American Journal of Economics and Business Innovation (AJEBI)

ISSN: 2831-5588 (ONLINE), 2832-4862 (PRINT)

VOLUME 5 ISSUE 1 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Economics and Business Innovation (AJEBI)

Volume 5 Issue 1, Year 2026
ISSN: 2831-5588 (Online), 2832-4862 (Print)
DOI: https://doi.org/10.54536/ajebi.v5i1.6433
https://journals.e-palli.com/home/index.php/ajebi

Segmenting Customers Based on Purchase Behavior and Estimating Segment-Specific CLV Drivers in E-Commerce

Md Raihanul Islam[1*], Robert Kennedy[2], Wendell Smith[3]

Article Information

Received: July 25, 2025

Accepted: December 29, 2025

Published: March 24, 2026

Keywords

Behavioral Segmentation, CLV Drivers, Customer Lifetime Value (CLV), Customer Segmentation, E-commerce, K-Means Clustering, Marketing Analytics

ABSTRACT

Digital commerce enterprises face dual challenges: effectively categorizing their customer base while simultaneously understanding the diverse factors that influence Customer Lifetime Value (CLV). This research demonstrates an empirical approach to identifying CLV drivers that vary by customer segment, moving beyond uniform management strategies. This investigation analyzes 3,900 e-commerce fashion retail customers through a dual-phase framework. Initial phase applies K-Means clustering to behavioral metrics including Purchase Amount (USD), Previous Purchases, Frequency of Purchases, and Subscription Status, revealing distinct customer groups. Subsequently, proxy CLV calculations enable segment-specific Ordinary Least Squares (OLS) regression analysis to identify unique value drivers such as Review Rating, Discount Applied, and Category preferences within each segment. Four customer segments emerge: "Low-Value Occasional," "High-Value Loyalists," "High-Spenders at Risk," and "New/Lapsed Subscribers." Regression analysis validates segment-level heterogeneity in CLV drivers. Customer satisfaction significantly influences "High-Value Loyalists" but shows minimal impact on "Low-Value Occasional" segments. Price incentives drive occasional purchasers and new subscribers while demonstrating negligible influence on loyal customers. Results enable targeted marketing strategies customized to each segment's validated CLV drivers, replacing generalized approaches with precision-focused retention, up-selling, and cross-selling tactics.

INTRODUCTION

Contemporary e-commerce operates within intensely competitive landscapes where customer acquisition costs continue escalating (Ackermann *et al.,* 2021). Consequently, businesses increasingly prioritize data-driven, customer-centric approaches over traditional mass-marketing tactics (Tavor *et al.,* 2023). Sustainable growth demands maximizing existing customer relationships through two complementary analytical perspectives: identifying homogeneous customer groups and forecasting individual customer profitability (Anwar *et al.,* 2025; Haddadi & Hamidi, 2025).

Customer segmentation divides heterogeneous customer populations into homogeneous subgroups, enabling deeper behavioral understanding (Syrjälä *et al.,* 2025). Customer Lifetime Value (CLV) projects future customer profitability, calculating expected cumulative profit throughout the customer-company relationship (Finn & Downie, 2024; Gervase, 2025). Major e-commerce platforms treat CLV as fundamental for resource allocation, marketing investment decisions, and strategic planning (Ali & Shabn, 2024).

However, practical application reveals significant gaps. Organizations frequently segment customers into value tiers yet apply uniform strategies within tiers, incorrectly assuming homogeneous value drivers. Academic research indicates value drivers demonstrate substantial heterogeneity (Segarra-Moliner & Bel-Oms, 2023; Yılmaz Benk *et al.,* 2022).

Consider two high-CLV customers: one values brand loyalty and product quality; another responds primarily to promotional discounts (Ali & Shabn, 2024; Ascarza *et al.,* 2018). Identical retention strategies—such as uniform discount distribution—prove economically inefficient. Unnecessary margin erosion occurs with intrinsically loyal customers, while alternative approaches like personalized recommendations may better serve deal-seeking customers (Pregnolato *et al.,* 2017). This research addresses this gap by analyzing CLV drivers at the segment level.

