International Journal of Information Engineering and Science

E-ISSN: 3048-1953
P-ISSN: 3048-1902

Research Article

Customer Data Management Analysis for Customer Segmentation Using K-Means Clustering Method

Andre Leto[1], Reza Aminullah[2*], Ani Dijah Rahajoe[3]

[1-3] Department of Computer Science, Faculty of Informatics Technology, Universitas Pembangunan Nasional Veteran Jawa Timur, Indonesia
* Corresponding Author: e-mail: aminullahreza65@gmail.com

Abstract: This study aims to examine customer segmentation through K-Means clustering from a customer data management perspective, emphasizing the interpretive value of analytical results rather than solely their computational outcomes. The research addresses a critical issue in contemporary data-driven organizations, where customer analytics is often reduced to technical modeling without sufficient translation into managerial insights. To respond to this gap, the study adopts a qualitative interpretive approach embedded within a quantitative clustering process, positioning clustering as part of a broader information management cycle. The empirical analysis is based on the Mall Customers Dataset obtained from Kaggle, consisting of 200 customer records with numerical attributes representing age, annual income, and spending score. Quantitative processing using K-Means clustering was employed to identify customer segments, while qualitative interpretation was applied to analyze the managerial meaning of each cluster. Data interpretation was supported by analytical documentation, visualization outputs, and reflective analysis of cluster characteristics. The findings reveal four distinct customer segments with different behavioral and economic profiles, each carrying specific strategic implications for customer relationship management and marketing decision-making. The study demonstrates that the primary value of clustering lies not merely in segment formation, but in its ability to transform raw customer data into actionable managerial knowledge. In conclusion, this research contributes to customer analytics literature by integrating data mining techniques with qualitative interpretation, offering a more human-centered and decision-oriented framework for customer data management. Future research is encouraged to extend this approach using organizational case studies or participatory decision-making contexts.

Keywords: Customer Data Management; Customer Segmentation; Data-Driven Decision Making; K-Means Clustering; Qualitative Interpretation

Received: June 13, 2025
Revised: August 8, 2025
Accepted: October 3, 2025
Published: November 28, 2025
Curr. Ver.: November 28,2025

1. Introduction

In today's digital era, the rapid advancement of information technology has caused exponential growth in the volume, velocity, and variety of customer data globally. This phenomenon fundamentally alters how organizations perceive data: no longer seen as passive records but as strategic assets capable of shaping competitive advantage. Across industries, organizations increasingly rely on customer data to understand behavior, inform marketing strategies, and improve organizational performance. Studies show that effective data utilization enhances personalization, increases customer loyalty, and strengthens decision-making capabilities, especially in markets with complex consumer interactions (Salminen et al., 2023). However, while data is abundant, the ability to manage and interpret it remains a central challenge for many businesses.

In the Indonesian business landscape, this challenge is particularly evident among small to medium enterprises (SMEs), where data accumulation does not necessarily translate into actionable insights. Many organizations collect transactional and demographic customer information but lack systematic frameworks to convert this raw data into strategic knowledge. This condition often results in segmentation practices that are superficial, focusing primarily on basic demographics rather than multidimensional customer behaviors. Such gaps limit the practical relevance of analytical outputs, which should inform customer relationship management and targeted marketing efforts. These practical constraints indicate that data complexity and poor analytical translation remain unresolved issues in real-world business settings.

Customer segmentation, as a method, seeks to partition markets into homogeneous groups to support tailored strategies that align with distinct customer needs. Clustering algorithms, particularly K-Means Clustering, have become popular in customer analytics due to their simplicity, scalability, and efficiency in handling large datasets (Salminen et al., 2023). Nonetheless, existing research on segmentation, while extensive in algorithmic evaluation, disproportionately emphasizes technical metrics such as cluster validity and performance indices. Such emphasis often obscures the interpretive processes necessary for transforming clusters into managerial meaning (Ghosh and Dubey, 2021). Without a nuanced understanding of segment characteristics, the segmentation risks becoming an end in itself rather than a means to strategic decision-making.

