# Satellite Imagery-Based Property Valuation

January 4, 2026

## Contents

# 1 Overview: Approach and Modeling Strategy

## 1.1 Problem Statement

A Real Estate Analytics firm aims to improve its valuation framework by developing a **Multimodal Regression Pipeline** that predicts property market value using both tabular data and satellite imagery. Traditional Automated Valuation Models (AVMs) rely on metadata like square footage, bedrooms, and location—but fail to capture the visual factors human appraisers naturally consider.

This project moves beyond standard data analysis by combining two different types of data—**numbers and images**—into a single predictive system. The goal is to build a model that accurately values assets by integrating "curb appeal" and neighborhood characteristics (like green cover or road density) into traditional pricing models.

## 1.2 Modeling Strategy: The Hybrid Architecture

We adopted a **Hybrid Multimodal** approach that separates the "visual understanding" and "valuation logic" into specialized components:

1. **Visual Feature Extraction (The "Eyes"):** We utilize **EfficientNet-B0**, a state-of-the-art Convolutional Neural Network, to process 64×64 pixel satellite images of each property. Instead of using the CNN for direct price prediction, we extract a high-dimensional **1280-length feature vector** representing the visual texture of the property and its surroundings.
2. **Tabular Feature Engineering:** We engineer robust features from the metadata, including `house_age`, `total_sqft`, and **Spatial KNN** features to capture micro-neighborhood pricing trends.
3. **Fusion Engine (The "Brain"):** The visual vectors from the CNN are concatenated with tabular features and fed into **CatBoost**, a gradient boosting library. This allows the model to learn complex, non-linear interactions that a simple neural network fusion head struggles to capture.

## 1.3 Rationale for the Hybrid Approach

Initial experiments with a pure Neural Network architecture revealed a phenomenon we term the "Luxury Ceiling"—the model systematically under-predicted prices for ultra-luxury properties (>$5M). Analysis indicated that the linear fusion head could not effectively extrapolate beyond the training distribution. By shifting the final valuation logic to CatBoost, which excels at handling outliers and tabular nuances through decision tree ensembles, we achieved significantly improved performance.

# 2 Exploratory Data Analysis and Preprocessing

## 2.1 Data Cleaning

The raw dataset contained approximately 21,000 property records from King County, Washington. Before modeling, the following cleaning steps were applied:

- **Duplicate Removal:** Rows with duplicate property IDs were dropped, retaining only the first occurrence.
- **Outlier Removal:** An erroneous record listing 33 bedrooms (a clear data entry error) was removed.
- **Date Parsing:** The sale date column was converted to datetime format to enable temporal feature extraction.

## 2.2 Feature Engineering

To enhance predictive power, several new features were engineered from the raw data:

**Table 1:** Engineered Features

| Feature | Formula | Rationale |
|---|---|---|
| `house_age` | `sales_year - yr_built` | Captures depreciation over time |
| `is_renovated` | `1 if yr_renovated > 0 else 0` | Binary renovation indicator |
| `total_sqft` | `sqft_living + sqft_lot` | Combined interior/exterior space |
| `zip_idx` | `LabelEncoder(zipcode)` | Integer index for embedding layer |

**Target Transformation:** The price column was transformed using `log1p()` to address extreme right-skewness, as detailed in Section 2.3.

## 2.3 Price Distribution Analysis

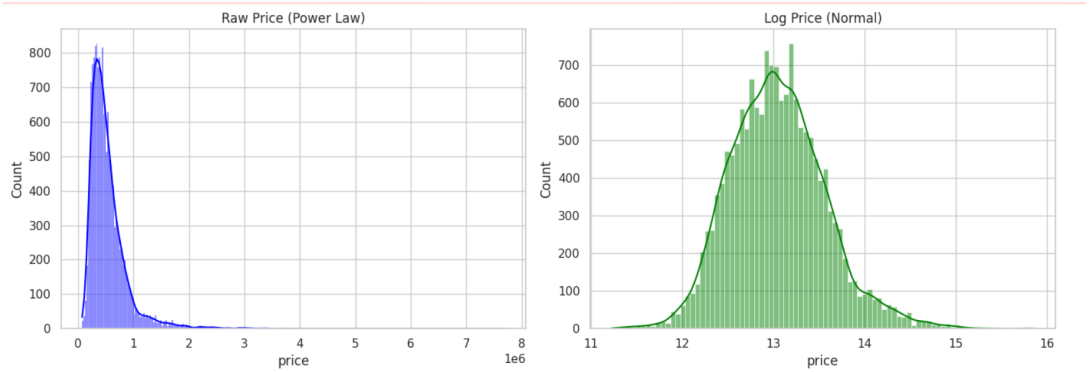


**Figure 1:** Price Distribution: Raw (left) vs. Log-Transformed (right). The raw distribution exhibits a severe Power Law shape with a long tail extending to \$7.7M+. After log transformation, the distribution approximates a Gaussian curve.

