Maeswara: Jurnal Riset Ilmu Manajemen dan Kewirausahaan
Volume 2 No 5 Oktober 2024
e-ISSN: 2988-5000; p-ISSN: 2988-4101, Hal 77-89
DOI : https://doi.org/10.61132/maeswara.v2i5.1283
Available online at: https://journal.arimbi.or.id/index.php/Maeswara

Leveraging the RFM Model for Customer Segmentation in a Software-as-a-Service (SaaS) Business Using Python

Andy Hermawan[1], Nila Rusiardi Jayanti[2], Aji Saputra[3], Army Putera Parta[4], Muhammad Abizar Algiffary Thahir[5], Taufiqurrahman[6]
Universitas Indraprasta PGRI, Jakarta, Indonesia[1,2]
Universitas Khairun, Jakarta, Indonesia[3]
Purwadhika Digital School, [6]Jakarta, Indonesia[4-6]
Email : andy.hermawan@unindra.ac.id[1], nilarusiardi.jayanti@unindra.ac.id[2], aji.saputra@unkhair.ac.id[3], armyputera@gmail.com[4], egifthahir@gmail.com[5], taufiqqurrahman382@gmail.com[6]

Korespondensi penulis : andy.hermawan@unindra.ac.id

Abstract. Customer segmentation plays a pivotal role in driving marketing strategies and improving customer retention across various industries. This study explores the application of the RFM (Recency, Frequency, Monetary) model for customer segmentation in a Software-as-a-Service (SaaS) business, using Python for efficient data processing and analysis. By analyzing one year of customer purchase data, we segmented customers into key groups such as "Champions," "Loyal Customers," and "At Risk." The results highlight that targeted discount strategies significantly affect profitability, especially for high-value customer segments. Furthermore, the research builds upon existing methodologies, demonstrating how Python-based implementations streamline RFM analysis and allow for scalable solutions in business contexts, as illustrated in prior works by Hermawan et al. (2024). This study offers actionable recommendations, including tailored discounting, loyalty programs, and personalized engagement strategies, to enhance customer retention and business profitability. The findings underscore the importance of data-driven marketing approaches for customer segmentation and engagement, reinforcing the relevance of the RFM model in modern business environments.

Keywords: Amazon, AWS, RFM Analysis, SaaS, Segmentation.

1. INTRODUCTION

In today's competitive business landscape, understanding customer behavior is essential for developing effective marketing strategies and enhancing customer relationship management. A key technique for this is customer segmentation, which divides a customer base into distinct groups based on specific characteristics. Among various segmentation models, the Recency, Frequency, Monetary (RFM) model is widely recognized for its simplicity and effectiveness in identifying valuable customers, predicting future behaviors, and enabling efficient resource allocation (Blattberg et al., 2008; Christy et al., 2021). This model's versatility has been demonstrated across industries such as retail, telecommunications, and e-commerce (Sabuncu et al., 2020; Ma, 2022), as well as in tourism and hospitality sectors (Dursun & Caber, 2016). Moreover, researchers like Khajvand et al. (2011) have applied the RFM model in estimating customer lifetime value, further illustrating its adaptability across different contexts.

Received: Agustus 11, 2024; Revised: September 19, 2024; Accepted: Oktober 03, 2024; Published: Oktober 08, 2024

For Software-as-a-Service (SaaS) businesses, customer retention and engagement are particularly crucial due to the subscription-based revenue model. High churn rates can significantly affect profitability and growth, yet there is a gap in research applying the RFM model specifically within the SaaS context. Advancements in data analytics, especially using Python, have facilitated more scalable and efficient RFM model implementations. Python's comprehensive libraries and user-friendly syntax make it optimal for handling large datasets and complex analyses, as demonstrated by Hermawan et al. (2024).

Building on these developments, this study aims to apply the RFM model within a SaaS business context, leveraging Python for analysis. By evaluating one year of customer purchase data, the research seeks to implement an efficient Python-based RFM model for effective customer segmentation. The study aims to identify key segments like "Champions," "Loyal Customers," and "At Risk," understanding their contribution to profitability. Additionally, the research examines the impact of discount strategies on different segments to optimize marketing efforts for customer retention and profitability.

This research contributes to existing literature by bridging the gap in RFM applications within the SaaS industry and demonstrating Python's effectiveness in such analyses. While previous studies have explored RFM in various sectors (Monalisa et al., 2023; Zamil & Vasista, 2021), its application in SaaS remains underexplored. Birant (2011) highlights how data mining techniques integrated with RFM can significantly improve customer segmentation accuracy, a notion further explored by Chang & Tsai (2011) who introduced group RFM analysis to uncover better customer behavior insights. Leveraging Python's analytical capabilities, this study intends to provide actionable insights for SaaS businesses to develop tailored marketing strategies, aligning with the principles of strategic database marketing (Stone & Jacobs, 1988; Hughes, 1994).

Understanding customer behavior through segmentation aids in identifying high-value customers and re-engaging inactive or at-risk customers, thereby enhancing customer satisfaction and loyalty (Wan et al., 2022).

2. THEORETICAL REVIEW

The RFM model, rooted in customer relationship management, measures customer value by evaluating their purchasing behavior across three dimensions: recency, frequency, and monetary value. This multidimensional approach allows businesses to target specific customer segments more effectively. According to Hughes (1994), segmenting customers based on their RFM scores can lead to more personalized marketing strategies, thereby increasing customer engagement and profitability.

Previous research in various industries has validated the RFM model's effectiveness. Christy et al. (2021) demonstrated its utility in identifying high-value customer groups, while Monalisa et al. (2023) applied the model to examine customer behavior in e-commerce. Aggelis & Christodoulakis (2005) demonstrated the use of RFM for customer clustering, which allows businesses to categorize customer groups for more targeted marketing. In the SaaS context, however, there is limited exploration of RFM applications, particularly with the enhanced capabilities provided by Python for data analysis.

Hermawan et al. (2024) emphasize the advantages of Python in RFM analysis, illustrating how its libraries, such as Pandas and NumPy, streamline data processing and segmentation. Furthermore, Python's Matplotlib and Seaborn libraries facilitate comprehensive visualizations of customer segments, aiding in actionable decision-making for marketing and retention strategies.

3. METHODOLOGY

Research Design

This research employs the RFM model for customer segmentation within a SaaS business. The study uses Python for implementing the segmentation process due to its efficiency in data handling and analysis (Hermawan et al., 2024). The research encompasses data collection, preprocessing, RFM score calculation, customer segmentation, and data analysis/visualization.

Data Collection and Preprocessing

The dataset, sourced from a SaaS company's customer transactions over one year, includes customer IDs, transaction dates, transaction frequency, and monetary values. Data preprocessing is essential to ensure accurate RFM analysis. Key preprocessing steps include:

- **Data Cleaning:** Missing values were addressed by imputation or removal, and duplicates were eliminated to maintain data integrity.

- **Data Formatting:** Transaction dates were standardized using Python's Pandas and datetime libraries (Hermawan et al., 2024).

These steps align with the importance of clean data for accurate segmentation as highlighted by Zamil & Vasista (2021).

RFM Score Calculation

RFM scores were calculated as follows:

- **Recency (R):** Days since the last transaction.
- **Frequency (F):** Total transactions within the observation period.
- **Monetary (M):** Total value of transactions.

Each metric was scored on a scale of 1 to 5, following Hughes (1994), where 5 indicates the highest value and 1 the lowest. The combined RFM score is the sum of these three individual scores, providing a comprehensive view of customer value (Christy et al., 2021; Monalisa et al., 2023).

Customer Segmentation

Based on RFM scores, customers were segmented into:

1. **Champions:** High recency, frequency, and monetary values.
2. **Loyal Customers:** High frequency with lower recency.
3. **At-Risk:** Previously active but with declining activity.
4. **Hibernating:** Low scores in all categories.

Python's Pandas and NumPy libraries facilitated this segmentation, following methodologies outlined by Hermawan et al. (2024). These segments provide insights into customer engagement and value, as similarly implemented in e-commerce contexts (Ma, 2022).

Data Analysis and Visualization

Post-segmentation, visualizations were created using Python's Matplotlib and Seaborn libraries to analyze customer distributions and contributions to revenue. The effect of discount strategies was evaluated by comparing performance between customers who received discounts and those who did not, drawing from Wan et al. (2022).

Tools and Implementation

The entire analysis utilized Python's data-handling capabilities. Critical libraries included Pandas for data manipulation, NumPy for numerical operations, and Matplotlib/Seaborn for visualization.

4. RESULTS AND ANALYSIS

Data Analysis

The data analysis section explores the insights derived from applying the RFM model to customer segmentation within a Software-as-a-Service (SaaS) business. Utilizing Python-based implementations, critical behaviors across various segments were analyzed, offering actionable insights for improving customer retention and profitability.

RFM Score Distribution

After calculating the RFM scores for each customer, the dataset was divided into distinct segments based on Recency, Frequency, and Monetary values. Customers were scored on a scale of 1 to 5 for each metric, with higher scores indicating greater recency, frequency, and monetary contribution.

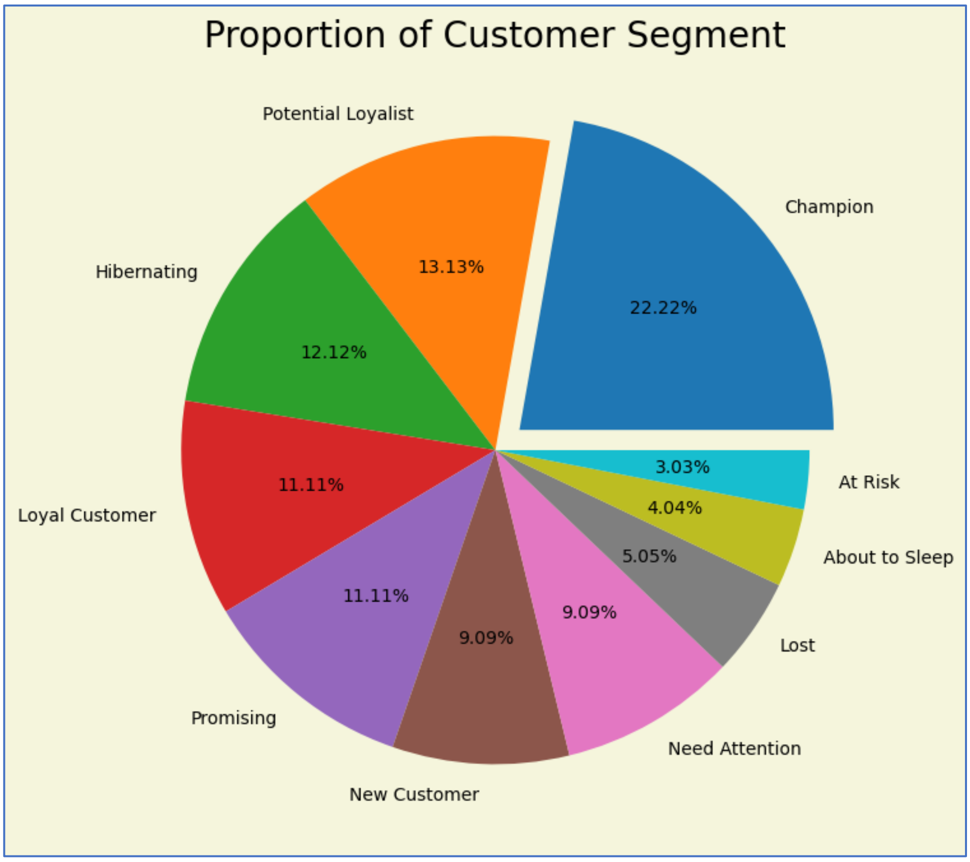

Figure 1. Customer Segment Proportion

The distribution of customer segments is as follows:

- **Champions (22.22%)**: Customers who frequently purchase, make recent transactions, and significantly contribute to revenue. These are the most valuable customers, requiring focused retention efforts (Hermawan et al., 2024).
- **Potential Loyalists (13.13%)**: Customers who exhibit promising buying behaviors and could evolve into Loyal Customers or Champions with targeted engagement.
- **Hibernating (12.12%)**: Customers with limited engagement over time, requiring reactivation efforts though their long-term revenue potential is low.
- **Loyal Customers (11.11%)**: Customers with high Frequency and Monetary scores but slightly lower Recency. They provide consistent revenue, though recent engagements have decreased.
- **Promising (11.11%)**: Customers showing potential for future engagement, needing nurturing to evolve into more valuable segments.

- **New Customers (9.09%)**: Recently engaged customers, requiring further nurturing to establish buying patterns.
- **Need Attention (9.09%)**: Customers showing reduced purchasing frequency and monetary contribution, indicating the need for targeted strategies to prevent churn.
- **Lost (5.05%)**: Customers who have disengaged, requiring significant re-engagement efforts.
- **About to Sleep (4.04%)**: Customers on the verge of inactivity, requiring immediate re-engagement to prevent transitioning to "Lost" or "Hibernating."
- **At-Risk (3.03%)**: Previously high-spending customers whose activity has sharply declined, requiring personalized re-engagement efforts.

Segment Contribution to Revenue

The analysis revealed that **Champions** and **Loyal Customers** collectively contributed nearly 50% of the company's revenue while representing only 33.33% of the customer base. This highlights the need to focus on these segments for marketing and retention strategies (Christy et al., 2021). Conversely, the **At-Risk** and **Lost** segments contributed less than 10% of revenue, indicating a decline due to customer churn (Wan et al., 2022).

Impact of Discount Strategies

The effect of discount strategies on profitability was evaluated across different segments by analyzing customer behaviors before and after discount campaigns:

1. **Champions**: Displayed stable purchase frequency and high monetary value, unaffected by discounts. Offering discounts to this group reduced profitability, as they are less price-sensitive (Hermawan et al., 2024).
2. **At-Risk**: Demonstrated moderate responsiveness to discounts with increased frequency but lower monetary contribution, suggesting that while discounts incentivize purchases, they do not fully restore high-value behaviors.

3. **Hibernating Customers**: Showed a notable increase in purchases post-discount, but their overall contribution to profitability remained low (Sabuncu et al., 2020; Wan et al., 2022).

The findings suggest that while discounts can temporarily increase sales, they do not always lead to sustainable revenue growth, particularly among high-value customers (Hughes, 1994; Stone & Jacobs, 1988).

Customer Engagement and Loyalty Programs

Loyal Customers and **Champions** are ideal candidates for loyalty programs, given their strong brand association. Offering non-monetary incentives, such as exclusive access or early product launches, can maintain their engagement without resorting to discounts (Zamil & Vasista, 2021). Python-based models can dynamically adjust loyalty programs based on real-time data, as illustrated by Hermawan et al. (2024).

Visualization and Interpretation of Results

Visualizations of RFM scores and customer segments were generated using Python's Matplotlib and Seaborn libraries. Pie charts and bar graphs depicted customer distribution across segments and the impact of discount strategies, offering insights into targeted marketing.

Profitability by Segment

1. **Champions and Loyal Customers**: Provided substantial profits without discounts (e.g., "Champions" contributed over 110,307.14 USD). Discounts eroded profitability (e.g., negative profit of -13,803.35 USD), suggesting that non-monetary incentives are preferable.

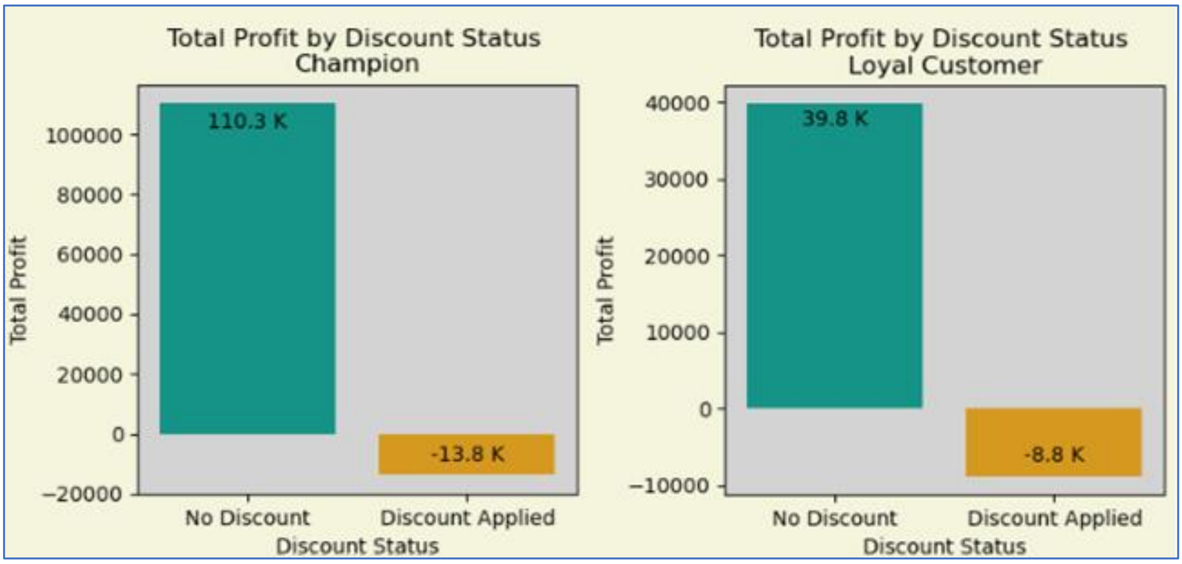

Figure 2. Champion and Loyal Customer Response

2. **Need Attention and At-Risk Segments**: Responded positively to discounts, with segments like **Need Attention** showing a profit increase (4,314.41 USD), indicating that targeted discounts can be effective.

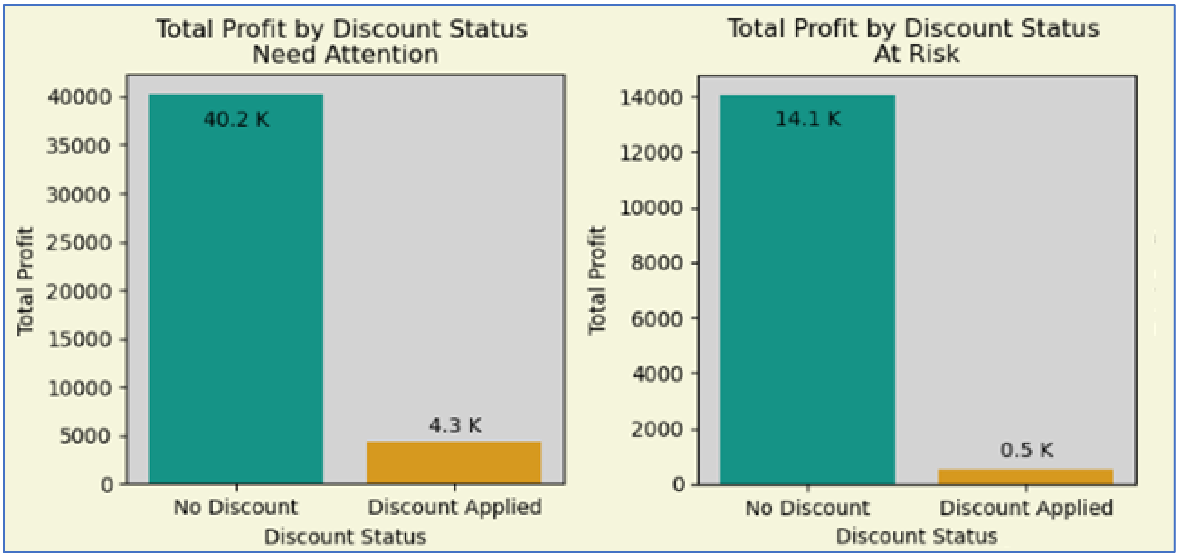

Figure 3. Need Attention and At-Risk Customer Response

3. **Promising and New Customers**: Demonstrated negative profits when discounts were applied, highlighting that such incentives are not effective for driving long-term loyalty.

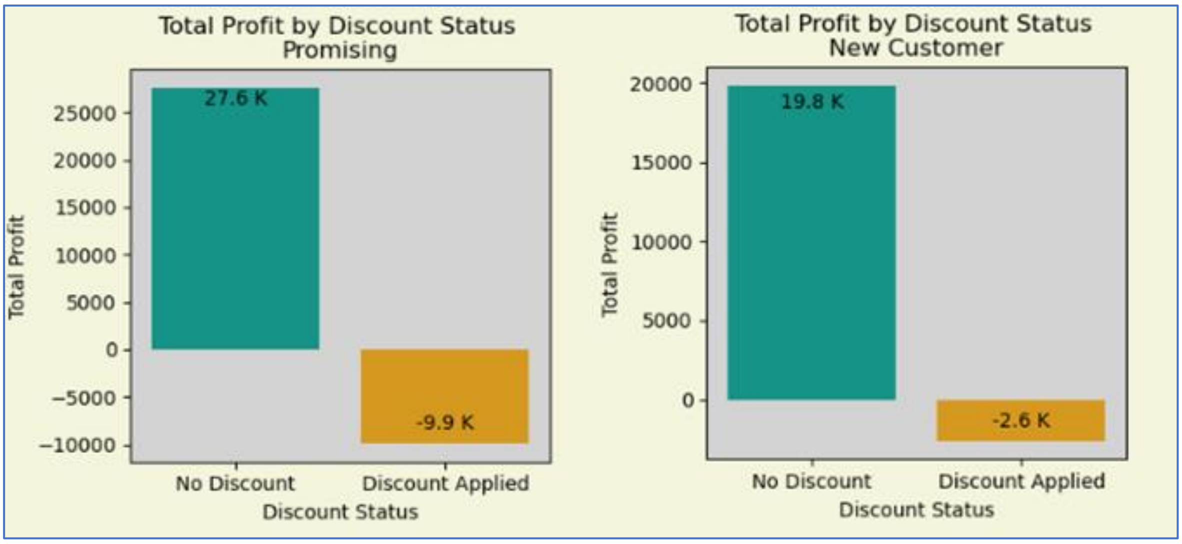

Figure 4. Promising and New Customer Response

These findings emphasize a need for differentiated engagement and discount strategies tailored to customer segments.

5. CONCLUSION AND SUGGESTIONS

This study applied the Recency, Frequency, and Monetary (RFM) model to effectively segment customers within a Software-as-a-Service (SaaS) business, offering actionable insights into customer behavior, engagement, and profitability. Leveraging Python for data analysis and segmentation, five key customer segments were identified: Champions, Loyal Customers, At-Risk, Hibernating, and Need Attention. Each segment was analyzed in terms of revenue contribution and responsiveness to discount strategies.

The analysis produced several key conclusions:

First, **Champions** and **Loyal Customers** were identified as high-value segments, contributing significantly to the company's revenue. These segments showed a high level of engagement and spending, making them ideal candidates for non-monetary engagement strategies such as loyalty programs and personalized services.

The minimal impact of discounts on these groups suggests that broad discounting could undermine profitability without enhancing engagement, a finding consistent with Hermawan et al. (2024).

Second, the **At-Risk** segment, which previously contributed significantly to revenue, demonstrated signs of potential churn.

Targeted discount strategies had a moderate effect on this group, implying that while financial incentives can prevent churn to some extent, more personalized re-engagement campaigns would likely be more effective. This is in line with Wan et al. (2022) and Zamil & Vasista (2021), who emphasized the value of tailored retention efforts for at-risk customers.

Third, the **Hibernating** segment exhibited the strongest response to discount offers, leading to short-term increases in purchasing behavior. However, their overall contribution to revenue remained low, indicating that while discounts may temporarily re-engage dormant customers, their long-term profitability is limited. Sabuncu et al. (2020) similarly noted that broad discounting does not result in sustainable engagement from low-value customers.

