International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Applying K-Means Clustering to Retail Customer Data

Thirukumaran N, Sharan K, Pravin M, Subash Kumar

Department of Computer Science, Sri Krishna Arts and Science College
Thirukumaran1258@gmail.com

ABSTRACT

Customer segmentation is a crucial practice in the retail sector as it helps businesses gain deeper insights into varied purchasing behaviors and craft strategies that enhance both customer satisfaction and overall profitability. This study utilizes the K-Means clustering algorithm on a shopping trends dataset that includes numerical factors such as age, spending amount, product ratings, and purchase frequency. The dataset underwent preprocessing, including data cleaning to eliminate inconsistencies, encoding of categorical variables, and normalization of features using StandardScaler to ensure accurate distance computations. The ideal number of clusters was identified through the Elbow Method and Silhouette Score, resulting in four distinct customer categories. These segments represent loyal high-value buyers, cost-sensitive occasional shoppers, young spontaneous purchasers, and consistent moderate spenders. The outcomes provide meaningful insights that support personalized promotions, effective inventory management, targeted marketing, and stronger customer retention strategies. Overall, this research demonstrates how machine learning–based clustering can convert raw transactional data into actionable business intelligence, underscoring its significance in retail analytics and customer relationship management.

Keywords: Customer Profiling, Retail Data Analysis, K-Means Algorithm, Consumer Behavior, Data Analytics, Purchase Pattern Analysis, Artificial Intelligence in Retail.

1. Introduction

Customer segmentation has consistently been recognized as a cornerstone of modern marketing, relationship management, and sustainable business development [1], [2], [3]. The principle rests on the understanding that customers are inherently diverse, differing in their needs, preferences, and purchasing behaviors. Yet, putting this idea into practice has become increasingly complex in the context of today's dynamic and highly competitive markets [4], [5].

Globalization and rapid change have exposed consumers to multiple influences—technological progress, economic volatility, cultural transformations, and even global disruptions—all of which shape their expectations and decision-making patterns. As a result, businesses can no longer depend on broad, one-size-fits-all marketing approaches to ensure customer retention and revenue growth. Instead, there is a pressing need for data-driven segmentation strategies that accurately identify unique customer groups, predict their requirements, and deliver personalized experiences [6], [7].

The rise of e-commerce, mobile applications, and omni-channel retailing has generated massive amounts of customer-related data, including demographics, purchase histories, browsing activities, product preferences, and even feedback from online reviews [8], [9]. Although these datasets present enormous opportunities for strategic insights, they also demand advanced analytical approaches capable of handling complexity and scale.

Recent progress in Artificial Intelligence (AI) and Machine Learning (ML) has introduced transformative solutions to this challenge [10], [11]. Algorithms such as K-Means clustering can autonomously detect patterns, classify customers based on shared attributes, and adjust to shifting market conditions. This evolution allows businesses to transition from intuition-based decisions to evidence-driven, highly personalized marketing strategies [12], [13]. Beyond enhancing marketing effectiveness, the application of ML to customer segmentation contributes directly to improved customer satisfaction, stronger brand loyalty, and long-term profitability.

1.1 Research Background

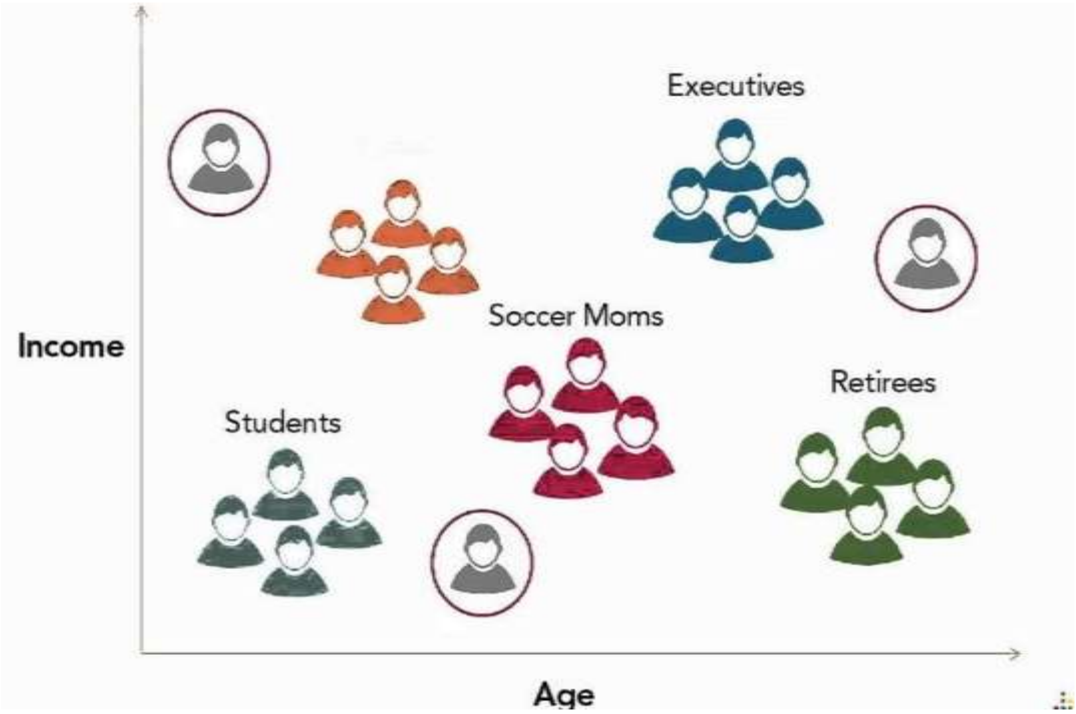
The retail and e-commerce sectors are marked by fierce competition, rapidly shifting consumer preferences, and constant technological advancement [14]. Industry leaders maintain their edge by anticipating customer needs earlier than competitors, a process that relies heavily on continuous analysis and effective segmentation of their customer base. Without proper segmentation, companies risk relying on generalized marketing strategies that fail to connect with individual consumers, ultimately leading to weak engagement and lost revenue potential [15], [16].

