Takealot Loyalty Program: A Data-Driven Strategy for Customer Retention in a Competitive E-Commerce Landscape.

Author: Mashabela Workers

Executive Summary

Overview

This report proposes a redesigned, data-driven loyalty program for Takealot, South Africa's leading e-commerce platform, to address slowing revenue growth (11% in FY2024) amid competition from Amazon, Temu, and Shein. Grounded in Social Exchange and Equity Theories, the program aims to enhance customer retention, increase purchase frequency, and boost customer lifetime value (CLV) through personalized, tiered rewards. Analysis of a synthetic dataset (370 records, cleaned to 211) reveals inefficiencies in current loyalty initiatives, such as TakealotMore and FNB eBucks, which lack personalization and inclusivity for price-sensitive shoppers in underserved areas like the Eastern Cape (30% participation vs. 50-60% in Gauteng/Western Cape).

Key Findings :

1. Comparative Analysis: T-tests indicate non-loyalty members spend more (R3,451.95 avg.) and report higher satisfaction (4.40) than loyalty members (R2,180.95 avg. spend; 3.99 satisfaction), with $p < 0.001$, suggesting the program attracts low-spenders and fails to satisfy members.

2. Regression Insights: A linear model shows a one-unit satisfaction increase correlates with R1,921.48 higher spending ($R^2 = 0.34$, $p < 0.01$), emphasizing satisfaction-driven incentives.
3. Customer Segmentation: K-means clustering (k=3) identifies: High-Value (high spend R4,286.09, low frequency 3.57, mostly non-members); Moderate (balanced metrics); and Low-Value (high frequency 8.29, low spend R410.26, 80% members with high cart abandonment).
4. Trends: High-frequency buyers (≥7 purchases) are all loyalty members but low-value; Electronics and Furniture dominate spending (R165,000 and R75,000 total).

Proposed Model

The tiered rewards system offers points (10% of purchase amount), with levels: Basic (0-500 points), Premium (501-1,500), and Elite (>1,500), providing escalating benefits like free shipping, discounts, and priority support. Personalized offers include 15-25% discounts for low-value customers and early sales access for high-value ones. Engagement incentives (e.g., bonus points for reviews) and regional campaigns (e.g., Mr D partnerships in Eastern Cape) target inclusivity.

Strategic Recommendations and Impact

1. Recruit high-value non-members with elite rewards to raise frequency from 3.57 to 5 (20% projected increase).

2. Reduce cart abandonment (80% in low-income segments) via targeted discounts (50% reduction goal).
3. Boost Eastern Cape enrollment from 30% to 50% through social media/Mr D campaigns (15% increase).
4. Enhance satisfaction with review bonuses (10% overall uplift, leveraging regression findings).
5. Projections demonstrate potential for significant ROI, supported by Python-based analytics (pandas, scikit-learn).

Conclusion

The model addresses Takealot's challenges by shifting focus from low-value retention to high-value recruitment and underserved market expansion, ensuring ethical compliance (POPIA via synthetic data). Limitations include small sample size and outliers; future enhancements suggest longitudinal data and advanced models like DBSCAN. Implementation could strengthen Takealot's market position, driving profitability in a competitive landscape.

Table of Contents

Introduction to the Business Case

Takealot, South Africa’s leading e-commerce platform since 2011, holds a significant 15–20% share of the national e-commerce market. In 2024, the company generated approximately R15 billion in online revenue. Headquartered in Cape Town, Takealot serves as a major online marketplace with a wide range of products, primarily targeting consumers in metropolitan areas.

However, the market is facing significant disruption from global competitors like Amazon, Temu, and Shein, which entered the market with aggressive pricing, rapid delivery, and extensive product ranges. This has led to a slowdown in Takealot's growth; for the fiscal year ending 31 March 2024, the company reported an 11% year-on-year growth in revenue and gross merchandise value (GMV), down from 15% the previous year. The Takealot Group also recorded a trading loss of R253 million, although Takealot.com itself achieved profitability.

The primary business challenge is to enhance customer retention and increase purchase frequency in this competitive landscape. Takealot's current loyalty initiatives, including the TakealotMore subscription service (Standard at R39/month, Premium at R99/month) and partnerships like FNB eBucks, offer benefits such as unlimited free delivery and complimentary subscriptions. However, these programmes lack sufficient personalisation and inclusivity, especially for price-sensitive shoppers in townships and rural areas. Key issues include low loyalty programme participation in regions like the Eastern Cape (30% vs. 50–60% in Gauteng and Western Cape) and high cart abandonment among low-income customers (80%).

