Urban Consumption Dynamics And Segmentation Rfm: Sociodemographic Characterization Of Customers In Crm Systems For Commercial Optimization

Jairo Eduardo Vargas Álvarez[1], Marcela Patricia Estrada Arango[2], Omar Alejandro Afanador Ortiz[3], Omar Hernán Nova Jaimes[4], Evelio Useda Melgarejo[5]

[1]Universidad de Investigación y Desarrollo (UDI), Colombia
ORCID: https://orcid.org/0000-0001-5320-0354
[2]Universidad de Investigación y Desarrollo (UDI), Colombia
[3]Universidad de Investigación y Desarrollo (UDI), Colombia
[4]Universidad de Investigación y Desarrollo (UDI), Colombia
[5]Cacique CC, Colombia

Abstract
This study examines the relationship between urban development and consumption patterns in Bucaramanga, focusing on the impact of the Centro Comercial Cacique, the region's largest shopping center. Using a quantitative approach grounded in Customer Relationship Management (CRM) system data, we analyze visitor flows, sales performance, and purchasing behaviors alongside surrounding urban dynamics. From a total universe of 823,734 transactional records, we examined a cleaned sample of 81,853 unique clients with complete demographic information. The results reveal significant changes in consumer habits and mobility patterns influenced by architectural design and the mall's strategic location. Additionally, a catalytic effect on nearby urban development is identified, particularly in the Provenza, Cabecera, and Mutis districts. The findings provide valuable insights for urban planners, policymakers, and commercial developers, emphasizing the importance of integrating real-time consumer data into sustainable urban planning strategies.
Keywords: urbanism; consumer behavior; shopping centers; urban development; RFM segmentation; Bucaramanga.

1. INTRODUCTION

The sustained growth of shopping malls in Latin America has reconfigured the urban structure and consumption patterns, consolidating them as epicenters of social and economic interaction (Baker, Grewal & Parasuraman, 2002). These infrastructures transcend their traditional commercial role to become multifunctional spaces for meeting, identity and cultural transformation, affecting both mobility and social relations in their immediate environment. In this context, the Cacique Shopping Center in the department of Santander represents a paradigmatic case: an urban node where the dynamics of consumption, architectural planning and local development converge.

The expansion of commercial spaces is not only an economic phenomenon, but part of a broader process of urbanization of consumption, where architecture and urban design acquire decisive symbolic and functional roles (Bressolles, Durrieu & Giraud, 2015). These structures redefine urban centrality and generate new forms of appropriation of space, acting as catalysts for change in intermediary cities. In Bucaramanga, these transformations are manifested in the reorganization of mobility flows, the redesign of the urban landscape, and the evolution of the consumption habits of its residents.

Our analysis examines the interrelationship between urbanism and consumption patterns in Bucaramanga, investigating how the strategic location, architectural design, and commercial management of the Cacique Shopping Center influence consumer behavior. Based on empirical data obtained from the CRM system, we performed multivariate analyses and RFM (Recency, Frequency, Monetary) segmentation, complemented by observations on urban dynamics. This methodological combination makes it possible to link individual consumer behavior with the spatial and economic transformations of the urban environment (Schiller & Voisard, 2004).

This study seeks to fill a gap in the literature on Latin American intermediary cities, where large shopping malls exert a more concentrated structural influence than in metropolises (Gao & Liu, 2014). Bucaramanga is, therefore, an ideal setting to explore how consumption spaces generate new centralities and modify social relations, mobility and the local economy.

General objective: To characterize the customer base of a CRM system through RFM segmentation, identifying sociodemographic and behavioral patterns that allow optimizing retention strategies and understanding the territorial impact of shopping centers in intermediary cities.

Research Question:

How are the sociodemographic characteristics and consumption behaviors of the customers of the Cacique Shopping Center related to urban transformations and commercial management strategies in an intermediate city like Bucaramanga?

2. Contextualization of the topic

During the last decades, shopping malls have established themselves as one of the main agents of urban transformation in Latin America, redefining the ways of inhabiting, consuming and interacting in public space (Wakefield & Baker, 1998). In intermediary cities, their impact is usually more visible than in large metropolises, due to their ability to concentrate economic, social and symbolic flows in a single urban node.

In Bucaramanga and its metropolitan area, the development of these infrastructures has substantially modified spatial and mobility dynamics, integrating residential, commercial and recreational areas into new urban circuits. The Cacique Shopping Centre is a paradigmatic case: not only because of its scale and strategic location, but also because of its role as an articulator of new urban centralities, influencing consumer behaviour, transport planning and land valorisation.

This urban context, reflected in indicators of the Metropolitan Area of Bucaramanga (2021-2024), allows us to understand the relationship between commercial expansion and territorial reorganization. The analysis of CRM data complements this vision, showing how consumption practices are aligned with the transformations of urban space, and how commercial architecture acts as an articulator of new forms of sociability and mobility.

Sociodemographic characterization

The sociodemographic analysis constitutes the basis for understanding the composition and behavior of the consumers of the Cacique Shopping Center, based on the records of the CRM system during the period 2021-2024. From the total universe of 823,734 transactional records generated in the CRM system, a curated sample of 81,853 unique customers was selected after eliminating duplicates, incomplete records and transactions of unidentified customers, ensuring the quality and validity of the subsequent analyses.

Of these, 95.3% belong to the department of Santander and 87.2% to Bucaramanga and its metropolitan area, confirming the strong local concentration of customers.

To optimize statistical representativeness, we performed an analysis of homogeneity of categorical variables, reducing the sample to 1% of cases with the lowest frequency in consumption categories. This procedure allowed the application of clustering algorithms that grouped the records into four homogeneous segments, with representativeness greater than 1%. The results show that the most significant differences are in gender, age, marital status, occupation and means of transport.

Main demographic profile: Adult woman (34-49 years), married or single in similar proportions, employed in the formal sector. Women represent 61.9% of the total number of customers analysed (updated data according to the filtered sample of the ACM 2024 model). In terms of age groups, adults of approximately 40% concentrate the largest segment, followed by the middle-aged population (50-65 years, 29%) and young adults (25-33 years, 22%).

Mobility patterns: 75.7% of customers access the Cacique Shopping Center by public transport or pedestrian mobility, without using a private vehicle, evidencing a high dependence on collective transport and indicating significant territorial accessibility. The remaining 21.3% use their own car, while 3% use other means of transport. This pattern, associated with the central location of the establishment and its accessibility by public transport, reinforces that the Cacique Shopping Centre operates as a metropolitan convergence point.

