



Medallion architecture and multiple correspondence analysis for the characterization of the montubio people in Guayas: a big data approach to intercultural tourism

.....

Arquitectura medallion y análisis de correspondencias múltiple para la caracterización del pueblo montubio en Guayas: una aproximación desde el big data al turismo intercultural

Roberto Marcelo Martínez Hinojosa*
Guillermo Enrique Valarezo Guzmán*
José Alfredo Castro Carrasco*
Richard Francisco García Alay*

Received: July 10, 2026

Approved: September 11, 2026

Martínez, R., Valarezo, G., Castro, J., García, R. (2026) Medallion architecture and multiple correspondence analysis for the characterization of the montubio people in Guayas: a big data approach to intercultural tourism. *Espiraless Revista Multidisciplinaria de investigación científica*, 10 (59), 89-100

Abstract

This article analyzes the potential of Big Data as a tool for valorizing the ancestral knowledge of the Montubio people in Guayas province, Ecuador. Through processing the 2022 Population and Housing Census microdata using a Bronze-Silver-Gold architecture on SQL Server 2022, 35,593 self-identified Montubio individuals were identified. Statistical analysis included chi-square tests, Cramér's V, and Multiple Correspondence Analysis (MCA) conditioned by area and canton. Results: 91.8% speak a native language, 64.3% have no formal education, 91.7% reside in urban areas. Five MCA dimensions explain 59.2% of total inertia. Strongest association: age group $\times$ economic activity ($\chi^2=150.5$; $V=0.073$; $p<0.001$).

Keywords: Big Data, intercultural tourism, Montubio people, ancestral knowledge, MCA, Guayas, 2022 Census.

* Doctor en ciencias economía Facultad de Ciencias económicas y sociales, Universidad de Zulia; Magister en Ciencias con Mención en Economía y Gestión Empresarial; Facultad de Ciencias Administrativas, Carrera de Gestión de la Información Gerencial; Universidad de Guayaquil; Guayaquil, Ecuador; roberto.martinez@ug.edu.ec;

<https://orcid.org/0000-0001-9759-3305>

* Magister en Seguridad Informática Aplicada; Ingeniero en Computación Especialización Sistemas Tecnológicos; Facultad de Ciencias Administrativas, Carrera de Gestión de la Información Gerencial; Universidad de Guayaquil; Guayaquil, Ecuador; guillermo.valarezog@ug.edu.ec;

<https://orcid.org/0009-0004-2211-8525>

* PhD Demografía; Máster en Matemáticas Aplicadas; Máster en Gestión Técnica de Seguros; Ingeniero en Estadística Informática; Facultad de Ciencias Administrativas; Carrera de Gestión de la Información Gerencial; Universidad de Guayaquil; Guayaquil, Ecuador; jose.castroca@ug.edu.ec;

<https://orcid.org/0009-0007-7693-7202>

* Máster Universitario en Análisis y Visualización de Datos Masivos; Ingeniero en Sistemas Computacionales; Facultad de Ciencias Administrativas, Carrera de Gestión de la Información Gerencial; Universidad de Guayaquil; Guayaquil, Ecuador; richard.garciaa@ug.edu.ec;

<https://orcid.org/0009-0003-2390-6534>

Resumen

El presente artículo analiza el potencial del Big Data como herramienta para la valorización de los saberes ancestrales del pueblo montubio en la provincia del Guayas, Ecuador, con énfasis en su integración a la oferta turística intercultural. Mediante el procesamiento de los microdatos del Censo de Población y Vivienda 2022 del INEC, utilizando una arquitectura de datos en capas Bronze-Silver-Gold sobre Microsoft SQL Server 2022, se identificaron 35.593 personas autoidentificadas como montubias en Guayas. El análisis estadístico multivariado incluyó pruebas de chi-cuadrado (χ^2), coeficiente V de Cramér y Análisis de Correspondencias Múltiple (ACM) general y condicionado por área y cantón. Los resultados revelan que el 91,8% habla lengua nativa, el 64,3% no posee instrucción formal y el 91,7% reside en zonas urbanas. Las cinco primeras dimensiones del ACM explican el 59,2% de la inercia total. La asociación más fuerte se registra entre grupo de edad y condición de actividad ($\chi^2=150,5$; $V=0,073$; $p<0,001$).

