

Segmentación del empleo en Ecuador con variables sociodemográficas: evidencia desde k-prototypes

Juan Carlos Pozo Fonseca*

La presente investigación tiene como objetivo examinar la segmentación del empleo en Ecuador a partir de un enfoque cuantitativo, utilizando la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU 2024) y aplicando el algoritmo k-prototypes para identificar patrones laborales según grupos etarios. Los resultados evidencian un mercado laboral altamente fragmentado, dividido en cuatro perfiles: adolescentes en condiciones de trabajo no remunerado y exclusión educativa; trabajadores adultos en empleo informal sin acceso a seguridad social; un segmento reducido con inserción formal y afiliación al IESS, y adultos mayores en condiciones de autoempleo precario. Asimismo, se observa que los segmentos identificados presentan niveles reducidos de ingreso per cápita, lo que refleja la debilidad estructural del mercado laboral y la limitada capacidad de generación de ingresos adecuados. En este sentido, el estudio evidencia que la educación y la experiencia no son condiciones suficientes para garantizar empleos de calidad en un entorno caracterizado por alta informalidad y bajo dinamismo económico.

Palabras clave: segmentación laboral, empleo, informalidad, k-prototypes, empleo adecuado

Códigos JEL: J21, J24, J46

DOI: 10.47550/RCE/36.1.6

Recibido: 2026-01-28 · **Aceptado:** 2026-06-03

*Universidad Católica de Cuenca, Facultad de Ciencias Económicas y Empresariales, Cuenca, Ecuador. ORCID: [0009-0008-4040-4153](https://orcid.org/0009-0008-4040-4153). CRediT: conceptualización, curación de datos, análisis formal, investigación, metodología, administración del proyecto, software, validación, visualización, redacción - borrador original, redacción - revisión y edición. Correo electrónico: juancarlos.pozo@ucacue.edu.ec.

Employment Segmentation in Ecuador with Sociodemographic Variables: Evidence from k-Prototypes

Juan Carlos Pozo Fonseca

The present study aims to examine labor market segmentation in Ecuador through a quantitative approach, using data from the National Employment, Unemployment, and Underemployment Survey (ENEMDU 2024) and applying the k-prototypes algorithm to identify labor patterns across age groups. The results reveal a highly fragmented labor market, divided into four distinct profiles: adolescents engaged in unpaid work and excluded from education; adult workers employed informally without access to social security; a small segment with formal employment and affiliation to the Ecuadorian Social Security Institute (IESS); and older adults engaged in precarious self-employment. Furthermore, none of the identified segments exhibit income levels substantially above the minimum wage, reflecting the structural weaknesses of the labor market and its limited capacity to generate adequate earnings. In this regard, the study shows that education and experience alone are not sufficient conditions to secure quality employment in an environment characterized by high informality and low economic dynamism.

Keywords: labor segmentation, employment, informality, k-prototypes, adequate employment

JEL Codes: J21, J24, J46

DOI: 10.47550/RCE/36.1.6

Received: 2026-01-28 · **Accepted:** 2026-06-03

1 Introducción

El mercado laboral ecuatoriano ha experimentado transformaciones relevantes en las últimas décadas, reflejando los efectos de la globalización, los cambios demográficos y la inestabilidad económica. A pesar de ciertos avances en indicadores de empleo, persisten profundas heterogeneidades que condicionan las oportunidades de inserción laboral. De acuerdo con los registros del Banco Central del Ecuador [BCE] (2024), el crecimiento económico promedio en los últimos diez años fue de apenas 1.5 %, lo que evidencia limitaciones estructurales para generar empleos de calidad y ampliar las oportunidades productivas. Esta problemática se vuelve más compleja al observar que el comportamiento del empleo varía entre distintos grupos etarios, lo que demanda un análisis segmentado y riguroso.

La necesidad de atender este fenómeno radica en su impacto social y económico, ya que el empleo constituye un determinante central del bienestar de los hogares y del dinamismo económico. Identificar patrones en la segmentación del empleo permite comprender de manera más precisa cómo influyen factores como la edad en la inserción laboral, lo cual resulta esencial para diseñar políticas públicas más efectivas. La trascendencia de este estudio se justifica porque provee evidencia empírica para la formulación de estrategias que promuevan mayor equidad y competitividad en el mercado laboral ecuatoriano.

El objetivo general de la investigación es analizar la segmentación del empleo en el Ecuador según grupos etarios, aplicando técnicas de *clustering* que permitan identificar tipologías diferenciadas. Los objetivos específicos incluyen: (i) determinar los principales factores que caracterizan a cada segmento laboral y (ii) discutir las implicaciones de los resultados de cara a la realidad socioeconómica del Ecuador. La hipótesis central plantea que los patrones de segmentación en el empleo no son aleatorios, sino que responden a estructuras socioeconómicas específicas vinculadas a la edad.

Para abordar el problema, se trabajó con la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) y se aplicó la técnica de *k-prototypes* en *Python*. Previamente, se evaluó la idoneidad del conjunto de datos mediante el estadístico de Hopkins, que arrojó un valor de 0.0107, confirmando la existencia de una clara tendencia a la formación de clústeres. Entre los principales hallazgos, se identificaron cuatro agrupamientos que revelan diferencias sustanciales en las condiciones laborales, especialmente entre jóvenes, adultos en edad productiva y personas mayores.

La contribución de este estudio se centra en ofrecer un marco empírico que combina técnicas de aprendizaje no supervisado con el análisis económico aplicado al empleo. Ello no solo aporta a la literatura académica, sino que también brinda herramientas útiles para los responsables de la política laboral en el país.

2 Revisión de la literatura

2.1 Teoría del ciclo de vida y trayectorias laborales

La teoría del ciclo de vida (Modigliani & Brumberg, 1954) permite comprender cómo las decisiones laborales de los individuos varían a lo largo de las distintas etapas de la vida. Este enfoque plantea que las personas ajustan su comportamiento económico en función de su edad, sus expectativas de ingreso y la acumulación de capital humano.

En las etapas iniciales, los individuos suelen enfrentar mayores dificultades de inserción laboral, caracterizadas por empleos de menor calidad, inestabilidad y bajos ingresos,

debido a la limitada experiencia. En la etapa adulta, se observa una mayor estabilidad laboral y mejores condiciones de empleo, asociadas a la acumulación de habilidades. Finalmente, en edades avanzadas, aumentan las restricciones de acceso a empleo formal, lo que puede derivar en informalidad o salida progresiva del mercado laboral.

En economías en desarrollo como Ecuador, estas dinámicas tienden a intensificarse debido a la alta informalidad y la limitada cobertura de seguridad social, lo que genera trayectorias laborales más inestables a lo largo del ciclo de vida.

Este enfoque resulta fundamental para el presente estudio, ya que permite interpretar las diferencias observadas entre los clústeres identificados mediante k-prototypes, los cuales reflejan una segmentación del mercado laboral asociada al ciclo de vida.

2.2 Factores sociodemográficos y su impacto en el empleo

Desde un enfoque teórico, los factores sociodemográficos son variables clave en el análisis del empleo, pues influyen directamente en la estructura y dinámica del mercado laboral. Variables como la edad, el género, el nivel educativo y la ubicación geográfica son determinantes tanto en el acceso al empleo como en su calidad.

Desde la teoría de la segmentación del mercado laboral ([Doeringer & Piore, 2020](#)), se plantea que el empleo no constituye un mercado homogéneo, sino que está dividido en segmentos con características diferenciadas en términos de estabilidad, salarios y acceso a beneficios. En este contexto, ciertos grupos —como jóvenes, mujeres y personas con menor nivel educativo— tienden a concentrarse en empleos de menor calidad, caracterizados por alta informalidad, bajos ingresos y limitada movilidad laboral.

Esta segmentación no responde únicamente a condiciones de oferta y demanda, sino también a barreras estructurales que limitan la movilidad entre segmentos. De este modo, los jóvenes pueden quedar confinados en empleos temporales debido a su falta de experiencia, mientras que las mujeres pueden concentrarse en sectores informales como resultado de restricciones sociales y culturales. Este enfoque resulta fundamental para el presente estudio, ya que permite interpretar las diferencias observadas entre los clústeres identificados, los cuales reflejan una segmentación del mercado laboral asociada a características sociodemográficas.

2.2.1 Capital humano y capital social

Desde la teoría del capital humano ([Becker, 1964](#)), se plantea que la educación y la capacitación incrementan la productividad de los individuos, mejorando sus oportunidades de acceso al empleo y sus niveles de ingreso. Desde esta perspectiva, el nivel educativo es un factor sociodemográfico determinante en la calidad del empleo, pues aquellos con mayor educación acceden a puestos más calificados y mejor remunerados. Complementando esta idea, la teoría del capital social destaca la importancia de las redes de contactos y el contexto social en el acceso al empleo. Este enfoque argumenta que los individuos que pertenecen a grupos con menor capital social —como algunas minorías o personas en áreas rurales— enfrentan más dificultades para acceder a empleos de calidad debido a la falta de redes que faciliten su inserción en el mercado formal ([Coleman, 1988](#)).

En el caso ecuatoriano, si bien la teoría del capital humano propone que mayores niveles de educación incrementan la productividad y el acceso a mejores oportunidades laborales ([Becker, 1964](#)), esta relación se ve limitada por la baja calidad del sistema educativo, la débil articulación entre el sistema de formación y las demandas del mercado, así como por las barreras estructurales que condicionan el acceso a empleos formales. Muchos jóvenes con formación técnica o universitaria no logran insertarse

en el mercado laboral de manera adecuada, enfrentando subempleo, informalidad o migración forzada. Esto revela que el capital humano, por sí solo, no garantiza movilidad social en un entorno económico como el ecuatoriano, donde la creación de empleo de calidad es limitada y donde predomina un aparato productivo de bajo valor agregado. La existencia de títulos sin empleo se convierte en un fenómeno frecuente, en especial entre jóvenes de sectores rurales o empobrecidos, cuyas credenciales académicas no se traducen en mejores condiciones laborales ni estabilidad económica.

