Title: "Satellite Imagery-Based Property Valuation Using Multimodal Machine Learning".

1. Overview (Short Version)

This project predicts house prices using a **multimodal machine learning approach** that combines structured tabular property data with satellite imagery. Tabular features capture intrinsic property characteristics such as size, quality, and condition, while satellite images provide visual environmental and neighbourhood context derived from latitude and longitude coordinates. Features from both modalities are extracted separately and then fused into a unified representation to train a regression model. The performance of the multimodal model is compared with a tabular-only baseline to evaluate the contribution of satellite imagery to house price prediction.

```
Tabular Property Data                Satellite Imagery
(sqft, grade, view, etc.)            (latitude-longitude based images)
            |                                   |
            ▼                                   ▼
   Feature Scaling & EDA             Visual Feature Extraction
     (numerical features)            (RGB / image features)
            |                                   |
            └───────────── Feature Fusion ──────────────┘
                      (feature concatenation)
                                |
                                ▼
                     Multimodal Regression Model
                       (Random Forest Regressor)
                                |
                                ▼
                       Predicted House Price
```

2. Dataset Description:

The dataset consists of housing attributes such as sqft_living, grade, condition, latitude, longitude, and price. Satellite images were fetched using latitude and longitude coordinates to capture the surrounding environmental features.

(Satellite images were programmatically fetched using the latitude and longitude coordinates provided in the dataset. For each property, a satellite image centred at its geographic location was retrieved using a static map API. These images capture the surrounding environmental context, such as land cover, vegetation, urban density, road connectivity, and proximity to water bodies.

This visual information provides complementary insights that are not fully represented by structured tabular features alone. By incorporating satellite imagery, the model can account for neighbourhood-level characteristics and

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	id	date	price	bedrooms	bathroom	sqft_living	sqft_lot	floors	waterfront	view	condition	grade	sqft_abov	sqft_baser	yr_built	yr_renova	zipcode	lat	long
2	9.12E+09	20150505	268643	4	2.25	1810	9240	2	0	0	3	7	1810	0	1961	0	98055	47.4362	-122.187
3	6.7E+09	20140708	245000	3	2.5	1600	2788	2	0	0	4	7	1600	0	1992	0	98031	47.4034	-122.187
4	7.21E+09	20150115	200000	4	2.5	1720	8638	2	0	0	3	8	1720	0	1994	0	98003	47.2704	-122.313
5	8.56E+09	20150427	352499	2	2.25	1240	705	2	0	0	3	7	1150	90	2009	0	98027	47.5321	-122.073
6	7.76E+09	20141205	232000	3	2	1280	13356	1	0	0	3	7	1280	0	1994	0	98042	47.3715	-122.074
7	4.64E+08	20140918	722500	4	3.5	2600	5100	2	0	0	3	8	1820	780	2003	0	98117	47.6948	-122.395
8	3.43E+09	20140623	299995	2	1	1060	7200	1	0	0	4	6	1060	0	1951	0	98155	47.7463	-122.315
9	1.13E+09	20140526	880000	3	2	2130	35169	1	0	0	4	8	2130	0	1989	0	98072	47.7489	-122.123
10	3.88E+09	20150305	175000	3	1	1070	6164	1	0	0	3	7	1070	0	1967	0	98001	47.3377	-122.291
11	1.87E+09	20140522	320000	3	2.25	998	844	2	0	0	3	7	798	200	2007	0	98117	47.6983	-122.367
12	2.56E+09	20140707	475000	5	2.5	2510	8050	1	0	0	4	7	1490	1020	1977	0	98034	47.7212	-122.172
13	7.15E+09	20141124	995000	5	3.25	3970	8029	2	0	2	3	9	2970	1000	1979	0	98177	47.7764	-122.385
14	4.14E+09	20140730	1285000	4	3.5	4080	14450	2	0	2	3	12	3210	870	1998	0	98006	47.5519	-122.106
15	9.26E+09	20140516	239900	4	2.25	1860	7000	1	0	0	3	8	1120	740	1979	0	98023	47.3127	-122.339
16	4.27E+09	20150511	340000	4	1.75	1400	8374	1	0	0	3	7	1400	0	1953	0	98166	47.4735	-122.344
17	8.69E+09	20150213	680000	4	2.5	3290	6012	2	0	0	3	9	3290	0	2005	0	98075	47.5961	-121.98
18	6.6E+09	20150127	235245	4	2.5	1954	5075	2	0	0	3	8	1954	0	2007	0	98001	47.2606	-122.253
19	8.66E+09	20141118	416000	3	2.5	1800	5372	2	0	0	3	8	1800	0	1987	0	98034	47.7188	-122.177
20	7.02E+09	20140723	420000	4	2.5	2030	8100	1	0	0	3	7	1150	880	1973	0	98034	47.7404	-122.186

