



## La productividad en el sector manufacturero ecuatoriano: una exploración espacial con *machine learning*

Juan González Pillalaza | Facultad Latinoamericana de Ciencias Sociales (Ecuador)

**RESUMEN:** Este estudio analiza la distribución espacial de la productividad manufacturera ecuatoriana en 2023 mediante *machine learning* no supervisado. Los algoritmos *K-means* y clusterización jerárquica identifican cuatro perfiles territoriales diferenciados según niveles de productividad laboral, especialización relativa y diversidad sectorial. Los resultados revelan una marcada heterogeneidad productiva, evidenciando que la relación entre aglomeración y desempeño productivo no responde a un patrón único, sino a configuraciones estructurales diversas. Se concluye que una política industrial basada en evidencias requiere diagnósticos territoriales diferenciados que permitan diseñar políticas acordes a las características de cada territorio.

**PALABRAS CLAVE:** Productividad manufacturera, economías de aglomeración, machine learning, política industrial, análisis espacial.

FECHA DE RECEPCIÓN 29/10/2025    FECHA DE REVISIÓN 6/03/2026    FECHA DE APROBACIÓN 12/3/2026

### Productivity in the Ecuadorian manufacturing sector: a spatial exploration with machine learning

**ABSTRACT:** This study analyzes the spatial distribution of Ecuadorian manufacturing productivity in 2023 using unsupervised machine learning. K-means and hierarchical clustering algorithms identify four distinct territorial profiles based on levels of labor productivity, relative specialization, and sectoral diversity. The results reveal marked productive heterogeneity, demonstrating that the relationship between agglomeration and productive performance does not follow a single pattern, but rather diverse structural configurations. The study concludes that an evidence-based industrial policy requires differentiated territorial analyses to design policies tailored to the characteristics of each region.

**KEYWORDS:** Manufacturing productivity, agglomeration economies, machine learning, industrial policy, spatial analysis.

JEL CODES R11, R12, C38, O54, D24

## INTRODUCCIÓN

El rol de Ecuador en la división internacional del trabajo, al igual que el de la mayoría de países de latinoamericanos, ha estado históricamente condicionado por la exportación de bienes primarios con escaso valor añadido y la importación de manufacturas y bienes de capital con alto valor agregado. Este patrón ha dado lugar a «procesos de industrialización tardíos y asincrónicos, lo que definió procesos de industrialización débiles o dependientes de los países ya industrializados» (Gonzales de Olarte, 2021, p. 254).

Esta trayectoria se materializa en el caso ecuatoriano a través de políticas industriales volátiles. Históricamente, el sector manufacturero transitó desde un enfoque de industrialización cepalina durante los años 1950 y 1970, pasando por el desmantelamiento del modelo de sustitución de importaciones entre 1982-2007, una política industrial neodesarrollista que prevaleció hasta 2016, para luego dar paso a un modelo industrial orientado hacia una economía más de mercado (Creamer, 2022; Acosta 2012). Este proceso ha estado marcado tanto por factores externos —volatilidad petrolera, crisis de deuda, y el impacto de la pandemia de covid-19— como por disyuntivas internas que han influido decisivamente al sector.

El legado de esta trayectoria histórica es un sector con un peso económico limitado y en decrecimiento. De acuerdo a información del Banco Central del Ecuador (BCE), entre 2020 y 2023, el valor agregado bruto (VAB) manufacturero promedió 14.000 millones de dólares anuales, equivalentes al 13,03% del PIB, un valor considerablemente inferior a los registros históricos de 16,70% en 1972 o el 19,72% en 1981 (Creamer, 2022). Ello ha evidenciado un proceso de desindustrialización relativa. Según el Ministerio de Producción, Comercio Exterior, Inversiones y Pesca (MPCEIP, 2024), en 2023 las ventas netas manufactureras alcanzaron los 35.361 millones de dólares, registrando una variación negativa de 0,7% interanual. En el ámbito laboral, la manufactura concentró apenas el 8,5% del empleo total en relación con la población económicamente activa, una cifra inferior al promedio latinoamericano, que lo ubicó en 11,1% en 2023 (Cepal, 2024; MPCEIP, 2023).

Este preámbulo pone en manifiesto un proceso de desindustrialización que conlleva la necesidad de un análisis renovado del sector manufacturero y de sus posibles estrategias de políticas públicas desde una perspectiva regional y contemporánea. La productividad, entendida como la relación entre insumos y producto total, se considera un excelente indicador de desempeño económico para alcanzar competitividad a largo plazo (Ruiz, 2013). Sin embargo, diversos estudios regionales postulan que la productividad no solo está determinada por atributos internos de las empresas, sino también por factores geográficos y espaciales que configuran condiciones locales de producción (Guevara, 2021; Amara y Thabet, 2019; Dvouletý y Blažková, 2021). Ante este panorama, ¿existen patrones espaciales a nivel cantonal que agrupen a las empresas manufactureras ecuatorianas según su productividad, y cómo se relacionan estos clústeres con el nivel de especialización y diversidad económica de los cantones? Responder este interrogante implica superar los análisis agregados para explorar la heterogeneidad territorial subnacional.

En este contexto, la presente investigación se propone analizar la distribución espacial de la productividad del sector manufacturero y generar una clasificación

cantonal en función de sus niveles de productividad, especialización y diversidad, mediante el uso de técnicas de *machine learning* (ML). Este enfoque metodológico, con un eje transversal en la política pública basada en evidencias (PPBE), busca generar diagnósticos empíricos que orienten políticas manufactureras y regionales diferenciadas. Al integrar la teoría de las economías de aglomeración con herramientas de ciencia de datos, la investigación aporta una caracterización empírica de los patrones territoriales de productividad manufacturera, que puede servir como insumo para el diseño de estrategias de competitividad territorial y una planificación pública sustentada en información verificable.

El artículo se estructura de la siguiente manera: tras esta introducción, la sección 2 presenta el marco teórico. La sección 3 detalla las fuentes de datos y la construcción de las variables. La sección 4 describe la estrategia metodológica, con especial atención en los modelos de aprendizaje no supervisado de ML. La sección 5 expone los resultados, mientras que la sección 6 discute dichos hallazgos y analiza sus implicaciones para la política pública. Finalmente, el artículo cierra con conclusiones y reflexiones en la sección 7.

## REVISIÓN DE LITERATURA

El análisis de la productividad manufacturera a nivel regional requiere un enfoque conceptual que trascienda los determinantes internos de la empresa e incorpore el papel del espacio geográfico. Este marco teórico se sustenta en la teoría de las economías de aglomeración, la cual explica las externalidades positivas que reciben empresas y personas al ubicarse en una determinada área geográfica (Glaeser *et al.*, 1992). La ubicación geográfica tiende a condicionar la productividad empresarial debido a diversos factores estructurales y contextuales que afectan costos operativos, oportunidades de innovación y expansión (Dvouletý y Blažková, 2021). Establecerse en regiones con adecuada infraestructura, cercanía a mercados clave o presencia de clústeres industriales permite reducir tiempos y costos logísticos, acceder a insumos, mano de obra calificada, y beneficiarse del aprendizaje colectivo y la transferencia de conocimiento que surgen de la proximidad entre empresas e instituciones (Martínez-Victoria *et al.*, 2017). Asimismo, los entornos urbanos o metropolitanos suelen ofrecer mejores servicios tecnológicos, financieros y administrativos, lo que favorece procesos de modernización productiva. La proximidad geográfica entre empresas genera externalidades positivas a través de intereses comunes e intercambios de información, que generan economías de escala externas, menores costes de transporte y transferencia de información, lo que fortalece la capacidad competitiva y por ende la productividad de las empresas.

En la literatura y en muchos estudios empíricos (Lin *et al.* 2011; Tao *et al.*, 2019; Singh, 2021), estas externalidades se dividen en dos tipos principales: economías de localización y de urbanización. La primera se refiere a la concentración de empresas de un solo tipo de industrias y han sido formalizadas en el enfoque conocido como MAR — por los estudios precursores de Marshall (1820), Arrow (1962) y Romer (1986)—. Estas economías se basan en la existencia de proveedores especializados, mercado de trabajo especializado y la presencia de externalidades del conocimiento,

factores denominados externalidades marshallianas (Lavoratori y Castellani, 2021). Por otro lado, la concentración de empresas pertenecientes a diferentes industrias en un mismo territorio se las denomina economías de urbanización. A partir del trabajo pionero de Jacobs (1969), una mayor diversidad empresarial, genera condiciones territoriales más dinámicas y propensas a la innovación, ya que la difusión de conocimiento es más favorable entre industrias que dentro de ellas (Sanfilippo y Seric, 2015; Cieřlik *et al.*, 2018). Sin embargo, la aglomeración puede generar externalidades negativas, como congestión, encarecimiento del suelo, aumento de los costos productivos y una competencia más intensa que puede desplazar a empresas menos competitivas y desincentivar nuevas localizaciones en regiones altamente concentradas (Beaudry y Schiffauerova, 2009).

En el caso ecuatoriano, Guevara *et al.* (2019) evidencian que la diversidad sectorial impulsa la productividad del sector servicios, mientras que Torres-Gutiérrez *et al.* (2019) y Guevara (2021) encuentran que, en la manufactura, la especialización es el principal factor explicativo que incide en la productividad, revelando así un comportamiento opuesto entre sectores. En consecuencia, esta heterogeneidad sectorial, a menudo contradictoria, subraya una limitación fundamental de los análisis agregados a nivel nacional: ocultan la heterogeneidad subnacional. Las políticas industriales tradicionales, basadas en promedios nacionales o en supuestos generales sobre las ventajas de la especialización o la diversidad sectorial, corren el riesgo de ser ineficaces o incluso contraproducentes cuando se aplican de manera uniforme en territorios con estructuras productivas, dotaciones de recursos y capacidades institucionales diferentes (Balland *et al.*, 2019). Esta crítica ha impulsado un cambio de paradigma hacia las PPBE, la cual exige diagnósticos precisos y específicos para cada contexto territorial.

En este contexto, la teoría de las economías de aglomeración proporciona el sustento conceptual para entender las fuentes regionales de la productividad, pero son las PPBE las que operan el marco de acción, requiriendo evidencia empírica robusta para su implementación. Sin embargo, la complejidad económica donde las disparidades espaciales y la productividad son el resultado de interacciones no lineales entre múltiples factores (Balland *et al.*, 2022) desafía los métodos econométricos tradicionales (Kopczewska, 2022). Es aquí donde el ML, y particularmente las técnicas como la clusterización se convierten en herramientas valiosas para identificar patrones espaciales ocultos. Singh (2021), por ejemplo, proporciona un análisis espacial detallado de los clústeres industriales en Hungría, revelando relaciones intrínsecas entre aglomeración, productividad laboral y urbanización, a su vez ofrece perspectivas concretas para un desarrollo industrial regional equilibrado. De forma similar, Cui *et al.* (2024) proponen un modelo para medir la eficiencia del uso del suelo industrial en China mediante técnicas de *clustering* y econometría espacial, destacando la influencia de la densidad manufacturera en la productividad regional. Por su parte, Balland *et al.* (2022) demuestran cómo los patrones de complejidad económica y especialización inteligente pueden analizarse con herramientas de ML para diseñar estrategias territoriales basadas en las capacidades tecnológicas locales.