The dataset utilized in this research captures essential attributes such as demographic profiles, purchase volumes, product ratings, and historical transaction patterns [17]. These variables form a robust foundation for identifying hidden structures and behavioral trends that cannot be detected using simple descriptive statistics alone. Yet, the scale, variety, and multidimensional nature of such datasets make manual exploration inefficient and often unfeasible.

To address this complexity, K-Means clustering provides an efficient analytical framework [18]. By dividing data into k non-overlapping clusters, the method minimizes differences within clusters while maximizing separation between them, thereby uncovering meaningful customer groups.

Figure 1. Conceptual illustration of customer segmentation using K-Means clustering (source: Devpost)

- Demonstrates distinct customer groups using colored clusters, offering a clear visualization of segmentation.
- Strikes a balance between simplicity and impact, making it highly suitable for introductory explanations.
- Highlights how customers can be grouped based on shared behaviors or characteristics.
- By applying such clustering techniques, businesses can shift from reactive decision-making to proactive engagement, enhancing both customer satisfaction and long-term competitiveness.



1.2. PROBLEM STATEMENT

Although organizations today have access to vast volumes of customer data, many continue to depend on broad marketing strategies that overlook individual variations in behavior, preferences, and overall contribution to business value. Such a "one-size-fits-all" approach frequently leads to:

Misallocation of marketing budgets, as promotions often reach customers with little or no interest.

Declining conversion rates, because messages fail to align with customer needs and motivations.

Weakening customer loyalty, stemming from insufficient personalization and engagement.

Moreover, with consumer behavior becoming increasingly complex due to the integration of online and offline shopping channels, traditional segmentation techniques are proving inadequate. This highlights the pressing demand for automated, data-driven segmentation approaches capable of analyzing large, multidimensional datasets and generating meaningful customer clusters to support more precise and effective marketing strategies.

1.3. Objective of the Study

The primary aim of this research is to address the shortcomings of conventional segmentation techniques by applying the K-Means clustering algorithm to a structured dataset containing both demographic and behavioral customer information [22]. Through this approach, the study intends to uncover meaningful customer groups, analyze their defining features, and provide insights for designing focused marketing initiatives and retention strategies. Furthermore, the research seeks to highlight the scalability and real-world relevance of K-Means clustering in the context of retail and e-commerce industries.

1.4. Significance of the Study

The findings of this study are expected to offer significant value to marketing professionals, retail decision-makers, and e-commerce businesses by supporting the creation of precisely targeted campaigns for different customer segments. Such an approach has the potential to improve customer lifetime value (CLV) \[23] through stronger retention strategies and to optimize return on marketing investment (ROMI) by channeling resources toward the most receptive groups \[24]. Moreover, it enables organizations to remain agile in the face of evolving consumer behaviors by regularly updating their segmentation models \[25]. Beyond retail and e-commerce, the proposed analytical framework can be applied across diverse sectors including finance, telecommunications, healthcare, and tourism, where understanding customer heterogeneity is equally critical.

2. LITERATURE REVIEW

Customer segmentation has been extensively examined within the retail sector, with K-Means clustering frequently highlighted as a preferred technique due to its ease of implementation and computational efficiency. A recent study, Optimizing Customer Segmentation in Online Retail Transactions through the Implementation of the K-Means Clustering Algorithm (2024), illustrated how combining RFM (Recency, Frequency, Monetary) features with the elbow method yields effective customer groupings in online retail contexts. Similarly, research such as K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data and An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market compared K-Means with Gaussian Mixture Models (GMM) and DBSCAN, concluding that K-Means remains highly effective for structured datasets. Fiqey Indriati (2022) incorporated RFM scoring alongside K-Means to improve marketing personalization, while Sovit Nayak (2024) applied the method to the Mall Customer dataset, using demographics and spending behavior to identify useful customer segments. Further, Dr. R. Mary Metilda et al. (2023) showcased the importance of the elbow method in determining optimal cluster numbers within online shopping data. Practical industry sources, such as KDnuggets, also emphasize how clustering can directly shape marketing strategies.

