Employing Machine Learning for Real Estate Price Prediction: A Case Study on Wazobia Real Estate

Overview - The Quick Pitch

In this study, our primary goal is to identify key factors influencing house prices in Wazobia's real estate market using advanced Machine Learning techniques. We will develop an accurate predictive model by analyzing features like location, bedrooms, bathrooms, house title, and parking space. This research will provide insights into housing market dynamics, empowering better decision-making for buyers, sellers, and investors."

The Problem

Here is where you describe the problem that you are solving.

- What are the significant factors that influence house prices in Wazobia's real estate market
- Can advanced Machine Learning techniques be effectively utilized to develop an accurate predictive model for house prices in Wazobia?

The Solution

.The solution include:

- Conducting Exploratory Data Analysis to gain insights into the factors influencing house prices
- Power predictive model to make accurate house prices predictions

Exploratory Data aAnalysis

Below are the graphs which indicate import results revealed from the analysis carried out to gain insights
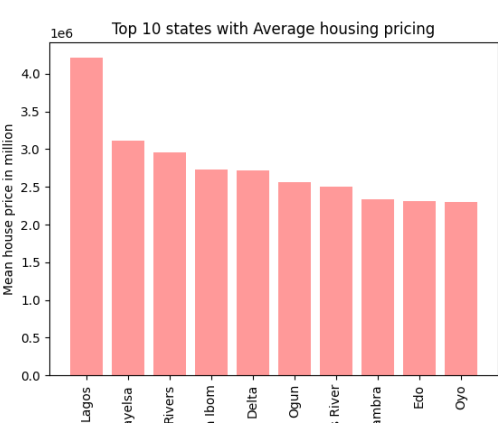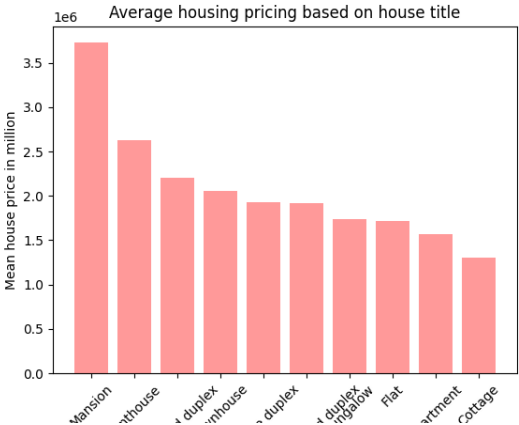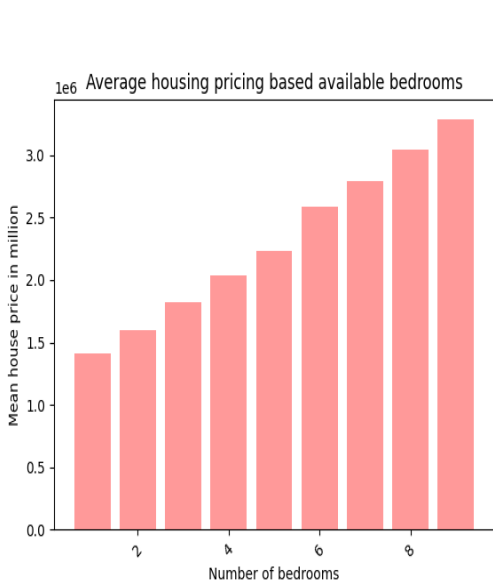

figure 1

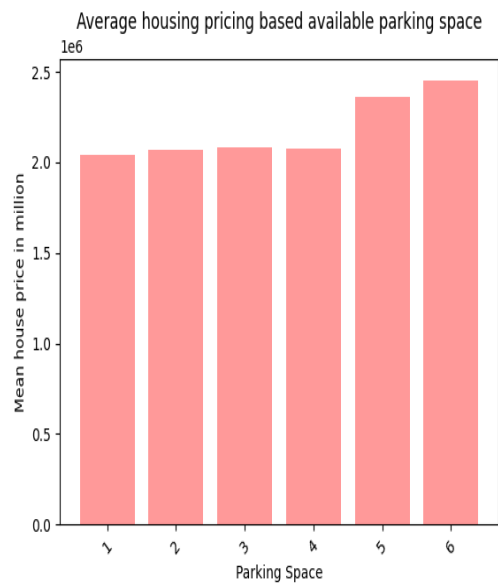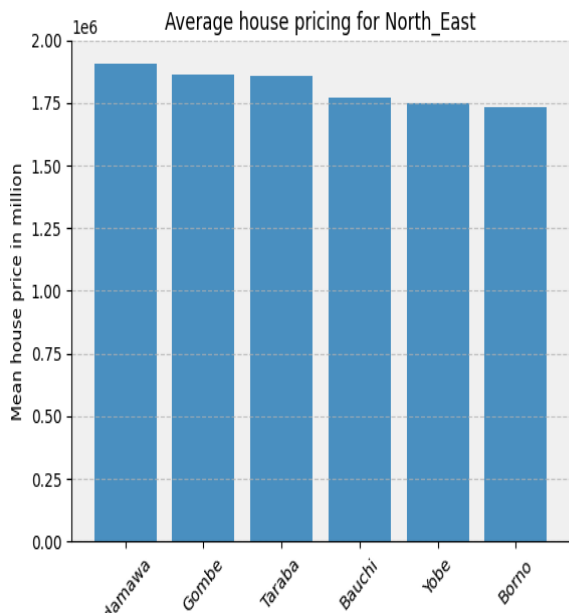*figure 2*