En Ecuador, aunque escasos, algunos estudios empiezan a aplicar metodologías similares. Matute y Muñoz (2024) emplean aprendizaje supervisado para clasificar pymes ecuatorianas según sus probabilidades de crecimiento, mientras que Rodríguez-Cruz (2024) realiza una clusterización de empresas manufactureras en función

de indicadores financieros. Si bien estos estudios constituyen un valioso antecedente, su enfoque se circunscribe predominantemente al nivel microempresarial, dejando de lado la dimensión regional. Frente a esta brecha, la presente investigación pretende aportar a la literatura con una clasificación cantonal a partir de variables de aglomeración y productividad, empleando el ML no como un fin en sí mismo, sino como una herramienta de diagnóstico territorial. El objetivo último es proveer evidencia robusta que sustente el diseño de políticas industriales diferenciadas y espacialmente contextualizadas en Ecuador.

## DATOS Y VARIABLES

Se utilizan los últimos datos del Registro Estadístico de Empresas (REEM) 2023 elaborado por el Instituto Nacional de Estadísticas y Censos (INEC). Esta base integra información procedente de las declaraciones tributarias anuales del Servicio de Rentas Internas (SRI), registros del Instituto Ecuatoriano de Seguridad Social (IESS), así como información de la Superintendencia de Compañías, la Superintendencia de Bancos y las encuestas empresariales realizados por el INEC. Su objetivo es cuantificar el comportamiento empresarial ecuatoriano en términos económicos, geográficos y empleo en empresas públicas y privadas formalmente constituidas.

La unidad de análisis es el par cantón-subsector a dos dígitos de la Clasificación Industrial Internacional Uniforme (CIIU). Este enfoque captura la heterogeneidad intra-cantonal al reconocer que un territorio puede albergar subsectores manufactureros con diferentes niveles de productividad. Se construyen variables que caracterizan las diferencias estructurales entre territorios. La productividad laboral se calcula a nivel de empresas como el logaritmo natural del cociente entre ventas totales y número de empleados, aproximando el valor agregado por trabajador. Esta medida se agrega a nivel cantonal-subsector, obteniendo la productividad media para cada combinación cantonal-sectorial, constituyendo la variable principal del análisis.

En cuanto a los cuantificadores de economías de aglomeración, estos se estiman a dos niveles complementarios. El primero corresponde al índice de especialización relativa, calculado mediante un Location Quotient (LQ) a nivel cantón-subsector manufacturero. Esta variable mide la concentración del empleo de un subsector en un cantón en relación con su participación a nivel nacional, permitiendo cuantificar el grado de economías de localización (Nakamura y Morrison, 2019). Se expresa como:

$$LQ_{i,r} = \frac{emp_{i,r}/emp_{i*}}{emp_{*r}/emp_{**}} \quad (1)$$

donde  $emp_{i,r}$  representa el número de empleados en el subsector  $i$  del cantón  $r$ ;  $emp_{i*}$  el total de empleados a nivel nacional del subsector  $i$ ;  $emp_{*r}$  el empleo manufacturero total en el cantón  $r$ ; y  $emp_{**}$  el empleo total de la economía en general. Un valor LQ mayor a uno, señala que la industria  $i$  está relativamente concentrada en el cantón  $r$ , evidenciando especialización (Zheng, 2013). Según Fracasso y Vittucci Marzetti (2018)

este indicador es el más utilizado en estudios empíricos de economías de aglomeración, ya que permite identificar aquellos cantones donde la actividad de un subsector supera lo esperado según la estructura nacional, sugiriendo ventajas comparativas locales y externalidades de especialización. A diferencia de medidas de concentración del empleo, el LQ incorpora una normalización respecto a la estructura productiva nacional, facilitando comparaciones entre territorios de distinta escala económica y evita que los resultados estén influenciados por el tamaño del cantón.

El segundo nivel corresponde al análisis de las economías de urbanización, medido a nivel cantonal. Para ello, se utiliza la inversa del índice de Herfindahl o denominada también índice de diversidad. Esta expresión formalizada matemáticamente por Durantón y Puga (2000) se expresada como:

$$div_r = \left[ \sum_i \left( \frac{emp_{i,r}}{emp_r} \right)^2 \right]^{-1} \quad (2)$$

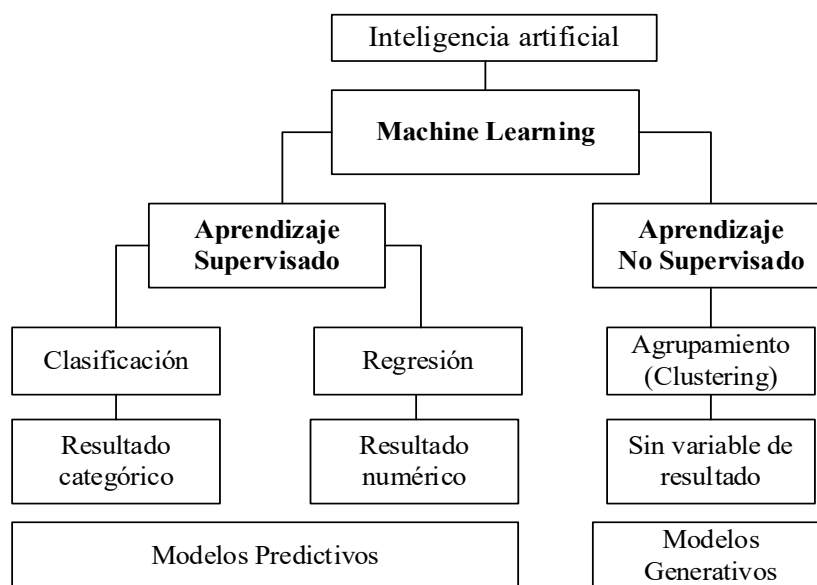
donde la sumatoria se realiza sobre todas las industrias manufactureras  $i$  en la región  $r$ , permitiendo agregar las contribuciones de cada sector al índice de diversidad. Un valor más alto de este indicador sugiere una mayor diversidad productiva y una distribución más equitativa del empleo entre varias industrias (Combes y Durantón 2015).

Estas variables —productividad laboral, especialización relativa y diversidad— fueron sometidas a un proceso de estandarización (z-scores) con el fin de homogeneizar sus escalas y evitar diferencias que puedan influir en los resultados de clusterización. Una vez estandarizadas estas variables, constituyen el insumo principal para la aplicación de algoritmos de ML, que permitan agrupar a las observaciones en clústeres homogéneos según sus niveles de productividad, especialización y diversidad económica. Cabe recalcar que la investigación se constituye sobre una unidad combinada cantón-subsector manufacturero, lo que permite capturar tanto las diferencias territoriales como las heterogeneidades sectoriales dentro de cada cantón. En total, se analizan cerca de 1242 observaciones, correspondientes a la articulación de 192 cantones del país con 23 subsectores manufactureros distintos.<sup>1</sup>

## METODOLOGÍA

El ML, por su traducción: *aprendizaje automático*, constituye la rama de la inteligencia artificial (IA) cuyo objetivo es la predicción estadística (Cerulli, 2023). Esta herramienta permite el desarrollo de técnicas computacionales capaces de identificar patrones, optimizar procesos y descubrir estructuras en función a datos o experiencias previas (Alpaydin, 2020; Zollanvari, 2023). A continuación, se presenta un esquema conceptual que sintetiza la estructura del ML dentro de la IA, así como sus los principales enfoques metodológicos y tipos de modelos (ver Figura 1).

**Figura 1.** Clasificación estándar de *machine learning* (ml)



Fuente: Cerulli (2023, p. 3).

El ML se divide esencialmente en dos: aprendizaje supervisado y no supervisado. El primero estima valores o categoría de salida a partir de variables previamente etiquetadas, y comprende principalmente dos tipos de modelos: regresión cuando la variable objetivo es numérica, y clasificación cuando esta corresponde a una categoría discreta (Cerulli, 2023; Müller y Guido, 2017). Por su parte, el aprendizaje no supervisado trabaja con datos no etiquetados y busca descubrir estructuras subyacentes o patrones latentes mediante la generación de nuevas representaciones o variables a partir de los propios datos (Cerulli, 2023; Zollanvari, 2023). Dentro de este enfoque se ubican los algoritmos de agrupamiento o análisis de conglomerados —clúster analysis—, los cuales permiten identificar grupos homogéneos de observaciones en función de su similitud multivariada, constituyendo así una herramienta para la exploración estadística y la caracterización regional. En este contexto, la clusterización consiste en agrupar un conjunto de datos en categorías denominadas clústers, de tal manera que las observaciones dentro de un mismo grupo sean muy similares entre sí, mientras que aquellas pertenecientes a distintos clústeres sean notablemente diferentes (Müller y Guido, 2017). Según Kopczevska (2022), el ML supervisado tiende a reemplazar progresivamente a la econometría espacial clásica, mientras que el aprendizaje no supervisado se orienta a identificar cómo los datos tienden a agruparse de manera natural. Este último enfoque constituye la base metodológica de la presente investigación. En particular, la técnica de clusterización resulta útil para descubrir patrones latentes en los datos y clasificar las unidades territoriales según sus características estructurales y su comportamiento relativo dentro del sistema analizado.

Entre los diversos métodos de clusterización, el presente estudio emplea dos algoritmos: *K-means* y clusterización jerárquica. El primero, también conocido como algoritmo de Lloyd, es uno de los más utilizados debido a su fácil implementación

y a su eficiencia computacional en comparación a otros métodos (Raschka *et al.*, 2022). Su objetivo consiste en identificar grupos de observaciones con alta similitud interna y clara diferenciación externa, utilizando la distancia euclidiana al cuadrado como medida de proximidad. El algoritmo busca minimizar la suma de los errores cuadráticos dentro de los clústeres (SSE), también denominada inercia, de modo que cada grupo esté representado por un centroide que resume sus características promedio. Matemáticamente, la función inercia se define como:

$$SSE = \sum_{i=1}^n \sum_{j=1}^k w_{ij} \|x_i - \mu_j\|^2 \quad (3)$$

donde  $x_i$  representa una observación  $i$ ,  $\mu_j$  el centroide del clúster  $j$ , y  $w_{ij}$  toma valor 1 si la observación  $i$  pertenece al clúster  $j$  o en caso contrario valor 0.

