Journal Applied of Economics, Mathematics, Computer Science, and Data Science in Business (ECOMMERCE)
Vol. 1 No. 1 (2025)

Customer Segmentation Using the K-Means Clustering Algorithm for Marketing Strategy Design

Muh. Arnesta Arnanda[1], Muhammad Rakib[2] , Valentino Aris[1,*]

[1]Digital Business Study Program, Faculty of Economics and Business, Makassar State University, Makassar, Indonesia
[2]Entrepreneurship Study Program, Faculty of Economics and Business, Makassar State University, Makassar, Indonesia

Abstract

This study aims to segment customers of the cat grooming service at Mocha Petshop using the K-Means Clustering algorithm and the RFM (Recency, Frequency, Monetary) model. Using the Elbow method and Davies Bouldin Index (DBI) evaluation, three optimal clusters were identified with a DBI value of 0.507. The results of the study show that there are three segments identified: Best Customers with high loyalty, Ordinary Customers with moderate loyalty, and Lost Customers with low loyalty. Based on this segmentation, a marketing strategy was designed using the STP (Segmentation, Targeting, Positioning) approach. For Best Customers, the strategy focuses on exclusive services and reward points, while for Ordinary and Lost Customers, the strategy includes education, discounts for first visits after inactivity, and discounts for subsequent services. Increasing transaction frequency is key to improving loyalty and moving customers to the Best Customer segment, which ultimately increases retention or transaction frequency at Mocha Petshop.

Keywords: Customer Segmentation, K-Means Clustering, RFM Model, Elbow Method, Davies Bouldin Index

1. Introduction

The Fourth Industrial Revolution, often termed Industry 4.0, is reshaping the foundations of competitive advantage through pervasive digitization, automation, and data-driven decision-making [1]. The rapid emergence of new industries and business models is disrupting incumbent sectors and reconfiguring global value chains [2]. In this context, information technology acts as a double-edged sword: it enables process efficiencies, innovation, and market expansion, yet it also intensifies rivalry as firms use digital tools to target the same customers in increasingly transparent markets. As a result, product or service quality, though necessary, is no longer sufficient for sustained differentiation [3], [4]. Firms must cultivate robust, trust-based customer relationships supported by integrated customer data, personalized engagement, and responsive service delivery to increase retention and customer lifetime value [5], [6]. To remain competitive, businesses should adopt coherent strategies that align digital investments with clear value propositions, build dynamic capabilities such as rapid experimentation, continuous learning, and agile operations, and implement governance that ensures cybersecurity, data privacy, and ethical use of artificial intelligence. These capabilities enable organizations not only to anticipate and absorb disruptive shocks but also to convert technological change into durable performance gains [7], [8].

Data-driven shifts in information technology and the internet are compelling firms to redesign marketing strategies to sustain performance and growth, with analytics-enabled decision-making now central to competitive advantage [9]. The capacity to collect and analyze behavioural data improves understanding of customer preferences and shopping journeys, enabling more relevant targeting, creative optimization, and offer design. Deeper audience insight should be translated into test-and-learn programs, value-based segmentation, and loyalty mechanics so that relevance scales and margins improve [4][6][8]. Relationship quality and rewards that are perceived as fair and personalized raise retention and customer lifetime value, with recent evidence showing positive performance effects from well-designed loyalty programs across categories and contexts [10], [11]. Yet a pronounced “personalization gap” remains: many brands overestimate how well they personalize relative to consumer perceptions, and the shortfall is most visible in marketing

*Corresponding author:
E-mail address: valentino.aris@unm.ac.id

communications [12]. For example, one recent retail study found that 71% of brands and retailers believe they excel at marketing personalization while only 34% of consumers agree, underscoring the need to better connect data, content, and channel execution [5]. Leaders that close this gap outgrow peers, with recent benchmarks linking advanced personalization to meaningfully higher revenue growth.

Personalized relationships require more than excellent service; they require an individual understanding of each customer, because personalization is now central to a satisfying experience and to building durable relationships [5], [13]. Data-driven insight helps firms recognize needs, tailor content and offers, and coordinate channels at the moment of choice [14]. Customer dissatisfaction has measurable costs for reputation and finance, since satisfaction and loyalty behaviors are systematically linked to retention, word-of-mouth, sales, and firm performance [11], [15]. When customers feel undervalued, they are more likely to switch, while well-designed personalization reliably raises engagement and purchase propensity across digital contexts [16]. Therefore, companies should operationalize personalization through trustworthy data use, continuous testing, and responsive service to sustain loyalty and profits [5], [13].

Personalized customer relationships require more than excellent service; they demand individual-level understanding grounded in data, because personalization is now central to customer experience quality and to building durable relationships [5]. Recent evidence shows strong consumer expectations for tailored interactions, with large-scale studies reporting that most customers expect personalization and become frustrated when they do not receive it, which in turn depresses engagement and purchase behaviour [9]. Empirical research also links effective personalization to higher satisfaction, loyalty, and firm performance, while dissatisfaction carries measurable reputational and financial costs [5], [11]. In practice, brands that connect trustworthy data with content and channels outperform peers, yet a sizable "personalization gap" persists between what brands believe they deliver and what customers perceive. Indonesia shows similarly high stakes: recent findings report that a large majority of consumers will switch or abandon purchases when experiences feel irrelevant, underscoring the need to operationalize real-time, transparent personalization at scale. Companies should systematize this through rigorous testing, privacy-safe data use, and responsive service so that relevance compounds into retention and profit growth.

