# Customer Segmentation in FMCG Sector with RFM Method and Digital Marketing Recommendations in E-Commerce

Ufuk CEBECİ[1], Feyzullah EĞERCİ[2] & Buse ALTAŞ[3]

**Abstract**
E-commerce volume in Turkey has increased significantly with the Covid-19 pandemic. In parallel, the number of sellers on e-commerce platforms increased approximately 8 times in a 3-year period. In this period of increasing competition, segmenting customers in detail is of great importance in order to manage customer relations effectively. Customers have different needs, preferences and behaviors; therefore, the same standard approach to every customer will not be enough to optimize customer satisfaction and sales performance. Companies have begun to attach great importance to data analysis in order to better manage customer relations. There are different method studies for customer segmentation in the literature. In this context, different segmentation methods were applied in the study and their results were compared. After the segmentation process was completed, appropriate digital marketing strategies were determined for each customer cluster in the selected method.

## 1. Introduction

In this study, methods that will increase the turnover of sellers in the e-commerce sector, whose numbers have increased significantly in recent years in Turkey, will be discussed. Customer relationship management (CRM) is important for every company today. Every customer has unique needs, preferences and behaviors; Therefore, the same standard approach to every seller will not be sufficient to optimize customer satisfaction and sales performance. Data analysis methods are used to classify sellers. RFM, one of the most popular of these methods, will be used in the study. To improve the RFM analysis, the RFMV method will be applied by adding the product variety variable (V) and will be compared with the classical method. In traditional RFM calculations, all customers of a business are evaluated with the same scales. Therefore, one of the most important points that are inadequate in the traditional RFM approach today is the segmentation of

---

[1] ORCID: 0000-0003-4367-6206. Istanbul Technical University, Industrial Engineering, Istanbul, Turkey. cebeciu@itu.edu.tr

[2] Corresponding Author. ORCID: 0009-0000-9954-9535. Istanbul Technical University, Industrial Engineering, Istanbul, Turkey. egerci19@itu.edu.tr

[3] ORCID: 0009-0005-6421-7465. Istanbul Technical University, Industrial Engineering, Istanbul, Turkey. altasb19@itu.edu.tr

customers according to RFM results. However, segmentation should be done first and then the RFM method should be applied for each segment. K-means and DBSCAN Algorithms will be used for segmentation in the study. The most successful method will be determined by comparing all results. The classes of the most successful method will be analyzed and marketing recommendations will be developed according to these analyses.

## 2. System Analysis

**Figure 1.** E-Commerce Volume in Turkey by Years (Billion TL).
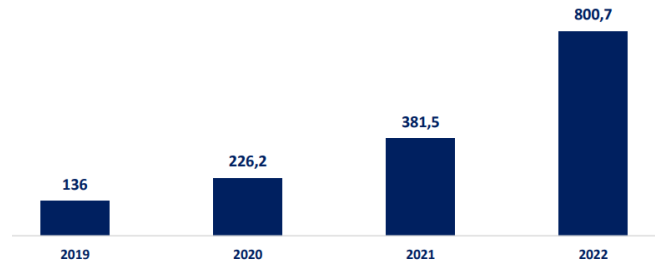


According to Figure 1, E-commerce volume in Turkey has increased significantly in recent years. In addition to the volume increase, according to ETBİS data, the number of sellers, which was 68,457 in 2019, increased to 548,688 in 2022. In other words, the number of users on e-commerce platforms has increased approximately 8 times in the last 3 years. Although the increase in e-commerce volume seems to continue, new sellers need to be careful in order to have a sustainable trade in the market. The level of competition increases in parallel with the number of sellers. Competition includes price wars and efforts to attract customers. The large number of sellers makes it difficult to create a sense of loyalty in customers. In addition, the large number of sellers increases customers' fears of fraud and fraud. Gaining customers' trust has become more difficult than ever. New sellers may face difficulties in promoting and branding themselves in the market. In an environment where competition is intense, it is important to create effective marketing strategies.

Despite these difficulties, e-commerce also has advantages over physical stores. E-commerce offers the opportunity to reach a wide customer base because it is not subject to geographical limitations. Anyone with internet access can become a potential customer. Compared to physical stores, e-commerce operating costs are generally lower. Costs such as store rent, staff salaries and energy expenses may decrease. E-commerce may be open 24 hours a day, every day. This gives customers the flexibility to shop whenever they want and eliminates time constraints. E-commerce platforms allow customers to implement tailored marketing strategies and provide personalized recommendations. E-commerce offers the opportunity to make detailed analyzes on customer behavior. This helps businesses understand shopping habits and preferences. It is possible to gain a place in the growing e-commerce market with an effective business strategy, target customer analysis and correct marketing strategies.

