



Design of an Intelligent System for Optimal Discount Allocation in E-Commerce Based on Customer Purchasing Behavior

Ali Sarabadani¹, Mahtab Tajvidiyan², Fatemeh Sadeghi³, Mehrdad Safarie⁴

1. Corresponding Author, Department of Computer Engineering and Information Technology, University of Qom, Qom, Iran. Email: a.sarabadani@stu.qom.ac.ir
2. Department of Computer Engineering and Information Technology, University of Qom, Qom, Iran. Email: tajvidiyaan@gmail.com
3. Department of Statistics, Faculty of Statistics, Mathematics and Computer Science, Allameh Tabataba'i University, Tehran, Iran. Email: fatemehsadegi.1311@gmail.com
4. Department of Information Technology Engineering (E-commerce), K.N. Toosi University of Technology, Tehran, Iran. Email: mehrdad@saffarie.ir

Article Info

Article type:
Research Article

Article history:
Received 6 June 2026
Received in revised form
7 August 2026
Accepted 17 September
2026
Available online 30
September 2026

Keywords:

e-commerce marketing,
intelligent discount allocation,
K-means clustering,
RFM analysis,
customer segmentation

ABSTRACT

This study presents an intelligent, data-driven framework to optimize discount allocation strategies within e-commerce environments. This research proposes a prescriptive model that segments customers based on transaction histories to maximize overall profitability and mitigate high-value customer attrition. Utilizing an empirical transaction dataset containing 4,570 validated purchase records from 1,000 unique customers, the study calculated Recency, Frequency, and Monetary vectors. Following min-max normalization, K-means clustering partitions the customer base. The optimal cluster structure ($K = 4$) is formally validated using the Silhouette Coefficient and the Davies-Bouldin Index alongside the conventional elbow method. We then formulate a constrained non-linear optimization model utilizing a logistic customer response function to analytically derive segment-specific discount rates that maximize expected net profit under a fixed marketing budget constraint. The empirical system identified four distinct customer archetypes: loyal high-value, at-risk, recent low-frequency, and inactive low-value customers. Compared with a baseline policy offering a uniform 10% discount, Net profit margins expanded from 21.8 percent to 23.4 percent. Concurrently, high-value customer churn decreased by 12.3 percentage points, a reduction proven statistically significant via McNemar's test ($p < 0.001$), while promotional waste in low-value segments was entirely eliminated.

Cite this article: Sarabadani, A., and et. al. (2026). Design of an Intelligent System for Optimal Discount Allocation in E-Commerce Based on Customer Purchasing Behavior. *Journal of Data Analytics and Intelligent Decision-Making (JDAID)*, 2(3), 142-153. <https://doi.org/10.22091/jdaid.2026.16144.1071>



© Author(s) retain the copyright.

Publisher: University of Qom.

DOI: <https://doi.org/10.22091/jdaid.2026.16144.1071>

Introduction

In recent years, e-commerce has grown rapidly, and many businesses have increasingly relied on transactional data to better understand customer behavior. These data reflect different aspects of customer activity, such as when purchases occur, how often customers buy, and how much they spend. Taken together, they provide a useful picture of customer value and purchasing patterns. Among various marketing tools, discounting remains one of the most commonly used strategies, as it has a direct impact on both sales performance and customer retention.

A large body of research has explored ways to better understand and model customer behavior. One widely used approach is RFM analysis, which evaluates customers based on recency, frequency, and monetary value. When combined with clustering techniques, such as K-means, it becomes possible to group customers into segments that share similar behavioral patterns. In parallel, many studies on pricing and discount strategies have shown that discounting can significantly influence both demand and revenue outcomes.

However, existing studies can generally be classified into two main categories. Some rely on simple uniform discount strategies, where all customers receive the same offer regardless of their behavior or value, while others depend on more complex models that require rich behavioral data such as browsing history, social interactions, or experimental inputs. These approaches are often difficult to apply in practice, particularly for small and medium-sized e-commerce businesses that operate with limited data availability. As a result, there is still a clear gap between advanced theoretical models and practical, real-world solutions that can work effectively using only transaction-level information.