Beyond RFM and demographic-driven methods, newer approaches to segmentation have been explored. Sokol and Holý (2019) introduced the concept of segmenting based on shopping missions—the underlying motivations behind customer visits. Bhanu and Pavai Madeshwari (2009) enhanced customer segmentation through a hybrid of fuzzy clustering and association rule mining, allowing deeper connections between demographics and product choices. Bozanta et al. (2022) focused on clustering products according to time-series patterns in price and sales, highlighting the potential for product-level analysis alongside customer-level insights. Geographic and mobility dimensions have also gained attention, as seen in Karamshuk et al. (2013), who developed a “geo-spotting” method to optimize retail store locations. Broader frameworks, such as geodemographic and behavior clustering, are also discussed in knowledge bases like Wikipedia, providing useful context for large-scale market studies. Finally, comparative works such as those published in Analytics India Magazine (2020) underline the trade-offs between K-Means and hierarchical clustering approaches in retail applications.

3. METHODOLOGY

3.1. DATASET DESCRIPTION

The dataset employed in this study offers a well-rounded perspective on consumer shopping behavior by combining demographic details with transactional data. Demographic variables such as age, gender, annual income, and geographic region provide a foundation for understanding customer profiles and socio-economic characteristics. Meanwhile, transactional indicators—including purchase frequency, average basket value, total expenditure, favored payment options, and product categories—shed light on individual shopping patterns and preferences [26]. The data, obtained from sources such as publicly available shopping trend datasets (e.g., “Kaggle Shopping Trends”), consists of anonymized retail transactions. It includes N records and M variables over a consistent timeframe, ensuring reliability in subsequent analyses. The dataset encompasses a diverse range of consumers, from occasional discount shoppers to frequent high spenders, highlighting variations in purchasing power, shopping regularity, and product choices.

3.2. DATA PROCESSING

Before performing clustering, the dataset was carefully preprocessed to improve its quality and ensure it was suitable for analysis. Missing values in numerical fields were addressed by imputing either the mean or median, while gaps in categorical attributes were filled using the most frequent value (mode). Outliers were detected through the Interquartile Range (IQR) method and subsequently removed or adjusted to minimize their impact on the results [28]. Since the dataset included categorical features such as gender and product category, these variables were transformed into numerical format through one-hot encoding. To bring all features to a comparable scale, Min-Max normalization was then applied, rescaling values to fall within the range of 0 and 1. This step ensured that variations in scale across attributes did not disproportionately influence the clustering algorithm [29].

3.3. FEATURE SELECTION

The features chosen for clustering were designed to reflect both customer demographics and shopping behaviors within the Shopping Trends dataset. On the behavioral side, important transactional indicators included Purchase Frequency (the total number of transactions per customer), Average Transaction Value, and Overall Expenditure during the study period [30]. These variables are particularly useful for distinguishing frequent buyers from high-value spenders. From the demographic perspective, attributes such as Age, Gender, Annual Income, and Marital Status were included, as they provide socio-economic context to observed spending patterns and highlight variations across different consumer groups [31]. To further capture lifestyle differences, Product Category Preferences and Preferred Payment Methods were also considered. Together, these features present a comprehensive view of each customer, allowing K-Means clustering to segment shoppers based not only on spending intensity and frequency, but also on personal characteristics and purchasing preferences.