## 3. Literature Review

According to the research conducted by Gomes and Meisen, there have been a total of 105 articles published in the field of e-commerce since the year 2000 that focus on customer segmentation and personalized customer marketing. Out of these articles (Figure 2), 47.6% of them involved manually selecting customer attributes, while the remaining 52.4% utilized feature selection methods. Among these methods, RFM analysis was found to be the most commonly used in the literature. (2023)

**Figure 2.** Usage prevalence of attribute selection methods in the literature from Gomes and Meisen
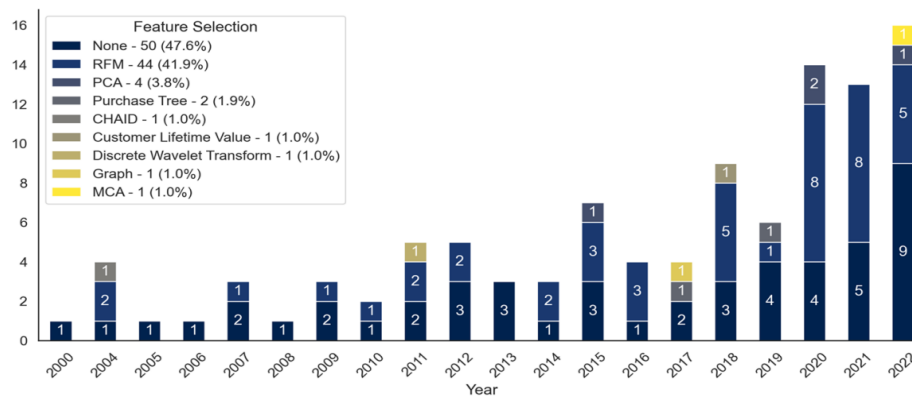


According to Zhang and Huang, the conventional RFM analysis is deemed inadequate for fully comprehending and evaluating customer behavior due to its reliance on just three elements. These features alone may pose limitations in providing a complete picture of customer relationships and shopping habits. Other important factors such as demographic characteristics, behavioral trends or product preferences are excluded from the RFM analysis. (Zhang and Huang ,2022)

For this reason, there are many different studies in the literature that go beyond RFM and combine machine learning and different data points to create a more detailed and effective customer segmentation and marketing strategies. Martinez et al. showed that a much clearer and easier segmentation was achieved using the Fuzzy Linguistic RFM model. It exemplifies the potential of this approach in customer segmentation by integrating 2-tuple and RFM analysis (2019). As another application example, Smaili and Hachimi incorporated the 'D' (diversity) factor into the RFM model, indicating the quantity of different products bought by the customer during the evaluated timeframe, and serving as a gauge of the customer's inclination to experiment with new products (2023).

## 4. Methodology

This study was prepared using real data from an FMCG company based in Turkey and actively operating in many countries in the field of personal care, personal cleansing, and hygiene products. The company sells its products through various retail markets and also offers them to consumers through online shopping channels.

The data received from the company to be used in our RFM analysis covers only sales made through e-commerce channels and includes basic sales information such as which customer purchased which product, in how many units, on what date, and at what price. It contains the company's e-sales information covering approximately two years, starting from September 24, 2019, until August 15, 2021. This information is very valuable for understanding and segmenting customer behavior. To protect data confidentiality and customer information, all data is masked and shared with unique codes. In this way, the information necessary to analyze the data will be provided while protecting customer confidentiality.

Our aim in this study is to perform clustering with the K-mean and DBSCAN algorithm using the company's sales data from previous years, and then perform RFM and RFMV analyses.

In the later stages of the project, it will be investigated how effective the RFMV method offers customer segmentation and will be compared with the traditional RFM method. Customer groups belonging to the created segments will be analyzed and the results of the analysis will be used in campaigns and marketing activities. In this way, it is aimed to help the company manage customer relations more effectively and make its marketing strategies more target-oriented by contributing to its CRM activities.

A diagram showing the planned methodology and summarizing the progress of our research is attached below.