A key recommendation from this study is the need for personalized customer retention strategies. While discounts can effectively boost engagement for some segments, non-monetary incentives—such as loyalty programs, personalized offers, and exclusive services—are more appropriate for high-value customers, like Champions and Loyal Customers. This approach aligns with findings by Christy et al. (2021) and Ma (2022), who demonstrated the positive impact of tailored strategies on customer satisfaction and retention.

Furthermore, the study demonstrated the efficiency of Python-based RFM implementation. The use of Python libraries, such as Pandas for data manipulation and Matplotlib for visualization, streamlined the analysis process, making it scalable and easy to interpret. This supports the methodology of Hermawan et al. (2024), who successfully employed Python to enhance customer segmentation in business contexts.

Implications for SaaS Businesses

For SaaS businesses, the study highlights the importance of leveraging data-driven models like RFM to understand customer behaviors and tailor marketing strategies accordingly. Focusing on the specific needs and behaviors of different customer segments can enhance customer retention, reduce churn, and increase profitability. Moreover, the findings reinforce the need to move away from blanket discounting strategies and instead adopt a more targeted, personalized approach to customer engagement.

REFERENCES

Aggelis, V., & Christodoulakis, D. (2005, July). Customer clustering using RFM analysis. In Proceedings of the 9th WSEAS International Conference on Computers (p. 2). doi: Not available

Birant, D. (2011). Data mining using RFM analysis. In Knowledge-oriented applications in data mining. IntechOpen. doi: Not available

Blattberg, R. C., Kim, B. D., & Neslin, S. A. (2008). RFM Analysis. In Database Marketing (Vol. 18, pp. 323–337). Springer, New York, NY. doi: 10.1007/978-0-387-72579-6_12

Chang, H. C., & Tsai, H. P. (2011). Group RFM analysis as a novel framework to discover better customer consumption behavior. Expert Systems with Applications, 38(12), 14499-14513. doi: 10.1016/j.eswa.2011.05.034

Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking – An effective approach to customer segmentation. Journal of King Saud University - Computer and Information Sciences, 33(10). doi: 10.1016/j.jksuci.2018.09.004

Dursun, A., & Caber, M. (2016). Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. Tourism Management Perspectives, 18, 153-160. doi: 10.1016/j.tmp.2016.03.001

Hermawan, A., Kahfi, R. A., Surya, E., Aini, U., & Hidayat, R. (2024). Penerapan Metode RFM dengan Python dalam Segmentasi Pelanggan. Jurnal Bisnis Inovatif Dan Digital, 1(3), 92–102. doi: 10.61132/jubid.v1i3.222

Hughes, A. M. (1994). Strategic Database Marketing. Probus Publishing.

Khajvand, M., Zolfaghar, K., Ashoori, S., & Alizadeh, S. (2011). Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study. Procedia Computer Science, 3, 57-63. doi: 10.1016/j.procs.2010.12.011

Ma, J. (2022). E-commerce customer segmentation based on RFM model. Lecture Notes in Electrical Engineering, 827, 118. doi: 10.1007/978-981-16-8052-6_118

Monalisa, S., Juniarti, Y., Saputra, E., Muttakin, F., & Ahsyar, T. K. (2023). Customer segmentation with RFM models and demographic variable using DBSCAN algorithm. Telkomnika (Telecommunication Computing Electronics and Control), 21(4). doi: 10.12928/TELKOMNIKA.v21i4.22759

Sabuncu, İ., Türkan, E., & Polat, H. (2020). Customer segmentation and profiling with RFM analysis. Turkish Journal of Marketing, 5(1), 22-36. doi: 10.30685/tujom.v5i1.84

Stone, B., & Jacobs, R. (1988). Successful Direct Marketing Methods. NTC Business Books.

Wan, S., Chen, J., Qi, Z., Gan, W., & Tang, L. (2022). Fast RFM model for customer segmentation. WWW 2022 - Companion Proceedings of the Web Conference 2022, 965 – 972. doi: 10.1145/3487553.3524707

Zamil, A. M. A., & Vasista, T. G. (2021). Customer segmentation using RFM analysis: Realizing through Python implementation. Pacific Business Review International, 13, 11 May 2021.