Analysis of Sales, Profitability, and Customer Behavior in the Superstore Dataset

Análisis de Ventas, Rentabilidad y Comportamiento del Cliente en el Dataset Superstore

DOI: http://dx.doi.org/10.17981/cesta.07.01.2026.02
Scientific research article
Date received: May 04, 2026, Date accepted: June 09, 2026.

Hugo Luis Salazar Jiménez
Universidad del Sinú Seccional Cartagena (Colombia)
hugo.salazar@unisinu.edu.co

How to cite:
H. Salazar Jiménez, "Analysis of Sales, Profitability, and Customer Behavior in the Superstore Dataset", J. Comput. Electron. Sci.: Theory Appl., vol. 7, no. 1, pp. 11-19, 2026. DOI: 10.17981/cesta.07.01.2026.02

Abstract

Introduction: The analysis of sales, profitability, and customer behavior has become essential for improving decision-making in the retail sector, especially through the use of business intelligence tools that allow data to be explored visually and analytically.
Objective: To analyze the factors that influence business performance in the retail sector using the Sample Superstore dataset, with emphasis on the effect of discounts on profitability, customer segment behavior, and geographic differences in sales and profit margins.
Method: A descriptive and analytical approach was applied using Microsoft Power BI as the main tool for data visualization and exploration. The analysis was structured around four research questions related to product category performance, discount impact, customer segment profitability, and regional opportunities for improvement.
Results: The results show that the Technology category generates the highest sales and profit levels, while some subcategories, such as Tables, present significant losses despite having considerable sales volume. A negative relationship was identified between discounts and profitability, especially when discounts exceed moderate levels. The Consumer segment generates the highest sales volume, whereas the Corporate segment shows more stable and profitable behavior. In addition, states such as Texas and Illinois were identified as areas with high sales volume but low profit margins.
Conclusions: The findings suggest the need to optimize pricing strategies, control discount policies, strengthen high-value customer segments, and improve regional commercial management in order to increase profitability and support more effective decision-making in the retail sector.

Keywords: Retail analytics, sales analysis, profitability, customer behavior, discounts, Power BI.

Resumen

Introducción: El análisis de ventas, rentabilidad y comportamiento del cliente es clave para fortalecer la toma de decisiones en el sector minorista mediante herramientas de inteligencia de negocios.
Objetivo: Analizar los factores que influyen en el desempeño empresarial del sector minorista a partir del conjunto de datos *Sample Superstore*, con énfasis en descuentos, segmentos de clientes y diferencias geográficas en ventas y rentabilidad.
Método: Se aplicó un enfoque descriptivo y analítico mediante Microsoft Power BI para visualizar y explorar los datos, considerando el desempeño por categoría, el impacto de los descuentos, la rentabilidad por segmento y las oportunidades regionales de mejora.
Resultados: La categoría *Technology* presentó los mayores niveles de ventas y rentabilidad, mientras que subcategorías como *Tables* registraron pérdidas pese a su alto volumen de ventas. Se evidenció una relación negativa entre descuentos y rentabilidad, especialmente ante descuentos elevados. El segmento *Consumer* concentró el mayor volumen de ventas, mientras que *Corporate* mostró un comportamiento más estable y rentable. Asimismo, Texas e Illinois presentaron altas ventas, pero bajos márgenes de ganancia.
Conclusiones: Los resultados resaltan la necesidad de optimizar precios, controlar descuentos, fortalecer segmentos de alto valor y mejorar la gestión comercial regional para incrementar la rentabilidad y apoyar decisiones más efectivas en el sector minorista.

Palabras clave: Analítica retail, Análisis de ventas, Rentabilidad, comportamiento del cliente, Descuentos, Power BI.