**Figure 3.** Planned Methodology and Summarizing the Progress
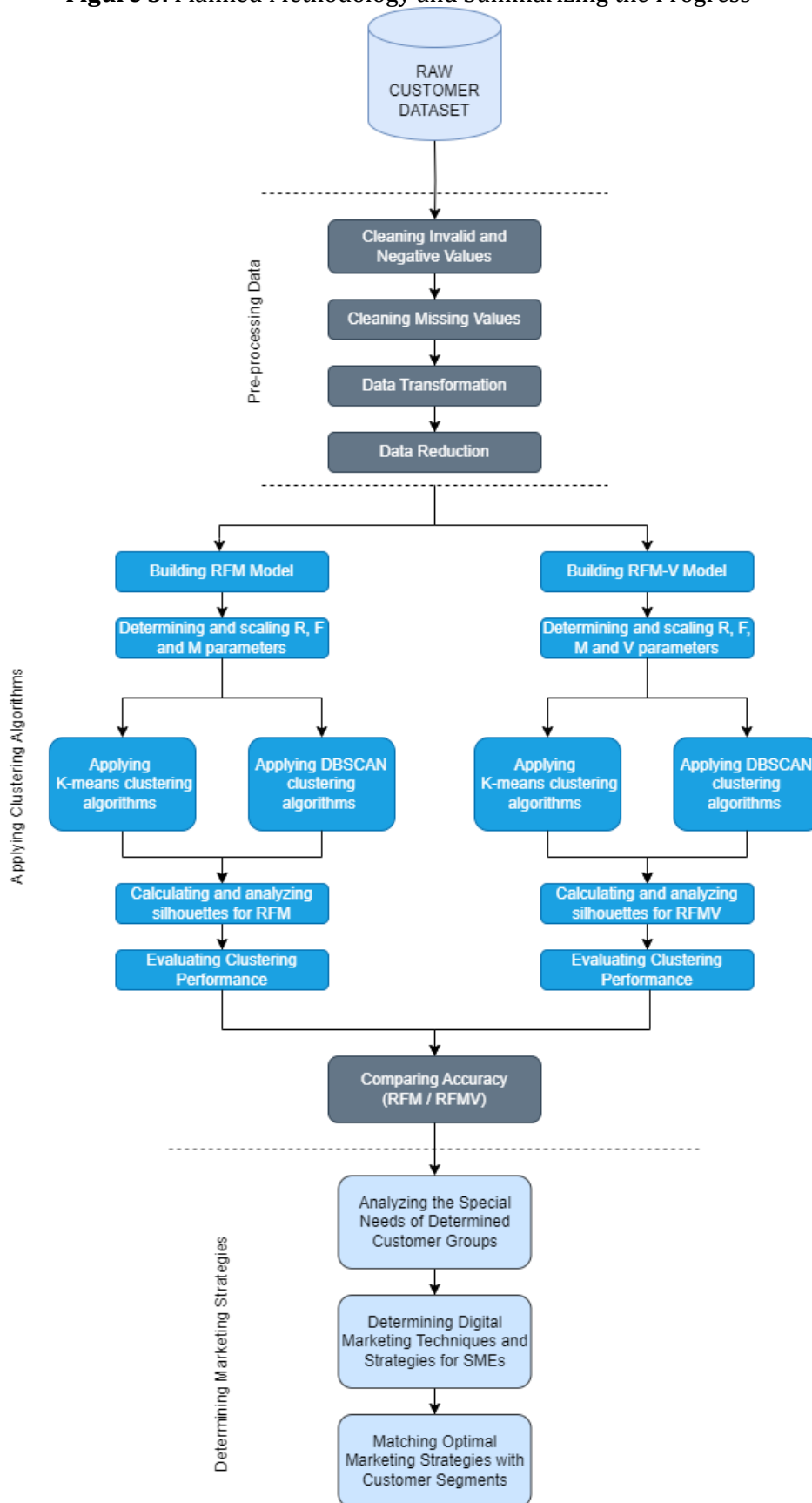
## 4.1. RFM Analysis

RFM (Recency, Frequency, Monetary) Analysis is a technique used for customer segmentation and direct marketing. RFM measures when (Recency), how often (Frequency) and how much money (Monetary) a customer spends. These three metrics are used together to determine how valuable a customer is to a business. RFM analysis ranks customers based on these three metrics and evaluates each with a specific score. These scores are then combined to create a customer review score. This score is used to segment customers and apply customized marketing strategies to each segment. In summary, RFM is a model that uses three features to separate important customers from large amounts of data (Sarvari, 2016).

## 4.2. RFMV Analysis

Adding the product variety (V) variable in addition to the three parameters can provide a more comprehensive and successful analysis for the company because the company has a total of 297 different products under three categories. According to Moghaddam et al., diversity in individual consumption may lead to increased demand for product diversity; Therefore, product assortment can be used along with other behavioral variables to better understand and categorize customer behavior (2016). In this study, it is aimed to cluster customers by adding the product variety variable (V), based on RFMV an improved RFM model.

At the end of our study, a comparative analysis test will be performed on the results of RFM and RFMV analyses.