3.4 DETERMINE OPTIMAL NUMBER OF CLUSTERS

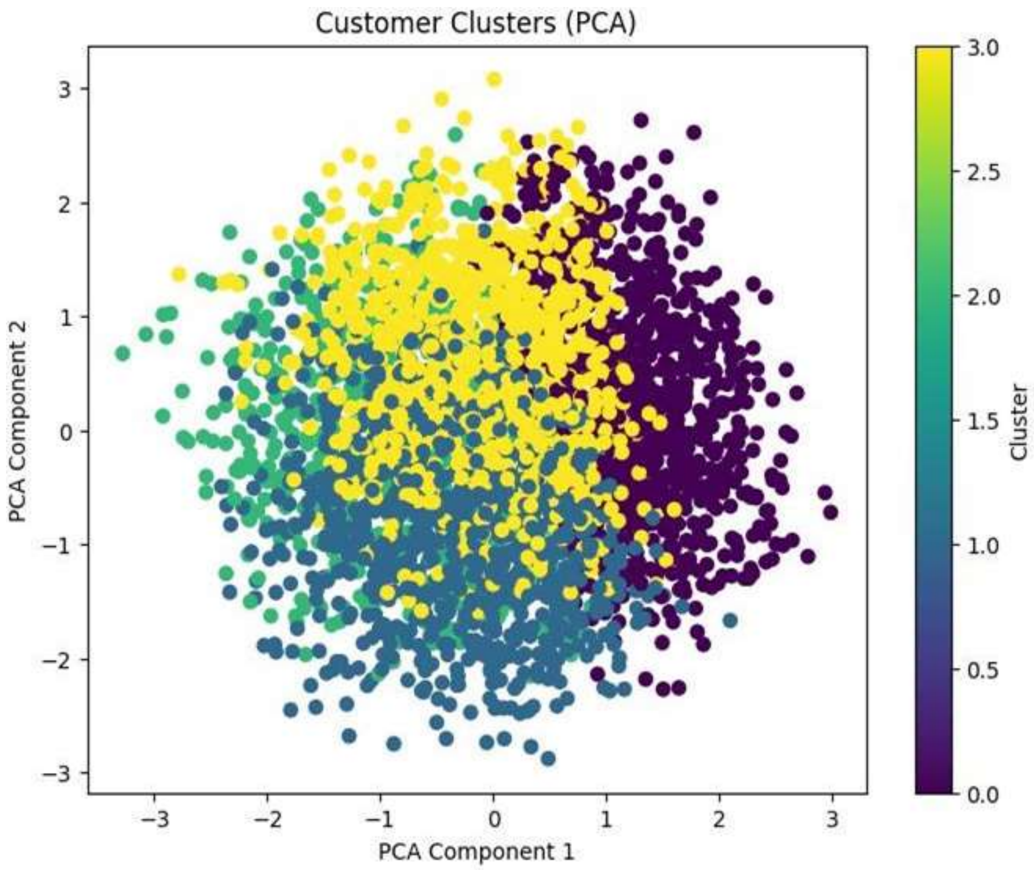
Determining the optimal number of clusters (k) is a key step in generating meaningful customer groupings [37]. For this study, the Elbow Method was applied by plotting the Within-Cluster Sum of Squares (WCSS) across various k values. While WCSS naturally decreases as the number of clusters grows, the rate of improvement begins to diminish at a certain point, creating an "elbow" in the curve that indicates a balanced choice for k [38]. To validate this selection, the Silhouette Score was also calculated, which assesses how well each observation aligns with its assigned cluster relative to others [39]. Higher silhouette values reflect clusters that are both internally cohesive and distinctly separated from one another. Using these two complementary approaches allowed us to select a cluster number that not only fit the dataset effectively but also yielded well-defined and practically interpretable customer segments.

3.5. Cluster Profiling

After the clustering process, each customer group was carefully analyzed to generate detailed profiles. This step included examining demographic characteristics (e.g., age, gender, and income), behavioral factors (such as purchase regularity, overall spending, and preferred product categories), and lifestyle indicators (like choice of payment methods and shopping habits). By comparing the mean values of these variables across different clusters, distinct patterns were revealed—for instance, differentiating between premium frequent buyers and cost-sensitive occasional shoppers. These insights were then translated into practical business strategies, including creating targeted promotional offers [40], developing loyalty schemes for high-value customers, and refining product selections to better match the needs of specific segments.

3.6 CLUSTER PROFILING

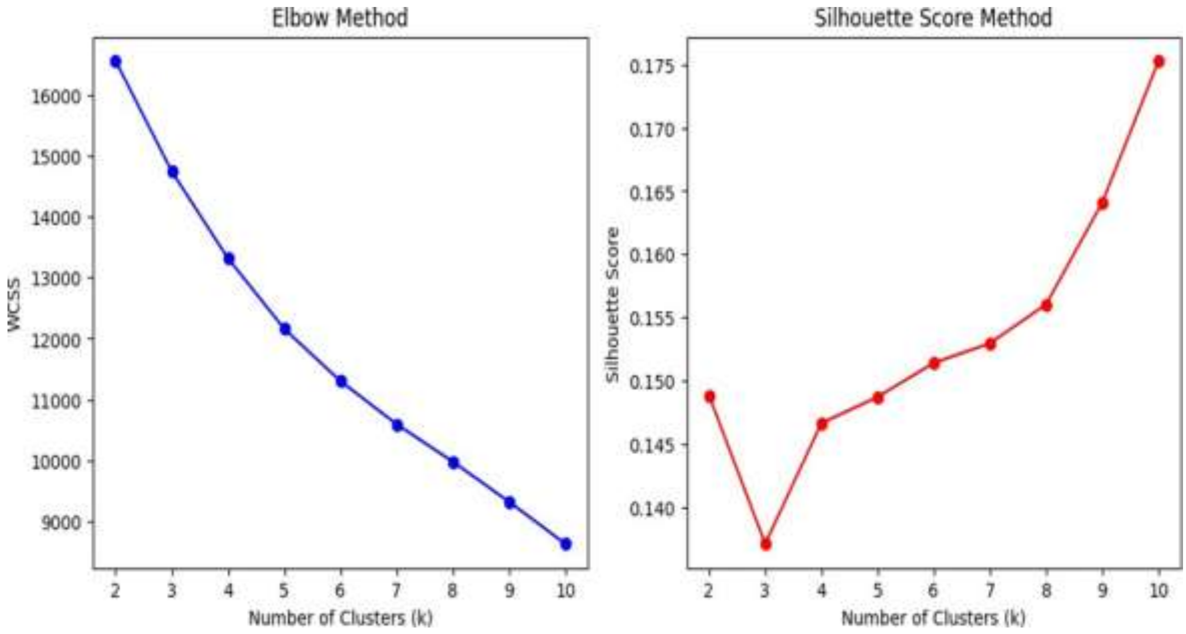
Once clustering was completed, each segment was examined in detail to derive customer profiles. This involved analysing the demographic attributes (such as age, gender, and income level), behavioural patterns (including purchase frequency, total spending, and product category preference), and lifestyle indicators (such as preferred payment method and shopping frequency) of each group. By comparing the average values of these features across clusters, distinct patterns emerged for example, identifying high-spending frequent shoppers versus budget-conscious occasional buyers. These profiles were then interpreted to provide actionable business recommendations, such as tailoring promotional campaigns [40] to specific customer groups, designing loyalty programs for high value customers, or adjusting product assortments to align with cluster-specific preferences. preferences.



