

Predicting Consumer Profiles to Enhance Targeted Marketing



Eleftheria Matta, George Stalidis , and Kyriaki Dimitriadou

Abstract This study addresses the need for behavior-based customer segmentation in the retail sector by introducing a novel methodological framework that combines multidimensional factor analysis with machine learning. The framework also yields strategic insights with direct implications for retail marketing, campaign management, and customer relationship development. Adopting a data-driven approach, the study uncovers behavioral patterns among supermarket customers in Greece. Using factor and clustering methods, six distinct shopper profiles were identified based on purchasing habits, store preferences, promotional responsiveness, and affinity for private label products. A predictive model was then developed to classify unknown customers into these profiles. The results provide practical tools for targeted marketing and customer engagement. Key business implications include enhanced personalization, improved customer loyalty strategies, and more efficient campaign planning, demonstrating the strategic value of integrating behavioral analytics into retail decision-making.

1 Introduction

Understanding consumer behavior is crucial for the success of modern marketing strategies, particularly in the competitive retail sector. This study explores the application of advanced customer profiling techniques to enhance marketing effectiveness, focusing on supermarket shoppers in Greece, both in-store and online. By delving into consumer purchase history and loyalty program data, our objective was to uncover the underlying behavioral factors that drive shopping decisions and to translate these findings into customer segments that can guide personalized marketing actions. Our aim was to move beyond traditional demographic profiling,

E. Matta · G. Stalidis (✉) · K. Dimitriadou
Department of Organization Management, Marketing and Tourism, International Hellenic University, Thessaloniki, Greece
e-mail: stalidgi@ihu.gr

focusing on behavioral variables, such as product category preferences and sensitivity to promotions, to provide a more detailed and insightful understanding of consumer decision-making.

2 Literature Review on Customer Profiling

Effective customer profiling and segmentation strategies are fundamental to achieving precision marketing. Across literature, numerous typologies, clusters, and segments have been developed using advanced analytical methods. Demographic and behavioral profiling remains a common approach in retail marketing research. For instance, Buchdadi [1] identified five distinct customer segments among Walmart customers based on age, income, and purchasing behavior. The segmentation guided targeted marketing strategies, including premium-focused campaigns for older consumers and digital engagement initiatives for younger audiences. Temporal and promotional behavior also play a significant role in customer differentiation. Namin and Dehdashti [2] utilized latent class modeling to identify consumer groups based on day-of-week shopping and response to promotions. For food and drink, 81% of shoppers preferred weekends and responded well to promotional offers, whereas a smaller group with larger families and lower income shopped primarily on Fridays. In the personal care category, consumers were grouped into price-sensitive, wealthy, and promotion-sensitive clusters, with each segment demonstrating distinct shopping timing and promotional responsiveness.

A dynamic perspective on behavioral segmentation was adopted by Abbasimehr and Sheikh Bagheri [3], who used recency, frequency, and monetary (RFM) variables to cluster grocery and supermarket consumers over time. Their model revealed evolving customer trajectories, such as stable middle-value, downward trending, growing low-value, and improving middle-to-high value groups. These trajectories enabled marketers to identify potential for growth or risk of attrition, supporting data-driven decisions for loyalty and retention campaigns. Expanding on the RFM framework, Peker et al. [4] introduced the LRFMP model, adding length and periodicity as dimensions to capture long-term engagement. Vanessa and Japutra [5] applied an RFM analysis to [BigBasket.com](https://www.bigbasket.com) customers, categorizing them into eight profiles: most valuable customers, barnacles, big basket shoppers, first timers, churned valuable customers, churned barnacles, churned big basket shoppers, and strangers. These profiles were used to optimize notification timing and campaign design, with the goal of improving engagement, retention, and profitability.

