Satellite Imagery-Based Property Valuation

PRIYANSHU NATH 23119029

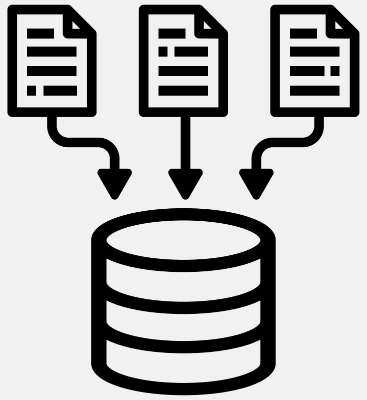
Overview

Objective :

The objective of this project is to develop a multimodal regression pipeline that predicts residential property prices by integrating tabular housing data with satellite imagery derived from geographic coordinates.

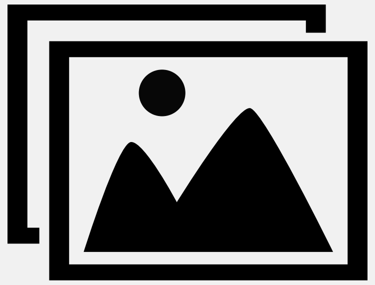
The system aims to extract meaningful visual representations of a property's environment neighborhood characteristics like green cover or road density and integrate it with traditional pricing models.

By bridging structured data and visual context, this project seeks to demonstrate the efficacy of multimodal learning in real-world real estate valuation compared to traditional approaches.



Tabular Data

The tabular dataset comprises structured numerical features describing intrinsic property attributes, including living area, bedroom and bathroom counts, construction grade, condition, and geographic coordinates. These variables represent the conventional feature space used in real estate price modeling. The target variable is the property sale price, transformed using a logarithmic function to address skewness and heteroscedasticity.



zoom 16 zoom 17 zoom 18

Satellite Image Data

Satellite imagery is programmatically acquired based on property latitude–longitude coordinates to capture spatial and environmental context. Images are collected at multiple zoom levels to encode both fine-scale structural details and broader neighborhood-level patterns. Visual features are extracted using a pretrained convolutional neural network, enabling integration of spatial information into the valuation model.

Overview

Modeling strategy:

- **Data Cleaning:** Conducted statistical analysis to analyse trends and patterns in the dataset.
- **Feature Scaling:** Applied standard scaling and normalization techniques to numeric features. This step was crucial to ensure convergence and prevent features with larger magnitudes from dominating the gradient boosting models.

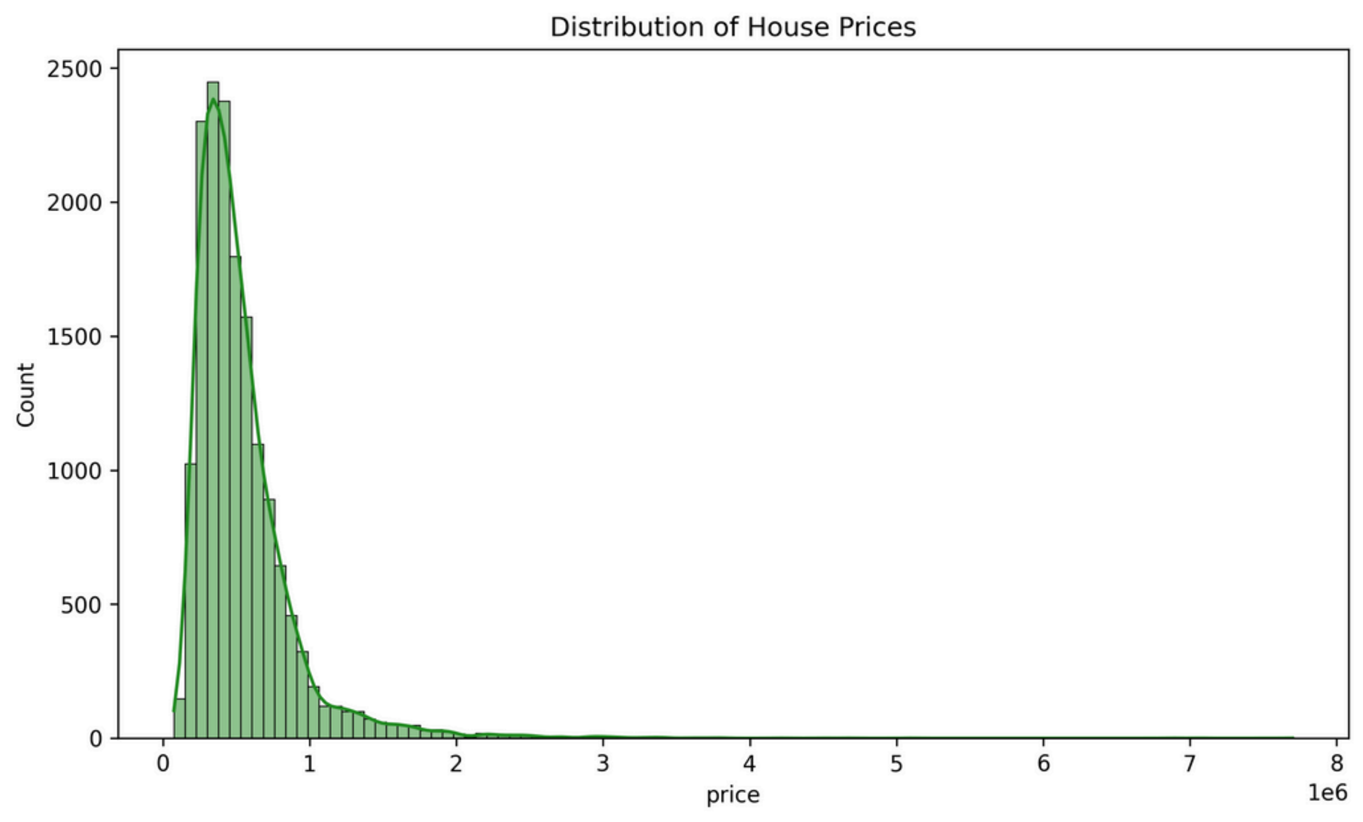
- **Model Selection:** Selected XGBoost and CatBoost as the primary regressors due to their superior handling of tabular data and resistance to overfitting.
- **Tuning Strategy:** Instead of using default parameters, we performed extensive Hyperparameter Tuning using RandomizedSearchCV. This allowed us to efficiently explore the search space for critical parameters (such as learning rate, tree depth, and L2 regularization) to minimize the Root Mean Square Error (RMSE).

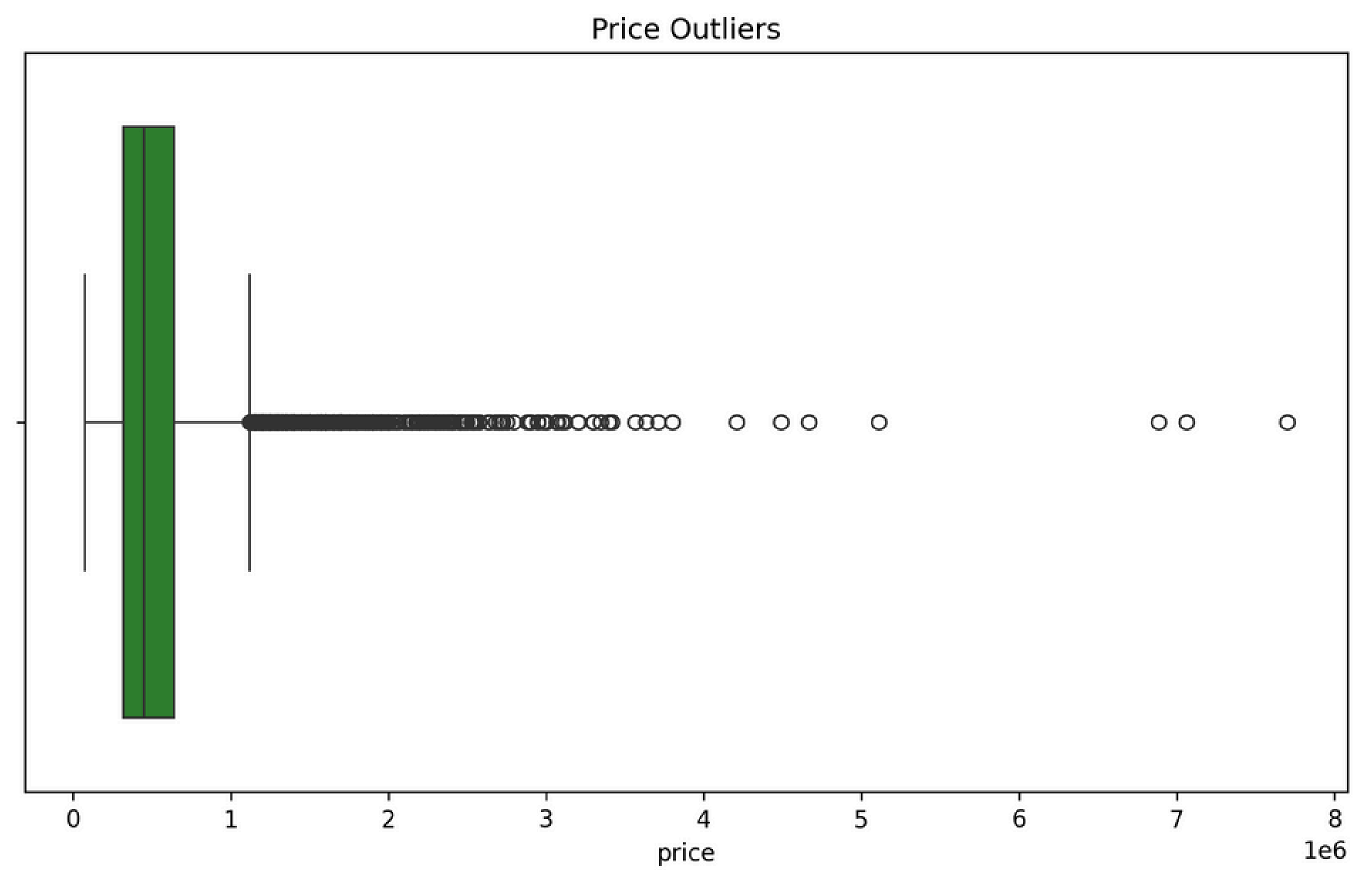
TABULAR PREPROCESSING → **DEEP VISUAL FEATURE EXTRACTION** → **HYPERPARAMETER OPTIMIZATION** → **LATE FUSION & FINAL PREDICTION**

- **Architecture:** Leveraged ResNet50, a Deep Convolutional Neural Network (CNN) pre-trained on ImageNet, utilizing Transfer Learning to extract high-level semantic features.
- **Multi-Scale Strategy:** Processed satellite imagery at three distinct zoom levels:
  - Zoom 16 (Macro): Neighborhood layout and green cover.
  - Zoom 17 (Street): Road proximity and local amenities.
  - Zoom 18 (Micro): Property-specific details (roof type, immediate vegetation).
- To prevent overfitting and ensure robust feature extraction, the deep learning architecture incorporated **Batch Normalization layers** (for stable gradient flow) and **Dropout layers** (to mitigate neuron co-adaptation).
- **Utilized dynamic Callbacks** (such as EarlyStopping and ModelCheckpoint) during the extraction/fine-tuning phase to monitor validation loss and save the optimal model weights.

