ROME BUSINESS SCHOOL

MASTER IN DATA SCIENCE MANAGEMENT

Customer Demographics Analyses for Product Recommendations

A Case Study of Bokku Stores

VELENZO GROUP 2

- IBEABUCHI IKEM OKWUCHUKWU | 24ODS2109II74
- NWAFOR PEREMOBO BLESSING | 24ODS3452NP52
- ADEOLA SIMISOLA ESTHER | 24ODS1011AS99

Presentation Outline

BACKGROUND OF THE STUDY

LITERATURE REVIEW

METHODOLOGY

RESULTS

CONCLUSION

REFERENCES

Background of the Study

Nigerian Retail & Bokku Stores

Retail sector transforming; e-commerce hitting **$14.1B by 2027** (103M+ users)

Bokku Stores: strong position but faces **overcrowding & unused customer data**

Solution: apply **demographics + machine learning** to personalize products, boost sales, optimize stock, and enhance satisfaction

Problem Statement

Bokku Stores collects extensive **demographic and transaction data**, but lacks a **personalized recommendation system**.

This gap leads to:

Overcrowding and poor customer flow in key outlets	**Generic recommendations** that ignore customer attributes	**Wasted inventory** and **missed upsell opportunities**

Literature Review

Customer Demographics & Retail:

- Age, gender, income, and location strongly influence consumer purchasing patterns.
- In Nigeria, targeted marketing is critical due to diverse socio-economic conditions.

Recommendation Systems:

- Traditional Methods: Collaborative & content-based filtering; issues with cold start & data sparsity.
- ML & Hybrid Models: Integrate demographics, outperform traditional methods; Park et al. (2019) showed a 17% targeting accuracy increase.

Literature Gaps:

- Few Nigerian/African studies using ML + demographics for recommendations.
- Need for localized, data-driven frameworks that match infrastructure realities.

Literature Review

Related works

Author(s) & Year	Problem Addressed	Approach Used	Gap / Limitation
Park et al. (2019)	Generic recommendations ignoring demographics	Integrated age, gender, location into recommender system	Focused on e-commerce; no retail store case study
Al-Gasawneh et al. (2021)	Underutilization of customer data in emerging markets	Customer segmentation using demographic attributes	Did not implement real-time recommendation system
Padlee et al. (2019)	Inefficient demographic targeting in retail marketing	Demographic segmentation to improve campaign response	Did not combine demographic and transaction data

Table 1: outlines prior studies, their focus, and gaps, highlighting the need for our demographic-based real-time recommendation system.

Methodology
Data & Preparation

Research Method

Design: Quantitative + predictive ML modeling

Data: POS, CRM, loyalty records from 6 branches (**38,765 transactions**)

Preprocessing: Clean missing values, encode categories, standardize prices, drop duplicates

EDA Focus: Age/gender distribution, top products & locations, spend patterns

Methodology

Data & Preparation

Research Design

- Quantitative + Predictive modeling using ML algorithms.

Data Collection

- POS, CRM, and loyalty program data from 6 Bokku Stores branches (38,765 transactions).

Data Preprocessing

- Remove missing values (critical fields: Product Price, Items).
- Encode categorical variables (Gender, Items, Location).
- Standardize prices (remove ₦ symbol).
- Drop duplicates & irrelevant columns.

Exploratory Data Analysis (EDA)

- Age & gender distribution
- Top products & locations
- Spend patterns

Methodology

Modeling & Tools

Model Development

- Logistic Regression
- Decision Tree
- Random Forest

Model Evaluation

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix.

Tools Used

- Python
- Pandas
- Scikit-learn
- Excel
- Matplotlib
- Seaborn.