In traditional RFM calculations, all customers of an organization are evaluated with the same scales, and information about which customers are valuable can be found as a result of the analysis. For this reason, one of the most important points that are inadequate in the traditional RFM approach is seen as segmenting customers according to RFM results. To solve this problem, segmentation should be done first and then the RFM method should be applied for each segment. As an example, in a study by Gustriansyah et al. (2020), an RFM analysis was combined with the application of k-Means with three clusters.

In our study, K-means and DBSCAN Algorithms will be used for clustering.

## 4.3. K-Means

The k-means algorithm is a common clustering method and requires at least two parameters. These are the number of clusters (k) and the distance measurement method. The algorithm calculates the distances of data points to cluster centers and determines clusters. This process is repeated until one of the specified stopping criteria is met. Stopping criteria are usually no change in cluster centers or reaching a certain number of iterations. Although K-means is a fast and effective method, factors such as random selection of starting centers and correct determination of the number of clusters can affect the results. Therefore, it is necessary to conduct more than one study and select the best result (James et al., 2013). The objective of K-Means (eq.4.3) clustering is to minimize total intra-cluster variance, or, the squared error function:

$$E = \sum_{\Sigma=1}^{k} \sum_{p \in C}^{c} \mathrm{d}(M_i, x)^2$$

**4.4 DBSCAN**

The DBSCAN algorithm is based on revealing the neighborhoods of data points with each other in two or multi-dimensional space. Since the database handles data from a spatial perspective, it is mostly used in the analysis of spatial data (Martin et all, 1998). Unlike K-means, the DBSCAN algorithm does not include every element of the database in a cluster and can filter exception data.

In order to detect possible contrary structures in our data set, the DBSCAN method, which produces very successful results especially in spatial applications, was also included in our study. DBSCAN is a non-parametric, density-based clustering algorithm that evaluates extreme values internally through the design of the algorithm and is a robust method against these values. DBSCAN, which allows complex structures to be separated from each other, provides solutions according to the parameters named "minPts" and "eps" determined by the user. (Monalisa et all. 2019) The total number of clusters and outliers in the data set are determined (indirectly) by these two parameters entered by the user.

**5. Application**

**5.1. Company Profile**

This project was implemented in cooperation with the e-commerce department of a leading FMCG company based in Turkey and operating internationally in the personal care, cleaning, and hygiene products sector. The company, which has strengthened its leading position in the sector with its innovative and sustainable production approach, also operates effectively in the international market.

**E-Commerce Department**

The company's e-commerce department carries out all activities related to the sales process of its products through online shopping channels. This department was established in order to quickly respond to the changing habits and expectations of customers in the digitalizing world order. It provided the company with the opportunity to reach a wide customer base through online sales platforms.

**Contributions to the Project**

The company hosted customer segmentation and digital marketing optimization studies and provided all necessary data sets within the scope of the project. Due to the great importance attached to data privacy, all customer data is processed in a masked and anonymized manner. It ensures that customer privacy is protected and that data analysis processes are carried out under ethical standards. The results of the project will be presented to the company's CRM and Marketing departments and will be used directly to manage customer relations more effectively and make marketing strategies more target-oriented. This collaboration will strengthen data-based decision-making processes in the company's marketing activities and increase its competitive advantage.

## 5.2. Preprocessing of Raw Customer Dataset

The implementation process started with raw customer data preprocessing. At this stage, invalid and lost data will be cleaned, and data transformation and reduction will be performed.

The company's raw sales data shared with us between September 24, 2019 and August 15, 2021 consists of 332728 rows and 31 columns.The dataset in .csv format was transferred to the Jupyter Notebook working environment with the Pandas library. When the data was examined, it was seen that 53786 different customers placed 112379 orders between the specified dates.

In order to create more effective and meaningful clusters with the data set, customers who shopped only once or twice within a two-year period were removed from the data set. The number of customers who shopped only once during this period is 2,324, and the number of customers who shopped only twice is 1,451. After removing these customers, the number of customers for whom RFM data was created and included in the cluster analysis decreased to 50011. All analyses carried out in this study were made on these 50011 unique customers. This process enabled the data set to be reduced to a more manageable size and cluster analysis to produce more meaningful and interpretable results.