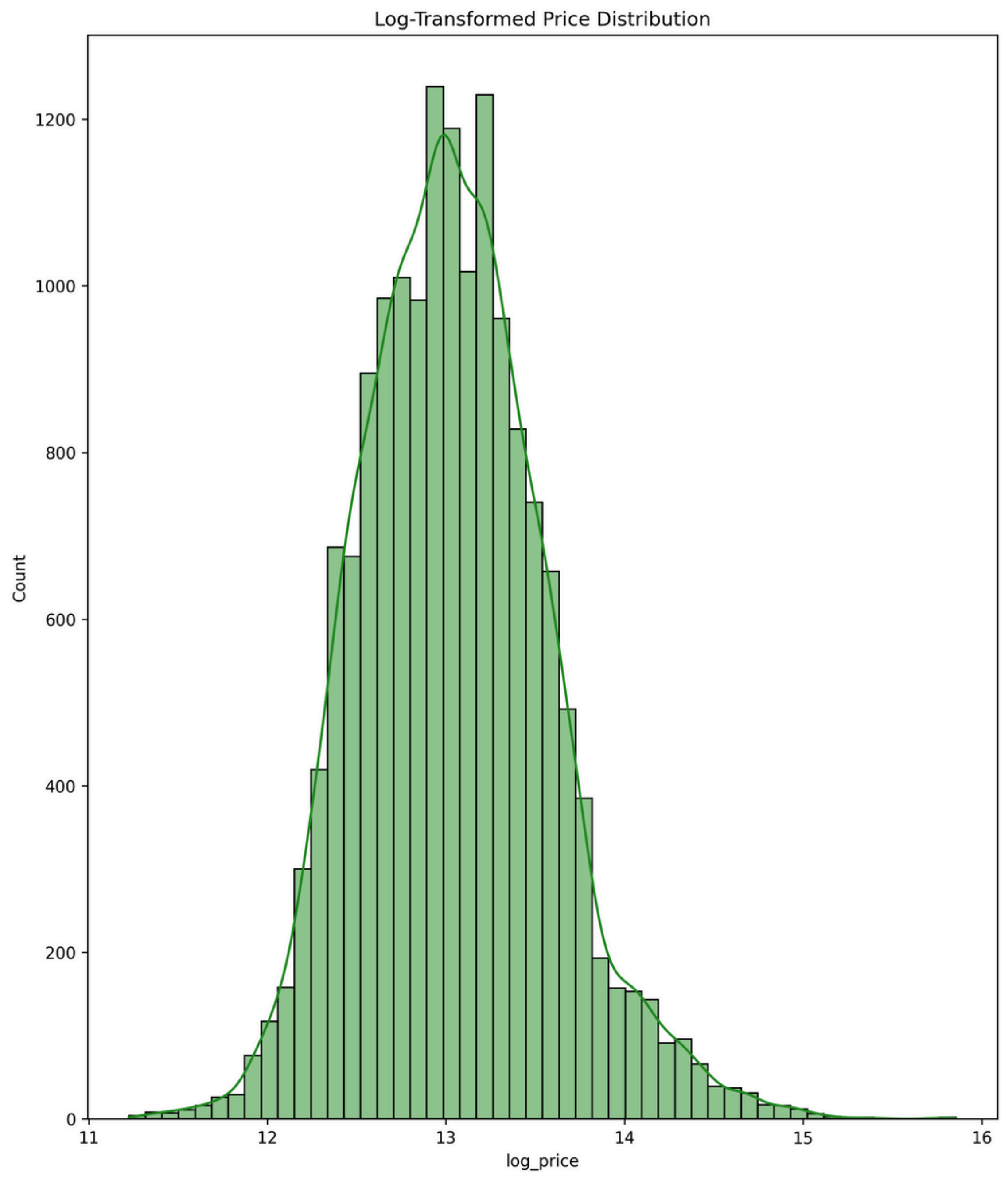
- **Fusion Mechanism:** Adopted a Late Fusion strategy where the high-dimensional visual embeddings (from the CNN) were concatenated with the processed tabular features to create a comprehensive "Master Feature Vector."
- **Final Regression:** The optimized CatBoost model was used to predict the price of the test dataset as it outperforms the fusion model due to its dominance on tabular features.

Exploratory Data Analysis





- House prices are heavily right-skewed, indicating a long luxury tail that inflates variance and motivates log-price transformation for modeling.
- A small number of extreme high-value properties act as leverage points, necessitating robust modeling or outlier handling to avoid bias.

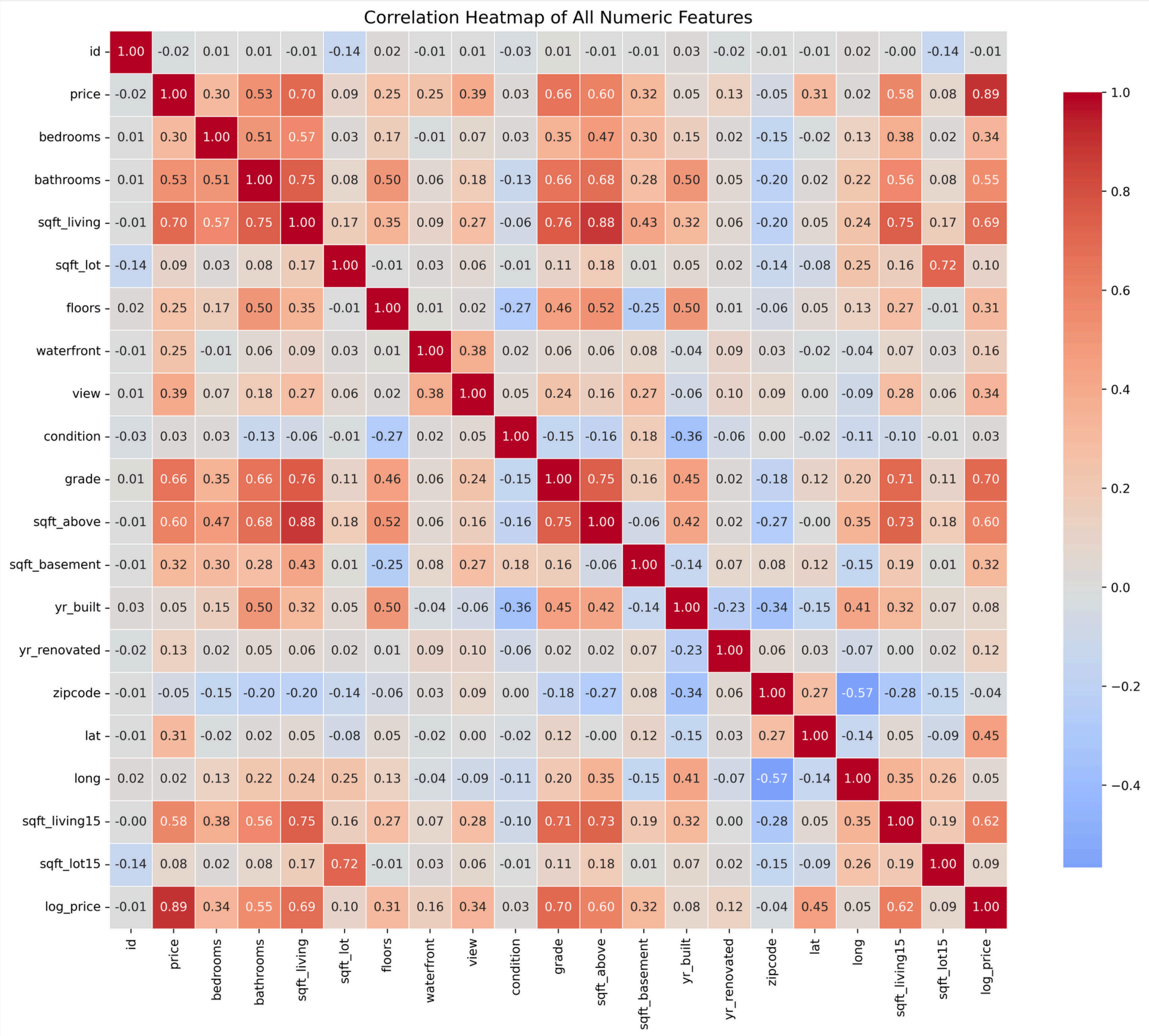


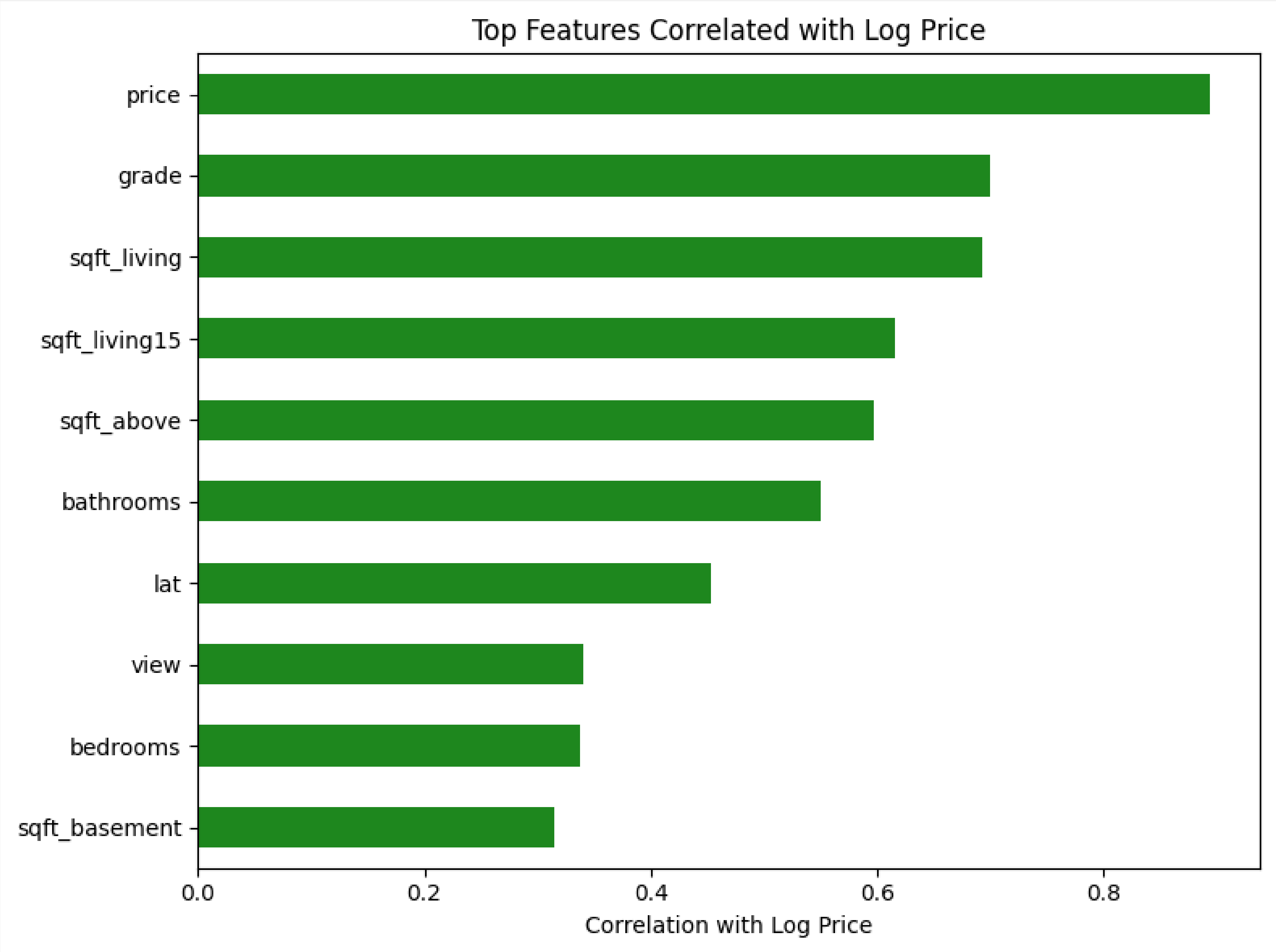
Log transformation produces a near-normal distribution.

