Tibuana
Journal of applied Industrial Engineering-University of PGRI Adi Buana
p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11641 *e*-ISSN 2622-2035

Customer Segmentation in a Campus-Area Convenience Store Using K-Means Clustering

Musaddad Alfani[1], and Putri Aysha Qalbi[2]
[1,2] Department of Industrial Engineering, Institut Teknologi Sumatera

Email: musaddad.alfani@ti.itera.ac.id

***Abstract*—Customer segmentation is an important strategy for understanding differences in customer characteristics and supporting more targeted marketing decisions. In campus areas, convenience store customers generally consist of students with diverse demographic, economic, and accessibility characteristics, resulting in different purchasing behaviors. This study aims to identify customer segments of a convenience store located in the Campus X area of Yogyakarta using the K-Means Clustering algorithm. The study employed a quantitative approach with data mining techniques using customer data from students. The variables analyzed included age, income, shopping intensity, and residential distance from the store. The optimal number of clusters was determined using the Elbow Method before applying the K-Means algorithm. The results identified three customer segments with distinct characteristics. The largest segment consisted of customers living relatively close to the store and exhibiting high shopping intensity, while another segment was characterized by lower purchasing activity and greater residential distance. The smallest segment showed the highest income level and the most diverse product preferences. Furthermore, differences in customer characteristics were reflected in product purchasing patterns across the identified segments. These findings indicate that accessibility and economic factors play important roles in shaping customer purchasing behavior in campus-area convenience stores. The results can serve as a reference for developing more targeted promotional and product management strategies.**

***Keywords:* Customer Segmentation, K-Means Clustering, Convenience Store, Consumer Behavior**

I. Introduction

Yogyakarta is widely recognized as one of Indonesia's major higher education centers, attracting students from various regions due to its high concentration of universities and vibrant academic environment [1]. This condition has stimulated the growth of economic activities that cater to student needs, including the modern retail sector, particularly convenience stores [2]. Convenience stores not only serve as providers of daily necessities but also constitute an integral part of student consumption activities, which are characterized by diverse purchasing patterns and consumer preferences [3]. Consequently, competition among convenience stores in campus areas is no longer solely determined by product availability but also by the ability of retailers to understand and identify customer characteristics more accurately [4].

Students, as the primary consumer group in campus areas, possess diverse social and economic backgrounds [5]. These differences contribute to variations in purchasing power and financial management capabilities among students [6]. Furthermore, student populations are heterogeneous not only in terms of socioeconomic characteristics but also in residential location. While some students reside close to campus and enjoy easier access to retail facilities, others live farther away. Such differences in economic conditions and accessibility may influence purchasing power, shopping frequency, and consumption patterns at convenience stores located around campus areas [7]. In addition, the spatial distribution of student residences within varying distances from campus makes residential distance an important factor that may affect customer mobility and shopping behavior toward nearby retail stores [8].

The management of convenience stores surrounding Campus X in Yogyakarta faces increasingly complex challenges due to customer mobility and substantial variations in purchasing

p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11641 *e*-ISSN 2622-2035

power across consumer groups. The growing diversity of customer characteristics and needs has resulted in a more heterogeneous market. This situation highlights the importance of market segmentation as an approach to grouping consumers into relatively homogeneous segments, enabling marketing strategies to be tailored according to the characteristics of each segment [9]. As a result, mass marketing approaches are becoming less effective because they fail to capture the specific needs of distinct customer groups. Therefore, a data-driven approach is required to map customer characteristics more comprehensively by considering not only demographic attributes but also spatial factors such as residential distance and customers' economic capacity [10].

Although previous studies have examined retail customer segmentation using demographic and purchasing behavior attributes, most have focused primarily on consumer characteristics without incorporating spatial dimensions that influence retail accessibility. In educational areas characterized by high student mobility, the distance between students' residences and convenience stores may significantly affect shopping frequency and consumption patterns. Better accessibility enables consumers to reach retail locations more easily, potentially increasing store visits and purchasing activities [11]. Therefore, integrating demographic, economic, and consumption-related variables into the customer segmentation process is essential for developing a more comprehensive market profile.

