

Customer Segmentation and Sales Patterns in Supermarket Retailing: A Data-Driven Case Study

Ayşenur ERDİL¹

ARTICLE INFO:

Copyright & License: The authors own the copyright of their work published in the journal and their work is published under the CC BY-NC 4.0 license.

Article Type: Research Article

Article Registration Date: 16.09.2026

Approval Date: 30.09.2026

Citation: : Erdil, A. (2026). Customer Segmentation and Sales Patterns in Supermarket Retailing: A Date-Driven Case Study . *Journal of Current Debates in Social Sciences*, 9(2), 108-115.

DOI: [10.29228/cudes.96629](https://doi.org/10.29228/cudes.96629)



Abstract

Segmenting customers allows businesses to identify separate customer groups based on differences in purchasing behaviour, allowing them to develop more targeted marketing campaigns. This research looks at sales trends and client segmentation using data from a retail firm in Turkey's Marmara Region. The study examines three product categories: fresh fruit, gourmet goods and infant care products, which cover three distinct retail shop configurations. The structural distribution of standardized sales data was investigated using descriptive analytical techniques. Furthermore, the usefulness of customer relationship management (CRM) data from loyalty programs was evaluated. Applying stochastic probability frameworks to different supply chain and operating environments improved this study even further. According to the annual numbers, fresh produce has the highest transaction volume inside the retail company. Meanwhile, gourmet items are becoming increasingly popular in the exclusive store model, catering mostly to people with higher spending power. Predictive statistical frameworks may enhance operational risk management, as demonstrated by the use of binomial and non-binomial distributions in probability modelling. In the long term, these empirical findings imply that client segmentation need to go beyond static descriptive classifications. Segmentation modelling should effectively guide store format decisions, tailored promotional appearances, and specialized positioning of goods in order to increase commercial performance.

Key Words: Customer relationship management; Customer segmentation; Retailing; Sales patterns; Supermarket

¹ Ph.D, Istanbul Medeniyet University, Türkiye, 0000-0002-6413-7482, runesyalidre@gmail.com.

1. INTRODUCTION

Consumer demands, tastes, buying patterns, and responses to marketing initiatives are all varied in retail sectors. By breaking up a large market into more homogeneous groupings, market segmentation emerged as a strategic way to deal with this variety and support management decision-making. Customer relationship management (CRM) and segmentation are strongly related in today's retail industry (Smith, 1956). Businesses can identify differences in client behavior and develop more specialized tactics thanks to the availability of transaction track record and loyalty scheme analytics.

Retailers that manage a variety of shop layouts and categories of products find that customer segmentation is particularly crucial. Customers who purchase fresh fruit, gourmet goods, or baby products may differ in terms of how frequently they shop, how involved they are in particular categories, how much they spend and how they react to sales. Consequently, segmentation facilitates optimal resource distribution, pricing decisions, promotion planning and product positioning. CRM should be seen as an all-encompassing procedure that incorporates information management, generating value, multidimensional integration and performance evaluation, according to Payne and Frow (2005). Accordingly, rather than being a stand-alone marketing tactic, customer segmentation functions as an analytical component of CRM.

Past research has shown that data from loyalty cards can effectively identify meaningful customer segments based on behavior. For instance, a study by Allaway and colleagues in 2006 discovered distinct behavioral segments in a retail loyalty-card system, demonstrating that specific marketing strategies could be crafted from customers' purchasing habits. Likewise, Reutterer and others in 2006 introduced a dynamic segmentation method grounded in retail purchase histories, highlighting the importance of consistent behavior and cross-category purchasing links. Additionally, Meyer-Waarden in 2008 observed that being part of a loyalty program correlates with variations in purchasing behavior among customers. These studies suggest that transactional and loyalty data serve as a valuable empirical basis for making customer-focused decisions in retail.

Through a case-oriented investigation of a retail sales dataset from a company in the Marmara Region of Turkey, this work builds on earlier ideas. Three retail shop concepts and three customer-product groups baby items, gourmet products and fruits and vegetables were examined in the first experiment. The primary goal was to use statistical analysis to identify trends in sales and support targeted marketing initiatives. The study also included probability exercises pertaining to quantity of rotten fruits and vegetables, conditional probabilities and manufacturing failures.

