SATELLITE IMAGERY-BASED PROPERTY VALUATION

-Prathvi Paliwal

PROJECT OVERVIEW & STRATEGY

GOAL

The objective of this project was to predict house prices by using both structured housing attributes and satellite imagery of the property location. Traditional features such as size, number of rooms, and building age were combined with visual information from satellite images, since property value is influenced not only by the house itself but also by its surrounding environment.

METHODOLOGY

- We first trained a **tabular XGBoost regression model** using the available structural and location-related features.
- In parallel, we developed a **Convolutional Neural Network (CNN)-based model** to learn price-relevant patterns directly from satellite imagery, such as neighbourhood layout and surrounding land characteristics.
- The outputs from both models were then **combined using a fusion model**, allowing the visual signals from the images to complement the structured data.
- Finally, we evaluated and compared the performance of:
    - the Tabular-only (XGBoost) model, and
    - the Tabular + Satellite Image fusion model.

The key objective was to see whether visual neighborhood context improves price prediction accuracy.

EXPLORATORY DATA ANALYSIS (EDA)

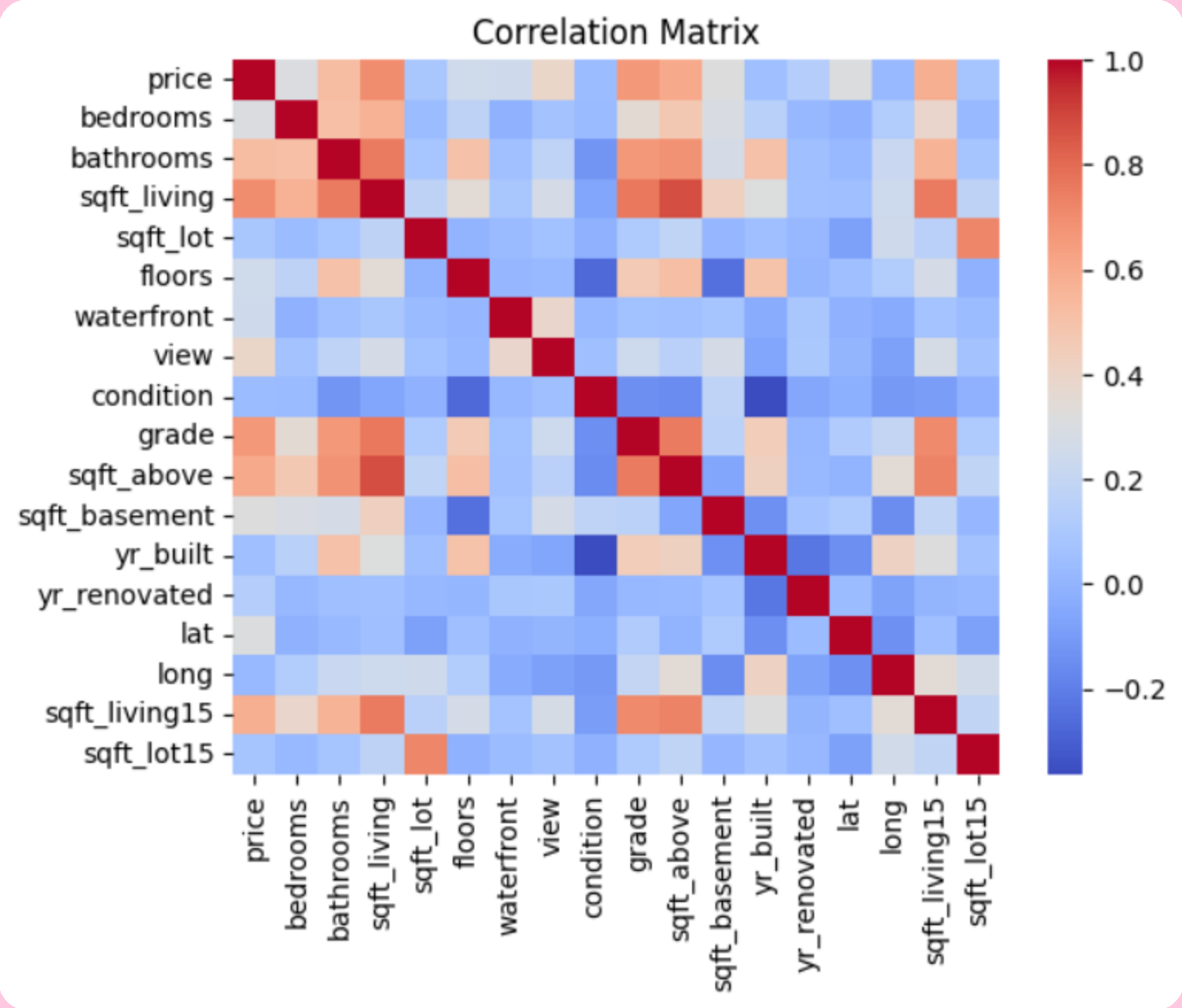
Before modelling, I carried out exploratory data analysis to understand the price distribution and visual characteristics of the properties.

