Capstone Project Report

AI-Driven Loyalty Program Optimization
Retail Supermarkets in Kenya

Business Understanding

Retail supermarkets in Kenya operate in an increasingly competitive environment, where evolving consumer behavior demands more sophisticated engagement strategies. Traditional loyalty programs, often static in design, have proven insufficient in utilizing the wealth of transactional and behavioral data collected at the point of sale. This shortfall has made it difficult for retailers to deliver tailored rewards that align with individual customer preferences, ultimately weakening retention efforts and diminishing the impact of promotional campaigns. This project leveraged artificial intelligence and machine learning to modernize loyalty systems—transforming them into intelligent, adaptive tools for personalized marketing and data-driven decision-making across product categories and customer segments.

Problem Statement

Although widely implemented, loyalty programs in Kenyan supermarkets have traditionally fallen short due to their limited ability to respond to evolving customer behaviors and purchase patterns. Retailers have historically lacked the analytical tools to understand customer churn, evaluate the true effectiveness of promotions, or craft personalized offers that drive spending at the category level. These gaps have led to generic marketing efforts, inefficient use of promotional budgets, and reduced customer engagement. Through this project, we addressed the need for a smarter, predictive framework that enables supermarkets to retain high-value customers and unlock revenue growth through tailored, data-driven loyalty strategies.

Project Objectives

- Developed machine learning models to segment customers based on purchase behaviour and loyalty metrics.
- Predicted customer churn to identify at-risk customers.
- Determined high value customers based on observed purchasing patterns for prioritization.

Data Preparation

The data preparation stage involved thorough cleaning and integration of the two datasets, including transaction logs, customer demographic profiles, loyalty card usage, and promotional activity records. Feature engineering was carried out to derive key variables such as recency, frequency, and monetary value (RFM), along with category-specific spending and reward redemption patterns, all aimed at improving model performance. Standard preprocessing methods were applied to address missing values, inconsistencies, and outliers. The data was then aggregated at both the customer and transaction levels to enable effective segmentation and predictive modeling.

Modeling

The project employed a combination of supervised and unsupervised machine learning techniques to derive actionable insights from the data. Clustering algorithms such as K-Means was utilized for customer segmentation, successfully uncovering distinct behavioral groups including loyal customers, price-sensitive shoppers, and occasional buyers. Supervised learning models—including Logistic Regression, Random Forest, Decision tree, SVM, Gradient boosting and XGBoost—were applied to predict customer churn with high accuracy. Additionally, we improved the churn prediction with RandomOverSampler and GridSearchCV. We also implemented a simple neural network model (MLP) with two hidden layers to predict loyal customers is implemented. All model pipelines were developed in a modular, scalable framework to ensure adaptability within dynamic retail settings.

Evaluation

1. **Best- Performing Model**:

- The XGBoost model (with regularization) emerged as the best-performing classifier for predicting loyal customers, achieving an accuracy of 87% and F1-score of 81%.

2. **Customer Loyalty Analysis**

- KMeans clustering using RFM (Recency, Frequency, Monetary) features successfully identified four distinct customer segments, supported by a silhouette score of 0.457, suggesting a meaningful cluster separation with clear differences in customer shopping behaviours and loyalty traits.
- The application of RandomOverSampler and GridSearchCV on the Random Forest model improved its ability to detect loyal customers, with a high ROC-AUC score of 0.92(92%), suggesting strong discriminative power between loyal and non-loyal customers.

3. **Churn Risk Identification**:

- The predictive models effectively identified Cluster 1 (customers at **high risk** of churn) — those with infrequent purchases, low spending, and reduced engagement.

4. **Shopping Behavior Insights**:

- Most customers make purchases below KSh 200, indicating frequent low-value shopping.
- Peak shopping days are Friday to Sunday, reflecting weekend shopping and stocking habits.
- High revenue is concentrated in a few key product categories and brands.

Recommendations

1. **Target Promotional and Marketing Campaigns as per Customer Cluster**:

Tailor the loyalty programs based on cluster profiles:

- Cluster 0: High-value loyal customers - reward this group with premium incentives.
- Cluster 1: At-risk customers - develop targeted campaigns especially on promotions to lure these group of customers back to normal purchasing.
- Cluster 2: New customers - onboard them with introductory offers.
- Cluster 3: Occasional spenders - upscale these customers to even spend more at each shopping trip.

2. **Optimize Promotions by Day**:

Align marketing efforts and promotional campaigns with peak shopping days (Fridays to Sundays) to maximize purchase impact.

3. **Personalize Customer Rewards**:

Leverage predicted loyalty scores and customer purchase patterns to offer personalized point-based rewards.

4. **Develop a Store-Level Strategy**:

Use registration location insights to implement location-specific campaigns, particularly in regions with high customer density.

5. **Ensure Continuous Model Monitoring**:

Retrain and recalibrate the classification and deep learning models periodically using latest data to adapt to changing consumer behaviour.

6. **Innovative Product Promotion**:

Promote top-performing product categories and consider bundling them in reward schemes to further boost loyalty.

7. **Upscale Customers to Premium Categories**

Ensure customers shop in premium categories by doing promotions to improve the customer spending (basket value).

CRISP-DM Workflow

- **Business Understanding**
- **Data Understanding**
- **Data Preparation**
- **EDA**
- **Modeling**
- **Evaluation**
- **Conclusions & Recommendations**
- **Deployment** (CRM, segmentation engine integration)

Expected Impact

Metric	Expected Uplift
Retention	+15%
Promo ROI	+25–40%
Basket Size	+10%
Category Growth	+12%
Private Label Sales	+15%