House Location Prediction

Using Machine Learning

Presented by Group 3
21st February, 2025

PROBLEM STATEMENT

- House hunting can be overwhelming—too many options, unclear pricing, and no easy way to find the best location within budget.

- This project uses machine learning to predict the best town for a house based on budget, property type, and key features.

Classification Problem

This is a classification problem as we are predicting a categorical outcome (Town) based on numerical and categorical features.

DATASET & FEATURES

Data Source: Kaggle.com

Feature Variables

- Price
- Number of bedrooms
- Number of bathrooms
- Number of toilets
- Parking space
- Title (property type)
- State

→

Target Variable

Town

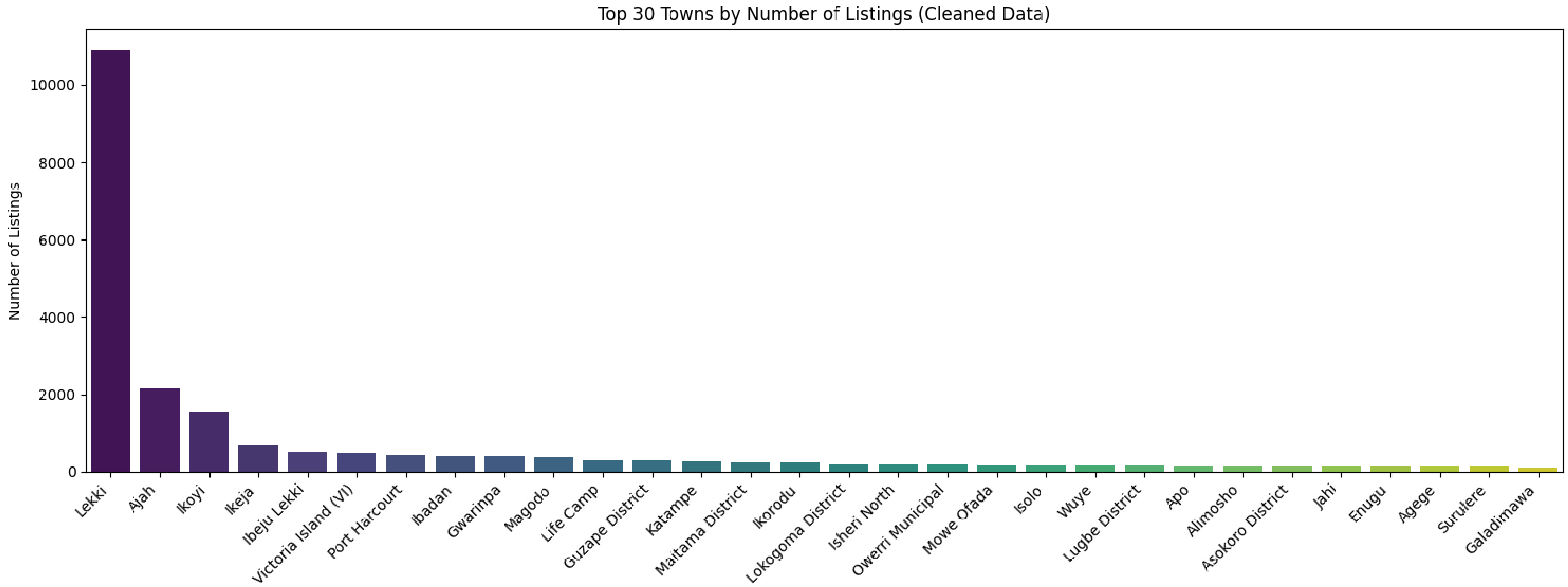
Data Cleaning and Preprocessing

The strategic process

- Checked for missing values and found that there were no missing values

Data Visualization

Distribution of listings across towns (Top 30)

