JRSSEM 2026, Vol. 5, no. 12, July
E-ISSN: 2807 - 6311
P-ISSN: 2807 - 6494

# RFM-Based Customer Segmentation Using K-Means Clustering to Recommend Marketing Strategies for Motorcycle Repair Shops

**Nanda Oberusti*, Sigit Anggoro, Tacbir Hendro Pudjiantoro**
Universitas Jenderal Achmad Yani, Indonesia
Email: ndo.22@si.unjani.ac.id*, sigit.anggoro@lecture.unjani.ac.id, tacbir.hendro@lecture.unjani.ac.id

**Keywords:**
Customer Segmentation; RFM; K-Means Clustering.

**Abstract**

Motorcycle repair workshops, as service-based businesses, require an understanding of customer characteristics to develop more effective marketing strategies. However, differences in customer transaction behavior make generalized marketing strategies less effective in meeting the needs of different customer groups. Customer segmentation is an approach that enables businesses to classify customers based on their transaction characteristics, allowing the identification of valuable customers and the development of more targeted marketing strategies. This study aims to apply the **Recency,** Frequency, Monetary (RFM) method and the K-Means Clustering algorithm to segment customers of AHASS Cimalaka Sumedang Motorcycle Workshop. The resulting customer segments are expected to support the formulation of marketing strategy recommendations, improve customer loyalty, and facilitate more effective marketing decision-making. Differences in customer transaction behavior reduce the effectiveness of generalized marketing strategies, highlighting the need for customer segmentation based on transaction data. To apply the Recency, Frequency, Monetary (RFM) method and the K-Means Clustering algorithm for customer segmentation and to develop marketing strategy recommendations. This study employed a quantitative approach consisting of data preprocessing, RFM calculation, data normalization, K-Means clustering, and cluster evaluation. The proposed approach successfully segmented customers into **five clusters** based on their RFM characteristics, namely **Potential Loyal Customer, At Risk Customer, Lost Customer, Inactive Customer, and Loyal Customer**, which served as the basis for prioritizing targeted marketing strategies. The combination of the RFM method and the K-Means Clustering algorithm effectively supports customer segmentation and provides a reliable basis for developing more targeted and effective marketing strategies.

## INTRODUCTION

The motor vehicle maintenance and repair services industry continues to grow along with the increasing number of motor vehicles in Indonesia. Based on data, the number of motor vehicles will increase from 166.46 million units in 2024 to 172.93 million units in 2025, so that the public's need for vehicle maintenance services is also increasing. This condition encourages competition between workshop service providers so that companies are not only required to acquire new customers, but also retain existing customers. Customer retention is an important factor in creating a competitive advantage, while an understanding of customer behavior is necessary to support more effective marketing decision-making (Akande et al., 2024; Badan Pusat Statistik, 2025; Pereira et al., 2025).

Customers are an important asset for business sustainability, but not all customers have the same characteristics and contribution to the company's revenue (Aldossary et al., 2024; De Mendonca & Zhou, 2019). Based on the results of the pre-research on the transaction data of the AHASS Cimalaka Sumedang Motorcycle Workshop for the period January 2024 to December 2025, out of 10,041 customers, as many as 5,538 customers (55.15%) only made one transaction, while 4,503 customers (44.85%) made at least two repeat transactions. This condition shows that the customer retention rate still needs to be improved and indicates that there are differences in customer behavior characteristics that need to be understood so that companies can develop a more targeted marketing strategy (Handojo et al., 2023).

One approach that can be used to understand customer characteristics is *customer segmentation* (Akinrinoye et al., 2020; Tabianan et al., 2022). The *Recency, Frequency, Monetary* (RFM) method is able to identify customer characteristics based on the last transaction time, transaction frequency, and transaction value, while the *K-Means Clustering* algorithm is used to group customers who have similar characteristics (Lulaj, 2023). Various studies have shown that the combination of RFM and K-Means Clustering methods is effectively used for customer segmentation in the retail, e-commerce, and various digital-based services sectors. Using the RFM method to identify customer value in the retail sector, while combining RFM and K-Means for e-commerce customer segmentation (Alves Gomes & Meisen, 2023). In the service sector, RFM and K-Means are applied to online transportation services to identify customer groups based on service usage patterns. The interpretation of cluster characteristics is generally only used to describe customer behavior, while its use as the basis for developing a prioritized marketing strategy based on the business value of each segment is still relatively limited (Chen & Gunawan, 2023; Perdhana & Heikal, 2024; Rahim et al., 2021).