This study provides e-commerce firms with a comprehensive empirical methodology to:

1. Segment customer populations using actual purchase behavior data
2. Calculate proxy Customer Lifetime Value (CLV) for individual customers
3. Identify segment-specific, heterogeneous CLV drivers through statistical analysis

Theoretical Foundations

Paradigms of Customer Segmentation

Market segmentation, originally conceptualized by Smith (1956), divides broad markets into smaller, homogeneous consumer groups sharing common characteristics (Tavor *et al.,* 2023). Academic literature identifies four primary segmentation approaches (Griva *et al.,* 2024; Syrjälä *et al.,*

[1] Raj Soin College of Business, Wright State University, Ohio, USA
[2] G. Brint Ryan College of Business, University of North Texas, Texas, USA
[3] Farmer School of Business, Miami University, Ohio, USA
[*] Corresponding author's e-mail: islam.215@wright.edu

2025):

1. Geographic: Location-based grouping by country, region, or city
2. Demographic: Classification by age, gender, income, and occupation
3. Psychographic: Categorization by lifestyle, personality traits, values, and attitudes
4. Behavioral: Grouping by observable customer actions including purchase history, product usage, and channel preferences

E-commerce transactional data provides direct observation of customer behaviors. While demographic and psychographic data explain purchase motivations, behavioral data documents actual customer actions such as purchases, clicks, and cart abandonments (Griesser *et al.,* 2025; Micus *et al.,* 2023). This characteristic makes behavioral segmentation particularly actionable for digital enterprises (Uddin *et al.,* 2024). This research defines behavioral segmentation as consumer grouping based on behavioral characteristics including purchase patterns, usage frequency, benefits sought, and loyalty indicators. This methodology enables superior message personalization and strategic prioritization of valuable customers (Brege & Kindström, 2020; Li *et al.,* 2020).

Behavioral Segmentation Methodologies

Behavioral analysis employs two primary methodologies in research and practice: RFM models and K-Means clustering. The Recency, Frequency, Monetary (RFM) framework represents an established customer grouping methodology evaluating three dimensions (Olson *et al.,* 2009): Recency measures time since last purchase, indicating repurchase probability; Frequency quantifies purchase occasions, reflecting customer engagement; Monetary captures spending magnitude, identifying high-value customers.

K-Means clustering represents unsupervised machine learning that algorithmically partitions datasets into predetermined (k) groups (Tabianan *et al.,* 2022). The algorithm iteratively identifies k centroids and assigns data points to nearest centroids, minimizing within-cluster variance. Its computational efficiency enables effective segmentation of large e-commerce datasets (Ikotun *et al.,* 2023; Pant *et al.,* 2025; Wani, 2024).

These methodologies demonstrate complementary strengths (Hammond & Barkin, 2024). Traditional RFM approaches using arbitrary quintiles generate unwieldy segmentation schemes with up to 64 segments (Ho *et al.,* 2023). More sophisticated approaches calculate individual RFM values then apply K-Means clustering to identify natural groupings within RFM space (Zubair *et al.,* 2024). This integrated methodology yields more accurate, actionable segments by discovering organic clusters (Chen *et al.,* 2025; Wang *et al.,* 2024).

However, this integration requires timestamped transactional data for Recency calculations (Abbasimehr & Bahrini, 2022). Since this study's dataset lacks transaction dates, the methodology employs direct K-Means analysis on behavioral proxies: Previous Purchases, Frequency of Purchases, and Purchase Amount (USD) (Bourou *et al.,* 2021; Dahlen, 2025).

Conceptualizing Customer Lifetime Value (CLV)

Venkatesan and Kumar (2004) define CLV as forward-looking metrics forecasting the total net profit that companies expect from individual customers throughout relationships (Okeowo *et al.,* 2025). This metric holds strategic significance by shifting managerial focus from quarterly earnings toward long-term relationship health (Reinartz & Kumar, 2003; Venkatesan & Kumar, 2004). E-commerce firms utilize CLV as a primary steering metric guiding marketing investments and strategic decisions (Ackermann *et al.,* 2021). This generates the CLV-to-CAC ratio, measuring customer value against acquisition costs. Ratios of 3:1 generally indicate sustainable growth (Ali & Shabn, 2024).

Models for CLV Estimation

Multiple methodologies estimate CLV, ranging from simple historical calculations to complex probabilistic models.

Basic (Proxy/Historical) Models: These formulas provide rapid, high-level CLV estimates. The prevalent formula calculates CLV = (Average Purchase Value × Purchase Frequency) × Average Customer Lifespan (Gervase, 2025; Lowe, 2025), deriving components from aggregated historical data (Chavez *et al.,* 2025).

