

**Ciencia Latina**  
Internacional

---

Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.  
ISSN 2707-2207 / ISSN 2707-2215 (en línea), noviembre-diciembre 2024,  
Volumen 8, Número 6.

[https://doi.org/10.37811/cl\\_rcm.v8i6](https://doi.org/10.37811/cl_rcm.v8i6)

**IDENTIFICACIÓN Y CARACTERIZACIÓN  
DE PERFILES DE INACTIVIDAD LABORAL  
EN ECUADOR MEDIANTE ANÁLISIS DE  
COMPONENTES PRINCIPALES Y CLÚSTERES  
K-MEANS**

IDENTIFICATION AND CHARACTERIZATION OF  
LABOR INACTIVITY PROFILES IN ECUADOR THROUGH  
PRINCIPAL COMPONENT ANALYSIS AND  
K-MEANS CLUSTERS

**Jorge Adrian Sarcos González**  
Universidad de Guayaquil, Ecuador

DOI: [https://doi.org/10.37811/cl\\_rcm.v8i6.15451](https://doi.org/10.37811/cl_rcm.v8i6.15451)

## Identificación y Caracterización de Perfiles de Inactividad Laboral en Ecuador mediante Análisis de Componentes Principales y Clústeres K-means

**Jorge Adrian Sarcos González<sup>1</sup>**[jorge.sarcosgon@ug.edu.ec](mailto:jorge.sarcosgon@ug.edu.ec)<https://orcid.org/0009-0007-6645-1567>Universidad de Guayaquil  
Ecuador

### RESUMEN

La inactividad laboral representa un desafío significativo para el desarrollo económico y social de Ecuador. Este estudio tiene como objetivo principal identificar y caracterizar los perfiles de inactividad laboral en el país utilizando datos de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) 2023. Se aplicaron técnicas estadísticas avanzadas, como el Análisis de Componentes Principales (ACP) y el Análisis de Clústeres K-means, para reducir la dimensionalidad de los datos y segmentar la población inactiva en grupos homogéneos. Los resultados revelaron tres clústeres predominantes: (1) "Trabajadores activos con baja educación", caracterizados por jóvenes urbanos con niveles educativos bajos; (2) "Inactivos con educación técnica o media", que incluyen una mezcla de personas activas e inactivas con formación técnica o media; y (3) "Diversos inactivos, estudiantes y trabajo no remunerado", compuesto por una amplia gama de individuos con diversas razones para su inactividad laboral. Estos perfiles proporcionan una base sólida para diseñar políticas públicas focalizadas que aborden las necesidades específicas de cada grupo. La combinación de ACP y K-means en el contexto ecuatoriano ofrece una comprensión más profunda de los factores que contribuyen a la inactividad laboral, facilitando la implementación de intervenciones efectivas.

**Palabras Clave:** inactividad laboral, Ecuador, análisis de componentes principales, clústeres k-means

---

<sup>1</sup> Autor principal

Correspondencia: [jorge.sarcosgon@ug.edu.ec](mailto:jorge.sarcosgon@ug.edu.ec)

# Identification and Characterization of Labor Inactivity Profiles in Ecuador through Principal Component Analysis and K-means Clusters

## ABSTRACT

Labor inactivity represents a significant challenge for the economic and social development of Ecuador. This study aims primarily to identify and characterize labor inactivity profiles in the country using data from the National Survey of Employment, Unemployment and Underemployment (ENEMDU) 2023. Advanced statistical techniques, such as Principal Component Analysis (PCA) and K-means Cluster Analysis, were applied to reduce the dimensionality of the data and segment the inactive population into homogeneous groups. The results revealed three predominant clusters: (1) “Active workers with low education,” characterized by urban youth with low educational levels; (2) “Inactive with technical or secondary education,” which include a mix of active and inactive individuals with technical or secondary training; and (3) “Diverse inactive, students and unpaid work,” composed of a wide range of individuals with diverse reasons for their labor inactivity. These profiles provide a solid basis for designing targeted public policies that address the specific needs of each group. The combination of PCA and K-means in the Ecuadorian context offers a deeper understanding of the factors that contribute to labor inactivity, facilitating the implementation of effective interventions.

**Keywords:** labor inactivity, Ecuador, principal component analysis, k-means clusters

*Artículo recibido 15 octubre 2024*

*Aceptado para publicación: 20 noviembre 2024*



## INTRODUCCIÓN

La inactividad laboral es un desafío persistente en el mercado laboral ecuatoriano, afectando significativamente el desarrollo económico y social del país. Comprender las características y factores asociados con la inactividad laboral es esencial para diseñar políticas efectivas que promuevan la inclusión en el mercado laboral.

El objetivo principal de este estudio es identificar y caracterizar perfiles de inactividad laboral en Ecuador utilizando datos de la ENEMDU 2023. Para ello, se emplean técnicas estadísticas avanzadas como el Análisis de Componentes Principales (ACP) y el Análisis de Clústeres K-means. Estas metodologías permiten reducir la complejidad de los datos y segmentar la población inactiva en grupos homogéneos, facilitando la comprensión de los factores que contribuyen a la inactividad laboral.

