






## Data Mining for Personalized Sales Strategies: A Clustering and Association Analysis Approach

Selçuk Alp<sup>1</sup>, Ebru Geçici<sup>2\*</sup>, Umut Rıfat Tuzkaya<sup>2</sup>, Ayhan Boyacıoğlu<sup>3</sup>, Yunus Taştutan<sup>3</sup>

<sup>1</sup>Yıldız Technical University, Faculty of Arts & Science, Department of Statistics, İstanbul, Türkiye.  
[ror.org/0547yzj13](http://ror.org/0547yzj13)

<sup>2</sup>Yıldız Technical University, Faculty of Mechanical Engineering, Department of Industrial Engineering, İstanbul, Türkiye.  
[ror.org/0547yzj13](http://ror.org/0547yzj13)

<sup>3</sup>F.I.T. Bilgi İşlem Sistemleri Servis Sanayi ve Ticaret A.Ş., İstanbul, Türkiye.

\*Corresponding author:  
[egecici@yildiz.edu.tr](mailto:egecici@yildiz.edu.tr)

Received: 14.01.2025

Revision: 05.11.2025

Accepted: 05.03.2026

Online: 19.04.2026

**Abstract:** Contemporary businesses must evaluate the performance of their sales personnel and refine their sales strategies. In this context, a variety of approaches are employed to develop strategies, including combining sellers based on their respective sales characteristics, to increase sales. Clustering, a machine learning approach, is used to derive inferences from sales data. The results are then used to inform future sales planning and determine priorities. To achieve this, the sellers are initially grouped (clustered) by similar characteristics based on specific criteria (such as sales volume and product information). This enables the identification of the typical strengths and weaknesses of sellers within each cluster. To illustrate, while sellers in a cluster with high sales volume and customer satisfaction scores may assume a pioneering role in the introduction of new products, it may be beneficial to investigate which products could be preferred in the region where sellers in a low-performing cluster are located, and what measures could be taken to increase sales of these products. By examining the sales performance of clustered sellers, it is possible to ascertain the relationships among the best-selling products across different applications. This approach enables the identification of products sold in conjunction, products that stimulate each other's sales, and products that appeal to disparate customer segments. Following the cluster analysis, an association analysis enables a more comprehensive investigation of the interrelationships among products. The results of this analysis permit the identification of product preferences among specific customer profiles. Based on the information mentioned above, more effective product recommendations and personalized marketing strategies can be formulated. An examination of sales within the identified clusters reveals pertinent information.

**Keywords:** Sales strategies, Clustering, RFM analysis, Association rule analysis

### I. INTRODUCTION

In today's competitive environment, performance evaluation is crucial. The performance evaluation sets objectives that are consistent with the organization's strategic goals and monitors and evaluates progress. Moreover, when the performance evaluation is considered, people's knowledge, skills, and abilities can be improved. Therefore, research on evaluating both employee and firm performance has occupied a prominent place in the literature from the past to the present. For instance, one of the most outstanding studies in the field is the evaluation of supplier performance in the supply chain. In this context, evaluations using various methods have been reported in the literature. Tong et al. (2020) provide an example of a fuzzy preference ranking organization method for enrichment evaluation (PROMETHEE) II, one of the multi-criteria decision-making (MCDM) methods for the evaluation of sustainable maintenance supplier performance in the petrochemical industry, while Giannakis et al. (2020) conduct a supplier sustainability performance evaluation. Several techniques can be employed to conduct these evaluations. MCDM (Ho et al., 2010; Agarwal et al., 2011; Giannakis et al., 2020; Behera & Beura, 2023), artificial intelligence and machine learning (Kohli, 2018; Dumitrascu et al., 2020), and a combination of these two methods (Liou et al., 2021; Tronnebati et al., 2022; Abdulla et al., 2023) are the most preferred methods today. Performance evaluation processes are also carried out for sellers, who are essential stakeholders of companies, such as suppliers. A seller's performance evaluation measures how closely a sales team's or a company's sales department's performance aligns with predetermined targets and how this performance impacts the process.

Sales performance evaluation provides information to identify strengths, improve uncovered areas, and reward outstanding contributions. Moreover, by examining the sellers' sales data, information such as which products are sold, where they are sold, and at what prices can be obtained. Thus, by using this information, future marketing strategies can be proposed. To this end, numerous studies have been conducted to evaluate the sellers. For example, Donassolo and Matos (2014) conduct a study to examine the factors affecting sales by conducting a survey on more than 300 sales personnel. Similarly, Kerr & Marcos-Cuevas (2022) conduct this study, in which they integrate objective and subjective evaluation criteria by applying a survey to 207 sales personnel and 39 sales managers. Using a naive Bayes model, Calixto

and Ferreira (2020) analyze three years of sales data from 594 salespeople at a global freight forwarding company to classify salespeople into predefined categories. Zohrehvandian et al. (2023) present a framework utilizing data mining and machine learning techniques to predict the effectiveness of salespersons' performance, aiming to guide club customers in selecting the most suitable shopping strategy.