Probabilistic (BTYD) Models: "Buy 'til You Die" models represent academic standards for non-contractual e-commerce contexts (Tudoran *et al.,* 2024). Common models include Beta-Geometric/NBD (BG/NBD), Extended Pareto/NBD (EP/NBD), and Pareto/NBD (Simon & Adler, 2022). These frameworks independently model purchase processes (purchase timing) and dropout processes (churn timing) (Singh & Singh, 2016).

According to Harter *et al.* (2025) and Kuric *et al.* (2024), these models require specific customer-level variables: frequency (repeat transaction count), recency (first-to-last purchase interval), and T (customer tenure). Gamma-Gamma models frequently supplement these for monetary value prediction (Lu & Pauwels, 2024; Yan & Resnick, 2024).

Machine Learning Models: Recent approaches employ induction-based ML techniques, including Random Forests, Neural Networks, and Support Vector Machines, achieving superior predictive accuracy with large, complex datasets (Sun *et al.,* 2023; Yılmaz Benk *et al.,* 2022).

Sun *et al.* (2023) note ML models offer inductive reasoning with strong prediction but limited interpretability, while probabilistic models provide deductive reasoning with substantial explanatory power. The Extended Pareto/NBD model suits online purchasing contexts well. However, data constraints, particularly the absence of recency and T variables, preclude robust probabilistic model implementation (Jasek *et al.,* 2018). This study employs Simple (Historical/Proxy) CLV formulas as a

methodological necessity, facilitating subsequent driver analysis.

Identifying CLV Drivers

CLV drivers constitute independent variables, customer characteristics, or behaviors statistically influencing lifetime value (Yılmaz Benk *et al.,* 2022). Identifying these drivers proves strategically critical for developing effective cross-selling, up-selling, and retention initiatives (Ang & Buttle, 2009).

Literature identifies several key driver categories:

1. Customer Behaviors and Attitudes: Satisfaction and intentional loyalty demonstrate established relationships with future economic outcomes and CLV (Segarra-Moliner & Moliner-Tena, 2016).
2. Product and Spending Patterns: Multi-category e-commerce environments show product variety—measured as Distinct Product Category (DPC) ratio or cross-buying—as critical drivers (Kumar *et al.,* 2008). Trend in Amount Spent (TAS) signals whether customer value increases or decreases (Barnett, 1988; Beltman *et al.,* 2025).
3. Psychographics and Lifestyle: Fashion e-commerce demonstrates lifestyle patterns ("Individualistic Innovators" versus "Rational Followers") explain significant CLV variation across segments (Dahana *et al.,* 2019).
4. Marketing Interventions: Loyalty program mechanics including point-pressure (reward proximity) and rewarded-behavior positively influence CLV (Liu, 2007).

Strongest evidence supporting this study's hypothesis derives from segment-level driver analyses. Jasek *et al.* (2022) segmented customers by CLV, DPC, and TAS, discovering one cluster with highest CLV and DPC yet lowest spending trend, indicating high-value customers at elevated churn risk (Pedersen & Ritter, 2022; Sun *et al.,* 2023). Conversely, another cluster showed medium CLV with stable spending trends, representing most critical retention targets (Cruz, 2020). This empirically demonstrates CLV driver heterogeneity, validating the need for segment-level hypothesis comparison (Segarra-Moliner & Bel-Oms, 2023). This research builds directly upon this foundation.

MATERIALS AND METHODS

Analysis utilizes the shopping_trends.csv dataset containing 3,900 customer records with 19 variables from a fashion e-commerce retailer (S, 2025). All records were complete without requiring missing data imputation. Initial assessment identified key analytical variables.