This improves :

1. Gradient Stability
2. Convergence
3. Robustness to outliers

Exploratory Data Analysis



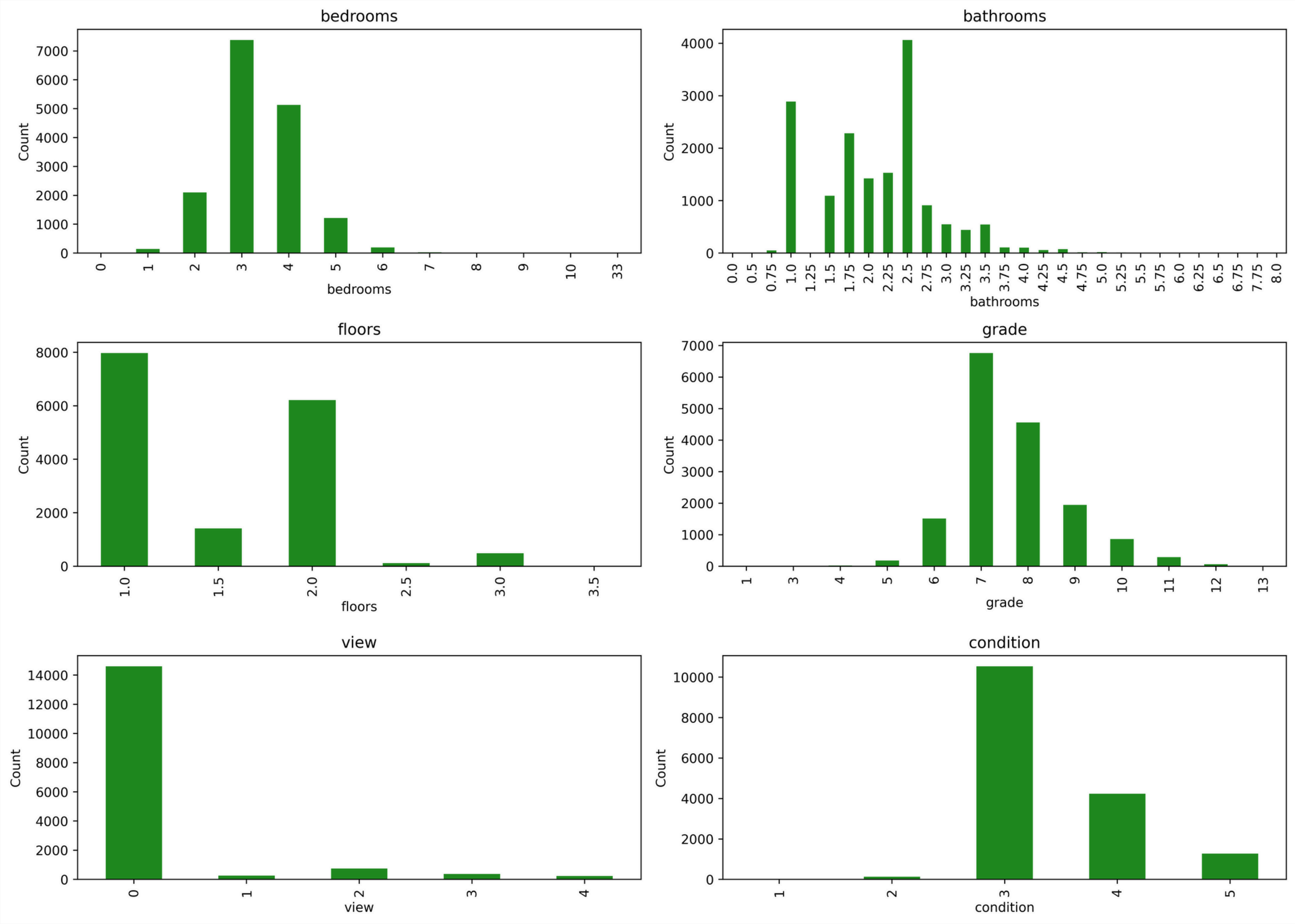


Size and quality dominate pricing: sqft_living and grade exhibit the strongest linear correlation with log price, confirming built area and construction quality as primary valuation drivers.

Spatial effects are significant but non-linear: Latitude and longitude show moderate correlation, indicating location effects that are better captured via spatial or image-based features rather than linear terms alone.

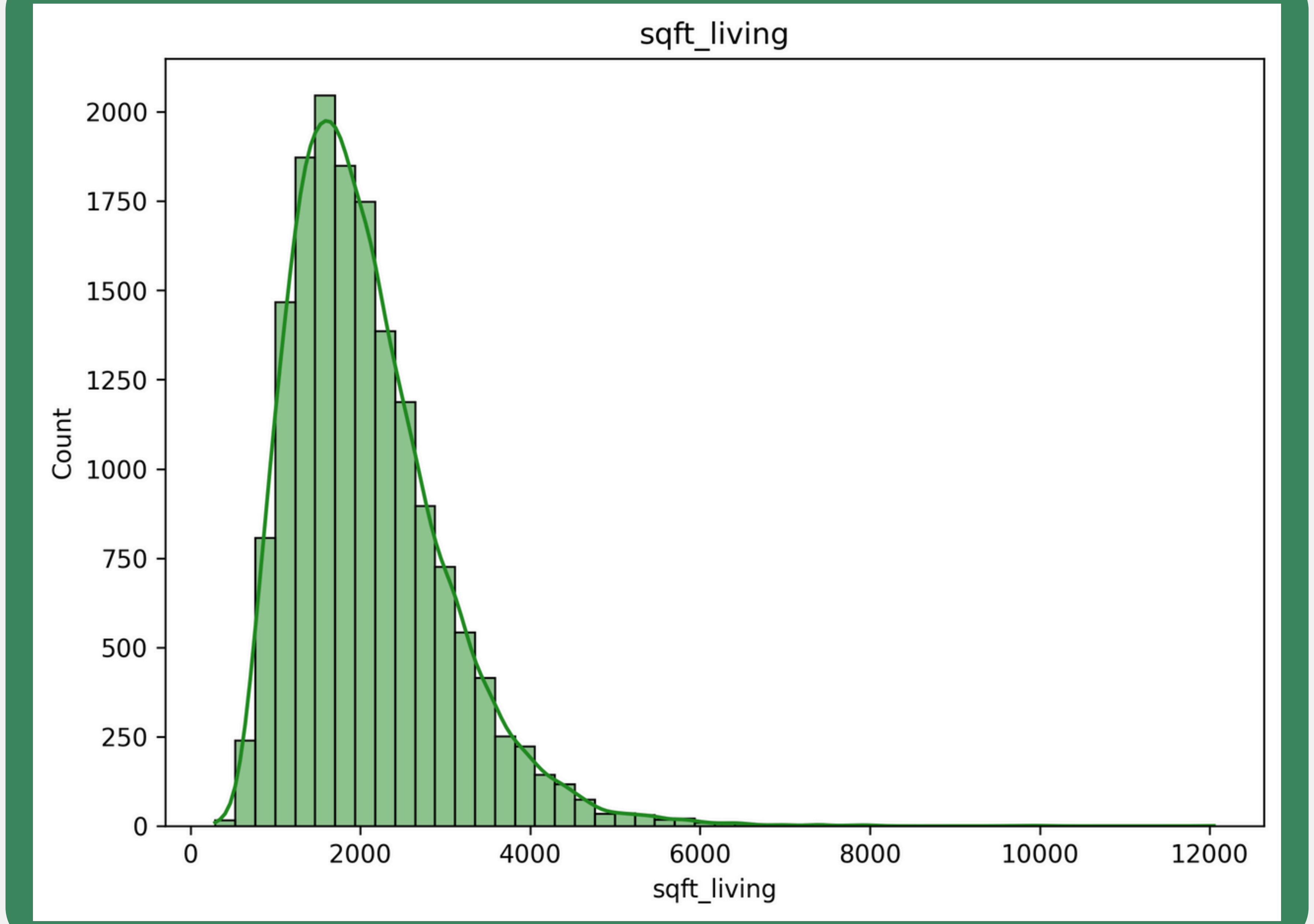
Amenity features are weakly linear: Variables like view and condition display limited linear correlation, suggesting their impact is interaction-driven or non-linear and may require advanced models (e.g., tree-based or CNN-enhanced).

