Journal of Posthumanism
2025
Volume: 5, No: 5, pp. 1195–1217
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
posthumanism.co.uk

DOI: https://doi.org/10.63332/joph.v5i5.1433

Classification and Typification of Costeño Cheese in the Colombian Caribbean Through Multivariate Analysis of Physico-chemical Parameters: Base Study for the department of Magdalena

Edwin Causado-Rodriguez[1], Jhon Jairo Vargas Sanchez[2], Andrés Mauricio Peñaloza Fernández[3]

Abstract

The analysis of the physicochemical parameters of Costeño cheese in the Magdalena department of the Caribbean region of Colombia presents unique challenges due to the presence of variability in these characteristics, influenced by climate, production technique, livestock type, and feed, among others, which in turn impact the sensory aspects of this type of dairy product. This study employs advanced statistical techniques that allow for a better understanding and classification of cheeses based on their properties, such as Principal Component Analysis (PCA), Cluster Analysis (CA) or hierarchical clustering, and K-Means clustering, facilitating dimensionality reduction, pattern identification, and data grouping. In addition, multivariate statistical analysis, PCA and CA, were used to determine the behavior of the study variables and characterize Costeño cheese from a group analysis. As a result, it was confirmed that multivariate statistical analysis can be used to improve the understanding of the quality of a product such as fresh cheese.

Keywords: *Cheese Costeño, Designation of Origin, Multivariate Analysis, Physicochemical Parameters, Typification.*

Introduction

The development of livestock farming in the Colombian Caribbean dates to the expansion of livestock lands in territories previously occupied by indigenous people and the adaptation to various climatic and topographic conditions (Mendoza-Ramos, 2009). During the 17th century, the adoption of European milking and preservation techniques significantly improved cheese production, transforming it into a salty product and consolidating it as a relevant business activity. In the 19th century, the coastal cheese trade expanded to the Caribbean islands, establishing a productive chain that sustained local economies (Narváez, La Torre & Ortiz, 1965; Elías-Caro & Vidal-Ortega, 2010; Causado-Rodriguez, E; Fonseca-Tovar, J & Galindo-Montero, A, 2023a).

In the 20th century, the mixing of livestock breeds and the implementation of European and American production techniques led to a significant increase in the quality and quantity of dairy

[1] Industrial Engineer, MSc Economics of the Environment and Natural Resources. Mg. Integrated Management Systems. PhD in Administration. Staff Teacher at the Faculty of Engineering of the University of Magdalena, Email: ecausado@unimagdalena.edu.co, https://orcid.org/0000-0001-6258-2307.
[2] Industrial Engineer. Master's in operations research and Statistics. PhD in Engineering. Staff Teacher at the Faculty of Engineering of the University of Magdalena, Email: jvargass@unimagdalena.edu.co, https://orcid.org/0000-0003-3799-9850
[3] Industrial Engineer, Universidad del Magdalena, Co-investigator of the "Queso Costeño" Project, Resource Management for Development Research Group - GIGRD, Santa Marta, Colombia, Email: apenalozaf@unimagdalena.edu.co, https://orcid.org/0009-0003-0570-0727.

products, including Costeño cheese (Sánchez-Mejía, 2015; 2019; 2022). This development culminated in 2011 with obtaining the first designation of origin for a fresh cheese such as Queso del Caquetá, recognizing the quality and authenticity of these traditional products (Lesmes Ramos, 2017). Appellations of origin have proven to be effective tools to add value in the agricultural sector, highlighting the main events and transformations in the coastal cheese supply chain.

Designation of Origin of Costeño Cheese

Costeño cheese, an emblematic product of the Caribbean region of Colombia, stands out for its distinctive characteristics that have positioned it as a cultural and gastronomic symbol. Its relevance lies in the tradition and quality that it has maintained over the years, making it recognizable both locally and nationally and internationally (Causado-Rodriguez, E; Romero-Borja, I & Galindo-Montero, A. (2023b).

Prieto, Sánchez and Rodríguez (2021) emphasize that the designation of origin is crucial for the positioning of the Costeño cheese brand in the region. According to their research, the designation of origin contributes to the homogeneity and recognition of the product, allowing consumers to identify and value quality based on tradition and regional production practices.

Costeño cheese can be classified into several categories depending on its production process. For example, acidic fresh cheese and non-acidic fresh cheese are produced in the municipality of Plato, Magdalena, where a double cream mozzarella type cheese is made, and in the municipality of Santa Ana, Magdalena (Causado-Rodriguez, E; Galindo-Montero, A & Peñaloza-Fernández, A. (2023c). This cheese is considered a gastronomic emblem of the Caribbean Coast due to its roots in various social classes. It is a staple food that is consumed both individually and accompanied, thanks to its characteristics that allow it. In addition, it has become an important input for the manufacture of many typical foods of the region (Gutiérrez Castañeda et al., 2017).

Theoretical Framework

Distinctive signs such as collective marks, certification marks and geographical indications are strategies used to certify the quality and origin of certain agricultural products (Gutiérrez et al., 2017; Garcia, 2017). According to the literature reviewed, these stamps, including the designation of origin and indications of provenance, have a collective interest. The designation of origin not only protects the product but also preserves the culture and tradition of the place of manufacture, generating economic and social benefits for the local population (Errázuriz Tortorelli, 2010; Granados, 2012). In this way, the designation of origin offers competitive advantages to the producer and assures the consumer a unique product with a seal of quality, covering the following aspects:

Natural Factors That Affect Product Quality

The bromatological composition of Costeño cheese is influenced by several natural factors, mainly environmental conditions and agricultural practices in the Caribbean Region of Colombia.

Climate and Precipitation

The Caribbean Region is in the Colombian low tropics, with average temperatures between 24°C and 30°C and annual rainfall that varies between 1,000 and 1,500 mm, concentrating mainly between May and November. The semi-arid conditions of areas such as La Guajira, with rainfall less than 300 mm per year, also influence agricultural and livestock production in the area (DANE, 2007). These climatic conditions directly affect the type and quality of forage available to livestock, impacting on the quality of the milk produced and, consequently, the cheese.

Grazing and Livestock Systems

The Colombian Caribbean allows the use of various grazing systems such as continuous, rotational and silvopastoral. The latter is especially beneficial, as it combines grasslands with trees and shrubs, creating favorable microclimates that improve forage availability and soil health (Causado, Galindo, & Peñaloza, 2023b).

Types of Livestock

The predominant livestock in the region, such as the Zebu Brahman and its crosses with European breeds (Holstein), are well adapted to the conditions of the tropics. These breeds are resistant to high temperatures and pests and are efficient in producing milk in hot climates. The Brahman Zebu breed is notable for its rapid growth, drought resistance and longevity, making it an ideal choice for the region (DANE, 2007)

Soils and Vegetation

The diversity of soils in the Caribbean Region, which includes less fertile sandy soils in coastal areas and nutrient-rich soils in inland areas, plays a crucial role in agricultural production.

Designation of Fresh Cheeses According to the Regulations

In Colombia, the designation and classification of fresh cheeses are regulated by the Colombian Technical Standard (NTC) 750 of 2009. This standard establishes that the classification of fresh cheeses is based mainly on the percentage of moisture and fat they contain. According to this regulation, cheeses are categorized as fresh when they meet the specific requirements for moisture and fat, thus guaranteeing their quality and compliance with the standards established for these dairy products (NTC 750, 2009).

In the case of Costeño cheese, it is recommended that the sodium content should be monitored, with limits established at 400 mg per 100 g in solids and 150 mg per 100 mL in liquids; Exceeding these limits requires a “HIGH SODIUM” warning seal. Added sugars must be below 10 g per 100 g in solids and 5 g per 100 mL in liquids to avoid the "HIGH IN ADDED SUGARS" seal. Finally, the saturated fat content should not exceed 4.0 g per 100 g in solids and 3.5 g per 100 mL in liquids; otherwise, "HIGH IN SATURATED FAT" should be indicated (Ministry of Health and Social Protection, 2021).