1.1. Subheadings Customer Segmentation

Customer segmentation is the methodical division of a heterogeneous market into groups based on shared traits. In 1956, Smith proposed segmentation as a different approach from seeing the market as homogeneous, arguing that different customer needs might be more effectively satisfied by customized marketing tactics. Behavioral, psychographic, geographic and demographic characteristics are among the segmentation factors. Because behavioral data represents actual purchasing behavior and reveals customer preferences, it is very revealing in the retail setting Smith (1956). Because purchase data may be used to identify groups with different product category preferences and purchasing behaviors, retailers find behavioral segmentation particularly useful. In 2006, research by Reutterer et al. showed how purchase histories may be used to generate permanent behavior-based groups for targeted marketing, while research by Allaway et al. showed how loyalty card data can create unique consumer profiles. The strategic significance of RFM-based and clustering approaches for identifying lucrative and strategically distinct consumer categories is still being highlighted by ongoing research (Doğan et al., 2018; Şentürk & Alp, 2023).

According to recent research, consumer segmentation is changing from static demographic groups to dynamic, data-driven methods that take into account variables including customer value, purchasing behavior, transaction frequency, recent transactions, monetary worth, and changes in customer behavior over time. Techniques like RFM-based segmentation and clustering have grown more popular in retail environments, assisting in the identification of various consumer groups and supporting the creation of customized marketing and CRM strategies (Abbasimehr & Bahrini, 2022; Alves Gomes & Meisen, 2023; Şentürk et al., 2024; Wang et al., 2024; Syahra et al., 2025).

1.2. Customer Relationship Management and Loyalty Programs

CRM is the organizational structure that converts customer data into useful information and management-relevant insights. CRM is a strategic, cross-functional process that encompasses strategy formulation, value generation, multichannel integration, information management and performance evaluation, according to Payne and Frow (2005). Loyalty programs can offer useful client information in this situation. Although the effect and duration of loyalty programs vary among consumer groups, studies indicate that members and non-members of loyalty programs have different purchasing habits (Meyer-Waarden, 2008; Liu, 2007). Therefore, rather than supposing that such programs naturally foster loyalty, loyalty-card data should be utilized as a tool for examining consumer behavior.

In order to develop unique and customized marketing tactics, modern retail CRM increasingly relies on consumer data. According to recent study, customer segmentation may improve focused CRM by grouping clients according to their spending patterns and company value. However, careful consideration of privacy, trust and the customers' desire to disclose their information is also necessary for the efficient use of consumer data (Pallant et al., 2022; Syahra et al., 2025).

2. PURPOSE

The focus of this study is to examine consumer segmentation and sales patterns across multiple product categories and retail shop formats, with the objective of identifying significant disparities in purchasing behaviors and distributions. It also investigates the possible use of loyalty program data and customer relationship management (CRM) information to develop more focused marketing tactics. The investigation aims to show how a combined review of product-level sales trends, store-format variations and customer segmentation may lead to a more organized understanding of retail performance. By studying these variables in combination, this research provides a foundation for identifying categories of products with distinct sales characteristics, recognizing variances among store configurations, and establishing more effective customer-oriented tactics. Furthermore, the research employs descriptive statistics and probability-based analyzes to provide further information about sales behavior, client segmentation and operational decision making.

3. METHOD

3.1. Research Design and Data Structure

Using the statistical project dataset from the original source, the study uses a descriptive case-study methodology. This dataset includes three retail shop types and three customer/product segments: infant items, gourmet products and fruits and vegetables. The retail company is situated in the Marmara Region of Turkey. The resea documentation states that sales data were examined in relation to typical sales numbers for the various shop formats, with loyalty-card data supporting consumer segmentation. The study incorporates probability computations, graphical analysis, consumer segmentation insights and descriptive statistics.

According to the source material, client segmentation was essential for differentiating between groups according to traits like spending habits and purchase inclinations. Families or consumers interested in infant products, those oriented toward gourmet goods and consumers of fruits and vegetables were the three primary categories found. The study also describes how CRM campaigns were implemented, including in-store interactions, social media and mobile app communications, customer-specific promotions, and product group-focused promotions.

