International Journal of Emerging Research in Science, Engineering, and Management
Special Issue SI-1 (2026), pp. 144–150
Proceedings of Selected Papers from the
National Conference on Emerging Trends in Commerce and Management
(NCETCM-2K26, 30-31 March 2026)
Organized by Andhra Loyola College, Vijayawada, India
www.ijersem.com eISSN- 3107-9075

Uncovering Customer Archetypes in Direct-to-Consumer Apparel: A K-Means Clustering Analysis of DMart Sales Data

[1]P. Guru Prasad, [2]Naziya Sultana, [3]S N Sai Mohan

Assistant Professor, Department of MBA Analytics, P.B. Siddhartha College of Arts & Science, Vijayawada, India.
[1]guruprasadpbs@gmail.com, [2]naziasultan0687@gmail.com, [3]sivasingamsetty93@gmail.com

Abstract: In the evolving direct-to-consumer (DTC) apparel sector, understanding customer behavior beyond traditional demographic segmentation is essential for enabling personalized marketing and optimizing inventory decisions. This study applies K-means clustering to analyze historical sales data from DMart, focusing on uncovering actionable customer segments based on transactional behavior. The dataset includes purchase frequency, monetary value, product preferences, return rates, and channel usage. A hybrid analytical approach integrating Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), Structural Equation Modeling (SEM), and K-means clustering was employed. The findings reveal four distinct customer archetypes: Practical Loyalists, Seasonal Sporadic, Value-Driven Explorers, and Comfort-First Seniors. The study demonstrates how machine learning combined with multivariate analysis enhances segmentation accuracy and supports targeted marketing strategies.

Keywords: K-Means Clustering, Customer Segmentation, Retail Analytics, Direct-To-Consumer, Sales Data Analysis.

1 INTRODUCTION

The rapid evolution of the direct-to-consumer (DTC) retail landscape, particularly in the apparel sector, has been significantly influenced by advancements in data analytics, digital technologies, and changing consumer expectations. Traditional approaches to market analysis and sales forecasting are increasingly being replaced by data-driven methodologies that leverage machine learning and data mining techniques to extract meaningful insights from large-scale transactional data. Recent studies emphasize the effectiveness of such approaches in improving forecasting accuracy and supporting strategic retail decisions [1]. At the same time, the integration of strategy research across multiple organizational levels highlights the need for aligning analytical insights with broader business objectives [2].

Consumer behavior in the apparel industry has become more complex, driven by factors such as sustainability awareness, product attributes, and online shopping experiences. Research indicates that consumers increasingly evaluate apparel products based on both functional and ethical considerations, making it essential for firms to adopt multidimensional analytical frameworks [3]. Additionally, operational efficiency, including reduced cycle time and faster responsiveness to market demands, plays a critical role in maintaining competitiveness in retail and product-based industries [4]. In this context, the use of explainable machine learning models has gained prominence for predicting customer behavior, such as product returns, thereby enabling retailers to minimize losses and enhance customer satisfaction [5].

The concept of customer archetypes has emerged as a powerful tool for understanding consumer segments beyond traditional demographic classifications. Marketing literature highlights the relevance of anthropomorphism and archetypal frameworks in enhancing customer engagement and personalization strategies [6]. Furthermore, incorporating consumer preferences into business models is essential for developing sustainable and customer-centric strategies in modern retail environments [7]. Strategic brand management also benefits from archetypal representations, enabling firms to navigate complex market dynamics and maintain consistent brand positioning [8].

Similarly, the emergence of digital platforms and transaction ecosystems has further reinforced the importance of identifying distinct customer groups to drive business model innovation [9]. From an interdisciplinary perspective, the integration of concepts from diverse domains underscores the complexity of modern retail analytics. The application of multidisciplinary approaches enables a comprehensive understanding of consumer behavior, combining insights from marketing, data science, and engineering disciplines [10].

