House Prices Prediction in Nigeria

Introduction

Wazobia Real Estate Limited is a distinguished real estate company operating in Nigeria, renowned for its extensive property portfolio. While aiming to provide precise and competitive house pricing, the company has encountered challenges in accurately forecasting house prices within the current market. In light of this, Wazobia Real Estate Limited has sought the expertise of data scientists to develop a robust predictive model that enhances its pricing accuracy. This executive report outlines the methods, algorithm, and performance evaluation carried out to achieve this task.

Data

The data used in this competition is the sole property of Zindi and the competition host. The data set is divided into a training dataset and a testing dataset. The training dataset contains 14000 rows and 7 columns, the columns are ID, loc, title, bedroom, bathroom, parking space and price which is the target column while the test data contain 6000 rows and 6 Columns. The dataset contains several missing data which subject the dataset to further cleaning.

Methodology

A thorough technique was rigorously followed in the pursuit of developing an accurate predictive model for Nigerian housing prices. The first phase included thorough data cleaning and preparation to ensure that the dataset was free of errors and missing values, providing the groundwork for subsequent research. A critical stage was feature selection, which was carried out using a statistical approach to identify the most influential characteristics that significantly contribute to price prediction.

This process was built on training a variety of machine learning models on the precisely produced dataset. This strategic ensemble included a variety of algorithms, each with unique strengths for capturing subtle patterns in data. The usefulness of these models was determined by a careful evaluation method centred on the root mean squared error (RMSE) statistic, which assesses forecast accuracy by comparing actual and expected prices.

The most capable model emerged from this rigorous evaluation, displaying higher predictive prowess. This top contender was carefully selected and trained to make precise predictions based on previously unknown test data. The culmination of our methodology assures that Wazobia Real Estate Limited has a strong forecasting tool, backed up by a methodical strategy that not only optimizes price accuracy but also enables educated decision-making in the volatile Nigerian real estate market. Our data-driven methodology demonstrates our dedication to providing significant solutions targeted to the real estate industry's particular difficulties.

Data Analysis and Preprocessing

The initial step involved loading and analyzing the training and test datasets. The training dataset, named "Housing_dataset_train.csv," comprises essential features such as location, title, bedrooms, bathrooms, parking space, and price. I also observed missing values in the data that required handling.

I conducted exploratory data analysis (EDA) to gain insights into the data distribution and relationships. The pairplot revealed correlations between various features, indicating potential predictive power. Additionally, countplots and scatterplots were utilized to visualize the distribution of categorical variables and relationships between numerical variables.

To prepare the data for modelling, we performed the following preprocessing steps:

- Encoded categorical variables (e.g., "loc" and "title") using label encoding to transform them into numerical representations.
- Filled missing values with either the mean or mode of the respective columns to ensure complete datasets for analysis.