Based on these conditions, this study not only applies the RFM method and K-Means Clustering algorithm to the motorcycle repair shop service industry but also interprets the characteristics of each cluster into marketing strategy recommendations that are prioritized according to the business value of each segment. The urgency of this research is underscored by the fact that Micro, Small, and Medium Enterprises (MSMEs) contribute more than 61% of Indonesia's Gross Domestic Product, yet most of them still face limitations in leveraging transactional data for customer retention strategies . The novelty of this research lies in the development of a framework that integrates RFM and K-Means segmentation with value-based marketing strategy prioritization, providing a more comprehensive and actionable approach for service-based businesses, particularly in the motorcycle repair shop industry. This approach is expected to produce more applicable information in supporting data-based marketing decision-making in service companies.

## METHOD

This study used a quantitative approach by applying the *Data Mining* to segment customers based on historical data on service transactions at the AHASS Cimalaka Sumedang Motorcycle Workshop. The data used is in the form of customer transaction data which includes customer identity, transaction date, transaction amount, and transaction value. The stages of the research carried out include:

1. Collection of historical data of customer transactions.

2. Data preprocessing, including *data cleaning*, deletion of duplicate data, incomplete data handling, and data transformation to be ready for analysis.
3. Recency , *Frequency*, *Monetary* (RFM) value calculation as a representation of customer characteristics.
4. RFM data normalization uses *Min-Max Scaling* to equalize the scale between variables.
5. Determination of the optimal number of *clusters* using *the Elbow Method*.
6. Customer segmentation uses *the K-Means Clustering* algorithm.
7. Evaluation of *clustering* results using *Sum of Squared Errors* (SSE) and *Silhouette Score*.
8. Interpretation of the characteristics of each *cluster*.
9. Preparation of marketing strategy recommendations based on customer segmentation results.
10. Drawing conclusions and compiling research reports.

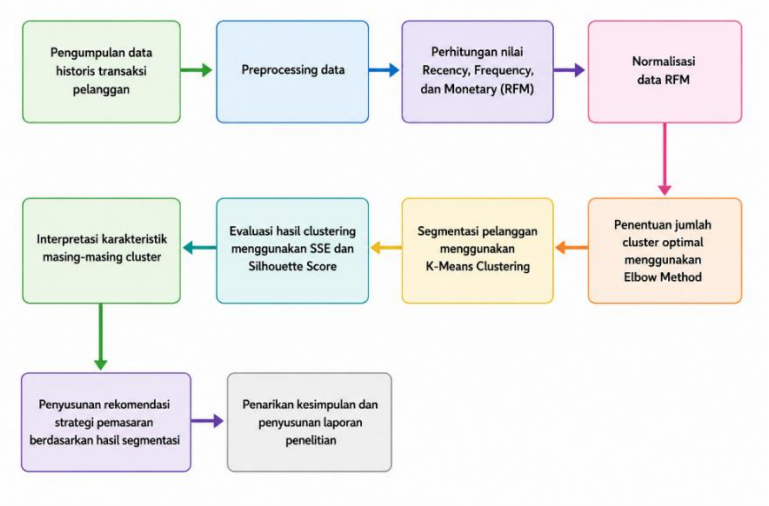


**Figure 1.** Research Stages

The obtained RFM value is used as input to the algorithm *K-Means Clustering* to segment customers based on the similarity of transaction characteristics. Quantity *cluster* Optimal is determined using *Elbow Method*, while the quality of the results *Clustering* evaluated using *Silhouette Score*. The final stage of the research is to interpret the characteristics of each *cluster* and develop marketing strategy recommendations tailored to customer segmentation results. The overall stages of the research are shown in Figure 1.

## RESULTS AND DISCUSSION

In the data collection process, 23,912 customer transaction data were obtained at the AHASS Cimalaka Sumedang Motorcycle Workshop during the period from January 2024 to December 2025 with 10,042 unique customers. The data used includes the attributes of NO_MESIN, TGL_SERVICE, and AMOUNT as the basis for calculating *the Recency, Frequency, and Monetary* (RFM) values. In the *preprocessing stage*, 42 duplicate transaction data were found which were then deleted as well as 2 data with invalid date format that were removed from the analysis process. After data cleaning, 23,868 transactions were obtained

from 10,041 customers, all of which had valid data formats and were ready to be used for RFM calculation and segmentation processes using *the K-Means Clustering* algorithm.

