

# Analyzing Fresh Food Customer Loyalty: A Clustering and Ordinal Regression Approach

Iman Radipour<sup>1</sup>, Roya Eshraghi Samani<sup>\*1</sup>, and Bitra Rahimi Badr<sup>1</sup>

## ABSTRACT

This study investigates customer loyalty in Iran's chain stores, with a particular emphasis on fresh food consumers. The research utilizes a combination of K-means clustering and a weighted Recency, Frequency, Monetary (RFM) model, and ordinal logistic regression to analyze customer behavior. Using real transaction data from 9,014 customers alongside questionnaire responses, the analysis categorizes customers into four distinct groups: very loyal, loyal, at-risk, and disloyal. The weighted RFM model indicates that recency is the most significant predictor of loyalty. Further, the ordinal logistic regression identifies several key factors influencing loyalty: age, marital status, income level, perceived food quality, preference for modern stores, and brand image. These all have positive effect on loyalty; on the contrary, the importance of price and a preference for packaging-free products negatively impact loyalty. These findings provide actionable insights for retail managers, enabling them to develop cluster-specific strategies that enhance customer loyalty and strengthen competitiveness in Iran's dynamic retail sector.

**Keywords:** Chain retailing, Customer segmentation, Purchase behavior.

## INTRODUCTION

In recent years, there has been a significant increase in the establishment of chain stores in Iran, which reflects a gradual shift in consumer purchasing patterns towards these retailers (Seifollahi *et al.*, 2020). This trend has increased competition within the retail sector, compelling businesses to prioritize customer loyalty and competitiveness (Noorani Kootenay *et al.*, 2021). To enhance customer loyalty, grocery retailers should align their brand image with the lifestyles of their target customers, clearly define consumers' social identities, and adapt to evolve consumer behaviors in the post-pandemic era (Liu *et al.*, 2022).

The prevailing direction of contemporary marketing is movement towards a customer-centric approach (Zintso *et al.*, 2023). To get adapted to these changing conditions,

organizations are expected to prioritize customer needs on one hand, and align their culture and processes with customer goals on the other hand (Akbar, 2024). Maintaining customer loyalty is a crucial strategy for companies seeking to thrive in a competitive environment. Customer loyalty plays a significant role in customer retention, enabling retail companies to compete effectively in various aspects i.e. market segmentation, service systems, pricing, and product quality (Suriasha, 2023).

The concept of consumer loyalty is defined as the willingness of customers to consistently engage in repeated purchases. Loyal customers show a positive attitude towards service providers, indicating a preference for them when seeking services (Sen *et al.*, 2023). Customer loyalty can be characterized as a consumer's steadfast commitment to regular purchase, or

---

<sup>1</sup> Department of Agricultural Economics, Ka.C, Islamic Azad University, Karaj, Islamic Republic of Iran.

\*Corresponding author; e-mail: ro.eshraghi@iaau.ac.ir



subscription to a product or service of interest on successive occasions, even in the face of enticing marketing offers (Suyanto and Femi, 2023). Loyalty behavior is associated with the repeated patronage of customers who consistently engage in purchasing a particular product or service (Ho and Wong, 2023).

Researches indicate that attaining customer loyalty is essential for chain stores seeking to establish a sustainable competitive advantage (Tanveer *et al.*, 2021). According to Kotler and Armstrong (2012), making customers loyal necessitates differentiating among different customer segments, and no company can afford to allocate the same level of attention to profitable customers as it does to unprofitable customers. Studies demonstrate that businesses can use clustering techniques to identify distinct customer groups and customize their strategies accordingly (De Sousa *et al.*, 2024). The consumer segmentation technique enables businesses to identify their most profitable customers (Salameh *et al.*, 2020). A key objective of this segmentation is to understand the attitudes of distinct customer groups toward specific products or service (Mohammadian and Makhani, 2019). Cluster analysis facilitates identification of similar customer groups (De Sousa *et al.*, 2024). Furthermore, the integration of machine learning algorithms with user data exemplifies an effective approach to customer segmentation, allowing businesses to uncover customer segments that may be challenging to identify through intuition and manual data analysis (Kumar, 2023).

K-means clustering is recognized as a more stable and suitable method for customer segmentation compared to other algorithms (Wilbert *et al.*, 2023). This technique has been successfully implemented across various industries, including finance (Rizkyanto and Gaol, 2023), retail (Chiou *et al.*, 2009; Baltas *et al.*, 2010; Gustriansyah *et al.*, 2020; Wilbert *et al.*, 2023; Tan Chun and Nurulhuda, 2021; 2020), Hospitality and Tourism (Chang *et*

*al.*, 2022). Utilizing K-means clustering in conjunction with RFM (Recency, Frequency, Monetary) data allows for the segmentation of customers into distinct categories based on their loyalty (Dey and Banerjee, 2023). These segmentation strategies empower businesses customize their approaches, enhance customer engagement, and foster growth by addressing the specific needs of each cluster (Kasem *et al.*, 2024; Wilbert *et al.*, 2023; Mensouri *et al.*, 2022). The combination of the RFM model and K-means clustering serves as an effective method analyzing customer behavior and improving marketing strategies (Mensouri *et al.*, 2022; Smaili and Hachimi, 2023; Barus *et al.*, 2023). The RFM model itself is a marketing technique designed to analyze customer behavior and segment customers based on their purchasing patterns (Alves Gomes and Meisen, 2023).

