ADVANCED CUSTOMER SEGMENTATION IN E-COMMERCE: CLUSTERING TECHNIQUES AND THEIR IMPACT ON MARKETING STRATEGY OPTIMIZATION

Alice LANGNER
International Burch University of Sarajevo, 71210, Bosnia and Herzegovina
alice.langner@stu.ibu.edu.ba
Muamer BEZDROB
International Burch University of Sarajevo, 71210, Bosnia and Herzegovina
muamer.bezdrob@ibu.edu.ba

Abstract
This study compares traditional demographic segmentation and K-means clustering to optimize customer segmentation in e-commerce. Using the "Customer Personality Analysis" dataset from a UK-based retailer, it evaluates the effectiveness of these methods based on behavioral variables, including product expenditures, promotional engagement, and purchase channels. To test the hypotheses, K-means clusters were compared with demographic clusters. ANOVA assessed spending differences, while MANOVA examined whether K-means clustering provided distinct and actionable insights. Findings confirm that K-means clustering identifies behaviorally distinct customer groups, offering deeper insights and better marketing applications than traditional segmentation. However, practical challenges may limit its adoption. This research underscores the value of data-driven clustering techniques for precise and effective customer segmentation, improving business strategies. Future research should explore additional behavioral variables and validate these findings in real-world marketing applications.

Key words: *Customer Segmentation, Demographic Segmentation, E-commerce, K-means Clustering, Marketing Strategy*

JEL Classification: *M30*

I. INTRODUCTION

The e-commerce sector has grown rapidly over the past two decades, revolutionizing business-consumer interactions. Early efforts focused on building online marketplaces and secure transaction processes, with companies like Amazon and eBay leading the way. Over time, the focus shifted to improving user experience and personalizing customer interactions, with customer segmentation becoming vital for tailoring marketing strategies. While traditional segmentation based on demographics has been useful, recent research has highlighted the potential of big data analytics and machine learning techniques, such as *K*-means clustering, for more precise segmentation (Johnson, 1967; Wang & Wang, 2006; Kim & Lee, 2015).

This study examines the effectiveness of *K*-means clustering in comparison to traditional segmentation methods. Although clustering techniques like *K*-means have become popular for customer segmentation, many e-commerce businesses still rely on traditional methods. Traditional approaches may fail to capture complex customer behaviors, limiting marketing effectiveness. This research seeks to explore whether clustering algorithms, especially *K*-means, offer superior customer segmentation and enable businesses to create more targeted marketing strategies. By comparing these methods, the study aims to help businesses make informed decisions about segmentation to improve customer satisfaction and revenue optimization.

To address the challenges and opportunities associated with segmentation methods, this study poses three key research questions:

1. *What factors influence the choice between clustering algorithms and traditional segmentation methods in e-commerce marketing?*
2. *What challenges do e-commerce businesses face when implementing these methods?*
3. *To what extent do clustering algorithms provide more actionable insights for customer segmentation in e-commerce compared to traditional methods?*

This study compares *K*-means clustering with traditional segmentation methods, using data that includes demographics and customer purchasing behavior. The goal is to identify distinct customer segments and assess the effectiveness of each method for targeted marketing.

II. Literature Review

Market segmentation is a foundational concept in marketing research and practice, offering critical frameworks for understanding and predicting consumer behavior (Hunt & Arnett, 2004). Traditional segmentation methods, relying on predefined variables such as demographics, have long been the standard. While studies validate their effectiveness (Kotler et al., 2018; Smith, 1956), the rise of e-commerce has made understanding consumer behavior more complex and essential.

Despite their widespread use, traditional segmentation methods face criticism for oversimplifying consumer motivations and failing to capture nuanced behaviors (Johnson, 1967; Kim & Lee, 2015). Demographic-based approaches often struggle to reflect evolving consumer preferences. However, their dominance persists due to hesitance in adopting alternative frameworks. Emerging techniques, such as unsupervised machine learning, present innovative solutions by identifying hidden patterns and relationships in data, addressing gaps left by traditional methods. This literature review explores the evolution of market segmentation, its limitations, and the potential of machine learning to refine marketing strategies in the data-driven era.

