Revista de Estudiantes de Economía / Número 9 / Enero-diciembre 2025

INTERCAMBIO

Analysis of unpaid care work patterns using K-Means: A Machine Learning application on ENUT 2021 data

Análisis de los patrones de trabajo de cuidado no remunerado mediante k-means: una aplicación de aprendizaje automático sobre datos de la ENUT 2021

Helen Granados Rodríguez
Juan Miguel Rodríguez Trujillo

E-ISSN 2619-6131

Facultad de Ciencias Humanas y Económicas
Sede Medellín

UNIVERSIDAD NACIONAL DE COLOMBIA

*Analysis of unpaid care work patterns using K-Means: A Machine Learning application on ENUT 2021 data**

Análisis de los patrones de trabajo de cuidado no remunerado mediante k-means: una aplicación de aprendizaje automático sobre datos de la ENUT 2021

*Helen Granados Rodríguez***
*Juan Miguel Rodríguez Trujillo****

Resumen

Este estudio aplica técnicas de aprendizaje automático y análisis de agrupamiento para identificar perfiles sociodemográficos de mujeres en Colombia según el tiempo dedicado al cuidado no remunerado. Empleando la Encuesta Nacional de Uso del Tiempo (ENUT) 2021, se construyeron variables sintéticas sobre trabajo doméstico, cuidado no remunerado, autocuidado, ocio y un índice de tradicionalismo. Se empleó el algoritmo k-means, junto con análisis de componentes principales (PCA) y criterios de validación como el método del codo y el índice de silueta, identificando seis perfiles con patrones contrastantes de organización del cuidado. Los

Intercamb. Rev. Estud. Economía. N°. 9
Enero-diciembre 2025
127 pp.
E-ISSN 2619-6131
pp. 44-86

* **Artículo recibido:** 20 de junio de 2025 | **Aceptado:** 15 de marzo de 2026 | **Modificado:** 6 de abril de 2026

** Economista y estudiante de Maestría en Estadística de la Universidad Nacional de Colombia. Correo electrónico: hgranados@unal.edu.co

*** Economista de la Universidad Externado. Correo electrónico: juan.rodriguez68@est.uexternado.edu.co

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

resultados confirman la persistencia de la división sexual del trabajo, con implicaciones para la política pública y la economía del cuidado, subrayando la necesidad de un enfoque interseccional.

Palabras clave: economía del cuidado, uso del tiempo, trabajo no remunerado, aprendizaje automático; k-means; interseccionalidad.

Clasificación JEL: J16, J22, C38

Abstract

This study applies machine learning and cluster analysis to identify sociodemographic profiles of women in Colombia, based on time dedicated to unpaid care work. Using data from the 2021 National Time Use Survey (ENUT), synthetic variables were constructed for domestic work, unpaid care, self-care, leisure, and a traditionalism index. The k-means algorithm, combined with principal component analysis (PCA) and validation criteria such as the elbow method and silhouette index, identified six profiles with contrasting care organization patterns. Results confirm the persistence of the sexual division of labor, with key implications for public policy and the care economy. They also highlight the importance of an intersectional approach to expose structural inequalities in the distribution of unpaid work burdens.

Key words: care economy, time use, unpaid work, machine learning, k-means, intersectionality.

JEL classification: J16, J22, C38

1. Introducción

In Colombia, as in most Latin American countries, women disproportionately bear the responsibilities of unpaid care work. Various studies, including analyses based on successive National Time Use Surveys (ENUT), have consistently documented those women dedicate significantly more time than men to caregiving tasks, domestic work, and the everyday reproduction of life (DANE, 2021; Castro & Ramírez, 2023; Beltran-Figueroa, 2025). Feminist economics has contributed to

making this invisible burden visible as a fundamental component of social sustainability, historically undervalued and absent from conventional economic metrics (Folbre, 2001; Carrasco, 2005).

However significant gaps remain in understanding the internal heterogeneity among women themselves with regard to care work. What types of women, in terms of age, educational level, ethnicity, household position, and gender attitudes, assume greater care burdens? What characteristics distinguish those who manage to combine caregiving with other spheres of life, such as paid employment or leisure? Answering these questions is key to designing public policies that are more responsive to the diverse realities and needs of women.

The aim of this study is to help close this gap through an exploratory analysis aimed at identifying differentiated profiles of women based on their time use, sociodemographic characteristics, and attitudes toward gender roles. To this end, an unsupervised machine learning approach is applied: the K-means algorithm, using data from the 2021 ENUT. The analysis considers dimensions such as domestic work, unpaid care, self-care, leisure, and a constructed traditionalism index.

Unlike previous studies focused on aggregated statistics, this approach makes it possible to reveal latent segments within the female population, providing a more detailed view of gender and intersectional inequalities in the social organization of care. The findings not only have empirical value but also offer relevant insights for the development of public policies aimed at a more equitable and sustainable distribution of care work.

The article is structured as follows. Section 2 describes the methodology and data used, including variable preprocessing and the logic of the K-means algorithm. Section 3 presents the segmentation results and identified profiles. Section 4 discusses the findings in light of the literature on the care economy and outlines their implications for public policy. Finally, Section 5 presents the study's limitations and possible directions for future research.

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

2. Methodology

This study is based on the 2021 National Time Use Survey (ENUT), a nationally representative statistical source for Colombia that provides detailed information on the distribution of time across various everyday activities. The ENUT collects information at the household, individual, and activity levels through different thematic modules. For the purpose of this analysis, modules ENUT_C1 to ENUT_C9 were used, covering aspects related to housing, sociodemographic characteristics, health status, educational attainment, employment, unpaid work, caregiving, personal activities, and the consequences of the COVID-19 pandemic.

The various modules were integrated using household and individual identification keys, which enabled the construction of a structured individual-level database. Since the analytical approach is exploratory and aims to characterize internal profiles within the sample rather than make population-level inferences. From the consolidated dataset, a set of variables was explicitly selected and constructed for the analysis (see Tables 1, 2 and 3). These include demographic and socioeconomic characteristics such as age, rural status (proxy for peasant condition), socioeconomic stratum, educational attainment, access to household assets (washing machine ownership), household headship, Afro-Colombian ethnic identification, employment status, and religiosity.

In addition, a traditionalism index was constructed to capture attitudes toward gender roles. Finally, time-use variables were aggregated into continuous measures (expressed in daily minutes), including personal self-care, food provision, clothing maintenance, cleaning activities, caregiving for others, and leisure time.

Table 1. Variables included in the analysis

Variable name	Type	Description
Afro-Colombian	Categorical (binary)	Ethnic self-identification
Age	Quantitative (continuous)	Age of the individual
Caregiving time	Quantitative (continuous)	Time spent caring for others
Cleaning activities time	Quantitative (continuous)	Time spent on household cleaning
Clothing maintenance time	Quantitative (continuous)	Time spent on clothing-related tasks
Educational attainment	Categorical (ordinal)	Highest level of education achieved
Employment	Categorical (binary)	Labor market participation
Food provision time	Quantitative (continuous)	Time spent preparing/providing food
Household headship	Categorical (binary)	Whether the individual is the household head
Leisure time	Quantitative (continuous)	Time allocated to leisure activities
Personal self-care time	Quantitative (continuous)	Time spent on personal care activities
Religiosity	Categorical (binary)	Self-reported religious affiliation/practice
Rural status	Categorical (binary)	Proxy for peasant condition
Socioeconomic stratum	Categorical (ordinal)	Household socioeconomic classification
Washing machine ownership	Categorical (binary)	Access to household asset

Note. Own elaboration based on ENUT 2021 (DANE).

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

Table 2. Summary statistics for the quantitative variables included in the analysis

Variable	Mean	Min	P25	Median	P75	Max
Age	46.07	18	32	44	59	115
Caregiving time	18.41	0.00	0.00	0.00	0.00	495
Cleaning activities time	62.59	0.00	0.00	60	100	395
Clothing maintenance time	30.93	0.00	0.00	0.00	60	270
Food provision time	109.32	0.00	45	110	160	445
Leisure time	66.65	0.00	0.00	60	120	420
Personal self-care time	613.87	0.00	540	600	675	1380

Note. Own elaboration based on ENUT 2021 (DANE). All values are expressed in minutes per day.

Table 3. Frequency distribution for the binary variables used in the analysis

Variable	No	Yes
Afro-Colombian	0.9188	0.0812
Employment	0.3586	0.6414
Household headship	0.6384	0.3616
Religiosity	0.0478	0.9522
Rural status	0.6987	0.3013
Washing machine ownership	0.2998	0.7002

Note. Own elaboration based on ENUT 2021 (DANE). Frequency distributions were computed using sampling weights (F_EXP) to ensure representativeness.

It is important to note that the analysis was limited to the female population, in accordance with the objective of study. For this reason, the sex variable was excluded from the multivariate analysis. Since there was no variability in this variable within the sample, it did not provide relevant additional information for the clustering process. Given the structure of the ENUT questionnaire, the first step consisted of building an

index that consolidated questions related to gender roles in time use. This approach allowed for the integration of multiple qualitative items into a single quantitative measure, ensuring greater consistency in the subsequent analyses.

The traditionalism index was derived from six items in the ENUT 2021 survey (questions P1107S1–P1107S6) that capture attitudes toward gender roles and family organization (see Appendix A). Respondents indicated their level of agreement on a four-point Likert scale (1 = strongly disagree, 4 = strongly agree). To maintain consistency in interpretation, the items expressing egalitarian views (S1, S2, and S5) were reverse-coded, so that higher scores uniformly reflected stronger adherence to traditional gender norms. The index was then calculated as the arithmetic mean of the six items, resulting in a continuous measure ranging from 1 to 4.

Although averaging Likert-scale items is not without methodological debate, given that the underlying scale is technically ordinal rather than continuous —this practice is widely accepted in fields such as psychology and the social sciences, where it is considered a reasonable approximation when the number of items is sufficient and the construct validity is supported (Norman, 2010; Sullivan & Artino, 2013). The resulting index is therefore treated as a continuous measure for analytical purposes.

Formally, the traditionalism index for each individual was defined as the simple average of their responses to the six items. To ensure validity, only respondents with at least three valid answers were included in the calculation. Observations with fewer than three responses were excluded, guaranteeing a minimum representation of the evaluated attitude and avoiding bias in the index construction.