A review of the literature reveals that performance evaluation processes have been monitored using various methods. This enables companies to assess their sales strategies and update their future recommendations based on these insights. This study aims to explore key metrics for evaluating sales performance, identify actionable insights for improvement, and establish a data-driven framework for marketing products. By integrating performance analytics with targeted marketing efforts, businesses can optimize resource allocation, improve customer satisfaction, and achieve sustained growth. For this purpose, this study integrates Recency, Frequency, Monetary (RFM) analysis and k-means clustering algorithms into the model. In literature, these algorithms are commonly used for customer segmentation; however, in this paper, they are applied to salesperson segmentation. Enhancing k-means clustering algorithms to group sellers with similar RFM characteristics into distinct segments. This machine learning approach helps identify natural patterns in sellers' behavior that might not be immediately apparent through traditional analysis. Moreover, the study of products sold together uses association rule analysis. Then, to improve sales strategies, suggestions can be made based on products that can be sold together and on sales staff performance to increase sales in the future. In this context, we can organize the continuation of the study as follows.

- While RFM analysis and *k*-means clustering are used for customer segmentation, this study applies them innovatively for salesperson segmentation.
- To provide insights into cross-selling and bundling strategies, the research integrates association rule analysis to identify products that are frequently sold together.
- By presenting performance evaluation and association rule analysis together, it demonstrates a holistic approach to sales strategies.

The rest of the paper is structured as follows. The second section presents the application of the salesperson's segmentation and its results. Section three concludes the paper.

## II. METHODOLOGY

The customer segmentation method, first developed by Smith in the 1950s (Smith, 1956), is based on dividing customers into subgroups according to their characteristics (Ho et al., 2023). Through the segments created, customer profiles are formed, and appropriate strategies are developed based on this information in the following processes. In other words, it is an unsupervised learning process that employs various clustering methods to categorize customer data by similarity using historical customer data (Alves Gomes & Meisen, 2023). Thus, integrating machine learning algorithms with user data can reveal customer segments and help businesses identify those that are challenging to identify through intuitive analysis and manual data inspection (Kumar, 2023). RFM analysis and k-means algorithms are presented, respectively, although various machine learning algorithms are used for customer segmentation in the literature.

### A. *Recency, Frequency, Monetary (RFM) Analysis for Customer Segmentation*

The RFM model, initially developed by Hughes (Hughes, 1994), aims to identify the most valuable customers from a set of customer data collected over a specified period. It represents one of the most prevalent behavioral segmentations (Alves Gomes & Meisen, 2023; Rungruang et al., 2024). The RFM model is a valuable tool for identifying customer relationships. It is widely employed in the development of discount and promotional strategies in marketing, as well as in the sales of loans and stocks by banks (Chang & Tsai, 2011; Hu et al., 2013).

The RFM algorithm uses customer data, including recency (R), frequency (F), and monetary value (M). Here, recency values represent the most recent purchase of the customer. In contrast, frequency values indicate the frequency of the customer's purchases, while monetary values represent the total amount spent by the customer. Based on these three attributes, a value between 1 and 5 is assigned to each attribute. That is, the RFM algorithm uses a quintile method, dividing customers into five quintiles. Then, based on the customer's group, a score is assigned to that customer. The first quintile with the highest F values is given a score of 5, and the remaining quintiles are given scores of 4, 3, 2, and 1 in rank order. The exact process is applied to M, then to R. Ultimately, all customers are ranked based on their R, F, and M values. Thus, by using this scoring process, the best customer group can achieve a score of 5-5-5, and the worst can achieve a score of 1-1-1. In this scoring system, customers with a score of 111 are considered lost customers. In contrast, customers with a score of 555 are considered loyal and are likely to be the most profitable for the company (Şentürk et al., 2024). The RFM score scale is shown in Table 1 (Cheng & Chen, 2009).

Table 1. The RFM score scale.

Score	Recency (%)	Frequency (%)	Monetary (%)
5	0-20	0-20	0-20
4	20-40	20-40	20-40
3	40-60	40-60	40-60
2	60-80	60-80	60-80
1	80-100	80-100	80-100

To facilitate consumer segmentation and classification, the RFM model has been combined with data mining techniques, including Self-Organizing Maps (SOM), K-means, and Decision Tree C4.5 (Gholamveisy et al., 2024).

### B. Clustering (*k*-Means Algorithm)

Machine learning methods are categorized as supervised and unsupervised, and this categorization varies with the dataset structure. If the data set contains the dependent variable expressed as  $Y$  in addition to the independent variable(s),  $X$ , the methods used to analyze this data set are called supervised learning. If the data set does not contain the  $Y$  variable, such learning problems are referred to as unsupervised learning. Most of the studies conducted under this learning approach are referred to as clustering. The aim here is to determine relationships among the observation values, discover hidden patterns, and group them.

Many different clustering algorithms can be applied to various fields, including medicine and finance. However, the most common clustering methods in the literature are  $k$ -means and hierarchical clustering. While the  $k$ -means method examines the relationship between observations using a predetermined number of clusters, hierarchical clustering uses a tree structure called a dendrogram to determine relationships among observations (James et al., 2013).

The next step is to explain the  $k$ -means the method to be used in the project. The reason for choosing this method is that the dataset is not suitable for hierarchical clustering, and the  $k$ -means the method is not appropriate. The  $k$ -means method attempting to partition the data set into  $k$  predetermined clusters. The following steps are followed for this separation process (James et al., 2013).