Results

EDA Insights

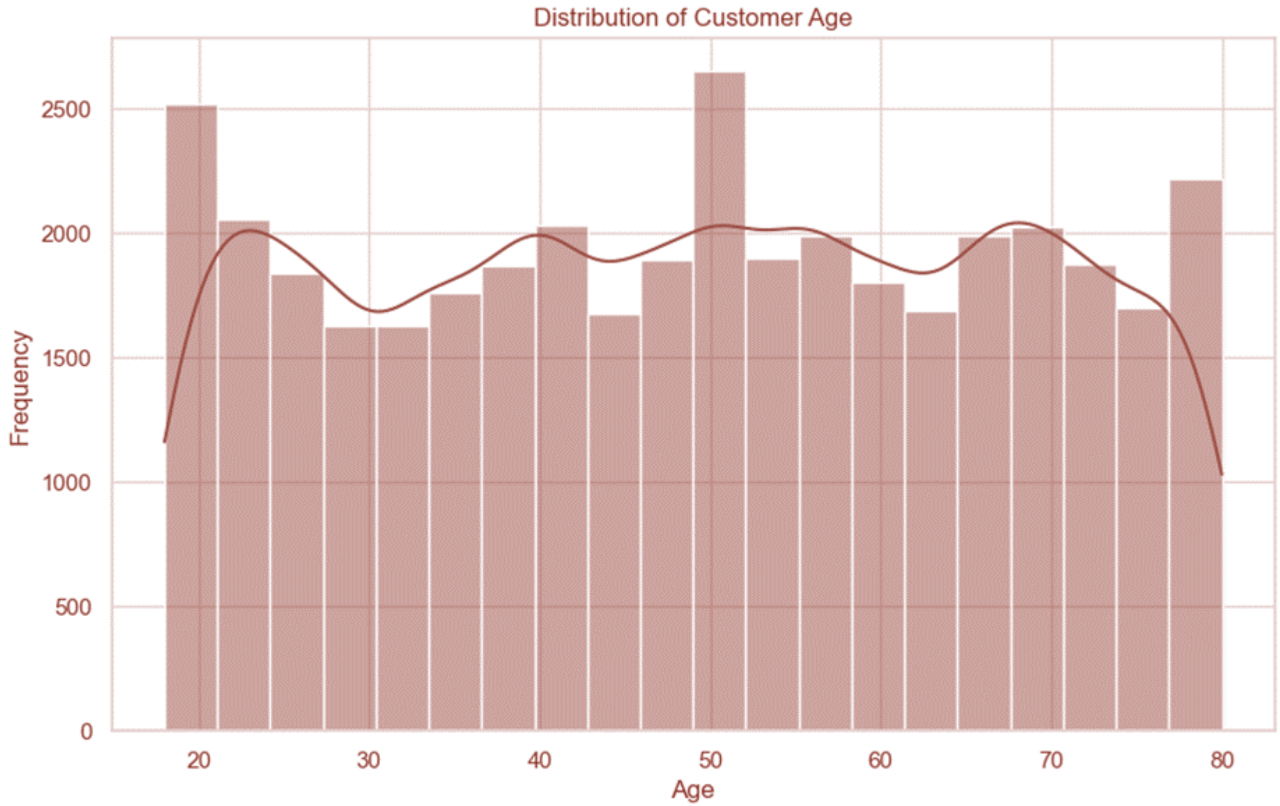

Fig 1 : Shopping frequency by age group.

The chart indicates that the highest shopping activity occurs among customers aged 25 – 45 years, identifying them as the store's core shopper segment.

Results
EDA Insights

Fig 2: Customer count by gender.

The gender distribution is nearly equal; however, male customers tend to spend more per visit compared to female customers.