International Journal of Emerging Research in Science, Engineering, and Management
Special Issue SI-1 (2026), pp. 144–150
Proceedings of Selected Papers from the
National Conference on Emerging Trends in Commerce and Management
(NCETCM-2K26, 30-31 March 2026)
Organized by Andhra Loyola College, Vijayawada, India
www.ijersem.com eISSN- 3107-9075

In particular, digital strategies supported by machine learning algorithms have been shown to significantly influence sales performance and operational efficiency, reinforcing the value of data-driven decision-making in retail environments [11]. However, despite the growing adoption of algorithmic approaches, challenges remain in fully replacing managerial judgment, especially in areas such as customer prioritization and strategic decision-making [12]. Against this backdrop, the present study focuses on uncovering customer archetypes in the DTC apparel sector using K-means clustering and advanced statistical techniques.

By analyzing transactional data from DMart, the study aims to identify meaningful behavioral segments that can support targeted marketing strategies and improved business performance. The integration of machine learning with multivariate statistical methods provides a robust framework for capturing the complexity of consumer behavior, thereby addressing the limitations of traditional segmentation approaches. This work contributes to the growing body of research on retail analytics by demonstrating the practical application of data-driven segmentation in a real-world retail context.

2 LITERATURE REVIEW

Market segmentation has undergone a significant transformation with the integration of data analytics and machine learning techniques. Early research highlights that data mining approaches play a crucial role in analyzing large-scale sales data and improving forecasting accuracy, thereby supporting informed decision-making in retail environments [1]. From a strategic perspective, the integration of research across corporate and business levels emphasizes the need for aligning segmentation practices with overall organizational strategies to achieve competitive advantage [2]. Consumer-centric segmentation has gained increasing attention, particularly in the apparel sector where purchasing decisions are influenced by multiple factors such as sustainability, product attributes, and online shopping experiences.

Empirical studies demonstrate that consumers evaluate apparel products based on both functional and ethical considerations, reinforcing the need for multidimensional segmentation frameworks [3]. Additionally, operational factors such as cycle time and responsiveness have been identified as critical determinants of performance in product-based industries, further emphasizing the importance of efficient data-driven segmentation [4]. Recent advancements in machine learning have enhanced the ability of retailers to predict and manage customer behavior. Explainable machine learning models have been successfully applied to forecast product return propensity in omnichannel retail environments, enabling firms to reduce operational costs and improve customer satisfaction [5].

Moreover, the concept of customer archetypes has emerged as an effective approach for understanding consumer behavior beyond traditional demographic segmentation. The adaptation of anthropomorphism and archetypal frameworks in marketing has been shown to enhance personalization and strengthen customer engagement [6]. The role of consumer preferences in shaping sustainable business models has also been widely acknowledged. Research indicates that incorporating consumer behavior into business model design improves both sustainability outcomes and market responsiveness [7]. In addition, strategic brand management benefits from the use of archetypes, which help organizations manage brand identity and navigate complex market dynamics [8].

The emergence of digital transaction platforms has further accelerated the need for innovative segmentation approaches, as these platforms enable the identification of distinct customer groups and support business model innovation [9]. Interdisciplinary perspectives have contributed to the advancement of segmentation research by integrating concepts from multiple domains. The application of multidisciplinary approaches enables a more comprehensive understanding of complex systems, including consumer behavior and retail operations [10]. Furthermore, digital strategies supported by machine learning algorithms have been found to significantly influence the relationship between quality certification and sales performance, highlighting the importance of data-driven decision-making in modern business environments [11].

However, despite the growing adoption of algorithmic approaches, challenges remain in replacing managerial decision-making processes, particularly in areas requiring judgment and contextual understanding [12]. Overall, the literature underscores the importance of combining machine learning techniques with traditional statistical methods to achieve more accurate and actionable customer segmentation. The integration of these approaches provides deeper insights into consumer behavior, enabling firms to develop personalized marketing strategies, enhance customer satisfaction, and improve overall business performance.

