Satellite Imagery Based Property Evaluation

Sudhanva Savyasachi S
21411034
Int.M.tech Geophysical Technology

Project Overview

1. Data Acquisition & Feature Extraction:

The primary objective was to build a **Multimodal Regression Pipeline** that integrates high-resolution satellite imagery with traditional structured data to enhance property valuation accuracy.

- **Satellite Imagery Acquisition:** High-resolution satellite images were retrieved using the **Google Maps Static API** for each property based on geographic coordinates. These images provide essential environmental context, such as neighborhood density, proximity to green spaces, and water bodies.

- **Visual Feature Extraction (DINOv2):** We employed a pre-trained **DINOv2 (base)** Vision Transformer (ViT) model. Unlike standard CNNs, DINOv2 utilizes self-supervised learning to capture intricate geometric and semantic features. Each satellite image was processed through the model to extract a **768-dimensional embedding**, serving as a comprehensive Visual Fingerprint of the property.

---

2. Feature Engineering & Pre-processing

To maximize the predictive power of the model, we engineered features to capture complex spatial and temporal relationships:

- **Geospatial Clustering:** We applied **K-Means Clustering** to the Latitude and Longitude coordinates to identify distinct "Geo-Clusters." This allows the model to account for localized market dynamics and neighborhood-specific price trends.

- **Temporal Encoding:** The Month feature was transformed using **Cyclical Encoding** (Sine and Cosine functions). This ensures the model recognizes the periodic nature of real estate seasonality (e.g., the proximity of December to January).

- **Log Transformation:** To handle the right-skewed distribution of property prices, the target variable was transformed using a **Log-Scale (log1p)**. This stabilized variance and allowed the model to maintain high accuracy across both standard and high-end property tiers.

---

3. Model Selection and Fusion Architecture

The modeling strategy followed a rigorous progression from a tabular baseline to a state-of-the-art multimodal system:

- **Tabular Baseline:** An initial **LightGBM Regressor** was trained on the 21 structured features, achieving a Cross-Validation **R^2 of 0.88**.

- **Multimodal Fusion via CatBoost:** We implemented a **CatBoost Regressor** for the final architecture, specifically leveraging its native **Embedding Features** support.

- **Native Embedding Strategy:** Rather than performing manual dimensionality reduction (like PCA) on the 768 visual dimensions, we passed the full DINOv2 vectors into CatBoost. This allowed the model to use its internal projection mechanisms to identify the most relevant visual signals for price prediction.

---

4. Validation and Regularization

I utilized a robust **5-Fold Cross-Validation** framework to ensure the model's reliability and generalizability.

- **Quantitative Results:** The integration of satellite embeddings resulted in a significant performance increase, raising the **R^2 from 0.88 to 0.9118**.

- **Overfitting Control:** To ensure the model generalized well to unseen data, we applied strategic **Regularization**. This included tuning parameters, such as l2_leaf_reg, to penalize complexity and adjusting the tree depth to maintain a healthy balance between training and validation scores.

