IDENTIFYING CUSTOMER SEGMENTATION AND PERSONA OF AMAZON CUSTOMER: AN APPROACH USING K-MEANS CLUSTERING

Rizki Agung Ramadani[1]
Universitas Bakrie, Jakarta, Indonesia
rizkiagungramadani@gmail.com

Tiyanda Hanti Arum Kusuma[2]
Universitas Bakrie, Jakarta, Indonesia
yandahanti@gmail.com

Rinta Agustiani Dwiputri[3]
Universitas Bakrie, Jakarta, Indonesia
rintaagustiani@gmail.com

Jerry Heikal[4]
Universitas Bakrie, Jakarta, Indonesia
jerry.heikal@bakrie.ac.id

Abstract

Technological developments have transformed traditional buying and selling practices into online transactions. Amazon, as one of the largest e-commerce platforms, continues to innovate to maintain its competitive advantage, offering a variety of products and services. This research uses the K-Means Clustering method to identify Amazon's customer segmentation and devise more effective marketing strategies. The analysis results show three main clusters: Price-Sensitive Browsers, Review-Driven Shoppers, and Quality Seekers. Cluster 2, which accounts for 47.49% of the total sample, is the most potential. Consumers in this cluster shop weekly, want better prices, pay attention to product reviews, and care about eco-friendly packaging. The right target segment for Amazon is women aged 24-26 who regularly shop weekly and care about environmental sustainability. By understanding consumer needs and preferences, Amazon can develop more effective marketing strategies, increase customer satisfaction and loyalty, and maintain its competitive advantage.

Keywords: Amazon, E-Commerce, K-Means, Clustering, Segmentation

INTRODUCTION

The development of technology around the world is so rapid that it has become an important element in daily activities. This technology has even affected people's habits and behavior, especially in buying and selling activities. People today are very familiar with the use of the internet, which makes it easier for them to access information related to the products they want to buy. Traditional buying and selling practices that prioritize direct meetings between sellers and buyers have begun to transform into online buying and selling systems. Consumer behavior, which is starting to be influenced by technology, has made a significant change from traditional transactions to online transactions through various e-commerce platforms.

E-commerce is a new concept that can be described as the process of buying and selling goods or services using the World Wide Web Internet or the process of buying and selling, or exchanging products, services, and information through information networks. According to Laudon and Laudon, e-commerce is a process of buying and selling products electronically by consumers and from company to company with computer media or the internet as an intermediary for business transactions. Turban also explained that e-commerce means electronic commerce, which includes the process of buying, selling, transferring, or exchanging products, services, or information through computer networks, including the internet. One of the largest and global e-commerce platforms is Amazon.

Amazon is an e-commerce giant founded by Jeff Bezos in 1994 and has become one of the most influential companies in the world. Amazon was originally an e-commerce that focused on buying and selling books online which allowed customers to buy books easily with various facilities, such as reviews and rating information of the book in question by only needing to make one click on the mouse to buy a book, and equipped with a search engine to facilitate book searches. However, Amazon is always innovating to maintain its competitive advantage, so that today Amazon is not only a site for buying and selling books but also offers a variety of products and services, including e-books, hardware, streaming services, cloud services, and daily necessities such as clothing, furniture, food, toys, and jewelry.

Being a company with a large consumer base, Amazon has varied consumer behaviors that will affect purchasing decisions and consumer loyalty to Amazon. Amazon

itself has achieved the highest customer loyalty rate among other e-commerce platforms, at 89%. Especially with the “Prime” membership program, which contributes quite high in supporting Amazon's customer loyalty. However, of the 85% of Amazon visitors who are Prime members, only 46% decide to make a purchase. This is due to consumer characteristics that can affect their consumer behavior.