## 5.3. Applying Clustering Algorithms

After Preproceesing phase, the project was carried out through 4 different alternative scenarios in the clustering algorithms application phase. In the first scenario, the RFM method and K-Means algorithm were used. On the other hand, in the second scenario, the RFM method was applied with the DBSCAN algorithm. The difference between these two scenarios and the traditional RFM method is that the RFM method is applied after the clusters are determined as a result of the clustering algorithms. In the 3rd scenario, the V parameter was added to the RFM method as an extra parameter and clusters were created again with the K-Means algorithm. In the 4th and last scenario, the RFMV method was applied together with the DBSCAN algorithm. All outputs were compared and which algorithm gave more accurate results and the contributions of the added parameter were analyzed.
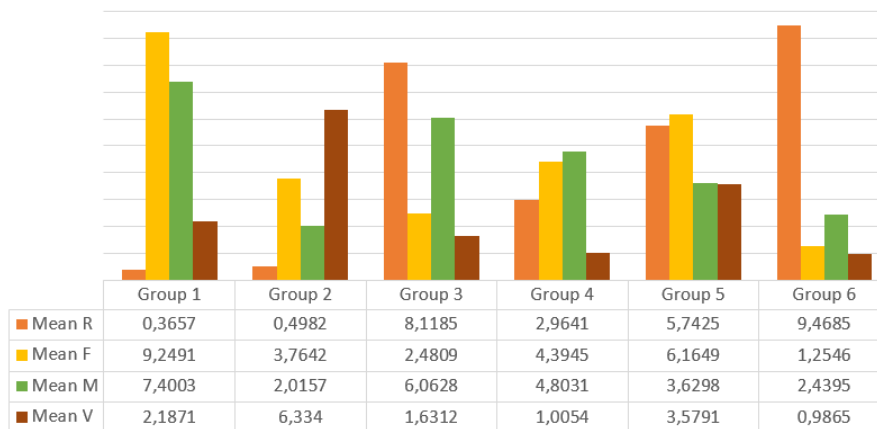
### 5.3.1 Silhouette Index Results

As a result of all these analyses, it was concluded that applying the RFMV Method and DBSCAN Algorithm together for the data set gave better results than other scenarios.

**Figure 5.** Silhouette Index Results

| Algorithm | Silhouette Index |
|---|---|
| Scenario 1 (RFM Method and K-Means) | 0,395 |
| Scenario 2 (RFM Method and DBSCAN) | 0,607 |
| Scenario 3 (RFMV Method and K-Means) | 0,548 |
| Scenario 4 (RFMV Method and DBSCAN) | 0,821 |

In the following stages of the study, customer segments were created according to the results of the RFMV and DBSCAN algorithms. According to the findings, it was concluded that dividing the customers into 6 segments would be sufficient and ideal.

**Figure 6.** R, F, M and V values of Customer Segments



CUSTOMER SEGMENTS

| | Group 1 | Group 2 | Group 3 | Group 4 | Group 5 | Group 6 |
|---|---|---|---|---|---|---|
| Mean R | 0,3657 | 0,4982 | 8,1185 | 2,9641 | 5,7425 | 9,4685 |
| Mean F | 9,2491 | 3,7642 | 2,4809 | 4,3945 | 6,1649 | 1,2546 |
| Mean M | 7,4003 | 2,0157 | 6,0628 | 4,8031 | 3,6298 | 2,4395 |
| Mean V | 2,1871 | 6,334 | 1,6312 | 1,0054 | 3,5791 | 0,9865 |

### 5.3.2 Customer Segmentation Findings

After applying the model, customer segments with different characteristics were obtained. In this way, customers were better understood and an idea was gained about what kind of approach should be taken to all customer groups and what marketing activities should be offered.

According to the results of the analysis, Group 1 constitutes the best customers. This group consists of 9852 people, that is, 19.7% of all customers. It has the highest Frequency and Monetary values compared to other customers. Recency values also showed that they have current shopping habits. However, it was determined that the Variety parameter of this group was at a low level. Therefore, directing customers in this group to other products can create an important source of income for the company. Specially defined discount coupons and different product advertisements should be offered to these customers through the

application and website. It is also recommended to give reminders and campaign news via personalized e-mails.

When Group 2 was examined, it was concluded that Frequency and Monetary values were low. However, the low Recency value of the group means that the last shopping dates are not far away. This shows that this group consists of new customers. Additionally, the group's high Variety value shows they are prone to purchasing different products. All these reasons mean that these customers show the potential to become future loyal customers of the company. Regular cart reminder push notifications should be offered. Additionally, personally defined discount coupons will be effective for these customers.

When the values of Group 3 were analyzed, it was concluded that this group was the old high-income group, but they were about to be lost. The group consists of 3801 customers with 7.6% of all customers. This group has high Monetary values. However, the same is not true for Frequency and Recency values. The aim should be to win back these customers. Therefore, reminder e-mails and application notifications should be provided regularly. Customers should be captured by sending advertisements for different product groups. SMS, MMS, and in-app marketing are recommended.