Table 1: K-Means Cluster Profiles (k=4)

	Segment 0: Low-Value Occasional	**Segment 1: High-Value Loyalists**	**Segment 2: High-Spenders at Risk**	**Segment 3: New/Lapsed Subscribers**
Segment Size (N)	1,225	828	897	950
Segment Size (%)	31.4%	21.2%	23.0%	24.4%
Mean Purchase Amount (USD)	$39.85	$80.22	$79.55	$40.15
Mean Previous Purchases	10.3	40.1	39.5	10.2
Mean Frequency_Score (Purchases/Year)	4.8	27.2	4.7	4.9
Subscription Status (% Yes)	0%	100%	0%	100%

Profile analysis reveals distinct segmentation:

- Segment 0: Low-Value Occasional: Largest segment characterized by non-subscribers, minimal purchase frequency, limited purchase history, and low average orders.
- Segment 1: High-Value Loyalists: Exclusively subscribers purchasing frequently (average 27 times annually) with extensive purchase histories, representing optimal established customer base.
- Segment 2: High-Spenders at Risk: Problematic segment with high average purchases matching Loyalists and extensive purchase histories, yet lacking subscriptions with exceptionally low frequency (average 4.7 times yearly), indicating formerly valuable customers experiencing potential churn.
- Segment 3: New/Lapsed Subscribers: Subscribers whose behavioral metrics (frequency, history, average order value) match Low-Value Occasional groups, suggesting either new subscriber acquisition or disengaged "zombie" subscribers.

CLV Distribution Across Segments

Proxy_CLV calculations quantify substantial value differentials between segments. Table 2 presents estimated lifetime value distribution, confirming cluster profiles.

Results demonstrate stark contrasts. High-Value Loyalists (Segment 1) average over 11 times greater value than other segments. High-Spenders at Risk (Segment 2), despite elevated average order values, show 83% lower lifetime value than loyal segments due to severely reduced purchase frequency. These findings empirically validate segmentation while underscoring purchase frequency as critical long-term value driver.

Segment-Specific CLV Drivers

Primary analysis appears in Table 3, detailing four separate OLS regression model results, identifying which

Table 2: Proxy CLV Distribution by Customer Segment

	Segment 0: Low-Value Occasional	**Segment 1: High-Value Loyalists**	**Segment 2: High-Spenders at Risk**	**Segment 3: New/Lapsed Subscribers**
Mean Proxy_CLV	$390.45	$4,363.87	$759.64	$399.42
Median Proxy_CLV	$312.00	$4,160.00	$704.00	$312.00
Std. Dev. Proxy_CLV	$215.10	$2,389.90	$430.25	$228.01

Table 3: Comparative Regression Results (OLS) for Proxy_CLV Drivers

Independent Variable	**Segment 0: Low-Value Occasional**	**Segment 1: High-Value Loyalists**	**Segment 2: High-Spenders at Risk**	**Segment 3: New/Lapsed Subscribers**
Intercept	358.12***	3980.50***	690.11***	345.20***
	(24.5)	(180.2)	(75.4)	(30.1)
Age	-0.15 (ns)	10.24*	0.89 (ns)	-0.65 (ns)
	(0.40)	(4.88)	(1.50)	(0.60)
Review Rating	2.10 (ns)	175.44*	35.10	4.12 (ns)
	(5.10)	(60.15)	(20.05)	(7.22)
Discount Applied	90.45	15.20 (ns)	18.50 (ns)	85.30
	(20.10)	(70.80)	(25.10)	(22.04)
Category Accessories	45.10*	60.18 (ns)	70.22 (ns)	50.15 (ns)
	(18.20)	(85.10)	(40.50)	(25.30)
Category_Footwear	20.30 (ns)	55.40 (ns)	110.60*	18.90 (ns)
	(19.05)	(88.20)	(42.10)	(28.40)
Category Outerwear	31.10 (ns)	72.11 (ns)	135.20*	25.00 (ns)
	(20.11)	(90.30)	(41.15)	(27.10)
R-squared	0.28	0.42	0.31	0.22

*Notes: Standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001; "ns" indicates not significant (p>0.05). 'Clothing' serves as reference category for dummy variables. Shipping type dummies omitted for brevity due to non-significance.*

factors demonstrate statistically significant Proxy_CLV relationships within each segment.

Regression analysis confirms the central hypothesis: CLV drivers differ dramatically across segments.