Seeing this phenomenon, it is important to conduct an STP (segmentation, targeting, positioning) analysis on Amazon. The purpose of STP analysis itself is to create value for target consumers. Segmentation needs to be done so that Amazon can divide the market into smaller segments that can be reached more efficiently and effectively with products that match the unique needs of consumers. Targeting can be done by Amazon to evaluate each segment's interest in the market and target more specific consumers. Meanwhile, positioning helps Amazon in structuring market offerings to place a competitive position with competitors that can be embedded in the minds of consumers.

Clustering is one of the methods used for customer segmentation. In clustering, the process of grouping data is done based on data characteristics, where the data will be placed into several clusters. According to Kadarsah and Heikal, clustering is a process of grouping data into groups naturally. This is usually used in marketing to divide customers into different homogeneous groups or also known as market segmentation. Wicaksono and Heikal explained that cluster analysis can be used to identify individuals or newly entered samples. With clustering, data that has the same pattern will be grouped into the same cluster, while data with different patterns will be grouped into different clusters. By doing clustering in this research, it can be determined what the Amazon consumer persona looks like based on their behavior, which in turn can help understand the needs, wants, and preferences of each segment, thus enabling the development of a more targeted marketing strategy.

There are various methods for performing clustering techniques; one of the most commonly used methods is K-means clustering. Saputra et al. explained that the application of K-means clustering can help companies to overcome challenges related to the difficulty in identifying differences in customer segmentation and making it a marketing strategy. With the K-means clustering method, companies can find different market segments and determine the marketing strategy design that suits each segment. This method allows companies to

group customers based on similar consumer preferences and behaviors, facilitating a deeper understanding of each segment, so that companies can develop more targeted, relevant, and efficient marketing strategies for each customer group.

Based on the above background, it can be concluded that identifying consumer segments is needed by companies, especially those with a variety of consumer behaviors, such as Amazon, to produce more effective and efficient marketing strategies. The purpose of this research is to identify and analyze the segmentation of Amazon customers by using the K-means clustering method to obtain personas and analyze customers who have high prospects and contributions to Amazon's targeting and positioning.

RESEARCH METHOD

The segmentation method used in this research is K-Means Clustering, which serves to group data into appropriate clusters. This algorithm is capable of processing data without category labels. The stages of this research can be seen in the flowchart (Figure 1), starting from the data collection process obtained from the site www.kaggle.com. Next is the data pre-processing stage, where the dataset that has been obtained is processed and adjusted to fit the format needed for data processing in the SPSS application. The next stage is data processing using K-Means Clustering in SPSS to obtain three clusters with respect to the existing variables. After that, proceed with selecting the most promising cluster and positioning within the cluster.

The following is the flowchart of this research:

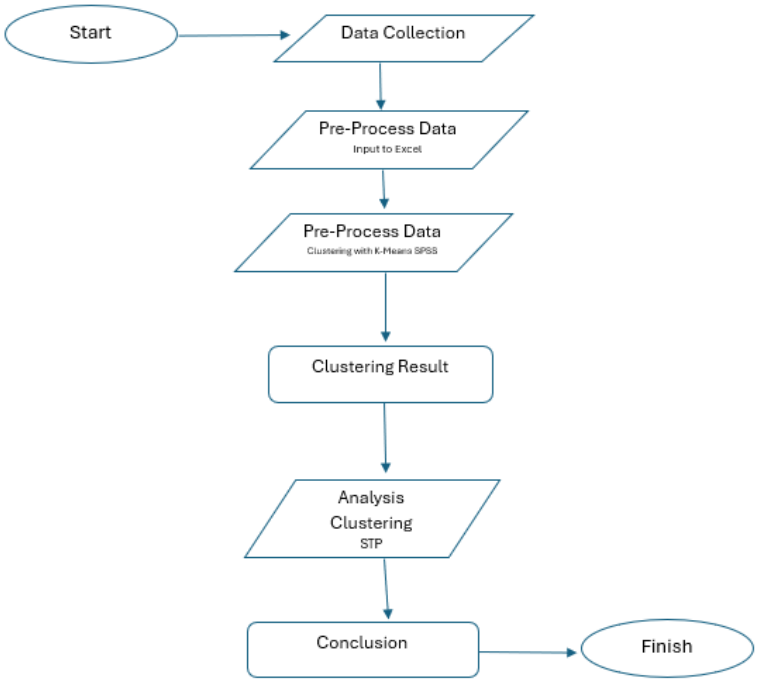

Figure 1.
Research Flowchart

RESULTS AND DISCUSSION

The population in this study is all Amazon users who shop for various types of goods in several cities in the United States, which is Amazon's largest market share, while the sample used is 598 respondents. In the dataset used, there are 17 criteria containing Customer Experience, Age, Gender, Personalized Recommendation, Browsing Frequency, Product Search Method, Add to Cart Browsing, Cart Abandonment Factors, Review Helpfulness, Recommendation Helpfulness, Purchase Frequency, Purchase Categories, Cart Completion Frequency, Save for later Frequency, Review Reliability, Service Appreciation and Improvement Areas. This data is analyzed with K-Means Clustering into 3 clusters to see the persona of each cluster. Of the 3 clusters analyzed, it can be seen that the dominant distribution of each cluster is dominated by cluster 2 as many as 284 respondents or 47.49%, while the lowest is in cluster 3 as many as 94 respondents or 15.71%.

Table 1.

Number of Cases in Each Cluster

Number of Case of each Cluster		
Cluster	**Jumlah**	**Persentase**
1	220	36,79
2	284	47,49
3	94	15,72
Total	**598**	**100**

Table 2.

Final Cluster Center

	Cluster	1	2	3
Customer Experience	Customer Reviews Importance	0,191	0,334	0,118
	Review Left	0,514	0,546	0,426
	Rating Accuracy	0,205	0,309	0,281
	Shopping Satisfaction	0,165	0,309	0,242
Age	24-26	0,241	0,373	0,426
	27-36	0,345	0,271	0,160
	37-67	0,305	0,246	0,202
Gender	Female	0,564	0,574	0,660
	Male	0,282	0,180	0,309
	Prefer Not To Say	0,136	0,194	0,032
Personalized Recommendation	No	0,432	0,313	0,691
	Sometimes	0,300	0,507	0,191
Browsing Frequency	Few Times A Month	0,295	0,363	0,330
	Few Times A Week	0,427	0,426	0,330
	Rarely	0,100	0,116	0,234
Product Search Method	Categories	0,264	0,458	0,351
	Filter	0,250	0,229	0,074
	Keyword	0,464	0,225	0,511
	Multiple Pages	0,732	0,701	0,840
Add to Cart Browsing	Maybe	0,277	0,581	0,213
	Yes	0,441	0,190	0,681
Cart Abandonment Factors	Changed My Mind Or No Longer Need The Item	0,423	0,415	0,309
	Found A Better Price Elsewhere	0,432	0,415	0,415
	High Shipping Costs	0,109	0,095	0,202

Review Helpfulness	Sometimes	0,264	0,528	0,181
	Yes	0,477	0,211	0,755
Recommendation Helpfulness	No	0,327	0,292	0,170
	Sometimes	0,286	0,553	0,532
Purchase Frequency	Few Times A Month	0,359	0,352	0,245
	Monthly	0,368	0,306	0,649
	Weekly	0,173	0,246	0,043
Purchase Categories	Beauty And Personal Care	0,091	0,275	0,074
	Clothing And Fashion	0,100	0,218	0,223
	Home And Kitchen	0,055	0,025	0,043
	Mixed	0,686	0,373	0,489
	Others	0,055	0,070	0,170
Cart Completion Frequency	Often	0,405	0,165	0,234
	Rarely-Never	0,095	0,201	0,160
	Sometimes	0,377	0,592	0,532
Save for later Frequency	Often	0,395	0,169	0,191
	Rarely-Never	0,159	0,222	0,457
	Sometimes	0,350	0,539	0,223
Review Reliability	Heavily	0,268	0,120	0,596
	Moderately	0,445	0,261	0,255
	Occasionally	0,214	0,461	0,117
Service Appreciation	Competitive Prices	0,382	0,264	0,234
	Product Recommendations	0,177	0,486	0,085
	User-friendly Web-App Interface	0,118	0,109	0,234
	Wide Product Selection	0,309	0,141	0,426
Improvement Areas	Customer Service Responsiveness	0,364	0,423	0,149
	Product Quality And Accuracy	0,286	0,190	0,447
	Reducing Packaging Waste	0,209	0,268	0,117
	Shipping Speed And Reliability	0,132	0,109	0,191