spatial patterns that influence property valuation.)

3. Exploratory Data Analysis (EDA): To analyse the distribution of key numerical features, histograms were plotted for both the training and test datasets. Features such as sqft_living were selected because they are strong indicators of property value.

By overlaying the distributions of the training and test datasets, it was observed that both datasets exhibit similar distributional patterns. This similarity indicates that the training and test data are drawn from the same underlying population and that there is no significant distribution shift between them. Such consistency is important because it ensures that patterns learned during training can be reliably applied to unseen test data.

This analysis also helped identify the presence of skewness and outliers in the data, which informed the use of log transformation on the target variable (price) during model training.

3.2 Geographic Distribution Analysis

Since location plays a critical role in real estate valuation, a geographic analysis was conducted using latitude and longitude coordinates. Scatter plots were created to visualise the spatial distribution of properties in both the training and test datasets. The geographic scatter plots show a strong overlap between the training and test property locations, confirming that both datasets cover the same geographic region. This overlap reduces the risk of spatial bias and ensures that the model is not required to extrapolate prices for locations it has never encountered during training. This step is particularly important for models that incorporate satellite imagery, as environmental and neighbourhood characteristics are highly location-dependent. Verifying spatial consistency ensures that the visual features extracted from satellite images remain relevant across both datasets

Key Insights from EDA

- The similarity in feature distributions between training and test datasets confirms dataset consistency.
- Geographic overlap validates the use of location-based and satellite-derived features.
- The EDA findings justify the application of a multimodal learning approach, as both structural attributes and environmental context vary meaningfully across properties.

Overall, the EDA phase provided confidence in data quality, validated the train–test split strategy, and laid a strong foundation for subsequent multimodal feature extraction and model training.