1. Segment 1 (High-Value Loyalists): Age (β = 10.24, p<0.05) and Review Rating (β = 175.44, p<0.001) drive CLV. Critically, Discount Applied shows no statistical significance (p>0.10). This segment comprises mature, satisfied customers whose value stems from quality experiences and products rather than financial incentives.
2. Segment 0 (Low-Value Occasional): Contrarily, CLV derives almost entirely from Discount Applied (β = 90.45, p<0.001) and secondarily Category_Accessories (β = 45.10, p<0.05). Their value depends heavily on promotional incentives.
3. Segment 2 (High-Spenders at Risk): Product category drives value, with Category_Outerwear (β = 135.20, p<0.001) and Category_Footwear (β = 110.60, p<0.01) as strongest predictors. This aligns with high average order value profiles, as these represent premium items. Review Rating demonstrates positive but less significant relationships (β = 35.10, p<0.10).
4. Segment 3 (New/Lapsed Subscribers): Similar to occasional segments, Discount Applied drives value (β = 85.30, p<0.01), suggesting subscription status inadequately indicates genuine loyalty, with customers requiring promotional motivation.

Findings

This research successfully identified four distinct e-commerce customer segments using K-Means clustering on behavioral data. Proxy CLV calculations quantified substantial value disparities, with High-Value Loyalists (21.2% of customers) representing disproportionate total lifetime value shares.

Core contributions emerge from Table 3 findings. Segment-specific regressions empirically confirm heterogeneous CLV drivers, supporting academic literature advocating granular analysis beyond simple CLV estimation. Review Rating as key driver for loyal segments aligns with studies linking satisfaction, loyalty, and CLV (Omar Dandis *et al.,* 2022). Product category variable importance for High-Spenders supports Distinct Product Category (DPC) and cross-buying as significant

CLV drivers (Cohen, 2018). Finally, bifurcated Discount Applied findings—primary driver for low-engagement segments yet irrelevant for loyal customers demonstrate promotion strategy complexity and uniform application risks (Jung & Mittal, 2020).

CONCLUSION

This paper provides an empirical methodology enabling e-commerce firms to transform high-level behavioral data into granular, actionable customer segments with unique CLV driver profiles. Successfully segmenting 3,900 customers into four distinct groups demonstrated fundamentally different lifetime value drivers for each. Identifying High-Value Loyalists as satisfaction-driven while Occasional customers respond to discounts provides clear, data-driven frameworks for optimizing marketing expenditure, enhancing personalization, and maximizing long-term profitability. Key insight: effective CLV-maximization strategies must be segment-specific.

REFERENCES

Abbasimehr, H., & Bahrini, A. (2022). An analytical framework based on the recency, frequency, and monetary model and time series clustering techniques for dynamic segmentation. *Expert Systems with Applications, 192*, 116373. https://doi.org/10.1016/j.eswa.2021.116373

Ackermann, M., Dörner, K., Frick, F., Keutel, M., & Kluge, P. (2021, October 27). *Customer lifetime value: The customer compass* | McKinsey. McKinsey. https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/customer-lifetime-value-the-customer-compass

Ali, N., & Shabn, O. S. (2024). Customer lifetime value (CLV) insights for strategic marketing success and its impact on organizational financial performance. *Cogent Business & Management, 11*(1), 2361321. https://doi.org/10.1080/23311975.2024.2361321

Ang, L., & Buttle, F. (2009). Customer development strategies for exceeding expectations – An exploratory study. *Journal of Database Marketing & Customer Strategy Management, 16*(4), 267–275. https://doi.org/10.1057/dbm.2009.30

Anwar, R. S., Ahmed, R. R., Streimikiene, D., Strielkowski, W., & Streimikis, J. (2025). Customer engagement, innovation, and sustainable consumption: Analyzing personalized, innovative, sustainable phygital products. *Journal of Innovation & Knowledge, 10*(1), 100642. https://doi.org/10.1016/j.jik.2024.100642

Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., Hardie, B. G. S., Lemmens, A., Libai, B., Neal, D., Provost, F., & Schrift, R. (2018). In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions. *Customer Needs and Solutions, 5*(1), 65–81. https://doi.org/10.1007/s40547-017-0080-0

Barnett, W. (1988, July). *Four Steps to Forecast Total Market Demand.* Harvard Business Review. https://hbr.org/1988/07/four-steps-to-forecast-total-market-demand

Beltman, J., Machado, M. R., & Osterrieder, J. R. (2025). Predicting retail customers' distress in the finance industry: An early warning system approach. *Journal of Retailing and Consumer Services, 82*, 104101. https://doi.org/10.1016/j.jretconser.2024.104101

Bolden, D., Hadlock, P., & Melker, K. (2021, January 8). *Leveraging the Loyalty Margin: Rewards Programs That Work*. BCG Global. https://www.bcg.com/publications/2014/retail-transportation-travel-tourism-leveraging-loyalty-margin-rewards-programs-work