Based on Twilio's States of Customer Engagement Report, brand personalization can be an effective tool for improving positive consumer behaviour because consumers respond very well to personalization. For example, consumers are more likely to make repeat purchases (48%), recommend the brand to friends and family (46%), click on links from the brand (38%), and leave positive reviews (35%). By personalizing their messages, products, and services, companies can build customer loyalty, encourage repeat purchases, boost word-of-mouth marketing, and increase customer engagement. Interestingly, the State of Customer Engagement Report (2024) mentions that 73% of consumers in Indonesia are willing to spend more money on personalization. This shows that consumers in Indonesia value personalized experiences and are willing to pay for them.

Personalization is important for every customer to achieve marketing accuracy. The aim is to market each product and service appropriately to each customer. One step in personalization is to identify or recognize all customers. One way to identify customers is through customer segmentation. Customer segmentation is one method of customer relationship management (CRM). By conducting customer segmentation, we can group customers based on specific characteristics, such as demographics, behaviour, and preferences. This will help us to better understand the needs and desires of each customer group, so that we can provide more effective personalization. The main objective of this segmentation process is to understand consumer behaviour and implement the right customer retention strategies. Thus, companies can increase profits and achieve long-term success.

Mocha Petshop has been offering grooming services and products for pet cats in Makassar since April 14, 2023. However, amid fierce competition from numerous similar businesses, Mocha Petshop faces two main challenges: competition that demands ways to stand out from competitors, and the use of a manual system for recording grooming service transactions. Currently, transaction recording is done using physical forms for new customers and communication via WhatsApp for customers who have previously transacted, without any customer segmentation or personalization. Marketing is also only conducted through Instagram. Mocha Petshop can enhance its services and customer loyalty by implementing Customer Relationship Management through personalization. Another challenge is retaining profitable customers. Without segmentation and marketing personalization, Mocha Petshop finds it difficult to understand and meet the specific needs of various customer groups. By adopting customer segmentation, Mocha Petshop can more effectively identify customer characteristics and behaviours, enabling the development of targeted marketing strategies. This step is expected to help Mocha Petshop compete better, improve operational efficiency, and retain customers who contribute significantly to revenue.

Data mining enables customer segmentation through the K-Means Clustering method. This grouping helps organizations formulate targeted and effective marketing strategies to reach the right target market [17]. This clustering method determines the desired number of clusters (groups), then groups customers into these clusters based on their similar characteristics [18]. Based on the above issues, this study will examine two main problems, namely, first, identifying customer segments at Mocha Petshop using the K-Means Clustering method and, second, designing an effective marketing strategy for Mocha Petshop based on the results of customer segmentation.

2. Method

This study is a descriptive study with a quantitative approach. quantitative research is based on the philosophy of positivism, in which research is conducted by taking samples from a specific population. The results of the research are presented in the form of numbers obtained through quantitative data collection. Descriptive research, on the other hand, aims to describe the status of human groups, objects, conditions, thoughts, or events that are occurring. This research was conducted at Mocha Petshop, Makassar, and focused on cat grooming service customers. The data used was customer transaction data, which included information such as transaction date, customer ID, type of grooming, total grooming, grooming price, and total price.

