Application of the K-Means Algorithm for Sales Data Clustering in the KartsPOS Application at PT Dizytech Bangun Asia

Andika Dwi Aryo H, Muhammad Syahputra Novelan

Abstract

The increasing availability of transaction data in point-of-sale applications creates opportunities to identify business patterns and support data-driven decision-making. This study aims to apply the K-Means clustering algorithm to group sales-related transaction data in the KartsPOS application at PT Dizytech Bangun Asia. The dataset consists of 925 transaction records collected over the observation period from October 2025 to October 2026. The research methodology includes data collection, data preprocessing, feature selection, determination of the number of clusters, K-Means implementation, and evaluation and interpretation of the clustering results. Two numerical attributes, namely POS/Webstore prices and invoice values, were selected and standardized before clustering. A total of 907 records with complete numerical attributes were processed using K-Means with three clusters. The monthly analysis revealed that September 2026 recorded the highest transaction volume, with 104 transactions, while August 2026 generated the highest invoice value of IDR 85,640,000. The clustering results produced three groups consisting of 807 low-value transactions, 88 medium-value transactions, and 12 high-value transactions. These findings demonstrate the potential of K-Means to organize transaction records according to their numerical characteristics and provide a structured overview of invoice-related activity. The resulting clusters can support transaction evaluation and business analysis at PT Dizytech Bangun Asia. However, the findings represent transaction segmentation rather than direct sales forecasting or profitability measurement.

***Keywords**: K-Means, data clustering, transaction data, sales analysis, KartsPOS*

[1]Andika Dwi Aryo H
Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: andhika.basisgb@gmail.com

[2]Muhammad Syahputra Novelan
e-mail: putranovelan@dosen.pancabudi.ac.id
2nd International Cofference on the Epicentrum of Economic Global Framework (ICEEGLOF)
Theme: Navigating The Future: Business and Social Paradigms in a Transformative Era.
https://proceeding.pancabudi.ac.id/index.php/ICEEGLOF

Introduction

The development of information technology has encouraged businesses to utilize digital systems to improve operational efficiency and support data-driven decision-making. One important aspect of business operations is sales management, which generates transaction data that can be analyzed to identify sales patterns, evaluate business performance, and determine appropriate marketing strategies. The increasing volume of transaction data provides opportunities for organizations to extract meaningful information through data mining and machine learning techniques. However, transaction data that is only stored as operational records may have limited strategic value if it is not processed and analyzed systematically. Therefore, an analytical approach is needed to transform sales transaction data into useful information for business development and decision-making [1] [2].

PT Dizytech Bangun Asia utilizes the KartsPOS application to support sales transaction recording and management. As a point-of-sale application, KartsPOS generates transaction data that can potentially provide valuable insights into sales performance. These data may include transaction frequency, product quantities sold, total sales value, and other relevant sales attributes. Analyzing these attributes can help identify differences in sales activity and reveal groups of transactions or sales periods with similar characteristics. Nevertheless, the availability of transaction data does not automatically provide a comprehensive understanding of sales performance. Without an appropriate analytical method, identifying sales patterns and distinguishing between high-, medium-, and low-performing sales categories can become difficult, particularly when transaction records continue to increase over time [3] [4] .

One method that can be applied to address this problem is K-Means clustering, a data mining algorithm that groups data objects according to similarities in their characteristics. The algorithm works by determining cluster centers, assigning data points to the nearest centroid, and iteratively updating the centroids until the clustering process converges. In sales analysis, K-Means can be used to group transaction records based on selected numerical attributes, such as transaction frequency, total sales value, and product quantities sold. The resulting clusters can provide a structured overview of sales characteristics and help distinguish groups with different performance levels. Compared with approaches that require labeled training data, K-Means offers a relatively straightforward implementation because it does not require predefined class labels. Its effectiveness, however, depends on the relevance of the selected attributes, appropriate data preprocessing, feature scaling when necessary, and the determination of a suitable number of clusters [5] [6] .

Previous research has demonstrated the usefulness of clustering techniques in extracting patterns from business and sales data. K-Means has been widely applied in customer segmentation, product grouping, sales performance analysis, and the identification of business activity patterns. These applications indicate that clustering can support the interpretation of large datasets by organizing records into groups with similar characteristics. However, the usefulness of a clustering model depends on the context in which it is implemented, the quality of the available data, and the business objectives of the analysis. In the context of the KartsPOS application at PT Dizytech Bangun Asia, a focused analysis is required to determine how transaction data can be grouped into meaningful sales categories. The resulting information is expected to support a more systematic understanding of sales characteristics rather than relying solely on aggregate transaction reports[7] [8] .