A ello se suma la relevancia del capital social como complemento explicativo. Según Coleman (1988), las redes sociales, los vínculos comunitarios y el acceso a círculos de confianza son determinantes en la posibilidad de acceder a oportunidades laborales, particularmente en contextos con altos niveles de informalidad. En Ecuador, donde gran parte del empleo se consigue por referencia, por conocidos o por recomendación, los jóvenes que provienen de hogares pobres, zonas rurales o familias sin conexiones laborales en el sector formal enfrentan desventajas estructurales. Esta carencia de redes funcionales al interior del sistema económico profundiza la segmentación del empleo y reproduce la desigualdad intergeneracional. Así, dos jóvenes con un mismo nivel educativo pueden tener trayectorias laborales muy distintas dependiendo del capital social que poseen.

Este enfoque resulta fundamental para el presente estudio, ya que permite interpretar las diferencias observadas entre los clústeres identificados, en los cuales el nivel educativo y las condiciones estructurales influyen directamente en la calidad del empleo.

2.2.2 Discriminación y desigualdad de género

En el ámbito teórico, la discriminación de género ha sido ampliamente analizada desde distintas perspectivas. Becker (1971) plantea que la desigualdad en el empleo responde, en parte, a sesgos y estereotipos que subvaloran la capacidad productiva de ciertos grupos, particularmente de las mujeres. Estos prejuicios pueden traducirse en menores oportunidades de acceso a empleos formales, así como en condiciones laborales más precarias y menores niveles de ingreso.

La evidencia reciente refuerza este enfoque, mostrando que las brechas de género en el mercado laboral persisten incluso después de controlar por niveles de educación y experiencia, lo cual evidencia la presencia de factores estructurales (World Bank, 2022). En particular, las mujeres tienden a concentrarse en sectores de baja productividad, empleo informal y ocupaciones vinculadas al trabajo de cuidado, lo que limita su acceso a empleos de calidad y a sistemas de protección social.

En contextos como el ecuatoriano, estas desigualdades se intensifican debido a la alta informalidad, la limitada generación de empleo formal y la persistencia de normas socioculturales que asignan roles diferenciados por género. Como resultado, muchas mujeres enfrentan trayectorias laborales más inestables y menores oportunidades de movilidad económica.

Este enfoque permite comprender que las diferencias en la inserción laboral no responden únicamente a características individuales, sino también a condiciones estructurales que afectan de manera diferenciada a hombres y mujeres. En este sentido, resulta clave para interpretar los patrones observados en los clústeres identificados en el presente estudio.

2.2.3 Migración y ubicación geográfica

La teoría de la migración y la localización geográfica sugiere que las disparidades en las oportunidades económicas entre regiones influyen en la distribución del empleo. Desde los aportes clásicos de Ravenstein (1885), se plantea que los individuos tienden a desplazarse desde zonas con menores oportunidades laborales hacia áreas con mayor dinamismo económico.

En este sentido, la ubicación geográfica ha sido identificada como un factor relevante en la inserción laboral, ya que las zonas rurales suelen presentar menor diversificación productiva, mayor informalidad y menores niveles de ingreso en comparación con las áreas urbanas (Ablaza et al., 2023). Esto genera patrones de movilidad interna que reflejan desigualdades estructurales en el acceso al empleo.

No obstante, en el presente análisis esta variable no evidenció un papel diferenciador en la conformación de los clústeres, lo cual es consistente con la naturaleza de los métodos de aprendizaje no supervisado, donde la agrupación depende de las variables con mayor capacidad de discriminación en los datos. En este sentido, es posible que los efectos asociados a la ubicación geográfica estén siendo capturados indirectamente por otras variables incluidas en el modelo, como el nivel educativo, el ingreso o la categoría ocupacional, tal como se ha documentado en estudios aplicados de segmentación socioeconómica (James et al., 2022).

Por tanto, si bien la literatura resalta la importancia de la dimensión territorial, los resultados sugieren que, en este contexto, otros factores sociodemográficos presentan un mayor peso en la segmentación del mercado laboral. Este enfoque resulta relevante para interpretar las diferencias observadas entre los clústeres identificados en el presente estudio.

2.3 Segmentación laboral y mercado dual

La teoría de la segmentación plantea una visión estructural del mercado laboral, donde este se divide en dos segmentos fundamentales con características y condiciones muy distintas: el mercado primario y el mercado secundario. Esta teoría fue desarrollada por economistas como Doeringer & Piore (2020) y sostiene que el mercado laboral no es un sistema homogéneo, sino que está segmentado de manera que ciertos grupos de trabajadores acceden a empleos con diferentes niveles de estabilidad, remuneración y posibilidades de progreso. Esta segmentación, según la teoría, se debe a factores institucionales y estructurales que van más allá de las cualificaciones individuales de los trabajadores, limitando la movilidad laboral y manteniendo la desigualdad en las oportunidades de empleo.

En contraste, el mercado secundario se caracteriza por empleos inestables, de baja remuneración, y sin beneficios adicionales, ubicados frecuentemente en sectores informales y ocupaciones de menor cualificación. La teoría del mercado dual argumenta que, debido a restricciones estructurales, ciertos colectivos (como mujeres, jóvenes, trabajadores migrantes y minorías étnicas) quedan atrapados en este segmento, con pocas oportunidades de movilidad hacia el mercado primario. Esta situación se explica por la existencia de factores discriminatorios y barreras que impiden que dichos grupos accedan a empleos de mejor calidad, a pesar de contar con capacidades o experiencia comparables.

Este enfoque resulta fundamental para el presente estudio, ya que permite interpretar los clústeres identificados como expresiones de una estructura laboral segmentada. En este sentido, los grupos obtenidos reflejan la coexistencia de segmentos con condiciones

de empleo diferenciadas, donde factores como la estabilidad laboral, el acceso a seguridad social y los niveles de ingreso permiten distinguir entre posiciones cercanas al mercado primario y otras claramente asociadas al segmento secundario.

2.4 Aplicaciones de técnicas de *clustering* en el análisis del mercado laboral

El uso de técnicas de aprendizaje automático en economía laboral ha experimentado un crecimiento significativo en los últimos años, particularmente en el análisis de la heterogeneidad del mercado de trabajo. Dentro de este campo, los métodos de aprendizaje no supervisado, como el *clustering*, han adquirido relevancia al permitir identificar estructuras latentes en los datos sin imponer clasificaciones previas (Athey & Imbens, 2019).

El *clustering* constituye una técnica fundamental del aprendizaje no supervisado cuyo objetivo es agrupar observaciones en función de su similitud, maximizando la homogeneidad intraclúster y la heterogeneidad interclúster (Hastie et al., 2009). En el contexto del mercado laboral, esta metodología permite capturar la complejidad estructural del empleo, caracterizada por múltiples dimensiones como ingresos, educación, estabilidad laboral y características sociodemográficas.

En la literatura reciente, el uso de *clustering* ha sido aplicado de manera creciente para analizar la segmentación laboral, especialmente en economías en desarrollo. A diferencia de los enfoques tradicionales que clasifican a los trabajadores en categorías predefinidas —como formal e informal—, los métodos de *clustering* permiten identificar segmentos de manera endógena, revelando patrones ocultos en los datos (Akay & Yüksel, 2017).

Diversos estudios han empleado estas técnicas para contrastar la hipótesis de segmentación del mercado laboral. Por ejemplo, Martin & Okolo (2022) utilizan algoritmos de *clustering* para analizar la heterogeneidad del mercado laboral en el Reino Unido, encontrando evidencia de agrupaciones diferenciadas que no se explican exclusivamente por el nivel educativo o la productividad.

En economías en desarrollo, la evidencia sugiere que la segmentación laboral es más pronunciada y puede ser capturada eficazmente mediante técnicas de aprendizaje automático. En este sentido, Rodríguez-Guerrero & Quintero (2024) aplican un enfoque de *clustering* basado en FAMD-k-means al mercado laboral colombiano, encontrando tres segmentos claramente diferenciados: un grupo mayoritario con condiciones laborales precarias, un segmento intermedio con características mixtas y un grupo reducido con empleos formales de alta calidad. Este enfoque permite superar las limitaciones de las definiciones tradicionales de informalidad, al incorporar múltiples dimensiones del empleo en la clasificación.

Desde una perspectiva metodológica, el uso de algoritmos como k-means, k-modes y k-prototypes ha facilitado el análisis de bases de datos con variables mixtas —numéricas y categóricas—, lo cual es particularmente relevante en estudios laborales (Grané & Sow-Barry, 2021). Asimismo, la incorporación de técnicas de reducción de dimensionalidad, como el análisis factorial de datos mixtos (FAMD), ha permitido mejorar la calidad de los clústeres al equilibrar la influencia de distintos tipos de variables.

En conjunto, esta literatura evidencia que el *clustering* no solo constituye una herramienta metodológica robusta, sino también un enfoque teóricamente consistente con la hipótesis de segmentación laboral. Al permitir que los datos determinen la estructura del mercado de trabajo, estas técnicas ofrecen una aproximación más flexible, multidi-

mensional y empíricamente fundamentada para el análisis del empleo en economías contemporáneas.