3.2. Statistical Analysis

Frequency distributions, relative and cumulative frequency data, histograms, pie charts, dot plots, bar charts and box plots were among the methods used in the descriptive analysis. These techniques assisted in identifying variations in the distribution and intensity of observations among various product/customer categories and retail types. The first research also included key summary metrics

including variance, standard deviation, mean and median. The scenario in which the third production failure occurs during the tenth process was examined using a negative-binomial model for the likelihood analysis. Given that the research required 10 fixed observations with a constant 0.02 probability of coming across a bad kilogram, a binomial model worked well to handle the fresh-produce problem. When there are a certain number of independent trials with two potential outcomes and a constant likelihood of success, the binomial probability function is appropriate (Illowsky & Dean, 2023).

4. RESULTS

4.1. Customer/Product Segment Sales

The study dataset's monthly observations yielded annualized totals that show clear distinctions across the three customer/product groupings and retail shop configurations. With 86,490 consumer observations across all three media, fruits and vegetables stand out as the largest category. Baby items make up 21,241.18 observations, while gourmet products come in second with 32,062.94. The retail company in the Marmara Region of Turkey provides 52,440 observations for the first store format, 50,034.12 for the second format and 37,320 for the third format when looking at the shop format level (see Table 1)

Table 1. Consumer observations by product segment and store format.

Customer/Product Segment	A Company	B Company	C Company	Total
Infant products, Baby-care products	11,370	6,990	2,881.18	21,241.18
Gourmet goods, Specialty foods	10,440	4,440	17,182.94	32,062.94
Fresh produce, Fresh fruits and vegetables	30,630	25,890	29,970	86,490
Total	52,440	37,320	50,034.12	139,794.12

The distribution differs depending on the type of retailer. All three constantly provide a large selection of fruits and vegetables, but a major gourmet food retail chain operating in Turkey is the main location for gourmet goods. Because diverse shop forms can attract customers with varying category preferences and purchasing patterns, this is consistent with the strategic purpose of diversifying store formats. As a result, a consistent marketing approach might not be appropriate for every format.

4.2. Customer Segmentation and CRM Implications

In order to prioritize product development, match items with target audiences, create specialized marketing tactics, and assist with price decisions, the project focuses on consumer segmentation. According to Payne and Frow (2005), these objectives are consistent with CRM literature that emphasizes the significance of integrating customer information for value generation and performance evaluation. Furthermore, as demonstrated by Allaway and colleagues in 2006, using loyalty-card purchase data is consistent with earlier research showing that retail loyalty programs provide behavioral insights useful in distinguishing consumer groups.

The research identifies three key consumer segments that are behaviorally significant. Baby-product customers are best engaged with family and child-centered promotions. Gourmet consumers may react favorably to specialist offers in their category and premium product positioning. In the meanwhile, advertisements emphasizing freshness, regular offers, seasonal availability, and value can be used to connect with consumers of fruits and vegetables. Because it enables a more precise distribution of promotional resources based on the unique purchasing patterns of each group, segment-specific techniques are more successful than a one-size-fits-all strategy.

4.3. Contingency Analysis

The contingent table in the proposal details the combined distribution of shop type and customer/product categories. Company A has the highest percentage of vegetables and fruits (0.22), followed by Company C (0.21) and Company B (0.19). The gourmet category is more concentrated at Company C (0.12), but the baby-related goods category is more evenly spread across all three store layouts.

The undertaking's table displays both client and product categories, as well as the combined shop format distribution. In the fruits and vegetables category, Company A has the greatest percentage (0.22), followed by Company C (0.21) and Company B (0.19). Although the baby-products category is pretty similarly distributed throughout the three store layouts, the gourmet department is more concentrated in Company C (0.12).

Table 2. Joint and marginal probabilities by store format and product category.