**Table 1.** Raw Data

| NO_MESIN | NOPOL | NAMA_PEMBAWA | NO_HP | NO_TELP | TGL_SERVICE | JENIS_SERVICE | AMOUNT |
|---|---|---|---|---|---|---|---|
| HOND4 00029XX | Z 2980 CM | asep | 1234 | 1234 | 04-08-2024 | SERVIS PAKET LENGKAP PERAWATAN PLUS | 75000 |
| HOND4 00065XX | Z 6539 CZ | evi | 87754657567 | 87754657567 | 14-08-2024 | SERVIS PAKET LENGKAP PERAWATAN | 75000 |
| HOND4 00065XX | Z 6539 CZ | evi | 87754657567 | 87754657567 | 15-08-2024 | SERVIS PAKET LENGKAP PERAWATAN PLUS | 75000 |
| HONDA 11143XX | Z 4367 CL | nendi | 87823990801 | 87823990801 | 03-08-2024 | SERVIS PAKET LENGKAP PERAWATAN | 80000 |
| HB21E 12766XX | Z 3768 AE | dodi | 0 | 0 | 08-08-2024 | PAKET SERVIS PUBLIC AREA | 65000 |
| JB51E 14236XX | Z 3456 CN | dewi maryati | 852 | 852 | 12-08-2024 | SERVIS PAKET LENGKAP PERAWATAN | 65000 |
| JB91E 14037XX | Z 2792 AS | ATEP SURYATNA B | 1234 | 1234 | 23-08-2024 | PAKET SERVIS PUBLIC AREA | 65000 |
| JB91E 23542XX | Z 4177 BB | RUBAYA DRS | 8122123811 | 8122123811 | 24-08-2024 | PAKET SERVIS PUBLIC AREA | 65000 |
| JB91E 24611XX | Z 5733 BC | SUKARYADI | 81394810786 | 81394810786 | 24-08-2024 | SERVIS PAKET LENGKAP PERAWATAN | 65000 |
| JBC1E 12912XX | Z 3174 AV | KRISTIAN ADI SUSILO | 1234 | 1234 | 20-08-2024 | PAKET SERVIS PUBLIC AREA | 65000 |

Value *Recency*, *Frequency*, and *Monetary* It is calculated for each customer based on the machine number (NO_MESIN). Variable *Recency* shows the time difference in months between the customer's last transaction and the research reference date, *Frequency* declare the number of service transactions during the study period, while *Monetary* is the accumulated value of transactions issued by the customer. The results of the calculation produced one row of data for each customer so that as many as 10,041 RFM data were obtained which were then used as input in the clustering process.

**Table 2.** RFM Calculation Results

| No. | NO_MESIN | *Recency* | *Frequency* | *Monetary* |
|---|---|---|---|---|
| **1.** | EF21E 10005XX | 14 | 2 | 0 |
| **2.** | F1234 45632XX | 15 | 2 | 238500 |
| **3.** | H0ND4 00025XX | 5 | 1 | 75000 |
| **4.** | H0ND4 00029XX | 17 | 1 | 75000 |
| **5.** | H0ND4 00029XX | 2 | 1 | 50000 |
| **6.** | H0ND4 00031XX | 7 | 1 | 85000 |
| **7.** | H0ND4 00032XX | 19 | 2 | 209500 |
| **8.** | H0ND4 00032XX | 4 | 3 | 225000 |
| **9.** | H0ND4 00035XX | 15 | 1 | 75000 |
| **10.** | H0ND4 00038XX | 19 | 1 | 129500 |

The histogram in Figure 3 shows that the three variables have different distribution characteristics. Variable *Recency* have a relatively spread distribution, although customers with a *Acknowledgments* low still dominates. This shows that some customers are still making transactions in a period relatively close to the research reference date.

The *Frequency distribution shows a* right-skewed *pattern*, where most customers only make one to two transactions, while only a small percentage of customers have a high transaction frequency. A similar pattern is also seen in the *Monetary* variable, where most customers have a relatively low total transaction value, while only a few customers contribute a very large transaction value.

The variation in distribution in the three variables shows that there are differences in customer behavior characteristics so that a segmentation method is needed to group customers with similar characteristics.