Traditional segmentation relies on geographic, demographic, psychographic, and behavioral variables. Demographic segmentation, categorizing consumers by age, gender, income, education, and occupation, assumes shared needs and behaviors (Beane & Ennis, 1987). While easy to implement, it often lacks precision. To address this, businesses develop customer personas—fictional characters representing key audience characteristics—to enhance marketing strategies (Dimitriadis et al., 2019).

Traditional segmentation strategies have limitations in fully capturing consumer behavior. Geographic variables have limited predictive power (Haley, 1968; Schoenwald, 2001), while psychographic segmentation, based on social class and personality traits, lacks strong theoretical links to behavior (Lesser & Hughes, 1986; Yankelovich & Meer, 2006). Behavioral segmentation, considering factors like benefits sought and purchase occasions, also faces accuracy challenges (Schoenwald, 2001; Haley, 1968). Demographic segmentation is particularly constrained by cultural shifts, technological advancements, and evolving societal norms (Beane & Ennis, 1987; Lugmayr et al., 2017). Customers are considered important because they determine the survival of a company, emphasizing the need for businesses to adopt customer-oriented strategies in all marketing activities (Kusumah, 2018).

E-commerce has significantly transformed consumer behavior, emphasizing convenience and accessibility (Vipin et al., 2021). Online shopping, fast delivery, and personalized experiences have driven this shift, requiring businesses to adopt agile strategies that integrate advanced technologies and data analytics. Understanding consumer psychology and employing innovative marketing techniques are critical for adaptation.

Wang et al. (2015) emphasize the importance of data in e-commerce, offering insights into customer demographics, browsing behavior, and purchase history. Big data, leveraged by companies like Yahoo, Google, and Facebook, has shifted marketing from an art to a science. Yaqoob et al. (2016) highlight the four Vs of data management: volume, variety, velocity, and veracity.

Data mining plays a crucial role in analyzing big data, using techniques such as clustering, classification, and anomaly detection to extract insights. Lugmayr et al. (2017) underscore the role of clustering techniques in organizing data points and identifying trends, aiding marketers in decision-making. However, challenges remain in managing large datasets from multiple sources (Sharda et al., 2021; Cao, 2023). Data silos hinder information flow, necessitating research into strategies for data integration, consistency, and sharing.

Advanced analytical techniques, such as unsupervised machine learning, enhance understanding of consumer behavior, offering a comprehensive view of market dynamics. Traditional segmentation retains value, but new methods promise more refined consumer targeting. Dimitriadis et al. (2018) highlight clustering's significance in identifying patterns within large, unlabeled datasets, particularly in e-commerce. Clustering algorithms, such as partitioning, hierarchical, and density-based methods, enable businesses to tailor strategies and optimize product offerings.

Rajput & Singh and Moore (2001) have contributed to research on clustering algorithms, particularly *K*-means. This method divides data into *K* groups, minimizing within-cluster variance. The iterative process continues until an optimal arrangement is reached, though initial results may vary, requiring multiple runs for accuracy (Yse, 2019). Despite its effectiveness, *K*-means has limitations, including manual cluster determination, variability in outcomes, and difficulty handling diverse datasets (Moore, 2001).

Ultimately, leveraging big data and machine learning refines decision-making and strengthens business strategies. By addressing traditional segmentation's shortcomings and embracing new analytical techniques, businesses can achieve a deeper and more precise understanding of modern consumer behavior.

III. RESEARCH DESIGN

This study compares traditional demographic segmentation with *K*-means clustering using behavioral data to optimize e-commerce marketing. A quantitative approach analyzes online retail transactions, assessing segmentation effectiveness based on accuracy, interpretation, and applicability. Demographic segmentation groups consumers by static factors like age and income, often leading to generalized marketing. In contrast, *K*-means clustering identifies customer patterns based on purchasing behavior, offering a more dynamic and precise method. By evaluating both approaches, this research determines which provides deeper insights for targeted marketing strategies, ensuring a more effective understanding of consumer behavior. The first hypothesis examines whether *K*-means clustering produces more effective segmentation than traditional demographic methods by capturing deeper insights into customer behavior. If clustering proves superior, it suggests that behavioral patterns offer a stronger basis for customer classification than demographic traits.