This report proposes a standalone, data-driven loyalty programme designed to address these issues by using predictive and prescriptive analytics. The programme is grounded in two key theoretical frameworks:

1. Social Exchange Theory, which highlights the importance of mutual benefit in customer relationships.
2. Equity Theory, which focuses on ensuring fairness in the allocation of rewards.

The programme aims to strengthen retention, increase customer lifetime value (CLV) with tailored rewards, expand accessibility to underserved shoppers, and differentiate Takealot with exclusive perks. Analysis of customer data reveals inefficiencies in the current loyalty offerings, justifying a redesigned approach to recruit high-value customers and improve outcomes for low-value ones.

Description of the Data

The analysis is based on `takealot_customer_dataset.csv`, a synthetic dataset with 370 transaction records designed to simulate Takealot's customer interactions. This dataset was generated using AI tools, ensuring ethical experimentation that complies with South Africa's Protection of Personal Information Act (POPIA) by avoiding the use of real customer data.

While key columns like `Purchase_Amount` and `Customer_Satisfaction` were clean, the `Customer_Loyalty_Program_Member` column had 211 missing values. These were filled with `False`, resulting in a final sample of 106 loyalty members and 105 non-loyalty members for comparative analysis.

Key Variables

The dataset includes 24 attributes, capturing a rich blend of demographic, transactional, behavioral, and loyalty-related metrics. Table 1 summarizes the key variables and their relevance to the loyalty program design.

Table 1: Key Dataset Variables and Their Relevance

Variable	**Description and Relevance**
Customer_ID	Unique identifier (e.g., ZA001-1234). Enables tracking of individual customer behavior for personalized marketing.
Age	Ranges from 22 to 50 years. Informs demographic segmentation for age-specific promotions.
Gender	Male or Female. Supports gender-based targeting for marketing campaigns.
Income_Level	Low, Middle, or High. Critical for identifying price-sensitive segments; 80% of low-income customers exhibit high cart abandonment ("Often").
Province	South African provinces (e.g., Gauteng, Eastern Cape). Enables regional targeting;

Variable	Description and Relevance
	Eastern Cape has 30% loyalty membership vs. 60% in Gauteng.
Purchase_Category	Electronics, Furniture, Groceries, etc. Electronics (R165,000 total) and Furniture (R75,000) dominate spending, guiding reward focus.
Purchase_Amount	Transaction value in ZAR, ranging from R200.10 to R8,600.25. T-test results surprisingly show that non-loyalty members spend significantly more on average (R3,451.95) than loyalty members (R2,180.95).
Frequency_of_Purchase	Number of purchases (2–10). High-frequency customers (≥7 purchases) are exclusively loyalty members, though they tend to have low spending.
Customer_Loyalty_Program_Member	A Boolean value indicating loyalty status. After data preparation, the dataset contains 106 members and 105 non-members.
Customer_Satisfaction	A rating from 1 to 5. T-tests show that non-loyalty members have higher satisfaction (4.40) than loyalty members (3.99). Regression analysis shows a one-unit increase in satisfaction is associated with a R1,921.48 increase in spending.
Cart_Abandonment_Frequency	Categorised as Never, Sometimes, or Often. This variable helps guide discount strategies for specific customer segments.
Discount_Sensitivity	Categorised as Not Sensitive, Somewhat, or Very. This informs pricing interventions for different customer clusters.

Data Preparation

To ensure analytical rigor, the dataset underwent meticulous preparation steps outlined below:

1. **Loading:** The CSV file was loaded using the `pandas` library.
2. **Cleaning:** Missing values in the `Customer_Loyalty_Program_Member` column were filled with `False` to create distinct loyalty and non-loyalty groups for analysis.
3. **Encoding:** Categorical variables such as `Income_Level`, `Discount_Sensitivity`, and `Cart_Abandonment_Frequency` were numerically encoded to be used in machine learning models like k-means clustering.
4. **Scaling:** Numerical features like `Age`, `Purchase_Amount`, and `Customer_Satisfaction` were standardised using `StandardScaler` to ensure they have equal weighting in the clustering algorithm.

