DECISION SCIENCES INSTITUTE

Leveraging Analytics & Modeling to Derive Insights for Point-based Loyalty Campaigns

Yifan Gu, Muskan Aggarwal, Sai Nithin Godi, Dhairya Dedhia, Harshal Amin, Nikitha Balaji, Rohan Ajay, Matthew A. Lanham

Purdue University Daniels School of Business
gu277@purdue.edu, aggar158@purdue.edu, sgodi@purdue.edu, ddedhia@purdue.edu, amin82@purdue.edu, balaji76@purdue.edu, rajay@purdue.edu, lanhamm@purdue.edu

ABSTRACT

This study examines how loyalty program design influences customer engagement and spending for a U.S. supercenter chain. We analyze the effectiveness of different offer types and delivery methods, controlling for pre-campaign behavior, to identify strategies driving incremental impact. Predictive models assess the ability of customer behavior to forecast campaign-period sales. Results indicate that per-visit incentives slightly outperform point multipliers, though differences are modest. A simple linear model performs comparably to more complex techniques, offering actionable insights for optimizing loyalty strategies. Future improvements could leverage granular segmentation and enriched data for enhanced program effectiveness.

KEYWORDS: loyalty programs, customer loyalty, predictive modeling, segmentation, retail marketing

INTRODUCTION

In the competitive retail landscape, loyalty programs are critical for customer retention and revenue growth. Shep Hyken (Forbes, 2025) notes that loyalty programs can boost customer spending by up to 20% through enhanced engagement. Deloitte (2025) further highlights that 72% of consumers are more likely to recommend brands with effective loyalty programs, emphasizing the need for personalized, value-driven experiences. Additionally, 65% of consumers prefer simple, accessible rewards, underscoring the importance of tailored benefits (WSJ, 2025).

Retailers globally leverage loyalty programs not only to retain customers but also to gather data for refining marketing strategies. Advanced analytics enable businesses to personalize offers based on consumer behavior, optimizing engagement strategies. This study explores the efficiency of a regional U.S. supercenter chain’s loyalty program, aiming to optimize offer characteristics to maximize engagement and drive behavioral change. The central research question is: How can loyalty program offers be optimized to enhance customer engagement and induce positive behavioral changes?

The analysis focuses on identifying which offer types and delivery methods most effectively drive engagement, distinguishing between transactional and incremental behavioral changes. Using data analysis and predictive modeling, we assess the impact of loyalty offers on sales, store visits, and department spending. Machine learning models are developed to forecast customer behavior, leveraging historical purchasing data.

This research integrates qualitative and quantitative analyses to evaluate the loyalty program's impact, offering actionable insights for program optimization. The findings contribute to retail marketing by demonstrating how data-driven decisions can improve business outcomes.

LITERATURE REVIEW

Customer loyalty programs (CLPs) are strategic tools for increasing retention and profitability. Point-based systems, prevalent in retail, hospitality, and financial services, reward customers with redeemable points for spending (Chen et al., 2021). Recent advancements in machine learning (ML) enable businesses to optimize CLPs by predicting customer responses and refining marketing strategies. This review examines the effectiveness of point-based loyalty programs, their impact on sales and profitability, and how predictive analytics can enhance loyalty strategies.

Effectiveness of Point-Based Loyalty Programs

Point-based programs encourage repeat purchases through delayed gratification, where customers accumulate points for rewards. Successful programs feature transparent redemption processes and meaningful rewards, driving engagement and spending (Gandomi & Zolfaghari, 2018). Conversely, complex structures, frequent point expirations, and high redemption thresholds can reduce effectiveness (Hofman-Kohlmeyer, 2016).

Loyalty programs also influence purchase behavior and price sensitivity. Enrolled customers become less price-sensitive, reducing the need for excessive discounting (Van Heerde & Bijmolt, 2005). However, their financial impact is mixed. While some studies show increased purchase frequency and lifetime value (Belli et al., 2022), others suggest they may not always yield sustainable profits (Daryanto et al., 2010; Meyer-Waarden et al., 2013).

