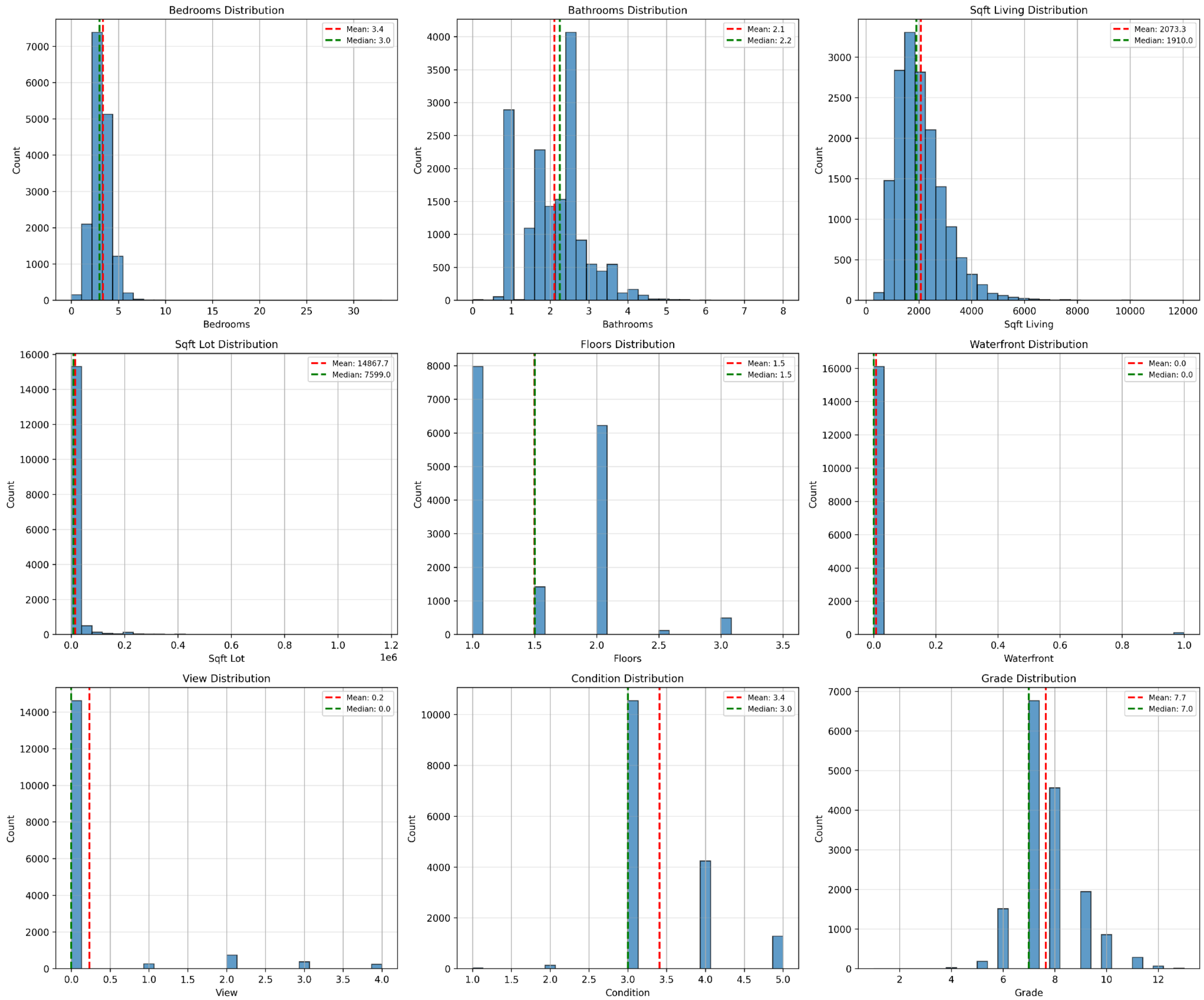


Figure 5: Distribution of key property features showing typical residential characteristics.