Palabras clave: Big Data, turismo intercultural, pueblo montubio, saberes ancestrales, ACM, Guayas, Censo 2022.

Introduction

Intercultural tourism has established itself as one of the fastest-growing forms of tourism in Latin America, driven by the demand for authentic experiences that connect visitors with local cultures (UNWTO, 2019). The Montubio people, recognized in Ecuador's 2008 Constitution as an ancestral people of the coast, possess a distinct cultural heritage that has not yet been systematically incorporated into Guayas' tourism offerings (CODEPMOC, 2020).

The Big Data revolution has transformed researchers' ability to understand large-scale social phenomena. Ecuador's 2022 Population and Housing Census (CPV 2022), with 16,938,986 national records and 86 variables per person, constitutes a paradigmatic example of Big Data in the social sciences (Mayer-Schönberger and Cukier, 2013). This article addresses the following question: How can Big Data analysis of the CPV 2022 contribute to the promotion of ancestral Montubio knowledge within Guayas' tourism offerings? To this end, the chi-square test, Cramér's V, and general and conditioned Multiple Correspondence Analysis (ACMc) are applied.

The Montubio is a mestizo inhabitant of the rural areas of the Ecuadorian coast, whose identity is forged at the confluence of Indigenous, African, and Spanish heritages (Muratorio, 2000). Their ancestral knowledge encompasses agroecological agriculture, medicinal ethnobotany, vernacular architecture using guadúa cane, the musical expressions of the amorfino, and horse training (CODEPMOC, 2020). This knowledge, transmitted orally, constitutes strategic intangible cultural heritage (UNESCO, 2003).

Intercultural tourism promotes genuine exchange between visitors and local communities, valuing cultural differences as experiential resources (Richards, 2018). The

UNWTO (2019) notes that destinations with cultural authenticity record an average tourist expenditure 23% higher than that of conventional destinations.

The Medallion architecture organizes processing into Bronze (raw data), Silver (cleaned data), and Gold (analytical data) layers. ACM, developed by Benzécri (1973) and systematized by Greenacre (2017), is the most appropriate multivariate technique for categorical variables. The ACMc controls for the effect of a supplementary variable to compare subgroups (Escofier and Pagès, 1998).

Materials and Methods

CPV 2022 Microdata (INEC): CPV_Population_2022_National.csv (16,938,986 records, 86 variables, 4.3 GB), CPV_Housing_2022_National.csv (6,611,555 records) and CPV_Household_2022_National.csv (5,193,548 records).

The Medallion architecture was implemented on Microsoft SQL Server 2022. Bronze: ingestion via BULK INSERT. Silver: Guayas filtering (I03=50, n=12,430,074) and Montubio subset (P22=5, n=35,593). Gold: analytical tables aggregated by canton and area.

For the ACM: seven active variables (area, language, education level, illiteracy, employment status, sex, age group) with 16 categories. Canton as a supplementary variable for the ACMc. Sequence applied: (1) descriptive statistics with IMP_VOPA expansion factors; (2) Pearson's chi-square (χ^2) between six pairs of variables; (3) Cramér's V; (4) ACM with five dimensions; (5) ACMc conditioned by area and by the four cantons with the highest concentration of Montubio. Analysis performed in Python 3.14 using Prince 0.19, SciPy 1.15, and Matplotlib 3.10.

Results

The 2022 CPV records 35,593 people who self-identify as Montubios in Guayas (0.29% of the provincial population). Table 1 shows the distribution by canton.

Table 1 *Distribution of the Montubio population by canton in Guayas (CPV 2022)*

Cantón	Código	Total personas	Pobl. expandida	Edad promedio
Guayaquil	901	6.760	13.174	39,9
Milagro	1701	6.145	11.831	40,7
Daule	2301	1.001	1.957	38,1
Balzar	1308	1.006	1.950	40,1
Babahoyo	1301	909	1.748	40,6
Otros cantones	—	19.772	38.665	39,6
Total Guayas	—	35.593	69.325	39,8

Note. Source: Author's own analysis based on the 2022 Population Census (INEC). SQL Server 2022, Silver tier. n=35,593.

Guayaquil accounts for 19.0% of the Montubio population (6,760 people), followed by Milagro (17.3%). The average age of 39.8 years indicates an adult population fully capable of transmitting knowledge and managing tourism ventures.