a. Persona Cluster 1: **Price-Sensitive Brows**

Cluster 1 has the name Price-Sensitive Brows which consists of 36.79% of the total sample. Customers in this cluster have the following characteristics:

Customer Experience : -

Age : 27 – 67 y.o

Gender : Women

Personalized Recommendation : Sometimes

Browsing Frequency	: Few times a week
Product Search Method	: Filter
Add to Cart Browsing	: -
Cart Abandonment Factors	: Change my mind or no longer need the item and found a better price elsewhere
Review Helpfulness	: -
Recommendation Helpfulness	: No
Purchase Frequency	: Few times a month
Purchase Category	: Home and Kitchen, Mixed
Cart Completion Frequency	: Often
Save for later Frequency	: Often
Review Reliability	: Moderately
Service Appreciation	: Competitive Price
Improvement Areas	: -

In Cluster 1 customers do not seem to place much importance on personalized reviews or recommendations. They have a low level of shopping satisfaction and rating accuracy. The majority of customers in this cluster are between 27-67 years old and the majority are female. Customers in this group occasionally use personalized recommendations and tend to search for products by typing keywords. In terms of shopping behavior, they often change their minds or cancel purchases because they no longer need the items in their cart. Even so, they have a high tendency to complete transactions in the cart

b. Persona Cluster 2: **Review-Driven Shoppers**

Cluster 2 has the name Review-Driven Shoppers which consists of 47.49% of the total sample. Customers in this cluster have the following characteristics:

Customer Experience	: Customer Review Importance; Review left; Rating Accuracy and Shopping Satisfaction
Age	: 24 – 26 y.o
Gender	: Prefer not to say
Personalized Recommendation	: Sometimes
Browsing Frequency	: Few time a month

Product Search Method	: Category
Add to Cart Browsing	: Maybe
Cart Abandonment Factors	: -
Review Helpfulness	: Sometimes
Recommendation Helpfulness	: Sometimes
Purchase Frequency	: Weekly
Purchase Category	: Beauty and Personal Care
Cart Completion Frequency	: Sometimes
Save for later Frequency	: Sometimes
Review Reliability	: Occasionaly
Service Appreciation	: Product Recommendations
Improvement Areas	: Customer Service Responsiveness; Reducing PackagingWaste

Cluster 2 consists of customers who pay close attention to product reviews and accept personalized recommendations at a moderate level. They are satisfied with the accuracy of the ratings, and their shopping satisfaction level is also moderate. This group is dominated by women aged 24-26. They tend not to use personalized recommendations and prefer to browse products through categories. In shopping behavior, these customers often put products in the cart for consideration but end up canceling the transaction after finding a better price elsewhere. They tend to shop weekly, find product reviews sometimes helpful, and highly value product recommendations provided by platforms. Customers in this cluster also have a particular concern for reducing packaging waste in the service.

c. Persona Cluster 3: **Quality Seekers**

Cluster 3 has the name Quality Seekers, which consists of 15.72% of the total sample. Customers in this cluster have the following characteristics:

Customer Experience	: -
Age	: 24 – 26 y.o
Gender	: Male and Female
Personalized Recommendation	: No