---

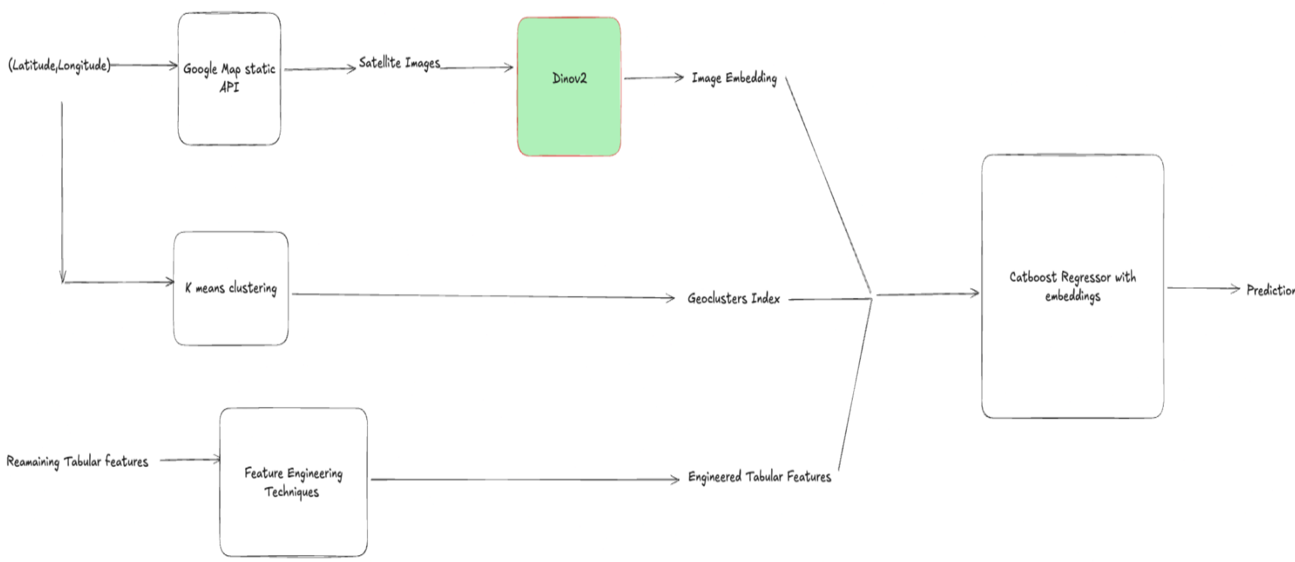
5. Model Explainability

To move beyond a "black box" approach and provide actionable insights, we utilized two primary explainability methods:

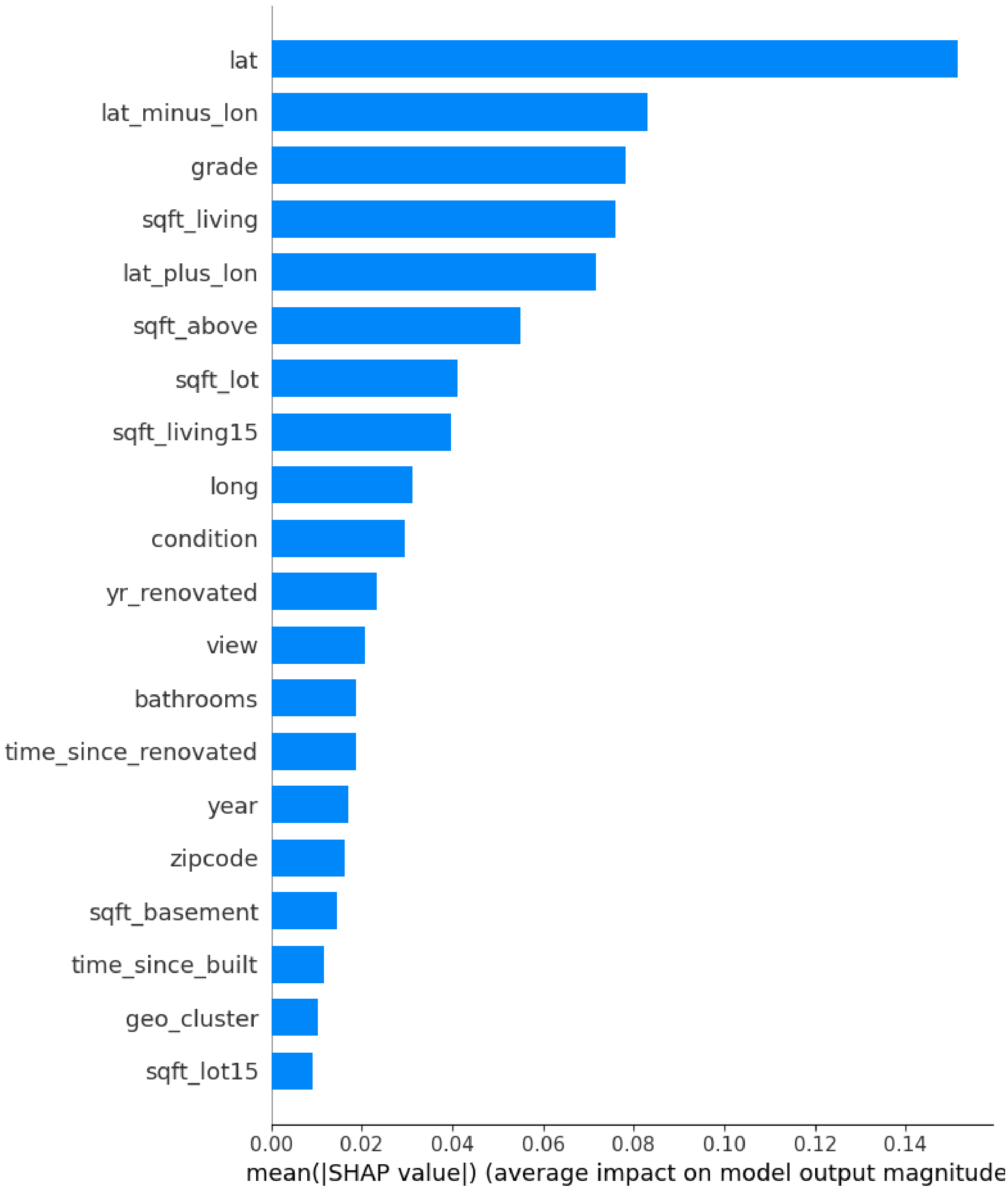
- **DINOv2 Self-Attention Maps:** We visualized the **Self-Attention heads** from the final layer of the Vision Transformer. These maps demonstrated that the model naturally attends to salient visual features, such as rooftops, driveways, and swimming pools, without explicit supervision.