Table 2 shows the distribution by area of residence.

Área	Total	Total expandido	Porcentaje (%)
Urbano	32.631	63.501	91,7
Rural	2.962	5.859	8,3
Total	35.593	69.360	100,0

Note. Source: Author's own calculations based on the 2022 Population Census (CPV) (INEC). Variable AUR.

91.7% of the Montubio population resides in urban areas, contrasting with the predominantly rural image associated with this people. This finding implies that tourism strategies must include both rural authenticity routes (Daule, Balzar) and urban cultural tourism circuits in Guayaquil and Milagro.

Table 3 Educational Level of the Montubio Population in Guayas (CPV 2022)

Nivel de instruction	Total	Porcentaje (%)
Ninguno	22.895	64,3
Alfabetización	11.904	33,4
Primaria	794	2,2
Total	35.593	100,0

Note. Source: Author's own analysis based on the 2022 CPV (INEC). Variable P08.

64.3% have no formal education, and 33.4% have only completed literacy training. This demonstrates that the Montubio knowledge system is fundamentally oral and experiential, which reinforces its value as a distinct form of ancestral knowledge for intercultural tourism.

Table 4 Everyday Language of the Montubio Population in Guayas (CPV 2022)

Idioma cotidiano	Total	Porcentaje (%)
Lengua nativa	32.664	91,8
Español	2.929	8,2
Total	35.593	100,0

Note. Source: Author's own analysis based on the CPV 2022 (INEC). Variable P15.

91.8% speak their native language as their everyday language. This exceptional indicator of linguistic vitality constitutes the main distinguishing asset for intercultural

tourism: the experience of interacting in an ancestral language is a value that no other destination can replicate.

Table 5 presents the results of the association tests between six pairs of variables. Figure 1 visually complements these results using a heat map.

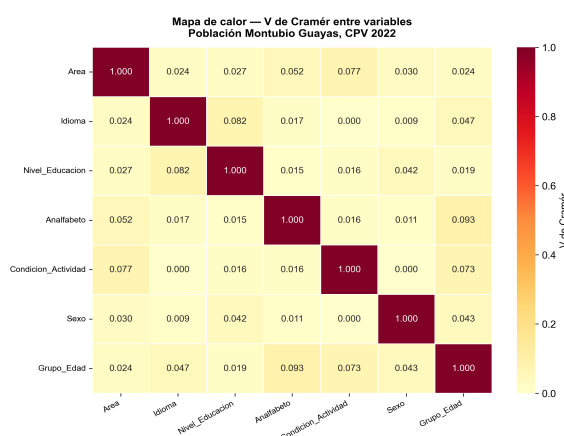
Variables	χ^2	gl	p-valor	Sig.	V Cramér	Asociación
Área × Nivel educación	21,1	2	$2,64 \times 10^{-5}$	***	0,027	Débil
Área × Idioma	15,8	2	$7,02 \times 10^{-5}$	***	0,024	Débil
Nivel educ. × Actividad	7,4	4	$2,48 \times 10^{-2}$	*	0,016	Débil
Idioma × Actividad	0,0	2	1,00	ns	0,000	Ninguna
Analfabetismo × Área	74,6	1	$5,81 \times 10^{-18}$	***	0,052	Débil
Grupo edad × Actividad	150,5	6	$2,07 \times 10^{-33}$	***	0,073	Débil

Note. Source: author's own calculations. ns=not significant; * $p < 0.05$; *** $p < 0.001$. n=27,969.

Five of the six pairs show a statistically significant association. The strongest is Age_Group × Activity ($V=0.073$; $p < 0.001$), and the weakest significant association is Education_Level × Activity ($V=0.016$; $p < 0.05$). Notably, Language × Activity shows no association ($V=0.000$; $p=1.000$), which has direct implications for the design of tourism products.

Figure 1

Heat map — Cramér's V between variables for the Montubio community, Guayas (CPV 2022)



Note. Created by the author using data from the CPV 2022 (INEC). n=27,969. The values indicate the strength of association (0=none, 1=perfect). Processed in Python 3.14.