Browsing Frequency	: Rarely
Product Search Method	: Keyword and Multiple Page
Add to Cart Browsing	: Yes
Cart Abandonment Factors	: Hight Shipping Cost
Review Helpfulness	: Yes
Recommendation Helpfulness	: -
Purchase Frequency	: Monthly
Purchase Category	: Clothing and Fashion
Cart Completion Frequency	: Rarely-Never
Save for later Frequency	: Rarely-Never
Review Reliability	: Heavily
Service Appreciation	: User-friendly web-app interface and wide product selection
Improvement Areas	: Product quality and Accuracy; Shipping speed and reliability

Cluster 3 does not place much importance on reviews but often receives personalized recommendations. They value high rating accuracy and are satisfied with their shopping experience. The majority of customers in this group are female, and they rarely use personalized recommendations to guide their shopping. They tend to search for products through categories, but often change their minds before finalizing a purchase. Their shopping frequency is monthly, and they rarely save items for later purchase and tend not to always complete transactions in the cart. They find product reviews helpful and rely on these recommendations, especially when appreciating a wide product selection. This cluster points to attention to product quality and accuracy as an area that Amazon can improve.

Based on the three clustering, the most potential target segmentation for Amazon is in Cluster 2 with the highest percentage of consumers at 47.49%. Consumers in Cluster 2 tend to have a high frequency of purchase, i.e. weekly shopping which shows high potential to increase short and long-term income. This cluster also shows specific needs such as better prices and cares about product reviews and there is a tendency to care about

environmentally friendly packaging, making it easier for Amazon to customize services and promotions.

Based on these characteristics, the right target for Amazon is women aged 24-26 who regularly do weekly shopping and pay attention to product reviews and care about environmental sustainability. Amazon's positioning for the Cluster 2 target segment can highlight friendly prices, sustainability values, and commitment to the needs of young consumers who are environmentally conscious. Reliability of product reviews is also an important positioning for this segment in order to increase consumer confidence.

CONCLUSION

Based on the segmentation analysis using the K-Means Clustering method, it's clear that **Cluster 2**, which focuses on young consumers, holds the greatest potential for boosting Amazon's revenue. This group, primarily women aged 24-26, exhibits high purchase frequency and specific needs, such as competitive prices, useful product reviews, and a strong concern for environmental sustainability. They tend to shop weekly, pay close attention to product reviews, and show a keen interest in eco-friendly packaging. To effectively target this segment, Amazon should implement marketing strategies that highlight competitive pricing to attract budget-conscious shoppers, promote eco-friendly packaging and sustainability initiatives to resonate with their environmental concerns, and ensure the availability of detailed and trustworthy product reviews to aid their purchasing decisions. By focusing on these aspects, Amazon can develop more relevant services and promotions, increase consumer loyalty, and strengthen its position in the e-commerce market. Understanding the shopping behavior of this segment allows Amazon to design personalized and relevant marketing campaigns. This could include offering competitive prices and attractive promotions. Given their concern for sustainability, Amazon can optimize eco-friendly packaging and promote sustainability initiatives across all aspects of their business, including marketing campaigns and loyalty programs.

REFERENCES

Anderson, E. W., & Sullivan, M. W. (1993). The Antecedents and Consequences of Customer Satisfaction for Firms. *Marketing Science*, *12*(2), 125–143.

https://doi.org/10.1287/mksc.12.2.125

A. S. M. AlQahtani. 2021. "Product Sentiment Analysis for Amazon Reviews," International
Journal of Computer Science and Information Technology. Vol. 13, No. 3, pp. 15–30, Jun. 2021, doi: 10.5121/ijcsit.2021.13302

Adhani, L. K., Dharmastiti, R., & Trapsilawati, F. (2020). Pengaruh Waktu Sebelum Dan Selama Pandemi Covid-19 Terhadap Perilaku Konsumen Belanja Online. Perspektif Keilmuan Teknik Industri Pada Era New Normal, 50–55.

Apriani, A., & Heikal, J. (2024). Segmentation Analysis Using K-Means Clustering Model with SPSS Case Study of Backpacker Jakarta Community Members. *Jurnal Indonesia Sosial Sains*, *5*(03), 441–460.