3 Materiales y métodos

Para este trabajo de investigación, se desarrolló una metodología basada en el análisis cuantitativo y el uso de técnicas de *machine learning* para agrupar patrones de empleo en la población ecuatoriana según grupos etarios.

3.1 Diseño de investigación

El presente estudio adopta un enfoque cuantitativo, descriptivo y exploratorio para analizar la segmentación del empleo en Ecuador mediante la aplicación de algoritmos de agrupamiento. Utilizando k-prototypes, se identificó patrones homogéneos de empleo dentro de grupos etarios, permitiendo observar cómo las características laborales varían entre diferentes rangos de edad. El análisis de *clustering* permitió segmentar la población en subgrupos significativos que comparten características laborales comunes, como estabilidad, remuneración y sector de empleo.

3.2 Fuente de datos

La fuente de datos principal es la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) 2024 (Instituto Nacional de Estadísticas y Censos [INEC], 2024). Esta encuesta proporciona datos detallados sobre la situación laboral de los habitantes de Ecuador, incluyendo variables sociodemográficas y económicas, tales como edad, género, nivel de educación, tipo de empleo (formal/informal), sector de actividad económica, nivel de ingreso, régimen laboral y horas trabajadas, ubicación geográfica (zona urbana o rural).

Estas variables permiten analizar el empleo en función de grupos de edad y otras características clave, lo cual facilita una segmentación detallada y relevante para el contexto ecuatoriano.

3.3 Preprocesamiento de datos

- **Limpieza y filtrado de datos:** se realizó una limpieza exhaustiva de los datos de la ENEMDU 2024, eliminando valores faltantes o inconsistentes y filtrando los datos para conservar registros con información válida en las variables sociodemográficas y laborales seleccionadas. Asimismo, se normalizaron las variables cuantitativas (como ingreso y horas trabajadas) y se codificaron las variables categóricas (como sector económico y tipo de empleo) para su procesamiento.
- **Estandarización de variables:** las variables cuantitativas se estandarizaron para garantizar que todas tengan una escala comparable. Este paso es crucial para evitar que variables con mayores valores absolutos dominen el análisis de agrupamiento.

Por otra parte, previo a la aplicación del algoritmo k-prototypes, se realizó un proceso de limpieza y depuración de la base de datos. En esta etapa se eliminaron registros con más del 50 % de valores faltantes, además de llevar a cabo un tratamiento inicial de datos atípicos. Asimismo, se descartaron variables redundantes que no aportaban información significativa al modelo, con el fin de reducir la dimensionalidad y mejorar la robustez de los clústeres obtenidos.

Para el resto de valores faltantes, se aplicó un proceso de imputación simple: se utilizó la media en el caso de variables numéricas y la moda para variables categóricas. Esta decisión se fundamenta en la naturaleza del algoritmo k-prototypes, el cual requiere una matriz completa de datos y combina simultáneamente variables de distinta naturaleza.

Si bien existen métodos más sofisticados de imputación, como la imputación múltiple o el uso de algoritmos basados en k-nearest neighbors (KNN), estos enfoques pueden introducir mayor complejidad y potencial sobreajuste, especialmente en contextos donde el objetivo principal es la identificación de patrones estructurales a nivel agregado más que la predicción individual.

En este sentido, se optó por un enfoque parsimonioso, que permite garantizar la estabilidad y reproducibilidad del modelo sin introducir supuestos adicionales sobre la estructura de los datos. No obstante, se reconoce que este tipo de imputación puede generar cierto grado de sesgo, por lo que futuras investigaciones podrían explorar métodos más avanzados con el fin de evaluar la robustez de los resultados obtenidos.

Adicionalmente, es importante señalar que la base de datos ENEMDU 2024 incluye información tanto de la población adulta (15 años y más) como de población menor de edad a través del módulo de trabajo infantil. En el presente estudio, no se realizó una restricción explícita de la muestra por edad en las etapas iniciales del procesamiento de datos. Como resultado, el algoritmo de *clustering* identificó de manera endógena un segmento caracterizado por individuos con edades significativamente menores al umbral de la población económicamente activa convencional. Este resultado no fue una decisión metodológica deliberada, sino un hallazgo emergente del análisis no supervisado.

Posterior al proceso de limpieza, depuración e imputación de los datos, la base final utilizada para la estimación del modelo quedó conformada por un total de 341394 observaciones correspondientes a la población económicamente activa. Este tamaño muestral garantiza una adecuada representatividad del mercado laboral ecuatoriano y permite la identificación robusta de patrones estructurales mediante técnicas de *clustering* no supervisado.

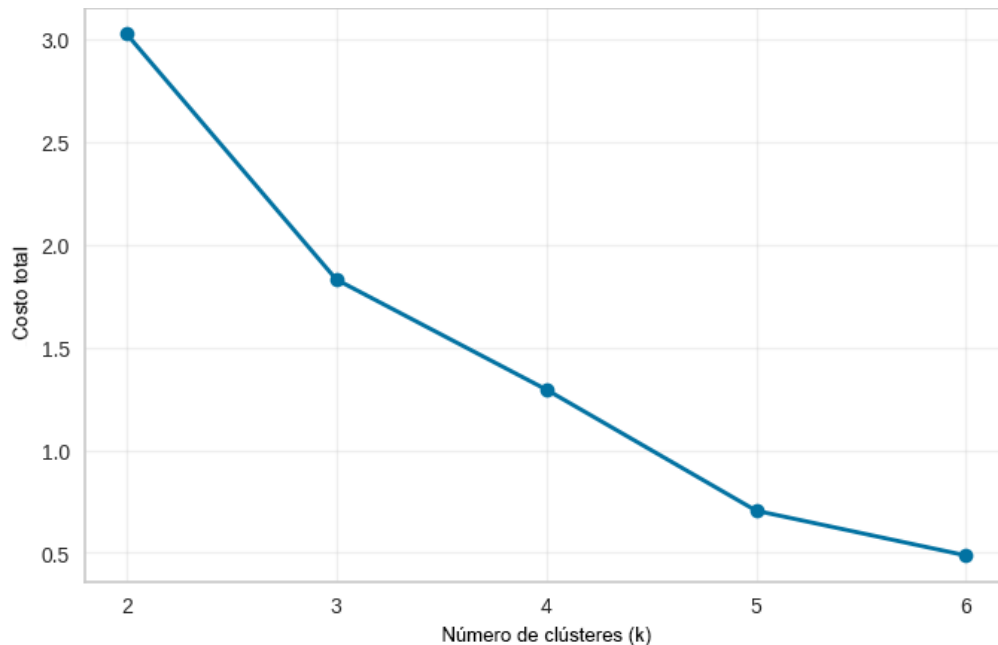
3.4 Definición de variables para *clustering*

Las variables seleccionadas se ajustan en función de su relevancia para el análisis de segmentación laboral por grupos etarios. Entre las variables principales están *edad*, (estratificada en grupos), *nivel de ingreso mensual*, *nivel de educación alcanzado*, *tipo de empleo*, (formal/informal), *sector económico*, *régimen laboral* y *horas trabajadas*. Estas variables sirven para definir los perfiles de empleo de cada grupo etario, permitiendo que el modelo k-prototypes identifique patrones específicos en función de estas características.

3.5 Aplicación del algoritmo k-prototypes

La determinación del número óptimo de clústeres (K) constituye un paso fundamental en la aplicación del algoritmo k-prototypes, ya que influye directamente en la capacidad del modelo para identificar patrones representativos dentro de los datos. Con el objetivo de seleccionar un valor adecuado de *k*, se empleó el método del codo, el cual permite evaluar la variación del costo total del modelo.

Figura 1. Método del codo para la selección del número óptimo de *clústeres* (*k*-prototypes)



Como se observa en la Figura 1, el costo total del modelo disminuye de manera pronunciada a medida que aumenta el número de clústeres desde $k = 2$ hasta $k = 4$. A partir de este punto, la reducción del costo se vuelve menos significativa, evidenciando un cambio en la pendiente de la curva. Este comportamiento indica la presencia de un punto de inflexión en $k = 4$, lo cual es consistente con el criterio del método del codo. En consecuencia, se seleccionó $k = 4$ como el número óptimo de clústeres para el análisis.

Se implementó el algoritmo de *k*-prototypes utilizando las variables seleccionadas (tanto continuas como categóricas), lo que permitió construir clústeres que reflejen perfiles heterogéneos del mercado laboral. Este procedimiento facilita la identificación de agrupamientos de individuos con características similares, describiendo patrones comunes en cada grupo etario y proporcionando una visión integral que combina factores cuantitativos y cualitativos. Para garantizar la reproducibilidad de los resultados, se fijó una semilla aleatoria en la implementación del algoritmo, de modo que los centroides iniciales se seleccionen de manera consistente en cada ejecución.

Por otra parte, el parámetro γ (*gamma*) del algoritmo *k*-prototypes incorpora un hiperparámetro cuya función es ponderar la contribución relativa de las variables numéricas y categóricas en el cálculo de la disimilitud total, lo que permite equilibrar ambos tipos de información dentro del proceso de agrupamiento. En la presente investigación, el valor de γ no fue fijado manualmente, sino que se dejó en su configuración automática según la implementación del algoritmo en *Python*. En este caso, el modelo estimó internamente un valor de $\gamma = 0.10$ durante el proceso de ajuste.

El valor estimado de $\gamma = 0.104$ sugiere un balance adecuado entre variables numéricas y categóricas, sin evidencia de dominancia excesiva de alguno de los dos tipos de información en el proceso de agrupamiento. Esta estimación automática se basa en la escala y dispersión de las variables numéricas, las cuales fueron previamente normalizadas mediante *MinMaxScaler*, lo que contribuye a un balance adecuado entre la parte numérica y categórica del modelo.