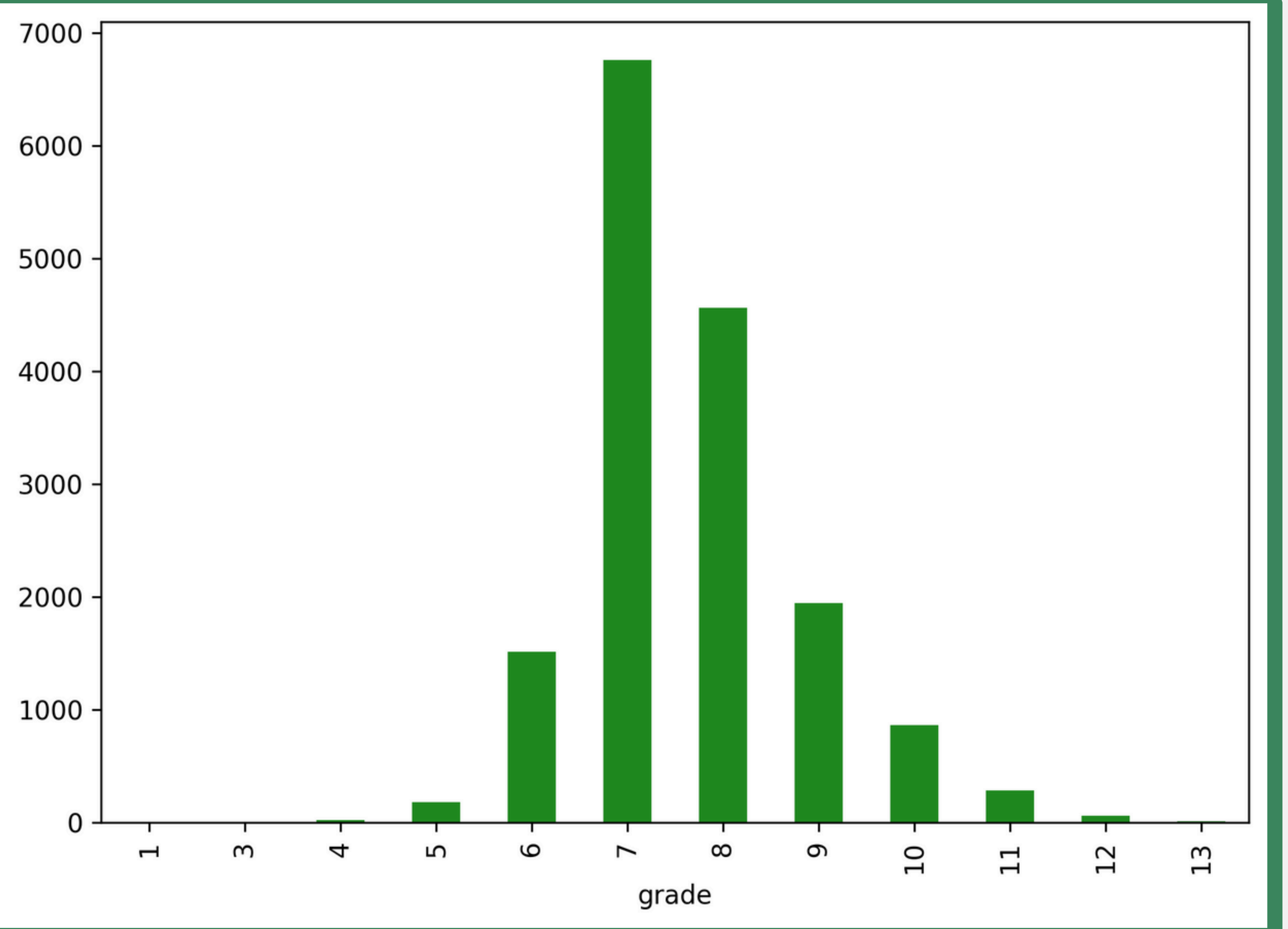
Exploratory Data Analysis



- **Structural features (bedrooms, bathrooms, floors)** show strong central clustering, indicating a largely homogeneous housing stock dominated by standard family homes.
- **Quality indicators (grade, condition)** are right-skewed toward average-to-above-average values, suggesting construction quality is a key differentiator rather than basic structure.
- **Amenity features (view)** are highly imbalanced, with premium views being rare, reinforcing their role as high-impact but low-frequency price drivers.
-

Exploratory Data Analysis

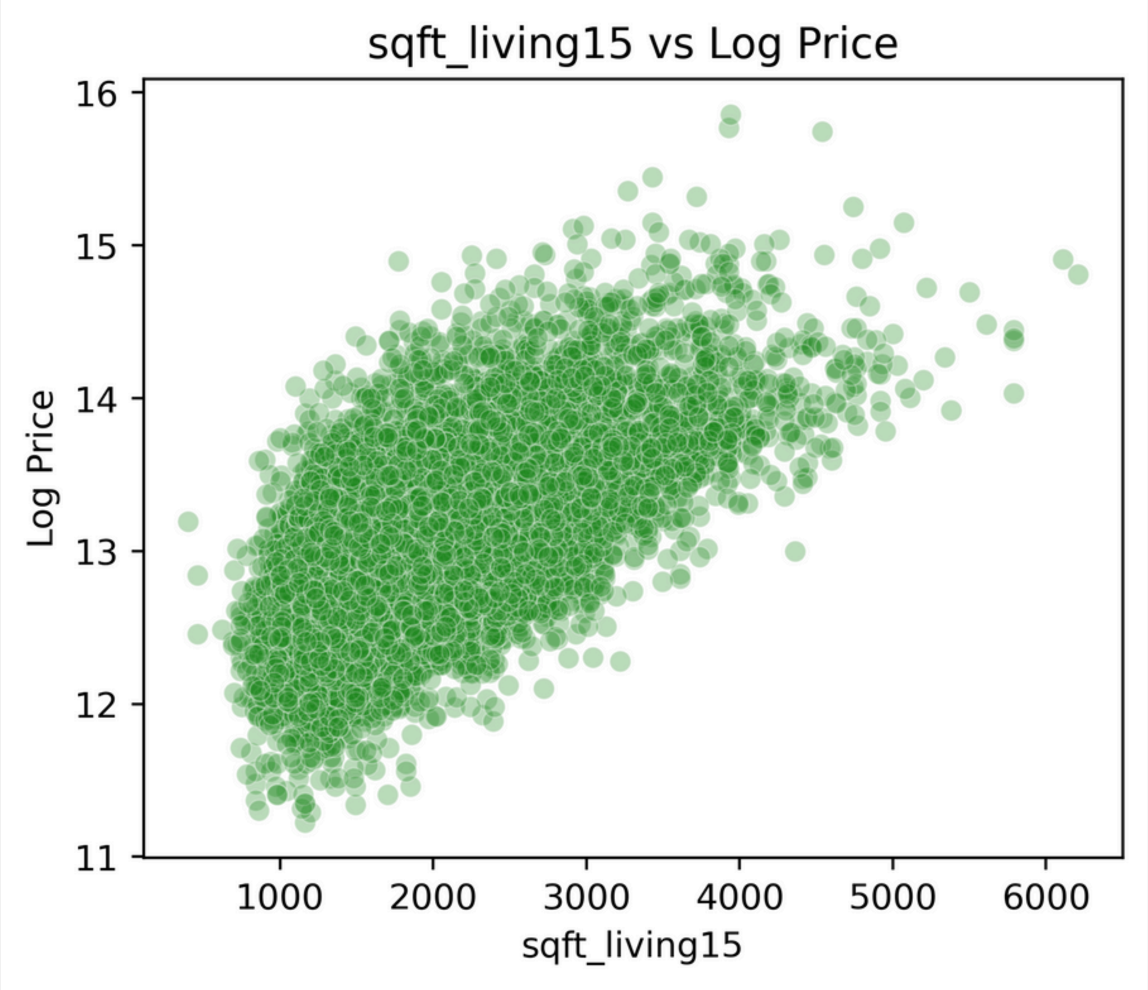


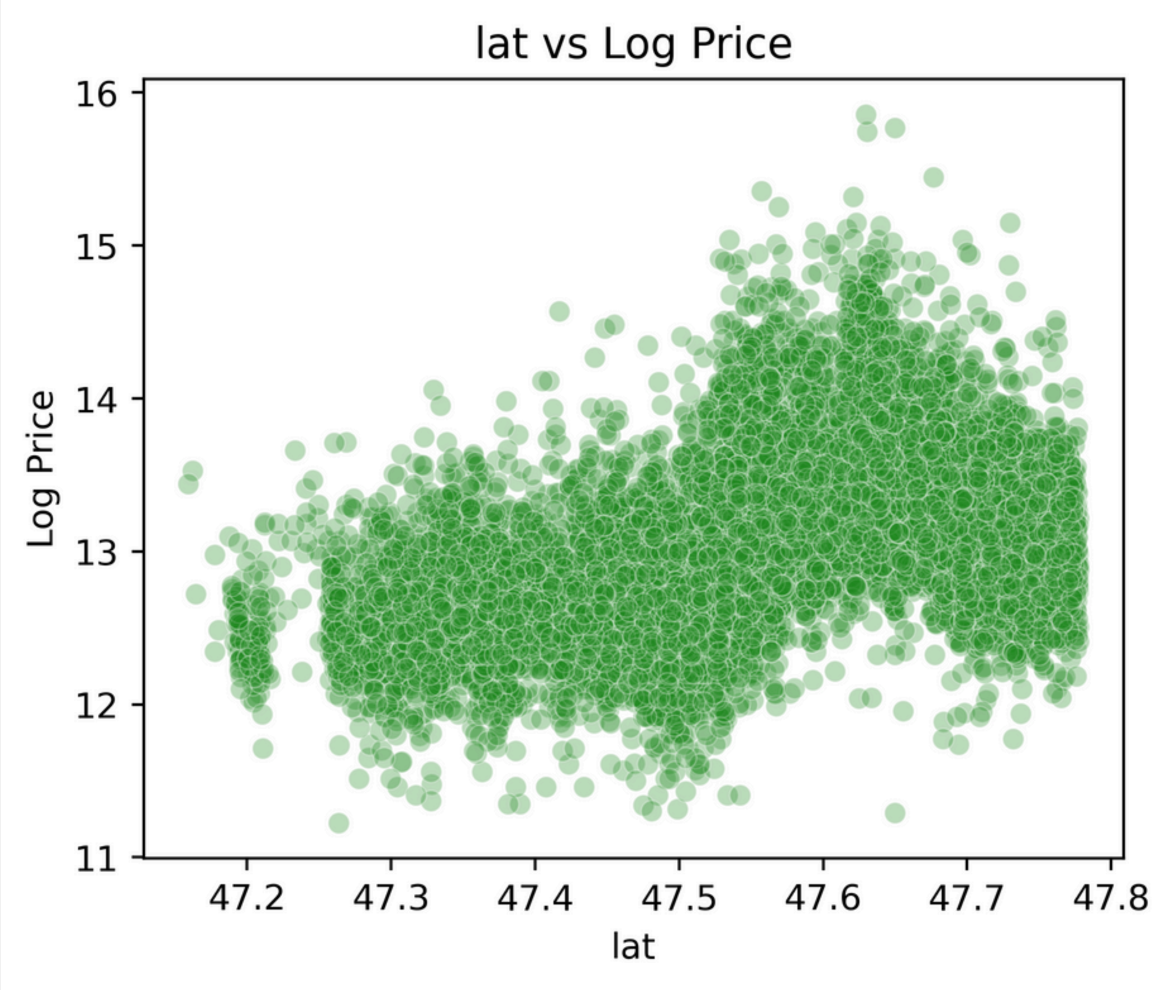


Strong right-skewed distribution indicates most homes are mid-sized, with a small tail of large properties driving high-end price variability.

Concentration around grades 7–8 reflects standardized construction quality, with higher grades acting as discrete upward valuation shifts.