In addition, the time-use variables were organized into synthetic dimensions that grouped similar activities: personal self-care, food provision, clothing maintenance, cleaning activities, support to other households, caregiving for others, and time dedicated to leisure. Each of these dimensions was constructed as the sum of the time reported for the corresponding activities, expressed in daily minutes:

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

$$D_{ik} = \sum_{l=1}^{n_k} t_{ilk}$$

Where D_{ik} is the total time individual i devoted to dimension k, t_{ilk} is the time reported for activity l within dimension k, and n_k is the number of activities included in that dimension.

Before the analysis, a rigorous data preparation process was conducted. Time-use variables, originally recorded in hours, were converted into minutes to ensure consistency in the dataset. Both the traditionalism index and the time-use dimensions were built as synthetic indices. Quantitative variables were then standardized using z-score normalization to ensure comparability within the clustering space:

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$$

where z_{ij} is the standardized value for individual i on variable j, u_j is the mean of variable j, and σ_j is its standard deviation. Standardization was performed using the StandardScaler from the scikit-learn library, developed by Pedregosa et al. (2011).

The multivariate analysis was conducted using a comprehensive set of 16 variables encompassing both continuous time-use measures and categorical sociodemographic characteristics. The continuous variables included: Age, Caregiving time, Cleaning activities time, Clothing maintenance time, Food provision time, Leisure time, Personal self-care time, and the constructed Traditionalism index. The categorical variables comprised: Afro-Colombian ethnic identification, Educational attainment, Employment status, Household headship, Religiosity, Rural status, Socioeconomic stratum, and Washing machine ownership.

Although Principal Component Analysis (PCA) is formally defined for continuous variables, its application to datasets including binary and ordinal variables is widely documented in applied social science research, particularly in the construction of socioeconomic indices. PCA has been extensively used to create composite measures

of household wealth and socioeconomic status based on discrete asset ownership variables and household characteristics (Vyas & Kumaranayake, 2006; Filmer & Pritchett, 2001). This empirical precedent, combined with methodological evaluations demonstrating the technique's robustness when variables are appropriately standardized (Kolenikov & Angeles, 2009), supports the application of PCA as a dimensionality reduction technique in mixed-data contexts where variables serve as proxies of underlying latent constructs.

In this study, categorical variables were re-coded into binary indicators following established practices for PCA-based analysis (Vyas & Kumaranayake, 2006): categorical variables were transformed into sets of binary dummy variables to avoid imposing artificial ordinal scales. All variables —both continuous time-use measures and binary sociodemographic indicators— were standardized prior to analysis to ensure equal weighting. Subsequently, a reduced set of principal components was selected based on cumulative explanatory power, retaining those components that together accounted for approximately 80% of the total variance, a threshold recommended in the literature to ensure adequate representation of the underlying data structure (Jolliffe, 2002; Jolliffe & Cadima, 2016). These components were then used as input for the K-means clustering algorithm.

While the standard PCA procedure applied to discrete data produces interpretable and empirically validated results (Filmer & Pritchett, 2001; Vyas & Kumaranayake, 2006), the underlying measurement assumptions regarding scale properties should be acknowledged. Kolenikov and Angeles (2009) note that more sophisticated alternatives —such as polychoric correlations— exist for handling ordinal data, though their simulations demonstrate that the widely-used dummy-variable approach remains a reliable method, particularly for dimensionality reduction purposes. In this application, where components are interpreted as latent dimensions reflecting women's socioeconomic positioning and time allocation patterns, the standard approach is appropriate and consistent with established practice in demographic and time-use research.

To focus the analysis on the adult population, the sample was restricted to

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

individuals aged 18 or older, consistent with the legal definition of adulthood in Colombia. This criterion ensures that the results reflect patterns of time use and gender attitudes among women who are likely to have greater autonomy in household decision-making. Only cases with complete information on the key variables were included.

Once the variables were standardized, a Principal Component Analysis (PCA) was applied to reduce the dimensionality of the dataset and condense information into fewer dimensions. This technique enables data representation in a more parsimonious latent space, improving computational efficiency and interpretability of clusters. For a detailed review of PCA, both the mathematical and applied perspectives of Jolliffe and Cadima (2016) and Shlens (2014) were followed.

In the social sciences, PCA serves a dual purpose: beyond its role in dimensionality reduction, it enables the identification of latent structures and underlying patterns of co-variation among observed variables (Joliffe & Cadima, 2016). In this study, PCA was applied with both objectives in mind —to condense the information into fewer dimensions for the clustering analysis, and to examine the latent axes along which women's time-use profiles and sociodemographic characteristics are organized.

Additionally, the number of principal components to retain was determined through two complementary criteria. First, a cumulative variance threshold of approximately 80% was applied, following standard recommendations in the literature (Jolliffe, 2002; Jolliffe & Cadima, 2016). Second, and more importantly, the selection was visually validated through a scree plot (Figure 2), which displays both individual and cumulative explained variance by component. The scree plot reveals a clear inflection point after the first few components, beyond which additional components contribute only marginal gains. The retention of the first ten components satisfies both the quantitative threshold and the visual criterion, providing a parsimonious yet representative reduction of the original data structure.

The PCA was performed on the previously standardized variables using the PCA class from the scikit-learn library of Python. The new latent components obtained through PCA were used directly as input for the K-means clustering analysis. Additionally,

to further verify the robustness of the analysis, the factor loadings of the variables on the first four principal components were examined. This analysis made it possible to identify which variables contributed most to the variance explained by each component, providing evidence of the structural coherence of the latent dimensions derived from PCA.

To identify homogeneous profiles within the population based on their demographic characteristics and time-use patterns, the K-means clustering algorithm —originally proposed by MacQueen (1967a)— was applied. This technique is based on minimizing intra-cluster variance. The selection of variables for this analysis included the previously standardized variables.

To determine the optimal number of clusters, two complementary techniques were implemented.

First, the elbow method was applied (see Figure 3), which involves plotting intra-cluster squared errors as a function of the number of groups k. This method seeks to identify the inflection point beyond which adding more clusters yields only marginal improvements in reducing within-group variance (Thorndike, 1953).

Second, the silhouette index proposed by Rousseeuw (1987) was calculated for values of k ranging from 2 to 10, with the goal of assessing internal cohesion and separation between clusters (see Figure 3). This coefficient quantifies how similar each observation is to its own cluster compared to other clusters. Values closer to 1 indicate better clustering structure. The silhouette index for observation i is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

where $a(i)$ is the average distance between i and all other observations in the same cluster, and $b(i)$ is the average distance between i and the observations in the nearest neighboring cluster.

Both metrics were considered jointly to narrow down the optimal range of clusters, and the final decision on the number of groups was complemented with substantive interpretability criteria of the resulting profiles.

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

K-means clustering was then applied to the dataset using the optimal number of clusters identified in the validation stage. Each observation was assigned to the nearest cluster centroid, generating labeled subsets that reflect the embedded structure of the data and enabling a detailed descriptive characterization of the resulting profiles (see Appendix B, Figure B1).

At its core, the K-means algorithm seeks to partition the dataset into k clusters by minimizing the within-cluster sum of squares (WSS), also known as intra-cluster inertia. This metric quantifies the compactness of clusters, ensuring that individuals assigned to the same group are as similar as possible according to Euclidean distance. Formally, inertia is defined as:

$$\mathrm{WSS} = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - \mu_k\|^2$$

where K is the number of clusters, C_k denotes the set of observations assigned to cluster k, x_i is the feature vector of observation i, and μ_k is the centroid of cluster k. The algorithm iteratively updates cluster centroids and reassigns observations until convergence is reached —that is, until centroid positions stabilize or changes fall below a predefined threshold.

To improve stability and avoid poor local minima, the k-means initialization method was used, which selects initial centroids in a way that maximizes their mutual separation (Arthur & Vassilvitskii, 2007). This step enhances convergence speed and increases the likelihood of obtaining a globally coherent solution.

A detailed schematic of the iterative K-means procedure is presented in Appendix B, where the initialization of centroids, the assignment of observations to the nearest cluster, and the iterative updates of cluster centroids are illustrated step by step. The process converges once centroid positions stabilize, yielding well-defined partitions that minimize within-cluster variance while maximizing between-cluster separation. Although K-means is sensitive to initial centroid placement, the k-means initialization method (Arthur & Vassilvitskii, 2007) significantly reduces this limitation by improving stability and convergence.

For each resulting cluster, measures of central tendency and dispersion were calculated for the original variables, enabling a substantive interpretation of the predominant sociodemographic and time-use patterns. Additionally, three-dimensional visualizations of the first three principal components were generated to facilitate the inspection of group separation. These complementary tools were useful both for the internal validation of the model and for the interpretation of the detected profiles.

All data processing and analysis were conducted in Python, using pandas (McKinney, 2010), scikit-learn (Pedregosa et al., 2011), matplotlib (Hunter, 2007), and seaborn (Waskom, 2021).

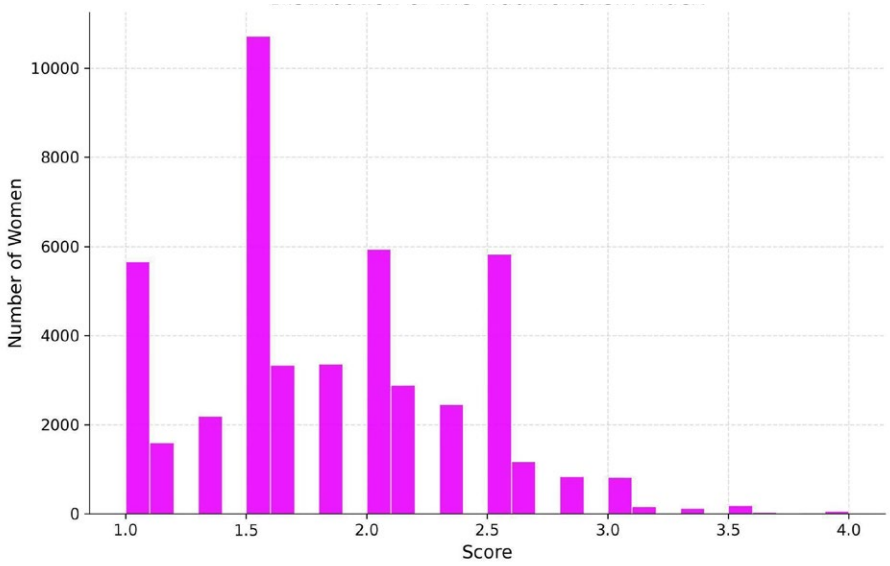
3. Results

The exploratory analysis aimed to synthesize the key dimensions of time use and sociocultural attitudes within the sample, based on the information provided by the 2021 ENUT. To this end, derived variables were constructed to capture integrated patterns of engagement in various daily activities, as well as a traditionalism index based on gender attitudes.

