Jurnal Serambi Engineering JSE
Volume X, No.2, April 2025 Hal 12665 - 12675
p-ISSN : 2528-3561
e-ISSN : 2541-1934

# Customer Segmentation Analysis with RFM Model (Recency, Frequency, Monetary) and K-Means Clustering: Case Study of Bottled Water Sales at PT XYZ

**Ema Rosary Sitorus*, Isna Nugraha**

Industrial Engineering Department, Universitas Pembangunan Nasional Veteran Jawa Timur, Surabaya
***Corresponding author**: 22032010177@student.upnjatim.ac.id

Received: January 3, 2024 Approved: January 11, 2025

**Abstract**
Customer segmentation is a crucial process in understanding consumer behavior patterns to support strategic decision making in marketing. The main challenge faced by companies is to accurately group customers based on transaction data. The purpose of this study is to find out and segment customers using the algorithm K-Means clustering based on RFM model (Recency, Frequency, Monetary) on Bottled Water sales transaction data at PT XYZ. The research method involves analysis of 111 customer data processed using software Orange Data Mining, with validation of results using Silhouette Score which is useful in determining the amount cluster ideal. This research produced four cluster customers, with Cluster 4 reflects customers with the highest level of loyalty, marked by a value Frequency And Monetary the dominant one, while Cluster 3 describes customers with low loyalty potential. The results of this study provide a scientific basis for the development of more focused and efficient data-based marketing strategies.
**Keywords:** *customer segmentation, data mining, k-means clustering, marketing strategy, rfm*

**Abstrak**
Segmentasi pelanggan adalah proses krusial dalam memahami pola perilaku konsumen untuk mendukung pengambilan keputusan strategis dalam pemasaran. Tantangan utama yang dihadapi perusahaan adalah mengelompokkan pelanggan secara akurat berdasarkan data transaksi. Penelitian ini bertujuan untuk melakukan segmentasi pelanggan menggunakan algoritma K-Means clustering berbasis model RFM (Recency, Frequency, Monetary) pada data transaksi penjualan AMDK di PT XYZ. Metode penelitian melibatkan analisis terhadap 111 data pelanggan yang diolah menggunakan software Orange Data Mining, dengan validasi hasil menggunakan Silhouette Score untuk menentukan jumlah cluster optimal. Penelitian ini menghasilkan empat cluster pelanggan, dengan Cluster 4 mencerminkan pelanggan dengan tingkat loyalitas tertinggi, ditandai oleh nilai Frequency dan Monetary yang dominan, sementara Cluster 3 menggambarkan pelanggan dengan potensi loyalitas rendah. Hasil penelitian ini memberikan landasan ilmiah untuk pengembangan strategi pemasaran berbasis data yang lebih terfokus dan efisien.
**Kata Kunci:** *data mining, k-means clustering, rfm, segmentasi pelanggan, strategi pemasaran*

## 1. Introduction

In the era of digitalization and increasingly fierce competition, understanding consumer behavior in depth is something that companies need to pay attention to in order to remain relevant, maintain, and expand market share. Consumer behavior refers to the actions and decisions of individuals in the process of purchasing or using products or services. This process is not only limited to purchasing activities, but also includes various decisions and factors that influence these decisions [1]. Market segmentation is the process of grouping the market into smaller parts, while consumer segmentation is an effort to divide the target market based on certain characteristics to design more relevant business strategies [2]. This segmentation aims to group consumers into unique, meaningful, and homogeneous subcategories based on their attributes and characteristics. Thus, companies can understand their customers more deeply and develop effective strategies to establish relationships according to the needs and characteristics of each group [3].

This study takes PT XYZ as the object of study. PT XYZ has various types of customers, ranging from household individuals to companies and institutions, with diverse backgrounds and consumption patterns. Despite having a broad customer base, the company has never implemented formal data-based customer segmentation. The marketing strategies used are still general and do not consider specific customer characteristics or preferences. This increases the risk of losing customers, especially if competitors offer something more attractive or more suited to their needs [4]. In an increasingly competitive

situation, the inability to manage customer segments properly can hinder companies from maintaining market share and facing pressure from competitors. Therefore, PT XYZ needs to implement a more focused approach to identifying customer needs and designing a marketing strategy that is able to adjust offers and promotions to the specific needs of each customer segment [5].