Model Design and Build

The loyalty programme is designed as a tiered rewards system, grounded in Social Exchange and Equity Theories. The model leverages predictive and prescriptive analytics to improve customer retention and CLV, with a focus on recruiting high-value customers and enhancing the experience for low-value segments.

Conceptual Framework: Tiered Rewards System

The model is designed to move beyond Takealot's existing initiatives by offering a structured, tiered rewards system that incentivizes repeat purchases and fosters long-term loyalty. Key components include:

1. **Points Structure:** Customers earn 10% of their `Purchase_Amount` in points (e.g., a R1,000 purchase earns 100 points). The programme features three tiers: Basic (0–500 points), Premium (501–1,500), and Elite (>1,500), with escalating benefits like free shipping, exclusive discounts, and priority support.
2. **Personalised Offers:** The model recommends targeted offers, such as 15–25% discounts for low-value customers who frequently abandon their carts, and premium rewards like early access to sales for high-value customers.

3. **Engagement Incentives:** To boost `Customer_Satisfaction`, the programme offers bonus points for activities like writing reviews, referring friends, or engaging on social media.
4. **Regional Targeting:** The model suggests using Mr D partnerships and social media campaigns to increase loyalty programme enrolment in the Eastern Cape from 30% to a target of 50%.

The model aligns with academic literature, which emphasizes tiered loyalty programs for enhancing CLV through behavioral stratification, ensuring theoretical robustness and practical relevance.

Predictive Analytics

The model employs advanced statistical and machine learning techniques to segment customers and predict their behavior and attitudes toward the brand.

1. **T-Tests for Group Comparison** Two-sample Welch's t-tests were used to compare the loyalty (n=106) and non-loyalty (n=105) groups. The results revealed that:

    a. **`Purchase_Amount`**: Non-loyalty members spend significantly more (mean = R3,451.95) than loyalty members (mean = R2,180.95), with $t = -3.88$ and $p = 0.0002$.

    b. **`Customer_Satisfaction`**: Non-loyalty members report higher satisfaction (mean = 4.40) compared to loyalty members (mean = 3.99), with $t = -4.11$ and $p = 0.0001$. **These counterintuitive findings suggest that the current loyalty programme may be inefficient**, potentially attracting lower-spending customers or failing to satisfy its members adequately.

2. **Linear Regression** A linear regression model was built to quantify the relationship between `Customer_Satisfaction` (independent) and `Purchase_Amount` (dependent).

    a. **Results**: The model produced a coefficient (β_1) of **1921.48** and an R^2 of **0.34** ($p < 0.01$).

   b. **Implication**: A one-unit increase in customer satisfaction is associated with an increase of R1,921.48 in spending, confirming that satisfaction-focused incentives can drive significant value.

3. **K-Means Clustering** K-means clustering (with k=3) segmented customers based on `Age`, `Income_Level`, `Purchase_Amount`, `Frequency_of_Purchase`, `Customer_Satisfaction`, `Brand_Loyalty`, and `Discount_Sensitivity`. The analysis identified three distinct clusters:

   a. **High-Value (Low Loyalty):** These customers have the highest average spend (R4,286.09) and satisfaction (4.98) but lower purchase frequency (3.57). They are primarily non-loyalty members and represent a key target for recruitment.
   b. **Moderate (Balanced):** This group has a balanced mix of loyalty and non-loyalty members with moderate spending (R2,623.68), frequency (4.20), and satisfaction (4.01).
   c. **Low-Value (High Frequency Loyalty):** These customers are frequent purchasers (8.29) but have very low spending (R410.26) and satisfaction (3.04). A large majority (80%) are loyalty members, suggesting the current programme retains high-frequency but low-spending customers.

Prescriptive Analytics

Based on the predictive results, the following actions are recommended:

1. **Recruit High-Value Customers:** Target the high-value cluster with elite-tier rewards to convert them into loyalty members.
2. **Improve Low-Value Customer Experience:** Offer discounts and user experience enhancements to boost satisfaction and spending for the low-value cluster.
3. **Launch Eastern Cape Campaigns:** Use targeted social media and Mr D campaigns to increase enrolment in this underperforming region.
4. **Incentivise Satisfaction:** Implement bonus points for reviews to leverage the strong positive relationship between satisfaction and spending identified in the regression model.