Principal Component Analysis

Regarding the analysis of food products, including cheeses and other classified agri-food products, statistical techniques such as Principal Component Analysis (PCA) and cluster analysis or similarity analysis are being used, which facilitate their classification. PCA has proven useful in dimensionality reduction and identifying patterns in complex data related to cheese quality.

Jolliffe (2002) highlights that PCA is a powerful technique for transforming original variables into principal components, facilitating the interpretation of complex data. In cheese studies, it has been used to classify and compare different types of cheese based on their physical and chemical characteristics. Likewise, conglomerate analysis or clustering, according to Ward (1963), allows data to be grouped into clusters that maximize homogeneity within each group and heterogeneity between groups. This technique has been applied in the analysis of food products to identify similar groups of products according to specific characteristics, such as the texture and flavor of cheese. For their part, similarity analysis and visualization using dendrograms are also valuable tools in food data analysis. Kaufman and Rousseeuw (1990) describe how hierarchical clustering and dendrograms can illustrate the structure and relationships between different samples, providing a clear view of how samples are grouped based on their characteristics Castell-Palou et al. 2010; Molina, 2019).

Methodology

The objective of this study is to determine the relationship between the bromatological variables of Costeño cheese, by studying the physicochemical variables of samples sent to the laboratory. Cheese samples were taken from 55 producers during the years 2022, 2023 and 2024, with which a database was built for analysis. The cheese samples, weighing 375 to 500 grams [g] as specified in NTC 750 of 2009, were sent to the laboratory for their respective physicochemical analysis (see table 1).

Designation according to consistency Moisture without fat	**(HSMG) [% m/m]**
Extra hard	<50,0
Hard	50-55
Firm/Semi-hard	56-68
Soft	> 68
Designation according to its fat content Fat matter in dry matter	(GES) [% m/m]
Extra fat	≥ 60,0
Fatty	≥45,0 - < 60,0
Semi-fat	≥25,0 - < 45,0
Semi-skimmed	≥10,0 - <25,0
Skim	< 10,0

Table 1. Designation of Cheese According to Its Consistency and Fat Content

Source: NTC 750 (2009).

Multivariate analysis is traditionally used in the quality evaluation of foods, as well as for wines or other products. PCA is applied to summarize the immense amount of data with minimum loss of information for the classification of samples (Vergara, A., & Silva, M., 2007).

In this sense, a multivariate statistical analysis was carried out with the quantitative and qualitative variables obtained in the project, to identify characteristics of groups based on the designation of fresh cheese for Costeño cheese. The quantitative variables are Fat [g/100g], Moisture [g/100g], Salt [g/100g], Protein [g/100g], Total Solids [g/100g], Moisture without Fat Matter (HSMG) [% m /m] and Fatty Matter in Dry Extract (GES) [% m/m]. The last two variables are obtained from the designation of the NTC 750 of 2009 as seen in Table 6. The qualitative variables of the study are the subregions (south, center and north), the municipalities where the samples were taken, the designation of HSMG and GES.

NTC 750 of 2009 designates its classification according to the HSMG and GES content, which are obtained through the equations (1) y (2):

$$HSMG = \frac{\%Moisture}{100\% - \%Fat} . 100 \tag{1}$$

$$GES = \frac{\%Fat}{100\% - \%Moisture} . 100 \tag{2}$$

The methodology for the analysis includes data collection, the use of R for statistical analysis, and the application of PCA, hierarchical clustering and K-Means clustering to identify patterns and similarities between the types of coastal cheeses from the Colombian Caribbean. The variable 'Total Solids' was excluded from the PCA to focus the analysis on the most relevant characteristics. Hierarchical analysis was performed using the Ward.D2 method and Euclidean distances, while K-Means clustering was used to define additional groups.

Study Area

In the municipalities of El Banco, Guamal, Santa Ana, Nueva Granada, Plato, Ariguaní, Fundación, Pivijay and Ciénaga, belonging to the department of Magdalena, the analysis and identification of similarities of coastal cheese in the Colombian Caribbean was carried out. The study focused on analyzing the key characteristics in the composition of this food product of dairy origin. To determine the key characteristics of Costeño cheese, NTC 750 of 2009 was used, with the purpose of grouping similarities.

In addition, the department was analyzed by subregions, which are divided into three: South, North and Center. Each subregion includes three municipalities. For the southern subregion, El Banco, Guamal and Santa Ana correspond; for the central subregion, Plato, Nueva Granada and Ariguaní; and for the northern subregion, Fundación, Pivijay and Ciénaga (see Figure 1).

Data Collection

A total of 54 data from Costeño cheeses were collected in the Magdalena region, Colombia, during the years 2022, 2023 and 2024. The data collection covered 9 municipalities and 3 subregions. The data was obtained through field visits to local producers, interviews and the collection of samples for physical-chemical analysis. These data included characteristics such as fat, moisture, salt, total solids protein, HSMG and GES as relevant parameters for the analysis.

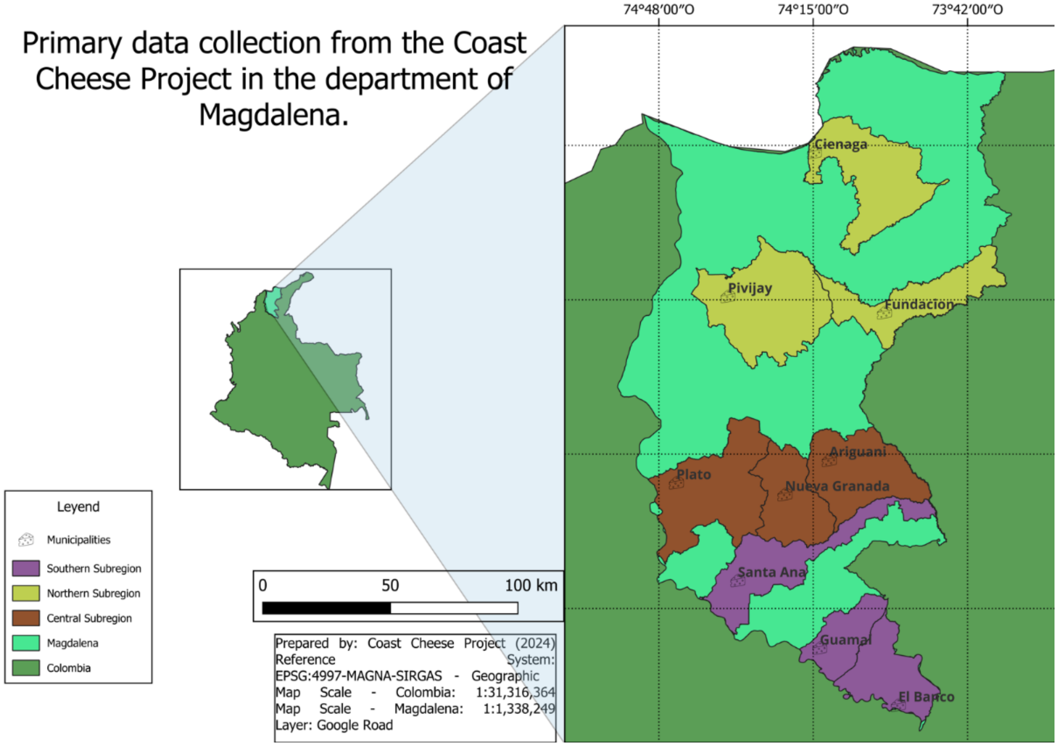

Figure 1. Sampling Area by Subregions of Coastal Cheese for the Department of Magdalena.

Source: The authors.