INTRODUCTION

In today's dynamic and fiercely competitive business landscape, the ability to understand and predict sales performance and profitability is not merely an advantage but the foundation upon which lasting business strategies are built [1], [2]. For this reason, we have turned to the extensive and rich Superstore dataset, which offers a wealth of detailed information from orders and customer profiles to geographic locations, applied discounts, and resulting profits, serving as the ideal starting point for conducting an in-depth exploratory analysis in the retail sector [3], [4]. Our primary objective throughout this study is to unravel the complex factors that explain these performance variations, analyzing how the discount mechanism directly impacts profitability, how purchasing behavior differs across customer segments, and where the true geographic opportunities lie for optimizing commercial management [1], [2]. To this end, we will go beyond superficial figures, addressing intermediate-complexity questions to simultaneously evaluate market dynamics and internal operational efficiency, all while leveraging the powerful visualization and modeling capabilities of Microsoft Power BI. As for the paper's structure, it begins by outlining the descriptive-analytical methodology employed, proceeds to present detailed results that answer our four research questions, then focuses on a discussion that integrates these answers for a holistic understanding of the phenomenon, and finally culminates in a series of conclusions and practical recommendations designed to translate the findings into concrete strategic actions and improve the overall profitability of the business [1], [2].

THEORETICAL FRAMEWORK

Retail sales and profitability analysis has become essential for strategic decision-making, especially when organizations manage large volumes of transactional data. Retail analytics allows companies to identify consumption patterns, evaluate product categories and subcategories, measure the impact of discounts, segment customers, and detect geographic opportunities for improvement. In this study, variables such as sales, profit, discount, category, subcategory, customer segment, region, and state are used to assess business performance from an integrated perspective.

Recent studies highlight that retail analytics is increasingly supported by artificial intelligence, big data, business intelligence, and data visualization tools. These approaches help transform transactional data into useful information for managerial decision-making [5]. In this context, sales and profit are central performance indicators, while category and subcategory allow the identification of profitable product lines and those that may negatively affect margins.

Business intelligence tools, such as dashboards and interactive visualizations, help users understand complex information and support better organizational decisions [6]. In this study, Power BI is used to relate sales, profit, discounts, customer segments, and geographic dimensions, making it possible to identify patterns such as high-sales products with low profitability or states with strong sales but reduced margins.

Profitability in retail depends not only on sales volume, but also on costs, discounts, pricing strategies, logistics, and portfolio composition. Therefore, high sales do not always imply high profit. This is particularly relevant when analyzing subcategories that generate considerable revenue but produce losses due to inappropriate pricing or excessive discounts.

Discounts are a key variable because they can stimulate demand but also reduce profitability when applied without proper control. Recent research suggests that pricing and discount strategies should consider customer segmentation, loyalty, and price sensitivity [7]. In this study, the discount variable is directly related to profit, allowing the evaluation of whether promotions increase business value or erode margins.

Customer segmentation is also essential in retail analytics. It allows companies to differentiate customers according to purchasing behavior, transaction value, frequency, and profitability. Studies using big data and clustering models show that customer segmentation supports more effective marketing strategies and improves long-term customer relationships [8]. In the Superstore dataset, the customer segment variable helps compare groups such as Consumer, Corporate, and Home Office in terms of sales volume and profitability.

Customer behavior can also be studied through predictive analytics. Machine learning models have been used to anticipate customer preferences, market trends, and purchasing behavior [9]. Although this study is mainly descriptive and analytical, variables such as order date, sales, profit, and customer segment could support future predictive studies focused on demand forecasting or expected profitability.

The temporal dimension is important because customer behavior and sales performance may change over time. Recent approaches based on time-series analysis allow researchers to identify variations in purchase frequency, transaction value, and customer behavior [10]. In this study, the order date variable supports the analysis of monthly or yearly changes in sales and profitability.

Geographic analysis is another relevant dimension in retail performance. Variables such as region, state, and city help identify territories with high sales volume but low profit margins. These differences may be associated with logistics costs, discount policies, local demand, or regional pricing strategies. Therefore, geographic analysis supports the design of differentiated strategies according to each market.

