Jurnal Scientia
Volume 13, Number 03, 2024, DOI 10.58471/ scientia.v13i03
ESSN 2723-7486 (Online)
https://infor.seaninstitute.org/index.php/pendidikan

Analysis Of Customer Segmentation In The Top Three Most Visited E-Commerce Platforms In Indonesia In 2023 Using RFM Model And Clustering Techniques

[1]Siti Wulansari, [2]Jerry Heikal
[1]Master of Management, Bakrie University Jakarta, [2]Master of Management Lecturers, Bakrie University Jakarta

Article Info
Keywords:
RFM,
K-Means Clustering,
E-commerce,
Digital Transformation,
Customer Segmentation,
Customer Behavior,
Marketing Strategy.

ABSTRACT
This study analyzes customer behavior on three most visited e-commerce platforms in Indonesia (Shopee, Tokopedia, and Lazada) using the RFM (Recency, Frequency, Monetary) model and K-Means clustering techniques. Amidst the digital transformation and rapid growth of e-commerce, significant changes in customer behavior highlight the importance of accurate and relevant customer segmentation. The study integrates demographic, psychographic, and purchasing behavior analysis, employing stratified random sampling to collect data over 6 months from 100 respondents on each e-commerce platform. The results revealed six customer clusters, each with distinct characteristics. These clusters were further analyzed to develop specific and effective marketing strategies tailored to customer behavior and preferences on each e-commerce platform.

Corresponding Author:
Siti Wulansari
Master of Management, Bakrie University Jakarta
sitiwulansari@hotmail.com

INTRODUCTION

Based on an analysis of the Indonesian e-commerce market, the size of the e-commerce market in Indonesia is estimated to reach USD 81.80 billion in 2024 and is expected to reach USD 168.10 billion in 2029, with a compound annual growth rate (CAGR) of 15.5% during the forecast period (2024-2029). The key factors driving the growth of e-commerce in Indonesia are the convenience factor offered by online shopping and the continuous focus of e-commerce platforms to improve the overall customer shopping experience. Shoppers across the country are now turning to e-commerce platforms to fulfill their diverse needs, from electronics and clothing to groceries. These factors have led to substantial growth in the number of online shoppers across the country over the past few years, significantly contributing to the size of the Indonesian e-commerce market. In addition, high smartphone use, and internet penetration stimulate the growth of the e-commerce industry in Indonesia. Increasing internet usage at affordable rates and increased use of smartphones has led to higher access to online shopping in the country. High mobile penetration in Indonesia makes it more convenient for consumers to shop for various products online, thereby positively influencing Indonesia's e-commerce market share. The government and central bank are also enthusiastic about facilitating digital payment systems in this country because these payments are more efficient and transparent. The proliferation of digital payments in this

Jurnal Scientia
Volume 13, Number 03, 2024, DOI 10.58471/ scientia.v13i03
ESSN 2723-7486 (Online)
https://infor.seaninstitute.org/index.php/pendidikan

country is also analyzed further to support the growth of e-commerce in Indonesia. With COVID-19, Indonesia's e-commerce industry has experienced significant expansion in recent years, mainly supported by high internet and smartphone penetration, a growing young and tech-savvy population, changing consumer shopping behavior, and a growing digital economy, coupled with the significant presence of major e-commerce vendors offering variety, convenience and competitive prices for online shopping. Apart from that, the proactive role of the Indonesian government in encouraging e-commerce in the digital economy has also further fueled the growth of the e-commerce industry in Indonesia (Mordor Intelligence, 2024).