Store format	Infant products, Baby-care products	Gourmet goods, Specialty foods	Fresh produce, Fresh fruits and vegetables	Total
A Company	0.07	0.07	0.22	0.36
B Company	0.06	0.04	0.19	0.29
C Company	0.02	0.12	0.21	0.35
Total	0.15	0.23	0.62	1.00

The Table 2 provides three conditional-probability calculations. First, there is a 0.07 chance of selecting an observation from both the gourmet category and the first retail shop type. Second, the probability that an observation selected from the second retail shop format belongs to the vegetables-and-fruits category is 0.655, computed as $0.19/0.29$. Finally, the likelihood that an observation relates to the gourmet category is 0.25, computed as $0.16/0.64$, as it does not originate from the first retail shop type.

4.4. Operational Probability Analysis

In the beginning instance, the probability of experiencing a manufacturing failure is set at 0.01. Production is halted after the third failure. Considering this, the likelihood that the third failure would occur during the tenth manufacturing run may be computed using the negative binomial distribution such as Equation (1):

$$[P(X = 10) = C(9, 2) \times (0.01)^3 \times (0.99)^7 = 0.00003355] \quad (1)$$

The predicted likelihood is around 0.00336%, which is incredibly low. This implies that encountering the third manufacturing failure, especially during the tenth process, is extremely unlikely given the predicted failure rate. Thus, suspending production at this point would be an unusual occurrence if the failure rate remained steady throughout the manufacturing process. In the framework of fresh product, (X) reflects the number of kilograms of ruined produce per ten kilos sold, with a 2% spoiling chance ($p = 0.02$). The binomial distribution is used to estimate the probability of different spoilage outcomes. These probabilities aid in calculating the estimated amount of ruined product in a ten-kilogram batch and show that high spoiling levels would be statistically uncommon given the estimated 2% degrading rate.

Table 3. Binomial probability distribution of product spoilage in fresh produce batches

Question	Probability	Result
At most 6 kg	$P(X \leq 6)$	0.9999999999 $\approx$ 100.00%
Exactly 4 kg	$P(X = 4)$	0.0000297643 $\approx$ 0.00298%
At least 3 kg	$P(X \geq 3)$	0.0008639063 $\approx$ 0.08639%

The findings highlight the managerial benefits of integrating customer segmentation with descriptive and probabilistic analysis (Table 3). The predominance of the vegetables and fruits category in the project's statistics shows that this segment may be a crucial driver of foot traffic across various shop types. Meanwhile, the concentration of gourmet goods at a major gourmet food retail chain operating in Turkey highlights the need of analyzing category success in connection to individual store formats rather than using a companywide aggregate.

Table 4. Descriptive statistics and quantile values for Infant products

Variable	Mean	SE	SD	Variance	Skewness	Kurtosis
Infant products, Baby-care products	590.00	50.15	300.90	90,545.70	0.10	-1.42
Variable	Minimum	Q1	Median	Q3	Maximum	IQR
Baby Products	150.00	270.00	570.00	887.50	1,050.00	617.50
Quantile	Value					
0.050	159.00					
0.950	1,020.00					

Note: SE = Standard Error; SD = Standard Deviation; Q1 = First Quartile (25th percentile); Q3= Third Quartile (75th percentile); QR = Interquartile Range.

Table 4 presents the descriptive statistics and percentile distribution for the Infant products dataset. The mean (590.00) and median (570.00) values are close together and the skewness coefficient (0.10) is low, indicating a fairly symmetric distribution. The dataset's variability is represented by a standard deviation of 300.90, a variance of 90,545.70 and an interquartile range (IQR) of 617.50, with minimum and maximum values of 150.00 and 1,050.00, respectively. Furthermore, the 5th (159.00) and 95th (1,020.00) quantiles establish the distribution's fundamental operational bounds while mitigating the impact of extreme outliers.

This observation aligns with existing segmentation literature. Market segmentation proves valuable when varied customer needs can be addressed through targeted managerial actions, as noted by Smith in 1956. Studies on retail loyalty cards also indicate that behavioral data can unveil distinct customer profiles, thus enabling tailored strategies (Allaway et al., 2006). Doğan et al. (2018) found that understanding consumer variety requires more than just expenditure, since frequency and recency also provide important information. As a result, this research paves the way for more complex segmentation algorithms.