La combinación de técnicas estadísticas como el Análisis de Componentes Principales (ACP) y el K-means ha demostrado ser una herramienta poderosa para la reducción de dimensionalidad y la segmentación en análisis de datos (Ding, C. & He, X., 2004)

Este enfoque metodológico es innovador en el contexto ecuatoriano, ya que pocos estudios han aplicado técnicas multivariantes para explorar patrones complejos en la población inactiva. Los hallazgos de este estudio aportan una comprensión más profunda de los perfiles de inactividad laboral, lo cual es crucial para el diseño de políticas públicas efectivas y focalizadas.

### Marco Teórico

La inactividad laboral, entendida como la ausencia de participación en el mercado laboral, representa un desafío importante para las economías en desarrollo. Según la teoría del capital humano, factores como la educación, la experiencia y las características individuales influyen en la capacidad de las personas para acceder al empleo y, por ende, en su decisión de participar en el mercado laboral. (Becker, 1964) argumenta que la inversión en capital humano a través de la educación y la capacitación es esencial para mejorar la productividad y la empleabilidad. En este contexto, los niveles educativos bajos limitan las oportunidades laborales, mientras que la educación superior puede contribuir a una inserción más exitosa en el mercado laboral. Sin embargo, las teorías de segmentación del mercado laboral proponen que, además de las características individuales, existen factores estructurales, como la desigualdad de género, las disparidades regionales y las barreras culturales, que limitan el acceso a

empleos de calidad, promoviendo así la inactividad o el empleo informal (Goodrich, L. G. W., Domínguez Mújica, J., & Parreño Castellano, J. M., 2017)

Por otro lado, las técnicas multivariantes como el Análisis de Componentes Principales (ACP) y el Análisis de Clústeres han demostrado ser herramientas útiles para analizar fenómenos complejos como la inactividad laboral.

El ACP se utiliza ampliamente en análisis multidimensionales para reducir la complejidad de los datos. Estudios recientes en Ecuador han validado la efectividad del ACP y K-means para analizar grupos poblacionales (Herrera, Arroyo, Á., Jiménez, A., & Herrero, Á., 2024)

Estas metodologías permiten reducir la dimensionalidad de los datos y segmentar poblaciones en grupos homogéneos, facilitando la identificación de patrones subyacentes y proporcionando una base sólida para el diseño de políticas públicas. Según Moreano y Castro (2022), estas técnicas son particularmente útiles para captar asociaciones entre variables sociodemográficas y resultados económicos, permitiendo una comprensión más profunda de los factores que influyen en la participación laboral.

### **Revisión De La Literatura**

Estudios previos han utilizado datos de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) para explorar las características de la inactividad laboral en Ecuador, destacando la importancia de factores como la educación, el estado civil y la ubicación geográfica. Goodrich et al. (2017) realizaron un análisis multidimensional de estudiantes en Ecuador, identificando patrones de asistencia y deserción escolar mediante técnicas de clústeres jerárquicos. Sus hallazgos subrayan la relevancia de las características sociodemográficas en la configuración de perfiles poblacionales. De manera similar, (Marchán, A. B., Delgado, R., & Stefos, E., 2017) analizaron a artesanos y trabajadores no calificados, utilizando ACP para identificar criterios de diferenciación clave, como el nivel educativo y las condiciones laborales. Estos estudios resaltan cómo las técnicas multivariantes pueden ser utilizadas para caracterizar grupos poblacionales y abordar problemas sociales complejos.

Otro aspecto crucial es el impacto de la educación en la inactividad laboral. (Aguirre, P. A. U., Martínez Valarezo, J. D., & Carrión Ayala, V. E., 2021) examinaron las disparidades salariales entre hombres y mujeres en Ecuador, destacando cómo el nivel educativo y el estado civil influyen en la participación laboral.

Este estudio también identificó una segregación ocupacional significativa, que afecta especialmente a las mujeres con niveles educativos bajos. Por su parte, (Armijos-Bravo, G. & Camino-Mogro, S. , 2023) analizaron el impacto de la pandemia de COVID-19 en la empleabilidad femenina, encontrando que las políticas de confinamiento exacerbaban las desigualdades laborales, particularmente entre las mujeres jóvenes y aquellas con menor nivel educativo. Estos hallazgos subrayan la importancia de considerar factores de género y educación en el análisis de la inactividad laboral.