The data from the physicochemical analysis for the implementation of multivariate statistical techniques (PCA and Cluster Analysis) were consolidated in a database with the characteristics of Costeño cheese in the different subregions of the Magdalena department. Figures and tables were prepared to visualize the variability in the properties of the cheese and facilitate the interpretation of the results (See Table 2).

Sample	Fat	Humidity	Salt	Protein	Total Solids	HSMG	GES
MSQ1_1	26,59	45,87	2,76	20,38	54,13	62,48	49,1
MSQ2_1	27,13	44,83	3,71	18,78	55,17	61,52	49,2
MSQ4_1	26,86	48,04	3,36	16,8	51,96	65,68	51,7
MSQ6_1	25,38	46,45	2,38	20,89	53,55	62,25	47,4
MSQ8_1	21,74	53,32	2,45	17,83	46,68	68,13	46,6
MCQ11_1	24,17	49,12	2,77	19,06	50,88	64,78	47,5
MCQ14_1	27,31	47,93	2,23	18,68	52,07	65,94	52,4
MCQ16_1	26,89	46,22	3,31	18,85	53,78	63,22	50
MCQ17_1	28,94	42,09	2,85	21,72	57,91	59,23	50
MNQ21_1	26,26	48,82	2,64	18,01	51,18	66,21	51,3
MNQ24_1	27,6	46,27	2,8	18,32	53,73	63,91	51,4

Sample	Fat	Humidity	Salt	Protein	Total Solids	HSMG	GES
MNQ25_1	22,81	53,01	1,86	17,78	46,99	68,67	48,6
MNQ27_1	23,01	51,14	2,86	18,15	48,86	66,42	47,1
MSQ2_2	25,48	46,52	3,97	18,1	53,48	62,43	47,6
MSQ3_2	26,4	49,4	1,81	18,76	50,6	67,12	52,2
MSQ4_2	26,81	46,57	5,07	14,93	53,43	63,63	50,2
MSQ8_2	25,25	44,3	5,47	19,2	55,7	59,26	45,3
MNQ16_2	28,21	41,49	4,17	20,93	58,51	57,79	48,2
MCQ22_2	30,28	38,54	3,78	23,27	61,46	55,28	49,3
MCQ23_2	30,5	38,15	3,7	23,5	61,85	54,89	49,3
MCQ25_2	26,92	46,76	2,71	18,53	53,24	63,98	50,6
MCQ27_2	21,4	55,82	2,15	15,48	44,18	71,02	48,4
MSQ3_1	22,08	51,3	3	18,09	48,7	65,84	45,3
MSQ7_1	26,64	46,51	2,46	19,53	53,49	63,40	49,8
MCQ10_1	20,71	52,09	2,36	19,75	47,91	65,70	43,2
MCQ15_1	24,93	47,83	2,64	19,47	52,17	63,71	47,8
MNQ23_1	21,98	52,6	2,83	17,09	47,4	67,42	46,4
MSQ1_2	25,3	47,89	2,3	19,42	52,11	64,11	48,5
MSQ6_2	24	53,1	1,1	18,7	46,9	69,87	51,17
MSQ7_2	22	48	2,19	21,6	52	61,54	42,31
MNQ11_2	20	44,6	4,02	24	55,4	55,75	36,10
MNQ12_2	24	48,8	2,44	19,2	51,2	64,21	46,88
MNQ13_2	30	43,5	0,11	19,7	56,5	62,14	53,10
MNQ15_2	28	47,9	3,54	20	52,1	66,53	53,74
MNQ18_2	28	48,1	1,55	18,4	51,9	66,81	53,95
MSQ5_1	28	44,6	5	18,7	55,4	61,94	50,54
MCQ12_1	24	51,1	3,6	17,1	48,9	67,24	49,08
MCQ13_1	24	52	2,21	17,4	48	68,42	50,00
MNQ26_1	26	52,2	1,1	17	47,8	70,54	54,39
MSQ9_2	29	50,9	2,28	15,2	49,1	71,69	59,06
MCQ11_2	28,8	49	3,2	16,7	51	68,82	56,47
MNQ22_2	34	41,3	2,96	18,1	58,7	62,58	57,92
MNQ24_2	27,7	49,4	2,95	17,8	50,6	68,33	54,74
MNQ25_2	22	47,3	3,47	22,3	52,7	60,64	41,75
MNQ27_2	22	49,4	3,85	15,4	50,6	63,33	43,48
MNQ28_2	26	49,5	2,53	18,1	50,5	66,89	51,49
MSQ5_2	24	50	1,75	18,9	50	65,79	48,00
MCQ12_2	20,5	52,6	3,5	17,8	47,4	66,16	43,25
MCQ15_2	21	54,8	2,42	16,4	45,2	69,37	46,46

Sample	Fat	Humidity	Salt	Protein	Total Solids	HSMG	GES
MCQ16_2	24	48,9	2,61	18,3	51,1	64,34	46,97
MCQ22_2	28	45,6	3,7	21	54,4	63,33	51,47
MNQ19_1	27	42,7	4,7	22,5	57,3	58,49	47,12
MCQ24_2	24	46	4,44	18,9	54	60,53	44,44
MNQ28_1	22	54	2,55	17,9	46	69,23	47,83
MNQ29_1	28	44,2	3	21,2	55,8	61,39	50,18

Table 2. Data used to carry out the ACP.

Source: The authors.

Statistical Analysis

Environment Cleaning and Package Installation: The work environment in R was started, cleaning previous variables and loading necessary packages such as vctrs, dplyr, and readr for data manipulation and analysis.

Data Preparation: The pivot_longer function was used to transform data from wide format to long format.

Kaiser-Meyer-Olkin (KMO): Sample Adequacy Test

The KMO Test measures the suitability of data for Principal Component Analysis (PCA). Evaluates whether partial correlations between variables are low, indicating that the data are suitable for analysis (Tirado-Malaver, M., González, J. M., & Blanco, J., 2018).

Acceptance Criteria:

KMO > 0.80: It is considered excellent. The data is very suitable for the ACP.

0.70 < KMO < 0.80: Appropriate. The data is sufficient for the ACP.

0.60 < KMO < 0.70: Moderate. The data may be acceptable, but there could be problems.

KMO < 0.60: Not suitable. The data is not suitable for the ACP.

Rejection Criteria:

A low KMO value (less than 0.60) suggests that the correlation between the variables is not sufficient to perform the PCA. This indicates that the variables may not be sufficiently correlated or that the sample is insufficient (Tirado-Malaver, M., González, J. M., & Blanco, J., 2018).

Bartlett's Test of Sphericity: Bartlett's Test of Sphericity verifies the null hypothesis that the correlation matrix is an identity matrix. That is, it evaluates whether the correlations between the variables are sufficiently different from zero for the PCA to be appropriate (Bartlett, 1954).

Acceptance Criterion: p-value < 0.05: Rejects the null hypothesis, indicating that the correlation matrix is not an identity matrix and that, therefore, the PCA is appropriate for the data (Bartlett, 1954).

Rejection Criterion: p-value > 0.05: The null hypothesis is not rejected, suggesting that the correlation matrix may be an identity matrix. This indicates that the correlations between the variables are low, and PCA may not be appropriate for the data (Kaiser, 1974)

Principal Component Analysis (PCA): Normalization of the data was performed using the scale function and then PCA was applied with the prcomp function. PCA focused on reducing the dimensionality of the data and identifying significant patterns (Jolliffe, 2002).

Hierarchical Clustering: Using the Ward.D2 method and Euclidean distances, hierarchical clustering was carried out to group the data according to the similarities in their characteristics (Borcard, D., Gillet, F., & Legendre, P., 2018). A dendrogram was generated to visualize the relationships between the clusters (Hennig, 2015).

K-Means Clustering: The K-Means algorithm was applied to group the data into a predefined number of clusters. To determine the optimal number of clusters, the silhouette method was used, which evaluates the quality of the grouping by calculating the measure of cohesion and separation between clusters. The silhouette method provided an evaluation of the internal consistency of the clusters and helped to select the appropriate number of clusters (Mac Queen, 1967).