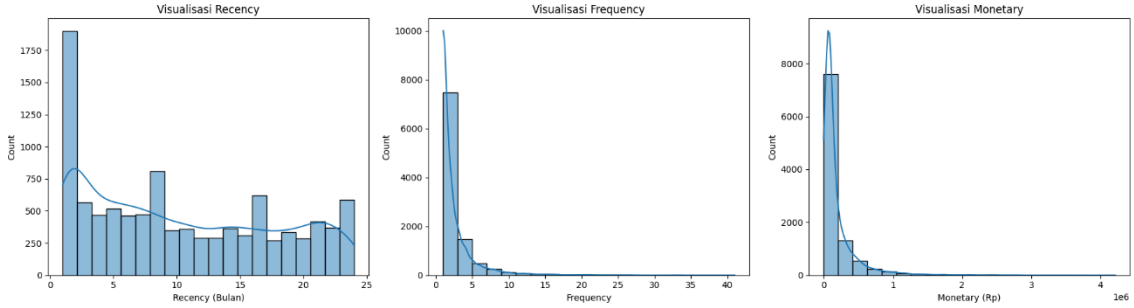


**Figure 2.** Recency, Frequency, and Monetary Visualization

The variation of the distribution on the three variables indicates that each attribute has different characteristics and value ranges. Variable *Monetary* has a much larger scale than *Recency* and *Frequency*, thus potentially dominating distance calculations in the K-Means algorithm. Therefore, before the clustering process is carried out, all variables are normalized using the Min-Max Scaling method so that they have the same value range, which is 0 to 1.

**Table 3.** RFM Data Normalization Results

| NO_MESIN | *Recency* | *Frequency* | *Monetary* |
|---|---|---|---|
| EF21E 10005XX | 0.553571 | 0.025 | 0.000000 |
| F1234 45632XX | 0.594780 | 0.025 | 0.056644 |
| H0ND4 00025XX | 0.196429 | 0.000 | 0.017813 |
| H0ND4 00029XX | 0.704670 | 0.000 | 0.017813 |
| H0ND4 00029XX | 0.060440 | 0.000 | 0.011875 |

Based on Table 3, the entire value of *Recency*, *Frequency*, and *Monetary* successfully transformed into a range of 0–1 without changing the relative relationship between the data. This normalization process ensures that each variable makes a balanced contribution to the calculation of Euclidean distances in the K-Means algorithm so that the segmentation results are not dominated by the variable with the largest value scale.

The value of Sum of Squared Errors (SSE) shown in Table 4 decreases as the number of clusters increases, from **1043,972** on **k = 1** become **41,717** on **k = 10**. The biggest decline

occurred in the number of initial clusters, then smaller as the value increased. **k**. This shows that the increase in the number of clusters is able to increase the homogeneity of the data within the cluster, but the increase becomes smaller after a certain number of clusters.

**Table 4.** SSE Value

| Number of *Clusters* (k) | SSE Value |
|---|---|
| 1 | 1043,972 |
| 2 | 304,420 |
| 3 | 185,396 |
| 4 | 135,185 |
| 5 | 92,349 |
| 6 | 74,902 |
| 7 | 60,837 |
| 8 | 52,293 |
| 9 | 46,619 |
| 10 | 41,717 |

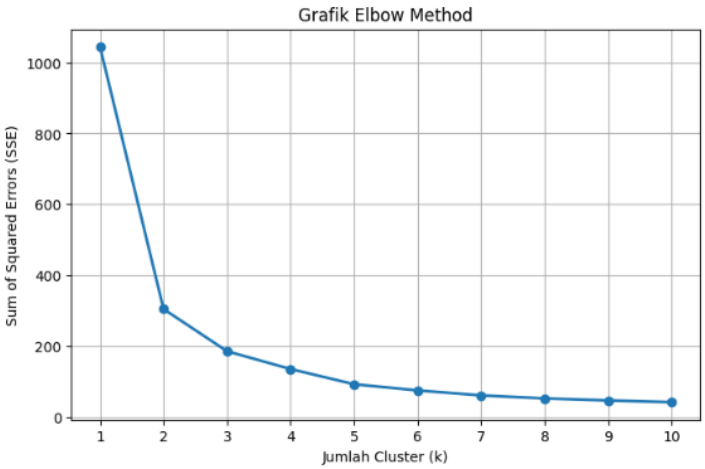


**Figure 2.** Elbow Method Chart

Based on Figure 3, the Elbow Method curve shows a change in the curve pattern that begins to slope after five clusters. These conditions suggest that the addition of clusters after k = 5 provides only a relatively small decrease in SSE values compared to the increase in model complexity.