Based on these considerations, this study aims to apply the K-Means algorithm to cluster sales data recorded in the KartsPOS application at PT Dizytech Bangun Asia. The research stages include collecting transaction data, selecting relevant attributes, preprocessing the dataset, determining the number of clusters, implementing the K-Means algorithm, and interpreting the characteristics of the resulting clusters. The quality of the clustering results can be evaluated using an appropriate internal validation measure, such as the Silhouette Coefficient, when the dataset and cluster structure support its application. The findings are

expected to provide an organized representation of sales transaction characteristics and practical information that can assist the company in evaluating sales performance and developing more targeted business strategies. This study also seeks to demonstrate the applicability of a relatively simple unsupervised machine learning method for extracting useful information from transaction data in a point-of-sale environment [9] .

Literature Review

2.1 Sales Transaction Data Analysis

Sales transaction data analysis is an important process for identifying sales patterns, evaluating business performance, and supporting data-driven decision-making. Transaction data generally include information such as transaction dates, product quantities, sales values, and transaction frequencies. When processed systematically, these data can provide insights into sales characteristics and variations in business activity. In point-of-sale (POS) systems, transaction records serve as a valuable source of business intelligence. However, the increasing volume of transaction data requires appropriate analytical methods to transform operational records into meaningful information. Therefore, data analysis techniques are needed to identify sales patterns and support the development of more effective business strategies [10] [11].

2.2 K-Means Clustering Algorithm

K-Means is an unsupervised machine learning algorithm used to group data into a predefined number of clusters based on similarities among observations. The algorithm assigns each observation to the nearest cluster center, known as a centroid, and repeatedly updates the centroids until the clustering process converges. In sales analysis, K-Means can be applied to numerical attributes such as total sales value, transaction frequency, and product quantities sold. The resulting clusters help identify groups with similar sales characteristics. Before implementation, data cleaning, relevant attribute selection, and feature scaling should be considered to improve the reliability of the clustering process. The number of clusters can be determined using approaches such as the Elbow Method, while appropriate validation measures can help assess the quality of the resulting clusters [12] [13] .

2.3 Application of K-Means for Sales Data Clustering

The application of K-Means in sales data analysis provides a systematic approach to identifying groups of transactions or sales periods with similar characteristics. The resulting clusters can help organizations distinguish sales activity levels and evaluate differences in transaction performance. In this study, K-Means is applied to sales transaction data recorded in the KartsPOS application at PT Dizytech Bangun Asia. The research process includes collecting transaction data, preprocessing the dataset, selecting relevant attributes, determining the number of clusters, implementing the algorithm, and interpreting the resulting cluster characteristics. Depending on the actual data, the clusters may represent different levels of sales activity, such as high, medium, and low sales. These categories are determined by examining the characteristics of each cluster rather than assuming predefined results. The findings are expected to support sales performance evaluation and provide useful information for business decision-making. Unlike predictive models, K-Means focuses on grouping existing data and identifying patterns rather than directly forecasting future sales values [14].

Research Methodology

The research methodology describes the systematic procedures used to apply the K-Means clustering algorithm for sales data clustering in the KartsPOS application at PT Dizytech Bangun Asia. The research process consists of six main stages: data collection, data

preprocessing, feature selection, determination of the optimal number of clusters, implementation of the K-Means algorithm, and evaluation and interpretation of clustering results. These stages are conducted sequentially to ensure that the sales transaction data are processed systematically and are suitable for clustering analysis. The research uses relevant numerical attributes, such as total sales value, transaction frequency, and product quantities sold, depending on the availability and characteristics of the dataset.

The implementation of these stages aims to identify groups of sales transactions with similar characteristics and reveal differences in sales activity. Data preprocessing and feature selection are conducted to improve data quality and ensure that the selected attributes align with the research objectives. Subsequently, the K-Means algorithm is applied to form clusters, while the resulting groups are evaluated and interpreted based on their characteristics. The findings are expected to provide a structured overview of sales performance and support data-driven decision-making at PT Dizytech Bangun Asia, particularly in evaluating sales patterns and developing appropriate business strategies.