Results and Discussion

The KMO index was carried out for the variables of Fat, Moisture, Salt, Protein, Total Solids, HSMF and GES, which returned a value of 0.5, which warns us that the correlation between the variables is not adequate. Therefore, the Bartlett test was applied, which reported a p-value = 0, with 21 degrees of freedom, but the Chi square value is infinite, which suggests that the data may have high collinearity. A correlation matrix was made to analyze the given variables (see table 3).

Variable	Fat	Moisture	Salt	Protein	Total solids	HSMG	GES
Fat	1,00						
Moisture	-0,65	1,00					
Salt	0,01	-0,35	1,00				
Protein	0,14	-0,67	0,13	1,00			
Total solids	0,65	-1,00	0,35	0,67	1,00		
HSMG	-0,21	0,88	-0,44	-0,78	-0,88	1,00	
GES	0,76	0,31	-0,28	-0,38	-0,31	0,47	1,00

Table 3. Correlation Matrix

Source: The authors.

The total solids variable is suggested to be not significant with the applied PCA model and the value of the KMO index was recalculated for the Fat, Moisture, Salt, Protein, HSMF and GES variables, which returned to a value of 0.61, so the data is moderately adequate. The Bartlett test was recalculated, which reported a p-value = 1.624322E-140, with 15 degrees of freedom, for a Chi square value of 704.9671, which indicates that the variables that have an impact on the study and they are highly related and therefore can be reduced. A correlation matrix was created again (see table 4).

Variable	Fat	Moisture	Salt	Protein	HSMG	GES
Fat	1,00					
Moisture	-0,65	1,00				
Salt	0,01	-0,35	1,00			
Protein	0,14	-0,67	0,13	1,00		
HSMG	-0,21	0,88	-0,44	-0,78	1,00	
GES	0,76	0,00	-0,28	-0,38	0,47	1,00

Table 4. Correlation Matrix Eliminating the Total Solids Variable

Source: The authors.

To determine the optimal number of components in the Principal Component Analysis (PCA) model, a scree plot was performed with the purpose of identifying the appropriate number of components that retain most of the variance in the data, while simplifying the model by reducing the dimensionality.

The scree plot (see figure 2) reveals that the number of components to retain is 4. This choice is based on the observation of the "elbow" in the figure, where the change in explained variance begins to stabilize. Selecting the number of components just before this stabilization allows you to capture the greatest amount of useful information from the data, while avoiding additional components that provide little added value and could increase the complexity of the model without substantial benefits.

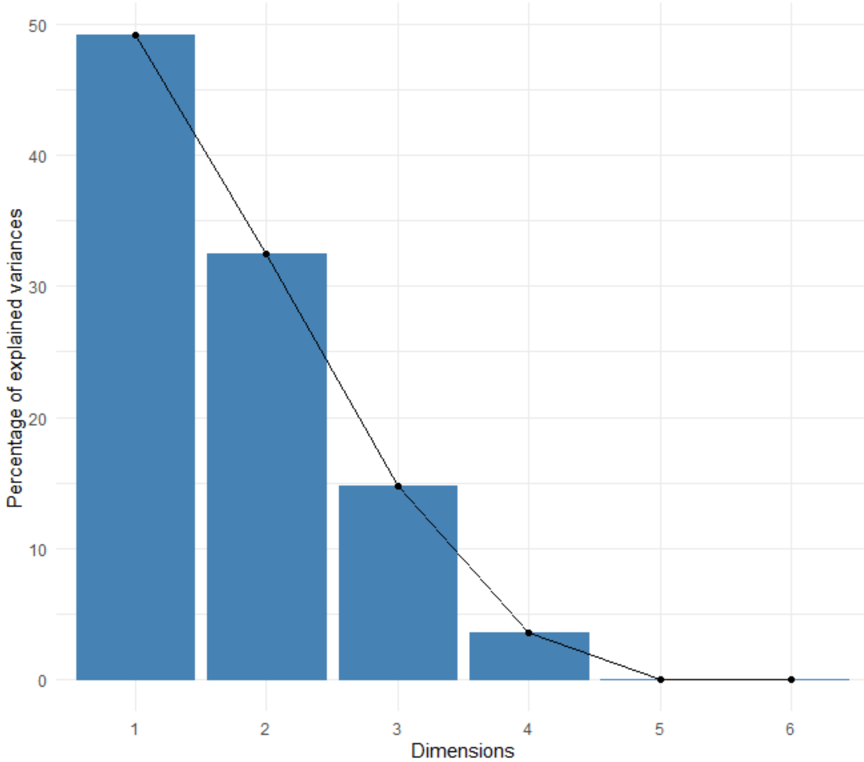

Figure 2. Scree Plot to Reduce the Number of Components

Source: The authors.

Table 5 provides a detailed summary of the variance explained by each component and the total cumulative variance:

Component	Explained variance [%]	Cumulative variance [%]
comp 1	49,16669088	49,16669088
comp 2	32,44986505	81,61655593
comp 3	14,80343229	96,41998822
comp 4	3,55039925	99,97038747
comp 5	0,01899593	99,98938340
comp 6	0,01061660	100,00000000

Table 5. Explained and Accumulated Variance by Principal Component in the PCA

Source: The authors.

According to the results in Table 5, the first four components explain 99.9704% of the total variance in the data. This means that by retaining these four components, virtually all relevant information contained in the original data set is captured, thus achieving a significant reduction in dimensionality without losing a significant amount of information. Therefore, the choice to retain four (4) components is appropriate to simplify the model while preserving the integrity of the information.

The PCA was performed with a rotation matrix of principal components using the Varimax method to identify the underlying dimensions in the variables related to physical-chemical parameters. This method allowed us to simplify the interpretation of the principal components by maximizing the variance explained by each component and minimizing the cross loading of the variables (see table 6).

Variable	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
Fat	-0,378	0,923			
Moisture	0,871	-0,319	-0,197	-0,317	
Salt	-0,203		0,975		
Protein	-0,5	-0,108		0,859	
HSMG	0,88	0,163	-0,244	-0,372	
GES	0,246	0,941	-0,148	-0,18	

Table 6. Factor Loadings After Varimax Rotation of the PCA

Source: The authors.

The first component (Dim.1) is mainly associated with the variables "Humidity" and "HSMG", showing loadings of 0.871 and 0.880, respectively. This indicates that this component captures a significant dimension related to moisture and characteristics associated with cheese quality.

The second component (Dim.2) has a high loading on "Fat" (0.923) and "GES" (0.941), suggesting that this component reflects aspects related to the fat content and other general characteristics of the cheese.

In the third component (Dim.3), "Salt" presents a prominent loading of 0.975, indicating that this component is mainly linked to the salt content in the cheese.

The fourth component (Dim.4) shows a high loading on “Protein” (0.859), suggesting that this component captures the variability in the protein content of the cheese.

The fifth component (Dim.5) does not present any variable with a significant loading (less than 0.5), which indicates that it does not provide additional relevant information in this analysis.

For the cluster analysis, a hierarchical clustering dendrogram was made with the objective of identifying similarities between the different groups. Figure 3 presents the dendrogram generated using the validated PCA variables. This dendrogram illustrates how observations are grouped based on similarities in the reduced variables, providing a clear visualization of the relationships between the clusters formed.