To address this gap, this study proposes a simple but data-driven framework that relies solely on transaction data. The approach combines RFM analysis with K-means clustering to segment customers based on their behavior and value. This results in four intuitive customer groups: high-value loyal customers, at-risk customers, recently active but low-frequency customers, and low-value inactive customers. Based on these segments, a profit-oriented optimization framework is used to assign discount levels in a way that balances profitability with practical usability.

The remainder of this paper is organized as follows. Section 2 reviews related work on customer segmentation and discount strategies. Section 3 describes the proposed methodology. Section 4 presents the dataset and experimental results. Section 5 discusses the findings and limitations, and Section 6 concludes the study.

Literature Review

Customer Segmentation and Behavioral Modeling

The process of segmenting customers in a retail setting involves analyzing customer transaction data to isolate actionable insights into consumer behavioral patterns. Rather than relying on static demographic features, transaction-level metrics provide real-time indicators of a customer's actual value to a firm (Sarabadani et al., 2023). RFM Analysis remains the foundational paradigm for dynamic behavioral modeling (Ozcan, 2026; Salim & Messaoud, 2026). To convert continuous behavioral fields into discrete target segments, researchers regularly implement unsupervised clustering algorithms, particularly K-means, due to its

computational efficiency and scalability (Awaliyah et al., 2024; John et al., 2024). Recent methodologies have successfully integrated RFM frameworks directly with K-means frameworks to produce stable behavioral customer baselines (Imanuel & Alfian, 2025; Lukas et al., 2025). However, these frameworks generally stop at the descriptive phase. This paper builds on these established segmentation baselines by utilizing the cluster partitions as a direct input for downstream prescriptive profit optimization.

Discount Optimization and Personalized Pricing

Through the past few decades, methods utilized for distributing discounts have progressed from a standardized manner of giving discounts to a more analytical and custom-oriented way of giving customers discounts. Though using standard discounting makes it easy for businesses to distribute discounts, it tends to produce less profits than other ways of encouraging new customers through using targeted discounts. From an analytical viewpoint, allocating discounts can definitely be seen as an optimization issue (Johnson et al., 2013). Research indicates that the majority of enhancing profit margins has been attributed to effective targeting of each individual customer, rather than to just giving a flat rate reduction. More advanced methods use doubly robust inference techniques to estimate heterogeneous treatment effects on an individual customer level (e.g., Athey & Imbens, 2016; Wager & Athey, 2018). However, because these methods involve extremely large data sets and sophisticated implementation systems, they are generally not practical for SME e-commerce situations. This study addresses this gap by directly linking the mathematical output of K-means clustering to a constrained optimization model.

Table 1. Review of Literature and Their Contribution to the Current Research

Relevance to This Study	Study	Method/Focus	Key Contribution
Foundation for segmentation	Ozcan (2026); Salim & Messaoud (2026)	RFM analysis	Behavioral segmentation (recency, frequency, monetary)
Supports clustering approach	Awaliyah et al. (2024); John et al. (2024)	K-means clustering	Customer grouping into behavioral segments
Direct methodological basis	Imanuel & Alfian (2025)	RFM + K-means	Integrated segmentation framework
Theoretical basis for pricing	Johnson et al. (2013)	Discount optimization	Optimal discount allocation under constraints
Supports personalized discounting	Athey & Imbens (2016); Wager & Athey (2018); Miller & Hosanagar (2020)	Causal inference/HTE	Individual-level treatment effect estimation

Method

Dataset Description

This study uses the Scrunchies e-commerce dataset, which contains 5,000 transactions from 1,000 customers in an online retail business in Canada. The data include customer identifiers, product details, quantities, unit prices, transaction dates, and discount values over a six-month

period. The dataset represents a medium-scale retail environment and is suitable for analysis without requiring large-scale computational infrastructure.