Unlike conventional market segmentation approaches that primarily focus on psychographic or demographic attributes, this study integrates a spatial dimension, represented by accessibility (distance), together with economic factors (income), using the K-Means clustering algorithm. This approach is considered particularly relevant in the context of convenience stores located in educational areas, where student purchasing behavior is strongly influenced by geographic proximity and varying levels of monthly financial resources. Furthermore, the use of real transactional data collected from a strategically located convenience store in Yogyakarta enhances the empirical validity of the resulting model and provides a more accurate representation of market dynamics. Accordingly, this study aims to identify and classify convenience store customers in a campus area using the K-Means clustering algorithm based on demographic, economic, and accessibility-related attributes.

This study is expected to contribute to the literature on retail customer segmentation by integrating demographic, economic, and spatial attributes within a data mining–based clustering framework. From a practical perspective, the findings may assist convenience store managers in designing more targeted marketing strategies, including differentiated promotional programs, product assortment adjustments, and service management initiatives tailored to the characteristics of the identified customer segments.

II. METHODS

This study employed a quantitative approach using data mining techniques to identify customer segmentation patterns in a convenience store located within a campus area in Yogyakarta. The dataset consisted of customer data collected from a convenience store operating near Campus X, with a total sample of 100 students. The variables analyzed included gender, age, income, shopping intensity, and residential distance from the store. These variables were selected because they represent customers' demographic, economic, and accessibility characteristics. In addition to customer profile data, this study utilized transaction data that were grouped into four product departments, namely snacks and bakery products (Dept 1), personal care products (Dept 2), beverages and dairy products (Dept 3), and household and toiletry products (Dept 4).

Customer segmentation was conducted using the K-Means algorithm, a non-hierarchical clustering technique that groups data based on the similarity of their characteristics [12]. The algorithm partitions observations into several clusters according to their proximity to cluster centers (centroids) [13]. The distance between data points and centroids was measured using Euclidean Distance, as expressed in Equation (1).

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

Where:
$d(x,y)$ = The Euclidean distance between a customer data point and a cluster centroid;

Tibuana
Journal of applied Industrial Engineering-University of PGRI Adi Buana
p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11641 e-ISSN 2622-2035

x_i = the value of the i -th attribute of the customer data point;
y_i= the value of the i -th attribute of the centroid;
i = denotes the attribute index.
n = the total number of attributes used in the distance calculation.

The analysis began with data preprocessing and standardization to ensure that all variables were measured on a comparable scale. Subsequently, the optimal number of clusters was determined using the Elbow Method based on the inertia value, also known as the Within-Cluster Sum of Squares (WCSS). After the optimal number of clusters had been identified, the K-Means algorithm was iteratively applied by assigning each customer to the nearest cluster and updating the centroid positions until stable cluster memberships were achieved. The resulting clusters were then analyzed to identify the characteristics of each customer segment based on age, income, shopping intensity, and residential distance. Finally, the clustering results were integrated with product transaction data to examine product category preferences across the identified customer segments.

III. RESULTS AND DISCUSSION

1. Determination of The Optimal Numbers of Clusters

The optimal number of clusters was determined using the Elbow Method by evaluating changes in inertia values across several cluster alternatives (k). This method was employed to identify the most representative number of clusters by observing the point at which the decrease in inertia began to slow down. The results of the Elbow Method are presented in Figure 1.

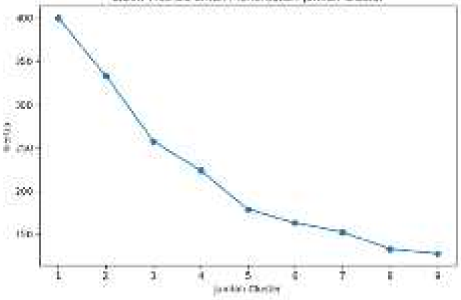

Figure 1. Elbow Method for Determining the Optimal Number of Clusters

As shown in Figure X, the inertia value decreases substantially up to k = 3 and then begins to level off for subsequent cluster numbers. This pattern indicates that increasing the number of clusters beyond k = 3 does not result in a significant reduction in inertia. Therefore, the elbow point can be identified at k = 3.

