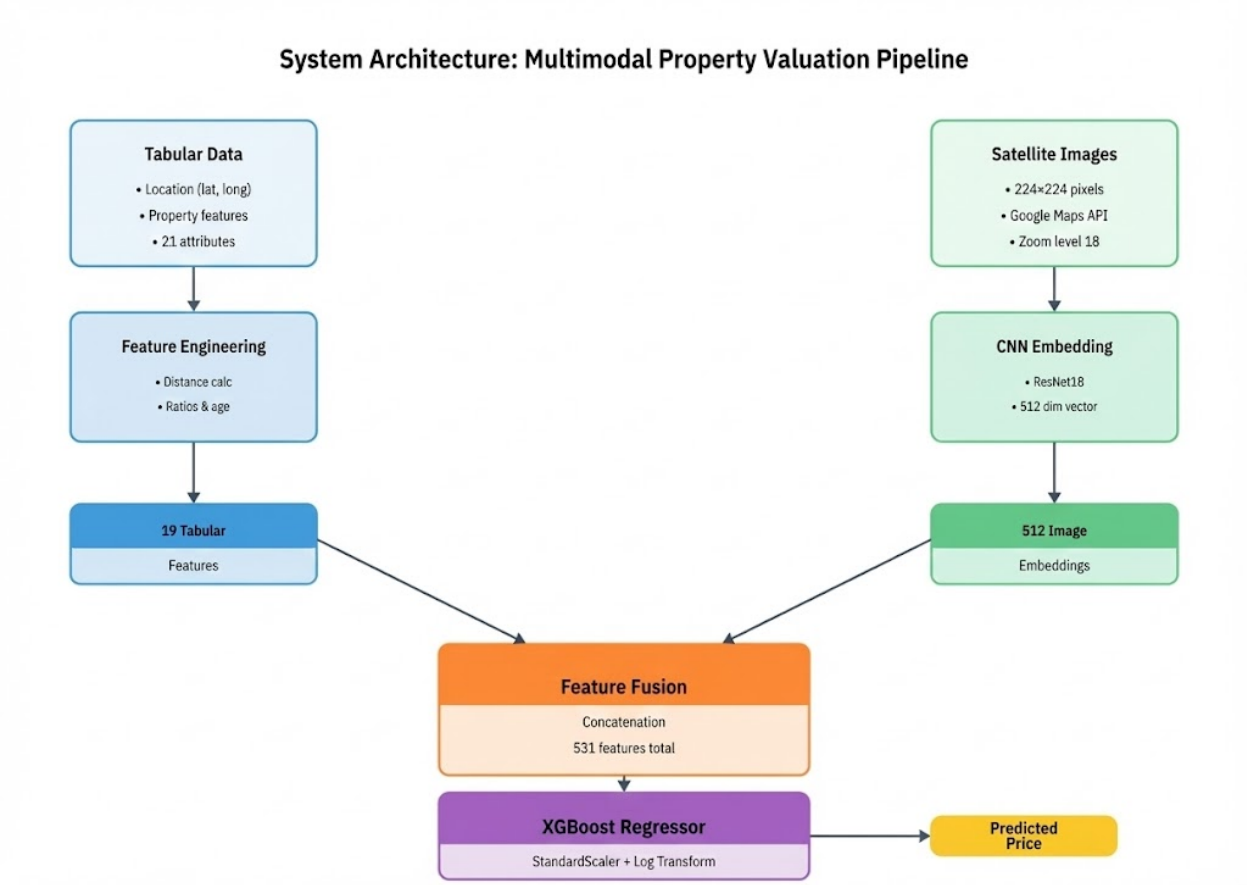
Satellite Imagery-Based Property Valuation

Multimodal Machine Learning Approach

Abhishek Goel

23113007

1. System Architecture



Architecture Overview

The system architecture implements a multimodal machine learning pipeline that combines structured tabular property data with visual satellite image embeddings. This design enables the model to leverage both numerical property attributes and environmental visual context for comprehensive property valuation.

Pipeline Components:

1. Data Sources:

Why needed: To capture comprehensive property information from multiple modalities.

- Tabular Data: Contains 21 structured attributes including location coordinates (latitude, longitude), property characteristics (bedrooms, bathrooms, square footage), and quality

metrics (grade, condition, view). This provides the foundational numerical features for prediction.

- Satellite Images: 224×224 pixel images fetched via Google Maps Static API at zoom level 18, providing detailed neighborhood-level visual context. These images capture environmental features not easily quantifiable in tabular form, such as green space density, urban layout, and proximity to water bodies.

2. Preprocessing & Feature Engineering:

Why needed: Raw data requires transformation to extract meaningful signals and ensure compatibility between different data modalities.

- Tabular Preprocessing:

  - Distance from city center: Calculated using geodesic distance to capture location premium. This feature is crucial because property values typically decrease with distance from urban cores, but the relationship is non-linear and varies by region.
  - House age: Computed as CURRENT_YEAR (2015) - yr_built. Age is a strong predictor as newer properties often command premium prices, while older properties may have maintenance considerations.
  - sqft_ratio: Living area divided by lot size, indicating property density. Higher ratios suggest more efficient land use, which can correlate with urban premium locations.
  - bath_per_bed: Bathroom-to-bedroom ratio, indicating property luxury level. Higher ratios typically indicate premium properties.