This research aimed to analyze the loyalty of fresh food consumers by examining their purchasing behaviors. This study investigates the factors that influence consumer loyalty at Shahrvand Chain Stores and proposes strategies to strengthen that loyalty. The findings of this research will provide valuable recommendations for improving customer loyalty, reducing customer attrition, and enhancing the competitiveness of supermarkets.

## MATERIALS AND METHOD

### Establishing the Level of Loyalty Using K-Means Clustering and RM Model

This stage utilizes a transaction dataset from Shahrvand Chain Store, compiled from multiple branches with similar input types to eliminate any specific store-related errors in the data and ensure uniformity across the dataset (Pradhan *et al.*, 2021). The data shall be assessed for completeness, consistency, and accuracy before undergoing a cleaning process. Once cleaned, the data will be transformed into RFM format. The recency

value represents time elapsed since the last transaction date relative to a specified period. The frequency value is derived from the total number of transactions made by the customer, while the monetary value reflects the total volume of transactions conducted by the customer. Subsequently, normalization of the attributes will be applied, as K-means uses clustering relies on distance measure, and the value range of the attribute values varies according to their nature (monetary can present values that may reach into the thousands, whereas other attributes typically range in the hundreds). This disparity can adversely affect the results (Wilbert *et al.*, 2023). According to Stone (1995), different weights should be assigned to the R, F, and M indicators based on the specific industry being analyzed, as the significance of each index varies according to the nature of the products of different sectors (Liang, 2010). Liu and Shih (2005) introduced an approach utilizing the Analytic Hierarchy Process (AHP) to determine the appropriate weights for each RFM variable, resulting in customized weights for each variable. Given that weight is a critical factor in the model, this approach is referred to as the weighted RFM model (Chouaten *et al.*, 2024).

After verifying the assigned weights, these weights will be multiplied by the corresponding values for each index within the Recency, Frequency, and Monetary Value model to calculate the final Weighted RFM (WRFM) for each client. The mathematical equation for calculating the WRFM value is as follows:

$$WRFM = W_R \times R + W_F \times F + W_M \times M \quad (1)$$

Where  $W_R$  is the Weight of Recency,  $W_F$  and  $W_M$  are the Weights of Frequency and the Weight of Monetary value, respectively.

When the variables and data preparation were established, customers were clustered with the K-Means algorithm using SPSS MODELLER 18 software. This method was initially introduced by MacQueen in 1967. In this algorithm, each pattern might have only one center at a time, so, in case  $X =$

$\{X_1, X_2, \dots, X_n\}$ , it includes a set of data. In this set,  $n$  represents the number of points. Also, if  $V = \{V_1, V_2, \dots, V_C\}$  is the set of cluster centers, in this set  $C$  shows the number of clusters. The purpose of the means-k clustering algorithm is to minimize the objective function  $j(v)$ , which is calculated according to Eq. 2:

$$j(V) = \sum_{i=1}^c \sum_{j=1}^{c_i} \|X_{ij} - V_j\|^2 \quad (2)$$

Where,  $\|x_{ij} - v_j\|^2$  is the Euclidean distance between  $x_{ij}$  and  $v_j$ .  $c_i$  is the data points of the  $i$ th cluster, and the  $i$ th center of  $v_i$  is calculated by the Equation (3):

$$V_i = \frac{1}{c_i} \sum_{j=1}^{c_i} X_{ij} \quad i = 1, 2, \dots, c \quad (3)$$

The means-k clustering algorithm is an unsupervised iterative process that partitions a dataset into  $K$  clusters. Data points are randomly allocated to each cluster; later, the distance from each point to the cluster center is computed, and points are reassigned to the nearest cluster center. The aforementioned steps are reiterated inasmuch as no other change points are remained (MacQueen, 1967).

The K-means algorithm, has been often validated via the method Silhouette Index (SI) (Rousseeuw, 1987; Gustriansyah *et al.*, 2020; Bagirov, 2023; Barus *et al.*, 2023). The silhouette index might be calculated using the following equation (Barus, *et al.*, 2023):

$$SI = \frac{b-a}{MAX(a, b)} \quad (4)$$

Where, "a" represents the average of the intra-cluster distance and "b" represents the distance between each sample and another cluster, the value of the silhouette index would be between -1 and 1. If one point has a silhouette index close to 1, it would be a good clustering. If the silhouette index closes to -1, poor clustering or misclassification would be indicated. If the silhouette index is close to zero, an intermediate case is indicated (Gustriansyah *et al.*, 2020).