Moreover, literature on customer segmentation primarily reports quantitative findings, with limited attention to how segmentation results are interpreted by practitioners within management contexts. Although clustering outcomes are visually represented and statistically valid, they rarely address the question of how decision-makers make sense of these patterns in relation to organizational goals, cultural context, and customer value propositions (Singh et al, 2023). This theoretical gap reflects a broader limitation in the field of data analytics: the separation between quantitative model output and qualitative interpretation, which is essential for translating analytical insight into effective business action (Salminen et al., 2023).

Qualitative inquiry, therefore, plays a critical role in bridging this divide by capturing managerial perspectives, interpretive meaning, and contextualized understanding of clustering results. By exploring how organizational actors perceive and utilize segmentation profiles, researchers can better understand the socio-technical processes that transform data into strategic wisdom. This approach aligns with contemporary discussions in marketing analytics, which advocate for integrating human interpretation with data-driven models to enhance practical relevance and theoretical depth (Kavya & Sumathi, 2025).

Given these considerations, the present study examines customer segmentation through a qualitative lens that complements quantitative clustering outputs. Rather than solely presenting segmented clusters, this research investigates how managers and analysts interpret customer group characteristics and apply these insights to decision-making in customer management practices (Aljarah et al., 2022). This study contributes to the literature by highlighting the interpretive dimension of customer analytics, providing insights into the interplay between quantitative clustering results and managerial reasoning.

The findings of this research are expected to contribute to theoretical development by offering a more holistic view of customer segmentation as both an analytical and interpretive process. Practically, the study aims to inform organizations on how to integrate segmentation results into strategic planning, thereby strengthening customer relationship management and enhancing marketing effectiveness.

2. Literature Review

Customer data management constitutes a core component of information management that emphasizes systematic processes of data collection, storage, processing, and utilization to support business decision-making. In contemporary management literature, customer data is increasingly viewed as a strategic organizational resource rather than a mere administrative by-product. Organizations that are capable of integrating customer data across operational and analytical systems are better positioned to generate insights that enhance competitiveness and customer-centric strategies. McAfee & Brynjolfsson (2021) argue that the strategic value of data lies not in its volume, but in the organization's ability to transform data into actionable knowledge through effective management practices.

Customer segmentation represents one of the most prominent applications of customer data management in marketing and business analytics. Traditionally, segmentation relied heavily on demographic and geographic variables, which provided a static representation of customer groups. While such approaches offered simplicity, they were limited in capturing the dynamic and behavioral dimensions of customers. The advancement of data analytics and machine learning has shifted segmentation practices toward data-driven and behavior-based approaches, enabling organizations to uncover latent patterns in customer behavior and preferences(Wedel & Kannan, 2021). This transition reflects a broader movement toward evidence-based marketing strategies grounded in empirical customer data.

Among various analytical techniques, clustering methods have been widely adopted for customer segmentation. K-Means Clustering, a non-hierarchical clustering algorithm, aims to minimize intra-cluster variance while maximizing inter-cluster separation. Due to its computational efficiency, scalability, and ease of implementation, K-Means remains one of the most frequently applied methods in customer analytics, particularly when dealing with numerical variables such as income level and spending behavior. Empirical studies demonstrate that K-Means is effective in identifying meaningful customer segments that can inform targeted marketing strategies (Salminen et al., 2023).

Despite its widespread application, existing research on K-Means Clustering in customer segmentation has largely concentrated on technical aspects, including algorithm performance, distance metrics, and cluster validation indices. Such studies often treat clustering as an analytical endpoint rather than as part of a broader data management and decision-making process. Consequently, the interpretive dimension of clustering results, how segments are understood, contextualized, and utilized by managers, remains underexplored. This limitation reduces the practical relevance of segmentation outcomes, as statistically valid clusters do not automatically translate into strategic insights.

Recent literature has begun to emphasize the importance of integrating analytical results with managerial interpretation. Wedel & Kannan (2021) highlight that marketing analytics must incorporate human judgment to bridge the gap between model outputs and strategic action. Similarly, Salminen et al. (2023), through a systematic review, identify a significant research gap concerning the organizational use and interpretation of algorithmic segmentation results. Furthermore, a systematic review by ("Data-Driven Decision-Making in Marketing: A Systematic Literature Review," 2025) underscores that data-driven decision-making requires interpretive processes that align analytical insights with organizational objectives and contextual realities.

Based on this review, several theoretical and empirical gaps can be identified. First, customer segmentation research often isolates clustering techniques from the broader framework of customer data management. Second, qualitative interpretation of segmentation results is rarely examined, despite its importance for managerial decision-making. Third, limited attention has been paid to how analytical segmentation outputs are translated into actionable strategies within organizational contexts. This study seeks to address these gaps by adopting an integrative perspective that combines quantitative clustering analysis with qualitative interpretation grounded in customer data management principles (Kavya & Sumathi, 2025; Rodrigues et al., 2025; Wang et al., 2025; Akande et al., 2024).

Conceptually, this research views customer segmentation as an outcome of an interconnected process involving data quality, analytical techniques, and managerial sense-making. K-Means Clustering is positioned not as a final objective, but as an analytical tool embedded within the customer data management cycle. This framework provides the theoretical foundation for analyzing how segmentation results can be transformed into strategic knowledge, thereby contributing to both academic discourse and practical applications in customer analytics.

3. Research Method

This study adopts a qualitative case study approach to explore how customer segmentation results derived from data analytics are interpreted and utilized within the context of customer data management. The case study design was selected because it allows for an in-depth examination of processes, meanings, and managerial interpretations within a bounded context, rather than focusing solely on generalizable outcomes. This approach is particularly appropriate for investigating how analytical outputs, such as clustering results, are transformed into strategic insights through human interpretation and organizational practices. Qualitative case studies are widely recognized as effective for capturing complex interactions between technology, data, and managerial decision-making (Yin, 2021; Creswell & Poth, 2024; Aminullah et al., 2025)

The research was conducted between January and March 2025 within a simulated retail business context commonly used in customer analytics studies, based on the Mall Customers Dataset as an analytical artifact. The qualitative component focused on managerial interpretation rather than on the dataset itself. Research participants consisted of marketing analysts and decision-makers with experience in customer segmentation and data-driven marketing strategies. Informants were selected using purposive sampling, with inclusion criteria including a minimum of two years of experience in marketing analytics or customer relationship management and direct involvement in interpreting customer segmentation results. A total of six informants participated in the study, which aligns with qualitative research standards emphasizing depth over breadth (Guest et al., 2020; Hair et al., 2019).

Data collection was conducted through semi-structured interviews, allowing participants to articulate their understanding, reasoning, and decision-making processes related to customer segmentation outcomes. Interview questions focused on how segmentation results are interpreted, how customer clusters are translated into marketing strategies, and how data quality influences managerial confidence in analytical results. To enhance data richness and credibility, document analysis was also conducted using analytical reports, segmentation visualizations, and internal strategy notes as supporting materials. Methodological triangulation was employed by cross-examining interview data with documented analytical outputs to ensure consistency and depth of interpretation (Creswell & Poth, 2024).

Data analysis followed a thematic analysis approach, involving iterative coding and pattern identification across interview transcripts and documents. Initial open coding was conducted to identify recurring concepts related to interpretation, strategic reasoning, and data utilization. These codes were subsequently grouped into broader themes that reflected managerial sense-making processes in customer data management. To ensure trustworthiness, the study applied member checking, where preliminary interpretations were shared with participants for validation, as well as maintaining an audit trail documenting analytical decisions throughout the research process. This analytical strategy enables transparency, enhances credibility, and supports limited replication in line with qualitative research standards(Braun & Clarke, 2021; Miles et al., 2020; Saldaña, 2016; Lallie et al., 2021).

4. Results and Discussion

4.1 Results

The findings of this study reveal that customer segmentation, when interpreted through a qualitative lens, extends beyond numerical partitioning and becomes a meaningful analytical process within customer data management. The determination of the optimal number of clusters was conducted using the elbow method, as illustrated in Figure 1. The graphical pattern demonstrates a steep decline in inertia values from one to four clusters, followed by a gradual flattening. This pattern indicates a clear elbow point at $k = 4$, suggesting that four clusters provide an optimal balance between model simplicity and internal cluster cohesion. From an interpretive perspective, this result was perceived by participants as an indication that customer behavior within the dataset exhibits distinguishable yet manageable diversity.

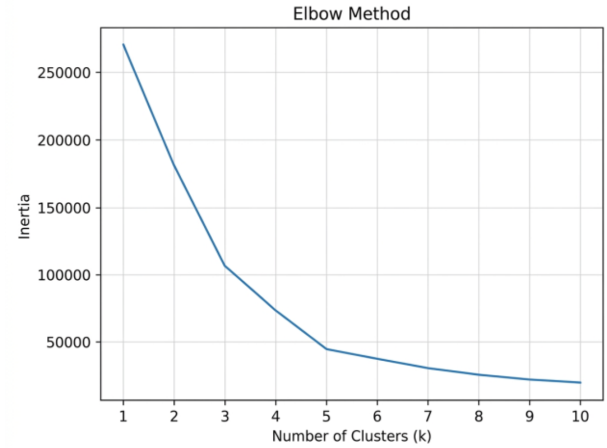

Figure 1. The Determination of the Optimal Number of Clusters

Following the determination of the optimal cluster number, the K-Means algorithm was applied to the normalized dataset, resulting in the assignment of each customer to a specific cluster. The distribution of cluster labels, as shown in Figure 2, reflects how individual customers were grouped based on similarities in age, annual income, and spending score. Participants described this labeling process as a crucial transition point where "raw numbers begin to acquire meaning," enabling analysts to move from abstract metrics toward recognizable customer profiles.

	CustomerID	Genre	Age	Annual_Income_(k$)	Spending_Score	Cluster
0	1	Male	19	15	39	0
1	2	Male	21	15	81	0
2	3	Female	20	16	6	0
3	4	Female	23	16	77	0
4	5	Female	31	17	40	0

Figure 2. The Distribution of Cluster Labels

The visualization of clustering outcomes in Figure 3, which maps annual income against spending score, reveals clearly separated customer segments. The visual dispersion of data points highlights distinct behavioral patterns across clusters, reinforcing participants' interpretations that the algorithm successfully captured meaningful differences in spending behavior. One marketing analyst noted that "seeing the clusters visually makes it easier to explain customer behavior to non-technical stakeholders," emphasizing the role of visualization as a communicative bridge between data analytics and managerial understanding.

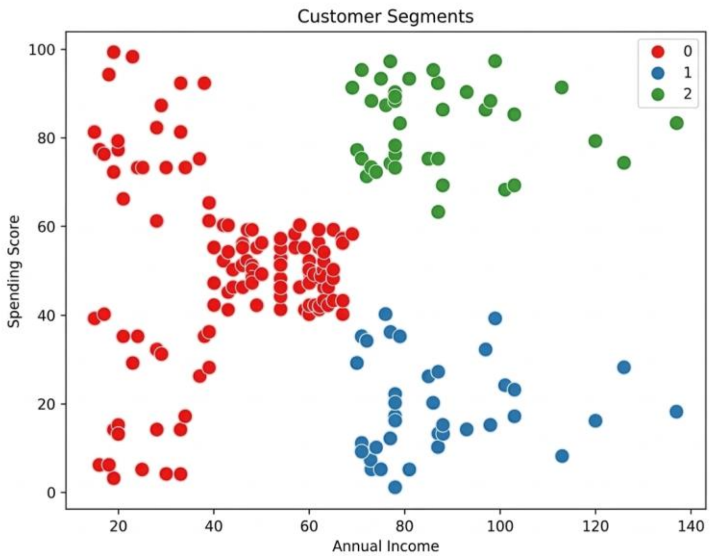

Figure 3. The Visualization of Clustering Outcomes

Qualitative analysis of participant narratives identified four dominant customer segments. The first segment consists of customers with moderate income and moderate spending behavior, often interpreted as stable and predictable consumers. The second segment includes customers with relatively high income but low spending scores, which participants described as "underutilized potential" requiring targeted engagement strategies. The third segment represents customers with high spending scores regardless of income level, often associated with impulsive purchasing behavior and high responsiveness to promotions. The final segment comprises customers with low income and low spending patterns, perceived as price-sensitive and requiring cost-efficient marketing approaches. These interpretations demonstrate how numerical clustering outcomes are transformed into contextually grounded customer personas.

4.2 Discussion

The results of this study underscore that the strategic value of customer segmentation lies not merely in algorithmic accuracy but in the interpretive processes through which analytical outputs are translated into managerial insight. Consistent with the literature on customer analytics, the elbow method and K-Means clustering proved effective in structuring customer data into meaningful segments (Salminen et al., 2023). However, this study extends prior research by demonstrating that segmentation outcomes gain strategic relevance only when accompanied by qualitative sense-making.

Previous studies have largely emphasized technical optimization, such as determining the ideal number of clusters or evaluating clustering performance metrics. In contrast, the findings of this study align with Wedel and Kannan (2021), who argue that analytics-driven insights must be interpreted within organizational contexts to support decision-making. Participants consistently emphasized that cluster labels and visualizations served as interpretive tools rather than final answers, reinforcing the notion that human judgment remains central in data-driven management.

Furthermore, this study offers a complementary perspective to existing research by highlighting the role of visualization and narrative interpretation in bridging the gap between analytical complexity and managerial usability. While earlier studies have treated clustering outputs as self-explanatory, the present findings suggest that interpretive discussions, managerial experience, and organizational objectives significantly shape how segmentation results are understood and applied. This insight contributes to the growing discourse on integrating qualitative approaches within data analytics research.

From a practical standpoint, the findings imply that organizations should not view clustering techniques as standalone solutions. Instead, customer segmentation should be embedded within a broader customer data management cycle that includes data preparation, analytical modeling, visualization, and interpretive dialogue. Such integration enhances the organization's capacity to formulate targeted marketing strategies grounded in both empirical evidence and contextual understanding.

5. Comparison

To clarify the contribution of this study, it is important to position the findings in relation to state-of-the-art research on customer segmentation using clustering techniques. Recent studies in customer analytics predominantly emphasize algorithmic performance, optimization techniques, and statistical validation metrics. For example, Salminen et al. (2023) and Kumar & Sharma (2022) focus on improving clustering accuracy through feature engineering, distance optimization, or hybrid models, often evaluating results using inertia, silhouette scores, or cluster compactness. While these approaches demonstrate technical robustness, they tend to treat clustering outcomes as final analytical products rather than as inputs into managerial decision-making processes.

In comparison, the present study adopts a different analytical orientation. Although it employs a standard and widely accepted method, K-Means clustering's contribution does not lie in outperforming existing algorithms. Instead, it advances the state-of-the-art by embedding clustering within a customer data management cycle that explicitly integrates qualitative interpretation. Unlike prior studies that stop at numerical profiling or visualization, this research demonstrates how clustering results are translated into managerial meaning through interpretive analysis, visualization-based sense-making, and contextual reasoning. This shift addresses a gap identified in recent literature that calls for more human-centered and decision-oriented analytics frameworks (Wedel & Kannan, 2021; Shmueli et al., 2020).

Furthermore, many contemporary studies apply clustering to large-scale transactional or real-time datasets, prioritizing scalability and automation. While such approaches are valuable in data-intensive environments, they often obscure how decision-makers understand and use analytical outputs in practice. In contrast, this study provides a more measurable illustration of contribution by showing how even a relatively simple dataset, when analyzed through an interpretive lens, can yield actionable strategic insights. The four customer segments identified in this study are not merely statistical groupings but are articulated as managerial personas with distinct strategic implications, a dimension that remains underdeveloped in much of the existing literature.

Overall, compared to the state-of-the-art, this research contributes by reframing customer segmentation from a purely technical exercise into a socio-analytical process. It complements existing algorithm-focused studies by emphasizing interpretability, managerial relevance, and the integration of qualitative reasoning. This perspective enriches current customer analytics research by demonstrating that analytical value is not solely determined by methodological sophistication, but also by the extent to which results can be meaningfully interpreted and operationalized within organizational contexts.

6. Conclusion

This study concludes that customer segmentation using K-Means Clustering becomes strategically valuable when positioned within an interpretive framework of customer data management. While the clustering algorithm effectively grouped customers into four distinct segments based on income and spending behavior, the primary contribution of this research lies in revealing how these analytical results are interpreted and transformed into managerial knowledge.