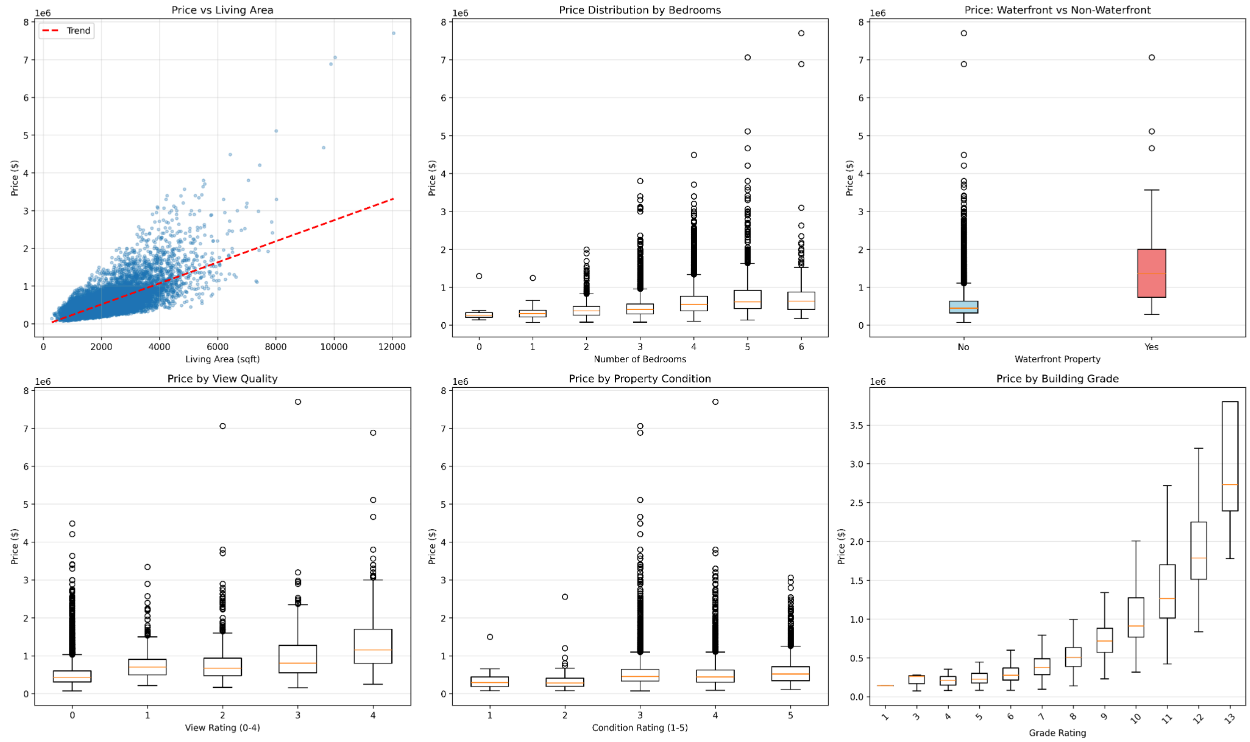


Figure 6: Relationships between price and key features showing clear trends.