3.6 Validación de resultados

Con el fin de validar los resultados del análisis de *clustering* y garantizar la solidez de los agrupamientos obtenidos, se aplicaron diversos métodos de validación estadística. En primer lugar, se utilizó el estadístico de Hopkins, cuyo valor fue 0.0107. Este resultado, al encontrarse cercano a cero, indica una fuerte tendencia a que los datos presenten una estructura no aleatoria, lo cual confirma que el conjunto de información posee una marcada propensión a conformar clústeres significativos y bien diferenciados.

Tabla 1. Métricas de validación del modelo de *clustering* mediante k-prototypes

Métrica	Valor	Interpretación general
Estadístico de Hopkins	0.0006	Alta no aleatoriedad/fuerte estructura
Coefficiente de silueta	0.143	Cohesión moderada con cierto solapamiento
Índice Davies-Bouldin	2.5286	Separación aceptable entre clústeres
Adjusted Rand index (promedio)	0.7287	Buena estabilidad del modelo
Parámetro γ (k-prototypes)	0.0891	Balance adecuado entre variables

Elaboración: autor

La Tabla 1 presenta las principales métricas de validación utilizadas para evaluar la calidad y robustez del modelo de *clustering* implementado mediante el algoritmo k-prototypes. En primer lugar, el estadístico de Hopkins (0.0006) evidencia que los datos presentan una estructura claramente no aleatoria, lo que justifica la aplicación de técnicas de segmentación. En cuanto a la calidad del agrupamiento, el coeficiente de silueta (0.143) sugiere una cohesión moderada entre los elementos de cada clúster, mientras que el índice de Davies-Bouldin (2.5286) indica una separación aceptable entre los grupos identificados.

Adicionalmente, el análisis de estabilidad basado en el *adjusted Rand index* (ARI) muestra un valor promedio de 0.7287, lo que refleja una buena consistencia en la asignación de los individuos a los clústeres frente a distintas inicializaciones del algoritmo. Finalmente, el valor del parámetro γ (0.0891) sugiere un balance adecuado entre la contribución de las variables numéricas y categóricas en el proceso de agrupamiento. En conjunto, estas métricas confirman la validez metodológica del modelo y su capacidad para identificar patrones estructurales relevantes dentro del mercado laboral analizado.

4 Resultados y limitaciones

La segmentación obtenida mediante el algoritmo k-prototypes permitió identificar cuatro grupos diferenciados dentro de la población analizada. En términos de tamaño, el clúster 0 concentra la mayor proporción de individuos, con un total de 130432 observaciones, lo que representa aproximadamente el 38.2 % del total de la muestra. En segundo lugar, el clúster 3 agrupa a 95254 individuos (27.9 %), seguido por el clúster 1, con 71786 observaciones (21.0 %). Finalmente, el clúster 2 constituye el segmento de menor tamaño, con 43922 individuos, equivalente al 12.9 % de la población analizada. Esta distribución evidencia una estructura heterogénea del mercado laboral, en la que ciertos segmentos concentran una mayor proporción de la población, mientras que otros representan grupos más específicos o nichos dentro de la dinámica ocupacional.

Tabla 2. Tamaño y proporción de los clústeres estimados

Clúster	Observaciones	Porcentaje
0	130432	38.2 %
1	71786	21.0 %
2	43922	12.9 %
3	95254	27.9 %
TOTAL	341394	100 %

Elaboración: autor

4.1 Caracterización de clústeres (centroides)

Con el objetivo de caracterizar los clústeres obtenidos mediante el algoritmo k-prototypes, se presentan los valores promedio para variables numéricas y las categorías predominantes para variables categóricas en cada grupo. Esta caracterización permite identificar los perfiles socioeconómicos asociados a cada clúster, facilitando la interpretación de los segmentos resultantes.

Tabla 3. Resultados del k-prototypes

Variable	Clúster 0	Clúster 1	Clúster 2	Clúster 3
Observaciones	130432	71786	43922	95254
Edad	14.031	38.259	41.560	61.126
Ingreso per cápita	225.899	281.128	607.936	295.506
Sexo	Mujer	Hombre	Mujer	Mujer
Relación de parentesco	Hijo/a	Jefe	Jefe	Jefe
Seguro social	Ninguno	Ninguno	IESS, seguro general	Ninguno
Estado civil	Soltero	Unión libre	Casado	Casado
Asiste a clases	Sí	No	No	No
Razón por la que no asiste	Falta de recursos económicos	Por trabajo	Terminó sus estudios	Edad
Nivel de instrucción	Educación básica	Secundaria	Superior universitario	Primaria
Cómo se considera	Mestizo	Mestizo	Mestizo	Mestizo
Trabajó la semana pasada	No	Sí	Sí	No
Categoría de ocupación	Trabajador del hogar no remunerado	Empleado privado	Empleado privado	Cuenta propia
Aporte a seguridad social	No aporta	No aporta	IESS general	No aporta
Recibió ingresos de capital	No	No	No	No

Elaboración: autor

4.1.1 Segmento 0: menores de edad en trabajo no remunerado, con privación educativa y alta vulnerabilidad estructural

El segmento 0 concentra 130432 observaciones, equivalentes al 38.2 % de la muestra, lo que lo convierte en el grupo de mayor tamaño. Su rasgo más distintivo es la edad promedio de 14.03 años, con un rango que va de 4 a 55 años. Aunque el rango máximo revela cierta heterogeneidad interna, la media etaria confirma que el perfil dominante corresponde a niños, niñas y adolescentes. Este dato es metodológica y sustantivamente crucial, porque desplaza la interpretación del clúster desde una simple categoría «juvenil» hacia una manifestación de inserción laboral temprana.

El perfil modal muestra a una mujer, hija dentro del hogar, soltera, mestiza, con educación básica, que asiste a clases, pero cuya principal razón para no hacerlo plenamente o mantenerse en el sistema educativo está asociada a la falta de recursos económicos. Esta combinación de características revela una tensión central entre escolarización y participación en actividades productivas o domésticas no remuneradas. No se trata solamente de un grupo con baja edad, sino de un segmento donde la trayectoria educativa aparece condicionada por la restricción económica del hogar.

En términos ocupacionales, este grupo se caracteriza por la categoría *trabajador del hogar no remunerado*, sin afiliación a la seguridad social, sin aportes previsionales y sin ingresos de capital. Su ingreso per cápita promedio es de 225.90, el más bajo entre los clústeres, con una elevada desviación estándar (228.05), lo que sugiere que incluso dentro de este segmento vulnerable existen diferencias de ingreso entre hogares. Sin embargo, la baja media de ingresos confirma que este clúster se ubica en condiciones materiales frágiles.

Desde una perspectiva analítica, este segmento no puede entenderse como una inserción laboral «normal» dentro del mercado de trabajo, sino como un espacio donde confluyen trabajo infantil, dependencia económica, desigualdad intrahogar y precariedad educativa. La relación de parentesco «hijo/a» y la ocupación no remunerada sugieren que la participación económica ocurre principalmente dentro de la unidad doméstica, más que en una relación laboral clásica. Esto tiene implicaciones importantes: la baja o nula monetización del trabajo no implica ausencia de explotación económica, sino invisibilización estadística del aporte productivo de menores.

La presencia mayoritaria de mujeres dentro de este segmento también introduce una dimensión de género relevante. Es plausible interpretar que una parte de este clúster esté vinculada a tareas de cuidado, apoyo doméstico o trabajo reproductivo no remunerado, lo que reproduce desigualdades tempranas en el uso del tiempo, la acumulación de capital humano y las oportunidades futuras de inserción laboral.

4.1.2 Segmento 1: adultos en empleo privado sin protección social: formalidad aparente e informalidad real

El segmento 1 agrupa 71786 observaciones (21.0 % de la muestra) y representa a individuos en plena edad laboral, con una edad promedio de 38.26 años. A diferencia del clúster 0, aquí ya no predomina la dependencia intrahogar, sino la condición de jefe de hogar, lo que sugiere una inserción económica activa y responsabilidad directa en la generación de ingresos del hogar. El perfil modal corresponde a un hombre, mestizo, en unión libre, con educación secundaria, que trabajó la semana pasada y cuya inasistencia al sistema educativo se explica por trabajo.

Esta última característica es especialmente reveladora: indica que la desvinculación educativa no responde a un cierre natural del ciclo formativo, sino a la necesidad

de participar en el mercado laboral. Esto sugiere trayectorias de entrada temprana al trabajo, con interrupción o truncamiento de la acumulación educativa. El promedio de 4.83 años aprobados y su relativamente baja dispersión (1.65) indican un perfil educativo más homogéneo que otros clústeres, concentrado en niveles intermedios y sin amplia expansión hacia niveles superiores.

Ocupacionalmente, este segmento se define por la categoría *empleado privado*, pero con una característica decisiva: no aporta a la seguridad social y su forma de aseguramiento es ninguna. Esta combinación es una de las más importantes del estudio, porque revela una situación de formalidad ocupacional aparente con informalidad institucional real. Es decir, los individuos trabajan como asalariados en el sector privado, pero sin acceso a los derechos laborales asociados a una relación formal de trabajo.

Su ingreso per cápita promedio es de 281.13, superior al del clúster 0, pero aún bajo en términos relativos, con una desviación estándar de 272.53. Esto indica que el segmento, aunque activo laboralmente, no alcanza niveles robustos de bienestar económico y presenta además dispersión significativa en sus resultados materiales. No se trata de un grupo excluido del mercado, sino de un grupo incorporado en condiciones de protección incompleta y vulnerabilidad persistente.