The raw price distribution exhibits a severe **Power Law** shape: the majority of homes cluster between \$300k–\$700k, while a long tail extends to \$7.7M+ for ultra-luxury properties. Training a regression model on this raw distribution causes gradient updates to be dominated by the few extreme high-value samples.

After applying $\log(1 + x)$ transformation, the distribution transforms into a near-perfect Gaussian curve. This ensures stable gradient updates and causes the model to optimize for relative (percentage) error rather than absolute dollar error.

## 2.4 Correlation Analysis

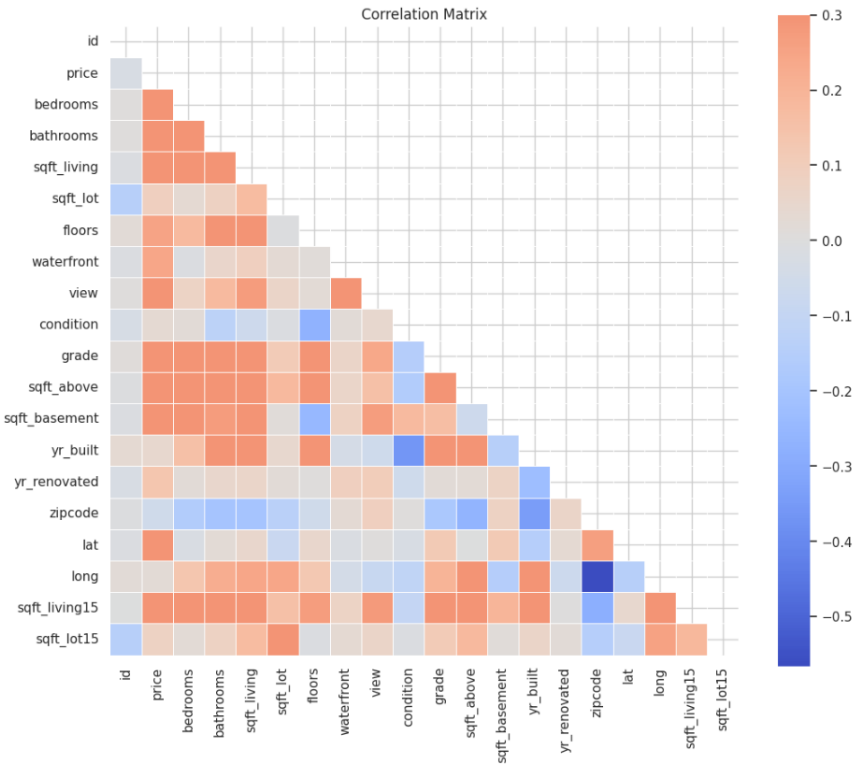


**Figure 2:** Feature Correlation Matrix. Strong positive correlations are observed between price and `sqft_living` (0.70), `grade` (0.67), and `bathrooms` (0.53).

Key observations from the correlation heatmap include:

- **Strongest Positive Correlations:** `sqft_living` (0.70), `grade` (0.67), and `bathrooms` (0.53).
- **Notable Patterns:** While `sqft_living` is strongly correlated with price, the relationship is not purely linear—a high-grade small home may command a higher price than a low-grade mansion. This non-linearity motivates the use of tree-based models.