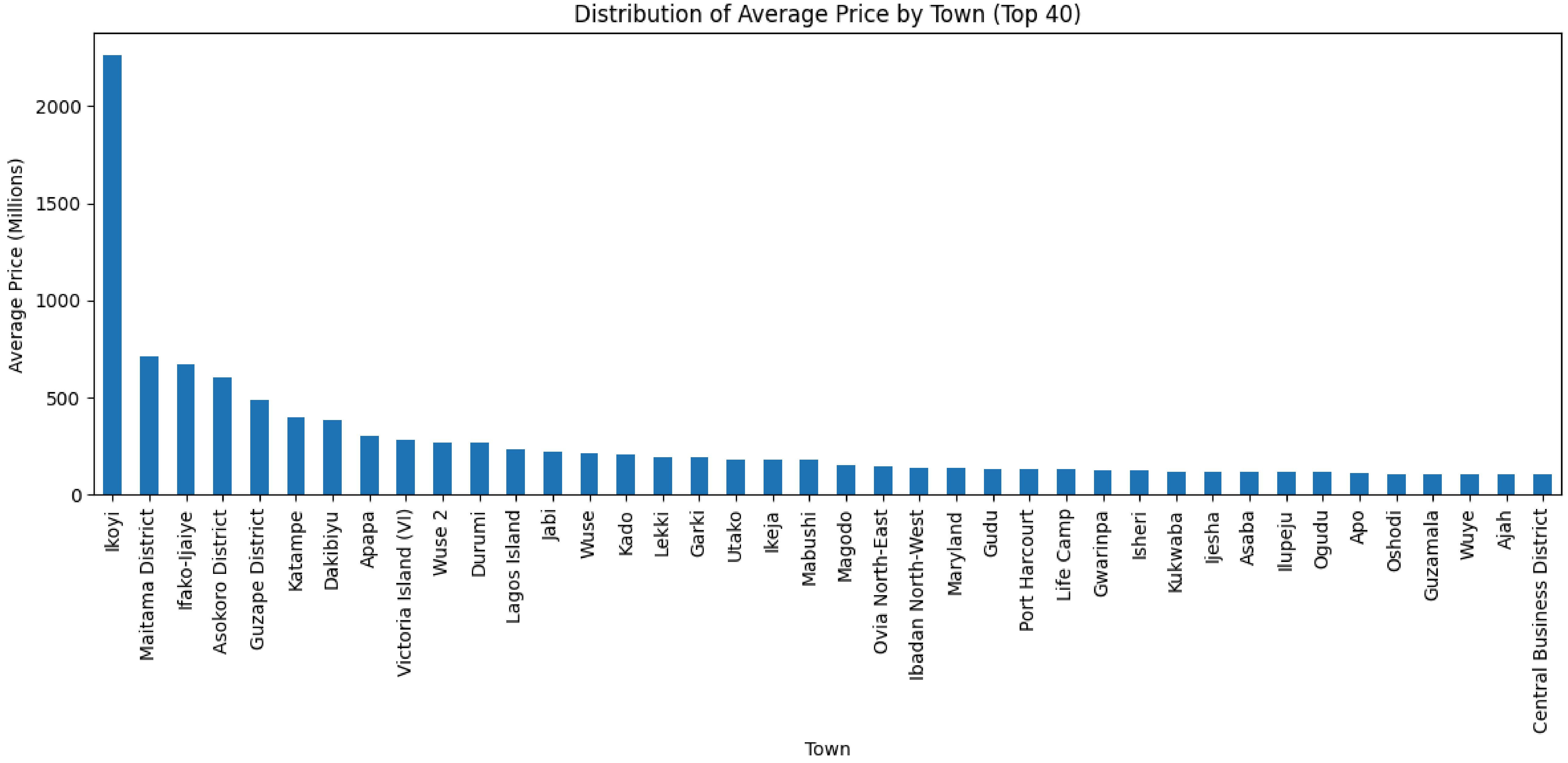


Observations

- Lekki has the highest number of listings, far more than other towns.
- Ajah, Ikoyi, and Ikeja follow but with significantly fewer listings.
- The distribution is highly imbalanced, with most listings concentrated in a few towns.
- Smaller towns have very few listings, which may affect model accuracy.

Data Visualization (cont.....)

Average Price Distribution by Town



Observations

- House prices vary significantly across towns.
- Lekki and Victoria Island have some of the highest average prices.

Our observations after EDA

Town Distribution:

The cleaned data still shows an imbalance in the number of listings across towns, which may affect our model's performance.