Results

EDA Insights

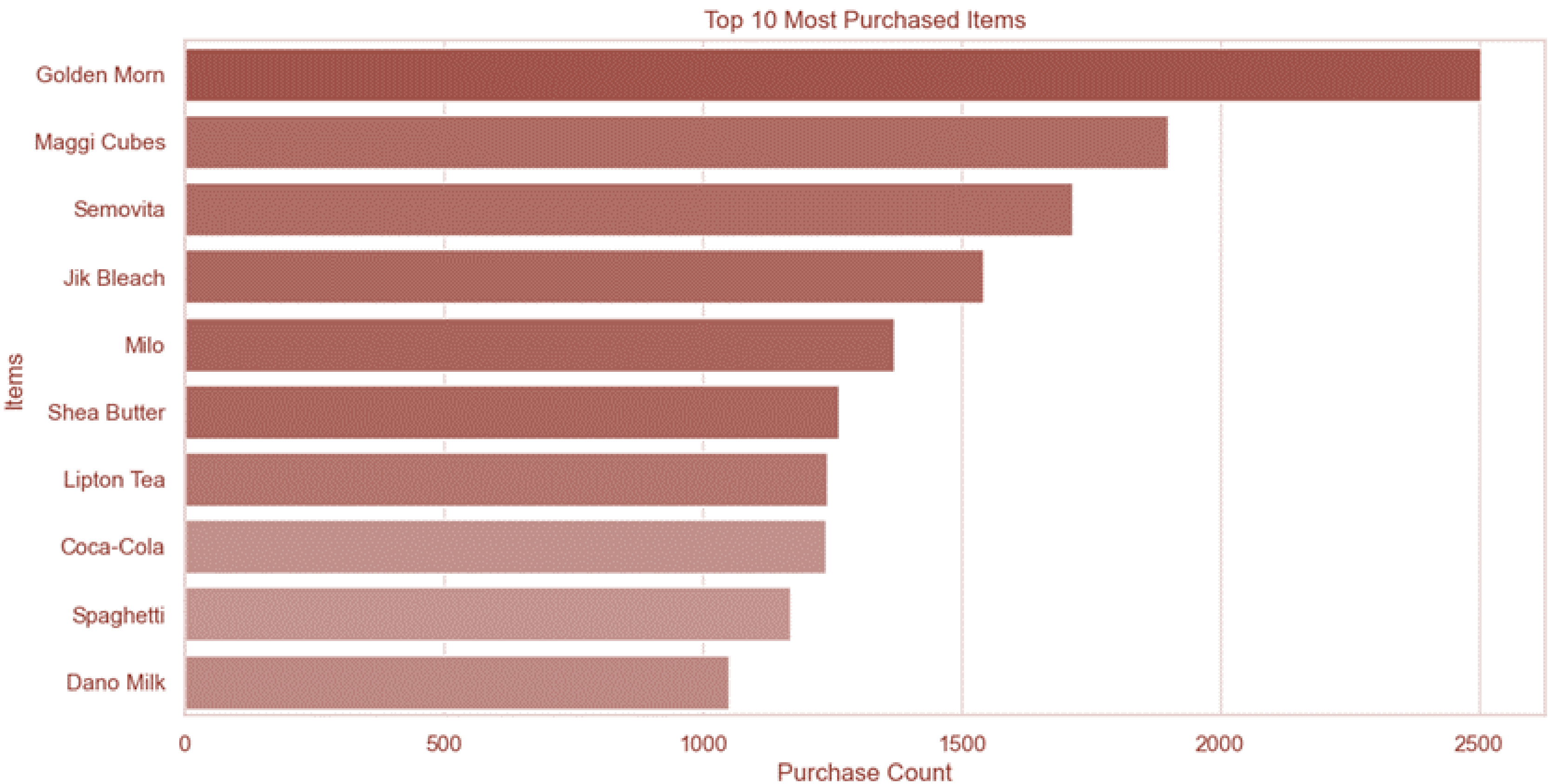

Fig 3: Purchase count by item.

The most frequently purchased products are Golden Morn, Maggi Cubes, and Semovita.