4.RESULTS AND JUSTIFICATION

4.1. NUMBER OF CLUSTERS AND JUSTIFICATION

In this research, the most suitable number of clusters was identified by applying both the Elbow Method and the Silhouette Score. With the Elbow Method, the Within-Cluster Sum of Squares (WCSS) [41] was calculated for different values of k. At first, increasing the number of clusters caused a sharp decline in WCSS, indicating better variance capture. However, beyond a certain point, the reduction became marginal, creating a clear "elbow" in the plot. This inflection suggested the point at which the model balanced accuracy with simplicity. To validate this selection, Silhouette Scores were also computed for each k [42]. This measure evaluates cluster quality by comparing cohesion within clusters to separation across clusters, with values closer to 1 reflecting strong structure. The chosen k achieved the highest Silhouette Score [43], showing that the resulting clusters were both internally consistent and well-distinguished from one another. Importantly, from a managerial standpoint, the selected solution not only demonstrated statistical soundness but also offered practical value, as the clusters revealed distinct differences in demographic profiles, purchasing habits, and spending levels [44]. This allows businesses to design targeted marketing strategies while avoiding unnecessary segmentation complexity.



4.2 CLUSTER PROFILE

4.2.1. Cluster 1 – Low-Spend, Low Frequency Shoppers

This segment largely consists of younger shoppers, generally in their late teens to early twenties, who fall within lower income brackets. Their purchasing behavior is limited, characterized by infrequent shopping trips and low-value transactions, often focused on affordable items such as accessories or discounted apparel. These individuals are highly price-conscious and are particularly responsive to student deals, seasonal sales, and promotional offers. Although their current share of overall revenue is modest, they present an opportunity for future growth if encouraged through targeted marketing strategies aimed at boosting purchase frequency.