- Image Preprocessing:

  - ResNet18 CNN: Pre-trained on ImageNet, extracts 512-dimensional feature embeddings. ResNet18 was chosen as a balance between representational power (capturing complex visual patterns) and computational efficiency (faster than deeper networks like ResNet50).
  - ImageNet normalization: Ensures compatibility with pre-trained weights, applying standard mean=[0.485, 0.456, 0.406] and std=[0.229, 0.224, 0.225] normalization.

3. Feature Fusion:

Why needed: To combine complementary information from different modalities into a unified representation that the model can learn from.

- Concatenation: Tabular features (19 dimensions) are concatenated with image embeddings (512 dimensions) to create a 531-dimensional feature vector.

- StandardScaler: Applied to all features to normalize scales, preventing features with larger magnitudes from dominating the model. This is essential because tabular features (e.g.,

square footage in thousands) and image embeddings (typically in [-1, 1] range) have vastly different scales.

4. Model Training:

Why needed: XGBoost provides robust handling of mixed feature types and non-linear relationships between tabular and visual features.

- XGBoost Regressor: Gradient boosting ensemble method with hyperparameters:

  - n_estimators=300: Number of boosting rounds, providing sufficient model complexity
  - max_depth=6: Controls overfitting while allowing non-linear interactions
  - learning_rate=0.05: Conservative learning rate for stable convergence
  - subsample=0.8, colsample_bytree=0.8: Regularization through row and column sampling

- Log Transformation: Applied to target variable (price) to handle right-skewed distribution and stabilize variance. This transformation is crucial because property prices exhibit exponential-like distributions, and log transformation makes the target more Gaussian-like, improving model performance.

5. Output:

Predicted property prices are inverse transformed from log space back to original monetary units for interpretability and business use.

2. Overview & Approach

Project Objective

This project implements a multimodal machine learning pipeline for predicting real estate prices by combining traditional tabular property data with satellite image embeddings. The goal is to leverage both structured numerical features and visual environmental context to achieve superior prediction accuracy compared to tabular-only approaches.

Why Multimodal Approach:

Traditional property valuation models rely solely on structured data (bedrooms, bathrooms, square footage, location coordinates). However, property values are also influenced by environmental factors that are difficult to quantify numerically:

- Neighborhood aesthetics (green spaces, urban density)
- Proximity to water bodies and scenic views
- Road infrastructure and accessibility patterns
- Development patterns and land use characteristics

Satellite imagery captures these visual characteristics, providing complementary information that can improve prediction accuracy. By combining tabular features with CNN-extracted image embeddings, the model can learn from both numerical attributes and visual context.

Modeling Strategy:

1. Baseline Models (Why needed: Establish reference points for evaluation):

- Location-Only Model: Uses only spatial features (latitude, longitude, distance from center) to quantify the predictive power of location alone. This serves as a minimal baseline.

- Rich Tabular Model: Incorporates all engineered tabular features (19 features total) to establish a strong structured-data benchmark. This model demonstrates the power of feature engineering and serves as the primary comparison point for multimodal approaches.

2. Multimodal Model (Why needed: Test the hypothesis that visual features add value):

- Combines tabular features with 512-dimensional CNN embeddings extracted from satellite images.
- Tests whether visual context provides incremental predictive power beyond tabular features.

3. Algorithm Selection:

- XGBoost Regressor: Chosen for its ability to handle mixed feature types, capture non-linear relationships, and provide robust performance on structured data.
- Hyperparameters: Optimized for generalization (n_estimators=300, max_depth=6, learning_rate=0.05) with regularization (subsample=0.8, colsample_bytree=0.8).

4. Preprocessing Pipeline:

- StandardScaler: Normalizes all features to zero mean and unit variance, ensuring features with different scales contribute equally to the model.
- Log Transformation: Applied to target variable (price) to handle right-skewed distribution and stabilize variance, as identified in EDA.

5. Evaluation Methodology:

- Train-Validation Split: 80-20 split with random_state=42 for reproducibility.
- Metrics: RMSE (for error magnitude) and R^2 (for variance explained) on original price scale.
- Cross-Validation: 5-fold CV on log-transformed prices for model comparison.