The heat map visually confirms the results in Table 5. The cell with the highest intensity corresponds to Age_Group × Activity_Status ($V=0.073$), reflected in the darkest shade. The Language × Activity_Status cell appears almost white ($V=0.000$), indicating independence between these variables. The cells of medium intensity—Illiterate × Area and Area × Education_Level—indicate that urban and rural conditions are related to differences in educational attainment. This pattern suggests that the spoken language does not determine the economic integration of the Montubio people, which opens up opportunities to design tourism products that promote the native language without compromising employability. The MCA was performed on 27,969 records with complete data. Table 6 presents the explained inertia by the first five dimensions.

Table 6 Explained inertia by dimension — MCA of the Montubio people, Guayas (CPV 2022)

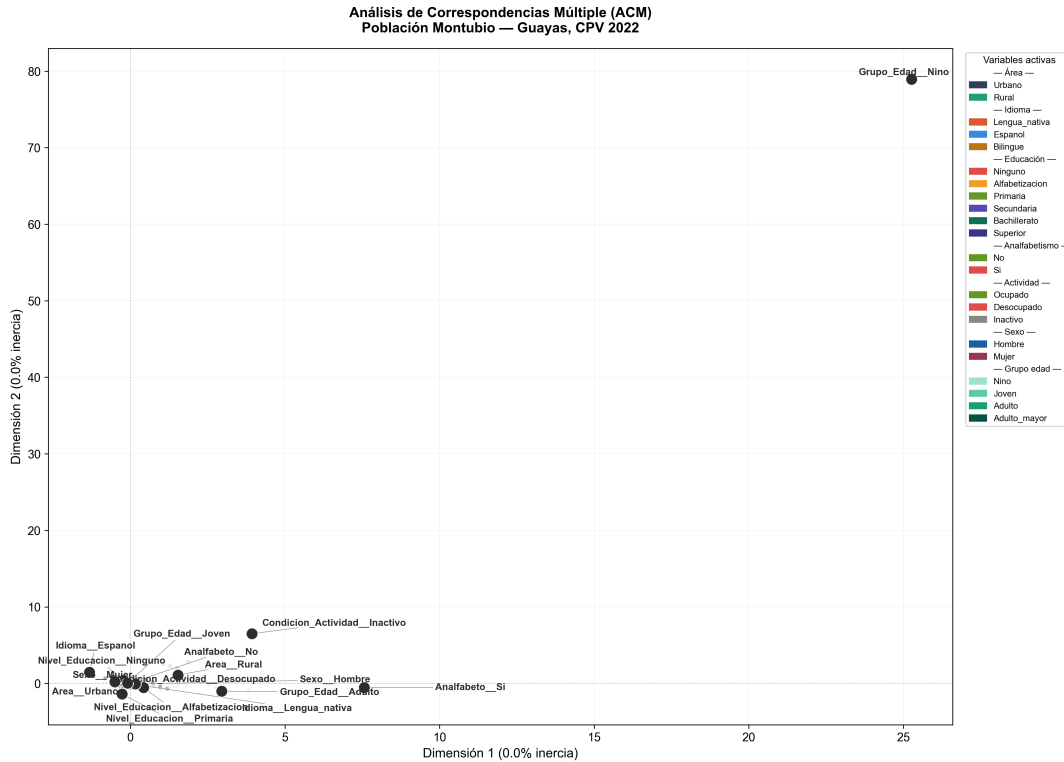
Dimensión	Valor propio	% de inercia	% acumulado
1	0,170	13,22%	13,22%
2	0,157	12,25%	25,47%
3	0,149	11,59%	37,06%
4	0,146	11,36%	48,42%
5	0,139	10,79%	59,21%

Note. Source: Author's own calculations. Python 3.14, prince 0.19. $n=27,969$.

The five dimensions explain 59.21% of the total inertia. The homogeneous distribution (13.22% to 10.79%) indicates that the variability of the Montubio phenomenon is not concentrated along a single axis but is multidimensional. Figures 2 and 3 present the factor plots.