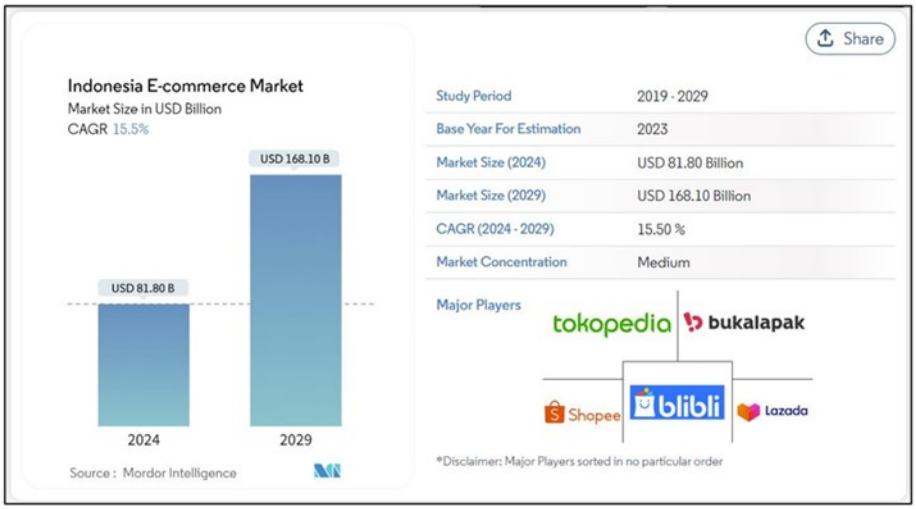

Figure 1. E-Commerce Market Size
Resource: (Hasil Olahan Penulis, 2024)

In 2023, there are the 10 most visited e-commerce in Indonesia, namely Shopee which dominates with around 2.3 billion visits throughout 2023 and recorded significant growth in visits of 41.39% compared to 2022. Tokopedia is in the second position with a total of 1.2 billion visits, however, visits to this site decreased by 21.08% throughout 2023. Lazada was in third place with 762.4 million visits but experienced a decrease of 46.72% compared to 2022. Blibli is in fourth place with 337.4 million visits and recorded visit growth of 25.18%. Bukalapak is in fifth position with 168.2 million visits but has experienced a decrease in visits of 56.5% compared to 2022. JD.ID is in sixth position with around 60 million visits. However, JD.ID stopped its operations in Indonesia in March 2023. Orami, which focuses on maternal and baby products, is in seventh place with more than 50 million visits throughout 2023. Zalora, an e-commerce platform that focuses on fashion, is in eighth position with more than 40 million visits throughout 2023. MatahariMall, part of Matahari Department Store, recorded more than 30 million visits throughout 2023, making it the ninth most popular e-commerce in Indonesia. Bhinneka, a B2B e-commerce platform that offers IT products and office equipment, is in tenth position with more than 1 million business customers per month. The following is a graph of the number of visitors from the 5 e-commerce sites with the highest number of visitors during 2023 (Annur, 2024).

Jurnal Scientia
Volume 13, Number 03, 2024, DOI 10.58471/ scientia.v13i03
ESSN 2723-7486 (Online)
https://infor.seaninstitute.org/index.php/pendidikan

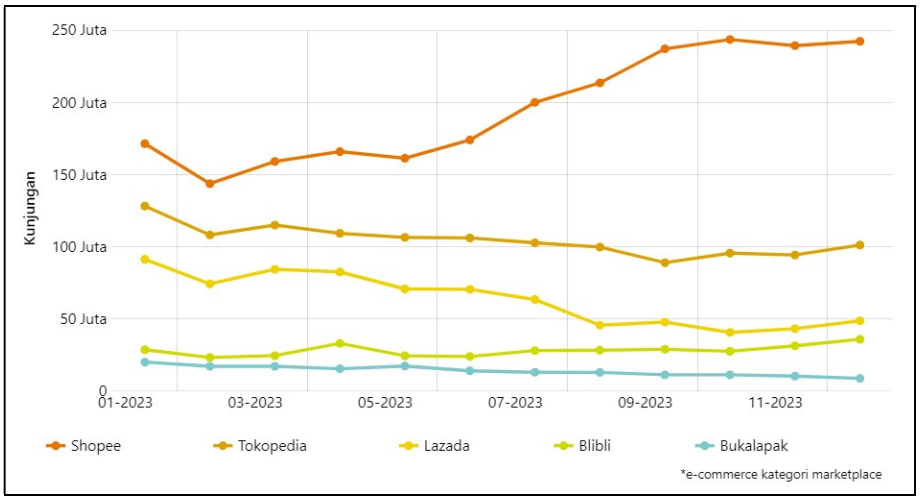

Figure 2. The number of visitors from the 5 e-commerce sites with the highest number of visitors in Indonesia during 2023
Resource: (Hasil Olahan Penulis, 2024)