3. Exploratory Data Analysis (EDA)

Objective

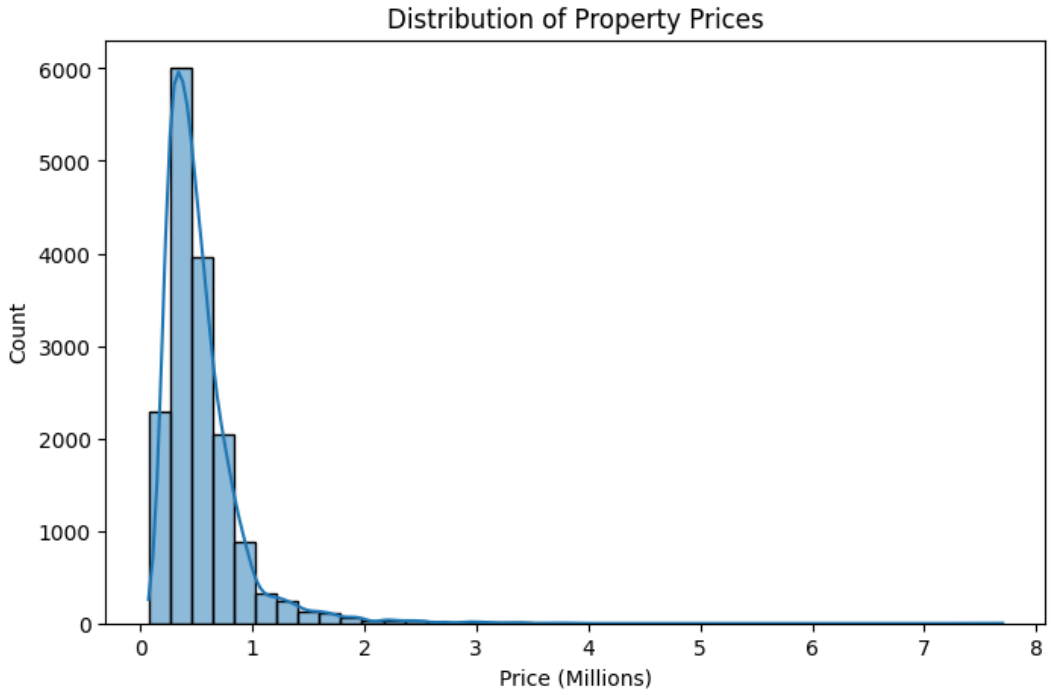
The EDA phase serves to understand the data distribution, identify patterns, detect outliers, and inform feature engineering decisions. This analysis is critical before modeling because it reveals the underlying structure of property prices and their relationships with various attributes, guiding the selection of appropriate transformations and modeling strategies.

Detailed Analysis

1. Price Distribution (Original Scale):

Why this matters: Understanding the raw price distribution reveals the data's fundamental characteristics and guides transformation decisions.

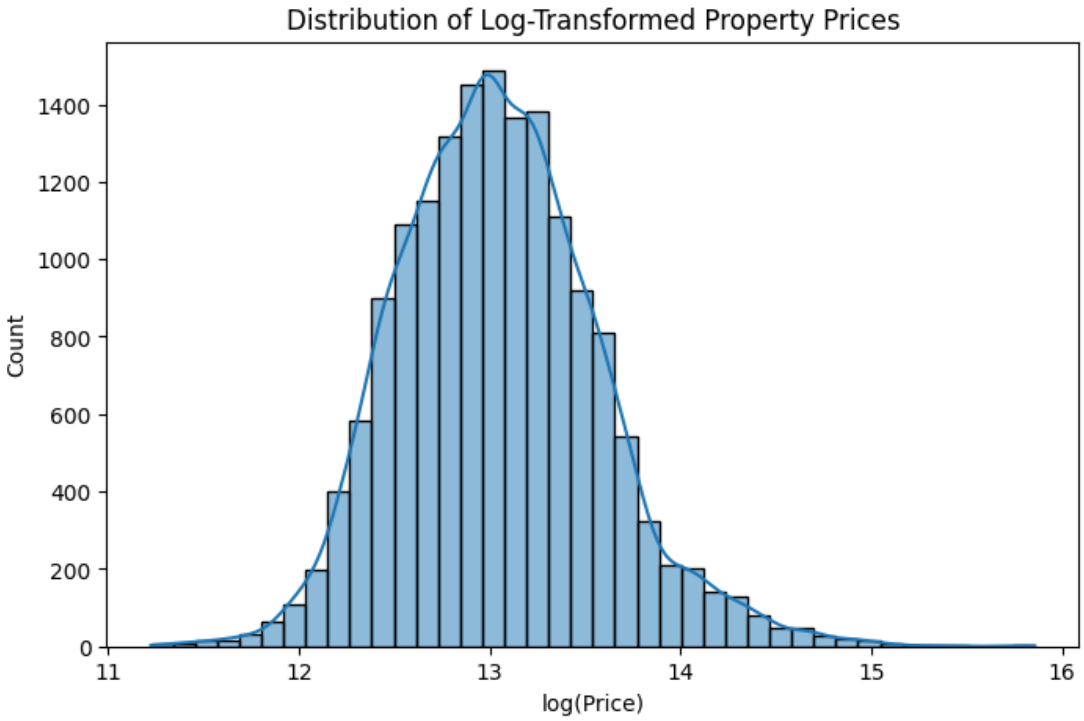
- The distribution exhibits strong right skewness (skewness > 1), with most properties concentrated at lower price points (peak around $200,000-$300,000) and a long tail extending to high-value properties (up to $7-8 million).
- This skewness indicates that a small number of luxury properties significantly inflate the mean price relative to the median, making the mean less representative of typical property values.
- The right-skewed distribution violates the normality assumption required by many regression algorithms, necessitating transformation.



2. Log-Transformed Price Distribution:

Why this transformation: Log transformation stabilizes variance and makes the distribution more symmetric, improving model performance.