Figure 2

ACM Factor Plot — Dimensions 1 and 2, Guayas Montubio community (CPV 2022)

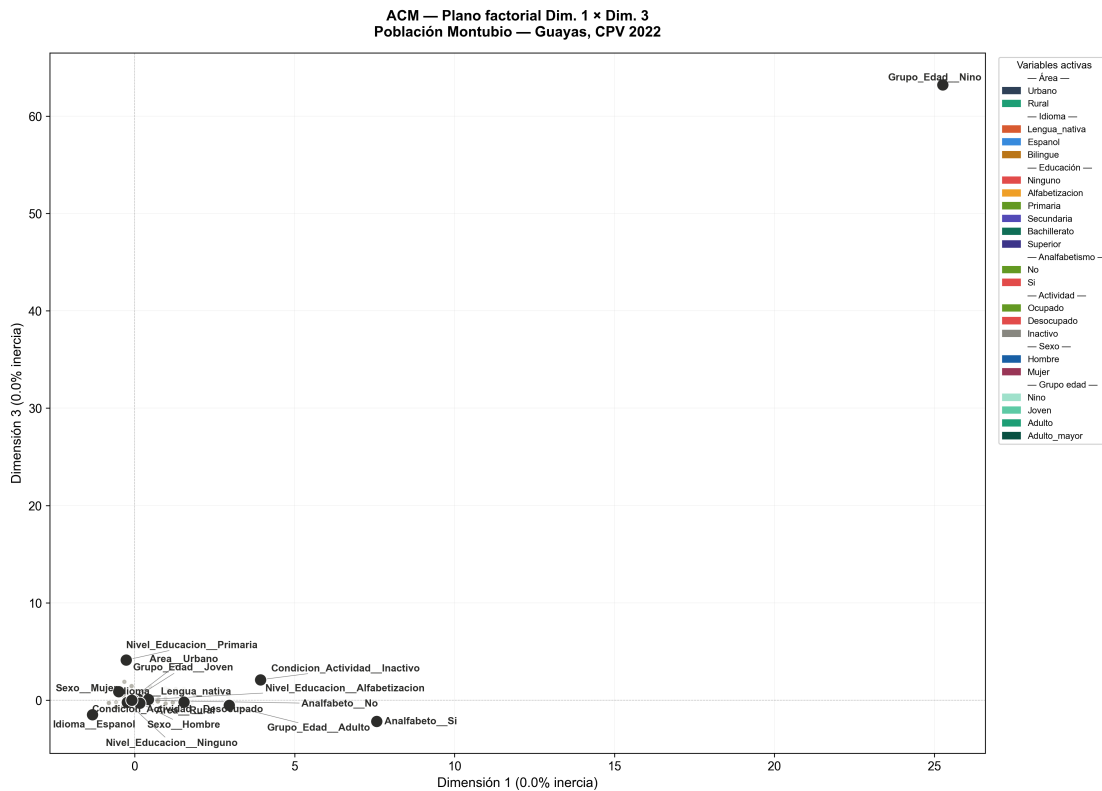


Note. Author's own work. The points represent active categories; proximity indicates association. Inertia Dim1+Dim2=25.47%. n=27,969.

The main factorial plot (25.47% cumulative inertia) reveals the central structure of association. The horizontal axis (Dim. 1, 13.22%) separates cultural profiles: at the positive pole, Native Language, None (educational level), and Rural are grouped, forming a profile of strong ancestral cultural identity; at the negative pole are Spanish, High School Diploma, and Urban, representing greater integration into the formal system. The vertical axis (Dim. 2, 12.25%) differentiates based on employment status and age group: Child and Inactive are associated at one pole, while Adult and Employed are grouped at the other, reflecting the work life cycle. The central position of Woman and Man indicates that gender has no preferential association with either pole of the Montubio cultural profile.

Figure 3

ACM Factor Plot — Dimensions 1 and 3, Montubio community in Guayas (CPV 2022)