To overcome this problem, PT XYZ needs to implement customer segmentation using the RFM model (Recency, Frequency, Monetary) based on the last date the customer made a transaction, the frequency of the customer making transactions and the number of transactions in a certain period [6]. Method data mining which is applied to customer segmentation is clustering, with the algorithm K-Means as the main approach. The data to be analyzed using the RFM model is product sales transaction data, after which a segmentation process will be carried out based on the algorithm K-Means [7]. The number of cluster the ideal is determined by the index Silhouette Score. Combination of RFM model and method K-Means facilitate grouping customers into the right categories and identifying their loyalty levels [8]. Through this approach, PT XYZ can identify customers with high risk churn higher, develop more effective retention strategies, and develop more optimal marketing strategies and adapt them to each customer segment. In a study conducted by Wahyuni, they successfully combined the RFM model with the method K-Means for customer segmentation in the printing sector. The results show that the combination of the two methods can produce more accurate and applicable segmentation. However, the study only used customer transaction data over a three month period, so it did not cover seasonal fluctuation patterns or long term promotional influences that could affect customer behavior [9]. This limitation is a gap that underlies the importance of further research with longer and more comprehensive data coverage.

This research is motivated by the need for PT XYZ to understand customer behavior in more deeply based on sales data for the last 10 months, namely from January to October 2024. The selection of this time period can include stable purchasing patterns as well as fluctuations that may occur throughout the year, including certain promotional periods or seasons. With the data from the last 10 months, PT XYZ can gain a more relevant and up to date understanding of customer behavior, so that the resulting segmentation can reflect actual and dynamic market conditions. Therefore, the purpose of this study is to analyze customer segmentation by integrating the RFM model and the method K-Means based on the latest sales data at PT XYZ. The resulting data based segmentation will allow PT XYZ to identify customer segments with high long term value, as well as understand the unique needs of each customer group. With these more specific analysis results, PT XYZ is expected to be able to increase customer loyalty through personalized strategies, while preventing the potential loss of high-value customers. Furthermore, understanding this segmentation can also strengthen the company's competitiveness in facing increasingly fierce competition in the business industry, as well as encourage increased profitability and smooth running of the company in the future [10].

## 2. Material and Methods

*Location and Time of Research*

This research was conducted in November at PT XYZ, which is located in the East Java area. PT XYZ is engaged in bottled water. The data used comes from sales of bottled water for the last ten months, from January to October 2024.

*Data Type and Resource*

Quantitative data is data used in this study, which is data that can be calculated and in nominal form. Bottled Water sales transaction data is included in the quantitative data category, because it is in the form of numbers that can be further processed for analysis. The data sources used in this study are primary and secondary data. Primary data were collected directly from research sources, namely customer transactions recorded in the sales system [11]. The data were collected to be used as a dataset that will be analyzed using data mining methods.

*Data Collection Methods*

To get pertinent data about client segmentation, this study employed a number of data gathering techniques. First, in order to obtain a thorough understanding of the customer transaction process, direct interviews with relevant parties particularly the sales unit were undertaken. Additionally, direct observation at the research site was utilized to gather data, particularly on the business operations occurring in the sales department [12]. Documentation is also used to gather notes or records from observations and interviews that will be utilized for additional analysis. Lastly, a review of the literature was done to find relevant references, such as earlier studies that dealt with the subject of consumer segmentation utilizing data mining

and the K-Means clustering technique. Together, these techniques offer a thorough understanding of consumer behavior and facilitate more precise analysis in this study.

*Data Mining Life Cycle*

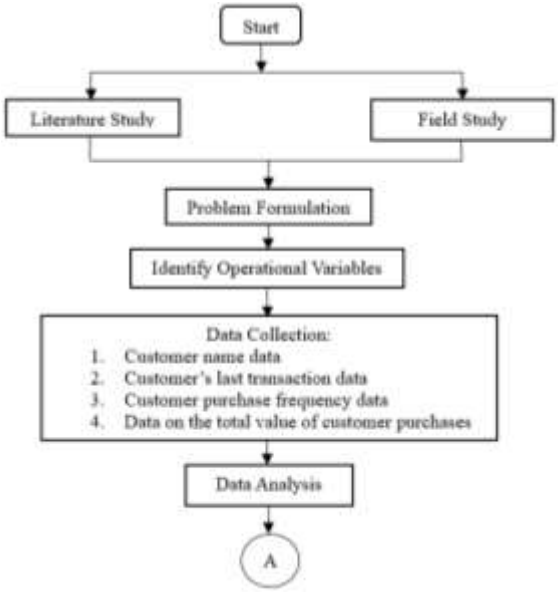
a. Data Cleaning, the collected data is processed to eliminate irrelevant or missing data so that analysis can be carried out with clean and accurate data [13].
b. Data Integration, data that has been collected from various sources from various sources is combined into data set to ensure that all necessary information is collected [14].
c. Data Selection, the process of selecting and sorting relevant data for analysis, in this study namely transaction data related to variables Recency, Frequency, and Monetary [15].
d. Transformation, the data is converted into a format that is appropriate for the analysis procedure to be performed using an algorithm K-Means clustering.
e. Data Mining, the algorithm K-Means clustering applied to segment customers into several groups based on their level of similarity. In this study, software Orange Data Mining is used to perform analysis clustering.
f. Pattern Evaluation, results clustering evaluated using the method Silhouette Score to determine the quality of the resulting groupings and ensure that the number of cluster what is created is optimal.