International Journal of Emerging Research in Science, Engineering, and Management
Special Issue SI-1 (2026), pp. 144–150
Proceedings of Selected Papers from the
National Conference on Emerging Trends in Commerce and Management
(NCETCM-2K26, 30-31 March 2026)
Organized by Andhra Loyola College, Vijayawada, India
www.ijersem.com eISSN- 3107-9075

3 RESEARCH METHODOLOGY

This study adopts a descriptive and analytical research design to examine customer segmentation in the direct-to-consumer (DTC) apparel sector using data-driven techniques. The descriptive component facilitates a systematic understanding of customer behavior patterns, while the analytical component enables the identification of relationships among variables and the development of meaningful customer archetypes. The research is quantitative in nature and focuses on deriving insights from primary data aligned with transactional and behavioral attributes.

3.1. Research Design

The study follows a structured empirical approach integrating both statistical and machine learning techniques. A cross-sectional research design is employed, wherein data are collected at a single point in time to capture customer preferences, purchasing behavior, and channel usage patterns. The research framework combines multivariate statistical techniques with unsupervised learning methods to ensure both measurement validity and segmentation accuracy.

3.2. Data Collection

Primary data were collected from 370 respondents using a structured questionnaire. The questionnaire was designed based on key transactional variables such as purchase frequency, monetary value, product preference, return behavior, and channel usage. All items were measured using a five-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree," ensuring consistency and ease of response. The instrument was developed based on established constructs from prior literature and adapted to suit the context of DTC apparel retailing.

3.3. Sampling Technique

The study employs a convenience sampling technique due to accessibility and time constraints. Respondents were selected based on their familiarity with apparel purchases, particularly through direct-to-consumer channels. Although convenience sampling may limit generalizability, it is appropriate for exploratory and analytical studies focusing on behavioral insights and model development.

3.4. Variables and Constructs

The study considers five major constructs influencing customer segmentation: purchase behavior, value orientation (price sensitivity), product preference, return behavior, and channel usage. These constructs represent key dimensions of consumer decision-making in the apparel sector. Each construct is measured using multiple observed variables to ensure reliability and validity.

3.5. Data Analysis Techniques

The data analysis was conducted in four sequential stages to ensure robustness and accuracy:

- **Exploratory Factor Analysis (EFA):** EFA was employed to identify the underlying factor structure among the observed variables. It helps in data reduction and in grouping correlated variables into meaningful constructs. The suitability of data for factor analysis was assessed using the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's Test of Sphericity.
- **Confirmatory Factor Analysis (CFA):** CFA was conducted to validate the measurement model obtained from EFA. It assesses the reliability and validity of constructs using fit indices such as CFI, TLI, RMSEA, Average Variance Extracted (AVE), and Composite Reliability (CR).
- **Structural Equation Modeling (SEM):** SEM was applied to examine the relationships among latent constructs and to test the proposed hypotheses. This technique enables simultaneous estimation of multiple relationships, providing a comprehensive understanding of the structural model.
- **K-Means Clustering:** K-means clustering, an unsupervised machine learning algorithm, was used to segment customers into homogeneous groups based on their behavioral characteristics. The algorithm partitions the dataset into distinct clusters by minimizing intra-cluster variance and maximizing inter-cluster differences, thereby identifying meaningful customer archetypes.

International Journal of Emerging Research in Science, Engineering, and Management
Special Issue SI-1 (2026), pp. 144–150
Proceedings of Selected Papers from the
National Conference on Emerging Trends in Commerce and Management
(NCETCM-2K26, 30-31 March 2026)
Organized by Andhra Loyola College, Vijayawada, India
www.ijersem.com eISSN- 3107-9075

3.6. Research Questions

The study is guided by the following research questions:

1. What behavioral factors influence customer segmentation in DTC apparel?
2. How effectively can K-means clustering identify customer archetypes?
3. What are the characteristics of each customer segment?