Además de la educación, las disparidades regionales y las condiciones geográficas juegan un papel fundamental en la segmentación del mercado laboral. (López-Cevallos & D. F., & Chi, C. , 2010) utilizaron ACP para evaluar las desigualdades en el acceso a servicios de salud en Ecuador, destacando cómo las disparidades económicas y étnicas limitan el acceso a oportunidades laborales en comunidades rurales e indígenas. Este análisis es relevante para comprender cómo la ubicación geográfica puede influir en la inactividad laboral y la desigualdad económica. Por otro lado, (Barbieri & Pan, W., 2013) analizaron factores que motivan a los colonos del Amazonas ecuatoriano a participar en empleos fuera del sector agrícola, identificando cómo las condiciones comunitarias y las características individuales interactúan para determinar la participación laboral.

En términos metodológicos, (Moreano, L. F. & Castro, L. U., 2022)revisaron el uso de métodos multivariantes en estudios educativos, destacando cómo el ACP y el análisis de clústeres permiten identificar patrones complejos en datos sociodemográficos. Su revisión concluyó que estas técnicas son esenciales para analizar fenómenos multidimensionales y proporcionar una base sólida para el diseño de políticas basadas en evidencia. Asimismo, Marchán et al. (2017) demostraron que el uso combinado de técnicas multivariantes y descriptivas puede revelar diferencias significativas entre grupos laborales, proporcionando información valiosa para intervenciones específicas.

La relación entre género, migración y participación laboral también ha sido objeto de estudio. (Brown, L., Pavri, F., & Lawson, V., 1998) examinaron cómo las políticas de ajuste estructural afectaron la organización del trabajo en Ecuador, encontrando que las mujeres migrantes enfrentaron mayores barreras para acceder a empleos formales durante la década de 1980. Este análisis histórico proporciona un contexto útil para entender las dinámicas actuales del mercado laboral ecuatoriano.

Por otro lado, Aguirre et al. (2021) destacaron la necesidad de fortalecer la participación de las mujeres en el mercado laboral, proponiendo políticas que promuevan la igualdad de género y reduzcan las barreras estructurales.

Finalmente, López-Cevallos y Chi (2010) sugieren que las desigualdades socioeconómicas en Ecuador limitan significativamente el acceso a servicios esenciales y oportunidades laborales, lo que perpetúa ciclos de pobreza e inactividad laboral. Este hallazgo es consistente con estudios que analizan la influencia de la educación, el género y la ubicación geográfica en la configuración de la inactividad laboral en el país.

## **METODOLOGÍA**

### **Descripción de los datos**

Este estudio utiliza datos de la Encuesta Nacional de Empleo, Desempleo y Subempleo Urbano y Rural (ENEMDU) correspondiente al año 2023. La base de datos original contiene 345,174 observaciones y 141 variables relacionadas con características sociodemográficas y económicas de la población ecuatoriana.

Para el análisis, se seleccionaron variables clave que se consideran relevantes para entender la inactividad laboral, incluyendo características personales (edad, estado civil), educativas (nivel educativo alcanzado), geográficas (zona urbana o rural) y socioeconómicas (condición de pobreza extrema).

### **Justificación de los métodos utilizados**

#### **Análisis de Componentes Principales (ACP)**

El Análisis de Componentes Principales (ACP) es una técnica estadística multivariante que permite reducir la dimensionalidad de un conjunto de datos con múltiples variables correlacionadas, transformándolas en un nuevo conjunto de variables no correlacionadas llamadas componentes principales. Cada componente principal es una combinación lineal de las variables originales y captura la máxima varianza posible.

La matriz de covarianza  $\Sigma$  se calcula como:

$$\Sigma = \frac{1}{n-1} X^T X$$

donde  $X$  es la matriz de datos original. El ACP implica encontrar los valores propios ( $\lambda$ ) y vectores propios ( $v$ ) de la matriz de covarianza  $\Sigma$ , resolviendo la ecuación:

$$\Sigma v = \lambda v$$

Los componentes principales son las proyecciones de los datos originales sobre los vectores propios, calculados como:

$$PC_k = Xvk \quad \text{para } k = 1, 2, \dots, m$$

donde  $m$  es el número de componentes seleccionados. La proporción de la varianza explicada por el  $k$ -ésimo componente principal está dada por:

$$\text{Proporción de Varianza} = \frac{\lambda_k}{\sum_{i=1}^p \lambda_i}$$

### **Justificación del uso del ACP**

**Reducción de complejidad:** Dada la gran cantidad de variables categóricas y continuas en el conjunto de datos, el ACP ayuda a simplificar el análisis al condensar la información en unos pocos componentes clave.

**Identificación de patrones ocultos:** El ACP permite descubrir relaciones subyacentes entre las variables que no son evidentes a simple vista.

**Facilitación de la segmentación:** Al transformar las variables originales en componentes principales, se mejora la eficacia del análisis de clústeres posterior, ya que se trabaja con variables no correlacionadas que capturan la esencia de los datos.

### **Análisis de Clústeres K-means**

El Análisis de Clústeres K-means es un método de segmentación que agrupa observaciones en  $k$  clústeres, de manera que cada observación pertenece al clúster con la media más cercana, minimizando la variación dentro de los clústeres y maximizando la variación entre ellos.