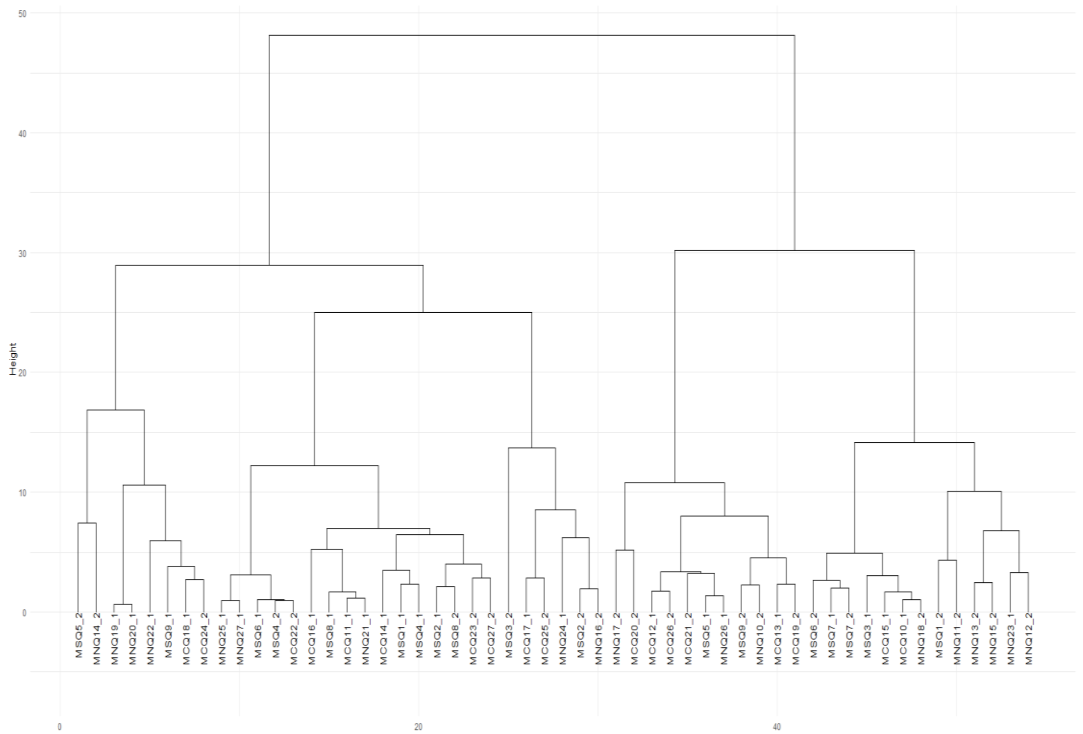

Figure 3. Dendrogram of Hierarchical Clustering Based on PCA Variables.

Source: The authors.

An analysis was performed to determine the optimal number of clusters using the silhouette index. This method allowed us to identify that the optimal number of clusters is 4, since it maximizes the separation between the groups and the cohesion within each group (see figure 4). The silhouette index graph confirms this choice (see figure 5).

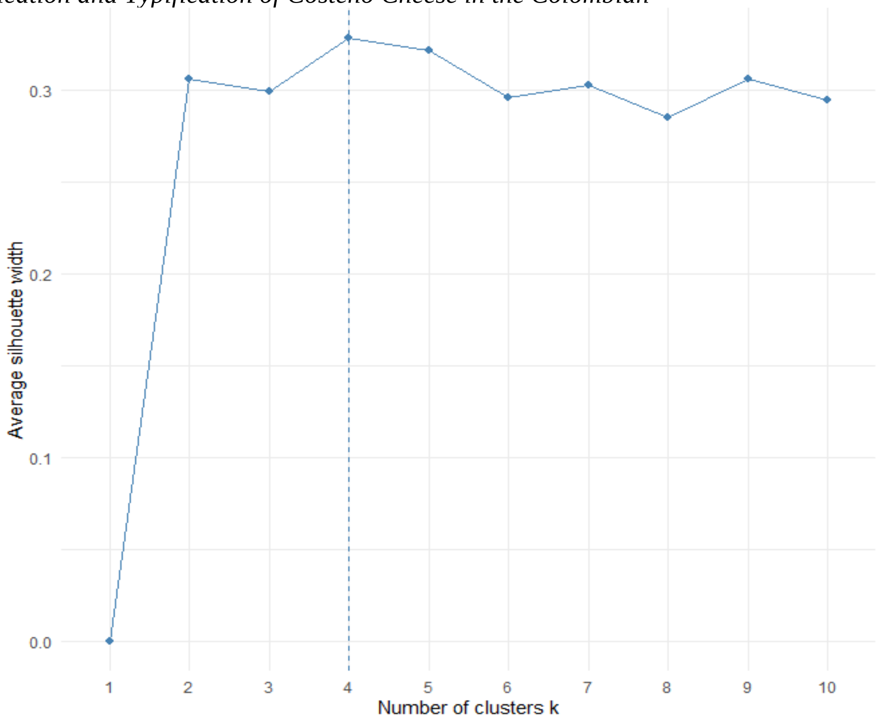

Figure 4. Optimal Number of Clusters According to the Silhouettes Index.

Source: The authors.

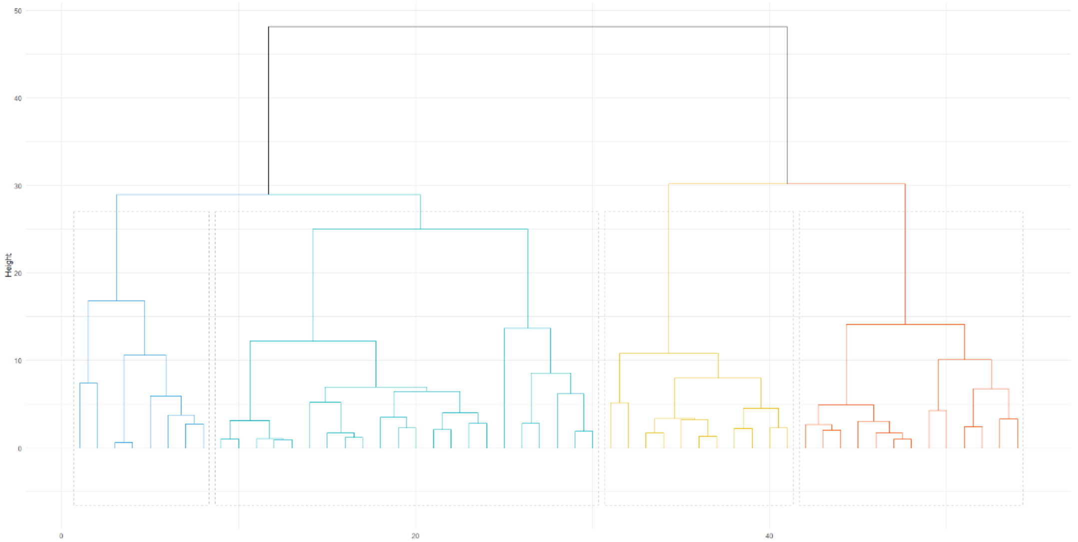

Figure 5. Hierarchical Clustering Dendrogram with an Optimal Value of 4 Clusters

Source: The authors.

To validate this selection, k-means clustering with 4 clusters was carried out. The cluster assignment was added to the data set, allowing clear identification of the groups formed. In the analysis of the hierarchical dendrogram, it was determined that the appropriate cutting height to obtain 4 clusters is 28.97 (see figure 6).