Psychographic segmentation has also gained attraction for its ability to capture consumers' underlying motivations and attitudes. Conlin and Labban [6] identified high involvement (HI) and low involvement (LI) grocery shoppers. HI consumers, often younger with children, approached grocery shopping as a pleasurable, identity-reinforcing activity, expressing enjoyment in bargain hunting and trying new products. In contrast, LI shoppers—typically older—were more price-conscious and less influenced by brand experience, favoring efficiency over

engagement. These findings support dual strategic approaches: experiential and emotional marketing for HI segments, and practical, efficiency-driven messaging for LI consumers. Loyalty, both attitudinal and behavioral, further refines customer segmentation. Hallikainen et al. [7] found that attitudinal loyalty was more strongly influenced by personalized price promotions than by generic product recommendations. Behavioral loyalty, characterized by repeat purchases, was less affected by personalization. This distinction suggests that reducing cognitive effort and customizing incentives are vital for establishing deeper emotional connections with customers. Additionally, Janssens et al. [8] employed value-based segmentation by applying Holbrook's typology to Dutch grocery shoppers, identifying values such as aesthetics, altruism, and efficiency as key predictors of satisfaction and repurchase behavior.

Emerging technologies have further enriched segmentation methods, particularly through machine learning and neural network-based approaches. Deep embedded clustering has been used to analyze Vietnamese supermarket data, identifying four high-dimensional customer types: daily essentials buyers, fashion and cosmetics shoppers, convenience store users, and processed goods consumers [9]. This neural approach enabled the extraction of nuanced behavioral patterns that traditional methods may overlook. Visit-level segmentation was also explored by analyzing market basket data to categorize mission-driven shopping behaviors, such as breakfast, meal preparation, snacks and beverages, or hygiene product missions [10]. These patterns varied by time of day and store format, allowing for tailored merchandising and promotional planning. Finally, regional and socioeconomic segmentation also contribute to deeper profiling. Theodoridis and Chatzipanagiotou [11] contributed a cultural dimension by profiling Greek supermarket shoppers into typical, unstable, social, and occasional segments, based on store image perception and satisfaction drivers, emphasizing the need for regional adaptation in marketing strategies. Building on the above studies, our work contributes a unified behavioral segmentation and prediction framework. Unlike prior models focusing on static attributes or metrics, our approach combines MCA, HCPC, and decision trees to enable interpretable, real-time customer classification based on behavioral variables.

3 Methods

In the proposed data-driven methodology, factors and clustering methods were applied to identify distinct shopper profiles. These were then integrated with a predictive model, which allows retailers to classify their customers into segments in real time and tailor their communication strategies accordingly. A multistage approach was employed, combining multiple correspondence analysis (MCA), hierarchical clustering on principal components (HCPC), and decision tree modeling for predictive classification (Fig. 1). This sequence allowed for both the identification of underlying consumer profiles and the ability to classify new customers based on their characteristics.