El objetivo es minimizar la suma de las distancias al cuadrado entre cada punto y el centroide de su clúster correspondiente:

$$\min_{C, \mu} \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

donde  $C = \{C_1, C_2, \dots, C_k\}$  es la partición de los datos en  $k$  clústeres y  $\mu_i$  es el centroide del clúster  $C_i$ .

El algoritmo K-means sigue un procedimiento iterativo que consta de los siguientes pasos:

1. **Inicialización:** Seleccionar k puntos iniciales como centroides.
2. **Asignación de Clústeres:** Asignar cada punto al clúster cuyo centroide está más cercano.
3. **Actualización de Centroides:** Recalcular los centroides como la media de los puntos asignados a cada clúster.
4. **Convergencia:** Repetir los pasos de asignación y actualización hasta que los centroides ya no cambien significativamente o se alcance un número máximo de iteraciones.

### **Justificación del uso de K-means**

Segmentación efectiva: K-means es especialmente útil para grandes conjuntos de datos y permite identificar grupos homogéneos basados en características múltiples.

Aplicabilidad práctica: La segmentación de la población inactiva en perfiles específicos facilita el diseño de políticas públicas focalizadas y efectivas.

Complementariedad con el ACP: Los componentes principales obtenidos del ACP proporcionan un espacio de variables adecuadas para aplicar K-means, mejorando la calidad de la segmentación.

El uso combinado de ACP y K-means ha mostrado resultados óptimos en contextos de segmentación compleja, como lo evidencian estudios en Kalimantan, Indonesia, sobre indicadores laborales (Rais, Goejantoro, R., & Prangga, S., 2021)

La integración del ACP para la reducción de dimensionalidad y K-means para la agrupación ha sido utilizada para optimizar procesos en análisis de datos complejos, incluso bajo escenarios de alta multicolinealidad (Afifa, R., Mazdadi, M. I., Saragih, T. H., Indriani, F., & Muliadi, M., 2024)

### **Preparación de los datos**

#### **Selección y transformación de variables**

Se realizó una selección de variables relevantes para el estudio, enfocándose en aquellas que influyen en la inactividad laboral. Entre las variables continuas se incluyó la Edad estandarizada. Las variables categóricas fueron transformadas en dummies e incluyeron las diferentes categorías de Estado Civil, tales como Soltero, Divorciado, Viudo, Separado, Casado, Unión Libre y Sin especificar, así como los distintos niveles educativos alcanzados, que abarcan desde Sin Educación Formal hasta Postgrado.

Además, se consideraron variables binarias que indicaban si la persona reside en área rural o urbana, y si se encontraba en pobreza extrema. La variable Condición de actividad laboral se mantuvo para analizar la relación entre los clústeres identificados y la condición laboral de los individuos.

### **Estandarización de variables**

La variable continua Edad fue estandarizada para eliminar efectos de escala y facilitar la comparación entre variables, generando así la variable Edad Estandarizada.

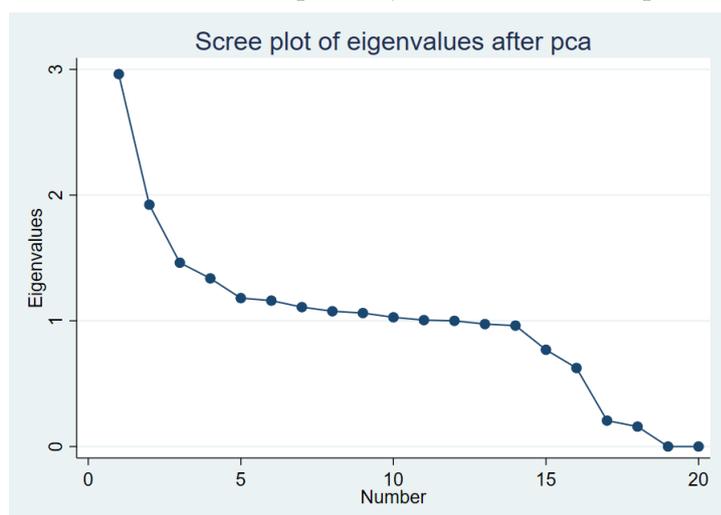
### **Análisis de Componentes Principales (ACP)**

#### **Ejecución del ACP**

Se aplicó el Análisis de Componentes Principales (ACP) a las variables Edad estandarizada, Estado Civil, Nivel Educativo y Área geográfica. La variable Pobreza extrema fue excluida del ACP debido a que presentaba varianza cero, es decir, no aportaba información discriminante al análisis. Se decidió extraer los tres primeros componentes principales basándose en el criterio de eigenvalores mayores a uno y en la proporción acumulada de la varianza explicada.

Se decidió extraer los tres primeros componentes principales basándose en el criterio de eigenvalores donde la tendencia se estabiliza y en la proporción acumulada de la varianza explicada.

**Figura 1.** Representación gráfica de los valores propios asociados a cada componente principal, indicando la varianza explicada y la selección de componentes clave basada en el criterio de codo.