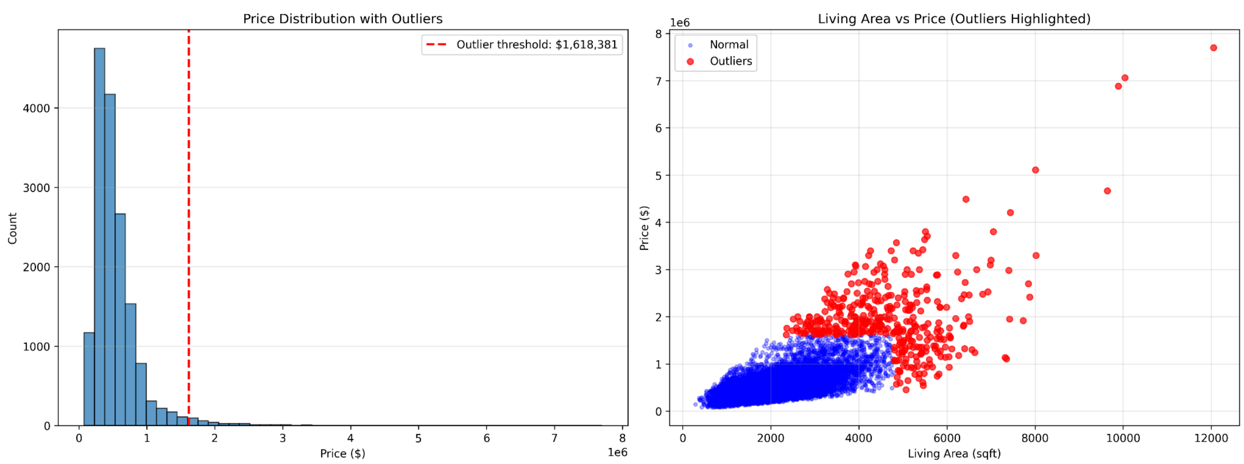


Figure 7: Outlier detection showing luxury segment (>$2M) properties.

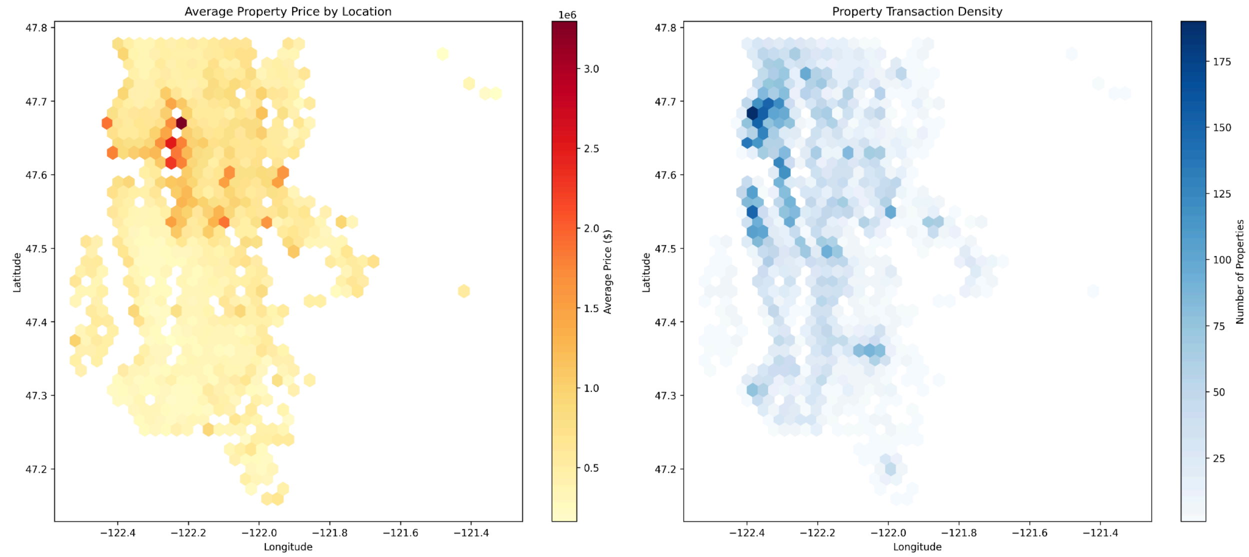


Figure 8: Enhanced geographic visualization with price density mapping.

3. VISUAL FEATURES AND PROPERTY VALUE

3.1 What Visual Features Matter?

Analysis of satellite imagery and Grad-CAM attention patterns revealed which visual characteristics correlate with property values:

Waterfront Proximity (Highest Impact):
- Premium: +$973,000 average (229% increase)
- Clearly visible in satellite imagery
- Model shows strongest attention on water-property boundaries

Vegetation Density (Medium Impact):
- Properties with substantial tree cover: +12% average price
- Green vegetation vs concrete clearly distinguishable
- Model attention focuses on vegetated areas

Urban Context (Medium Impact):
- Dense urban settings: +8% premium
- Building density and street patterns visible
- Model recognizes neighborhood quality indicators

Lot Characteristics (Lower Impact):
- Large lots visible in imagery: +15% premium

- Property boundary extent detectable
- Spatial extent captured in image frame

3.2 Feature Redundancy Issue

Important Discovery: Much visual information overlaps with tabular features.