H1: K-means clustering provides more effective customer segmentation than traditional demographic segmentation methods.

In today's data-driven market, businesses require more sophisticated segmentation techniques to remain competitive. Traditional methods, though widely used, may fail to account for evolving consumer preferences and behaviors. Clustering techniques, by identifying hidden patterns, offer the potential for more relevant and actionable insights.

The second hypothesis tests whether clustering techniques provide more valuable insights for marketing strategy development than traditional demographic segmentation. If proven, this would reinforce the need for businesses to adopt machine learning-based segmentation for improved customer targeting and engagement.

H2: Clustering techniques offer more actionable insights for marketing strategy development compared to traditional methods.

By comparing these approaches, this study contributes to the ongoing discussion on the effectiveness of segmentation techniques, highlighting the role of advanced data analytics in modern marketing.

IV. DATA AND METHODOLOGY

Chikkaswamygowda (2023) used Kaggle.com to collect data for a study on customer segmentation in e-commerce. The selected dataset, "Customer Personality Analysis," includes anonymized demographic and behavioral information from 2,241 online retail customers between December 1st, 2009, and December 9th, 2011. The data is classified as primary or secondary, with secondary data collected from sources like publications, personal records, and census data. Key variables relate to customer demographics, purchase behavior, and promotional engagement. These variables form the foundation for segmentation analysis, allowing for an exploration of how demographic and behavioral factors influence customer spending patterns and promotional responses.

For the traditional segmentation approach, key demographic variables such as *Age* and *Income* were recoded into three distinct groups. *Age* was segmented into three generational groups based on a generation report (Stillman & Stillman, 2017), while *Income* was divided into three Socio-Economic Groups (SEG) based on an individual's or household's social and economic position.

The children variable was recoded into three categories: "No children," "Families with a child," and "Larger families." Education was recoded into three groups: "Basic education," "Educated," and "High education." Marital status was recoded into three groups: "Single," "Partnered," and "Separated." This allowed for the creation of demographic segments based on age, income, children, education, and marital status.

ation, and marital status.

The original dataset was divided into two subsets: training and test datasets. The training dataset was the foundation for model development using the appropriate machine-learning algorithms. Conversely, the test dataset has been reserved for evaluating model performance and assessing its generalizability.

To test the first hypothesis, which suggests that clustering techniques provide more effective customer segmentation than traditional demographic methods, a rigorous analytical approach was adopted. This involved segmenting the dataset using both traditional demographic criteria and *K*-means clustering based on behavioral data.

The study used *K*-means clustering to identify clusters with high expenditures on products associated with each demographic group. The optimal *K* value was selected based on cohesion and interpretability. The ANOVA is used to compare traditional demographic segmentation and *K*-means clustering. The ANOVA test assessed the effectiveness of *K*-means clustering compared to traditional segmentation.

For the second hypothesis, the study used a data-driven approach to segmentation, using *K*-means clustering to target specific customer groups with distinct purchasing behaviors. Four distinct segmentation runs were performed, identifying customer groups such as *All-around spenders*, *Deal offers searchers*, *Online wine buyers*, and *Store meat buyers*. MANOVA was used to assess the effectiveness of *K*-means clustering in deriving actionable customer segments for marketing strategies. The primary objective was to examine whether *K*-means clustering could distinguish between different customer behaviors that could directly influence the design of targeted marketing campaigns. The analysis involved four separate segmentation runs, each focusing on different sets of behavioral variables. Two comparison groups were constructed within the test dataset. Group 1 contained customers whose demographic characteristics matched those identified by the *K*-means clustering process, and Group 2 listing all other customers. The results of this analysis contribute to understanding how advanced segmentation methods can outperform traditional demographic-based approaches.

V. RESULTS AND DISCUSSION

1) Traditional Demographic Segmentation Results

The study compared traditional demographic segmentation and *K*-means clustering to identify customer groups and their behaviors for marketing strategies. Based on extant literature (Stewart et al., 2021; Bruwer et al., 2014; Bauder, 2023; Public Health England, 2015; Govzman et al., 2020; Scheelbeek et al., 2020; Stillman & Stillman, 2017), traditional segmentation differentiated three main demographic groups: young adults with medium income, adults with medium income, and elders. Young adults spent the most on wine, meat, and sweet products, while adults with medium income spent more on wine, fruits, and fish. Elders prioritized fruits, fish, and gold products, reflecting a focus on health-conscious consumption and high-quality goods (Table 1).