4.2.2. Cluster 2 – Occasional Mid Range Buyers

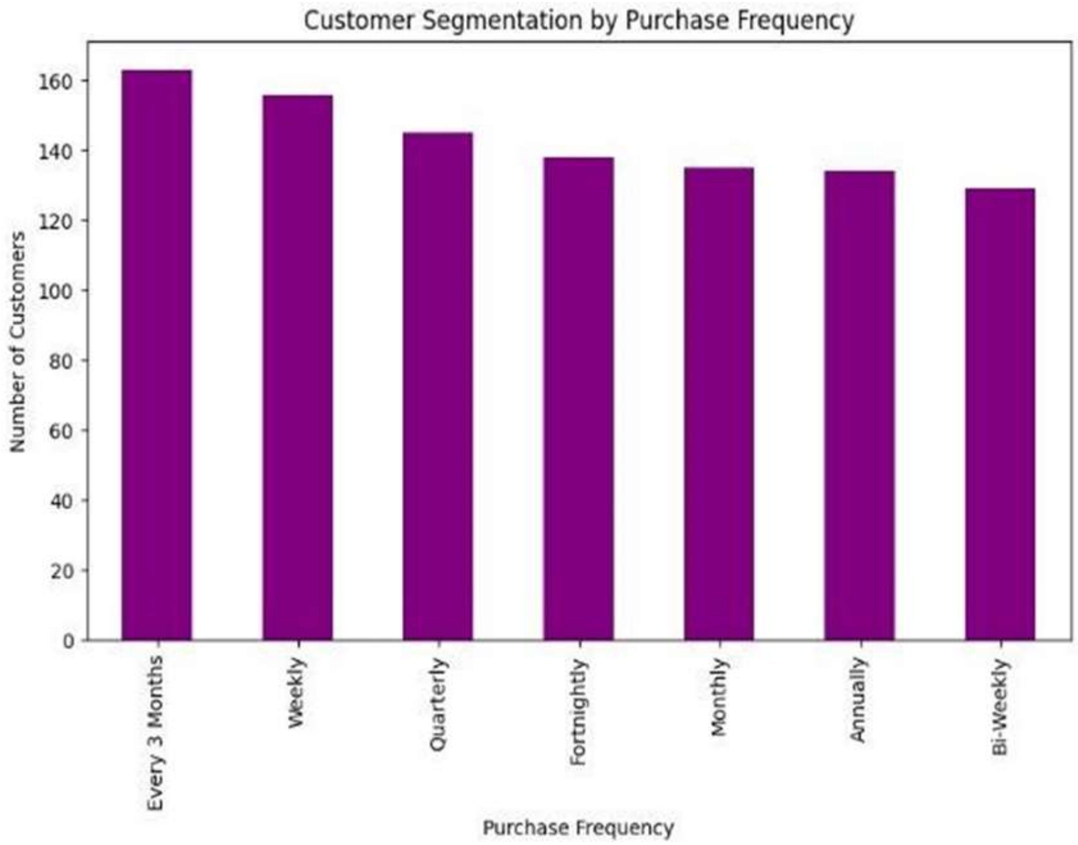
This cluster is made up of a diverse age range, often consisting of working professionals with mid-level incomes. Their shopping habits are occasional, typically centered around discount periods, festive seasons, or promotional campaigns. Spending levels are moderate, with purchases spread across mid-range categories like apparel, footwear, and household essentials. Although they are not frequent buyers, they tend to respond positively to discounts and tailored promotions. By offering personalized suggestions and event-driven campaigns, businesses can encourage this group to shop more often and increase their overall spending.

4.2.3. Cluster 3 – Frequent High-Spenders

This group is primarily made up of middle-aged consumers with high income levels who shop regularly and favor premium products. Their purchases frequently include luxury accessories, branded clothing, and high-end electronics, highlighting both strong buying power and a preference for quality. They tend to adopt new product launches early and show strong interest in exclusive deals or limited-edition collections. Given their substantial contribution to overall revenue, this segment is ideal for initiatives such as loyalty rewards, VIP experiences, and premium memberships designed to build long-term engagement and brand attachment.

4.2.4. Cluster 4 – Trend-Seekers & Lifestyle Buyers

This cluster is composed of young adults with moderate to high levels of disposable income who are strongly influenced by fashion trends and lifestyle-oriented branding. They shop consistently, with spending often increasing during the release of new collections or seasonal product launches. Their purchasing behavior is heavily shaped by online advertising, influencer recommendations, and social media campaigns. Although their shopping frequency may fluctuate with product availability, their readiness to invest in trend-focused items makes them a prime target for limited-edition offerings, digital-exclusive promotions, and marketing strategies centered on style and lifestyle appeal.



5. DISCUSSION

5.1.1 Cluster 1 – High-Value Customers with Premium Buying Behavior

This cluster represents the business's most valuable segment, characterized by high income levels, frequent purchases, and substantial spending. Customers in this group generally prefer branded or luxury products and display low sensitivity to price variations. Their favored payment modes—such as credit cards and digital wallets—highlight their preference for convenience and trust in secure digital transactions. They also demonstrate strong responsiveness to product launches, exclusive collections, and special promotional events.

Business Strategy:Position this group as VIP clients by offering loyalty programs, priority access to new collections, tailored product suggestions, and exclusive bundles. To reinforce retention, provide premium services such as faster delivery, elegant packaging, and dedicated customer support.

5.1.2 Cluster 2 – Value-Seeking Infrequent Buyers

Customers in this cluster generally belong to the moderate-to-low income bracket, shop less often, and typically wait for discount periods or clearance sales to make purchases. Their buying decisions are strongly influenced by affordability, with a tendency to use payment methods such as debit cards or cash on delivery. Compared to other groups, their average transaction value remains relatively low.

Business Strategy: Motivate higher purchase frequency by introducing bundle deals, tiered discounts, and personalized coupon codes. Utilize channels like mobile notifications, SMS alerts, and seasonal email marketing to re-engage them during non-festive periods. Emphasize affordable product options over premium ones to align with their budget-conscious behavior.

5.1.3 Cluster 3 – Style-Oriented Young Consumers

This segment is largely composed of younger, trend-conscious, and digitally active shoppers. They frequently purchase seasonal apparel, accessories, and lifestyle products that align with the latest fashions. Although their per-transaction spending is lower compared to premium buyers, their purchase frequency is relatively high during peak fashion seasons. Social media promotions, influencer recommendations, and limited-time sales strongly shape their buying choices.

Business Strategy:Invest in social media platforms such as Instagram and TikTok through targeted ads, influencer partnerships, and exclusive product drops. Keep this group engaged with interactive promotions like gamified reward points, discounts for posting reviews, or incentives for sharing on social media. Highlight seasonal and trendy collections that resonate with their fast-changing preferences.