Geographic segmentation: Four main areas of influence were identified where most of the records are concentrated: Cacique, Cañaveral, Provenza and Único Outlet. The Cañaveral, Provenza, Real de Minas, Cabecera and San Francisco neighborhoods top the Top 10 by number of customers, highlighting the correlation between residential location and frequency of consumption.

2. Objectives and research question

The general objective of this research is to characterize the customer base through RFM segmentation, identifying sociodemographic and behavioral patterns that allow optimizing retention strategies and understanding the territorial impact of shopping centers in intermediary cities.

Specific objectives:

1. Identify customer segments using RFM (Recency, Frequency, Monetary) analysis.
2. Characterize the sociodemographic profile of each segment.
3. Analyze urban mobility patterns and their relationship with purchasing behavior.
4. Evaluate the impact of the shopping center on the reorganization of local urban dynamics.
5. Propose marketing strategies based on predictive analytics.

4. Relevance and expected impact

The relevance of this research lies in its contribution to the understanding of the role played by large shopping centers in the transformation of intermediary cities, particularly in Latin American contexts. Where most studies have focused on global metropolises (Mexico City, São Paulo, Buenos Aires), the case of Bucaramanga offers a necessary perspective to analyze how large-scale commercial infrastructures influence urban structure, mobility, and consumption practices in smaller territories.

Expected impact:

Academic: To provide empirical evidence that strengthens debates on the relationship between urbanism and consumption, proposing a methodological framework that can be replicated in other intermediary cities in the region.

Practical: The results will guide sustainable urban planning, land management and mobility policies, offering useful information to formulate strategies that harmoniously integrate commercial infrastructures within the urban fabric.
Decisional: The findings are expected to influence decisions by developers and local authorities, promoting commercial growth models that maximize positive economic impact without compromising social cohesion or urban functionality.

5. THEORETICAL FRAMEWORK

5.1 Urbanism and its influence on consumption patterns

Contemporary urbanism has evolved from being an exercise in physical ordering to becoming an interdisciplinary field where territorial planning, mobility and social dynamics converge (Baker, Grewal & Parasuraman, 2002). In this framework, consumption is configured as a spatial practice: purchasing decisions, travel routes and frequency of visits are closely linked to the urban structure that hosts them.

Well-planned urban spaces facilitate access, stimulate permanence and reinforce the consumer's identity with the territory. Wakefield and Baker (1998) demonstrated that the layout of the physical environment, lighting, accessibility, and architectural elements directly influence the buyer's emotions and perceptions, shaping more satisfying consumer experiences. For their part, Bressolles, Durrieu, and Giraud (2015) argue that modern commercial architecture functions as a device of symbolic interaction, where the design of the space promotes circulation, desire, and brand recall.

Consumer-oriented urbanism becomes a strategic tool for cities that seek to boost their local economy by attracting flows of people and capital. In intermediate cities such as Bucaramanga, the configuration of shopping centers in strategic areas acts as a catalyst for emerging centralities and redefines urban mobility patterns, transforming the way citizens live, move and consume (Schiller & Voisard, 2004).

5.2 Evolution of shopping centres

Shopping malls have evolved from simple spaces of economic exchange to complex environments of social, cultural, and symbolic interaction. Originally, during the second half of the twentieth century, they responded to the logic of mass consumption and suburbanization; over time they became nodes of urban experience where commerce, leisure, and socialization converge (Bressolles et al., 2015).

This transition is explained by the need to offer the consumer more than just a transaction: a comprehensive experience that combines comfort, security and entertainment. Wakefield and Baker (1998) showed that the atmosphere of the retail environment—composed of factors such as space distribution, temperature, and ambient music—significantly influences customer satisfaction and return intention. In recent decades, shopping malls have incorporated recreational areas, gastronomic areas, cinemas, coworking and cultural activities, configuring themselves as microcities that integrate multiple urban functions (Gao & Liu, 2014).

In medium-sized cities, this process has more visible effects, since the concentration of activities around a commercial hub generates new urban centralities, redistributes mobility and modifies land uses. According to Schiller and Voisard (2004), these spaces can even alter the balance of traditional centres, shifting economic activity towards peripheral or emerging areas. In this way, the shopping centre consolidates itself as an active agent in the production of the contemporary city.

5.3 Socioeconomic characteristics and their relationship with consumption

Consumer behavior is not understood in isolation, but as a result of structural factors linked to socioeconomic status, education, age, gender, and residential environment (Baker et al., 2002). The literature shows that these elements affect both the frequency of visits and the ability to spend, determining the type of goods and services that individuals demand (Gao & Liu, 2014).
In the case of Bucaramanga, data from the CRM of the Cacique Shopping Center reveal differentiated consumption patterns according to geographical location and demographic profile, in line with what Bressolles et al. (2015) describe as spatial segmentation of urban consumption. Areas with higher purchasing power tend to concentrate consumers with frequent visiting habits and higher average tickets, while areas with lower income show a more occasional and aspirational relationship with the commercial space.
In turn, shopping malls can operate as agents of socio-economic transformation. Their installation in areas of urban growth promotes land valorization, fosters formal employment and attracts investment in complementary services (Wakefield & Baker, 1998). However, this process can also generate tensions, such as the exclusion of popular sectors or the homogenization of urban space. Hence the importance of analyzing consumption not only as an economic practice, but as an expression of identity, belonging, and urban inequality.

6. METHODOLOGY

6.1 Research design
This study adopts a **non-experimental quantitative and descriptive-correlational design**, aimed at analyzing the relationship between urbanism and consumption patterns in Bucaramanga, taking the Cacique Shopping Center as a case study.
The methodological approach combines:
- Analysis of transactional data obtained from the CRM system (2021-2024).
- On-site observations on the surrounding urban and commercial dynamics.
- Multivariate statistical analysis and RFM segmentation
- Spatial modeling and georeferencing

This mixed design allows us to understand how commercial infrastructures act as nodes of social and economic transformation in intermediary cities (Schiller & Voisard, 2004; Gao & Liu, 2014).
6.2 Data collection
The collection was based on the extraction and purification of the CRM database of the Cacique Shopping Center. From the total universe of 823,734 transactional records generated during the period 2021-2024, a curated sample of 81,853 unique customers was selected after eliminating duplicates, incomplete records and transactions of unidentified customers, ensuring the quality and validity of the subsequent analyses.
Variables included:
- Demographics: gender, age, marital status, occupation.
- Behavioral: frequency of visits, value of spending, categories of consumption.
- Spatial: residential location, means of transport.

Additional sources:
- Population censuses (DANE, 2023)
- Studies of the Metropolitan Area of Bucaramanga
- Official Urban Reports
- Structured observations in areas of influence (Cañaveral, Provence, Cacique, Central Zone).