Group 4 consists of 24.3% of all customers, that is, 12153 people. This customer group is considered average customers. Frequency, Monetary and Recency values of the group are at sufficient scores. These customers should not be reduced to lower categories or lost. Because the largest number of people is here compared to other groups. Therefore, automatic reminder e-mails should be sent regularly. Additionally, offering in-app marketing ads and discount coupons will be beneficial to retain this group.

According to the analysis results, Group 5 consists of 6401 people. Although the Frequency value of this group is high, it has been observed that the Monetary value is slightly lower. For this reason, customers in this group are considered as customers who are prone to shopping at affordable prices. They tend to buy more affordable products and marketing activities should be determined considering this condition. It is recommended to send reminder e-mails and notifications to this group for affordable product groups. Additionally, discount coupons will be important to keep this customer group active.

Group 6 consists of 9602 people with 19.2%. The Recency value of this group is very high, meaning these customers have not been shopping for a long time. Additionally, Frequency, Monetary and Variety values are low. This group is considered as inactive customers. It is suggested that efforts should be made to regain this group because they consist of a large number of people. E-mail and SMS reminders and promotional notifications should be presented to this group regularly and efforts should be made to regain them.

**Figure 7.** Marketing Recommendations for Identified Customer Segments (Summary Table)

| Customer Groups | Mean R | Mean F | Mean M | Mean V | Marketing Recommendations |
|---|---|---|---|---|---|
| Group 1 | 0,3657 | 9,2491 | 7,4003 | 2,1871 | Best customers group (good level of R,F,M value)<br>Product variety is low (V parameter is low)<br>Personalized Email marketing<br>Specially defined discount coupons<br>Intended to direct to other product groups |
| Group 2 | 0,4982 | 3,7642 | 2,0157 | 6,334 | New customer category (low F, M values)<br>R value is good<br>Tendency to buy various products (high V value)<br>Cart reminder push notifications<br>Personally defined discount coupons<br>Aiming to attract loyal customers and not lose them |
| Group 3 | 8,1185 | 2,4809 | 6,0628 | 1,6312 | Former high income group<br>Generated a good level of income (high M value)<br>Haven't shopped in a long time (R value is not good)<br>Reminder e-mail<br>SMS, MMS and in-app marketing<br>The aim is to win back these customers |
| Group 4 | 2,9641 | 4,3945 | 4,8031 | 1,0054 | Average customers<br>Fve M values are average<br>R value is sufficient<br>In-app marketing<br>Automated Targeted Emails<br>It should not be relegated to lower categories. |
| Group 5 | 5,7425 | 6,1649 | 3,6298 | 3,5791 | Affordable shopping customers<br>Frequency of shopping is high (F value is high)<br>They tend to buy cheaper products (M is the average value)<br>push notifications<br>In-app intrusive ads<br>Discounts and deals should be offered for various product ranges |
| Group 6 | 9,4685 | 1,2546 | 2,4395 | 0,9865 | Inactive Customers<br>No regular shopping habits (low R,F,M value)<br>E-mail, SMS ads<br>Push notifications<br>Group that needs attention because it is large in number |

## 6. Conclusion

The study emphasizes the importance of using data analysis and segmentation methods in order to increase the competitive advantages of sellers in the e-commerce sector in Turkey. The study, conducted using the RFMV method to develop RFM analysis, aims to evaluate customers not only by their purchasing frequency, time and expenditure, but also by product variety. Since the traditional RFM approach does not make customer segmentation specific to data sets, segmentation using K-means and DBSCAN algorithms helped determine the most effective method. The results of the study, together with analyzing the classes of the most successful method determined and developing marketing recommendations based on these analyses, can contribute to sellers achieving sustainable success in the market by increasing customer satisfaction. The study supports sellers operating in e-commerce to adopt customer-oriented strategies and increase their competitive advantages. In addition, it was designed to help sellers adapt more effectively to the dynamics in the e-commerce sector.

# References

Almukhtar, F., Mahmoodd, N., & Kareem, S. (2021). Search engine optimization: a review. *Applied Computer Science*, *17*(1), 70-80.

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*, 1-44.

Arslan İ. (2019). Python ile Veri Bilimi (4. bs). Pusula.

Atar, R., Jadhav, B. T., & Wagh, J. A. (2021). An Overview of Digital Marketing. *Available at SSRN 3980984*.

Balaji, S., & Srivatsa, S. K. (2012). Customer segmentation for decision support using clustering and association rule based approaches. *International Journal of Computer Science & Engineering Technology*, *3*(11), 525-529.

Baykasoğlu, A., Gölcük, İ., & Özsoydan, F. (2018). Improving fuzzy c-means clustering via quantumenhanced weighted superposition attraction algorithm. Hacettepe Journal of Mathematics and Statistics, 48(3), 859-882. https://doi.org/10.15672/hjms.2019.657.

Belas, J., Gavurova, B., Novotna, L., & Smrčka, L. (2022). Examination of differences in using marketing tools in the management of SMEs in the Visegrad Group countries. *Amfiteatru Economic*.

Berman, R., & Katona, Z. (2013). The role of search engine optimization in search marketing. *Marketing Science*, *32*(4), 644-651.

Brahmana, R. S., Mohammed, F. A., & Chairuang, K. (2020). Customer segmentation based on RFM model using K-means, K-medoids, and DBSCAN methods. *Lontar Komput. J. Ilm. Teknol. Inf*, *11*(1), 32.

Bramer, M. (2007). Clustering. *Principles of Data Mining*, 221-238.

Chaffey, D. (2013). Definitions of E-marketing vs Internet vs Digital marketing, Smart Insight Blog.

Chen, X., Guo, C., Fang, Y., & Mao, R. (2019). Structured Spectral Clustering of PurTree Data. In *Database Systems for Advanced Applications: 24th International Conference, DASFAA 2019, Chiang Mai, Thailand, April 22–25, 2019, Proceedings, Part II 24* (pp. 485-501). Springer International Publishing.

Cui, M. (2020). Introduction to the k-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance*, *1*(1), 5-8.

Deng, Y., & Gao, Q. (2020). A study on e-commerce customer segmentation management based on improved K-means algorithm. *Information Systems and e-Business Management*, *18*, 497-510.

Erdoğmuş, İ. E., & Cicek, M. (2012). The impact of social media marketing on brand loyalty. *Procedia-Social and behavioral sciences*, *58*, 1353-1360.

Cebeci, U., Eğerci, F. & Altaş, B. (2024). Customer Segmentation in FMCG Sector with RFM Method and Digital Marketing Recommendations in E-Commerce

Ester, M., & Wittmann, R. (1998). Incremental generalization for mining in a data warehousing environment. In *Advances in Database Technology—EDBT'98: 6th International Conference on Extending Database Technology Valencia, Spain, March 23–27, 1998 Proceedings 6* (pp. 135-149). Springer Berlin Heidelberg.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd* (Vol. 96, No. 34, pp. 226-231).

Gustriansyah, R., Suhandi, N., & Antony, F. (2020). Clustering optimization in RFM analysis Based on k-Means. Indonesian Journal Of Electrical Engineering And Computer Science, 18(1), 470-476. https://doi.org/10.11591/ijeecs.v18.i1.pp470-477

Ho, T., Nguyen, S., Nguyen, H., Nguyen, N., Man, D. S., & Le, T. G. (2023). An Extended RFM Model for Customer Behaviour and Demographic Analysis in Retail Industry. *Business Systems Research: International journal of the Society for Advancing Innovation and Research in Economy*, *14*(1), 26-53.

Ithawi, A. M. (2022). *Sosyal medya pazarlamasının KOBİ müşterilerinin artmasında etkileri: Hazır giyim sektöründe bir araştırma* (Master's thesis, İstanbul Gelişim Üniversitesi Lisansüstü Eğitim Enstitüsü).

Jain, R. (2015). Impact of social media marketing on SME business. *SAMVAD*, *8*.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning, Vol. 112, p. 18. New York: Springer.

Karimi, S., & Naghibi, H. S. (2015). Social media marketing (SMM) strategies for small to medium enterprises (SMEs). *International Journal of Information, Business and Management*, *7*(4), 86.

Khan, I., & Iftikhar, F. (2017). The Impact of Email Marketing, Mobile Marketing and Retargeting on online Consumer Buying Behavior.

Marisa, F., Ahmad, S. S. S., Yusof, Z. I. M., Hunaini, F., & Aziz, T. M. A. (2019). Segmentation model of customer lifetime value in small and medium enterprise (SMEs) using K-means clustering and LRFM model. *International Journal of Integrated Engineering*, *11*(3).

Martinez, R. G., Carrasco, R. A., García-Madariaga, J., Gallego, C. P., & Herrera-Viedma, E. (2019). A comparison between fuzzy linguistic RFM model and traditional RFM model applied to campaign management. case study of retail business. Procedia Computer Science, 162, 281–289. https://doi.org/10.1016/j.procs.2019.11.286

Mckinney, W. (2011). pandas: a Foundational Python Library for Data Analysis and Statistics. Python for high performance and scientific computing, 14(9), 1-9. http://pandas.sf.net

Monalisa, S., & Kurnia, F. (2019). Analysis of DBSCAN and K-means algorithm for evaluating outlier on RFM model of customer behaviour. *Telkomnika (Telecommunication Computing Electronics and Control)*, *17*(1), 110-117.

Morris, N. (2009). Understanding digital marketing: marketing strategies for engaging the digital generation. *Journal of Direct, Data and Digital Marketing Practice*, *10*(4), 384-387.

Olonde, J. O. (2017). *Impact of digital marketing on sales growth of small and medium enterprises in Nairobi, Kenya* (Doctoral dissertation, University of Nairobi).

Omuya, E. O., Okeyo, G. O., & Kimwele, M. W. (2021). Feature selection for classification using principal component analysis and inform

Onyango, K. E. N. N. E. D. Y. (2016). Influence of digital marketing strategies on performance of cutflowers exporting firms in Kenya.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825- 2830. http://scikit-learn.sourceforge.net.

Peter, M. K., & Dalla Vecchia, M. (2021). The digital marketing toolkit: a literature review for the identification of digital marketing channels and platforms. *New trends in business information systems and technology: Digital innovation and digital business transformation*, 251-265.

Pradipta, I. M. D., Eka, A., Wahyudi, A., & Aryani, S. (2018). Fuzzy c-means clustering for customer segmentation. *Int. J. Eng. Emerg. Technol*, *3*(1), 18-22.

Prahalad, C. K., & Ramaswamy, V. (2004). Co-creating unique value with customers. *Strategy & leadership*, *32*(3), 4-9.

Qadaki Moghaddam, S., Abdolvand, N., & Rajaee Harandi, S. (2017). A RFMV model and customer segmentation based on variety of products. *Information Systems & Telecommunication*.

Randles, B. M., Pasquetto, I. V., Golshan, M. S., & Borgman, C. L. (2017). Using the Jupyter notebook as a tool for open science: An empirical study. ACM/IEEE Joint Conference on Digital Libraries (JCDL), 1-2.

Rugova, B., & Prenaj, B. (2016). Social media as marketing tool for SMEs: opportunities and challenges. *Academic Journal of Business*, *2*(3), 85-97.

Sahni, N. S., Wheeler, S. C., & Chintagunta, P. (2018). Personalization in email marketing: The role of noninformative advertising content. *Marketing Science*, *37*(2), 236-258.

Sarvari, P. A., Ustundag, A., & Takci, H. (2016). Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis. Kybernetes, 45(7), 1129– 1157. https://doi.org/10.1108/k-07-2015- 0180.

Shirole, R., Salokhe, L., & Jadhav, S. (2021). Customer segmentation using rfm model and k-means clustering. *Int. J. Sci. Res. Sci. Technol*, *8*, 591-597.

Starczewski, A., & Krzyżak, A. (2015). Performance evaluation of the silhouette index. In *Artificial Intelligence and Soft Computing: 14th International Conference, ICAISC 2015, Zakopane, Poland, June 14-18, 2015, Proceedings, Part II 14* (pp. 49-58). Springer International Publishing.

Taiminen, H. M., & Karjaluoto, H. (2015). The usage of digital marketing channels in SMEs. *Journal of small business and enterprise development*, *22*(4), 633-651.

Tosi, S. (2009). Matplotlib for Python developers. Packt Publishing Ltd.

Upadhyay, A. (2023). *An analysis of e-commerce purchase behaviour across the UK and Brazil* (Doctoral dissertation, Dublin, National College of Ireland).

Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation.

Wan, S., Chen, J., Qi, Z., Gan, W., & Tang, L. (2022, April). Fast RFM model for customer segmentation. In *Companion Proceedings of the Web Conference 2022* (pp. 965-972).

Wan, S., Chen, J., Qi, Z., Gan, W., & Tang, L. (2022, April). Fast RFM model for customer segmentation. In Companion Proceedings of the Web Conference 2022 (pp. 965-972).

Wei, J. T., Lin, S. Y., & Wu, H. H. (2010). A review of the application of RFM model. African Journal of Business Management, 4(19), 4199.

Youssef SMAILI, M., HACHIMI, H., New RFM-D Classification Model for Improving Customer Analysis and Response Prediction, Ain Shams Engineering Journal (2023), doi: https://doi.org/ 10.1016/j.asej.2023.102254

Yuliari, N. P. P., Putra, I. K. G. D., & Rusjayanti, N. K. D. (2015). Customer segmentation through fuzzy C-means and fuzzy RFM method. *Journal of Theoretical and Applied Information Technology*, *78*(3), 380

Zhang, D., & Huang, M. (2022). A precision marketing strategy of e-commerce platform based on consumer behavior analysis in the era of big data. *Mathematical Problems in Engineering*, *2022*.