**Table 5** Clustering Results

| *Cluster* | *Recency* (Month) | *Frequency* (Frequency) | *Monetary* (Rp) |
|---|---|---|---|
| **0** | 8,31 | 2,09 | 160.390,91 |
| **1** | 21,03 | 1,40 | 86.255,46 |
| **2** | 2,68 | 2,41 | 199.376,36 |
| **3** | 14,52 | 1,64 | 105.979,55 |
| **4** | 2,02 | 13,03 | 1.256.933,67 |

Measurements using Silhouette Score were carried out to evaluate the quality of clustering results as an internal validation metric. The calculation results showed a Silhouette Score value of 0.5024 in the K-Means model with five clusters. This value shows that the grouping results are of fairly good quality, characterized by a high level of similarity between members in the same cluster (*cohesion*) and adequate separation of other clusters (*separation*). Thus, the results of clustering are considered suitable for use as a basis for the interpretation of customer characteristics and the preparation of marketing strategies.

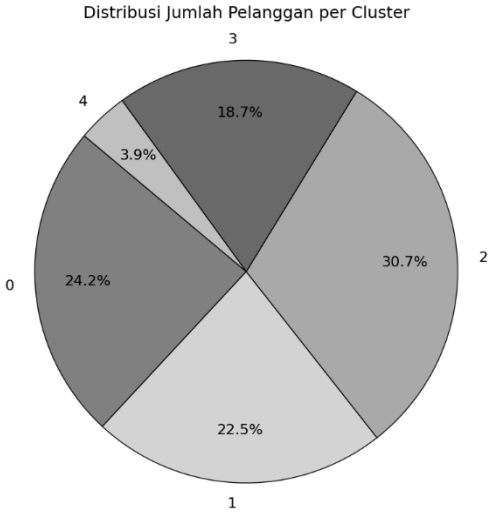


**Figure 3.** Distribution of Cluster Member Grouping Results

The clustering results using the K-Means algorithm resulted in five customer segments with different RFM characteristics as shown in Table 5 and Figure 4. Based on customer distribution, **Cluster 2 (Potential Loyal Customer)** is the largest segment by proportion **30,7%** of total customers, followed by **Cluster 0 (At Risk Customer)** by **24,2%**, **Cluster 1 (Lost Customer)** by **22,5%**, **Cluster 3 (Inactive Customer)** by **18,7%**, and **Cluster 4 (Loyal Customer)** by **3,9%**. Despite having the least number of customers, Cluster 4 performs best with value ***Recency*** lowest (2.02 months), ***Frequency*** highest (13.03 times), and ***Monetary*** (IDR 1,256,933.67), making it the segment with the highest business value. In contrast, Cluster 1 has ***Recency*** highest (21.03 months) and the value of ***Frequency*** and ***Monetary*** lowest, so that it is identified as **Lost Customer**. Meanwhile, Cluster 2 has a *Acknowledgments* low (2.68 months) with *Frequency* and *Monetary* which is relatively high so that it is categorized as **Potential Loyal Customer**, while Cluster 0 and Cluster 3 are interpreted as **At Risk Customer** and **Inactive Customer**. Visualization of segmentation results in three-dimensional RFM space is presented in Figure 5.