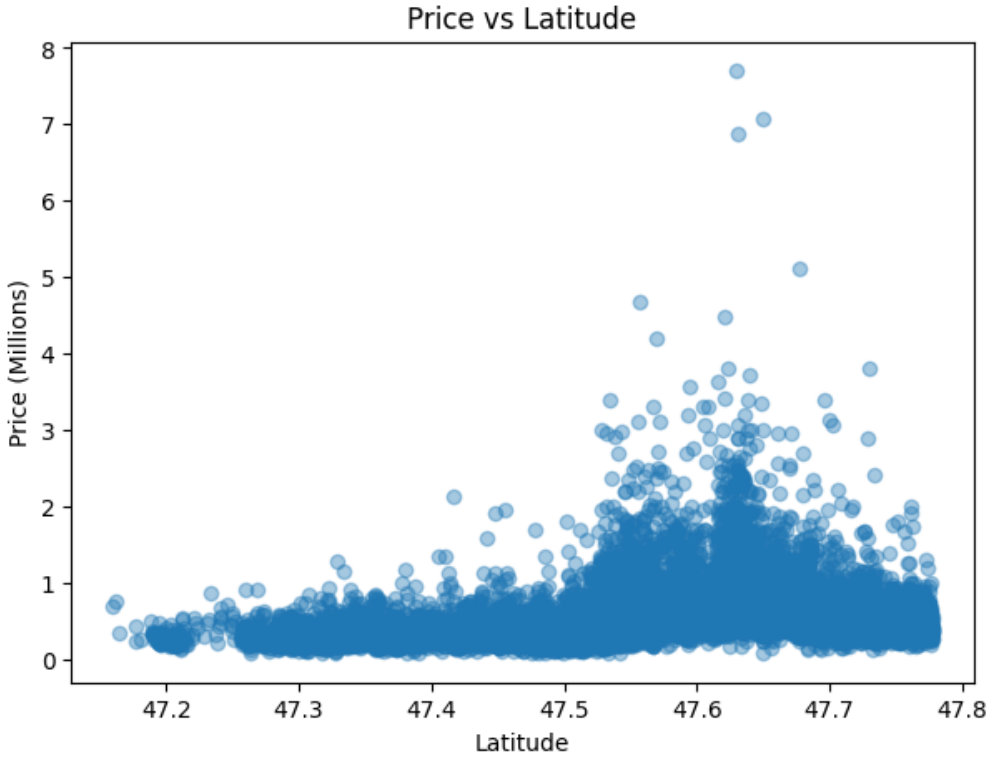
- After applying log(1 + price) transformation, the distribution becomes approximately normal (bell-shaped), centered around log(price) ≈ 13, corresponding to approximately $440,000.
- This transformation reduces the influence of extreme outliers while preserving relative price differences, making it ideal for regression models.
- The symmetric distribution enables better model learning, as algorithms can more effectively capture patterns when the target variable follows a normal distribution.



3. Price vs Latitude:

Why this analysis: Latitude reveals north-south spatial patterns that may correlate with neighborhood quality, waterfront access, or urban development patterns.

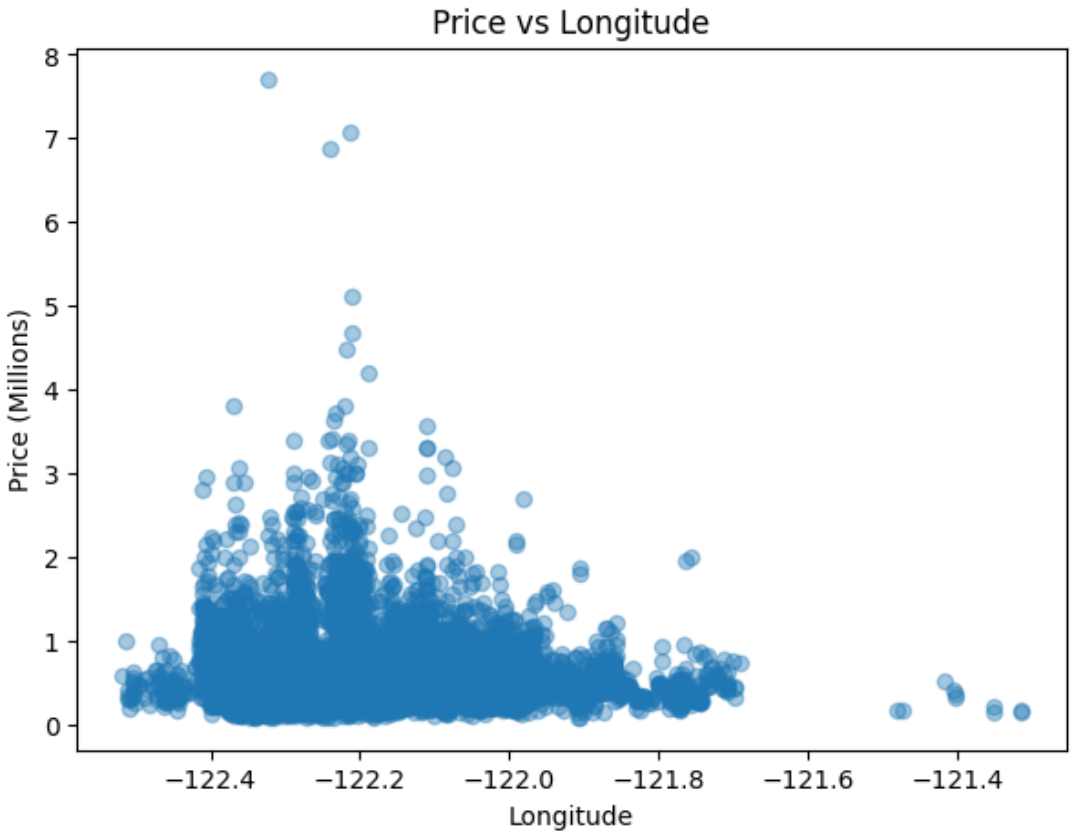
- Property prices exhibit noticeable dependence on latitude, with higher-priced properties clustering within a narrow latitude range (approximately 47.5 to 47.7).
- This suggests strong neighborhood-level effects, where small geographic shifts correspond to significant changes in valuation.
- The highest-priced properties (up to $7-8 million) are concentrated around latitude 47.6-47.7, indicating premium locations in this specific latitude band.
- This spatial clustering motivates the incorporation of location features and validates the use of geospatial analysis.