Este segmento es en especial importante porque probablemente representa una fracción amplia del mercado laboral ecuatoriano contemporáneo: trabajadores insertos en relaciones asalariadas, pero sin acceso pleno a la institucionalidad laboral. Desde una perspectiva de segmentación, esto confirma que la dicotomía formal/informal no siempre coincide con la dicotomía asalariado/no asalariado. Se puede estar en el sector privado, tener jefe, ocupar un puesto aparentemente moderno, y aun así permanecer fuera del sistema de seguridad social.

En términos estructurales, este clúster refleja una economía donde el empleo privado no garantiza derechos, y donde la inserción laboral está marcada por una precariedad silenciosa. La combinación de jefatura de hogar, educación media y ausencia de protección social también implica un alto riesgo de transmisión intergeneracional de la vulnerabilidad: el individuo trabaja, sostiene el hogar, pero no acumula suficiente seguridad institucional para estabilizar su trayectoria de largo plazo.

4.1.3 Segmento 2: trabajadoras formales con mayor ingreso relativo: una minoría integrada al núcleo protegido del mercado laboral

El segmento 2 reúne 43922 observaciones (12.9 % de la muestra), siendo el grupo más pequeño, pero también el más aventajado en términos de ingreso y formalidad. Su edad promedio es de 41.56 años, con rango entre 18 y 76 años, lo que sugiere una población adulta madura y plenamente inserta en el mercado laboral. El perfil modal corresponde a una mujer, jefa de hogar, casada, mestiza, con educación superior universitaria, que trabajó la semana pasada, no asiste a clases porque terminó sus estudios, trabaja como empleada privada, y está afiliada al IESS, seguro general.

Este conjunto de rasgos convierte a este segmento en el más claramente vinculado al núcleo formal y protegido del mercado laboral. A diferencia de los clústeres 0, 1 y 3, aquí la educación no solo está presente, sino culminada; el empleo no solo existe, sino que está institucionalmente respaldado, y la jefatura de hogar no se sostiene desde la precariedad, sino desde una posición relativamente más estable.

El dato económico más contundente es su ingreso per cápita promedio de 607.94, el más alto de todos los clústeres. Sin embargo, la desviación estándar también es elevada (549.66), lo que indica que incluso dentro del segmento formal existen diferencias importantes de ingreso. Esta dispersión sugiere que la formalidad no elimina la desigualdad

interna, pero sí desplaza a este grupo hacia un umbral de bienestar claramente superior al de los demás segmentos.

Un hallazgo particularmente potente es que el grupo con mayores ingresos esté encabezado modalmente por una mujer y no por un hombre. Esto puede interpretarse como la existencia de un subconjunto femenino con fuerte acumulación de capital humano, integración al empleo asalariado formal y rol económico principal dentro del hogar. En términos analíticos, este resultado rompe parcialmente con la idea de una segmentación laboral puramente masculina en el espacio de la formalidad, aunque no debe exagerarse: el tamaño reducido del clúster indica que este perfil es minoritario en la estructura general.

Desde la teoría del capital humano, este es el clúster donde la relación entre educación, empleo e ingreso parece funcionar con más claridad. No obstante, desde una lectura crítica, también revela el carácter excluyente del mercado formal: solo una fracción relativamente pequeña de la muestra logra combinar educación superior, afiliación al IESS y mayores ingresos. Por tanto, más que representar la norma, este clúster representa la excepción integrada dentro de un sistema laboral mayoritariamente precario.

4.1.4 **Segmento 3: adultas mayores en autoempleo sin protección: subsistencia prolongada y fragilidad en la vejez**

El segmento 3 comprende 95254 observaciones (27.9 % de la muestra) y está claramente definido por su edad promedio de 61.13 años, la más alta de todos los clústeres. El perfil modal corresponde a una mujer, jefa de hogar, casada, mestiza, con educación primaria, que no asiste a clases por edad, no trabajó la semana pasada, se desempeña como cuenta propia, no posee seguro social y no aporta al sistema de seguridad social.

Este segmento es particularmente importante porque representa una forma de inserción laboral extendida hacia edades avanzadas, pero sin respaldo institucional. A diferencia del clúster 2, donde la edad madura se combina con formalidad, aquí la madurez se asocia a bajo nivel educativo, autoempleo y exclusión previsional. El promedio educativo de 4.84 años aprobados es casi idéntico al del clúster 1, pero la estructura ocupacional es distinta: ya no se trata de empleo privado asalariado, sino de trabajo por cuenta propia, lo que sugiere una trayectoria laboral construida en la informalidad o fuera del núcleo formal de protección.

Su ingreso per cápita promedio es de 295.51, apenas superior al clúster 1 y muy inferior al clúster 2, con una dispersión alta (325.41). Esto significa que, pese a la extensa trayectoria vital, este grupo no logra traducir experiencia o permanencia en el tiempo en mejores resultados económicos. Más bien, parece tratarse de un segmento donde la continuidad en el mercado laboral responde a necesidad de subsistencia, no a acumulación exitosa.

La combinación de edad avanzada más cuenta propia y más no aporte sugiere una inserción laboral prolongada por ausencia de mecanismos de retiro adecuados. En otras palabras, no estamos frente a un autoempleo necesariamente emprendedor o voluntario, sino posiblemente frente a un autoempleo forzado por la insuficiencia de pensiones, ahorros o redes de protección. La condición de jefa de hogar refuerza esta hipótesis: la permanencia económica de estas personas no depende de un ingreso pasivo garantizado, sino de su capacidad de seguir generando recursos.

La predominancia femenina vuelve a introducir una dimensión estructural relevante. Este clúster puede estar reflejando el resultado acumulado de trayectorias laborales femeninas marcadas por baja escolaridad, informalidad histórica y exclusión previsional, cuyas consecuencias se manifiestan con mayor dureza en edades avanzadas.

En términos de segmentación laboral, este grupo representa el extremo de la informalidad persistente en la vejez. Es un segmento donde el mercado laboral no opera como mecanismo de movilidad, sino como espacio de supervivencia prolongada. Su existencia cuestiona la capacidad del sistema laboral y de protección social para garantizar una salida digna del ciclo productivo.

4.2 Discusión integradora de los cuatro segmentos

Los cuatro clústeres identificados permiten reconstruir una tipología del mercado laboral ecuatoriano organizada por tres ejes estructurales: ciclo de vida, nivel educativo y grado de formalización laboral.

El segmento 0 concentra exclusión temprana, dependencia y trabajo no remunerado; el segmento 1 representa inserción activa pero precaria en el sector privado sin protección; el segmento 2 condensa la minoría formal con educación superior y mayor ingreso, y el segmento 3 refleja la persistencia de la informalidad en edades avanzadas, especialmente a través del autoempleo sin cobertura.

La lectura conjunta de estos resultados muestra que el mercado laboral ecuatoriano no está segmentado solo entre formal e informal, sino entre trayectorias vitales profundamente desiguales. El acceso al empleo, la protección social y el ingreso no se distribuyen de forma uniforme a lo largo del ciclo de vida. Al contrario, la precariedad aparece tempranamente, persiste en la adultez activa y se prolonga en la vejez para amplios grupos de la población.

Además, la formalidad laboral emerge como un espacio reducido y selectivo. El único clúster claramente protegido es el 2, y representa apenas el 12.9 % de la muestra. Esto implica que la mayoría de la población se ubica en segmentos donde el trabajo no asegura seguridad social, estabilidad ni ingresos suficientes. Desde esta perspectiva, el estudio no solo identifica clústeres estadísticos, sino que revela una estructura laboral estratificada, donde las oportunidades de acumulación de capital humano y bienestar económico están distribuidas de manera profundamente desigual.

4.3 Variabilidad interna de los *clústeres*

Tabla 4. Estadísticos descriptivos (media, desviación estándar y rango) de variables numéricas por clúster

Clúster	Edad				Año aprobado				Ingreso per cápita			
	Media	Desv. est.	Mín.	Máx.	Media	Desv. est.	Mín.	Máx.	Media	Desv. est.	Mín.	Máx.
0	14.03	7.71	4	55	4.31	2.91	0	10	225.90	228.05	0.60	6000
1	38.26	11.95	14	76	4.83	1.65	1	10	281.13	272.53	0.60	8000
2	41.56	12.12	18	76	3.85	1.48	0	10	607.94	549.66	0.75	13250
3	61.13	12.30	24	76	4.84	1.61	0	10	295.51	325.41	0.60	15800

Elaboración: autor

4.3.1 Clúster 0: inserción laboral temprana y alta heterogeneidad

El clúster 0 presenta una edad promedio de 14.03 años, con una desviación estándar de 7.71 y un rango que va desde los 4 hasta los 55 años. En términos educativos, registra un promedio de 4.31 años aprobados, con alta dispersión ($\sigma = 2.91$), mientras que el ingreso per cápita promedio es de 225.90, también con elevada variabilidad ($\sigma = 228.05$).

Este clúster evidencia una fuerte presencia de población menor de edad, lo que sugiere dinámicas de trabajo infantil o inserción laboral temprana. No obstante, el amplio rango de edad indica que el grupo no es completamente homogéneo, incorporando también individuos adultos. La elevada dispersión tanto en educación como en ingresos refuerza la idea de un segmento altamente heterogéneo y vulnerable, donde coexisten distintas trayectorias laborales, posiblemente asociadas a informalidad y baja estabilidad económica.