1. Clusters are created that contain the indices of the observations in each cluster by using  $C_1, C_2, \dots, C_K$ .
2. Each observation value is assigned to at least one cluster,  $C_1 \cup C_2 \cup \dots \cup C_K = \{1, \dots, n\}$ . Here  $n$  represents the total number of observations.
3. It is not allowed for an observation value to be present in more than one cluster,  $C_k \cap C_{k'} = \emptyset$  for each  $k \neq k'$ .

If the observation  $i$  is in a cluster, it should be denoted as  $i \in C_k$ . The primary purpose of the K-means method is to minimize the within-cluster variation, i.e., the deviation within the created clusters.  $W(C_k)$ . Thus, to solve the problem, the within-cluster variation in each cluster ( $\min_{C_1, C_2, \dots, C_K} \{\sum_{k=1}^K W(C_k)\}$ ) is tried to be minimized. While different measurement methods can be used to calculate within-cluster variation, the preferred distance metric is the Euclidean Distance. Accordingly, the Euclidean distance is expressed as  $W(C_k) = \frac{1}{|C_k|} \sum_{i, i' \in C_k} \sum_{j=1}^p (x_{ij} - x_{i'j})^2$  and the objective function is given by Eq. 1.

$$\min_{C_1, C_2, \dots, C_K} \left\{ \sum_{k=1}^K \frac{1}{|C_k|} \sum_{i, i' \in C_k} \sum_{j=1}^p (x_{ij} - x_{i'j})^2 \right\} \quad (1)$$

The number of clusters  $k$ , which is predetermined as a hyperparameter, is the most important parameter required for the model to work. Therefore, determining the appropriate number of clusters is crucial. One of the most used methods in the literature for the determination process is the *elbow curve*. To create the elbow graph, the within-cluster variation value is calculated for each  $k$  value. Then, the calculated values are plotted with the cluster number  $k$  on the x-axis and the obtained value on the y-axis. In the obtained graph, the point at which the slope of the transition between the clusters begins to decrease is identified as the appropriate cluster number.

### C. Association Rules Analysis

Data mining methods that analyze the co-occurrence of events are called association rules. These methods reveal the rules of co-occurrence with certain probabilities. Association rules are an approach that supports future studies by analyzing past data and identifying association behaviors within these data. The Apriori algorithm, developed by Agrawal and Srikant (1994), is the most widely known and used algorithm in association rule mining. The name of the algorithm comes from the word "prior", as it takes the information from the previous step. Apriori is an algorithmic method for identifying frequent item sets and generating association rules from transaction data (Silva et al., 2019).

In the association rule, the association between elements is measured by support and confidence. The support criterion indicates how often items in the data are correlated, whereas the confidence criterion indicates the probability that a product Y will be together with the product X. Calculations of the criteria can be provided as in Table 2.

Table 2. Calculation of the support and confidence criteria.

Criterion	Explanation
Support	X and Y are different products
	Support for product X is the proportion of product X in all purchases: $SUPPORT(X) = \text{The number of } X / \text{Total Number of Purchases}$
	Support for products X and Y is the probability that X and Y occur together in all purchases: $SUPPORT(X, Y) = \text{The number of } (X, Y) / \text{Total Number of Purchases}$
Confidence	X and Y are different products
	$CONFIDENCE(X, Y) = \text{The number of } (X, Y) / \text{Number of Purchases Containing } X$ $CONFIDENCE(X \Rightarrow Y) = SUPPORT(X, Y) / SUPPORT(X)$

The reliability of the rules obtained is directly proportional to the support and trust values.

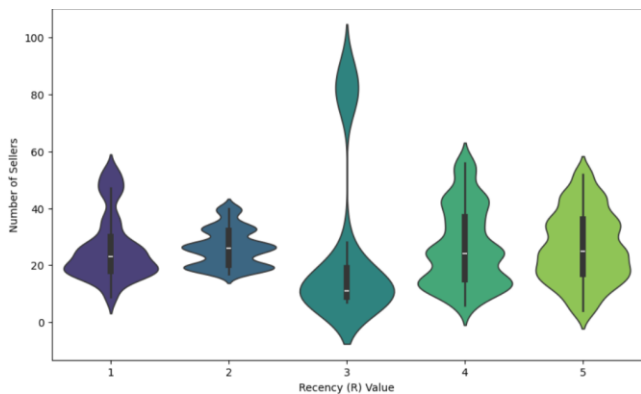
### III. APPLICATION AND RESULTS

Customer segmentation has evolved into a fundamental practice in marketing and data-driven decision-making. It involves dividing consumers into distinct subgroups based on behavioral, demographic, and transactional attributes, thereby enabling firms to design tailored strategies that enhance customer engagement and profitability. With the increasing availability of large-scale customer data, traditional segmentation approaches have been enriched by machine learning techniques, particularly unsupervised learning, which allow the detection of hidden patterns beyond the scope of intuitive or manual analysis. Among the most prominent methods are RFM analysis, which categorizes customers based on their purchasing behavior; k-means clustering, which groups observations into optimal clusters based on similarity measures; and association rule mining, which identifies co-occurrence patterns using algorithms such as Apriori. Together, these approaches provide a robust analytical framework for understanding consumer heterogeneity and formulating evidence-based marketing strategies.

