

SPSS-BASED CUSTOMER RETENTION ANALYSIS FOR A LOYALTY PROGRAM IN RETAIL

1. Background and Problem Statement

A national retail chain observed a stagnation in its customer loyalty program engagement. Although the program had over 50,000 enrollees, the repeat purchase rate was lower than expected. The company wanted to understand the key drivers of retention and assess the effectiveness of loyalty points, promotional emails, and demographic factors in retaining customers. The goal was to leverage SPSS for data-driven insights that would shape future campaign strategies and resource allocation.

2. Objectives

- To analyze how loyalty program features influence customer retention
- To examine demographic factors (age, gender, income) as predictors of retention
- To identify customer segments with the highest retention risk
- To recommend targeted improvements to the loyalty program structure

3. Methodology

3.1 Data Preparation and Cleaning in SPSS

- Dataset: 2,500 loyalty program members' transactional data over 12 months
- Variables:
 - Customer ID (string)
 - Retention (binary: 1 = active, 0 = inactive)
 - Loyalty Points Earned (continuous)
 - Loyalty Points Redeemed (continuous)
 - Email Promotions Received (numeric)
 - Age (continuous), Gender (categorical), Income Bracket (ordinal)
- Actions:
 - Recoding gender, creating dummy variables for income

- Filtering out members with less than 3 months of data

3.2 Statistical Analyses in SPSS

- Descriptive analysis to summarize central trends in points, email campaigns, and demographics
- Bivariate analysis using Chi-square for categorical predictors (e.g., gender $\times$ retention)
- Logistic regression with retention as the dependent variable
- Model diagnostics including Hosmer-Lemeshow test and classification tables
- Segment analysis using cluster analysis based on redemption behavior and age

4. Key Results

- **Descriptive Stats:**
 - Average age: 37.6 years; 58% female; 45% in mid-income bracket
 - Average emails received per year: 9.4; Average redemption rate: 62%
- **Chi-square Tests:**
 - Gender not significantly associated with retention
 - Income bracket significantly associated ($p = .021$), with higher retention in upper-middle tier
- **Logistic Regression:**
 - Points Redeemed ($\beta = 0.84$, $p < .001$) and Emails Received ($\beta = 0.38$, $p = .002$) significantly increased odds of retention
 - Age negatively associated ($\beta = -0.06$, $p = .03$) — older users showed lower retention
- **Clustering:**
 - Three clusters identified:
 1. “Frequent Redeemers” – high retention
 2. “Passive Members” – low points usage, low retention
 3. “Young Shoppers” – redeem occasionally, moderate retention

5. Interpretation and Insights

- Higher point redemption correlates strongly with higher retention, indicating that ease and relevance of redemption options are key.
- Email campaigns have a significant but weaker impact, suggesting the need for better targeting or personalization.
- Older customers disengage more quickly, requiring specific reactivation strategies.
- Demographics alone are weak predictors; behavioral data (usage and redemption) drive stronger insights.

6. Recommendations

- Increase the ease of redeeming loyalty points, possibly integrating smaller reward tiers
- Personalize promotional emails based on customer redemption behavior
- Implement re-engagement campaigns specifically for senior age groups
- Design targeted offers for “Passive Members” cluster to boost activity

7. Deliverables

- Cleaned SPSS dataset (.sav) with variable labels and computed fields
- Syntax file with regression and cluster procedures
- Full APA-style report with tables, visuals, and interpretation
- Optional executive summary for business stakeholders

8. Stakeholder Relevance

Academic:

- Ideal for illustrating predictive analytics and customer segmentation using SPSS
- Demonstrates real-world application of logistic regression and cluster analysis

Corporate:

- Supports loyalty program teams in refining segmentation, reward design, and marketing outreach
- Provides a template for ongoing monitoring and reporting of customer retention metrics