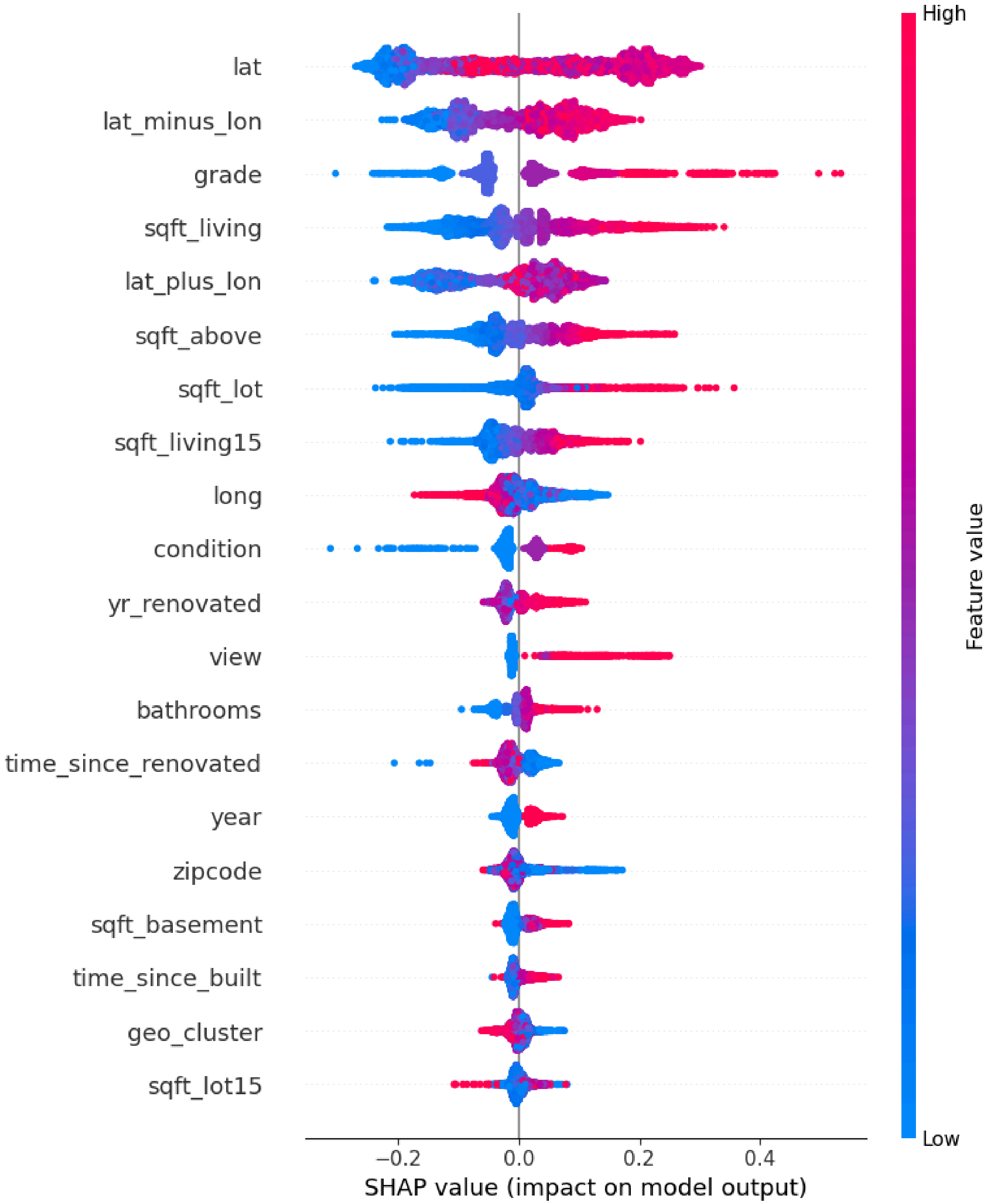
- **SHAP (SHapley Additive exPlanations) Analysis:** We conducted **SHAP analysis** to quantify the contribution of each feature to individual price predictions. This provided a transparent breakdown of how tabular data (like square footage) and visual embeddings interacted to arrive at the final market value.