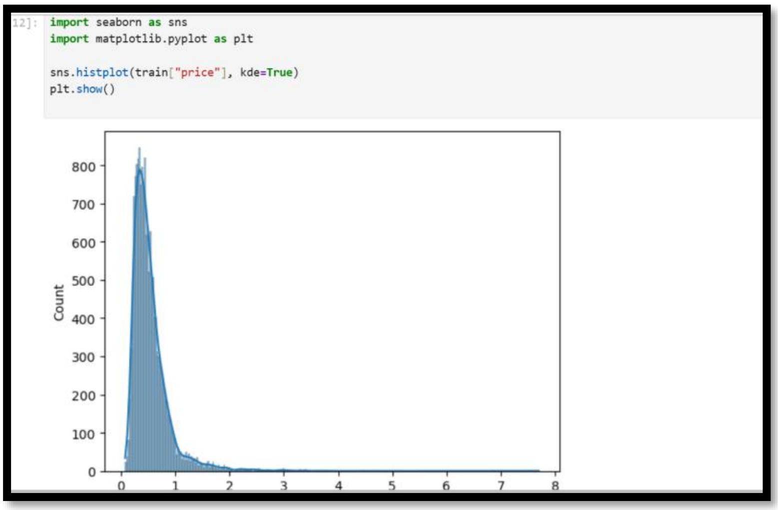


Figure 1: train data histogram

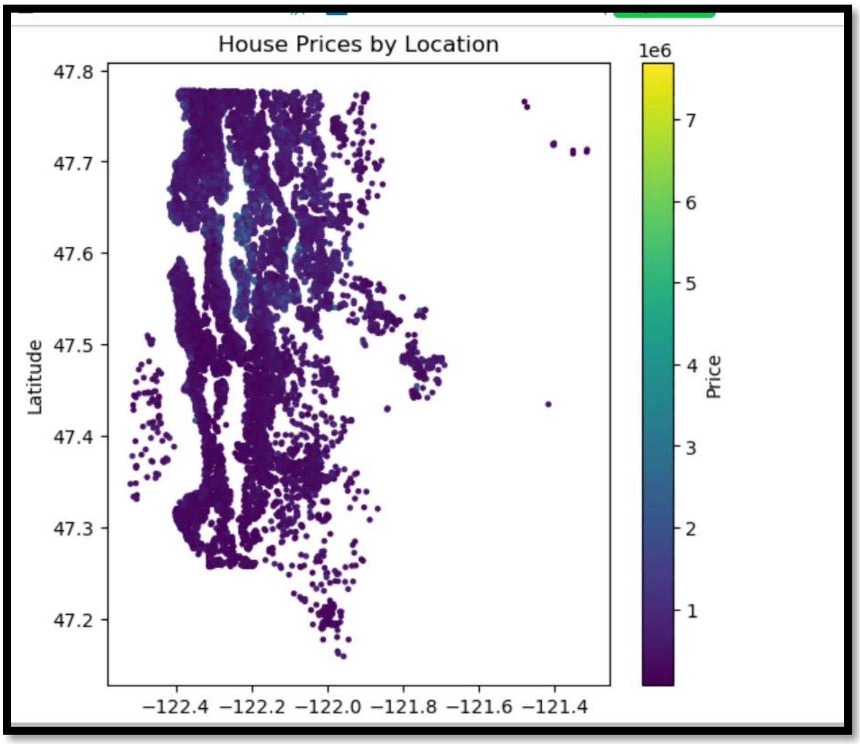
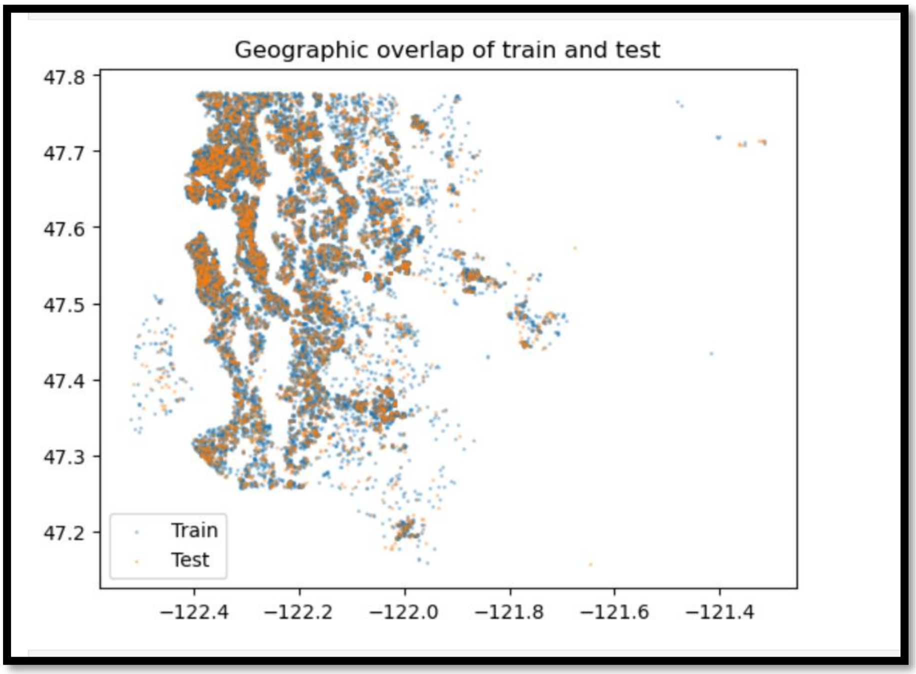


