IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 4, August 2026, pp. 3402~3410
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i4.pp3402-3410

An empirical comparison of clustering approaches for recency, frequency, and monetary customer segmentation

Upekkha Lau, Meditya Wasesa
School of Business and Management, Institut Teknologi Bandung, Bandung, Indonesia

Article Info

Article history:

Received May 15, 2025
Revised May 25, 2026
Accepted Jun 19, 2026

Keywords:

Business analytics
Customer segmentation
DBSCAN clustering
Hierarchical clustering
K-means clustering

ABSTRACT

This study evaluates effectiveness of three clustering techniques—k-means, hierarchical clustering, and density-based spatial clustering of applications with noise (DBSCAN)—applied to the recency-frequency-monetary (RFM) model for customer segmentation in the retail sector. Using sales transaction data from a distributor of computer accessories and printing products. The results show that k-means achieved the best clustering validation scores and effectively identified high-value customers, hierarchical clustering generated less meaningful groupings than k-means, and DBSCAN misclassified key customers as noise. These findings highlight k-means as the most suitable technique for RFM-based segmentation in this retail business context. The study offers practical insights for retail and distribution businesses aiming to adopt data-driven customer strategies and suggests future research to enhance segmentation robustness and refine the RFM framework.

Corresponding Author:

Upekkha Lau
School of Business and Management, Institut Teknologi Bandung
Bandung, Indonesia
Email: upklau2121@gmail.com

1. INTRODUCTION

Competing on analytics has become an increasingly popular strategy for businesses that generate large volumes of transactional data [1]. This study explores customer segmentation in the distribution business using the recency-frequency-monetary (RFM) model combined with clustering analysis. Customer segmentation is a marketing strategy that groups customer with similar characteristics into the same category, which they respond to marketing efforts in a similar way [2]. Clustering is a machine learning technique that groups a set of physical or abstract objects into clusters based on similarity [3]. It is particularly useful for identifying patterns in unlabeled data. In the context of customer segmentation, the "objects" represent customers, and similarity is typically measured based on purchasing behavior, demographics, or other relevant attributes. By leveraging clustering techniques, businesses can design targeted marketing campaigns, provide personalized recommendations, and implement tailored customer service strategies for each identified customer segment.

The RFM model is a popular approach for customer segmentation. It evaluates customer behavior based on three key dimensions: recency, frequency, and monetary value derived from sales data [4], [5]. By clustering these RFM variables, businesses can identify distinct customer groups based on their engagement patterns. This study applies three clustering techniques—k-means, hierarchical clustering, and density-based spatial clustering of applications with noise (DBSCAN)—to determine the most suitable method for analyzing master dealer transactional data. The k-means algorithm partitions data into n clusters by minimizing intra-cluster variance, also known as inertia or within-cluster sum of squares (WCSS) [6]. Hierarchical clustering, on the other hand, creates a hierarchy of clusters without requiring a predefined number of

clusters [4]. DBSCAN groups data points based on density, allowing it to detect clusters of arbitrary shape and identify noise or outliers, without needing to specify the number of clusters in advance [7].

This research addresses the problem of determining the most suitable clustering method for customer segmentation in a retail distribution business, where understanding customer patterns is crucial for strategic decision-making. The study compares three widely used clustering techniques—k-means, hierarchical, and DBSCAN—to evaluate their effectiveness in segmenting customers based on RFM analysis. To ensure the reliability and validity of the clustering results, the study employs three internal validation metrics: silhouette score, Davies-Bouldin index (DBI), and Calinski-Harabasz index (CHI). Based on the evaluation outcomes, the research not only identifies the best-performing clustering method but also provides an interpretation of the resulting customer segments to offer practical insights for the business.

Table 1 summarizes prior studies on customer segmentation employing clustering techniques. Among the ten reviewed studies, seven applied k-means clustering, four used hierarchical clustering, and two employed DBSCAN. Although k-means appears to be the most frequently used method, its prevalence does not imply methodological superiority. Rather, k-means is often adopted due to its simplicity, scalability, and general applicability across diverse segmentation contexts.