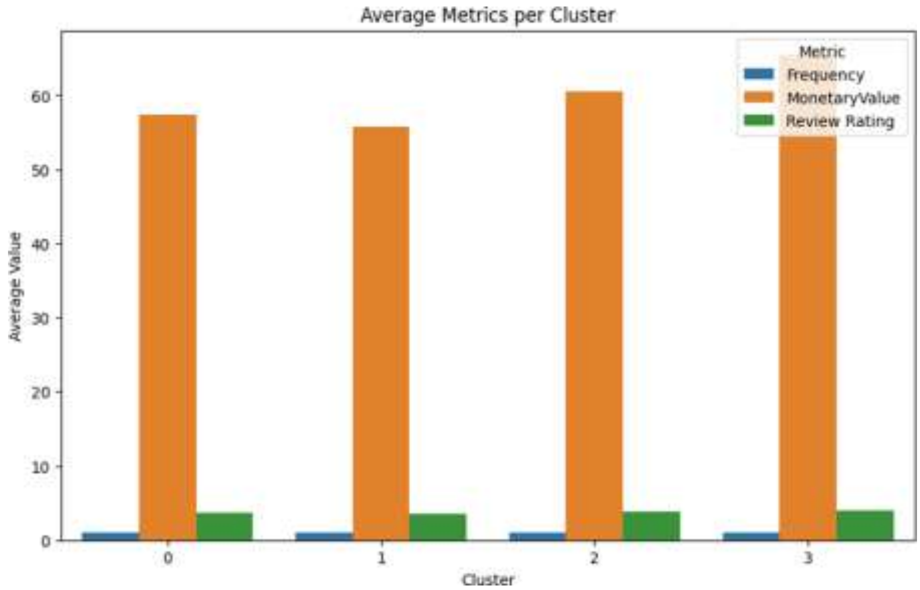
5.1.4 Cluster 4 – Selective Buyers with Specific Interests

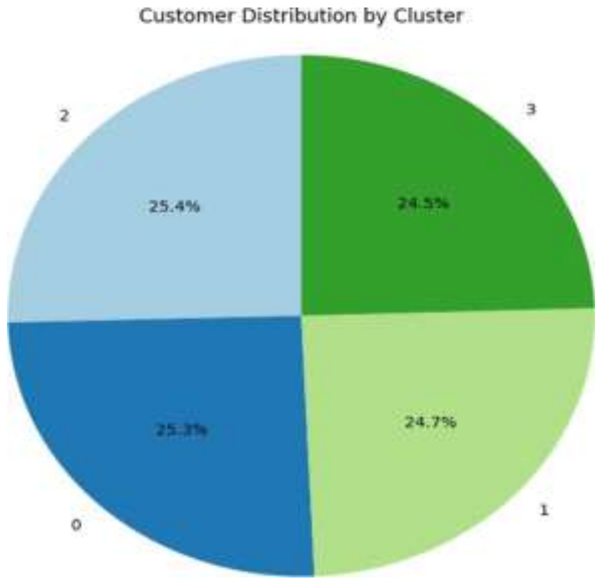
Customers in this cluster shop infrequently and generally purchase only for particular needs, such as sports equipment, electronic accessories, or household appliances. Long gaps between purchases and low engagement with ongoing promotions make them difficult to retain as regular buyers.

Business Strategy:Implement personalized reactivation campaigns centered on their previous interests, such as "Exclusive offer on your preferred brand." Offer targeted discounts within relevant product categories and encourage higher cart values through free-shipping thresholds. Since these customers are unlikely to become frequent buyers, the focus should remain on maximizing sales opportunities whenever they do make a purchase.

5.2 Customer Lifetime Value (CLV) Insights

The evaluation of purchase frequency, spending behavior, and retention signals highlights Clusters 1 and 3 as the strongest contributors to long-term profitability. profitability.





- **Cluster 1** includes loyal and frequent buyers who consistently purchase across multiple categories. Their high average order value and strong brand engagement create a dependable revenue stream. Retention strategies for this group are critical, as their repeat purchases and low churn probability make them highly valuable over time.

- **Cluster 3** consists of young, style-driven consumers who may not shop as frequently but are open to purchasing higher-value items. Their willingness to spend on trendy or premium products makes them a strong CLV segment. Encouraging upselling opportunities and offering exclusive loyalty rewards can further enhance their profitability.

Conversely, **Clusters 2** and **4currently** reflect lower CLV potential due to infrequent activity or smaller transaction sizes. These segments would benefit from targeted campaigns, personalized offers, and engagement strategies aimed at increasing both frequency and spending. By prioritizing retention and upselling for high-CLV clusters while applying nurturing strategies to low-CLV ones, businesses can strengthen profitability and build sustainable growth over time.