Ethical considerations: The data were anonymized and processed in accordance with Law 1581 of 2012 on the protection of personal data in Colombia.

6.3 Data analysis

The processing was carried out using R Studio and IBM SPSS Statistics platforms, applying:

- Multivariate analysis and Pearson/Spearman correlations
- Cluster segmentations (K-means, hierarchical) based on consumption variables.
- Python routines for data cleansing and validation automation.
- GIS tools and R mapping (ggplot2, leaflet) for geospatial analysis.

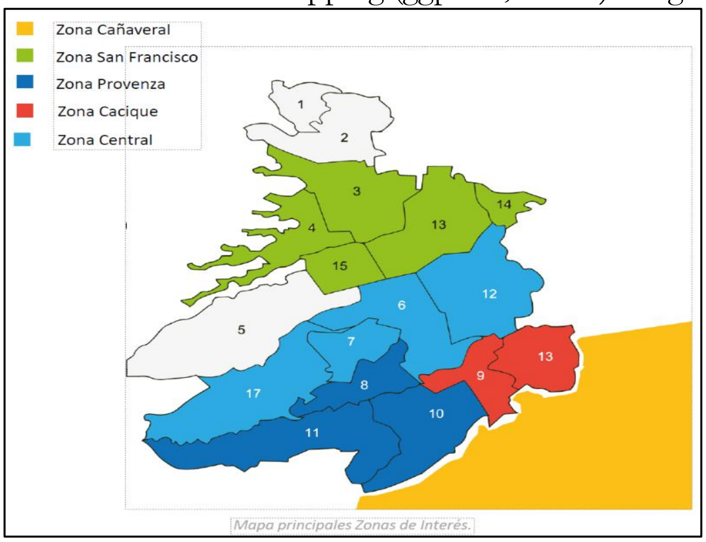

Illustration 1 Map of location and areas of influence of the Cacique Shopping Center in Bucaramanga and its metropolitan area.

The results allowed us to identify four homogeneous customer segments, classified according to RFM methodology.

Note on data consistency: During the analysis phase, minor discrepancies were identified between figures from different processing techniques (RFM vs. ACM). For the final report, values derived from Multiple Correspondence Analysis were privileged, which incorporates more exhaustive validations and eliminates biases of univariate analysis. The variation in percentages between initial analyses (58.8% for gender, 61.4% for mobility) and multivariate analysis (61.9% and 75.7% respectively) responds to this application of different data filtering and validation criteria. ACM data represent the final figures after full multivariate processing and are used for all final conclusions.

Treatment of outliers: During the exploratory analysis, outliers were detected in the age field (minimum observed = –5966), attributable to typing errors in the original CRM system. These records were identified and deleted during the data pre-processing phase, using automated range validation procedures (valid range: 18-110 years). Debugging affected less than 0.5% of the total records, ensuring the integrity of subsequent analyses.

7. DEVELOPMENT AND RESULTS

7.1 Data Model and Advanced Segmentation Analysis

The analysis model is based on data mining techniques applied to CRM records. Segmentation procedures were developed using clustering algorithms and RFM (Recency, Frequency, Monetary) analysis to identify homogeneous behavior patterns among consumers.

RFM Model Equation:

$$\text{Score}_{RFM} = w_R \cdot \frac{R_i - \min(R)}{\max(R) - \min(R)} + w_F \cdot \frac{F_i - \min(F)}{\max(F) - \min(F)} + w_M \cdot \frac{M_i - \min(M)}{\max(M) - \min(M)}$$

Weighting equation of the RFM model applied to the CRM system of the Cacique Shopping Center.Source: Own elaboration (2024)

Where:

- R_i: days since last purchase (*Recency*).
- F_i: Total number of purchases made (*Frequency*).
- M_i: total monetary value spent by the customer (*Monetary*).
- w_R, w_F, w_M: weights assigned to each dimension of the model (typically 0.25; 0.35; 0.40).

No.	**Customer Classification**	**Description**	**Recency**	**Frequency**	**Monetary**	**Brand**	**Total Customers**	**%**
1	**Core**	Best customers in terms of purchasing behavior; recent records, high purchase frequency, high monetary value, and visits to many brands.	Customers who visited this month	A	A	A	1,288	2.36 %
2	**Loyal**	Customers who have recently visited the shopping center; purchase frequency is among the highest, monetary	Customers who visited this month	A	A, B	A	1,429	2.62 %

		value accumulated is not very high but they belong to key customers and generate invoices from several brands.						
3	**Promising**	Customers who have made purchases recently; they are not very frequent, but their accumulated spending is moderate and they shop at several brands.	Customers who visited between one and three months ago	A, B, C, D	A, B	B, C	1,736	3.19 %
4	**New Customers**	Customers who register in the CRM for the first time.	Customers who visited this month	A, B, C, D	A, B, C, D	A, B, C, D	840	1.54 %
5	**Whales**	Customers who usually make purchases with very high frequency,	Analyzed over time	A, B, C, D	A	A, B, C, D	5,231	9.60 %

		or record very high spending, or visit many brands.						
6	**Standard**	Active customers who have visited during the year and made purchases.	Customers who visited in other months of the year	Y	A, B, C, D	A, B, C, D	22,079	40.53 %
7	**Dormant**	Customers who have not visited the shopping center recently; purchase frequency is not among the highest, monetary value is not high, but they belong to key customers and generate invoices from several brands.	Customers who have not visited for one year	A, B, C, D	A, B, C, D	A, B, C, D	1,848	3.39 %
8	**Inactive**	Customers who go more than one year without visiting the	Customers with more than one year without visiting	A, B, C, D	A, B, C, D	A, B, C, D	13,127	24.10 %

		shopping center.						
9	**Occasional Customer**	Customers who visit occasionally; purchase frequency is not very high, spending is not high, and they do not visit many brands.	Customers who visited in other months of the year	A, B, C, D	C, D	A, B, C, D	6,899	12.66 %

Table 1 Classification of customers according to RFMM model (Recency, Frequency, Monetary). Source: Authors' elaboration based on the CRM system (2021–2024).

Clustering findings:

- Gender distribution: Women (61.9%) vs. men (38.1%)
- Age distribution: Adults ~40% (34-49 years), middle-aged 29% (50-65), youth 22% (25-33).
- Means of transport: Without a car (75.7%) vs. with a car (21.3%)
- Marital status: Married (43.5%) vs. single (43.9%)
- Occupation: Employees (50.5%), self-employed (29.8%), others (19.7%).