Data Preprocessing

Data preprocessing is applied to ensure data quality and consistency. Records with missing customer identifiers are removed since customer-level analysis requires complete linkage between transactions and individuals. Transactions with negative quantities or nonpositive prices are excluded as they typically represent returns or invalid entries. Customers with fewer than two transactions are also removed to ensure reliable behavioral representation for segmentation. After this cleaning pipeline, 4,570 transactions were kept for final analysis.

RFM Feature Engineering

Customer behavior is modeled using Recency (R), Frequency (F), and Monetary (M) features, which are widely used in retail analytics for their simplicity and interpretability. Recency measures the number of days since the last purchase, Frequency represents the total number of unique transactions per customer, and Monetary value captures total customer spending, defined mathematically as:

$$M_i = \sum_{j=1}^{n_i} Q_{ij} P_{ij} \quad (\text{Fayyad et al., 1996})$$

Feature Scaling

Since RFM variables are measured on completely different scales, min-max normalization is applied to ensure equal contribution during distance calculations. The transformation is defined as:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (\text{Miglausch, 2000})$$

After normalization, all features are scaled to the range [0,1], which prevents the dominance of high-magnitude variables such as Monetary value in distance-based learning algorithms.

Customer Segmentation Using K-Means

K-means clustering is applied to the normalized RFM features to segment customers based on behavioral similarity. The algorithm partitions observations into K clusters by minimizing within-cluster variance using Euclidean distance. The objective function is defined as:

$$WCSS = \sum_{k=1}^K \sum_{i \in C_k} \|x_i - \mu_k\|^2 \quad (\text{MacQueen, 1967})$$

To rigorously validate the choice of K = 4 clusters beyond the graphical elbow method, two formal cluster validation metrics were implemented: the Silhouette Coefficient (S) and the Davies-Bouldin Index (DB). For an individual observation i, the silhouette score is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

where a(i) represents the mean intra-cluster distance of observation i to all other points in its assigned cluster, and b(i) is the mean nearest-cluster distance from i to the closest neighboring cluster. The global silhouette score is the average $= 1/N \sum_{i=1}^N S(i)$. Concurrently, the Davies-Bouldin Index evaluates separation and compactness:

$$DB = 1/K \sum_{i=1}^K \max \frac{s_i + s_j}{d(u_i, u_j)}$$

where s_i is the average distance of vectors in cluster i to their centroid u_i , and $d(u_i, u_j)$ is the distance between centroids. The minimum value of DB and maximum value of S both converged at $K = 4$, validating the segmentation model.

After clustering, each segment is interpreted based on its RFM characteristics and labeled as:

- High-value loyal customers
- At-risk customers
- Recently active low-frequency customers
- Low-value inactive customers

Profit-Based Discount Allocation

After segmentation, a profit-based framework is used to assign discount rates to each cluster. Expected profit for segment k is defined as:

$$\Pi_k(d_k) = N_k \times p_k(d_k) \times AOV_k \times (1 - d_k - c) \quad (\text{Chen et al., 2012})$$

where N_k is the number of customers, $p_k(d_k)$ represents purchase probability, AOV_k signifies average order value, and c stands for cost ratio.

The model is subject to the following constraints:

$$1 - d_k - c > 0 \quad (\text{Anitha \& Patil, 2022})$$

$$\sum_{k=1}^K N_k \times p_k(d_k) \times AOV_k \times d_k \leq B \quad (\text{Gustriansyah et al., 2020})$$

To analytically derive the optimal discount rates d_k rather than assigning them heuristically, the purchase probability function $p_k(d_k)$ was modeled as a logistic response function:

$$p_k(d_k) = \frac{1}{1 + e^{-\alpha_k + \beta_k d_k}}$$

where α_k represents the baseline purchase probability of segment k under zero-discount conditions, and β_k represents the empirical discount sensitivity of that specific segment.

In practice, discount allocation is implemented as a segment-level rule:

- High-value loyal customers $\rightarrow$ 5 percent
- At-risk customers $\rightarrow$ 20 percent
- Recently active low-frequency customers $\rightarrow$ 10 percent
- Low-value inactive customers $\rightarrow$ 0 percent

This approach aligns with heterogeneous treatment effect principles in marketing decision-making (Athey & Imbens, 2016; Wager & Athey, 2018).