Exploratory Data Analysis



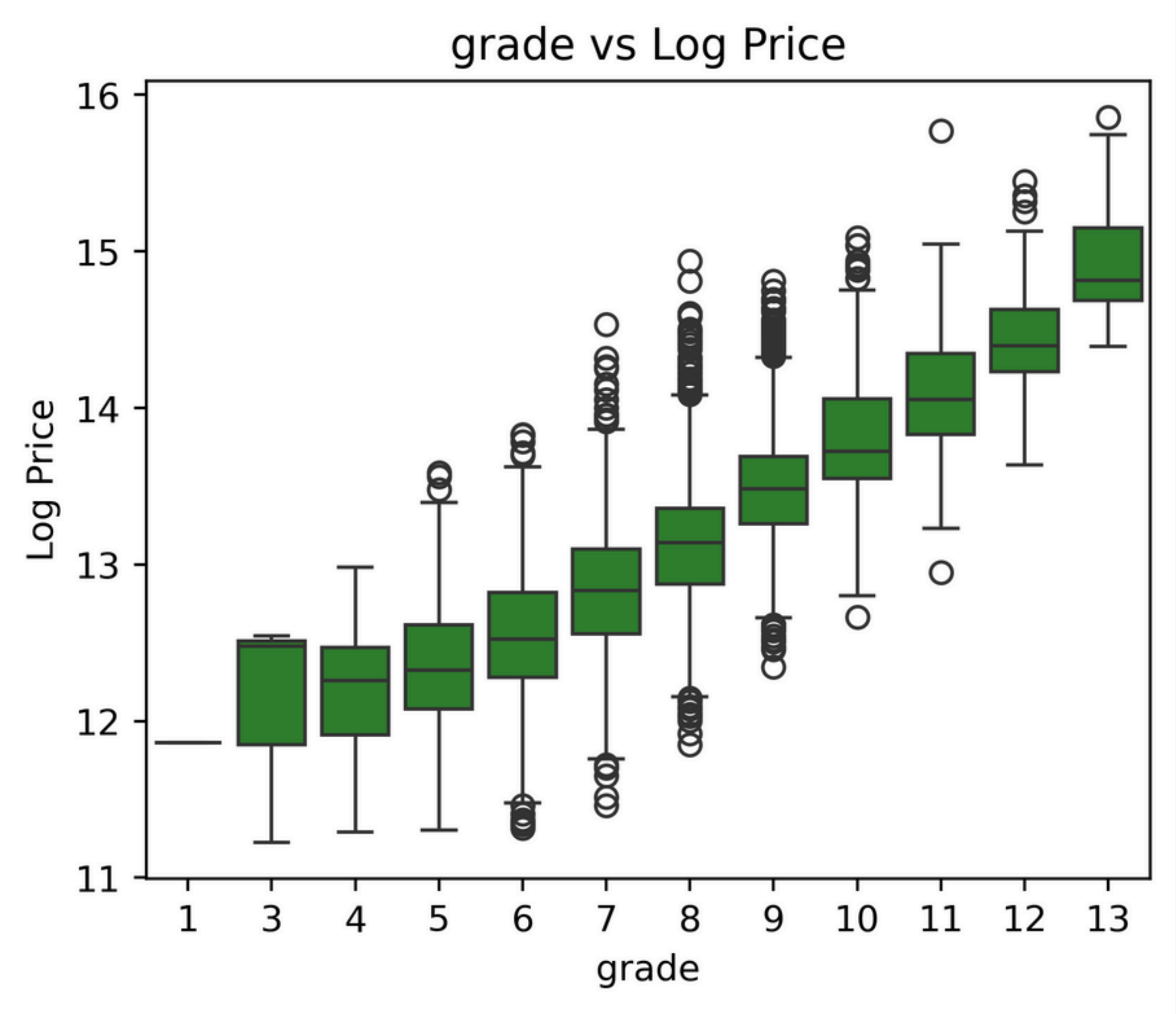


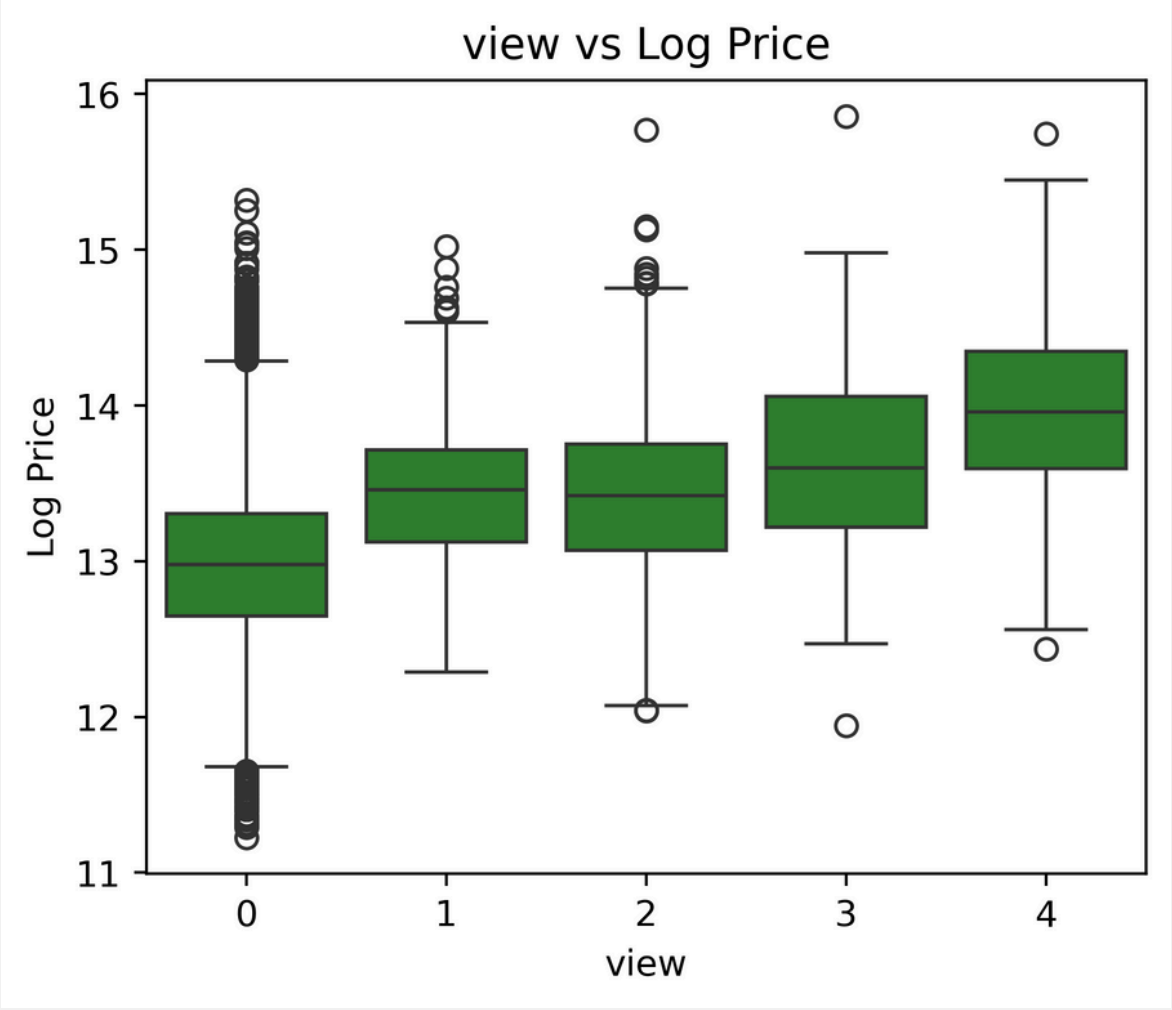
Living area shows a strong positive relationship with log-price, with diminishing returns for very large houses, indicating that price growth slows beyond a certain size.

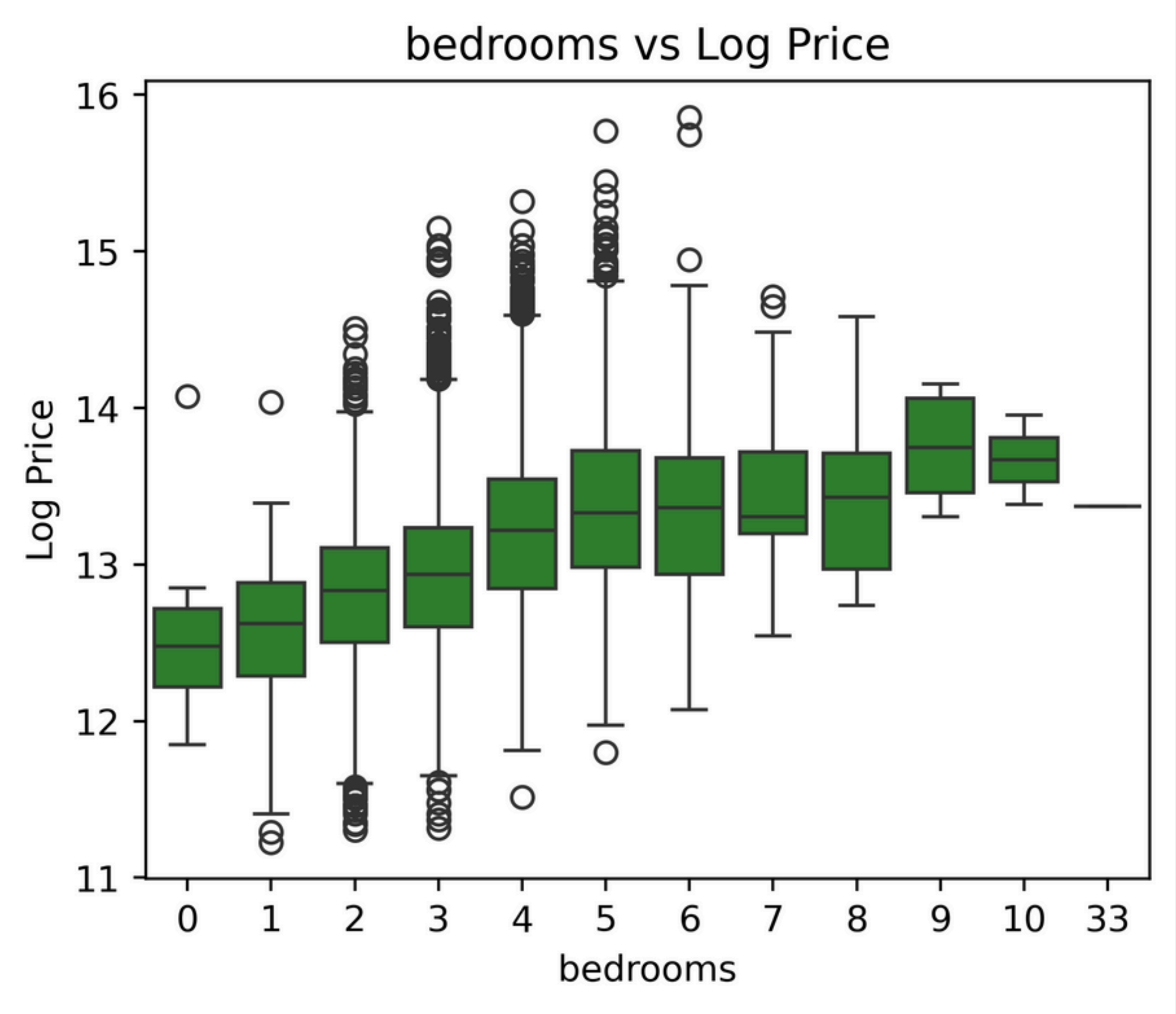
Log price increases with surrounding living area (sqft_living15), indicating a clear positive neighborhood size effect on property value.