Finally, data-driven marketing integrates customer, sales, and performance information to improve decision-making, resource allocation, personalization, and business profitability. Recent literature shows that quantitative data and analytical models are increasingly used to support strategic and operational marketing decisions [11]. Therefore, the combined analysis of sales, profit, discounts, customer segments, and geographic location provides a solid basis for improving pricing strategies, reducing unprofitable discounts, strengthening high-value segments, and optimizing regional performance.

In summary, this theoretical framework supports the study through five main dimensions: commercial performance, represented by sales and profit; product portfolio management, represented by category and subcategory; promotional strategy, represented by discounts; customer behavior, represented by customer segment and time; and geographic performance, represented by region, state, and city.

METHODOLOGY

This study follows a descriptive and analytical approach aimed at identifying patterns in sales, profitability, customer behavior, discount application, and geographic performance. The analysis was conducted using the Sample Superstore dataset in .xlsx format, which contains transactional information related to orders, sales, profits, discounts, product categories, customer segments, and geographic locations.

Microsoft Power BI Desktop was used as the main tool for data processing, modeling, visualization, and interpretation. The dataset was imported into Power BI, where the structure of the data was reviewed and the relevant fields were selected according to the objectives of the study. The main variables considered were Sales, Profit, Discount, Category, Subcategory, Segment, Region, State, City, and Order Date.

The analysis was organized into four dimensions. First, product performance was evaluated through the relationship between Category, Subcategory, Sales, and Profit in order to identify the product groups with the highest contribution to business performance. Second, the impact of discounts on profitability was examined by comparing Discount and Profit values across products and subcategories. Third, customer behavior was analyzed by comparing sales and profit levels among the Consumer, Corporate, and Home Office segments. Fourth, the geographic dimension was explored through Region, State, and City to identify territories with high sales volume but low profit margins.

Several visualizations were created in Power BI, including bar charts, scatter plots, segment comparison charts, and geographic maps. These visualizations allowed the identification of relevant patterns, such as profitable categories, unprofitable subcategories, the negative effect of excessive discounts, differences among customer segments, and regional areas requiring commercial improvement. The results obtained from the visual analysis were interpreted in relation to the research questions and supported the development of the discussion, conclusions, and recommendations.

RESEARCH QUESTIONS

- Which category and subcategory generate the highest sales, and what factors explain their performance relative to the rest of the portfolio?
- What is the impact of the discount on profitability, and is there a point at which the discount ceases to be effective?
- Which customer segments contribute the most profitability, and how does their purchasing behavior vary over time?
- Which regions or states represent opportunities for improvement due to high sales volume but low profit margin?.