In this study, the first stage uses customer segmentation, which involves dividing customers into segments based on sales data collected by sellers. Then, using the same data set, the relationships among sold products are examined to provide sellers with effective sales strategies to increase sales. Note that details of the methods used are presented in the Appendix.

Sales data of the personnel who market and sell digital accounting applications within the SOVOS company are used to conduct the specified studies. The models are implemented using a dataset of sales data from 3,280 sales employees, along with Python libraries. We want to note that the analysis is conducted using a computer equipped with an 11th-generation Intel(R) Core(TM) i5-1135G7 processor and 8 GB of RAM.

The RFM value is calculated using data obtained from sales made by sellers. Then, a density graph is created to show the number of sellers for each "R-F-M" trio. In this context, the graph type known as a "violin plot" (see Figure 1), which also includes a box plot (showing the quartiles in the data distribution), is used to illustrate the distributions.



(a)

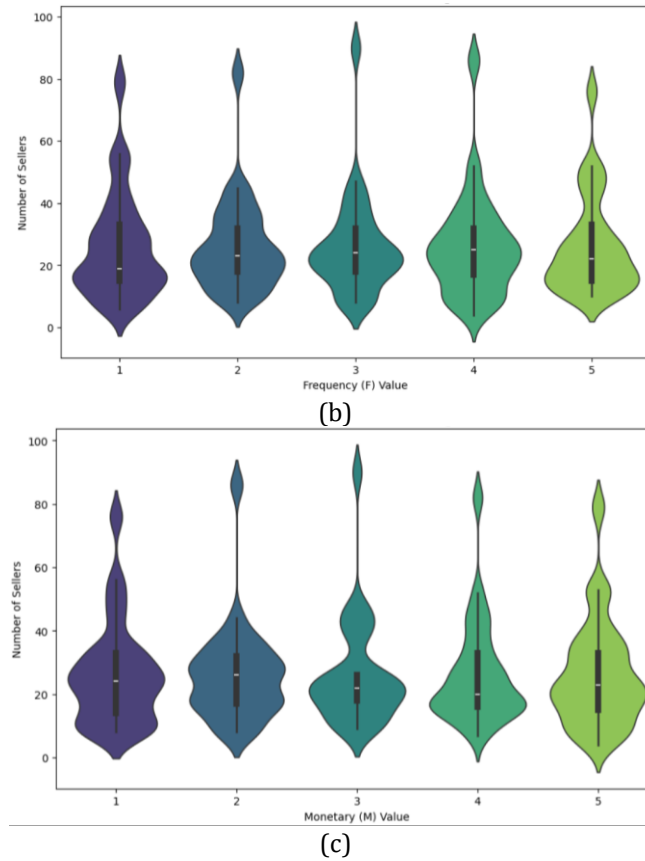


Figure 1. Distribution of the values: (a) recency (R) values, (b) frequency (F) values, and (c) monetary (M) values

The distribution graphics in Figure 1 examine how the number of sellers is distributed across RFM values, namely R, F, and M. Thus, the distribution of each RFM value among the seller groups and how it changes are shown. When we examine the sellers according to the R values (Figure 1(a)), which we measure the time between the last sale made by the sellers and the date taken as the basis in the analysis, it is observed that a similar structure exists for 4 and 5 (approximately the same average value and the most significant and most minor value). In contrast, the differentiation is greater for the value three compared to the others. This differentiation is due to the difference between the largest and the lowest value. Additionally, the middle of the data is lower, and sellers primarily concentrate between 0 and 20. In the graphs given for F and M values in Figure 1(b) and Figure 1(b), a distinction as significant as in R values is not seen. The F and M values, corresponding to the number of sales made by the sellers and the monetary value of those sales, respectively, exhibit parallel characteristics. While most sellers are concentrated between 0 and 60 for the F values, they are more evenly distributed in this range for F values of 1 and 4. Similarly, for the M value, the sellers are distributed between 0 and 60. The difference in this graph is observed in the value of 3, which has the most significant and smallest values.

Afterwards, the existing data is used to classify the sellers with the k-means algorithm. Within the scope of this classification, 120 models are created that differ in features such as the number of iterations, the initial algorithm, and the number of times the k-means the algorithm is run with different centroid seeds. These models are then run using the grid search algorithm. Then the model with the appropriate parameter values is found. As mentioned earlier, in k-means algorithms, the k value must be specified as an external input to the model. In this context, the graphic method known as the elbow graph, illustrated in Figure 2, is employed.

