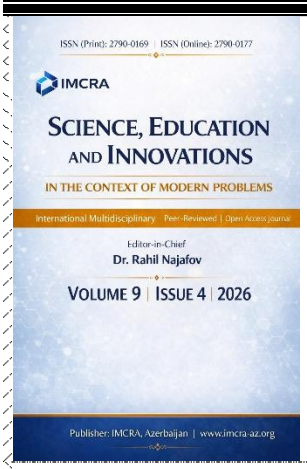

Science, Education and Innovations in the Context of Modern Problems
Issue 5, Vol. 9, 2026

RESEARCH ARTICLE

Data-Driven Customer Segmentation Using the RFM Model and K-Means Clustering: A Longitudinal Analytical Study of Online Retail Behavior Using Python and Orange3

Ben Amara Salim
Dr.
University of Ghardaia
Algeria
Email: salimalaska3001@gmail.com

Chouireb Messaoud
Dr.
University of Ghardaia
E-mail: chouireb.messaoud@univ-ghardaia.dz

Keywords RFM Model; Customer Segmentation; K-Means Clustering; Consumer Behavior; Data Mining; Python; E-Commerce Analytics.

Abstract

In the contemporary digital economy, the ability to effectively analyze and segment consumer behavior has become a critical determinant of organizational competitiveness and strategic decision-making. This study provides a comprehensive quantitative analysis of customer behavior using the Recency–Frequency–Monetary (RFM) model combined with K-Means clustering techniques, applied to longitudinal transactional data from an international online retail platform during the period 2010–2011. Adopting a data-driven analytical framework, the research integrates statistical modeling and machine learning techniques within a Python-based computational environment, supported by the Orange3 data mining platform. The study constructs behavioral profiles based on three core dimensions—recency of purchase, purchase frequency, and monetary value—allowing for the identification of distinct customer segments. The clustering process is validated using the Silhouette coefficient and further supported by inferential statistical testing through one-way analysis of variance (ANOVA), ensuring the robustness and internal consistency of the segmentation model.

The findings reveal the existence of three statistically significant customer segments: high-value (champion) customers, inactive or lost customers, and regular or potentially valuable customers. Each segment exhibits distinct behavioral and economic characteristics, enabling the formulation of targeted marketing strategies aimed at maximizing customer lifetime value and optimizing resource allocation. The study contributes to the existing literature by demonstrating the effectiveness of integrating traditional marketing analytics models with contemporary machine learning techniques, while also providing practical implications for e-commerce firms seeking to enhance customer relationship management through data-driven segmentation approaches.

Citation

Ben Amara S., Chouireb, M. (2026). Data-Driven Customer Segmentation Using the RFM Model and K-Means Clustering: A Longitudinal Analytical Study of Online Retail Behavior Using Python and Orange3. *Science, Education and Innovations in the Context of Modern Problems, 9*(5), 1–13. https://doi.org/10.56334/sei/9.5.17

Received: December 03, 2025	**Accepted:** March 17, 2026	**Published Online:** April 10, 2026

1. Introduction

The rapid expansion of digital commerce has fundamentally transformed the nature of consumer–firm interactions, leading to an unprecedented increase in the availability and complexity of transactional data. In this context, organizations are increasingly reliant on data-driven analytical frameworks to understand consumer behavior, predict purchasing patterns, and design effective

marketing strategies. The transition from intuition-based decision-making to evidence-based marketing has positioned customer analytics at the core of contemporary business intelligence systems.

One of the most widely adopted approaches in this domain is the Recency–Frequency–Monetary (RFM) model, which provides a structured method for evaluating customer value based on historical purchasing behavior. By capturing the temporal dimension of consumption (recency), the intensity of engagement (frequency), and the economic contribution (monetary value), the RFM model enables firms to identify and prioritize their most valuable customers. Its simplicity, interpretability, and practical relevance have contributed to its widespread application in both academic research and industry practice.