## 2.5 Feature-Price Relationships

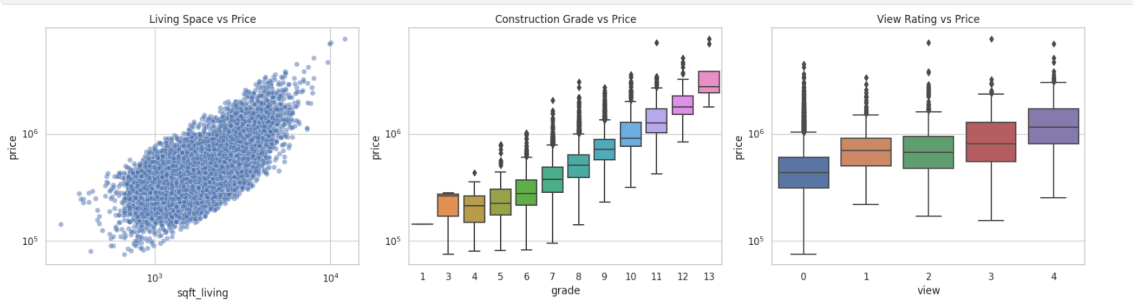


**Figure 3:** Key Feature Relationships: (Left) Living Space vs. Price showing log-linear correlation, (Center) Construction Grade vs. Price showing exponential value increase at higher grades, (Right) View Rating vs. Price demonstrating the "view premium."

Three critical relationships emerge from the feature analysis:

- **Living Space (`sqft_living`):** A clear log-linear relationship exists between interior square footage and price. However, variance increases significantly at higher sizes, indicating that size alone does not determine luxury pricing.
- **Construction Grade:** The box plots reveal an *exponential* relationship—each grade increment yields progressively larger price increases. Grade 13 (luxury custom) homes command median prices exceeding $2M, while Grade 5 homes cluster around $200k.
- **View Rating:** Properties with higher view ratings (3–4) show elevated median prices and longer upper tails, confirming that scenic views command a significant premium. Notably, even a rating of 1 (minimal view) slightly outperforms 0 (no view).

## 2.6 Geospatial Analysis

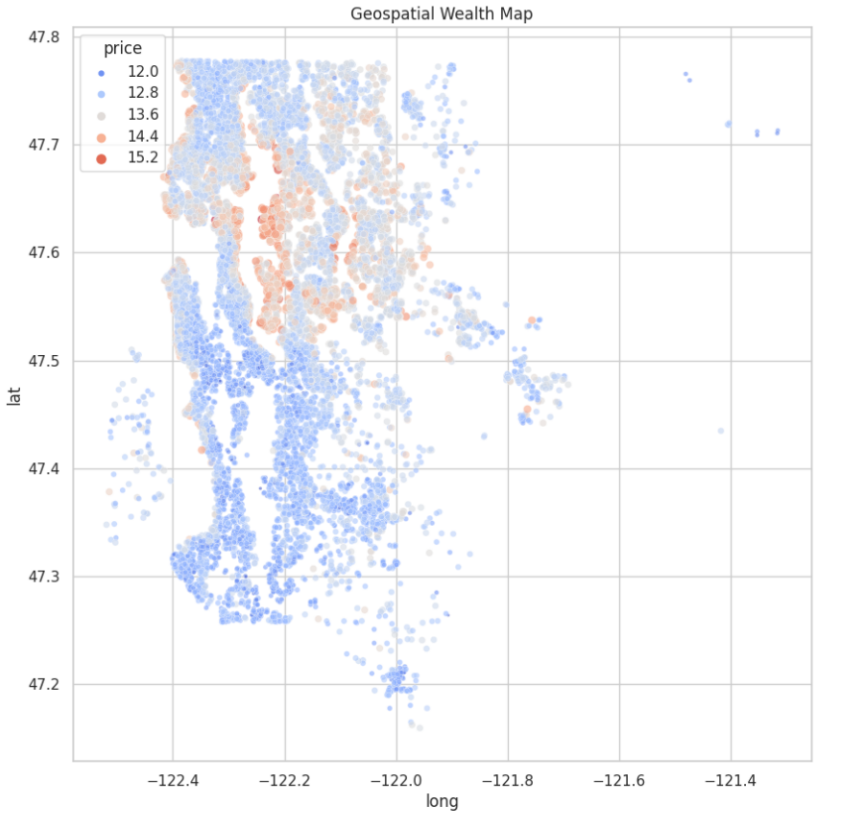


**Figure 4:** Geospatial Wealth Map. Each point represents a property, colored by log price. Distinct "wealth clusters" (red) are visible along the Lake Washington shoreline.

The scatter plot of latitude and longitude colored by price reveals distinct **wealth clusters** concentrated along the Lake Washington shoreline. Notably, these clusters often cross official zipcode boundaries, creating micro-neighborhoods that simple zipcode encoding cannot capture.

This finding directly motivated our use of **Spatial K-Nearest Neighbors (KNN)**: by computing the average price of a property's geographic neighbors, we capture the "neighbor effect"—a house surrounded by expensive homes is likely to be more valuable than its metadata alone suggests.