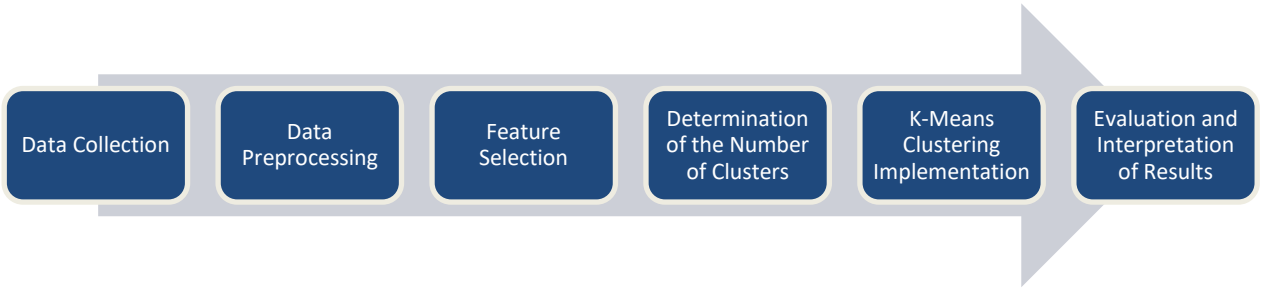


Figure 1. Research Methodology

1. Data Collection
   The data collection stage involves obtaining sales transaction records from the KartsPOS application at PT Dizytech Bangun Asia. The collected data may include transaction dates, product names, quantities sold, transaction frequencies, and total sales values. The data serve as the primary input for the clustering process. The completeness and relevance of the transaction records are examined to ensure that the dataset is suitable for further analysis.
2. Data Preprocessing
   Data preprocessing is performed to prepare the transaction dataset before applying the K-Means algorithm. This stage includes checking missing values, identifying duplicate records, correcting inconsistent data formats, and handling invalid values. Numerical attributes are also examined for consistency, and feature scaling is applied when necessary because K-Means relies on distance calculations. This process helps improve data quality and reduce the influence of inconsistencies on the clustering results.
3. Feature Selection
   Feature selection involves identifying the numerical attributes most relevant to the sales clustering objective. The selected attributes may include total sales value, product quantities sold, and transaction frequency. These attributes are chosen according to their availability, data quality, and relevance to the company's sales evaluation needs. Selecting appropriate features is important because the characteristics of the input variables directly influence the formation and interpretation of the resulting clusters.
4. Determination of the Number of Clusters
   This stage determines the number of clusters to be generated by the K-Means algorithm. The Elbow Method can be used to examine changes in the within-cluster sum of squared distances at different values of \(K\). The appropriate number of clusters is selected by considering the resulting pattern, cluster interpretability, and business requirements. The objective is to obtain meaningful groups that represent differences in sales characteristics without creating unnecessarily complex clusters.

5. K-Means Clustering Implementation
   The K-Means algorithm is applied to group sales records according to similarities in their selected attributes. The process begins with the initialization of cluster centroids, followed by calculating distances between observations and centroids. Each observation is assigned to the nearest centroid, and the centroids are recalculated based on the assigned observations. These steps are repeated until the centroids stabilize or the stopping criterion is met. The output consists of cluster assignments that group sales records with similar characteristics.
6. Evaluation and Interpretation of Results
   The final stage evaluates and interprets the resulting clusters to understand the characteristics of each sales group. Cluster centroids and the distribution of observations are examined to identify differences in sales value, quantities sold, and transaction frequency. When appropriate, the Silhouette Coefficient can be used to assess cluster cohesion and separation. Based on the observed characteristics, the clusters may be interpreted as high-, medium-, or low-sales groups if the data support these classifications. The findings provide information that can assist PT Dizytech Bangun Asia in evaluating sales performance and developing more targeted business strategies.

Results

The analysis was conducted using sales transaction records from the KartsPOS application at PT Dizytech Bangun Asia. The dataset contains 925 transaction records covering the period from 12 October 2025 to 12 October 2026 and includes merchant information, subscription details, POS/Webstore prices, and invoice values. Across the dataset, 759 unique merchants were identified and the recorded invoice value totaled IDR 636,436,000. These records provide a basis for examining transaction activity and differences in invoice values over the observed period. Because the available fields primarily describe subscription and renewal transactions, the analysis focuses on invoice-based transaction characteristics rather than product-level sales.

Monthly analysis was used to identify changes in transaction volume and invoice value throughout the observation period. The highest transaction volume occurred in September 2026, with 104 records, whereas August 2026 recorded the highest monthly invoice value at IDR 85,640,000. This difference indicates that the month with the largest number of transactions did not necessarily generate the highest total invoice value. Examining both measures therefore provides a more complete view of the activity represented in the KartsPOS dataset.