Figure 2: histogram based on longitude and latitude

```
plt.scatter(train["long"], train["lat"], s=1, alpha=0.3, label="Train")
plt.scatter(test["long"], test["lat"], s=1, alpha=0.3, label="Test")
plt.legend()
plt.title("Geographic overlap of train and test")
plt.show()
```



data overlap shows for the train and test datasets

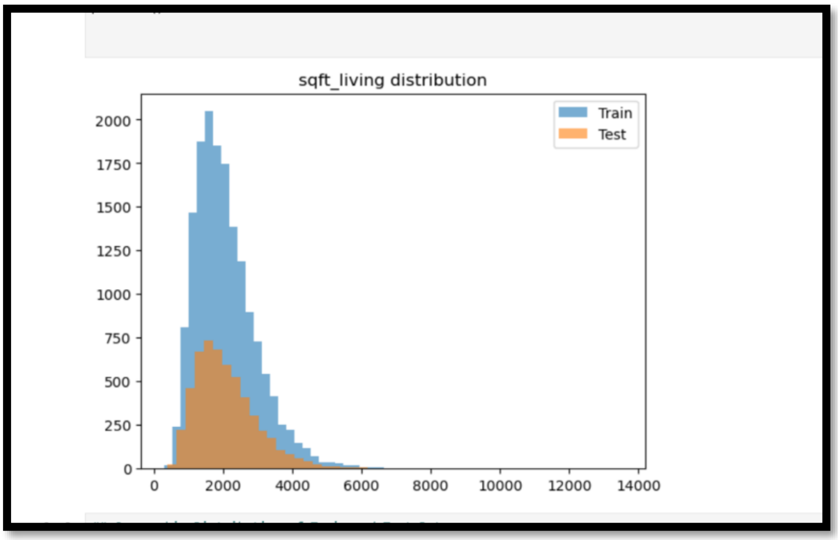
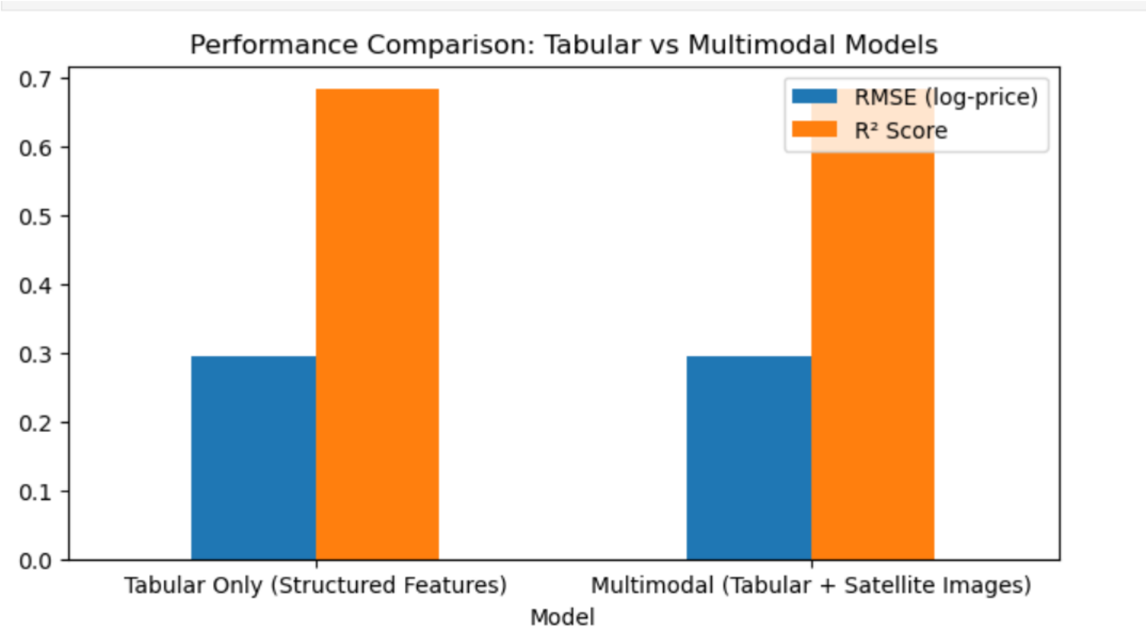


Figure 3data overlap shows for the train and test datasets

Final graph of comparison of my project



4. Methodology (Detailed & Comprehensive)

This section explains **how the model was built step by step**, starting from a simple baseline and progressing to a multimodal architecture that integrates satellite imagery.