Estos componentes fueron fundamentales para captar las dimensiones clave que diferencian a los individuos inactivos.

## RESULTADOS

### Descripción detallada de los componentes principales

Los tres componentes principales obtenidos explican en conjunto el 31.74% de la varianza total, lo cual es adecuado considerando la naturaleza multidimensional de los datos.

### Interpretación integrada de los componentes

El primer componente principal, denominado "Edad y Nivel Educativo", explica el 14.81% de la varianza total. Este componente refleja un gradiente de edad y nivel educativo. Las cargas factoriales positivas significativas incluyen la edad estandarizada (0.5370), Educación Superior Incompleta (0.2913) y Educación Superior Completa (0.2038). Por otro lado, las cargas factoriales negativas significativas están asociadas con Educación Primaria Incompleta (-0.3696) y Sin Educación Formal (-0.2483). Este patrón indica que las puntuaciones altas en este componente corresponden a individuos de mayor edad con niveles educativos superiores, mientras que las puntuaciones bajas se asocian con personas más jóvenes con bajos niveles educativos. Las cargas positivas en niveles educativos superiores y edad sugieren que, a medida que aumenta la edad, es más probable que las personas hayan alcanzado niveles educativos más altos. Las cargas negativas en niveles educativos bajos indican que los individuos jóvenes tienden a tener menor educación, lo cual puede influir en su inactividad laboral.

El segundo componente principal, denominado "Estado Civil y Educación Media", explica el 9.62% de la varianza total. Este componente distingue entre diferentes estados civiles y niveles educativos intermedios. Las cargas factoriales positivas significativas incluyen ser Soltero (0.4496) y Sin Educación Formal (0.3589), mientras que las cargas factoriales negativas significativas están asociadas con Unión Libre (-0.5811) y Educación Secundaria Completa (-0.4274). Este patrón sugiere que las puntuaciones altas en este componente están asociadas con individuos solteros y con muy bajos niveles educativos, mientras que las puntuaciones bajas corresponden a personas en unión libre con educación secundaria completa. Este componente indica que el estado civil y la educación media están relacionados con la inactividad laboral de maneras específicas, posiblemente reflejando responsabilidades familiares o diferencias culturales en la participación laboral.

El tercer componente principal, denominado "Área Geográfica y Educación Técnica", explica el 7.31% de la varianza total. Este componente captura diferencias geográficas y educativas, reflejando

disparidades entre áreas rurales y urbanas en términos de educación y acceso al empleo. Las cargas factoriales positivas significativas incluyen residir en área rural (0.4537), Educación Superior Incompleta (0.4614) y Educación Técnica o Tecnológica (0.3211). Las cargas factoriales negativas significativas están asociadas con Educación Superior Completa (-0.4171) y Postgrado (-0.3010). Este patrón indica que las puntuaciones altas en este componente corresponden a individuos en áreas rurales con educación técnica o superior incompleta, mientras que las puntuaciones bajas corresponden a personas en áreas urbanas con educación superior completa o postgrado.

**Tabla 1.** Variables utilizadas en el análisis.

Variable	Descripción	Tipo
Edad estandarizada	Edad estandarizada	Continua
Estado Civil 1-7	Dummies de estado civil	Categóricas (variables dummy)
Nivel Educativo 1-11	Dummies de nivel educativo alcanzado	Categóricas (variables dummy)
Área geográfica	Área geográfica (1 = Rural, 0 = Urbana)	Binaria
Pobreza extrema	Pobreza extrema (1 = Sí, 0 = No)	Binaria
Condición de actividad laboral	Condición de actividad laboral	Categórica

**Tabla 2.** Distribución de frecuencia de las variables categóricas

**Tabla 2A.** Distribución del estado civil

Estado Civil	Frecuencia	Porcentaje
Sin especificar	56,140	16.26%
Soltero	85,554	24.79%
Divorciado	16,670	4.83%
Viudo	10,031	2.91%
Separado	14,734	4.27%
Casado	45,595	13.21%
Unión Libre	116,450	33.74%
<b>Total</b>	<b>345,174</b>	<b>100.00%</b>

Nota: Los códigos de estado civil no especificados corresponden a categorías no identificadas en los datos.

**Tabla 2B.** Distribución del nivel educativo.

Nivel Educativo	Frecuencia	Porcentaje
Sin especificar	20,232	5.86%
1 (Sin Educación Formal)	9,512	2.76%
2	396	0.11%
3	1	0.00%
4 (Educación Secundaria Incompleta)	67,128	19.45%
5 (Educación Técnica o Tecnológica)	66,965	19.40%
6 (Educación Superior Incompleta)	71,576	20.74%
7 (Educación Primaria Incompleta)	40,782	11.81%
8 (Educación Superior Completa)	8,394	2.43%
9 (Educación Secundaria Completa)	53,267	15.43%
10 (Postgrado)	6,921	2.01%
<b>Total</b>	<b>345,174</b>	<b>100.00%</b>

Nota: Los códigos de nivel educativo no especificados corresponden a categorías no identificadas en los datos.