House prices showed a right-skewed distribution, with most properties clustered in the lower to mid price range and a smaller number of high value outliers.

A sample of satellite images was also reviewed to observe visual differences across locations, such as greenery, building density, and surrounding land use. These observations supported the idea that neighbourhood context may influence price, motivating the use of satellite imagery in the model.

EXPLORATORY DATA ANALYSIS (EDA)

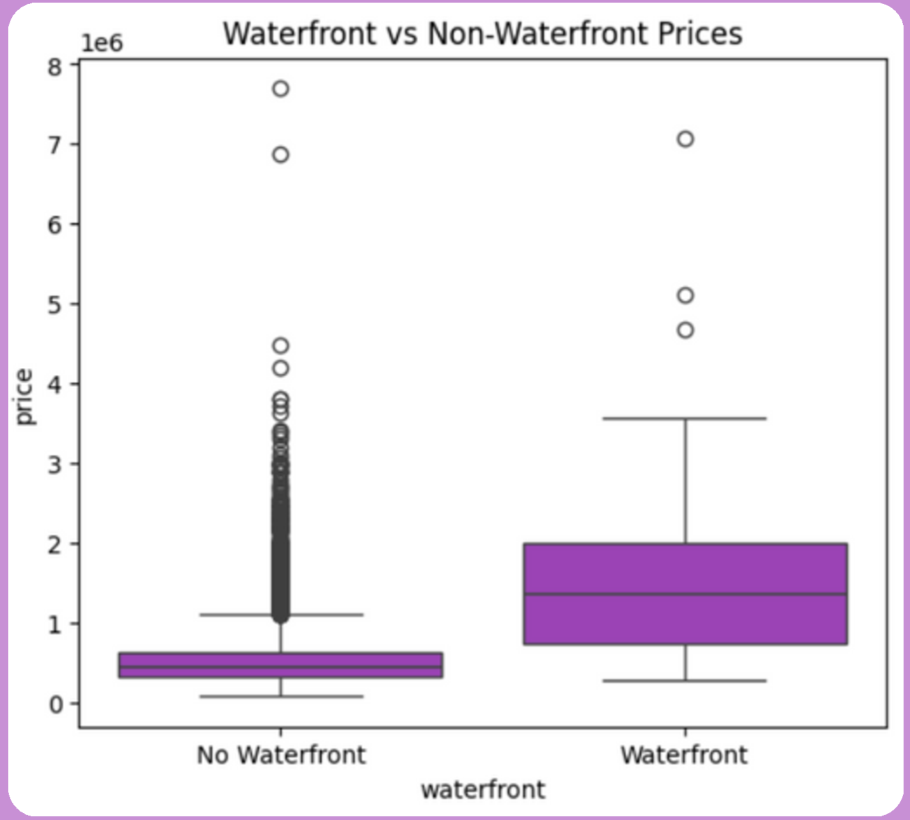
Core Feature Relationships:



The correlation heatmap shows that price is most strongly associated with living area, overall grade, and bathrooms. Location coordinates and some structural features show weaker relationships. This confirms that size and quality play a major role in pricing.

There is a clear positive relationship between living area and price. Larger homes generally sell for more, although price variation increases at higher sizes, suggesting other factors also influence value.

Waterfront homes are consistently more expensive than non-waterfront homes, with a much higher median price and upper range. This indicates a clear price premium for water-adjacent properties.