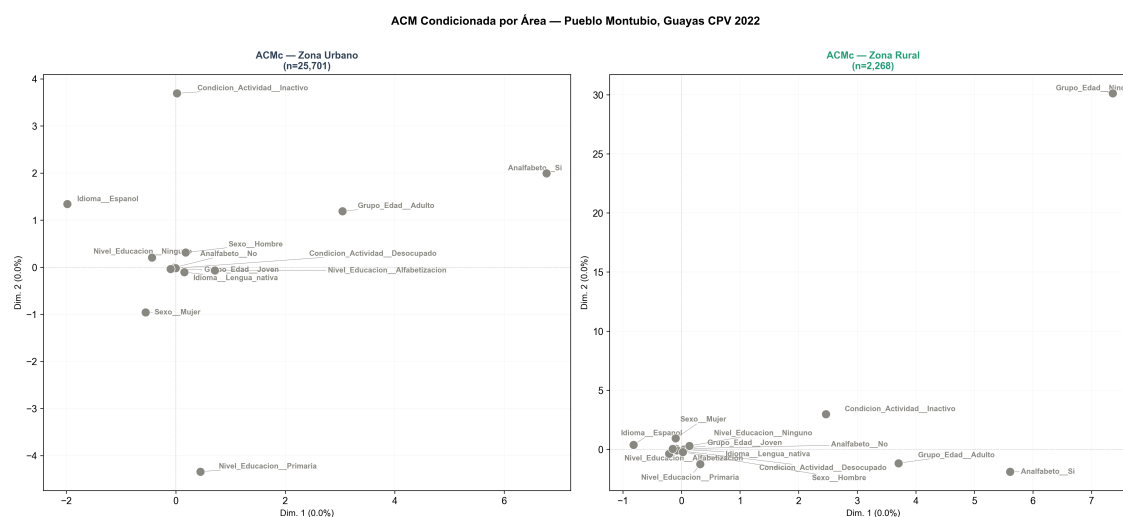
Note. Author's own work. Inertia Dim1+Dim3=24.81%. n=27,969.

The factor plot of Dimensions 1 and 3 (24.81% cumulative inertia) complements the interpretation of the principal ACM. Dimension 3 (11.59%) introduces an additional axis related primarily to intermediate levels of education. The categories "Literacy" and "Elementary" are separated from "None" on this axis, suggesting that Dim. 3 captures the difference between those who have had at least some minimal contact with the formal education system and those who have had none. This finding is relevant to tourism: knowledge bearers with basic literacy have a greater capacity to interact with visitors in structured formats (workshops, written guides), while those without formal education are the most authentic bearers of the oral tradition.

The ACMc analyzes the structure of associations within each subgroup. Figure 4 presents the ACMc by area of residence.

Figure 4

Conditional ACM by Area of Residence — Montubio People, Guayas (CPV 2022)



Note. Author's own work. Left: urban area (n=25,701). Right: rural area (n=2,268). Python 3.14, prince 0.19.

The ACMc by area reveals structural differences between the urban and rural Montubio populations. In the urban area (left panel), the categories show greater dispersion, reflecting the socioeconomic heterogeneity typical of urban environments: Montubios with different educational levels, diverse employment statuses, and a greater presence of Spanish coexist. In rural areas (right panel), the categories show greater concentration and proximity to one another, highlighting the cultural homogeneity of rural communities: native language, no formal education, and formal inactivity form a culturally cohesive profile. For intercultural tourism, rural areas offer more authentic experiences, while urban areas allow for larger-scale and more accessible tourism products.

Table 7 presents the cantons selected for the territorial ACMc, and Figure 5 shows the results.