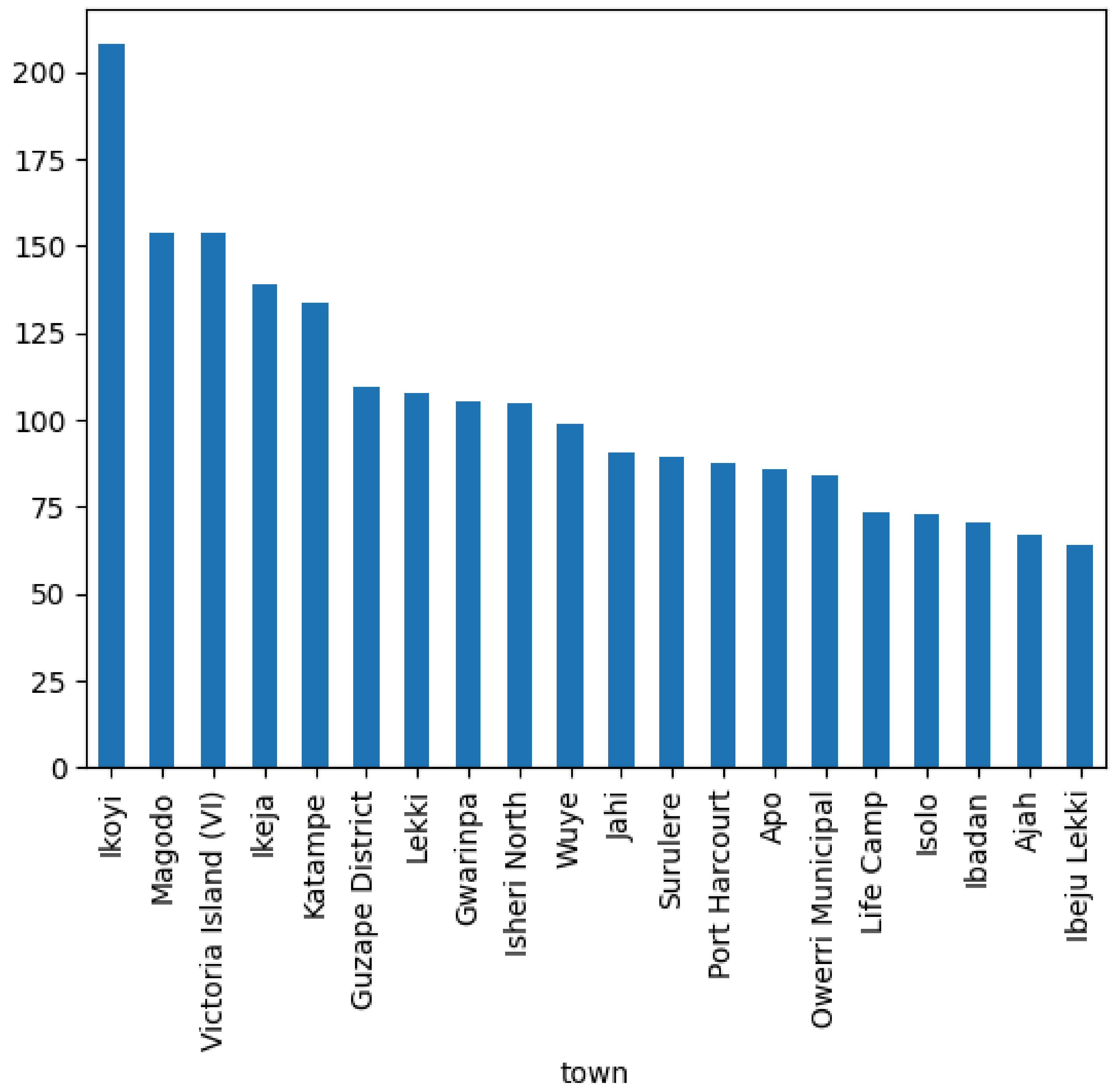
Numerical Features:

The descriptive statistics confirm that price values are in millions, ranging from ₦0.09 million to ₦322 million. Other features like bedrooms, bathrooms and toilets have consistent ranges without extreme outliers.

Remedy

To remedy for the imbalances observed in our town distribution, we created a boolean mask where prices of houses in the various towns is between the 15th and 85th percentiles.

Result



200
175
150
125
100
75
50
25
0
Ikoyi
Magodo
Victoria Island (VI)
Ikeja
Katampe
Guzape District
Lekki
Gwarinpa
Isheri North
Wuye
Jahi
Surulere
Port Harcourt
Apo
Owerri Municipal
Life Camp
Isolo
Ibadan
Ajah
Ibeju Lekki
town

Model Selection & Training

Data were splitted into training (70%) and testing (30%) sets.

Random Forest Classifier Model

- Random Forest Model trained with 100 estimators
- Achieved 71% accuracy
- Model performed well for frequent towns but struggled with less common ones.