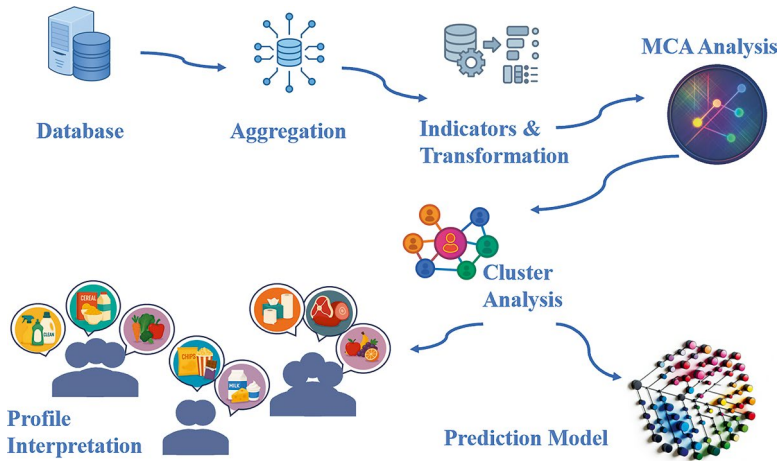


Fig. 1 The data analysis process scheme

The dataset was derived from a rolling year's sales (1/4/2021–31/3/2022) of a large supermarket chain in Greece. It was a small sample of the company's clientele that consisted of 61,895 individuals (physical store, online, and hybrid customers). The number of baskets in the dataset was approximately two million. In order to capture the customers' buying behavior, the purchase data were aggregated at the customer and product group levels. The quantities that each customer purchases from each product group were transformed to three-level qualitative variables, expressing the preference of the customer for the product group (low, average, high), in relation to other product groups and to the average customer's basket. Furthermore, 17 indicators were evaluated for each customer, e.g., their preference toward private label products and products on promotion [12]. These indicators were derived from purchase history and converted into categorical variables using thresholds.

The primary method was MCA, a dimensionality-reduction method tailored for datasets comprising categorical variables [13, 14]. MCA is particularly effective for visualizing and interpreting complex relationships in multidimensional data. It transforms the original data into a reduced number of dimensions (principal components), facilitating the identification of latent structures without imposing prior assumptions [15]. The factors revealed by MCA were used as input to HCPC, which combines the interpretability of hierarchical clustering with the efficiency of dimensionality reduction, enabling robust groupings of observations. This step allowed us to identify distinct customer clusters characterized by their purchasing behavior, product preferences, and sociodemographic profiles [16]. Both MCA and HCPC analyses were conducted using the FactoMineR packages in R.

Following the factor and clustering analysis, the results of the clustering method were used to train a decision tree model to classify customers into one of the identified clusters. The goal was to create an interpretable model capable of predicting the

profile to which an unknown customer belongs, based on a set of input variables. Decision trees are widely applied in customer segmentation for their ability to classify individuals into target groups through a transparent and visually interpretable rule-based structure [17]. The model not only predicts group membership but also reveals which variables are most influential in determining customer profiles [18]. This supervised learning step bridges the gap between exploratory segmentation and operational decision-making, allowing retailers to automate customer classification for targeted marketing.

4 Analysis Results

The two most important dimensions resulting from MCA are visible on the factorial plane F1 × F2 (Fig. 2). The 1st factor (32.1% of inertia) differentiated the limited purchases of occasional customers from the average purchases of regular customers. The vertical axis (2nd factor—3.4% of inertia) differentiated high purchasing of food products from nonfood groups and was thus the food vs nonfood factor. The 3rd factor differentiated home cooking from snacking behaviors, the 4th factor was e-shop vs physical store, and the 5th factor was interpreted as quality vs economy.

In the next analysis stage, the application of HCPC resulted in six clusters. In order to associate clusters with customer profiles, they were projected on the factorial planes F1 × F2 and F3 × F4, as shown in Fig. 3. Each cluster was interpreted considering its position on the factors. For example, on the factorial plane F1 × F2 (occasional vs regular customers x food vs nonfood), we observed that Clusters 1 and 2 are on the left side of the 1st factor and correspond clearly to occasional customers with low spending, while Cluster 5 is on the regular customers’ side and is linked to nonfood products (top side). The six resulting clusters are shown in Table 1.