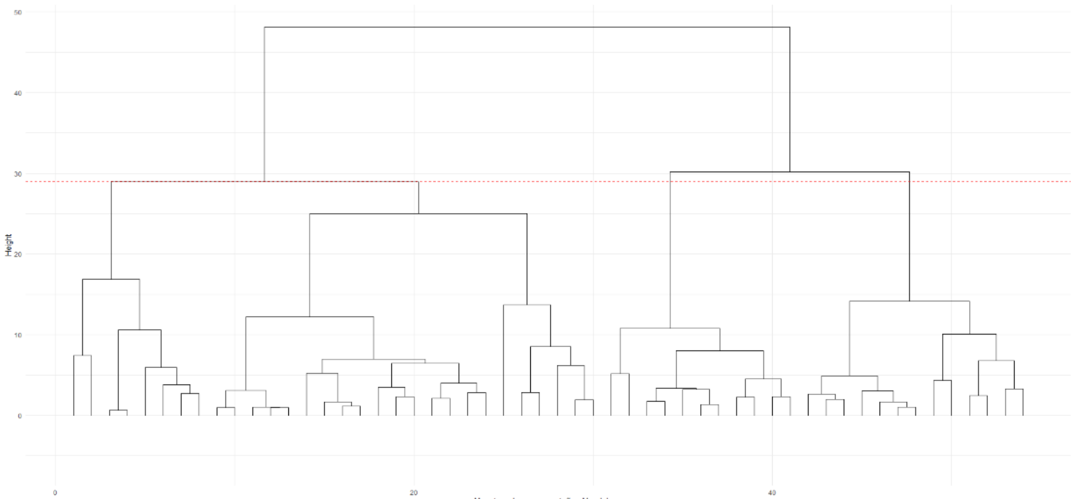

Figure 6. Dendrogram of Hierarchical Clustering with cut height n = 28,97

Source: The authors.

The cutoff height of 28.97 corresponds to a similar percentage of 96.23% within the height range of the dendrogram. This indicates that the clusters obtained are very close to the groups represented in the hierarchical dendrogram, which validates the choice of the number of clusters and their grouping (see figure 7).

Additionally, clustering table 7 shows the assignment of the Costeño cheese samples to different clusters, evidencing significant patterns in the physicochemical characteristics. The samples are grouped into four distinct clusters, each with distinctive characteristics in terms of fat, moisture, salt, protein, and other parameters.

Cluster 1 includes samples with relatively high levels of fat and moisture, and a moderate concentration of salt and protein. These samples tend to be smoother and less salty, which could correlate with a creamier and less salty flavor and texture profile.

Cluster 2 groups samples with high levels of fat and moisture, but with a lower amount of salt. This cluster stands out for its higher fat content, which could indicate a richer texture and greater density of flavor, despite the lower salinity.

Cluster 3 contains samples with more balanced values in fat, moisture and salt. The protein in these samples varies, but overall, the cluster shows a combination of characteristics that could be associated with an optimal balance between softness and salinity.

Cluster 4 presents samples with higher levels of salt and lower moisture and fat. These samples, being drier and saltier, could offer a firmer texture and a more robust flavor profile.

This analysis demonstrates that differences in physicochemical parameters, such as fat, humidity, salt content, HSGM and GES are essential to classify Costeño cheese samples into different groups, each with its own characteristics that can influence its quality. and market acceptance. These findings provide a solid basis for the evaluation and potential improvement of the quality of Costeño cheese from the department of Magdalena based on its physicochemical attributes (see table 7).

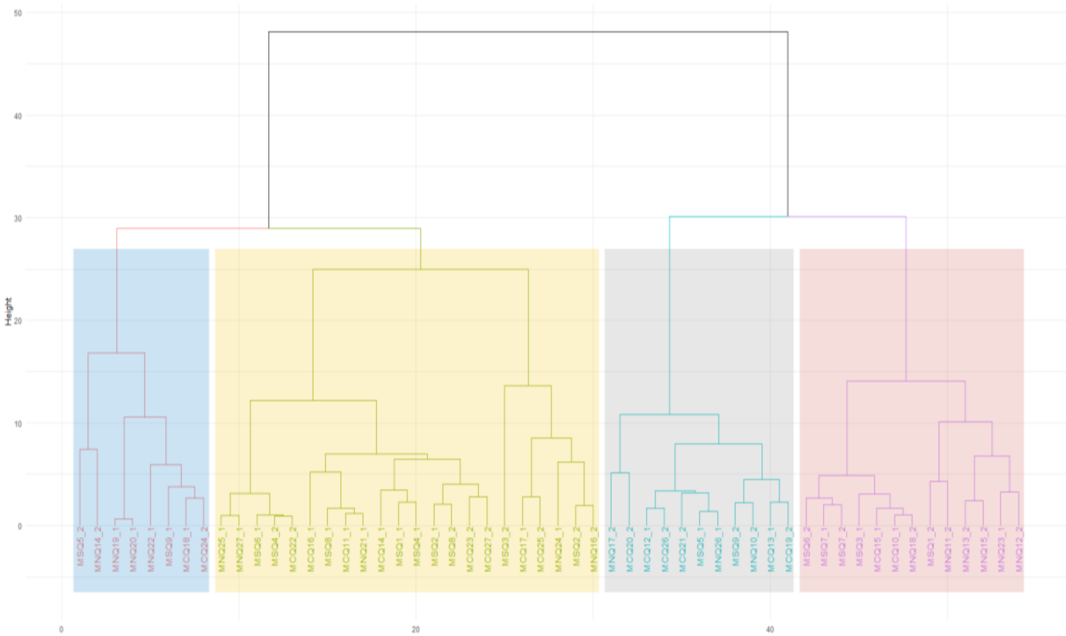

Figure 7. Hierarchical Clustering Dendrogram with an optimal value of 4 clusters and n = 28.97 with 96.23% similarity.

Source: The authors.

Sample	Fat	Moisture	Salt	Protein	Total solids	HSMG	GES	CLÚSTER
MSQ1_1	26.59	45.87	2.76	20.38	54.13	62.48	49.12	1
MSQ2_1	27.13	44.83	3.71	18.78	55.17	61.52	49.18	1
MSQ4_1	25.38	46.45	2.38	20.89	53.55	62.25	47.39	1
MSQ6_1	24.17	49.12	2.77	19.06	50.88	64.78	47.50	1
MSQ8_1	26.89	46.22	3.31	18.85	53.78	63.22	50.00	1
MCQ11_1	27.60	46.27	2.80	18.32	53.73	63.91	51.37	1
MCQ14_1	25.48	46.52	3.97	18.10	53.48	62.43	47.64	1
MCQ16_1	26.81	46.57	5.07	14.93	53.43	63.63	50.18	1
MCQ17_1	25.25	44.30	5.47	19.20	55.70	59.26	45.33	1
MNQ21_1	26.92	46.76	2.71	18.53	53.24	63.98	50.56	1

Sample	Fat	Moisture	Salt	Protein	Total solids	HSMG	GES	CLÚSTER
MNQ24_1	23.00	47.60	0.00	18.00	52.40	61.82	43.89	1
MNQ25_1	24.93	47.83	2.64	19.47	52.17	63.71	47.79	1
MNQ27_1	25.30	47.89	2.30	19.42	52.11	64.11	48.55	1
MSQ2_2	22.00	48.00	2.19	21.60	52.00	61.54	42.31	1
MSQ3_2	20.00	44.60	4.02	24.00	55.40	55.75	36.10	1
MSQ4_2	24.00	48.80	2.44	19.20	51.20	64.21	46.88	1
MSQ8_2	28.00	44.60	5.00	18.70	55.40	61.94	50.54	1
MNQ16_2	22.00	47.30	3.47	22.30	52.70	60.64	41.75	1
MCQ22_2	24.00	48.90	2.61	18.30	51.10	64.34	46.97	1
MCQ23_2	28.00	45.60	3.70	21.00	54.40	63.33	51.47	1
MCQ25_2	24.00	46.00	4.44	18.90	54.00	60.53	44.44	1
MCQ27_2	28.00	44.20	3.00	21.20	55.80	61.39	50.18	1
MSQ3_1	26.86	48.04	3.36	16.80	51.96	65.68	51.69	2
MSQ7_1	27.31	47.93	2.23	18.68	52.07	65.94	52.45	2
MCQ10_1	26.26	48.82	2.64	18.01	51.18	66.21	51.31	2
MCQ15_1	26.40	49.40	1.81	18.76	50.60	67.12	52.17	2
MNQ23_1	28.10	50.90	0.00	16.00	49.10	70.79	57.23	2
MSQ1_2	24.00	53.10	1.10	18.70	46.90	69.87	51.17	2
MSQ6_2	28.00	47.90	3.54	20.00	52.10	66.53	53.74	2
MSQ7_2	28.00	48.10	1.55	18.40	51.90	66.81	53.95	2
MNQ11_2	26.00	52.20	1.10	17.00	47.80	70.54	54.39	2
MNQ12_2	29.00	50.90	2.28	15.20	49.10	71.69	59.06	2
MNQ13_2	28.80	49.00	3.20	16.70	51.00	68.82	56.47	2
MNQ15_2	27.70	49.40	2.95	17.80	50.60	68.33	54.74	2
MNQ18_2	26.00	49.50	2.53	18.10	50.50	66.89	51.49	2
MSQ5_1	21.74	53.32	2.45	17.83	46.68	68.13	46.57	3
MCQ12_1	22.81	53.01	1.86	17.78	46.99	68.67	48.54	3
MCQ13_1	23.01	51.14	2.86	18.15	48.86	66.42	47.09	3
MNQ26_1	21.98	52.60	2.83	17.09	47.40	67.42	46.37	3