The authors [8], [9] both applied the RFM model with k-means clustering to segment customers based on purchasing behavior, aiming to enhance data-driven marketing and identify high-value customers. Akande *et al*. [9] used online retail data from the UCI Machine Learning Repository and identified five customer categories—platinum, gold, silver, bronze, and bad—using the elbow method to determine the optimal number of clusters. Similarly, Christy *et al.* [10] utilized RFM with k-means, fuzzy c-means, and repetitive median based k-means (RM k-means) clustering, finding RM k-means to be the most effective in terms of convergence and cluster compactness, reinforcing the value of RFM-based segmentation in customer relationship management. Meanwhile, Sutramiani *et al.* [11] compared k-means and DBSCAN for clustering micro, small, and medium enterprises (MSMEs) based on asset value and turnover. K-means outperformed DBSCAN in cluster compactness (DBI), though DBSCAN was superior in identifying outliers and did not require a predefined number of clusters. This study contributes novelty by systematically comparing k-means, hierarchical clustering, and DBSCAN using multiple internal validation metrics (silhouette, DBI, and CHI) within a real-world distribution business context to identify the most suitable method for actionable customer segmentation.

Table 1. Previous studies on customer segmentation clustering analysis

Article	Business context	Features	Clustering method	Validation method
[4]	Retail	RFM	Hierarchical	DBI and silhouette
[12]	Retail	Market basket	K-means	Sum of squared error (SSE), silhouette, and gap statistic
[8]	Online retail	RFM	K-means	Silhouette
[10]	Online retail	RFM	K-means, RM k-means, and fuzzy c-means	Silhouette
[13]	Online retail	RFM interpurchase time	Hierarchical	CHI and DBI
[14]	Internet banking	RFM	K-means and DBSCAN	DBI
[15]	Retail	RFM (time)	Hierarchical and k-shape	Silhouette and CHI
[16]	Online retail	RdFdMd	K-means	DBI
[17]	Retail	RFM	Density peak (DP)-genetic algorithm (GA)-possibilistic fuzzy c-means (PFCM) and k-means	ARI, NMI, and Dunn-Bonferroni
[18]	Distribution and retail	RFM	Fuzzy c-means	Partition coefficient
This study	Distribution	RFM	K-means, hierarchical, and DBSCAN	Silhouette, DBI, and CHI

2. METHOD

Figure 1 outline the research framework of this study. It is adapted from previous research [19]. The framework consists of four stages: transactional data collection, data preprocessing and preparation, clustering and validation, and labelling and analysis. In this study, the transactional dataset was already in a clean and structured format, with no missing values or inconsistencies identified; therefore, no additional data cleaning was required. Preprocessing was conducted by transforming the raw data into recency, frequency, and monetary variables.

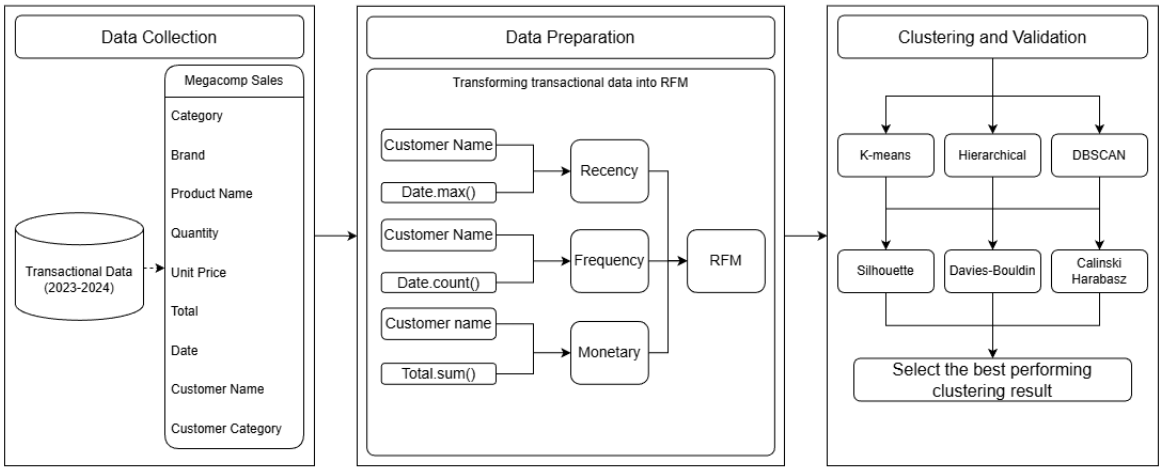


Figure 1. The research framework

2.1. Clustering and validation

2.1.1. Clustering

This study employs three widely used clustering techniques—k-means, hierarchical, and DBSCAN—selected for their distinct underlying principles: centroid-based (k-means), hierarchical (agglomerative), and density-based (DBSCAN).