Category-specific factors also play a role. For example, Lin & Bowman (2022) found that loyalty programs initially boost sales, but effects diminish over time. High-frequency categories like dairy sustain growth, while low-frequency categories like grooming items decline. This highlights the need for tailored strategies aligned with customer purchasing habits.

Behavioral mechanisms, such as the goal gradient effect (where customers spend more as they near a reward threshold), further drive engagement (Septianto et al., 2019). However, unredeemed points can lead to disengagement, emphasizing the need to balance aspirational rewards with attainable benefits (Gandomi & Zolfaghari, 2018).

Machine Learning for Predictive Loyalty Program Optimization

Traditional methods for evaluating loyalty programs, such as historical sales data and surveys, often lack predictive accuracy. ML models, however, enable data-driven optimization by forecasting customer responses and refining promotional strategies. For example, Usman et al. (2023) demonstrated ML's advantages in sales forecasting, with XGBoost outperforming other models like Linear Regression and ARIMA. ML allows businesses to segment customers based on purchasing behavior and tailor rewards, ensuring cost-effective and impactful promotions.

Sales Forecasting and Loyalty Program Performance

Sales forecasting is critical for loyalty program management, enabling businesses to anticipate demand fluctuations and adjust marketing efforts. ML models analyze transaction data to identify patterns traditional methods often miss (Usman et al., 2023). Predictive analytics can refine incentive structures by dynamically adjusting point accumulation and redemption values based on demand.

Lin & Bowman (2022) found that category penetration and purchase frequency are key to loyalty program success. Similarly, Verhoef (2003) emphasized targeting high-repeat-purchase segments. ML-driven price optimization further enhances loyalty programs by refining reward structures without unnecessary revenue loss (Van Heerde & Bijmolt, 2005). By integrating ML, businesses can move beyond traditional retention strategies, adopting data-driven approaches that maximize engagement and long-term profitability.

DATA

This study used a dataset comprising 400,662 distinct customer-level observations, including twelve variables capturing both pre-campaign and during-campaign customer behaviors. Each record is uniquely identified by a customer identifier and is assigned to one of five customer group categories. Behavioral features include total sales and store visits recorded separately for the pre-campaign (e.g., PreSales, PreTrips) and campaign periods (e.g., DuringSales, DuringTrips). Digital engagement is captured through the number of days customers visited the website, viewed offer pages, and claimed offers, also split into pre-campaign (e.g., PreWebVisitDays, PreOfferPageViewDays, PreClaimDays) and during-campaign (e.g., DuringWebVisitDays, DuringOfferPageViewDays, DuringClaimDays) metrics.

The dataset was constructed by merging three sources: customer group assignments, pre-campaign data, and during-campaign data, using the unique customer identifier. Initial validation confirmed the dataset was clean, with 4,259 rows of missing during-campaign values treated as zeros (i.e., no observed activity) after confirming with the retail company.

Customers were segmented into five groups based on combinations of offer delivery method, offer type, and campaign duration as shown in Table 1. Delivery methods included a newer, more informative format and a traditional coupon-style approach. Two groups received offers that rewarded additional points per qualifying visit, while another two received multipliers on base loyalty points earned per purchase. Four groups participated in a three-month campaign, while one group received a standard offer over a shorter, four-week period.

Table 1: Customer Treatment Groups

Group Number	**Delivery Type**	**Offer Type**	**Duration**
Group 1	New	PointsPerVisit	ThreeMonth
Group 2	New	PointsMultiplier	ThreeMonth
Group 3	Old	PointsPerVisit	ThreeMonth
Group 4	Old	PointsMultiplier	ThreeMonth
Group 5	Old	Normal Targeting	FourWeek

METHODOLOGY

Offer Effect Analysis

We first examined whether different combinations of offer types and delivery methods (captured through customer group assignments) impacted campaign-period spending. To isolate the effect of marketing treatments, we controlled pre-campaign sales (PreSales) using a multiple linear regression model. DuringSales was regressed on PreSales with one-hot encoded CustomerGroup indicators (Group 1 used as the baseline group). This allowed us to assess whether different offer delivery methods or offer types significantly influenced sales, holding historical purchasing behavior constant.