Ariati, I., Norsa, R. N., Akhsan, L., & Heikal, J. (2023). Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Pelanggan Uht Milk Greenfield. *Cerdika: Jurnal Ilmiah Indonesia*, *3*(7), 729–743.

Bagus, D. (2010). E-commerse: Definisi Jenis, Tujuan, Manfaat dan Ancaman Menggunakan ECommerce. Diakses dari http://jurnal-sdm.blogspot.com/2009/08/e-commerse-definisi-jenistujuan.html

Benhardy, K. A., Hardiyansyah, Putranto, A., & Ronadi, M. (2020). Brand image and price perceptions impact on purchase intentions: Mediating brand trust. *Management Science Letters*, *10*(14), 3425–3432. https://doi.org/10.5267/j.msl.2020.5.035

Creswell, J. . (2014). *Penelitian Kualitatif & Desain Riset*. Pustaka Pelajar.

David. (2023). ANALISIS PENGARUH UI/UX APLIKASI TOKOPEDIA TERHADAP KEPUASAN PENGGUNA. *Jurnal Ilmu Komputer dan Sistem Informasi*, *11*(1). https://doi.org/10.24912/jiksi.v11i1.24088

Farhan, M., & Heikal, J. (2024). Used Car Customer Segmentation Using K-Means Clustering Model With SPSS Program: Case Study Caroline. Id. *Jurnal Indonesia Sosial Sains*, *5*(03).

Hizbandyah B.N., Silvia L., Nurahma S., Prawira I. F. A. (2023). Inovasi Amazon dalam Menerapkan Teknologi sebagai Strategi Bisnis terhadap Keunggulan Bersaing. *Jurnal Teknik Informatika dan Sistem Informasi*. 10 (4): 361-373.

Indriantoro, N., & Supomo. (2002). *Metodologi Penelitian Bisnis Untuk. Akuntansi dan Manajemen*. BPFE.

Kadarsah, D., & Heikal, J. (2024). Customer Segmentation With K-Means Clustering Suzuki Mobil Bandung Customer Case Study. *Jurnal Indonesia Sosial Teknologi*, *5*(3), 768–774.

Maulana, E. B., Hadiana, A. I., & Umbara, F. R. (2022). Segmentasi Pengunjung Pusat Perbelanjaan Menggunakan Metode K-Means Clustering. Seminar Nasional Sistem Informasi (Senasif) 2022, 6(1), 3093–3102.

Murpratiwi, S. I., Agung Indrawan, I. G., & Aranta, A. (2021). Analisis Pemilihan Cluster Optimal Dalam Segmentasi Pelanggan Toko Retail. Jurnal Pendidikan Teknologi Dan Kejuruan, 18(2), 152. https://doi.org/10.23887/jptk-undiksha.v18i2.37426

Saputra, T. B. C., Fadhilah, S. M., Mangkuto, S. M., Heikal, J. (2024). Segmentation, targeting and positioning analysis using k- means clustering model: A case study of the laptop market in Indonesia. *International Journal of Applied Finance and Business Studies, 12 (2) : 195-203.*

Sugiyono. (2007). *Metode Penelitian Kuantitatif Kualitatif Dan R&D*. Alfabeta.

Sugiyono. (2017). *Metode Penelitian Evaluasi (Pendekatan Kuantitatif, Kualitatif, dan Kombinasi)*. Alfabeta.

Turban, E., Rainer, R. K., Potter, R. E. (2005). Introduction to Information Technology. US: John Wiley and Sons, inc.

Wicaksono, M. H. S., & Heikal, J. (2024). Identifying Customer Segmentation and Persona of Soft Drink Industry in India: An Approach Using K-Means Clustering. *Jurnal Ekonomi, Koperasi, & Kewirausahaan*, *15(3), 1-10.*

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The Behavioral Consequences of Service Quality. *Journal of Marketing*, *60*(2), 31. https://doi.org/10.2307/1251929