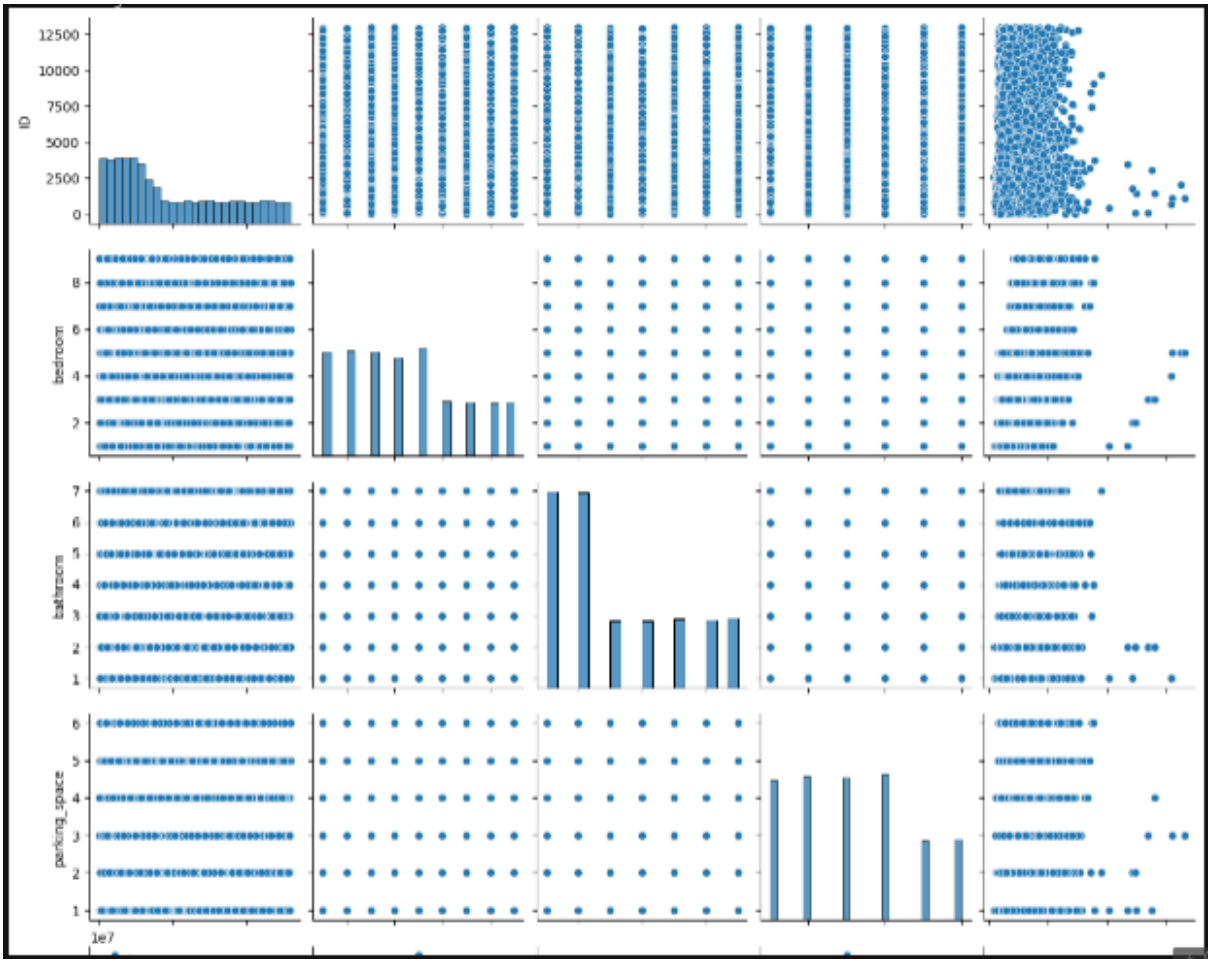

Figure 1.0: pair plot of the training data

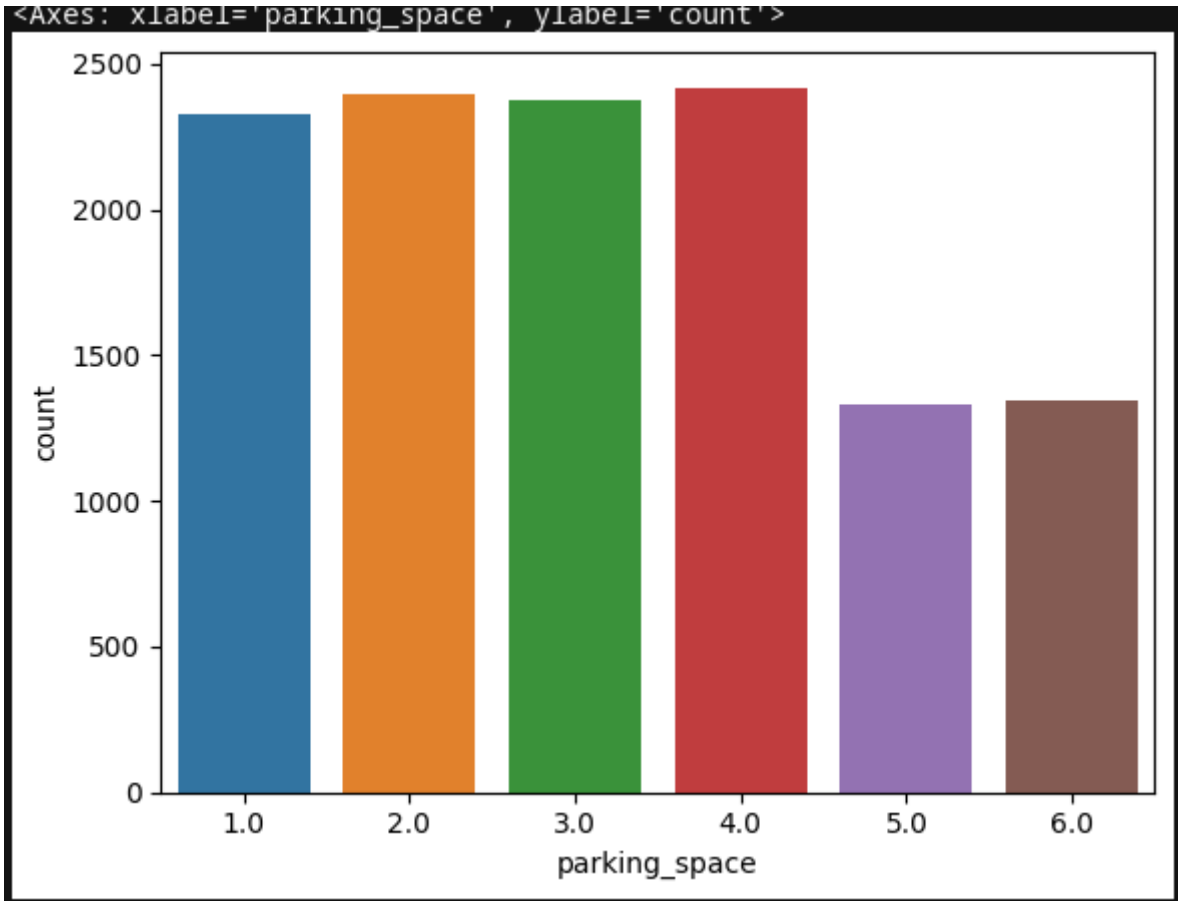

Figure 1.1: count plot of the `parking_space` column

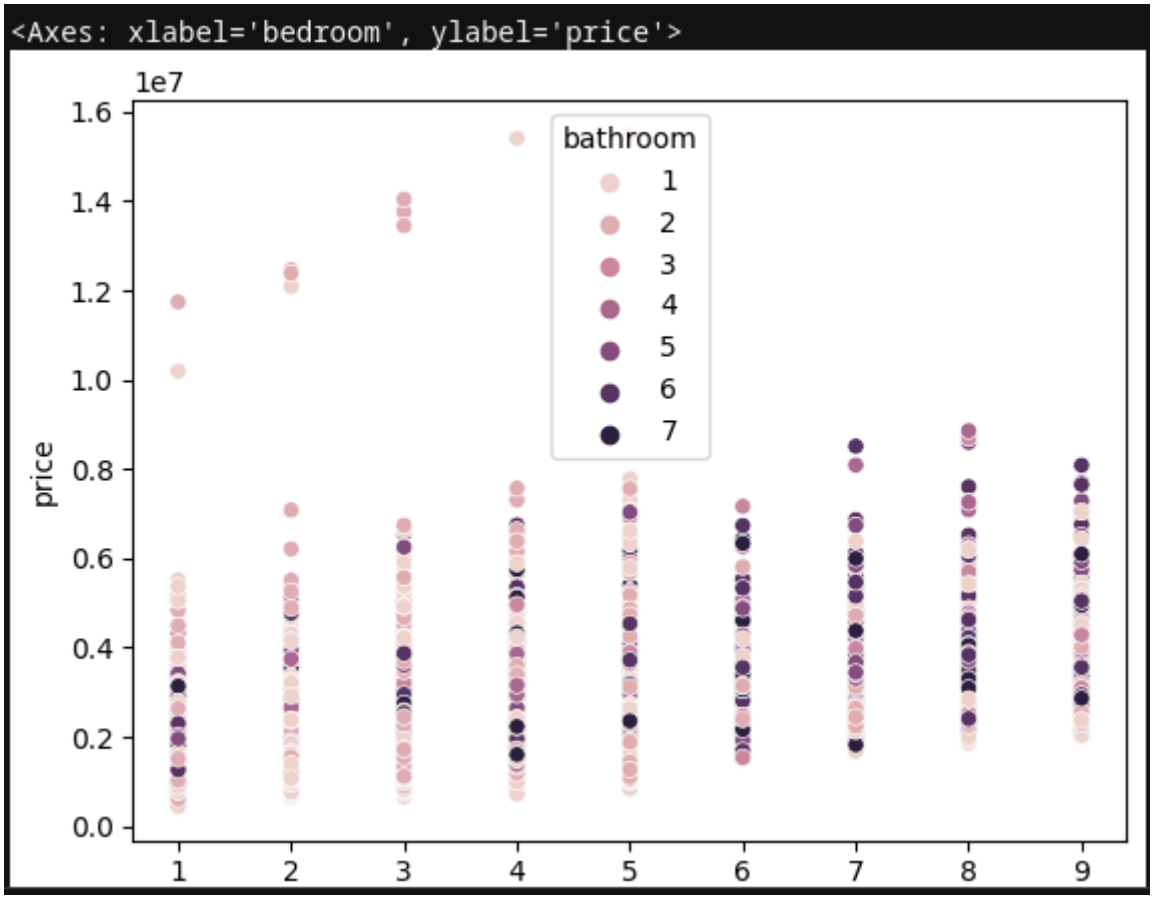

Figure 1.2: scatterplot of the `price` column vs. the `bedroom` column, with the `bathroom` column as hue

Model Selection and Evaluation

For the predictive modelling phase, we opted for various regression algorithms known for their efficacy in forecasting housing prices. The selected models include:

- CatBoostRegressor
- RandomForestRegressor
- GradientBoostingRegressor
- LinearRegression
- DecisionTreeRegressor

- XGBRegressor
- KNeighborsRegressor
- LGBMRegressor

I split the preprocessed training data into training and testing sets (70% training and 30% testing) to train and evaluate the models. The CatBoostRegressor was chosen as the primary model due to its superior performance and ability to handle categorical features without extensive preprocessing. The model's hyperparameters were tuned to optimize its predictive capabilities.

CatBoost is a machine learning library that uses gradient boosting to build predictive models. It is designed for use on problems like classification and regression that have a large number of independent features. CatBoostRegressor is a regression model that is built on top of CatBoost. It is specifically designed for problems where the target variable is continuous, such as house prices or stock prices. CatBoostRegressor uses a variety of techniques to improve its predictive accuracy, including:

- Feature importance: CatBoostRegressor automatically identifies the most important features in the dataset and uses them to build the model.