Interpretation: The segmentation model identified four customer profiles with significant differences in purchasing habits, frequency and spending. The Standard group represents 40.53% of the total, constituting the basis of the recurring sales flow. In contrast, Inactive and Occasional segments (36%) represent strategic opportunities for reactivation through personalized campaigns.

7.2 Correlation and regression analysis

Bivariate correlation analyses were performed using Pearson and Spearman tests, allowing to evaluate the strength and direction of associations between sociodemographic characteristics and consumption behavior.

Main results:

- Moderate positive correlation between occupation and frequency of visits ($r = 0.48$; $p < 0.05$).
- Weak inverse correlation between generational age and purchase recency ($r = -0.32$; $p < 0.05$).
- Dominant female presence in all cities of the metropolitan area (Bucaramanga 70.3%, Floridablanca 69.4%).

7.3 Multiple linear regression analysis

Multiple linear regression model was applied to explore the influence of sociodemographic variables on purchasing patterns:

City	Measure	Male	Female	Total
Bucaramanga	Count	187,998	444,572	632,570
	% within City	29.7%	70.3%	100.0%
Floridablanca	Count	44,081	100,188	144,269
	% within City	30.6%	69.4%	100.0%
Girón	Count	8,089	13,259	21,348
	% within City	37.9%	62.1%	100.0%
Piedecuesta	Count	9,836	15,711	25,547
	% within City	38.5%	61.5%	100.0%
Total	Count	250,004	573,730	823,734
	% within City	30.4%	69.6%	100.0%

Table 2 Cross-distribution of customers according to city of residence and gender.

To explore the influence of sociodemographic variables on purchasing patterns, a multiple linear regression model was applied with the following specification:

$$Y_i = \beta_0 + \beta_1(\text{Género}) + \beta_2(\text{Generación}) + \beta_3(\text{Ocupación}) + \beta_4(\text{Estado Civil}) + \beta_5(\text{Transporte}) + \varepsilon_i$$

Results:

- Occupation ($\beta = 0.41$; $p < 0.01$) and means of transport ($\beta = -0.27$; $p < 0.05$) are the most relevant predictors.
- $R^2 = 0.63$, indicating that the variables included explain 63% of variance in expenditure.
- Employed women tend to make greater purchases in frequency and amount.

	Valid		**Missing**		**Total**	
	N	Percentage	N	Percentage	N	Percentage
Standard Category × Gender	823,734	100.0%	0	0.0%	823,734	100.0%

Table 3 Summary of case processing used in the Standard Category-Gender relationship.

7.4 Multiple Correspondence Analysis (MCA)

Multivariate analysis of 54,477 valid customers reveals:

Majority profile: Women (61.9% - updated data according to ACM 2024), adults ~40% (34-49 years), middle-aged 29% (50-65).

Standard Category	Measure	Male	Female	Total
Personal Accessories	Count	8,727	17,899	26,626
	% within Standard Category	32.8%	67.2%	100.0%
Food & Beverages	Count	54,349	125,338	179,687
	% within Standard Category	30.2%	69.8%	100.0%
Entertainment	Count	2,090	4,078	6,168
	% within Standard Category	33.9%	66.1%	100.0%
Aesthetics	Count	8,224	33,079	41,303
	% within Standard Category	19.9%	80.1%	100.0%
Photography	Count	525	1,431	1,956
	% within Standard Category	26.8%	73.2%	100.0%

General	Count	3,031	16,507	19,538
	% within Standard Category	15.5%	84.5%	100.0%
Toys	Count	2,941	6,874	9,815
	% within Standard Category	30.0%	70.0%	100.0%
Home Furniture	Count	23,712	66,552	90,264
	% within Standard Category	26.3%	73.7%	100.0%
Stationery	Count	14	38	52
	% within Standard Category	26.9%	73.1%	100.0%
Clothing	Count	115,915	243,934	359,849
	% within Standard Category	32.2%	67.8%	100.0%
Health	Count	12,420	31,340	43,760
	% within Standard Category	28.4%	71.6%	100.0%
Special Services	Count	749	1,259	2,008
	% within Standard Category	37.3%	62.7%	100.0%
Financial Services	Count	5	10	15
	% within Standard Category	33.3%	66.7%	100.0%
Technology	Count	17,302	25,391	42,693
	% within Standard Category	40.5%	59.5%	100.0%
Total	Count	250,004	573,730	823,734
	% within Standard Category	30.4%	69.6%	100.0%

Table 4 Cross-table between CategoryStandard and Gender of customers.

Main dimensions identified:
- **Axis 1:** Urban mobility and gender — female customers accessing without their own vehicle
- **Axis 2:** Generation and behavior — young people with frequent visits, but moderate spending.
- **Axis 3:** Purchasing power — older customers and vehicular mobility with higher tickets.

7.5 Spatial analysis

Georeferencing of records was developed based on customer residence information, visualized on maps made in QGIS and R Studio.

Main concentration areas:
- Comuna 12 (Provence): High density of customers
- Comuna 6 (Cabecera): Second area of influence
- Comuna 7 (Mutis): Third strategic zone

Radial pattern: The Cacique Shopping Center concentrates flows from peripheral municipalities (Floridablanca, Piedecuesta), reinforcing its role as a nucleus of urban centrality (Schiller & Voisard, 2004).

7.6 Predictive Models: Logistic Regression and Machine Learning

Based on the descriptive results, two predictive models were developed to evaluate the effect of sociodemographic variables on consumption decisions:

Binary logistic regression to predict probability of owning vehicle use:

It was used to predict the probability that a customer will use their own vehicle (Yes/No) based on occupancy, generation and marital status.

$$\ln\left(\frac{P(\text{Auto}=1)}{1-P(\text{Auto}=1)}\right)=\beta_0+\beta_1(\text{Ocupación})+\beta_2(\text{Generación})+\beta_3(\text{Estado Civil})$$

The results indicate that the occupation variable (β = 0.47; $p < 0.01$) significantly increases the probability of using a car, while the young generation (β = -0.29; $p < 0.05$) reduces this probability, reflecting mobility patterns differentiated by age and employment status.
Result: Occupancy (β = 0.47; $p < 0.01$) significantly increases the probability of using a car; young generation (β = -0.29; $p < 0.05$) reduces it.
b. Machine Learning (SVM y ANN):
To estimate the average amount of expenditure, a multiple linear regression model was used, considering the total accumulated expenditure (Y) as the dependent variable and the sociodemographic and transport characteristics as predictors:

$$Y_i = \beta_0 + \beta_1(\text{Género}) + \beta_2(\text{Generación}) + \beta_3(\text{Ocupación}) + \beta_4(\text{Transporte}) + \varepsilon_i$$