**Tabla 3.** Resultados del Análisis de Componentes Principales (ACP)

Cargas factoriales de los tres primeros componentes principales

Variable	Comp1	Comp2	Comp3	Comunalidad
Edad estandarizada	0.537	0.0435	0.0642	0.864
Soltero	-0.4102	0.4496	-0.0262	0.8882
Divorciado	0.3237	0.1695	-0.1532	0.3999
Unión Libre	-0.1877	-0.5811	0.1315	0.7792
Sin Educación Formal	-0.2483	0.3589	-0.0846	0.4408
Educación Superior Incompleta	0.2913	0.1794	0.4614	0.6246
Educación Primaria Incompleta	-0.3696	0.1037	0.1018	0.4405
Educación Superior Completa	0.2038	0.0483	-0.4171	0.3818
Educación Secundaria Completa	-0.1199	-0.4274	0.1987	0.4517
Postgrado	0.0611	-0.2142	-0.301	0.2318
Área geográfica	-0.0054	0.0975	0.4537	0.3193

Nota: Solo se incluyen las variables con cargas significativas en al menos uno de los componentes.

**Tabla 4.** Estadísticas descriptivas de los componentes principales

Componente	N	Media	Desviación estándar	Mínimo	Máximo
PC1	345,174	0	1.721	-3.0076	3.0496
PC2	345,174	0	1.386	-2.577	2.9228
PC3	345,174	0	1.209	-2.5261	3.8434

Nota: Las medias son aproximadamente cero debido a la estandarización de las variables.

## Resultados del análisis de clústeres

### Análisis de Clústeres (K-means)

Se aplicó el Análisis de Clústeres K-means utilizando las puntuaciones de los tres componentes principales (Edad y Nivel Educativo, Estado Civil y Educación Media, y Área Geográfica y Educación Técnica). Se decidió segmentar la población en tres clústeres, basándose en la interpretación de los componentes principales y en métricas de validación como el Método del Codo y el Índice de Silhouette, que indicó un valor promedio de 0.52, lo que sugiere una buena calidad de la segmentación.

**Tabla 5.** Resultados de la regresión de condición de actividad laboral sobre los componentes principales.

Variable	Coefficiente	Error estándar	Valor t	Valor p	Intervalo de confianza (95%)
Edad y Nivel Educativo (PC1)	0.8279	0.003	276.98	0	[0.8221, 0.8338]
Estado Civil y Educación Media (PC2)	-0.8406	0.0037	-226.61	0	[-0.8479, -0.8334]
Área Geográfica y Educación Técnica (PC3)	0.4905	0.0043	115.28	0	[0.4822, 0.4989]
Constante	3.8435	0.0051	747.08	0	[3.8334, 3.8536]

**Tabla 6.** Resultados del análisis de clústeres (K-means)

Clúster	Frecuencia	Porcentaje
1	56,140	16.26%
2	118,746	34.40%
3	170,288	49.33%
Total	345,174	100.00%

## Relación con la inactividad laboral

La variable **condición de actividad laboral** mostró patrones diferenciados entre los clústeres, destacando que los motivos de inactividad varían significativamente según la edad, educación y ubicación geográfica.

**Tabla 7.** Estadísticas descriptivas de los componentes principales por clúster

Tabla 7A. Clúster 1

Componente	N	Media	Desviación estándar	Mínimo	Máximo
PC1	56,140	-2.7571	0.2184	-3.0076	-1.2045
PC2	56,140	1.962	0.6118	1.4394	2.9228
PC3	56,140	-0.087	0.5717	-0.8124	1.9317

Tabla 7B. Clúster 2

Componente	N	Media	Desviación estándar	Mínimo	Máximo
PC1	118,746	-0.8566	0.7291	-1.8848	1.161
PC2	118,746	-1.5545	0.7184	-2.577	0.6461
PC3	118,746	0.241	0.8617	-1.2683	2.9201

Tabla 7C. Clúster 3

Componente	N	Media	Desviación estándar	Mínimo	Máximo
PC1	170,288	1.5062	0.5761	-0.6429	3.0496
PC2	170,288	0.4371	0.5197	-1.6916	1.3543
PC3	170,288	-0.1394	1.5085	-2.5261	3.8434

**Tabla 8.** Distribución de los clústeres según área geográfica

Clúster	Urbana (0)	Rural (1)	Total
1	39,809	16,331	56,140
2	87,668	31,078	118,746
3	124,986	45,302	170,288
Total	252,463	92,711	345,174

**Tabla 9.** Medias de variables seleccionadas por clúster

Variable	Clúster 1	Clúster 2	Clúster 3
PC1 (Edad y Nivel Educativo)	-2.7571	-0.8566	1.5062
PC2 (Estado Civil y Educación Media)	1.962	-1.5545	0.4371
PC3 (Área Geográfica y Educación Técnica)	-0.087	0.241	-0.1394
Área geográfica (1 = Rural)	0.2909	0.2617	0.266
Sin Educación Formal	0.3604	0	0
Educación Superior Incompleta	0	0.0353	0.3696
Educación Primaria Incompleta	0.6252	0.2648	0.0025
Educación Superior Completa	0	0.0944	0.3545
Educación Secundaria Completa	0	0.3426	0.0006
Postgrado	0	0.2069	0.1685

Nota: Los valores representan proporciones dentro de cada clúster para las variables categóricas.