K-Means clustering was applied to records with complete values for POS/Webstore price and invoice value. After removing records with missing values in these two features, 907 records remained for analysis. The two numerical features were standardized before clustering, and K-Means was configured to generate three clusters. The resulting groups were ordered by their mean invoice value and interpreted as low-value, medium-value, and high-value clusters. These labels describe the observed invoice-value patterns and should not be interpreted as direct measures of merchant profitability or future sales.

1. Dataset Overview

Table 1 summarizes the main characteristics of the source dataset and identifies the number of records available for the clustering stage. These indicators establish the scope of the analysis and make clear the difference between the full transaction dataset and the subset used by the algorithm.

Table 1. Overview of the KartsPOS transaction dataset.

Indicator	Result
Observation period	12 Oct 2025–12 Oct 2026
Total transaction records	925
Unique merchants	759
Total recorded invoice value	IDR 636,436,000
Records retained for clustering	907
Missing invoice values	17
Missing POS/Webstore prices	3
Number of clusters	3

The dataset contains 925 transaction records associated with 759 unique merchants over the stated observation period. The total invoice value of IDR 636,436,000 represents the sum recorded in the source dataset, not necessarily recognized revenue or realized cash receipts. Some records lack one or both numerical features required for clustering: 17 records have missing invoice values and three have missing POS/Webstore prices. After excluding records with missing values in either selected feature, 907 observations remained. This represents approximately 98.1% of the original records, allowing the analysis to retain most of the available data while avoiding imputation assumptions that could alter the distance-based clustering results. The table also records the decision to form three clusters; this configuration is the model setting used for the reported grouping and should be considered alongside validation or business interpretability when evaluating whether it is the most suitable choice.

2. Monthly Transaction Activity

Table 2 reports monthly transaction counts and the corresponding total invoice value. The monthly view makes it possible to compare the frequency of transactions with their monetary contribution and to identify periods with comparatively high or low activity.

Table 2. Monthly transaction volume and invoice value.

Month	Transactions	Invoice value
2025-10	35	IDR 21,800,000
2025-11	58	IDR 42,946,000
2025-12	61	IDR 46,680,000
2026-01	73	IDR 54,620,000
2026-02	62	IDR 44,320,000
2026-03	50	IDR 32,460,000
2026-04	77	IDR 48,847,000
2026-05	97	IDR 64,163,000
2026-06	87	IDR 53,960,000
2026-07	85	IDR 44,300,000
2026-08	99	IDR 85,640,000
2026-09	104	IDR 70,760,000
2026-10	37	IDR 25,940,000
Total	925	IDR 636,436,000

The monthly totals show variation in both the number of transactions and their invoice value. Transaction volume reached its highest point in September 2026, with 104 records, followed by August 2026 with 99 and May 2026 with 97. March 2026 recorded 50 transactions, while October 2025 had 35. In financial terms, August 2026 recorded the highest monthly

invoice value at IDR 85,640,000, followed by September 2026 at IDR 70,760,000 and May 2026 at IDR 64,163,000. This difference between the peaks is important: the number of transactions does not alone determine the invoice total, because the amount associated with each transaction can vary. October 2025 and October 2026 are partial months under the specified observation window, so their totals should not be compared directly with full months as if they represented equivalent exposure periods. The monthly results describe recorded invoice activity and should not be interpreted as a forecast or as evidence of a causal business effect.