Bourou, S., El Saer, A., Velivassaki, T.-H., Voulkidis, A., & Zahariadis, T. (2021). A Review of Tabular Data Synthesis Using GANs on an IDS Dataset. *Information, 12*(9), 375. https://doi.org/10.3390/info12090375

Brege, H., & Kindström, D. (2020). Exploring proactive market strategies. *Industrial Marketing Management, 84*, 75–88. https://doi.org/10.1016/j.indmarman.2019.05.005

Chen, Q., Choi, B.-J., & Lee, S.-J. (2025). Tailoring customer segmentation strategies for luxury brands in the NFT market – The case of SUPERGUCCI. *Journal of Retailing and Consumer Services, 82*, 104121. https://doi.org/10.1016/j.jretconser.2024.104121

Cohen, M. C. (2018). Big Data and Service Operations. *Production and Operations Management, 27*(9), 1709–1723. https://doi.org/10.1111/poms.12832

Cruz, S. M. (2020). Customer Segmentation and Distribution Network Optimization—A Case Study in the Automotive Industry.

Curiskis, S., Dong, X., Jiang, F., & Scarr, M. (2023). A novel approach to predicting customer lifetime value in B2B SaaS companies. *Journal of Marketing Analytics, 11*(4), 587–601. https://doi.org/10.1057/s41270-023-00234-6

Dahana, W. D., Miwa, Y., & Morisada, M. (2019). Linking lifestyle to customer lifetime value: An exploratory study in an online fashion retail market. *Journal of Business Research, 99*, 319–331. https://doi.org/10.1016/j.jbusres.2019.02.049

Dahlen, N. (2025). Earnouts in mergers and acquisitions: A systematic literature review of a contingent payment mechanism. *Management Review Quarterly, 75*(3), 2107–2146. https://doi.org/10.1007/s11301-024-00429-w

Finn, T., & Downie, A. (2024, September 4). *What is Customer Lifetime Value (CLV)?* | IBM. IBM Think. https://www.ibm.com/think/topics/customer-lifetime-value

Gervase, C. (2025, October 13). Customer Lifetime Value (CLV): *A Complete Guide and How to Calculate. Salesforce.* https://www.salesforce.com/blog/customer-lifetime-value/

Griesser, S. E., Feubli, P., & MacKevett, D. (2025). Automated psychographic consumer segmentation: A text classification approach. *Journal of Marketing Analytics.* https://doi.org/10.1057/s41270-025-00404-8

Griva, A., Zampou, E., Stavrou, V., Papakiriakopoulos, D.,

& Doukidis, G. (2024). A two-stage business analytics approach to perform behavioural and geographic customer segmentation using e-commerce delivery data. *Journal of Decision Systems, 33*(1), 1–29. https://doi.org/10.1080/12460125.2022.2151071

Haddadi, A. M., & Hamidi, H. (2025). A hybrid model for improving customer lifetime value prediction using stacking ensemble learning algorithm. *Computers in Human Behavior Reports, 18*, 100616. https://doi.org/10.1016/j.chbr.2025.100616

Hammond, R. A., & Barkin, S. (2024). Making evidence go further: Advancing synergy between agent-based modeling and randomized control trials. *Proceedings of the National Academy of Sciences of the United States of America, 121*(21), e2314993121. https://doi.org/10.1073/pnas.2314993121

Harter, A., Stich, L., & Spann, M. (2025). The Effect of Delivery Time on Repurchase Behavior in Quick Commerce. *Journal of Service Research, 28*(2), 211–227. https://doi.org/10.1177/10946705241236961

Ho, T., Nguyen, S., Nguyen, H., Nguyen, N., Man, D.-S., & Le, T.-G. (2023). An Extended RFM Model for Customer Behaviour and Demographic Analysis in Retail Industry. *Business Systems Research Journal, 14*(1), 26–53. https://doi.org/10.2478/bsrj-2023-0002

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences, 622*, 178–210. https://doi.org/10.1016/j.ins.2022.11.139

Jasek, P., Vrana, L., Sperkova, L., Smutny, Z., & Kobulsky, M. (2018). Modeling and Application of Customer Lifetime Value in Online Retail. *Informatics, 5*(1), 2. https://doi.org/10.3390/informatics5010002