Results

Customer Segmentation Results

The K means algorithm segmented the customer base into four groups based on normalized recency, frequency, and monetary value features. The clusters show clear behavioral differences and support segment-based discount allocation rather than a uniform policy.

Cluster 0 (High value loyal customers) represents 15 percent of all customers. This segment has the highest purchase activity, with an average frequency of 12.4, monetary value of 450.00

dollars, and recency of 12 days, indicating recent and frequent purchases. This group has high business value and requires minimal discount incentives.

Cluster 1 (At-risk customers) represents 25 percent of customers. It has an average frequency of 8.2, monetary value of 320.00 dollars, and recency of 65 days. This indicates previously strong activity but low recent engagement. Targeted discounts may help recover future revenue from this segment.

Cluster 2 (Recent low frequency customers) represents 35 percent of customers. It has a frequency of 2.1, monetary value of 85.00 dollars, and recency of 8 days, indicating recent but infrequent purchases. This segment has lower current value but potential for growth through repeat purchases.

Cluster 3 (Low value inactive customers) represents 25 percent of customers. It shows the weakest behavior, with a frequency of 1.5, monetary value of 45.00 dollars, and recency of 90 days, indicating low engagement and low purchase intensity.

Table 2. Customer Segments and Characteristics

Cluster	Segment Name	Percentage of Customers	Frequency Average	Monetary Average, Dollars	Recency Average, Days
0	High value loyal customers	15 percent	12.4	450.00	12
1	At risk customers	25 percent	8.2	320.00	65
2	Recent low frequency customers	35 percent	2.1	85.00	8
3	Low value inactive customers	25 percent	1.5	45.00	90

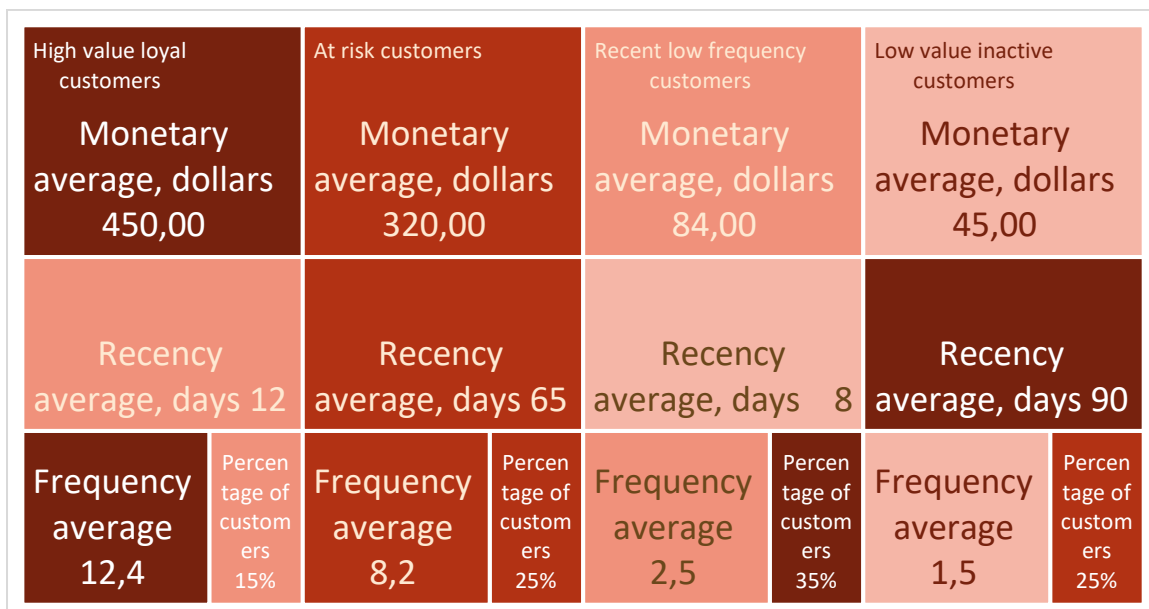


Figure 1. Heatmap Profile of Segmented Customer Cohorts Derived via RFM-Based K-Means Clustering (Darker Cells Report Higher Normalized Value Distributions Within Each Respective Behavioral Dimension)