RESULTS

A. Performance by Category an Subcategory

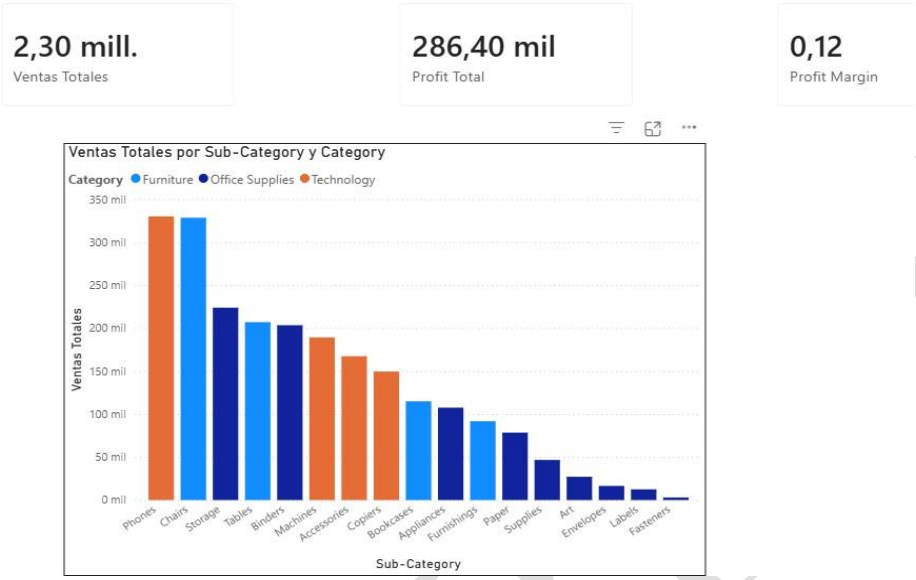

Figure 1: Sales and Profit by Product Category and Subcategory

Our analysis reveals that the Technology category clearly leads in sales, followed closely by Furniture and Office Supplies. We observe that, within Technology, specific subcategories such as Phones, Accessories, and Copiers not only account for a high sales volume but also maintain significantly high profit margins. However, we find a concerning contrast in the Tables subcategory, which, despite generating considerable sales, compels us to highlight its alarming level of losses, making it a critical focus of attention for the company. We interpret that Technology leads the market thanks to its high turnover and premium product pricing, while the situation with Tables suggests serious cost issues or, possibly, a very poorly implemented discount policy, which is directly affecting its profitability.

B. Performance by Category and Subcategory

When examining the relationship between discount and profit, the correlation analysis confirmed a markedly negative relationship: in practice, as the discount increases, profitability systematically decreases, even resulting in significant losses. After a detailed segmentation, we observed that discounts in the 20% to 40% range are frequently associated with losses in various subcategories, and even moderate discounts on products such as tables or bookcases lead to negative profitability. This leads us to the clear conclusion that there is no “optimal point” at which a discount drives sales without harming profitability, so its application must be managed with much greater caution and selectivity.

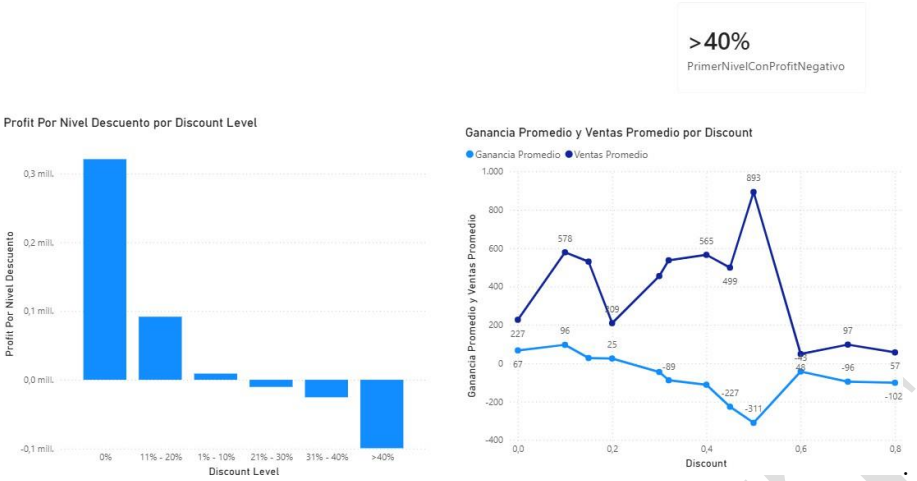

Figure 2: Relationship Between Discount and Profit

C. Profitability by Customer Segment

When classifying our customers into the Consumer, Corporate, and Home Office segments, we noticed that the Consumer segment generates the highest sales volume, followed by Corporate. However, it is the Corporate segment that exhibits consistently more stable margins, while Home Office shows the widest fluctuations in profitability. The key findings suggest that, while Consumer is the largest segment, it is also the most susceptible to discounts; on the other hand, Corporate comprises more consistent customers, with business purchases that naturally tend to be more profitable; finally, the Home Office segment clearly needs incentives and differentiated value strategies to effectively increase its share.