*Algorithm Process K-Means*

To perform the clustering process, the first step is to determine the desired number of groups (K). After that, initialize the cluster centers randomly. Next, calculate the distance of each data from the cluster center using the Euclidean Distance formula. The data is then classified into clusters based on the closest distance to the cluster center. After classification, the position of the cluster center is updated by calculating the average of the data in the cluster. This process is repeated from the step of calculating the distance to updating the cluster center position until the center position does not change anymore, indicating that the clustering process has reached convergence [16].

*Flow Chart*

**Figure 1** Flow Chart is a visual representation of steps or processes in the form of a diagram consisting of certain symbols. In this context, flow chart aims to illustrate the stages of research completion from start to finish. The research began with a literature review and field observation, then continued with formulating the problem and identifying operational variables. Furthermore, customer data was collected which included information on name, last transaction, purchase frequency, and total purchase value. The collected data was analyzed using Orange Data Mining software with an approach K-Means Cluster and Silhouette Score to evaluate the validity of the results. If the results are valid, the next step is discussion and preparation of conclusions and suggestions. The research process ends after all steps are completed.



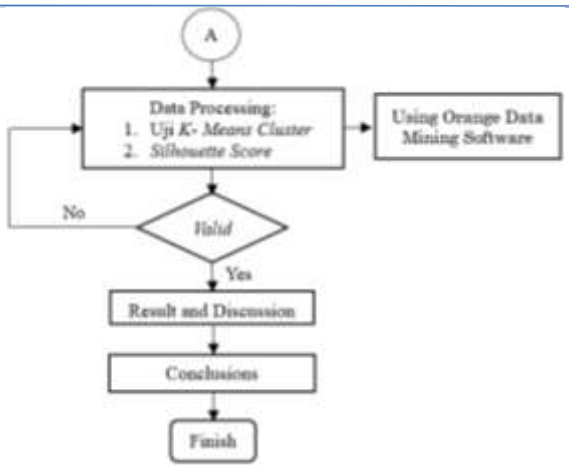



**Fig. 1**: Flow Chart
Source: Researcher (2025)

## 3. Results and Discussion

*Customer Segmentation*

Segmentation is the activity of dividing customers into several groups based on the level of customer loyalty to design an effective marketing strategy. This is the first step in developing a business model [8]. Customer segmentation aims to divide customers into groups with similar needs and behaviors. Marketing activities, such as communication, product or service offerings, and designed programs, become more effective because they are tailored to the similarities in the needs and behavior of each segment [17].

*Data Collection and Cleaning*

Transactions in a certain period are collected for analysis. By utilizing transaction summary data online from website companies in the period from January to October 2024 obtained from the Bottled Water sales division. This process is carried out so that unavailable data or attributes can be eliminated [18]. It can be seen in **Figure 2**. Bottled Water Transaction Data at PT XYZ A total of 2,845 transaction recaps were successfully collected during this period.

| | A | B | C |
|---|---|---|---|
| 1 | Nama Toko | Satuan | Harga |
| 2 | Pembelian_1 | 20 | 297.297,20 |
| 3 | Pembelian_2 | 120 | 153.189,60 |
| 4 | Pembelian_3 | 21 | 283.783,71 |
| 5 | Pembelian_4 | 25 | 337.837,75 |
| 2842 | Pembelian_2841 | 15 | 236.486,55 |
| 2843 | Pembelian_2842 | 20 | 315.315,40 |
| 2844 | Pembelian_2843 | 25 | 360.360,25 |
| 2845 | Pembelian_2844 | 23 | 331.531,43 |

**Fig. 2**: Bottled Water Transaction at PT XYZ
Source: Analysis Results (2025)

*Data Integration*

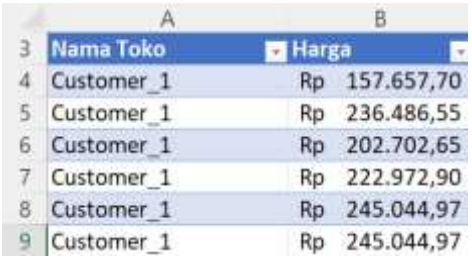
At this stage, merging transaction data from the same customer and customers who make transactions more than once is a stage of the data integration process [19]. Before merging, the author first recorded important information, such as the date of the last transaction made by each customer, the total number of transactions issued by each customer, and the total shopping value collected during the analysis period, namely from January to October 2024. This step aims to ensure that the merged data accurately reflects customer transaction activity, which can be seen in **Figure 3**. Transactions from the Same Customer.