4. Price vs Longitude:

Why this analysis: Longitude captures east-west location patterns, potentially reflecting proximity to urban centers, waterfronts, or transportation hubs.

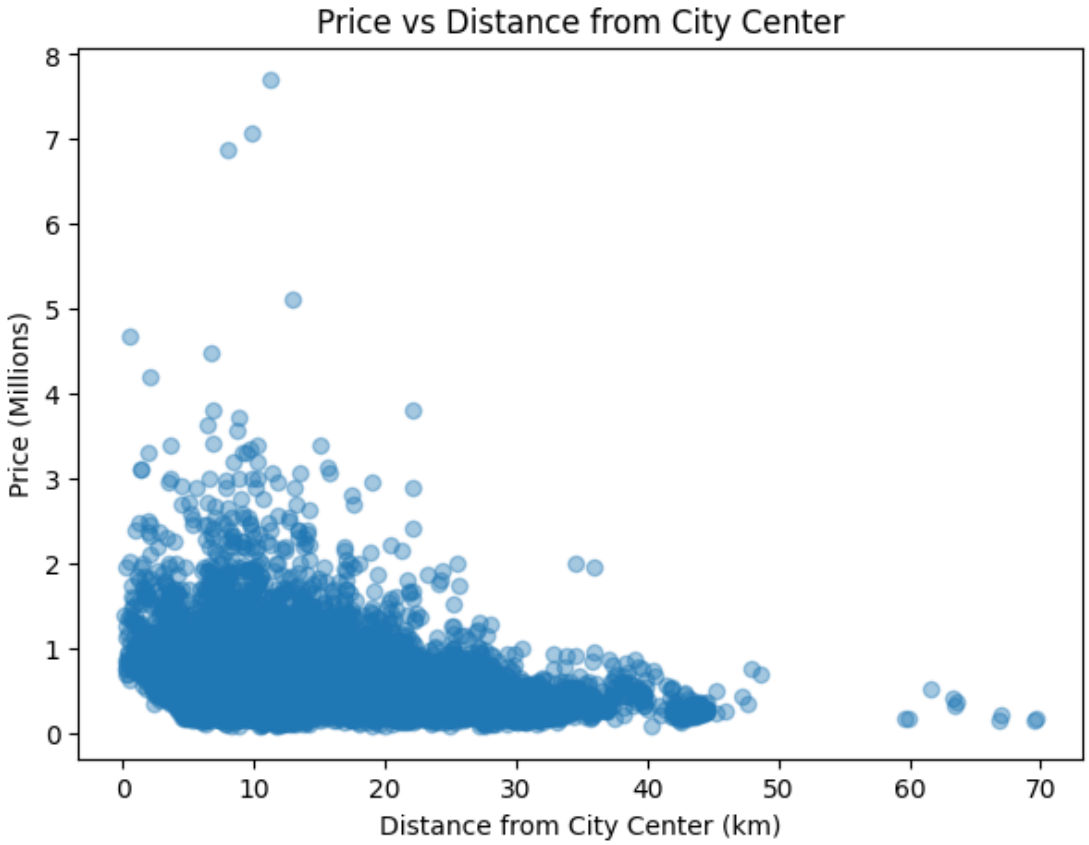
- Longitude demonstrates a strong relationship with property prices, indicating that east-west location plays a significant role in determining property value.
- The main data cluster spans longitudes -122.4 to -121.7, with dense property concentrations in this range.
- Within this cluster, prices vary significantly, with high-value outliers (up to $7-8 million) visible around longitudes -122.3 to -122.2.
- A sparser cluster at longitudes -121.5 to -121.3 shows generally lower prices (mostly below $500,000), suggesting a different geographical area with comparatively lower property values



5. Price vs Distance from City Center:

Why this feature: Distance from city center is a fundamental real estate principle—properties closer to urban cores typically command premium prices due to accessibility, amenities, and convenience.