Non-linear spatial dependence observed: log prices cluster at higher latitudes, indicating strong location-specific valuation effects.

Exploratory Data Analysis





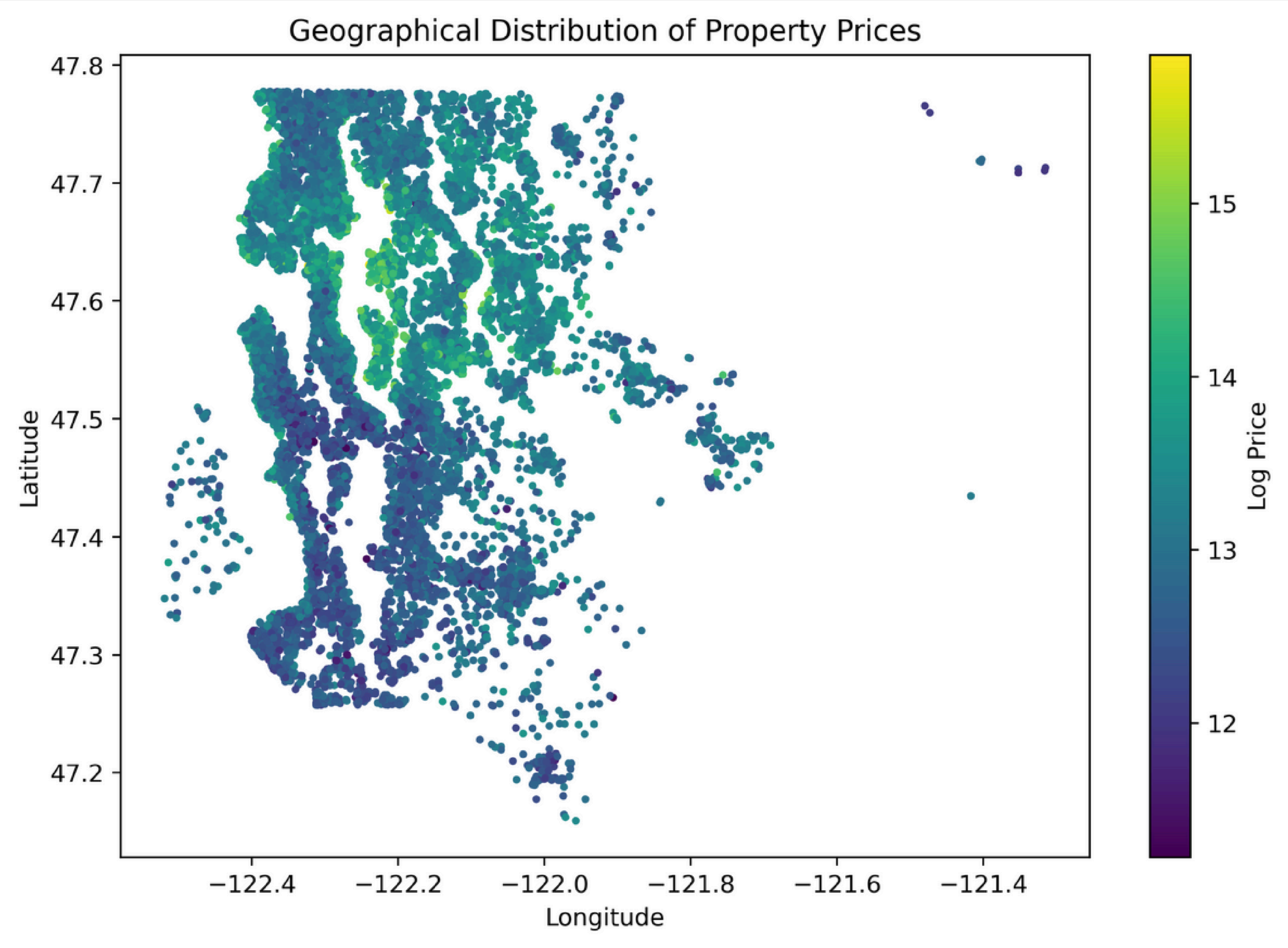


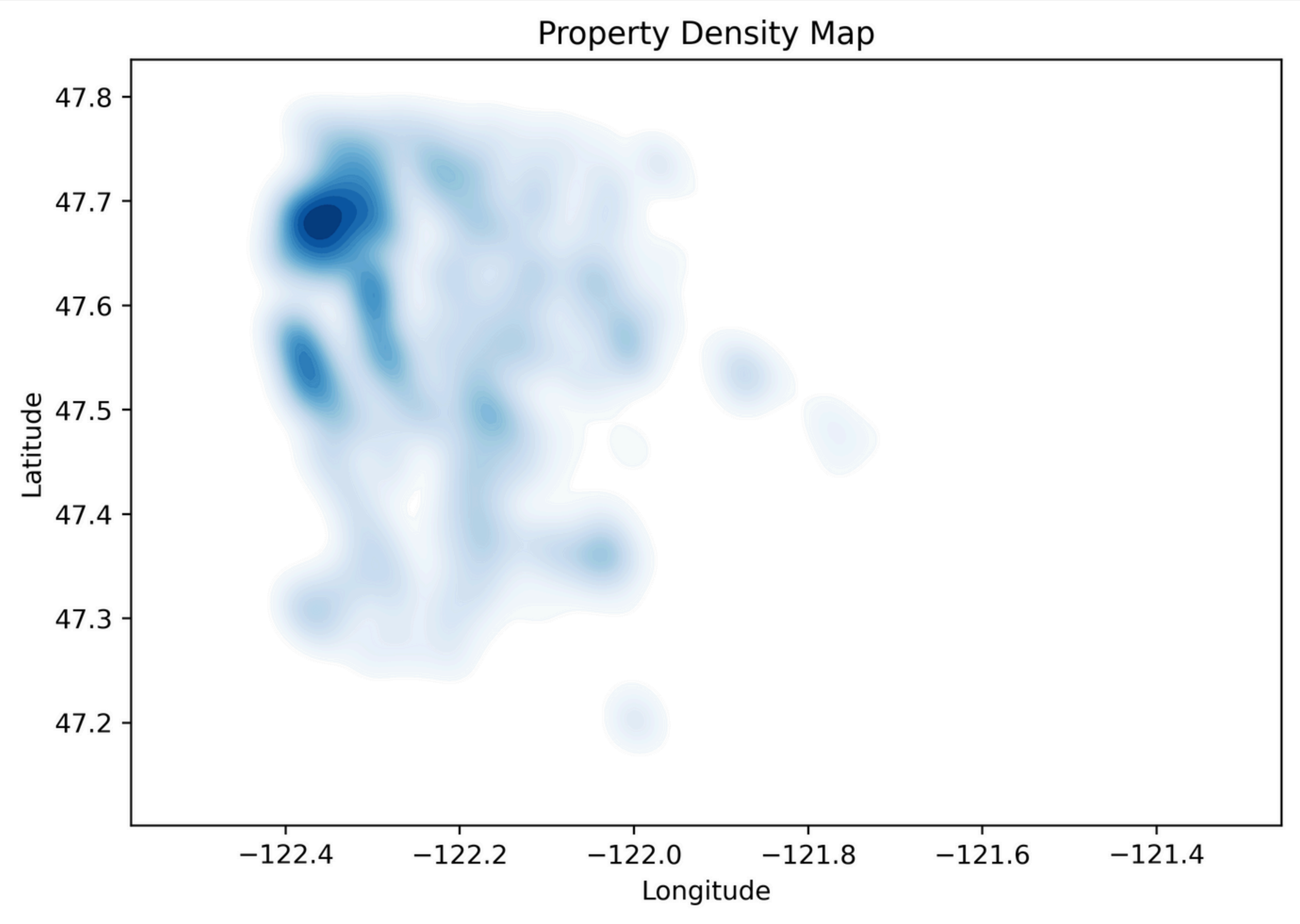
Monotonic increase in median log price with grade indicates strong positive correlation between construction quality and property valuation.

Higher view ratings correspond to higher median log prices, reflecting significant amenity-based valuation effects

Log price increases with bedroom count up to a point, after which marginal gains diminish, indicating non-linear returns to additional bedrooms.

Exploratory Data Analysis





Property prices exhibit clear spatial clustering, with higher log prices concentrated in specific latitude–longitude corridors, highlighting strong localized market effects.

Residential properties are unevenly distributed across space, with pronounced density hotspots indicating spatial heterogeneity that must be accounted for in valuation models.

