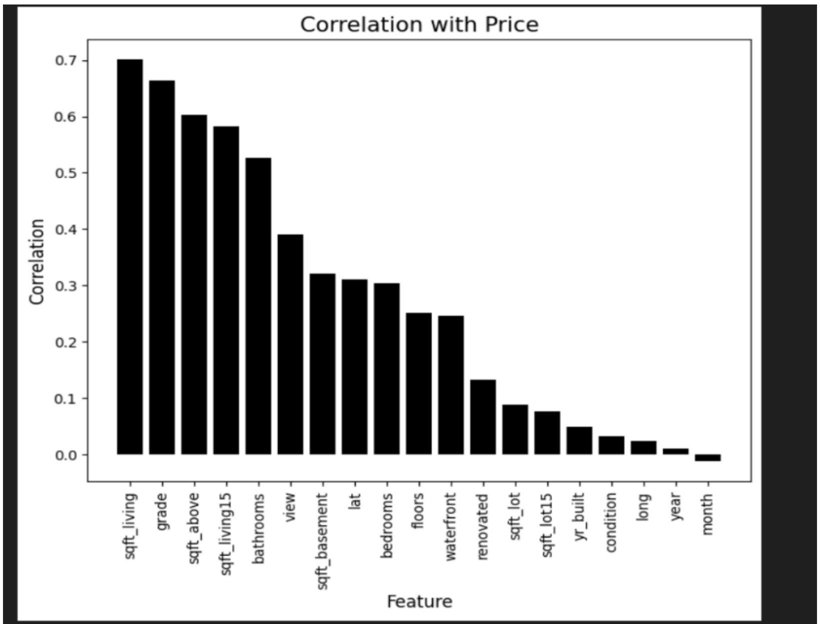
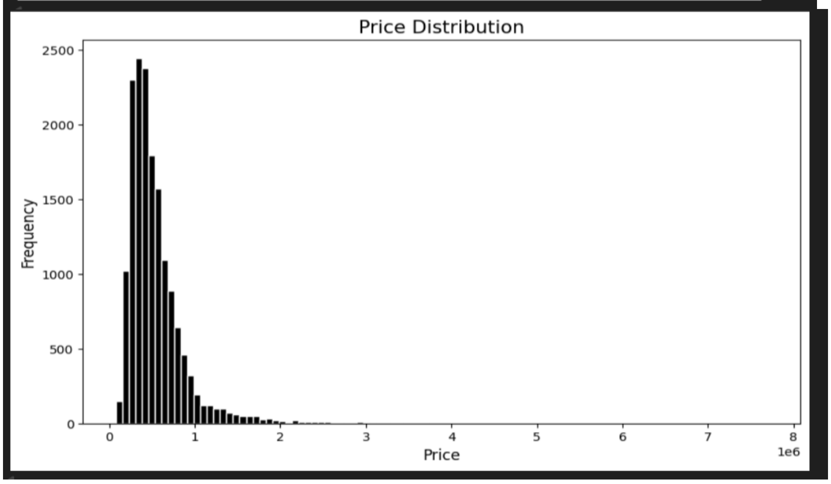
# OVERVIEW

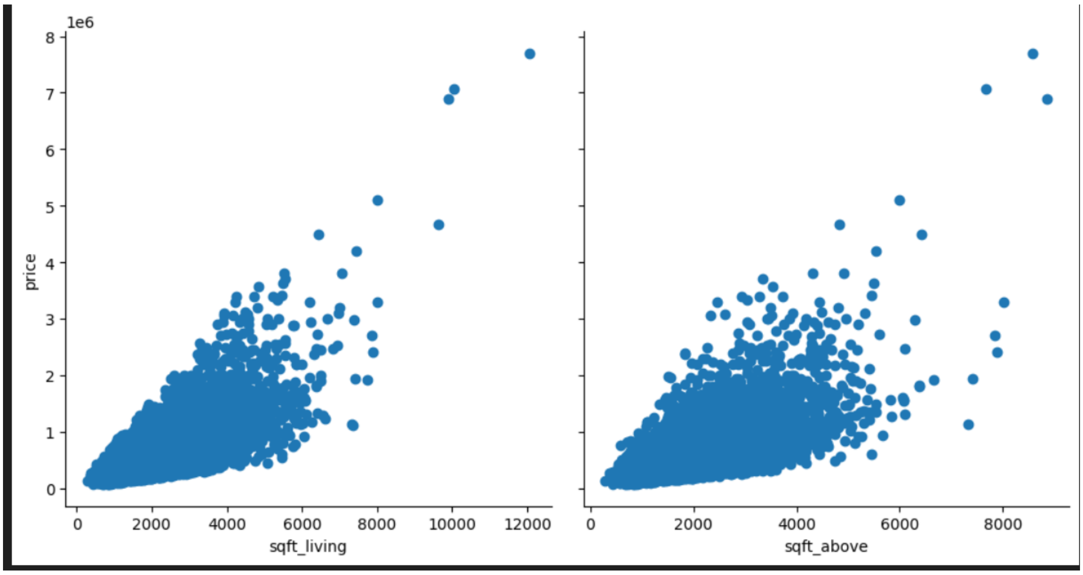
I built a multi-modal deep learning model that predicts house prices by combining two sources of information: property details and satellite images of the surrounding area. First, I used Esri's ArcGIS to use the latitude and longitudes to get the satellite images. Then I used a ResNet-18 CNN to analyse satellite images. This allowed the model to capture important environmental factors such as neighbourhood layout, road patterns, and greenery that influence property value. At the same time, I processed structured property features using a Multi-Layer Perceptron (MLP). To address the price skewness common in real estate markets and improve training stability, we applied a log transformation to the target price. The outputs from both branches were then combined through a fusion layer, enabling the model to balance physical house attributes with visual neighbourhood context when making predictions. The model was trained for 20 epochs with a learning rate of 1e-4, achieving a strong $R^2$ score of 0.76 and RMSE OF $166,183. This fine-tuning was done at least 10 times to increase the r2 score I used this trained model to predict the prices of the test data given, and the results are saved in the prediction file. I used Grad-CAM visualisations, which highlight the specific regions in satellite images that most influenced each prediction. .