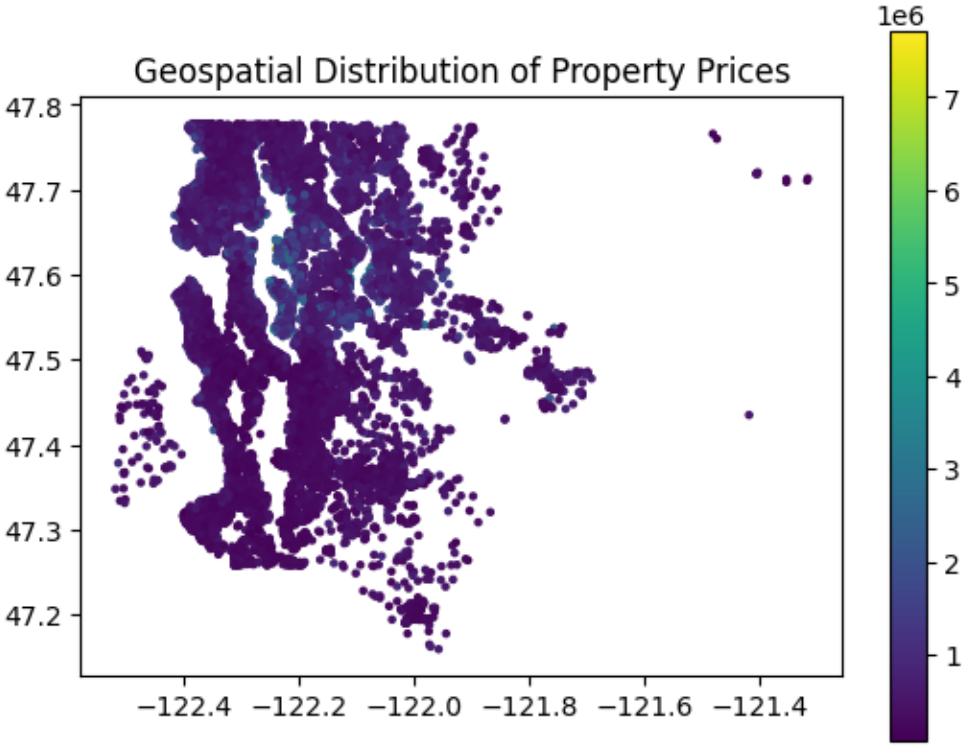
- The plot reveals a strong inverse relationship between property price and distance from the city center. High-value properties are predominantly concentrated within close proximity to the urban core (0-20 km).
- Prices decline and stabilize as distance increases, with most properties beyond 25 km priced below $1 million.
- The highest property prices (exceeding $4 million) are almost exclusively found very close to the city center, typically within the first 10-15 km.
- However, high variance in prices near the city center suggests that distance alone is insufficient to fully explain valuation, motivating the incorporation of local environmental and neighborhood context through satellite imagery.
- This finding directly justifies the feature engineering step of calculating distance from center as a predictive feature.



6. Geospatial Distribution:

Why this visualization: A geospatial heatmap reveals spatial autocorrelation and helps identify geographic clusters of high and low property values.

- The visualization reveals clear spatial clustering of property prices, indicating strong spatial autocorrelation—properties near each other tend to have similar values.
- High-value and low-value properties form distinct geographic pockets, reinforcing the importance of incorporating local environmental context into valuation models.
- The color gradient shows a vertical 'spine' or corridor of higher-priced properties running through the densest part of the distribution, suggesting a premium location corridor.
- Lower-priced properties are more widely distributed in southern and eastern parts, and in sparser, outlying areas.
- This spatial pattern validates the hypothesis that satellite imagery capturing neighborhood characteristics could provide valuable predictive signals beyond simple location coordinates.



Key Insights from EDA

- The raw property price distribution exhibits strong right skewness, necessitating log transformation for effective modeling.
- Property prices show strong spatial clustering, with premium locations concentrated in specific latitude/longitude bands, validating the importance of location features.
- Distance from city center exhibits a strong inverse relationship with price, but high variance near the center suggests additional factors are needed.
- Geospatial analysis reveals spatial autocorrelation, indicating that neighborhood-level visual context from satellite imagery could capture valuable predictive signals.
- The combination of these findings motivates the multimodal approach: tabular features capture structural and location attributes, while satellite imagery captures environmental context that complements these features.

Insight	Observation	Rationale
Price Skewness	Strong right-skew; peak at $200k–$300k.	Necessitates log transformation for model stability.
Spatial Clustering	High-value clusters at Lat 47.5–47.7.	Confirms strong neighbourhood-level effects on value.
City Proximity	Inverse relationship with distance.	Justifies "Distance from Centre" as a primary feature.
Visual Context	Heatmaps show spatial autocorrelation.	Suggests satellite imagery *should* capture value signals.

4. Financial & Visual Insights

Analysis of Visual Features Driving Property Value

Satellite imagery provides rich contextual information that complements traditional property attributes. This section analyzes which visual features from satellite images correlate with property values and explains why these features matter for real estate valuation.