Figure 1 shows the distribution of the traditionalism index in the female sample. A higher concentration of cases can be observed around the intermediate values of the index, with a notable peak near 1.5, suggesting a moderate tendency toward traditionalist positions. Although there are also cases at the lower and upper extremes, these are considerably less frequent. The skewness of the distribution —with a longer tail toward higher index values— indicates that while many women express low or moderate levels of traditionalism, there is a smaller segment that strongly adheres to traditional gender norms. This distribution suggests heterogeneity in gender attitudes, with most women concentrated at low and intermediate index values, reflecting generally moderate or less traditionalist positions. The smaller frequency of high index values indicates that only a minority strongly adheres to traditional gender norms.

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

Figure 1. Distribution of the Traditionalism Index for Women



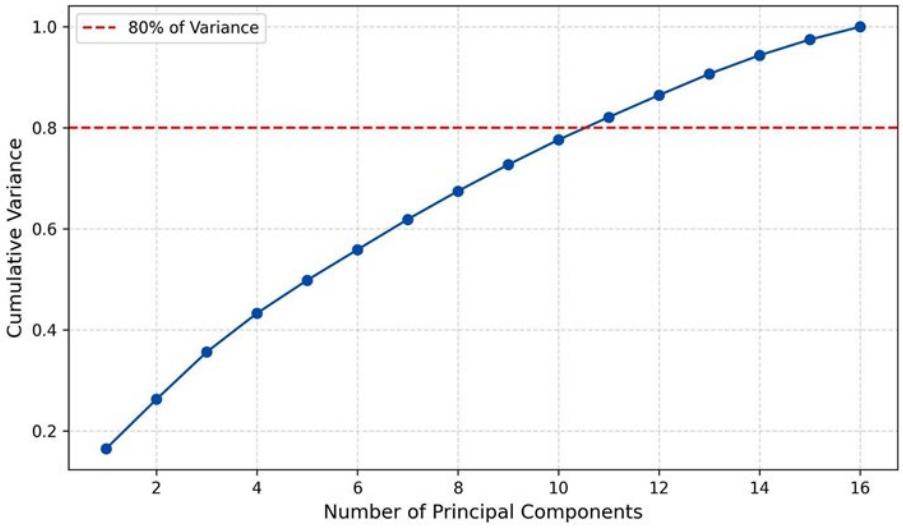
Note. Own elaboration based on ENUT 2021 (DANE). Most women concentrate around intermediate values (peak near 1.5), indicating a moderate tendency toward traditional gender roles.

Complementarily, synthetic variables were constructed to integrate the time reported (in daily minutes) for key activities: personal self-care, food provision, clothing maintenance, household cleaning, support to other households, caregiving for others, and leisure. These dimensions allow for a more precise characterization of the daily time organization across different profiles of women.

In this regard, a Principal Component Analysis (PCA) was conducted using both the time-use variables that capture daily time dedicated to key dimensions of everyday life, and sociodemographic characteristics that contextualize women's positioning within household and labor structures.

Figure 2 presents the scree plot with the cumulative explained variance. This indicates that the dimensionality of the original set of variables can be reduced while retaining most of the relevant information, which facilitates the characterization of women's profiles according to their time-use patterns.

Figure 2. Scree Plot of Cumulative Explained Variance (PCA)

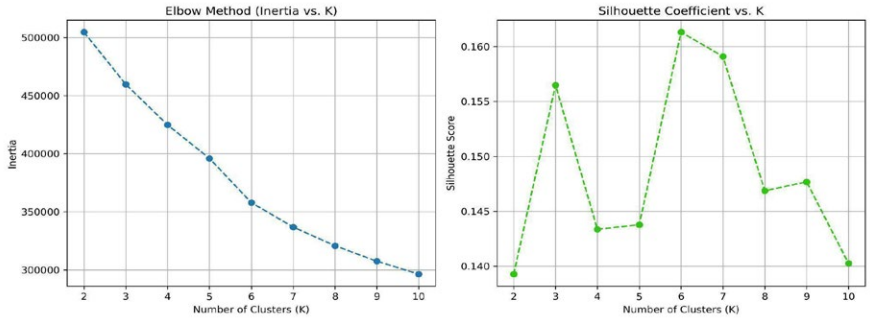


Note. Own elaboration based on ENUT 2021 (DANE). The first 10 principal components together explain approximately 80% of the total variance, justifying their selection for the clustering analysis.

Subsequently, for the k-means analysis, an exploratory assessment was conducted to determine the optimal number of clusters using two complementary techniques: the elbow method and the silhouette coefficient, applied over the same range of values. The results, illustrated in Figure 3, revealed an inflection point around , a criterion that was confirmed by a peak in the silhouette coefficient. Consequently, a final segmentation into six clusters was adopted.

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

Figure 3. Determination of Optimal Number of Clusters Using the Elbow Method and Silhouette Coefficient



Note. Own elaboration based on ENUT 2021 (DANE). Both metrics converge around , which was selected as the optimal number of clusters for the segmentation.

Based on these results, and considering the interpretability of the emerging profiles, a k-means model with six clusters (k = 6) was ultimately selected to segment the sample. The final solution produced a balanced and coherent distribution of records across the clusters. Membership in each cluster was determined by the proximity of individual observations to the respective centroid, which represents the mean profile of the group in the multidimensional space defined by the principal components included in the analysis. The resulting allocation of cases within the sample is presented below.

Table 4. Distribution of Records by Cluster

Cluster	N° of cases	% of Total
Cluster 0	4,271	9,01%
Cluster 1	2,260	4,77%
Cluster 2	13,982	29,50%
Cluster 3	11,867	25,04%
Cluster 4	12,289	25,93%
Cluster 5	2,723	5,75%

Note. Own elaboration based on ENUT 2021 (DANE). This table shows the number and percentage of women in each cluster identified by the k-means algorithm.

The resulting cluster structure was visualized through a Principal Component Analysis (PCA). Figure 4 shows the two-dimensional projection, while Figure 5 provides a three-dimensional representation —both illustrating the spatial differentiation among the groups. These visualizations reveal clear spatial patterns of separation between clusters. Although some degree of overlap is expected due to the complexity of the social variables involved, distinct clusters with well-defined shapes and positions can also be identified, which reinforces the robustness of the clustering results.

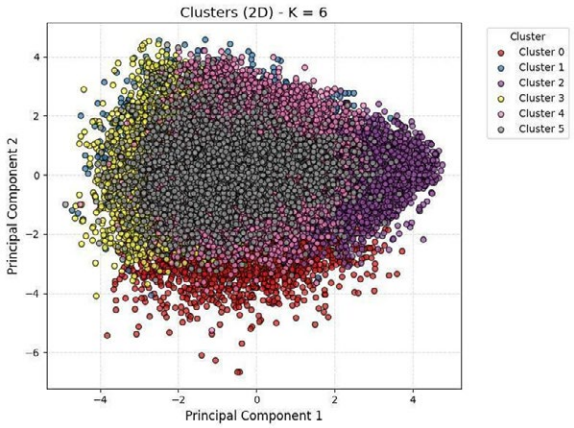
These visualizations reveal the spatial distribution of the six clusters along the principal components. As is common in clustering analyses applied to social survey data, some degree of overlap is observable, particularly in the central region of the projection. This is an expected feature when the underlying population does not form perfectly discrete groups, but rather occupies a continuum of sociodemographic and behavioral characteristics. It is important to note that two-dimensional projections inherently involve information loss, as they capture only a fraction of the total variance. The three-dimensional projection (Figure 5) partially addresses this limitation by revealing additional separation along the third component. The decision to retain six clusters was not made solely on geometric grounds, but was grounded in the substantive interpretability of the resulting profiles: each cluster maps onto a theoretically coherent pattern of time use, traditionalism, and sociodemographic positioning that would be obscured by a more parsimonious solution. Robustness checks with and (presented in Appendix C and D) confirm that the six-cluster solution offers a meaningful balance between statistical fit and theoretical richness.

Figure 4 shows the two-dimensional projection of the data onto the first two principal components. Each point represents an observation -a woman from the sample-, colored according to the cluster to which it was assigned. A high density is observed in the central region, where several clusters partially overlap. However, distinct concentrations and shapes associated with specific groups can also be identified, indicating a certain degree of structural differentiation. Visually, six groups can be distinguished, represented by the following color:

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

- Red (Cluster 0): Concentrated in the lower part of the graph, extending horizontally with noticeable density. This group shows a well-defined location with less overlap compared to other clusters.
- Dark Blue (Cluster 1): Scattered across the upper area of the plane, it has lower density and is situated at the upper extreme of the dataset.
- Purple (Cluster 2): Located on the right side of the graph, with a relatively compact and elongated shape. It is clearly distinguishable due to its concentration in that quadrant.
- Yellow (Cluster 3): Positioned toward the left side of the graph, forming a relatively dense cloud of points, though more dispersed at the edges.
- Pink (Cluster 4): Distributed in intermediate and peripheral areas, partially surrounding other groups. This suggests it may occupy a transitional position between more clearly defined clusters.
- Dark Gray (Cluster 5): Appears as the most central and dense group, occupying the core of the projection. Its location suggests overlap with multiple clusters, which could reflect a more heterogeneous or intermediate group.