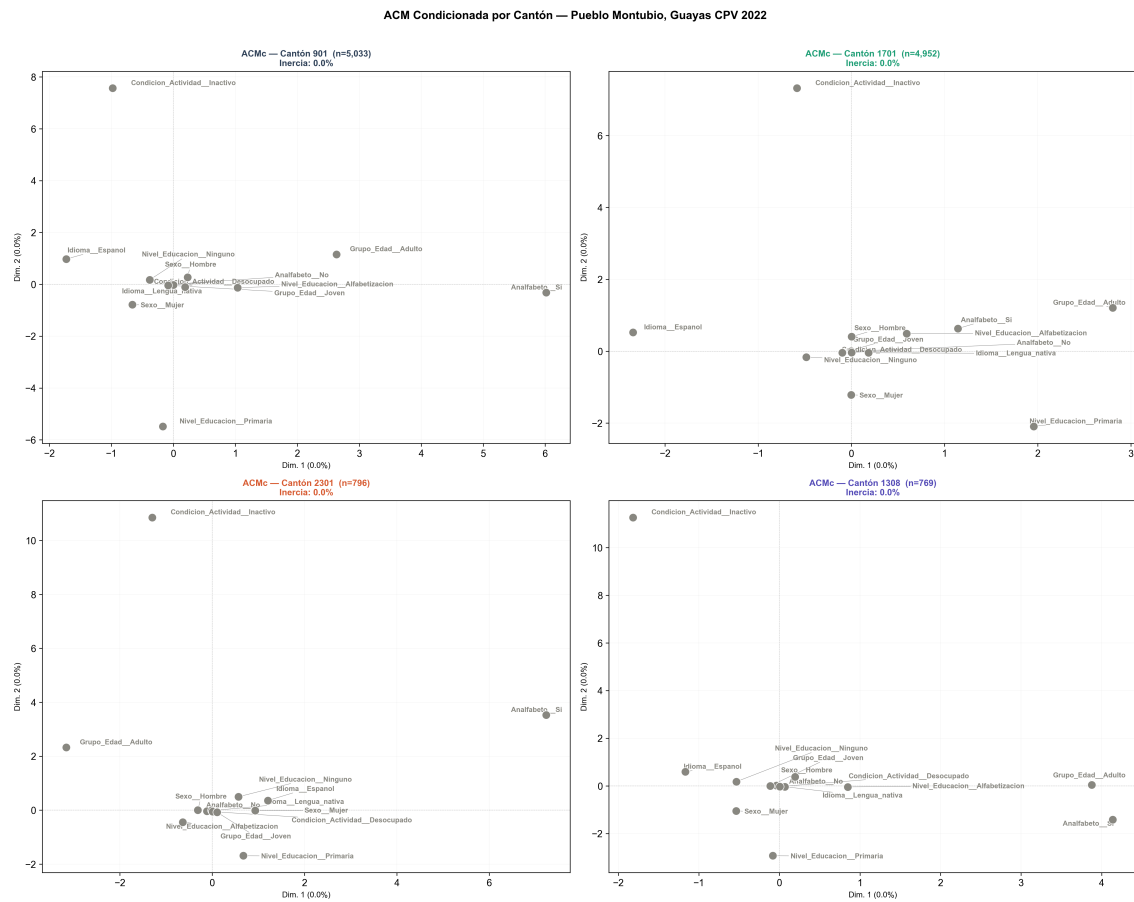
Table 7 Cantons selected for the territorial Conditional ACM (CPV 2022)

Cantón	Código	n (muestra)	Incluido en ACMc
Guayaquil	901	5.033	Sí
Milagro	1701	4.952	Sí
Daule	2301	796	Sí
Balzar	1308	769	Sí

Note. Source: Author's own work. The four cantons represent 49.0% of the Montubio population in Guayas.

Figure 5

Conditional ACM by Canton — Montubio People, Guayas (CPV 2022)



Note. Compiled by the author. 901=Guayaquil (n=5,033), 1701=Milagro (n=4,952), 2301=Daule (n=796), 1308=Balzar (n=769).

The ACMc by canton reveals distinct cultural profiles. Guayaquil and Milagro (upper panels) show greater dispersion of categories, consistent with the socioeconomic diversity of these urban centers: groups with different levels of education, varied employment statuses, and a greater presence of Spanish are observed. In contrast, Daule and Balzar (lower panels) show a concentration of categories such as Native_Language, None, and Inactive at one end of the spectrum, forming a profile of high ancestral cultural coherence. For tourism policy: Daule and Balzar are the cantons with the greatest potential for cultural authenticity tourism routes, while Guayaquil and Milagro, due to their critical mass of population, are better suited for large-scale urban cultural tourism. It is recommended to design differentiated products for each cantonal typology.

The results confirm and enrich the characterization of the Montubio people in Guayas. The absence of an association between language and economic status ($V=0.000$; $p=1.000$) is the most relevant finding for intercultural tourism policies: speaking a native language does not constitute a barrier to economic participation. This opens up

possibilities for designing tourism products that incorporate the linguistic experience as added value without compromising the employability of cultural bearers.

The homogeneous distribution of inertia in the ACM reveals that Montubio identity is a multidimensional phenomenon. The territorial ACMc identifies Daule and Balzar as the cantons with the highest level of cultural preservation and, therefore, the greatest potential for authenticity-based tourism. Methodologically, the Bronze-Silver-Gold architecture processed 16.9 million records in approximately 8 minutes, demonstrating its feasibility for research using conventional computational resources.

Conclusions

The Big Data analysis of the 2022 CPV allows for an empirically precise characterization of the Montubio people of Guayas: 35,593 individuals, 91.8% of whom are native language speakers and 64.3% of whom have no formal education, demonstrating a robust system of oral ancestral cultural transmission.