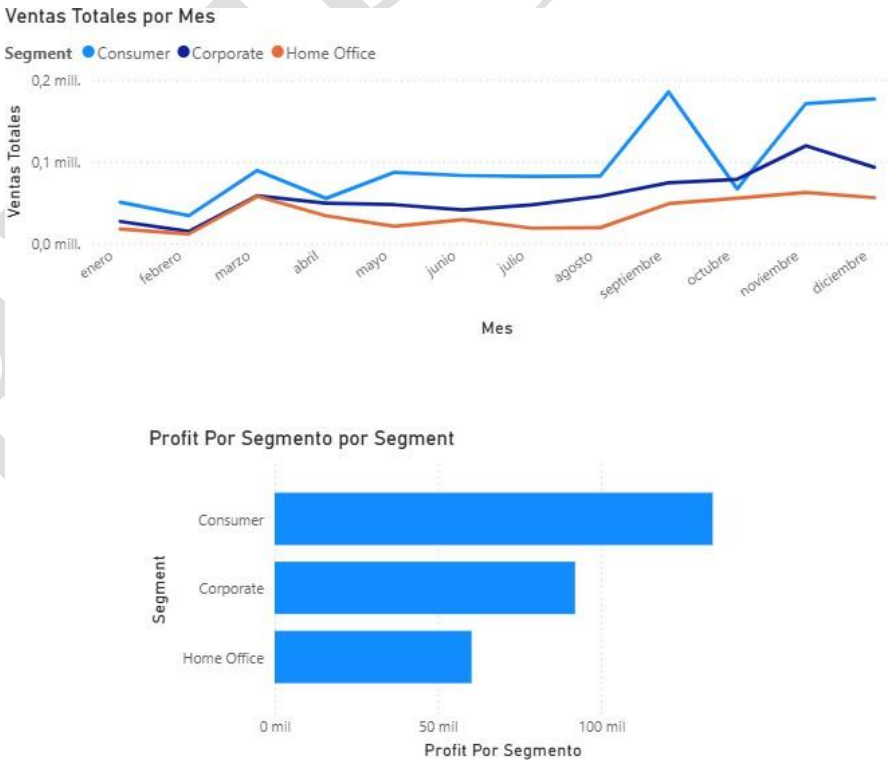

Figure 3: Sales and Profit by Customer Segment

D. Profitability by Customer Segment

Using a heat map in Power BI was revealing, showing that although major states like California, New York, and Texas lead in sales volume, some of them have much lower margins than expected. A notable example is Texas, which, despite its high sales volume, exhibits reduced profits—a situation likely linked to higher discount levels or less efficient logistics costs; we also observe inconsistencies between sales and cumulative profit in states such as Illinois and Pennsylvania. This underscores significant opportunities for improvement, including reevaluating state-specific pricing and discount strategies, optimizing shipping and storage costs, and even seeking more efficient logistics partnerships in regions where losses are most pronounced.

Ventas Totales	Suma de Profit	Profit Margin	% Ordenes Devueltas	State
19.510,64	5.786,83	0,30	8,71	Alabama
35.282,00	-3.427,92	-0,10	2,74	Arizona
11.678,13	4.008,69	0,34	10,96	Arkansas
457.687,63	76.381,39	0,17	0,29	California
32.108,12	-6.527,86	-0,20	3,75	Colorado
13.384,36	3.511,49	0,26	6,58	Connecticut
27.451,07	9.977,37	0,36	6,73	Delaware
2.865,02	1.059,59	0,37	74,00	District of Columbia
89.473,71	-3.399,30	-0,04	1,48	Florida
49.095,84	16.250,04	0,33	3,25	Georgia
4.382,49	826,72	0,19	26,91	Idaho
80.166,10	-12.607,89	-0,16	1,07	Illinois
53.555,36	18.382,94	0,34	4,05	Indiana
4.579,76	1.183,81	0,26	17,41	Iowa
2.914,31	836,44	0,29	21,14	Kansas
36.591,75	11.199,70	0,31	4,85	Kentucky
9.217,03	2.196,10	0,24	14,10	Louisiana
1.270,53	454,49	0,36	98,67	Maine
23.705,52	7.031,18	0,30	6,43	Maryland
28.634,43	6.785,50	0,24	4,77	Massachusetts
76.269,61	24.463,19	0,32	2,53	Michigan
29.863,15	10.823,19	0,36	6,73	Minnesota
10.771,34	3.172,98	0,29	11,38	Mississippi
22.205,15	6.436,21	0,29	9,87	Missouri
5.589,35	1.833,33	0,33	37,00	Montana
7.464,93	2.037,09	0,27	12,87	Nebraska
16.729,10	3.316,77	0,20	12,87	Nevada
7.292,52	1.706,50	0,23	17,41	New Hampshire
2.297.200,86	**286.397,02**	**0,12**	**0,06**	