El procedimiento es iterativo y consta de cuatro pasos: (i) selecciona aleatoriamente  $k$  centroides iniciales, (ii) asignar cada observación al centroide más cercano, (iii) recalculan los centroides como el promedio de los elementos asignados, y (iv) repetir los pasos 2 y 3 hasta alcanzar la convergencia (Zollanvari, 2023). Aunque el método es computacionalmente eficiente y ampliamente aplicado, requiere definir previamente el número de clústeres. Para ello, se utiliza el método del codo, que evalúa la relación entre el número de clústeres y la varianza intra grupal. El punto óptimo se identifica donde la reducción marginal de la inercia se estabiliza—formando un «codo» en la gráfica—, lo cual indica el equilibrio entre homogeneidad interna y parsimonia del modelo. En este estudio, dicho criterio se emplea para determinar el número adecuado de conglomerados cantonales según los niveles de productividad, especialización y diversidad económica, garantizando una clasificación empíricamente robusta y territorialmente coherente.

No obstante, el método *K-means* presenta una limitación importante: requiere definir de antemano el número de clústeres ( $k$ ) que estructurará la partición de los datos. Para superar este inconveniente y validar la robustez de los resultados, la investigación incorpora un segundo enfoque: la clusterización jerárquica, la cual permite construir una descomposición progresiva del conjunto de observaciones y determinar de manera más flexible el número óptimo de agrupamientos (Horobet *et al.*, 2021). Su principio consiste en evaluar las similitudes entre las observaciones y fusionarlas o dividir las sucesivamente según su proximidad, medida a través de una función de distancia o similitud (Alpaydin, 2020). Para esta investigación se utilizó la distancia euclidiana, la cual cuantifica la proximidad entre dos puntos  $x$  y  $y$  en un espacio  $n$ -dimensional. Su expresión matemática es:

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

donde  $x = (x_1, x_2, \dots, x_n)$  y  $y = (y_1, y_2, \dots, y_n)$  representan dos observaciones en el espacio euclidiano de  $n$  dimensiones. Esta medida evalúa el grado de similitud entre instancias, siendo los valores menores indicativos de una mayor proximidad entre las observaciones.

En este estudio, se aplicó el método de Ward (1963), una técnica de enlace jerárquico que fusiona iterativamente los grupos que provocan el menor incremento en la varianza interna o suma de errores cuadrados (SSE), produciendo agrupamientos compactos y homogéneos. Este procedimiento busca minimizar la varianza total dentro de los grupos tras cada fusión, según la expresión:

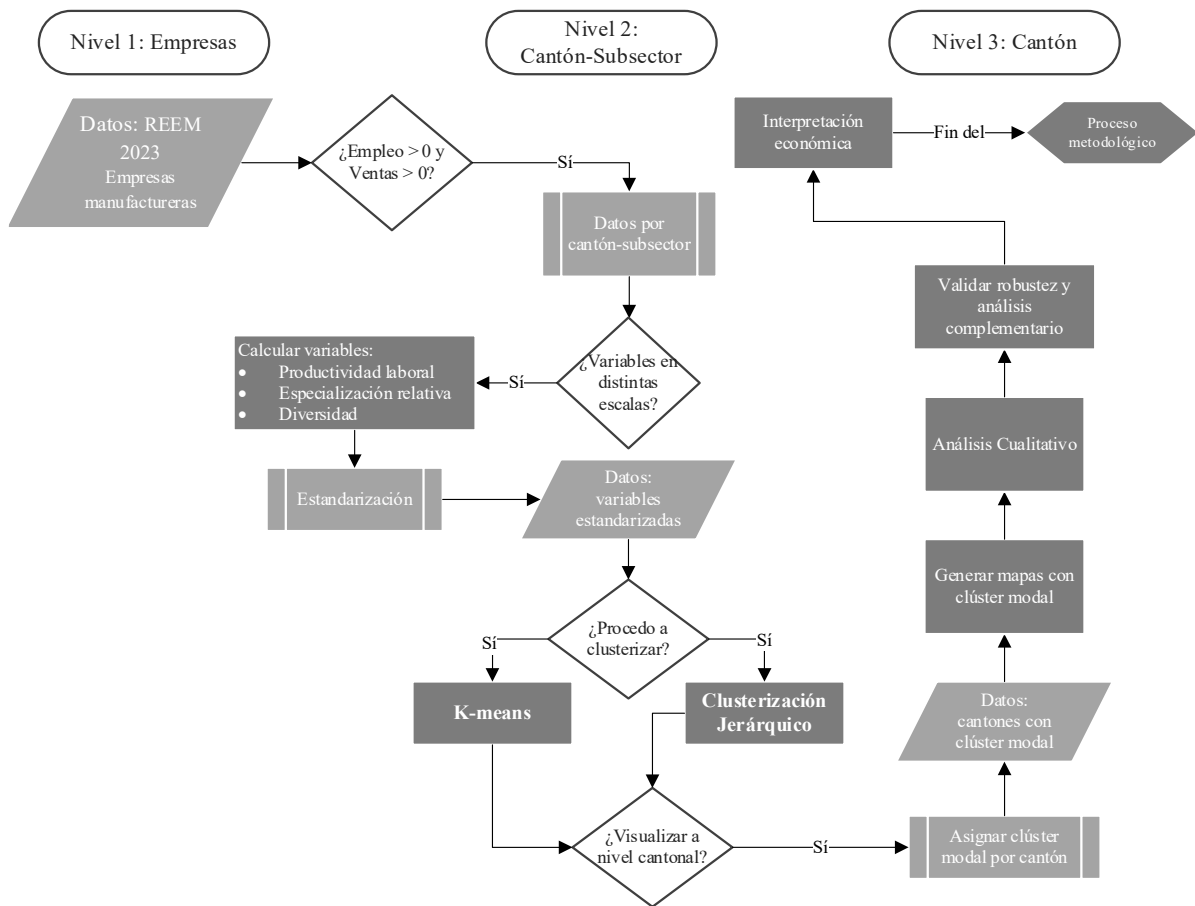
$$d(A, B) = \frac{|C| + |B|}{|C| + |B| + |D|} d(C, B) + \frac{|D| + |B|}{|C| + |B| + |D|} d(D, B) - \frac{|B|}{|C| + |B| + |D|} d(C, D) \quad (5)$$

Donde  $|A|$ ,  $|B|$ ,  $|C|$  y  $|D|$  representan el número de elementos en cada grupo. Este criterio permite combinar iterativamente las unidades más similares en términos de varianza, optimizando la cohesión interna del agrupamiento (Horobet *et al.*, 2021; Jagódka, 2025).

El resultado del proceso se representa mediante un dendrograma, una estructura en forma de árbol que ilustra las relaciones jerárquicas entre las observaciones y la secuencia en la que se agrupan. Esta representación permite visualizar las uniones progresivas entre los elementos, y a su vez, identificar de manera empírica el número de clústeres más adecuado para los datos (Raschka *et al.*, 2022). Para determinar dicho número óptimo de clústeres, se combinan dos criterios complementarios: el análisis de distancias de fusión y la inconsistencia jerárquica, que permiten identificar el punto de corte más estable en el dendrograma. Finalmente, la calidad y consistencia del modelo se evaluaron mediante los índices de Silhouette y Calinski-Harabasz.

En este sentido, la clusterización jerárquica complementa el análisis con *K-means*, ofreciendo una perspectiva más exploratoria sobre la estructura espacial y productiva de los cantones y subsectores manufactureros en Ecuador. El proceso metodológico completo —desde el tratamiento de microdatos empresariales, agregación a nivel cantón-subsector, cálculo y estandarización de variables, aplicación de algoritmos de clusterización, asignación por moda a nivel cantonal para representación cartográfica, validaciones de robustez e interpretación económica— se sintetiza en el Diagrama de Flujo Metodológico (ver Figura 2). Este esquema, permite comprender las decisiones analíticas adoptadas en cada etapa y la articulación entre las diferentes fases del estudio.

**Figura 2.** Diagrama de flujo metodológico

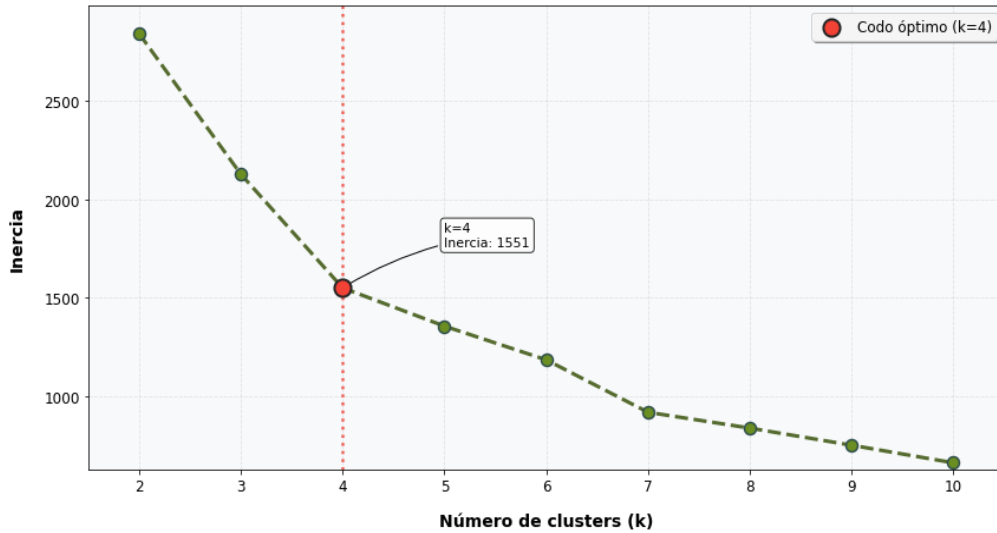


## RESULTADOS

En primer lugar, dentro del nivel de observación cantón-subsector, se aplicó el algoritmo *K-means* para identificar agrupamientos con comportamientos productivos similares. La elección del número óptimo de clústers se lo realizó mediante un análisis de sensibilidad entre el rango  $k = 2$  a  $k = 10$  utilizando dos métricas complementarias: el método del codo (basado en la inercia) y dos métricas de validación: Silhouette y Calinski-Harabasz, ampliamente utilizados en la literatura (Kirilyuk y Senko, 2020; Wang y Xu, 2019). El método del codo (ver Figura 3) evidenció un punto de inflexión en  $k = 4$ , a partir del cual la reducción de la varianza intraclúster se estabiliza. A su vez, esta solución registró un Silhouette Score de 0.346 y un Calinski-Harabasz de 578.6 (ver Figura 4). Si el primero es superior a 0.30 indica una estructura de agrupamiento razonablemente definida, mientras que el segundo confirma que la partición en cuatro clústers maximiza la relación entre dispersión intergrupo y cohesión intragrupo. Adicionalmente, se evaluó la estabilidad de la solución frente a distintas inicializaciones aleatorias del algoritmo. Utilizando siete semillas diferentes, el Índice Rand Ajustado (ARI) promedio entre las particiones obtenidas fue de 0.9368 ( $\pm 0.0549$ ), lo que demuestra una alta estabilidad del