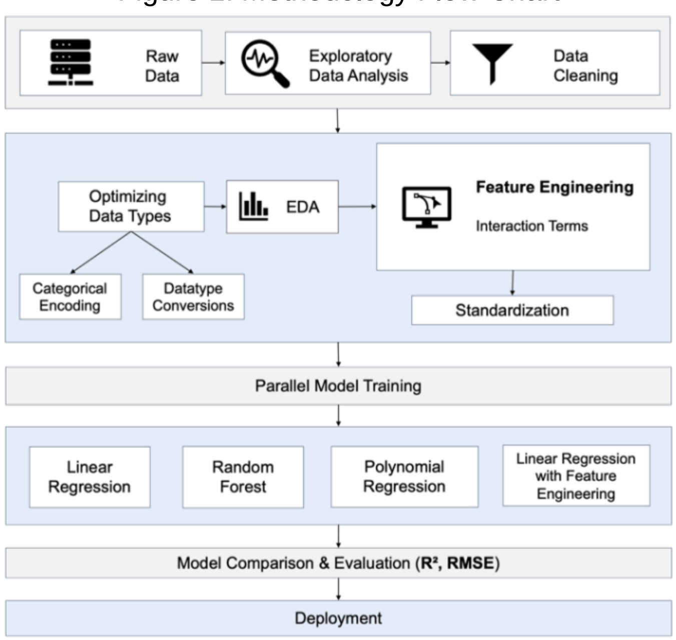
Modeling with All Pre-Campaign Features

We explored whether pre-campaign behavior alone could predict campaign-period sales using five pre-campaign predictors: PreSales, PreTrips, PreOfferPageViewDays, PreClaimDays, and PreWebVisitDays, reflecting in-store and digital engagement. Four modeling approaches were evaluated: linear regression, Random Forest, polynomial regression (Degree 2), and linear regression with interaction terms. Models were trained using an 80/20 train-test split, with performance assessed using RMSE and R^2 to balance complexity and predictive value.

Modeling with Best Selected Features

Beyond pre-campaign features, we expanded the feature set to include both pre- and during-campaign variables, identifying the most impactful predictors of campaign-period sales. The same four machine learning models are applied. Figure 1 represents the overall design of our explanatory analysis and predictive modeling:

Figure 1: Methodology Flow Chart



MODEL

We used a multiple linear regression (controlled offer group comparison) model as shown in Equation 1.

Equation 1: multiple linear regression used for our offer effect analysis:

$$DuringSales_i = \beta_0 + \beta_1 \cdot \boldsymbol{PreSales_i} + \Sigma k = 2^5 \boldsymbol{\beta_k} \cdot \boldsymbol{Group_k} + \epsilon_i$$

where:
- $\boldsymbol{PreSales_i}$ captures customer i's prior behavior
- $\boldsymbol{Group_k}$ are binary indicators for Customer Groups 2 through 5 (Group 1 as baseline)
- Coefficients $\boldsymbol{\beta_k}$ represent the adjusted difference in sales compared to Group 1

The variables were standardized for comparability and evaluated significance using p-values.

We developed the following additional models using pre-campaign features using the five selected pre-period variables:
- Linear Regression (OLS) to establish a simple, interpretable baseline.
- Random Forest (n = 100) to account for potential nonlinearities and interactions.
- Polynomial Regression (Degree = 2) to capture curvilinear relationships between predictors.
- Feature-Engineered Linear Regression, using second-degree interaction terms (without squared terms), to test for synergistic effects between variables.

The four models developed had the same structure to those with pre-period features, only that the model inputs were expanded to include pre- and during-campaign variables.