There are also substantial consequences for CRM. Customer data should serve more than just reporting purposes; it must inform decisions on product selection, promotions, communication, and customer retention efforts. Payne and Frow (2005) underline that CRM is most effective when information management is tied to value generation and performance evaluation. The suggestions of this issue for regional campaigns, targeted promotions, social media strategies, mobile apps, and in-store communication align with a differentiated CRM approach. From an operational standpoint, probability analysis supplements customer segmentation by assisting in the evaluation of operational uncertainties while emphasizing demand side distinctions. By combining these techniques, merchants may create a comprehensive decision-support system that leverages quantitative information to balance marketing goals against operational risks.

4. CONCLUSION

A retail sales and customer-segmentation effort are reorganized into a well-structured empirical essay in this study. Significant discrepancies between infant products, gourmet goods, and fruits and vegetables, as well as variances across store formats of a retail business operating in the Marmara Region of Türkiye, are shown by the investigation. The dataset's largest portion is made up of fruits and vegetables, and the retail business operating in the Marmara Region of Türkiye has a noticeable concentration of gourmet goods. These results support the use of different consumer and store-format methods rather than standard promotional guidelines. Additionally, the study emphasizes how probability analysis and descriptive statistics might improve consumer segmentation in retail decision-making. Diverse insights for administrative choices are offered by tools including frequency distributions, graphical analysis, conditional probabilities, binomial computations, and negative-binomial reasoning. Building on this basis, future studies should investigate transaction-level behavioral segmentation and carry out long-term assessments of campaign efficacy.

The variances noted might provide useful information for decisions about product diversity, promotional activities, store-targeted initiatives, and marketing budget allocation. Probability analyzes improve this evaluation by demonstrating how uncertainty and operational risk may be included into retail decisions. The negative-binomial calculation estimates the likelihood of experiencing a specific number of failures within a production or operational cycle, whereas the binomial analysis anticipates the probability of varied levels of product loss. Although these analyzes do not directly monitor consumer behavior, they demonstrate how probability-based reasoning may help managers assess operational occurrences and associated dangers. When paired with sales distributions and customer segmentation, these statistical tools give a more complete picture of retail performance.

This research has a few drawbacks. First, the study is dependent on the project dataset provided in the source material and the underlying sampling and data-generation techniques are not completely explained. Second, the study is essentially descriptive and does not quantify the causal impacts of CRM campaigns or loyalty program membership. Third, the three customer/product segments were predetermined; so, the study does not investigate whether these categories arise statistically via an unsupervised clustering approach. Fourth, because the research includes historical store-format data, the results should not be construed as proof of present retail success. Future study might include bigger transaction-level datasets, RFM analysis, K-means or hierarchical clustering, estimating customer lifetime value, and evaluating longitudinal campaigns. The findings suggest that consumer segmentation should be evaluated with variations in product attributes and shop layouts rather than as a single classification assignment. The importance of fruits and vegetables in overall sales statistics, the distinct position of gourmet items in the major gourmet food retail chain operating in Turkey structure and the comparatively low sales volume of baby products all indicate that customer needs and purchasing patterns can change dramatically among retail environments. As a result, managers may find it helpful to create strategies adapted to certain categories and stores rather of depending entirely on conventional advertising approaches. Future study might build on this methodology by include transaction-level customer data, accounting for characteristics such as purchase frequency, recency, monetary value and basket composition, and investigating how consumer categories change over time. longitudinal investigations of promotional programs may also provide further information on whether segment-based methods lead to long-term increases in sales and customer involvement.

This study has a lot of limitations. To begin, the analysis is based on the project dataset provided in the source material, although the sampling and data production processes are not completely stated. Second, the study is mostly descriptive and does not examine the causal impact of CRM activities or participation in loyalty programs. Third, the three customer/product categories were chosen, therefore the study does not look at whether these groups would develop spontaneously using an unsupervised clustering approach. Finally, because the study relied on historical store-format data, the findings should not be taken as indicative of current retail performance. Future research could gain advantages from using larger transaction-level data sets, including RFM (Recency, Frequency, Monetary) analysis, K-means or hierarchical clustering, estimating customer lifetime value and evaluating advertising data that is longitudinal.