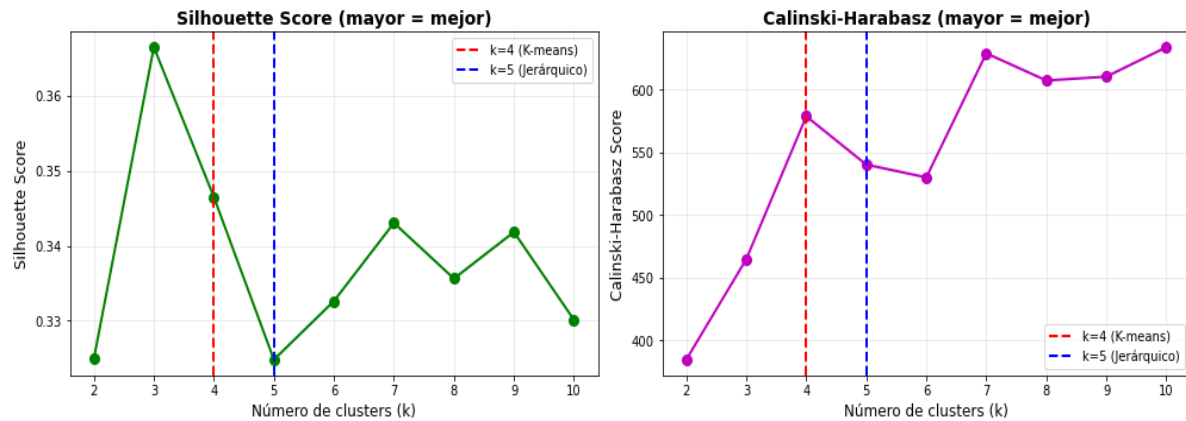
algoritmo y descarta que los resultados dependan de una configuración inicial particular. En conjunto, estos resultados respaldan la elección de  $k$  como una solución moderada y estadísticamente consistente, adecuada para capturar la heterogeneidad territorial sin sobredimensionar el modelo.

**Figura 3.** Método del Codo para determinar  $k$  óptimo



Nota: Elaborado en Python 3.12.4. Fuente: REEM (2023).

**Figura 4.** Validación de clusterización, métricas para rango de clústers K

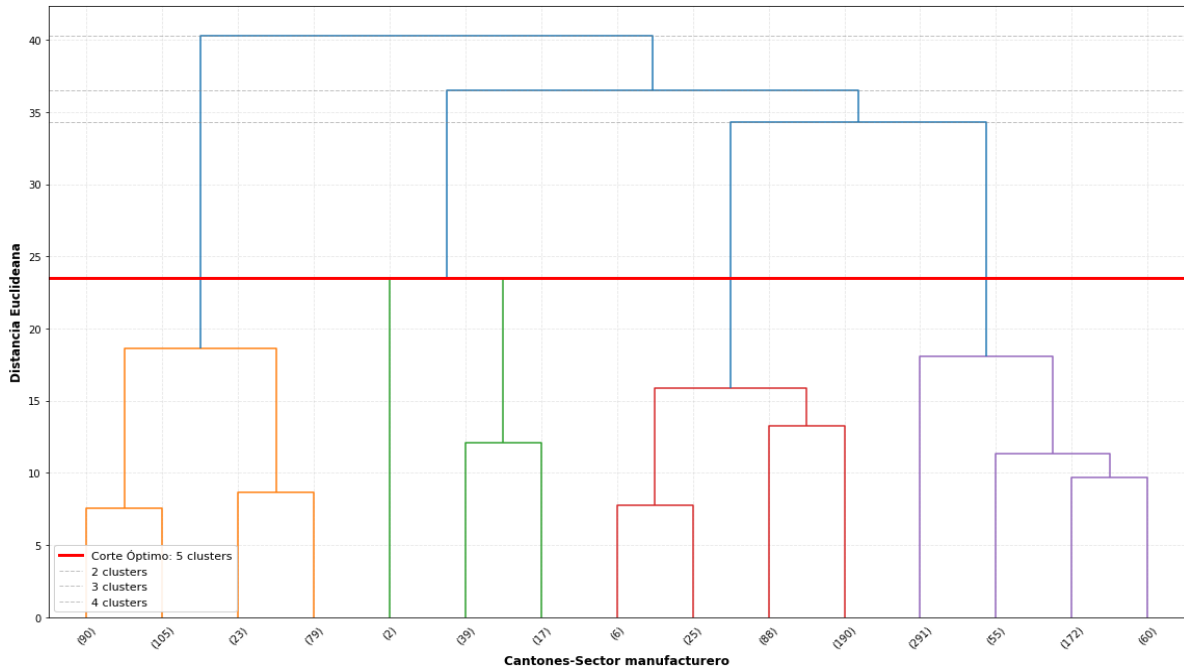


Nota: Elaborado en Python 3.12.4. Fuente: REEM (2023).

Para contrastar la robustez de *K-means*, se aplicó un análisis de clusterización jerárquica con enlace de Ward. El dendrograma resultante (ver Figura 5) fue evaluado a partir de dos criterios formales: el análisis de las distancias de fusión<sup>2</sup> (que sugiere un corte en  $k = 4$ ) y el estadístico de inconsistencia jerárquico<sup>3</sup> (que identifica una estructura binaria  $k = 2$ ). Considerando la magnitud de los saltos en las distancias de fusión y la morfología del dendrograma, se optó por una solución de cinco clústers como último nivel estructuralmente estable antes de la consolidación de grupos

internamente diferenciados. Esta decisión permite preservar la heterogeneidad sustantiva del entramado manufacturero sin incurrir en fragmentación excesiva. Como validación complementaria, esta solución registró un Silhouette Score de 0.331 y un Calinski-Harabasz de 523.6 (ver Figura 4), valores consistentes con una separación moderada pero interpretable entre grupos.

**Figura 5.** Dendrograma clusterización jerárquica



Nota: Elaborado en Python 3.12.4. Fuente: REEM (2023).

Para facilitar la interpretación, se elaboraron dos tipos de clasificaciones. La primera presenta los estadísticos descriptivos de cada clúster calculados a partir de las 1242 observaciones (ver Tabla 1). Complementariamente, se realizó una clasificación cualitativa (ver Tabla 2) basada en la posición relativa de sus centroides. Dado que las variables se encuentran estandarizadas, los centroides representan el perfil promedio de cada grupo en el espacio multivariante. Sobre este último se aplicó un criterio de terciles: aquellos clústeres cuyos centroides se ubicaron en el tercio superior de cada dimensión fueron calificados como alto; en el tercio inferior, como bajo; y en el intervalo intermedio, como medio. Ello permite comparar el desempeño de los clústeres entre sí, evitando el uso de umbrales arbitrarios y reconociendo que la significancia de los valores estandarizados depende del contexto del sistema analizado. Complementariamente, se identificó el subsector manufacturero predominante en cada clúster como aquel con mayor frecuencia relativa de observaciones.

**Tabla 1.** Estadísticos descriptivos con valores estandarizados

Tipo	Número clústeres	Número cantones	Observaciones	Productividad laboral	Especialización	Diversidad sectorial
K-means	0	83	431	- 0.936 (0.757)	- 0.148 (0.393)	- 0.343 (0.542)
	1	20	274	0.276 (0.604)	0.163 (0.296)	1.513 (0.847)
	2	18	48	- 0.229 (1.243)	3.816 (2.701)	- 0.603 (0.569)
	3	71	489	0.693 (0.619)	- 0.153 (0.409)	- 0.486 (0.438)
Jerárquico	0	23	297	0.320 (0.535)	- 0.166 (0.288)	1.397 (0.894)
	1	1	2	- 0.760 (0.693)	14.834 (4.381)	- 0.798 (0.356)
	2	24	56	0.084 (0.839)	2.922 (1.375)	- 0.522 (0.531)
	3	51	309	- 1.209 (0.793)	- 0.107 (0.533)	- 0.262 (0.643)
	4	93	578	0.476 (0.732)	- 0.192 (0.311)	- 0.525 (0.387)

Nota: Los valores corresponden a sus promedios, entre paréntesis la desviación estándar. Fuente: REEM (2023).

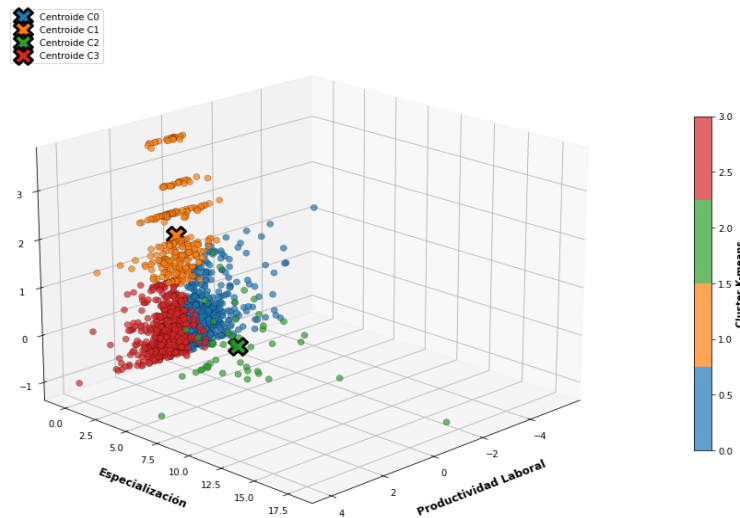
**Tabla 2.** Clasificación cualitativa de los clústeres según el algoritmo de clusterización

Tipo	Número clústers	Productividad laboral	Especialización	Diversidad sectorial	Subsector principal
K-means	0	Bajo	Alto	Alto	C14: Fabricación de prendas de vestir
	1	Alto	Bajo	Alto	C33: Reparación e instalación de maquinaria y equipo
	2	Medio	Alto	Bajo	C11: Elaboración de bebidas
	3	Alto	Medio	Medio	C10: Elaboración de productos alimenticios
Jerárquico	0	Medio	Bajo	Alto	C25: Fabricación de productos elaborados de metal
	1	Bajo	Alto	Bajo	C32: Otras industrias manufactureras
	2	Medio	Alto	Medio	C16: Industria de la madera y corcho
	3	Bajo	Medio	Alto	C14: Fabricación de prendas de vestir
	4	Medio	Bajo	Bajo	C10: Elaboración de productos alimenticios

Fuente: REEM (2023).