Longitudinal Simulation

A simulation using `Frequency_of_Purchase` revealed that customers with high purchase frequency (≥7) are all loyalty members (100%). However, this group aligns with the low-value cluster, showing low spending and satisfaction. This confirms that while the programme is effective at driving repeat purchases, it primarily retains a less profitable customer segment.

Implementation

The model was implemented in **Python** using libraries such as `pandas`, `scikit-learn`, `Seaborn`, and `Matplotlib`. The code was designed for reproducibility in a Google Colab notebook, generating all statistical outputs and visualisations to support the loyalty programme's design.

Application to Decision-Making

The loyalty program model translates analytical insights into actionable marketing strategies, directly addressing Takealot's business challenges, providing specific, feasible, and evidence-based recommendations. The strategies are designed to enhance retention, reduce abandonment, and expand market reach, supported by robust data insights.

Strategic Recommendations

1. **Increase Repeat Purchases for High-Value Customers:** Offer elite-tier rewards to the high-value cluster to encourage them to increase their purchase frequency from 3.57 to a target of 5.
2. **Reduce Cart Abandonment:** Provide targeted discounts to the low-value cluster, where 80% of low-income customers report frequent cart abandonment.
3. **Boost Programme Enrolment:** Launch targeted campaigns in the Eastern Cape to address low participation rates.
4. **Enhance Customer Satisfaction:** Leverage the regression finding (β_1 = 1921.48) by offering bonus points for reviews, which can directly translate into higher spending.

Visualisations and Exploration of Customer Trends

The below visualisations aim to provide a clear, data-driven foundation for Takealot's loyalty program by highlighting critical trends in customer behaviour.

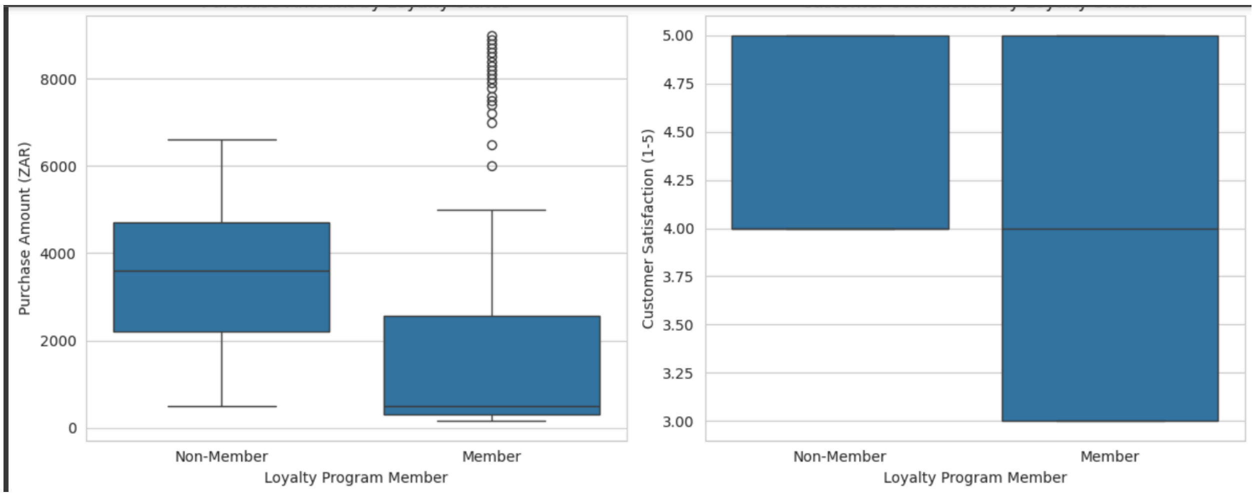

Figure 1: Boxplots: Highlight the need to redesign the loyalty program to match non-members' higher spending and satisfaction, addressing low Eastern Cape participation (30%) and high cart abandonment (80% among low-income customers).