In this study, k = 3 was selected because it provides a balance between the model's ability to capture data variability and the interpretability of the resulting customer segments. A larger number of clusters may produce more specific customer groups; however, it does not necessarily yield more meaningful insights. Therefore, the use of three clusters was considered sufficient to represent the differences in customer characteristics clearly and in accordance with the objectives of this study.

2. Customer Segmentation Results

The optimal number of clusters was determined using the Elbow Method by evaluating changes in inertia values across several cluster alternatives (k). This method was employed to identify the most representative number of clusters by observing the point at which the decrease in inertia began to slow down. The results of the Elbow Method are presented in Figure 2.

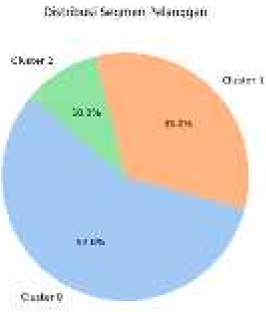

Figure 2. Distribution of Customers Across Clusters

Cluster 0 represents the largest segment, accounting for 57% of the total customers, followed by Cluster 1 with 33% and Cluster 2 with 10%. This distribution indicates that the majority of customers are concentrated within a single dominant segment, while the remaining customers are distributed across two smaller

p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11641 *e*-ISSN 2622-2035

segments. The dominance of Cluster 0 suggests the presence of a customer group with relatively common characteristics compared to the other segments.

Meanwhile, the existence of Cluster 1 and Cluster 2 reflects variations in customer characteristics that are not fully represented by the dominant segment. In particular, Cluster 2, which accounts for the smallest proportion of customers, indicates the presence of a more specific customer group with distinctive characteristics that differ from those of the majority.

Overall, the segmentation results demonstrate that although one segment dominates the customer base, the customer structure remains heterogeneous and can be effectively classified into three major groups. These findings provide a foundation for further analysis of the characteristics and purchasing behavior associated with each customer segment in the following section.

3. Characteristics of Customer Segments

Based on the clustering results, customers can be classified into three distinct segments, each exhibiting different characteristics across the observed variables. The profile of each customer segment is presented in Table 1.

Table 1. Profile of Customer Segments

Cluster	Number of Customers	Percentage (%)	Average Age	Average Income	Average Shopping Intensity	Average Distance
0	57	57%	20.24	1.80	4.30	1.81
1	33	33%	20.36	1.45	2.09	3.61
2	10	10%	22.40	4.45	3.40	4.60

Cluster 0 represents the largest segment, accounting for 57% of the total customers. This segment is characterized by relatively short residential distances, the highest shopping intensity, and moderate-income levels. These characteristics suggest that the segment is largely composed of students living near the campus area, such as in boarding houses, which provides easier access to the convenience store. This geographical proximity may contribute to more frequent visits, which is reflected in the relatively high level of shopping activity. In addition, the average age of customers in this cluster indicates a tendency toward younger students, who generally exhibit routine purchasing behavior to meet their daily needs.

Cluster 1 accounts for 33% of customers and is characterized by greater residential distances, the lowest shopping intensity, and relatively lower income levels compared to the other clusters. These findings suggest that customers in this segment tend to be more selective in their purchasing decisions due to limitations in accessibility and purchasing power. Within the campus context, this segment may represent students living farther from the campus area or residents who do not rely heavily on the convenience store for their daily needs. The relatively low shopping intensity observed in this segment indicates that both distance and budget constraints may play an important role in shaping purchasing behavior.

Cluster 2 represents the smallest segment, comprising 10% of the total customers, yet it exhibits characteristics that differ considerably from the other clusters. This segment has the highest income level, a relatively older average age, and the greatest residential distance, while maintaining a moderate level of shopping intensity. These characteristics indicate the presence of customers with greater purchasing power and higher mobility. In the context of a campus area, this segment may consist of senior students or individuals with additional sources of income. However, since this study did not explicitly examine socio-economic background variables, further interpretation regarding the underlying factors of this segment should be made with caution.