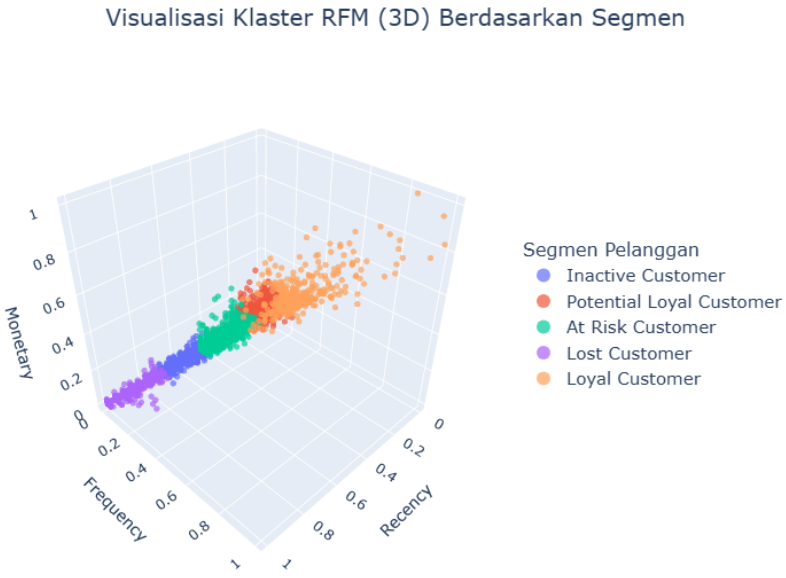


**Figure 4.** Visualization of RFM Clusters by Segments

The results of the study showed that the combination of the RFM method and the K-Means algorithm was able to group customers of the AHASS Cimalaka Sumedang Motorcycle Workshop into five segments based on transaction behavior characteristics. These findings are in line with research that states that the *Recency*, *Frequency*, and *Monetary* able to represent customer transaction behavior so that it is effectively used as a basis for customer segmentation. In addition, the results of this study are also in line with research that shows that the combination of RFM and K-Means is able to produce customer segmentation as the basis for the preparation of marketing strategies. However, unlike the study, this study not only produces customer segmentation, but also prepares marketing strategy priorities based on a combination of RFM values, customer proportions, and potential business contributions of each segment so that the recommendations produced are more appropriate to support marketing decision-making in the workshop service industry.

Based on the results of customer segmentation, each cluster has different business characteristics and values so it requires a customized marketing strategy. Prioritizing strategies takes into account value *Recency*, *Frequency*, *Monetary* (RFM), the proportion of customers in each cluster, and their potential contribution to increasing loyalty and company revenue. This approach is in line with the concepts of customer loyalty, customer segmentation, and data-driven marketing that emphasizes the importance of marketing decision-making based on customer value and data analysis results (Buttle & Maklan, 2019; Payne & Frow, 2015; Wedel & Kamakura, 2012; Wedel & Kannan, 2016).

**Table 6.** Recommended Marketing Strategies Based on Customer Segmentation Results

| Priorities | Cluster | Characteristics | Marketing Strategy |
|---|---|---|---|
| **1** | **Potential Loyal Customer** | Active customers with the potential to become loyal and are the largest segment (30.7%). | Send service reminders based on the last service history and offer periodic service packages or bundling relevant parts to increase loyalty. |
| **2** | **At Risk Customer** | Transaction activity is starting to decline, but it still has a fairly good transaction value. | Remind the service schedule with a service promo or replacement of components that are estimated to have entered the maintenance period based on the transaction history. |
| **3** | **Inactive Customer** | Customers have not transacted for a long time but still have the potential to be reactivated. | Offer a comeback package in the form of a free vehicle inspection or a full service discount based on the last service history. |
| **4** | **Loyal Customers** | Customers with the best RFM value and the highest revenue contribution despite the small number (3.9%). | Develop customers into *customer advocates* through referral programs, giving appreciation to loyal customers, testimonials, and priority access to exclusive services and promos to encourage word-of-mouth recommendations and increase new customer acquisition. |
| **5** | **Lost Customer** | Customers with the lowest transaction activity and value and low chance of return. | Not being the target of an active marketing campaign. Conduct a survey to find out why customers stop transacting. |

Marketing strategy recommendations are compiled according to the characteristics and priorities of each customer segment as presented in Table 5. The top priority is given to **Potential Loyal Customers** because they are the largest segment and have a high chance of being upgraded to loyal customers. Furthermore, **At Risk Customer** is the focus of retention

so as not to move to inactive customer segments, while **Inactive Customer** requires a reactivation strategy to increase transaction frequency again. Although **Loyal Customers** only make up a small percentage of customers, this segment still needs to be maintained through loyalty programs because it contributes the largest amount of revenue. Meanwhile, **Lost Customers** are the last priority because they have a relatively low probability of returning to transactions. Thus, the results of the segmentation of RFM and K-Means are not only able to identify customer characteristics, but also provide an objective basis for workshops to develop more effective and efficient marketing strategies in accordance with the values and potential of each customer segment.