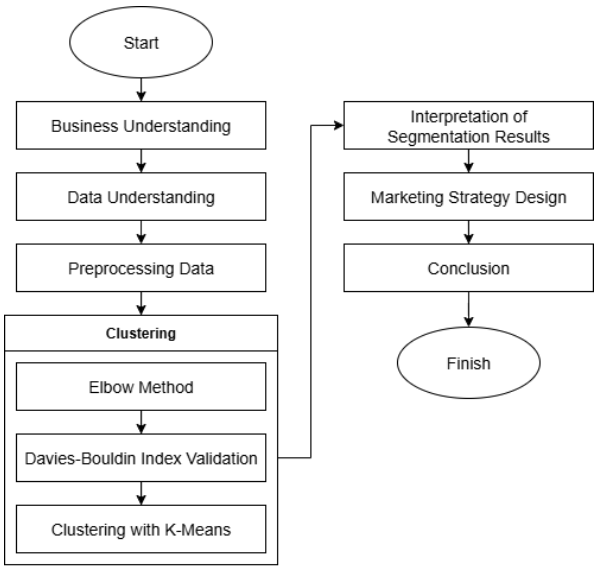

Fig. 1. Research Design

The research design used in this study is as follows First, Business Understanding, this stage focuses on understanding the objectives and problems of Mocha Petshop's business, such as manual grooming service transactions and competition with two nearby stores. In-depth interviews with the owner and staff revealed that the marketing strategy was only through Instagram without personalization. The purpose of the research is to improve marketing effectiveness through customer segmentation. Second, Data Understanding, this stage involves describing and exploring transaction and customer data for grooming services. Initial data analysis helps to understand customer characteristics and behavior, with only relevant and high-quality data selected for further analysis. Next, Preprocessing Data where data is prepared for mining using the RFM (Recency, Frequency, Monetary) model for customer labeling. This process uses Excel for digitization and Python for data selection and scoring, ensuring that the variables used can produce optimal clusters. Fifth, Clustering using the Elbow method is used to determine the optimal number of clusters, followed by validation with the Davies Bouldin Index. The K-Means Clustering algorithm is applied to group customers based on similar characteristics, ensuring that customer segments are well formed. The last, Interpretation of Results and Marketing Strategy, the results of the research are in the form of customer clustering data analysis and marketing strategy recommendations. The recommendations aim to increase revenue and customer satisfaction, as well as support the growth of Mocha Petshop's business amid competition.

Data analysis techniques encompass various methods and approaches designed to process, analyze, and understand data. This process aims to extract valuable information and identify patterns, relationships, and trends contained within

the data. Thus, data analysis helps generate relevant conclusions to answer research questions or support the achievement of predetermined study objectives.

2.1. RFM Analysis (Recency, Frequency, Monetary)

RFM analysis is a method used to categorize customers based on three main aspects, namely Recency, the period of time since the last purchase, the shorter this period, the greater the likelihood of repeat purchases. Frequency, the number of purchases made within a certain period; a higher frequency indicates a greater level of loyalty. Monetary, the amount of money spent during a certain period; a higher value indicates that the company needs to pay more attention to these customers[19], [20].

Table 1. RFM Model Attributes

Attributes	Description
Recency (R)	Recency is the time between the last transaction made by a customer with Mocha Petshop during the analysis period. Recency = Date of Research - Date of Last Transaction Example: If the last transaction date is June 30, 2024, and the analysis period ends on July 1, 2024, then the recency is 1 day. Therefore, the recency value is 1.
Frequency (F)	Frequency measuring the number of transactions made by customers at Mocha Petshop during the analysis period. Example: If customer M003 made 8 transactions at Mocha Petshop during the analysis period, then the frequency value is 8.
Monetary (M)	Monetary refers to the total amount of money spent by customers at Mocha Petshop during the analysis period. Example: If customer M003 spends Rp. 500,000 on transactions at Mocha Petshop during the analysis period, then the monetary value is Rp. 500,000.

The next step is scoring by converting the values of the R, F, and M attributes into scores from 1 to 4, where 4 is the highest score. Scoring is done by dividing the data for each attribute into quantiles, which are four equal groups. For the recency attribute, the smallest value (most recent) receives the highest score, while the frequency and monetary attributes are sorted from largest to smallest to receive the highest score. These scoring results enable Mocha Petshop to identify customer segmentation and design more effective marketing strategies.

2.2. Elbow Method

The elbow method uses the principle of determining the best cluster value based on a significant decrease in the sum of square error (SSE) that forms an elbow-like pattern[21], [22]. Thus, the quality of the resulting clustering becomes more optimal and can maximize the more dominant groups. The percentage difference in results for each cluster value can be visualized through a graph as a source of information. If the first cluster value compared to the second cluster value forms an angle on the graph or shows the most significant decrease, then that cluster value is considered the best.

$$SSE = \sum_{i=1}^{k} \sum_{x_i \in C_i} D(x_i, C_i)^2$$

where:
k = is the number of clusters
$x_i \in C_i$ = is the membership value of data point x_i to cluster center C_i
C_i = is the center of the i-th cluster
$D(x_i, C_i)$ = is the distance from point x_i to the group C_i that is followed

2.3. K-Means Clustering

The K-Means algorithm is a popular clustering technique in unsupervised learning that divides data into clusters based on similarity of characteristics[23]. The goal is to minimize variance within clusters and maximize variance between clusters. Unlike hierarchical methods, K-Means directly separates data into desired clusters by determining initial centroids, then grouping data based on Euclidean distance to the nearest centroid. Designed for numerical data, this

algorithm calculates the distance between data points and groups them based on their similarity. The advantages of K-Means lie in its efficiency in handling large-scale data and its resistance to outliers, making it an ideal choice for complex numerical data analysis[23], [24]. To perform the K-means algorithm, the following basic algorithm is generally used (1) Determine the number of clusters or k value to be used by employing the elbow method, (2) Determine the centroid randomly in the data, and (3) Calculate the distance of the object to the centroid. The distance between each data point and the centroid of each cluster is calculated using the Euclidean Distance formula with the equation below:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

where:
x_i = is the i-th coordinate of data point x
y_i = is the i-th coordinate of data point y
n = is the number of dimensions (number of coordinates)

1) Group each data point into clusters with the nearest centroid (smallest distance).
2) Recalculate the centroid position by taking the average of the data in each cluster using the formula below:

$$C_k = \frac{1}{n_k}\sum d_i$$

where:
n_k = is the amount of data in cluster k
d_i = is the sum of the distance values in each cluster

The K-Means process is repeated until the centroid positions and data clustering are stable. At each iteration, the distance between the data and the centroid is calculated using the Euclidean formula, then the data is grouped into clusters with the nearest centroid. The centroid position is updated based on the average position of the data in the cluster, and this process continues until there is no change in the centroid position or data clustering.

2.4. Davies-Bouldin Index

The Davies-Bouldin Index (DBI) is one of the methods used to measure the validity of grouping in clustering methods. Measurements using the Davies-Bouldin Index aim to maximize the inter-cluster distance while minimizing the intra-cluster distance[25], [26]. Maximum inter-cluster distance indicates that the characteristics between clusters have low similarity, making the differences between clusters more apparent. Meanwhile, minimal intra-cluster distance indicates that each object within a cluster has a high level of similarity in characteristics. Clustering results are obtained from the determination of the centroid, which is then evaluated using the DBI method. A smaller DBI value indicates better cluster quality. This makes it possible to determine the correlation between the centroid determination method based on Sum of Squared Error (SSE) and the improvement in cluster quality as measured by the DBI value. The calculation of the Davies-Bouldin Index (DBI) begins with calculating the sum of squares within clusters (SSW), which is used to determine the cohesion value in cluster i. The formula is as follows:

$$SSW_i = \frac{1}{m_i}\sum_{j=1}^{m_i} d(x_j, c_i)$$

where:
m_i = is the amount of data in cluster i
c_i = is the centroid of the i-th cluster
$d(x_j, c_i)$ = is the distance from data x_j to *centroid* c_i
The next step is to calculate the separation between clusters using the equation below:

$$SSB_{ij} = d(i, j)$$

where:
$d(i, j)$ = is the Euclidean distance between data point i and data point j

Next, calculate the ratio of the SSW and SSB calculations using the equation below:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}}$$

For the final step, use the equation below to obtain the Davies Bouldin Index (DBI) value:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max\, i \neq j\, (R_{i,j})$$

where:

k = is the number of clusters counted

The smaller the DBI value obtained, i.e., a value close to 0 or equal to 0, the better the quality of the clusters from the grouping results. The DBI value will be calculated for K-Means. Then, this DBI value will be used to evaluate the performance of the clustering method.

3. Results and Discussion

3.1. Business Understanding

Mocha Petshop, an SME engaged in cat grooming products and services, is in Makassar City. Mocha Petshop offers a variety of cat supplies and services such as sterilization, vaccination, and various types of cat grooming. From June 2023 to June 2024, Mocha Petshop has served 35 customers for grooming services with varying frequencies, especially increasing on weekends. However, Mocha Petshop faces several significant challenges. Manual transaction systems and marketing only through Instagram without segmentation or personalization, as well as fierce competition with two similar businesses within a 2-kilometer radius and many other competitors in Makassar, hinder efforts to understand and meet specific customer needs. Adopting customer segmentation using the K-Means Clustering method can help Mocha Petshop better identify customer characteristics and behaviors. This step is expected to develop more targeted marketing strategies, increase customer satisfaction, build stronger loyalty, and improve operational efficiency in a competitive market.

3.2. Data Understanding

Mocha Petshop transaction data was originally in physical document form and needed to be digitized into Excel format. The data columns provided by the owner included Date, Customer Name, Grooming Type, Number of Grooming Sessions, and Price per Grooming Type. After reorganization, the data columns became Date, Customer ID, Grooming Type, Number of Grooming Sessions, Price per Grooming Type, and Total Price.

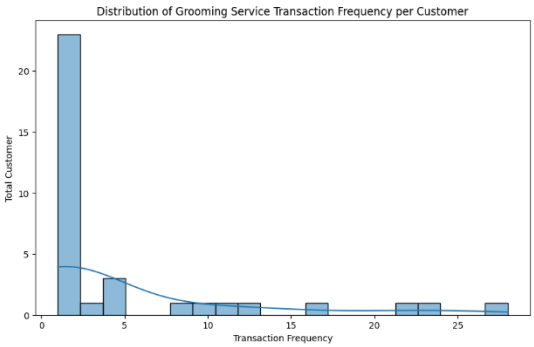

Fig. 2. Grooming Frequency Distribution per Customer

Data analysis shows that most customers only make one grooming transaction, with transaction frequency decreasing dramatically thereafter. Only a few customers make more than five transactions, indicating that most are infrequent users of grooming services or are trying the service for the first time.

3.3. Preprocessing Data

To perform clustering, Mocha Petshop only requires a few key attributes, which are selected using the RFM (Recency, Frequency, Monetary) model to reduce relevant data. The Recency attribute is obtained from the date column, Frequency from the Customer ID column to protect privacy, and Monetary from the Total Price column.

Table 2. Transaction Data in RFM Model Form

Customer ID	Recency	Frequency	Monetary
M001	19	28	4120000
M002	323	2	70000
M003	28	23	1895000
…	…	…	…
M035	153	1	85000

The Recency value is calculated by measuring the distance between the customer's last transaction date and one day after the most recent date in the data. Frequency is calculated based on the number of transactions in a year, and Monetary is calculated from the total customer spending on grooming services in a year. This process is carried out for each customer to obtain the Recency, Frequency, and Monetary values.

Table 3. RFM Scoring

Customer ID	R score	F Score	M Score
M001	4	4	4
M002	1	1	1
M003	3	4	3
…	…	…	…
M035	1	1	1

After the RFM values are obtained, an assessment is carried out using the quantile method (25, 50, 75, 100) with if-else logic in Python. For Recency, a score of 4 is given if the value is small, indicating a recent transaction, and a score of 1 if the value is large. Conversely, for Frequency and Monetary, a score of 4 is given if the value is large, indicating frequent transactions and large expenditures, and a score of 1 if the value is small. This process is repeated to obtain the RFM score for each Mocha Petshop customer.

3.4. Clustering

Modeling using data mining techniques, specifically clustering with the K-Means algorithm, was applied to 173 Mocha Petshop customer transaction data. K-Means Clustering grouped customers based on similar characteristics, using RFM (Recency, Frequency, Monetary) model attribute scores. Before clustering, the elbow method was used to determine the optimal k value or number of clusters, which was k=3, resulting in three customer segments.

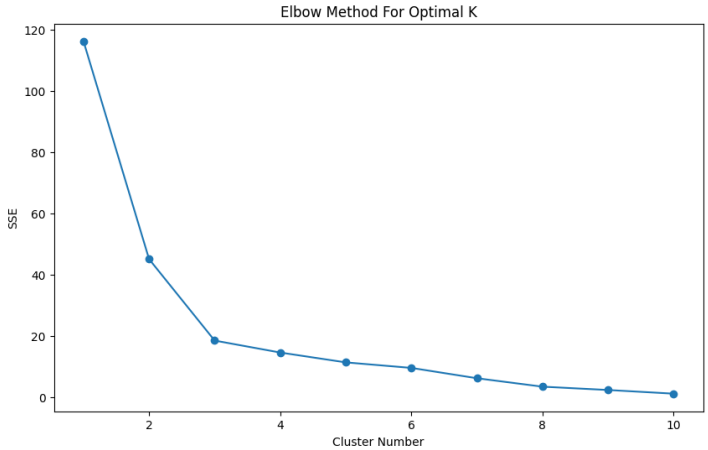

Fig. 3. Elbow Method

The elbow method determines the optimal number of clusters by calculating the Sum of Squared Errors (SSE) for various values of k. The SSE decreases rapidly at first and slows down after a certain point, forming an “elbow” on the graph. In the case of Mocha Petshop, k=3 was chosen because the decrease in SSE began to slow down at this point, resulting in three efficient customer segments without adding complexity.

Table 4. Davies-Bouldin Index Evaluation Results

Number of Clusters	*Davies-Bouldin Index*
2	0.727
3	0.507

Table 4 shows The Davies–Bouldin Index to evaluates cluster quality on the principle that lower values indicate tighter, better-separated groups; in your results, k = 2 yields a DBI of 0.727, suggesting at least one cluster remains diffuse or

close to its neighbor, while k = 3 reduces the DBI to 0.507, an improvement of roughly thirty percent in the compactness-versus-separation trade-off. This clear drop implies that splitting the population into three clusters resolves previously mixed behavior into more homogeneous and distinct segments, making the profiles more stable and interpretable. Practically, this supports selecting three clusters as the optimal solution for customer segmentation and using them to tailor strategies—for example, protect-and-grow for the highest-value segment, repeat-purchase activation for the middle segment, and win-back or low-cost outreach for the at-risk segment.

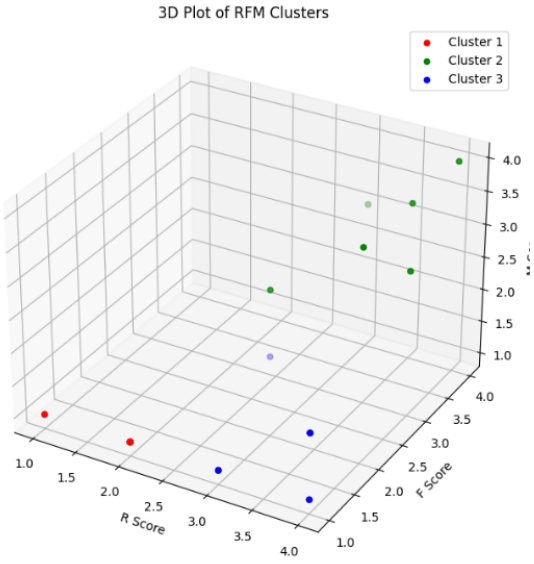

Fig. 4. Clustering Results with 3D Plot

The 3D plot visualization shows the clustering results with 3 clusters, based on RFM attribute scores. This cluster consist of three groups: red (Cluster 1), green (Cluster 2), and blue (Cluster 3). The green points concentrate at higher Frequency and higher Monetary with medium-to-high Recency, indicating loyal, high-value customers who purchase often and spend more; the blue points have medium-to-high Recency but lower Frequency and Monetary, suggesting newer or recently active customers who have not yet built purchasing depth; and the red points lie at lower Frequency and Monetary with lower Recency, representing lapsed or at-risk customers with infrequent, low-value activity. The clear separation among colors implies that RFM features effectively distinguish behavior, making these clusters actionable for targeted strategies such as protect-and-grow for green, nurture and repeat-purchase activation for blue, and win-back campaigns for red.

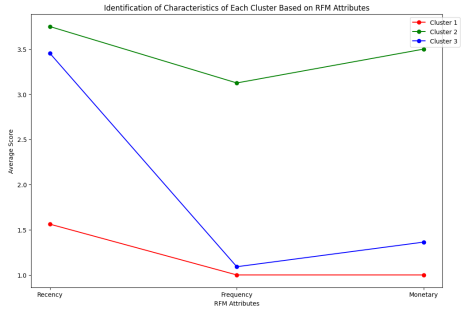

Fig. 5. Identify the Characteristics of Each Cluster

The line plot profiles three clusters by their average RFM scores, letting you compare how recently they buy, how often they buy, and how much they spend. The green line (Cluster 2) stays consistently high—its Recency is near the top of the scale and, although Frequency dips relative to its own Recency and Monetary, it remains above the other clusters, Monetary rises again to the highest level. This is your "high-value" segment: customers purchased recently, buy relatively often, and spend the most. The red line (Cluster 1) sits low across all attributes, with modest Recency and the lowest Frequency and Monetary, indicating lapsed or low-engagement customers whose purchases are infrequent and low value. The blue line (Cluster 3) starts high on Recency, drops sharply on Frequency, then recovers slightly on Monetary, suggesting newly acquired or recently reactivated customers who have shown interest or made a recent

purchase but have not yet built repeat behavior or consistent spend. Read left-to-right, the vertical gaps between lines quantify how segments differ by attribute: the largest gap appears at Monetary (green > blue ≥ red), confirming where value concentrates; the steep decline in the blue line from Recency to Frequency highlights a conversion opportunity from first to repeat purchase; and the consistently low red line flags a win-back challenge. In practice, prioritize "protect and grow" tactics for Cluster 2 (VIP benefits, tailored bundles), onboarding and repeat-activation programs for Cluster 3 (timely reminders, cross-sell), and reactivation or cost-controlled outreach for Cluster 1 (limited-time offers, feedback prompts).

3.5. Interpretation of Segmentation Results

The average RFM score is used to determine the potential ranking of each cluster. With clusterization line plots, RFM attributes are ranked to assess the potential of each cluster.

Table 5. RFM Attribute Rank in Each Cluster

Rank	Cluster	Symbol Rank	Label	Potentiall
1	2	R ↑ F ↑ M ↑	*Best Customer*	High
2	3	R ↑ F ↓ M ↓	*Ordinary Customer*	Medium
3	1	R ↓ F ↓ M ↓	*Lost Customer*	Low

Customers are analyzed based on their RFM attribute ratings. Cluster 2, with high ratings on all attributes, is called "Best Customer." Cluster 3, with high ratings on R and low ratings on F and M, is called "Ordinary Customer." Cluster 1, with low ratings on all attributes, is called "Lost Customer.".

Table 6. Characteristics of Each Cluster

Cluster	Characteristics	Label
2	1. Customers in cluster 2 are customers with a very high level. 2. Customers who last made a transaction within the last 16-76 days. 3. Customers who have made 10-28 purchases. 4. Customers with a transaction amount above Rp 2,025,000.	*Best Customer*
3	1. Customers who last made a transaction within the past 16-76 days. 2. Customers who have made 1-10 purchases. 3. Customers who only have transactions totaling Rp. 680,000 or less. 4. Customers with average loyalty levels or new customers in 2024.	*Ordinary Customer*
1	1. Customers who last made a transaction between 125 and 323 days ago. 2. Customers who have made a maximum of 4 purchases. 3. Customers who have only made transactions totaling Rp. 170,000 or less. 4. Customers who have never made a transaction at Mocha Petshop again.	*Lost Customer*

The customer segmentation table groups buyers into three actionable clusters—Best Customers with high transactions and loyalty, Ordinary Customers with moderate activity who need retention tactics, and Lost Customers with infrequent purchases and elevated churn risk—and the study operationalizes these insights through the classic Segmentation, Targeting, and Positioning framework to craft strategy for Mocha Petshop [27]. Using K-Means is appropriate for this task because it partitions numerical RFM features into behaviorally coherent groups at scale and is widely validated in recent research[20], [28]. Consistent with STP logic, Targeting emphasizes the high-value cluster, here cluster 2, to protect and grow revenue, while Positioning and programs for the Ordinary segment aim to raise purchase frequency and spend, and reactivation programs for the Lost segment focus on improving all RFM attributes and reducing churn[11], [15].

In the positioning stage, Mocha Petshop's marketing strategy will focus on cluster 2 (Best Customers) by promoting itself as the best and most exclusive grooming service provider. Customers in this segment will receive special offers, exclusive discounts, priority access, and highly personalized services. This is expected to maintain their loyalty. Other segments, such as "Ordinary Customers" and "Lost Customers," will also be given attention to improve their RFM attributes. Marketing strategies for cluster 2 (Best Customers) include exclusive services such as priority access to grooming slots, reward points, and flexible reservations. For cluster 3 (Ordinary Customers) and cluster 1 (Lost Customers), the strategy focuses on educating them about the importance of routine care, reward point programs, and significant discounts. Increasing retention or transaction frequency is considered key to improving customer loyalty, as increased transaction frequency will also improve the Recency and Monetary attributes, ultimately turning customers into "Best Customers".

4. Conclusion

This study identifies three actionable customer segments at Mocha Petshop using RFM-based K-Means clustering and validates the three-cluster solution with internal metrics, then maps the segments to an STP strategy. Best Customers exhibit high loyalty and value; the recommended focus is protecting and growing this group through exclusive services, priority booking, and reward points that reinforce advocacy and repeat purchases. Ordinary Customers show medium loyalty; the strategy emphasizes frequency-building tactics such as education on routine care, personalized reminders, and stepwise incentives to raise visit cadence and basket size. Lost Customers have low loyalty and elevated churn risk; reactivation is pursued with welcome-back offers, discounts on the first visit after inactivity, and follow-on grooming discounts. The overarching lever is increasing transaction frequency so that improvements in R, F, and M move customers into the Best Customer tier, lifting retention and total revenue.

The research has several limitations. Results come from a single setting and a cross-sectional snapshot, so findings may not generalize across time, locations, or product mixes. Feature space is restricted to RFM, excluding variables such as service category, channel, demographics, or promotion exposure; K-Means also assumes roughly spherical clusters in Euclidean space and can be sensitive to scaling and outliers. Model validation relied primarily on internal indices, limiting causal interpretation of marketing effects. Future work should incorporate multi-period data, test alternative algorithms and initializations, and triangulate validity with Silhouette and Calinski–Harabasz scores and stability checks across seeds. Most importantly, run controlled field experiments or A/B tests on the proposed offers to estimate incremental lift and profitability by segment, and extend the feature set to include service usage, campaign history, and predicted churn or lifetime value to refine targeting and budgeting.

Acknowledgements: The authors would like to express sincere gratitude to the Digital Business Study Program, Faculty of Economics and Business, Universitas Negeri Makassar, for its continuous support and encouragement throughout the development of this research—from its initial conception to the completion of this article.

References

[1] M. Z. Yaqub and A. Alsabban, "Industry-4.0-Enabled Digital Transformation: Prospects, Instruments, Challenges, and Implications for Business Strategies," *Sustain.*, vol. 15, no. 11, 2023, doi: 10.3390/su15118553.

[2] M. M. H. Emon and T. Khan, "The transformative role of Industry 4.0 in supply chains: Exploring digital integration and innovation in the manufacturing enterprises," *J. Open Innov. Technol. Mark. Complex.*, vol. 11, no. 2, 2025, doi: 10.1016/j.joitmc.2025.100516.

[3] L. Liang and F. Ye, "Information transparency and business mode selection in online platforms under manufacturer competition," *J. Bus. Res.*, vol. 170, p. 114324, 2024, doi: https://doi.org/10.1016/j.jbusres.2023.114324.

[4] P. Liu, F. Zhang, Y. Liu, S. Liu, and C. Huo, "Enabling or burdening?—The double-edged sword impact of digital transformation on employee resilience," *Comput. Human Behav.*, vol. 157, p. 108220, 2024, doi: https://doi.org/10.1016/j.chb.2024.108220.

[5] T. Teepapal, "AI-driven personalization: Unraveling consumer perceptions in social media engagement," *Comput. Human Behav.*, vol. 165, p. 108549, 2025, doi: https://doi.org/10.1016/j.chb.2024.108549.

[6] M. de S. Pereira *et al.*, "Factors of Customer Loyalty and Retention in the Digital Environment," *J. Theor. Appl. Electron. Commer. Res.* , vol. 20, no. 2, pp. 1–21, 2025, doi: 10.3390/jtaer20020071.

[7] J. Mökander, M. Sheth, M. Gersbro-Sundler, P. Blomgren, and L. Floridi, "Challenges and best practices in corporate AI governance: Lessons from the biopharmaceutical industry," *Front. Comput. Sci.*, vol. 4, 2022, doi: 10.3389/fcomp.2022.1068361.

[8] D. Ellström, J. Holtström, E. Berg, and C. Josefsson, "Dynamic capabilities for digital transformation," *J. Strateg. Manag.*, vol. 15, no. 2, pp. 272–286, 2022, doi: 10.1108/JSMA-04-2021-0089.

[9] N. Arora *et al.*, "The value of getting personalization right—or wrong—is multiplying," mckinsey.com. Accessed: Jul. 17, 2025. [Online]. Available: https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-value-of-getting-personalization-right-or-wrong-is-multiplying?utm_source=chatgpt.com#/

[10] C. Lin and D. Bowman, "The impact of introducing a customer loyalty program on category sales and profitability," *J. Retail. Consum. Serv.*, vol. 64, p. 102769, 2022, doi: https://doi.org/10.1016/j.jretconser.2021.102769.

[11] V. Mittal *et al.*, "Customer satisfaction, loyalty behaviors, and firm financial performance: what 40 years of

research tells us," *Mark. Lett.*, vol. 34, no. 2, pp. 171–187, 2023, doi: 10.1007/s11002-023-09671-w.
[12] Sailthru and Coresight Research, "Retail Personalization in 2022: Balancing Trust , Data Collection and Privacy," 2022.
[13] J. Weidig, M. Weippert, and C. Kuehnl, "Personalized touchpoints and customer experience: A conceptual synthesis," *J. Bus. Res.*, vol. 177, p. 114641, 2024, doi: https://doi.org/10.1016/j.jbusres.2024.114641.
[14] P. Kotler, H. Kartajaya, and I. Setiawan, *Marketing 5.0: Technology for Humanity*. Wiley, 2021. [Online]. Available: https://books.google.co.id/books?id=ANfzyQEACAAJ
[15] M. Guenther and P. Guenther, "The complex firm financial effects of customer satisfaction improvements," *Int. J. Res. Mark.*, vol. 38, no. 3, pp. 639–662, 2021, doi: https://doi.org/10.1016/j.ijresmar.2020.10.003.
[16] J. Yin, X. Qiu, and Y. Wang, "The Impact of AI-Personalized Recommendations on Clicking Intentions: Evidence from Chinese E-Commerce," *J. Theor. Appl. Electron. Commer. Res.* , vol. 20, no. 1, 2025, doi: 10.3390/jtaer20010021.
[17] S. N. Lathifah and Z. F. Azzahra, "AI-Driven Customers Segmentation Using K-Means Clustering," *G-Tech J. Teknol. Terap.*, vol. 9, no. 1, pp. 320–329, 2025, doi: 10.70609/gtech.v9i1.6202.
[18] S. Wahyuni and F. Fahrullah, "Segmentasi Pelanggan Berdasarkan Analisis Recency, Frequency, Monetary Menggunakan Algoritma K-Means," *J. Simantec*, vol. 12, no. 1, pp. 29–36, 2023, doi: 10.21107/simantec.v12i1.23164.
[19] F. M. Talaat, A. Aljadani, B. Alharthi, M. A. Farsi, M. Badawy, and M. Elhosseini, "A Mathematical Model for Customer Segmentation Leveraging Deep Learning, Explainable AI, and RFM Analysis in Targeted Marketing," *Mathematics*, vol. 11, no. 18, 2023, doi: 10.3390/math11183930.
[20] M. A. Rahim, M. Mushafiq, S. Khan, and Z. A. Arain, "RFM-based repurchase behavior for customer classification and segmentation," *J. Retail. Consum. Serv.*, vol. 61, p. 102566, 2021, doi: https://doi.org/10.1016/j.jretconser.2021.102566.
[21] T. Rahmawati, Y. Wilandari, and P. Kartikasari, "Analisis Perbandingan Silhouette Coefficient Dan Metode Elbow Pada Pengelompokkan Provinsi Di Indonesia Berdasarkan Indikator Ipm Dengan K-Medoids," *J. Gaussian*, vol. 13, no. 1, pp. 13–24, 2024, doi: 10.14710/j.gauss.13.1.13-24.
[22] I. Herdiana, M. A. Kamal, Triyani, M. N. Estri, and Renny, "A More Precise Elbow Method for Optimum K-means Clustering," pp. 1–22, 2025, [Online]. Available: http://arxiv.org/abs/2502.00851
[23] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (Ny).*, vol. 622, pp. 178–210, 2023, doi: https://doi.org/10.1016/j.ins.2022.11.139.
[24] B. Brejna, M. Pietranik, and A. Kozierkiewicz, "The New K-Means Initialization Method," in *Computational Collective Intelligence*, N. T. Nguyen, B. Franczyk, A. Ludwig, M. Núñez, J. Treur, G. Vossen, and A. Kozierkiewicz, Eds., Cham: Springer Nature Switzerland, 2024, pp. 372–381.
[25] F. Ros, R. Riad, and S. Guillaume, "PDBI: A partitioning Davies-Bouldin index for clustering evaluation," *Neurocomputing*, vol. 528, pp. 178–199, 2023, doi: https://doi.org/10.1016/j.neucom.2023.01.043.
[26] M. D. Kartikasari, "Self-Organizing Map Menggunakan Davies-Bouldin Index dalam Pengelompokan Wilayah Indonesia Berdasarkan Konsumsi Pangan," *Jambura J. Math.*, vol. 3, no. 2, pp. 187–196, 2021, doi: 10.34312/jjom.v3i2.10942.
[27] P. Kotler, K. L. Keller, and A. Chernev, *Marketing Management*. in Always learning. Pearson, 2022. [Online]. Available: https://books.google.co.id/books?id=_-2hDwAAQBAJ
[28] H. Abbasimehr and A. Bahrini, "An analytical framework based on the recency, frequency, and monetary model and time series clustering techniques for dynamic segmentation," *Expert Syst. Appl.*, vol. 192, p. 116373, 2022, doi: https://doi.org/10.1016/j.eswa.2021.116373.