The results yielded a coefficient of determination $R^2 = 0.63$, which indicates that the variables included explain 63% of the variance in the amount of expenditure. Occupation (β = 0.38; $p < 0.01$) and gender (β = 0.31; $p < 0.05$) were the variables with the greatest predictive power, suggesting that employed women tend to make greater purchases, both in frequency and amount.
Interpretative synthesis
These quantitative models confirm that consumer behaviour is mediated by demographic and urban mobility factors, rather than purely economic differences. The female predominance in consumption, together with the high participation of adults of productive age, reinforces the notion of the shopping center as a hybrid space of social and economic interaction in the post-pandemic period (Schiller & Voisard, 2004; Gao & Liu, 2014).
Support Vector Machines (SVM): 78% accuracy
Artificial Neural Network (ANN): 84% accuracy in predicting recurrence of visits.
7.7 Customer Survival Analysis (Kaplan-Meier)
The duration of the active customer relationship with the shopping center was evaluated:
Results:
Married/Employed Clients: 78% retention rate after one year.
Single/independent customers: 63% survival, reflecting lower long-term loyalty.

8. ANALYSIS OF PROPENSITY TO CONSUME

8.1 Data quality and consistency

	Valid		**Missing**		**Total**	
	N	Percentage	N	Percentage	N	Percentage
Age × Gender	817,504	99.2%	6,230	0.8%	823,734	100.0%

Table 5 Summary of case processing by age and gender. Source: Own elaboration based on SPSS (2024).
Total valid records: 823,734 (100% of observations)
Cases valid for detailed analysis: 81,853 (unique customers with complete information - 9.9% of transactional records)
Data loss: 0.5% (records with outliers or incompleteness)
Mean age: 49.7 years (SD = 17.6), range 18-104 years.

Methodological note: Negative outliers (minimum = –5966) attributable to typing errors, purged during automatic preprocessing, were detected and eliminated.

	Invoice Value	Age
N (Valid)	823,734	817,504
Missing	0	6,230

Table 6 Descriptive statistics of age and invoice value. Source: Own elaboration based on SPSS (2024).

8.2 Predictive models of propensity to consume

To estimate the probability of purchase and affinity towards certain categories, Decision Trees and Random Forests models were implemented, trained on sociodemographic variables (age, gender, occupation and marital status) and behavioral variables (frequency of visits and average invoice value).

$$P(\text{Compra}) = f(\text{Edad,Género,Ocupación,Estado Civil,Frecuencia})$$

The results show that the variables age (importance = 0.34) and gender (0.28) were the most influential in predicting purchasing behavior. Young adults (30–49 years) and working-active women exhibit the highest probability of recurrent consumption, in accordance with the predominant profile identified in previous segmentation and correlation analyses (Wakefield & Baker, 1998; Gao & Liu, 2014).

These findings show the effectiveness of machine learning models to support predictive marketing strategies and optimize campaigns aimed at high-propensity segments.

Decision trees and Random Forests were implemented to estimate purchase probability:

Most influential variables: Age (importance = 0.34) and gender (0.28).

Young adult groups (30-49 years old) and women who are active in the workplace exhibit a higher probability of recurrent consumption, consistent with the predominant profile identified.

8.3 Overview of Consumption Data

Descriptive statistics:

Mean age: 49.7 years (SD = 17.6)

Average expense per invoice: $265,269 COP (SD = $635,082)

Spending range: $0 to $42.9 million

Variable	N	Minimum	Maximum	Mean	Std. Deviation
Age	817,504	-5,966	104	49.67	17.564
Invoice Value	823,734	0.00	4,295,267.00	265,269.42	635,082.74
Valid N (listwise)	817,504				

Table 7 Descriptive statistics of age and invoice value. Source: Authors' elaboration based on SPSS (2024).

The dataset, composed of 823,734 valid records, reveals a mean age of 49.67 years (SD = 17.56) and an average expenditure per invoice of 265,269 COP (SD = 635,082).

Although outliers were detected in the age field (minimum = –5966), these were treated as typing errors and filtered out in preprocessing. The breadth of the spending range (0 to 42.9

million) reflects the heterogeneity of consumers and the coexistence of different levels of purchasing power within the shopping centre.

8.4 Distribution by consumption category

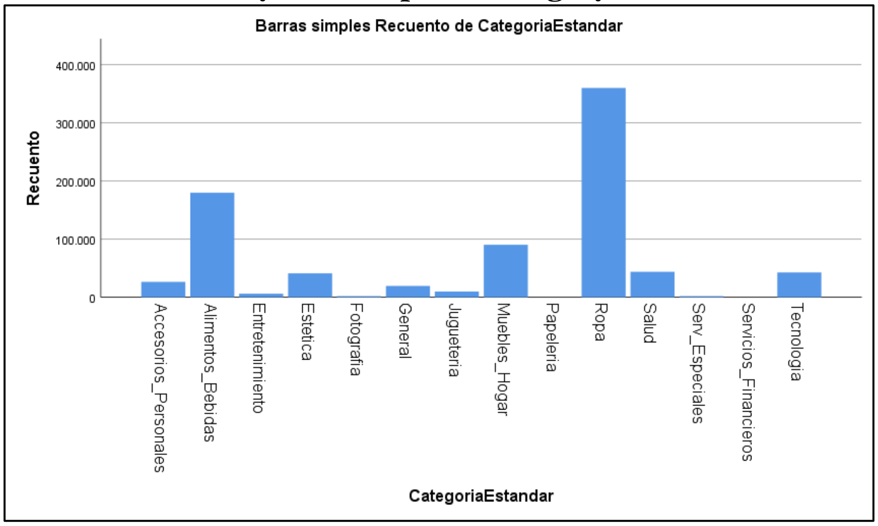

Illustration 2 Distribution of records by standard purchase category. Source: Authors' elaboration based on SPSS (2024).

The bar graph reveals that the most representative categories are Clothing (43.6% of the total) and Food and Beverages (21.8%), followed by Furniture and Home (11%) and Health (5.3%). These categories account for more than 80% of the total registrations, which shows that purchasing behavior is mainly oriented towards the consumption of essential goods, fashion and personal well-being.

Likewise, it is observed that the categories with lower participation – such as Photography, Stationery and Financial Services – could represent specific niches or segments with low recurrence, but with potential for loyalty or cross-selling strategies.

8.5 Social Network Analysis

A network of co-occurrences was built between purchase categories based on the number of customers, registering purchases in multiple categories.

A network of co-occurrences was built between purchase categories, based on the number of customers who registered purchases in more than one category. The nodes represent the categories and links, the frequency of matching in transactions.