### Analyzing Customer Loyalty Using Ordinal Logistic Regression

Ordinal regression models, due to their effectiveness in handling ordered categorical data, have become increasingly important in marketing research. These models are particularly useful for analyzing customer preferences and choices (Gutiérrez *et al.*, 2016). Ordinal logistic regression is an excellent approach for modeling customer loyalty, especially when dealing with ordinal output variables. Certain researches have examined the influence of different elements on loyalty (Zikienė *et al.*, 2024; Figler *et al.*, 2011; McMullan and Gilmore, 2008). This method enables the estimation of odds ratios and confidence intervals, offering valuable insights into the significance of predictors. The coefficients in ordinal logistic regression are interpreted as log odds, which help determine the likelihood of an event being categorized in or above a specific category compared to the category below that. The odds ratio derived from the model reflects both the strength and direction of the relationship between independent variables and the likelihood of the dependent variable. The ordinal logistic regression model can be defined as follows (Wang *et al.*, 2022):

$$\log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right) = \alpha_j - (\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad (5)$$

Where,  $j = 1, 2, \dots, J-1$  represent the levels of the ordinal loyalty category variable  $Y$ ,  $J$  is an integer denoting the number of categories of  $Y$ ,  $\alpha$  is the intercept, and  $\beta$  is the slope for all  $x$  predictors. The estimated coefficients derived from the ordinal logit model are not easily interpretable in a direct manner. To extract more meaningful insights from the model, it is essential to compute the marginal effects for each customer group. A widely accepted approach for interpreting such models, as described by Wooldridge (2016), involves estimating the regression coefficients, transforming these coefficients into their exponential form, and, subsequently, calculating the coefficients in terms of the

odds ratios associated with advancement to a higher category. Based on Stat's estimates, the odds ratio for  $X$  is expressed as  $(-\beta_p)$ . The selection of independent variables was guided by theoretical principles (Qin and Hu, 2024; Barbosa and Shojaei, 2023). Assessing the adequacy of a model is essential to prevent misleading inferences (Sarkar and Midi, 2010). Researchers should utilize multiple goodness-of-fit measures to thoroughly evaluate logistic regression models prior to drawing conclusions (Ailobhio and Ikughur, 2024). This study uses the Hosmer-Lemeshow test, a widely recognized goodness-of-fit measure, to assess the alignment between the predicted probabilities and observed outcomes (Hosmer and Lemeshow, 2013). Furthermore, the Cox and Snell and Nagelkerke  $R^2$  metrics are used to evaluate the explanatory power of the models (Sarkar and Midi, 2010). The presence of multicollinearity in the regression models is assessed using the Variance Inflation Factor (VIF) (Akinwande *et al.*, 2015).

The dataset employed for the cluster analysis is the real-world dataset comprising sales data from Shahrvand Chain Stores in Iran. It includes transactional records spanning from May 22 to March 22, 2023. The dataset contained fresh food products and a total of 870,145 sales transactions from 9,014 customers. The data consists of various details as to the quantity and value of customers' purchases, the brand and product purchased, the time and date of purchase, and the available relevant demographic information (Table 1).

The data regarding the variables influencing customer loyalty was collected via a structured questionnaire. The measures utilized in this phase were drawn from the existing literature and were appropriately adapted to fit the context of the study. The study population consisted of fresh food consumers from the Shahrvand Chain Store, identified at the beginning stages of the research. After clustering the customers and ranking these clusters based on loyalty, a stratified random sampling method with

**Table 1.** Sample information in customer records.

Titles	Information
User code	A seven-digit code
Unique invoice number	17-Digit code
National code	10 Digits
Date of birth	Day/Month/Year
Product Type	Mutton, chicken, fish, mushroom, ...
Brand	Ronak, Aida, Sina and....
Weight	1000 Grams
Sales value	\$
Selling price	\$
Discount	Percentage

proportional assignment was employed. From a total of 9,014 customers, 380 samples were selected in accordance with the guidelines established by Karjesi and Morgan (1970). The questionnaire was designed based on the variables outlined in Table 2. The questionnaire's validity was confirmed by experts, and its reliability was assessed using the coefficient of Cronbach's alpha ( $\alpha \geq 0.8$ ). Finally, customer loyalty was predicted using an ordinal logistic regression model.