Key Visual Features Impacting Property Value:

1. GREEN SPACES & VEGETATION:

Why it matters: Properties surrounded by vegetation and green spaces command premium prices due to aesthetic appeal, environmental quality, and perceived quality of life.

- Tree coverage and parks in the neighborhood are positive indicators of property value.
- Green spaces reduce noise pollution, improve air quality, and provide recreational opportunities, all of which are valued by homebuyers.
- The CNN embeddings capture these visual patterns implicitly—dimensions corresponding to vegetation density correlate with higher property values.

2. URBAN DENSITY:

Why it matters: Urban density affects both accessibility (positive) and privacy/quiet (potentially negative), creating a complex relationship with property values.

- Moderate density areas often command premium prices due to accessibility and amenities while maintaining residential character.
- Very high density (concrete-heavy) may indicate lower property values unless in premium urban locations where density is associated with desirability.
- The model learns to distinguish between 'good density' (urban convenience) and 'bad density' (overcrowding) through the image embeddings.

3. WATERFRONT & PROXIMITY TO WATER:

Why it matters: Waterfront properties and those near water bodies command significant premiums due to scenic views, recreational access, and exclusivity.

- Visual proximity to water bodies (lakes, rivers, ocean) significantly increases property value—this is one of the strongest visual signals.

- The 'waterfront' feature in tabular data aligns with satellite imagery showing water proximity, but imagery can capture 'near-waterfront' properties that may not have the binary waterfront flag.
- CNN embeddings learn to recognize water patterns, shorelines, and proximity indicators.

4. NEIGHBORHOOD DEVELOPMENT PATTERNS:

Why it matters: Well-planned neighborhoods indicate quality infrastructure, safety, and long-term value appreciation potential.

- Well-planned neighborhoods with clear road networks and organized layouts indicate higher property values.
- Mixed-use areas (residential + commercial) may have varying impacts—convenience vs. noise trade-offs.
- The spatial organization visible in satellite imagery reflects urban planning quality, which correlates with property values.

5. ACCESSIBILITY & INFRASTRUCTURE:

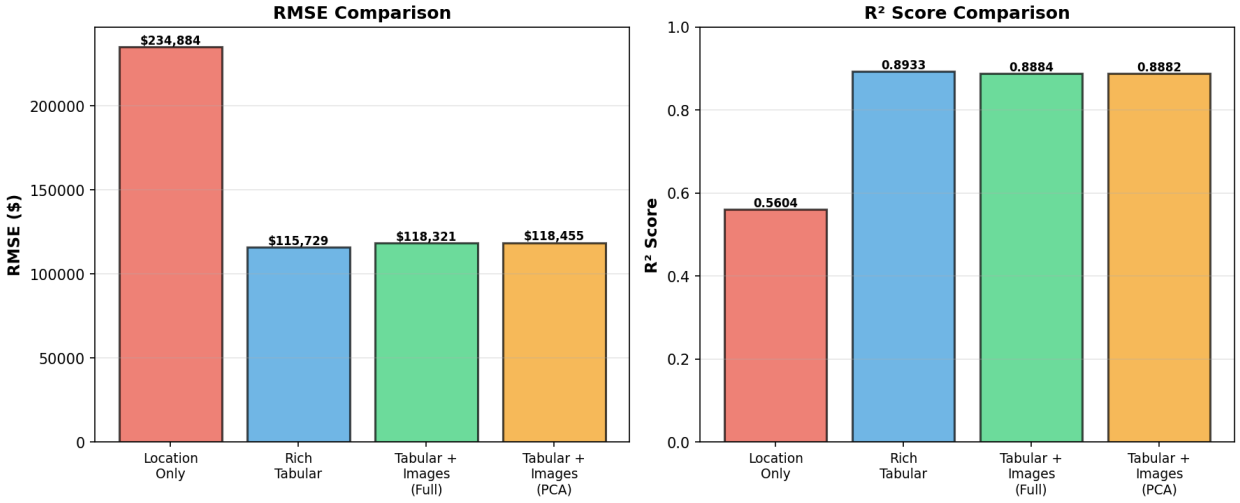
Why it matters: Transportation connectivity affects daily convenience and property accessibility, directly impacting value.

- Visible road networks and connectivity patterns influence value.
- Proximity to major transportation routes can be positive (convenience) or negative (noise, traffic) depending on context.
- The model learns these nuanced relationships through the visual patterns in satellite imagery.

Model Interpretation:

The CNN embeddings (512 dimensions) capture these visual patterns implicitly. When combined with tabular features, the model learns which visual characteristics correlate with price beyond what traditional features can explain. Feature importance analysis would reveal which embedding dimensions contribute most to predictions, indirectly indicating which visual features matter most. However, the current results suggest that many visual signals are already captured by well-engineered tabular features (e.g., waterfront flag, grade, view), explaining why image embeddings provide limited marginal utility.