i) K-means clustering: the k-means algorithm partitions data into k clusters by minimizing the spread within each cluster [20]. It starts by randomly selecting k centroids and assigning each data point to the nearest one. Then, it iteratively updates centroids by calculating the mean of assigned points and reassigns points until centroid movement falls under a threshold, indicating cluster stability.

$$J = \sum_{i=1}^{k} \sum_{x_j \in C_i} \left\| x_j - \mu_i \right\|^2 \quad (1)$$

Where k is number of clusters; C_i is set of points in cluster i; x_j is data point in cluster i; μ_i is the centroid of cluster i; and $\left\| x_j - \mu_i \right\|^2$ is squared Euclidean distance between a data point and its centroid.

ii) Hierarchical clustering: agglomerative (hierarchical) clustering uses a bottom-up approach, merging individual data points into clusters successively to form a tree-like structure called a dendrogram [20]. Unlike k-means, it doesn't require specifying the number of clusters in advance. The process starts with each point as its own cluster, merging the closest pairs until all points form a single cluster. This paper uses Ward's method, which minimizes the sum of squared differences within clusters.

$$D(A, B) = \frac{|B|}{|A|+|B|} \|\mu_A - \mu_B\|^2 \quad (2)$$

Where $D(A, B)$ is distance between cluster A and B; $|A|$ and $|B|$ is number of points in cluster A and B; μ_A and μ_B is centroid of cluster A and B; and $\|\mu_A - \mu_B\|^2$ is squared Euclidean distance between the centroids.

iii) DBSCAN: DBSCAN groups dense regions of data while treating sparse areas as noise [11]. Unlike k-means, it can detect clusters of arbitrary shapes without requiring the number of clusters in advance. It starts with an unvisited point and checks how many neighbors lie within a specified distance (eps). If this number meets or exceeds min_samples, the point becomes a core sample and initiates a cluster. The cluster expands by recursively including neighboring core and border points. This process continues until all points are classified.

$$N_\epsilon(p) = \{q \in D | d(p, q) \leq \epsilon\} \quad (3)$$

Where $N_\epsilon(p)$ is neighborhood of point p within radius ϵ; $d(p, q)$ is distance between points p and q; ϵ is neighborhood radius parameter; and A point p is core point if $|N_\epsilon(p)| \geq minPts$.

2.1.2. Validation

For validation, this study utilizes introduces: silhouette score, DBI, and CHI. As shown in Table 1, these metrics were selected for their widespread use in customer segmentation and their availability in the scikit-learn library, ensuring easy reproducibility. Furthermore, prior research identifies silhouette, DBI, and CHI as among the top-performing indices within a group of recommended clustering validation metrics [21].

i) Silhouette score: the silhouette score evaluates clustering quality by comparing how well a data point fits within its own cluster relative to the closest neighboring cluster [22]. This is done by measuring the distance between a point and all other points in its cluster, as well as its distance to the nearest cluster. Calculating silhouette scores requires two things, the clustering result from a chosen algorithm and the pairwise proximities between samples [23].

$$s(i) = \frac{b(i)-a(i)}{max\{a(i),b(i)\}} \quad (4)$$

Where $a(i)$ is mean intra-cluster distance; $b(i)$ is mean nearest-cluster distance; and $max\{a(i), b(i)\}$ is normalize the score so that score remain within -1 and 1.

ii) DBI: the DBI evaluates clustering quality based on cluster compactness and separation [24]. It measures how tightly grouped the points in each cluster are and how distinct the clusters are from one another. Lower DBI values indicate better clustering.

$$DB(k) = \frac{1}{k}\sum_{i=1}^{k} \underset{i\neq j}{max} \frac{S_i+S_j}{d(\bar{x}_i,\bar{x}_j)} \quad (5)$$

Where S_i and S_j is measure the compactness of cluster i and j; and $d(\bar{x}_i, \bar{x}_j)$ is measure how far apart cluster i and j.

iii) CHI: the CHI, or variance ratio criterion, measures clustering quality by comparing between-cluster separation to within-cluster compactness [25]. A higher CHI indicates denser, well-separated clusters.

$$CH(k) = \frac{BGSS/(k-1)}{WGSS/(n-k)} \quad (6)$$

Where between-cluster sum of squares (BCSS) is measures how far each cluster centroid is from the global centroid; and WCSS is measures how compact the clusters are by summing the squared distances of points to their cluster centroid.