## RESULTS

Prior to clustering, data cleaning, transformation, and scaling were performed. Following data preprocessing, the weights of Recency (R), Frequency (F), and Monetary (M) values were calculated using the methodology established by Liu and Shih (2005), where the weights of all the three variables are calculated. The averaging relative weights for a sample of 9014 customers were determined as follows: WR=

**Table 2.** The questions included in the questionnaire.

Variable	Classification–description
Gender	Male=0; Female=1
Age	18 ≤ Age < 30; 30 ≤ Age < 40 40 ≤ Age < 50; 50 ≤ Age < 60 60 ≤ Age
Marital status	Single = 0; Married = 1
Education	Under Diploma = 1; Diploma = 2 Post Diploma = 3; M Bachelor = 4 Master's degree and above = 5
Household income (Monthly)	Income ≤ 200\$ = 1; 200 < Income ≤ 400\$ = 2 400 < Income ≤ 600\$ = 3; 600 < Income ≤ 800\$ = 4; 800 < Income\$ = 5
Importance of brands variety	Very low = 1; Low = 2; Medium = 3; Much = 4; Very much = 5
Attention to Food quality	Very low = 1; Low = 2; Medium = 3; Much = 4; Very much = 5
Preferring packaging-free product	Very low = 1; Low = 2; Medium = 3; Much = 4; Very much = 5
Perceived service quality (Customer respect, comfort facilities, environmental hygiene, proper air conditioning, proper arrangement of products, responsiveness)	Very low = 1; Low = 2; Medium = 3; Much = 4; Very much = 5
Preference to buy from modern stores	No = 0; Yes = 1
Price Importance	Very low = 1; Low = 2; Medium = 3; Much = 4; Very much = 5
Brand Image Importance (The perspectives, interpretations, and sentiments held by customers to chain store)	Very low = 1, Low = 2, Medium = 3, Much = 4, Very much = 5



0.476; WF= 0.323, and WM= 0.201

To validate the accuracy of these weights, the inconsistency index of the comparisons made was calculated as 0.048, which is below the threshold of 0.1. This indicates that the results derived from the hierarchical analysis process can be considered reliable

In the following, the data was clustered by determining the optimal value of K using the Silhouette score (Table 3). According to the silhouette index, the best number of clusters was 4. Thus, the K-means clustering algorithm was employed to segment the customers into 4 distinct clusters.

The cluster names were selected based on distinct behavioral patterns and the insights of subject-matter experts, categorizing customers as very loyal, loyal, at-risk loyalty, and disloyal.

### Very Loyal Customers

These customers exhibit a brief interval between their most recent purchase from the chain store, a relatively high frequency of purchases, and a significant purchase value. Only 3% of the 9,014 customers belong to this cluster.

### Loyal

What is evident is that loyal customers exhibit a short interval between their most recent purchases from the chain store, demonstrating a relatively high frequency of purchases. However, in terms of purchase value, these transactions are not particularly high. Approximately 49% of the customers fall into this category.

### At-Risk Loyalty

This group of customers has a high frequency of purchases; however, they have not done any recent purchase. Their purchase value might be either high or low. There is a risk that these customers may discontinue their patronage in the future, and they represent 32% of the total customer base.

### Disloyal Customers

The customers in this category are the ones who have not shopped with the store for an extended period, and their purchase frequency is low. Some of them might have

**Table 3.** The result of the silhouette index through the implementation of K-means clustering.

Number of clusters	Silhouette
2	0.5
3	0.6
4	0.7
5	0.5

**Table 4.** Level of customer loyalty based on K-means clustering.

Cluster	Recency	Frequency	Monetary Value	Number of customers	Percentage of customers
Loyal	1 or 2 or 3 or 4	3 or 4	1 or 2 or 3 or 4	4417	49
Disloyal	1 or 2 or 3 or 4	1	1 or 2 or 3 or 4	1442	16
At- Risk	3 or 4	2 or 3	1 or 2 or 3 or 4	2884	32
Very Loyal	1 or 2 or 3 or 4	1 or 2	4	270	3
Total				9014	100

Source: The research findings.

made relatively high-value purchases during their limited interactions, though these transactions have not been repeated. This group of customers tend to shopping sporadically. Approximately 16% of the customers are in this cluster.

According to the characteristics of these clusters, the loyalty variable is measured on an ordinal scale with the following categories: Disloyal= 1; At-risk loyalty= 2; Loyal = 3; and Very loyal= 4.

the average age of the participants in this study was 39.74 years. The demographic profile of the respondents indicates that the majority were male (60%), married (54.47%), and held a bachelor's degree (38.94%). Additionally, the most common monthly income level among participants was classified as level 3, accounting for 56.52% of the respondents. In terms of the importance of food quality and safety, a significant majority (55.78%) rated its importance as high. Furthermore, when asked about their preference for modern retail stores, 61% of participants also selected the high option.

The model's goodness-of-fit and explanatory power were assessed using the Hosmer-Lemeshow test and  $R^2$  statistics (Table 5). The non-significant p-value obtained from the Hosmer-Lemeshow test indicates a good fit for the model. The  $R^2$  values indicate the model's classification accuracy.