Figure 4: Geographic Distribution of Sales and Profit by State

DISCUSSION

The results of this study show that retail performance cannot be evaluated only through sales volume, since profitability depends on the interaction between product category, discount policy, customer segment, and geographic location. This finding is consistent with recent approaches in retail analytics, which emphasize the use of transactional data, business intelligence tools, and visual analytics to support strategic decision-making [5], [6].

Regarding product performance, the Technology category shows the strongest contribution to both sales and profit. This behavior may be associated with higher-value products, stronger demand, and better profit margins within subcategories such as Phones, Accessories, and Copiers. However, the Tables subcategory presents a contrasting situation: although it generates considerable sales, it also produces significant losses. This suggests that high demand does not necessarily imply positive profitability and that certain product lines may require a review of costs, pricing strategies, and promotional practices.

The analysis of discounts reveals a negative relationship between discount levels and profitability. In several cases, higher discounts are associated with lower profit or even losses, particularly in subcategories such as Tables and Bookcases. This result supports the idea that discount strategies should not be applied uniformly across all products. Instead, they should be designed according to product margins, customer sensitivity, and segment behavior. Recent

literature also highlights that pricing and discount decisions should consider customer segmentation and price sensitivity to avoid margin erosion [7].

Customer segment analysis indicates that the Consumer segment contributes the highest sales volume, while the Corporate segment shows more stable and profitable results. This suggests that sales volume and customer value should be analyzed separately. Although individual consumers may generate more transactions, corporate customers may offer better profitability due to more consistent purchasing behavior and lower sensitivity to discounts. Therefore, differentiated marketing strategies should be developed for each segment, prioritizing customer value rather than only transaction volume.

The geographic analysis also reveals relevant differences in commercial performance. States such as Texas and Illinois show high sales volume but low profit margins, which may be related to high discount levels, logistics costs, regional pricing policies, or product mix. These findings indicate that regional performance should be analyzed not only by revenue but also by profitability. Geographic visualization in Power BI provides useful support for identifying areas that require operational or commercial adjustments.

Overall, the findings demonstrate the value of combining sales, profit, discount, customer, product, and geographic variables in a single analytical model. The use of Power BI made it possible to identify patterns that may remain hidden in traditional tabular analysis. Therefore, the study provides useful evidence for improving pricing policies, reducing unprofitable discounts, strengthening high-value customer segments, and optimizing regional strategies. Future studies could expand this analysis by incorporating predictive models, demand elasticity analysis, logistics cost evaluation, and customer lifetime value estimation.

RECOMMENDATIONS

To translate these findings into action and improve overall profitability, we suggest the following actions. First, it is essential to reduce or completely eliminate discounts in those subcategories that have shown historical losses, such as Tables, and in those where the negative impact is greatest. Second, the company should move toward implementing dynamic pricing, which would be based on actual, region-specific market behavior. Third, we recommend designing promotions specifically targeted by segment, always focusing on customer value rather than solely on increasing volume. Fourth, it is crucial to carry out additional studies on demand elasticity and logistics costs to understand price limits in each subcategory. Finally, we advise segmenting marketing strategies, giving priority to the Corporate segments and to customers classified as Consumer Premium.

CONCLUSIONS

This study shows that sales volume alone is not enough to evaluate business performance, since high revenue does not always lead to high profitability. The analysis of the Superstore dataset revealed that profitability depends on the relationship between sales, discounts, product category, customer segment, and geographic location. The Technology category showed the strongest overall performance, with relevant contributions to both sales and profit. However, some subcategories, such as Tables, presented important losses despite generating considerable sales. This indicates the need to review pricing strategies, costs, and discount policies for products with low or negative margins. The results also confirm that discounts have a significant impact on profitability. Although discounts may increase sales volume, excessive or poorly targeted discounts can reduce profit and even generate losses. Therefore, promotional strategies should be applied more selectively, considering the profitability of each product category and customer segment.

