**INTRODUCTION**

**Background:**

The Nigerian real estate market is rapidly evolving, with property prices influenced by factors such as location, property size, infrastructure, and economic conditions. However, the lack of a standardized, data-driven approach to property evaluation often makes results price inconsistence, making it difficult for buyers, sellers, and investors to make informed decisions.

This project aims to develop a machine learning model to predict house prices in Nigeria using Ridge Regression. By analyzing key property features and market trends, the model will provide accurate price estimates, helping to bring transparency and efficiency to the real estate sector.

**Objectives:**  
The main objective of this study is to develop a machine learning model that predict the relationship between the property features and house prices in Nigeria.

**Scope:**  
This dataset contains Houses listings in Nigeria and their prices based on Location and other parameters such as number of bedrooms, number of toilets, number of bathrooms, house type and parking space.

**DATA COLLECTION AND PREPARATION**

**Data Source:**

This dataset was sourced from gaggle an online repository, stored in a csv file. This dataset provides detailed of house prices, location, type of house and property features.

**Variables in the dataset include:**

* **bedrooms:** number of bedrooms in the houses
* **bathrooms:** number of bathrooms in the houses
* **toilets:** number of toilets
* **parking space**
* **title:** house type
* **town:** town in which the house is located
* **state:** state within Nigeria in which the house is located and finally
* **price:** the target column.

**Data Processing:**

* **Handling Missing Values:** No missing value was detected all columns and rows was checked and no missing value was found.
* **Handling Outliers:** Outliers in the "Price" columns were detected and adjusted to improve the model's accuracy.

**MODEL SELECTION AND TRAINING**

**Model Selection:**

Ridge Regression was chosen as the predictive model for house price prediction in Nigeria. Ridge Regression is a variation of linear regression, which helps prevent overfitting by penalizing large coefficients. The dataset contains categorical features such as "Title” comprises of house type (Detached Duplex, Terraced Duplexes, Semi Detached Duplex, Detached Bungalow, Block of Flats, Semi Detached Bungalow, Terraced Bungalow) which were encoded using one-hot encoding. Ridge Regression performs well with such encoded categorical variables.

**Train-Test Split:**

**The dataset was divided into two parts:**

* Training Set (80%): Used to train the model.
* Test Set (20%): Used to evaluate the model’s performance.

The independent variables (X) for the model are: "Bedrooms", “Parking Space”, “Title” columns, while the dependent variable (y) was the "Price".

**MODEL PERFORMANCE & EVALUATION**

The **Mean Absolute Error (MAE)** was used as the primary metric for model evaluation. MAE measures the average magnitude of errors in the model’s predictions, making it a good choice for this regression problem.

**Results:**

* **Baseline Model MAE: 77,032,009 (a simple guess or average of the target values).**
* **Model (Train) MAE: 62,352,238**
* **Model (Test) MAE: 62,422,800**

**The test MAE of 62,422,800 indicates that the model’s predictions are closer to the actual values than the baseline model (77,032,009), demonstrating improved predictive accuracy.**

**LIMITATIONS**

1. **Data Availability and Quality – The accuracy of the model is limited by the availability and reliability of real estate data in Nigeria. Missing or inconsistent data could affect predictions.**
2. **Feature Representation – The categorical feature "title" (which includes house type) was encoded, but different types of houses may have varying price determinants that are not fully captured.**
3. **Market Fluctuations – House prices are influenced by external factors such as inflation, interest rates, and government policies, which are not included in the model.**
4. **Limited Geographical Scope – this model primarily focuses on specific regions after the data was trimmed, predictions may not generalize well to other Nigerian cities with different real estate dynamics.**

**CONCLUSION & RECOMMENDATIONS**

**The Ridge Regression model provided a significant improvement over the baseline, demonstrating that machine learning can be effectively used to predict house prices in Nigeria. Some features are highly correlated. For example, the more the bedrooms, the more the bathrooms, the number of bathrooms is not needed for the model to predict the price of a house. The number of parking spaces in a house contributes to the overall price of the house with one parking space contributing about 3 million naira to the overall price of the house.**

**While the model performs well in its current form, incorporating additional features and expanding the dataset can enhance accuracy and generalizability. This study highlights the potential of data-driven approaches in real estate pricing, offering a more objective and transparent valuation method for buyers, sellers, and investors.**

**Recommendations:**

1. **Incorporate Additional Features – Including economic indicators (e.g., inflation rate, exchange rate) and location-specific attributes (e.g., proximity to schools, roads) could enhance prediction accuracy.**
2. **Feature Engineering – Consider transforming highly correlated features into composite variables (e.g., total number of rooms) to reduce redundancy.**
3. **Data Expansion – Collecting more real estate data from multiple sources can improve the model’s robustness and generalizability.**
4. **Regular Model Updates – The real estate market changes over time, so periodic retraining of the model with updated data will improve accuracy.**
5. **Exploring Other Regression Models – Testing models like Gradient Boosting, Random Forest, or Neural Networks may yield better predictive performance.**