The Variance Inflation Factor (VIF) was used to check multicollinearity among the independent variables, ensuring the reliability of the coefficients. The VIF for all variables in the model was less than 5, indicating that multicollinearity was not a concern for this model.

The ordinal logistic regression analysis was conducted to examine the factors influencing customer loyalty. The results indicated that, among demographic variables, age had a positive and statistically significant effect. That is, each increase in the age category was associated with a 20.9% increase in the odds of being in a higher loyalty category ( $P= 0.012$ ). Similarly, marital status demonstrated a substantial positive effect. Married individuals were 80.3% more likely to exhibit greater loyalty compared to their unmarried counterparts ( $P= 0.020$ ). Income level emerged as another highly significant predictor, with an increase in income category corresponding to a 34.9% rise in the odds of higher loyalty ( $P < 0.001$ ). In contrast, gender was found to have a significant negative effect, with females being 23.7% less likely than males to fall into a higher loyalty category ( $P= 0.012$ ). Education level showed a negative coefficient, suggesting a potential inverse relationship with loyalty; however, this effect did not reach statistical significance ( $P= 0.081$ ), rendering its impact inconclusive within the current dataset. Perception-based and preference-related factors also contributed significantly to customer loyalty. Perceived food quality positively influenced loyalty, increasing the odds of being in a higher loyalty category by 23.4% ( $P= 0.001$ ). Similarly, a preference for modern stores was associated with a 41.9% greater likelihood of higher loyalty ( $P < 0.001$ ), which underscored the importance of store modernity in enhancing customer retention. The brand image proved to be a critical determinant, with improvements in brand perception linked to

**Table 5.** Model Validation and Fit Statistics.

Test	Statistic	P-value
Hosmer-Lemeshow Test	7.82	0.415
Cox and Snell $R^2$	0.49	-
Nagelkerke $R^2$	0.57	-

Source: research findings.

**Table 6.** Estimation results of the ordinal logit regression for predicting fresh food buyers' loyalty.

Variable	Coeff	Wald test	Significance	Odds ratio
Age	0.19	5.53	0.012	1.209
Education level	-0.51	-2.37	0.081	0.600
Gender	-0.27	-2.91	0.012	0.763
Marital status	0.59	4.14	0.020	1.803
Income level	0.30	6.05	0.000	1.349
Perceived service quality	0.14	3.96	0.010	1.150
Food quality and safety	0.21	3.48	0.001	1.234
Perceived usefulness of shopping	0.16	5.093	0.006	1.174
Price Importance	-0.47	-6.795	0.001	0.625
Preference modern stores	0.35	0.431	0.000	1.419
Preferring packaging-free product	-0.48	-3.12	0.001	0.619
Brand Image	0.55	2.04	0.041	1.730

Source: Research findings.

a substantial 73% increase in the odds of higher loyalty ( $P= 0.041$ ). Conversely, the importance of price exhibited a significant negative association, with customers who prioritized price being 37.5% less likely to demonstrate higher loyalty ( $P= 0.001$ ). This indicates that price sensitivity may undermine loyalty. Additionally, a preference for packaging-free products was associated with a 38.1% reduction in the odds of being in a higher loyalty category ( $P= 0.001$ ), suggesting a potential conflict between this preference and loyalty to chain stores.

## DISCUSSION

The application of K-means clustering has successfully segmented customers into four distinct groups. Very loyal customers demonstrated frequent purchases with high monetary value and recent activity, while disloyal customers demonstrated sporadic engagement with low frequency. One participant's purchasing behavior aligns with Suyanto and Femi's (2023) definition of loyalty as a consistent commitment to repurchase despite external temptations. The WRFM model, which prioritizes recency over frequency and monetary value, highlights the significance of recent engagement in defining loyalty. This finding

is consistent with Dey and Banerjee (2023), who advocate RFM-based segmentation to develop tailored marketing strategies.

Demographic factors significantly influenced loyalty outcomes. Older customers showed a 20.9% increase in loyalty odds, married individuals were 80.3% more likely to demonstrate loyalty, potentially reflecting stable household needs. Higher income levels also correlated with a 34.9% increase in loyalty, suggesting that financial security fosters loyalty. Conversely, females were 23.7% less loyal than males, which may be attributed to differing shopping priorities or sensitivities, diverging from the findings of Sen *et al.* (2023). Perceptions of food quality and brand image emerged as strong drivers of loyalty, with a 23.4% and 73% increase in loyalty odds, respectively. A Preference for modern stores further enhanced loyalty by 41.9%, aligning with Liu *et al.*'s (2022) emphasis matching retail formats with consumer lifestyles. However, price sensitivity and a preference for packaging-free products reduced loyalty odds by 37.5% and 38.1%, respectively, highlighting a tension between cost consciousness and loyalty. This finding is consistent with Zhang *et al.*'s (2023) observations regarding the challenges of differentiation in homogeneous product markets, emphasizing loyalty as a competitive advantage, and

indicating that price sensitivity may disproportionately challenge chain stores in price-competitive environments. The study reflects the intensifying competition within Iran's retail sector, as noted by Seifollahi *et al.* (2020). The focus on service quality and operational strategies is aligned with the insights of Qin and Hu (2024), underscoring the shift toward customer-centric approaches in contemporary marketing (Zintso *et al.*, 2023). The positive influence of service quality and brand image on loyalty resonates with the findings of de Sousa *et al.* (2024), reinforcing their critical roles in fostering repeat patronage and advocacy.