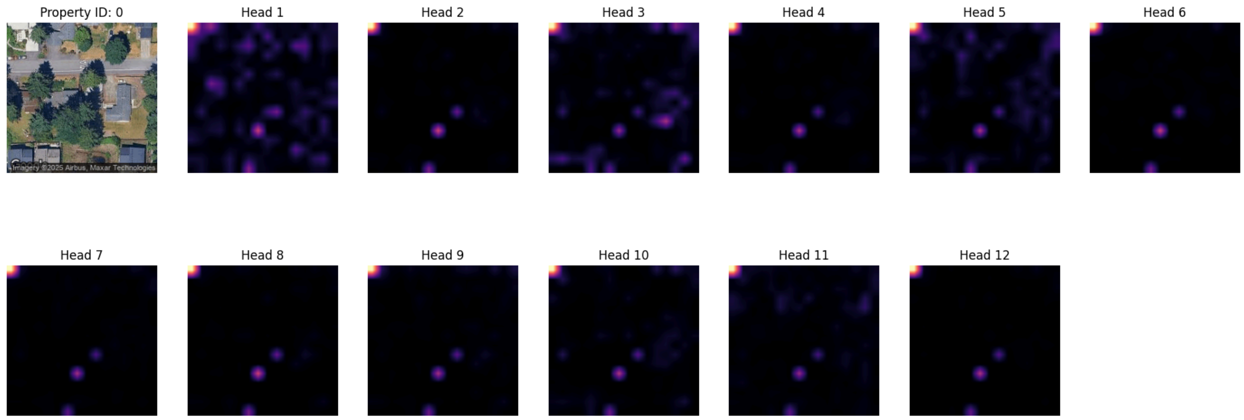
6. Architecture Diagram:



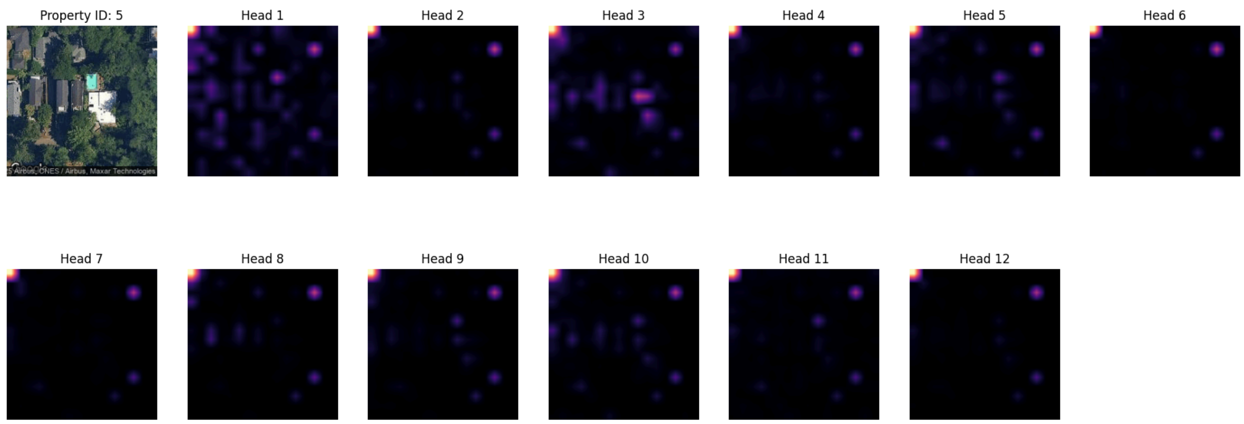
7. Shap Analysis and Attention Maps


lat
lat_minus_lon
grade
sqft_living
lat_plus_lon
sqft_above
sqft_lot
sqft_living15
long
condition
yr_renovated
view
bathrooms
time_since_renovated
year
zipcode
sqft_basement
time_since_built
geo_cluster
sqft_lot15
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
mean(|SHAP value|) (average impact on model output magnitude

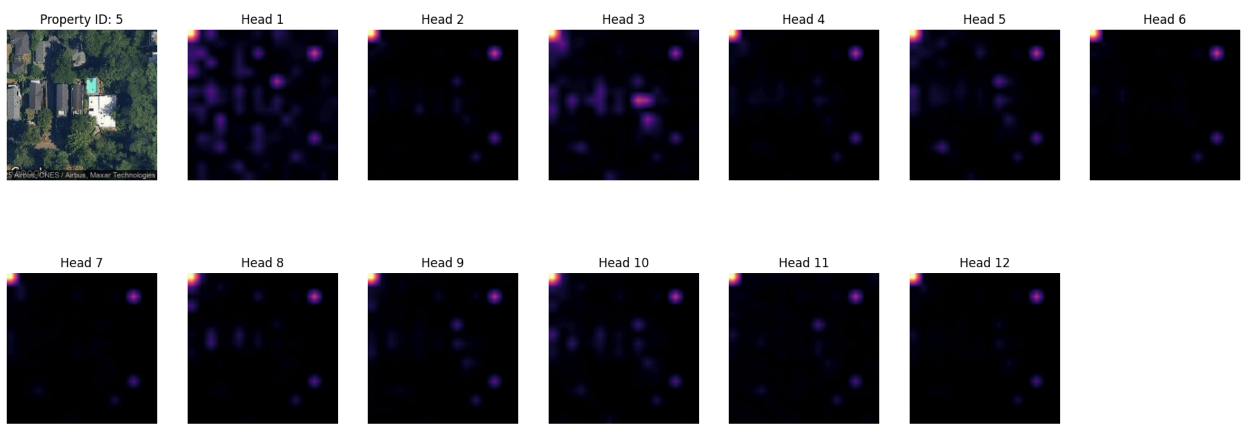

lat
lat_minus_lon
grade
sqft_living
lat_plus_lon
sqft_above
sqft_lot
sqft_living15
long
condition
yr_renovated
view
bathrooms
time_since_renovated
year
zipcode
sqft_basement
time_since_built
geo_cluster
sqft_lot15
−0.2
0.0
0.2
0.4
SHAP value (impact on model output)
High
Low
Feature value



Property ID: 0
Head 1
Head 2
Head 3
Head 4
Head 5
Head 6
Head 7
Head 8
Head 9
Head 10
Head 11
Head 12