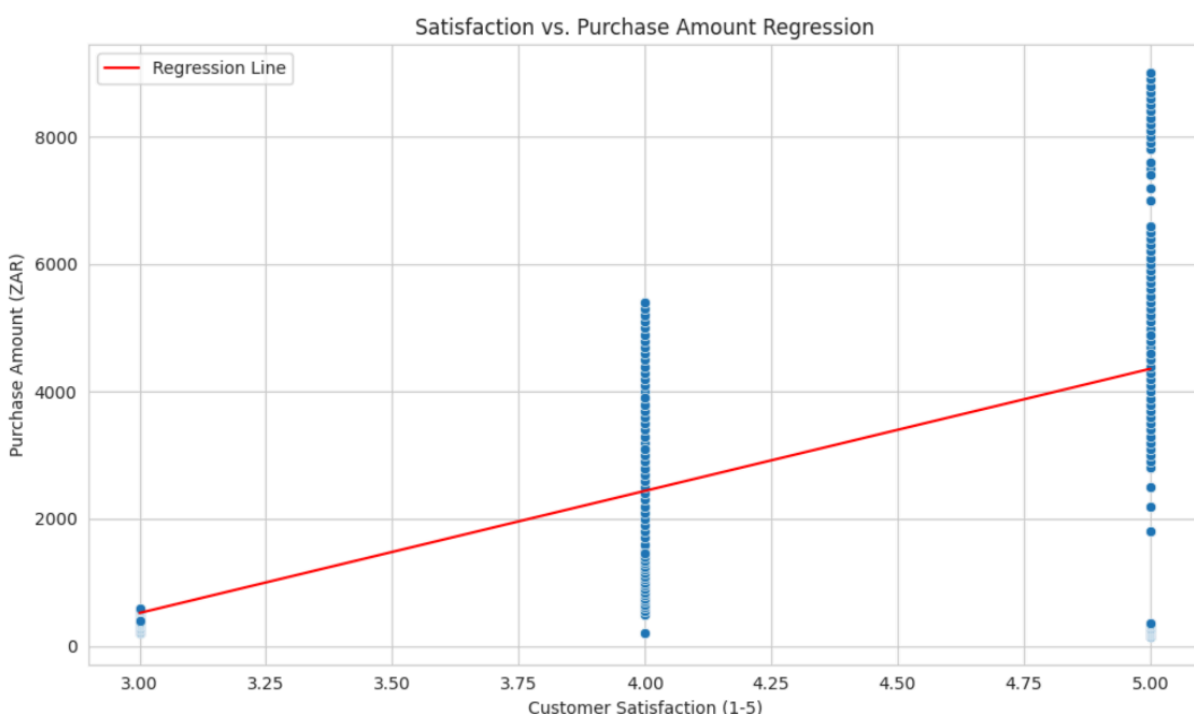

Figure 2: Regression Scatterplot: Supports satisfaction-focused incentives to increase CLV, particularly for township and rural shoppers.