To enhance customer loyalty, Shahrvand Chain Stores can implement strategies such as encouraging frequent shopping, targeting marketing, improving food quality and brand image, minimizing price sensitivity, modernizing store experiences, and balancing packaging preferences. These strategies can be implemented through time-limited promotions, targeted marketing campaigns, and personalized reminders for inactive customers. By adopting these approaches, Shahrvand can effectively tailor their strategies to distinct customer segments, thereby enhancing retention and strengthening its competitive position in Iran's evolving retail sector.

## REFERENCES

1. Ailobhio, D. T. and. Ikughur, J. A. 2024. A Review of Some Goodness-of-Fit Tests for Logistic Regression Model. *Asian J. Prob. Stat.*, **26(7)**: 75-85.
2. Akbar, M. A. 2024. Customer-Centric Strategies: Navigating the Dynamics of Marketing Management for Competitive Advantage. *ABIM*, **2(2)**: 96-109.
3. Akinwande, M., Dikko, H. and Samson, A. (2015) Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. *Open J. Stat.*, **5**: 754-767.
4. Alves Gomes, M. and Meisen, T. 2023. A Review on Customer Segmentation Methods for Personalized Customer Targeting in E-commerce Use Cases. *Inf. Syst. e-Bus. Manag.*, **21(3)**: 527-570.
5. Bagirov, A. M., Aliguliyev, R. M. and Sultanova, N. 2023. Finding Compact and Well-Separated Clusters: Clustering Using Silhouette Coefficients. *Pattern Recognit.*, **135**: 1-15.
6. Baltas, G., Argouslidis, P. C. and Skarmas, D. 2010. The Role of Customer Factors in Multiple Store Patronage: A Cost-Benefit Approach. *J. Retail.*, **86(1)**: 37-50.
7. Barbosa, B., Shabani Shojaei, A. and Miranda, H. 2023. Packaging-Free Practices in Food Retail: The Impact on Customer Loyalty. *Balt. J. Manag.*, **18(4)**: 474-492.
8. Barus, O. P., Nathasya, C. and Pangaribuan, J. J. 2023. The Implementation of RFM Analysis to Customer Profiling Using K-Means Clustering. *Math. Model. Eng. Probl.*, **10(1)**: 298-303.
9. Chang, I. C., Horng, J. S., Liu, C. H., Chou, S. F., and Yu, T. Y. 2022. Exploration of Topic Classification in the Tourism Field with Text Mining Technology: A Case Study of the Academic Journal Papers. *Sustainability*, **14(7)**: 4053.
10. Chiou, J. S., Wu, L. Y. and Sung, Y. P. 2009. Buyer Satisfaction and Loyalty Intention in Online Auctions: Online Auction Web Site versus Online Auction Seller. *J. Serv. Manag.*, **20(5)**: 521-543.
11. Chouaten, K., Rodriguez Rivero, C., Nack, F. and Reckers, M. 2024. Unlocking High-Value Football Fans: Unsupervised Machine Learning for Customer Segmentation and Lifetime Value. *Front. Sports Act. Living*, **6**: 1362489.
12. De Sousa, A., Moro, S., and Pereira, R. 2024. Cluster-Based Approaches Toward Developing a Customer Loyalty Program in a Private Security Company. *Appl. Sci.*, **14(1)**: 78.
13. Dey, D. and Banerjee, K. 2023. AI Driven Customer Segmentation and Recommendation of Product for Super Mall. *J. Mines Met. Fuels*, **71(5)**: 656-660.
14. Figler, S. A., Sriraj, P. S., Welch, E. W. and Yavuz, N. 2011. Customer Loyalty and Chicago, Illinois, Transit Authority Buses:



- Results from 2008 Customer Satisfaction Survey. *Transp. Res. Rec*, **2216(1)**: 148-156.
15. Gustriansyah, R., Suhandi, N. and Antony, F. 2020. Clustering Optimization in RFM Analysis Based on K-Means. *Indones. J. Electr. Eng. Comput. Sci.*, **18(1)**: 470-477.
  16. Gutiérrez, M. Pérez-Ortiz, J. Sánchez-Monedero, F., Fernández N. and Hervás-Martínez, C. 2016. Ordinal Regression Methods: Survey and Experimental Study. *IEEE Trans. Knowl. Data Eng.*, **28(1)**: 127-146.
  17. Ho, S. P. S. and Wong, A. 2023. The Role of Customer Personality in Premium Banking Services. *J. Financ. Serv. Mark.*, **28(2)**: 285-305.
  18. Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X., 2013. *Applied Logistic Regression*. 3rd Edition, John Wiley and Sons, Hoboken, New Jersey.
  19. Kasem, M.S., Hamada, M. and Taj-Eddin, I. 2024. Customer Profiling, Segmentation, and Sales Prediction Using AI in Direct Marketing. *Neural Comput. Appl.*, **36**: 4995-5005.
  20. Kotler, P. and Armstrong, G. 2012. *Principles of Marketing*. 14th Edition, Pearson Education Limited, Essex, England.
  21. Krejcie, R.V. and Morgan, D.W. 1970. Determining Sample Size for Research Activities. *Educ. Psychol. Meas.*, **30(3)**: 607-610.
  22. Kumar, A. 2023. Customer Segmentation of Shopping Mall Users Using K-Means Clustering. Chapter 13. In: *Advancing SMEs Toward E-Commerce Policies for Sustainability*. IGI Global, PP. 248-270.
  23. Liang, Y.H. 2010. Integration of Data Mining Technologies to Analyze Customer Value for the Automotive Maintenance Industry. *Expert Syst. Appl.*, **37(12)**: 7489-7496.
  24. Liu, D. R. and Shih, Y. Y. 2005. Integrating AHP and Data Mining for Product Recommendation Based on Customer Lifetime Value. *Inf. Manag.*, **42(3)**: 387-400.
  25. Liu, H., Mu, Y., Fu, X. and Liu, Y. 2022. Passionately Attached or Properly Matched? The Effect of Self-Congruence on Grocery Store Loyalty. *Br. Food J.*, **124(11)**: 4054-4071.
  26. MacQueen, J. B. 1967. Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, University of California Press, Berkeley, PP. 281-296.
  27. McMullan, R. and Gilmore, A. 2008. Customer Loyalty: An Empirical Study. *Eur. J. Mark.*, **42(9/10)**: 1084-1094.
  28. Mensouri, D., Azmani, A. and Azmani, M. 2022. K-Means Customers Clustering by their RFMT and Score Satisfaction Analysis. *Int. J. Adv. Comput. Sci. Appl.*, **13(6)**: 469-476.
  29. Mohammadian, M., and Makhani, I. 2019. RFM-Based Customer Segmentation as an Elaborative Analytical Tool for Enriching the Creation of Sales and Trade Marketing Strategies. *Int. Acad. J. Acc. Fin. Manag. (IAJAFM)*, **6(1)**: 102-116.
  30. Noorani kootenae, M., Rezaei Dolatabadi, H. and Mohammad Shafiee, M. 2021. Developing a Model for Retailer Brand Competitiveness with the Grounded Theory Approach. *Q. J. Brand Manag.*, **8(3)**: 44-15.
  31. Pradhan, S., Patel, G. and Priya, P. 2021. Measuring Customer Lifetime Value: Application of Analytic Hierarchy Process in Determining Relative Weights of LRFM. *Int. J. Anal. Hierarchy Process*, **13(3)**: 526-547.
  32. Qin, W. and Hu, M. 2024. Dennis Supermarket Chain Customer Loyalty Study. *FHSS*, **4(7)**: 349-358.
  33. Rizkyanto, H. and Gaol, F. L. 2023. Customer segmentation of personal credit using Recency, Frequency, Monetary (RFM) and K-means on financial industry. *Int. J. Adv. Comput. Sci. Appl.*, **14(4)**: 149-159.
  34. Rousseeuw, P. J. 1987. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math. J. Comput. Appl. Math.*, **20**: 53-65.

