

# Report: Building a Predictive Model for House Price Forecasting

By Bashir Amina Garba

## Introduction:

I am pleased to present the report detailing the process of building a predictive model using linear regression to forecast house prices based on various features of residential properties. This project aimed to leverage the housing price dataset from Kaggle to develop a robust model that can accurately predict house prices, thereby assisting in decision-making processes related to real estate investments and valuation.

## Dataset:

The dataset provided contains information about residential properties, including features such as squarefeet, bedrooms, bathrooms, price, etc. The target variable is the sale price of the properties, which we aimed to predict using the available features.

## Methodology:

### 1. Importing Libraries and Dataset:

- Utilized Python programming language along with libraries such as pandas, numpy, matplotlib, seaborn, and scikit-learn for data manipulation, visualization, and modeling.
- Loaded the housing price dataset from Kaggle into the notebook using the `read_csv()` function provided by the pandas library.

### 2. Exploratory Data Analysis (EDA):

- Conducted thorough exploratory data analysis to understand the dataset's structure and gain insights into the features and target variable.
- Checked for missing values, summarized statistics, visualized the distribution of the target variable (sale price) using histograms, correlation between two features using scatterplot and a correlation Heatmap.

### 3. Data Preprocessing:

- Encoded categorical variables into numerical format using one-hot encoding to prepare the dataset for modeling.
- Split the dataset into features (X) and the target variable (y) and further split it into training and testing sets.

### 4. Model Building:

- Created a linear regression model using scikit-learn's **LinearRegression** class.
- Fitted the model to the training data to learn the coefficients for each feature that best fit the training data.

### 5. Model Evaluation:

- Evaluated the performance of the linear regression model on the testing set using metrics such as Mean Squared Error (MSE) and R-squared.
- Calculated the MSE and R-squared to measure the accuracy and goodness-of-fit of the model.
- The deployed model can be utilized to forecast house prices based on various

features of residential properties in real-world scenarios.

**Conclusion:**

In conclusion, the project successfully developed a predictive model using linear regression to forecast house prices based on the housing price dataset from Kaggle. The model demonstrated promising performance in accurately predicting house prices, thereby providing valuable insights for real estate investment and valuation decisions.