**Tabla 10.** Distribución de la condición de actividad laboral (conduct) por clúster

Clúster	conduct = 0	conduct = 1	conduct = 2	conduct = 3	conduct = 4	conduct = 5	conduct = 6	conduct = 7	conduct = 8	conduct = 9	Total
1	56,140	0	0	0	0	0	0	0	0	0	56,140
2	19,119	20,768	11,706	1,156	12,205	6,771	127	4,747	461	41,686	118,746
3	1	49,859	17,755	1,922	34,657	8,616	388	2,515	443	54,132	170,288
Total	75,260	70,627	29,461	3,078	46,862	15,387	515	7,262	904	95,818	345,174

Nota: Los códigos de condición de actividad laboral no especificados corresponden a categorías de actividad laboral no identificadas en los datos.

## Caracterización detallada de los clústeres

### Clúster 1: "Trabajadores activos con baja educación"

Este clúster comprende 56,140 individuos (16.26%) de la muestra total y se caracteriza por ser exclusivamente trabajadores activos con bajos niveles educativos. Predominantemente jóvenes y residentes en áreas urbanas, este grupo muestra una alta proporción de personas solteras. A pesar de su baja educación, estos individuos están integrados en el mercado laboral, posiblemente ocupando empleos informales o de baja cualificación. Las limitaciones educativas pueden restringir sus oportunidades de crecimiento económico, aunque su condición de solteros facilita su participación en

el mercado laboral sin las responsabilidades familiares que podrían limitar su disponibilidad para trabajar.

### **Clúster 2: "Inactivos con educación técnica o media"**

Este clúster incluye 118,746 individuos (34.40%) y presenta una mezcla de personas activas e inactivas, destacando a estudiantes y quienes realizan trabajos no remunerados. Con niveles educativos técnicos o medios y un equilibrio entre residentes en áreas rurales y urbanas, este grupo refleja una diversidad en la condición de actividad laboral. Algunos individuos pueden estar en transición hacia el mercado laboral o enfrentando obstáculos específicos que les impiden una plena participación, como desajustes entre su formación y las demandas del mercado laboral local.

### **Clúster 3: "Diversos inactivos, estudiantes y trabajo no remunerado"**

Compuesto por 170,288 individuos (49.33%), este clúster es el más grande y heterogéneo. Incluye a estudiantes, personas en trabajos no remunerados y quienes enfrentan barreras estructurales significativas. Con una alta diversidad en niveles educativos, incluyendo educación superior, y una presencia tanto en áreas urbanas como rurales, este grupo enfrenta variadas razones de inactividad laboral. Las barreras estructurales, como la falta de oportunidades o la discriminación, son más pronunciadas en este clúster, afectando su capacidad para participar activamente en el mercado laboral.

## **CONCLUSIÓN**

El presente estudio ha logrado identificar y caracterizar distintos perfiles de inactividad laboral en Ecuador mediante la aplicación conjunta del Análisis de Componentes Principales (ACP) y el Análisis de Clústeres K-means. Los tres clústeres identificados—"Trabajadores activos con baja educación", "Inactivos con educación técnica o media" y "Diversos inactivos, estudiantes y trabajo no remunerado"—reflejan la diversidad de factores que influyen en la inactividad laboral en el país, tales como el nivel educativo, el estado civil y la ubicación geográfica. Estos hallazgos subrayan la necesidad de diseñar políticas públicas específicas y focalizadas que aborden las particularidades de cada grupo. Por ejemplo, para los trabajadores activos con baja educación, se recomiendan programas de educación continua y mejora de condiciones laborales; para los inactivos con educación técnica o media, es crucial implementar orientaciones profesionales y promover la alineación entre la oferta educativa y las demandas del mercado laboral; y para el grupo más heterogéneo de diversos inactivos, estudiantes y

trabajadores no remunerados, se sugieren programas de inclusión laboral y reconocimiento del trabajo no remunerado.

Los resultados de la segmentación en este estudio son consistentes con análisis previos en América del Sur, donde el ACP y los clústeres jerárquicos identificaron características claves entre diferentes grupos de países (Martín-Barreiro, 2021)

Además, la metodología empleada en este estudio, que combina ACP y K-means, demuestra ser una herramienta eficaz para analizar fenómenos complejos y multidimensionales como la inactividad laboral. Esta aproximación permite una comprensión más detallada y estructurada de los factores subyacentes, facilitando así la implementación de intervenciones más efectivas.