3. RESULTS AND DISCUSSION

3.1. Results

The analysis begins with data visualization to understand customer behavior. Recency values were reversed for consistency with frequency and monetary metrics, where higher values indicate better customers. Figure 2 shows that over half of the customers have high recency, indicating strong loyalty in the past two years. Figure 3 reveals most customers have low transaction frequency, with few having high frequency. Similarly, Figure 4 shows most customers have low transaction values. To protect sensitive information, the customer monetary values distribution plot is normalized as requested by the data owner. Figure 5 show the correlation between recency and frequency (Figure 5(a)), recency and monetary (Figure 5(b)), and frequency and monetary (Figure 5(c)) value.

i) Recency vs. frequency: the scatter plot shows that most customers have low frequency values, regardless of their recency. However, there is a group of customers with high recency values who also exhibit slightly higher frequency, suggesting that customers who have not engaged recently tend to have lower interaction levels.

ii) Recency vs. monetary: similar to the recency-frequency relationship, the majority of customers with high recency scores have low monetary values. A small portion of customers with high recency scores also have higher monetary values, indicating that while most inactive customers have low spending, a few exceptions exist.

iii) Frequency vs. monetary: the plot reveals a positive correlation, where customers with higher purchase frequencies tend to have higher monetary values. However, a concentration of customers with low frequency and low monetary values shows a significant portion of the customer base consists of infrequent and low-spending buyers.

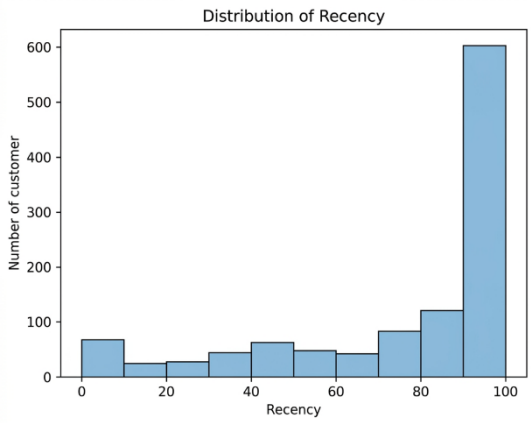


Figure 2. Distribution of customer recency values

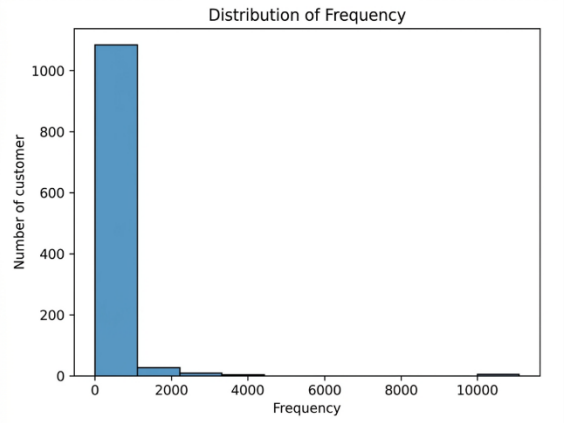


Figure 3. Distribution of customer frequency values

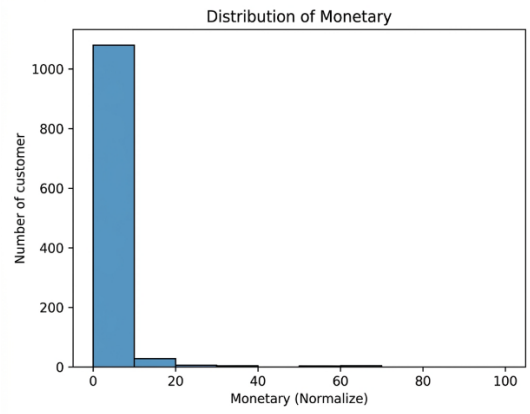
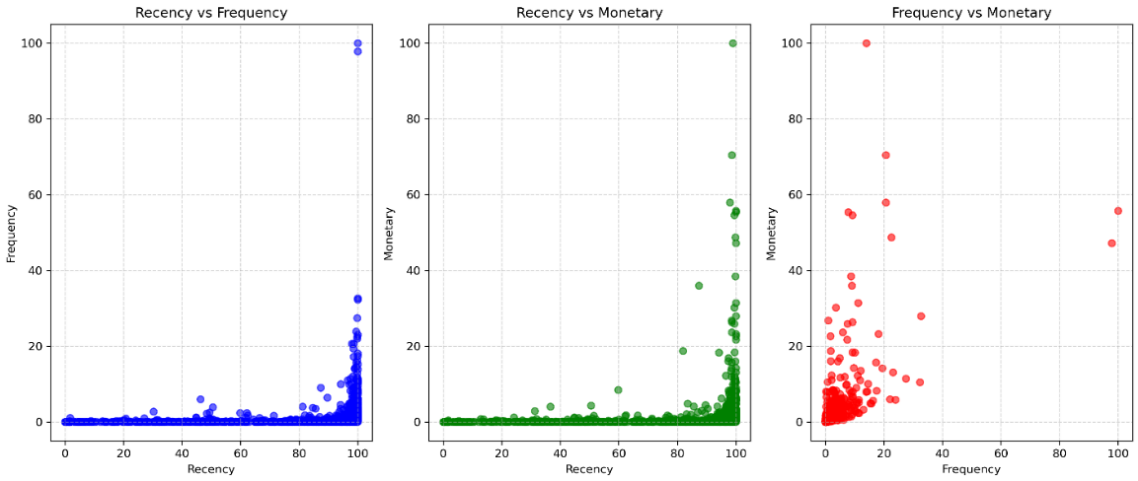