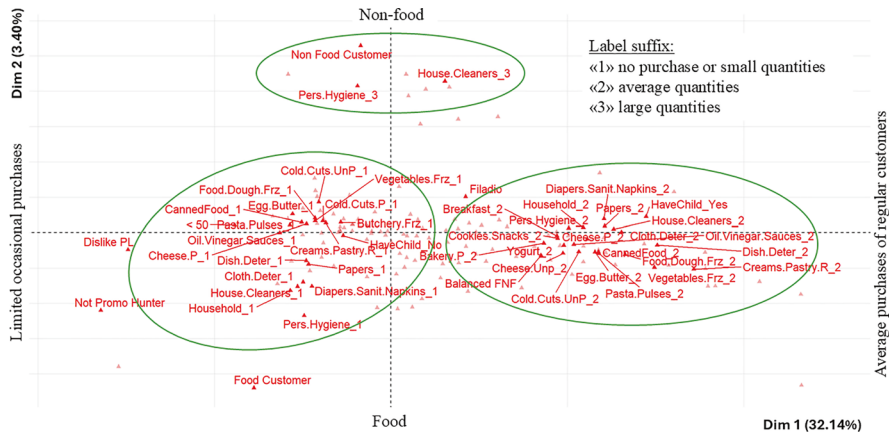


Fig. 2 The factorial plane F1 × F2

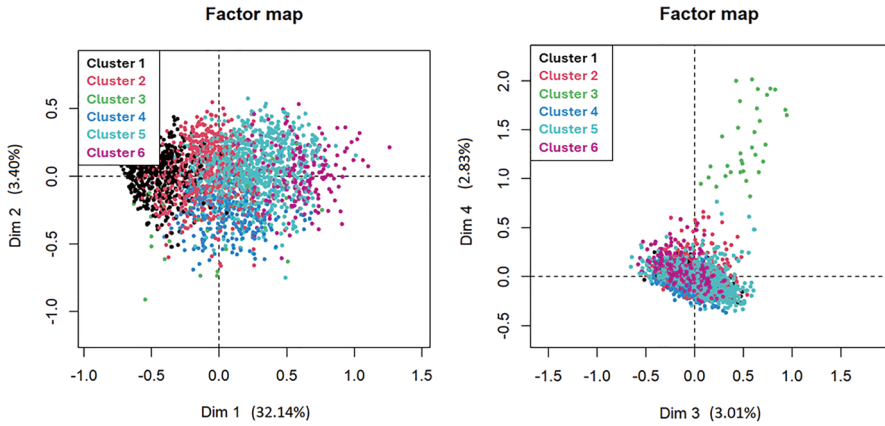


Fig. 3 Projection of clustered individuals on the factorial planes (images 1 and 2)

Table 1 The identified clusters and their interpretation

Cluster	Description	# customers	% customers
Cluster 1	Occasional customers who prefer private-label products and actively seek promotions	14,038	23
Cluster 2	Non-loyal customers who are not interested in private label products	14,880	24
Cluster 3	E-shop customers with moderate spending, who prefer packaged over bulk products	817	1
Cluster 4	Selective customers with moderate spending, who prefer physical stores, buy ready meals, and show little interest in private label products	6,239	10
Cluster 5	Moderate value consumers, private label enthusiasts	20,945	34
Cluster 6	Loyal premium customers with high spending	4,976	8

Following the segmentation phase, a predictive model was developed to automate the classification of individual customers into one of the six identified profiles. A decision tree model was trained using selected behavioral variables as input features and cluster membership as the variable to be predicted. As input features, we kept only the indicators that can be estimated by the short-term purchase history of a customer (e.g., purchasing of Bio or PL products) and are thus expected to be available at the time of applying personalized marketing actions.

In Fig. 4, an extract of the resulting decision tree is shown, where leaf nodes are the predicted profile and the branching nodes are input features. For instance, the highlighted path shows that a customer who is indifferent to private label products, does not buy many bio products, is more promotion-sensitive than average, is a parent, and shops in the morning is predicted to belong to Cluster 6: loyal premium customer with high spending.