4.3.2 Clúster 1: población laboral activa con condiciones intermedias

El clúster 1 presenta una edad promedio de 38 años, con una desviación estándar de 11.95 y un rango amplio (14 a 76 años). En educación, el promedio es de 4.83 años aprobados, con menor dispersión ($\sigma = 1.65$), mientras que el ingreso per cápita promedio alcanza 281.13, con alta variabilidad ($\sigma = 272.53$).

Este clúster representa una población en edad laboral activa, con un nivel educativo relativamente homogéneo en comparación con otros grupos. Sin embargo, la elevada dispersión en ingresos sugiere que, a pesar de compartir características educativas similares, los individuos enfrentan resultados económicos desiguales. Esto es consistente con un mercado laboral segmentado, donde la educación no garantiza necesariamente mejores ingresos, especialmente en contextos de informalidad.

4.3.3 Clúster 2: segmento de mayor ingreso con desigualdad interna

El clúster 2 muestra una edad promedio de 41.56 años, con desviación estándar de 12.12 y un rango entre 18 y 76 años. Presenta un promedio educativo de 3.85 años aprobados, ligeramente inferior a otros clústeres, pero destaca por un ingreso per cápita promedio de 607.94, el más alto entre todos los grupos, acompañado de una elevada dispersión ($\sigma = 549.66$).

Este clúster se caracteriza por concentrar a los individuos con mayores niveles de ingreso, lo que sugiere una mejor inserción laboral o acceso a actividades más rentables. Sin embargo, la alta variabilidad interna indica una fuerte desigualdad dentro del grupo, donde coexisten individuos con ingresos elevados y otros significativamente menores. Además, el menor nivel educativo promedio refuerza la idea de que en este contexto el ingreso no depende exclusivamente del capital humano formal, sino posiblemente de factores como experiencia, tipo de ocupación o acceso a recursos productivos.

4.3.4 Clúster 3: población envejecida con ingresos moderados y dispersos

El clúster 3 presenta la edad promedio más alta (61 años), con desviación estándar de 12.30 y un rango entre 24 y 76 años. En educación, el promedio es de 4.84 años aprobados, mientras que el ingreso per cápita promedio es de 295.51, con una dispersión considerable ($\sigma = 325.41$).

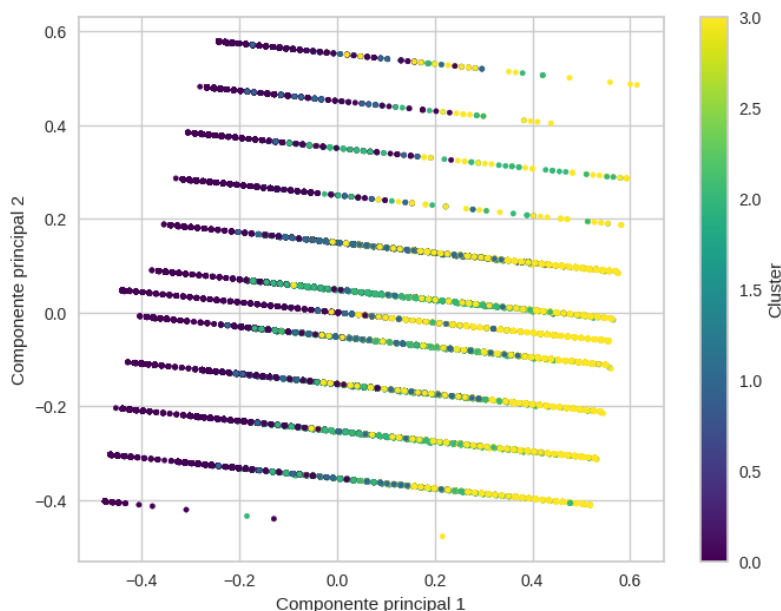
Este clúster agrupa principalmente a individuos de mayor edad, lo que sugiere una población cercana o en etapa de retiro. La variabilidad en ingresos puede reflejar la coexistencia de diferentes fuentes de sustento, como pensiones, trabajo informal o dependencia económica. A pesar de tener niveles educativos similares al clúster 1, los ingresos no son significativamente superiores, lo que sugiere limitaciones estructurales en la generación de ingresos para la población adulta mayor.

4.4 Visualización de clústeres (PCA)

El gráfico de PCA permite observar la distribución de los individuos en un espacio bidimensional construido a partir de las principales componentes de las variables numéricas. Se identifican cuatro grupos diferenciados correspondientes a los clústeres estimados mediante el algoritmo k-prototypes. Se evidencia una separación parcial entre clústeres, particularmente en el eje horizontal (componente principal 1), donde ciertos grupos tienden a concentrarse en rangos específicos. Sin embargo, también se observa

superposición entre clústeres, lo cual es consistente con la naturaleza continua y multi-dimensional de los datos socioeconómicos.

Figura 2. Visualización de clústeres mediante análisis de componentes principales (PCA)



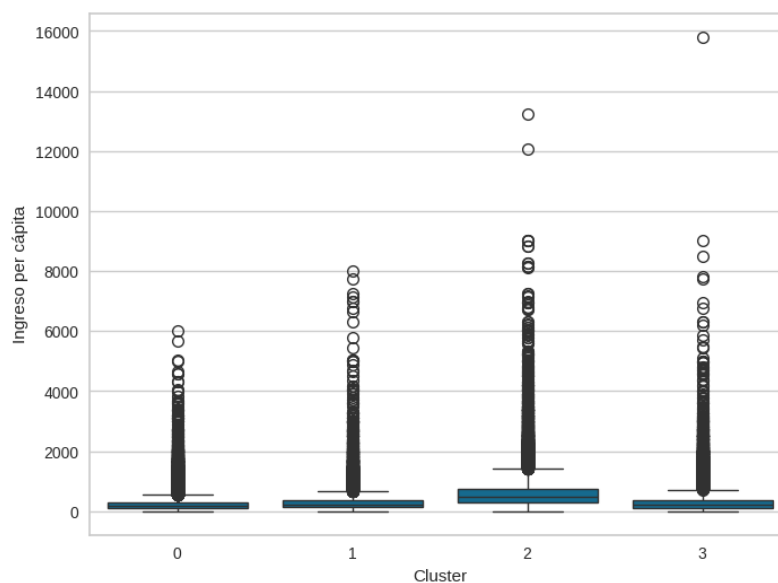
Elaboración: autor

La presencia de solapamiento entre clústeres sugiere que, si bien existen patrones diferenciados, las fronteras entre los grupos no son completamente rígidas. Este comportamiento es esperado en datos reales, donde variables como ingreso, educación y edad presentan distribuciones continuas. Asimismo, la estructura en bandas paralelas observada en el gráfico indica posibles efectos de discretización en algunas variables (por ejemplo, años de educación), lo que influye en la forma de la proyección en el espacio PCA. En conjunto, el gráfico respalda la validez del *clustering*, mostrando que los grupos identificados capturan estructuras relevantes, aunque con cierto grado de heterogeneidad interna.

El diagrama de caja (Figura 3) muestra la distribución del ingreso per cápita en cada clúster. Se observa que el clúster 2 presenta la mediana más alta, así como una mayor dispersión, indicando que este grupo concentra a los individuos con mayores niveles de ingreso.

Por otro lado, los clústeres 0, 1 y 3 presentan medianas significativamente menores, aunque con presencia de valores atípicos elevados en todos los grupos. La amplia dispersión del ingreso en todos los clústeres refleja la desigualdad interna dentro de cada grupo, especialmente en el clúster 2, donde coexisten individuos con ingresos altos y muy altos. La presencia de numerosos valores atípicos en todos los clústeres sugiere que el ingreso per cápita presenta una distribución altamente sesgada, característica común en variables económicas. Además, la clara diferencia en la mediana del clúster 2 refuerza la idea de que este grupo representa un segmento económicamente más favorecido, mientras que los demás clústeres agrupan poblaciones con menor capacidad de generación de ingresos.

Figura 3. Distribución del ingreso per cápita por clúster



Elaboración: autor

4.5 Validación estadística (ANOVA/chi-cuadrado)

Con el fin de evaluar la significancia estadística de las diferencias entre clústeres, se aplicaron pruebas ANOVA (Tabla 5) para las variables numéricas clave. Los resultados evidencian que todas las variables analizadas presentan diferencias altamente significativas entre los clústeres ($p < 0.001$). En particular, la variable *edad* muestra el mayor valor del estadístico F, lo que indica que la segmentación captura de manera clara la estructura etaria de la población. Asimismo, las variables relacionadas con educación e ingreso per cápita también presentan diferencias significativas entre grupos, lo que confirma que los clústeres identificados reflejan patrones socioeconómicos diferenciados. En conjunto, estos resultados validan estadísticamente la segmentación obtenida mediante el algoritmo k-prototypes, evidenciando que los clústeres no son producto del azar, sino que responden a estructuras subyacentes en los datos.

Tabla 5. Pruebas ANOVA de diferencias entre *clústeres* en variables socioeconómicas clave

Variable	Estadístico F	p-valor
Edad	365450	< 0.001
Años de educación	2872	< 0.001
Ingreso per cápita	15642	< 0.001

Elaboración: autor

Con el objetivo de evaluar la asociación entre los clústeres y las variables categóricas, se aplicaron pruebas de chi-cuadrado de independencia (Tabla 6). Los resultados muestran que todas las variables analizadas presentan asociaciones estadísticamente significativas con la pertenencia a los clústeres ($p < 0.001$). En particular, variables como el nivel de instrucción, la categoría ocupacional y el estado civil presentan valores elevados del

estadístico chi-cuadrado, lo que indica una fuerte dependencia entre estas características y la segmentación obtenida. Asimismo, variables relacionadas con condiciones socioeconómicas, como pobreza y pobreza extrema, también muestran diferencias significativas entre clústeres, lo que evidencia que los grupos identificados capturan desigualdades estructurales en la población. En conjunto, estos resultados refuerzan la validez del modelo de *clustering*, confirmando que los clústeres no son aleatorios, sino que reflejan patrones socioeconómicos y demográficos claramente diferenciados.