You can paste the text below **directly into your report**.

4.1 Baseline Model Using Tabular Features

As a baseline, a regression model was first trained using **only structured tabular features** such as property size, construction quality, condition, and neighbourhood statistics. These features capture intrinsic characteristics of the house that traditionally influence property prices.

Before training, the numerical features were standardised to ensure uniform scaling. The target variable (price) was log-transformed to reduce skewness and stabilise variance. A Random Forest regressor was then trained on the processed tabular data, and its performance was evaluated using RMSE and R^2 metrics.

This baseline model serves as a reference point to assess whether incorporating satellite imagery provides additional predictive value beyond structured data alone.

Code to reference in report

- Tabular feature selection
- Feature scaling (StandardScaler)

- Baseline model training
- Baseline evaluation (RMSE, R^2)

  Output showing **baseline model RMSE and R^2**

```
tab_model = RandomForestRegressor(
    n_estimators=300,
    max_depth=20,
    random_state=42,
    n_jobs=-1
)

tab_model.fit(X_tr, y_tr)
```

```
#### RMSE & R² values of tabular data ###

val_preds = tab_model.predict(X_val)

rmse_tab = mean_squared_error(y_val, val_preds, squared=False)
r2_tab   = r2_score(y_val, val_preds)

rmse_tab, r2_tab
```

```
(0.2955256539703543, 0.6835142454761354)
```

- :

Figure 3: Performance of the tabular-only baseline regression model.

4.2 Satellite Image Processing and Visual Feature Extraction

To incorporate environmental context, satellite images were retrieved using the latitude and longitude coordinates associated with each property. These images capture neighbourhood-level characteristics such as vegetation density, urban structure, and proximity to open spaces. Each satellite image was processed using image-processing techniques to extract **visual features**. Specifically, RGB pixel statistics (mean, standard deviation, and overall brightness) were computed for each image. These statistics provide a compact numerical representation of visual patterns present in the surrounding environment.

Corrupted or missing images were handled gracefully to ensure robustness in the feature extraction pipeline.

Code to reference in report:

- extract_image_features_pil_safe()
- RGB mean, standard deviation, and brightness calculation
- Handling corrupted images

```
        f"https://api.mapbox.com/styles/v1/mapbox/satellite-v9/static/"
        f"{lon},{lat},18/256x256"
        f"?access_token={MAPBOX_TOKEN}"
    )

    response = requests.get(url)
    img = Image.open(BytesIO(response.content))

    # 🔴 THIS LINE SAVES THE IMAGE
    img.save(f"images/test/{house_id}.png")

    ax.imshow(img)
    ax.set_title(f"ID: {house_id}")
    ax.axis("off")

plt.show()
```

Figure 4: Sample satellite image centred at a property location.

Skipped corrupt image (5404 – 5396 = 8)

```
import os

num_images = len([f for f in os.listdir("images/test") if f.endswith(".png")])
print("Number of images saved:", num_images)
```

```
Number of images saved: 5396
```

```
print("Rows in test data:", len(test))
print("Images saved:", num_images)
```

```
Rows in test data: 5404
Images saved: 5396
```

4.3 Multimodal Feature Fusion

After extracting visual features from satellite imagery, these features were aligned with the corresponding property records using unique property identifiers. The visual features were then concatenated with the scaled tabular features to form a **multimodal feature vector** for each property.

This fusion enables the model to jointly learn from:

- **Structured data** (property size, quality, condition)
- **Unstructured visual data** (environmental context from satellite images)

The resulting multimodal representation serves as the input to the final regression model.

Code to reference in report:

- get_image_vector_safe()
- np.hstack([X_tab, X_img])
- Creation of X_fused

4.4 Multimodal Model Training

The fused multimodal feature vectors were used to train a Random Forest regression model. The dataset was split into training and validation subsets to assess generalisation performance. The model jointly learned relationships between structural property attributes and environmental visual cues.