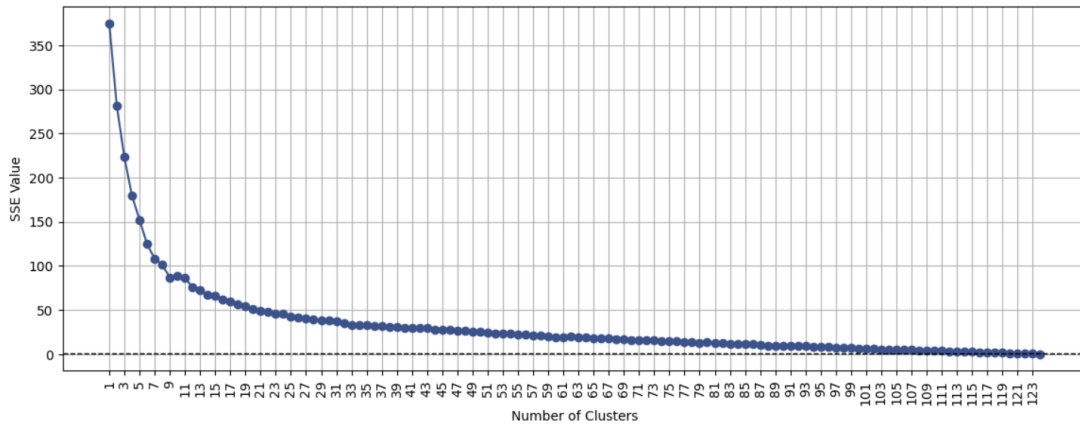


Figure 2. Elbow graph: Determination of the  $k$ -value for the  $k$ -means algorithm

To draw the specified elbow graph, the cluster numbers are shown on the  $X$ -axis, and the SSE values obtained from the  $k$ -means algorithm run using these cluster numbers are shown on the  $Y$ -axis. When the elbow graph in Figure 2 is examined, it is evident that the number of clusters is truncated at 9. The  $k$ -means algorithm was run with nine clusters. Additionally, to determine the adequacy of the number of clusters, the algorithm was run using different numbers of clusters. In this context, the model was run with  $k$  values of 3, 4, and 9.

When we run the model with  $k = 9$ , the center of the formed clusters is obtained, as shown by the purple triangles in Figure 3. In addition, when we examine these centers, we observe that some are very close to each other, and clusters can be identified as four if desired. In this context, clusters are created with four centers (represented by green circles) and three centers (represented by yellow rhombi).

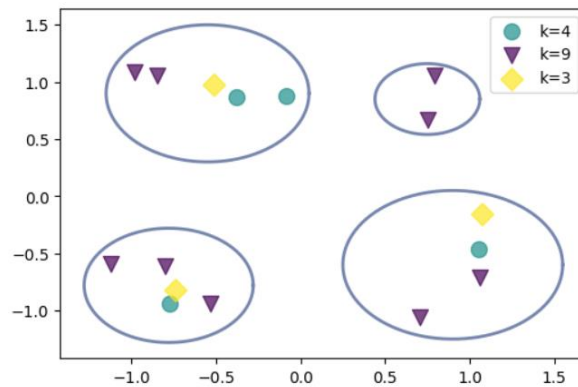


Figure 3. Centers of the clusters with respect to different numbers of clusters:  $k=3$ ,  $k=4$ , and  $k=9$

The change in the number of sellers in the clusters, as specified by the cluster numbers, is presented in Table 3. In this table, “#” shows the number of components in each cluster, whereas “Cs.” The column presents the number of sellers in each element. For example, for  $k = 3$ , the first cluster contains 46 components, and the total number of sellers in that cluster is equal to 1271. Moreover, “Ave.,” “Std.,” “Min” and “Max”, which are information of each component, are also presented in Table 3.

Table 3. Statistical information of clusters with respect to  $k=3$ ,  $k=4$ , and  $k=9$

Feature\Cluster	1	2	3	4	5	6	7	8	9
#	46	40	39						
Cs.	1271	917	1012						
Ave.	26.1	26.14	25.66						
Std.	18.24	18.00	14.45						
Min.	4	6	7						
Max.	90	90	82						

Table 3 (cont.). Statistical information of clusters with respect to  $k=3$ ,  $k=4$ , and  $k=9$

Feature\Cluster	1	2	3	4	5	6	7	8	9
$k=4$	#	37	30	33	25				
	Cs.	1028	617	895	740				
	Ave.	26.52	26.10	26.36	26.25				
	Std.	18.27	16.81	14.43	16.76				
	Min.	4	4	14	4				
	Max.	52	90	82	90				
$k=9$	#	18	13	15	10	12	16	17	12
	Cs.	463	304	329	314	382	445	401	338
	Ave.	25.72	23.83	21.93	31.4	31.83	27.81	23.59	28.17
	Std.	13.02	20.08	14.45	20.03	21.45	16.64	16.48	10.29
	Min.	6	7	7	7	6	9	4	9
	Max.	56	90	82	90	82	86	86	53

When the clusters are examined by the number of sellers they contain, the largest clusters are formed with  $k = 3$ , with cluster sizes ranging from 39 to 46 sellers. Additionally, the most balanced distribution is observed in the clusters formed with  $k = 4$ . According to the results, as the number of clusters increases, cluster sizes decrease markedly (ranging from 10 to 18 sellers), and variation between them increases. In addition, no significant difference is observed in the structures of the clusters at smaller  $k$  values. That is, when examining average values across clusters, it is observed that for  $k = 3$  and  $k = 4$ , the averages remain close to each other (around 25–26), indicating no significant structural difference. However, with  $k = 9$ , the averages vary more widely (from 21.93 to 31.83), showing that higher cluster counts introduce greater heterogeneity in the group structures. This pattern indicates that while a smaller number of clusters captures general patterns without significant structural differences, increasing the number of clusters can over-segment the data, thereby reducing interpretability. From a managerial perspective, the  $k = 4$  solution may be preferable, as it strikes a balance between interpretability and actionable differentiation. Sales managers could, for example, tailor training or incentive schemes based on these four relatively balanced groups, ensuring that managerial resources are allocated equitably while still recognizing performance diversity.

