Dubai Property Price & Valuation Tool: Summary of Findings

Executive Summary

This project successfully developed a high-accuracy model to predict property prices in Dubai and leveraged it to build a functional valuation tool. Through exploratory data analysis, we identified the primary drivers of property value. By comparing three distinct machine learning algorithms, a champion model was selected that explains **84.5% of price variance**. The resulting tool provides users with an objective, data-driven assessment of whether a property is underpriced, fairly priced, or overpriced.

Key Data Insights

- **Primary Price Drivers:** Property `size` and `neighbourhood` were conclusively identified as the most significant factors influencing a property's price.
- **Critical Role of Location:** Analysis revealed extreme price disparities across different neighbourhoods, with median prices in some areas being several times higher than in others. This confirms that location is a non-negotiable feature for accurate valuation.
- **Impact of Data Cleaning:** The initial dataset contained extreme outliers (the top 1%) that made preliminary models highly inaccurate. Removing these outliers was the single most important step in achieving a reliable and predictive model.

Prediction Model Performance

Three models were trained and evaluated to ensure the highest accuracy. The **Random Forest Regressor** was selected as the champion model for its superior performance.

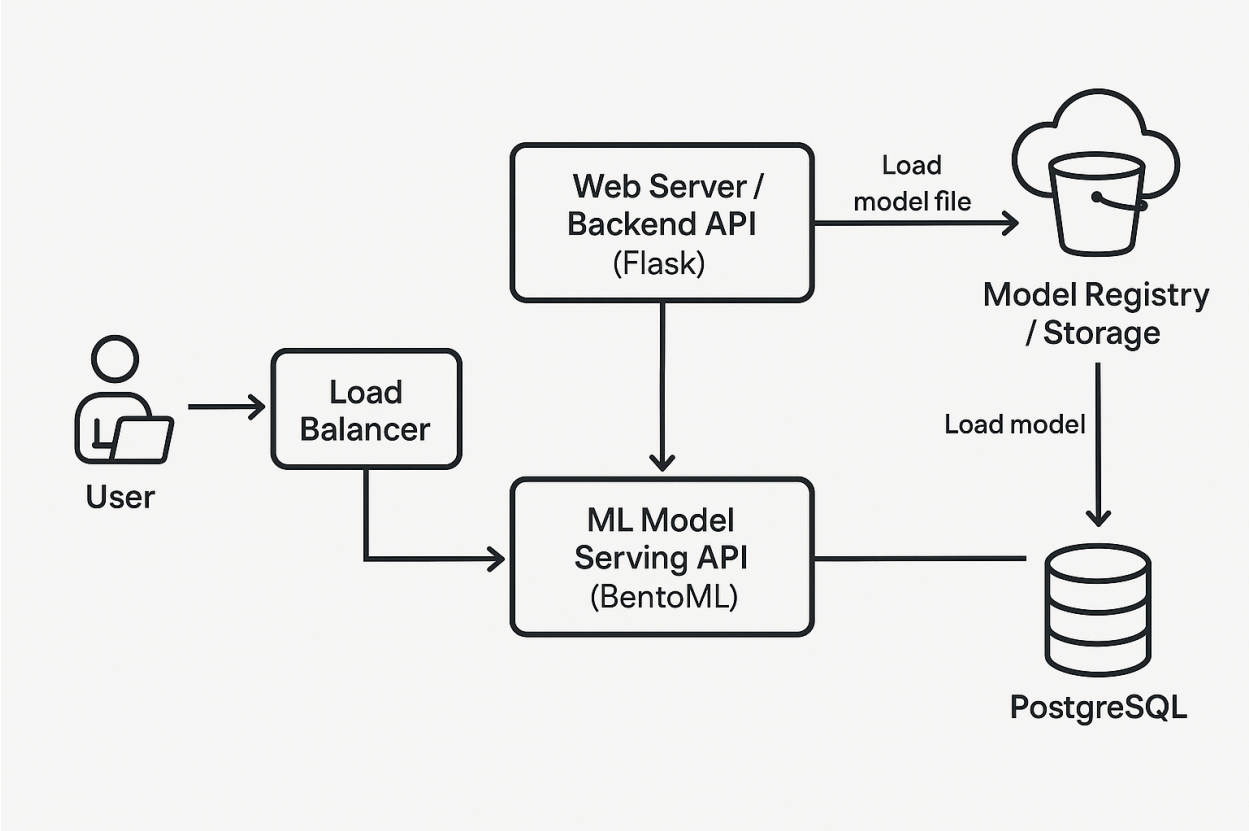
- **Champion Model:** Random Forest Regressor
- **R-squared (R^2) Score: 0.845**
    - *This means the model successfully explains 84.5% of the variation in property prices based on the provided features.*
- **Mean Absolute Error (MAE): ~13,400 AED**
    - *This means, on average, the model's price prediction is within 13,400 AED of the actual listing price.*

Valuation Tool Functionality

A tool was designed to provide users with an instant property valuation based on the champion model.

- **Core Logic:** The tool takes a property's features and its listing price as input. It uses the Random Forest model to predict a "fair market price."
- **Output:** It compares the listing price to the predicted price and classifies the property as:
    - **Underpriced** (if >10% below the prediction)
    - **Fairly Priced** (if within +/- 10% of the prediction)
    - **Overpriced** (if >10% above the prediction)

Architectural diagram of how we can serve this tool to the users.



Conclusion & Next Steps

The project delivered a robust, data-driven method for property price prediction and valuation. For real-world deployment, a scalable cloud architecture was designed. To ensure long-term accuracy, it is recommended that the model be continuously monitored for performance degradation (model drift) and retrained periodically on new market data.