Figure 4. Distribution of customer monetary values



(a) (b) (c)

Figure 5. Pairwise correlations among (a) recency vs. frequency, (b) recency vs. monetary, and (c) frequency vs. monetary

Table 2 presents the iterative results of the k-means clustering, validated using three evaluation metrics for each number of clusters. To identify the optimal clustering solution, the validation scores are normalized and assigned equal weights of 0.33. Equal weights were assigned to the normalized silhouette, DBI, and CHI to ensure a balanced evaluation, as each metric captures complementary aspects of clustering

quality and no single metric is theoretically dominant for the given application [21]. The cluster configuration with the highest combined weighted score is selected as the best-performing solution. Although k =2 yields the highest combined weighted score, this solution provides limited analytical insight due to the overly coarse segmentation. Consequently, k =7 is selected as the optimal number of clusters, achieving a combined weighted score of 0.49. Figure 6 illustrates the resulting k-means clustering solution.

Similar to Table 2, Table 3 presents the results of the hierarchical clustering iterations, evaluated for cluster numbers ranging from two to ten. For each configuration, the validation metrics, normalized validation scores, and combined weighted scores are reported. Based on these results, k =3 is selected as the optimal number of clusters for the hierarchical clustering analysis. Figure 7 illustrates the hierarchical clustering solution.

Table 2. K-means clustering results

k	Silhouette	DBI	CHI	Silhouette_norm	DBI_norm	CHI_norm	Weighted_score
2	0.725640681	0.43804690	3380.43216	1	1	0.20328174	0.72708297
7	0.618539941	0.46839858	3869.64461	0.26400870	0.764641245	0.45971310	0.49115980
6	0.620620343	0.46440856	3502.42214	0.27830513	0.795581374	0.26722545	0.44256694
9	0.594375815	0.53529723	4852.05627	0.09795395	0.245882872	0.97466557	0.43510579
3	0.664009699	0.49082345	3160.21805	0.57647476	0.590750008	0.08785172	0.41417524

Table 3. Hierarchical clustering results

k	Silhouette	DBI	CHI	Silhouette_norm	DBI_norm	CHI_norm	Weighted_score
2	0.70044913	0.48555912	2980.290047	1	1	0.044730974	0.67476122
3	0.65798431	0.49189968	3121.721262	0.794662519	0.962598208	0.123692485	0.62071456
10	0.50243429	0.55424105	4691.311752	0.042504732	0.594857766	1	0.54032962
9	0.51976614	0.55066791	4033.047773	0.126312397	0.615935075	0.632489063	0.45366305
4	0.67108371	0.57117187	2900.170557	0.858004308	0.494985887	0	0.44648676