Table 1. Mean spending on products per group

Product / Group	Mean spending (GBP)		
	Young adults	Adults	Elders
Meat	*54.08*	-	-
Sweets	*11.17*	-	-
Wine	*93.67*	-	-
Wine	-	*142.17*	-
Fruits	-	*8.52*	-
Fish	-	*12.72*	-
Fruits	-	-	*8.60*
Fish	-	-	*14.74*
Gold	-	-	*31.77*

2) K-Means Clustering Results

The *K*-means clustering technique was used to identify optimal clusters for each spending set. For the first spending set (wine, sweets, and meat), the cluster that emerged as the highest-spending group (*Cluster 1*) was predominantly represented by adults living together, not having children, graduates, and having a high income. The highest-spending cluster (*Cluster 2*) for the second spending set (wine, fruits, and fish) was predominantly represented by families living together, without children, and with high income. Finally, for the third spending set (fruits, fish, and gold), the highest-spending cluster (*Cluster 3*) was predominantly represented by elders living together, not having children, and graduates (Table 2).

Table 2. Mean spending on products per spending set and cluster

		Mean spending (GBP)		
Spending set	Products	Cluster 1	Cluster 2	Cluster 3
Set 1	Wine	*959*	-	-
	Sweets	*50*	-	-
	Meat	*422*	-	-
Set 2	Wine	-	*939*	-
	Fruits	-	*55*	-
	Fish	-	*75*	-
Set 3	Fruits	-	-	*99*
	Fish	-	-	*182*
	Gold	-	-	*148*

The ANOVA was applied to the test dataset to compare the total spending on selected spending sets between the customer groups obtained by two segmentation approaches. For all three spending sets, the ANOVA results showed a significant difference in spending between the traditional demographic segment (*Group 1*) and the *K*-means cluster (*Group 2*). Table 3 presents the obtained results for all three ANOVA testing procedures, which are graphically presented in Figure 1.

Table 3. ANOVA tests for group differences

Spending set	Source	Σ of sq.	df	Mean sq.	*F*	Sig.
Set 1	Model (Between groups)	7271603.306	1	7271603.306	99.946	0.001
	Error (Within groups)	1236843.641	17	72755.508		
Set 2	Model (Between groups)	8623711.268	1	8623711.268	128.785	0.001
	Error (Within groups)	5825693.114	87	66961.990		
Set 3	Model (Between groups)	134168.596	1	134168.596	13.771	0.001
	Error (Within groups)	370217.804	38	9742.574		

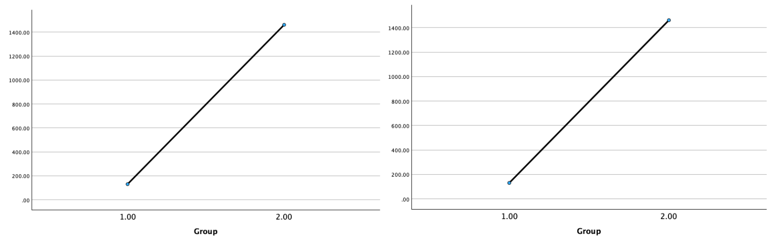

Figure 1 – Graphical display of the group differences

The obtained results from ANOVA tests for all three spending sets showed a statistically significant difference in spending between the traditional demographic segments (Group 1) and the *K*-means clusters (Group 2). These results support the first hypothesis, suggesting that clustering techniques outperform traditional demographic approaches in customer segmentation.

3) MANOVA Results

The MANOVA test was conducted to assess whether the means of selected dependent variables differed significantly between the two groups for each segmentation run. The two groups compared were: Group 1, representing customers corresponding to the demographic characteristics of the identified clusters, and Group 2, representing all other customers. Below are the results for each segmentation run.