**Table 2:** Top 5 Most Expensive Zipcodes

| Zipcode | Count | Mean Price | Median Price |
|---|---|---|---|
| 98039 (Medina) | 36 | \$2.09M | \$1.91M |
| 98004 (Bellevue) | 233 | \$1.33M | \$1.10M |
| 98040 (Mercer Island) | 204 | \$1.20M | \$997k |
| 98112 (Capitol Hill) | 206 | \$1.10M | \$930k |
| 98102 (Eastlake) | 79 | \$934k | \$690k |

## 2.7 Satellite Imagery Analysis


**Figure 5:** Satellite Image Comparison: High-Value Properties (top row, \$6.9M–\$7.7M) vs. Low-Value Properties (bottom row, \$75k–\$81k). Note the stark differences in vegetation density, lot size, and building layout.

A visual comparison of top-tier versus entry-level properties reveals fundamental differences:

**Table 3:** Visual Feature Comparison

| **Feature** | **High-Value Properties** | **Low-Value Properties** |
|---|---|---|
| Density | Low (large setbacks) | High (lot lines touching) |
| Vegetation | Heavy tree canopy | Sparse or absent |
| Road Layout | Curvilinear, private | Grid pattern, street-facing |
| Amenities | Pools, tennis courts visible | None visible |

These visual cues are *not captured in tabular data*. For instance, `sqft_lot` does not distinguish between a paved lot and a forested lot. This observation validates our core hypothesis: satellite imagery contains unique value signals.

# 3 Financial and Visual Insights

## 3.1 Visual Explainability Using Grad-CAM

To interpret how satellite imagery influences property price predictions, we employ **Gradient-weighted Class Activation Mapping (Grad-CAM)**. This technique highlights spatial regions in satellite images that contribute most strongly to the convolutional neural network's output.

- **Warmer colors (red/yellow):** Regions with higher positive influence on the predicted price.
- **Cooler colors (blue):** Regions with minimal or negative impact.

Grad-CAM visualizations were generated for a representative set of 20 samples, including random properties, highest-value predictions ($3M–$10M), and lowest-value predictions ($150k–$165k). This stratified sampling enables comprehensive analysis of how the model's visual focus shifts across the property value spectrum.

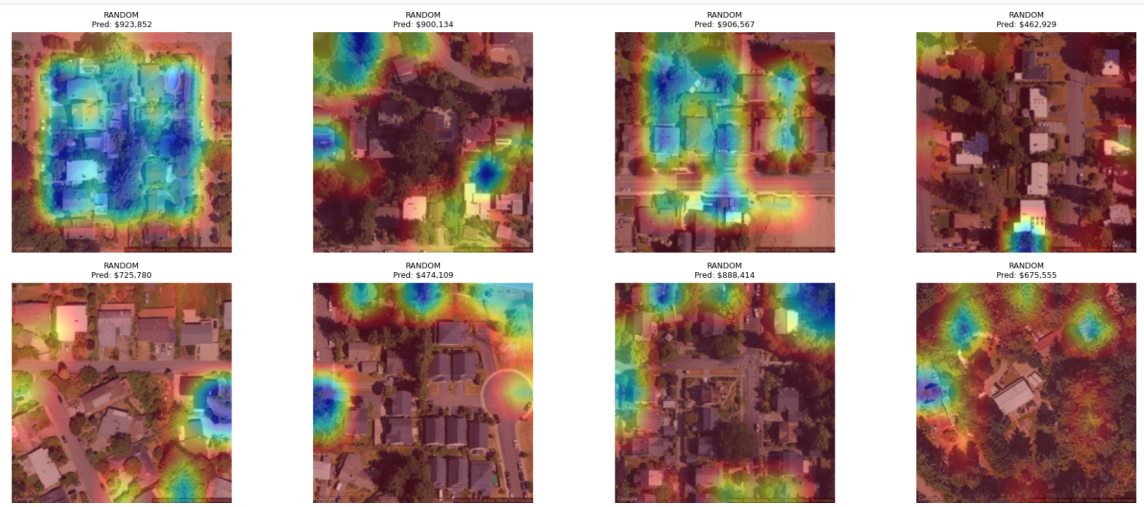


**Figure 6:** Grad-CAM Visualizations: Random Sample Predictions. These baseline samples demonstrate typical model attention patterns across the general property distribution.