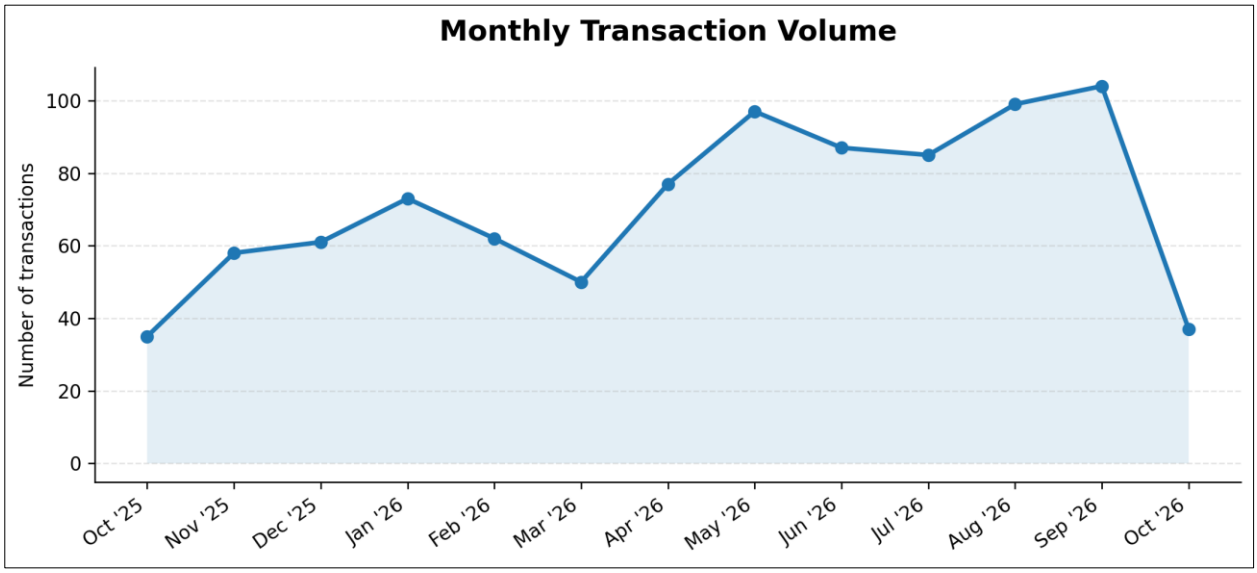


Figure 1. Monthly transaction volume.

Figure 1 visualizes the monthly transaction counts as a time series, making the rises and declines easier to compare than in a table alone. The curve increases from the initial partial month of October 2025 toward the end of 2025, then fluctuates during the first quarter of 2026. Activity rises again in April and May, eases slightly in June and July, and reaches its highest point in September 2026. The final October 2026 value is lower because the dataset ends on 12 October, rather than at the end of the month. The figure therefore indicates that transaction activity was not uniform across the year. For operational planning, these variations may justify reviewing the timing of subscription renewals, customer onboarding, or sales campaigns, but the graph alone cannot establish why the fluctuations occurred. A separate analysis of customer cohorts, package types, and renewal dates would be needed to investigate possible explanations.

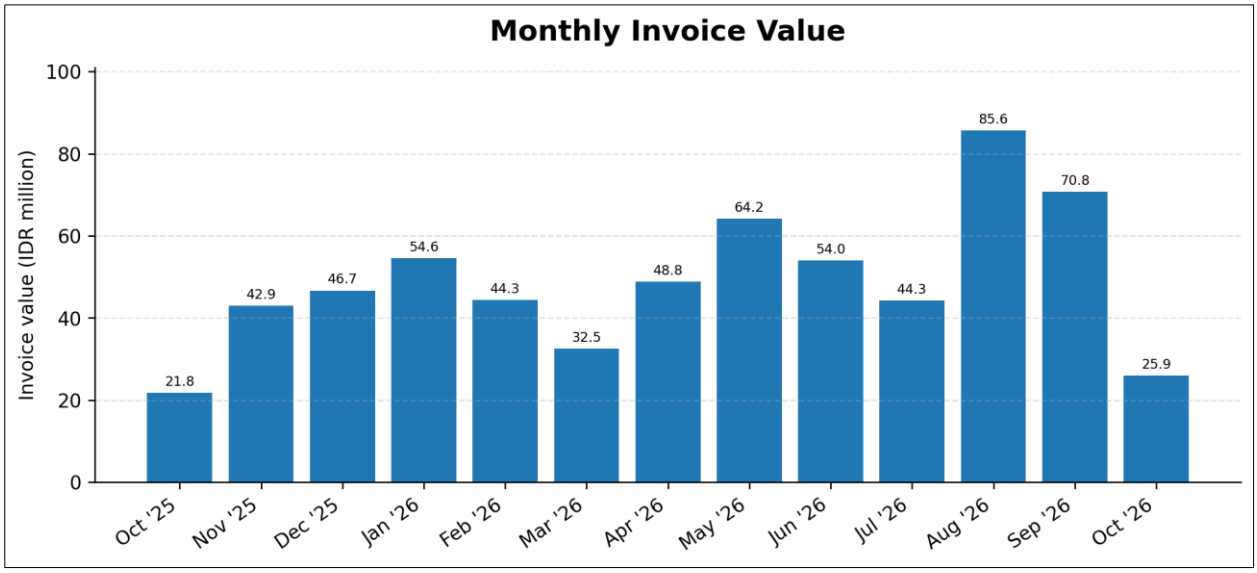


Figure 2. Monthly invoice value (IDR million).