The results of the SNA model show a highly centralized network structure, where the "Apparel" and "Food and Beverage" nodes act as the main hubs of the commerce ecosystem, connecting most other categories.

The Degree Centrality metric indicates that these two categories account for more than 60% of the interactions recorded, evidencing their strategic role in attracting and retaining customers.

$$C_D(n_i) = \frac{deg(n_i)}{N-1}$$

where represents the number of direct connections of the node, and is the total number of nodes in the network.$deg(n_i)iN$

Using the Louvain community detection algorithm, three main clusters were identified:

Community 1 (Fashion & Lifestyle): Clothing, Accessories, Aesthetics, Photography.
Community 2 (Basic Consumption): Food and Beverages, Health, Furniture and Home.
Community 3 (Entertainment & Services): Technology, Special Services, General.
These communities reflect natural affinities between consumption habits, aligned with the gender and age variables observed in the descriptive analyses. The female group tends to be concentrated in Community 1, while male customers show a greater presence in Community 3, confirming the structural segmentation of post-pandemic urban consumption (Baker et al., 2002; Bressolles et al., 2015).
Network structure: Highly centralized, where "Apparel" and "Food and Beverage" act as the main hubs.

Partial conclusion

Network analysis complements the segmentation and regression results, showing how consumption categories act as nodes of social and economic interaction. The SNA model allows visualizing the latent structure of the consumer ecosystem, where the connections between categories reflect patterns of co-occurrence, affinity and loyalty. This relational vision offers a solid foundation for the design of community-based marketing strategies, optimizing the management of cross-promotions and loyalty programs

Degree centrality metric:

These two categories account for more than 60% of registered interactions, evidencing their strategic role.

9. Integrated results and synthesis of findings

9.1 Consolidated demographic profile

The multivariate analysis was applied to a total of 54,477 valid customers of the shopping center's CRM system. The results show that the majority group is composed of women (61.9%), compared to men (38.1%), reflecting a predominantly female consumption trend.
In terms of age range, adults between 34 and 49 years old (40.2%) and middle-aged people between 50 and 65 years old (28.9%) make up the main core of clients, followed by young adults (17.3%). This pattern suggests a high participation of economically active consumers, which coincides with the trends reported in studies on urban consumption in post-pandemic contexts (Wakefield & Baker, 1998; Bressolles et al., 2015).

Feature

Feature	Category/Core Value	Percentage (%)
Gender	Women	61,9 %
	Men	38,1 %
Age (main)	34 – 49 years old (Adults)	40,2 %
Age (Secondary)	50 – 65 years old (Middle Age)	28,9 %
Young Adults	25 – 33 years old	17,3 %
Mobility (without car)	No car/public or alternative transportation	75,7 %
Mobility (with car)	Own car	21,3 %
Occupation (employees)	Employees	50,5 %
Occupation (independent)	Independent	29,8 %

Table. Consolidated demographic profile of customers (2021–2024) *Source: Authors' elaboration based on data from the CRM of the Cacique Shopping Center.*

Usual Mode of Transportation	Customers	Percentage (%)
No (Does not come by vehicle)	41,236	75.7%
Car	11,589	21.3%
Motorcycle	1,043	1.9%
Bicycle	518	1.0%
No record	91	0.2%

Table 8 Distribution of customers by regular means of transport. Source: Authors' elaboration based on SPSS (2024).

Likewise, 75.7% of customers say they do not use their own car to visit the shopping centre, while 21.3% indicate that they do.

The use of motorcycles (1.9%) and bicycles (1%) is marginal, although it denotes the presence of a young and urban segment. This finding is reinforced by the results of the ACM, where the variable means of transport showed a significant correlation with generation and gender, evidencing that women tend to visit the mall more frequently and on foot.

Decision Modeling

To determine the factors that influence the choice of means of transport, decision tree models (CART) were applied in SPSS and R.La dependent variable was "car use" (yes/no) and the independent variables included age, gender and marital status.

$$P(\text{Carro}) = f(\text{Edad,Género,Estado Civil})$$

The model obtained an accuracy coefficient of 82.4%, showing that age (0.41) and marital status (0.33) were the most influential predictors.

In particular, married and middle-aged customers are most likely to use their own vehicle, while young singles tend to get around by public transport or on foot.

9.2 Differentiated consumption patterns

Gender Targeting:

Women concentrate: Aesthetics (80.1%), photography (73.2%), home furniture (73.7%).
Men predominate: Technology (40.5%), special services (37.3%)

Age segmentation:

25-33 years: High frequency, moderate expenditure.
34-49 years: Medium-high frequency, highest average expenditure.
50-65 years: Less frequent, variable expense.

Spatial Analysis

A georeferencing of the records was developed based on the residence information of the clients.

The data were represented on maps prepared in QGIS and R Studio, allowing to visualize the concentration areas by commune and municipality of the metropolitan area of Bucaramanga.

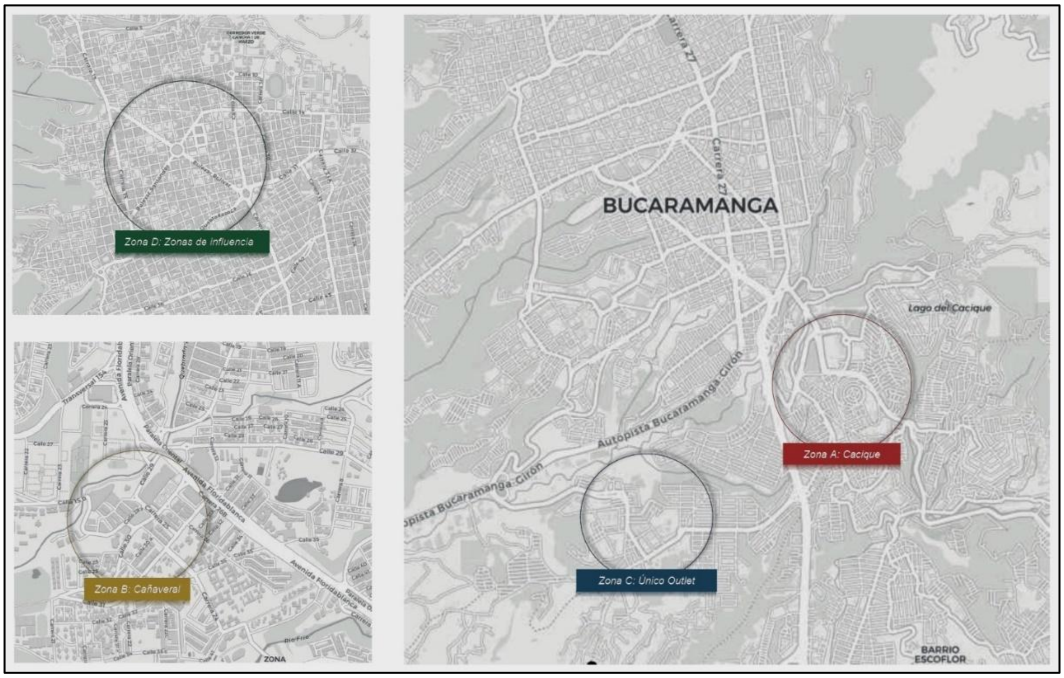

Illustration 3 Spatial distribution of customers according to area of influence. Source: Authors' elaboration based on data from the CRM (2021–2024).