After the RFM analysis and clustering, the association analysis for the products sold in these sales transactions is complete. As a result of this process, the conclusion is reached on which products are sold together. In this context, for example, the relationships between product 4 and products 1-2-3 and 8, as well as the relationship between product 1 and product 2, are examined in the product pool, which includes eight different products. For these product combinations, “Support” and “Confidence” values are calculated with equations given in Appendix Table 3. Then, using these values, network diagrams are created for each calculation, as presented in Figures 4 and 5.

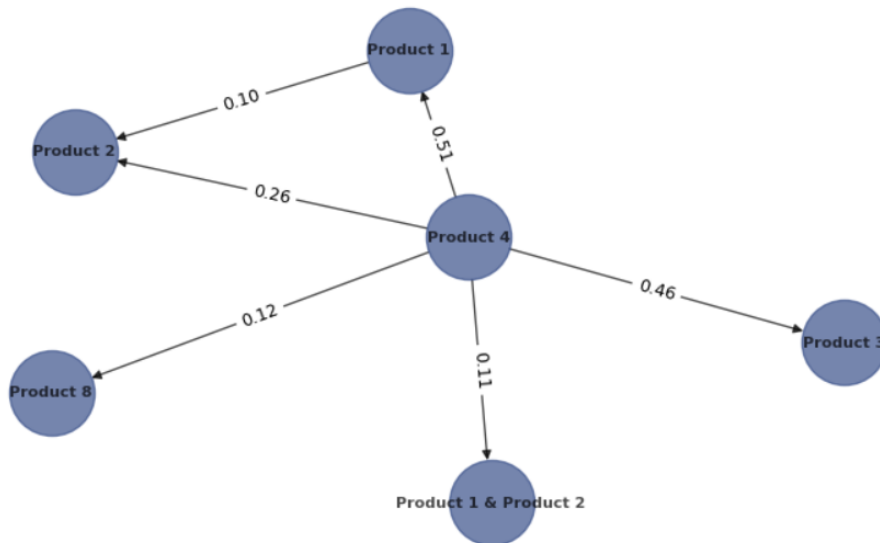


Figure 4. Association rule network diagram by confidence

When the trust relationship between the selected products is examined, it is evident that a strong relationship exists between Product 1 and Product 4, compared to the other relationships. If Product 4 is sold, it can be emphasized that the probability of selling Product 1 is high compared to other sales that can be made together. The lowest trust score is created between Product 1 and Product 2. Similarly, the trust value between Products 4 and 8 is low. In this case, it would

not be appropriate to generalize, such as stating that Product 4 is sold, Product 8 is also sold, or that Product 1 is sold and Product 2 is also sold. In summary, it would be appropriate to consider high-confidence matches when generalizing to products that can be sold together.

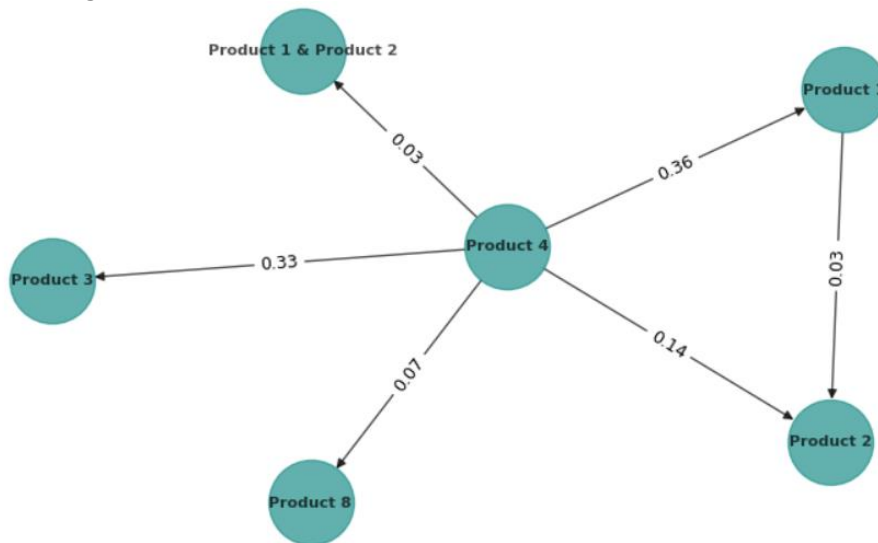


Figure 5. Association rules network diagram by support

The relationship between the products based on the support score is presented in Figure 5. Here, the highest Support score, as in Trust, is observed between Products 4 and 1. The support score between Product 4 and Product 1, which is higher than the other pairs', indicates that this pair is observed more frequently in sales than the other pairs. According to the association rule analyses based on support and confidence, the connections that Product 4 has established with several other products allow us to name this product the central product. Finally, if Product 4 is to be sold, Products 3 and 1 can be offered as alternatives. For product relationships with low trust scores, it would be appropriate to suggest different sales strategies.