35. Salameh, A. A., Hatamleh, A., Azim, M. S. and Kanaan, A. G. 2020. Customer-Oriented Determinants of e-CRM Success Factors. *Uncertain Supply Chain Manag.*, **8(4)**:713-720.
36. Sarkar, S. K. and Midi, H. 2010. Importance of Assessing the Model Adequacy of Binary Logistic Regression. *J. Appl. Sci.*, **10**: 479-486.
37. Seifollahi, N., Rahimi clever, H. and Shirnezhad Moghanlou, M. 2020. The Effect of Response to Point of Purchase Advertising on Store Brand Loyalty with the Mediating Role of the Customer's Self-Concept. *Q. J. Brand Manag.*, **7(1)**: 213-260.
38. Sen, S. S., Alexandrov, A., Jha, S., Mcdowell, W. C. and Babakus, E. 2023. Convenient Competitive? How Brick-And-Mortar Retailers Can Cope with Online Competition. *Rev. Manag. Sci.*, **17(5)**: 1615–1643.
39. Smaili, M.Y. and H. Hachimi. 2023. New RFM-D Classification Model for Improving Customer Analysis and Response Prediction. *Ain Shams Eng. J.*, **14(12)**: 102254.
40. Stone, B. 1995. *Successful Direct Marketing Methods*. 5th Edition, NTC Business Books, Lincolnwood, IL.
41. Suriانشa, R. 2023. The Role of Customer Loyalty on Customer Retention in Retail Companies. *Int. J. Multidiscipl. Res. Anal.*, **6 (7)**: 3359-3365.
42. Suyanto, A. and Femi, S. R. 2023. Analysis of the Effect of Impulsive Purchase and Service Quality on Customer Satisfaction and Loyalty in Beauty E-Commerce. *Qual. Access Success*, **24(194)**: 18 - 23.
43. Tan C. K. and Nurulhuda F. M. A. 2021. Customer Profiling for Malaysia Online Retail Industry using K-Means Clustering and RM Model. *Int. J. Adv. Comput. Sci. Appl.*, **12(1)**: 106-113.
44. Tanveer, M., Ahmad, A. R., Mahmood, H. and Ul Haq, I. 2021. Role of Ethical Marketing in Driving Consumer Brand Relationships and Brand Loyalty: A Sustainable Marketing Approach. *Sustainability*, **13(12)**: 1-17.
45. Taulant K., Brunela T. and Shkelqim F. 2024. Understanding Customer Satisfaction Factors: A Logistic Regression Analysis. *J. Educ. Soc. Res.*, **14(2)**: 218
46. Wang, H., Quintana, F. G., Lu, Y., Mohebujjaman, M. and Kamronnahr, K. 2022. How Are BMI, Nutrition, and Physical Exercise Related? An Application of Ordinal Logistic Regression. *Life*, **12(12)**: 2098.
47. Wilbert, H. J., Hoppe, A. F., Sartori, A., Stefenon, S. F. and Silva, L. A. 2023. Recency, Frequency, Monetary Value, Clustering, and Internal and External Indices for Customer Segmentation from Retail Data. *Algorithms*, **16(9)**: 396.
48. Wooldridge, J. 2016. *Introductory Econometrics: A Modern Approach*. 6th Edition, Standalone Book, Cengage Learning.
49. Zhang, D., Shen, Z. and Li, Y. 2023. Requirement Analysis and Service Optimization of Multiple Category Fresh Products in Online Retailing Using Importance-Kano Analysis. *J. Retail. Consum. Serv.*, **72**:103253.
50. Zikienė, K., Kyguolienė, A. and Kisieliauskas, J. 2024. The Influence of Creative and Innovative Loyalty Programs' Features on Customers' Attitudinal, Aonative, and Behavioural Loyalty. *Creat. Stud.*, **17(1)**: 254–273.
51. Zintso, Y., Fedorishina, I., Zaiachkovska, H., Kovalchuk, O. and Tyagunova, Z. 2023. Analysis of Current Trends in The Use of Digital Marketing for The Successful Promotion of Goods and Services in Ukraine. *Financ. Credit Act.: Probl. Theory Pract.* **3(50)**: 174-184.



## تحلیل وفاداری مشتریان مواد غذایی تازه: رویکرد خوشه بندی و رگرسیون ترتیبی

ایمان رادی پور، رویا اشراقی سامانی، و بیتا رحیمی بدر

### چکیده

پژوهش حاضر به بررسی وفاداری مشتریان مواد غذایی در فروشگاه‌های زنجیره‌ای می‌پردازد. روش مورد استفاده در این مطالعه ترکیبی از خوشه‌بندی به روش K-means با استفاده از مدل WRFM و رگرسیون لجستیک ترتیبی برای تحلیل وفاداری مشتریان است. با بهره‌گیری از داده‌های واقعی تراکنش‌های 9014 مشتری و پاسخ‌های 381 پرسشنامه، مشتریان از نظر وفاداری به چهار خوشه بسیار وفادار، وفادار، وفادار ریسکی و غیر وفادار تقسیم شدند. نتایج مدل WRFM نشان داد که شاخص تازگی خرید (Recency) مهم‌ترین پیش‌بینی‌کننده وفاداری است. رگرسیون لجستیک ترتیبی نشان داد که سن، وضعیت تأهل، سطح درآمد، کیفیت غذایی ادراک‌شده، ترجیح به فروشگاه‌های مدرن و تصویر برند، تأثیر مثبتی بر وفاداری خریداران مواد غذایی فروشگاه‌های زنجیره‌ای دارد، اما حساسیت به قیمت و ترجیح برای محصولات بدون بسته‌بندی تأثیر منفی دارند. این یافته‌ها استراتژی‌های خاص خوشه را برای مدیران خرده‌فروشی فراهم می‌کند تا وفاداری مشتریان را بهبود بخشند و رقابت‌پذیری را در چشم‌انداز پویای خرده‌فروشی ارتقا دهند.