| | A | B | C | |
|---|---|---|---|---|
| 3 | Nama Toko | Satuan | Harga | |
| 4 | Customer_1 | 10 | Rp | 157.657,70 |
| 5 | Customer_1 | 15 | Rp | 236.486,55 |
| 6 | Customer_1 | 15 | Rp | 202.702,65 |
| 7 | Customer_1 | 15 | Rp | 222.972,90 |
| 8 | Customer_1 | 17 | Rp | 245.044,97 |
| 9 | Customer_1 | 17 | Rp | 245.044,97 |

**Fig. 3**: Transactions from the Same Customer
Source: Analysis Results (2025)

*Data Selection*

Researchers select data that is considered relevant and can be used as variables for analysis [20]. In **Figure 4,** transaction Data Recap Selection, it can be seen that the attributes that will be used in this study are taken from the customer name and transaction amount attributes.



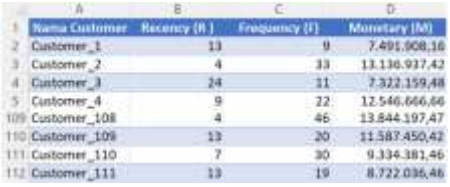
| | A | B | |
|---|---|---|---|
| 3 | Nama Toko | Harga | |
| 4 | Customer_1 | Rp | 157.657,70 |
| 5 | Customer_1 | Rp | 236.486,55 |
| 6 | Customer_1 | Rp | 202.702,65 |
| 7 | Customer_1 | Rp | 222.972,90 |
| 8 | Customer_1 | Rp | 245.044,97 |
| 9 | Customer_1 | Rp | 245.044,97 |

**Fig. 4**: Transaction Data Recap Selection
Source: Analysis Results (2025)

*Data Transformation*

At this stage, the data is converted into a format that is ready for further processing in data mining. The author adds three new attributes, namely Recency(day), Frequency (number of transactions) and Monetary (total spending), which is used to build the RFM model [21].

a. Recency, calculated by measuring the time difference between the customer's last transaction date and the research date, which describes the time elapsed since the last transaction.
b. Frequency, describes how many times a customer made a transaction during the analysis period.
c. Monetary refers to the total customer expenditure throughout the analyzed period, calculated by adding up all payments made by the customer [22].



| | A | B | C | D |
|---|---|---|---|---|
| 1 | Nama Customer | Recency (R ) | Frequency (F) | Monetary (M) |
| 2 | Customer_1 | 13 | 9 | 7.491.908,16 |
| 3 | Customer_2 | 4 | 33 | 13.136.937,42 |
| 4 | Customer_3 | 24 | 11 | 7.322.159,48 |
| 5 | Customer_4 | 9 | 22 | 12.546.666,66 |
| 109 | Customer_108 | 4 | 46 | 13.844.197,47 |
| 110 | Customer_109 | 13 | 20 | 11.587.450,42 |
| 111 | Customer_110 | 7 | 30 | 9.334.381,46 |
| 112 | Customer_111 | 13 | 19 | 8.722.036,46 |

**Fig. 5**: Attribute Creation Recency, Frequency and Monetary
Source: Analysis Results (2025)

After that, transaction data is merged to reduce redundancy. It can be seen in **Figure 5**. Attribute Creation Recency, Frequency and Monetary, the data that originally amounted to 2.845 transactions, where some customers were recorded more than once, after being combined into 111 transactions representing 111 unique customers. This merging process aims to avoid double recording and ensure that each transaction data used represents a different customer, so that the analysis clustering which is done can produce more accurate and optimal modeling.

*Data Mining*

In this study, the technique data mining applied using algorithms K-Means clustering, with the model implemented through Orange Data Mining software. **Figure 6**, process Import Data is the first stage that must be carried out, namely the researcher adds File Widget into the worksheet and import files data that will be used for in depth research software Orange Data Mining.