Figure 2 compares the total invoice value recorded each month, expressed in millions of Indonesian rupiah. The chart highlights a pronounced peak in August 2026, at IDR 85.64 million, followed by September 2026 at IDR 70.76 million. May 2026 also shows a comparatively high total of IDR 64.16 million. Lower totals occur in October 2025 and March 2026, although the partial coverage of October affects the comparability of those months. The invoice-value trend does not perfectly follow the transaction-volume trend: September has the highest count, but August has the highest invoice total. This suggests that the monetary value per recorded transaction differs across months. The figure can support monitoring of invoice totals, but it should not be treated as a direct measure of profitability, collected cash, or recurring revenue without confirmation of payment status, discounts, taxes, refunds, and accounting treatment.

3. K-Means Clustering Results

Table 3 summarizes the three groups generated from the 907 records with complete POS/Webstore price and invoice values. The cluster labels are assigned after ordering the clusters by their mean invoice value; they are descriptive labels rather than predefined business categories.

Table 3. K-Means clustering summary.

Cluster	Records	Mean invoice	Invoice range	Invoice contribution
Low-value	807	IDR 513,351	IDR 150,000–1,100,000	65.15%
Medium-value	88	IDR 1,751,727	IDR 1,150,000–3,000,000	24.24%
High-value	12	IDR 5,617,500	IDR 3,500,000–16,800,000	10.60%
Total	907	—	—	100.00%

The low-value cluster contains the largest share of observations, with 807 records, while the medium-value cluster contains 88 and the high-value cluster contains 12. The mean invoice value increases from IDR 513,351 in the low-value group to IDR 1,751,727 in the medium-

value group and IDR 5,617,500 in the high-value group. The high-value group therefore has the largest average invoice, although it represents only a small proportion of the records. Its invoice contribution is approximately 10.60% of the total invoice value among clustered observations, while the low-value group contributes approximately 65.15% and the medium-value group contributes approximately 24.24%. These proportions indicate that the low-value group dominates by count and cumulative invoice value, while the smaller high-value group is characterized by larger invoice amounts. The ranges in the table are descriptive and may overlap in the joint feature space because the algorithm uses both POS/Webstore price and invoice value, rather than invoice value alone, to determine cluster membership. Accordingly, the groups should be interpreted as multivariate transaction segments, not merely as fixed invoice bands.

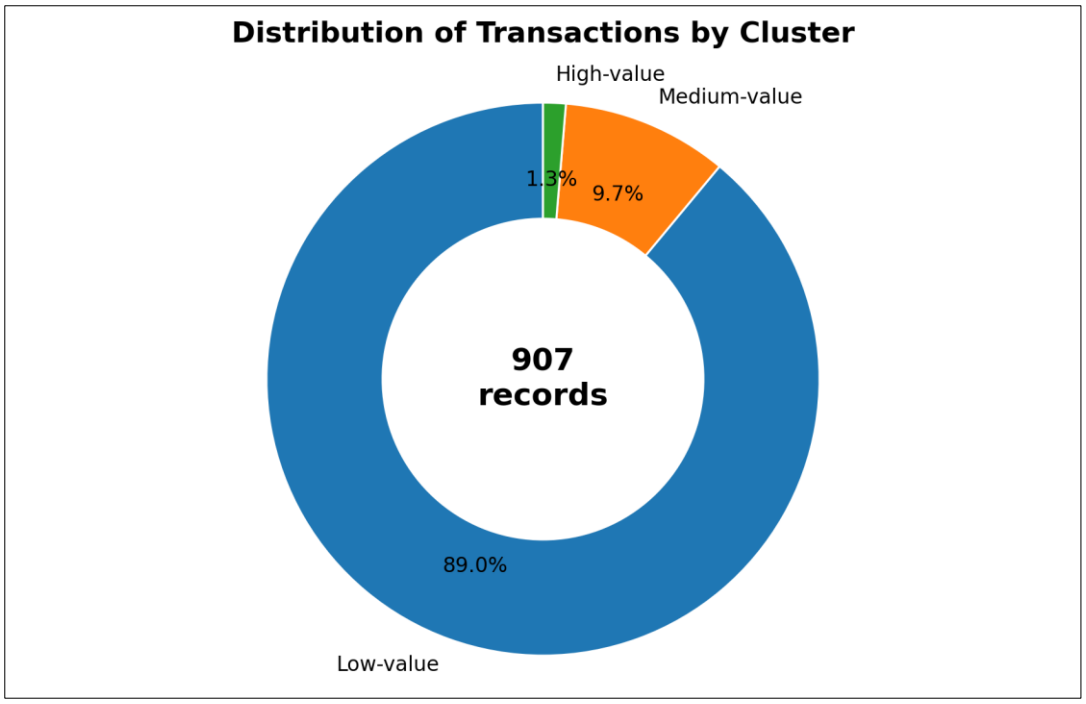


Figure 3. Distribution of transactions by cluster.