Overall, the segmentation results indicate that customer behavior in convenience stores located within campus areas is influenced by a combination of accessibility and economic capacity. Customers living closer to the store tend to exhibit higher shopping intensity, whereas those located farther away generally

Tibuana
Journal of applied Industrial Engineering-University of PGRI Adi Buana
p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11641 *e*-ISSN 2622-2035

show lower purchasing activity, except for customers with relatively high purchasing power. These findings suggest that integrating accessibility and economic variables into the segmentation process can provide a more comprehensive understanding of customer consumption patterns.

4. Customer Cluster Visualization

The customer segmentation results obtained using the K-Means algorithm can be visualized based on the relationship between income and shopping intensity, as presented in Figure 3.

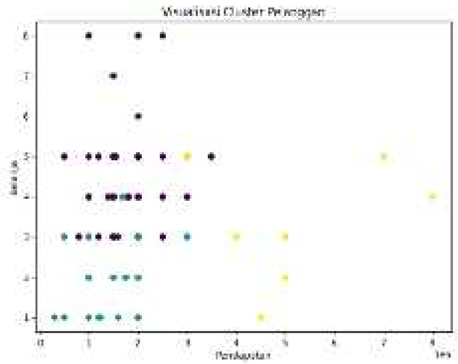

Figure 3. Customer Cluster Visualization

Figure 3 illustrates the distribution of customers according to income and shopping intensity. Each color represents a different cluster, allowing the distribution pattern of customers to be observed visually. In general, a separation can be identified between customers with low-to-moderate income levels and those with relatively higher incomes.

Customers with higher income levels tend to be located on the right side of the graph and exhibit moderate to high shopping intensity. In contrast, most customers are concentrated within the low-to-moderate income range, with a wider variation in shopping intensity. This pattern is consistent with the segmentation results discussed previously, indicating differences in economic characteristics and purchasing behavior among customer segments. Although some overlap between clusters is still visible, the visualization demonstrates that the K-Means algorithm was able to identify customer groups with relatively distinct characteristics. These results further support the effectiveness of the clustering approach in distinguishing customer segments within the context of a convenience store located in a campus area.

5. Product Preferences Across Customer Segments

Based on the customer segmentation results, each cluster exhibits different purchasing tendencies across product categories. The distribution of product purchases for each customer segment is presented in Figure 4.

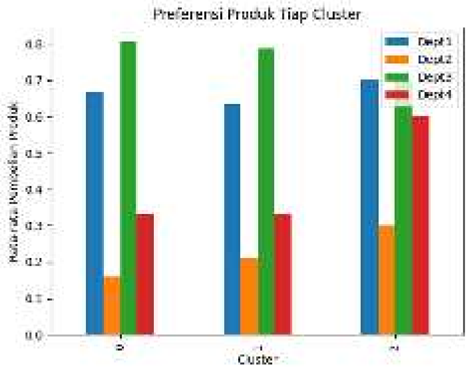

Figure 4. Product Preferences Across Customer Segments

Figure 4 shows variations in purchasing levels across product categories among the identified customer segments. In general, products in Dept 1 and Dept 3 exhibit relatively higher purchase levels than the other categories. In contrast, Cluster 2 demonstrates a more diverse consumption pattern, with relatively high purchase levels across almost all product categories.

To provide a more detailed understanding of product preferences within each customer segment, the average purchase values for each product department are presented in Table 2.

Table 2. Average Product Purchase Values by Customer Segment

Cluster	Dept 1	Dept 2	Dept 3	Dept 4
0	0,67	0,16	0,80	0,33
1	0,63	0,21	0,79	0,33
2	0,70	0,30	0,70	0,60

Overall, Dept 3 represents the most frequently purchased product category among customers in Cluster 0 and Cluster 1, with average values of 0.80 and 0.79, respectively. This finding suggests that customers in these two segments share relatively similar needs and tend to focus on routine consumption products commonly purchased by students. In addition, Dept 1 also

p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11641 *e*-ISSN 2622-2035

records relatively high purchase levels within both segments.

