

Executive Report

From the Exploratory Data Analysis (EDA) conducted, it was observed that the train data has 1,400 rows and 7 distinct features comprising;

- i. ID: unique identity of each property (integer)
- ii. loc: the 36 states of the federation (object)
- iii. title: kinds of properties as tenable in the country (object)
- iv. bedroom: the number of bedrooms in a property (float)
- v. bathroom: the number of bathrooms in a property (float)
- vi. parking_space: number of parking spaces in a property (float)
- vii. price: the property price and target variable (float)

Similarly, the table of descriptive statistics showed that;

- i. An average property does comprise 4 bedrooms, 3 bathrooms, 3 parking space and cost around N2million, on the average.
- ii. The least expensive property does have 1 bedroom, 1 bathroom, 1 parking space and mostly cost a little above N400,000
- iii. Whereas the most expensive properties do comprise 9 bedrooms, 7 bathrooms and 6 parking space, with price exceeding over N16million.

From the Skewness analysis of the train data, it could be seen that the data is positively skewed with all features having majority of their observations to the right-hand side of the distribution. Similarly, the Kurtosis function indicates a Platykurtic distribution among features with the data mostly stretched around the mean.

The target variable (price), however, exhibits a Leptokurtic distribution with the data points having long tails and presence of outliers. This outliers in the prices variable indicates the existence of prices of some other properties which are exorbitant and way more expensive from regular property prices within the country.

Similarly, running cross-tabulation of property types by prices, it's obvious that Mansions are the most expensive property types within the country with average prices close to about N4million. This is closely followed by Penthouses costing over N2.6million and Detached duplexes costing over N2million on average. The least expensive properties are Flats with average price around N1.7million, Apartments at N1.6million and Cottages costing around N1.3million on average.

Location of a property seem to be another major factor to prices of the property. For instance, properties in Lagos, Bayelsa and Rivers tend to be much expensive than from any other part of the country. Similarly, Sokoto, Ebonyi and Kebbi tend to complete the bottom list of states with the least expensive property prices, on average.

Deep-diving to a more granular level to see property prices by property type across locations, it could be seen that Mansions in Lagos tend to be more expensive than any other property type across the country. This is closely followed by Mansions in Rivers, Bayelsa and Delta states. Similarly, Penthouses, Mansions and Detached duplexes across Lagos, Ogun, Cross River, Akwa Ibom and Anambra makes the list of the top 10 states with the most expensive properties by average prices. On the flip side, Cottages in Kogi, Yobe,

Gombe, Kaduna, Abia and Kebbi tend to top the list of the states with the least expensive properties by average price.

Consequently, the Correlation matrix and Heatmap of the numeric features shows a couple of relationships among variables. For instance, there exists a moderate positive correlation between number of bedrooms and property prices – meaning the higher the number of bedrooms in a property, the more likely the price of the property will be higher. Similarly, there exists a weak positive correlation between number of bathrooms in a property and price of the property.

Looking at number of bedrooms, bathrooms and parking spaces as a function of price, it was obvious that properties with 3 to 5 bedrooms tend to be more expensive than properties with much or less number of bedrooms. Similarly, some of the most expensive properties only have about 1 to 2 bathrooms some of the most expensive properties tend to have only about 1 to 3 parking spaces.

Inspecting distribution of bedrooms, bathrooms and parking spaces across property types within the country, it is obvious that there are mostly between 1 to 5 bedrooms in most properties, with very few other properties having more than 5 bedrooms. Similarly, most properties do have only about 1 to 2 bathrooms, whereas only about few others do have more than 2 bathrooms. Parking space is another feature which is observed to be just between 1 to 4 for most property types, with few others having more than 4 parking spaces.

Looking also at distribution of property types across states of the federation, there seem to be more of Flats, Apartments and Townhouse in the country. With the presence of few Cottages being owned among property owners in the country. Similarly, there seem to be more properties located in states like Kaduna, Anambra and Benue states. With least number of properties seen across states like Cross River, Zamfara and Imo states, respectively.

Behind preprocessing the train data by label-encoding categorical variables and using the SimpleImputer function to handle missing values, then comes the choice of a befitting regression model to deploy in building the predictive model. Although several models such as Linear, RandomForest and Gradient Boosting were attempted and their performances evaluated, XGB Regressor was eventually deployed based on the fact that it has fast execution speed, higher model performance and based on its ability to learn from non-linear patterns (this is not to neglect its ability to easily overfit). The XGB model was able to predict prices of properties with a root mean squared error of over N570,000.

Consequently, a bagging ensembling techniques was similarly deployed over the XGB model to improve its performance. The Bagging model showed some improvements in model performance by predicting property prices with an error of N560,000.

Since the Bagging predictive model tend to perform better, this model was deployed in predicting property prices on the Test set.

Similarly, the actual property price and the predicted price from the model built when compared, both tend to exhibit some linear positive correlation and tended towards same plane.

Looking at the R^2 score value of the model, about 75% variations in prices was shown to have been explained by features on the train data. An indication of how well the model fits the dataset.

With these analysis and discovery, I am confident of the ability of the Wazobia Real Estate Limited to be able to make informed data-driven decision around house prices and the real estate sector, within the country.