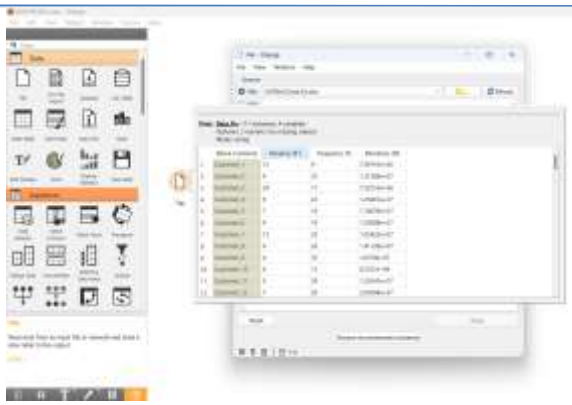


**Fig. 6**: Process Import Data
Source: Analysis Results (2025)

The next steps can be seen in **Figure 7**. Process Preview Data, at this stage the researcher added Data Table Widget to check the preview of the data to be processed and ensure that no data is lost.

**Fig. 7**: Process Preview Data
Source: Analysis Results (2025)

After verifying that the data is complete, the researcher proceeds to add K-Means widget which is connected with Data Table Widget to carry out the clustering process using the algorithm K-Means. The attributes used in this clustering include Recency, Frequency and Monetary. At this stage, the researcher chooses the setting K-Means+ and enable the option maximum iterations with a maximum iteration limit of 300 which can be seen in **Figure 8**. Settings K-Means Widget.


**Fig. 8**: Settings K-Means Widget
Source: Analysis Results (2025)

After the clustering process in **Figure 8**. Settings K-Means Widget above with data grouping into 4 categories cluster, obtained group results (cluster) 1 contains 57 data, group (cluster) 2 contains 46 data, group (cluster) 3 contains 5 data, and group (cluster) 4 contains 3 data. Complete information regarding the membership of each cluster can be seen in **Table 1**. Membership Table of Each Cluster which shows the respective memberships cluster. Then the clustering results are presented using 3 methods, namely Scatter Plot **Figure 9**, Silhouette Plot in **Figure 10**, and Box Plotin **Figures 11, 12,** and **13**.

**Table 1**. Membership Table of Each Cluster

| Customer Cluster | Member Cluster |
|---|---|
| Cluster 1 | 1, 4, 5, 6, 7, 10, 13, 14, 15, 16, 17, 20, 21, 24, 28, 29, 30, 31, 32, 33, 35, 38, 39, 41, 42, 43, 44, 45, 48, 49, 52, 56, 57, 58, 60, 61, 62, 63, 67, 70, 73, 74, 75, 77, 82, 85, 86, 87, 92, 94, 97, 101, 103, 104, 106, 109, 111. **(57)** |
| Cluster 2 | 2, 8, 9, 11, 12, 18, 19, 22, 23, 25, 26, 27, 34, 36, 37, 40, 46, 47, 50, 51, 53, 54, 55, 64, 65, 66, 68, 68, 71, 76, 77, 79, 80, 81, 84, 88, 89, 90, 91, 95, 96, 99, 100, 102, 108, 110. **(46)** |
| Cluster 3 | 3, 83, 93, 105, 107. **(5)** |
| Cluster 4 | 59, 72, 98. **(3)** |

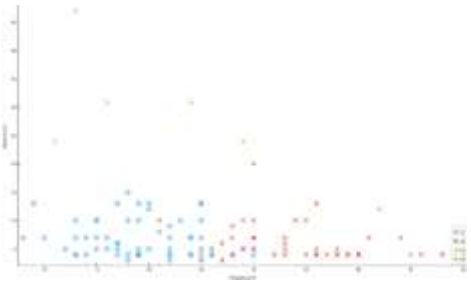


**Fig. 9**: Scatter Plot Results Cluster
Source: Analysis Results (2025)

Based on the visualization results scatter plot clustering using the RFM model in **Figure 9**. Scatter Plot Results Cluster, customer data is grouped into several cluster which represents the characteristics of customer behavior. On the X-axis (Frequency), the data reflects how often customers make purchases, while the Y-axis (Recency) shows the time since the last transaction. The analysis results show that each cluster has a unique pattern. Cluster the first cluster (C1) which is dominated by the blue circle is in the low to medium frequency area and recency low. This shows that customers in this cluster have been actively transacting recently but with a purchase intensity that is not too frequent. The second cluster (C2) represented by the red cross symbol indicates customers with low frequency but have variation recency higher. This indicates that customers in cluster this makes infrequent purchases with a more widely spread transaction time pattern. Meanwhile, the third cluster (C3) represented by the green triangle is seen to be in the high frequency area and recency high. Customers in cluster it has a significant purchasing pattern but has not made any recent transactions for quite some time. This cluster is of particular concern for customer reactivation strategies.