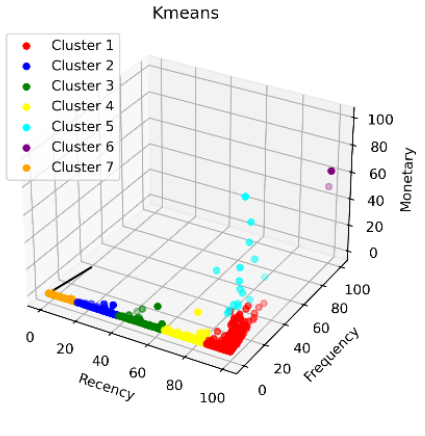


Figure 6. K-means clustering result

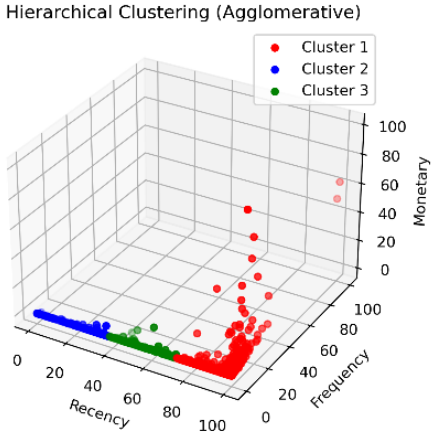


Figure 7. Hierarchical clustering result

The DBSCAN configuration differs from those of k-means and hierarchical clustering, as it does not require prior specification of the clusters number. Instead, DBSCAN relies on two parameters, ε (epsilon) and min_samples. In this study, ε was varied from 0.1 to 5.0 with increments of 0.1, while min_samples ranged from 6 to 20 with increments of 1. Table 4 presents the top ten parameter combinations obtained from this iterative process, from which ε=1.6 and min_samples =7 were identified as producing the best clustering performance. The resulting DBSCAN clustering structure based on these parameters is illustrated in Figure 8.

Table 4. Top 10 DBSCAN clustering results

eps	min_samples	Clusters	Silhouette	DBI	CHI	Silhouette norm	DBI_norm	CHI_norm	Weighted_ score
1.6	7	3	0.48902	1.13440	1706.92	0.93387	0.73158	1	0.879602
2.1	12	3	0.53151	1.19885	1496.54	0.97025	0.71105	0.874985	0.843576
2	10	3	0.51923	1.24785	1464.34	0.95973	0.69544	0.855853	0.828642
1.4	6	4	0.46177	1.04308	1389.07	0.91055	0.76067	0.811123	0.819175
1.9	9	3	0.53032	1.29533	1393.93	0.96923	0.68032	0.814012	0.812976

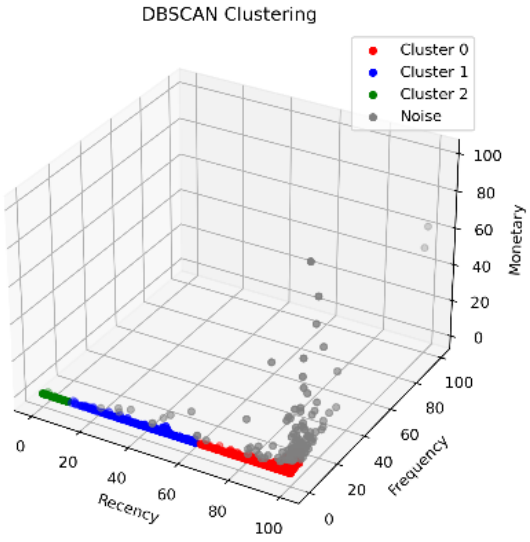


Figure 8. DBSCAN clustering result

3.2. Discussion

This subsection evaluates the clustering results using validation metrics and extends the analysis to strategic customer insights. Figures 2 to 4 show that the frequency and monetary variables are left-skewed, while recency is right-skewed—an expected pattern aligned with the Pareto principle, which states that a small proportion of customers contributes to the majority of revenue [26]. In this dataset, 84% of the revenue is generated by just 16% of the customers. Identifying these high-value "Pareto customers" is essential for effective segmentation and targeted marketing. Clustering techniques that can accurately isolate these customers provide significant business value.

The experimental results, shown in Table 5, indicate that k-means achieved the best performance on the DBI and CHI, while hierarchical clustering had the highest silhouette score. DBSCAN performed the worst across all metrics. Visualizations in Figures 6 to 8 support these findings: k-means is the only method that clearly separates Pareto customers into distinct clusters (e.g., purple and cyan points with high values). Hierarchical clustering fails to distinguish these customers, grouping them with lower-value segments. Although DBSCAN is effective at identifying noise, in this context it misclassifies Pareto customers as outliers—limiting its usefulness when high-value customers may appear as sparse but critical cases in data.

Table 5. Comparison of clustering performance across methods

Algorithm	k	Silhouette	DBI	CHI
K-means	7	0.61853994	0.46839858	3.869.644.617
Hierarchical	3	0.65798431	0.49189968	3.121.721.262
DBSCAN	3	0.48902765	113.440.359	1.706.924.696

This study has several limitations. First, the analysis is based on a single dataset from one distribution business, which may limit the generalizability of the findings to other industries or customer structures. Future research should validate the proposed approach across multiple datasets and sectors to enhance external validity. Second, the study employs equal weighting across clustering validation metrics without conducting sensitivity analysis; although this ensures a balanced evaluation, the absence of robustness testing may influence the stability of the final clustering selection under alternative weighting schemes. Future work should incorporate sensitivity analysis to further assess the consistency of the results.