The first run aimed to identify customers who exhibit the highest spending across all product categories. Six clusters were generated, with the fifth being "all-around spenders." To evaluate differences between all-around spenders (Group 1) and other customers (Group 2), a MANOVA test was conducted on the test dataset (16 cases for Group 1, 188 cases for Group 2). The multivariate test results showed significant differences for all dependent variables combined across the two customer groups, where all commonly used multivariate tests are statistically significant at $p < 0.001$ (e.g., $\Lambda = 0.82$, $F(7, 196) = 6.02$, $p < 0.001$). This indicates that Group 1 exhibits distinct spending patterns compared to Group 2. Univariate tests (Table 4) revealed statistically significant differences ($p < 0.05$) for all spending variables.

Table 4. Univariate Tests – All-around Spenders

Source	Variable	Σ of sq.	df	Mean sq.	*F*	Sig.
Model	Wines	1973816.711	1	1973816.711	20.163	0.001
	Fruits	16184.486	1	16184.486	13.606	0.001
	Meat	1616240.637	1	1616240.637	32.786	0.001
	Fish	53935.163	1	53935.163	24.643	0.001
	Sweet	14017.277	1	14017.277	11.037	0.001
	Gold	27795.747	1	27795.747	9.590	0.002
Error	Wines	19774862.936	202	97895.361		
	Fruits	240276.745	202	1189.489		
	Meat	9957918.319	202	49296.625		
	Fish	442107.347	202	2188.650		
	Sweet	256540.660	202	1270.003		
	Gold	585495.410	202	2898.492		

A comparison of the mean values for each dependent variable further supported the findings. For example, Group 1 exhibited higher mean spending on wines, fruits, and meat products compared to Group 2. This consistent pattern of higher spending among Group 1 customers demonstrates the effectiveness of the clustering-based segmentation in identifying high-value customer segments. The MANOVA results for all-around spenders strongly support the second hypothesis, showing that clustering techniques provide more practical and effective outcomes for marketing strategies than traditional demographic segmentation methods. Group 1, identified through clustering, exhibited behaviors of higher spending across most products. These findings suggest that clustering-based segmentation methods are not only statistically robust but also offer valuable insights for strategic marketing decisions.

The second run aimed to identify customers who frequently search for deal offers. To evaluate differences between Deal Offers Searchers (Group 1) and other customers (Group 2), a MANOVA test was conducted on the test dataset (25 cases for Group 1, 179 cases for Group 2). The multivariate test results did not reveal statistically significant differences between groups based on their responses to deal offers. All commonly used multivariate tests are statistically insignificant ($p > 0.05$).

While the MANOVA results suggest no significant differences between Deal Offers Searchers and other customers, this may indicate that deal-seeking behaviors are more uniformly distributed across customer segments. Further refinement of clustering variables or inclusion of additional behavioral metrics may yield more distinct results in future analyses. These findings do not support the second hypothesis (H2) that clustering-based segmentation provides more practical and effective outcomes for marketing strategies compared to traditional demographic segmentation methods. However, the insights gained emphasize the importance of testing variable combinations to uncover meaningful patterns in customer behavior.

The third analysis focused on identifying patterns among online wine buyers. Out of five identified clusters, Cluster 4 was selected for detailed analysis, consisting of 79 cases. Like for the first run, the multivariate test results showed significant differences for all dependent variables combined across the two customer groups. All commonly used multivariate tests are statistically significant at $p < 0.001$ (e.g., $\Lambda = 0.82$, $F(4, 199) = 10.64$, $p < 0.001$). This indicates that Group 1 exhibits distinct spending patterns compared to Group 2. Univariate tests (Table 5) revealed statistically significant differences ($p < 0.05$) for all spending variables.

Table 5. Univariate Tests – Online Wine Buyers

Source	Variable	Σ of sq.	df	Mean sq.	*F*	Sig.
Model	NumWebPurchases	63.282	1	63.282	9.409	0.002
	NumCatalogPurchases	135.158	1	135.158	12.962	0.000
	NumWebVisitsMonth	160.830	1	160.830	27.529	0.000
	MntWines	2635784.408	1	2635784.408	27.857	0.000
Error	NumWebPurchases	1358.639	202	6.726		
	NumCatalogPurchases	2106.352	202	10.427		
	NumWebVisitsMonth	1180.126	202	5.842		
	MntWines	19112895.239	202	94618.293		

For Online Wine Buyers, the MANOVA results indicated significant differences between the identified customer segments. The analysis demonstrated that spending behaviors varied notably across groups, with Cluster 4 emerging as a distinct segment. The multivariate test results were statistically significant ($p < 0.001$), confirming that the identified clusters exhibit meaningful differences in purchasing patterns. These findings support the second hypothesis, suggesting that clustering-based segmentation provides more practical and effective outcomes for marketing strategies compared to traditional demographic segmentation methods. The results highlight the value of refining clustering approaches and selecting relevant behavioral variables to enhance segmentation accuracy and uncover actionable customer insights.