Financial / Visual Insights

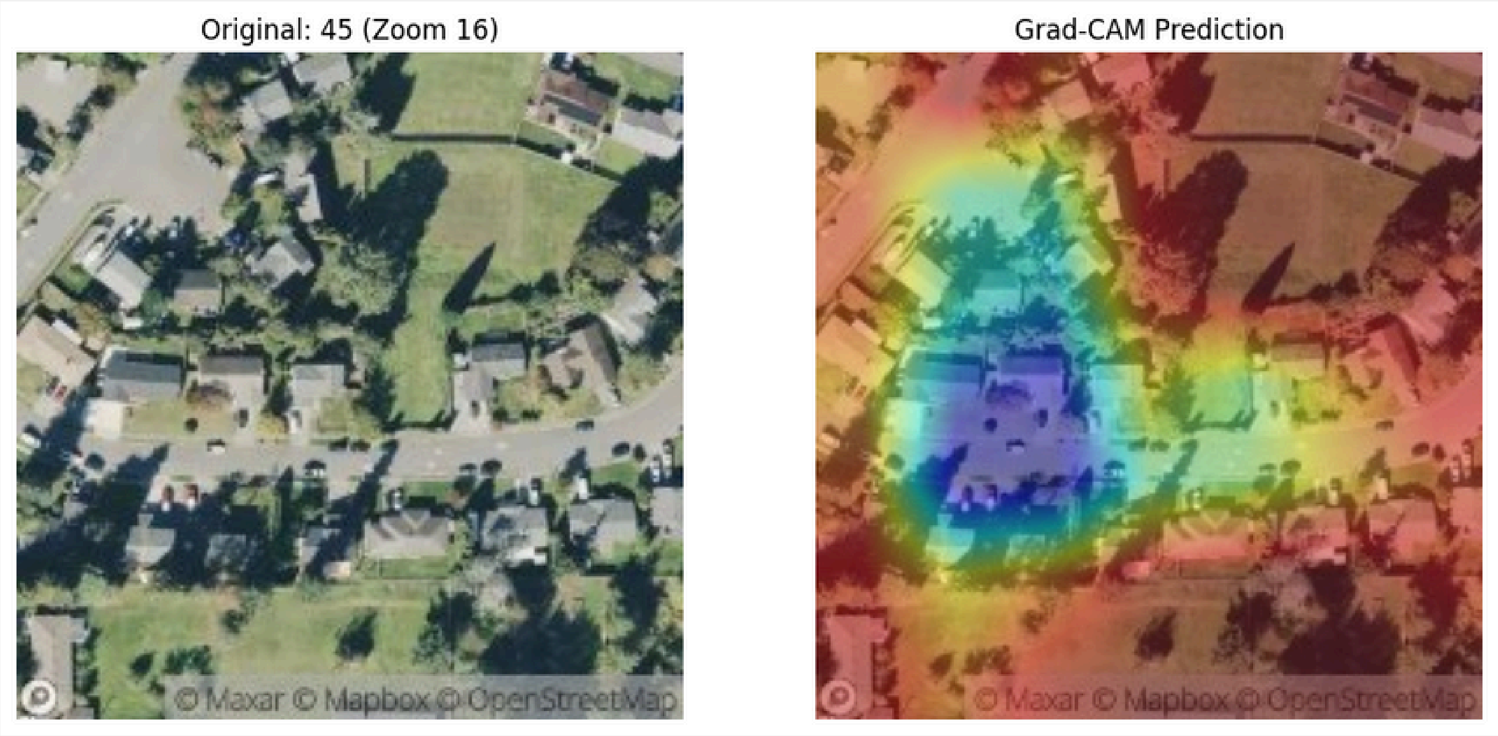
Low-Priced Properties

- Located in visually less structured neighborhoods with irregular layouts and limited spatial planning.
- Surroundings show sparse amenities or proximity to major roads, indicating lower residential desirability.
- Environmental context appears utility-driven rather than lifestyle-oriented, reflected in muted visual quality.
- Visual patterns remain consistent across zoom levels, reinforcing systemic neighborhood disadvantages.

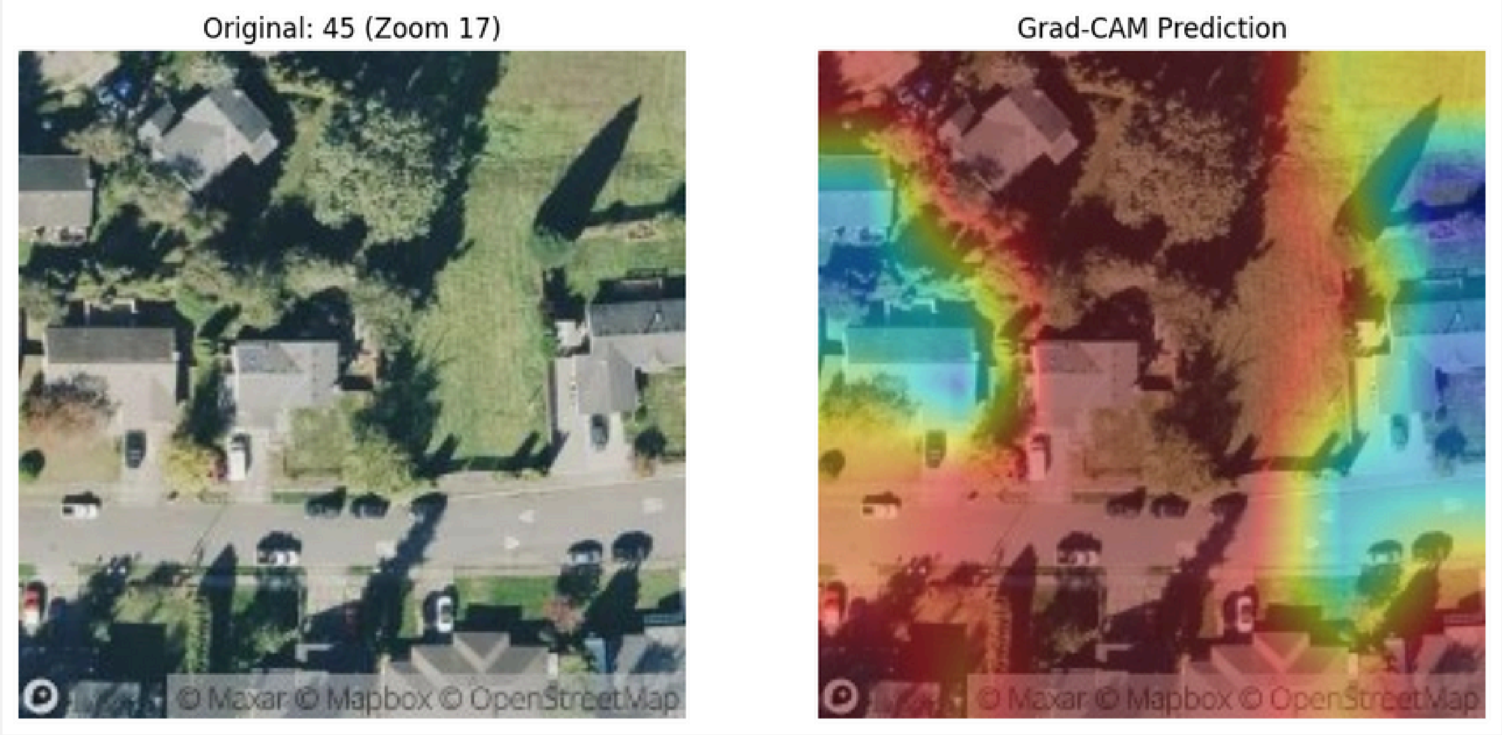
High-Priced Properties

- Situated in well-planned neighborhoods with organized road networks and clear plot boundaries.
- Consistently exhibit high vegetation density, open spaces, and premium amenities such as waterfront access.
- Visual environment reflects low congestion and high livability, aligning with luxury valuation signals.
- Multi-scale consistency across zoom levels indicates strong neighborhood-level price premiums.

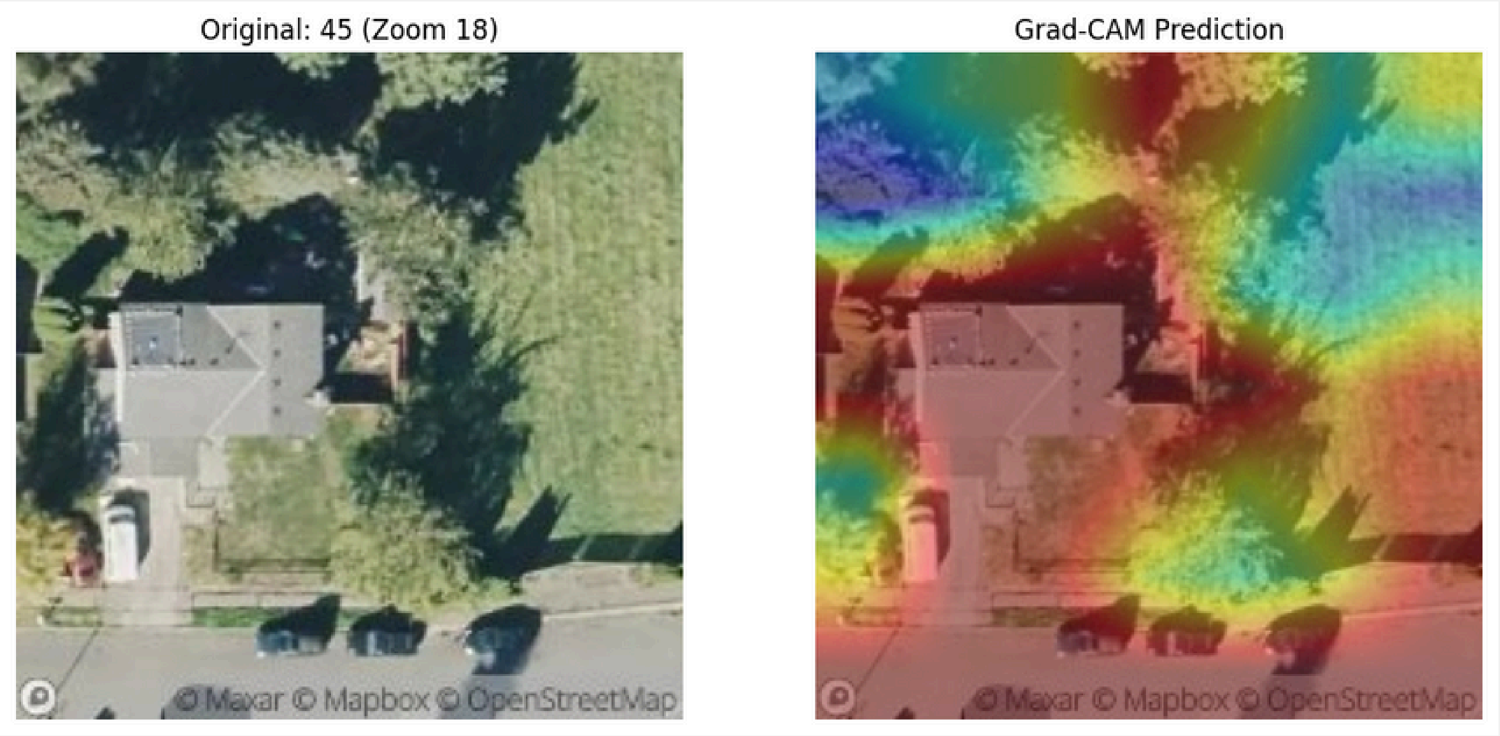
Grad-CAM Analysis



Zoom 16
Model attention is distributed across the broader residential cluster, indicating sensitivity to neighborhood layout, open spaces, and surrounding land use.

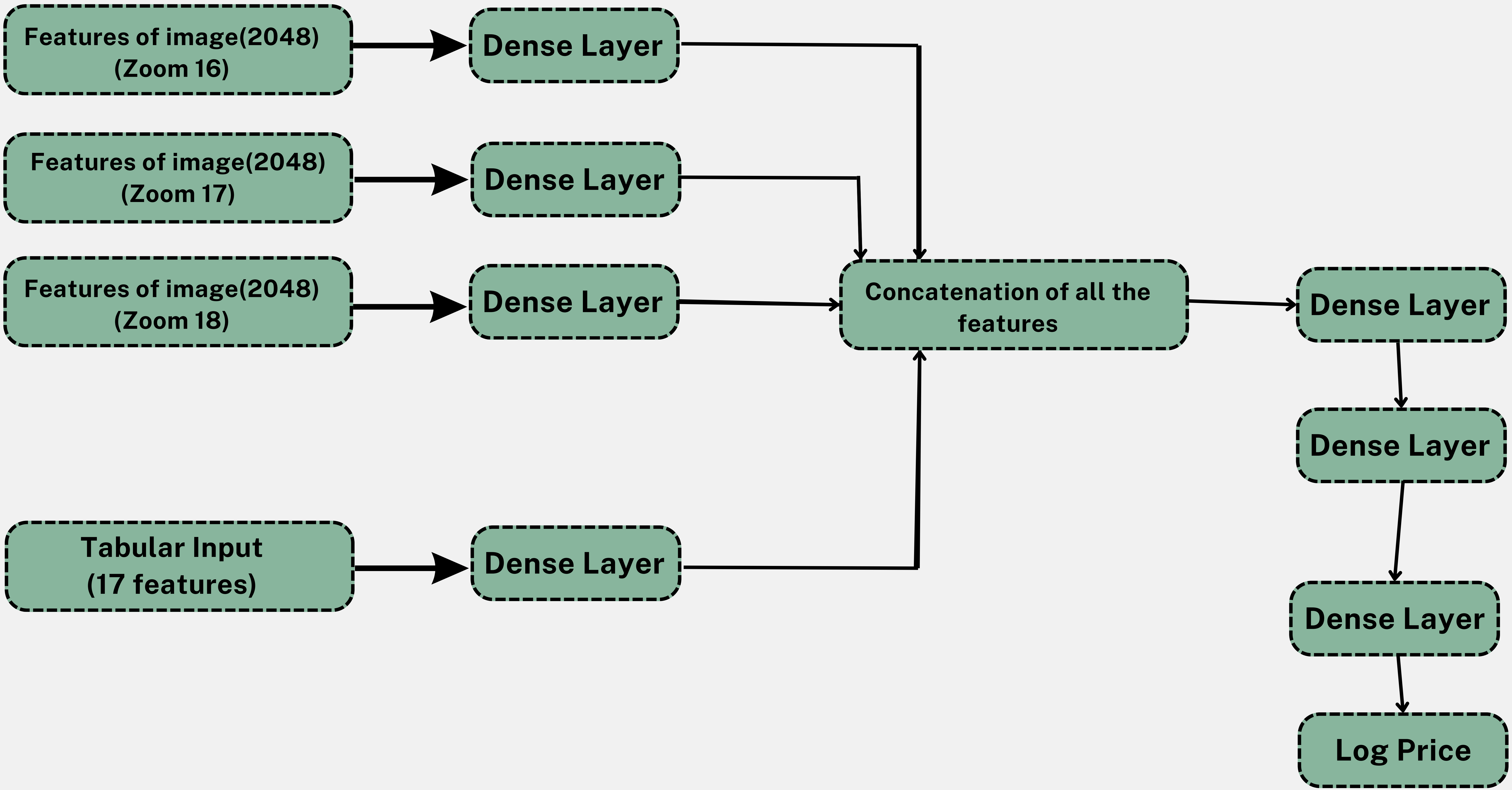


Zoom 17
Activation concentrates around adjacent buildings, road connectivity, and nearby greenery, highlighting the importance of immediate neighborhood infrastructure.



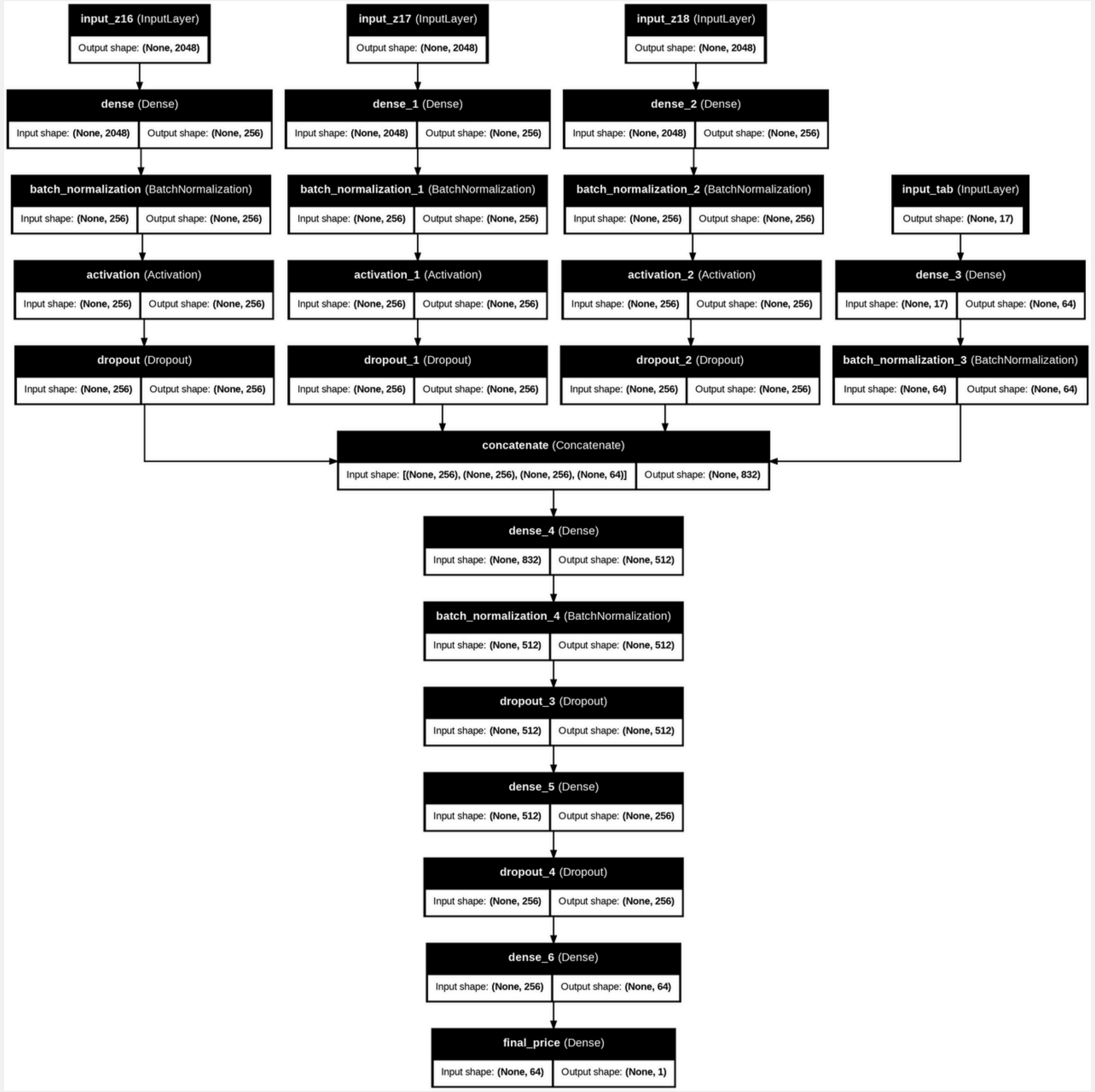
Zoom 18
Grad-CAM strongly focuses on the property footprint and surrounding vegetation, suggesting that fine-grained structural and environmental features directly influence valuation.

Architecture Diagram



Architecture Diagram

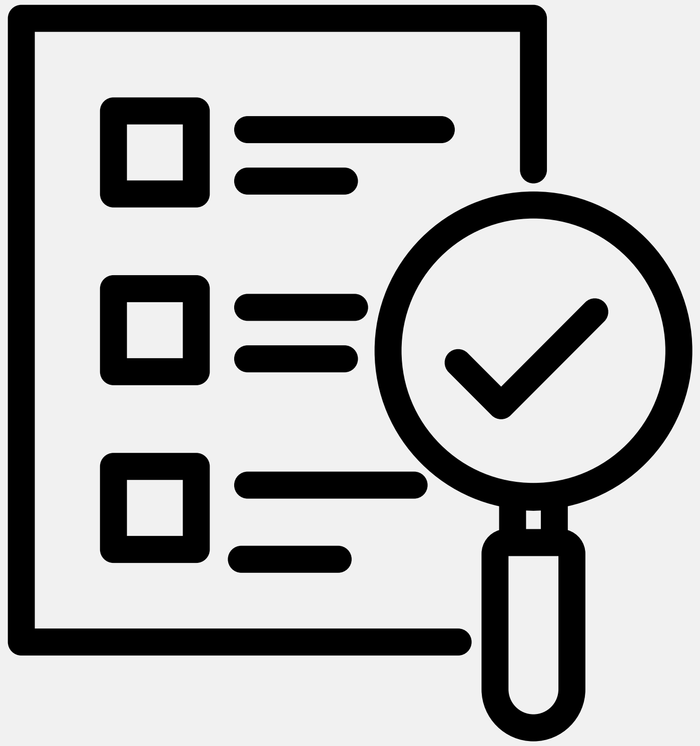
- **Pre-Extracted Visual Embeddings:** Leveraged ResNet50 to extract 2048-dimensional feature vectors offline for all three zoom levels, allowing the model to directly ingest high-level semantic representations without the computational overhead of retraining the CNN backbone.
- **Multimodal Late Fusion:** Implemented a neural network that processes the pre-computed image features (Zoom 16, 17, 18) and tabular data via separate input branches before merging them for the final prediction.
- **Feature Compression Towers:** Designed dedicated "Image Towers" to compress the high-dimensional ResNet embeddings (2048) into compact 256-sized vectors using L2 regularization to retain only essential visual signals.
- **Robust Non-Linearity:** Utilized ELU (Exponential Linear Unit) activations and Batch Normalization throughout the deep fusion block (512 → 256 → 64 neurons) to ensure stable gradients and faster convergence.
- **Regularization Strategy:** Integrated aggressive Dropout (0.3–0.4) and L2 penalties to specifically target and prevent overfitting given the high dimensionality of the combined feature set.
- **Dynamic Optimization:** Employed the Adam optimizer with an adaptive training loop, using ReduceLROnPlateau to lower learning rates when stuck and EarlyStopping to halt training automatically when validation loss stabilized.



ACTUAL NETWORK plotted using Keras

Results

Model	RMSE	R-squared (R2)
Catbooost	114044.98	0.8964
XgBoost	115149.76	0.8943
Fusion Model	132897.39	0.8593



- **Tree-based models outperform due to strong tabular signals:** CatBoost and XGBoost achieve lower RMSE and higher R^2 because the dataset contains highly informative structured features (e.g., living area, grade, location), which gradient-boosted trees are particularly effective at modeling.
- **Fusion model complexity vs. data scale:** The multimodal fusion model introduces a significantly higher number of parameters, requiring substantially more data to generalize effectively. Given limited training samples, this leads to reduced performance relative to tabular-only models.
- **Visual features provide complementary but weaker signals**: While satellite imagery captures neighborhood context, its contribution is more subtle compared to dominant tabular predictors, resulting in modest gains that may not offset added model complexity.
- **Suboptimal image–tabular alignment:** Minor spatial misalignment between satellite imagery and property coordinates can introduce noise, reducing the effective contribution of visual features.
- **Optimization and training constraints:** The fusion model likely requires longer training, careful regularization, and more aggressive hyperparameter tuning to fully exploit multimodal information.

Thank You