Tabla 6. Resultados de la prueba chi-cuadrado para variables categóricas por clúster

Variable	Chi-cuadrado	p-valor
Sexo	23752.25	< 0.001
Estado civil	157382.75	< 0.001
Nivel de instrucción	186992.96	< 0.001
Categoría de ocupación	87991.12	< 0.001
Condición de pobreza	12080.41	< 0.001
Condición de pobreza extrema	4536.57	< 0.001

Elaboración: autor

Cabe señalar que, debido al elevado tamaño muestral del estudio, la significancia estadística de las pruebas ANOVA y chi-cuadrado debe interpretarse junto con la magnitud y el sentido sustantivo de las diferencias observadas entre clústeres. Por ello, los valores *p* se consideran como evidencia complementaria de validación, y no como único criterio de interpretación de los segmentos identificados.

5 Conclusiones y discusión

5.1 Segmentación estructural con trayectorias laborales diferenciadas

Los resultados confirman la existencia de una segmentación estructural del mercado laboral ecuatoriano, donde los individuos se distribuyen en trayectorias diferenciadas desde etapas tempranas del ciclo de vida. El clúster 0 refleja una inserción temprana en actividades no remuneradas vinculadas al hogar; el clúster 1 evidencia empleo asalariado sin protección social; el clúster 2 representa el único segmento con inserción formal y acceso al IESS, mientras que el clúster 3 muestra dinámicas de autoempleo precario en edades avanzadas.

Esta configuración sugiere que la desigualdad laboral no es un fenómeno estático, sino acumulativo, donde las condiciones iniciales condicionan las trayectorias futuras de inserción laboral. En este sentido, se requiere un enfoque de política pública basado en trayectorias, que intervenga en momentos críticos del ciclo de vida para evitar la consolidación de esquemas de informalidad y exclusión.

5.2 Políticas diferenciadas por segmento: de la evidencia a la intervención concreta

Los resultados permiten diseñar intervenciones específicas para cada grupo, superando enfoques homogéneos que han demostrado ser insuficientes. En el caso del clúster 0, conformado por menores en condiciones de exclusión educativa y participación en trabajo no remunerado, la problemática central radica en la combinación de baja escolaridad, inactividad laboral formal y restricciones económicas del hogar que impulsan su inserción temprana en actividades domésticas. Frente a ello, se propone la implementación de transferencias monetarias condicionadas a la asistencia escolar —como una versión reforzada del bono de desarrollo humano— complementadas con programas de alimentación escolar y seguimiento territorial, con el objetivo de reducir la deserción educativa.

Por su parte, el clúster 1, integrado por trabajadores adultos en condiciones de informalidad sin acceso a seguridad social, evidencia una inserción laboral activa pero estructuralmente precaria. En este contexto, se plantea como política concreta la creación de un esquema de formalización progresiva obligatoria en el empleo privado, sustentado en incentivos tributarios para empleadores que registren a sus trabajadores, junto con mecanismos de sanción efectivos ante la no afiliación.

En relación con el clúster 2, que representa el segmento formal del mercado laboral con acceso a seguridad social, las políticas deben orientarse no solo a preservar este grupo, sino a potenciar su capacidad de generación de valor. En este sentido, se recomienda impulsar estrategias de retención y escalamiento del empleo formal mediante incentivos a la productividad empresarial, tales como acceso a crédito productivo, programas de innovación y capacitación laboral certificada. Asimismo, es necesario fortalecer mecanismos de movilidad laboral ascendente que permitan evitar el estancamiento en posiciones de baja productividad dentro del sector formal, promoviendo trayectorias laborales más dinámicas y sostenibles.

Finalmente, el clúster 3, compuesto por adultos mayores en condiciones de autoempleo precario y sin cobertura institucional, refleja una problemática asociada a la ausencia de mecanismos de protección en etapas avanzadas del ciclo de vida. Frente a ello, se propone la creación de esquemas de seguridad social no contributiva o semi-contributiva adaptados a trabajadores independientes, complementados con pensiones mínimas focalizadas que garanticen un nivel básico de bienestar. De manera adicional, se recomienda el desarrollo de programas de reconversión productiva y apoyo a microemprendimientos, orientados a mejorar los ingresos en actividades de subsistencia y reducir la vulnerabilidad económica de este grupo.

5.3 La seguridad social como eje crítico de intervención estructural

La seguridad social emerge como el eje crítico de intervención estructural dentro del mercado laboral ecuatoriano, en tanto constituye la principal línea de diferenciación entre los segmentos identificados. El análisis evidencia que, más allá de las diferencias en ingreso o nivel educativo, el acceso a protección social define de manera decisiva las condiciones de estabilidad y bienestar de los trabajadores. En este sentido, mientras el clúster 2 presenta afiliación al IESS y, por tanto, acceso a mecanismos de protección frente a riesgos económicos y sociales, los demás grupos se encuentran excluidos del

sistema, lo que incrementa significativamente su vulnerabilidad ante eventos como enfermedad, pérdida de ingresos o envejecimiento.

Esta brecha institucional pone de manifiesto una falla estructural en el modelo de protección social vigente, el cual no logra adaptarse a la heterogeneidad del mercado laboral, caracterizado por altos niveles de informalidad, autoempleo y trayectorias laborales discontinuas. En particular, la exclusión de trabajadores informales, independientes y de aquellos en etapas tempranas del ciclo de vida limita la capacidad del sistema para cumplir su función redistributiva y de aseguramiento social.

En este contexto, resulta necesario avanzar hacia un modelo de seguridad social más flexible e inclusivo, capaz de responder a las dinámicas reales del empleo. Esto implica el diseño de esquemas de afiliación proporcional a los ingresos, que permitan la incorporación gradual de trabajadores con baja capacidad contributiva; la implementación de mecanismos simplificados de afiliación dirigidos a trabajadores informales y de pequeñas unidades productivas, reduciendo barreras administrativas y costos de acceso; y la ampliación de programas de cobertura no contributiva orientados a grupos vulnerables, particularmente en etapas de la vida donde la inserción laboral es limitada.

5.4 Limitaciones del estudio

El presente estudio presenta algunas limitaciones que deben ser consideradas al interpretar los resultados. En primer lugar, el análisis se basa en datos de corte transversal, lo que impide establecer relaciones causales entre las variables y limita la interpretación de los resultados a asociaciones observadas. En segundo lugar, la utilización de técnicas de aprendizaje no supervisado, como k-prototypes, implica que la conformación de los clústeres depende de la estructura de los datos y de las variables incluidas en el modelo, por lo que la exclusión de ciertas dimensiones —como variables familiares o territoriales— podría influir en los patrones identificados. Asimismo, algunas variables utilizadas provienen de información autodeclarada, lo que puede introducir sesgos de medición. Finalmente, si bien el enfoque de *clustering* permite capturar la heterogeneidad del mercado laboral, no sustituye el análisis econométrico tradicional, por lo que futuros estudios podrían complementar estos resultados mediante modelos causales o longitudinales que permitan profundizar en las dinámicas de la segmentación laboral.

Anexos

Anexo 1. Estadísticos descriptivos

La Tabla 7 presenta los estadísticos descriptivos de las principales variables numéricas utilizadas en el análisis. En promedio, la edad de los individuos es de 35.81 años, evidenciando una población predominantemente adulta. No obstante, la elevada dispersión y la presencia de valores mínimos muy bajos sugieren una alta heterogeneidad etaria.

En cuanto al nivel educativo, medido a través de los años aprobados, se observa un promedio de 4.51 años, lo que refleja una escolaridad relativamente baja en la población analizada. Este resultado sugiere limitaciones en el capital humano, con una distribución moderadamente dispersa entre individuos.

Por su parte, el ingreso per cápita presenta una media de 306.07 y una desviación estándar superior a la media, lo que evidencia una fuerte desigualdad en la distribución de ingresos. La presencia de valores máximos elevados confirma la existencia de atípicos y una distribución asimétrica, característica de economías en desarrollo.

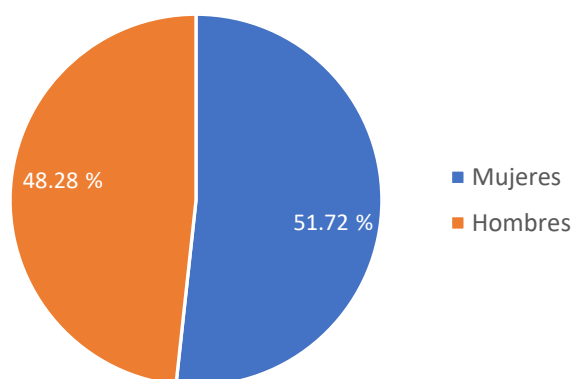
Tabla 7. Estadísticos descriptivos de las variables numéricas utilizadas en el análisis

Variable	Media	Desv. Std.	Mínimo	Máximo
Edad	35.81	21.89	4.0	76.0
Año aprobado	4.51	2.18	0.0	10.0
Ingreso por cápita	306.07	343.84	0.6	15800.0

Fuente: ENEMDU 2024, INEC

En la Figura 4 se puede observar que la distribución por sexo muestra una ligera predominancia femenina, con un 51.72% de mujeres frente a un 48.28% de hombres. Este resultado evidencia una composición equilibrada de la población analizada, lo que sugiere que no existen sesgos importantes en la representación por género. Desde una perspectiva analítica, esto permite evaluar diferencias en el mercado laboral sin distorsiones asociadas a una sobrerrepresentación de un grupo específico. Además, la ligera mayoría femenina puede ser relevante en contextos donde la participación laboral de las mujeres presenta dinámicas diferenciadas, particularmente en términos de informalidad y calidad del empleo.