Based on the explanation above regarding the overall future of e-commerce and the global retail industry, which has already begun to take shape from 2022 until now, therefore, these platforms need to establish effective strategies and business steps to remain competitive with the increasing number of new e-commerce platforms emerging as online product and services sales continue to grow. One way to do this is to attract customer attention and try to retain it. As a company's main asset and the key to business success, companies need to understand customer data better in all aspects. Detecting similarities and differences among customers, predicting their behavior, proposing better choices and opportunities for customers has become crucial for customer and company engagement. Customer segmentation according to their data is important in this case. RFM (Recency, Frequency, and Monetary) values have been used for years to identify which customers are valuable to the company, which customers need promotional activities, etc. (Dogan, Aycin, & Bulut, 2018).

Apart from that, there are three main types of market segmentation that you should know, which include demographic, psychographic and behavioral segmentation. It is important to understand what these three segmentations are if a company is to achieve continued success. Each consumer segment created will have specific wants and needs, which can be accommodated effectively after conducting proper research. By utilizing market segmentation, companies will benefit from being able to use their resources more efficiently and make more informed marketing decisions. RFM models and clustering techniques have been used extensively in previous research to identify valuable customer segments and optimize marketing strategies (Fader, Hardie, & Lee, 2005). For example, research by (Dogan, Aycin, & Bulut, 2018), shows that the use of the RFM model can help companies understand customer behavior and significantly increase customer retention. Based on this description, and supported by the phenomenon of rapid e-commerce development and the importance of customer segmentation in increasing sales and profits

Jurnal Scientia
Volume 13, Number 03, 2024, DOI 10.58471/ scientia.v13i03
ESSN 2723-7486 (Online)
https://infor.seaninstitute.org/index.php/pendidikan

for companies, especially those running businesses in the e-commerce sector, the author wishes to conduct research with the title, "Analysis of Customer Segmentation in the Top Three Most Visited Commerce Platforms in Indonesia in 2023 Using the RFM Model and Clustering Technique". This research aims to determine customer characteristics and group customers based on the same characteristics into their respective groups using survey data regarding customer past transaction history using the clustering method with the K-Means algorithm and a combination of the RFM model. This is very important considering that customer segmentation analysis is an effort to predict and target potential customers, attract new customers, implement appropriate marketing strategies, identify the relationship between customers and the company, and increase the company's expected profitability.

METHODS

This research is a quantitative descriptive analysis which aims to describe a phenomenon, event or incident that occurs factually, systematically and accurately using numbers that describe the characteristics of the subject under study (Apuke, 2017). In this research, data mining techniques in the form of clustering will be used. By using descriptive analysis in the form of clustering, this research will be able to provide clusters or segments of customers along with the characteristics of each cluster based on Recency, Frequency and Monetary data from customers in the form of numbers (numerical). In this case, the clustering technique used is the K-Means technique, namely a distance-based clustering technique, namely the Recency, Frequency and Monetary distance between one customer and another, so that customers can be divided into a number of clusters or groups. In this research, only data or features in the form of Recency, Frequency, and Monetary will be obtained for every customer who shops at e-commerce Shopee, Tokopedia, and Lazada. Even though there is no initial label or segment for each customer, using the K-Means technique for clustering will still be able to group customers into several segments, namely by calculating the Recency, Frequency and Monetary distance between all customers. A close distance between two customers indicates that the two customers are in the same segment group, while a long distance between two customers indicates that the two customers are in different segment groups.

Data shows that visitors from Shopee, Tokopedia, and Lazada already covers more than 80% of the total number of visitors from the 10 most visited e-commerce sites in Indonesia in 2023, therefore, the researcher will focus on segmenting customers within these top three platforms. The number of visitors from each e-commerce that will be studied will be the population in this research. The number of respondents in the study was calculated using the Slovin formula. Based on the sample calculation using the Slovin formula above, and with a tolerable error percentage of 10%, the number of samples that could be used as respondents in this study was 100 respondents from each e-commerce company. The sampling technique used is stratified random sampling, which is a sampling technique by first dividing the population into several subpopulations (layers, strata) and then taking samples from each population. In this way, the overall sample (of n) will generally be more reflective of the population than directly taking n units from the

Jurnal Scientia
Volume 13, Number 03, 2024, DOI 10.58471/ scientia.v13i03
ESSN 2723-7486 (Online)
https://infor.seaninstitute.org/index.php/pendidikan

population, which could result in a sample that only represents some groups in the population (Asra & Prasetyo, 2015).