Figure 3 presents the proportion of records assigned to each cluster. The low-value group occupies the largest segment, representing approximately 89.0% of the 907 clustered observations. The medium-value group accounts for around 9.7%, while the high-value group represents approximately 1.3%. The uneven distribution shows that most records share relatively low invoice-value characteristics, whereas only a small number are associated with higher invoice values. This pattern can help the company prioritize routine monitoring of the dominant low-value group while separately reviewing high-value transactions, where individual records may have a larger monetary impact. Nevertheless, a small cluster is not automatically an anomaly or a more valuable customer segment. Further checks against subscription packages, pricing rules, discounts, and transaction validity are needed before making operational decisions based on these groupings.

Discussion

The results demonstrate that the application of the K-Means clustering algorithm provides a systematic approach to identifying differences in transaction characteristics within the KartsPOS application at PT Dizytech Bangun Asia. The analysis of 925 transaction records revealed variations in monthly transaction volume and invoice value, indicating that transaction frequency does not necessarily correspond to the highest monetary contribution. September 2026 recorded the highest transaction volume, with 104 transactions, whereas August 2026

generated the highest invoice value of IDR 85,640,000. This finding highlights the importance of evaluating transaction frequency and invoice value as complementary indicators when examining business activity. By considering these indicators together, the company can obtain a more comprehensive understanding of transaction patterns and identify periods that may require further operational evaluation.

The K-Means clustering results further revealed three distinct groups based on POS/Webstore prices and invoice values among the 907 records with complete numerical attributes. The low-value cluster contained 807 transactions, the medium-value cluster contained 88 transactions, and the high-value cluster contained 12 transactions. This distribution indicates that most recorded transactions were concentrated in the low-value group, while a relatively small number of transactions exhibited higher average invoice values. These findings demonstrate the potential of clustering to organize transaction records into more interpretable categories and support the evaluation of sales-related activity. However, the resulting groups should be interpreted according to their selected numerical features and should not be considered direct indicators of profitability, customer loyalty, or future sales performance. For further development, the company could evaluate alternative cluster configurations using appropriate validation measures and incorporate additional transaction attributes where available to improve the analytical value of the segmentation results.

Conclusion

This study demonstrates the application of the K-Means clustering algorithm for grouping sales-related transaction data in the KartsPOS application at PT Dizytech Bangun Asia. Based on 925 transaction records, the monthly analysis identified variations in transaction volume and invoice value, with September 2026 recording the highest transaction volume of 104 transactions and August 2026 generating the highest invoice value of IDR 85,640,000. These findings indicate that transaction frequency and invoice value provide complementary perspectives for evaluating business activity.

The K-Means algorithm grouped 907 records with complete numerical attributes into three clusters: low-value, medium-value, and high-value transactions, comprising 807, 88, and 12 records, respectively. The resulting clusters provide a structured overview of transaction characteristics and demonstrate the potential of unsupervised machine learning to support transaction analysis and business evaluation. Nevertheless, the clustering results depend on the selected attributes and the specified number of clusters. Future research should evaluate cluster quality using appropriate validation measures and incorporate additional relevant attributes, such as subscription packages and transaction status, where available. The findings can serve as a basis for improving data-driven evaluation and supporting more informed business decisions at PT Dizytech Bangun Asia.

References

[1] H. W. Dhany, F. Izhari, and S. Komputer, “Jurnal ICT : Information and Communication Technologies, 15 (2) (2024) 48-54 Prediction analysis condition animal use algorithm (SVM+KNN),” 2024. [Online]. Available: www.ejournal.marqchainstitute.or.id/index.php/JICT

[2] Z. Sitorus, D. Prayogi, M. A. Rizko, A. G. Suteja, and M. R. Harahap, “Implementation of the Insertion Sort Algorithm to Sort Positive Integers in Ascending Order Using Flowgorithm,” *Journal of Information Technology, computer science and Electrical Engineering (JITCSE)*, vol. 1, no. 3, pp. 323–328, 2024, doi: 10.30596/jitcse.