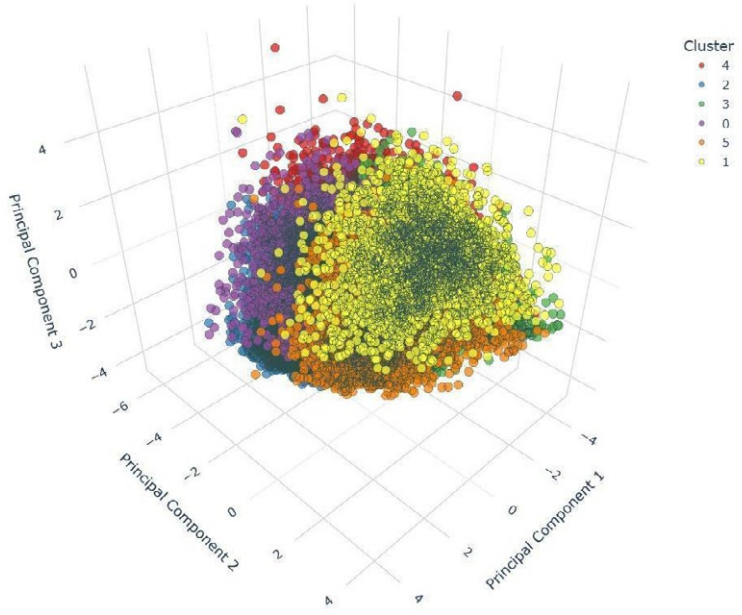
Figure 4. Two-Dimensional Projection of Clusters (PCA 2D)



Note. Own elaboration based on ENUT 2021 (DANE). The projection shows the six clusters distributed along the first two principal components, evidencing partial overlap but clear concentration zones that justify the segmentation.

However, the three-dimensional projection based on the first three principal components allows for a richer observation of the geometric configuration and relative spatial distribution of the clusters. Unlike the two-dimensional plane, this representation reveals depth, volume, and overlapping patterns that are not readily apparent at first glance.

Figure 5. Three-Dimensional Projection of Clusters (PCA 3D)



Note. Own elaboration based on ENUT 2021 (DANE). The 3D projection highlights the separation between clusters more clearly than the 2D view, showing distinct groupings along the three principals.

- Red (Cluster 0): Occupies a lower position along the vertical axis, forming a broad and slightly flattened base at the bottom of the graph. Its shape suggests a relatively stable grouping, expanding horizontally while showing more defined vertical boundaries.
- Dark Blue (Cluster 1): Appears distributed in a higher and somewhat peripheral region of the three-dimensional space, with more dispersed and less concentrated points. Its marginal location and loose distribution may indicate greater internal variability or separation from the core of the data.

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

- Purple (Cluster 2): Projects toward a lateral edge of the space, where it forms a dense, elongated structure, with no clear mixing with other groups. Its spatial orientation seems to follow a well-defined axis, indicating a cohesive and directional cluster.
- Yellow (Cluster 3): Notable for its volume, as it expands significantly in multiple directions. It occupies a deep, lateral zone of the graph, with high density in the center and somewhat diffuse edges. This almost spherical shape suggests a broad and potentially diverse structure.
- Pink (Cluster 4): Rather than concentrating in a specific region, it is dispersed around several other clusters, adopting a more enveloping or peripheral geometry. This layout reinforces the idea that it may function as a liminal group, closer to the outer edges of the set.
- Gray (Cluster 5): Tightly concentrated in the center of the three-dimensional space, with a dense and compact shape. It is the most embedded group among the others, suggesting an intermediate or shared position across multiple profiles. Its central cohesion may also reflect characteristics common to several groups.

Once these groupings were identified, a characterization was carried out based on a sociodemographic profile, household condition and labor situation by cluster, and traditionalism index and time use by cluster.

Table 5 presents a statistical summary of the sociodemographic profile of the identified clusters, based on six variables: age, rural occupation (as farmer), socioeconomic stratum, educational level, and ethnic affiliation (Afro-Colombian). For each cluster, the average values and standard deviations are reported, allowing for the observation of distinct patterns within the analyzed population.

Table 5. Sociodemographic Profile of Clusters

Variable	C0	C1	C2	C3	C4	C5
Age (years)	32.30	54.33	37.86	47.52	57.67	44.47
Age (std. dev.)	10.32	16.28	11.97	17.26	15.61	16.82
% Rural (farmer)	0.25	0.38	0.09	0.72	0.19	0.44
(std. dev.)	0.43	0.49	0.29	0.45	0.39	0.50
Socioeconomic stratum	1.75	2.04	2.57	1.33	2.54	1.56
(std. dev.)	0.86	1.05	1.17	0.52	1.06	0.88
Education level	5.00	4.57	7.08	3.39	4.50	4.86
(std. dev.)	2.15	2.53	3.00	1.20	2.23	2.63
% Afro-Colombian	0.00	0.07	0.00	0.00	0.00	1.00
(std. dev.)	0.06	0.26	0.01	0.00	0.00	0.00

Note. Own elaboration based on ENUT 2021 (DANE). This table summarizes the mean sociodemographic characteristics and standard deviations for each of the six clusters.

Cluster C0 groups relatively young individuals (average age 32.3), with a medium level of education (5.00), a low proportion of rural workers (25%), and a low socioeconomic stratum (1.75). In contrast, C1 represents an older group (54.3 years), with lower educational attainment (4.57), a higher share of rural workers (38%), and a slightly higher stratum (2.04).

C2 is characterized by the highest level of education (7.08) and the highest socioeconomic stratum (2.57), with low rural representation (9%) and a relatively young population (37.9 years). In contrast, C3 comprises individuals from a very low stratum (1.33), the highest percentage of rural workers (72%), and the lowest education level among all groups (3.39), with a middle-aged population (47.5 years).

C4 shares similarities with C2 in terms of socioeconomic stratum (2.54) but has lower educational attainment (4.50) and a higher average age (57.7 years). Finally, C5 is distinguished by exclusively comprising Afro-Colombian individuals (100%), with a mid-range average age (44.5 years), low socioeconomic stratum (1.56), moderate education level (4.86), and a significant proportion of rural workers (44%).

These results reflect a clear segmentation of the population in sociodemographic terms, revealing the coexistence of contrasting profiles with respect to key aspects such as age, education, ethnic background, rural status, and socioeconomic level.

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

Next, Table 6 presents a statistical summary of household conditions and employment status, based on four variables: head-of-household status, access to household goods (washing machine), employment situation, and religious affiliation.

Table 6. Household Condition and Employment Status

Variable	C0	C1	C2	C3	C4	C5
% Head of household	0.25	0.42	0.37	0.41	0.39	0.46
(std. dev.)	0.43	0.49	0.48	0.49	0.49	0.50
% Household with washer	0.64	0.71	0.87	0.25	0.97	0.66
(std. dev.)	0.48	0.45	0.33	0.43	0.18	0.47
% Employed	0.16	0.22	0.84	0.20	0.07	0.33
(std. dev.)	0.37	0.41	0.37	0.40	0.25	0.47
% Religious	0.00	1.00	0.00	0.00	0.00	0.00
(std. dev.)	0.04	0.00	0.00	0.00	0.00	0.00

Note. Own elaboration based on ENUT 2021 (DANE). This table summarizes household characteristics and employment status across the six identified clusters.

The percentage of individuals identified as heads of household varies across clusters, being lowest in C0 (25%) and highest in C5 (46%). This pattern may be related to the average age observed in each group. Regarding household conditions, access to a washing machine serves as a relevant indicator of material well-being, showing high coverage in C4 (97%) and C2 (87%), in contrast to C3 (25%), which reveals more precarious living conditions.

With respect to employment status, Cluster C2 has the highest proportion of individuals reporting employment (84%), which aligns with its previously observed higher level of education and socioeconomic status. Conversely, C4 and C0 show the lowest employment rates (7% and 16%, respectively), which could be linked to factors such as older age in C4 or youth and economic dependency in C0.

As for religious affiliation, only Cluster C1 reports 100% religious' identification, while this trait is absent in all other clusters. This may suggest a specific cultural or community identity within that particular group.

Taken together, these results demonstrate significant differences in household conditions and labor trajectories across clusters, associated with other previously described sociodemographic factors. This reinforces the structural diversity of the analyzed population.

Lastly, Table 7 presents a statistical summary of women's time use and the traditionalism index across clusters. This analysis includes seven variables: the traditionalism index, time allocated to personal self-care, food provision, clothing maintenance, cleaning activities, caregiving for others, and leisure time.

Table 7. Traditionalism and Time Use

Variable	C0	C1	C2	C3	C4	C5
Traditionalism Index	1.83	2.01	1.55	2.01	1.89	2.02
(std. dev.)	0.53	0.56	0.47	0.54	0.55	0.51
Personal self-care (min)	603.11	639.03	573.86	638.51	634.44	615.10
(std. dev.)	114.18	127.20	107.45	144.41	138.30	166.73
Food provision (min)	142.60	122.78	59.52	131.02	132.82	101.17
(std. dev.)	78.11	80.69	59.20	80.12	79.26	71.95
Clothing maintenance (min)	39.32	36.14	11.82	45.05	33.06	40.50
(std. dev.)	52.55	49.77	28.16	57.58	47.77	58.16
Cleaning activities (min)	80.13	70.88	35.42	64.56	84.66	59.57
(std. dev.)	60.95	64.06	45.89	55.62	71.19	56.90
Care for others (min)	137.74	9.23	7.40	6.52	2.87	17.20
(std. dev.)	66.18	33.24	22.43	19.23	12.72	39.33
Leisure time (min)	56.16	80.12	61.22	68.87	70.62	72.37
(std. dev.)	67.67	76.34	73.52	77.04	78.18	74.97

Note. Own elaboration based on ENUT 2021 (DANE). This table presents the average traditionalism index and daily time allocation (in minutes) to key activities for each cluster, along with standard.

Clusters C1, C3, and C5 exhibit the highest levels of traditionalism (above 2.0), which is associated with greater dedication to domestic and caregiving tasks, while C2 records the lowest value on this index (1.55), suggesting a less traditional profile. In terms of time devoted to personal self-care, values are relatively homogeneous across clusters, averaging close to 10 hours per day, although C2 shows the lowest level (573.9 minutes).

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

Regarding domestic tasks, Cluster C2 also stands out for spending significantly less time on food provision (59.5 minutes), clothing maintenance (11.8 minutes), and cleaning (35.4 minutes), reinforcing its less traditional profile. In contrast, C0 and C3 dedicate more time to these activities—particularly caregiving for others, where C0 reports the highest value (137.7 minutes), followed distantly by C5 (17.2 minutes), while C4, C3, and C2 show minimal levels in this category.

With respect to leisure time, differences among clusters are less pronounced, although C1, C4, and C5 show the highest values (above 70 minutes), which may reflect greater availability of free time in these groups. In contrast, C0 presents the lowest level of leisure time (56.2 minutes), possibly due to its heavy caregiving burden.