Logistic classifier Mode

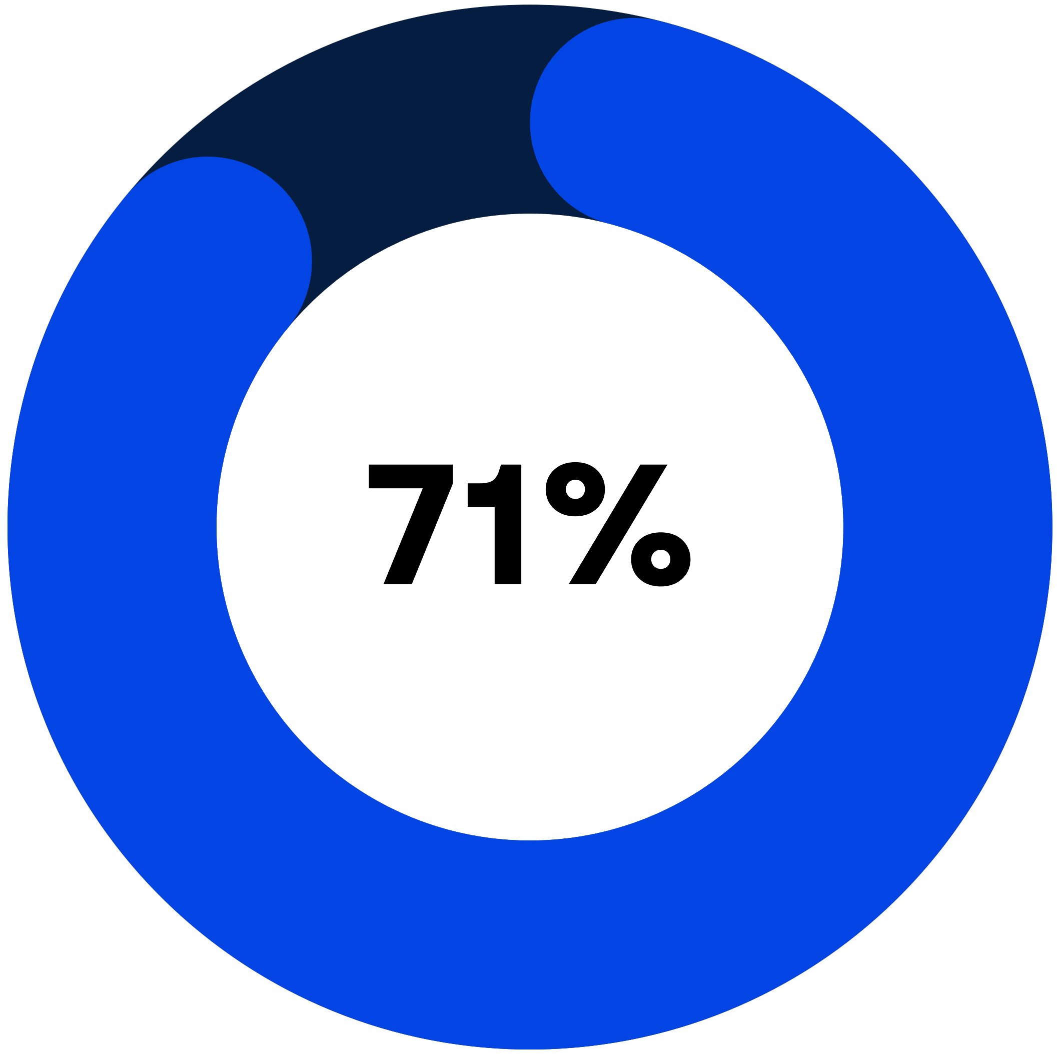
- Achieved 70% accuracy, slightly lower than Random Forest.
- Struggled with imbalanced data, leading to lower recall for some towns.

Support Vector Machine (SVM) Model

- Accuracy 69 % : Slightly lower than Random Forest
- Performed well on highly populated towns but poorly on rare cases.

K-Nearest Neighbors model

- Struggled with classification due to distance-based approach.
- Accuracy 69 % : lower than Random Forest.



Hyperparameter Tuning

- Increased estimators from 100 to 500 in Random Forest.
- Slight improvement but thesame accuracy and increased computation time.

Conclusion & Recommendations

Random Forest was the best-performing model

Challenges faced:

Some towns had too few listings for reliable predictions.

Future Work:

- Collect more data to balance town distributions.
- Add State as a target variable alongside Town
- Experiment with deep learning for improved performance

THANK YOU!