RESULTS

Offer Effect Analysis

The regression analysis shows in Table 2 that pre-campaign sales (PreSales) is the strongest predictor of campaign-period sales, while differences across customer groups were generally modest. Group 2 and Group 4 were found to have statistically significant negative coefficients, indicating lower sales compared to Group 1. Group 3 and Group 5 were not significantly different from Group 1. We also observed that per-visit offers (Groups 1 and 3) outperformed point multipliers (Groups 2 and 4), suggesting that rewarding specific shopping behaviors may be more effective than applying generalized point multipliers.

Table 2: Multiple Linear Regression Summary

Predictor	Coefficient	P-value	Interpretation
*constant	589.9122	0.000	-
Standardized PreSales	364.9537	0.000	-
Group 2	-2.8596	0.028	New Delivery: PointsPerVisit vs PointsMultiplier
Group 3	-1.1072	0.394	PointsPerVisit: New Delivery vs Old Delivery

Group 4	-3.1436	0.016	New Delivery PointsPerVisit vs Old Delivery PointsMultiplier
Group 5	-2.1214	0.081	New Delivery PointsPerVisit vs Old Delivery Normal Target

Model Comparison (Pre-Campaign Behavior)

Models using only pre-campaign behavioral features showed moderately strong predictive power (R^2 ~0.67, RMSE ~250) as shown in Table 3. The linear regression performed nearly as well as more complex models, stressing the effectiveness of simple, behavior-based predictors like store visits and digital engagement.

Table 3: Model Comparison with Pre-Campaign Features

Model	R^2	RMSE
Linear Regression	0.6761	251.1766
Random Forest (n=100)	0.6553	259.0880
Polynomial Regression (Degree=2)	0.6766	250.9664
Feature Engineered Linear Regression (2nd Degree Interactions)	0.6767	250.9433

Model Comparison (Best Selected Features)

The final models included three best features: pre-campaign sales, pre-campaign trips, and during-campaign trips. A summary of the model results is presented in Table 4. Polynomial Regression (Degree=2) was found to achieve the best R^2 and lowest RMSE, but improvements over Linear Regression were marginal. Increasing model complexity (e.g., higher-degree polynomials) was shown to degrade performance, suggesting linear relationships dominate.

Table 4: Model Comparison with Best Selected Features

Model	Predictors Used	R^2	RMSE
Best Incremental Linear Regression	DuringTrips, PreSales, PreTrips	0.8098	192.4681
Random Forest (n=100)	DuringTrips, PreSales, PreTrips	0.8102	192.2517
Polynomial Regression (Degree=2)	DuringTrips, PreSales, PreTrips	0.8245	184.8838
Feature Engineered Linear Regression (2nd Degree Interactions)	Same as LG + interactions	0.8131	190.8038

DISCUSSION

Measurable differences between groups were observed in the Offer Effect Analysis, but the effects were relatively small. Prior research suggests that loyalty program success depends on category-specific factors like purchase frequency and product type (Lin & Bowman, 2022), and not all customer segments respond equally to the same incentives (Verhoef, 2003). A more granular analysis at the retail department level is recommended to uncover where point multipliers or per-visit rewards are most effective. For example, high-frequency categories (e.g., grocery, produce) may sustain engagement better than impulse-driven ones. This aligns with the literature's call for predictive, category-aware program optimization.

Across both pre-campaign and final models, a simple linear regression model performed nearly as well as more complex alternatives. Despite testing nonlinear models (e.g., Random Forests, polynomial regression), no significant performance gains were observed. This suggests that, within the current dataset, the relationship between customer behavior and sales is largely linear, with past spending and store visits being the most predictive variables.

From a business perspective, the linear model offers a balance of accuracy, simplicity, and transparency, making it suitable for loyalty program planning, customer targeting, and internal forecasting. However, the analysis highlights a key limitation: model performance may be constrained by the available feature set. Advanced models like XGBoost and LSTM nerual networks may only outperform simpler methods when applied to richer datasets with granular behavioral and contextual variables (Usman et al., 2023).