The final segmentation run aimed to identify customers who exhibit the highest spending on meat products through store purchases. Six clusters were generated, among them, 87 cases in Cluster 5, labeled as "high meat spenders." Again, the multivariate test results showed significant differences for all dependent variables combined across the two customer groups. All commonly used multivariate tests are statistically significant at $p < 0.001$ (e.g., $\Lambda = 0.73$, $F(4, 199) = 17.88$, $p < 0.001$). This indicates that Group 1 exhibits distinct spending patterns compared to Group 2. Univariate tests (Table 6) revealed that statistically significant differences ($p < 0.05$) for all spending variables.

Table 6. Univariate Tests – In-store Meat Buyers

Source	Variable	Σ of sq.	df	Mean sq.	*F*	Sig.
Model	NumDealsPurchases	18.430	1	18.430	5.137	0.024
	NumStorePurchases	86.759	1	86.759	10.048	0.047
	MntMeatProducts	2322273.407	1	2322273.407	50.703	0.001
Error	NumDealsPurchases	724.727	202	3.588		
	NumStorePurchases	1744.079	202	8.634		
	MntMeatProducts	1334.793	202	45801.414		

The clustering-based segmentation identifies distinct spending behaviors with significant group differences and meaningful effect sizes. These findings align with the second hypothesis, showing that clustering techniques offer robust and actionable insights beyond traditional demographic segmentation methods, particularly for targeting high-value customer segments such as store meat buyers. The MANOVA results for store meat buyers strongly support the second hypothesis, showing that clustering techniques provide more practical and effective outcomes for marketing strategies than traditional demographic segmentation methods. Group 1, identified through clustering, exhibited behaviors of higher spending across most products. These findings suggest that clustering-based segmentation methods are not only statistically robust but also offer valuable insights for strategic marketing decisions.

4) Discussion

This study examined two hypotheses to compare traditional demographic segmentation with K-means clustering for customer segmentation and marketing strategies. The first hypothesis suggested that K-means clustering, based on behavioral data, would generate more detailed and predictive customer segments than traditional demographic methods. The second hypothesis proposed that the clusters derived from behavioral data would show statistically significant differences in spending patterns, supporting K-means clustering's effectiveness for targeted marketing.

Traditional demographic segmentation grouped customers by age, income, and other factors, producing broad categories like young adults with high income and elders with medium income. These segments showed distinct but general spending patterns across product categories. However, the granularity was insufficient to capture deeper behavioral insights.

To test the first hypothesis, K-means clustering was applied to behavioral variables, such as recency of purchases and promotional responses, with multiple cluster values tested. The results revealed more nuanced customer groups, capturing complex spending behaviors that traditional segmentation missed.

For the second hypothesis, total spending differences were compared across clusters using MANOVA. The analysis found significant spending disparities among the K-means clusters, while differences in traditional demographic segments were less pronounced. This confirmed that behavioral clustering was not only more detailed but also more effective for identifying high-spending groups.

The study findings support both hypotheses, demonstrating that K-means clustering provides a deeper understanding of customer behavior and more actionable insights for marketing strategies. It outperformed traditional segmentation by uncovering hidden patterns in spending, validating its practical value for targeted marketing.

5) Limitations

This study has several limitations, including reliance on a single UK-based retailer dataset, subjective selection of demographic groups, and the *K*-means clustering approach. It also omits psychographic and geospatial factors that could refine customer segmentation. Additionally, time and resource constraints restrict the scope of analysis, preventing further exploration of alternative clustering techniques or extended validation across multiple datasets.