figure 3

figure 4

figure 5

figure 6

fi

Keys findings from EDA

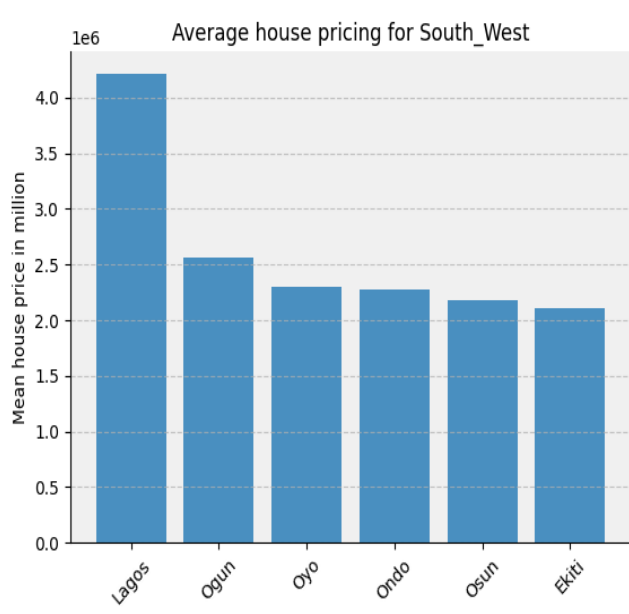These are the main findings from the exploratory data analysis

- Fig 1 illustrates a significant price disparity between houses in Lagos and other locations. This contrast is likely attributed to factors like high demand, limited housing supply, economic prominence, and its status as a major urban center. The same trend is evident when comparing Lagos to other South-West states (see Figure 6).

- Figure 3 reveals that the number of bedrooms exerts a more significant influence on house prices than the availability of parking spaces (as depicted in Figure 4). This effect can be attributed to the direct correlation between the number of bedrooms and the overall size and accommodation capacity of the house, which plays a crucial role in determining the pricing.
- Figure 2 reveals that mansion-type houses have the highest prices, while cottage-type houses are the most affordable. Additionally, prices of duplex-type houses show relatively little variation.

Model Building approach

- Import feature Engineering approaches
- Machine learning models used and their performances
- Model interpretations with feature importance

Import Feature Engineering approaches

1. **Careful handling of missing values**
   - To mitigate the potential introduction of noise to the data, the missing values in the number of locations, house title, and number of bedrooms were removed from the dataset instead of filling them
   - Missing values in the number of bathrooms and parking spaces were filled using the mode based on the house title to ensure data consistency and preserve the most frequent values for these features.use title
2. **Feature Creation**
   - New features like geopolitical zone and sum of bedrooms,bathrooms and parking spaces were introduced
   - The new feature, sum of the ranking of the location and house title, exhibits a strong positive correlation of about 0.73 with the target variable (house prices), signifying its potential importance in enhancing the model's predictive performance.
3. **Feature Transformation**
   - Features like locations, title(house type e.g mansion) are encoded using mean encoding from the target variable
   - log transformation was performed on the target variable to normalize it

Machine learning models used and their performances

Few models are selected and evaluated using root mean square(RMSE) as the evaluation metric
The table below shows the generated results

Model	Score(RMSE)
Gradient Boosting Regressor	367148.845194
light gradient boosting	372179.046952
catboost	389999.180707
Linear Regression	391778.466745
xgboost	410578.889002

Comment : The result above shows Gradient boosting machine outperforms others after a total of 10 folds of cross validation on the validation data

Hypermeter optimization was later carried out xgboost which boasts a score of about 360000 rmse outperforming the rest