Chi-square and Cramér's V tests confirm significant associations in five of the six pairs analyzed. Principal component analysis (PCA) reveals a multidimensional structure with 59.21% of the variance explained across five dimensions. Cluster analysis (CA) identifies Daule and Balzar as the cantons with the greatest potential for cultural authenticity tourism, and Guayaquil and Milagro for large-scale urban cultural tourism.

It is recommended that future research integrate logistic regression to model the probability of tourist participation and incorporate geospatial data for territorial mapping in relation to existing tourist corridors.

.....

References

- Benzécri, J.-P. (1973). L'analyse des données: Vol. 2. L'analyse des correspondances. Dunod.https://books.google.com.ec/books/about/L_analyse_des_donn%C3%A9es_Benz%C3%A9cri_J_P_et.html?hl=fr&id=sDTwAAAAMAAJ&redir_esc=y
- Consejo de Desarrollo del Pueblo Montubio de la Costa Ecuatoriana y Zonas Subtropicales de la Región Litoral. (2014). Modelo de desarrollo del pueblo montubio. CODEPMOC.<https://redinvestigadoresidentidadesnacionales.wordpress.com/2014/07/31/consejo-de-desarrollo-del-pueblo-montubio-de-la-costa-ecuatoriana-y-zonas-subtropicales-de-la-region-litoral-codepmoc/>
- Databricks. (2021). What is medallion architecture? Databricks. <https://www.databricks.com/blog/what-is-medallion-architecture>
- Escofier, B., y Pagès, J. (1998). Analyses factorielles simples et multiples: Objectifs, méthodes et interprétation (3.^a ed.). Dunod. <https://media.electre-ng.com/extraits/extrait->

id/982b2cef0173f5009ee1877be8fde8b73a0542e60ead94306ba6880c7a8a76
a4.pdf

- Greenacre, M. (2017). Correspondence analysis in practice (3.^a ed.). Chapman and Hall/CRC. ISBN: 978-1-4987-3177-5.
<https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.1201/9781315369983&type=googlepdf>
- IBM. (2021). What is big data? IBM Cloud Learn Hub. <https://www.ibm.com/topics/big-data>
- Instituto Nacional de Estadística y Censos. (2022). VIII Censo de Población y VII de Vivienda 2022: Microdatos [conjunto de datos]. INEC.
<https://www.censoecuador.gob.ec/resultados-censo/#tabulados>
- Mayer-Schönberger, V., y Cukier, K. (2013). Big data: La revolución de los datos masivos. Turner Publicaciones. ISBN: 978-84-15-83210-2.
<http://catedradatos.com.ar/media/3.-Big-data.-La-revolucion-de-los-datos-masivos-Noema-Spanish-Edition-Viktor-Mayer-Schonberger-Kenneth-Cukier.pdf>
- Ministerio de Turismo del Ecuador. (2021). Plan Nacional de Turismo 2030. MINTUR.
<https://www.turismo.gob.ec/plan-de-turismo-2030-herramienta-que-marca-un-hito-en-la-historia-turistica-del-ecuador/>
- Muratorio, B. (2000). Identidades de mujeres indígenas y política de reproducción cultural en la Amazonía ecuatoriana. En I. Cervone y F. Rivera (Eds.), Ecuador racista (pp. 251-280). FLACSO.
<https://www.flacso.edu.ec/biblio/catalog/resGet.php?resId=43460>
- Organización Mundial del Turismo. (2019). Tourism and culture synergies. OMT/UNWTO. <https://www.unwto.org/tourism-and-culture>
- Richards, G. (2018). Cultural tourism: A review of recent research and trends. Journal of Hospitality and Tourism Management, 36, 12-21.
<https://doi.org/10.1016/j.jhtm.2018.03.005>
- Ruiz, E., y Solís, D. (Coords.). (2007). Turismo comunitario en Ecuador: Desarrollo y sostenibilidad social. Abya-Yala / Universidad de Cuenca.
<https://dialnet.unirioja.es/servlet/libro?codigo=996856>
- UNESCO. (2003). Convención para la Salvaguardia del Patrimonio Cultural Inmaterial. UNESCO. <https://ich.unesco.org/es/convención>