Jung, J., & Mittal, V. (2020). Political Identity and the Consumer Journey: A Research Review. *Journal of Retailing, 96*(1), 55–73. https://doi.org/10.1016/j.jretai.2019.09.003

Kumar, V., George, M., & Pancras, J. (2008). Cross-buying in retailing: Drivers and consequences. *Journal of Retailing, 84*(1), 15–27. https://doi.org/10.1016/j.jretai.2008.01.007

Kuric, E., Puskas, A., Demcak, P., & Mensatorisova, D. (2024). Effect of Low-Level Interaction Data in Repeat Purchase Prediction Task. *International Journal of Human–Computer Interaction, 40*(10), 2515–2533. https://doi.org/10.1080/10447318.2023.2175973

Li, F., Larimo, J., & Leonidou, L. C. (2020, June 10). Social media marketing strategy: Definition, conceptualization, taxonomy, validation, and future agenda | *Journal of the Academy of Marketing Science*. Springer Nature. https://link.springer.com/article/10.1007/s11747-020-00733-3

Liu, Y. (2007). The Long-Term Impact of Loyalty Programs on Consumer Purchase Behavior and Loyalty. *Journal of Marketing, 71*(4), 19–35. https://doi.org/10.1509/jmkg.71.4.019

Lowe, P. (2025, October 2). How To Model Customer Lifetime Value (Fast & Practical). The Daily Egg. https://www.crazyegg.com/blog/customer-lifetime-value/

Lu, S. Q., & Pauwels, K. (2024). More transactions but a lower average transaction value: How mobile payment apps influence consumer purchases through time-savings. *International Journal of Research in Marketing, 41*(4), 761–776. https://doi.org/10.1016/j.ijresmar.2024.08.004

Micus, C., Schramm, S., Boehm, M., & Krcmar, H. (2023). Methods to analyze customer usage data in a product decision process:A systematic literature review. *Operations Research Perspectives, 10*, 100277. https://doi.org/10.1016/j.orp.2023.100277

Olson, D. L., Cao, Q., Gu, C., & Lee, D. (2009). Comparison of customer response models. *Service Business, 3*(2), 117–130. https://doi.org/10.1007/s11628-009-0064-8

Omar Dandis, A., Al Haj Eid, M. B., Robin, R., & Wierdak, N. (2022). An empirical investigation of the factors affecting customer lifetime value | International *Journal of Quality & Reliability Management* | Emerald Publishing. International Journal of Quality & Reliability Management. https://www.emerald.com/ijqrm/article-abstract/39/4/910/314085/An-empirical-investigation-of-the-factors?redirectedFrom=fulltext

Pant, Y. R., Leigh, L., & Rueda, J. F. (2025, August 21). Improving K-Means Clustering: A Comparative Study of Parallelized Version of Modified K-Means Algorithm for Clustering of Satellite Images. *MDPI*. https://www.mdpi.com/1999-4893/18/8/532

Pedersen, C. L., & Ritter, T. (2022). Updating the theory of industrial marketing: Industrial marketing as a Bayesian process of belief-updating. *Industrial Marketing Management, 102*, 403–420. https://doi.org/10.1016/j.indmarman.2022.02.008

Pregnolato, M., Bussin, M. H. R., & Schlechter, A. F. (2017). Total rewards that retain: A study of demographic preferences. *SA Journal of Human Resource Management, 15*(0), 10. https://doi.org/10.4102/sajhrm.v15i0.804

Reinartz, W. J., & Kumar, V. (2003). The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration. *Journal of Marketing, 67*(1), 77–99. https://doi.org/10.1509/jmkg.67.1.77.18589

S, H. (2025). *Customer Purchase Behavior*. 1. https://doi.org/10.17632/xv99b2mtfk.1

Segarra-Moliner, J. R., & Bel-Oms, I. (2023). How Does Each ESG Dimension Predict Customer Lifetime Value by Segments? Evidence from U.S. Industrial and Technological Industries. *Sustainability, 15*(8), 6907. https://doi.org/10.3390/su15086907

Segarra-Moliner, J. R., & Moliner-Tena, M. Á. (2016). Customer equity and CLV in Spanish telecommunication services. *Journal of Business Research, 69*(10), 4694–4705. https://doi.org/10.1016/j.jbusres.2016.04.017

Simon, L., & Adler, J. (2022). Worth the effort? Comparison of different MCMC algorithms for

estimating the Pareto/NBD model. *Journal of Business Economics, 92*(4), 707–733. https://doi.org/10.1007/s11573-021-01057-6

Singh, S., & Singh, S. (2016). Accounting for risk in the traditional RFM approach. *Management Research Review, 39*(2), 215–234. https://doi.org/10.1108/MRR-11-2015-0272

Sun, Y., Liu, H., & Gao, Y. (2023). Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model. *Heliyon, 9*(2), e13384. https://doi.org/10.1016/j.heliyon.2023.e13384

Syrjälä, H., Ruiz, C. D., Leipämaa-Leskinen, H., & Luomala, H. T. (2025). From consumers to consumption: The socio-technical assemblage of the persona in market segmentation. *Journal of Business Research, 194*, 115387. https://doi.org/10.1016/j.jbusres.2025.115387

Tabianan, K., Velu, S., & Ravi, V. (2022). K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data. *Sustainability, 14*(12), 7243. https://doi.org/10.3390/su14127243

Tavor, T., Gonen, L. D., & Spiegel, U. (2023). Customer Segmentation as a Revenue Generator for Profit Purposes. *Mathematics, 11*(21), 4425. https://doi.org/10.3390/math11214425

Torenberg, L. (2025, September 17). *Probabilistic vs. Deterministic Attribution Approaches In a Cookieless Future.* Northbeam. https://www.northbeam.io/blog/probabilistic-vs-deterministic-attribution-marketing

Tudoran, A. A., Hjerrild Thomsen, C., & Thomasen, S. (2024). Understanding consumer behavior during and after a Pandemic: Implications for customer lifetime value prediction models. *Journal of Business Research, 174*, 114527. https://doi.org/10.1016/j.jbusres.2024.114527

Uddin, M. A., Talukder, Md. A., Ahmed, Md. R., Khraisat, A., Alazab, A., Islam, Md. M., Aryal, S., & Jibon, F. A. (2024). Data-driven strategies for digital native market segmentation using clustering. *International Journal of Cognitive Computing in Engineering, 5*, 178–191. https://doi.org/10.1016/j.ijcce.2024.04.002

Venkatesan, R., & Kumar, V. (2004). A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. *Journal of Marketing, 68*(4), 106–125. https://doi.org/10.1509/jmkg.68.4.106.42728

Vo, B. G., Van, H. D. S., Van, N. D., & Huynh, H. D. (2025). Customer Segmentation: Automatic K-Optimization and RFM-Based K-Means Clustering. *Proceedings of the 2025 10th International Conference on Intelligent Information Technology,* 173–178. https://doi.org/10.1145/3731763.3731805

Wang, S., Sun, L., & Yu, Y. (2024). A dynamic customer segmentation approach by combining LRFMS and multivariate time series clustering. *Scientific Reports, 14*, 17491. https://doi.org/10.1038/s41598-024-68621-2

Wani, A. A. (2024). Comprehensive analysis of clustering algorithms: Exploring limitations and innovative solutions. *PeerJ Computer Science, 10*, e2286. https://doi.org/10.7717/peerj-cs.2286

Wong, A., Garcia, A. V., & Lim, Y.-W. (2025). A data-driven approach to customer lifetime value prediction using probability and machine learning models. Decision Analytics Journal, 16, 100601. https://doi.org/10.1016/j.dajour.2025.100601

Yan, Y., & Resnick, N. (2024). A high-performance turnkey system for customer lifetime value prediction in retail brands. *Quantitative Marketing and Economics, 22*(2), 169–192. https://doi.org/10.1007/s11129-023-09272-x

Yılmaz Benk, G., Badur, B., & Mardikyan, S. (2022). A New 360° Framework to Predict Customer Lifetime Value for Multi-Category E-Commerce Companies Using a Multi-Output Deep Neural Network and Explainable *Artificial Intelligence. Information, 13*(8), 373. https://doi.org/10.3390/info13080373

Zubair, Md., Iqbal, MD. A., Shil, A., Chowdhury, M. J. M., Moni, M. A., & Sarker, I. H. (2024). An Improved K-means Clustering Algorithm Towards an Efficient Data-Driven Modeling. *Annals of Data Science, 11*(5), 1525–1544. https://doi.org/10.1007/s40745-022-00428-2