Taken together, the results show a clear differentiation between clusters in terms of gender roles and time allocation, with implications for how daily life, domestic responsibilities, and personal opportunities are experienced —especially for women. These findings help identify more or less traditional profiles, potentially influenced by the previously described sociodemographic factors.

The analysis of the emerging profiles thus made it possible to identify distinct configurations of time organization and gender attitudes. The resulting clusters reveal both generational contrasts and patterns associated with labor participation, rurality, and ethnic identity. Below is a qualitative summary of the six resulting profiles.

These six profiles reveal the heterogeneity in how women organize their time, shaped by factors such as age, labor force participation, rurality, educational attainment, and ethnic identity. The results confirm that gender inequalities in time use are intricately connected to other axes of social inequality.

Cluster 0 (Young women with a high caregiving burden): This group consists mostly of young women (average age 32.3), from low socioeconomic backgrounds, with a medium level of education and low labor participation (16%). Despite their age, they dedicate significant time to caregiving for others (137.7 minutes daily), as well as domestic activities like food provision (142.6 min) and cleaning (80.1 min). Their traditionalism index is moderately high (1.83), suggesting the persistence of

traditional gender roles even in younger contexts. This profile reflects the overburden of unpaid care work among young women, possibly within extended family structures.

Cluster 1 (Older religious women with strong traditional orientation):

With an average age of 54.3 and 100% religious' affiliation, this group shows one of the highest levels of traditionalism (2.01). Although labor force participation is low (22%), they spend substantial time on domestic tasks such as food provision (122.8 min) and household maintenance. Notably, caregiving time is very low (9.2 min), which may relate to their life stage, where they are no longer responsible for dependent children. This profile reflects a daily life shaped by traditional norms, but with reduced direct caregiving responsibilities.

Cluster 2 (Young, educated women with greater autonomy): This cluster includes women in their late 30s (average 37.9 years), with the highest education level (7.08), highest socioeconomic status, and the highest labor force participation (84%). It is marked by the lowest traditionalism index (1.55) and a time-use distribution reflecting greater autonomy: minimal time spent on caregiving (7.4 min) and other domestic tasks such as cleaning (35.4 min) and food preparation (59.5 min). This profile represents a break from traditional gender models, with a time organization focused on productive activities and reduced involvement in unpaid care work.

Cluster 3 (Rural women in vulnerable situations): With a high proportion of rural workers (72%), low education (3.39), and very low socioeconomic status (1.33), this group reflects structural vulnerability. Despite a high traditionalism index (2.01), caregiving time is low (6.5 min), possibly due to time constraints or external labor demands. However, they carry significant burdens in domestic work, such as food provision (131.0 min) and clothing maintenance (45.0 min). This profile shows how poverty and rurality influence a time-use pattern focused on household subsistence and reproduction, within traditional roles but lacking institutional or community support.

Cluster 4 (Older women with material well-being but disconnected from caregiving): This group has the highest average age (57.7 years) and the greatest access to household goods (97% have a washing machine). They show a relatively high

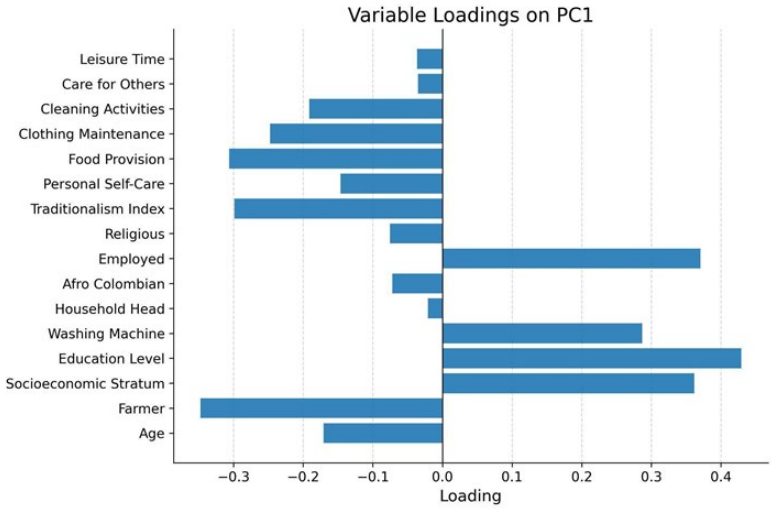
Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

traditionalism index (1.89), yet very low labor participation (7%). Despite their favorable socioeconomic profile, they dedicate very little time to caregiving (2.9 min), likely due to age-related detachment from such responsibilities. Domestic tasks remain elevated, particularly cleaning (84.7 min). This profile suggests a traditional domestic role division, but without active caregiving duties.

Cluster 5 (Afro-Colombian women with a double burden of care and work): This group is composed entirely of Afro-Colombian women (100%), from low socioeconomic strata (1.56) and with medium education. They show intermediate labor participation (33%) and a strong presence in household headship roles (46%). Despite a high traditionalism index (2.02), they bear a double burden: significant time is dedicated to domestic tasks (food: 101.2 min; cleaning: 59.6 min) as well as caregiving (17.2 min). This profile reveals tension between caregiving responsibilities and the need to earn income, likely shaped by structural and racial inequalities.

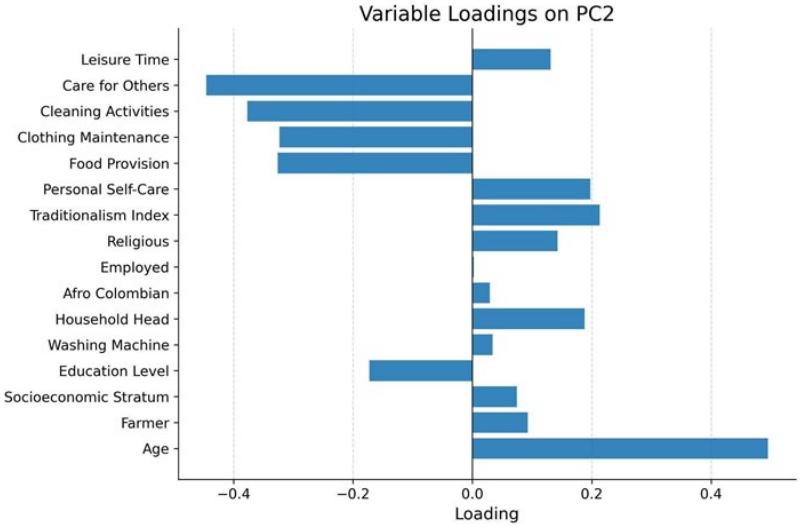
Finally, the analysis was complemented by a detailed PCA of factor loadings. Figures 6 to 9 show the latent structure of the variables, highlighting that the first principal component primarily reflects socioeconomic and demographic dimensions, while the subsequent components capture variability in time-use patterns. These results confirm that differences in time use among the various profiles of women in the sample are structured around social and gender factors, consistently reflected across the different analytical methods applied.

Figure 6. Variable Loadings on Principal Components (PCA)



Note. Own elaboration based on ENUT 2021 (DANE). The figure displays the variable contributions to the first two principal components, highlighting that education level, socioeconomic stratum, and employment are the main drivers of PC1, while rurality (farmer) and age dominate PC2.

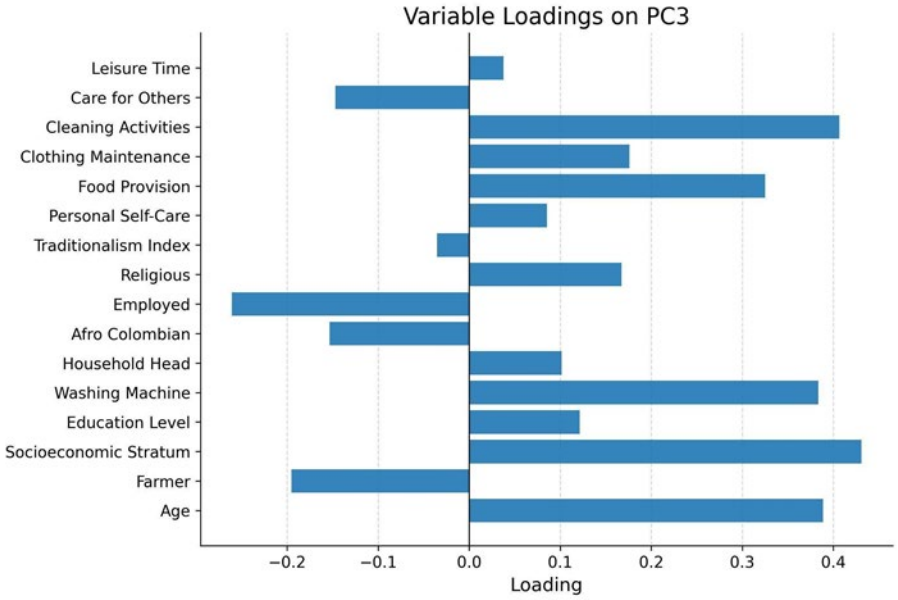
Figure 7. Variable Loadings on the Second Principal Component (PC2)



Note. Own elaboration based on ENUT 2021 (DANE). This figure shows that leisure time, care for others, and cleaning activities have the highest positive or negative loadings on PC2, indicating that this component primarily contrasts time spent on unpaid care and leisure-related activities.