4. CONCLUSION

This study evaluated k-means, hierarchical clustering, and DBSCAN for RFM-based customer segmentation using the silhouette score, DBI, and CHI. The results show that k-means consistently outperforms other methods and is uniquely effective in identifying high-value Pareto customers, making it particularly suitable for data-driven marketing applications. Academically, the study contributes a systematic

comparison of clustering techniques that integrates quantitative validation metrics with business interpretability, emphasizing the importance of aligning methodological choices with analytical objectives. Practically, the findings demonstrate that k-means offers a reliable and actionable approach for customer segmentation to support targeted marketing and resource allocation decisions. Nevertheless, the generalizability of the results is limited by the dataset and methods used, and future research should explore additional clustering algorithms and enhanced RFM models tailored to other business contexts.

ACKNOWLEDGMENTS

The researchers would like to express their appreciation to the owner of Megacomp for granting permission to use the company's sales data.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Upekkha Lau	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Meditya Wasesa	✓	✓		✓						✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data that support the findings of this study are available from Megacomp. Restrictions apply to the availability of these data, which were used under permision for this study. Data are available from the authors with the permission of Megacomp.

REFERENCES

[1] T. H. Davenport, "Competing on analytics," *Harvard Business Review*, vol. 84, no. 1, pp. 98–107, 2006.
[2] B. Cooil, L. Aksoy, and T. L. Keiningham, "Approaches to customer segmentation," *Journal of Relationship Marketing*, vol. 6, no. 3–4, pp. 9–39, Jan. 2008, doi: 10.1300/J366v06n03_02.
[3] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*. Massachusetts, United States: Morgan Kaufmann, 2012, doi: 10.1016/C2009-0-61819-5.
[4] C. Rungruang, P. Riyapan, A. Intarasit, K. Chuarkham, and J. Muangprathub, "RFM model customer segmentation based on hierarchical approach using FCA," *Expert Systems with Applications*, vol. 237, Mar. 2024, doi: 10.1016/j.eswa.2023.121449.
[5] T. Ho, S. Nguyen, H. Nguyen, N. Nguyen, D.-S. Man, and T.-G. Le, "An extended RFM model for customer behaviour and demographic analysis in retail industry," *Business Systems Research Journal*, vol. 14, no. 1, pp. 26–53, Sep. 2023, doi: 10.2478/bsrj-2023-0002.
[6] L. Buitinck *et al.*, "API design for machine learning software: experiences from the scikit-learn project," in *European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases*, 2013, pp. 1–15.
[7] P. Bíró, B. B. H. Kovács, T. Novák, and M. Erdélyi, "Cluster parameter-based DBSCAN maps for image characterization," *Computational and Structural Biotechnology Journal*, vol. 27, pp. 920–927, Jan. 2025, doi: 10.1016/j.csbj.2025.02.037.
[8] P. Anitha and M. M. Patil, "RFM model for customer purchase behavior using k-means algorithm," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1785–1792, May 2022, doi: 10.1016/j.jksuci.2019.12.011.