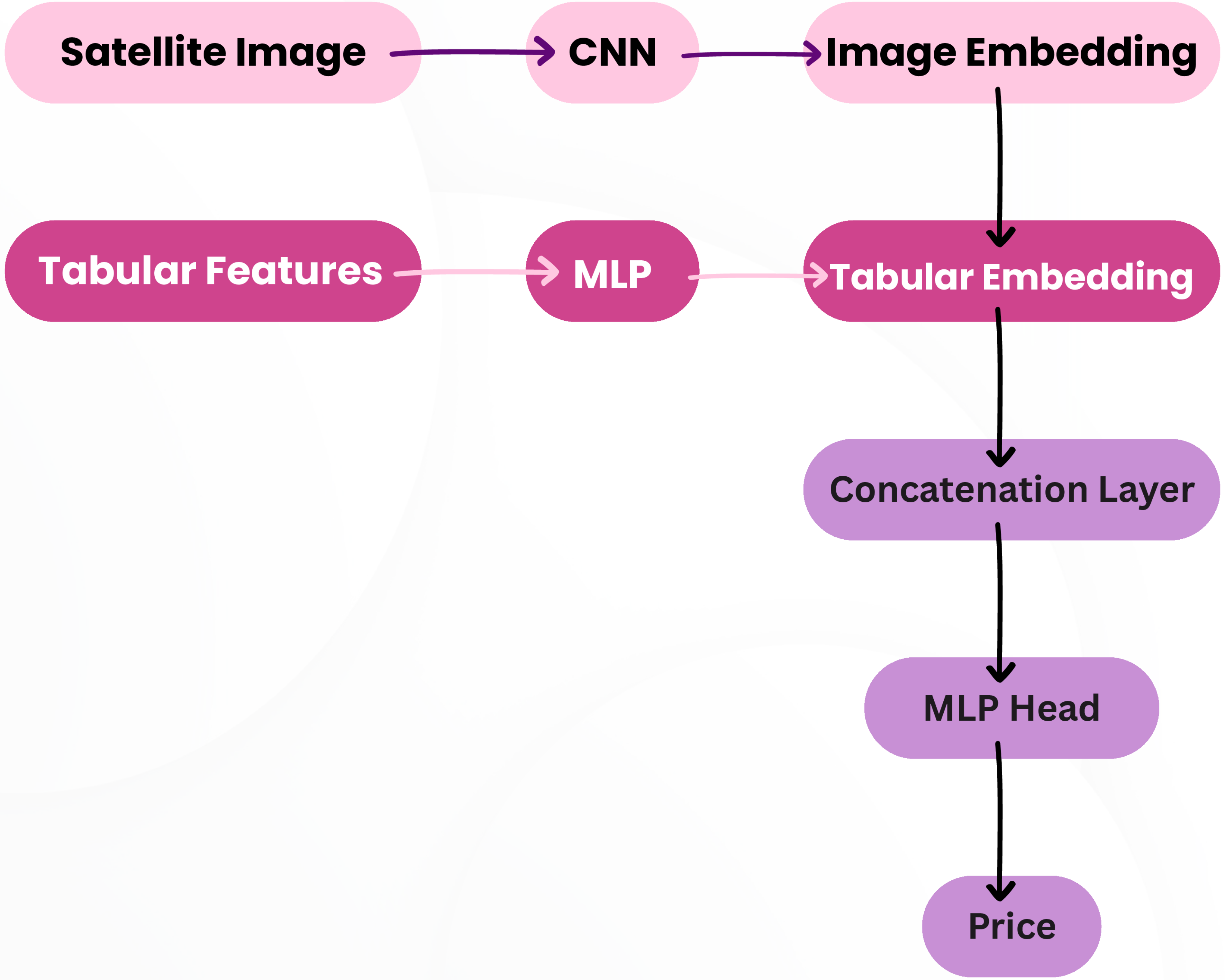
FINANCIAL & VISUAL INSIGHTS

KEY INSIGHTS

- **High-priced properties** showed strong activation around **large standalone buildings**, **open land**, and **dense greenery**, suggesting that space and environmental quality are important value drivers.
- **Mid-priced homes** displayed a mixed pattern, with attention shared between the house structure and surrounding neighbourhood layout.
- **Lower-priced properties** often showed activation near **roads**, **intersections**, and **compact housing clusters**, indicating higher density and limited private land.
- These results align with economic intuition:
  - More space + greenery → higher perceived value
  - Dense, road-dominated surroundings → lower value

Overall, the satellite images helped capture neighbourhood quality and land characteristics that are not observable from tabular housing attributes alone.

ARCHITECTURE DIAGRAM



The final model combines information from the satellite images and the structured housing data into a single price prediction pipeline.

- The **satellite image** is first processed through a **CNN**, which converts the visual information into a compact feature representation (image embedding).
- In parallel, the **tabular housing features** are passed through a **neural network (MLP)** to learn a representation of the structured data.
- The two learned representations are then **merged into a single feature vector**.
- This combined representation is fed into a final MLP layer, which outputs the **predicted house price**.

This setup allows the model to use both neighbourhood visuals and structured property attributes jointly during prediction.

RESULTS AND EVALUATION

To evaluate the impact of satellite images on price prediction, we compared three models: a tabular-only XGBoost model, a CNN trained only on images, and a fusion model that combines both data sources.

Model Type	Description	RMSE	R^2
CNN-only Model	Uses only satellite images to predict price	251689.412	**0.4951**
Tabular-only Model (XGBoost)	Uses only structured housing attributes	117430.146	**0.8901**
Fusion Model (Tabular + CNN)	Combines tabular features with satellite-image embeddings	115441.879	**0.8938**

KEY TAKEAWAYS

- The **CNN-only model performs worse** than the XGBoost tabular model, showing that images alone are not sufficient to capture price variation.
- The **tabular XGBoost model performs strongly**, as structural and location-based features already explain much of the price signal.
- The **fusion model performs best**, meaning that visual neighbourhood context adds **additional predictive value** beyond structured data.
- Satellite images help capture environmental quality like greenery, spacing, roads, and surroundings which are **not fully reflected in tabular features**.
- The benefit is **most noticeable for high-priced homes**, where neighbourhood setting plays a larger role.

Combining tabular data with satellite-image insight leads to the most accurate and realistic price predictions.

THANK YOU