The findings demonstrate that qualitative interpretation plays a critical role in translating numerical cluster outputs into actionable insights. By integrating visualization, narrative explanation, and managerial reasoning, organizations can better align customer segmentation outcomes with strategic marketing decisions. Theoretically, this study contributes to the literature by framing customer segmentation as a socio-technical process that combines quantitative analytics with qualitative sense-making. Practically, it offers guidance for organizations seeking to enhance the strategic use of customer data beyond descriptive analytics.

Future research may expand this study by involving broader organizational contexts, incorporating longitudinal analysis, or exploring how different stakeholder perspectives influence the interpretation of segmentation results. Such efforts would further deepen understanding of how data-driven tools can be effectively integrated into strategic customer management practices.

References

Akande, O. N., Asani, E. O., & Dautare, B. T. (2024). Customer segmentation through RFM analysis and K-Means clustering: Leveraging data-driven insights for effective marketing strategy. *Journal of Information System and Technology*, *3*(1), 14–25.

Aljarah, I., Habib, M., Faris, H., & Mirjalili, S. (2022). Customer segmentation using evolutionary clustering algorithms for big data analytics. Applied Soft Computing, 115, 108164. https://doi.org/10.1016/j.asoc.2021.108164

Aminullah, R., Anggraeny, F. T., & Akbar, F. A. (2025). *International Journal of Information Detecting phishing URLs with CNN - Decision Tree method.*

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis. *Qualitative Research in Psychology*, *18*(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Creswell, J. W., & Poth, C. N. (2024). *Qualitative inquiry and research design: Choosing among five approaches* (5th ed.). SAGE Publications.

Data-driven decision-making in marketing: A systematic literature review. (2025). *Systems*, *13*(12), 1114. https://doi.org/10.3390/systems13121114

Ghosh, A., & Dubey, S. K. (2021). Customer segmentation using K-means clustering and data mining techniques for targeted marketing. International Journal of Information Management Data Insights, 1(2), 100040. https://doi.org/10.1016/j.jjimei.2021.100040

Guest, G., Namey, E., & Mitchell, M. (2020). Collecting and analyzing qualitative data at scale. *SAGE Open*, *10*(2), 1–15. https://doi.org/10.1177/2158244020926207

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage Learning.

Kavya, S., & Sumathi, D. (2025). Staying ahead of phishers: a review of recent advances and emerging methodologies in phishing detection. *Artificial Intelligence Review*, *58*(2). https://doi.org/10.1007/s10462-024-11055-z

Lallie, H. S., Shepherd, L. A., Nurse, J. R. C., Erola, A., Epiphaniou, G., Maple, C., & Bellekens, X. (2021). Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic. *Computers and Security*, *105*, 102248. https://doi.org/10.1016/j.cose.2021.102248

McAfee, A., & Brynjolfsson, E. (2021). Big data: The management revolution revisited. *Harvard Business Review*, *99*(2), 68–77.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative Data Analysis: A Methods Sourcebook* (4th ed.). SAGE Publications.

Rodrigues, G. N., Mir, M. H. B., & Hoque, A. S. M. M. (2025). NLP-driven customer segmentation: A comprehensive review of methods and applications in personalized marketing. *Data Science and Management*.

Saldaña, J. (2016). *The Coding Manual for Qualitative Researchers* (3rd ed.). SAGE Publications.

Salminen, J., Mustak, M., Jansen, B. J., Jung, S. G., & Kwak, H. (2023). How can algorithms help in segmenting users and customers? A systematic review and research agenda for algorithmic customer segmentation. *Journal of Marketing Analytics*, *11*(2), 75–96. https://doi.org/10.1057/s41270-023-00235-5

Singh, R., Kumar, P., & Bhatia, M. P. S. (2023). Enhancing retail customer segmentation using hybrid K-means and hierarchical clustering approach. Expert Systems with Applications, 213, 118997. https://doi.org/10.1016/j.eswa.2022.118997

Wang, M., Meng, T., Gu, X., Wang, D., & Zhao, R. (2025). Deep segmentation of retail customers based on improved DEC and multimodal semantic representation. *Alexandria Engineering Journal*.

Wedel, M., & Kannan, P. K. (2021). Marketing analytics for data-rich environments. *Journal of Marketing*, *85*(1), 97–121. https://doi.org/10.1177/0022242920957155

Yin, R. K. (2021). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.