Model interpretation with Feature importance

A feature importance analysis was performed on the trained Gradient Boosting Machine to predict house prices. The graph shows the relative importance of features in influencing house prices. Higher bars indicate more significant impact, guiding us to focus on the most relevant features for accurate predictions.
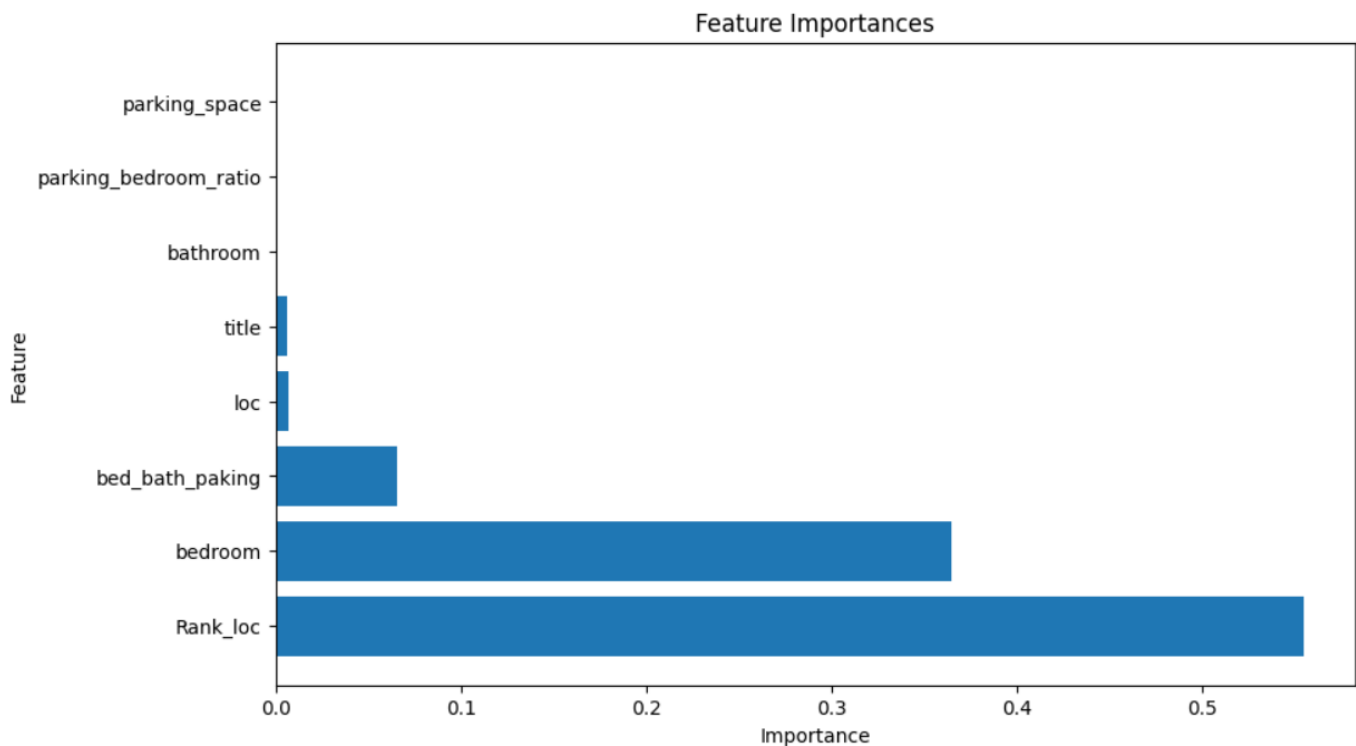
The feature importance analysis from the model indicates that three features play a pivotal role in predicting the house prices. These features are:

1. Number of bedrooms
2. The sum of the ranking of the location and house title
3. The sum of the number of bedrooms, bathrooms, and parking spaces

According to the trained model, these factors carry the most significant influence when determining the price of a house. As a result, when using this model for predictions, it is crucial to give special attention to these key features, as they have demonstrated strong predictive power in the context of house pricing.

Conclusion

Advanced Machine Learning techniques developed an accurate predictive model for Wazobia's real estate prices. Key factors impacting prices were identified: number of bedrooms, location and house title ranking, and total bedrooms, bathrooms, and parking spaces. Gradient Boosting Machine outperformed other models, with optimized XGBoost achieving the best result (RMSE ~360,000).

Recommendations

1. Prioritize key features: Focus on bedrooms, location ranking, and combined bedrooms, bathrooms, and parking spaces for accurate pricing.

2. Target high-demand areas: Consider properties in prominent zones like Lagos with substantial price differences.

3. Tailor to house types: Customize marketing based on demand for different house types.

4. Use optimized model: Adopt the XGBoost for precise house price predictions.

5. Maintain data quality: Minimize missing values and apply feature engineering techniques.

6. Regular updates: Keep the model relevant with timely data incorporation.