[3] K. Khairul, A. H. Nasyuha, A. Ikhwan, M. H. Aly, and A. Ahyanuardi, "Implementation of Multiple Linear Regression to Estimate Profit on Sales of Screen Printing Equipment," *JURNAL INFOTEL*, vol. 15, no. 2, pp. 55–61, Jun. 2023, doi: 10.20895/infotel.v15i2.934.

[4] M. S. Novelan and S. Aryza, "OPTIMIZATION CVRP WITH MACHINE LEARNING FOR IMPROVED CLASSIFICATION OF IMBALANCED DATA FOOD DISTRIBUTION," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 4, pp. 917–925, Jun. 2025, doi: 10.33480/jitk.v10i4.6467.

[5] R. Farta Wijaya, F. Kurniawan, R. R. Putrai, A. Alvin, F. Sains, and D. Teknologi, "OPTIMIZING THE MANAGEMENT OF VILLAGE ACTIVITIES THROUGH INFORMATION SYSTEMS FOR TRANSPARENCY AND ACCOUNTABILITY IN PERTUMBUKAN VILLAGE, WAMPU DISTRICT," 2024.

[6] M. M. Maulana, A. I. Suroso, Y. Nurhadryani, and K. B. Seminar, "The Smart Governance Framework and Enterprise System's Capability for Improving Bio-Business Licensing Services," *Informatics*, vol. 10, no. 2, Jun. 2023, doi: 10.3390/informatics10020053.

[7] A. Putera Utama Siahaan, N. Azizah Harahap, R. Yuni Simanullang, and P. Wanny, "Analysis of Inpatient Data Using Cluster Analysis on Simulation Dataset," *Bulletin of Information Technology (BIT)*, vol. 6, no. 1, pp. 33–39, 2025, doi: 10.47065/bit.v5i2.1830.

[8] M. S. Novelan, S. Efendi, P. Sihombing, and H. Mawengkang, "VEHICLE ROUTING PROBLEM OPTIMIZATION WITH MACHINE LEARNING IN IMBALANCED CLASSIFICATION VEHICLE ROUTE DATA," *Eastern-European Journal of Enterprise Technologies*, vol. 5, no. 3(125), pp. 49–56, 2023, doi: 10.15587/1729-4061.2023.288280.

[9] M. Iqbal and S. Efendi, "Data-Driven Approach for Credit Risk Analysis Using C4.5 Algorithm," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 14, no. 1, pp. 11–20, May 2023, doi: 10.21512/comtech.v14i1.8243.

[10] J. Asharf, N. Moustafa, H. Khurshid, E. Debie, W. Haider, and A. Wahab, "A review of intrusion detection systems using machine and deep learning in internet of things: Challenges, solutions and future directions," Jul. 01, 2020, *MDPI AG*. doi: 10.3390/electronics9071177.

[11] W. Cetera, W. Gogołek, A. Żołnierski, and D. Jaruga, "Potential for the use of large unstructured data resources by public innovation support institutions," *J. Big Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00610-6.

[12] G. Rajeshkumar *et al.*, "Smart office automation via faster R-CNN based face recognition and internet of things," *Measurement: Sensors*, vol. 27, Jun. 2023, doi: 10.1016/j.measen.2023.100719.

[13] A. Rejeb *et al.*, "The Internet of Things (IoT) in healthcare: Taking stock and moving forward," Jul. 01, 2023, *Elsevier B.V.* doi: 10.1016/j.iot.2023.100721.

[14] M. Overton, S. Larson, L. J. Carlson, and S. Kleinschmit, "Public data primacy: the changing landscape of public service delivery as big data gets bigger," *Global Public Policy and Governance*, vol. 2, no. 4, pp. 381–399, Dec. 2022, doi: 10.1007/s43508-022-00052-z.

[15] J. Ilmiah, E. & Erudisi, A. Zahara, M. I. Kabullah, and R. E. Putera, "JIEE: Effectiveness of the OSSRBA (Online Single Submission Risk Based Approach) System in Business Licensing Services in Payakumbuh City DPMPTSP," vol. 3, pp. 22–40, 2023.

[16] W. Jin, "Challenges and Innovative Countermeasures Faced by Public Administration in the Context of Big Data and Internet of Things," *Math. Probl. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/8949365.