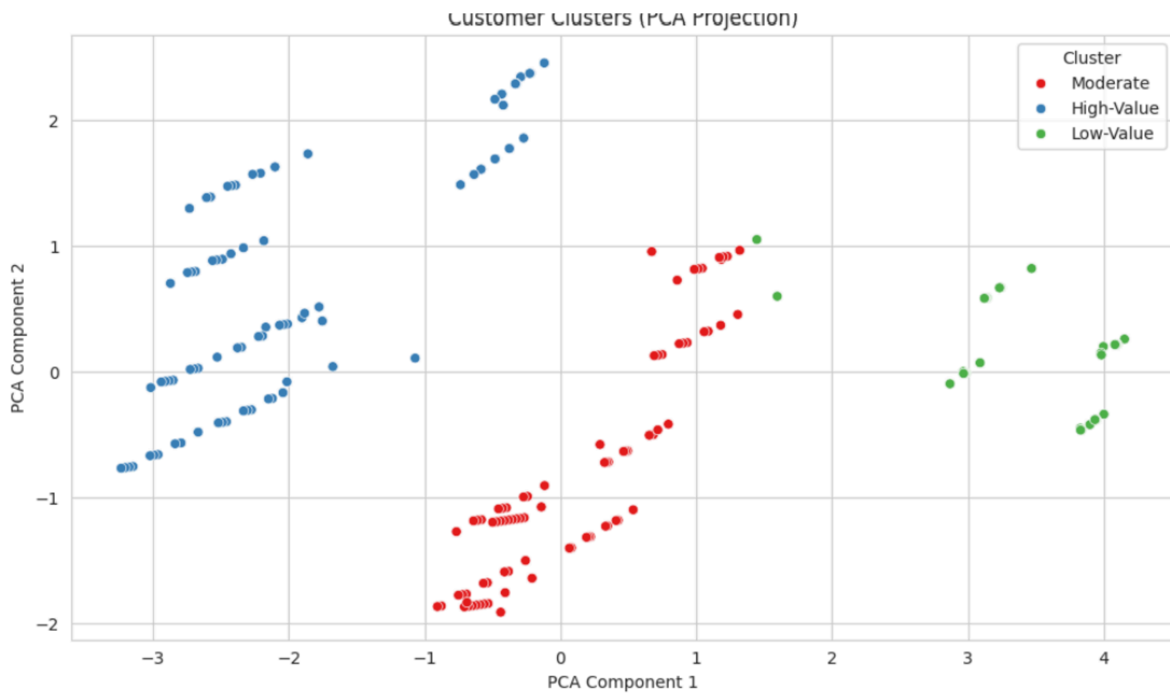

Figure 3: PCA Cluster Plot: Enables personalized rewards for High-Value (recruitment) and Low-Value (improvement) segments, targeting underserved markets.

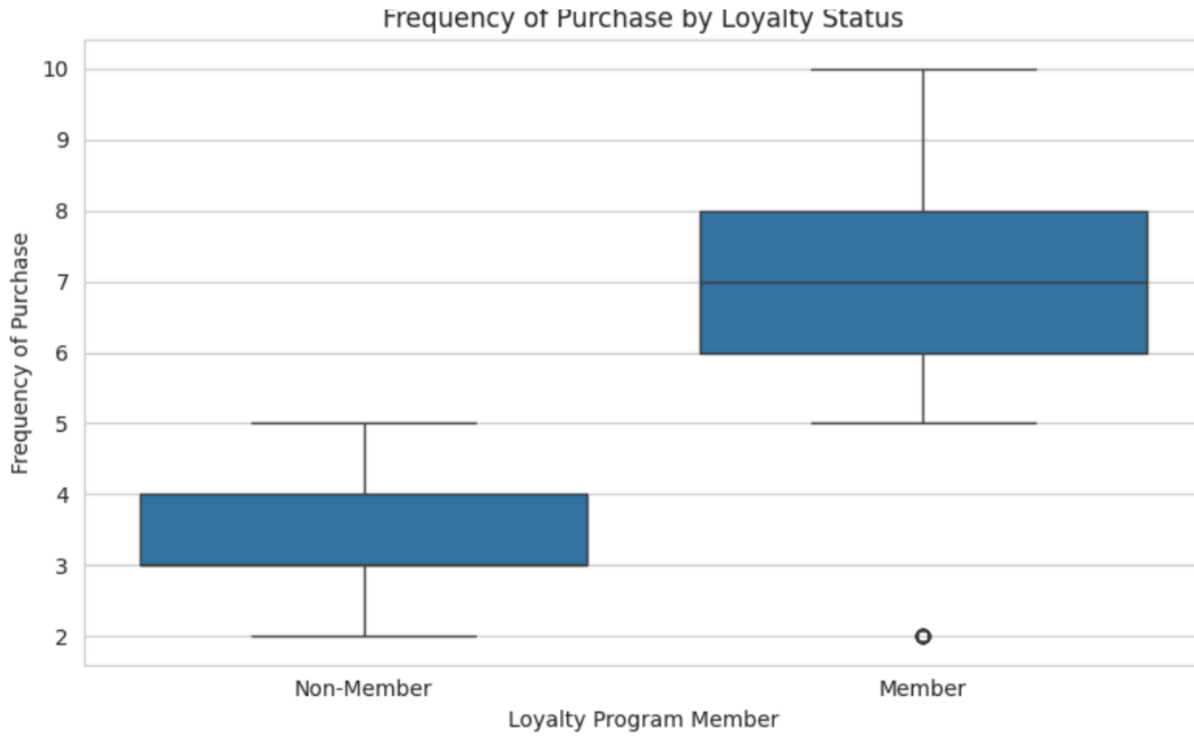

Figure 4: Frequency Boxplot: Confirms retention success but underscores the need to enhance Low-Value customer outcomes to counter competitors like Amazon and Shein.