## CONCLUSION

Customer segmentation using *the Recency, Frequency, Monetary* (RFM) method and the K-Means Clustering algorithm succeeded in grouping 10,041 customers of the AHASS Cimalaka Sumedang Motorcycle Workshop into five segments, namely Potential Loyal Customer (30.66%), At Risk Customer (24.17%), Lost Customer (22.52%), Inactive Customer (18.75%), and Loyal Customer (3.90%). The results of segmentation show that Potential Loyal Customers are the largest segment with a high chance of being developed into loyal customers, while Loyal Customers, although only amounting to 3.90%, provide the highest transaction value and frequency of visits so that they become customers with the largest business value. Based on these characteristics, marketing strategies are recommended with priority on Potential Loyal Customers, followed by At Risk Customer, Inactive Customer, Loyal Customer, and Lost Customer. Thus, the implementation of RFM and K-Means is able to produce customer segmentation that supports the preparation of a more effective and targeted marketing strategy for the AHASS Cimalaka Sumedang Motorcycle Workshop. The research findings show that the RFM and K-Means approaches are not only able to generate customer segmentation, but also support the prioritization of customer value-based marketing strategies so that companies can allocate marketing resources more effectively.

## REFERENCES


Aldossary, M., Alyahya, M., & Agag, G. (2024). How and when does engaging customers in environmental sustainability pay off? The role of business strategy. *Sustainability*, *16*(12), 4924.

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.

Akande, O. N., Akande, H. B., Asani, E. O., & Dautare, B. T. (2024). Customer segmentation through RFM analysis and K-means clustering: Leveraging data-driven insights for effective marketing strategy. In *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)* (pp. 1–8). IEEE. https://doi.org/10.1109/SEB4SDG60871.2024.10630052

Akinrinoye, O. V., Kufile, O. T., Otokiti, B. O., Ejike, O. G., Umezurike, S. A., & Onifade, A. Y. (2020). Customer segmentation strategies in emerging markets: a review of tools, models, and applications. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *6*(1), 194–217.

Badan Pusat Statistik. (2025). *Perkembangan jumlah kendaraan bermotor menurut jenis kendaraan di Indonesia*. https://www.bps.go.id

Buttle, F., & Maklan, S. (2019). *Customer relationship management* (4th ed.). Routledge. https://doi.org/10.4324/9781351016551

Chen, A. H.-L., & Gunawan, S. (2023). Enhancing retail transactions: A data-driven recommendation using modified RFM analysis and association rules mining. *Applied Sciences, 13*(18), Article 10057. https://doi.org/10.3390/app131810057

De Mendonca, T. R., & Zhou, Y. (2019). Environmental performance, customer satisfaction, and profitability: A study among large US companies. *Sustainability*, *11*(19), 5418.

Handojo, A., Pujawan, N., Santosa, B., & Singgih, M. L. (2023). A multi-layer recency, frequency, monetary method for customer priority segmentation in online transactions. *Cogent Engineering, 10*(1). https://doi.org/10.1080/23311916.2022.2162679

Lulaj, E. (2023). A sustainable business profit through customers and its impacts on three key business domains: technology, innovation, and service (TIS). *Business, Management and Economics Engineering*, *21*(1), 19–47.

Payne, A., & Frow, P. (2015). *Strategic customer management: Integrating relationship marketing and CRM*. Cambridge University Press.

Pereira, M. de S., et al. (2025). Factors of customer loyalty and retention in the digital environment. *Journal of Theoretical and Applied Electronic Commerce Research, 20*(2), Article 71. https://doi.org/10.3390/jtaer20020071

Perdhana, R. B., & Heikal, J. (2024). *Enhancing customer segmentation in online transportation services: A comprehensive approach using K-means clustering and RFM model*. [Manuscript].

Rahim, M. A., Mushafiq, M., Khan, S., & Arain, Z. A. (2021). RFM-based repurchase behavior for customer classification and segmentation. *Journal of Retailing and Consumer Services, 61*, Article 102566. https://doi.org/10.1016/j.jretconser.2021.102566

Tabianan, K., Velu, S., & Ravi, V. (2022). K-means clustering approach for intelligent customer segmentation using customer purchase behavior data. *Sustainability*, *14*(12), 7243.

Wedel, M., & Kamakura, W. A. (2012). *Market segmentation: Conceptual and methodological foundations* (2nd ed.). Springer.

Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, *80*(6), 97–121. https://doi.org/10.1509/jm.15.0413