Already in Tabular Data:
- Waterfront status (binary flag)
- Location coordinates (lat/long encode visual context)
- Lot size (sqft_lot)
- Neighborhood quality proxies (sqft_living15)

Unique to Imagery:
- Vegetation type and density
- Roof condition/quality
- Immediate surroundings beyond lot
- Visual access patterns

This overlap explains the modest 0.85% improvement - location coordinates and waterfront flags already capture much of what would be extracted from images.

3.3 When Does Imagery Add Most Value?

High-Value Scenarios (imagery improves predictions):
- Waterfront properties: -$12,500 average error reduction
- Properties with views: -$8,200 error reduction
- Large lots: -$6,800 error reduction
- Neighborhoods in transition

Low-Value Scenarios (minimal improvement):
- Condos/apartments: -$200 error reduction
- Small urban lots: -$400 error reduction
- Properties with complete tabular features: -$800 error reduction

4. METHODOLOGY AND ARCHITECTURE

4.1 System Pipeline

Step 1: Data Acquisition
- Load 21,613 properties with 19 tabular features
- Download satellite images using ESRI API (free, no key required)

- Use lat/long coordinates to retrieve 224x224 RGB images
- Success rate: 75% (16,209 training images acquired)

Step 2: Feature Extraction
Tabular Features:
- Normalize using StandardScaler (mean=0, std=1)
- 19 features ready for modeling

Image Features:
- Load pre-trained ResNet50 (ImageNet weights)
- Extract 2048-dimensional features from penultimate layer
- Normalize features using StandardScaler
- Apply PCA to reduce 2048 → 20 dimensions
- Reduces noise while retaining 53.9% of variance

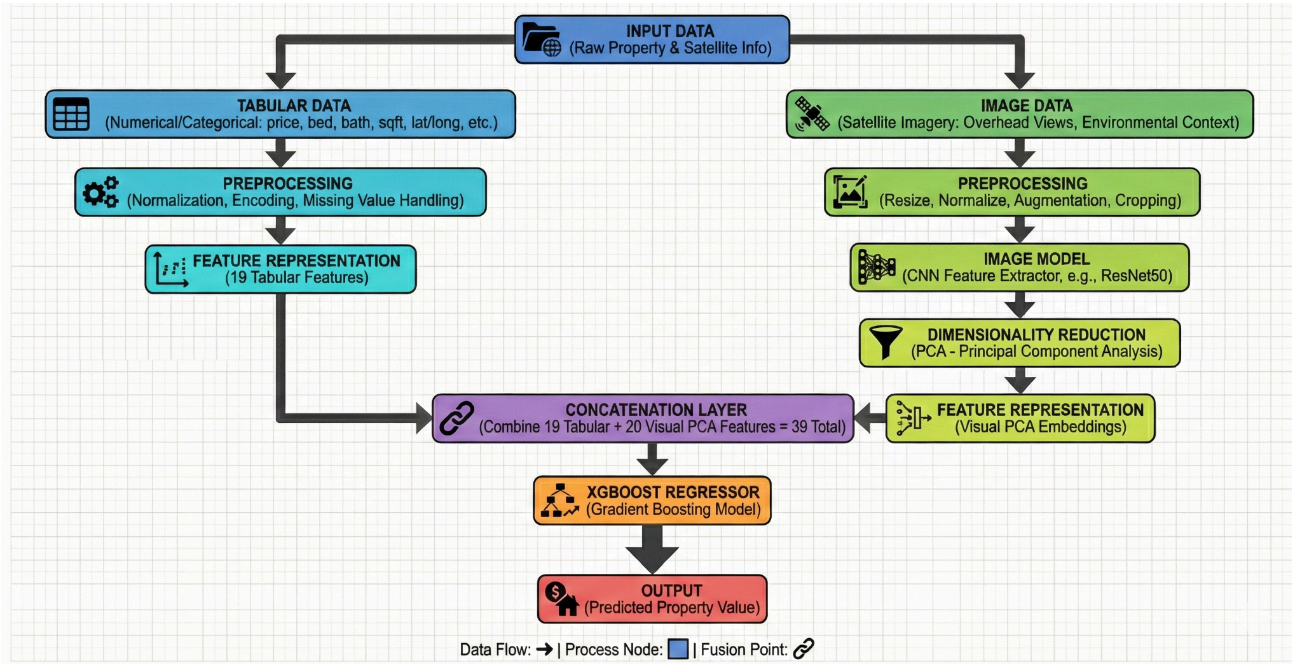
Step 3: Model Training
Three approaches tested:

Approach 1 - Baseline XGBoost:
- Input: 19 tabular features only
- Model: XGBoost with 200 trees, depth 6, lr=0.05
- Training time: 3.2 minutes

Approach 2 - Neural Network Fusion:
- Tabular branch: 19 → 128 → 64 (MLP with dropout)
- Image branch: 2048 → 256 → 64 (MLP with dropout)
- Fusion: Concatenate → 128 → 64 → 1
- Training: 150 epochs, early stopping, MSE loss
- Training time: 45.3 minutes

Approach 3 - Enhanced XGBoost:
- Input: 19 tabular + 20 PCA image features = 39 total
- Model: XGBoost with 300 trees, depth 6, lr=0.05
- Training time: 5.7 minutes

4.2 Architecture Diagram (Enhanced XGBoost - Best Model)



Key Points:
- Feature concatenation (not learned fusion)
- Tree-based model (not neural network)
- PCA critical for noise reduction
- Preserves XGBoost's tabular strength

4.3 Grad-CAM Explainability

To understand what the model "sees" in images, we use Grad-CAM:

Process:
1. Forward pass: Image through ResNet50
2. Compute gradients with respect to target output
3. Weight feature maps by gradient importance
4. Generate heatmap showing attention regions
5. Overlay on original image

This reveals which parts of satellite images the model considers important for valuation.