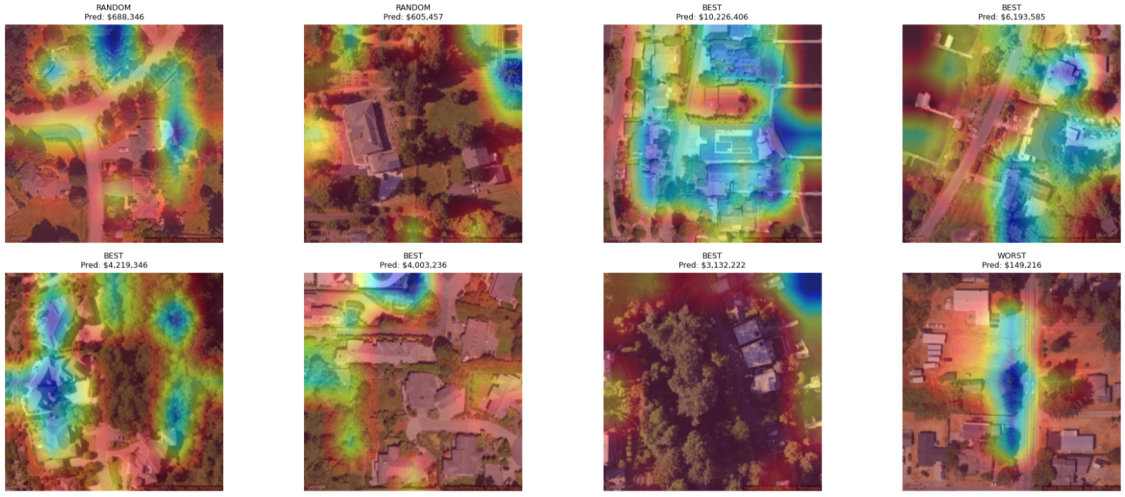


**Figure 7:** Grad-CAM Visualizations: Best Predictions (High-Value Properties, $3M–$10M). Note the diffuse, contextual attention spreading outward to encompass vegetation, lawns, and setbacks—indicating the model prioritizes "estate context" for luxury valuation.

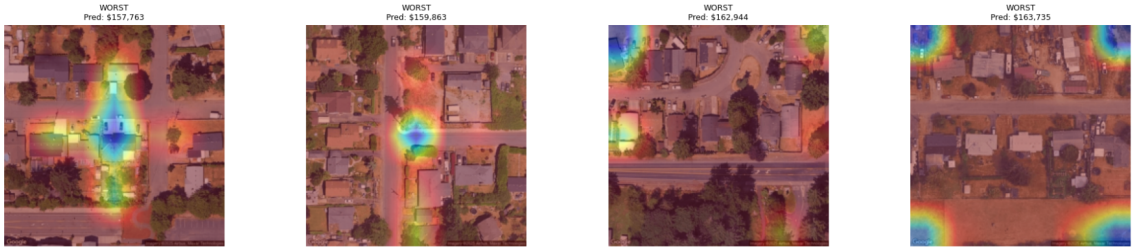


**Figure 8:** Grad-CAM Visualizations: Worst Predictions (Low-Value Properties, $150k–$165k). Attention contracts inward, tightly hugging building footprints and highlighting density constraints where structures meet streets or neighbors.

## 3.2 Dual-Mode Attention: Context vs. Structure

Analysis of the Grad-CAM overlays reveals a distinct behavioral shift in model attention based on predicted property value.

**Table 4:** Model Attention Patterns by Value Tier

| Value Tier | Attention Focus | Visual Signature |
|---|---|---|
| High-Value (>$3M) | Contextual/Environmental | Large, diffuse heatmaps |
| Low-Value (<$200k) | Structural/Density | Compact, localized heatmaps |

For **luxury properties**, the model has learned that surrounding space—privacy, vegetation, land availability—is a primary value driver. The building itself is assumed valuable; context determines *how* valuable. For **entry-level properties**, the model focuses on structure size and density constraints, with the absence of "green spread" signaling a lower value ceiling.

### 3.3 Built Infrastructure vs. Vegetation Influence

Grad-CAM activation analysis reveals clear patterns:

- **High-Value Drivers:** Tree canopy, lawns, setbacks, and curvilinear roads (hallmarks of luxury communities).
- **Low-Value Drivers:** Grid street patterns, concrete dominance, and minimal lot boundaries (density signals).