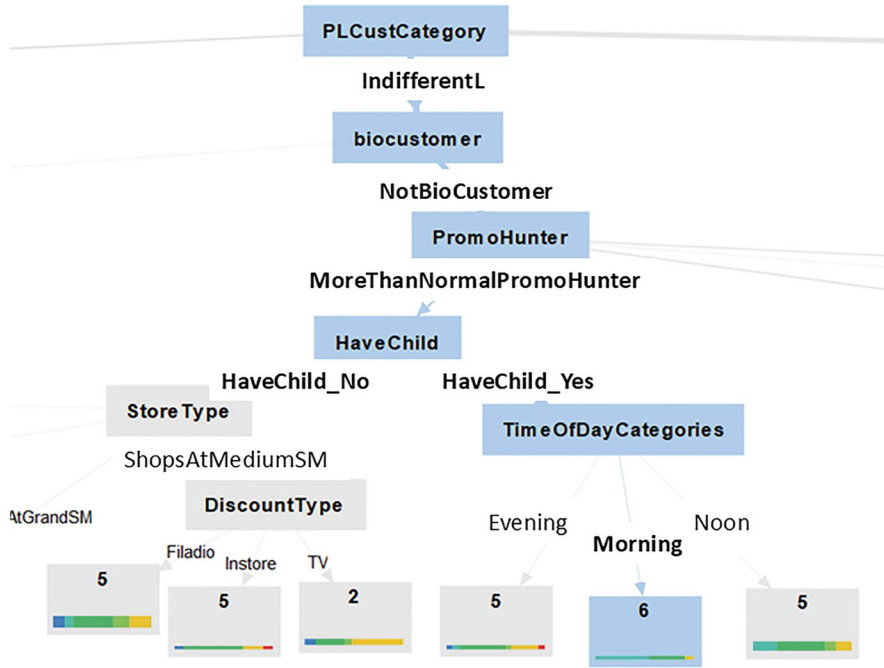


Fig. 4 Extract of the decision tree model

Table 2 Confusion matrix of decision tree classifier

Accuracy: 70.04%							
	True 1	True 6	True 5	True 4	True 2	True 3	Precision (%)
Pred.1	449	3	68	27	94	0	70.05
Pred.6	1	154	26	9	13	2	75.12
Pred.5	48	31	649	45	140	1	71.01
Pred.4	16	7	32	146	29	0	63.48
Pred.2	53	6	70	25	323	0	67.71
Pred.3	0	0	1	0	2	30	90.91
Recall (%)	79.19	76.62	76.71	57.94	53.74	90.91	

The prediction model was evaluated on a random test set of 2500 customers, achieving 70% accuracy. As shown in the confusion matrix (Table 2), it performed particularly well in identifying Cluster 3 (e-shop) and Cluster 6 (loyal premium) customers. However, it showed some difficulty in accurately distinguishing Cluster 2 (non-loyal) and Cluster 4 (moderate selective), often confusing these customers with the profiles of Cluster 1 (occasional), and Cluster 5 (moderate value). The graph also shows that the preference for private label and for biological products, as well as promotion hunting, are important features in predicting a customer’s profile.

5 Discussion and Conclusion

We combined factor and cluster analysis with decision-tree classifiers to generate actionable consumer segments. Moving beyond conventional demographic profiling, we identified six distinct shopper profiles that support more targeted marketing strategies. The integration of clustering and predictive modeling aligns with prior literature [3, 4], while the use of decision trees enhances interpretability and real-time application. Key predictors, such as private label preference, promotional sensitivity, and store format choice, confirm the value of behavior-based profiling, aligning with previous findings [6, 7]. Business implications are substantial: retailers can improve customer loyalty, optimize promotional targeting, and better allocate resources by tailoring strategies to each segment. The methodology is adaptable to various retail settings and highlights the importance of continuous data analysis in an evolving retail landscape.