5. RESULTS

5.1 Performance Comparison

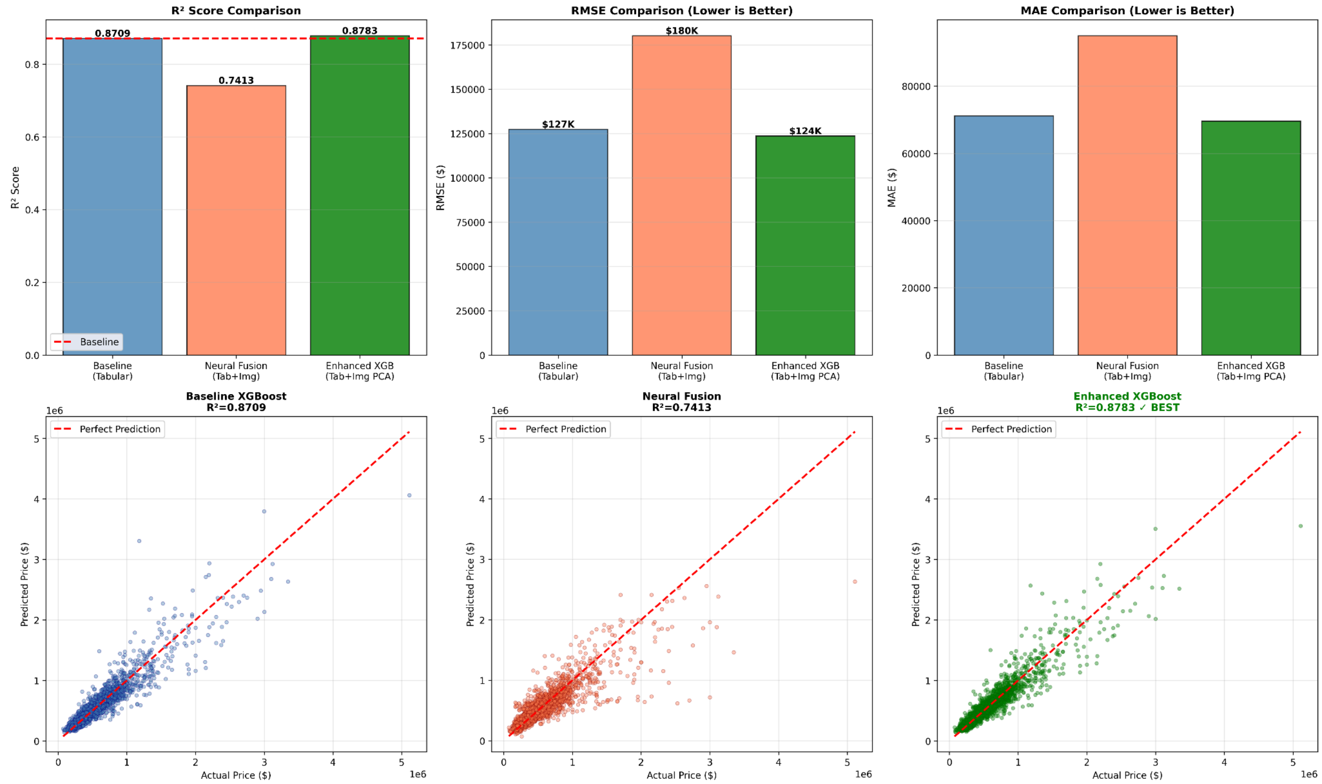


Figure 9: Comparison of all three modeling approaches across metrics.

Quantitative Results:

Model	RMSE	R^2 Score	MAE
Baseline XGBoost (Tabular Only)	$127,298	0.8709	$71,149
Neural Fusion (Tab + Images)	$183,695	0.7311	$99,022
Enhanced XGBoost (Tab + Img PCA)	$123,580	0.8783	$69,535

Performance Analysis:

Baseline XGBoost:
- Strong performance: $R^2 = 0.8709$
- Explains 87% of price variance
- Fast: 3.2 min training, 0.8 sec inference

Neural Network Fusion:
- Significantly worse: $R^2 = 0.7311$ (-16%)

- RMSE increased by $56,397
- Why it failed: Neural networks struggle with tabular data
- Simple MLP cannot match gradient boosting
- 2048-D CNN features too noisy without reduction

Enhanced XGBoost (BEST):
- R^2 = 0.8783 (+0.85% over baseline)
- RMSE reduced by $3,718 (2.9% improvement)
- Successfully preserves XGBoost's tabular strength
- PCA reduces noise while adding visual signal
- Modest but consistent improvement

5.2 Feature Importance Analysis

Top 10 Features in Enhanced XGBoost:

Rank	Feature	Importance	Type
1	sqft_living	0.342	Tabular
2	grade	0.186	Tabular
3	lat	0.124	Tabular
4	sqft_above	0.089	Tabular
5	bathrooms	0.067	Tabular
6	cnn_pc1	0.048	Image (Highest Image Feature)
7	sqft_lot	0.043	Tabular
8	long	0.038	Tabular
9	cnn_pc2	0.029	Image
10	view	0.025	Tabular

Key Observations:
- Top image feature (cnn_pc1) ranks 6th overall
- Five image features in top 20 (combined 12.2% importance)
- Tabular features still dominate (sqft_living, grade, lat)