6) Recommendations for Future Research

The study's limitations can be addressed by expanding upon its findings. Testing different *K* values for *K*-means clustering could provide valuable insights into the precision and relevance of segmentation. Including psychographic factors and geospatial data could refine customer profiles, providing deeper insights into consumer motivations and preferences. Integrating RFM analysis could strengthen the segmentation process by linking purchasing behavior to customer value. Expanding the dataset to diverse sources and industries could improve generalizability, allowing the model to be tested across different markets and customer bases. Using big data analytics, including real-time data processing, could enable more dynamic and personalized marketing strategies.

VI. Conclusion

The study explored the comparison between traditional demographic segmentation with *K*-means clustering for customer segmentation and marketing strategy development. The first hypothesis suggested that *K*-means clustering would produce more nuanced customer groups with greater predictive power for spending habits than traditional demographic segmentation. The second hypothesis suggested that these clusters would lead to statistically significant differences in total spending patterns among the groups, validating the practical application of *K*-means clustering for targeted marketing. The results showed that *K*-means clustering offered a deeper understanding of customer behavior compared to traditional demographic segmentation, and that it outperformed traditional segmentation in terms of identifying high-spending groups and offering actionable insights for marketing strategies.

The findings underscore the practical and theoretical value of data-driven customer segmentation approaches. This study demonstrated the efficacy of K-means clustering over traditional demographic segmentation in uncovering hidden patterns and refining customer segmentation. However, limitations such as dataset specificity and data quality must be acknowledged. Future research should address these by incorporating diverse data sources and advanced algorithms like deep learning. Additionally, hybrid approaches combining demographic and behavioral data, along with real-time analytics, could further enhance segmentation accuracy and scalability. Despite these challenges, clustering algorithms hold significant potential for improving personalized and effective e-commerce marketing strategies.