This research was carried out at the clustering process stage or grouping customers using the K-Means method. The data used in this stage is normalized data. After knowing the clustering results that have been used, the next step is to test the validity with WCSS. WCSS is used to analyze variations in the distance between the centers of the clusters formed.

RESULTS AND DISCUSSION

RFM Attribute Value

At this stage, manual calculation tests were carried out on K-Means Clustering model, the data used was data that had been transformed in the previous stage.

Table 1. RFM Attribute Values

Kategori	*Recency*	*Frequency*	*Monetary*
A	Tinggi	Tinggi	Tinggi
B	Sedang	Sedang	Sedang
C	Rendah	Rendah	Rendah

Resource: (Hasil Olahan Penulis, 2024)

The following is an explanation for each category based on a combination of Recency, Frequency and Monetary values:

Category A: High-High-High

Characteristics:

1. High Recency: These customers haven't interacted with the business in a while, indicating that they may have lost interest or found alternatives.
2. High Frequency: These customers frequently make transactions, indicating high loyalty and engagement.
3. High Monetary: These customers own high transaction value, either because they frequently purchase expensive products or because they buy in large quantities.

   Even though they have high Recency (which means they haven't interacted for a long time), they are still very valuable customers because they have high Frequency and high Monetary. They may be at risk of churn or are no longer active. Businesses can consider drawing them back with reactivation campaigns, special offers, or promotions to remind them of the value of the product or service offered. If it is not possible to reactivate them, resources might be better allocated to more valuable categories.

Category B: Fair-Moderate

Characteristics:

1. Medium Recency: These customers last interacted with the business some time ago, but not too long ago.

Jurnal Scientia
Volume 13, Number 03, 2024, DOI 10.58471/ scientia.v13i03
ESSN 2723-7486 (Online)
https://infor.seaninstitute.org/index.php/pendidikan

2. Medium Frequency: These customers make purchases with moderate frequency, not too often but not infrequently either.
3. Medium Monetary: These customers contribute a moderate amount of revenue, either in terms of purchase quantity or transaction value.
   This category of customers shows steady but not outstanding engagement. They may need additional encouragement to increase their purchase frequency or purchase value. Businesses can use upselling, cross-selling, or offer additional incentives to increase the activity of these customers and encourage them to move towards category A.

Category C: Low-Low-Low

Characteristics:

1. Low Recency: These customers has recently interacted with the business, indicating that they are still active and engaged.
2. Low Frequency: These customers have a low purchasing frequency, which means they rarely make transactions.
3. Low Monetary: These customers contribute little to revenue, either because they purchase infrequently or because they they buy low-priced products.
   This category shows customers who have recently interacted but have low engagement in terms of purchase frequency and value. Efforts should be focused on increasing their transaction frequency and value so they can move into more profitable categories by providing products or services that match their preferences, as well as using loyalty programs to increase transaction value and purchase frequency.

Clustering Analysis with K-Means

In analyzing this data, one of the important steps is determining the optimal number of clusters for grouping the data. In this research, researcher used the K-Means Clustering method. This method is a popular technique in machine learning that is used to divide data into a number of clusters based on the similarities between them. To determine the optimal number of clusters, the author applies the Elbow Method. Elbow Method is a technique that involves calculating the Within-Cluster Sum of Squares (WCSS) for various numbers of clusters. Within-Cluster Sum of Squares (WCSS) is used to measure the validity of clustering in the K-Means method. The point at which the WCSS reduction begins to slow down ("elbow") indicates the optimal number of clusters, where adding additional clusters does not provide a significant benefit in WCSS reduction. WCSS is defined as the sum of the squares of the distance between each data point and its nearest cluster centroid. The following are the WCSS values for each number of Clusters k:

Jurnal Scientia
Volume 13, Number 03, 2024, DOI 10.58471/ scientia.v13i03
ESSN 2723-7486 (Online)
https://infor.seaninstitute.org/index.php/pendidikan

Table 2. Within-Cluster Sum of Squares (WCSS) values per Cluster

Jumlah *Cluster* (k)	WCSS Shopee	WCSS Tokopedia	WCSS Lazada
1	23,033	19,132	15,736
2	14,308	12,817	10,471
3	9,284	8,298	7,165
4	7,614	6,416	5,437
5	6,246	5,412	4,408
6	5,253	4,642	3,810
7	4,220	4,054	3,374
8	3,744	3,444	2,936
9	3,291	2,951	2,580
10	2,945	2,609	2,248

Resource: (Hasil Olahan Penulis, 2024)

WCSS measures the total squared distance between data points and their cluster centers. In general, a lower WCSS (Within-Cluster Sum of Squares) indicates better clustering. The following is an interpretation of the clustering results shown in Table 2.

1. k1 – With only one cluster, all data is grouped into a single cluster, resulting in a very high WCSS because all data variation is calculated within one cluster.
2. k:2 – With two clusters, the WCSS decreases significantly, indicating that the data can be more efficiently grouped into two clusters.
3. k:3 – With three clusters, the WCSS decreases further, indicating a better separation of data into three groups.
4. k:4 – The decline continues, showing that four clusters can better capture the data structure.
5. k:5 – Five clusters show a further reduction in WCSS, but the decrease begins to slow down.
6. k:6 – With six clusters, WCSS decreases again, but the reduction is smaller compared to previous decreases.
7. k:7 – With seven clusters, there is a further reduction in WCSS, but the decrease starts to flatten out.
8. k:8 – Eight clusters show a further reduction in WCSS, but the change is not significant compared to the previous decrease.
9. k:9 – Nine clusters result in an even smaller reduction in WCSS.
10. k:10 – With ten clusters, WCSS reaches the lowest reported point.

From the results of the K-Means clustering, it can be concluded that:

1. “Elbow” Point: One way to determine the optimal number of clusters is to look for the point where the WCSS drop becomes flatter, known as the “elbow”. From the data provided, the elbow point appears to be around k = 4 to k = 6, where the decline in

Jurnal Scientia
Volume 13, Number 03, 2024, DOI 10.58471/ scientia.v13i03
ESSN 2723-7486 (Online)
https://infor.seaninstitute.org/index.php/pendidikan

WCSS starts to be slower. Choosing a value of k in this range may offer a balance between minimizing WCSS and model complexity.

2. Diminishing Returns: After a certain point, increasing the number of clusters does not provide a significant decrease in WCSS. This suggests that adding more clusters no longer provides a substantial advantage in reducing variation in clusters.
3. Interpretation Practical: In addition to using mathematical criteria such as WCSS, it is also important to consider practical interpretation and the desire for easy-to-understand segmentation. If the model is too complex (with too many clusters), its interpretation can become difficult, and there may not be significant differences in actions taken based on different groups. Overall, choosing the optimal number of clusters is about striking a balance between a simple but fairly descriptive model, and this analysis suggests that the sweet spot is likely to be around 4 to 6 clusters. The following will show a visualization of the elbow curve for each e-commerce:

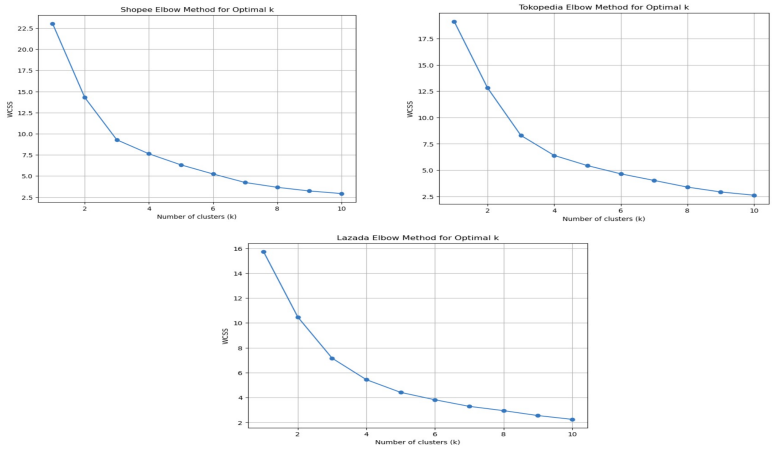

Figure 3. Elbow Curve Shopee, Tokopedia, dan Lazada
Resource: (Hasil Olahan Penulis, 2024)

Here is the comparison of the characteristics of the six clusters based on RFM data and mapping for each e-commerce platform (Shopee, Tokopedia, Lazada):

a. Engaged Bargain Hunters is a customer cluster that actively seeks the best deals and often shops with a focus on discounted or low-priced products. Although they have varying shopping frequencies, they tend to be more responsive to promotions and special offers that provide greater value for their money.
    1. Shopee Customers: They have relatively recent interactions and shop with moderate frequency, but their transaction values tend to be low. They are interested in discount offers and promotions.
    2. Tokopedia Customers: They shop infrequently but may make large purchases at discounted prices, indicating a tendency to seek higher transaction value.
    3. Lazada Customers: Although they shop rarely and have low frequency and transaction values, they are still interested in products with special offers.

    Marketing Strategy Recommendation: Promotions and Discounts, Exclusive Offer Campaigns, and Loyalty Programs.

b. High-Spending Recents is a customer cluster that has recently made purchases with high transaction values. They are customers who have recently interacted with the e-commerce platform and made significant purchases.
   1. Shopee Customers: Although they have relatively recent interactions, they tend to make purchases with low frequency but with high transaction values.
   2. Tokopedia Customers: Despite making purchases recently with high frequency, their transaction values remain significant.
   3. Lazada Customers: They make purchases with high transaction values and tend to have high frequency, indicating active engagement with the offers or products they buy.

   Marketing Strategy Recommendation: Exclusive and Premium Offers, Personalized Campaigns, and Membership or VIP Programs.

c. Loyal Big Spenders is a customer cluster that exhibits highly profitable shopping patterns with high frequency and substantial transaction values, while maintaining consistent engagement with the e-commerce platform. They are customers who not only shop frequently but also tend to spend significant amounts of money each time they make a transaction.
   1. Shopee Customers: They may shop frequently with high transaction values, indicating loyalty and a tendency to purchase more expensive or larger quantities of products.
   2. Tokopedia Customers: They also show high frequency and large transaction values, signaling their commitment to the platform and potential for purchasing premium or bulk items.
   3. Lazada Customers: They are active with high frequency and transaction values, demonstrating strong engagement and a tendency to shop for premium-priced products.

   Marketing Strategy Recommendation: Exclusive and Premium Offers, Loyalty and Rewards Programs, Personalized Campaigns, and Special Events.

d. Frequent Value Seekers is a customer cluster that shops frequently but tends to look for deals or discounts. They are active buyers who make transactions regularly but prioritize value or price in their purchasing decisions
   1. Shopee Customers: They are active shoppers with high frequency but are more price-sensitive and seek out the best deals.
   2. Tokopedia Customers: They shop frequently but focus on getting the best value for their money, often taking advantage of discounts or offers.
   3. Lazada Customers: They also shop frequently and look for good value in their transactions, frequently leveraging promotions or discounted prices.

   Marketing Strategy Recommendation: Offers and Discounts, Price-Based Campaigns, Discount-Based Loyalty Programs, and Seasonal and Special Event Offers.

e. Infrequent High-Value Shoppers is a customer cluster that makes infrequent purchases but tends to spend large amounts of money when they do. They prioritize high value in each purchase, even though they do not shop regularly.

Jurnal Scientia
Volume 13, Number 03, 2024, DOI 10.58471/ scientia.v13i03
ESSN 2723-7486 (Online)
https://infor.seaninstitute.org/index.php/pendidikan

1. Shopee Customers: They make high-value purchases but do so infrequently, suggesting they may only buy large or expensive items occasionally.
2. Tokopedia Customers: They tend to spend a lot per transaction despite their low shopping frequency, indicating a tendency to purchase high-value items sporadically.
3. Lazada Customers: They have a similar pattern, with infrequent purchases but high transaction values, suggesting they may buy premium or large-quantity items.

Marketing Strategy Recommendation: Exclusive and Premium Offers, Special Loyalty Programs, Personalized Campaigns, and Exclusive Events and Limited Offers.

f. Lapsed Premium Spenders is a customer cluster that previously had a habit of purchasing high-value products but now makes purchases infrequently. They are customers who were once active with significant spending but have recently stopped or rarely transacted.
   1. Shopee Customers: They used to shop frequently with high transaction values but now rarely make purchases. They may have previously bought premium or expensive items on Shopee but have been inactive recently.
   2. Tokopedia Customers: They had a similar purchasing pattern, spending a lot per transaction on Tokopedia in the past, but their shopping frequency has decreased.
   3. Lazada Customers: They also have a history of high-value purchases on Lazada but have not been active recently, suggesting they may have stopped buying premium products on this platform.

   Marketing Strategy Recommendation: Reactivation with Premium Offers, Personalized Premium Product Campaigns, Loyalty Programs and Incentives, Educational Outreach.

CONCLUSIONS

The conclusions from the research entitled “Analysis of Customer Segmentation in the Top Three Most Visited Commerce Platforms in Indonesia in 2023 Using the RFM Model and Clustering Technique" are: (1) Based on the RFM analysis using the K-Means clustering method which was carried out, this research identified six customer behavior clusters on Shopee, Tokopedia, and Lazada, namely Engaged Bargain Hunters, High-Spending Recents, Loyal Big Spenders, Frequent Value Seekers, Infrequent High-Value Shoppers, and Lapsed Premium Spenders. Each cluster shows different shopping characteristics. The significant variation in shopping behavior across these three platforms emphasizes the need for tailored marketing strategies for each e-commerce site. (2) Mapping customer characteristics across these three platforms shows that while the same clusters can be identified across the three platforms, the characteristics and behaviors of customers within these clusters can vary significantly, reflecting different strategies, product offerings, and demographics on each e-commerce site. These differences highlight variations in customer behavior patterns, shopping preferences, and transaction values across platforms. Therefore, even though the number of clusters formed is nominally the same (six clusters), each e-commerce platform has unique customer dynamics and characteristics. This

highlights the importance of a tailored approach to marketing strategy and service development for each e-commerce platform, based on specific and unique RFM analysis for each customer segment in order to maximize customer retention and transaction value. Personalized approaches and relevant offers are key to optimizing customer spend across these platforms.

REFERENCE

Annur, C. M. (2024, Januari 10). Jumlah Kunjungan ke 5 Situs E-Commerce Terbesar di Indonesia (Januari-Desember 2023). Diambil kembali dari databoks: https://databoks.katadata.co.id/datapublish/2024/01/10/ini-persaingan kunjungan-5-e-commerce-terbesar-di-indonesia-sepanjang-2023

Apuke, O. D. (2017, October 12). Quantitative Research Methods: A Synopsis Approach. Arabian Journal of Business and Management Review (Kuwait Chapter), 6(10). doi:10.12816/0040336

Asra, A., & Prasetyo, A. (2015). Pengambilan Sampel dalam Penelitian Survey. Jakarta: PT RajaGrafindo Persada.

Dogan, O., Aycin, E., & Bulut, Z. A. (2018). Customer Segmentation by Using RFM Model and Clustering Methods: A Case Study in Retail Industry. International Journal of Contemporary Economics and Administrative Science, 8(1), 1-19.

Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. Journal of Marketing Research, 42(4), 415 430. doi:10.1509/jmkr.2005.42.4.415

Fox, G. (2020, October 25). Customer Segmentation – How To Segment Customers A Step By Step Guide For Growth. Diambil kembali dari Gary Fox Strategy and Innovation Consultant:https://www.garyfox.co/customersegmentation/#examples_of_geographic_segmentation

Gupta, A. (2014). E-Commerce: Role of E-Commerce in Today's Business. International Journal of Computing and Corporate Research, 4(1).

Mordore Intelligence. (2024). INDONESIA E-COMMERCE MARKET SIZE & SHARE ANALYSIS - GROWTH TRENDS & FORECASTS (2024 - 2029). Diambil kembali dari Mordore Intelligence:https://www.mordorintelligence.com/industry-reports/global-ecommerce-market

Scridon, M. A. (2008). Understanding Customers - Profiling and Segmentation. Management & Marketing, 175-184.