### 3.4 Financial Interpretation

From a real estate valuation perspective, these findings align with established economic principles:

1. **The "Privacy Premium":** For luxury properties, the marginal value of additional land and privacy exceeds that of additional interior square footage. A 5,000 sqft home on 0.5 acres is worth less than the same home on 2 forested acres.
2. **The "Density Penalty":** In lower-value markets, high density signals urban congestion, reduced privacy, and lower desirability.
3. **The "Curb Appeal Effect":** The model values neighborhood context—not just building footprint. This is a signal that tabular data cannot fully capture.

### 3.5 Key Insight

The visual explainability analysis confirms that satellite imagery contributes **economically relevant information** to the pricing model. The CNN functions as a **Socio-Economic Context Extractor**, learning the "Visual Texture of Wealth" (greenery, privacy, curvilinear layouts) versus the "Visual Texture of Density" (grid patterns, concrete, tight boundaries).

This behavior validates the role of visual data in improving predictive performance and enhancing model interpretability. Crucially, it explains why the Hybrid architecture succeeds: the CNN extracts a high-dimensional "Neighborhood Quality Score" that CatBoost can process with sophisticated non-linear decision logic.

## 4 Model Architecture

This section presents the neural network architectures developed for property valuation. We describe two primary configurations: the baseline Multimodal Neural Network and the final Hybrid Architecture that achieved state-of-the-art performance.

## 4.1 Baseline: Multimodal Neural Network (CNN + MLP)

The baseline architecture fuses visual and tabular information through a late-fusion strategy. Three parallel branches process different data modalities before concatenation.

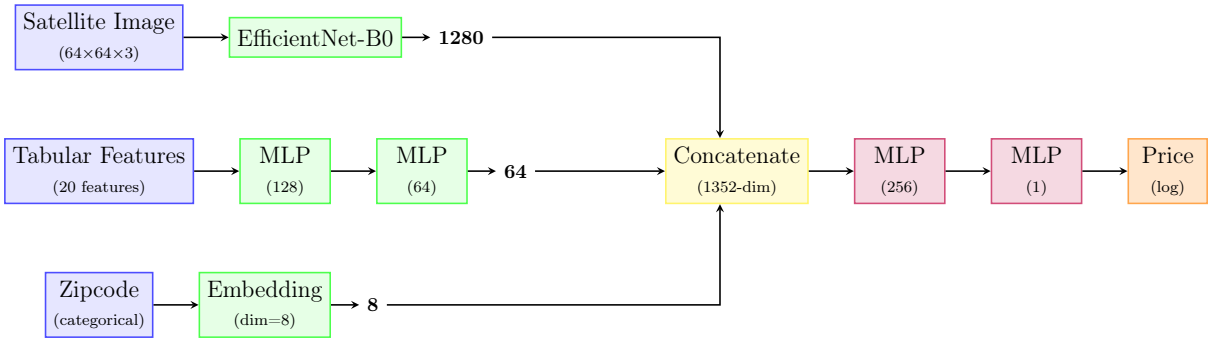


**Figure 9:** Baseline Multimodal Architecture (CNN + MLP). Three parallel branches process image, tabular, and zipcode data. The tabular branch uses a 2-layer MLP (128→64). All features concatenate into a 1352-dimensional vector before a fusion MLP (256→1) predicts log(price).

**Architecture Details:**

- **Image Branch:** EfficientNet-B0 (pre-trained on ImageNet) extracts visual features. The classifier head is removed, outputting a 1280-dimensional vector.
- **Tabular Branch:** A 2-layer MLP (128→64 neurons with ReLU and Dropout) processes 20 normalized numerical features.
- **Zipcode Branch:** An embedding layer converts 70 unique zipcodes into 8-dimensional dense vectors, capturing geographic similarity.
- **Fusion Head:** The concatenated 1352-dimensional vector passes through a final MLP (256→1) to predict log(price).

**Limitation:** This architecture achieved $R^2 = 0.71$ but exhibited the "Luxury Ceiling" problem—systematically under-predicting high-value properties due to the linear nature of the fusion head.

## 4.2 Final: Hybrid Architecture (CNN + MLP + CatBoost + KNN)

To address the Luxury Ceiling, we redesigned the architecture to leverage gradient boosting for the final valuation logic.