La efectividad del ACP y K-means en el análisis de datos multidimensionales es ampliamente respaldada por la literatura, destacando su utilidad en la optimización de políticas públicas y toma de decisiones basada en datos. (Tamonob, A. M., Saefuddin, A., & Wigena, A., 2020)

No obstante, futuras investigaciones podrían expandir este análisis incorporando variables adicionales y considerando cambios temporales en el mercado laboral. En conclusión, este estudio aporta una perspectiva novedosa y detallada sobre la inactividad laboral en Ecuador, ofreciendo una base sólida para el desarrollo de estrategias que promuevan una mayor inclusión y participación en el mercado laboral.

## REFERENCIAS BIBLIOGRAFICAS

- Afifa, R., Mazdadi, M. I., Saragih, T. H., Indriani, F., & Muliadi, M. (2024). Implementasi Principal Component Analysis (PCA) dan Gap Statistic untuk Clustering Kanker Payudara pada Algoritma K-Means. *SISTEMASI*. doi: <https://doi.org/10.32520/stmsi.v13i5.4015>
- Aguirre, P. A. U., Martínez Valarezo, J. D., & Carrión Ayala, V. E. (2021). Diferencias salariales y segregación ocupacional en el mercado laboral del Ecuador. *INNOVA Research Journal*, 6(1), 129-144. doi: <https://doi.org/10.33890/innova.v6.n1.2021.1454>
- Armijos-Bravo, G., & Camino-Mogro, S. . (2023). Covid-19 lockdown in Ecuador: Are there gender differences in unemployment? *The Journal of Development Studies*, 833-853. doi:10.1080/00220388.2023.2192573

- Barbieri, A., & Pan, W. (2013). People, land, and context: Multilevel determinants of off-farm employment in the Ecuadorian Amazon. *Population, Space, and Place*, 19(5), 558-579. doi: <http://dx.doi.org/10.1002/psp.1733>
- Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press. doi:<https://ssrn.com/abstract=1496221>
- Brown, L., Pavri, F., & Lawson, V. (1998). Gender, migration and the organisation of work under economic devolution: Ecuador, 1982-90. *International Journal of Population Geography*, 4(3), 259-274. doi: [https://doi.org/10.1002/\(sici\)1099-1220\(199809\)4:3%3C259::aid-ijpg107%3E3.0.co;2-p](https://doi.org/10.1002/(sici)1099-1220(199809)4:3%3C259::aid-ijpg107%3E3.0.co;2-p)
- Ding, C., & He, X. (2004). K-means clustering via principal component analysis. *Proceedings of the twenty-first international conference on Machine learning*. doi: <https://doi.org/10.1145/1015330.1015408>
- Goodrich, L. G. W., Domínguez Mújica, J., & Parreño Castellano, J. M. (2017). An analysis of the social profile of 15 to 17-year-old students in Ecuador regarding secondary school attendance and truancy. *Review of European Studies*, 9(2), 91. doi: <https://doi.org/10.5539/RES.V9N2P91>
- Herrera, A., Arroyo, Á., Jiménez, A., & Herrero, Á. (2024). Exploratory techniques to analyse Ecuador's tourism industry. *Logic Journal of the IGPL*. doi:<https://doi.org/10.1093/jigpal%2Fjzae040>
- López-Cevallos, & D. F., & Chi, C. . (2010). Health care utilization in Ecuador: A multilevel analysis of socio-economic determinants and inequality issues. . *Health Policy and Planning*, 25(3), 209-218. doi: <https://doi.org/10.1093/heapol/czp052>
- Marchán, A. B., Delgado, R., & Stefos, E. (2017). A multidimensional statistical analysis of artisans and unskilled workers of Ecuador. *Review of European Studies*, 9(2), 237. doi: <http://dx.doi.org/10.5539/res.v9n2p237>
- Martín-Barreiro, C. R.-F.-S.-V. (2021). Disjoint and functional principal component analysis for infected cases and deaths due to COVID-19. *Sensors*. doi: <https://doi.org/10.3390/s21124094>
- Moreano, L. F., & Castro, L. U. (2022). Review of multivariate statistical methods applied to the results of the Ser Bachiller exam in Ecuador. *Minerva*. Doi : <https://doi.org/10.47460/minerva.v1iSpecial.86>



- Rais, M., Goejantoro, R., & Prangga, S. (2021). Optimalisasi K-Means Cluster dengan Principal Component Analysis. *EKSPONENSIAL*. doi: <https://doi.org/10.30872/eksponensial.v12i2.805>
- Tamonob, A. M., Saefuddin, A., & Wigena, A. (2020). Principal component analysis with successive interval in K-Means Cluster Analysis. *IOP Conference Series: Materials Science and Engineering*. doi: <https://doi.org/10.1088/1757-899X/823/1/012055>