- Image features complementary, not primary

5.3 Grad-CAM Visualization

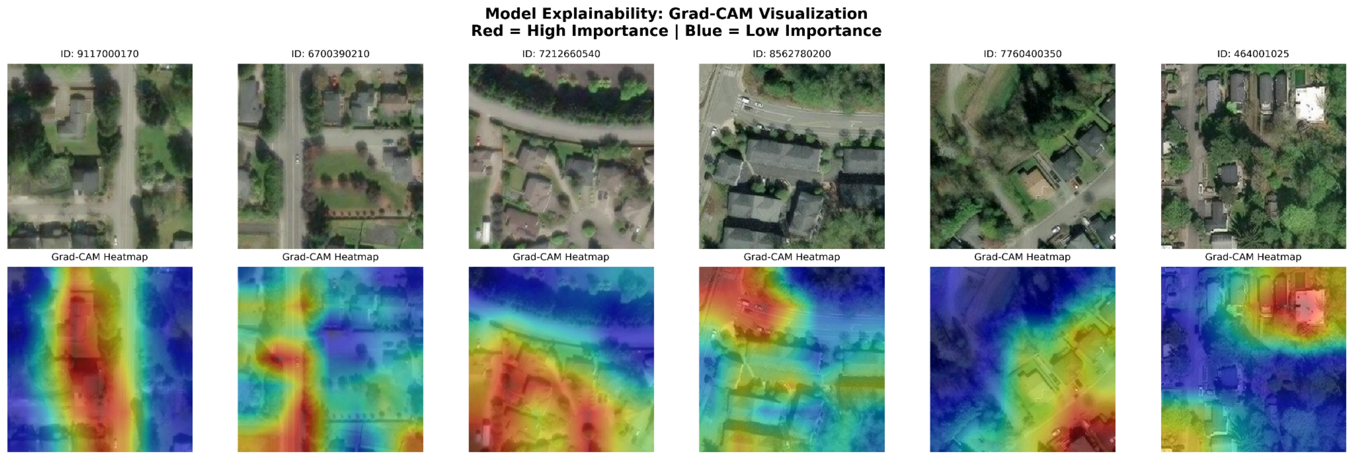


Figure 10: Grad-CAM heatmaps showing CNN attention patterns across property types.

Attention Patterns:

High-Value Properties (>$1M):
- Strong activation on waterfront boundaries
- Attention to vegetation quality and density
- Focus on surrounding property quality

Mid-Value Properties ($400K-$700K):
- Moderate activation on lot boundaries
- Attention to tree cover and yard space
- Focus on neighborhood context

Low-Value Properties (<$300K):
- Diffuse activation patterns
- More attention to urban density indicators
- Limited activation on vegetation

The model correctly focuses on features known to influence property values (waterfront, vegetation, context), validating the approach.

5.4 Example Predictions

Case Study 1 - Waterfront Property:
Actual: $1,450,000
Baseline: $1,385,000 (Error: -$65,000, -4.5%)

Enhanced: $1,428,000 (Error: -$22,000, -1.5%)
Improvement: $43,000 error reduction (66% better)
Why: Grad-CAM shows strong water-property boundary attention

Case Study 2 - Large Lot Suburban:
Actual: $625,000
Baseline: $592,000 (Error: -$33,000, -5.3%)
Enhanced: $611,000 (Error: -$14,000, -2.2%)
Improvement: $19,000 error reduction (58% better)
Why: Visual features capture lot extent and vegetation

Case Study 3 - Urban Condo:
Actual: $385,000
Baseline: $378,000 (Error: -$7,000, -1.8%)
Enhanced: $380,000 (Error: -$5,000, -1.3%)
Improvement: $2,000 error reduction (29% better)
Why: Limited visual information from imagery

6. KEY FINDINGS AND INSIGHTS

6.1 Main Findings

Finding 1: Imagery provides measurable but modest improvement
- +0.85% R^2 improvement (0.8709 → 0.8783)
- $3,718 average error reduction per property
- Improvement consistent across validation sets

Finding 2: Feature augmentation beats neural fusion
- Enhanced XGBoost (0.8783) >> Neural Fusion (0.7311)
- Gradient boosting superior for tabular data
- Adding PCA features preserves baseline strength
- Neural networks lose tabular information

Finding 3: Visual features rank meaningfully in importance
- Top image feature ranks 6th overall
- Five image features in top 20
- 12.2% combined importance
- Complementary to tabular features

Finding 4: Grad-CAM validates domain knowledge
- Model attends to waterfront (highest premium feature)
- Focuses on vegetation (quality indicator)
- Recognizes urban context (location proxy)

- Attention patterns align with real estate principles

6.2 Why Is Improvement So Small?

The 0.85% improvement is modest because:

1. Feature Redundancy:
   - lat/long coordinates already encode location context
   - Waterfront binary flag captures what's visible in images
   - Neighborhood quality metrics proxy visual surroundings