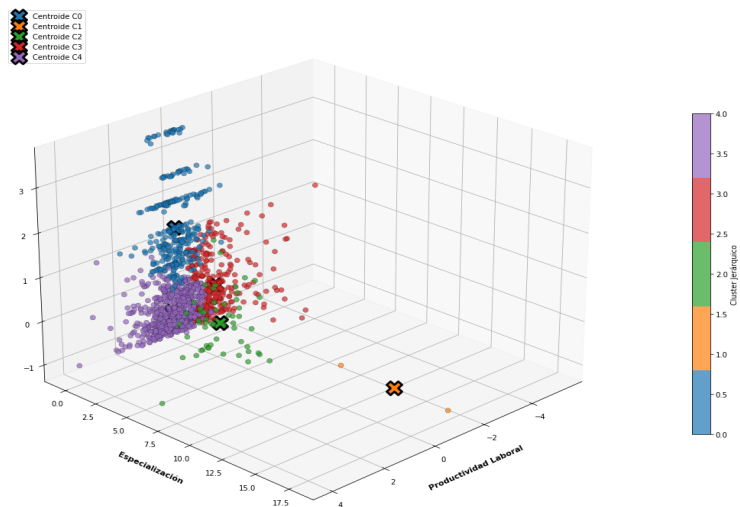
En base a lo anterior, la representación tridimensional del *K-means* (ver Figura 6), permite apreciar la conformación de los cuatro clústers resultantes; del mismo modo que lo hace la proyección tridimensional para la clusterización jerárquica (ver Figura 7). Para evaluar la concordancia entre ambos algoritmos, se empleó el ARI,<sup>4</sup> una medida ampliamente utilizada en la literatura para comparar clústers provenientes de diferentes algoritmos (Albuquerque *et al.*, 2016; Warrens y van der Hoef, 2022). El valor obtenido de 0.582 indica una concordancia moderada pero sustancial entre ambas soluciones. Este resultado, lejos de sugerir discrepancia, confirma que los cuatro clústeres de *K-means* constituyen una desagregación coherente de la estructura más estable identificada en el análisis jerárquico.

**Figura 6.** Visualización 3D de clústeres K-means: productividad laboral, especialización relativa y diversidad sectorial



Nota: Elaborado en Python 3.12.4. Fuente: REEM (2023).

**Figura 7.** Visualización 3D de clústeres jerárquicos: productividad laboral, especialización relativa y diversidad sectorial

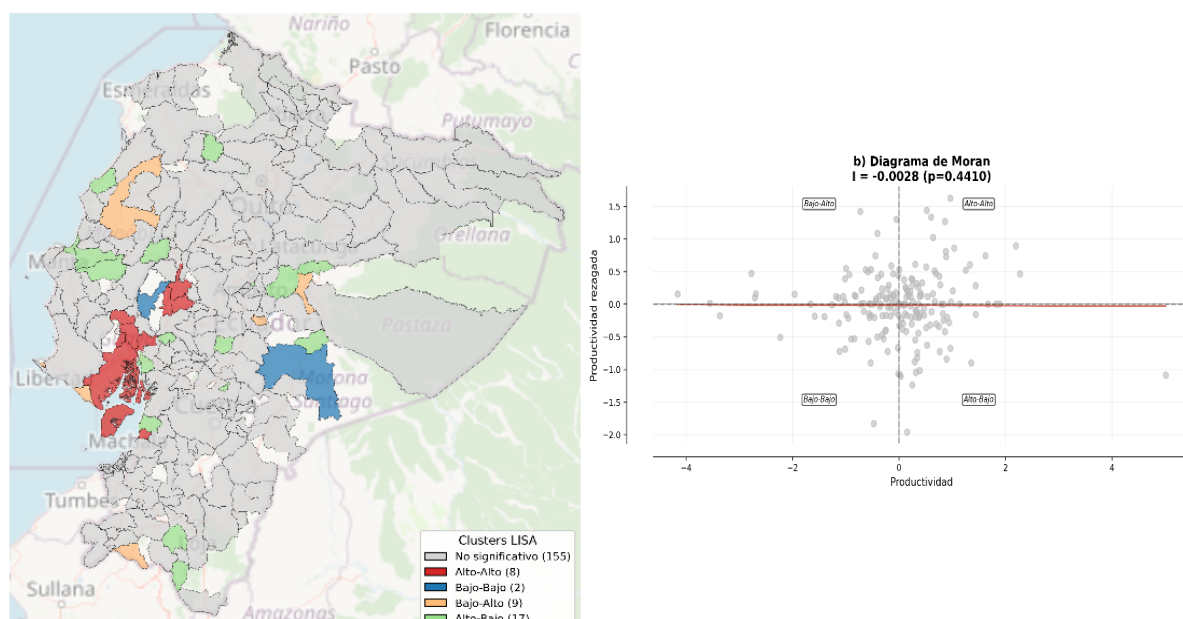


Nota: Elaborado en Python 3.12.4. Fuente: REEM (2023).

Una vez validada la consistencia entre los algoritmos de clusterización a nivel cantón-subsector, se generó una clasificación cantonal con fines de representación cartográfica. Para ello, se asignó a cada cantón el clúster predominante según la frecuencia relativa (moda) de los subsectores localizados en su territorio, sintetizando así la heterogeneidad productiva interna. La validez de esta agregación se evaluó mediante un análisis de pureza cantonal, que arrojó una pureza promedio ponderada del 72,1% para *K-means* y 75,1% para la jerárquica. Estos valores indican que, en promedio, al menos siete de cada diez subsectores dentro de un cantón pertenecen al clúster que se le ha asignado. Si bien se identificó que un 70,8% de cantones presentan una composición mixta —albergan más de un tipo de clúster intracantonal—, la alta pureza promedio evidencia una clara concentración en torno a un clúster dominante en cada territorio. Este hallazgo justifica empíricamente la agregación por moda y otorga solidez a la tipología cantonal resultante, que será la base para el análisis en la siguiente sección sobre implicaciones descriptivas de política pública.

Complementariamente, se examinó la autocorrelación espacial de la productividad laboral manufacturera a nivel cantonal mediante el I de Moran. El resultado de  $-0.0028$  ( $p=0.475$ ) indica ausencia de dependencia espacial significativa. Este hallazgo sugiere que la productividad no se organiza en el espacio de forma simple y continua, sino que responde a estructuras complejas de especialización y diversidad que la clusterización permite detallar. Esta ausencia de autocorrelación espacial ya fue identificada por Guevara (2021) en pymes manufactureras ecuatorianas. No obstante, el análisis LISA, identifica clústers locales de interés: siete cantones en regímenes alto-alto (alta productividad rodeada de alta productividad) y dos en bajo-bajo (baja productividad rodeada de baja productividad). Estos focos puntuales de concentración espacial, aunque no determinan una estructura global, complementan la clasificación cantonal y enriquecen la lectura territorial de los resultados (ver Figura 8).

**Figura 8.** Análisis de autocorrelación espacial: Mapa LISA y diagrama de dispersión



Nota: Elaborado en Python 3.12.4. Fuente: REEM (2023).

## DISCUSIÓN

La caracterización de los clústers productivos adquiere pleno significado económico al trasladar los resultados desde las variables estandarizadas —necesarias para la clusterización— hacia sus valores en escala original (ver Tabla 3). Esta conversión, conjuntamente con su respectivo mapeo, permite visualizar la distribución territorial de los perfiles identificados y dotarlos de interpretación económica, revelando así la complejidad del entramado manufacturero ecuatoriano. Como sostiene Porter (2009), las políticas empresariales o sectoriales deben dar paso a políticas basadas en clústeres, por ser más eficientes, menos distorsionadoras de la competencia y más coherentes con la naturaleza de la competencia en la economía moderna. Por lo tanto, la clusterización no solo clasifica regiones, sino que permite estructurar lineamientos diferenciados de política pública en caminados al cambio estructural (Cimoli *et al.*, 2009).

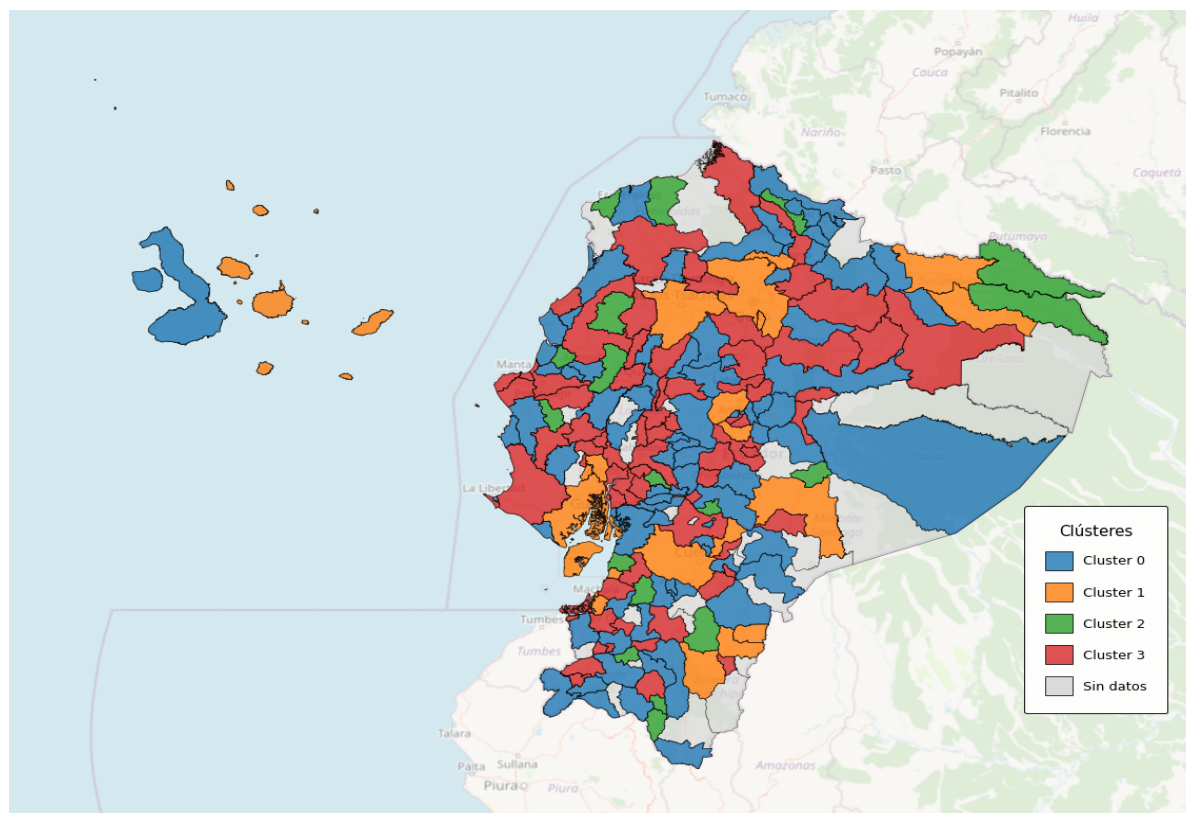
**Tabla 3.** Valores promedios a escala original

Tipo	Número clústers	Productividad laboral (logaritmo natural)	Especialización relativa (LQ)	Diversidad sectorial (Inversa HH)
K-means	0	9.442	2.025	2.484
	1	10.900	1.913	6.388
	2	10.293	33.208	1.939
	3	11.402	1.990	2.184
Jerárquico	0	10.953	1.889	6.143
	1	9.654	119.872	1.529
	2	10.669	26.172	2.109
	3	9.114	2.352	2.656
	4	11.141	1.680	2.104

Nota: Las observaciones están a nivel cantón-subsector que conforman cada clúster. Fuente: REEM (2023).