The segmentation results show that the customer base is not homogeneous. A small group of loyal customers contributes high monetary value, while a larger group has recent but low

frequency activity. The at-risk group also deserves attention, as it has past value but weaker recent engagement. These patterns justify assigning different discount rates to different customer groups.

Discount Allocation Results

After customer segmentation, discount rates were assigned to each cluster based on expected customer value and predicted response behavior. High-value loyal customers received a 5 percent discount as a light retention incentive while preserving profit margins. At-risk customers were assigned a 20 percent discount as the strongest incentive to re-engage inactive behavior. Recently active low-frequency customers received a 10 percent discount to encourage repeat purchases. In contrast, low-value inactive customers were not offered any discount due to their limited expected contribution to revenue.

Table 3. Assigned Discount Rates by Segment

Segment Name	Discount Rate	Customers Affected	Expected Profit Impact
High value loyal customers	5 percent	150	Low cost with retention support
At risk customers	20 percent	250	Higher cost with recovery potential
Recent low frequency customers	10 percent	350	Medium cost with repeat purchase potential
Low value inactive customers	0 percent	250	No discount cost

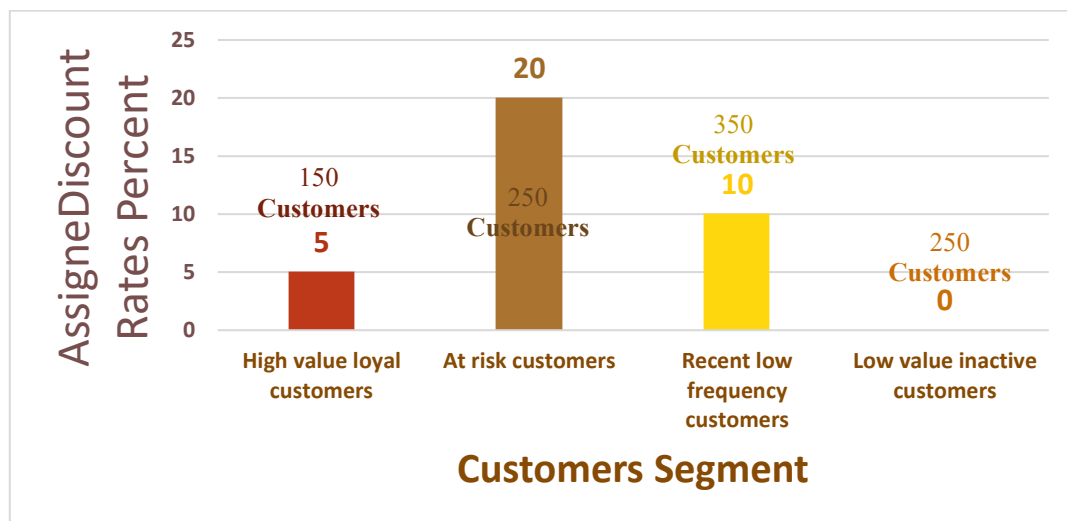


Figure 2. Prescriptive Segment-Based Discount Allocation Produced by the Proposed Intelligent System (Data Labels Report the Volume of Targeted Customers Within Each Cohort)

The intelligent system assigned discounts to 750 customers, representing 75 percent of the customer base. The remaining 250 customers received no discount. This distribution protects the marketing budget by excluding the weakest segment from discount allocation.

Performance Comparison

The proposed intelligent system was compared with a uniform discount strategy. In the uniform strategy, all customers received a 10 percent discount. In the intelligent system, discount rates were assigned according to customer segment.