References

1. Buchdadi AD (2024) Segmenting Walmart customers for personalized marketing strategies using MiniBatchKMeans clustering and decision trees: an analysis of purchasing behavior. *J Digit Mark Digit Curr* 1:204–224. <https://doi.org/10.47738/jdmdc.v1i13.18>
2. Namin A, Dehdashti Y (2019) A “hidden” side of consumer grocery shopping choice. *J Retail Consum Serv* 48:16–27. <https://doi.org/10.1016/j.jretconser.2019.01.009>
3. Abbasimehr H, Sheikh Baghery F (2022) A novel time series clustering method with fine-tuned support vector regression for customer behavior analysis. *Expert Syst Appl* 204:117584. <https://doi.org/10.1016/j.eswa.2022.117584>
4. Peker S, Kocyigit A, Eren PE (2017) LRFMP model for customer segmentation in the grocery retail industry: a case study. *Mark Intell Plan* 35:544–559. <https://doi.org/10.1108/MIP-11-2016-0210>
5. Vanessa N, Japutra A (2017) Contextual marketing based on customer buying pattern in grocery e-commerce: the case of bigbasket.com (India). *ASEAN Mark J* IX:56–67. <https://doi.org/10.21002/amj.v9i1.9286>
6. Conlin R, Labban A (2019) Clustering attitudes and behaviors of high/low involvement grocery shopper. *J Food Prod Mark* 25:647–667. <https://doi.org/10.1080/10454446.2019.1629558>
7. Hallikainen H, Luongo M, Dhir A, Laukkanen T (2022) Consequences of personalized product recommendations and price promotions in online grocery shopping. *J Retail Consum Serv* 69:103088. <https://doi.org/10.1016/j.jretconser.2022.103088>
8. Janssens K, Lambrechts W, Keur H, Semeijn J (2020) Customer value types predicting consumer behavior at Dutch grocery retailers. *Behav Sci* 10:127. <https://doi.org/10.3390/BS10080127>
9. Nguyen SP (2021) Deep customer segmentation with applications to a Vietnamese supermarkets’ data. *Soft Comput* 25:7785–7793. <https://doi.org/10.1007/s00500-021-05796-0>
10. Griva A, Bardaki C, Pramatarı K, Papakiriakopoulos D (2018) Retail business analytics: customer visit segmentation using market basket data. *Expert Syst Appl* 100:1–16. <https://doi.org/10.1016/j.eswa.2018.01.029>
11. Theodoridis PK, Chatzipanagiotou KC (2009) Store image attributes and customer satisfaction across different customer profiles within the supermarket sector in Greece. *Eur J Mark* 43:708–734. <https://doi.org/10.1108/03090560910947016>

12. Matta E, Stalidis G (2024) Profiling online and physical supermarket customers using factor and clustering methods. In: Reis JL, Zelený J, Gavurová B, dos Santos JPM (eds) Marketing and smart technologies, ICMarkTech 2023. Smart innovation, systems and technologies, vol 386. Springer, Singapore. https://doi.org/10.1007/978-981-97-1552-7_15
13. Greenacre M (2013) Contribution biplots. *J Comput Graph Stat* 22:107–122. <https://doi.org/10.1080/10618600.2012.702494>
14. Beh EJ, Lombardo R (2019) Multiple and multiway correspondence analysis. *Wiley Interdiscip Rev Comput Stat* 11:1–11. <https://doi.org/10.1002/wics.1464>
15. Bejaei M, Cliff MA, Singh A (2020) Multiple correspondence and hierarchical cluster analyses for the profiling of fresh apple customers using data from two marketplaces. *Foods* 9:873. <https://doi.org/10.3390/foods9070873>
16. Vrhovac V, Orošnjak M, Ristić K, Sremčev N, Jocanović M, Spajić J, Brkljač N (2024) Unsupervised modelling of e-customers' profiles: multiple correspondence analysis with hierarchical clustering of principal components and machine learning classifiers. *Mathematics* 12:3794. <https://doi.org/10.3390/math12233794>
17. Ernawati E, Baharin SSK, Kasmin F (2021) A review of data mining methods in RFM-based customer segmentation. *J Phys Conf Ser* 1869:012085. <https://doi.org/10.1088/1742-6596/1869/1/012085>
18. Shrivastava A, Jain JK, Chauhan D (2023) Literature review on tools & applications of data mining. *Int J Comput Sci Eng* 11:46–54. <https://doi.org/10.26438/ijcse/v11i4.4654>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