Future research should expand the dataset to include additional attributes, such as offer interactions, temporal patterns, or external engagement indicators, to better leverage advanced machine learning techniques. For now, a well-tuned linear regression model remains a practical and effective solution, offering strong performance and interpretability for decision-makers.

CONCLUSION

This study explored how offer types, delivery methods, and customer behaviors influence sales during a loyalty campaign, with the goal of optimizing program design through data-driven insights. Our offer effect analysis showed that per-visit incentives slightly outperformed point multipliers, but the differences across groups were modest, highlighting that offer format alone does not drive substantial impact when prior behavior is accounted for.

We also built predictive models using both pre- and during-campaign behavioral data. Across all models tested, a simple linear regression model performed nearly as well as more complex alternatives, suggesting a largely linear relationship in the current dataset. Past spending and store visits emerged as the most influential predictors.

These findings are actionable for businesses aiming to enhance loyalty performance while keeping models interpretable and efficient. However, performance gains from more advanced models may require richer, more granular data, pointing to future opportunities in feature enrichment and segmentation at the category level to better tailor marketing strategies.

REFERENCES

Belli, A., Geyskens, I., Dekimpe, M. G., & Steenkamp, J.-B. E. M. (2022). 40 years of loyalty programs: How effective are they? *Journal of the Academy of Marketing Science, 50*(1), 147–173.

Chen, Y., Fay, S., & Wang, Q. (2021). Three decades of research on loyalty programs: A literature review and future research agenda. *Journal of Business Research, 124*, 179–197.

Daryanto, A., de Ruyter, K., Wetzels, M., & Patterson, P. G. (2010). Service firms and customer loyalty programs: A regulatory fit perspective of reward preferences in a health club setting. *Journal of the Academy of Marketing Science, 38*(5), 604–616.

Deloitte. (n.d.). *Transcending expectations in the new loyalty landscape.* Deloitte. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/us-transcending-expectations-in-the-new-loyalty-landscape.pdf

Deloitte. (2024, November 19). What loyalty program benefits do consumers want? *The Wall Street Journal.* https://deloitte.wsj.com/cmo/what-loyalty-program-benefits-do-consumers-want-4e0a6511

Gandomi, A., & Zolfaghari, S. (2018). To tier or not to tier: An analysis of multitier loyalty programs' optimality conditions. *Omega, 74*, 20–36.

Hofman-Kohlmeyer, M. (2016). Customer loyalty program as a tool of customer retention: Literature review. *CBU International Conference Proceedings, 4*, 202–206.

Hyken, S. (2025, February 9). The hidden revenue engine: Loyalty programs are more than just repeat business. *Forbes.* https://www.forbes.com/sites/shephyken/2025/02/09/the-hidden-revenue-engine-loyalty-programs-are-more-than-just-repeat-business

Lin, C., & Bowman, D. (2022). The impact of introducing a customer loyalty program on category sales and profitability. *Journal of Retailing and Consumer Services, 64*, 102769.

Meyer-Waarden, L., Benavent, C., & Castéran, H. (2013). The effects of purchase orientations on perceived loyalty programmes' benefits and loyalty. *International Journal of Retail & Distribution Management, 41*(3), 201–225.

Septianto, F., Sinha, R. K., & Sadarangani, P. (2019). The similar versus divergent effects of pride and happiness on the effectiveness of loyalty programs. *Journal of Business Research, 99*, 12–22.

Usman, S., Malik, K., Rauf, A., & Zahid, A. (2023). Predictive analysis of retail sales forecasting using machine learning techniques. *ResearchGate.*

Van Heerde, H. J., & Bijmolt, T. H. A. (2005). Decomposing the promotional revenue bump for loyalty program members versus nonmembers. *Journal of Marketing Research, 42*(4), 443–457.

Verhoef, P. C. (2003). Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing, 67*(4), 30–45.