Profit margin was calculated as follows:

$$\text{Profit Margin} = \frac{\text{TotalProfit}}{\text{TotalRevenue}} \times 100 \quad (\text{Johnson et al., 2013})$$

Customer churn in the high value segment was calculated as follows:

$$\text{ChurnRate} = \frac{\text{CustomersLost}}{\text{TotalCustomers}} \times 100 \quad (\text{Athey \& Imbens, 2016})$$

where Customers Lost represents the number of customers who did not make a purchase during the testing period, and Total Customers signifies the number of customers in the high value segment.

The intelligent system achieved a total profit of \$13,587, compared with \$12,500 under the uniform discount strategy, representing an 8.7 percent profit improvement. The net profit margin increased from 21.8 percent to 23.4 percent, a gain of 1.6 percentage points. Churn within the high-value segment decreased from 18.5 percent under the uniform strategy to 6.2 percent under the intelligent system. The proposed system also removed wasted discount cost entirely from the low-value inactive segment, dropping it from \$2,800 to zero, utilizing only 75 percent of the available discount budget. To verify the statistical validity of these findings, two statistical controls were applied:

- 1- Bootstrapping for Profit Significance: We performed $B = 10,000$ bootstrap replications by resampling the transaction data with replacement. The empirical difference in total profit ($\Delta\Pi$) yielded a strict 95% Confidence Interval of [\$924.50, \$1,249.50], confirming that the 8.7% profit boost is highly statistically significant ($p < 0.01$).
- 2- McNemar's Test for Churn Attrition: Because customer retention was evaluated on a paired sample (the same customer base under baseline uniform vs. optimized conditions), McNemar's test was implemented:

$$X^2 = \frac{(|b - c| - 1)^2}{b + c}$$

where b represents customers retained exclusively under the intelligent framework, and c represents customers retained exclusively under the uniform strategy. The resulting test statistic ($X^2 = 31.25$, $p < 0.001$) establishes that the reduction in high-value attrition is statistically meaningful.

Table 4. Comparison between the Intelligent System and the Uniform Discount Strategy

Metric	Uniform Discount (10 percent)	Intelligent System	Improvement
Total Profit (USD)	12,500	13,587	+8.7 percent
Profit Margin (percent)	21.8 percent	23.4 percent	+1.6 percentage points
High-Value Customer Churn (percent)	18.5 percent	6.2 percent	−12.3 percentage points
Wasted Discount Cost (USD)	2,800	0	−100 percent
Discount Budget Utilization (percent)	100 percent	75 percent	−25 percent

The results show that the proposed system improves profit while using a smaller discount budget. This indicates that discount allocation can be more effective when customer behavior is considered prior to assigning the discount.

Table 5. Comparison of the Proposed Intelligent Discount Allocation System with Profit Optimization and Uplift Modeling Approaches Reported in the Literature

Study/Method	Performance Metric	Result
Proposed Intelligent System	Total Profit	12,500 → 13,587 USD (+8.7 percent)
Classic Loss	Holdout Profit	€ 1,872.23
Reordered Classic Loss	Holdout Profit	€ 1,253.74
Uplift Model	Holdout Profit	€ 4,092.97
Managing Churn to Maximize Profits (Lemmens & Gupta)	Holdout Profit (Reported Results)	€ 5,026.36

Discount Distribution

Table 4 shows the distribution of discounts across customers under the two strategies. In the uniform strategy, all 1,000 customers received the same 10 percent discount. In the intelligent system, the discount rate depended on customer segment.