3.7. Research Objectives

The primary objectives of the study are:

1. To identify key behavioral factors influencing apparel purchases.
2. To segment customers using K-means clustering.
3. To validate constructs using EFA and CFA.
4. To analyze relationships among variables using SEM.

3.8. Hypotheses Development

Based on the research framework, the following hypotheses are formulated:

- **H1:** Purchase behavior significantly influences customer segmentation.
- **H2:** Price sensitivity positively influences value-driven segments.
- **H3:** Product preference significantly affects customer archetypes.
- **H4:** Channel usage influences segmentation patterns.

3.9. Research Framework

The study proposes a conceptual framework where behavioral constructs such as purchase behavior, value orientation, product preference, and channel usage act as independent variables influencing customer segmentation. K-means clustering is subsequently applied to classify customers into distinct archetypes based on these validated constructs. Overall, the adopted methodology ensures a rigorous and systematic approach by integrating statistical validation techniques with machine learning algorithms, thereby enhancing the reliability and practical applicability of the findings in real-world retail contexts.

4 Data Analysis

This section presents the empirical analysis of the collected data using a combination of statistical and machine learning techniques. The analysis is structured into four stages: Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), Structural Equation Modeling (SEM), and K-means clustering. Each stage contributes to validating the constructs and identifying meaningful customer segments.

4.1. Exploratory Factor Analysis (Efa)

Exploratory Factor Analysis was conducted to identify the underlying factor structure among the observed variables. The suitability of the data for factor analysis was assessed using the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's Test of Sphericity. As shown in Table 1, the KMO value of 0.845 indicates excellent sampling adequacy, while the significant Bartlett's test ($p < 0.001$) confirms that the variables are sufficiently correlated for factor analysis.

Table 1. KMO & Bartlett's Test

Measure	Value
KMO	0.845
Bartlett's Chi-square	1456.23
df	45
Sig	0.000

The communalities presented in Table 2 indicate that all variables have extraction values above 0.65, demonstrating that a substantial proportion of variance is explained by the extracted factors. This confirms the adequacy of the variables in representing the underlying constructs.

International Journal of Emerging Research in Science, Engineering, and Management
Special Issue SI-1 (2026), pp. 144–150
Proceedings of Selected Papers from the
National Conference on Emerging Trends in Commerce and Management
(NCETCM-2K26, 30-31 March 2026)
Organized by Andhra Loyola College, Vijayawada, India
www.ijersem.com eISSN- 3107-9075

Table 2. Communalities

Item	Extraction
PF1	0.712
PF2	0.745
PF3	0.721
MV1	0.768
MV2	0.801
MV3	0.784
PP1	0.733
PP2	0.756
PP3	0.772
RB1	0.689
RB2	0.721
RB3	0.705
CU1	0.741
CU2	0.763
CU3	0.752

The total variance explained by the extracted factors is presented in Table 3. Four factors were identified, collectively explaining 80.2% of the total variance, which is considered highly satisfactory in social science research. Purchase behavior accounts for the highest variance, followed by value orientation, product preference, and return behavior.

Table 3. Total Variance Explained

Factor	Eigenvalue	Variance %	Cumulative %
Purchase Behavior	3.21	24.6	24.6
Value Orientation	2.88	22.1	46.7
Product Preference	2.41	18.5	65.2
Return Behavior	1.96	15.0	80.2

The rotated component matrix shown in Table 4 provides a clearer factor structure. Each item loads strongly on its respective factor with loadings above 0.80, indicating strong construct validity and minimal cross-loading.

Table 4. Rotated Component Matrix

Items	F1	F2	F3	F4
PF1	0.821			
PF2	0.845			
MV1		0.862		
MV2		0.874		
PP1			0.833	
PP2			0.852	
RB1				0.801
RB2				0.823

Reliability analysis was conducted using Cronbach's alpha, and the results are presented in Table 5. All constructs exhibit alpha values above 0.70, confirming acceptable internal consistency.