- Ensemble learning: CatBoostRegressor builds multiple models and then combines their predictions to improve accuracy.
- Gradient boosting: CatBoostRegressor uses gradient boosting to iteratively improve the model's predictions.

How does CatBoost work?

CatBoost works by first creating a decision tree for each feature in the dataset. The decision trees are then combined to form a model. The model is trained by

iteratively adjusting the weights of the decision trees to minimize the error in the training data.

CatBoost is a powerful machine-learning library that can be used to solve a variety of problems. It is easy to use and can be trained on large datasets. CatBoost is a good choice for problems where the target variable is continuous and there are a large number of independent features.

Benefits of using Catboost

- Accuracy: CatBoost is a very accurate machine-learning library. It can achieve state-of-the-art results on a variety of problems.
- Speed: CatBoost is a fast machine learning library. It can train models on large datasets quickly.
- Ease of use: CatBoost is easy to use. It has a simple API that makes it easy to build and train models.
- Flexibility: CatBoost is flexible. It can be used to solve a variety of problems.

Result

On the test set, the CatBoostRegressor's performance was assessed using the root mean squared error (RMSE) measure. The smaller the RMSE, the more accurate the model's ability to predict housing prices. I acquired a competitive RMSE score after training and fine-tuning the CatBoostRegressor, suggesting its efficiency in predicting housing prices.

Recommendations

The following recommendations are made for future work on this project:

1. The dataset could be expanded to include more houses and more features.

2. Other machine learning models could be evaluated to see if they can provide better results.
3. The model could be deployed in production to make predictions on real-time data.

Conclusion

The project's conclusion highlights a critical breakthrough: the development of a highly durable forecasting model adapted to the intricate dynamics of Nigerian housing values. The evidence resoundingly validates the viability of such a model's building through a painstaking amalgamation of data exploration, feature engineering, and advanced modelling approaches. The ramifications are significant, indicating a transformative opportunity for Wazobia Real Estate Limited to rethink its pricing methods. Wazobia Real Estate Limited could benefit from the model's computational power. The program has the potential to recalculate the company's pricing accuracy to unprecedented levels by integrating detailed patterns from the data. The dynamic insights gained from the model's predictive prowess provide an opportunity for Wazobia Real Estate Limited to not only modify existing pricing approaches but also introduce a new level of competition. With the ability to decipher the intricate interplay of variables influencing house prices, the company can seamlessly tailor its offerings to align with market trends and customer expectations, propelling it to the forefront of the Nigerian real estate sector with a potent blend of accuracy and strategic agility.