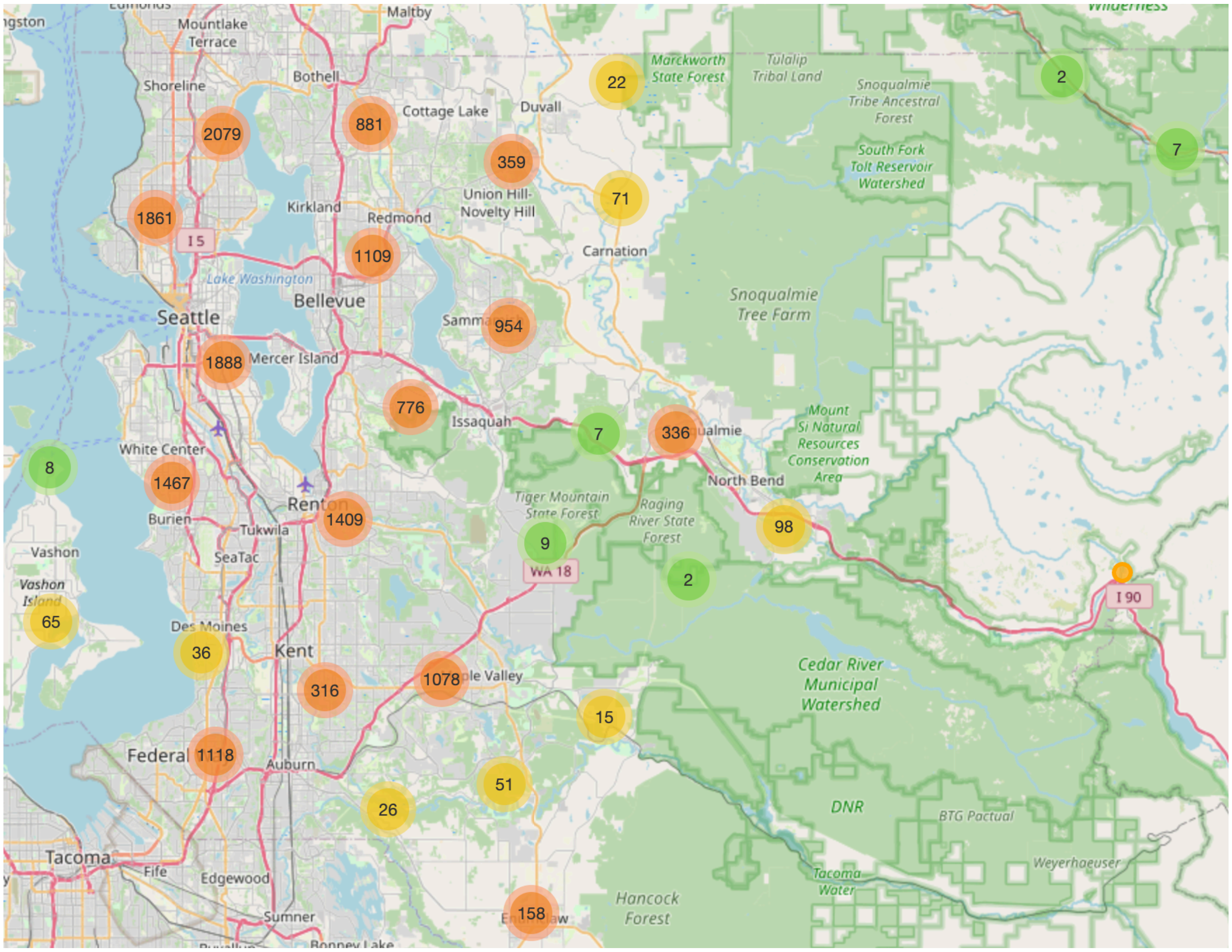


Property ID: 5
Head 1
Head 2
Head 3
Head 4
Head 5
Head 6
Head 7
Head 8
Head 9
Head 10
Head 11
Head 12