Según *K-means* (ver Figura 9), el clúster 0, denominado *base manufacturera tradicional extensiva*, presenta la productividad laboral más baja del sistema, equivalente a USD 12.600 por trabajador, con niveles moderados de especialización y diversidad sectorial. Su subsector principal es la fabricación de prendas de vestir. Este perfil, extensamente distribuido por todo el país, sugiere la persistencia de trayectorias productivas históricas que no logran traducir su especialización en ventajas competitivas. Un caso paradigmático es el cantón Antonio Ante, cuya industria textil ha desarrollado una identidad productiva persistente en el tiempo (Arghoty, 2013), reflejando lo que Boschma (2015) define como «especialización institucionalizada»: una dependencia histórica de un sector que refuerza la estabilidad del sistema, pero también limita su capacidad adaptativa. Desde la perspectiva de política pública, estos cantones requerirían intervenciones orientadas a la modernización tecnológica y la cualificación del capital humano, más que a reforzar la especialización existente.

**Figura 9.** Clústers K-means: productividad laboral, especialización relativa y diversidad sectorial



Nota: La lista completa de cantones que conforman cada clúster se detalla en la Tabla 4. Elaborado en Python 3.12.4. Fuente: REEM (2023).

**Tabla 4.** Clusterización K-means por cantones, Ecuador año 2023

CLÚSTER	CANTONES
0	Girón; Gualaceo; San Fernando; Santa Isabel; Chillanes; Chimbo; San Miguel; Caluma; Biblián; La Troncal; El Tambo; Déleg; Tulcán; Bolívar; Espejo; Montúfar; San Pedro de Huaca; La Maná; Pujilí; Saquisilí; Sigchos; Alausí; Guamote; Penipe; Arenillas; Pasaje; Piñas; Zaruma; Esmeraldas; Balzar; El Triunfo; Naranjal; Naranjito; Pedro Carbo; Playas; Coronel Marcelino Maridueña; General Antonio Elizalde; Antonio Ante; Cotacachi; Pimampiro; San Miguel de Urququi; Loja; Calvas; Catamayo; Cécica; Macará; Paltas; Zapotillo; Montalvo; Quevedo; Vinces; Buena Fé; Mocache; Bolívar; Jipijapa; Junín; Rocafuerte; Sucre; Pedernales; Puerto López; San Vicente; Gualaquiza; Limón Indanza; Palora; Santiago; Tena; Quijos; Pastaza; Pedro Moncayo; Puerto Quito; Baños de Agua Santa; Cevallos; Patate; San Pedro de Pelileo; Tisaleo; Chinchipe; Isabela; Gonzalo Pizarro; Cascales; La Joya de los Sachas; La Concordia; La Libertad.
1	Cuenca; Azogues; Guano; Machala; Guayaquil; Daule; Nobol; Otavalo; Morona; Quito; Rumiñahui; Ambato; Zamora; Yanzatza; El Panguí; San Cristóbal; Santa Cruz; Lago Agrio; Shushufindi; Santo Domingo.
2	Pucara; Mira; Chunchi; Atacames; Rioverde; Balao; Simón Bolívar; Chaguarpamba; Espíndola; Quilanga; Flavio Alfaro; Pichincha; Tosagua; 24 de Mayo; Huamboya; Yacuambí; Putumayo; Cuyabeno.

3	<p>Paute; Sigsig; Chordeleg; Camilo Ponce Enríquez; Echeandía; Cañar; Latacunga; Pangua; Salcedo; Riobamba; Colta, Chambo; Pallatanga; Cumandá; Atahualpa; Balsas; El Guabo; ; Huaquillas; Marcabelí; Portovelo; Santa Rosa; Quinindé; San Lorenzo; Alfredo Baquerizo Moreno (Juján); Colimes; Durán; El Empalme; Milagro; Palestina; Samborondón; Santa Lucía; Salitre (Urbina Jado); San Jacinto de Yaguachi; Lomas de Sargentillo; Ibarra; Gonzanamá; Puyango; Saraguro; Pindal; Bahoyo; Puebloviejo; Urdaneta; Ventanas; Valencia; Quinsaloma; Portoviejo; Chone; El Carmen; Manta; Montecristi; Paján; Santa Ana; Jama; Jaramijó; Sucúa; Archidona; El chaco; Carlos Julio Arosemena Tola; Mera; Cayambe; Mejía; San Miguel de los Bancos; Pedro Vicente Maldonado; Mocha; Quero; Santiago de Pillaro; Centinela del Cóndor; Orellana; Loreto; Santa Elena; Salinas</p>
---	--

Fuente: REEM (2023).

El clúster 1, denominado *núcleos dinámicos diversos*, exhibe productividad laboral promedio de USD 54.000, diversidad sectorial alta y una especialización relativa cercana a la media nacional. El subsector referente a reparación e instalación de maquinaria y equipo, sugiere una orientación hacia actividades de mayor contenido tecnológico. Su concentración espacial en las principales ciudades —Quito, Guayaquil, Cuenca, Ambato y Santo Domingo— sugiere una posible asociación entre diversidad sectorial y alta productividad en regiones con alta densidad poblacional, patrón consistente a lo que Jacobs (1969) describe como economías de urbanización, aunque el diseño del estudio no permite establecer relaciones causales. La evidencia para Guayaquil confirma este patrón, al registrar avances en inversión, modernización productiva y empleo formal asociados a empresas con alta capacidad de absorción tecnológica (Rendón *et al.*, 2025). Este dinamismo no es casual, en regiones europeas estos efectos son más intensos cuando existe variedad relacionada, es decir, sectores diversos con bases de conocimiento afines (Frenken *et al.*, 2007), lo que sustenta el enfoque de especialización inteligente para identificar nichos de diversificación vinculados a capacidades existentes (McCann y Ortega, 2013). Para este grupo, las políticas deberían orientarse a fortalecer los encadenamientos productivos y la innovación, aprovechando la diversidad sectorial como activo para la generación de conocimiento intersectorial.

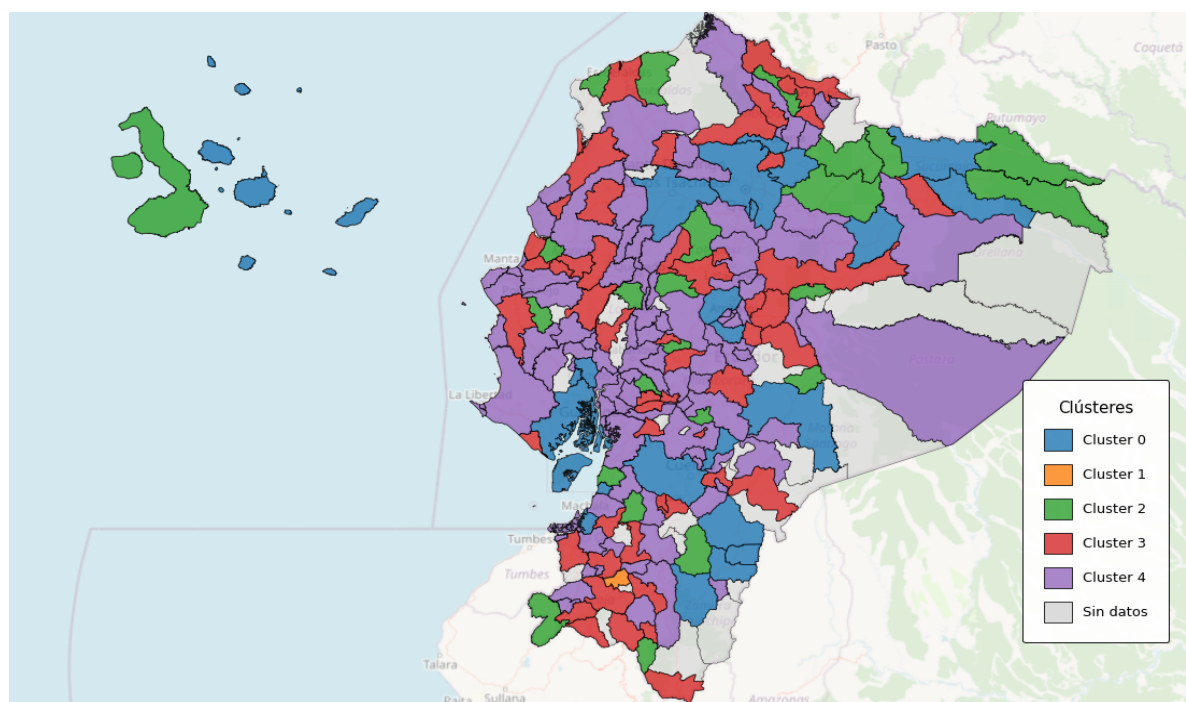
El clúster 2, titulado *enclaves hiperespecializados*, constituye el caso más extremo de especialización relativa, en elaboración de bebidas, el más alto registrado. Sin embargo, su productividad, de USD 29.500 por trabajador es inferior a la de los clústeres 1 y 3, y su diversidad sectorial es la más baja. Este clúster ilustra un patrón donde la alta especialización relativa no garantiza por sí sola mayores niveles de eficiencia. Espacialmente dispersos, estos territorios corresponden a lo que Phelps *et al.* (2015) denominan «enclaves tradicionales», donde el debilitamiento de las economías de localización y la escueta capacidad de innovación generan vulnerabilidad frente a cambios tecnológicos. Estos resultados concuerdan con Torres-Gutiérrez *et al.* (2019), quienes destacan que los cantones ecuatorianos tienden a especializarse en actividades con gran arraigo regional, aunque dicha especialización no siempre se traduce en ventajas productivas sostenidas. Frente a este escenario, la política pública no debería reforzar la especialización existente, sino orientar una reconversión productiva que diversifique las fuentes de empleo y promueva la transición hacia actividades con mayor contenido tecnológico (Kiesel y Dannenberg, 2023).