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

***Figure 8. Variable Loadings on the Third Principal Component* (PC3)**

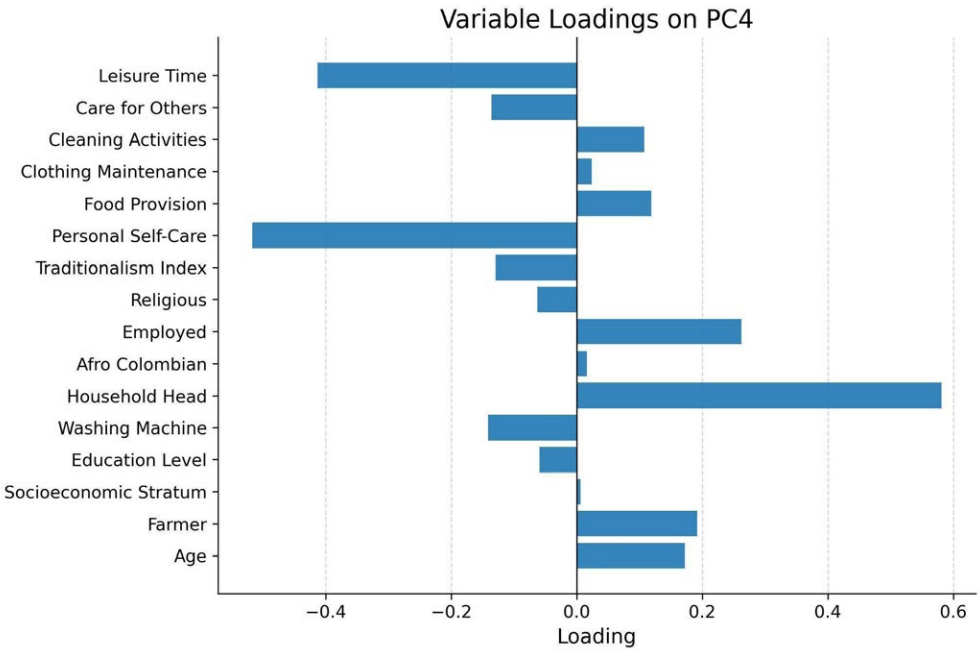


Note. Own elaboration based on ENUT 2021 (DANE). This figure shows that clothing maintenance, food provision, washing machine availability, socioeconomic stratum, and age are the main contributors to PC3, highlighting that this component captures household infrastructure and domestic work intensity.

The principal component analysis revealed the latent dimensions structuring variability in time-use patterns and sociodemographic characteristics. Examination of the factor loadings for the first four components provides differentiated interpretative axes.

The first component is characterized by strong positive loadings on educational attainment, socioeconomic stratum, access to household goods (washing machine), and employment status. In contrast, it shows negative loadings on farming occupation, age, food provision, and cleaning activities. This pattern indicates that PC1 synthesizes an axis of socioeconomic status and human capital, distinguishing women with higher educational and material resources —along with greater labor force participation— from those associated with lower strata, rurality, and heavier engagement in domestic work. It therefore captures structural inequalities related to class and territory.

Figure 9. Variable Loadings on the Fourth Principal Component (PC4)



Note. Own elaboration based on ENUT 2021 (DANE). This figure highlights that household headship, employment, and personal self-care have the strongest contributions (positive and negative) to PC4, suggesting this component captures contrasts in household responsibility and individual time allocation.

The second component displays notable negative loadings on care for others, food provision, clothing maintenance, and cleaning activities, and positive loadings on leisure time, personal self-care, traditionalism, and religiosity. This structure reflects an axis of reproductive workload versus personal well-being and traditional attitudes. High scores on this component are associated with profiles where leisure and self-care predominate over direct caregiving responsibilities, framed within traditional values, while low scores denote greater dedication to reproductive and domestic tasks.

The third component reveals positive loadings on cleaning activities, food provision, personal self-care, and access to household goods, along with associations with socioeconomic stratum and age. Negative loadings appear for employment status and, to a lesser extent, for Afro-Colombian identification. This suggests an axis contrasting labor market

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

participation and income diversification with a pattern of time-intensive domestic tasks and household maintenance. It highlights the tension between paid work and domestic responsibilities, showing how increases in one dimension tend to exclude the other.

The fourth component is dominated by very high positive loadings on household headship, accompanied by associations with employment and access to goods, while showing negative loadings on personal self-care, leisure time, and age. This pattern indicates an axis of domestic responsibility and family leadership versus individual well-being. Women scoring highly on this component are identified as main household providers, simultaneously assuming productive and reproductive responsibilities at the expense of their own personal time.

4. Discussion

The results make visible how the organization of women's time in Colombia remains structured around a persistent sexual division of labor, shaped by socioeconomic, territorial, and ethno-racial inequalities. From the perspective of feminist economics, this segmentation reflects not only individual practices but also a social architecture of care sustained by family arrangements and structural inequalities (Carrasco, 2006; Pérez Orozco, 2012).

One of the most revealing findings is the coexistence of profiles with unequal unpaid care burdens. For example, Cluster 0 groups young women (average age 32), with low labor force participation (16%) and high caregiving time (137 minutes daily), along with a relatively high level of traditionalism (1.83) (Table 7). This configuration reflects what feminist theory has described as the "sustainability of life at the expense of women" (Pérez Orozco, 2012), where the youngest, with less economic autonomy, perform essential tasks for the well-being of others without social or economic recognition.

In contrast, Cluster 2 represents women with the highest educational level (7.08) and the highest labor force participation (84%). They dedicate significantly less time to domestic and caregiving work (only 7.4 minutes daily for caregiving and 35.4

minutes for cleaning). They also have the lowest traditionalism index in the entire sample (1.55) (Tables 5 and 7). This pattern can be interpreted as a break from traditional gender mandates. However, as Fraser (2011) warns, the outsourcing of care work to other (often more precarious) women may reproduce new forms of inequality if no structural redistribution takes place.

The analysis also reveals how structural factors such as rurality, education, and social class combine to shape differentiated trajectories. Cluster 3, composed mostly of rural women (72%), with the lowest level of education (3.39) and socioeconomic status (1.33), exhibits high levels of traditionalism (2.01) and a notable domestic workload (131 minutes on food preparation, 45 on clothing maintenance), but low interpersonal caregiving time (6.5 minutes) (Tables 5 and 7). This may reflect a context where available time is focused on subsistence, with little institutional or community support for reproductive tasks.

The ethno-racial dimension is strongly represented in Cluster 5, composed exclusively of Afro-Colombian women, who, in addition to performing caregiving tasks (17.2 minutes daily) and domestic work (101 minutes on food preparation), also show notable labor participation (33%) and high head-of-household status (46%) (Tables 5, 6 and 7). Despite their high traditionalism index (2.02), this group embodies the double burden of paid and unpaid labor under structurally vulnerable conditions, confirming the importance of considering racialization as a vector of inequality in the organization of care (Crenshaw, 1991; Hill Collins, 2000).

At the opposite end, Cluster 4 comprises older women (57.7 years), from higher socioeconomic strata, with substantial access to household goods (97% own a washing machine), and low labor force participation (7%). Despite moderate levels of education and traditionalism (1.89), they spend very little time on caregiving (only 2.9 minutes daily), though they still dedicate relevant time to household tasks like cleaning (84.7 minutes) (Tables 6 and 7). This suggests a continuity of reproductive roles in the absence of direct dependents. It also confirms that access to material well-being alone does not necessarily lead to a transformation in the sexual division

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

of labor (Carrasco, 2006).

The patterns of time use, traditionalism, and household conditions are also synthesized in the latent structure of the principal components (Figures 6 to 9), which confirms that differences between profiles are organized around key social axes: socioeconomic level, education, ethnicity, and gender.

In sum, the results provide empirical confirmation of what feminist theory has long argued: that the sustainability of life in Colombia rests on a network of unequal arrangements in which women —particularly those in more vulnerable positions— subsidize the state and the market with their time (Benería, 2003; Moreno Salamanca, 2017). The segmentation identified reveals not only heterogeneity but also clearly demonstrates how power relations and inequalities operate in everyday life, significantly shaping the possibilities for agency, rest, and autonomy.

This evidence calls for public policies that are sensitive to the intersection of multiple inequalities, promote a fair redistribution of care work, recognize its social value, and ensure it does not continue to be a privatized and feminized burden.

5. Conclusions

This study applied unsupervised machine learning techniques —specifically the K-means algorithm— to identify differentiated profiles of women in Colombia based on their time use, sociodemographic characteristics, and attitudes toward gender roles. Using data from the 2021 National Time Use Survey (ENUT) and constructing synthetic variables, including a traditionalism index, six clusters were identified with contrasting patterns in the organization of domestic and unpaid care work.

The results confirm that although all women in the sample are subject to a structurally embedded sexual division of labor, this does not manifest uniformly. On the contrary, sharply differentiated profiles emerge based on age, educational attainment, employment status, rurality, and ethno-racial background. These findings reinforce the need to adopt an intersectional perspective to understand the distribution of care work and to design appropriate public policies.

The analysis highlighted, for instance, the situation of young women with a high care burden (Cluster 0), as well as the profile of Afro-Colombian women facing a double burden of caregiving and employment under conditions of structural disadvantage (Cluster 5). A group of women with greater autonomy and fewer reproductive responsibilities (Cluster 2) was also identified, although this does not necessarily imply a fair redistribution of care, but may instead involve delegation mechanisms to other, more precarious women.

The empirical findings support the claims of feminist economics, as they demonstrate how care work —unpaid, invisible, and deeply feminized— remains a fundamental pillar of life's sustainability, yet goes unrecognized in traditional economic metrics. Moreover, they show that access to material or educational resources alone does not guarantee a transformation in gender patterns, as seen in the case of older women with relative material well-being but adherence to traditional domestic roles (Cluster 4).

From a methodological standpoint, the use of clustering techniques on time-use data offers a powerful tool to capture internal heterogeneity among women, moving beyond aggregate approaches that tend to homogenize their experiences. The identification of latent profiles provides valuable input for advancing toward public policies that are more responsive to diversity and oriented toward the real redistribution of care work.

Finally, the study acknowledges its limitations. As an exploratory analysis based on cross-sectional data, it does not allow for causal inference or the capture of temporal dynamics. Future research could incorporate longitudinal methods, cross-country comparisons, or qualitative approaches that deepen the understanding of individual trajectories and the meanings attributed to care.

In sum, this research contributes both empirical and conceptual evidence to redirect public debate toward a more just, democratic, and inequality-conscious social organization of care —one that reflects the complex everyday realities of women in Colombia.