6. CONCLUSION AND MARKETING RECOMMENDATION

The K-Means clustering analysis revealed four distinct customer groups, each displaying unique behaviors, purchasing capacities, and levels of engagement, providing valuable insights for targeted marketing and efficient resource allocation.

Cluster 1 represents loyal, frequent shoppers who engage across multiple categories and consistently contribute to revenue. To strengthen their connection, businesses should focus on loyalty rewards, early access to new collections, and personalized appreciation offers that reinforce long-term retention.

Cluster 2 consists of cost-conscious buyers who shop less frequently but respond strongly to price incentives. Promotional discounts, bundle deals, and tailored seasonal campaigns can encourage greater purchase frequency while ensuring profitability.

Cluster 3 includes premium customers who make fewer transactions but spend significantly more per purchase, often favoring luxury items. This group should be targeted with VIP experiences, premium memberships, personalized upselling, and exclusive product launches to maximize their contribution.

Cluster 4 comprises occasional and low-engagement shoppers with minimal spending and weak brand association. Reactivation strategies, such as welcome-back incentives, engaging email campaigns, and curated product suggestions, can help convert them into more active participants.

From a Customer Lifetime Value (CLV) perspective, Clusters 1 and 3 hold the greatest long-term revenue potential and should be prioritized in retention-driven strategies. Meanwhile, Clusters 2 and 4 require nurturing approaches aimed at increasing engagement and lifetime value.

In conclusion, by leveraging these insights, businesses can adopt differentiated marketing approaches—retention programs for high-value customers and activation campaigns for low-engagement groups—ultimately enhancing customer satisfaction, boosting sales, and achieving sustainable revenue growth.

7. REFERENCE

[1].Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering —

A decade review. *Information Systems*, 53, 16–38.

[2].Arora, A., & Aggarwal, A. (2020). Customer segmentation using RFM and K-means clustering for e-commerce. *International Journal of Advanced Science and Technology*, 29(4), 1234–1245.

[3].Baesens, B., Van den Poel, D., & Vlasselaer, V. (2009). *Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS*. Wiley.

[4].Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*.

Plenum Press.

[5].Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer. [6].Bozanta, A., Berry, S., Cevik, M., Bulut, B., Yigit, D., Gonen, F. F., & Basar, A.

(2022). Time series clustering for grouping products based on price and sales patterns.

arXiv preprint arXiv:2204.08334.

[7].Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

[8].Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1–38.

[9].Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of KDD* (pp. 226–231).

[10]. Fader, P. S., & Hardie, B. G. S. (2009). Probability models for customer-base analysis. *Journal of Interactive Marketing*, 23(1), 61–69.

[11]. Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing customers. *Journal of Marketing Research*, 41(1), 7–18.

[12]. Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*

(3rd ed.). Morgan Kaufmann.

[13]. Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100–108.

[14]. Hughes, A. M. (1994). *Strategic Database Marketing*. Probus Publishing. [15]. Hruschka, E. R., Ebecken, N. F. F., & Campello, R. J. G. B. (2010). A scalable

k-means algorithm for clustering relational data. *Information Sciences*, 180(22), 4403–4415.

[16]. Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666.

[17]. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3), 264–323.

[18]. Karamshuk, D., Noulas, A., Scellato, S., & Mascolo, C. (2013). Geo-spotting: Mining online location-based services for optimal retail store placement. In *Proceedings of the 22nd International Conference on World Wide Web* (WWW '13 Companion).

[19]. Kim, H., & Park, J. (2014). A study on customer segmentation using clustering for online marketplace. *Journal of Retailing and Consumer Services*, 21(6), 846–852.

[20] K-Means original reference: MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281– 297). University of California Press.

[21]. Liao, S. H. (2005). Clustering of time series data — A survey. *Pattern Recognition*, 38(11), 1857–1874.

[22]. Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137.

[23]. Lu, J., & Hose, K. (2017). Customer segmentation in retail: A survey and directions for future research. *International Journal of Retail & Distribution Management*, 45(3), 1–27.

[24]. Ma, Y., & Deng, K. (2021). Enhanced RFM model for e-commerce customer segmentation. *Journal of Electronic Commerce Research*, 22(2), 85–101.

[25]. Martínez, V., & Ortega, F. (2020). Integrating demographic and behavioral data for customer segmentation in retail. *Journal of Retailing*, 96(3), 44–59.

[26]. McCarthy, J., & Talbott, K. (2018). RFM (Recency, Frequency, Monetary) modeling for segmentation and churn prediction. *Journal of Database Marketing & Customer Strategy Management*, 25(4), 243–257.

[27]. Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2), 159–179.

[28]. Mohammad, A. R., & Rahman, M. (2021). Comparative analysis of clustering techniques for customer segmentation in retail. *Journal of Big Data*, 8(1), 88.

[29]. Nanda, S., & Rani, S. (2019). Customer segmentation in retail using machine learning techniques. *International Journal of Computer Applications*, 178(35), 1–8.

[30]. Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), 2592–2602.

[31]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.