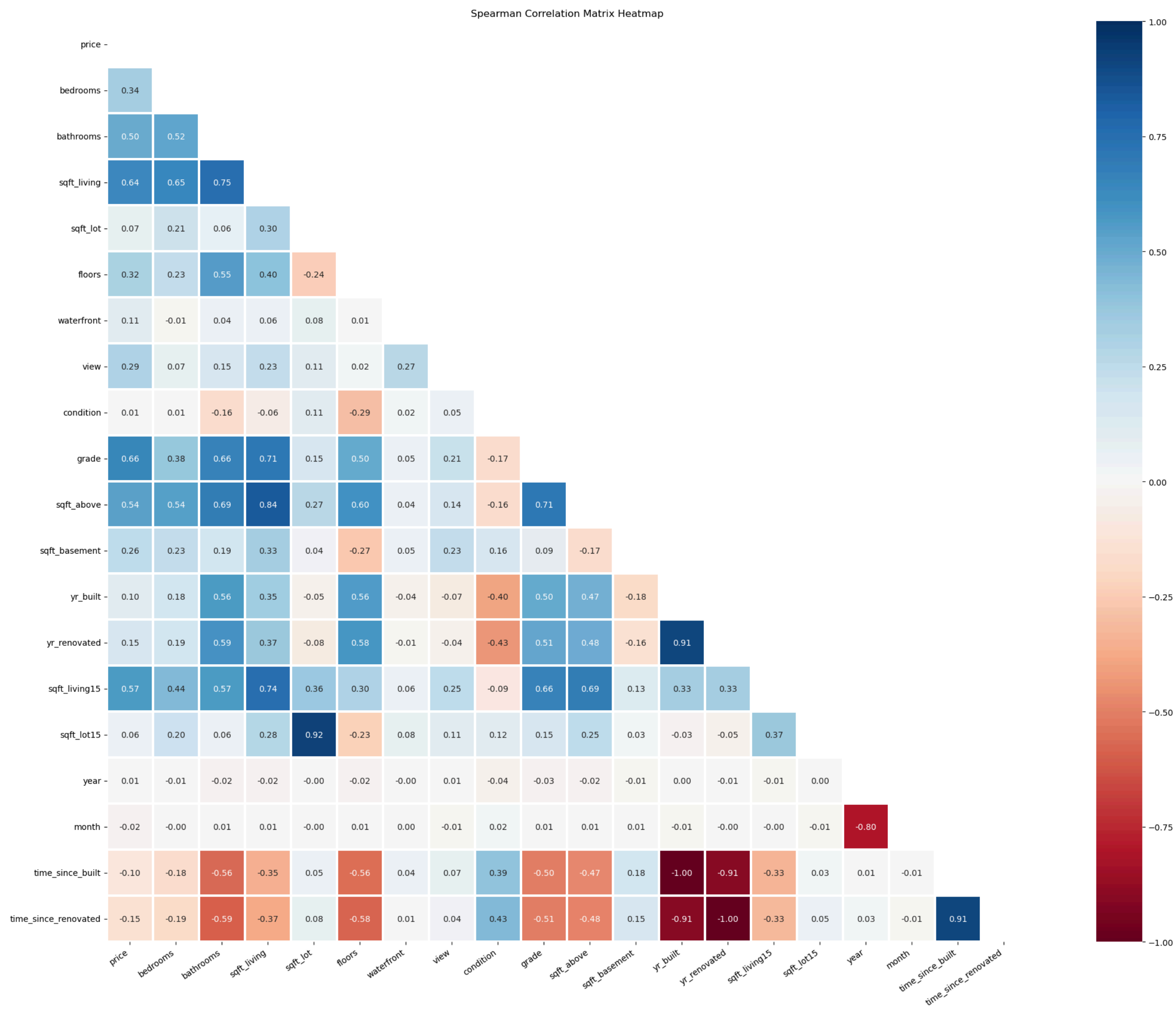
Property ID: 5
Head 1
Head 2
Head 3
Head 4
Head 5
Head 6
Head 7
Head 8
Head 9
Head 10
Head 11
Head 12

8. Basic EDA



The numbers are number of houses in that particular area

Individual Feature's EDA is in the notebook of eda_tabular. Below i have attached the correlation heatmap:


Spearman Correlation Matrix Heatmap
price
bedrooms
bathrooms
sqft_living
sqft_lot
floors
waterfront
view
condition
grade
sqft_above
sqft_basement
yr_built
yr_renovated
sqft_living15
sqft_lot15
year
month
time_since_built
time_since_renovated
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00