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

6. References

1. Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (pp. 1027–1035). Society for Industrial and Applied Mathematics.
2. Benería, L. (2003). Gender, development, and globalization: Economics as if all people mat-tered. Routledge.
3. Carrasco, C. (2006). Invisibles y fundamentales: El trabajo de cuidados en Europa. Icaria. Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. Stanford Law Review, 43(6), 1241-1299.
4. Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data —or tears: An application to educational enrollments in states of India. Demography, 38(1), 115–132. https://doi.org/10.1353/dem.2001.0003
5. Fraser, N. (2011). Fortunas del feminismo: Del capitalismo gestionado por el Estado a la crisis neoliberal. Traficantes de Sueños.
6. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media.
7. Hill Collins, P. (2000). Black feminist thought: Knowledge, consciousness, and the politics of empowerment. Routledge.
8. Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90-95.
9. Jolliffe, I. T. (2002). Principal Component Analysis (2nd ed.). Springer.
10. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society: A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.
11. Kolenikov, S., & Angeles, G. (2009). Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer? Review of Income and Wealth, 55(1), 128–165. https://doi.org/10.1111/j.1475-4991.2008.00309.x

12. MacQueen, J. (1967a). Some Methods for Classification and Analysis of Multiva- riate Observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, 281-297.
13. MacQueen, J. (1967b). Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, 1, 281-297.
14. McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, 51-56.
15. Moreno Salamanca, E. N. (2017). Organización social del cuidado y medición del trabajo no remunerado en Colombia. Colombia Internacional, (91), 23-50.
16. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Courna- peau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.
17. Pérez Orozco, A. (2012). Subversión feminista de la economía: Aportes para un debate sobre el conflicto capital-vida. Traficantes de Sueños.
18. Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53-65. https://doi.org/10.1016/0377-0427(87)90125-7
19. Shlens, J. (2014). A tutorial on principal component analysis [arXiv:1404.1100]. https://arxiv.org/abs/1404.1100
20. Thorndike, R. L. (1953). Who belongs in the family? Psychometrika, 18(4), 267-276. https://doi.org/10.1007/BF02289263
21. Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: How to use principal components analysis. Health Policy and Planning, 21(6), 459–468. https://doi.org/10.1093/heapol/czl029
22. Waskom, M. (2021). Seaborn: Statistical Data Visualization. Journal of Open Source Software, 6(60), 3021.

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

Appendix

This section presents supplementary material to the main analysis, including visualizations and descriptive tables associated with the results of the K-means model for different values of K, applied to the female population. It includes distributions, spatial projections of the clusters, and a detailed characterization of the identified profiles.

Appendix A

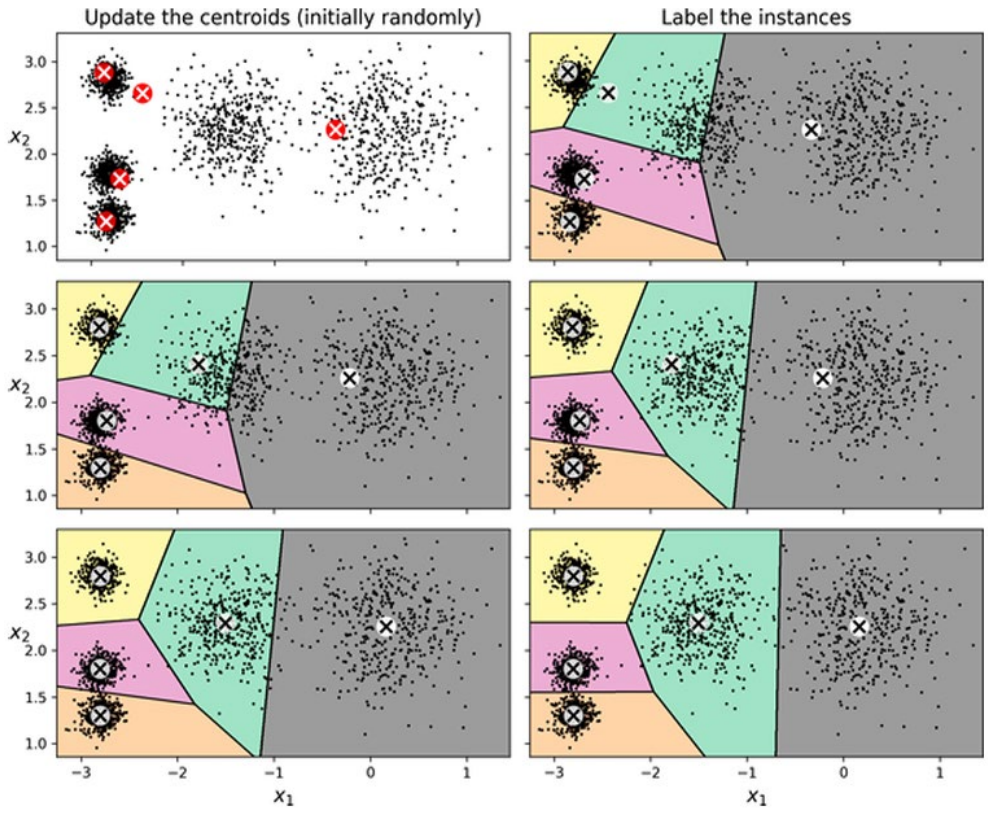
Survey Items Used to Construct the Traditionalism Index Respondents expressed their degree of agreement on a 1–4 Likert scale (1 = strongly disagree, 4 = strongly agree). For items S1, S2, and S5, values were reverse-coded (5 – response) so that higher scores consistently reflect stronger adherence to traditional gender norms.

1. P1107S1. "A mother who works outside the home is as good a mother as one who only works at home" (reverse-coded).
2. P1107S2. "Both men and women should contribute to household income" (reverse coded).
3. P1107S3. "A woman's main goal is to marry and have children".
4. P1107S4. "Women are better suited for domestic work than men".
5. P1107S5. "Women have the same right as men to go out for leisure" (reverse-coded).
6. P1107S6. "The household head should be a man".

Appendix B. Illustration of the K-means Algorithm

This appendix provides a schematic example of the iterative procedure followed by the K-means algorithm, using a hypothetical dataset for illustrative purposes. The aim is to visually demonstrate the steps involved in partitioning observations into clusters.

Figure B1. Schematic illustration of the iterative K-means process



Note. Adapted from Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media.

Step 1. Initialization. Centroids are placed in the data space. In the standard algorithm, they are chosen randomly, although in practice the k-means initialization (Arthur & Vassilvitskii, 2007) is recommended to improve stability and convergence.

Step 2. Assignment. Each observation is assigned to the nearest centroid based on Euclidean distance, forming provisional clusters. This produces a partition of the data into Voronoi regions.

Step 3. Update. Centroids are recalculated as the mean position of the points assigned to each cluster. This update reflects the central tendency of the group.

Step 4. Iteration. Steps 2 and 3 are repeated iteratively. With each cycle, centroids move closer to stable positions, and cluster assignments adjust accordingly.

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

Step 5. Convergence. The algorithm stops when centroid positions no longer change significantly or when a pre-defined maximum number of iterations is reached. At this point, clusters are considered stable.

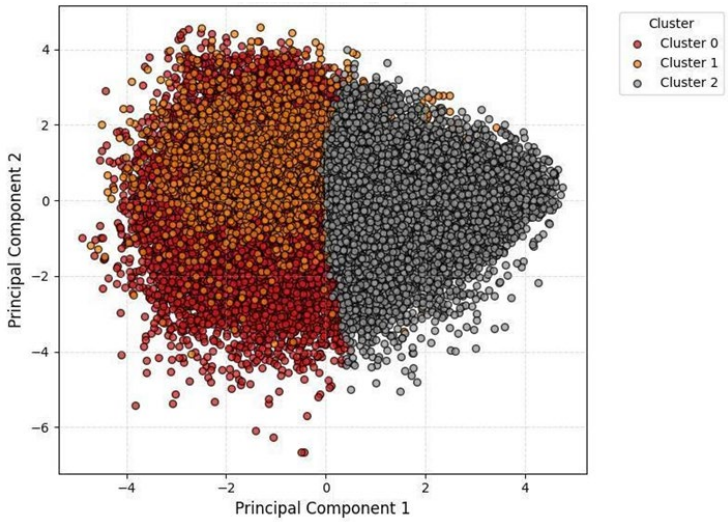
This iterative process ensures that the within-cluster sum of squares (WSS) is minimized, thereby maximizing the compactness of groups and the separation between them. Although K-means is sensitive to outliers and the initial placement of centroids, the use of k-means and multiple random restarts greatly improves the reliability of the solution.

Appendix C. Clustering Results with K = 3

Table C1. Distribution of records by Cluster (K = 3)

Cluster	N° of cases	% of total
Cluster 0	25,138	53,04%
Cluster 1	2,266	4,78%
Cluster 2	19,988	42,18%

Figure C1. Two-Dimensional projection of Clusters (PCA 2D), K = 3



Note. Own elaboration. Source: ENUT 2021, DANE.

Table C2. Sociodemographic characterization by Cluster

Variable	C0	C1	C2
Sex (1 = Male)	1.00	1.00	1.00
(std. dev.)	0.00	0.00	0.00
Age	49.35	54.31	41.03
(std. dev.)	17.95	16.28	14.71
Farmer	0.50	0.38	0.09
(std. dev.)	0.50	0.48	0.29
Socioeconomic status	1.66	2.04	2.64
(std. dev.)	0.79	1.05	1.20
Educational level	3.71	4.58	6.79
(std. dev.)	2.57	2.63	2.00

Table C3. Household condition and labor situation

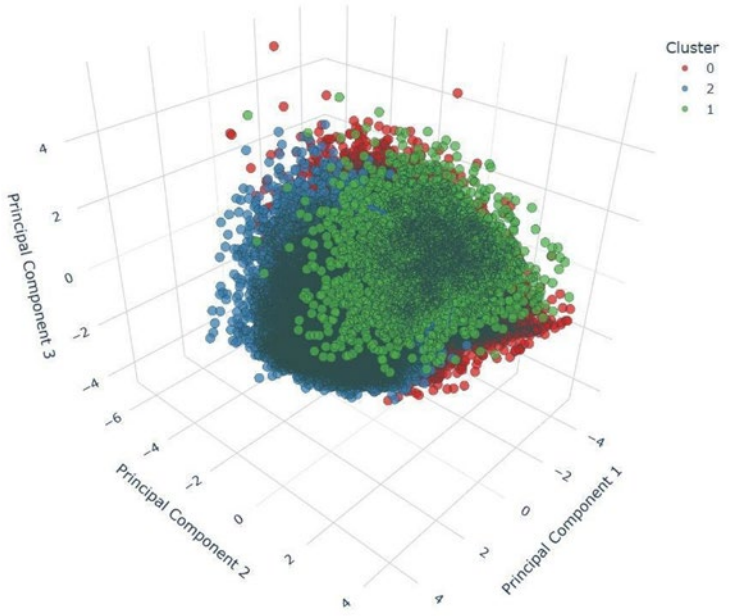
Variable	C0	C1	C2
% Head of household	0.38	0.42	0.38
(std. dev.)	0.49	0.49	0.48
% Household with washing machine	0.56	0.71	0.88
(std. dev.)	0.50	0.45	0.33
% Employed	0.11	0.22	0.68
(std. dev.)	0.32	0.41	0.47
% Religious	0.00	1.00	0.00
(std. dev.)	0.00	0.00	0.00