Figures 4 and 5 further demonstrate that rules involving Product 4 not only achieve the highest confidence (e.g., Product 4 → Product 1 with 0.51) but also the highest support (e.g., Product 4 → Product 1 with 0.36), confirming its central role in the network. While confidence highlights the reliability of individual rules, support reveals their overall frequency across the dataset.

#### IV. CONCLUSION

Salesperson/seller segmentation is a strategic approach to categorizing salespeople or sellers by their transactional behavior and value to the organization. To track sales, companies maintain digital records of sales made by their sellers. Then, using this information, the processes for determining sales strategies and promotions for the following periods are carried out. Within the scope of this study, the analysis of sales data from a company that sells digital accounting applications is conducted. In this context, as in the customer segmentation literature, sellers are divided into segments. For this process, RFM analysis is applied first, and then the data are clustered using the *k*-means clustering algorithm, a popular unsupervised learning technique. In the second stage, the products sold are examined, and association rule analysis is performed to identify strategies for selling products together.

RFM Analysis and *k*-means methods are considered during the segmentation phase for the sellers. Based on the RFM analysis, it is concluded that the sellers' sales differed in recency but shared similar characteristics in frequency and monetary value. When evaluated using cluster analysis results, the sellers can be categorized into 9 groups based on sales; however, it is also appropriate to determine them as 3 or 4 groups. Moreover, this segmentation method provides a data-driven foundation for developing targeted seller management strategies and making informed procurement decisions.

In the second step, to understand which products are preferred together in the sales made, and then to market these products together, the association rule analysis concluded that there is a strong relationship between Product 1 and Product 4 according to the trust and support scores, and that these two products are sold more compared to other research. Similarly, the relationship between Product 4 and Product 3 could also be considered. Regarding different products, no significant relationship is observed. These findings highlight the central role of Product 4 in the product network and suggest that bundling it with Products 1 and 3 could increase sales effectiveness. From a managerial perspective, such associations provide actionable insights for cross-selling strategies, enabling sales managers to design targeted campaigns that focus on the most frequently co-purchased products while avoiding unprofitable pairings.

Future studies can further evaluate these analyses and integrate them with different clustering methods. Additionally, time series analysis can be incorporated into RFM analysis to track periodic changes in salesperson behavior. In association rule analysis, lift can be evaluated alongside support and confidence scores. Finally, the results of these analyses can be integrated into sales and marketing strategies by developing a suggestion system. A decision support system can be developed, allowing management and employees to easily track the application of methods to increase both salesperson's performance and customer satisfaction.

## DECLARATIONS

**Acknowledgments:** The author/authors do not wish to acknowledge any individual or institution.

**Author Contributions:** Conceptualization, S.A., E.G. and U.R.T; Methodology, S.A., E.G. and U.R.T.; Software, S.A., E.G., U.R.T, A.B. and Y.T.; Validation, S.A., E.G., U.R.T, A.B. and Y.T.; Formal Analysis, S.A., E.G., U.R.T, A.B. and Y.T.; Investigation, S.A., E.G., U.R.T, A.B. and Y.T.; Resources, S.A., E.G., U.R.T, A.B. and Y.T.; Data Curation, S.A., E.G., U.R.T, A.B. and Y.T.; Writing—Original Draft, S.A., E.G., and U.R.T; Writing—Review & Editing, S.A., E.G., U.R.T, A.B. and Y.T.; Visualization, S.A. and E.G; Supervision, S.A. All authors have read and approved the final version of the manuscript.

**Conflict of Interest:** The authors declare no conflict of interest.

**Supporting Institutions:** This research received no external funding.

**Ethical Approval:** This study does not involve human or animal participants. All procedures followed scientific and ethical principles, and all referenced studies are appropriately cited.

**Plagiarism Statement:** This article has been evaluated for plagiarism, and no instances of plagiarism were detected.

**Use of AI Tools:** The author/authors declare that no Artificial Intelligence (AI) tools were used in the creation of this article.