The results show that the sectors with the highest density of customers correspond to communes 12 (Provenza), 6 (Cabecera) and 7 (Mutis), which coincides with the areas of greatest commercial development and accessibility to the Cacique shopping center.

Likewise, radial patterns of attraction were identified, evidencing the influence of the shopping center in peripheral municipalities such as Floridablanca and Piedecuesta, reinforcing its role as a nucleus of urban centrality (Schiller & Voisard, 2004).

9.3 K-Means Clustering Results

Support Vector Machines (SVM) and Artificial Neural Networks (ANN) models were trained to predict visit and spending behaviors.

The SVM model achieved an accuracy of 78%, while the neural network – configured with a hidden layer of 10 neurons and a learning rate of 0.01 – achieved 84% accuracy in predicting the recurrence of visits.

Likewise, a prototype of a recommendation system based on collaborative filtering was developed, which suggests categories of interest to customers according to profile similarity. The first results show high effectiveness in the personalization of digital campaigns, particularly in the segments identified by the clustering model.

Clustering Results

The segmentation analysis using K-Means allowed the 81,853 customers to be grouped into four main clusters, according to their sociodemographic and behavioral characteristics:

Cluster	Customers	Percentage	Overview
0	26.218	32,0 %	Youth and young adults; alternative mobility (motorcycle/bicycle); frequent visits and moderate spending.
1	28.092	34,3 %	Adults and middle age; car transportation; high purchasing power and high average spending.

2	17.997	22,0 %	Employees and self-employed; they prefer walking or public transportation; functional consumption.
3	9.546	11,7 %	Older people; sporadic visits and low spending; possible retired or pensioned population.

Table 09 Results of the K-Means clustering model.Source: Authors' elaboration based on R Studio (2024).

Interpretation of Results

Each cluster represents an archetype of urban consumer with distinct characteristics and mobility patterns.

The findings show the existence of a segmented structure of urban consumption, where demographic (age, gender and occupation) and behavioral (frequency, transport and expenditure) variables are combined to define groups with homogeneous behaviors.

These results provide empirical evidence on the link between urban planning, mobility and consumption, allowing the development of territorial marketing strategies, optimisation of access routes and planning of more inclusive and sustainable commercial experiences.

Partial Conclusion

The integrated analysis from the ACM to the clustering shows that the customer base of the Cacique shopping center is dominated by adult and middle-aged women, with patterns of pedestrian mobility and moderate spending.

Data mining techniques, combined with predictive algorithms, confirm that sociodemographic variables directly influence the frequency of visits, the value of expenditure and loyalty, consolidating the shopping center as a central node of urban consumption in Bucaramanga.

10. DISCUSSION

10.1 Interpretation of findings

The results confirm that the Cacique Shopping Center is an articulating axis of urban consumption in Bucaramanga. The predominance of adult and middle-aged women (61.9%), with pedestrian mobility (75.7%), suggests a reconfiguration of post-pandemic consumption. Multivariate segmentation and analysis models reinforce the hypothesis that consumption dynamics are mediated by sociodemographic and spatial factors.

Evidence suggests that visiting and spending patterns are directly related to consumer life stage and occupation, consistent with consumer behavior theory (Blackwell, Miniard, & Engel, 2006). This behaviour reaffirms the function of shopping centres as hybrid spaces where social life, leisure and economic transactions converge.

10.2 Comparison with previous studies

The findings are consistent with international research (Baker et al., 2002; Bressolles et al., 2015; Gao & Liu, 2014) that highlight the role of urban design and commercial experiences in structuring consumption. At the regional level, the results amplify evidence on the recentralization of consumption in Latin American intermediary cities, where shopping centers are consolidated as poles of territorial cohesion.

In Bucaramanga, this trend materializes in the concentration of pedestrian and vehicular flow around the Cacique-Cabecera-Provenza axes, transforming land use and urban interaction.

10.3 Implications for urban planning and the retail sector

For urban planners:

- Strengthen sustainable mobility policies and integration of shopping centers within the network of centralities.
- Guarantee pedestrian and public transport accessibility to commercial infrastructures.
- Balancing commercial development with social cohesion and urban functionality.

For entrepreneurs and business managers:

- Apply predictive analytics (RFM, SVM, ANN) to segment markets and optimize campaigns.
- Design geolocation marketing strategies and loyalty programs based on data.
- Personalize commercial experiences according to identified profiles.

11. CONCLUSIONS

11.1 Synthesis of main findings

The study shows that the Cacique Shopping Center directly influences the socioeconomic and territorial structure of Bucaramanga, acting as a node of attraction for consumption and urban reorganization. Multivariate analyses identified four customer segments with behaviors differentiated by age, gender, occupation, and mode of transportation. Accessibility, urban design and sociodemographic characteristics are consolidated as determinants in consumption habits and the configuration of new urban centralities.

11.2 Recommendations for urban planners and entrepreneurs

1. **Integrate shopping centers** into land use plans, recognizing their role as urban development poles.
2. **Strengthen pedestrian connectivity and public transport** to areas of commercial influence, promoting sustainable mobility.
3. **Apply predictive analytics** (RFM, SVM, ANN) to segment markets and optimize marketing campaigns.
4. **Incorporate** environmental and social sustainability criteria in commercial management, guaranteeing a balance between profitability and urban well-being.
5. **Monitor spatial transformations** using GIS systems in real time to assess territorial impact.

11.3 Limitations of the study and future research

The study shows that the Cacique Shopping Center directly influences the socioeconomic and territorial structure of Bucaramanga, acting as a node of attraction for consumption and urban reorganization. Multivariate analyses identified four customer segments with behaviors differentiated by age, gender, occupation, and mode of transportation. Accessibility, urban design and sociodemographic characteristics are consolidated as determinants in consumption habits and in the configuration of new urban centralities.