Table C4. Traditional values and time allocation

Variable	C0	C1	C2
Traditionalism Index	2.01	2.01	1.58
(std. dev.)	0.54	0.57	0.47
Personal care (min)	638.18	638.85	580.47
(std. dev.)	143.44	127.45	113.16
Food provision (min)	138.91	122.87	70.58
(std. dev.)	78.99	80.80	65.41
Clothing maintenance (min)	44.38	36.28	13.42
(std. dev.)	56.16	49.86	29.64
Cleaning activities (min)	77.35	71.02	43.09
(std. dev.)	63.61	64.12	52.91
Caring for others (min)	22.02	10.02	14.79
(std. dev.)	51.75	36.61	41.89
Leisure time (min)	69.87	80.05	61.10
(std. dev.)	76.67	76.32	73.68

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

Figure C2. Three-Dimensional projection of Clusters (PCA 3D), K = 3



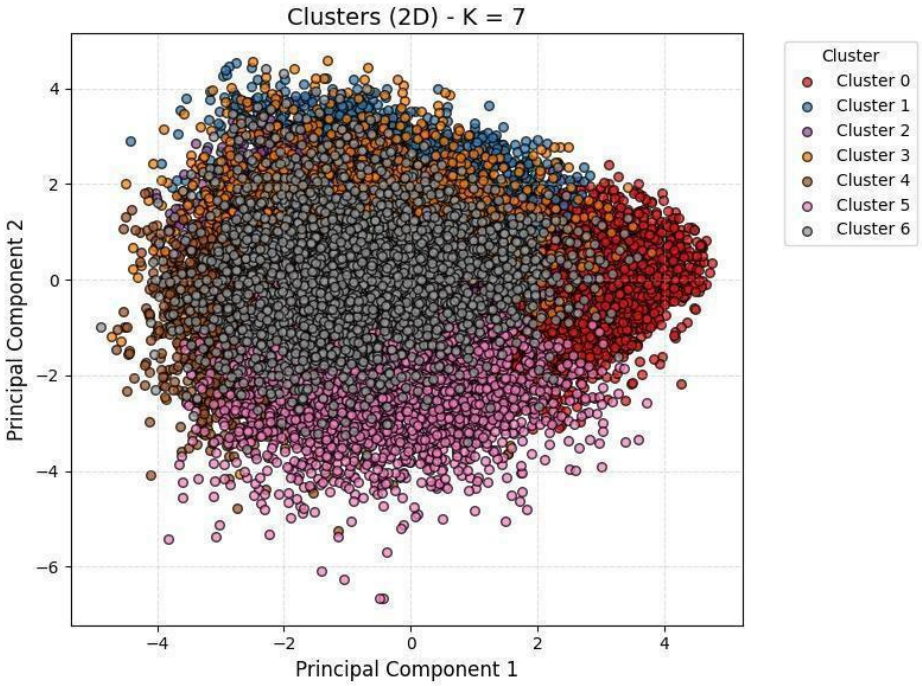
Note. Own elaboration. Source: ENUT 2021, DANE.

Appendix D. Clustering Results with

Table D1. Distribution of Records by Cluster (K = 7)

Cluster	N° of cases	% of total
Cluster 0	12,812	27,04%
Cluster 1	8,463	17,86%
Cluster 2	9,191	19,40%
Cluster 3	2,259	4,77%
Cluster 4	8,096	17,09%
Cluster 5	3,842	8,11%
Cluster 6	2,721	5,74%

Figure D1. Two-Dimensional Projection of Clusters (PCA 2D), K = 7



Note. Own elaboration. Source: ENUT 2021, DANE.

Table D2. Sociodemographic characterization by Cluster

Variable	C0	C1	C2	C3	C4	C5	C6
Sex (1 = Male)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
(std. dev.)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Age	38.33	64.06	44.64	54.34	45.89	49.27	39.31
(std. dev.)	12.13	14.39	16.30	16.28	14.37	17.60	13.37
Farmer	0.05	0.25	0.68	0.38	0.40	0.50	0.07
(std. dev.)	0.23	0.43	0.47	0.49	0.49	0.50	0.26
Socioeconomic status	2.75	2.48	1.37	2.04	1.83	1.66	2.65
(std. dev.)	1.18	1.11	0.55	1.05	0.85	0.79	1.20
Educational level	7.50	3.94	3.58	4.57	4.23	3.71	6.79
(std. dev.)	1.47	2.35	2.49	2.64	2.42	2.57	2.00

Intercamb. Rev. Estud. Economía. N. 9
Medellín, enero-diciembre 2025, 127 pp. E-ISSN 2619-6131, pp. 44-86

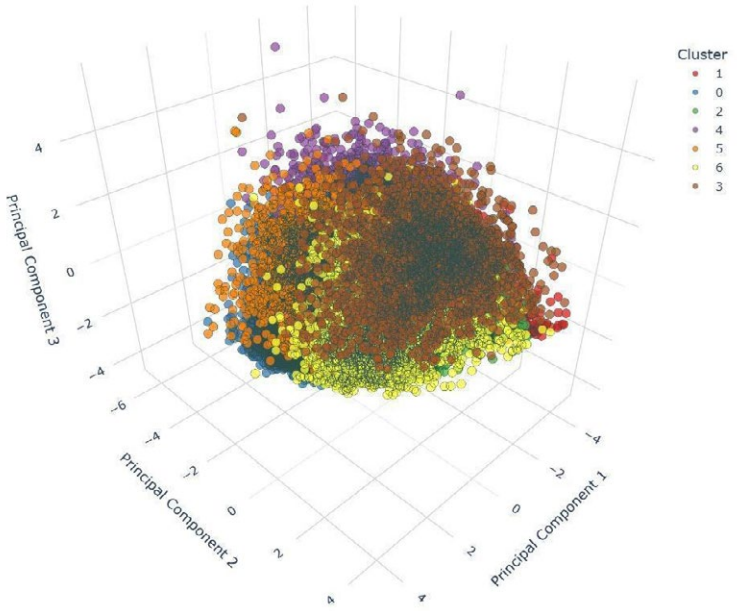
Table D3. Household Condition and labor situation

Variable	C0	C1	C2	C3	C4	C5	C6
% Head of household	0.35	0.44	0.47	0.42	0.29	0.25	0.46
(std. dev.)	0.48	0.50	0.50	0.49	0.46	0.43	0.50
% Household with washing machine	0.93	0.92	0.21	0.71	0.70	0.64	0.66
(std. dev.)	0.25	0.27	0.41	0.45	0.46	0.48	0.47
% Employed	0.80	0.05	0.38	0.22	0.11	0.16	0.33
(std. dev.)	0.40	0.22	0.48	0.41	0.31	0.37	0.47
% Religious	0.00	0.00	0.00	1.00	0.00	0.00	0.00
(std. dev.)	0.00	0.00	0.00	0.00	0.00	0.04	0.00

Table D4. Traditional values and time allocation

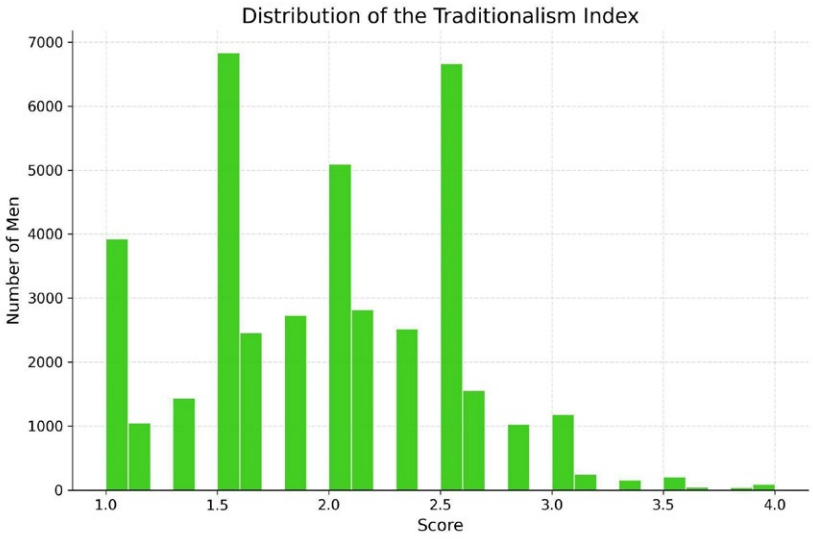
Variable	C0	C1	C2	C3	C4	C5
Traditionalism Index	1.52	2.04	1.92	2.01	1.89	1.83
(std. dev.)	0.45	0.57	0.52	0.56	0.53	0.53
Personal care (min)	575.37	673.17	599.86	639.08	625.73	604.48
(std. dev.)	106.09	163.13	131.55	127.20	112.32	116.52
Food provision (min)	63.58	101.90	102.84	122.75	181.26	140.10
(std. dev.)	61.93	74.79	69.29	80.70	73.33	77.13
Clothing maintenance (min)	11.54	11.98	19.21	36.16	89.42	32.31
(std. dev.)	26.03	24.87	32.92	49.77	60.21	45.76
Cleaning activities (min)	39.34	50.46	45.68	70.89	123.25	76.85
(std. dev.)	48.58	50.48	42.10	64.08	69.63	58.73
Caring for others (min)	7.43	1.96	7.52	9.13	8.39	144.71
(std. dev.)	22.82	11.00	20.48	32.89	23.08	65.48
Leisure time (min)	60.87	79.86	61.64	80.15	66.76	56.69
(std. dev.)	72.37	85.29	75.35	76.34	70.75	68.18

Figure D2. Three-Dimensional Projection of Clusters (PCA 3D), K = 7



Note. Own elaboration. Source: ENUT 2021, DANE.

Figure D3. Distribution of the Traditionalism Index for Men



Note. Own elaboration. Source: ENUT 2021, DANE.