## REFERENCES

- Abdulla, A., Baryannis, G., & Badi, I. (2023). An integrated machine learning and MARCOS method for supplier evaluation and selection. *Decision Analytics Journal*, 9, Article 100342. <https://doi.org/10.1016/j.dajour.2023.100342>
- Agarwal, P., Sahai, M., Mishra, V., Bag, M., & Singh, V. (2011). A review of multi-criteria decision making techniques for supplier evaluation and selection. *International Journal of Industrial Engineering Computations*, 2(4), 801–810. <https://doi.org/10.5267/j.ijiec.2011.06.004>
- Agrawal, R. & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)* (pp. 487–499).
- 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(3), 527–570. <https://doi.org/10.1007/s10257-023-00640-4>
- Behera, D. K., & Beura, S. (2023). Supplier selection for an industry using MCDM techniques. *Materials Today: Proceedings*, 74(4), 901–909. <https://doi.org/10.1016/j.matpr.2022.11.291>
- Calixto, N., & Ferreira, J. (2020). Salespeople performance evaluation with predictive analytics in B2B. *Applied Sciences*, 10(11), Article 4036. <https://doi.org/10.3390/app10114036>
- Chang, H. C., & Tsai, H. P. (2011). Group RFM analysis as a novel framework to discover better customer consumption behavior. *Expert Systems with Applications*, 38(12), 14499–14513. <https://doi.org/10.1016/j.eswa.2011.05.034>
- Cheng, C. H., & Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, 36(3), 4176–4184. <https://doi.org/10.1016/j.eswa.2008.04.003>
- Donassolo, P. H., & de Matos, C. A. (2014). The predictors of sales performance: A study with wholesale sellers. *Revista Brasileira de Gestão de Negócios*, 16(52), 448–465. <https://doi.org/10.7819/rbgv.v16i52.1686>
- Dumitrascu, O., Dumitrascu, M., & Dobrotă, D. (2020). Performance evaluation for a sustainable supply chain management system in the automotive industry using artificial intelligence. *Processes*, 8(11), Article 1384. <https://doi.org/10.3390/pr8111384>
- Giannakis, M., Dubey, R., Vlachos, I., & Ju, Y. (2020). Supplier sustainability performance evaluation using the analytic network process. *Journal of Cleaner Production*, 247, Article 119439. <https://doi.org/10.1016/j.jclepro.2019.119439>
- Gholamveisy, S., Homayooni, S., Shemshaki, M., Sheykhan, S., Boozary, P., Tanhaei, H. G., & Akbari, N. (2024). Application of data mining technique for customer purchase behavior via extended RFM model with focus on BCG matrix from a data set of online retailing. *Journal of Infrastructure, Policy and Development*, 8(7), Article 4426. <https://doi.org/10.24294/jipd.v8i7.4426>
- 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. <https://doi.org/10.2478/bsrj-2023-0002>
- Ho, W., Xu, X., & Dey, P. K. (2010). Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. *European Journal of Operational Research*, 202(1), 16–24. <https://doi.org/10.1016/j.ejor.2009.05.009>
- Hu, Y. H., Huang, T. C. K., & Kao, Y. H. (2013). Knowledge discovery of weighted RFM sequential patterns from customer sequence databases. *Journal of Systems and Software*, 86(3), 779–788. <https://doi.org/10.1016/j.jss.2012.11.016>
- Hughes, A. M. (1994). *Strategic database marketing*. Probus Publishing Company.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.
- Kerr, P. D., & Marcos-Cuevas, J. (2022). The interplay between objective and subjective measures of salesperson performance: Towards an integrated approach. *Journal of Personal Selling & Sales Management*, 42(3), 225–242. <https://doi.org/10.1080/08853134.2022.2044344>
- Kohli, M. (2018). Supplier evaluation model on SAP ERP application using machine learning algorithms. *International Journal of Engineering & Technology*, 7(28), 306–311. <https://doi.org/10.14419/ijet.v7i2.28.12951>
- Kumar, A. (2023). Customer segmentation of shopping mall users using K-means clustering. In *Advancing SMEs toward e-commerce policies for sustainability* (pp. 248–270). IGI Global. <https://doi.org/10.4018/978-1-6684-5727-6.ch013>
- Liou, J. J. H., Chang, M. H., Lo, H. W., & Hsu, M. H. (2021). Application of an MCDM model with data mining techniques for green supplier evaluation and selection. *Applied Soft Computing*, 109, Article 107534. <https://doi.org/10.1016/j.asoc.2021.107534>
- Rungruang, C., Riyapan, P., Intarasit, A., Chuarkham, K., & Muangprathub, J. (2024). RFM model customer segmentation based on hierarchical approach using FCA. *Expert Systems with Applications*, 237, Article 121449. <https://doi.org/10.1016/j.eswa.2023.121449>
- 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. [https://doi.org/10.17932/IAU.IAUSB.2021.021/iausbd\\_v16i4004](https://doi.org/10.17932/IAU.IAUSB.2021.021/iausbd_v16i4004)
- Silva, J., Varela, N., López, L. A. B., & Millán, R. H. R. (2019). Association rules extraction for customer segmentation in the SMEs sector using the apriori algorithm. *Procedia Computer Science*, 151, 1207–1212. <https://doi.org/10.1016/j.procs.2019.04.173>
- 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>
- Tong, L., Pu, Z., Chen, K., & Yi, J. (2020). Sustainable maintenance supplier performance evaluation based on an extend fuzzy PROMETHEE II approach in petrochemical industry. *Journal of Cleaner Production*, 273, Article 122771. <https://doi.org/10.1016/j.jclepro.2020.122771>
- Tronnebati, I., El Yadari, M., & Jawab, F. (2022). A review of green supplier evaluation and selection issues using MCDM, MP and AI models. *Sustainability*, 14(24), Article 16714. <https://doi.org/10.3390/su142416714>
- Zohrehvandian, K., Ghaffarian, H., & Mahmoudi, A. (2023). Predicting the level of salesperson's performance in encouraging customers to use appropriate shopping strategies in sports clubs. *Interdisciplinary Journal of Management Studies*, 17(1), 169–183. <https://doi.org/10.22059/ijms.2023.342973.675100>