Results

EDA Insights

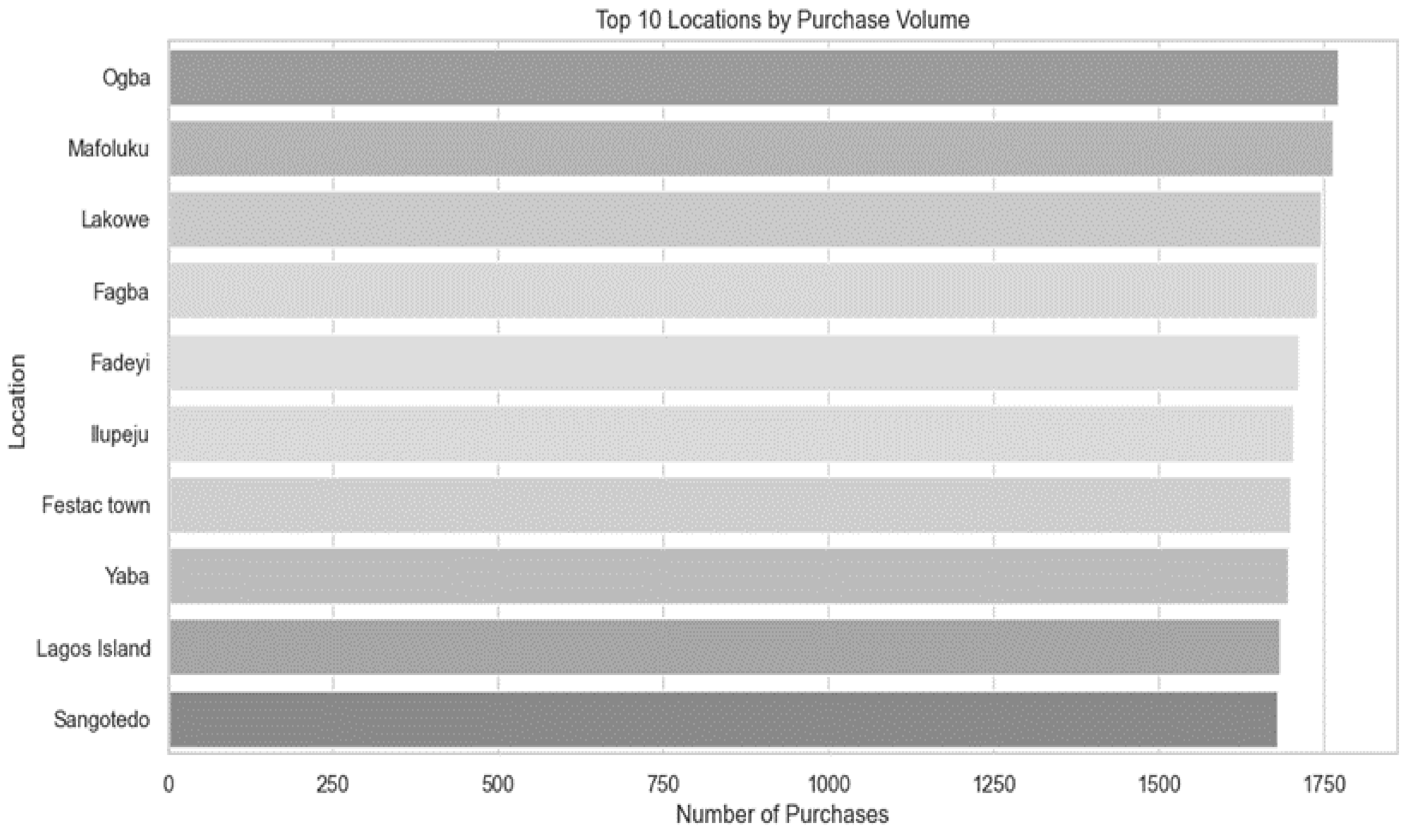

Fig 4 : Number of purchases by location.

The highest purchase volumes are recorded in Ogba, Mafoluku, and Lakowe, identifying them as key sales hotspots.

Results

Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	72.6%	Moderate	Low	Fair
Decision Tree	83.44%	High	High	Excellent
Random Forest	80.29%	High	Consistent	Balanced

Table 2: compares model performance metrics, showing Decision Tree as the most accurate and Random Forest as the most deployment-ready.

Results

Model Performance Comparison (Pictorial View)