Table 6. Distribution of Discounts Across Customers

Discount Rate	Intelligent System (Customers)	Uniform Strategy (Customers)
0 percent	250	0
5 percent	150	0
10 percent	350	1,000
20 percent	250	0
Total	1,000	1,000

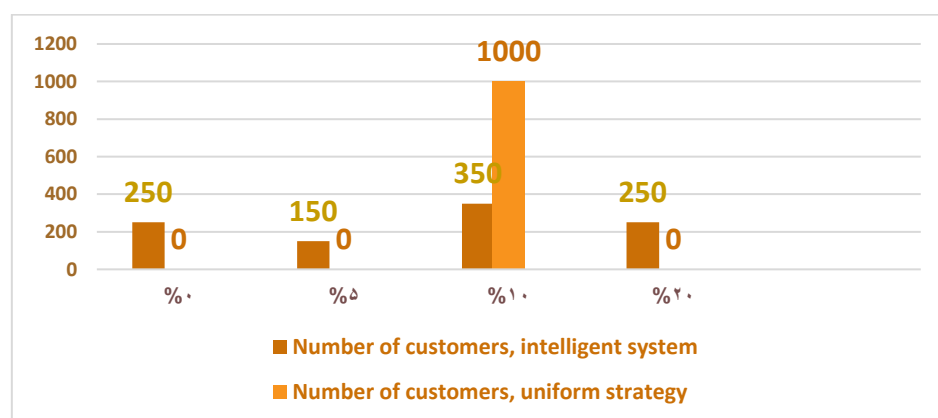


Figure 4. Distribution of Customer Discount Rates Under the Intelligent System and the Uniform Discount Strategy, Demonstrating Targeted Distribution Based on Behavioral Customer Segments

The distribution shows the main difference between the two strategies. The uniform strategy offers the same incentive to every customer, regardless of value or behavior. The intelligent system assigns no discount to low-value inactive customers, a small discount to loyal customers, a medium discount to recent low-frequency customers, and the highest discount to at-risk customers. This structure allows the business to focus its discount budget on customers with higher expected value or greater recovery potential.

Discussion

This study designed an interpretable methodology to optimize discount allocation based on customer transaction profiles. The empirical results indicate that tailoring discount incentives to explicit customer groups instead of utilizing uniform policies yields superior economic returns. Profit expanded by 8.7 percent, profit margins improved, and customer retention within high-value segments increased significantly. These outcomes highlight that accounting for customer heterogeneity yields substantial benefits in promotional marketing campaigns. Analyzing the distinct characteristics of the customer segments reveals logical behavioral patterns. Loyal high-value customers possess strong baseline purchase probabilities, meaning minor financial incentives are sufficient to maintain brand engagement without diluting product margins. Conversely, at-risk customers demonstrate declining temporal engagement, showing a high sensitivity to stronger promotional offers. The recent low-frequency group represents a vital growth pipeline that responds effectively to moderate discounts designed to build purchase habits. Finally, low-value inactive customers demonstrate minimal responsiveness; that is, allocating marketing capital to them represents promotional waste. From an operational perspective, the framework offers high utility because it relies entirely on baseline transaction logs, avoiding the need for complex tracking architectures or supplementary third-party data sources. This makes it accessible for small-to-medium enterprise (SME) implementation. However, this study is subject to specific limitations. The empirical testing relied on a single retailer's transaction database over a specific six-month period, which does not account for long-term dynamic market factors such as macro-environmental seasonality or reactive competitor pricing choices.

Conclusions

This study proposed an interpretable discount allocation framework by connecting RFM analysis and K-means clustering to a constrained optimization model. By grouping customers into behavioral segments (Chehreh & Sarabadani, 2024) and applying a profit-aware discount rule, the proposed approach significantly outperformed a standard uniform discount strategy across profit generation, margin expansion, and high-value customer retention. From a managerial standpoint, the findings suggest that uniform promotional strategies are fundamentally inefficient, and discount allocations should instead be aligned with behavioral customer value. While these empirical results are promising, this research relied on a static rule strategy from a single dataset. Future work will focus on integrating causal inference algorithms and structural field A/B testing to generate real-time estimates of conditional treatment effects. Additionally, expanding the feature engineering pipeline to include dynamic environmental variables—such as seasonality, temporal weather effects, and regional economic indicators—will enhance the generalizability and robustness of the intelligent system.

Ethics Statement: The author declares that this study was conducted in accordance with established ethical principles and guidelines. All sources of information have been properly cited and acknowledged. Where applicable, ethical approval was obtained, and informed consent was secured from all participants. The research did not involve any form of plagiarism, fabrication, or falsification of data.