# EDA and VISUALISATION

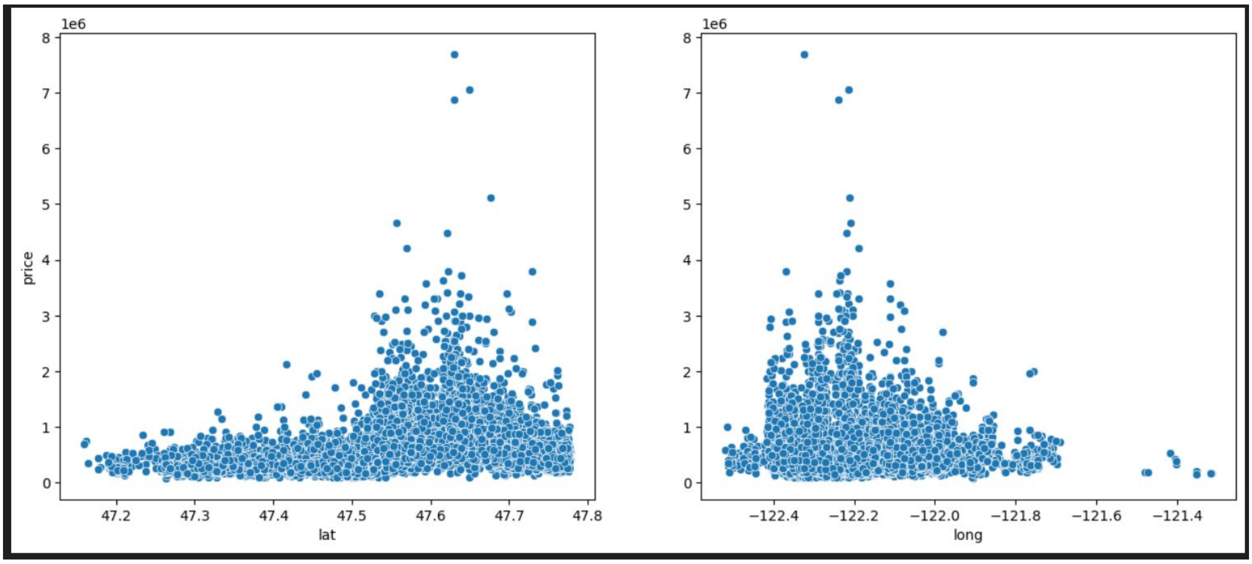


scatter plots of price with features that have a high correlation with it.




in some geographical latitudes and longitudes, the prices of houses are higher.



Sample satellite images(Seattle) – out of 16209 images from which I extracted 16153 images.

# GRAD-CAM OUTPUT


From here, we can conclude that the model neighbourhood and the tree density influenced the prices significantly. This is only an example.

```
Evaluating: 100%|██████████| 101/101 [00:56<00:00,  1.79it/s]
--- Model Performance ---
RMSE: $166,183.80
R² Score: 0.7639
```

THIS IS MODEL PERFORMANCE IN TABULAR+IMAGES

```
RMSE: 130014.45630390491
MAE: 77332.859375
Explained Variance Score: 0.8709380626678467
R2 Score: 0.8646866679191589
```

And this is the model performance using only tabular data

This indicates that visual noise can interfere with clean numerical facts; the model successfully identified high-value environmental contexts. Grad-CAM visualisations prove that the network focuses on surrounding greenery and neighbourhood density, which the tabular model lacks. This strategy demonstrates that while images introduce complexity, they provide critical spatial Comparables that help ground property valuations in real-world geography.