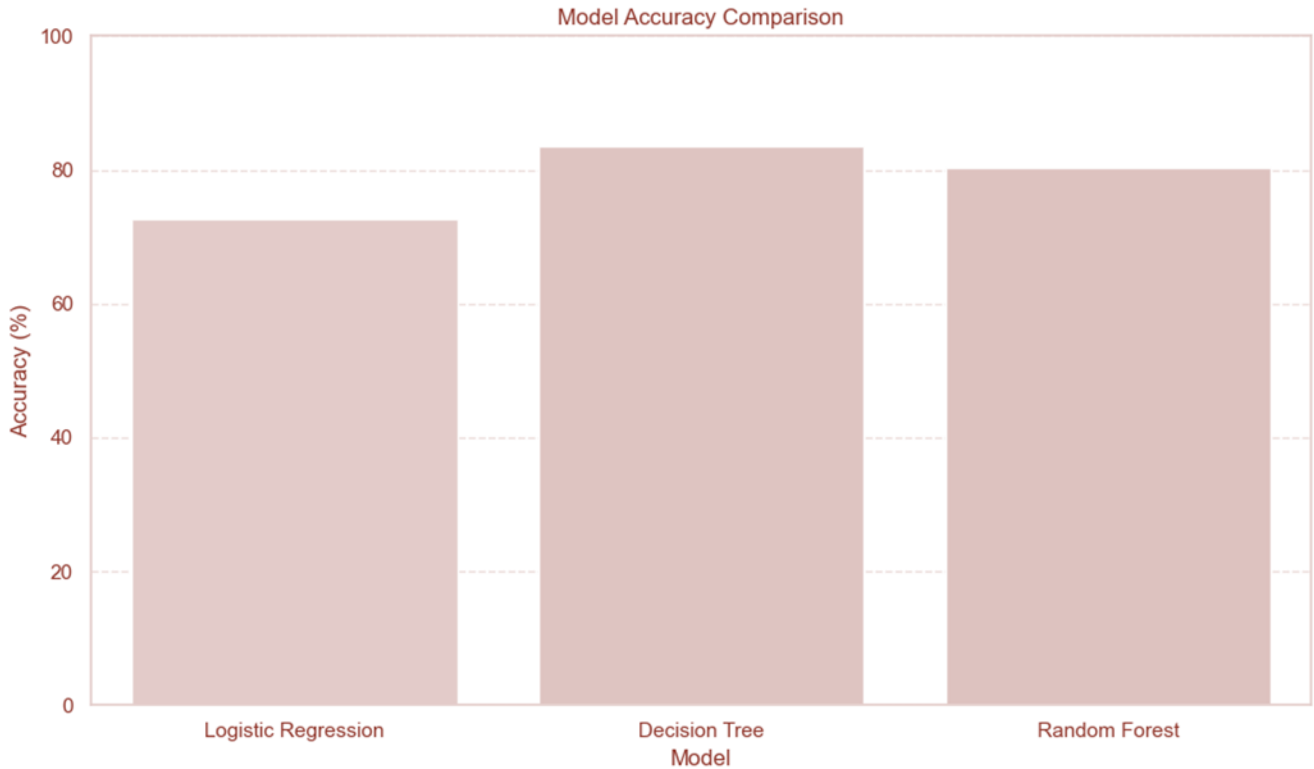

Fig 5: Model accuracy comparison.

Among all models tested, Decision Tree achieved the highest accuracy, followed closely by Random Forest.

Results

Confusion Matrix – Random Forest Model

Metric	Result
Accuracy	85%
Precision	0.89 (Golden Morn)
Recall	0.84
F1 Score	0.86

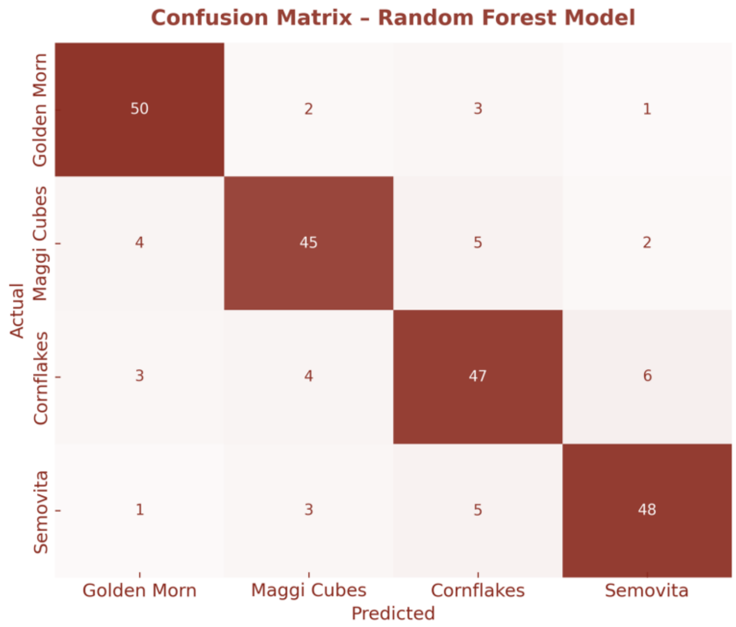

Table 3 & Fig 6: Confusion matrix for the Random Forest model.

This model showed strong and balanced classification performance, making it more reliable for real-world deployment.

Results

Business Insights from Model Output

Analysis confirms that age and location are the most influential predictors of purchase behavior.

High-demand branches (Ogba, Mafoluku, Lakowe) should get inventory priority.

Middle-aged male shoppers show potential for bulk purchase promotions, while essentials like Golden Morn and Maggi are ideal loyalty reward triggers.

Integrating the recommendation system into POS can enable real-time suggestions at checkout.

Conclusion

This study proves that even with basic demographic data, Bokku Stores can accurately predict customer preferences and make smarter business decisions. By choosing the Random Forest model, we’ve prioritized not just accuracy, but long-term reliability. The next step is to integrate this system into real-time operations transforming every checkout into an opportunity for smarter stocking, targeted marketing, and stronger customer loyalty.

References

Park, S., Kim, Y., & Choi, H. (2019). Enhancing recommendation systems with demographic data. Journal of Retail Analytics, 15(3), 45–58.

Al-Gasawneh, J., Anuar, M. M., & Dacko-Pikiewicz, Z. (2021). Customer data utilization in emerging markets: A retail perspective. Marketing Intelligence & Planning, 39(5), 625–641.

Padlee, S. F., Thaw, C. Y., & Abdul Rashid, Z. (2019). Demographic segmentation and its influence on purchase decisions. International Journal of Retail & Distribution Management, 47(11), 1161–1177.

(Full reference list available in project report)

Thank You!