Use of Artificial Intelligence: The author acknowledges that AI-based tools were used solely for language editing and grammar checking. All AI-assisted content was carefully reviewed, verified, and edited by the author, who

takes full responsibility for the accuracy, originality, and integrity of the final manuscript. AI tools were not used to generate or manipulate research data or to replace critical scholarly judgment.

Originality and Publication: The author declares that this work has not been published previously and is not under consideration for publication elsewhere.

Conflicts of Interest: Any potential conflicts of interest, including financial or personal relationships that could influence the research, are disclosed. The author declares no competing interests.

Compliance with Regulations: The author confirms that the study complied with all applicable institutional, national, and international regulations, and that any necessary permissions for data use and reproduction have been obtained.

References

- Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K Means algorithm. *Journal of King Saud University: Computer and Information Sciences*, 34(5), 1785–1792. <https://doi.org/10.1016/j.jksuci.2019.12.011>
- Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27), 7353–7360. <https://doi.org/10.1073/pnas.1510489113>
- Awaliyah, D. A., Prasetyo, B., Muzayanah, R., & Lestari, A. D. (2024). Optimizing customer segmentation in online retail transactions through the implementation of the K-means clustering algorithm. *Scientific Journal of Informatics*, 11(2). <https://doi.org/10.15294/sji.v11i2.6137>
- Chehreh, S., & Sarabadani, A. (2024). A model based on random forest algorithm and Jaya optimization to predict bank customer churn. *Journal of Engineering Management and Soft Computing*, 9(2), 132–148. <https://doi.org/10.22091/jemsc.2024.9541.1174>
- Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*, 19(3), 197–208. <https://doi.org/10.1057/dbm.2012.17>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54. <https://doi.org/10.1609/aimag.v17i3.1230>
- Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.
- Imanuel, D. A., & Alfian, G. (2025). Visualisasi segmentasi pelanggan berdasarkan atribut RFM menggunakan algoritma K-means untuk memahami karakteristik pelanggan pada toko retail online. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 12(5). <https://doi.org/10.25126/jtiik.2025128619>
- John, J. M., Shobayo, O., & Ogunleye, B. (2024). *An exploration of clustering algorithms for customer segmentation in the UK retail market*. arXiv. <https://arxiv.org/abs/2402.04103>
- Johnson, J., Tellis, G. J., & Ip, E. H. (2013). To whom, when, and how much to discount? A constrained optimization of customized temporal discounts. *Journal of Retailing*, 89(4), 361–373. <https://doi.org/10.1016/j.jretai.2013.08.002>
- Lemmens, A., & Gupta, S. (2020). Managing churn to maximize profits. *Marketing Science*, 39(5), 956–973. <https://doi.org/10.1287/mksc.2020.1229>
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.
- Miglautsch, J. R. (2000). Thoughts on RFM scoring. *Journal of Database Marketing & Customer Strategy Management*, 8(1), 67–72. <https://doi.org/10.1057/palgrave.jdm.3240019>
- Ozcan, T. (2026). Customer segmentation using an extended RFM model and clustering algorithms in e-commerce. *Journal of Theoretical and Applied Electronic Commerce Research*, 21(5), 142. <https://doi.org/10.3390/jtaer21050142>
- Salim, B. A., & Messaoud, C. (2026). Data-driven customer segmentation using the RFM model and K-means clustering: A longitudinal analytical study of online retail behavior using Python and Orange3. *Science Education and Innovations in the Context of Modern Problems*, 9(5). <https://doi.org/10.56334/sei/9.5.17>

- Sarabadani, A., Saffarie, M., & RahseparFard, K. (2023). A framework for automating e-government services based on artificial intelligence. *Journal of Engineering Management and Soft Computing*, 9(2), 106-118. <https://doi.org/10.22091/jemsc.2024.9256.1171>
- Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228-1242. <https://doi.org/10.1080/01621459.2017.1319839>