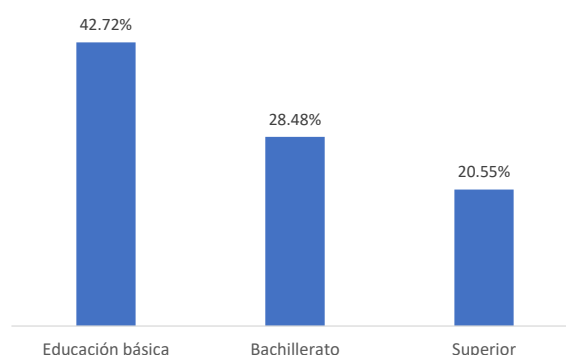
Figura 4. Sexo de las personas



Fuente: ENEMDU 2024, INEC

En cuanto al nivel educativo, el 42.72 % de los individuos posee educación básica, seguido del 28.48 % con educación media o bachillerato y el 20.55 % con educación superior. La concentración en niveles educativos intermedios refleja limitaciones estructurales en la acumulación de capital humano avanzado. A pesar de la presencia de individuos con educación superior, su proporción sigue siendo relativamente baja, lo que podría restringir la productividad laboral y el acceso a empleos formales de mayor calidad. Este resultado es particularmente relevante para el análisis, ya que sugiere que la educación, por sí sola, no garantiza una inserción laboral favorable, especialmente en contextos caracterizados por alta informalidad. Cabe señalar que se presentan únicamente las categorías con mayor frecuencia relativa, por lo que los porcentajes no suman el 100 % del total.

Figura 5. Nivel de instrucción

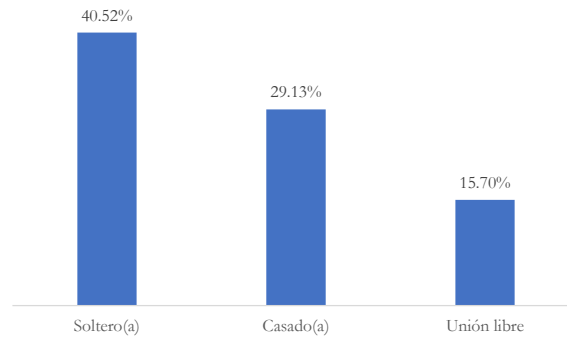


Fuente: ENEMDU 2024, INEC

Respecto al estado civil (Figura 6), se observa una mayor proporción de individuos solteros, seguida de personas casadas y en unión libre. Esta distribución sugiere una población relativamente joven o en etapas tempranas del ciclo de vida. Desde una perspectiva laboral, este perfil demográfico puede estar asociado a trayectorias ocupacionales más inestables, mayor rotación laboral y una mayor propensión al empleo informal. Asimismo, las responsabilidades familiares, que suelen ser menores en población soltera, pueden influir en la toma de decisiones laborales, especialmente en

términos de flexibilidad y tipo de empleo. Cabe señalar que se presentan únicamente las categorías con mayor frecuencia relativa, por lo que los porcentajes no suman el 100 % del total.

Figura 6. Estado civil

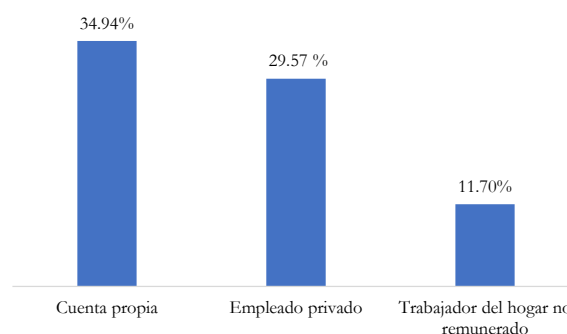


Fuente: ENEMDU 2024, INEC

La categoría ocupacional evidencia una participación importante de trabajadores por cuenta propia y empleados del sector privado (Figura 7). Este resultado refleja claramente la dualidad del mercado laboral ecuatoriano, donde coexisten formas de empleo formal e informal. La alta presencia de trabajadores por cuenta propia es un indicador clásico de informalidad estructural, lo que sugiere que una parte significativa de la población recurre al autoempleo como mecanismo de subsistencia ante la limitada generación de empleo formal. Este hallazgo es central para el análisis, ya que permite entender la segmentación del mercado laboral y las diferencias en calidad del empleo.

La magnitud de los trabajadores no remunerados dentro del hogar es un hallazgo particularmente relevante, ya que refleja formas de trabajo invisibilizadas dentro de las estadísticas tradicionales del mercado laboral. Este grupo suele estar asociado a actividades domésticas o de cuidado, realizadas principalmente por mujeres, lo que introduce una dimensión de desigualdad de género en la estructura ocupacional.

Figura 7. Categoría ocupacional

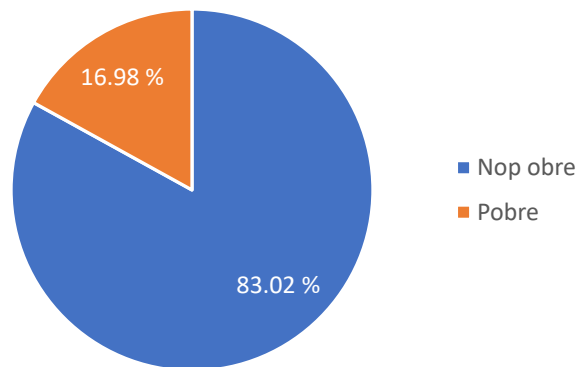


Fuente: ENEMDU 2024, INEC

En términos de pobreza, el 83.02 % de la población no se encuentra en condición de pobreza, mientras que el 16.98 % sí lo está. En la ENEMDU, la condición de pobreza se determina en función del ingreso per cápita del hogar en relación con la línea de pobreza

oficial definida por el INEC. Es decir, un individuo es considerado pobre si el ingreso de su hogar no alcanza el nivel mínimo necesario para cubrir una canasta básica de bienes y servicios. Bajo esta definición, aunque la mayoría de la población se ubica fuera de la pobreza, la proporción observada de individuos pobres sigue siendo relevante, lo que evidencia limitaciones en la capacidad del mercado laboral para generar ingresos suficientes de manera generalizada.

Figura 8. Pobreza



Fuente: ENEMDU 2024, INEC

Bibliografía

- Ablaza, C., Alladi, V., & Pape, U. (2023). *Indonesia's Informal Economy: Measurement, Evidence, and a Research Agenda* (Policy Research Working Paper No. 10608). World Bank. <https://doi.org/10.1596/1813-9450-10608>
- Akay, Ö., & Yüksel, G. (2017). Clustering the mixed panel dataset using Gower's distance and k-prototypes algorithms. *Communications in Statistics - Simulation and Computation*, 47(10), 3031-3041. <https://doi.org/10.1080/03610918.2017.1367806>
- Athey, S., & Imbens, G. W. (2019). Machine Learning Methods That Economists Should Know About. *Annual Review of Economics*, 11(1), 685-725. <https://doi.org/10.1146/annurev-economics-080217-053433>
- Banco Central del Ecuador [BCE]. (2024). *Cuentas nacionales trimestrales*. https://contenido.bce.fin.ec/documentos/informacioneconomica/SectorReal/ix_SectorRealPrin.html
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press.
- Becker, G. S. (1971). *The Economics of Discrimination*. <https://doi.org/10.7208/chicago/9780226041049.001.0001>
- Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, 94, S95-S120. <https://doi.org/10.1086/228943>
- Doeringer, P. B., & Piore, M. J. (2020). *Internal Labor Markets and Manpower Analysis*. Routledge. <https://doi.org/10.4324/9781003069720>
- Grané, A., & Sow-Barry, A. A. (2021). Visualizing Profiles of Large Datasets of Weighted and Mixed Data. *Mathematics*, 9(8), 891. <https://doi.org/10.3390/math9080891>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Instituto Nacional de Estadística y Censos. (2024). *Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU): Indicadores laborales, diciembre 2024* [Boletín Técnico]. INEC. https://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/2024/Diciembre_2024/202412_Mercado_Laboral.pdf
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2022). *An Introduction to Statistical Learning: with Applications in R* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
- Martin, C., & Okolo, M. (2022). *Heterogeneity in the UK Labour Market: Using Machine Learning To Test Macroeconomic Models* (Bath Economics Research Papers No. 93/22). University of Bath, Department of Economics.
- Modigliani, F., & Brumberg, R. (1954). Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data. En K. K. Kurihara (Ed.), *Post-Keynesian Economics* (pp. 388-436). Rutgers University Press. <https://archive.org/details/postkeynesianeco0000kuri>
- Ravenstein, E. G. (1885). The Laws of Migration. *Journal of the Statistical Society of London*, 48(2), 167. <https://doi.org/10.2307/2979181>
- Rodríguez-Guerrero, D. A., & Quintero, J. E. (2024). Are Labour Markets Segmented in Developing Economies? A Clustering Approach for Colombian Workers. *Ensayos de Economía*, 34(65), 69-93. <https://doi.org/10.15446/ede.v34n65.110808>
- World Bank. (2022). *Women, Business and the Law 2022*. World Bank. <https://doi.org/10.1596/978-1-4648-1817-2>