Por último, el clúster 3, designado *núcleo productivo eficiente equilibrado*, registra la productividad laboral más alta, equivalente a USD 89.000. A su vez, compagina

niveles intermedios de especialización y diversidad sectorial, con predominio del subsector de elaboración de alimentos. Su localización en cantones próximos a las principales urbes sugiere la existencia de micro aglomeraciones flexibles (Pike *et al.*, 2017), capaces de alcanzar altos niveles de eficiencia a través de procesos de innovación adaptativa y aprovechamiento de nichos de mercado específicos. Este patrón es consistente con el concepto de *borrowed size*, según el cual regiones intermedias aledañas a sistemas urbanos mayores pueden beneficiarse de economías de escala, acceso a mercados y flujos de conocimiento sin asumir plenamente costos de congestión (Meijers y Burger, 2017). En términos de política, estos casos requieren fortalecer capacidades institucionales locales y conectividad territorial, aprovechando sus ventajas espaciales, evitando que estos espacios dinámicos permanezcan aislados.

Finalmente, la representación espacial de la clusterización jerárquica (ver Figura 10) desagrega estos perfiles con mayor nivel de detalle. Destaca particularmente el clúster jerárquico 1, con la especialización relativa más alta en el subsector otras industrias manufactureras y la productividad laboral más baja, aproximadamente USD 15.500. Este caso atípico, con apenas dos observaciones en un cantón, resulta inadvertido para *K-means* por su escaso peso relativo, pero ejemplifica situaciones de clúster extremo que pueden requerir atención política específica. Asimismo, los clústeres jerárquicos 0 y 4 desagregan un perfil de alta productividad en dos variantes: una con alta diversidad y otra con baja diversidad, revelando la coexistencia de trayectorias diferenciadas. Este nivel de detalle presentado por el jerárquico no invalida a *K-means*, sino que la complementa: la concordancia moderada entre ambos métodos indica que, pese a sus diferencias, capturan estructuras subyacentes comunes, ofreciendo perspectivas igualmente válidas sobre una misma realidad compleja.

**Figura 10.** Clústers jerárquicos: productividad laboral, especialización relativa y diversidad sectorial



Nota: La lista completa de cantones que conforman cada clúster se detalla en la Tabla 5. Elaborado en Python 3.12.4. Fuente: REEM (2023).

**Tabla 5.** Clusterización jerárquica por cantones, Ecuador año 2023

Clúster	Cantones
0	Cuenca; Azogues; Guano; Machala; Guayaquil; Daule; Nobol; Otavalo, Morona; Gualaquiza; Quito; Cayambe; Rumiñahui; Ambato; Zamora; Yanzatza; El pangui; San Cristóbal; Santa Cruz; Lago Agrio; Shushufindi; Loreto; Santo Domingo.
1	Chaguarpamba.
2	Pucara; Chimbo; Mira; Pangua; Sigchos; Chunchi; Atacames; Rioverde; Balao; Simón Bolívar; Espíndola; Zapotillo; Mocache; Tosagua; 24 de Mayo; Huamboya; El chaco; Carlos Julio Arosemena Tola; Yacuambí; Isabela; Gonzalo Pizarro; Putumayo; Cascales; Cuyabeno.
3	Girón; Gualaceo; San Fernando; San Miguel; Caluma; La Troncal; El Tambo; Tulcán, Bolívar, Espejo; La Maná; Pujilí; Saquisilí, Guamote, Penipe; Arenillas; Pasaje; Piñas; Zaruma; Esmeraldas; Balzar; Naranjito; Playas; Coronel Marcelino Maridueña; General Antonio Elizalde; Cotacachi; San Miguel de Urcuquí; Calvas; Catamayo; Céllica; Macará; Paltas; Quilanga; Vinces; Bolívar; Flavio Alfaro; Jipijapa; Junín; Pichincha; Rocafuerte; Sucre; Pedernales; Limón Indanza; Palora; Tena; Pedro Moncayo; Puerto Quito; Baños de Agua Santa; Patate; Chinchipe; La Joya de los Sachas.
4	Paute; Santa Isabel; Sigsig; Chordeleg; Camilo Ponce Enríquez; Guaranda; Chillanes; Echeandía; Biblián; Cañar; Déleg; Montúfar; San Pedro de Huaca; Latacunga; Salcedo; Riobamba; Alausí; Colta; Chambo; Pallatanga; Cumandá; Atahualpa; Balsas; El Guabo; Huaquillas; Marcabelí; Portovelo; Santa Rosa; Quinindé; San Lorenzo; Alfredo Baquerizo Moreno (Juján); Colimes; Durán; El Empalme; El Triunfo; Milagro; Naranjal; Palestina; Pedro Carbo; Samborondón; Santa Lucía; Salitre (Urbina Jado); San Jacinto de Yaguachi; Lomas de Sargentillo; Ibarra; Antonio Ante; Pimampiro; Loja; Gonzanamá; Puyango; Saraguro; Pindal; Babahoyo; Montalvo; Puebloviejo; Quevedo; Urdaneta; Ventanas; Buena Fé; Valencia; Quinsaloma; Portoviejo; Chone; El Carmen; Manta; Montecristi; Paján; Santa Ana; Puerto López; Jama; Jaramijó; San Vicente; Santiago; Sucúa; Archidona; Quijos; Pastaza; Mera; Mejía; San Miguel de los Bancos; Pedro Vicente Maldonado; Cevallos; Mocha; Que-ro; San Pedro de Pelileo; Santiago de Píllaro; Tisaleo; Centinela del Cóndor; Orellana; La Concor-dia; Santa Elena; La Libertad; Salinas

Fuente: REEM (2023)

En este punto, conviene retomar la advertencia de Wolman y Hincapie (2015), donde el enfoque de clústeres no debe entenderse como un conjunto de políticas prescriptivas, sino como una lente analítica que permite examinar las economías regionales de forma más integrada, superando la visión tradicional de sectores aislados. Bajo esta perspectiva, la discusión presentada constituye precisamente ese lente: una herramienta que permite comprender la complejidad territorial, sobre la cual pueden articularse, con mayor sentido y eficacia, políticas de desarrollo económico más tradicionales pero focalizadas territorialmente.

## CONCLUSIONES

La presente investigación utilizó técnicas de aprendizaje automático no supervisado para explorar y caracterizar la distribución espacial de la productividad

manufacturera ecuatoriana a nivel cantonal. La aplicación de los algoritmos *K-means* y clusterización jerárquica, permitió identificar cuatro perfiles productivos: base manufacturera tradicional extensiva, núcleos dinámicos diversos, enclaves hiperespecializados y núcleo productivo eficiente equilibrado; cuya validez fue respaldada por análisis de estabilidad por semillas, pureza cantonal y consistencia entre métodos. Los hallazgos revelan que la relación entre aglomeración y productividad en Ecuador no responde a patrones únicos, sino a configuraciones territoriales heterogéneas donde especialización y diversidad interactúan de manera diferenciada. La principal contribución de este trabajo radica en ofrecer una clasificación empírica robusta que pueda orientar al diseño de políticas públicas industriales territorialmente diferenciadas. Más que establecer relaciones causales, el análisis proporciona un marco interpretativo que permite alinear estrategias de desarrollo con las capacidades y trayectorias productivas de cada territorio, demostrando el potencial del aprendizaje no supervisado como herramienta para fortalecer la planificación pública basada en evidencia.

No obstante, existen limitaciones que deben considerarse. Su carácter transversal, restringido al año 2023, impide capturar dinámicas temporales o trayectorias de cambio. Asimismo, al tratarse de ejercicio descriptivo, los clústers identificados no deben interpretarse como evidencia causal, especialmente ante la ausencia de covariables como: tamaño empresarial, capital físico, nivel educativo, inversión, infraestructura o factores institucionales. Finalmente, aunque el nivel cantón-subsector permite capturar heterogeneidad intracantonal, puede omitir dinámicas regionales y efectos de vecindad que trascienden los límites tratados aquí.

Futuras líneas de investigación podrían incorporar modelos econométricos espaciales con variable de control para identificar los determinantes de la pertenencia a cada clúster. Asimismo, enfoques de redes productivas y técnicas de densidad espacial permitirían explorar encadenamientos e interacciones territoriales más específicas. En conjunto, estos avances permitirían profundizar en la comprensión de las dinámicas productivas territoriales y orientar política basada en evidencias que promueva un desarrollo sostenible.

## NOTAS

<sup>1</sup> De los 221 cantones del país, 29 no registran actividad manufacturera formal en el REEM 2023 y fueron excluidos. Los 192 cantones restantes conforman la base efectiva para las 1242 observaciones cantón-subsector con empleo positivo. Esta depuración es necesaria para garantizar el correcto cálculo de los indicadores de productividad, especialización y diversidad, evitando la división entre cero o la inclusión de estructuras productivas inexistentes que distorsionarían los resultados.

<sup>2</sup> Es criterio examina los incrementos sucesivos en la altura del dendrograma para identificar saltos abruptos que indican la unión de conglomerados relativamente heterogéneos. La determinación del número óptimo de clústers suele ubicarse inmediatamente antes del mayor incremento en la distancia (Milligan y Cooper, 1985)

<sup>3</sup> Compara la altura de cada fusión con el promedio y la desviación estándar de fusiones previas en el mismo subárbol, permitiendo detectar uniones estructuralmente inestables. Valores elevados sugieren agregaciones entre grupos internamente disímiles (Murtagh y Contreras, 2011).

<sup>4</sup> Corrige el efecto del azar mediante un ajuste basado en el valor esperado bajo independencia de los clústers. El índice toma valores entre -1 y 1, donde 1 indica acuerdo perfecto, 0 representa el nivel de concordancia esperado por azar, y valores negativos señalan acuerdos menores a los esperados aleatoriamente (Albuquerque *et al.*, 2016).