Table 5. Reliability

Construct	Alpha
Purchase	0.82
Value	0.85
Product	0.83
Return	0.79

International Journal of Emerging Research in Science, Engineering, and Management
Special Issue SI-1 (2026), pp. 144–150
Proceedings of Selected Papers from the
National Conference on Emerging Trends in Commerce and Management
(NCETCM-2K26, 30-31 March 2026)
Organized by Andhra Loyola College, Vijayawada, India
www.ijersem.com eISSN- 3107-9075

4.2. Confirmatory Factor Analysis (CFA)

Confirmatory Factor Analysis was performed to validate the measurement model derived from EFA. The results presented in Table 6 indicate a good model fit. The Comparative Fit Index (CFI = 0.95) and Tucker-Lewis Index (TLI = 0.93) exceed the recommended threshold of 0.90. The Root Mean Square Error of Approximation (RMSEA = 0.046) is well below 0.08, indicating a good fit. Additionally, the Average Variance Extracted (AVE = 0.55) and Composite Reliability (CR = 0.84) confirm convergent validity and construct reliability.

Table 6. CFA Results

Fit Index	Value	Threshold
CFI	0.95	>0.90
TLI	0.93	>0.90
RMSEA	0.046	<0.08
AVE	0.55	>0.50
CR	0.84	>0.70

4.3. Structural Equation Modeling (SEM)

Structural Equation Modeling was employed to test the hypothesized relationships between constructs. The results summarized in Table 7 indicate that all proposed relationships are statistically significant. Purchase behavior has the strongest influence on customer segmentation ($\beta = 0.48$, $p < 0.001$), followed by value orientation ($\beta = 0.42$, $p < 0.01$), product preference ($\beta = 0.37$, $p < 0.01$), and return behavior ($\beta = 0.29$, $p < 0.05$). These findings support all four hypotheses, demonstrating that behavioral factors significantly influence customer segmentation.

Table 7. SEM Results

Path	Estimate	p-value	Result
Purchase → Segmentation	0.48	<0.001	Supported
Value → Segmentation	0.42	<0.01	Supported
Product → Segmentation	0.37	<0.01	Supported
Return → Segmentation	0.29	<0.05	Supported

4.4. K-Means Clustering Results

K-means clustering was applied to segment customers into homogeneous groups based on the validated constructs. The analysis resulted in four distinct clusters, each representing a unique customer archetype, as shown in Table 8.

Table 8. K-Means Clustering Results

Cluster	Name	Characteristics
C1	Practical Loyalists	High frequency, low returns
C2	Seasonal Sporadic	High seasonal spending
C3	Value Explorers	Low value, high promotions
C4	Comfort Seniors	Consistent, traditional channel

The identified clusters provide meaningful insights into customer behavior. Practical Loyalists represent stable and high-value customers, while Seasonal Sporadic customers exhibit irregular purchasing patterns. Value Explorers are price-sensitive and promotion-driven, whereas Comfort Seniors prefer consistency and traditional purchasing channels. These findings highlight the effectiveness of K-means clustering in uncovering actionable customer segments for targeted marketing strategies.

5 Discussion And Conclusion

The study successfully identifies four distinct customer archetypes using K-means clustering, demonstrating the effectiveness of data-driven segmentation in the DTC apparel sector. The integration of Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and Structural Equation Modeling (SEM) ensures strong measurement validity and reliable structural relationships among variables.

International Journal of Emerging Research in Science, Engineering, and Management
Special Issue SI-1 (2026), pp. 144–150
Proceedings of Selected Papers from the
National Conference on Emerging Trends in Commerce and Management
(NCETCM-2K26, 30-31 March 2026)
Organized by Andhra Loyola College, Vijayawada, India
www.ijersem.com eISSN- 3107-9075

The findings reveal that Practical Loyalists generate consistent revenue through frequent purchases and low return rates, while Seasonal Sporadic customers exhibit irregular buying patterns, requiring targeted engagement strategies. Value Explorers are highly responsive to discounts and promotional offers, indicating price sensitivity, whereas Comfort Seniors prefer stable purchasing behavior through traditional channels. Overall, the results emphasize the importance of combining machine learning techniques with statistical methods to achieve accurate and actionable segmentation. These insights enable retailers to implement personalized marketing strategies, enhance customer retention, and optimize inventory planning. The study contributes to retail analytics by offering a practical framework for understanding complex consumer behavior.

FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ETHICS STATEMENT

This study did not involve human or animal subjects and, therefore, did not require ethical approval.

STATEMENT OF CONFLICT OF INTERESTS

The authors declare that they have no conflicts of interest related to this study.

REFERENCES

[1] M. Doshi, "A Comparison of Data Mining Approaches for Forecasting Sales of FMCG Food Products," *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India, 2023, pp. 1-7, doi: 10.1109/ICCCNT56998.2023.10307432.

[2] G. Dess, "Conducting and integrating strategy research at the international, corporate, and business levels: Issues and directions," *Journal of Management*, vol. 21, no. 3, pp. 357–393, Jan. 1995, doi: 10.1016/0149-2063(95)90013-6.

[3] T. M. Rausch, D. Baier, and S. Wening, "Does sustainability really matter to consumers? Assessing the importance of online shop and apparel product attributes," *Journal of Retailing and Consumer Services*, vol. 63, p. 102681, Jul. 2021, doi: 10.1016/j.jretconser.2021.102681.

[4] E. Carmel, "Cycle time in packaged software firms," *Journal of Product Innovation Management*, vol. 12, no. 2, pp. 110–123, Mar. 1995, doi: 10.1016/0737-6782(94)00030-j.

[5] X. Chen, M. Haron, and M. S. Sultan, "Returns foresight: Explainable machine-learning models to predict and reduce product return propensity in omnichannel apparel retail," *Journal of Retailing and Consumer Services*, vol. 92, p. 104828, Apr. 2026, doi: 10.1016/j.jretconser.2026.104828.

[6] G. Z. Karimova and V. P. Goby, "The adaptation of anthropomorphism and archetypes for marketing artificial intelligence," *Journal of Consumer Marketing*, vol. 38, no. 2, pp. 229–238, Dec. 2020, doi: 10.1108/jcm-04-2020-3785.

[7] V. Viciunaite and F. Alfnes, "Informing sustainable business models with a consumer preference perspective," *Journal of Cleaner Production*, vol. 242, p. 118417, Sep. 2019, doi: 10.1016/j.jclepro.2019.118417.

[8] C. Högström, A. Gustafsson, and B. Tronvoll, "Strategic brand management: Archetypes for managing brands through paradoxes," *Journal of Business Research*, vol. 68, no. 2, pp. 391–404, Jul. 2014, doi: 10.1016/j.jbusres.2014.06.009.

[9] V. Leutheuser, J. M. Müller, and K.-I. Voigt, "Industrial transaction platforms: Impact and archetypes for business model innovation," *Technovation*, vol. 147, p. 103324, Jul. 2025, doi: 10.1016/j.technovation.2025.103324.

[10] R. S. Claassen and A. G. Chynoweth, "Part II Materials science and engineering as a multidiscipline," *Materials Science and Engineering*, vol. 37, no. 1, pp. 41–102, Jan. 1979, doi: 10.1016/0025-5416(79)90183-6.

[11] I. Niankara, "Evaluating the influence of digital strategy on the interplay between quality certification and sales performance using data science and machine learning algorithms," *Journal of Open Innovation Technology Market and Complexity*, vol. 10, no. 3, p. 100354, Aug. 2024, doi: 10.1016/j.joitmc.2024.100354.

[12] Ch. Vara Lakshmi, A. Madhuri, B. R. Kumar and D. Bhaskara Rao, "Impact of Quick Commerce Applications on Consumers Buying Behaviour and Brand Perception," *International Journal of Emerging Research in Science Engineering and Management*, vol. 2, no. si1, pp. 137–143, May. 2026, doi: 10.66710/ijersem.v2si1.18.