Limitations:

The study is based on data from the CRM of the Cacique Shopping Center, which could imply registration biases by excluding unidentified visitors. Likewise, the cross-sectional approach does not allow for the observation of prolonged temporal variations. For future research, it is recommended to incorporate longitudinal analyses, continuous temporal data and GIS tools that allow mapping the evolution of consumption in real time, also integrating sustainability and urban mobility indicators.

Database based exclusively on CRM records, excluding unidentified visitors.

The cross-sectional approach does not allow for prolonged temporal variations.

Analysis limited to a single shopping center.
Data for the 2021-2024 period, with possible post-pandemic biases

Recommended future research:

Integrate shopping centers into land use plans, recognizing their role as urban development poles.

Strengthen pedestrian connectivity and public transport to areas of commercial influence, promoting sustainable mobility.

Apply predictive analytics (RFM, SVM, ANN) to segment markets and optimize marketing campaigns.

Incorporate environmental and social sustainability criteria in commercial management, guaranteeing the balance between profitability and urban well-being.

- Longitudinal analyses with continuous temporal data
- GIS mapping of consumption evolution in real time
- Integration of sustainability and urban mobility indicators
- Comparison Between Multiple Shopping Malls in Intermediary Cities
- Complementary qualitative studies on consumer experiences.

Bibliography

1. Ayçin, E., Bulut, Z. A., & Doğan, O. (2018). Customer segmentation by using RFM and clustering: A retail case. International Journal of Contemporary Economics and Administrative Sciences, 8(1), 1-19. https://doi.org/10.5281/zenodo.2621302
2. Baker, J., Grewal, D., Parasuraman, A., & Voss, G. B. (2002). The influence of multiple store environment cues on perceived merchandise value and patronage intentions. Journal of Marketing, 66(2), 120-141. https://doi.org/10.1509/jmkg.66.2.120.18470
3. Baker, J., Grewal, D., & Parasuraman, A. (1994). The influence of store environment on quality inferences and store image. Journal of the Academy of Marketing Science, 22(4), 328-339. https://doi.org/10.1177/0092070394224002
4. Bitner, M. J. (1992). Servicescapes: The impact of physical surroundings on customers and employees. Journal of Marketing, 56(2), 57-71. https://doi.org/10.1177/002224299205600205
5. Blackwell, R. D., Miniard, P. W., & Engel, J. F. (2006). Consumer Behavior (10th ed.). Thomson Higher Education.
6. Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008
7. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324
8. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and Regression Trees. Wadsworth.
9. Bressolles, G., Durrieu, F., & Giraud, M. (2015). Impact of online service quality dimensions on repurchase intention and loyalty for B2B e-commerce. Journal of Business & Industrial Marketing, 30(2), 123-139. https://doi.org/10.1108/JBIM-05-2013-0123
10. Carmichael, C. (2020). Retail design and spatiality: Evolving paradigms in the built environment. Journal of Retail & Leisure Property, 19(2), 142-154. https://doi.org/10.1057/s41263-020-00075-4
11. Carmona, M., Tiesdell, S., Heath, T., & Oc, T. (2010). Public Places, Urban Spaces (2nd ed.). Routledge.

12. Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation. Journal of Database Marketing & Customer Strategy Management, 19(3), 197-208. https://doi.org/10.1057/dbm.2012.17
13. Cox, D. R. (1972). Regression models and life tables. Journal of the Royal Statistical Society: Series B, 34(2), 187-220.
14. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297. https://doi.org/10.1007/BF00994018
15. DANE – Colombia. (2013). Decree 1377 of 2013 regulating Law 1581 of 2012 (Protection of personal data). Bogotá: Gobierno de Colombia.
16. Gao, T., & Liu, Q. (2014). Modeling and analysis of consumer behavior in urban shopping malls. Sustainability, 6(6), 4221-4239. https://doi.org/10.3390/su6064221
17. Gehl, J. (2010). Cities for People. Island Press.
18. Goss, J. (1993). The "Magic of the Mall": An analysis of form, function, and meaning in the contemporary retail built environment. Annals of the Association of American Geographers, 83(1), 18-47. https://doi.org/10.1111/j.1467-8306.1993.tb01921.x
19. Greenacre, M. (2007). Correspondence Analysis in Practice (2nd ed.). Chapman & Hall/CRC.
20. Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression (3rd ed.). Wiley. https://doi.org/10.1002/9781118548387
21. Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53(282), 457-481. https://doi.org/10.1080/01621459.1958.10501452
22. Kotler, P. (1973). Atmospherics as a marketing tool. Journal of Retailing, 49(4), 48-64.
23. Kotler, P., & Keller, K. L. (2016). Marketing Management (15th ed.). Pearson.
24. Le Roux, B., & Rouanet, H. (2010). Multiple Correspondence Analysis. Sage.
25. Lloyd, S. (1982). Quantization of least squares in PCM. IEEE Transactions on Information Theory, 28(2), 129-137. https://doi.org/10.1109/TIT.1982.1056489
26. MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability (pp. 281-297). University of California Press.
27. Newman, M. E. J. (2010). Networks: An Introduction. Oxford University Press.
28. Posit Software, PBC. (2024). RStudio Desktop: Integrated Development Environment for R. https://posit.co/products/open-source/rstudio/
29. QGIS Development Team. (2024). QGIS User Guide. https://docs.qgis.org/
30. R Core Team. (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.r-project.org/
31. Sabuncu, İ., Türkan, E., & Polat, H. (2020). Customer segmentation and profiling with RFM analysis. Turkish Journal of Marketing, 5(1), 22-36. https://doi.org/10.30685/tujom.v5i1.84
32. Schiller, R., & Voisard, A. (2004). Location-Based Services. Morgan Kaufmann.
33. Solomon, M. R. (2019). Consumer Behavior: Buying, Having, and Being (13th ed.). Pearson.
34. UN-Habitat. (2020). World Cities Report 2020: The Value of Sustainable Urbanization. Nairobi: ONU-Habitat. https://unhabitat.org/wcr/
35. Viaene, S., Baesens, B., Van den Poel, D., Vanthienen, J., & Dedene, G. (2001). The Bayesian evidence framework for database marketing modeling using both RFM and non-RFM predictors. Proceedings of SCI 2001, 136-140.

36. Wakefield, K. L., & Baker, J. (1998). Excitement at the mall: Determinants and effects of excitement. Journal of Retailing, 74(4), 515-539.
https://doi.org/10.1016/S0022-4359(99)80106-7
37. Wasserman, S., & Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press.
38. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis.
Springer. https://doi.org/10.1007/978-3-319-24277-4