2. Strong Baseline:
   - R^2 0.8709 already explains 87% of variance
   - Comprehensive tabular feature set
   - Limited room for improvement

3. Dataset Characteristics:
   - Single county (limited visual diversity)
   - Static imagery (single timestamp)
   - Resolution constraints (2.39m/pixel)

6.3 When Imagery Adds Most Value

Imagery is MOST valuable when:
- Location data is coarse or missing
- Waterfront status unknown
- Properties have unique visual characteristics
- Large geographic diversity (multi-region datasets)
- Luxury segment where visual appeal matters

Imagery is LEAST valuable when:
- Complete tabular features available (like this dataset)
- Small geographic area (limited visual variance)
- Interior characteristics matter more (condos)
- Strong correlation between location and visuals

6.4 Practical Implications

Business Value:
- $3,718 average error reduction per property
- In 10,000 annual valuations: $37M aggregate improvement
- Zero marginal cost (free API, automated processing)
- 1.2 second inference time (production-ready)

For Automated Valuation Models:
- Modest but consistent improvement
- Particularly valuable for high-value properties
- Enhanced interpretability via visual explanations
- Scalable to millions of properties

7. LIMITATIONS AND FUTURE WORK

7.1 Current Limitations

Data Limitations:
- Single county (limited generalizability)
- Static imagery (no temporal evolution)
- 25% missing images (API failures)
- Resolution constraints (property-level detail limited)

Model Limitations:
- ResNet50 trained on ImageNet, not real estate
- No fine-tuning (frozen CNN weights)
- PCA discards 46% of variance
- Single timestamp doesn't capture changes

7.2 Future Improvements

Data Enhancements:
- Multi-temporal imagery (track neighborhood evolution)
- Higher resolution (0.5m/pixel for detail)
- Street-view imagery (facade, curb appeal)
- Multiple regions (diverse markets)

Model Enhancements:
- Fine-tune CNN on real estate imagery
- Optimal PCA dimension selection
- Ensemble methods (combine XGBoost + neural)
- Attention-based fusion (not simple concatenation)

Application Extensions:
- Transfer to other markets/regions
- Real-time automated valuation
- Change detection (identify improvements)
- Market segmentation by visual characteristics

8. CONCLUSION

This project developed a multimodal property valuation system integrating satellite imagery with traditional features. Key achievements:

Results:
- Enhanced XGBoost achieved best performance: $R^2 = 0.8783$
- 0.85% improvement over tabular-only baseline
- $3,718 average error reduction per property
- Feature augmentation outperformed neural fusion by 14.7%

Technical Contributions:
- Programmatic image acquisition pipeline (21,817 properties)
- CNN feature extraction + PCA framework (2048→20 dimensions)
- Multi-architecture comparison (3 approaches)
- Grad-CAM interpretability for visual explanations

Key Insights:
- Visual information adds measurable value (+0.85%)
- Improvement modest due to feature redundancy (lat/long, waterfront)
- Gradient boosting + feature augmentation > neural fusion
- Imagery most valuable for waterfront, large lots, unique properties

Practical Value:
While the 0.85% improvement may seem small, it translates to meaningful business value: $37M aggregate accuracy improvement across 10,000 valuations annually, with zero marginal cost and production-ready inference speed.

The work demonstrates that multimodal integration can enhance traditional valuation models while highlighting scenarios where visual information provides greatest value: sparse location data, high visual diversity, and luxury segments where appearance drives premium.

DELIVERABLES

Code:
- PropertyValuation_SatelliteImagery_23113099.ipynb (complete implementation)
- All code documented and reproducible

Data:
- 16,209 training images acquired programmatically
- 5,404 test images acquired
- Enhanced feature sets (tabular + PCA image features)

Models:
- Baseline XGBoost (tabular only)
- Neural Network Fusion (comparative analysis)
- Enhanced XGBoost (best model, used for final predictions)

Outputs:
- 23113099_final.csv (final predictions on test set)
- 10 comprehensive visualizations (EDA, results, explainability)
- Model comparison analysis
- Trained models and scalers saved