[9] O. N. Akande, H. B. Akande, E. O. Asani, and B. T. Dautare, "Customer segmentation through RFM analysis and k-means clustering: leveraging data-driven insights for effective marketing strategy," in *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)*, 2024, pp. 1–8, doi: 10.1109/SEB4SDG60871.2024.10630052.
[10] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, "RFM ranking – an effective approach to customer segmentation," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 10, pp. 1251–1257, Dec. 2021, doi: 10.1016/j.jksuci.2018.09.004.
[11] N. P. Sutramiani, I. M. T. Arthana, P. F. Lampung, S. Aurelia, M. Fauzi, and I. W. A. S. Darma, "The performance comparison of DBSCAN and k-means clustering for MSMEs grouping based on asset value and turnover," *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 1, pp. 13–24, Feb. 2024, doi: 10.20473/jisebi.10.1.13-24.
[12] A. Griva, C. Bardaki, K. Pramatari, and D. Papakiriakopoulos, "Retail business analytics: customer visit segmentation using market basket data," *Expert Systems with Applications*, vol. 100, pp. 1–16, Jun. 2018, doi: 10.1016/j.eswa.2018.01.029.
[13] J. Zhou, J. Wei, and B. Xu, "Customer segmentation by web content mining," *Journal of Retailing and Consumer Services*, vol. 61, Jul. 2021, doi: 10.1016/j.jretconser.2021.102588.
[14] M. Hosseini, N. Abdolvand, and S. R. Harandi, "Two-dimensional analysis of customer behavior in traditional and electronic banking," *Digital Business*, vol. 2, no. 2, 2022, doi: 10.1016/j.digbus.2022.100030.
[15] H. Abbasimehr and A. Bahrini, "An analytical framework based on the recency, frequency, and monetary model and time series clustering techniques for dynamic segmentation," *Expert Systems with Applications*, vol. 192, Apr. 2022, doi: 10.1016/j.eswa.2021.116373.
[16] S. Haghighatnia, N. Abdolvand, and S. R. Harandi, "Evaluating discounts as a dimension of customer behavior analysis," *Journal of Marketing Communications*, vol. 24, no. 4, pp. 321–336, May 2018, doi: 10.1080/13527266.2017.1410210.
[17] R. J. Kuo, M. N. Alfareza, and T. P. Q. Nguyen, "Genetic based density peak possibilistic fuzzy c-means algorithms to cluster analysis- a case study on customer segmentation," *Engineering Science and Technology, an International Journal*, vol. 47, Nov. 2023, doi: 10.1016/j.jestch.2023.101525.
[18] S. Monalisa, P. Nadya, and R. Novita, "Analysis for customer lifetime value categorization with RFM model," *Procedia Computer Science*, vol. 161, pp. 834–840, 2019, doi: 10.1016/j.procs.2019.11.190.
[19] H. H. Hamidi and B. Haghi, "An approach based on data mining and genetic algorithm to optimizing time series clustering for efficient segmentation of customer behavior," *Computers in Human Behavior Reports*, vol. 16, Dec. 2024, doi: 10.1016/j.chbr.2024.100520.
[20] Scikit Learn, "2.3. Clustering," *scikit-learn.org*. Accessed: Mar. 22, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/clustering.html
[21] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognition*, vol. 46, no. 1, pp. 243–256, Jan. 2013, doi: 10.1016/j.patcog.2012.07.021.
[22] L. Lovmar, A. Ahlford, M. Jonsson, and A.-C. Syvänen, "Silhouette scores for assessment of SNP genotype clusters," *BMC Genomics*, vol. 6, no. 1, Mar. 2005, doi: 10.1186/1471-2164-6-35.
[23] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.
[24] F. Ros, R. Riad, and S. Guillaume, "PDBI: a partitioning Davies-Bouldin index for clustering evaluation," *Neurocomputing*, vol. 528, pp. 178–199, Apr. 2023, doi: 10.1016/j.neucom.2023.01.043.
[25] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, Jan. 1974, doi: 10.1080/03610927408827101.
[26] R. Sanders, "The pareto principle: Its use and abuse," *Journal of Product & Brand Management*, vol. 1, no. 2, pp. 37–40, 1992, doi: 10.1108/eb024706.

BIOGRAPHIES OF AUTHORS

Upekkha Lau graduated with a degree in Informatics from Parahyangan Catholic University in 2022. He began his professional journey as a data analyst at a retail distribution company, where he gained hands-on experience in leveraging data to support business decisions. His work involved analyzing sales trends and monitoring inventory performance. In 2025, he pursued further education by enrolling in the Master of Science Management program at the School of Business and Management, Institut Teknologi Bandung, with a focus on business analytics. His current academic interests include data-driven decision making, clustering, and the integration of analytics into strategic business processes. He can be contacted at email: upklau2121@gmail.com.

Meditya Wasesa is an associate professor at the School of Business and Management, Institut Teknologi Bandung. He holds a Ph.D. in Management Information Systems from Rotterdam School of Management, Erasmus University, Netherlands, an M.Sc. in Logistics Engineering from Duisburg-Essen University, Germany, and a Bachelor of Mechanical Engineering (S.T.) from Institut Teknologi Bandung. In Indonesia, he has served as an executive, consultant, and trainer for a diverse international and national organizations. His primary interest revolves around leveraging advanced information systems to improve business decisions, with a particular emphasis on business analytics. His work has been featured in renowned i.e., IEEE Access, Journal of Enterprise Information Management, Journal of Management Analytics, Resources Policy, Journal of Cleaner Production, and Decision Support Systems. He can be contacted at email: meditya.wasesa@itb.ac.id.