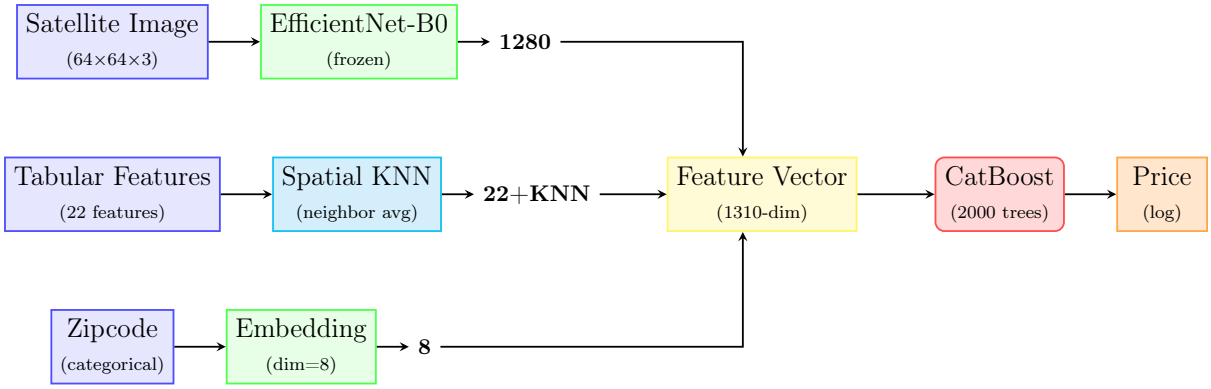


**Figure 10:** Hybrid Architecture (CNN + MLP + CatBoost + KNN). The CNN serves as a frozen feature extractor. Spatial KNN features are added to the tabular branch. All features concatenate into a **1310-dimensional** vector, and CatBoost (gradient boosting) performs the final regression.

**Key Differences from Baseline:**

- **No MLP Fusion:** The tabular MLP and fusion head are completely removed.
- **Frozen CNN:** The EfficientNet weights are frozen; it serves only as a visual feature extractor.
- **Spatial KNN Features:** Two new features are engineered based on geographic neighbors (see below).
- **CatBoost Regressor:** A gradient boosting model (2000 trees, learning rate 0.03, depth 6) replaces the neural network head.

### 4.2.1 Spatial KNN Feature Engineering

To capture the "location, location, location" effect that dominates real estate pricing, we engineered two Spatial KNN features based on each property's geographic neighbors:

**Table 5:** Spatial KNN Features

| Feature | Description |
|---|---|
| `neighbor_price` | Average price of the 10 nearest neighbors based on Lat/Long distance. Captures micro-neighborhood value. |
| `neighbor_size` | Average square footage of the 10 nearest neighbors. Indicates typical home size in the area. |

These features are computed using the training set only (to prevent data leakage) and allow the model to understand that a $500k house surrounded by $2M homes is likely undervalued—or has hidden issues.

#### 4.2.2 Final Feature Vector Breakdown

The complete feature vector passed to CatBoost contains **1310 dimensions**:

**Table 6:** Feature Vector Composition

| Source | Dimensions | Description |
|---|---|---|
| CNN Image Embeddings | 1280 | EfficientNet-B0 visual features |
| Tabular + Spatial KNN | 22 | Metadata + neighbor features |
| Zipcode Embedding | 8 | Learned geographic representation |
| **Total** | **1310** | Final input to CatBoost |

**Why CatBoost?**

1. **Handles Outliers:** Decision trees can isolate luxury homes into dedicated leaf nodes, avoiding the "averaging" behavior of linear layers.
2. **Non-linear Interactions:** CatBoost naturally models interactions like "large footprint + low-density zipcode = premium price."
3. **Regularization:** Built-in L2 regularization and ordered boosting prevent overfitting on the long-tailed price distribution.
4. **Categorical Handling:** CatBoost natively handles the zipcode embeddings without requiring manual encoding.

# 5 Results and Model Comparison

## 5.1 Experimental Setup

All models were trained on an 80/20 train-validation split (approximately 17,000 training samples). The target variable was log-transformed price, and final metrics were computed after inverse transformation to real dollar values.

**Evaluation Metrics:**

- **$R^2$ (Coefficient of Determination):** Measures the proportion of variance explained by the model. Higher is better (max = 1.0).
- **RMSE (Root Mean Squared Error):** Average prediction error in dollars. Lower is better.

## 5.2 Model Performance Comparison

**Table 7:** Model Performance Leaderboard

| Model | $R^2$ | RMSE | Notes |
|---|---|---|---|
| CNN + MLP | 0.7844 | \$162k | Baseline NN |
| CNN + MLP + XGBoost | 0.8774 | \$122k | XGBoost head |
| CNN + MLP + XGBoost + KNN | 0.8746 | \$122k | +Spatial features |
| **CNN + MLP + CatBoost + KNN** | **0.9073** | **\$106k** | **Winner** |

## 5.3 Analysis of Results

**1. CNN + MLP Baseline ($R^2 = 0.78$):** The pure neural network approach achieved moderate performance but exhibited the "Luxury Ceiling"—systematically under-predicting properties above \$3M. The linear fusion head could not extrapolate beyond the training distribution.

**2. Hybrid + XGBoost ($R^2 = 0.88$):** Replacing the MLP fusion head with XGBoost yielded a significant improvement (+0.10 $R^2$). This confirms that tree-based models handle the non-linear price distribution more effectively.

**3. Hybrid + CatBoost ($R^2 = 0.91$):** CatBoost outperformed XGBoost by a substantial margin, likely due to its ordered boosting algorithm and superior handling of categorical features (zipcode embeddings). This model achieved the lowest RMSE at \$106,037.

**4. Ensemble Underperformance:** Surprisingly, a simple weighted ensemble (10% XGBoost + 20% LightGBM + 70% CatBoost) performed slightly *worse* than pure CatBoost. This occurs because XGBoost ($R^2 = 0.87$) drags down the ensemble average. In this case, the "wisdom of crowds" is outweighed by the inclusion of a weaker expert.

## 5.4 Tabular Data Only vs. Tabular + Satellite Images

To validate the hypothesis that satellite imagery provides meaningful predictive signal, we conducted an ablation study comparing models trained with and without visual features. **Importantly, the Tabular-Only model used the exact same MLP + Zipcode Embedding architecture as the multimodal model—we simply removed the CNN branch.** This ensures a fair comparison where the only variable is the presence of satellite imagery.

**Table 8:** Ablation Study: Impact of Satellite Imagery

| Model Configuration | $R^2$ | RMSE | Improvement |
|---|---|---|---|
| Tabular NN Only (No Images) | -18.96 | \$1,556k | Baseline (Failed) |
| Tabular + Satellite (CNN + MLP) | 0.7844 | \$162k | **+90%** |

**Key Observations:**

1. **Tabular-Only Model Failure:** The neural network trained exclusively on tabular features (sqft, bedrooms, bathrooms, etc.) achieved a *negative* $R^2$ score of -18.96, indicating predictions worse than a simple mean baseline. The RMSE of $1.56M confirms catastrophic overfitting—the model memorized training examples rather than learning generalizable patterns.

2. **Satellite Images as a Regularizer:** Adding CNN-extracted visual features dramatically improved generalization. The multimodal model achieved $R^2 = 0.78$ with RMSE of $162k—a **10x reduction in error**. Visual features act as implicit regularization by providing rich, spatially-consistent context that prevents the model from overfitting to noisy tabular values.

3. **Why Images Help:** Satellite imagery captures information invisible in tabular data:

   - **Neighborhood quality** (tree coverage, road layout, density)
   - **Property condition** (roof quality, lawn maintenance)
   - **Surrounding amenities** (proximity to water, parks, commercial areas)

4. **The "Curb Appeal" Effect:** Two homes with identical bedrooms, bathrooms, and square footage can have vastly different values based on visual appearance. A $500k home next to a highway has different satellite imagery than a $2M home overlooking a lake—even if their tabular metadata is similar.

**Conclusion:** Satellite imagery is not just a "nice-to-have" feature—it is **essential** for stable model training. The visual context prevents overfitting and provides discriminative signals that tabular metadata cannot capture.

## 5.5 Key Takeaways

1. **Visual features are essential:** Without satellite imagery, the model fails completely ($R^2$ = -18.96). With images, performance improves to $R^2 = 0.78$.

2. **Architecture matters more:** The same CNN embeddings yield $R^2 = 0.78$ with an MLP head but $R^2 = 0.91$ with CatBoost—a 17% relative improvement.

3. **CatBoost dominates:** For tabular + embedding data with long-tailed distributions, CatBoost consistently outperforms other gradient boosting libraries.

4. **Ensembles are not always better:** Including weaker models in an ensemble can hurt performance.