In addition, cluster others such as C4 (symbol "+") show unique distributions and may reflect segments with certain characteristics, such as high loyalty or significant contribution to revenue. The existence of outliers some cluster, such as points that are far from the center cluster, also indicates that there are customers with very different purchasing behavior compared to other customers. This could be an opportunity to develop a special personalization strategy. Overall, the results clustering helps PT XYZ understand customer behavior based on the RFM model and identify key segments. More targeted marketing strategies, such as loyalty programs for cluster with high frequency or reactivation campaigns to cluster with recency high, can be applied to improve sales performance.

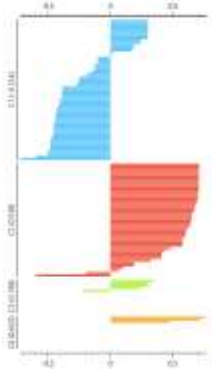


**Fig. 10**: Silhouette Plot Results Cluster
Source: Analysis Results (2025)

The results of the analysis using silhouette plot in Figure 10. Silhouette Plot Results Clustering above shows the quality of separation between cluster in customer segmentation based on RFM model and methods K-Means. Plot this illustrates the value silhouette for each data on each cluster, which reflects the level of internal cohesion cluster and separation between cluster. Cluster1 (C1), which is marked with a blue bar, has a value of silhouette positive which is quite high and consistent. This shows that the members in cluster these are well grouped, with spacing intra-cluster small and distance inter-cluster which is significant to cluster other. This cluster can be interpreted as a group of customers with homogeneous characteristics. Cluster 2 (C2), which is represented by the red bar, also has a value silhouette positive which is quite good. However, there is a greater variation in values compared to C1, indicating that some members cluster has a closer distance to other cluster.

This indicates the possibility of customers in C2 having similar characteristics to customers from other cluster. Meanwhile, cluster smaller ones like C3 (yellow) and C4 (green) have shorter stems and higher values silhouette which varies. This cluster appears to contain customer groups with less data and more unique characteristics. Although they have positive silhouette, it should be noted that the internal cohesion of these clusters may not be as strong as other cluster. Overall, most of the silhouette values are positive indicating that the clustering method has produced a fairly good separation among the customer segments. However, some cluster members with silhouette values close to zero or negative need to be further analyzed as they indicate possible overlap between clusters. This provides an opportunity to re-evaluate the optimal number of clusters or clustering parameters to improve customer segmentation accuracy. These results provide strategic insights for PT XYZ to develop a more targeted approach to customers based on their segments, focusing on improving customer experience and more personalized marketing strategies.

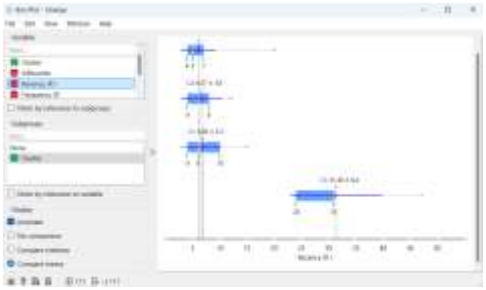


**Fig. 11**: Box Plot of Recency Attribute Results
Source: Analysis Results (2025)

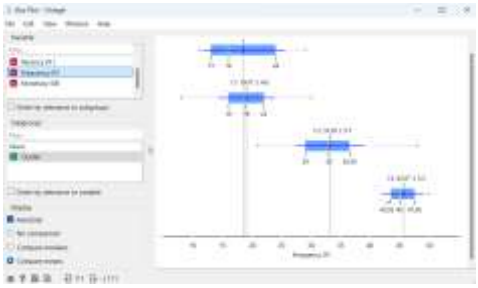


**Fig. 12**: Box Plot of Frequency Attribute Results
Source: Analysis Results (2025)

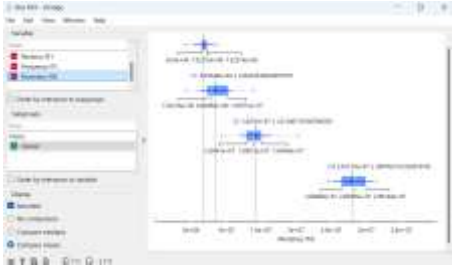


**Fig. 13**: Box Plot of Monetary Attribute Results
Source: Analysis Results (2025)

**Table 2**. Centroid Value of Each Cluster

| Cluster | Recency | Frequency | Monetary | Rank |
|---|---|---|---|---|
| C1 | 6,86 | 19,07 | Rp. 8.556.480 | 3 |
| C2 | 6,00 | 33,20 | Rp. 14.255.000 | 2 |
| C3 | 31,40 | 18,60 | Rp. 6.829.230 | 4 |
| C4 | 6,67 | 45,67 | Rp. 28.121.300 | 1 |

Based on the results listed in **Table 3**, Cluster 4 is the highest ranked customer group, with an average recency value of 6,67; a frequency of 45,67 and a monetary value of Rp. 28.121.300. Cluster 2, which is ranked second, has an average recency variable value of 6,00; a frequency of 33,20 and a monetary value of Rp. 14.255.000. Meanwhile, Cluster 1 which is ranked third, has an average frequency value of 6,86; an average frequency of 19,07 and a monetary value of Rp. 8.556.480. Finally Cluster 3, which is ranked fourth, has an average frequency value of 31,40; an average frequency of 18,60 and a monetary value of Rp. 6.829.230. Clustering with the RFM model results in customer segmentation in each cluster of the 111 Bottled Water sales transaction data analyzed. From the RFM model, customers who fall into the very loyal category are in Cluster 4, with 3 customers. This is due to the fact that they have the highest frequency value (number of transactions) and the largest amount of spending (total). Cluster 1 which consists of 57 customers, is considered a less potential customer as it has a lower RFM value than the customers in Cluster 4, and Cluster 2 which consists of 46 customers, is considered a loyal customer as it has a comparable RFM value to the customers in Cluster 4.

*Pattern Evaluation*

The final stage is evaluation (pattern evaluation), which is the process of evaluating data that has been visualized or modeled to assess the resulting patterns, ensure that the patterns are in accordance with the analysis objectives, and identify the extent to which the patterns provide insight relevant, accurate, and applicable in supporting effective decision making or strategy formulation [23]. In this study, the evaluation of the clustering results was carried out using the method Silhouette Score. The purpose of this examination is to find out whether the number of selected clusters provides the best segmentation results or not. The main principle of the method Silhouette Scores that the higher score value indicates better clustering quality, as it shows that the data in one cluster are more similar to each other and further away from each other.

**Fig. 14**: Evaluation Silhouette Score
Source: Analysis Results (2025)

Researchers conducted analysis on various types of cluster, ranging from two to five. The results of the analysis showed that the group with 4 cluster have value Silhouette Score highest at 0.438, compared to other groups. The results of this evaluation show that the division of data into four cluster resulting in more discrete customer segmentation and higher uniformity within each cluster. Based on these results, the most ideal configuration for the analyzed data is the division cluster into four groups. **Figure 14** above shows the details of the evaluation results using the method Silhouette Score.

## 4. Conclusion

To separate customers based on RFM features, K-Means clustering algorithm was successfully used. Out of 111 data, dividing the data into 4 clusters with the Sillhouette Score method gave the best results. Cluster 4 consists of highly loyal customers with high Monetary and Frequency scores, while Cluster 3 contains customers with lower Monetary and Frequency scores. Cluster 2 includes loyal customers with characteristics similar to Cluster 4, and Cluster 1 consists of customers with lower potential. These clustering results help management design marketing and business development strategies. This research

has several limitations, such as the use of data that only covers the period January to October 2024 and the dependence of K-Means on the selection of the number of clusters. For future research, it can be considered to use other clustering algorithms and extend the data period, including external factors such as promotions or seasons, to enrich the customer segmentation analysis.

## 5. Acknowledgment

The authors would like to thank all those who supported this research. Particularly to the management of PT XYZ for their permission and cooperation in providing the sales data on which this research is based. The authors would also like to thank the administrative team for helping with the facilities and resources needed during the research, which facilitated the implementation of this research. The authors would also like to thank the lecturers who provided essential guidance, direction, and input, as well as the colleagues who provided moral and professional support throughout the process of preparing this journal. All forms of contribution provided are very valuable for the completion of this research.

## 6. References

[1] Z. Setiawan, Y. Zebus, Rony, Sandra, D. Suprayitno, S. Hamid, Rahmad, V. Islami, and A. Marsyaf, *Buku Ajar Perilaku Konsumen*. Jambi: PT. Sonpedia Publishing Indonesia, 2024.

[2] E. F. L. Awalina and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail," *J. Teknol. dan Inf.*, vol. 13, no. 2, pp. 122–137, 2023, doi: 10.34010/jati.v13i2.10090.

[3] A. Perdana, Satria, F. Florentin, Sara, and A. Santoso, "Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift," *Sebatik*, vol. 26, no. 2, pp. 446–457, 2022, doi: 10.46984/sebatik.v26i2.2134.

[4] M. G. H. Fawzi, A. S. Iskandar, H. Erlangga, Nurjaya, and S. Denok, *Strategi Pemasaran*. Tangerang Selatan: Pascal Book, 2021.

[5] E. Yunus, *Manajemen Strategis*. Yogyakarta: CV Andi Offset, 2016.