VII. REFERENCES

1. Bauder, S. (2023, May). 10.8% of Americans Own Gold, while 11.6% own silver, according to a new survey. Gold IRA Guide. https://goldiraguide.org/10-8-of-americans-own-gold-while-11-6-own-silver-according-to-a-new-survey/#comments
2. Beane, T. P., & Ennis, D. M. (1987). Market Segmentation: A Review. *European Journal of Marketing*, *21*(5), 20–42. https://doi.org/10.1108/EUM0000000004695
3. Bruwer, J., Jiranek, V., Halstead, L., & Saliba, A. (2014). Lower alcohol wines in the UK market: Some baseline consumer behaviour metrics. British Food Journal, 116(7), 1143–1161. https://doi.org/10.1108/BFJ-03-2013-0077
4. Cao, J. (2023). *E-Commerce Big Data Mining and Analytics*. Springer Nature Singapore. https://doi.org/10.1007/978-981-99-3588-8
5. Chikkaswamygowda, A. (2023). *New RFMT Model for Segmentation* [Dasaset; CSV]. https://www.kaggle.com/code/vionaclarissa/new-rfmt-model-for-segmentation-1st-in-kaggle/input
6. Dimitriadis, N., Dimitriadis, N. J., & Ney, J. (2018). Advanced marketing management: Principles, skills and tools (1 Edition). Kogan Page.
7. Govzman, S., Looby, S., Wang, X., Butler, F., Gibney, E. R., & Timon, C. M. (2021). A systematic review of the determinants of seafood consumption. *British Journal of Nutrition*, *126*(1), 66–80. https://doi.org/10.1017/S0007114520003773
8. Haley, R. I. (1968). Benefit Segmentation: A Decision-oriented Research Tool. *Journal of Marketing*, *32*(3), 30–35. https://doi.org/10.1177/002224296803200306
9. Hunt, S. D., & Arnett, D. B. (2004). Market Segmentation Strategy, Competitive Advantage, and Public Policy: Grounding Segmentation Strategy in Resource-Advantage Theory. *Australasian Marketing Journal*, *12*(1), 7–25. https://doi.org/10.1016/S1441-3582(04)70083-X
10. Johnson, S. C. (1967). Hierarchical Clustering Schemes. *Psychometrika*, *32*(3), 241–254. https://doi.org/10.1007/BF02289588
11. Kim, K. Y., & Lee, B. G. (2015). Marketing insights for mobile advertising and consumer segmentation in the cloud era: A Q–R hybrid methodology and practices. *Technological Forecasting and Social Change*, *91*, 78–92. https://doi.org/10.1016/j.techfore.2014.01.011
12. Kotler, P., Armstrong, G., & Opresnik, M. O. (2018). *Principles of marketing* (17th Global Ed.). Harlow: Pearson.
13. Kusumah, E. P. (2018). Customer Loyalty Model: Customer Satisfactionas Intervining Variable. EcoForum, 7(2). https://journalsfeaa.usv.ro/ojs/index.php/eco/article/view/2121/2064
14. Lugmayr, A., Stockleben, B., Scheib, C., & Mailaparampil, M. A. (2017). Cognitive big data: Survey and review on big data research and its implications. What is really "new" in big data? *Journal of Knowledge Management*, *21*(1), 197–212. https://doi.org/10.1108/JKM-07-2016-0307
15. Moore, A. (2001). *K-means and Hierarchical Clustering*. CMU School of Computer Science. https://www.cs.cmu.edu/~cga/ai-course/kmeans.pdf
16. Public Health England. (2015). Sugar reduction: the evidence for action. *Annexe 2: A Mixed Method Review of Behavior Changes Resulting From Experimental Studies That Examine the Effect of Fiscal Measure Targeted at High Sugar Food and Non-Alcoholic Drink.*
17. Scheelbeek, P. F. D., Moss, C., Kastner, T., Alae-Carew, C., Jarmul, S., Green, R., Taylor, A., Haines, A., & Dangour, A. D. (2020). United Kingdom's fruit and vegetable supply is increasingly dependent on imports from climate-vulnerable producing countries. *Nature Food*, *1*(11), 705–712. https://doi.org/10.1038/s43016-020-00179-4
18. Rajput, L., & Singh, S. N. (2023). Customer Segmentation of E-commerce data using K-means Clustering Algorithm. 2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 658–664. https://doi.org/10.1109/Confluence56041.2023.10048834
19. Schoenwald, M. (2001). Psychographic segmentation: Used or abused. Brandweek, 42(4), 34-38.
20. Sharda, R., Delen, D., & Turban, E. (2021). *Analytics, data science, & artificial intelligence: Systems for decision support* (Eleventh edition, global edition). Pearson.
21. Smith, W. R. (1956). Product Differentiation and Market Segmentation as Alternative Marketing Strategies. *Journal of Marketing*, *21*(1), 3–8. https://doi.org/10.1177/002224295602100102
22. Stewart, C., Piernas, C., Cook, B., & Jebb, S. A. (2021). Trends in UK meat consumption: Analysis of data from years 1–11 (2008–09 to 2018–19) of the National Diet and Nutrition Survey rolling programme. *The Lancet Planetary Health*, *5*(10), e699–e708. https://doi.org/10.1016/S2542-5196(21)00228-X
23. Stillman, D. & Stillman, J. (2017). *Gen Z Work: How the Next Generation Is Transforming the Workplace* (1st ed). HarperCollins Publishers.
24. Vipin, J., Bindoo, M., & Satyendra, A. (2021). An Overview of Electronic Commerce (e-Commerce). *Journal of Contemporary Issues in Business and Government*, *27*(3). https://doi.org/10.47750/cibg.2021.27.03.090
25. Wang, C., & Wang, Z. (2006). *The impact of Internet on service quality in the banking sector* [Master Thesis, Lulea University of Technology]. https://www.diva-portal.org/smash/get/diva2:1029746/FULLTEXT01.pdf
26. Wang, R. J.-H., Malthouse, E. C., & Krishnamurthi, L. (2015). On the Go: How Mobile Shopping Affects Customer Purchase Behavior. Journal of Retailing, 91(2), 217–234. https://doi.org/10.1016/j.jretai.2015.01.002
27. Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., & Vasilakos, A. V. (2016). Big data: From beginning to future. *International Journal of Information Management*, *36*(6), 1231–1247. https://doi.org/10.1016/j.ijinfomgt.2016.07.009
28. Yse, D. L. (2019, May). *Guide to K-Means Clustering Algorithm*. KDNuggets. https://www.kdnuggets.com/2019/05/guide-k-means-clustering-algorithm.html