Although Cluster 0 and Cluster 1 exhibit similar product preferences, the characteristics of customers within these segments differ. Cluster 0, which is dominated by customers living closer to the convenience store and displaying higher shopping intensity, demonstrates slightly higher purchase levels than Cluster 1. This finding suggests that accessibility may influence the consumption level of specific product categories.

In contrast to the other segments, Cluster 2 exhibits a more diverse purchasing pattern. Besides showing relatively high purchase levels in Dept 1 and Dept 3, customers in this segment also demonstrate a greater tendency to purchase products from Dept 2 and particularly Dept 4, with average values of 0.30 and 0.60, respectively. These values are considerably higher than those observed in Cluster 0 and Cluster 1, which range from 0.16–0.21 for Dept 2 and approximately 0.33 for Dept 4.

This pattern indicates that customers in Cluster 2 possess a broader range of consumption needs than those in the other segments. The findings are consistent with the characteristics of Cluster 2, which has the highest income level and a relatively older average age. With greater purchasing power, customers in this segment are not only focused on basic consumption products but also tend to purchase products from a wider variety of categories.

Overall, the results indicate that differences in customer characteristics identified through the clustering process are reflected in product purchasing patterns. These findings suggest that product management and promotional strategies can be tailored according to the characteristics of each customer segment. For the dominant segment (Cluster 0), maintaining the availability of products in Dept 1 and Dept 3 should be prioritized. Meanwhile, customers in Cluster 2 may represent a promising target for cross-selling and multi-category promotional strategies due to their tendency to purchase a more diverse range of products.

IV. CONCLUSIONS

This study identified three customer segments of a convenience store located in the Campus X area of Yogyakarta using the K-Means Clustering algorithm based on age, income, shopping intensity, and residential distance attributes. The results indicate that the largest segment consists of customers who live relatively close to the store and exhibit high shopping intensity, while the remaining segments display distinct characteristics in terms of accessibility, purchasing power, and product preferences. Customers with the highest income levels tend to demonstrate more diverse consumption patterns than other customer groups. These findings suggest that accessibility and economic capacity play important roles in shaping customer purchasing behavior within a campus environment.

From a theoretical perspective, this study contributes to the customer segmentation literature by integrating demographic, economic, and accessibility attributes into the clustering process, thereby providing a more comprehensive customer profiling approach than segmentation methods that rely solely on demographic characteristics. From a practical perspective, the segmentation results can assist convenience store managers in developing more targeted marketing strategies, including the design of promotional programs, product category management, and service strategies tailored to the characteristics of each customer segment. Future research is recommended to incorporate additional variables that may provide a deeper understanding of customer differences across segments, such as academic major, year of study, residential status, and socio-economic background. Furthermore, the application of alternative clustering techniques may be explored to compare segmentation consistency and to gain a more comprehensive understanding of customer consumption patterns in educational environments.

REFERENCES

[1] M. Devinta, N. Hidayah, and G. Hendrastomo, "Fenomena Culture Shock (Gegar Budaya) Pada Mahasiswa Perantauan di Yogyakarta," *E-Societas: Jurnal Pendidikan Sosiologi*, vol. 5, no. 3, pp. 1–15, 2016, doi: 10.21831/e-societas. v5i3.3946.

[2] T. A. Widyadhana *et al.*, "Analisis Perilaku dan Preferensi Mahasiswa terhadap Pengalaman Belanja Online dan Offline," *PAJAMKEU: Pajak dan*

p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11641 *e*-ISSN 2622-2035

Manajemen Keuangan, vol. 1, no. 2, pp. 01–13, 2024, doi: https://doi.org/10.61132/pajamkeu.v1i2.77.

[3] A. N. Auliya and H. Hartini, "PENGARUH SALES PROMOTION, STORE ATMOSPHERE DAN KEPRIBADIAN KONSUMEN TERHADAP IMPULSE BUYING PADA KONSUMEN ALFAMART DI KECAMATAN SUMBAWA," *Jurnal Nusa Manajemen*, vol. 3, no. 1, pp. 11–26, 2026, doi: doi.org/10.62237/jnm.v3i1.386.