By highlighting important trends, such as the higher spending and contentment of non-loyalty members (boxplots), the association between spending and satisfaction (regression scatterplot), the various customer segments (PCA plot), and loyalty-driven retention (frequency boxplot), the visualizations successfully aid in decision-making. They support a revamped loyalty program aimed at High-Value recruiting (Elite-tier prizes) and Low-Value improvement (discounts, user experience improvements), which are in line with the business case's objectives of increasing frequency and retention.

Impact Assessment

The proposed strategies are projected to have the following impact:

1. **Tiered Rewards:** A 20% increase in purchase frequency for the high-value cluster.
2. **Targeted Discounts:** A 50% reduction in cart abandonment for the low-value cluster.

3. **Regional Campaigns:** A 15% increase in loyalty programme enrolment in the Eastern Cape.
4. **Engagement Incentives:** A 10% increase in overall customer satisfaction.

These projections, grounded in analytical insights, demonstrate the model's potential to drive significant business impact.

Critical Reflection

The proposed loyalty program model is well-justified, it incorporates ethical considerations and identifies areas for improvement. Offers a strong, data-driven framework to boost Takealot's customer retention and strengthen its position against competitors. The key strengths, limitations and future enhancements are outlined below:

Strengths

1. **Robust Analytics:** The model is built on statistically significant findings from t-tests ($p < 0.001$) and regression analysis ($R^2 = 0.34$), providing a strong evidence base.
2. **Targeted Strategies:** The use of k-means clustering allows for tailored rewards and interventions for distinct customer segments.
3. **Ethical Compliance:** The use of synthetic data ensures compliance with POPIA, avoiding privacy risks associated with real customer data.

Limitations

1. **Unexpected Findings:** The discovery that non-loyalty members exhibit higher spending and satisfaction is a significant limitation of the *current* programme and indicates clear inefficiencies that need to be addressed.
2. **Small Sample Size:** After cleaning, the analysis was based on only 211 valid records, which limits the precision of the statistical inferences.
3. **Outlier Impact:** The analysis identified five high-value outliers (non-members spending >R7,000) that may have skewed the t-test and regression results.

4. **Simulation Assumptions:** The longitudinal simulation relied on `Frequency_of_Purchase` as a proxy for retention, which is less reliable than actual longitudinal data.

Future Improvements

To address these limitations and enhance the model's effectiveness, the following improvements are proposed:

1. Collect and analyse longitudinal data to track customer behaviour over time for more accurate retention and churn analysis.
2. Use more advanced clustering algorithms like DBSCAN to better handle outliers and identify non-spherical clusters.
3. Implement churn prediction models using techniques like logistic regression to proactively identify and retain at-risk customers.
4. Analyse `Points_Earned` and `Points_Redeemed` data more deeply to evaluate the effectiveness of specific rewards..

Ethical Considerations

The use of a synthetic dataset aligns with ethical standards by avoiding the use of real customer data, thus ensuring compliance with POPIA and mitigating privacy risks. Future implementations should maintain transparency in data usage, clearly communicate data practices to customers, and ensure that personalisation strategies respect customer consent and preferences.

Conclusion

The analysis reveals significant inefficiencies in Takealot's current loyalty initiatives, as non-loyalty members demonstrate higher spending and satisfaction. The redesigned data-driven loyalty programme addresses this by shifting focus to recruiting high-value non-members and improving the experience for the existing low-value, high-frequency members. By implementing targeted, evidence-based strategies, Takealot can enhance customer retention, boost profitability, and strengthen its competitive position in the South African e-commerce market.

REFERENCES

1. Ailawadi, K. L., Gedenk, K., Lutzky, C., & Neslin, S. A. (2007b). Decomposition of the sales impact of promotion-induced stockpiling. *Journal of Marketing Research*, *44*(3), 450–467.
2. Bijmolt, T. H. A., & Verhoef, P. C. (2017). Loyalty Programs: Current Insights, Research Challenges, and Emerging Trends. In B. Wierenga & R. van der Lans (Eds.), *Handbook of Marketing Decision Models* (2nd ed., pp. 143–165). Springer.
3. Homans, G. C. (1958). Social Behavior as Exchange. *American Journal of Sociology*, *63*(6), 597–606.
4. Steinhoff, L., & Palmatier, R. W. (2016). Understanding the effectiveness of loyalty programs: Managing target and bystander effects. *Journal of the Academy of Marketing Science*, *44*, 88–107.
5. Takealot Group. (2024). *Annual Financial Report*.