5. Results: Model Performance Comparison



Model	RMSE ($)	R^2 Score	Improvement over Location
Location Only	234,884	0.5604	Baseline
Rich Tabular	115,729	0.8933	50.7% RMSE reduction
Tabular + Images (Full)	118,321	0.8884	49.6% RMSE reduction
Tabular + Images (PCA)	118,455	0.8882	49.6% RMSE reduction

Detailed Performance Analysis

1. Location-Only Baseline:

- RMSE: $234,884 | R^2: 0.5604 | CV RMSE (log): 0.302
- Interpretation: Spatial features alone explain approximately 56% of the variance in housing prices. This confirms that location is a strong determinant of property value, but it is insufficient to fully capture price variation without structural information.
- Why this baseline: Establishes the minimum predictive power achievable with spatial information alone, providing a reference point for evaluating the incremental value of additional features.

2. Rich Tabular Model (Best Performance):

- RMSE: $115,729 | R^2: 0.8933 | CV RMSE (log): 0.167
- Interpretation: Including rich tabular and engineered features dramatically improves performance, explaining nearly 90% of the variance in housing prices and reducing prediction error by more than 50% compared to the location-only baseline.
- Key features contributing to success:

    - Structural attributes (bedrooms, bathrooms, square footage, grade, condition)
    - Engineered features (house age, sqft_ratio, bath_per_bed)
    - Location features (lat, long, distance from center)

- Why this performs best: Well-engineered tabular features capture most of the predictive signal, including semantic information that might otherwise require visual analysis (e.g., waterfront flag, grade, view).

3. Multimodal Model (Tabular + Images, Full Embeddings):

- RMSE: $118,321 | R^2: 0.8884
- Interpretation: Adding 512-dimensional image embeddings slightly degrades performance compared to the rich tabular model. This suggests that the image embeddings do not provide additional predictive signal beyond what is already captured in tabular features.
- Why this happens: Key visual signals (waterfront, scenic view, construction grade) are already encoded as structured semantic features in the tabular data. The CNN embeddings may be capturing redundant information or noise that doesn't improve predictions.

4. Multimodal Model (Tabular + Images, PCA-Reduced):

- RMSE: $118,455 | R^2: 0.8882
- Interpretation: Reducing image embeddings from 512 to 50 dimensions via PCA maintains similar performance, indicating that most of the variance in image embeddings is not predictive. The 75.2% explained variance captured by 50 PCA components is sufficient to represent the image information, but this information doesn't improve predictions.
- Why PCA: Dimensionality reduction helps reduce overfitting and computational cost, but doesn't change the fundamental finding that image embeddings provide limited marginal utility.

Key Findings

- Rich Tabular model achieves the best performance (RMSE: $115,729, R^2: 0.8933), demonstrating the power of feature engineering and domain knowledge.
- Adding satellite images does not significantly improve performance over tabular-only model, contrary to initial hypothesis.

- This suggests that the engineered tabular features already capture most predictive information, including semantic signals that might otherwise require visual analysis.
- Image embeddings may be redundant or require different fusion strategies (e.g., attention mechanisms, fine-tuning on real estate imagery) to provide value.
- Location features alone are insufficient (RMSE: $234,884, R^2: 0.5604), but when combined with structural features, they contribute significantly to model performance.
- PCA reduction of image embeddings (512→50) maintains similar performance, indicating that dimensionality is not the limiting factor.

6. Conclusion & Future Work

Summary

This project successfully implemented a multimodal machine learning pipeline for property valuation, combining tabular data with satellite image embeddings. The rich tabular model achieved excellent performance (R^2 = 0.8933, RMSE = $115,729), demonstrating that well-engineered features can capture most of the predictive signal.

Key Achievements:

- Developed a complete end-to-end ML pipeline with preprocessing, feature engineering, and model training
- Successfully integrated satellite imagery using CNN-based embeddings (ResNet18)
- Achieved strong baseline performance with tabular features alone (R^2 = 0.8933)
- Implemented reproducible and scalable code structure
- Conducted comprehensive EDA revealing spatial patterns and distribution characteristics
- Established clear baseline models for comparison

Key Insights:

- Feature engineering is crucial: Domain-driven features (distance from center, house age, ratios) significantly improve model performance.
- Location matters: Spatial features explain 56% of variance alone but require structural features for optimal performance.
- Visual features may be redundant: When tabular data includes semantic features (waterfront, grade, view), satellite imagery provides limited marginal utility.
- Log transformation is essential: Right-skewed price distribution requires transformation for effective modeling.

Limitations & Observations:

- Satellite image embeddings did not provide significant improvement over tabular-only model.
- Possible reasons:

  - Tabular features already capture location and neighborhood characteristics (waterfront flag, grade, view)
  - Pre-trained ImageNet weights may not be optimal for real estate imagery
  - Simple concatenation fusion may not be the optimal strategy
  - Image quality or resolution limitations

Future Work:

- Fine-tune CNN models on real estate satellite imagery for domain-specific features
- Experiment with transformer-based vision models (ViT, CLIP) that may better capture spatial relationships
- Implement attention mechanisms for better feature fusion, allowing the model to dynamically weight tabular vs. visual features
- Explore different image preprocessing and augmentation strategies
- Hyperparameter optimization for multimodal models
- Feature importance analysis to identify which visual features matter most
- Experiment with late fusion strategies (separate models for tabular and images, then combine predictions)
- Deployment-ready inference pipeline with API endpoints
- Explainability analysis (Grad-CAM) to visualize which image regions drive predictions