Regarding customer behavior, the Consumer segment generated the highest sales volume, while the Corporate segment showed more stable and profitable results. This suggests that the company should strengthen strategies aimed at high-value customers and design differentiated marketing actions according to each segment. The geographic analysis identified regions and states with high sales but low profit margins, such as Texas and Illinois. These findings suggest opportunities to review logistics costs, regional pricing strategies, and discount policies in specific locations. In conclusion, the use of Power BI and data analytics allowed the identification of key business patterns that can support strategic decision-making. The company should move from a strategy focused only on increasing sales toward a more balanced approach based on profitability, customer segmentation, product performance, and regional optimization.

AUTHOR CONTRIBUTION

The author's contributions to this article are as follows:

Hugo Luis Salazar Jiménez: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Project administration, Writing–original draft, Writing–review and editing.

The author reviewed the results and approved the final version of the article.

CONFLICT OF INTERESTS

The authors declare that they have no interests or financial relationships that could have influenced this work.

REFERENCES

[1] P. Kopalle, C. Mela, and L. C. Marsh, "The Dynamic Effect of Discounting on Sales: Empirical Analysis and Normative Pricing Implications," Marketing Science, vol. 18, no. 3, pp. 275–295, 1999.

[2] D. K. Gauri, B. Ratchford, J. Pancras, and D. Talukdar, "An Empirical Analysis of the Impact of Promotional Discounts on Store Performance," Journal of Retailing, vol. 93, no. 3, pp. 283–303, 2017.

[3] P. Anitha and M. Patil, "RFM model for customer purchase behavior using K-Means algorithm," Journal of King Saud University – Computer and Information Sciences, vol. 31, no. 4, pp. 528–536, 2019.

[4] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," Sustainability, vol. 14, no. 12, p. 7032, 2022.

[5] J. D. Velásquez, "An analysis of trends, challenges, and opportunities in retail analytics," Journal of Advertising Research, vol. 67, no. 4, 2025, doi: 10.1177/14707853251315585.

[6] S. Hjelle, P. Mikalef, N. Altwaijry, and V. Parida, "Organizational decision making and analytics: An experimental study on dashboard visualizations," Information & Management, 2024, doi: 10.1016/j.im.2024.104011.

[7] L. D. Gonen, T. Tavor, and U. Spiegel, "Unlocking Market Potential: Strategic Consumer Segmentation and Dynamic Pricing for Balancing Loyalty and Deal Seeking," Mathematics, vol. 12, no. 21, p. 3364, 2024, doi: 10.3390/math12213364.

[8] R. H. Lin et al., "Applied Big Data Analysis to Build Customer Product Recommendation Model," Sustainability, vol. 13, no. 9, p. 4985, 2021, doi: 10.3390/su13094985.

[9] H. GhorbanTanhaei, P. Boozary, S. Sheykhan, M. Rabiee, F. Rahmani, and I. Hosseini, "Predictive analytics in customer behavior: Anticipating trends and preferences," Results in Control and Optimization, vol. 17, p. 100462, 2024, doi: 10.1016/j.rico.2024.100462.

[10] H. Hamidi and B. Haghi, "An approach based on data mining and genetic algorithm to optimizing time series clustering for efficient segmentation of customer behavior," Computers in Human Behavior Reports, vol. 16, p. 100520, 2024, doi: 10.1016/j.chbr.2024.100520.

[11] R. N. Cruz and A. T. Rosário, "Data-Driven Decision-Making in Marketing: A Systematic Literature Review of Emerging Themes and Research Gaps," Systems, vol. 13, no. 12, p. 1114, 2025, doi: 10.3390/systems13121114.

Hugo Luis Salazar Jiménez. Systems Engineering student at Universidad del Sinú Seccional Cartagena.