Sample	Fat	Moisture	Salt	Protein	Total solids	HSMG	GES	CLÚSTER
MSQ9_2	24.00	51.10	3.60	17.10	48.90	68.96	49.31	3
MCQ11_2	20.65	50.20	2.45	17.55	49.80	66.94	47.62	3
MNQ22_2	22.00	49.30	3.35	18.00	50.70	68.11	48.55	3
MNQ24_2	20.40	51.70	2.53	18.35	48.30	66.79	46.78	3
MNQ25_2	22.40	52.00	2.05	18.50	48.00	68.20	48.64	3
MNQ27_2	22.90	50.20	2.88	18.70	49.80	68.78	47.31	3
MNQ28_2	22.50	51.00	2.53	17.90	49.00	67.89	46.87	3
MSQ5_2	23.00	47.90	2.71	19.32	52.10	64.85	46.42	3
MCQ12_2	23.00	48.00	2.64	19.23	52.00	65.32	47.18	3
MCQ15_2	23.00	47.70	2.55	19.45	52.30	64.76	46.75	3
MCQ16_2	23.00	47.50	2.78	19.54	52.50	65.90	47.32	3
MCQ22_2	25.50	46.00	3.85	18.40	54.00	64.68	48.12	4
MNQ19_1	25.00	43.80	4.40	19.70	56.20	63.80	47.14	4
MCQ24_2	27.50	42.50	4.00	20.00	57.50	60.21	46.43	4
MNQ28_1	28.00	41.20	3.75	21.00	58.80	62.94	48.52	4
MNQ29_1	27.00	43.50	4.10	20.50	56.50	61.29	47.79	4

Table 7. Assignment of Costeño Cheese Samples to Clusters Based on Clustering Analysis

Source: The authors.

Cluster	Fat	Moisture	Salt	Protein	HSMG	GES
1	25.25	46.56	3.22	19.51	62.31	47.23
2	27.11	49.63	2.18	17.70	68.09	53.84
3	22.46	52.18	2.72	17.43	67.29	46.97
4	29.55	41.45	2.78	21.55	58.85	50.48

Table 8. Average Value of Each Cluster

Source: The authors.

The clustering analysis has allowed the cheeses to be segmented into four different groups based on their physical-chemical characteristics. This segmentation has the objective of identifying patterns and differences between cheese groups for a better understanding and classification according to established parameters. The results of the analysis are as follows (see table 8 and 9):

Cluster 1: Includes cheese with moderate fat (25.25%) and protein (19.51%) content, and low humidity (46.56%). These cheeses have a high salt concentration (3.22) and an HSMG classification of Firm/Semi-Hard and GES of Fatty, indicating that they are suitable for applications that require a firm texture and robust flavor.

Cluster 2: Cheeses in this group have moderate fat and protein content (27.11% and 17.70%, respectively) and higher moisture (49.63%). The low salt concentration (2.18) and the HSMG Soft and GES Fatty classification suggest cheeses with a softer, creamier texture, ideal for uses that require a softer consistency.

Cluster 3: Presents cheese with low fat content (22.46%) and moderate humidity (52.18%), with high levels of salt (2.72) and protein (17.43%). The HSMG classification of Firm/Semi-Hard and GES of Fat indicates cheeses with a good combination of characteristics that make them suitable for applications that require a firm texture and pronounced flavor.

Cluster 4: This group includes cheeses with moderate fat (29.55%) and protein (21.55%) content, but with low moisture (41.45%). They have a high salt level (2.78) and are classified as HSMG Firm/Semi-Hard and GES Fatty, indicating cheeses with a firm texture and robust flavor, suitable for applications requiring good consistency and flavor.

Cluster	Fat classification	Salt classification	HSMG classification	GES classification
1	High	High	Firm/ Semi-hard	Fat
2	High	High	Soft	Fat
3	Low	High	Firm/ Semi-hard	Fat
4	Low	High	Firm/ Semi-hard	Fat

Table 9. Classification of Each Cluster According to Resolution 821 Of 2021 of the Ministry of Health and Social Protection (2021) Y NTC 750

Source: The authors.

Conclusions

The initial Kaiser-Meyer-Olkin (KMO) index for the Fat, Moisture, Salt, Protein, Total Solids, HSMG and GES variables was 0.50, indicating inadequacy in the data for PCA analysis. A KMO value less than 0.60 suggests that the correlations between the variables are not strong enough, which could compromise the reliability of the analysis. This insufficiency was confirmed by Bartlett's test of sphericity, which presented a p-value of 0 and an infinite Chi square value. This indicates high collinearity between the variables, suggesting that they could be excessively correlated, making it difficult to identify significant components.

The exclusion of the 'Total Solids' variable, due to its lack of significance in the model, allowed the KMO index to be recalculated, which resulted in 0.61. Although this value is still moderate, it shows an improvement in the suitability of the data set for PCA after the removal of the problematic variable. Subsequently, the Bartlett test for the reduced set of variables showed an extremely low p-value (1.624322E-140) and a Chi square value of 704.9671. These results suggest that the remaining variables are significantly correlated and are suitable for PCA, reducing high collinearity and allowing more effective identification of principal components

Interpretation of the Components

Component 1: With high loadings on “Moisture” (0.871) and “HSMG” (0.880), this component reflects a crucial dimension related to moisture and the characteristics associated with cheese quality. The high loading on these variables suggests that humidity and associated characteristics are determining factors in cheese variability.

. Component 2: The high loadings on "Fat" (0.923) and "GES" (0.941) indicate that this component captures aspects related to the fat content and general characteristics of the cheese. This suggests that fat content and other general characteristics are important in the second dimension.

. Component 3: The high loading on "Salt" (0.975) shows that this component is mainly related to the salt content in the cheese, highlighting the importance of this factor in the physical-chemical composition.

. Component 4: The high loading on "Protein" (0.859) suggests that this component reflects the variability in the protein content of the cheese. This indicates that protein is an important dimension in principal component analysis.

The clustering analysis reveals that, despite the variations in the bromatological composition of Costeño cheese in the Magdalena region, the cheese shows high homogeneity in terms of its main physical-chemical parameters. This consistency is due to similar environmental conditions in the region, such as soil type and livestock feed, which contribute to a uniform quality in the cheese produced.

Finally, the clustering analysis and classification provide a detailed view of Costeño cheese in the department of Magdalena - Colombia, highlighting its homogeneity and allowing a precise classification that reinforces the authenticity and quality of the product according to the standards established by Resolution 821 of 2021 from the Ministry of Health and Social Protection (2021) and NTC 750. This analysis offers a solid basis to optimize production and classification processes, helping to improve the quality of the cheese and its positioning in the market.