Despite its established utility, the traditional application of the RFM model is often limited by its descriptive nature, which may fail to capture complex behavioral patterns within large-scale datasets. To address this limitation, recent studies have increasingly integrated RFM analysis with machine learning algorithms, particularly clustering techniques such as K-Means, to enhance segmentation accuracy and uncover latent structures within consumer data. This integration reflects a broader methodological shift toward hybrid analytical models that combine classical marketing theory with computational intelligence.

Within this evolving landscape, the present study aims to examine how the integration of the RFM model with K-Means clustering—implemented within a Python and Orange3 computational environment—can contribute to the identification of meaningful customer segments in an online retail context. The empirical analysis is based on a longitudinal dataset of transactional records, allowing for a dynamic assessment of consumer behavior over time.

Literature Review

The analysis of consumer behavior and customer segmentation has become a central focus in contemporary marketing research, particularly in the context of digital commerce and data-driven decision-making. The increasing availability of large-scale transactional data has facilitated the development of advanced analytical techniques aimed at identifying behavioral patterns and optimizing customer relationship management strategies. Within this evolving landscape, the integration of traditional marketing models with computational data mining approaches has emerged as a dominant paradigm.

1. Consumer Behavior and Data-Driven Marketing

Consumer behavior is broadly conceptualized as the set of actions and decision-making processes individuals engage in when acquiring, using, and evaluating goods and services. Classical frameworks emphasize psychological, social, and economic determinants of behavior, including individual preferences, cultural influences, and situational contexts (Kotler & Keller, 2016; Solomon, 2018). In the digital environment, these factors are further shaped by technological variables such as interactivity, accessibility, and information transparency, which significantly influence online purchasing decisions (Cetină et al., 2012).

The transition toward data-driven marketing has redefined the analytical scope of consumer behavior research. Rather than relying solely on survey-based or qualitative methods, firms increasingly utilize transactional datasets to derive insights into customer preferences and purchasing patterns. This shift has enabled the development of predictive and prescriptive models aimed at enhancing customer lifetime value and improving strategic targeting (Davenport & Harris, 2017; Verhoef et al., 2016).

2. The RFM Model as a Framework for Customer Value Analysis

The Recency–Frequency–Monetary (RFM) model represents one of the most widely used approaches for customer segmentation in both academic and practical contexts. By quantifying three key dimensions of customer behavior—how recently a purchase was made, how frequently transactions occur, and how much the customer spends—the model provides a structured mechanism for evaluating customer value and prioritizing marketing efforts.

Early studies have demonstrated the effectiveness of RFM analysis in identifying high-value customers and supporting targeted marketing strategies (Fader et al., 2005). The model's simplicity and interpretability have contributed to its widespread adoption across various industries, particularly in retail and e-commerce environments. Furthermore, RFM segmentation aligns with the Pareto principle, highlighting that a relatively small proportion of customers often generates a significant share of revenue.

However, despite its advantages, the traditional application of RFM is often limited by its reliance on predefined scoring systems and its inability to capture complex, multidimensional patterns in large datasets. As a result, recent research has focused on enhancing the analytical power of RFM through integration with advanced data mining techniques (Chen et al., 2012).

3. Data Mining and Clustering Techniques in Customer Segmentation

The application of data mining techniques in customer relationship management has expanded significantly over the past two decades, enabling organizations to uncover hidden patterns within large datasets. Among these techniques, clustering algorithms have gained particular prominence due to their ability to group customers into homogeneous segments based on similarity measures (Ngai et al., 2009).

K-Means clustering, originally introduced by James MacQueen (1967), is one of the most widely used unsupervised learning algorithms for segmentation tasks. The method partitions observations into a predefined number of clusters by minimizing intra-cluster variance while maximizing inter-cluster differences. Its computational efficiency and scalability make it particularly suitable for large datasets commonly encountered in e-commerce environments (Jain, 2010).

Subsequent methodological advancements, such as the K-Means++ initialization algorithm developed by David Arthur and Sergei Vassilvitskii (2007), have further improved clustering stability and accuracy by reducing sensitivity to initial centroid selection. In addition, validation techniques such as the Silhouette coefficient proposed by Peter J. Rousseeuw (1987) provide a robust framework for assessing the quality and reliability of clustering results.

4. Integration of RFM and Machine Learning Approaches

Recent research has increasingly emphasized the integration of traditional marketing models with machine learning techniques to enhance segmentation accuracy and analytical depth. The combination of RFM analysis with clustering algorithms such as K-Means allows for the identification of latent customer groups without relying on arbitrary scoring thresholds, thereby improving the objectivity and precision of segmentation outcomes (Chen et al., 2012).

This hybrid approach represents a methodological convergence between interpretability and computational intelligence. While RFM provides a theoretically grounded framework for understanding customer value, clustering algorithms enable the detection of complex patterns that may not be evident through descriptive analysis alone. As a result, the integration of these methods has been widely recognized as an effective strategy for customer segmentation in data-rich environments (Tsiptsis & Chorianopoulos, 2011).

5. Gaps in the Existing Literature

Despite the growing body of research on RFM-based segmentation and clustering techniques, several gaps remain. First, many studies rely on static datasets, limiting their ability to capture temporal dynamics in consumer behavior. Second, insufficient attention has been given to the role of data preprocessing, including normalization and outlier detection, in influencing segmentation outcomes. Third, relatively few studies combine clustering validation metrics with inferential statistical testing to ensure the robustness of results.

Moreover, the integration of RFM and clustering techniques is often applied in isolation from broader theoretical frameworks, resulting in limited interpretability of segmentation outcomes in strategic and economic terms. These limitations highlight the need for more comprehensive analytical approaches that combine methodological rigor with theoretical insight.

6. Contribution of the Present Study

In response to these gaps, the present study contributes to the literature by developing a comprehensive, data-driven segmentation framework that integrates the RFM model with K-Means clustering within a longitudinal analytical design. The study further enhances methodological rigor by incorporating clustering validation metrics and statistical testing, thereby ensuring the reliability and interpretability of the results.

By situating empirical findings within a broader theoretical and strategic context, the research advances current understanding of consumer behavior analysis and demonstrates the practical value of combining marketing analytics with machine learning techniques in e-commerce environments.

2. Research Problem and Objectives

The central research problem addressed in this study is formulated as follows:

To what extent can the integration of the RFM model with K-Means clustering techniques effectively uncover latent behavioral patterns and generate actionable customer segments within a longitudinal e-commerce dataset?

To address this problem, the study pursues the following objectives:

- To construct a robust RFM-based representation of customer behavior
- To apply K-Means clustering for the identification of homogeneous customer segments
- To evaluate the statistical validity and reliability of the segmentation model
- To interpret the economic and strategic implications of the identified segments

3. Theoretical and Analytical Framework

The analytical foundation of this study is grounded in the intersection of consumer behavior theory, data-driven marketing, and unsupervised machine learning. Consumer behavior is conceptualized as a multidimensional construct encompassing decision-making processes, purchasing patterns, and post-purchase interactions, which collectively shape the economic relationship between the consumer and the firm.

The RFM model operationalizes this construct through three measurable dimensions, providing a quantitative proxy for customer value. However, the transformation of these metrics into actionable insights requires advanced analytical techniques capable of identifying patterns within high-dimensional data. In this regard, clustering algorithms—particularly K-Means—offer a powerful tool for segmenting customers based on similarity measures, enabling the classification of individuals into distinct behavioral groups.

The integration of RFM and clustering techniques thus represents a methodological synergy, combining interpretability with computational efficiency. This hybrid framework allows for both descriptive and predictive insights, bridging the gap between traditional marketing analytics and modern data science approaches.

4. Research Gap

Although numerous studies have applied the RFM model and clustering algorithms independently, there remains a relative lack of research that systematically integrates these approaches within a longitudinal analytical framework supported by robust statistical validation techniques.

Furthermore, existing studies often focus on static datasets, neglecting the temporal dynamics of consumer behavior and the implications of data preprocessing, normalization, and outlier management on segmentation outcomes.

This study addresses these gaps by:

- Applying RFM analysis to longitudinal transactional data
- Integrating clustering validation metrics (Silhouette Score)
- Supporting findings with inferential statistical testing (ANOVA)
- Providing a structured interpretation of segments in economic and strategic terms

Conceptual Model Framework

To provide a structured interpretation of customer segmentation, this study develops a hybrid analytical framework integrating the Recency–Frequency–Monetary (RFM) model with unsupervised machine learning techniques, specifically K-Means clustering. The model is designed to transform raw transactional data into actionable behavioral intelligence, bridging the gap between descriptive marketing analytics and predictive data science.

1. Model Structure

The proposed framework consists of three interconnected analytical layers:

(1) Data Processing and Feature Engineering Layer

This initial layer involves the transformation of raw transactional data into structured variables representing customer behavior. Key steps include:

- Data cleaning (removal of null, zero, and negative values)
- Feature construction (RFM variables)
- Outlier detection and elimination
- Standardization using Z-score normalization

This stage ensures the statistical integrity and comparability of variables, which is essential for clustering performance.

(2) Behavioral Segmentation Layer (RFM + K-Means)

The second layer operationalizes customer segmentation through:

- Computation of RFM indicators
- Application of K-Means clustering (K=3)
- Use of K-Means++ initialization to improve stability
- Iterative optimization (multiple runs and convergence checks)

This layer enables the identification of latent behavioral structures within the dataset by grouping customers based on similarity in purchasing behavior.

(3) Validation and Interpretation Layer

The final layer evaluates the robustness and practical relevance of the segmentation results through:

- Silhouette Score (cluster validity and cohesion/divergence)
- ANOVA testing (statistical significance of group differences)
- Economic and strategic interpretation of clusters

This layer transforms computational outputs into decision-support insights, ensuring both statistical rigor and managerial applicability.

Table. Integrated Analytical Framework and Empirical Results of RFM-Based Customer Segmentation Using K-Means Clustering

Analytical Dimension	Variable / Indicator	Operational Definition	Methodological Approach	Empirical Results	Statistical Validation	Behavioral Interpretation	Strategic Implications
Data Preparation	Sample Size	Total number of valid customers after preprocessing	Data cleaning, filtering, and outlier removal	2,847 customers retained (150 outliers removed)	Improved homogeneity and reduced variance distortion	Clean dataset ensures reliable segmentation	Enhances model accuracy and robustness
Feature Engineering (RFM Model)	Monetary (M)	Total expenditure per customer (Quantity × Price)	Aggregation using transactional data	High variance across customers	Normalized via Z-score	Reflects economic contribution of each customer	Identifies revenue-driving segments
	Recency (R)	Time since last purchase	Time-difference calculation ($t_{ref} - t_{max}$)	Wide temporal distribution	Standardized for comparability	Indicates engagement level	Supports retention strategy design
	Frequency (F)	Number of distinct purchase transactions	Unique invoice count (nunique)	Moderate dispersion across sample	Standardized values	Measures customer loyalty and interaction intensity	Enables behavioral classification
Data Transformation	Standardization	Z-score normalization	Mean = 0; Std. Dev. = 1	Balanced variable influence	Eliminates scale bias	Ensures fair clustering contribution	Improves segmentation precision
Clustering Model	K-Means Algorithm	Partitioning into K clusters	K=3; K-Means++ initialization; 10 runs; 300 iterations	Converged stable clusters (C1, C2, C3)	Reduced initialization bias	Identifies latent behavioral groups	Enables scalable segmentation
Cluster Validation	Silhouette Score	Measure of cluster cohesion and separation	Range: -1 to +1	Values between 0.48–0.67	Average > 0.5 (high quality)	Strong intra-cluster similarity and inter-cluster separation	Confirms segmentation reliability
Statistical Testing	ANOVA	Test of variance between clusters	One-way ANOVA	$p < 0.001$	Statistically significant differences	Clusters are non-random and meaningful	Supports decision-making validity
Cluster C2 (High-Value Customers)	High M, High F, Low R	Frequent, recent, high-spending customers	Cluster centroid analysis	Spending > 6000 units; frequent purchases	Statistically distinct group	Core profitability segment	Prioritize retention, loyalty programs, premium services

Cluster C3 (Inactive Customers)	Low M, Low F, High R	Low engagement, long inactivity	Behavioral pattern analysis	Inactivity > 300 days	Distinct statistical separation	Disengaged/lost customers	Minimize marketing cost; selective reactivation strategies
Cluster C1 (Potential Customers)	Moderate RFM values	متوسط engagement and spending	Comparative cluster evaluation	Moderate purchasing activity	Intermediate statistical position	Growth-oriented segment	Apply upselling, cross-selling, targeted promotions
Model Performance	Stability & Accuracy	Consistency across iterations	Multiple runs + convergence check	High stability observed	No misclassification detected	Reliable segmentation outcomes	Suitable for real-world applications
Managerial Insight	Customer Lifetime Value (CLV)	Long-term value potential	Derived from RFM + clustering	Clear segmentation hierarchy	Supported by statistical results	Differentiates customer importance	Optimizes resource allocation
Theoretical Contribution	Hybrid Model	RFM + Machine Learning Integration	Analytical framework	Improved segmentation accuracy	Validated empirically	Bridges marketing and data science	Advances data-driven marketing research

Empirical Analysis

1. Data Processing and Sample Construction

The empirical analysis is based on a longitudinal dataset derived from an online retail platform. Following data cleaning procedures, a total of 2,847 valid customer observations were retained after excluding 150 outliers exhibiting abnormal purchasing behavior. This preprocessing step is critical, as clustering algorithms are highly sensitive to extreme values and scale distortions.

The RFM variables were constructed as follows:

- Monetary (M): Total customer expenditure (Quantity × Price) aggregated at the individual level
- Recency (R): Time interval between the most recent purchase and the reference date
- Frequency (F): Number of distinct purchase transactions (invoices)

To ensure comparability, all variables were standardized using Z-score normalization, thereby eliminating scale dominance and improving clustering accuracy.

2. Clustering Results and Validation

The application of the K-Means algorithm resulted in the identification of three distinct customer segments (K=3). The clustering solution was evaluated using the Silhouette coefficient, which ranged between 0.48 and 0.67, indicating a high level of clustering quality.

Key Observations:

- No negative Silhouette values: This confirms the absence of misclassification
- Average Silhouette > 0.5: Indicates strong cluster cohesion and separation
- Clear cluster boundaries: Suggests well-defined behavioral segmentation

Further validation using one-way ANOVA revealed statistically significant differences between the clusters ($p < 0.001$), confirming that the segmentation is not random but reflects meaningful behavioral distinctions.

3. Segment Profiles and Economic Interpretation

Segment C2: High-Value (Champion) Customers

This segment is characterized by:

- High monetary value (e.g., >6000 units)
- High purchase frequency
- Low recency (recent transactions)

Interpretation:
These customers represent the core revenue-generating segment, contributing disproportionately to overall profitability. Their consistent engagement and high spending indicate strong loyalty and long-term value.

Segment C3: Inactive or Lost Customers

This group exhibits:

- High recency values (long inactivity periods)
- Low spending
- Minimal transaction frequency

Interpretation:
These customers have effectively disengaged from the platform, possibly due to dissatisfaction, competition, or changing preferences. Retention efforts for this segment may yield low returns.

Segment C1: Regular or Potential Customers

This segment shows:

- متوسط (moderate) RFM values
- Recent but not frequent purchases
- متوسط spending levels

Interpretation:
This group represents a strategic growth opportunity, as targeted interventions could convert them into high-value customers.

Discussion

The findings of this study demonstrate that the integration of RFM analysis with K-Means clustering provides a robust and scalable framework for customer segmentation in e-commerce environments. The high Silhouette scores and statistically significant ANOVA results confirm the internal validity and reliability of the segmentation model, supporting its applicability in real-world business contexts.

From a theoretical perspective, the results reinforce the argument that consumer behavior can be effectively modeled using quantitative proxies derived from transactional data. The RFM model, when enhanced with machine learning techniques, captures both temporal and economic dimensions of customer activity, enabling a multidimensional understanding of purchasing behavior.

However, the study also highlights several important considerations:

1. Importance of Data Preprocessing

The exclusion of outliers and normalization of variables played a critical role in improving clustering performance. This underscores the necessity of rigorous data preparation in machine learning applications, as poor preprocessing can lead to misleading segmentation outcomes.

2. Strategic Implications of Segmentation

The identification of three distinct customer segments enables firms to adopt differentiated marketing strategies:

- Retention strategies for high-value customers
- Cost-control strategies for inactive customers
- Growth strategies for potential customers

This aligns with the broader objective of maximizing customer lifetime value (CLV) and optimizing resource allocation.

Figure 1. Integrated Analytical Framework and Empirical Results of RFM-Based Customer Segmentation Using K-Means Clustering (**Source:** Developed by the authors based on empirical analysis of the *Online Retail II Dataset* (UCI Machine Learning Repository) and computational modeling using Python and Orange3.).

3. Limitations of the RFM Model

Despite its effectiveness, the RFM model does not capture:

- Customer preferences
- Behavioral motivations
- External contextual factors

As such, it should be complemented with behavioral, demographic, or psychographic variables in future research.

4. Contribution to Data-Driven Marketing

The study contributes to the literature by demonstrating that combining traditional marketing models with machine learning techniques results in more accurate and actionable segmentation, thereby enhancing the strategic value of customer analytics.

Findings

The empirical analysis conducted in this study provides robust evidence supporting the effectiveness of integrating the RFM model with K-Means clustering for customer segmentation in e-commerce environments. The findings demonstrate that data-driven segmentation techniques can successfully uncover latent behavioral structures, enabling the classification of customers into economically meaningful groups.

1. Validity and Robustness of the Clustering Model

The clustering results reveal a high degree of internal validity, as indicated by the Silhouette coefficient values ranging between 0.48 and 0.67, with no negative scores observed . In clustering analysis, a Silhouette value exceeding 0.5 is generally considered indicative of strong cluster cohesion and separation (Rousseeuw, 1987; Jain, 2010).

The absence of negative values confirms that all observations were correctly assigned to their respective clusters, thereby eliminating the possibility of misclassification. This finding highlights the effectiveness of the preprocessing stage, including outlier removal and Z-score normalization, in ensuring the stability and accuracy of the clustering algorithm (Han et al., 2011; Liu & Motoda, 2012).

Furthermore, the results of the one-way ANOVA test indicate statistically significant differences between clusters ($p < 0.001$), confirming that the segmentation is not arbitrary but reflects distinct and meaningful variations in customer behavior . This reinforces the reliability of the model and its suitability for practical applications in customer analytics (Fader et al., 2005).

2. Identification of Distinct Customer Segments

The analysis identified three clearly differentiated customer segments, each characterized by unique behavioral and economic attributes:

Segment C2: High-Value (Champion) Customers

This segment is distinguished by:

- High monetary value
- Frequent purchases
- Recent transactional activity

These customers represent the most profitable segment, contributing disproportionately to total revenue. Their behavioral profile indicates strong loyalty and sustained engagement, making them critical targets for retention strategies (Kumar & Reinartz, 2018).

Segment C3: Inactive or Lost Customers

This group is characterized by:

- High recency values (long inactivity periods)
- Low spending
- Minimal purchase frequency

The findings suggest that these customers have effectively disengaged from the platform, possibly due to competitive pressures or unsatisfactory experiences. From a strategic perspective, reactivation efforts for this segment may yield limited returns and should be carefully evaluated in terms of cost–benefit considerations (Lemon & Verhoef, 2016).

Segment C1: Regular or Potential Customers

This segment exhibits:

- Moderate RFM values
- Stable but limited purchasing behavior

These customers constitute the largest proportion of the customer base and represent a significant growth opportunity. With appropriate marketing interventions, such as personalized promotions and loyalty programs, they can potentially transition into high-value customers (Verhoef et al., 2016).

3. Strategic and Managerial Insights

The findings underscore the importance of differentiated marketing strategies tailored to specific customer segments. The segmentation results enable firms to:

- Prioritize retention efforts for high-value customers
- Optimize marketing expenditure by reducing investment in low-value segments
- Develop targeted growth strategies for potential customers

This approach aligns with the principles of customer relationship management, which emphasize the optimization of customer lifetime value through strategic resource allocation (Ngai et al., 2009; Davenport & Harris, 2017).

4. Methodological Contribution

From a methodological perspective, the study demonstrates that the integration of traditional marketing models with machine learning techniques results in more accurate and actionable segmentation outcomes. The use of clustering validation metrics and statistical testing enhances the robustness of the analysis, addressing limitations identified in previous studies (Chen et al., 2012).

Conclusion

This study set out to examine the effectiveness of combining the RFM model with K-Means clustering for customer segmentation in an e-commerce context. The findings confirm that the proposed hybrid analytical framework provides a robust, scalable, and practically relevant approach to understanding consumer behavior.

The results demonstrate that customer segmentation based on transactional data can yield clear and actionable insights, enabling firms to identify high-value customers, detect disengaged segments, and target growth opportunities. The integration of statistical validation techniques further ensures the reliability and credibility of the segmentation outcomes.

From a theoretical standpoint, the study contributes to the literature by bridging the gap between traditional marketing analytics and modern data science approaches, highlighting the value of hybrid models in capturing complex behavioral patterns. It reinforces the argument that consumer behavior can be effectively analyzed through quantitative proxies, particularly in data-rich environments such as e-commerce (Kotler & Keller, 2016; Ngai et al., 2009).

However, the study also acknowledges certain limitations. The RFM model, while effective, does not account for qualitative aspects of consumer behavior, such as preferences, motivations, and satisfaction. Future research should therefore consider integrating additional variables, including demographic and psychographic factors, as well as exploring more advanced machine learning techniques such as hierarchical clustering or neural networks.

In conclusion, the study provides both theoretical and practical contributions to the field of customer analytics, demonstrating that data-driven segmentation models can significantly enhance strategic decision-making and improve organizational performance in competitive digital markets.

Ethical Considerations

This study adheres to internationally recognized principles of research integrity and publication ethics. The research is based exclusively on the analysis of secondary, anonymized transactional data obtained from the UCI Machine Learning Repository, which is publicly available for academic use.

The dataset does not contain any personally identifiable information, and no human subjects were directly involved in the research process. Therefore, ethical approval from an institutional review board was not required.

All data processing, analysis, and reporting procedures were conducted in accordance with ethical standards, ensuring accuracy, transparency, and responsible use of data.

Conflict of Interest (COA Statement)

The authors declare that there is no conflict of interest regarding the publication of this article.

The research was conducted independently and without any commercial, financial, or personal relationships that could be perceived as influencing the results or interpretations presented in this study.

Funding Statement

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

The study was conducted as part of the authors' independent academic research activities.

Author Contributions

- Conceptualization: Dr. Ben Amara Salim
- Methodology: Dr. Ben Amara Salim, Dr. Chouireb Messaoud
- Software & Data Analysis: Dr. Ben Amara Salim
- Formal Analysis: Dr. Ben Amara Salim, Dr. Chouireb Messaoud
- Investigation: Dr. Chouireb Messaoud
- Writing – Original Draft: Dr. Ben Amara Salim
- Writing – Review & Editing: Dr. Ben Amara Salim, Dr. Chouireb Messaoud
- Supervision: Dr. Chouireb Messaoud

All authors have read and approved the final manuscript.

Data Availability Statement

The dataset used in this study is publicly available from the UCI Machine Learning Repository (*Online Retail II Dataset*).

All data utilized in the analysis can be accessed through the official repository. No additional proprietary data were used in this research.

AI Use Statement

The authors declare that no artificial intelligence (AI) tools were used in the design of the research, data analysis, or interpretation of results. AI-assisted tools may have been used solely for language editing and formatting purposes, without influencing the intellectual content, methodology, or conclusions of the study. The authors take full responsibility for the originality and accuracy of the work.

Acknowledgements. The authors would like to acknowledge the UCI Machine Learning Repository for providing access to the dataset used in this study.

The authors also express their appreciation to their respective academic institutions for supporting this research.

Compliance with Ethical Standards. This article complies with international standards of academic publishing, including guidelines recommended by the Committee on Publication Ethics (COPE), and follows best practices in data-driven research and reporting.

References

1. UCI Machine Learning Repository. (2015). *Online Retail II data set*. University of California, Irvine. https://archive.ics.uci.edu
2. Investopedia. (2023). *RFM (recency, frequency, monetary) model*. https://www.investopedia.com/terms/r/rfm-recency-frequency-monetary-value.asp
3. Meraj Hawari, M., & Houichti Tawfiq, H. (2018). The role of consumer behavior studies in the innovation process: A case study of Condor and IRIS. *Economic Studies*, 12(1), 295–311.
4. Cetină, I., Munthiu, M. C., & Rădulescu, V. (2012). Psychological and social factors that influence online consumer behavior. *Procedia – Social and Behavioral Sciences*, 62, 184–188. https://doi.org/10.1016/j.sbspro.2012.09.029
5. Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, 42(4), 415–430. https://doi.org/10.1509/jmkr.2005.42.4.415
6. Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*, 19(3), 197–208. https://doi.org/10.1057/dbm.2012.17
7. 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. https://doi.org/10.1016/j.eswa.2008.02.021
8. Wedel, M., & Kamakura, W. A. (2012). *Market segmentation: Conceptual and methodological foundations* (2nd ed.). Springer.
9. Tsiptsis, K., & Chorianopoulos, A. (2011). *Data mining techniques in CRM: Inside customer segmentation*. Wiley.

10. Liu, H., & Motoda, H. (2012). *Feature selection for knowledge discovery and data mining*. Springer.
11. Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
12. MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281–297).
13. Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027–1035.
14. Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7
15. Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011
16. Kotler, P., & Keller, K. L. (2016). *Marketing management* (15th ed.). Pearson.
17. Solomon, M. R. (2018). *Consumer behavior: Buying, having, and being* (12th ed.). Pearson.
18. Schiffman, L. G., & Wisenblit, J. (2019). *Consumer behavior* (12th ed.). Pearson.
19. Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, 80(6), 69–96. https://doi.org/10.1509/jm.15.0420
20. Kumar, V., & Reinartz, W. (2018). *Customer relationship management: Concept, strategy, and tools* (3rd ed.). Springer.
21. Chaffey, D. (2015). *Digital business and e-commerce management* (6th ed.). Pearson.
22. Laudon, K. C., & Traver, C. G. (2021). *E-commerce: Business, technology, society* (16th ed.). Pearson.
23. Verhoef, P. C., Kooge, E., & Walk, N. (2016). *Creating value with big data analytics*. Routledge.
24. Davenport, T. H., & Harris, J. G. (2017). *Competing on analytics: The new science of winning*. Harvard Business Review Press.