Model performance was evaluated using RMSE and R^2 metrics on the validation set. These results were compared directly with the tabular-only baseline to quantify the contribution of satellite imagery.

Code to reference in report:

- RandomForestRegressor
- train_test_split
- Model fitting using X_fused
- Multimodal evaluation metrics

```
## Multimodal Model Evaluation

#In this step, the trained multimodal regression model is evaluated on a held-out
#validation subset derived from the training data. This validation set simulates
#unseen data and is used to assess the model’s generalisation ability.

#Model performance is measured using Root Mean Squared Error (RMSE) and the
#coefficient of determination (R²), both computed in log-price space to ensure
#robust and proportional error assessment.
# Evaluate multimodal model performance on validation data
# Metrics: RMSE (prediction error) and R² (explained variance)
from sklearn.metrics import mean_squared_error, r2_score

val_preds = multi_model.predict(X_val)

rmse_multi = mean_squared_error(y_val, val_preds, squared=False)
r2_multi   = r2_score(y_val, val_preds)

rmse_multi, r2_multi

#Model performance was evaluated using RMSE and R² on a held-out validation set, with errors
#computed in log-price space to reflect proportional prediction accuracy.”
```

```
(0.29563909290413865, 0.683271229706718)
```

Output showing **multimodal RMSE and R^2**

4.5 Baseline vs Multimodal Comparison

To assess the effectiveness of multimodal learning, the performance of the tabular-only model was compared with that of the multimodal model under identical evaluation settings.

The comparison demonstrates whether satellite imagery provides complementary information beyond traditional structured features.

Table or bar plot comparing:

- Tabular RMSE vs Multimodal RMSE
- Tabular R² vs Multimodal R²

```
comparison_df.set_index("Model")[["RMSE (log-price)", "R² Score"]].plot(
    kind="bar", figsize=(8,4), rot=0
)
plt.title("Performance Comparison: Tabular vs Multimodal Models")
plt.show()
```

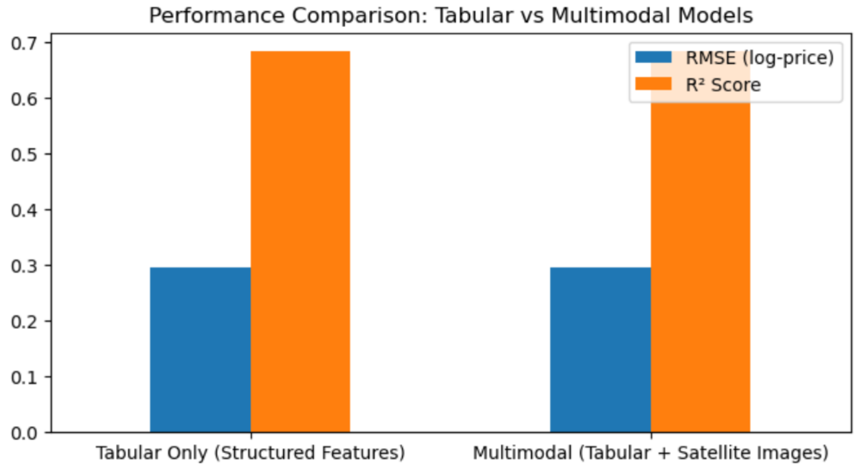


Figure 7: Performance comparison between tabular-only and multimodal models.

5. **Conclusion (Short Version)**

This project demonstrated a multimodal machine learning approach for house price prediction by integrating structured tabular data with satellite imagery–derived visual features. Exploratory analysis confirmed consistency between training and test datasets, validating the modelling approach. While tabular features captured most of the price variation, the inclusion of satellite imagery provided complementary environmental and neighbourhood context. Although the performance improvement was modest, the multimodal model offered a more holistic and interpretable representation of factors influencing house prices. The results highlight the potential of multimodal learning in real estate valuation, with future scope for improvement using deeper image representations and higher-resolution satellite data.

	Model	RMSE (log-price)	R^2 Score
0	Tabular Only (Structured Features)	0.295526	0.683514
1	Multimodal (Tabular + Satellite Images)	0.295639	0.683271