## REFERENCIAS

- Acosta, A. (2012). *Breve historia económica del Ecuador*. Corporación Editora Nacional.
- Albuquerque, M., Nunes de Oliveira, K., Fernandes, J. and Caraciolo, R. (2016). Determination and validation of group numbers in a cluster analysis: A case study applied to forestry science. *Acta Scientiarum. Technology*, 38(3), 339-344.
- Alpaydin, E. (2020). *Introduction to machine learning*. The MIT Press.
- Amara, M. and Thabet, K. (2019). Firm and regional factors of productivity: a multilevel analysis of Tunisian manufacturing. *Annals of Regional Science*, 63(1), 25-51. <https://doi.org/10.1007/S00168-019-00918-X/TABLES/12>
- Arghoty, A. (2013). Nivel de utilización de tecnologías de información y comunicación (TIC) en las pymes de Atuntaqui. En H. Jácome y K. King (Eds.), *Estudios industriales de la micro, pequeña y mediana empresa* (pp. 248-299). Flacso, Sede Ecuador y Ministerio de Industrias y Productividad.
- Arrow, K. (1962). The economic implications of learning by doing. *The Review of Economic Studies*, 29(3), 155-173. <https://doi.org/10.2307/2295952>.
- Balland, P., Boschma, R., Crespo, J. and Rigby, D. (2019). Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies*, 53(9), 1252-1268. <https://doi.org/10.1080/00343404.2018.1437900>
- Balland, P., Broekel, T., Diodato, D., Giuliani, E., Hausmann R., O'Clery, N. and Rigby, D. (2022). The new paradigm of economic complexity. *Research Policy*, 51(3), 104-450. <https://doi.org/10.1016/j.respol.2021.104450>
- Beaudry, C. y Schiffauerova, A. (2009). Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy*, 38(2), 318-337.
- Boschma, R. (2015). Towards an evolutionary perspective on regional resilience. *Regional Studies*, 49(5), 733-751.
- Cepal. (2024). *Anuario Estadístico de América Latina y el Caribe 2024/ Statistical Yearbook for Latin America and The Caribbean 2024*. <https://repositorio.cepal.org/server/api/core/bitstreams/9e2080ce-41d9-4d2d-9386-fb4a1f13a071/content>
- Cerulli, G. (2023). *Fundamentals of supervised machine learning: with applications in Python, R and Stata*. Springer Cham.
- Cimoli, M., Dosi, G. and Stiglitz J.E. (Eds.). (2009). *Industrial policy and development: The political economy of capabilities accumulation*. Oxford University Press.
- Combes, P. and Gobillon, L. (2015). The Empirics of Agglomeration Economies. En G. Duranton, J. Henderson y W. Strange (Eds.), *Handbook of Regional and Urban Economics* 5 (pp. 247-348). Elsevier.
- Creamer, C. (2022). Estado e industrialización en el Ecuador, 1948-2021. *RIRA*, 7(1), 57-122. <https://doi.org/10.18800/revistaira.202201.003>
- Cui, Y., Niu, Y., Ren, Y., Zhang, S. and Zhao, L. (2024). A model to analyze industrial clusters to measure land use efficiency in China. *Land*, 13(7). <https://doi.org/10.3390/land13071070>
- Duranton, G. and Puga, D. (2001). Nursery cities: urban diversity, process innovation, and the life cycle of products. *American Economic Review*, 91(5), 1454-1477.
- Dvoutetý, O. and Blažková, I. (2021). Exploring firm-level and sectoral variation in total factor productivity (TFP). *International Journal of Entrepreneurial Behaviour & Research*, 27(6), 1526-1547. <https://doi.org/10.1108/IJEER-11-2020-0744>

- Fracasso, A. and Vittucci Marzetti, G. (2018). Estimating dynamic localization economies: the inadvertent success of the specialization index and the location quotient. *Regional Studies*, 52(1), 119-132. <https://doi.org/10.1080/00343404.2017.1281388>
- Frenken, K., Van Oort, F. and Verburg, T. (2007). Related Variety, unrelated variety and regional economic growth. *Regional Studies*, 41(5), 685-697. <https://doi.org/10.1080/00343400601120296>
- Glaeser, E., Kallal H., Scheinkman J. and Shleifer A. (1992). Growth in cities. *Journal of Political Economy*, 100(6), 1126-1152. <http://www.jstor.org/stable/2138829>
- Gonzales de Olarte, E. (2021). *Economía regional y urbana: el espacio importa*. Fondo editorial Pontificia Universidad Católica del Perú.
- Guevara, G. (2021). Determinants of manufacturing micro firms' productivity in Ecuador. Do industry and canton where they operate matter? *Regional Science Policy and Practice*, 13(4), 1215-1248. <https://doi.org/10.1111/rsp3.12399>
- Guevara, G., Riou, S. and Autant-Bernard, C. (2018). Agglomeration externalities in Ecuador: do urbanization and tertiarization matter? *Regional Studies*, 53(5), 706-719. <https://doi.org/10.1080/00343404.2018.1470325>
- Horobet, A., Vrinceanu, G., Popescu, C. and Belascu, L. (2021). Business dynamics in recovery times: a comparative perspective on manufacturing firms' performance in the European Union. *Management Dynamics in the Knowledge Economy*. 9(1), 122-136. <https://www.managementdynamics.ro/index.php/journal/article/view/404>
- Jacobs, J. (1969). *The economies of cities*. Random House.
- Jagódka, M. (2025). Typification of polish regions based on human capital and innovativeness: a cluster analysis approach. *Transforming government: People, Process and Police*, 19(3), 614-637. <https://doi.org/10.1108/TG-02-2025-0050>
- Kiesel, C. and Dannenberg, P. (2023). Special economic zones in the global South: between integrated spaces and enclaves —a literature review. *DIE ERDE— Journal of the Geographical Society of Berlin*, 154(1-2), 5-19. <https://doi.org/10.12854/erde-2023-606>
- Kirilyuk, I.L. and Senko, O. V. (2020). Assessing the validity of clustering of panel data by Monte Carlo methods (using as example the data of the Russian regional economy). *Computer Research and Modeling*, 12(6), 1501-1513.
- Kopczewska, K. (2022). Spatial machine learning: new opportunities for regional science. *The Annals of Regional Science*, 68, 713-755. <https://doi.org/10.1007/s00168-021-01101-x>
- Lavoratori, K. and Castellani, D. (2021). Too close for comfort? Microgeography of agglomeration economies in the United Kingdom. *Regional Science*, 61(5), 1002-1028. <https://doi.org/10.1111/jors.12531>
- Lin, H., Li, Y. and Yang, C. (2011). Agglomeration and productivity: Firm-level evidence from China's textile industry. *China Economic Review*, 22(3), 313-329. <https://doi.org/10.1016/j.chieco.2011.03.003>
- Marshall, A. (1890). *Principles of Economics*. Macmillan.
- Martínez-Victoria, M., Maté-Sánchez, M. and Lansink, A. (2017). Spatial determinants of productivity growth on agri-food Spanish firms: a comparison between cooperatives and investor-owned firms: a comparison between cooperatives and investor-owned firms. *Agricultural Economics*, 49(2), 213-223. <https://doi.org/10.1111/agec.12410>
- Matute, K. y Muñoz, E. (2024). Análisis multivariante de factores socioeconómicos en pymes: modelos de regresión machine learning. *Universidad, Ciencia y Tecnología*, 28(125), 142-152. <https://doi.org/10.47460/uct.v28i125.864>

- McCann, P. and Ortega, R. (2015). Smart specialization, regional growth and applications to European Union cohesion policy. *Regional Studies*, 49(8), 1291-1302. <https://doi.org/10.1080/00343404.2013.799769>
- Meijers, E. and Burger, M. (2017). Stretching the concept of 'borrowed size'. *Urban Studies*, 54(1), 269-291.
- Milligan, G. and Cooper, M. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50, 159-179.
- Ministerio de Producción, Comercio Exterior, Inversiones y Pesca. (2023). *Boletín de cifras del sector productivo, diciembre 2023*. Gobierno de Ecuador <https://www.produccion.gob.ec/wp-content/uploads/2023/12/Boletin-Cifras-Productivas-DIC-2023.pdf>
- Ministerio de Producción, Comercio Exterior, Inversiones y Pesca. (2024). *Boletín de cifras del sector productivo, diciembre 2024*. Gobierno de Ecuador <https://www.produccion.gob.ec/wp-content/uploads/2024/12/VFBoletinProduccion-DIC2024.pdf>
- Müller, A. and Guido, S. (2017). *Introduction to machine learning with Python*. Sebastopol: O'Reilly.
- Murtagh, F. and Contreras, P. (2011). Algorithms for hierarchical clustering: an overview. *WIREs Data Mining Knowl Discov*.
- Nakamura, R. and Morrison Paul, C. (2019). Measuring agglomeration. In R. Capello y P. Nijkamp (Eds.), *Handbook of Regional Growth and Development Theories* (2.<sup>a</sup> ed., pp. 389-412). Edward Elgar Publishing.
- Phelps, N., Atienza, M. and Arias, M. (2015). Encore for the enclave: the changing nature of the industry enclave with illustrations from the Mining Industry in Chile. *Economic Geography*, 91(2), 119-146. <https://doi.org/10.1111/ecge.12086>
- Pike, A., Rodríguez-Pose, A. and Tomaney, J. (2017). *Local and Regional Development*. Routledge.
- Porter, M. (2009). *Clusters and economic policy: Aligning public policy with the new economics of competition*. Harvard Business School, Institute for Strategy and Competitiveness.
- Raschka, S., Liu, Y. and Mirjalili, V. (2022). *Machine learning with PyTorch and Scikit-Learn*. Birmingham: Packt Publishing Ltd.
- Rendón, J., Pilacuan, L., Ruiz, H. and Macías, P. (2025). Technological gaps and opportunities in urban productive sectors: The case of Guayaquil in the framework of sustainable development. *E3S Web of Conferences*, 658.
- Rodríguez-Cruz, X. (2024). Análisis de la eficiencia y rentabilidad de las empresas manufactureras ecuatorianas: Una aplicación de algoritmos no supervisados. *Cuestiones Económicas*, 34(2), 164-212, <https://doi.org/10.47550/34.2.5>
- Romer, P. (1986). Increasing Returns and Long Run Growth. *Journal of Political Economy*, 94(5), 1002-1037
- Ruiz, P. (2013). Indicadores de productividad de la industria ecuatoriana. En H. Jácome y K. King (Eds.), *Estudios industriales de la micro, pequeña y mediana empresa* (pp. 45-80). Flacso y Ministerio de Industrias y Productividad.
- Sanfilippo, M. and Seric, A. (2015). Spillovers from agglomerations and inward FDI: a multilevel analysis on sub-Saharan African firms. *Review of World Economics*, 152, 147-176. <https://doi.org/10.1007/s10290-015-0237-6>

- Singh, D. (2021). Cluster space among labor productivity, urbanization, and agglomeration of industries in Hungary. *Journal of the Knowledge Economy*, 13, 1008-1027. <https://doi.org/10.1007/s13132-021-00726-9>
- Tao, J., Ho, C., Luo, S. and Sheng Y. (2019). Agglomeration economies in creative industries. *Regional Science and Urban Economics*, 77, 141-154. <https://doi.org/10.1016/j.regsciurbeco.2019.04.002>
- Torres-Gutiérrez, T., Correa-Quezada, R., Álvarez-García, J. and Del Río-Rama, M. (2019). Agglomeration Economies: An Analysis of the Determinants of Employment in the Cities of Ecuador. *Symmetry 2019*. <https://doi.org/10.3390/SYM11111421>
- Wang, X. and Xu, Y. (2019). An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index. *IOP Conf Ser Mater Sci Eng.* 2019; 569(5): 052024.
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236-244.
- Warrens, M. J., y Van der Hoef, H. (2022). Understanding the adjusted rand index and other partition comparison indices based on counting object pairs. *Journal of Classification*, 39, 487-509.
- Wolman, H. and Hincapie, D. (2015). Clusters and Cluster-Based Development Policy. *Economic Development Quarterly*, 29(2), 135-149.
- Zheng, T. (2013). Measuring Agglomeration using the standardized location quotient with a Bootstrap method. *Journal of Regional Analysis & Policy*, 43(2), 186-197.
- Zollanvari, A. (2023). *Machine learning with Python: theory and implementation*. Springer.