[6] W. A. Taqwim, N. Y. Setiawan, and A. Bachtiar, Fitra, "Analisis Segmentasi Pelanggan Dengan RFM Model Pada Pt . Arthamas Citra Mandiri Menggunakan Metode Fuzzy C-Means Clustering," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 2, pp. 1986–1993, 2019.

[7] K. Auliasari and M. Kertaningtyas, "Penerapan Algoritma K-Means Untuk Segmentasi Konsumen Menggunakan R," *J. Teknol. Manaj. Inform.*, vol. 5, no. 1, pp. 42–49, 2019, doi: 10.59134/jsk.v5i2.386.

[8] B. E. Adiana, I. Soesanti, and A. E. Permanasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering," *JUTEI*, vol. 2, no. 1, pp. 23–32, 2018, doi: 10.21460/jutei.2017.21.76.

[9] S. Wahyuni, T. T. Wulansari, and F. Fahrullah, "Segmentasi Pelanggan Berdasarkan Analisis Recency, Frequency, Monetary Menggunakan Algoritma K-Means Pada CV. Toedjoe Sinar Group," *J. Rekayasa Teknol. Inf.*, vol. 7, no. 2, pp. 180–187, 2023, doi: 10.30872/jurti.v7i2.8748.

[10] T. D. Nursanti *et al.*, *Entrepreneurship*. Jambi: PT Sonpedia Publishing Indonesia, 2024.

[11] E. Bykova, I. Vasiliev, V. Bosneaga, and V. Suslov, "Application of Neural Networks for Forecasting Energy Security Indicators," *Int. J. Sci. Eng. Res.*, vol. 12, no. 11, pp. 5–11, 2024.

[12] B. M. Wildemuth, *Application of Social Reserch Methods to Questions in Information and Library Science*. California: ABC-CLIO, LLC, 2017.

[13] J. W. Osborne, *Best Practice in Data Cleaning*. California: SAGE Publications, 2013.

[14] A. Hermawan *et al.*, "Optimalisasi Strategi Pemasaran Melalui Analisis RFM pada Dataset Transaksi Ritel Menggunakan Python," *J. Manaj. Ris. Inov.*, vol. 2, no. 4, pp. 254–267, 2024.

[15] M. Reyes, *Consumer Behavior and Marketing*. London: Intech Open, 2020.

[16] N. Trianasari and T. A. Permadi, "Analysis of Product Recommendation Models at Each Fixed Broadband Sales Location Using K-Means, DBSCAN, Hierarchical Clustering, SVM, RF, and ANN," *J. Appl. Data Sci.*, vol. 5, no. 2, pp. 636–652, 2024, doi: 10.47738/jads.v5i2.210.

[17] V. R. Hananto, A. D. Churniawan, and A. P. Wardhanie, "Perancangan Analytical CRM untuk Mendukung Segmentasi Pelanggan di Institusi Pendidikan," *J. Ilm. Teknol. Inf. Asia*, vol. 11, no. 1, pp. 79–88, 2017, doi: 10.32815/jitika.v11i1.55.

[18] K. M. Shaliha, Angelyna, A. A. Nugraha, M. H. Wahisyam, and T. K. Sandi, "Implementasi K-Means Clustering pada Online Retail berdasarkan Recency , Frequency , dan Monetary," *Gunung Djati Conf. Ser.*, vol. 3, no. 1, pp. 99–106, 2021.

[19] M. Rizki, D. Devrika, F. S. Lubis, Silvia, and I. H. Umam, "Aplikasi Data Mining dalam Penentuan Layout Swalayan dengan Menggunakan Metode MBA," *J. Has. Penelit. dan Karya Ilm.*, vol. 5, no. 2, pp. 130–138, 2019.

[20] H. Kusumo, E. Sediyono, and M. Marwata, "Analisis Algoritma Apriori Untuk Mendukung Strategi Promosi Perguruan Tinggi," *Walisongo J. Inf. Technol.*, vol. 1, no. 1, pp. 51–62, 2019.

[21] R. Gustriansyah, N. Suhandi, and F. Antony, "Clustering optimization in RFM Analysis Based on K-Means," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 18, no. 1, pp. 470–477, 2020, doi: 10.11591/ijeecs.v18.i1.pp470-477.

[22] I. Sabuncu, E. Turkan, and H. Polat, "Customer Segmentation and Profiling with RFM," *Turkish J. Mark.*, vol. 5, no. 1, pp. 22–36, 2020.

[23] N. C. Sastya and I. Nugraha, "Penerapan Metode CRISP-DM dalam Menganalisis Data untuk Menentukan Customer Behavior di MeatSolution," *J. Pendidik. dan Apl. Ind.*, vol. 10, no. 2, pp. 103–115, 2023, doi: 10.33592/unistek.v10i2.3079.