[4] F. Lakuy and Y. Sopacua, "Antara Harga, Promosi, dan Gaya Hidup: Faktor-Faktor Pembentuk Perilaku Konsumtif Mahasiswa di Era Marketplace Digital," *Populis: Jurnal Ilmu Sosial dan Ilmu Politik*, vol. 19, no. 2, pp. 170–184, May 2025, doi: 10.30598/populis.19.2.170-184.

[5] R. Pangesti, S. Z. Aliah, N. Nazela, V. V. Sununianti, I. Istiqomah, and D. A. Kurniawan, "Budaya Konsumtif Mahasiswa dalam Mengikuti Tren: Analisis Perspektif Karl Marx," *RISOMA: Jurnal Riset Sosial Humaniora dan Pendidikan*, vol. 4, no. 3, pp. 275–266, May 2026, doi: 10.62383/risoma.v4i3.1690.

[6] S. Hidajat and W. T. Wardhana, "Pengaruh Literasi Keuangan dan Sikap Keuangan Terhadap Pengelolaan Keuangan Mahasiswa," *Journal of Economics and Business UBS*, vol. 12, no. 2, pp. 1036–1048, 2023, doi: doi.org/10.52644/joeb.v12i2.200.

[7] H. Holil, N. Susanti, and R. T. Yanti, "The Influence of Price, Location and Service on Purchase Decisions at the Air Sebakul NRL Minimarket, Bengkulu City," *Jurnal Fokus Manajemen*, vol. 1, no. 2, pp. 48–54, 2021, doi: https://doi.org/10.37676/jfm.v1i2.1880.

[8] D. D. Eisenring, "PERTUMBUHAN AREA PERKOTAAN 2 DI SEKITAR KAMPUS PERGURUAN TINGGI UNIVERSITAS TADULAKO PALU," *RUANG : JURNAL ARSITEKTUR*, vol. 16, no. 1, pp. 39–50, 2022, doi: https://doi.org/10.22487/ruang.v16i1%20Maret.42.

[9] C. C. Ishano, N. Adiarni, and M. Najamuddin, "SEGMENTASI PASAR KONSUMEN MAKANAN DI JAKARTA, INDONESIA DENGAN PENDEKATAN FOOD-RELATED LIFESTYLE," *AGRIBUSINESS JOURNAL*, vol. 10, no. 2, pp. 149–166, 2016, doi: doi.org/10.15408/aj.v10i1.9228.

[10] A. W. A. Wibowo, "Analisis Klaster Perilaku Konsumen Produk Mocaf Menggunakan K-Means Clustering," *JABis: Jurnal Administrasi Bisnis*, vol. 23, no. 1, pp. 1–15, 2025.

[11] K. M. Chelviani, M. A. Meitriana, and I. A. Haris, "ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI PEMILIHAN LOKASI TOKO MODERN DI KECAMATAN BULELENG," *Jurnal Pendidikan Ekonomi Undiksha*, vol. 9, no. 2, pp. 257–266, 2019, doi: https://doi.org/10.23887/jjpe.v9i2.20051.

[12] M. Norshahlan, H. Jaya, and R. Kustini, "Penerapan Metode Clustering Dengan Algoritma K-means Pada Pengelompokan Data Calon Siswa Baru," *JURSI TGD : Jurnal Sistem Informasi Triguna Dharma*, vol. 2, no. 6, pp. 1042–1053, 2023, doi: https://doi.org/10.53513/jursi.v2i6.9148.

[13] S. Sindrawati, D. Syaripudin, and A. R. Raharja, "PENERAPAN ALGORITMA K-MEANS CLUSTERING PADA DATA NILAI SISWA UNTUK MENENTUKAN KELOMPOK PENERIMA BEASISWA," *SISINFO : Jurnal Sistem Informasi dan Informatika*, vol. 6, no. 2, pp. 47–52, 2024, doi: https://doi.org/10.37278/sisinfo.v6i2.900.