REFERENCES

- Akgöz, E. ve Gürsoy, Y. (2014). Turizm Eğitiminde Yabancı Dil Öğrenme, İstek ve Kararlılıkları: Selçuk Üniversitesi Beyşehir Örneği. *Journal of Tourism and Gastronomy Studies*, 2(1), 21-29.
- Abbasimehr, H., & Bahrini, A. (2022). An analytical framework based on the recency, frequency, and monetary model and time series clustering techniques for dynamic segmentation. *Expert Systems with Applications*, 192, 116373. <https://doi.org/10.1016/j.eswa.2021.116373>.
- Allaway, A. W., Gooner, R. A., Berkowitz, D., & Davis, L. (2006). Deriving and exploring behavior segments within a retail loyalty card program. *European Journal of Marketing*, 40(11/12), 1317-1339. <https://doi.org/10.1108/03090560610702830>.
- Alves Gomes, M., & Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Information Systems and e-Business Management*, 21, 527-570. <https://doi.org/10.1007/s10257-023-00640-4>.
- Doğan, O., Ayçin, E., & Bulut, Z. A. (2018). Customer segmentation by using RFM model and clustering methods: A case study in retail industry. *International Journal of Contemporary Economics and Administrative Sciences*, 8(1), 1-19. Retrieved from <https://www.ijceas.com/ijceas/article/view/174>.
- Illowsky, B., & Dean, S. (2023). *Introductory Business Statistics 2e*. OpenStax. <https://openstax.org/books/introductory-business-statistics-2e/>
- Kumar, N. (2025). Intelligent customer segmentation: Unveiling consumer patterns with machine learning. *Journal of Umm Al-Qura University for Engineering and Architecture*, 16, 774-783. <https://doi.org/10.1007/s43995-025-00180-7>.
- Liu, Y. (2007). The long-term impact of loyalty programs on consumer purchase behavior and loyalty. *Journal of Marketing*, 71(4), 19-35. <https://doi.org/10.1509/jmkg.71.4.019>.
- Meyer-Waarden, L. (2008). The influence of loyalty programme membership on customer purchase behaviour. *European Journal of Marketing*, 42(1/2), 87-114. <https://doi.org/10.1108/03090560810840925>.
- Pai, P.-Y., Lin, S.-W., & Lu, W.-M. (2026). Integration of association rule mining and RFM analysis with machine learning for e-commerce customer value segmentation: A sustainable retail perspective. *Quality & Quantity*, 60, 87-125. <https://doi.org/10.1007/s11135-025-02259-8>.
- Pallant, J. I., Pallant, J. L., Sands, S. J., Ferraro, C. R., & Afifi, E. (2022). When and how consumers are willing to exchange data with retailers: An exploratory segmentation. *Journal of Retailing and Consumer Services*, 64, 102774. <https://doi.org/10.1016/j.jretconser.2021.102774>.
- Payne, A., & Frow, P. (2005). A strategic framework for customer relationship management. *Journal of Marketing*, 69(4), 167-176. <https://doi.org/10.1509/jmkg.2005.69.4.167>.
- Reutterer, T., Mild, A., Natter, M., & Taudes, A. (2006). A dynamic segmentation approach for targeting and customizing direct marketing campaigns. *Journal of Interactive Marketing*, 20(3-4), 43-57. <https://doi.org/10.1002/dir.20066>.
- Şentürk, H., & Alp, S. (2023). Customer segmentation with RFM analysis in the retail industry. *Turkish Journal of Entrepreneurship*, 7(13).
- Şentürk, H., Geçici, E., & Alp, S. (2024). Customer segmentation with clustering methods in the retail industry. *İstanbul Aydın Üniversitesi Sosyal Bilimler Dergisi*, 16(4), 551-573.
- Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing*, 21(1), 3-8. <https://doi.org/10.1177/002224295602100102>.
- Syakra, Y., Fadlil, A., & Yuliansyah, H. (2025). Customer segmentation using RFM and K-Means clustering to support CRM in retail industry. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 9(3), 1120-1131. <https://doi.org/10.33395/sinkron.v9i3.14907>.
- Wang, S., Sun, L., & Yu, Y. (2024). A dynamic customer segmentation approach by combining LRFMS and multivariate time series clustering. *Scientific Reports*, 14, 17491. <https://doi.org/10.1038/s41598-024-68621-2>.