Acknowledgments

This article is part of the project "Strengthening the Productive and Commercial Capacity of the Coastal Cheese Supply Chain in the Subregions of the Colombian Caribbean, Departments of Magdalena, Córdoba, and La Guajira," with BPIN code 2020000100116 - General Royalties System (SGR). We would like to thank in advance the Ministry of Science, Technology, and Innovation (Minciencias), the National Planning Department (DNP), the SGR, the Universities of Magdalena, Córdoba, and La Guajira; and, especially, the stakeholders in the coastal cheese supply chain of the Colombian Caribbean.

Contribution of the Authors

Edwin Causado Rodriguez worked on the methodological process of analysis Contribution: writing - first draft writing - review and editing. Jhon Jairo Vargas Sanchez contributed: writing - first draft writing - review and editing; and Andres Peñaloza Fernandez contributed: writing - first draft writing - review and editing. The three authors contributed to the interpretation of the results and to the preparation and writing of the document.

Conflict of Interest

The authors declare that there is no conflict of interest in relation to the publication of this manuscript. Additionally, ethical aspects, including plagiarism, informed consent, data fabrication and/or falsehood, duplicate, and redundant publication were observed and verified by the authors.

References

Bartlett, M. S. (1954). A note on the multiplying factors for various chi square approximations. Journal of the Royal Statistical Society, 16(2), 296-298. https://doi.org/10.1111/j.2517-6161.1954.tb00122.x

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons.

Borcard, D., Gillet, F., & Legendre, P. (2018). Numerical Ecology with R (2nd ed., Vol. 2). Springer.

Calliope, S. R., Segundo, C. N., & Molina, E. A. (2019). Evaluación de la calidad de quesos frescos artesanales de tres regiones de la provincia de Jujuy. Repositorio CONICET. https://ri.conicet.gov.ar/handle/11336/121427

Castell-Palou, A., Rosselló, C., Femenia, A., & Simal, S. (2010). Application of Multivariate Statistical Analysis to Chemical, Physical and Sensory Characteristics of Majorcan Cheese. International Journal of Food Engineering, 6(2). https://doi.org/10.2202/1556-3758.1417

Causado-Rodriguez, E; Fonseca-Tovar, J & Galindo-Montero, A. (2023a). Gestión Empresarial y Asociatividad Productiva para la Cadena de Suministro del Queso Costeño en el Caribe Colombiano. Editorial de la Universidad de La Guajira. ISBN: 978-628-7619-74-6.

Causado-Rodriguez, E; Romero-Borja, I & Galindo-Montero, A. (2023b). Higiene y Manipulación de Alimentos: Queso Fresco – Queso Costeño. Cadena de Suministros del Queso Costeño. Editorial de la Universidad de La Guajira. 978-628-7619-75-3.

Causado-Rodriguez, E., Galindo, A., & Peñaloza, A. (2023c). Buenas Prácticas de Ordeño – BPO y Buenas Prácticas de Manufactura – BPM Para la Elaboración y Manejo del Queso Costeño. Cadena de Suministros del Queso Costeño. 1era Edición - Editorial Uniguajira.

Departamento Administrativo Nacional de Estadística (2007). Informe Nacional de Agricultura. DANE. https://www.dane.gov.co

Departamento Administrativo Nacional de Estadística (2021). Sistema de Información de Precios y Abastecimiento del Sector Agropecuario – SIPSA. Disponible en: https://www.dane.gov.co/index.php/servicios-al-ciudadano/servicios-informacion/sipsa

Elías-Caro, J., & Vidal-Ortega, M. (2010). Historia del Queso Costeño y su Impacto en la Economía Regional. Universidad de los Andes.

Ensumesa. (n.d.). Características de los Suelos en la Región Caribe de Colombia. Universidad Nacional de Colombia. https://www.ensumesa.unal.edu.co

Errázuriz Tortorelli, C. (2010). Indicaciones geográficas y denominación de origen: propiedad intelectual en progreso. Revista Chilena de Derecho, Vol. 37(n. 2)., 207-239. doi: http://dx.doi.org/10.4067/S0718-34372010000200002

Garciá, I. (2017). Los signos distintivos: implicaciones de una marca colectiva frente a una denominación de origen. Advocatus, Vol 14(No 29). Universidad Libre., 105- 121. doi: https://doi.org/10.18041/0124-0102/advocatus.29.1709

Granados Aristizabal, J. I. (2012). Las denominaciones de origen en la industria agrícola: una herramienta de distinción y competitividad. Revista Producción + Limpia. vol., 7(N° 2), ISSN 1909-0455., 95-105.

Gutiérrez Castañeda, C., Quintero Peñaranda, R., Burbano Caicedo, I. y, & Simancas Trujillo, R. (2017). Modelo de quesería artesanal bajo un signo distintivo en el Caribe colombiano: caso Atlántico. Revista

Lasallista de Investigación, Vol.14(. Núm.1), 72-83. https://doi.org/10.22507/rli.v14n1a6
Hennig, C. (2015). Clustering criteria. En Flexible and Adaptive Clustering (pp. 1-27). Springer. https://doi.org/10.1007/978-3-319-71404-2_1
Jolliffe, I. T. (2002). Principal Component Analysis. Springer Series in Statistics.
Kaiser, H. F. (1958). "The varimax criterion for analytic rotation in factor analysis."
Psychometrika, 23(3), 187-200. https://doi.org/10.1007/BF02289233
Kaiser, H. F. (1974). An index of factorial simplicity. Psychometrika, 39(1), 31-36. https://doi.org/10.1007/BF02291575
Kaufman, L., & Rousseeuw, P. J. (1990). Finding Groups in Data: An Introduction to Clúster Analysis. Wiley.
Lesmes Ramos, R. (2017). Evolución y Actualidad de las Denominaciones de Origen en Colombia. Editorial Universitaria.
MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. En Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and
Probability (Vol. 1, pp. 281-297).
MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. En Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, pp. 281-297).
Martínez, J. A., Rodríguez, R., & Pérez, E. (2021). Evaluación de la calidad del queso costeño mediante Análisis de Clúster. Revista Colombiana de Ciencias Pecuarias, 34(2), 123-135. https://doi.org/10.32006/rccp.2021.34.2.123
Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. Psychometrika, 50(2), 159-179. https://doi.org/10.1007/BF02294245
Ministerio de Salud y Protección Social. (2021). Resolución 821 de 2021. Recuperado de https://www.minsalud.gov.co/Normativa_Nacional/Resolucion_821_de_2021.pdf
Narváez, M., La Torre, C., & Ortiz, J. (1965). Historia de la Producción Láctea en Colombia. Editorial Universitaria.
NTC 750 (2009). Norma Técnica Colombiana para la Clasificación de Quesos Frescos. Instituto Colombiano de Normas Técnicas y Certificación.
R Core Team. (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/
Prieto, A. V., Sánchez, A. C. A., & Rodríguez, N. Q. (2021). La denominación de origen como estrategia de posicionamiento de marca del queso de capa del municipio de Mompox. Cuadernos Latinoamericanos de Administración, 17(32), 1-19.
Vergara, A., & Silva, M. (2007). Análisis multivariante aplicado a los resultados físico-químicos de los vinos tintos de Ica. Anales Científicos UNALM 68 (N° 68).
Sánchez-Mejía, A. (2015). Análisis Multivariado de la Calidad del Queso Costeño. Universidad de los Andes.
Sánchez-Mejía, A. (2019). Optimización de la Producción de Queso Costeño en el Caribe Colombiano. Universidad de los Andes.
Sánchez-Mejía, A. (2022). Estudio Comparativo de Métodos para la Clasificación del Queso Costeño. Universidad de los Andes.
Tirado-Malaver, M., González, J. M., & Blanco, J. (2018). Pruebas de adecuación de muestra para el análisis multivariante. Revista de Estadística Aplicada, 13(2), 45-56. https://doi.org/10.1234/revistaestadistica2018
Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. Journal of the American

Statistical Association, 58(301), 236-244. https://doi.org/10.2307/2282967
Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer.