Identificación de casos de trampa de ingresos medios (MIT) mediante clustering dinámico

JORGE LUIS JIMÉNEZ TOLEDO *

Resumen

Este trabajo estudia la Trampa de Ingreso Medio (MIT), una condición estructural que impide a ciertos países alcanzar altos niveles de desarrollo pese a haber superado la pobreza. Combinando teoría estructuralista con técnicas avanzadas de análisis de datos, se construyó una base panel (1990–2020) con más de veinte indicadores económicos y sociales. Las variables fueron utilizadas en un proceso de segmentación mediante clustering unidimensional y multidimensional con Dynamic Time Warping (DTW). La metodología permitió identificar patrones dinámicos y estructurales más precisos que los enfoques basados exclusivamente en umbrales de ingreso, detectando países con altos ingresos, pero con debilidades estructurales propias de la MIT. La clasificación resultante distingue dos grupos: desarrollados, y en transición. Se sugiere que futuras investigaciones incorporen indicadores cualitativos y enfoques mixtos para abordar casos atípicos y explorar subclusterizaciones dentro de la MIT. Desde el punto de vista práctico, el trabajo ofrece una herramienta replicable para organismos multilaterales y gobiernos que busquen diseñar políticas públicas diferenciadas basadas en evidencia. Las implicancias sociales incluyen la posibilidad de identificar trayectorias de desarrollo sostenible y orientar intervenciones estructurales. Este estudio representa un aporte relevante a la economía aplicada, al proponer una metodología robusta y adaptable para comprender y superar la trampa del ingreso medio.

*Facultad de Ciencias Económicas Escuela de Negocios y Administración Pública, Maestría en Economía Aplicada, Buenos Aires, Argentina; email: jorgejimenez222@gmail.com

1. Introducción

La trampa de países de ingreso medio es un fenómeno económico crucial que afecta a numerosos países en desarrollo, impidiéndoles avanzar hacia niveles más altos de desarrollo y prosperidad económica. Este problema se manifiesta cuando un país alcanza un nivel de ingreso medio, pero no logra mantener un crecimiento sostenible que le permita superar este umbral y avanzar hacia un estatus de ingreso alto. El estudio de esta trampa es esencial para entender los desafíos económicos que enfrentan los países en desarrollo y para formular políticas que faciliten su ascenso económico.

Uno de los enfoques más comunes para identificar esta situación es el uso de umbrales de ingreso per cápita, como los propuestos por el Banco Mundial, que dividen a los países en tramos de ingreso bajo, medio (bajo y alto), y alto. Estos umbrales, aunque ampliamente utilizados, presentan limitaciones relevantes: no consideran la evolución estructural ni las trayectorias temporales de los países, y pueden conducir a clasificaciones erróneas o poco representativas de la realidad económica. Por ello, diversos autores han propuesto enfoques alternativos que incorporen variables estructurales, sociales y dinámicas para capturar mejor las condiciones que determinan la permanencia en la trampa de ingresos medios.

En este trabajo se propone un enfoque metodológico alternativo basado en técnicas de clustering estructural y dinámico, que busca superar las limitaciones de las clasificaciones tradicionales. No obstante, la propuesta no se plantea en oposición directa, sino en diálogo con el enfoque convencional: se toma como punto de partida la clasificación tradicional basada en umbrales de ingreso per cápita, aplicada al mismo conjunto de datos para el período 1990–2024. Esta clasificación servirá como referencia (“benchmark”) para contrastar los resultados obtenidos con la metodología alternativa, y así evaluar su capacidad para representar de forma más precisa y coherente la situación estructural y dinámica de los países.

1.1. Hipótesis

Esta investigación parte de una serie de hipótesis orientadas a mejorar la comprensión del fenómeno de la trampa de ingresos medios (MIT) a partir de métodos cuantitativos de clasificación. La hipótesis principal sostiene que los países atrapados en la MIT presentan

ciertos patrones estructurales comunes, tales como baja diversificación económica, altos niveles de desigualdad (reflejados en un índice de Gini elevado), y una limitada inversión en investigación y desarrollo (I+D). Estos patrones pueden ser identificados a partir del análisis multivariado de indicadores estructurales mediante técnicas de clustering no supervisado.

Como hipótesis complementaria, se plantea que la incorporación de técnicas de *clustering* dinámico, como *Dynamic Time Warping* (DTW) —en línea con la noción de clasificación dinámica desarrollada por Diday (1971)— permite capturar trayectorias temporales diferenciadas entre los países atrapados en la MIT y aquellos que lograron superarla. Estas diferencias no siempre son visibles en las clasificaciones estructurales estáticas y ofrecen una perspectiva adicional sobre la evolución de los países en el tiempo.

De acuerdo con Diday (1971) y Diday & Simon (1976), la *cohesión*[1] se refiere al grado de proximidad o similitud interna entre los elementos de un mismo grupo, mientras que la *separación* mide la distancia o disimilitud entre grupos distintos. Este diseño metodológico también busca favorecer la estabilidad de la partición —entendida como la capacidad de mantener la clasificación ante variaciones menores en la muestra o en las condiciones iniciales—, lo que permitiría distinguir de forma consistente y replicable a los países que comparten trayectorias y condiciones estructurales similares, superando las limitaciones de los enfoques que dependen únicamente de umbrales de ingreso per cápita. En lugar de un enfoque unidimensional usando solo PBIpc como la literatura para identificar países MIT, *cluster* k-medias tiene la ventaja de poder identificar grupos de países según una gran dimensionalidad macroeconómica de los países. En su versión estándar, K-medias utiliza distancia euclidiana y se aplica a datos de corte transversal, sin incorporar la dimensión temporal. Para capturar similitudes en series a lo largo del tiempo, se emplea DTW, que permite medir patrones considerando la secuencia y variabilidad temporal. Por lo que, la metodología lo contrasta con las recientes innovaciones de dicho método que tiene en cuenta la temporalidad de todas las variables en *cluster* dinámico (DWT).

[1] En este trabajo, el término *separación* se emplea de manera explícita, mientras que la noción de *cohesión* se refleja implícitamente al analizar la homogeneidad y dispersión interna de los grupos. Por su parte, el término *estabilidad* se empleará de manera intercalada con el de *robustez y este mismo significado.*

1.2. Contribuciones

Este estudio busca aportar varias contribuciones significativas al análisis de la trampa de ingresos medios (MIT). En primer lugar, se propone una metodología reproducible para identificar países atrapados en la MIT, basada en técnicas de *clustering* y *Dynamic Clustering*, mediante un enfoque multidimensional que combine variables estructurales, económicas y sociales. Esta metodología no solo facilita la clasificación de los países, sino que también ofrece un enfoque sistemático y escalable que puede ser aplicado en futuros estudios.

Además, el análisis generará evidencia empírica que complementará la literatura existente al incorporar métodos avanzados de agrupamiento. El uso de *clustering* y *Dynamic Clustering* permitirá identificar patrones y trayectorias que trasciendan las clasificaciones tradicionales basadas exclusivamente en el PBI per cápita. Esto ampliará la comprensión de los factores que contribuyen a la permanencia o superación de la MIT, ofreciendo una perspectiva más rica y holística que la que se encuentra en los enfoques tradicionales.

2. Revisión de la Literatura

El abordaje de la trampa de ingresos medios (MIT) ha sido históricamente dominado por clasificaciones basadas en umbrales de ingreso per cápita. Este enfoque tradicional propone identificar a los países atrapados en la MIT a partir de su incapacidad para superar cierto umbral de ingreso durante un período prolongado. Si bien este criterio ha permitido realizar una primera segmentación global, presenta limitaciones sustantivas para capturar las dinámicas estructurales que impiden el desarrollo sostenido. Por ello, en los últimos años se han introducido enfoques alternativos que incorporan dimensiones estructurales y metodologías dinámicas, ampliando el análisis más allá del ingreso.

2.1. Clasificación tradicional: definición y fundamentos

Los métodos tradicionales de clasificación de países se basan generalmente en los umbrales definidos por organismos internacionales como el Banco Mundial, que segmentan las economías en niveles de ingreso bajo, medio-bajo, medio-alto y alto. Sin embargo, autores como Felipe et al. (2012) han propuesto ajustes relevantes a estos

rangos, señalando que los países con ingresos per cápita cercanos a los 12.000 USD, que no logran superar dicha barrera en el largo plazo, son los más propensos a quedar atrapados. En este trabajo, se adopta una referencia intermedia, considerando como umbral de MIT el intervalo entre 4.000 y 12.000 USD de PBI per cápita promedio (corriente) durante el período 1990–2024. Esta elección no sólo se sustenta en la literatura especializada, sino también en la necesidad de excluir el efecto de eventos atípicos como la pandemia y de asegurar la consistencia temporal de los datos.

2.2. Caracterización empírica tradicional: benchmark de ingreso

A partir de esta definición, se construyó una clasificación empírica aplicando dicho rango al promedio de ingreso per cápita por país. Los resultados muestran que una proporción significativa de países permanece estancada en el rango medio, sin lograr alcanzar el umbral de alto ingreso. América Latina se presenta como una región paradigmática: países como Argentina (8.473 USD), Brasil (6.600 USD), México (7.997 USD), y Perú (3.769 USD), se encuentran en el rango medio-alto o medio-bajo, sin consolidar el tránsito a economías desarrolladas. De acuerdo con la clasificación del presente análisis, más de 100 países están ubicados en categorías distintas al “alto ingreso”, siendo 76 países clasificados como de ingreso “medio-bajo” y 38 como “medio-alto”. Entre los países que sí han superado el umbral se encuentran Corea del Sur (19.019 USD), Irlanda (43.777 USD) y la mayoría de los países de Europa Occidental, Norteamérica y Oceanía.

Este análisis de *benchmark* permite establecer una línea base para contrastar con los enfoques alternativos. Desde un punto de vista cualitativo, los países atrapados en la MIT tienden a compartir ciertas características estructurales: menor capacidad de innovación, desigualdad elevada, dependencia de exportaciones primarias y menor complejidad económica, en línea con los hallazgos de Felipe, Abdon y Kumar (2012)

2.3. Críticas y límites del enfoque tradicional

Si bien el enfoque de umbrales permite una clasificación simple y estandarizada, ha sido objeto de múltiples críticas. La principal limitación radica en su carácter estático: no considera la trayectoria o la dinámica temporal del crecimiento, ni las causas estructurales que impiden el avance económico. Además, el uso de umbrales únicos ignora la heterogeneidad entre países y regiones, lo que dificulta una comprensión más matizada

del fenómeno. Autores como Hausmann y Klinger (2006) y Sachs (2005) han señalado la necesidad de incorporar otras dimensiones —como la estructura productiva, la inversión en I+D, la calidad institucional y la equidad distributiva— para una caracterización más precisa de los países atrapados.

2.4. Enfoques no tradicionales: integración de dimensiones estructurales y dinámicas

En respuesta a estas críticas, se han desarrollado métodos que combinan múltiples variables estructurales y técnicas de análisis multivariado. En particular, este trabajo propone una metodología que integra técnicas de *clustering* estático (K-*means*) y dinámico (*Dynamic Time Warping*, DTW), sobre un panel de más de 15 indicadores económicos, sociales e institucionales. Esta batería de variables incluye desigualdad (Gini), productividad, exportaciones por complejidad económica, inversión en I+D, tasas de urbanización, indicadores de salud, empleo, educación, entre otros. Estas variables fueron seleccionadas con base en la literatura especializada (Acemoglu y Robinson, 2012; Glaeser, 2011; Aghion y Howitt, 1998) y reflejan los factores estructurales que inciden en el desarrollo sostenido.

La ventaja metodológica de los enfoques dinámicos es que permiten observar la evolución de las economías a lo largo del tiempo, capturando estancamientos, convergencias y trayectorias divergentes. Por ejemplo, países como Corea del Sur o Irlanda no solo superaron el umbral de ingreso, sino que lo hicieron manteniendo trayectorias ascendentes y consistentes en múltiples dimensiones estructurales. Por el contrario, países como Argentina, Sudáfrica o Ecuador presentan patrones de estancamiento o retroceso en al menos una parte del período, pese a haber alcanzado en ocasiones valores cercanos al umbral de alto ingreso.

2.5. Resultados cuantitativos y cualitativos del estado del arte

El cruce entre la clasificación tradicional por ingreso y los resultados de los enfoques estructurales permite observar un patrón: la mayoría de los países con bajo o medio ingreso no sólo presentan estancamiento en el PBI per cápita, sino también rezagos en las variables estructurales mencionadas. El análisis empírico evidencia que países como Bangladesh (762 USD), Bolivia (1.686 USD), y Honduras (1.533 USD) siguen atrapados

en niveles bajos de ingreso y presentan, según la literatura, altos niveles de desigualdad y baja inversión en innovación.

Desde una perspectiva cualitativa, la incorporación de dimensiones como educación, salud y composición exportadora revela diferencias sustantivas en las capacidades estructurales de los países. Por ejemplo, países con ingresos similares como Thalandia (7.35k USD) y Sud África (6.25k USD) presentan trayectorias completamente distintas en términos de crecimiento, productividad y complejidad económica. Esto sugiere que el ingreso por sí solo no captura adecuadamente el fenómeno de la MIT, y que es necesario considerar patrones estructurales y dinámicos para una clasificación más precisa.

3. Datos

En primera instancia, se desarrolla un marco analítico que define rigurosamente los conceptos clave que articulan la lógica del estudio, tales como desarrollo económico, estructura económica, diversificación productiva, innovación, desigualdad, entre otros. Estos conceptos no solo tienen sustento teórico en la literatura económica y del desarrollo, sino que también son operacionalizados mediante indicadores cuantitativos extraídos de fuentes como el Banco Mundial y la base de datos World Development Indicators, lo que permite vincular directamente la dimensión teórica con la estrategia empírica.

3.1. Proceso de normalización

El procedimiento técnico completo aplicado para la normalización de las series temporales se detalla en el **Apéndice C, sección E**. Allí se explican los pasos seguidos para segmentar, estandarizar y preparar las series sin exclusión de variables ni imputación de datos faltantes, utilizando un enfoque basado en el z-score.

3.2. Transformación de variables con correlación alta

Una matriz única del panel completo puede inducir sesgos (heterogeneidad no observada, dependencia temporal, tendencias comunes), por lo que se adoptó un enfoque específico para panel: (i) correlaciones "*within*" (descentrando por país) y (ii) correlaciones en primeras diferencias (Δ). Estas se usan con fines diagnósticos, no causales: identificar redundancias y mitigar multicolinealidad antes del *clustering*; cuando $|\rho|$ es alto de forma

consistente, se prioriza el indicador con mejor cobertura y relevancia teórica, evitando eliminar dimensiones nucleares y recurriendo a combinación/transformación si es necesario. Este procedimiento mejora la estabilidad e interpretabilidad del multidimensional y es coherente con el esquema unidimensional posterior (DTW por variable → moda ponderada con pesos de *Random Forest*), donde la contribución de cada dimensión queda explícita.

Adicionalmente, se construyó una vía alternativa de análisis sin aplicar transformaciones ni reducción de variables, con el fin de evaluar la estabilidad de los resultados frente a distintas configuraciones del conjunto de datos. Para una revisión detallada de esta vía sin transformación, ver Apéndice C, sección K.

3.3. Acotación muestra 1990 - 2024

El análisis se centra en un conjunto de series temporales que representan variables económicas clave de 276 países en el período comprendido entre 1960 y 2024. Las series temporales fueron normalizadas para garantizar la comparabilidad entre países y variables, teniendo en cuenta la heterogeneidad en las economías analizadas.

Sin embargo, tras realizar pruebas de *clusterización* con distintos períodos de datos, se observó que los resultados mejoran significativamente al considerar únicamente la información 1990 - 2024. Cuando se emplea la serie completa desde 1960, la segmentación tiende a generar solo dos grandes grupos, perdiendo granularidad y capacidad explicativa. Por otro lado, al realizar la *clusterización* únicamente con los datos del último año disponible, se obtuvo un número elevado de *clusters*, lo que sugiere que la estructura de los datos cambia considerablemente en función del período analizado. En cambio, al limitar el análisis al intervalo 1990-2024, la *clusterización* produce una segmentación más equilibrada, capturando tanto la evolución histórica reciente como la diversidad entre países de manera coherente con la realidad económica contemporánea. Por este motivo, la presente investigación se centra en el período 1990-2024.

4. Enfoque de Cluster y conceptos clave

Sobre esta base conceptual, se construye un enfoque de agrupamiento (*clustering*) que combina métodos estáticos y dinámicos para clasificar economías en función de sus características estructurales y trayectorias de desarrollo. En particular, se utilizan dos técnicas complementarias: por un lado, el algoritmo de K-Medias. que permite la segmentación de países según variables de corte transversal estructurales y socioeconómicas, y un enfoque de agrupamiento dinámico de *cluster* DTW, el cual permite agrupar trayectorias mediante un enfoque temporal.

4.1. Cluster de K-Medias

El algoritmo K-Means (K-Medias), propuesto por MacQueen (1967), constituye el primer enfoque de agrupamiento aplicado en este estudio, con el objetivo de segmentar a los países en conglomerados homogéneos en función de variables estructurales y socioeconómicas clave. Para el *clustering* estático de K-medias, se construyen dos configuraciones complementarias: una que incorpora el PBI per cápita como variable principal y otra que lo excluye, enfocándose exclusivamente en dimensiones estructurales. Esta distinción metodológica permite evaluar si los patrones de desarrollo detectados responden a criterios exógenos de ingreso o a dinámicas más profundas asociadas a la estructura económica. En ambos casos, el número óptimo de *clusters* se determina mediante técnicas como el método del *Elbow* y el coeficiente de silueta, asegurando la robustez estadística de los resultados.

Complementariamente, se lleva a cabo un análisis comparativo detallado entre las distintas clasificaciones generadas, con el fin de evaluar la consistencia y estabilidad de los resultados obtenidos, que permitieron condensar la información temporal de cada país en un único vector representativo. Esta transformación respondió tanto a la necesidad analítica de sintetizar las trayectorias económicas como a las exigencias técnicas de los algoritmos utilizados, particularmente K-Medias, que requiere datos transversales. Para esto se adoptaron dos estrategias: una basada en el valor del último año disponible (Estado Fijo) y otra en el promedio histórico de las variables (Media Histórica)[2]. Esta

[2] En esta investigación, se utilizó la estrategia de media histórica para contrastar resultados con *clustering dinámico* (DTW), lo que permitió explorar el grado de coincidencia entre metodologías y detectar posibles discrepancias asociadas a factores coyunturales o estructurales.

combinación de enfoques da lugar a cuatro clasificaciones complementarias, que permiten observar cómo varían los patrones de agrupamiento cuando se priorizan diferentes conjuntos de variables. El detalle completo del procedimiento estático con K-*Means*, así como sus respectivas configuraciones, número óptimo de *clusters* y resultados, puede consultarse en el **Apéndice C, sección J.**

4.2. Cluster Dinámico

Mientras el *clustering* simple basado en promedios históricos resalta características estructurales persistentes, DTW prioriza la evolución temporal y la detección de patrones dinámicos, proporcionando una visión más consolidada y estable de las diferencias entre países. Esta complementariedad es clave para una comprensión integral de la trampa de ingresos medios (MIT), ya que las agrupaciones detectadas mediante el análisis de k-medias pueden validarse y contextualizarse al contrastarlas con las trayectorias dinámicas obtenidas mediante DTW, el cual es más estable ante fluctuaciones coyunturales. De esta forma, la combinación de ambos enfoques —uno centrado en las trayectorias dinámicas y otro en los perfiles estructurales consolidados— podría ayudar a distinguir con mayor claridad entre países atrapados en la MIT, aquellos en transición y economías que han logrado superarla.

El enfoque DTW fue aplicado de manera unidimiensional con el fin de alinear y comparar series temporales por variable, capturando similitudes en la forma de las trayectorias aun cuando existan desfases temporales y distintas velocidades de ajuste. Este enfoque, permite agrupar países según similitudes en la forma de sus trayectorias de ingreso per cápita, integrando posteriormente ponderadores derivados de técnicas como *Random Forest* para construir una clasificación sintética robusta, basada en la moda ponderada.

Finalmente, las clasificaciones obtenidas son contrastadas con los umbrales de ingreso establecidos por organismos internacionales como el Banco Mundial, lo que permite evaluar en qué medida las segmentaciones empíricas coinciden, discrepan o enriquecen los enfoques tradicionales.

4.3. Definición de conceptos clave

Esta sección resume los conceptos fundamentales utilizados en el modelo de análisis de la trampa de ingresos medios (MIT) y su operacionalización empírica con indicadores

provenientes del Banco Mundial y otros organismos. Las definiciones conservan el sentido original y se vinculan directamente con las variables usadas en el panel 1990–2024.

(i) Desarrollo económico: proceso multidimensional de mejora sostenida del bienestar que incluye ingreso per cápita, educación, salud, equidad y productividad. Operativamente se aproxima con PBI per cápita (ajustado por poder adquisitivo), tasa de urbanización, productividad laboral y reducción de la desigualdad. (ii) Estructura económica: composición sectorial y asignación de factores entre actividades primarias, secundarias y terciarias, asociada a complejidad productiva. Se operacionaliza con empleo en manufactura, apertura comercial y diversificación exportadora. (iii) Diversificación productiva: ampliación y complejización de la base productiva y exportadora, del primario hacia manufacturas y servicios. Se mide con el índice de diversificación del comercio y *proxies* de composición del valor agregado. (iv) Innovación: capacidad de generar y aplicar nuevo conocimiento que impulse productividad y competitividad. Se aproxima mediante gasto en I+D como % del PBIpc, con series del Banco Mundial y complementos de la UNESCO. (v) Capital humano: habilidades, conocimientos y experiencia acumulados. Se mide por tasas de educación secundaria y terciaria y gasto público en educación, capturando la aptitud para adoptar y crear tecnología. (vi) Desigualdad estructural: barrera para el desarrollo sostenido e inclusivo. Se mide con el coeficiente de Gini e impacta cohesión social, acumulación de capacidades y oportunidades. (vii) Tasa de urbanización: indicador de transición estructural hacia sectores más productivos y de cambios en demanda e infraestructura urbana. (viii) Tasa de crecimiento poblacional: refleja presiones demográficas sobre empleo, educación e infraestructura que condicionan la trayectoria de desarrollo. (ix) Términos de intercambio: relación precios de exportación/importación que captura la restricción externa y ayuda a explicar ciclos de estancamiento o expansión. (x) Tasa de inversión en activos fijos: formación bruta de capital fijo como % del PBIpc; indica acumulación de capital físico e impulso de la capacidad productiva de largo plazo. (xi) Tasa de pobreza: porcentaje bajo el umbral internacional (p. ej., 1,90 USD PPP/día); permite evaluar si el crecimiento es inclusivo o deja amplios grupos vulnerables. (xii) Tasa de desempleo: porcentaje de la fuerza laboral sin empleo y en búsqueda activa; refleja capacidad de absorción de trabajo y posibles rigideces del mercado laboral. (xii) Tasa de inversión extranjera directa (IED): IED neta como % del PBIpc; proxy de

integración internacional y confianza de inversionistas. (xiii) Tasa de deuda pública: obligaciones del gobierno central como % del PBIpc; condiciona el espacio fiscal y la vulnerabilidad macroeconómica ante shocks. (xiv) Tasa arancelaria: nivel promedio de impuestos a las importaciones; aproxima el grado de apertura/proteccionismo y la integración a mercados globales.

4.4. Relación Concepto-Variable

El concepto de desarrollo económico se aproxima mediante la variable PBI per cápita (gdp), ampliamente utilizada en la literatura como indicador agregado del nivel de bienestar material y capacidad productiva de un país. La desigualdad de ingresos se representa a través del índice de Gini (gini_value), reconocido como una medida estándar de la concentración de ingresos dentro de una economía (Banco Mundial). La educación se operacionaliza mediante la tasa de inversión en educación respecto al PBIpc (ed_value), lo cual permite capturar el esfuerzo público por mejorar el capital humano. La diversificación productiva se refleja en la variable DivEcoTec_value, un indicador que aproxima la heterogeneidad sectorial y la reducción de dependencia de un solo rubro económico (Hausmann y Hidalgo, 2011). La innovación, por su parte, se captura con la tasa de innovación (TasaInn_value), definida por el número de patentes por millón de habitantes, reflejando la capacidad de una economía para generar conocimiento nuevo y aplicarlo productivamente. La estructura económica se aproxima mediante múltiples indicadores, entre ellos la tasa de inversión en activos (TasaInvAct_value), que da cuenta del dinamismo en la acumulación de capital, y la deuda pública (DeudaPub_value), como reflejo de las capacidades fiscales. La dinámica demográfica, importante en la caracterización de países en desarrollo, se representa por la tasa de crecimiento poblacional (Tasa_Crec_Pob_value) y la tasa de urbanización (Urban_value), ambas relevantes para entender los cambios estructurales en la demanda agregada y en la presión sobre servicios públicos. A su vez, la tasa de pobreza (Pobreza_value) y la tasa de desempleo (Desempleo_value) capturan dimensiones del desarrollo social y la inclusión. Finalmente, las variables Indicador de Crecimiento Económico (Indicador_Crecimiento) y Apertura Económica (Indicador_Apertura_Economica) permiten evaluar la trayectoria de crecimiento sostenido y la integración al comercio internacional. Esta explicitación refuerza la consistencia metodológica del trabajo y permitirá, en la sección de hipótesis,

clarificar de forma más precisa cómo se espera que estas dimensiones se vinculen con la condición de estar atrapado o no en la trampa de ingresos medios.

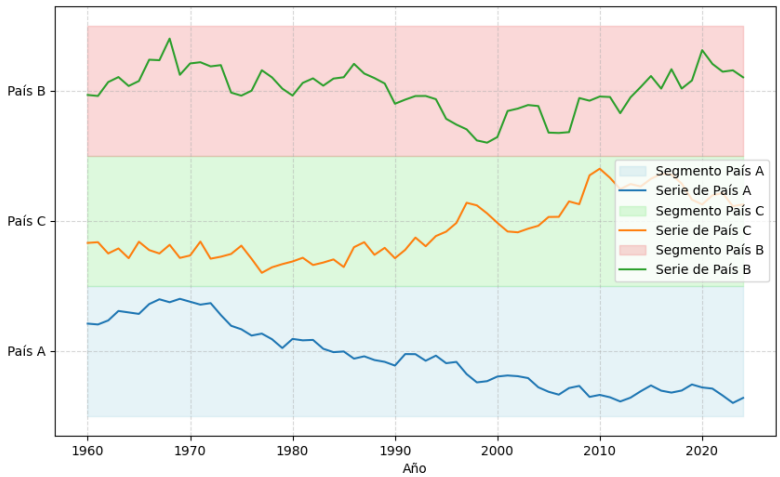
4.4.1. Segmentación y Estructuración de Datos para el Análisis

La segmentación de series temporales es una técnica fundamental para dividir las observaciones en segmentos homogéneos, facilitando la aplicación de métodos como el *clustering*. Un ejemplo gráfico de esta segmentación (Imagen 1) ilustra cómo se agrupan los valores de dos series temporales (Serie 1 y Serie 2) en función del país y del tiempo. Los valores faltantes (*NaN*) son manejados mediante técnicas de imputación o exclusión, según corresponda, para garantizar la robustez del análisis.

En términos de estructuración, los datos se preparan en un formato JSON, donde cada segmento incluye el identificador del segmento y del país asociado, las variables económicas disponibles para cada período, y un registro de las columnas no nulas, permitiendo identificar la consistencia de los datos dentro de cada segmento. Esta estructura permite integrar las series temporales segmentadas en técnicas de análisis avanzado, como DTW, para medir la similitud entre segmentos y realizar agrupaciones jerárquicas más precisas. Para una mejor comprensión de esta organización de datos, se presenta una ilustración detallada en el anexo del *paper*.

Figura 1

Ejemplo de Segmentación Temporal con DTW Clustering



Nota. Esta figura ilustra el uso de *Dynamic Time Warping* (DTW) *Clustering* para segmentar series temporales sin depender de variables específicas. DTW permite

comparar trayectorias en el tiempo, identificando patrones de similitud incluso cuando los puntos de las series no están alineados temporalmente

4.4.2. Explicación técnica del cálculo de distancias

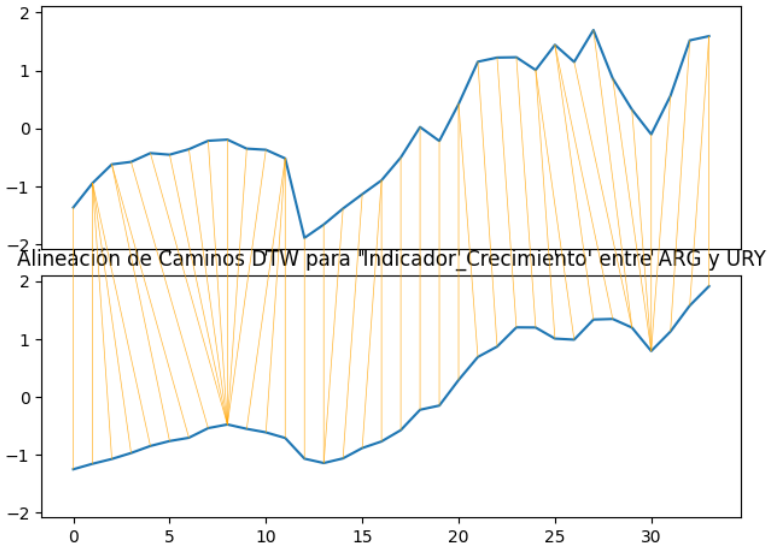
Para calcular las distancias entre series temporales, se utilizó el método de Dynamic Time Warping (DTW), que permite encontrar la mejor alineación entre series que pueden estar desfasadas temporalmente. Esta técnica es especialmente útil en contextos donde las trayectorias pueden tener distintas longitudes o dinámicas. Se aplicó la función dtw.distance() de la librería dtaidistance sobre cada atributo y par de segmentos definidos, y los resultados se almacenaron en un diccionario estructurado en formato JSON, permitiendo su organización, reproducibilidad y escalabilidad para futuros estudios.

El resultado fue una matriz de distancias DTW por atributo, útil como insumo para técnicas de agrupamiento como el *clustering* jerárquico o *k-medoides*. Estas distancias permiten identificar similitudes estructurales entre países a lo largo del tiempo y son especialmente relevantes para analizar la trampa de ingresos medios.

Un ejemplo ilustrativo se presenta mediante la alineación entre dos segmentos (ARG y URY) para el atributo Tasa_Crec_Pob_value. El gráfico resultante muestra alineaciones densas en ciertos periodos, lo que indica trayectorias similares en el crecimiento poblacional, aunque con diferencias en la velocidad y la temporalidad de sus cambios, posiblemente debido a factores económicos o demográficos comunes.

Figura 2

*Ejemplo de a*lineación de Caminos DTW para "Tasa_Crec_Pob_value" entre Segmento ARG y Segmento URY



*Nota***.** Las líneas naranjas en el gráfico representan las alineaciones entre los puntos de ambas series, indicando qué valores han sido emparejados en función de la similitud de su trayectoria temporal. La parte superior del gráfico muestra la serie temporal correspondiente al segmento 1 (Arg**)**, mientras que la parte inferior representa la serie del segmento 7 (Ury)**.**

4.4.3. Formalización del *cluster* dinámico

Formalmente, la distancia DTW se define como la alineación óptima entre dos secuencias temporales $Q = \{q_1, q_2, \dots, q_n\}\ y\ C = \{c_1, c_2, \dots, c_m\}$, donde se busca minimizar el costo total de alineación. Para ello, se considera la función de costo:

$$C(Q, C) = \sum_{i,j} M_{i,j} C_{i,j} \tag{1}$$

donde $\mathrm{C_{i,j}}$ representa la diferencia entre los valores de las secuencias en las posiciones i y j, definida comúnmente como la distancia euclidiana al cuadrado:

$$\mathrm{Ci, j = (qi - cj)2} \tag{2}$$

La matriz M denota todas las alineaciones posibles entre las secuencias, y la solución óptima M^* se obtiene minimizando la función de costo sobre el conjunto de alineaciones viables:

$$M = \arg\min_{\mathrm{M} \in \mathcal{M}} \sum_{\mathrm{i,j}} \mathrm{M_{i,j} C_{i,j}} \tag{3}$$

Computacionalmente, esta optimización se resuelve mediante programación dinámica, construyendo una matriz acumulativa de costos $\mathrm{DTW(i, j)}$, definida recursivamente como:

$$D\mathrm{TW(i, j)} = \mathrm{d}\big(q_i, c_j,\big) + \min \begin{cases} \mathrm{DTW(i-1, j)} \\ \mathrm{DTW(i, j-1)} \\ \mathrm{DTW(i-, j-1)} \end{cases} \tag{4}$$

Esta ecuación indica que la alineación óptima en (i, j) se obtiene sumando el costo de alineación actual $\mathrm{d}(i_i, c_j)$ y el mínimo costo acumulado de los tres caminos posibles: inserción, eliminación o alineación directa. Una vez calculada la matriz $\mathrm{DTW}(\mathrm{i}, \mathrm{j})$, el camino óptimo se reconstruye aplicando arg min, seleccionando en cada paso la transición con menor costo.

Para aplicar este enfoque en la clasificación de países en la MIT, primero se construye una matriz de distancia DTW entre las series temporales. Posteriormente, se emplea un algoritmo de *clustering* DTW, utilizando criterios de enlace para definir las agrupaciones óptimas. Esto permite agrupar economías que han seguido trayectorias de crecimiento similares, incluso si han experimentado diferencias en la velocidad de convergencia.

4.4.4. Acotación de la muestra de países mediante enfoque multidimensional.

En primer lugar, se realizó una filtración inicial de la muestra, excluyendo países con trayectorias de desarrollo considerablemente más bajas, que distorsionaban los resultados del *clustering* por su alto nivel de pobreza y debilidad estructural. Esta selección permitió centrarse en países con trayectorias relativamente comparables, facilitando un análisis más fino sobre la transición hacia el desarrollo.

Posteriormente, se aplicó el Clustering DTW de Primer Nivel sobre los países considerados "desarrollados" en función de sus variables estructurales promedio. Sin embargo, este primer análisis generó solo dos grupos amplios, con una segmentación poco informativa para los fines del estudio.

A partir de esta observación, se decidió aplicar un Clustering DTW de Segundo Nivel a la misma muestra, sin excluir países con trayectorias menos claras. Esta segunda aplicación permitió una mayor granularidad en la clasificación, identificando subgrupos con trayectorias diferenciadas: países nórdicos, América Latina, Europa Occidental, Europa del Este, entre otros.

Esta estructura de doble nivel de análisis permitió consolidar una muestra de países estructuralmente comparables, pero con suficientes diferencias dinámicas como para enriquecer el análisis. Además, facilitó la posterior aplicación del *clustering* DTW

unidimensional sobre variables específicas, permitiendo identificar con mayor precisión patrones asociados a la trampa de ingresos medios (MIT).

4.4.5. Justificación del Clustering Unidimensional

El uso de clustering unidimensional aplicado de forma independiente a cada variable estructural responde a la necesidad de capturar patrones específicos y consistentes dentro de cada dimensión económica, sin imponer restricciones asociadas a combinaciones multivariadas. Esta estrategia permite identificar, para cada atributo temporal, agrupamientos homogéneos en función de su trayectoria histórica, utilizando distancias DTW que reflejan similitud en la evolución temporal y no solo en niveles absolutos.

La principal ventaja de este enfoque es que permite construir una grilla de clasificación dicotómica, donde cada país recibe una asignación por variable que indica su pertenencia relativa a un grupo más desarrollado o a uno de ingreso medio, en función del comportamiento temporal observado. Esta grilla sirve como base para un análisis integral, en el cual se consideran todas las variables relevantes sin necesidad de reducir dimensionalidad o perder especificidad. En lugar de buscar un único modelo que capture todas las interacciones posibles entre variables, se opta por respetar la estructura individual de cada serie, permitiendo observar cómo se posiciona cada país dentro de cada dimensión económica clave. Así, se genera un marco empírico sólido para la posterior construcción de un indicador resumen mediante moda ponderada, que sintetice la pertenencia estructural global de cada país.

4.4.6. Construcción de grilla de Clustering mediante clasificación DTW Unidimensional

A partir de los resultados obtenidos en el análisis de *clustering* DTW unidimensional por variable, se procedió a construir una grilla binaria de clasificación estructural**,** con el objetivo de integrar la información proveniente de cada dimensión económica. En esta grilla, cada fila representa un país y cada columna una variable clave del análisis (como PBI per cápita, desigualdad, inversión, educación, innovación, entre otras). A cada país se le asignó un valor de 0 o 1 por variable, en función del grupo al que fue clasificado mediante el *clustering*: el valor 1 corresponde al grupo estructuralmente más cercano a

las economías desarrolladas, mientras que el valor **0** indica mayor similitud con trayectorias de ingreso medio. Esta matriz de clasificación permite visualizar el comportamiento agregado de cada país en múltiples dimensiones, y fue utilizada como base para calcular la moda ponderada por fila, incorporando ponderadores previamente estimados con *Random Forest*. Este procedimiento permite generar una clasificación final integrada, que resume de forma estructural y multidimensional la condición de cada país respecto a la trampa de ingresos medios.

4.4.7. Homologación de la Grilla: Corrección de Etiquetas según Desarrollo Económico

Dado que la asignación de etiquetas en el *clustering* no tiene un significado intrínseco, es posible que la numeración inicial no refleje correctamente la división entre países desarrollados y de ingreso medio. Para garantizar una interpretación coherente, se aplica un proceso de homologación de etiquetas, el cual busca asegurar que *Cluster* 0 represente a los países desarrollados y *Cluster* 1 a los de medio ingreso. Para ello, se determina cuál de las dos etiquetas contiene una mayor proporción de países desarrollados, a través del PBI per cápita, siguiendo las convenciones establecidas en la literatura. Si la mayoría de los países desarrollados fueron asignados al *Cluster* 1, se procede a invertir las etiquetas en toda la grilla, intercambiando los valores de 0 y 1 según corresponda.

Este ajuste permite corregir inconsistencias en la clasificación y garantizar que la grilla de *clustering* refleje fielmente la estructura económica global. Como resultado, la grilla ajustada proporciona una herramienta más precisa para el análisis comparativo entre países, permitiendo identificar patrones económicos y sociales de manera coherente con la realidad económica.

Tabla 2

Ejemplo de la Grilla Ajustada para Argentina y Uruguay

Country	gdp	TasaInv	TasaCrecPob	Desemp	DeudaPub	TasaInn	TasaCrecEcon	Ind_Crec	Ind_Desig	Ind_Apert	ed	Term_Inter	gini	Pobreza	DivEcoTec	Cluster
ARG	1	2.0	2	2	2.0	1.0	1	1	2	2	1.0	2.0	1.0	1.0	2.0	2
URY	2	2.0	2	2	–	1.0	1	1	1	1	1.0	2.0	1.0	2.0	1.0	1

4.4.8. Obtención de ponderadores sobre la media 1990-2024, de la segmentación de países.

En el análisis de *clustering*, la correcta asignación de países a distintos grupos depende de la relevancia de las variables utilizadas en la segmentación. Sin embargo, no todas las variables contribuyen en la misma medida, por lo que es fundamental establecer un esquema de ponderación que refleje su importancia relativa.

En este estudio, se evaluará la importancia de las variables a través de *Feature Importance* en un modelo de *Random Forest*, el cual mide la contribución de cada variable a la segmentación en función de la reducción de impureza dentro de los árboles de decisión. Paralelamente, se analizarán los valores SHAP (*SHapley Additive exPlanations*), que permiten una interpretación más granular al descomponer el impacto individual de cada variable en la predicción y mostrar tanto su relevancia como la dirección de su efecto. **Apéndice C, sección N**, donde se examina su aplicación para comparar los resultados obtenidos con *Featured Importance*

4.4.8.1. Random Forest con Featured Importance.

Se utilizó *Random Forest* como método de análisis de importancia de variables. Primero, se asignaron los clusters generados a partir del dendrograma a cada país, estableciendo un umbral de distancia para la formación de grupos. Luego, se transformaron las series temporales en una matriz de entrenamiento, donde cada fila representaba un país y cada columna contenía el valor promedio de una variable a lo largo del tiempo. Se implementó un manejo de valores faltantes, reemplazando NaN con 0 o la media de la columna para evitar distorsiones en el modelo. Una vez estructurados los datos, se entrenó un *Random Forest Classifier* con 100 árboles de decisión (*n_estimators*=100), utilizando los valores económicos como predictores (X) y los *clusters* asignados como variable objetivo (y). Para obtener un análisis consolidado, se normalizó la importancia dentro de cada iteración y se sumaron los valores a lo largo de todas las variables evaluadas. Para evitar problemas de escala, se aplicó un factor de ajuste (x1000) a la importancia final antes de la visualización. Adicionalmente, se corrigieron valores negativos o infinitos en la matriz de distancias de DTW para evitar errores en la construcción de *clusters*. Para mejorar la

estabilidad del modelo, se verificó que la suma total de importancia en cada iteración de *Random Forest* fuera mayor a 0, evitando normalizaciones incorrectas. También se controló que las importancias no fueran desproporcionadamente pequeñas mediante un ajuste en la escala de valores. Finalmente, se validó la coherencia del análisis mediante la comparación de los clusters generados con las tendencias observadas en las variables más influyentes.

4.4.9. Aplicación de la Moda Ponderada como Criterio de Clasificación Final

Una vez construida la grilla de *clustering* unidimensional por variable, se procedió al cálculo de la moda ponderada por país, para cada una de estas, según la relevancia obtenidas mediante técnica de *Random Forest*, esto, con el objetivo de generar una clasificación integrada que refleje la pertenencia estructural de las variables a la categoría de economías desarrolladas o atrapadas en la trampa de ingresos medios (MIT). La elección de la moda ponderada como método de agregación responde a su capacidad para sintetizar múltiples dimensiones en una única etiqueta estructural sin perder la heterogeneidad de los atributos considerados.

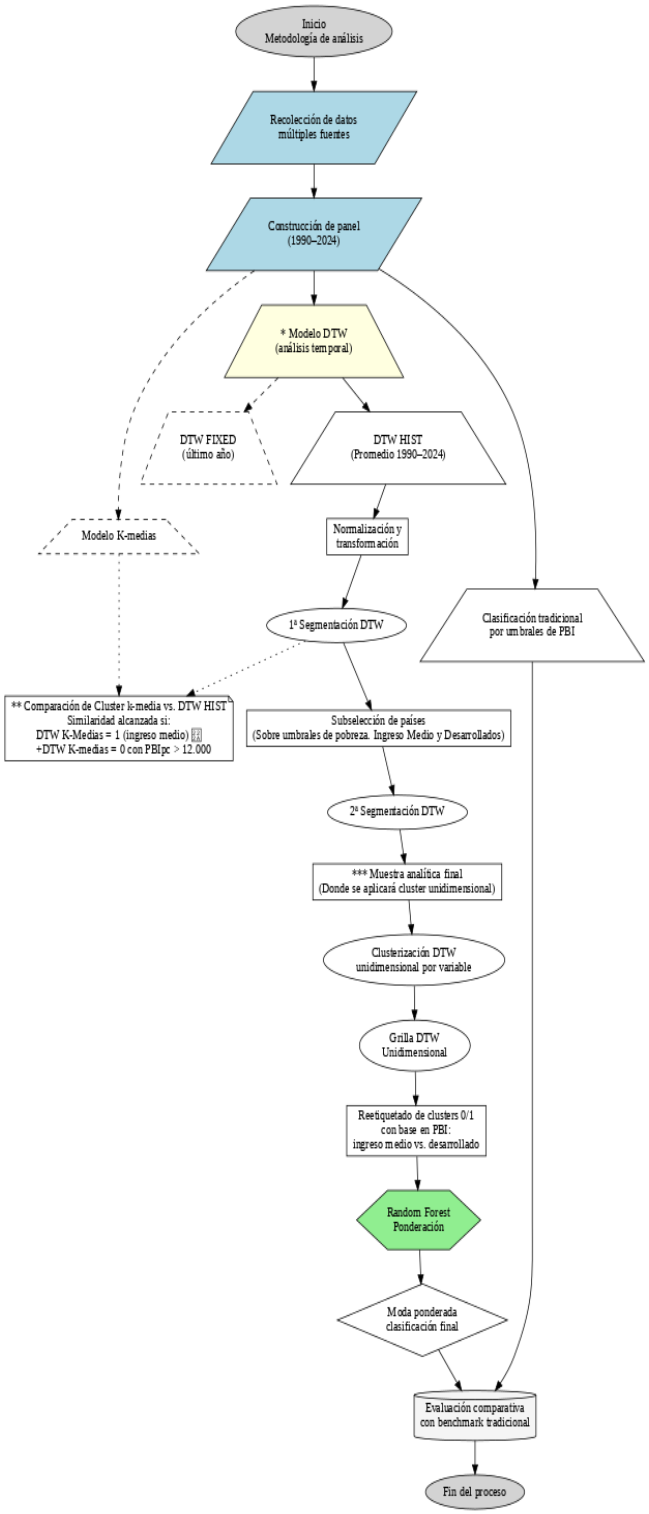
Este enfoque metodológico permite evaluar la condición estructural de cada país en función de la consistencia entre variables económicas clave, buscando superar los enfoques tradicionales. A través de esta estrategia, se busca identificar posibles casos MIT o economías en transición. Esta clasificación integrada será luego utilizada para analizar coincidencias y divergencias respecto a los umbrales internacionales convencionales.

4.5. Flujograma metodológico: descripción y justificación de etapas

El flujograma metodológico constituye la figura central del capítulo de Metodología, permitiendo visualizar de forma estructurada y secuencial todas las etapas del análisis aplicado en esta investigación.

Figura

Flujograma metodológico



*. DTW FIXED se refiere al cluster dinámico que utiliza únicamente el último año, mientras que DTW HIST es el cluster dinámico que contiene toda la temporalidad del panel. El análisis DTW FIXED se presenta de manera complementaria en el **Apéndice C, sección F**, donde se examina su aplicación y se comparan los resultados obtenidos con los de DTW HIST.

**. Resultados similares bajo la siguiente condición: K-Medias cluster 1 que corresponde a países de ingreso medio, y K-Medias cluster 0 que incluye únicamente países con PBIpc superior a 12.000 USD. Esto ya que k-medias agrupa ingresos medios por un lado (cluster 1), e ingresos altos y bajos por el otro (cluster 0).

***. Dado que la distinción más desafiante es entre países de ingresos altos y medios, se eliminan los países de ingreso bajos, cuya separación entre grupos (medio vs. bajo) es más clara cuando se agrega mayor dimensionalidad a simplemente el umbral del PBIpc

5. Resultados Empíricos

Se describen las fuentes de datos y criterios de selección utilizados para construir el panel, seguidos por la metodología de clasificación preliminar según umbrales de ingreso. Finalmente, se compara esta clasificación con resultados obtenidos mediante *clustering* y técnicas dinámicas, evaluando su consistencia y complementariedad.

5.1. Datos

La presente investigación se basa en una base de datos panel país-año que abarca el período 1960–2024. El objetivo central es captar tanto las trayectorias de largo plazo como las tendencias recientes de los países en relación con la Trampa de Ingreso Medio (MIT). Para ello, se integraron fuentes primarias del Banco Mundial y datos complementarios de organismos internacionales y nacionales, priorizando la continuidad y calidad de las series temporales.

Tras un proceso de depuración, la base final incluye 138 países, 3.450 observaciones y 25 variables clave asociadas al desarrollo económico y social, entre ellas: PBI per cápita, desigualdad, pobreza, educación, inversión, crecimiento poblacional, productividad, innovación, comercio y deuda pública. Esta información permite aplicar umbrales de la literatura para clasificar preliminarmente a los países según su situación respecto a la MIT, así como sustentar los análisis exploratorios y de *clustering*. Para más detalles sobre orígenes y definiciones de variables, véase Apéndice A. El procesamiento se realizó en Stata, y la descripción completa de las variables, fuentes y herramientas utilizadas se presenta en los Apéndices A y B.

5.1.1. Características de Panel

La construcción del panel comenzó con un universo de 276 países, reducido a 138 luego de aplicar filtros de completitud y consistencia. El análisis se centró en variables estructurales clave, organizadas luego en tres indicadores compuestos: crecimiento económico, desigualdad y pobreza, y flujos de capital. Estos fueron definidos con base en correlaciones conceptuales y baja colinealidad estadística, facilitando una lectura más sintética de los determinantes de la MIT.

Respecto a la disponibilidad de datos, variables como PBI per cápita y crecimiento poblacional presentan baja proporción de datos faltantes (<20%), mientras que desigualdad y diversificación tecnológica superan el 80%, especialmente en décadas anteriores. El 93% de los países tiene menos del 50% de datos faltantes, y el 45% menos del 10%, lo que permitió aplicar técnicas de imputación selectiva para asegurar representatividad y solidez en los análisis posteriores.

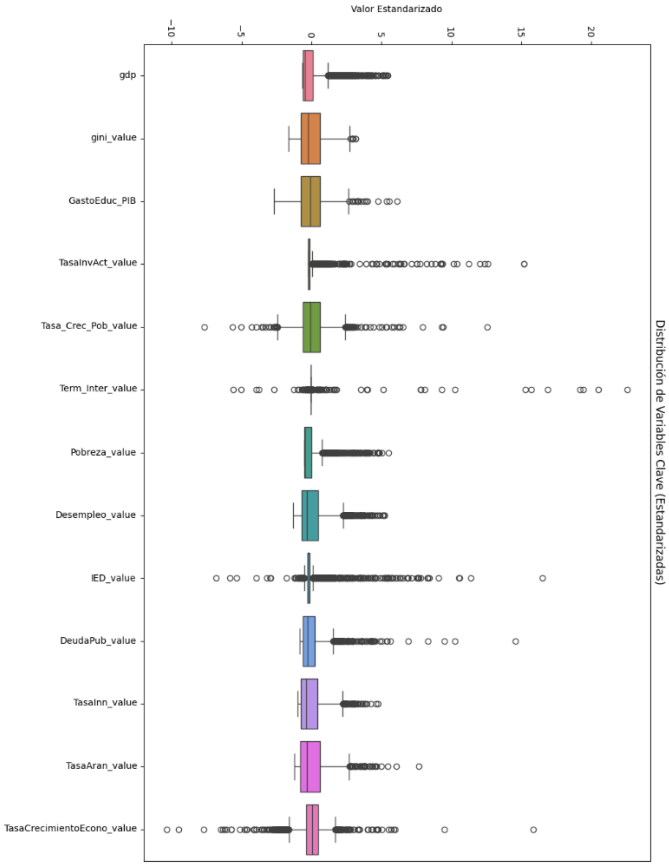
5.1.2. Estadísticas Descriptivas

El análisis descriptivo reveló patrones estructurales relevantes: los países en la MIT presentan en promedio menor PBI per cápita, mayor desigualdad y menor inversión en educación respecto a los países fuera de la trampa. Estas diferencias se reflejan también en la dispersión estadística de las variables. Los gráficos de *boxplots* muestran alta variabilidad y presencia de valores atípicos en variables como PBI per cápita y desigualdad, y mayor homogeneidad en educación e inversión.

Para mayor detalle sobre datos faltantes, estadísticas descriptivas y representaciones gráficas, véase Apéndice C, sección C.

Figura 3

Distribución de variables clave



Nota. La caja representa el *rango intercuartil* (IQR), que va desde el *cuartil 25%* hasta el *cuartil 75%*. 2. La *línea central* dentro de la caja es la *mediana* de los datos (cuartil 50%).

3. Los *bigotes* se extienden hasta el valor máximo dentro de 1.5 veces el IQR desde los cuartiles 25% y 75%. 4. Los puntos *fuera de los bigotes* son los *valores atípicos* (*outliers*), que están significativamente alejados de la mayoría de los datos. 5. Los *colores* de las cajas ayudan a diferenciar las distintas variables representadas. Este gráfico permite comparar la distribución, la dispersión y los valores atípicos de las variables estandarizadas

5.1.3. Distribución del PBIpc

El análisis de las distribuciones del PBI per cápita en 1990 y 2024 (ver Apéndice C, sección D) evidencia un desplazamiento general hacia mayores niveles de ingreso promedio, aunque con persistente concentración en el rango de ingresos medios. Si bien la densidad global en 1990 se concentraba cerca del umbral de ingreso bajo, en 2024 se observa una mejora agregada; sin embargo, la mayoría de los países continúa ubicada entre los tramos de ingreso bajo y medio-bajo (USD 1.135 a USD 4.465).

Esta persistencia sugiere que el crecimiento del ingreso per cápita no ha sido suficiente para romper con las barreras estructurales que caracterizan a la Trampa de Ingreso Medio (MIT). La evidencia respalda la aplicación de metodologías como el clustering para distinguir trayectorias diferenciadas, así como la necesidad de ir más allá del ingreso promedio al analizar patrones de desarrollo. Para una exploración adicional sobre la relación entre el PBIpc y la desigualdad, véase Apéndice C, sección C.

5.2. Análisis simple Clústeres: Caracterización Socioeconómica de Países Basada en Indicadores de Desarrollo y Crecimiento

Las variables MIT_Fixed y MIT_Hist representan, respectivamente, el valor del PBIpc en el último año disponible y el promedio histórico de la muestra. El análisis muestra que MIT_Hist tiende a registrar un PBIpc más bajo que MIT_Fixed, lo cual es esperable dado el crecimiento económico sostenido en muchos países. Esta diferencia refleja cómo los niveles actuales superan a los promedios del pasado, cuando el PBI per cápita y total eran generalmente inferiores por razones estructurales y contextuales.

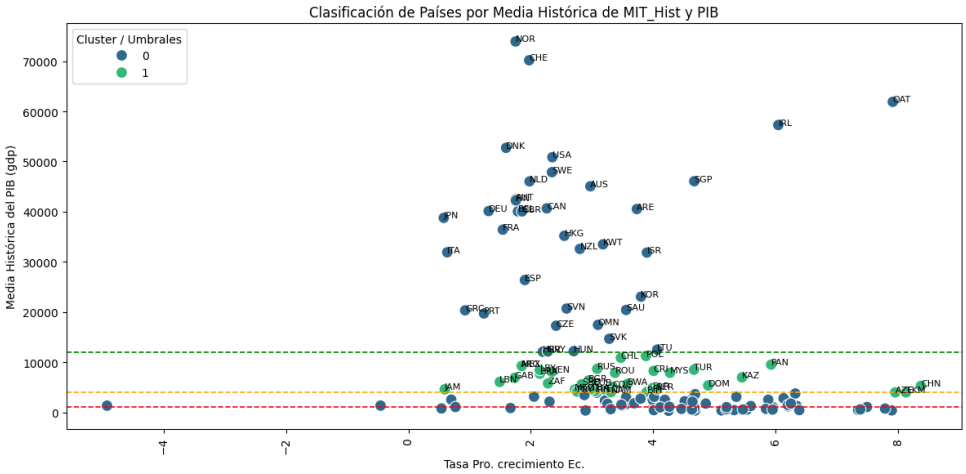
5.2.1. Enfoque principal (MIT_Hist).

El modelo de *clustering* principal (MIT_Hist), con variables normalizadas, segmentó a los países en dos grupos —uno con ingresos medios y debilidades estructurales

persistentes, y otro con economías más desarrolladas—, siendo el PBI per cápita histórico la variable de mayor poder discriminante; a partir de esta partición se aisló un subconjunto relevante de economías desarrolladas e ingreso medio alto para examinar casos en posible situación de MIT**.** MIT_Hist, que incorpora el promedio histórico del PBI per cápita, constituye el enfoque central de la investigación para clasificar países respecto de la MIT al capturar trayectorias de largo plazo, mientras que MIT_Fixed, basado en valores recientes, brinda una perspectiva alternativa pero más sensible a fluctuaciones coyunturales. El análisis comparativo entre MIT_Hist y MIT_Fixed se presenta de manera complementaria en el Apéndice C, sección F.

Figura 4

Agrupación de países por media historia de MIT_Hist y PBIpc



5.2.2. Determinación del Número Óptimo de Clusters basado en MIT_Hist

Para validar la elección del número de clusters en los análisis basados en MIT_Hist, se utilizaron métricas estadísticas estándar como el método del codo (*Elbow*), el coeficiente de Silueta y el índice de Calinski-Harabasz. Los tres métodos coincidieron en señalar que la segmentación en dos *clusters* representa una solución óptima, al balancear simplicidad con capacidad explicativa. Los resultados detallados de este análisis, incluyendo gráficos y puntuaciones específicas, se presentan en el Apéndice C, sección G.

5.2.3. Agrupación de Países Mediante Clustering K-*Means*: Análisis de Desarrollo Socioeconómico (Sin PBIpc)

Con el objetivo de evaluar la relevancia del PBIpc como variable estructural, se realizó un análisis de *clustering* K-Means excluyéndolo del conjunto de variables históricas (MIT_Hist). Los resultados evidencian una pérdida clara de estructura en los agrupamientos, lo que confirma la importancia del rol del PBIpc como discriminador entre niveles de desarrollo. Para una descripción detallada del procedimiento, los gráficos correspondientes y su comparación con variantes de cuatro grupos, véase Apéndice C, sección G y H (ya existentes), así como Apéndice C, sección I, donde se presenta el análisis completo sin la variable de PBIpc.

5.3. Introducción al Cluster Dinámico (DTW)

El análisis de agrupamiento basado en *Dynamic Time Warping* (DTW) permite evaluar similitudes entre series temporales, incluso cuando presentan distintas longitudes o desfases temporales. Esta propiedad resulta especialmente útil para analizar trayectorias económicas heterogéneas, como crecimiento, desigualdad, inversión o diversificación, que evolucionan de forma no sincronizada entre países. En este trabajo, DTW se utiliza como métrica de distancia para construir *clusters* que capturen patrones dinámicos entre economías. El procedimiento metodológico incluye la estandarización de series temporales por país, el cálculo de distancias mediante DTW y la posterior aplicación de *clustering* jerárquico. Este enfoque permite identificar grupos de países con dinámicas estructurales similares, superando las limitaciones de métodos basados en promedios o correlaciones simples. Así, el *clustering* con DTW se configura como una herramienta robusta para explorar las características comunes de las economías en desarrollo y aquellas atrapadas en la Trampa de Ingreso Medio (MIT)

5.3.1. Análisis de Correlación: Exploración de Relaciones Entre Variables Económicas

Se realizó un análisis de correlación para examinar las relaciones entre las principales variables estructurales consideradas en el estudio. Este ejercicio permitió identificar asociaciones significativas que fundamentaron la posterior construcción de indicadores compuestos utilizados en los análisis multivariados. Para una descripción detallada de las correlaciones observadas, así como la metodología aplicada y los indicadores derivados, véase Apéndice C, sección K.

5.3.2. Implementación de Cluster DTW Multidimensional

El análisis de las series temporales segmentadas incluyó la implementación de la técnica *Dynamic Time Warping* (DTW) para calcular distancias entre segmentos.

5.3.2.1. **Atributos Considerados**.

Se han considerado un total 18 variables clave para el análisis. Estas variables abarcan distintos aspectos fundamentales del desarrollo económico, la desigualdad, la inversión, la pobreza y la apertura comercial.

En primer lugar, se consideran indicadores de crecimiento y desarrollo económico, tales como el PBIpc (*gdp*), la tasa de crecimiento económico (TasaCrecimientoEcono_value) y un indicador específico de crecimiento (Indicador_Crecimiento). En relación con la distribución del ingreso y la desigualdad, se incluyen el coeficiente de Gini (gini_value), la tasa de pobreza (Pobreza_value) y un indicador de desigualdad y pobreza (Indicador_Desigualdad_Pobreza).

Asimismo, se analizan variables relacionadas con la inversión y el mercado laboral, como la tasa de inversión activa (TasaInvAct_value), la inversión extranjera directa (IED_value), la tasa de desempleo (Desempleo_value) y la deuda pública (DeudaPub_value). La diversificación económica también es un aspecto relevante dentro del análisis, representado por indicadores de diversificación comercial *(DivEcoCom_value)*, tecnológica (DivEcoTec_value) e inversión (DivEcoInv_value).

Por otro lado, se examinan factores relacionados con la apertura comercial y la competitividad, incluyendo la tasa de innovación (TasaInn_value), los aranceles comerciales (TasaAran_value) y un indicador general de flujo de capital (Indicador_Flujo_Capital). Finalmente, se incorporan variables institucionales como la duración de los mandatos (Term_Inter_value) y la tasa de crecimiento poblacional (Tasa_Crec_Pob_value).

5.3.2.2. Clustering Jerárquico DTW: agrupamiento inicial y selección del grupo de análisis.

El clustering jerárquico basado en la métrica Dynamic Time Warping (DTW) permitió agrupar a los países en dos grandes grupos según similitudes en sus trayectorias

económicas normalizadas. A través del dendrograma resultante, se identificaron patrones diferenciados en las dinámicas estructurales de desarrollo.

El Grupo Naranja (*Cluster* 1) incluye economías con estructuras diversificadas y mayor relevancia analítica para el estudio de la Trampa de Ingreso Medio (MIT). Está conformado por países desarrollados y emergentes como Argentina, Brasil, Estados Unidos, Alemania, Japón y Australia, entre otros, que presentan diferencias en niveles de ingreso, pero comparten características institucionales, productivas y sociales que los hacen comparables.

En contraste, el Grupo Verde (*Cluster* 2) agrupa economías atípicas, como microestados, países rentistas o con estructuras altamente concentradas (por ejemplo, Mónaco, Qatar, Kuwait), que, si bien presentan elevados niveles de ingreso per cápita, no poseen trayectorias estructurales representativas del proceso de desarrollo que caracteriza a la MIT. La fuerte dependencia de sectores específicos, la baja diversificación y las condiciones institucionales excepcionales justifican su exclusión del análisis principal.

Por lo tanto, el Grupo Naranja fue seleccionado como núcleo para los análisis posteriores. Sobre este grupo se aplicará un segundo nivel de *clustering* jerárquico y un análisis DTW unidimensional por variable, lo que permitirá identificar subgrupos con comportamientos similares, examinar su evolución temporal, y analizar el impacto de variables críticas como la desigualdad, la diversificación y la inversión externa.

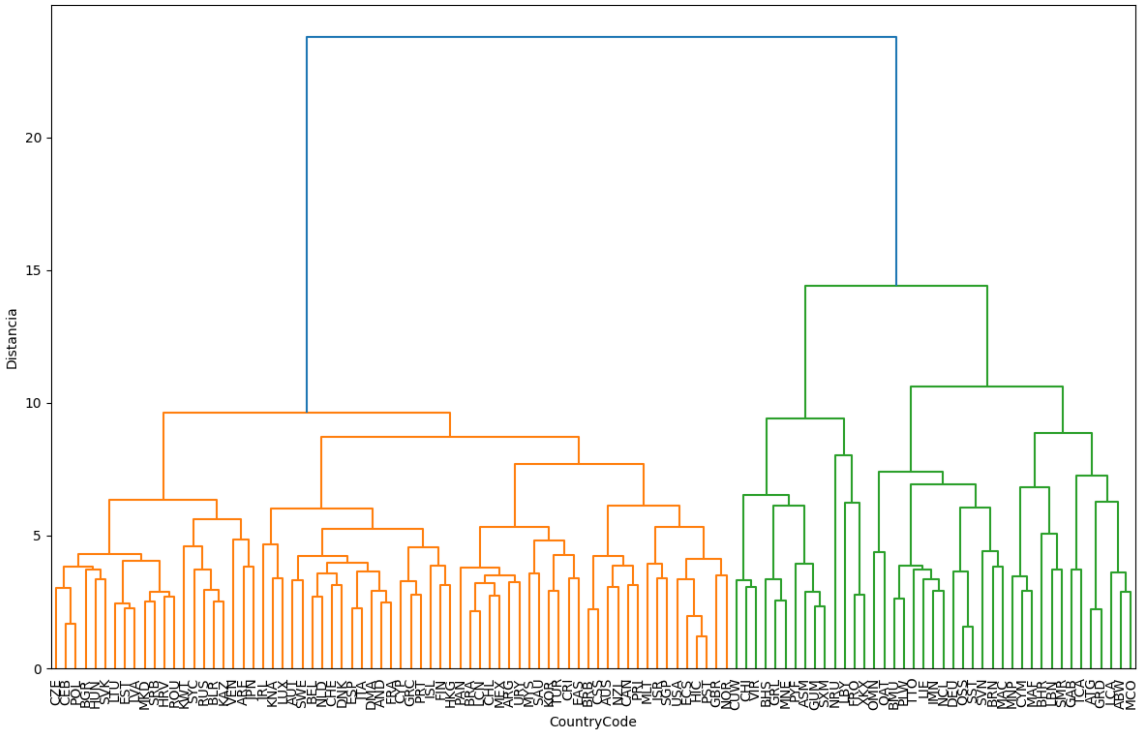
Tabla 4

*Distribución de países por cluster***.**

Cluster	Países
1 (Grupo Naranja)	Argentina (ARG), Australia (AUS), Austria (AUT), Bélgica (BEL), Brasil (BRA), Canadá (CAN), España (ESP), Francia (FRA), Alemania (DEU), Italia (ITA), Japón (JPN), México (MEX), Reino Unido (GBR), Suiza (CHE), Estados Unidos (USA), Albania (ALB), Armenia (ARM), Bulgaria (BGR), Bielorrusia (BLR), Chile (CHL), Croacia (HRV), Chipre (CYP), República Checa (CZE), Dinamarca (DNK), Estonia (EST), Finlandia (FIN), Grecia (GRC), Hungría (HUN), Irlanda (IRL), Israel (ISR), Kazajistán (KAZ), Letonia (LVA), Lituania (LTU), Malasia (MYS), Macedonia del Norte (MKD), Polonia (POL), Portugal (PRT), Rusia (RUS), Eslovaquia (SVK), Eslovenia (SVN), Suecia (SWE).
2 (Grupo Verde)	Aruba (ABW), Andorra (AND), Emiratos Árabes Unidos (ARE), Samoa Americana (ASM), Antigua y Barbuda (ATG), Bahréin (BHR), Bahamas (BHS), Bermuda (BMU), Brunéi (BRN), Gabón (GAB), Granada (GRD), Groenlandia (GRL), Guam (GUM), Isla de Man (IMN), San Cristóbal y Nieves (KNA), Kuwait (KWT), Líbano (LBN), Libia (LBY), Santa Lucía (LCA), Curazao (CUW), Islas Caimán (CYM), Liechtenstein (LIE), Macao (MAC), Mónaco (MCO), Montenegro (MNE), Nueva Caledonia (NCL), Nauru (NRU), Omán (OMN), Palau (PLW), Polinesia Francesa (PYF), Catar (QAT), San Marino (SMR), Sint Maarten (SXM), Islas Turcas y Caicos (TCA), Trinidad y Tobago (TTO), Islas Vírgenes de EE. UU. (VIR).

Figura 8

Dendrograma de Clustering basado en DTW



5.3.2.3. Enfoque en el Grupo Naranja para Análisis Posteriores

Tras excluir el Grupo Verde, se aplicó un segundo nivel de *clustering* jerárquico mediante DTW multidimensional al conjunto de países MIT y desarrollados, con el objetivo de lograr una segmentación más precisa entre trayectorias económicas comparables. Esta clasificación refinada permitió identificar subgrupos homogéneos en función de dinámicas estructurales y niveles de desarrollo, mejorando la coherencia analítica del análisis posterior.

El dendrograma resultante definió seis *clusters* (C1 a C6), con características económicas diferenciadas. Mientras que algunos grupos (como C2 y C5) agrupan economías con instituciones sólidas y alto desarrollo productivo, otros (como C1 y C4) reflejan estructuras más frágiles, dependencia externa y menor estabilidad institucional. Se observó además la presencia de *outliers* en ciertos *clusters*, como Venezuela y Turquía, lo que sugiere posibles limitaciones en la capacidad de algunas variables para reflejar adecuadamente casos específicos.

La segmentación muestra también un componente geográfico, aunque no determinante, ya que países de distintas regiones comparten patrones estructurales comunes. Esto refuerza la idea de que las similitudes económicas trascienden la proximidad regional y responden a factores estructurales.

Tabla 5

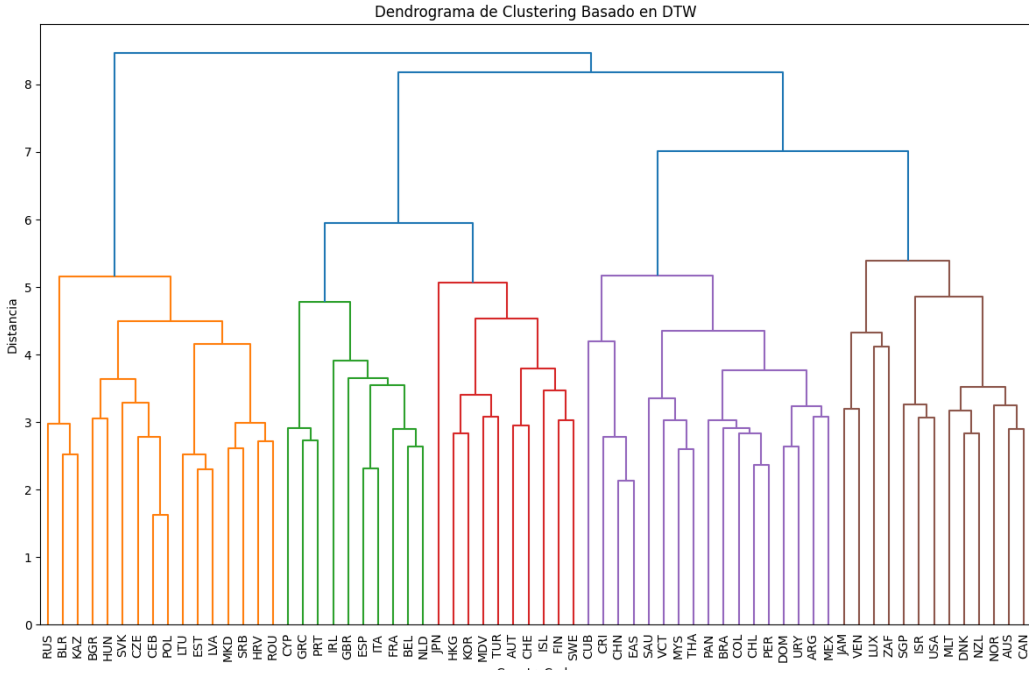
Clasificación Cluster segundo nivel

Cluster	Países
C1	Bulgaria (BGR), Bielorrusia (BLR), Europa Central y los Balcanes (CEB), República Checa (CZE), Estonia (EST), Croacia (HRV), Hungría (HUN), Kazajistán (KAZ), Lituania (LTU), Letonia (LVA), Macedonia del Norte (MKD), Polonia (POL), Rumania (ROU), Rusia (RUS), Serbia (SRB), Eslovaquia (SVK).
C2	Bélgica (BEL), Chipre (CYP), España (ESP), Francia (FRA), Reino Unido (GBR), Grecia (GRC), Irlanda (IRL), Italia (ITA), Países Bajos (NLD), Portugal (PRT).
C3	Austria (AUT), Suiza (CHE), Finlandia (FIN), Hong Kong (HKG), Islandia (ISL), Japón (JPN), Corea del Sur (KOR), Maldivas (MDV), Suecia (SWE), Turquía (TUR).
C4	Argentina (ARG), Brasil (BRA), Chile (CHL), China (CHN), Colombia (COL), Costa Rica (CRI), Cuba (CUB), República Dominicana (DOM), Asia Oriental y Sudeste Asiático (EAS), México (MEX), Malasia (MYS), Panamá (PAN), Perú (PER), Arabia Saudita (SAU), Tailandia (THA), Uruguay (URY), San Vicente y las Granadinas (VCT).
C5	Australia (AUS), Canadá (CAN), Dinamarca (DNK), Israel (ISR), Jamaica (JAM), Luxemburgo (LUX), Malta (MLT), Noruega (NOR), Nueva Zelanda (NZL), Singapur (SGP), Estados Unidos (USA), Venezuela (VEN), Sudáfrica (ZAF).

Nota. C1 (Europa del Este y Asia Central en desarrollo): economías pequeñas con estructuras acotadas, alta dependencia de recursos naturales, turismo o servicios financieros *offshore*; C2 (Europa Occidental diversificada): economías desarrolladas con mercados altamente diversificados (manufactura, tecnología, turismo, finanzas), aunque con heterogeneidad en estabilidad fiscal y crecimiento (p. ej., Italia, Grecia); C3 (avanzadas y especializadas): países con liderazgo en innovación/tecnología y alto bienestar (Japón, Corea del Sur, Suiza, Suecia, Finlandia, Islandia); C4 (emergentes y en desarrollo): América Latina, Asia y Oriente Medio con crecimiento aún apoyado en materias primas, manufactura o turismo; C5 (altamente desarrolladas y globalizadas): altos ingresos, estabilidad financiera e integración comercial, con diversificación y liderazgo sectorial (EE. UU., Canadá, Australia) y polos financieros/comerciales (Singapur, Luxemburgo).

Figura 9

Dendrograma de Clustering Basado en DTW segundo nivel con variables combinadas



Nota. Para más detalles del cluster DTW Multidimensional en Segundo sin variables combinadas, ver Apéndice A, sección L: " DTW Multidimensional en Segundo Nivel: Exclusión del Cluster Verde y Sin Variables Combinadas

5.3.3. Conclusiones y Comparación con el Clustering Anterior: Mejoras o empeoramiento en la *clusterización*

El DTW de segundo nivel entrega agrupaciones coherentes e interpretables, con mayor granularidad y mejor cohesión respecto del primer nivel, manteniendo patrones regionales y sectoriales plausibles para economías avanzadas, emergentes y en transición; en conjunto, esta validación confirma que la muestra y la parametrización son adecuadas para avanzar al esquema unidimensional, donde se evaluará con mayor transparencia la contribución de cada dimensión y se señalarán casos en posible situación de MIT.

Tabla 7

Cuadro comparativo de resultados con y sin variables combinadas.

Aspecto	Con variables combinadas	Sin variables combinadas
Claridad en la diferenciación entre economías desarrolladas y emergentes	✓ Mejor segmentación de economías avanzadas y emergentes.	✗ Menos diferenciación, economías emergentes en C1.
Clasificación de países en transición (MIT, Europa del Este)	✓ Europa del Este y Asia Central en un grupo claro (C4).	✓ Mantiene la agrupación en C4, aunque sin grandes mejoras.
Casos atípicos (Venezuela, Turquía, Sudáfrica, Arabia Saudita, Jamaica)	✗ Venezuela y Turquía siguen mal clasificados.	✗ Persisten los errores con Venezuela, pero ahora Jamaica y Arabia Saudita aparecen en C5.
Coherencia con estructuras económicas reales	✓ Mayor precisión al reflejar diferencias estructurales entre economías.	✗ Depende más de valores absolutos, lo que genera agrupaciones menos explicativas.

5.3.3.1. Limitaciones del multidimensional

El *clustering* multidimensional ofrece una agrupación coherente, pero no permite identificar con precisión casos en posible MIT**:** la agregación simultánea de muchas variables induce colinealidad y dominancia (PBI per cápita), y la interpretabilidad es limitada en casos fronterizos; por ello se complementa con el enfoque unidimensional que explicita la contribución de cada dimensión y mejora la transparencia del diagnóstico frente al *benchmark* por umbrales.

5.3.4. Clustering DTW Unidimensional: Análisis de Estadísticos Descriptivos y Significancia Estadística

El proceso de *clustering* se lleva a cabo utilizando DTW de manera unidimensional para cada variable analizada. (i) La variable PBI per cápita (gdp) presenta una diferencia significativa entre los grupos, con una T-*Statistic* de 13.7066 y un P-*Value* extremadamente bajo (3.8070e-12), lo que indica que la separación entre *clusters* en términos de PBIpc es altamente significativa. El *Cluster* 1 tiene una media de 43,078.70 USD, con una mediana de 39,299.06 USD, mientras que el *Cluster* 2 presenta un PBI per cápita promedio de 3,723.18 USD, con una mediana de 2,205.88 USD. Está marcada diferencia valida la segmentación en términos de desarrollo económico, sugiriendo que el modelo ha capturado correctamente la brecha estructural entre economías avanzadas y en desarrollo. (ii) El índice de Gini (gini_value), también muestra diferencias significativas entre los *clusters* (T-*Statistic* de -15.4633, P-*Value* de 1.9735e-29). El *Cluster* 1 presenta una media de 32.97, con una desviación estándar de 3.33, mientras que el *Cluster* 2 tiene un valor medio de 45.04, con una dispersión mayor (desviación estándar de 5.81). (iii) La Tasa de inversión en educación (ed_value) muestra una diferencia altamente significativa (T-*Statistic* de -12.0422, P-*Value* de 4.7737e-20), con el *Cluster* 1 teniendo una media de 3.37 años de educación superior y el *Cluster* 2 con 6.14 años en promedio. Esta relación aparentemente inversa puede reflejar un sesgo en la distribución de la educación en países en desarrollo, donde una parte menor de la población accede a niveles educativos avanzados, lo que impacta en la dispersión y la mediana dentro del grupo. (iv) La Tasa de Inversión en Activos (TasaInvAct_value) presenta diferencias considerables (T-*Statistic* de -9.07, P-*Value* de 7.9189e-06), con una mediana significativamente mayor en el *Cluster* 2 (3.86 billones USD) en comparación con el *Cluster* 1 (7.92 mil millones USD). Esto indica que, si bien las economías menos

desarrolladas pueden tener un PBIpc menor, su nivel de inversión en activos es proporcionalmente alto, posiblemente debido a estrategias de desarrollo económico o dependencia de inversiones extranjeras. (v) La tasa de crecimiento poblacional (Tasa_Crec_Pob_value) refuerza la segmentación del modelo, con una T-*Statistic* de -19.6396 y un P-*Value* de 6.3001e-43, lo que implica diferencias altamente significativas. El *Cluster* 1 tiene un crecimiento poblacional casi nulo (0.27% en promedio), mientras que el *Cluster* 2 muestra un crecimiento poblacional considerablemente mayor (2.17% en promedio), lo que es consistente con patrones demográficos globales en las economías avanzadas. (vi) La variable pobreza (Pobreza_value) muestra una clara diferenciación entre los *clusters*, con un T-*Statistic* de 18.8116 y un P-*Value* de 2.2822e-24, indicando que la distribución de la pobreza es marcadamente distinta en ambos grupos. Mientras que el *Cluster* 1 presenta una media de 51.04% de población en pobreza, el *Cluster* 2 tiene una media significativamente menor de 8.48%, lo que confirma la validez del *clustering* en términos de bienestar económico. (vii) El desempleo (Desempleo_value), presenta una tasa de 4.83% en Cluster 1, mientras que el *Cluster* 2 muestra un valor más elevado de 13.48%. Aquí la significancia estadística es alta (T-*Statistic* de -12.9689, P-*Value* de 7.5942e-20), reforzando la idea de que las economías menos desarrolladas tienen mayores dificultades en la generación de empleo. (viii) La deuda pública (DeudaPub_value) revela diferencias notables entre los *clusters* (T-*Statistic* de -9.9003, P-*Value* de 1.6704e-13), con un valor promedio de 2.59 en el *Cluster* 1 y 6.98 en el *Cluster* 2, lo que sugiere que las economías en desarrollo tienen una mayor dependencia de financiamiento externo o menores niveles de sostenibilidad fiscal. (ix) La Tasa de Innovación (TasaInn_value) presenta una diferencia altamente significativa (T-*Statistic* de 13.7643, P-*Value* de 1.1927e-11), con el *Cluster* 1 mostrando un promedio de 2.53, mientras que el *Cluster* 2 tiene una media de 0.47. Esto es consistente con la realidad económica global, donde las economías más desarrolladas presentan mayores niveles de inversión en innovación y tecnología. (x) El crecimiento económico (Indicador_Crecimiento) muestra una diferencia sustancial entre los *clusters* (T-*Statistic* de 13.718, P-*Value* de 3.7757e-12), con el *Cluster* 1 exhibiendo una media de 14,360.91 y el *Cluster* 2 con 1,233.26. Esto sugiere que los países en el *Cluster* 1 han experimentado un desarrollo sostenido en términos de crecimiento del PBIpc, mientras que los países en el Cluster 2 aún presentan dificultades para consolidar un crecimiento económico estable. (xi) El Indicador de Flujo de Capital (Indicador_Flujo_Capital) muestra diferencias

significativas (T-*Statistic* de -8.9853, P-*Value* de 8.5732e-06), lo que indica que el grado de apertura de la economía es un factor diferenciador clave entre los *clusters*. Las economías más integradas al comercio global presentan valores significativamente mayores en comparación con aquellas más cerradas o dependientes de recursos naturales.

Tabla 8

Resultados del Clustering DTW Unidimensional: Estadísticos descriptivos y pruebas de significancia

Variable	Media Cluster 1	Media Cluster 2	Diferencia de Medias	Significancia (P-Value)
PIB per cápita (gdp)	43,078.70	3,723.18	39,355.52	3.81×10^{-12}
Coeficiente de Gini (gini_value)	32.98	45.05	−12.07	1.97×10^{-29}
Educación (ed_value)	3.37	6.14	−2.78	4.77×10^{-20}
Tasa de inversión activos / Formación bruta de capital fijo (TasaInvAct_value) **(en millones)**	71,900.22	3,783,921.28	−3,712,021.06	7.92×10^{-6}
Crecimiento poblacional (Tasa_Crec_Pob_value)	0.27	2.17	−1.90	6.30×10^{-43}
Término Intercambio comercial (Term_Inter_value) **(en millones)**	562.71	1,418,389.30	−1,417,826.59	–
Nivel de pobreza (Pobreza_value)	51.04	8.48	42.57	2.28×10^{-24}
Tasa de desempleo (Desempleo_value)	4.83	13.48	−8.65	7.59×10^{-20}
Deuda pública (DeudaPub_value)	2.59	6.99	−4.40	1.67×10^{-13}
Índice de innovación (TasaInn_value)	2.53	0.47	2.06	1.19×10^{-11}
Diversificación tecnológica (DivEcoTec_value)	0.87	−0.94	1.81	5.17×10^{-33}
Crecimiento económico (TasaCrecimientoEcono_value)	3.64	2.85	0.79	2.22×10^{-2}
Indicador de crecimiento (Indicador_Crecimiento) **(en millones)**	14.36	1.23	13.13	3.78×10^{-12}
Indicador de desigualdad y pobreza (Indicador_Desigualdad_Pobreza)	2.24	8.90	−6.66	9.15×10^{-34}
Indicador Flujo de Capital (Indicador_Flujo_Capital) **(en millones)**	23,571.19	1,347,878.01	−1,324,306.82	8.57×10^{-6}

Nota. Las variables marcadas con (en millones) fueron transformadas para facilitar su interpretación. Los valores originales provienen de bases de datos como el Banco Mundial y corresponden a cifras agregadas. En términos de la evaluación de la significancia estadística, la presencia de P-*Values* extremadamente bajos en la mayoría de las variables confirma que la segmentación es estadísticamente significativa. Sin embargo, ciertos

indicadores, como la inversión en activos y la deuda pública, presentan valores atípicos y alta dispersión dentro de los *clusters*, lo que sugiere la posibilidad de ajustar la metodología de agrupamiento.

5.3.4.1. Construcción de grilla de Clustering mediante clasificación DTW Unidimensional

A partir del *clustering* DTW unidimensional aplicado a variables económicas relevantes, se elaboró una grilla dicotómica que clasifica a cada país según su similitud con economías desarrolladas (valor 1) o con países en la Trampa de Ingreso Medio (valor 0). Esta matriz fue integrada mediante una moda ponderada por país, utilizando los coeficientes de importancia obtenidos con *Random Forest.* El resultado es una clasificación final que sintetiza múltiples dimensiones estructurales, brindando una visión robusta sobre la pertenencia relativa de los países al desarrollo avanzado o a la MIT.

Tabla 9

Grilla dicotómica

País	PIB	Inversión	Crec. Pobl.	Desemp.	Innov.	Educación	Gini
ARG	0	1	0	0	0	0	0
CHL	1	1	0	1	1	0	0
BRA	0	0	0	0	0	0	0
USA	1	1	1	1	1	1	1
BEL	1	1	1	1	1	1	1
AUS	1	1	1	1	1	1	1
CHE	1	1	1	1	1	1	1
DEU	1	1	1	1	1	1	1
ESP	1	0	1	1	1	1	1
FRA	1	1	1	1	1	1	1
ITA	0	0	0	1	1	1	1
JPN	1	0	0	1	1	1	1
KOR	1	1	1	1	1	1	1
CAN	1	1	1	1	1	1	1
MEX	0	1	0	0	0	0	0
COL	0	0	0	1	0	0	1
URY	1	1	0	1	1	1	1
ZAF	0	0	0	0	0	0	0

Nota. Esta tabla incluye una muestra representativa de países y no refleja la totalidad de la base de datos.

5.3.4.2. **Resultado de importancia de variables para aplicar ponderación.**

A continuación, se presentan los resultados del análisis de importancia de variables obtenidos mediante *Random Forest*. Estos resultados permiten identificar qué

dimensiones económicas tienen mayor peso en la segmentación de países, y sirven como base para asignar ponderadores diferenciados en el cálculo de la moda ponderada. Se destacan las variables más influyentes, considerando tanto su capacidad para discriminar entre *clusters* como su contribución a la segmentación estructural del conjunto de países.

5.3.4.3. Random Forest con Featured Importance.

Los resultados obtenidos muestran que ciertas variables económicas y sociales tienen un impacto significativo en la segmentación de países, mientras que otras presentan una relevancia comparativamente menor en la diferenciación de grupos. Las variables con mayor peso en la segmentación fueron la inversión en activos (TasaInvAct_value – 1109.4), la diversificación económica tecnológica (DivEcoTec_value – 1142.3) y el crecimiento económico (TasaCrecimientoEcono_value – 1054.8). Estos resultados sugieren que factores vinculados a la estructura productiva, la innovación tecnológica y el dinamismo económico son altamente determinantes en la formación de clusters, diferenciando a países con capacidades productivas más sofisticadas y sostenidas en el tiempo. En un segundo nivel de impacto, aunque todavía elevado, se destacan variables como la tasa de innovación (TasaInn_value – 967.9), la tasa arancelaria (TasaAran_value – 993.7), y la diversificación económica por inversión (DivEcoInv_value – 976.5). También adquieren relevancia indicadores compuestos como el Indicador de Crecimiento (917.9) y el Indicador de Desigualdad y Pobreza (888.2), así como el PBI per cápita (gdp – 897.5). Esto indica que, aunque los factores sociales siguen siendo importantes, aspectos estructurales de apertura, inversión y productividad explican una parte sustancial de la heterogeneidad entre países.

En la categoría de impacto moderado, se ubicaron variables como la educación (ed_value – 800.4), el desempleo (Desempleo_value – 696.1), y el Indicador de Flujo de Capital (762.6). A pesar de su reconocida relevancia en el desarrollo, su menor peso relativo en esta segmentación puede deberse a su correlación con otras variables más estructurales o a una menor variabilidad entre países en desarrollo y desarrollados.

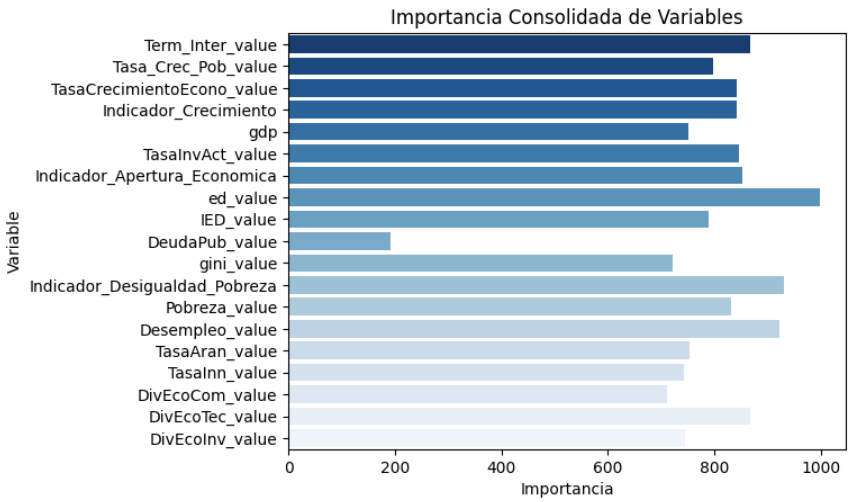
Por otro lado, la deuda pública (DeudaPub_value – 426.1) resultó ser la variable con menor incidencia en la formación de clusters, lo que sugiere que los niveles de endeudamiento no generan una diferenciación clara entre los países analizados. Es posible que distintas economías, independientemente de su nivel de desarrollo, mantengan

estructuras de deuda similares, disminuyendo su capacidad discriminante. Dado que el enfoque DTW ya captura la evolución histórica de las principales variables, los resultados de importancia deben interpretarse considerando no solo su peso estático en la segmentación, sino también las trayectorias temporales que han seguido dichas variables en cada país.

En conclusión, la segmentación de países no está determinada exclusivamente por el nivel de ingreso per cápita, sino que responde a una combinación de factores estructurales, sociales y económicos. La inversión productiva, la diversificación tecnológica y el crecimiento económico surgen como los principales ejes de diferenciación, sugiriendo que las políticas públicas que promuevan estos pilares pueden ser decisivas para el desarrollo sostenido. La incorporación de *Random Forest* permite la evaluación de la importancia de variables, a través de un enfoque flexible que identifica patrones complejos en las trayectorias de desarrollo de los países.

Figura 10

Relevancia en porcentaje de cada variable en la asignación de Clusters según Random Forest con Featured Importance



5.3.4.4. Evaluación del Clustering Unidimensional con Ponderación para la Identificación de Economías en la Trampa de Ingresos Medios (MIT)

El análisis del *clustering* unidimensional con ponderación permitió una segmentación que, en términos generales, distingue de manera efectiva entre países desarrollados y

aquellos atrapados en la trampa de ingresos medios (MIT). Esta clasificación, construida a partir de agrupamientos independientes por variable y luego integrados mediante una moda ponderada, incorpora múltiples dimensiones económicas y sociales —como desigualdad, inversión, diversificación y crecimiento— lo que ofrece una visión más integral que la basada exclusivamente en PBI per cápita. Si bien la separación entre grupos es clara en la mayoría de los casos, persisten ciertos solapamientos y asignaciones discutibles que reflejan trayectorias económicas atípicas o transicionales.

Al comparar los resultados con los criterios tradicionales utilizados por organismos internacionales, se observa que esta metodología captura mejor las limitaciones estructurales que impiden el desarrollo sostenido, identificando a países que, pese a tener ingresos relativamente altos, presentan estancamiento productivo, alta desigualdad o baja innovación. Este enfoque evidencia que el desarrollo económico trasciende el ingreso per cápita, requiriendo una mirada más amplia. Sin embargo, la heterogeneidad interna de los casos MIT sugiere la necesidad de futuras segmentaciones más finas. Para mejorar la precisión, se sugiere complementar esta aproximación con variables institucionales y de movilidad social, que permitan capturar aspectos cualitativos del desarrollo que no siempre se reflejan en indicadores cuantitativos.

Tabla 10

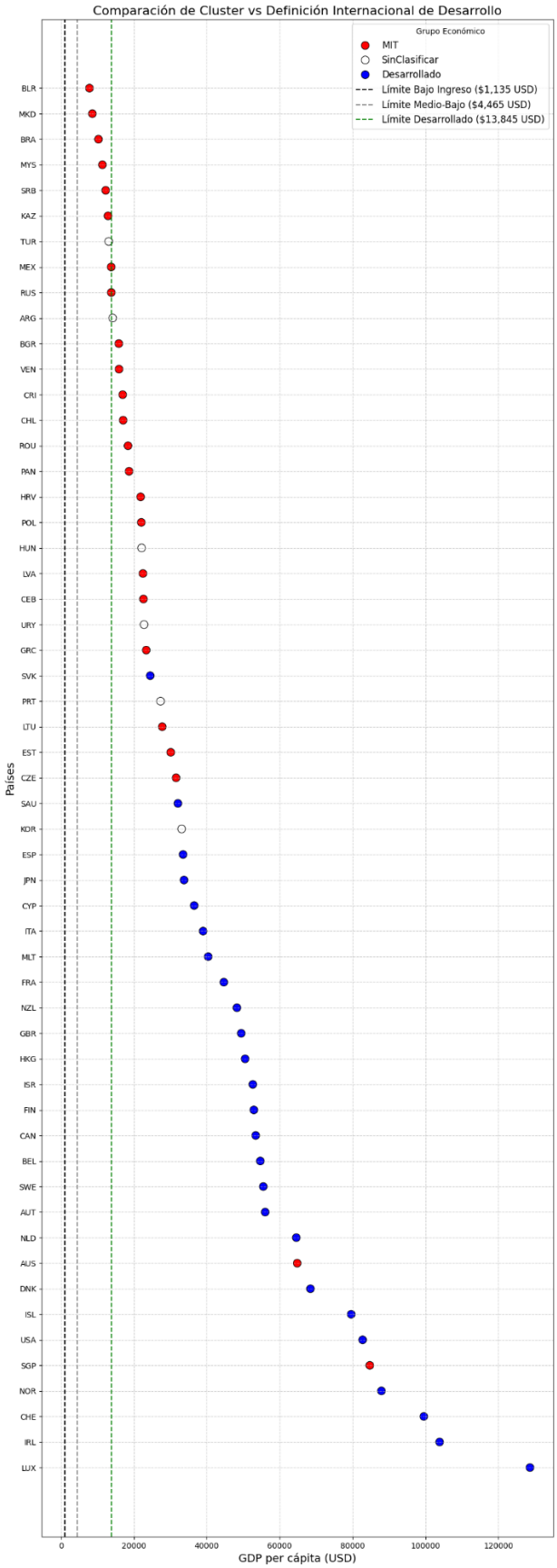
Grilla dicotómica por país y variable (1 = desarrollado, 0 = ingreso medio)

País	PIB (5.49)	Inv. Act. (5.56)	Crec. Pobl. (5.72)	Desemp. (5.66)	Innov. (4.53)	Educación (5.00)	Gini (5.05)	Moda Ponderada
ARG	0	1	0	0	0	0	0	**0**
CHL	1	1	0	1	1	0	0	**1**
BRA	0	0	0	0	0	0	0	**0**
USA	1	1	1	1	1	1	1	**1**
BEL	1	1	1	1	1	1	1	**1**
AUS	1	1	1	1	1	1	1	**1**
CHE	1	1	1	1	1	1	1	**1**
DEU	1	1	1	1	1	1	1	**1**
ESP	1	0	1	1	1	1	1	**1**
FRA	1	1	1	1	1	1	1	**1**
ITA	0	0	0	1	1	1	1	**1**
JPN	1	0	0	1	1	1	1	**1**
KOR	1	1	1	1	1	1	1	**1**
CAN	1	1	1	1	1	1	1	**1**
MEX	0	1	0	0	0	0	0	**0**
COL	0	0	0	1	0	0	1	**0**
URY	1	1	0	1	1	1	1	**1**
ZAF	0	0	0	0	0	0	0	**0**

Nota. Esta tabla incluye una muestra representativa de países y no refleja la totalidad de la base de datos. Las cifras entre paréntesis indican la ponderación relativa de cada variable según el modelo *Random Forest* (\%).

Figura 12

Comparación de Clusters vs Definición Internacional de Desarrollo



5.3.4.5. Interpretación y Conclusiones

El enfoque propuesto —pre-filtrado multidimensional, *clustering* DTW unidimensional por variable y agregación mediante moda ponderada con pesos obtenidos por *Random Forest*— ofrece una lectura consistente de la MIT frente al *benchmark* de umbrales por PBI per cápita. En la Figura: XX se observa una separación visual clara entre economías clasificadas como desarrolladas y aquellas con patrones propios de MIT; la Tabla XX muestra que variables como innovación, educación, desigualdad e inversión en activos concentran el mayor poder discriminante en la clasificación final.

Los resultados confirman la hipótesis principal: los países identificados como MIT comparten patrones estructurales de menor diversificación, mayor desigualdad y baja intensidad en I+D, incluso cuando su PBI per cápita se aproxima a umbrales de "alto ingreso". La clasificación integrada por moda ponderada sintetiza estas dimensiones y evita depender de un único indicador. Asimismo, la hipótesis complementaria se ve respaldada: el DTW permite capturar trayectorias y desfases temporales que no afloran en cortes transversales, identificando países con evoluciones similares, aunque con calendarios distintos de adopción tecnológica o reformas.

El contraste con el criterio internacional por umbrales indica una alta concordancia global y, simultáneamente, revela discrepancias informativas: casos con PBI per cápita relativamente elevado, pero con Gini alto, baja I+D o crecimiento poblacional que tensiona el mercado laboral, quedan clasificados como MIT; a la inversa, economías con ingreso medio-alto, pero con trayectorias sostenidas de innovación y capital humano se aproximan al grupo desarrollado. Estas discrepancias no emergen como "errores" del método, sino como evidencia de que el ingreso es condición necesaria pero no suficiente para el desarrollo sostenido.

En suma, la evidencia respalda que (i) un esquema multidimensional + dinámico representa mejor la realidad estructural que el umbral de ingreso per cápita aislado; (ii) la ponderación basada en importancia de variables mejora la coherencia de la etiqueta final; y (iii) existen sub-patrones dentro de MIT (p. ej., "MIT con renta alta y desigualdad alta" versus "MIT con rezago en I+D y capital humano") que justifican segmentaciones posteriores. Con ello, el estudio confirma sus hipótesis y ofrece una clasificación reproducible y estable para guiar análisis comparativos y diagnósticos de política.

6. Discusión

Los hallazgos se alinean con la crítica clásica al enfoque de umbrales: el PBI per cápita captura nivel, no estructura ni trayectoria. Al integrar K-medias con DTW y agregar por moda ponderada según importancia de variables, la clasificación converge con el *benchmark* en la mayor parte de los casos, pero enriquece la lectura allí donde el ingreso "miente": economías con estancamiento productivo, alta desigualdad o debilidad en I+D/educación no muestran condiciones suficientes para transitar al desarrollo, aun cuando superen umbrales de renta. Este resultado dialoga con la literatura que enfatiza diversificación, instituciones e innovación como motores del catch-up, y con la noción de cohesión y separación (Diday) aplicada a trayectorias temporales.

El uso de DTW unidimensional aporta estabilidad temporal y legibilidad. La homologación de etiquetas asegura consistencia semántica de la grilla, mientras que la ponderación derivada de *Random Forest* privilegia dimensiones realmente informativas. Esto reduce el riesgo de sobre-ajuste a un indicador y atenúa la multicolinealidad.

Las discrepancias con el umbral de ingreso deben interpretarse como diagnóstico estructural: (i) economías con PBIpcpc alto + Gini alto y/o mercados laborales frágiles: requieren paquetes de productividad + redistribución (competencia, formalización, impuestos/protecciones bien diseñados); (ii) economías con rezago en I+D/educación: piden políticas de innovación (I+D empresarial, vínculos universidad-empresa) y capital humano; (iii) economías con demografía presionante: necesitan empleo juvenil, participación femenina y políticas de habilidades. Estas líneas son coherentes con los patrones observados en la Tabla 10 y con la literatura sobre complejidad/innovación.

El método es descriptivo/explicativo, no causal. La cobertura histórica de datos condiciona el poder discriminante de algunas series; aunque el z-score facilita comparabilidad, las brechas de datos y decisiones sobre ventanas DTW pueden afectar casos fronterizos. Dado que la MIT es heterogénea, conviene profundizar en sub-clusterizaciones dentro de MIT (p. ej., "renta alta-desigualdad alta" vs "baja I+D"), evaluar sensibilidades de pesos y parámetros DTW y/o, una métrica de concordancia con el benchmark (p. ej., % de match o Kappa). Todo ello puede hacerse con los insumos ya generados en el paper (grilla, etiquetas, benchmark), sin requerir nueva recolección de datos.

Bibliografía

Acemoglu, D., & Robinson, J. A. (2012). *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. Crown Business.

Aghion, P., & Howitt, P. (1998). *Endogenous Growth Theory*. MIT Press.

Barro, R. J. (1991). *Economic Growth in a Cross Section of Countries*. *Quarterly Journal of Economics, 106*(2), 407-443. https://doi.org/10.2307/2937943

Breiman, L. (2001). *Random Forests*. Machine Learning, 45(1), 5–32. Lundberg, S. M., & Lee, S. I. (2017). *A Unified Approach to Interpreting Model Predictions*. In Advances in Neural Information Processing Systems (pp. 4765–4774).

Christensen, C. M. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business Review Press.

Diday, E. (1971). La méthode des nuées dynamiques. *Revue de Statistique Appliquée, 19(2), 19-33.*

Easterly, W. (2001). *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*. MIT Press.

Felipe, J., Abdon, A., & Kumar, U. (2012). Tracking the middle-income trap: What is it, who is in it, and why? *Levy Economics Institute, Working Paper, (715).*

Glaeser, E. L. (2011). *The Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. Penguin Press.

Hausmann, R., & Klinger, B. (2006). *The Diversification Advantage*. *Center for International Development Working Paper No. 128*. Harvard University. https://doi.org/10.2139/ssrn.939649

Krugman, P. (1989). *Debt and Development*. *International Economic Review, 30*(1), 3-21. https://doi.org/10.2307/2526306

Lederman, D., & Maloney, W. F. (2007). *Natural Resources: Neither Curse Nor Destiny*. Stanford University Press.

MacQueen, J. (1967). Multivariate observations. In *Proceedings ofthe 5th Berkeley Symposium on Mathematical Statisticsand Probability (Vol. 1, pp. 281-297).*

Mounk, Y. (2017). *The Rise and Fall of Nations: Forces of Change in the Post-Crisis World*. W.W. Norton & Company.

Piketty, T. (2014). *Capital in the Twenty-First Century*. Harvard University Press.

Sachs, J. D. (2005). *The End of Poverty: Economic Possibilities for Our Time*. Penguin Press.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing, 26(1), 43-49.

Stiglitz, J. E. (2002). *Globalization and Its Discontents*. W.W. Norton & Company.

Stiglitz, J. E. (2012). *The Price of Inequality: How Today's Divided Society Endangers Our Future*. W.W. Norton & Company.

Wilkinson, R., & Pickett, K. (2009). *The Spirit Level: Why Greater Equality Makes Societies Stronger*. Bloomsbury Press.

Apéndice

Apéndice A:

a) Series y fuentes de datos.

Variable	Literatura	Fuente de Datos	Fecha Inicio	Fecha Inicio	Fecha de Inicio de Datos para Cada País	Enlace
PBI per cápita	- "Why Nations Fail" (Acemoglu y Robinson) - "The Rise and Fall of Nations" (Mounk, Y. (2017)) - "The End of Poverty" (Sachs)	Banco Mundial: GDP per capita	1960	2024	Datos disponibles desde 1960 para todos los países mencionados en Banco Mundial.	Descargar Datos
Índice de Gini	- "The Spirit Level" (Wilkinson, R., & Pickett, K. (2009)) - "The Price of Inequality" (Stiglitz, J. E. (2012))	Banco Mundial: Gini Index	1960	2024	Datos disponibles desde 1960 para todos los países mencionados en Banco Mundial.	Descargar Datos
Tasa de educación	- "Capital in the Twenty-First Century" (Piketty) - "The Rise and Fall of Nations" (Mounk, Y. (2017)) - "The End of Poverty" (Sachs)	Banco Mundial: Years of schooling UIS (para más detalles)	Aprox. 1970	2024	Datos disponibles desde aproximadamente 1970 para los países mencionados en Banco Mundial y UIS.	Descargar Datos (Banco Mundial) Descargar Datos (UIS)
Tasa de inversión activos (Formación bruta de capital)	- "Growth Theory" (Barro, R. J. (1991)) - "The End of Poverty" (Sachs)	Banco Mundial: Gross capital formation	1960	2024	Datos disponibles desde 1960 para todos los países mencionados en Banco Mundial.	Descargar Datos
Tasa de crecimiento de población	- "The Rise and Fall of Nations" (Mounk, Y. (2017)) - "The End of Poverty" (Sachs) - "Globalization and Its Discontents" (Stiglitz, J. E. (2002))	Banco Mundial: Population growth	1960	2024	Datos disponibles desde 1960 para todos los países mencionados en Banco Mundial.	Descargar Datos
Tasa de urbanización	- "The Urbanization of the World" (Glaeser) - "Cities and the Wealth of Nations" (Glaeser)	Banco Mundial: Urban population	1960	2024	Datos disponibles desde 1960 para todos los países mencionados en Banco Mundial.	Descargar Datos
Término Intercambio comercial	- "The End of Poverty" (Sachs) - "The Rise and Fall of Nations" (Mounk, Y. (2017))	Banco Mundial: Labor productivity	2000	2024	Datos disponibles desde 2000 para todos los países mencionados en Banco Mundial.	Descargar Datos
Tasa de pobreza	- "The End of Poverty" (Sachs) - "Development as Freedom" (Sen)	Banco Mundial: Poverty headcount ratio	1990	2024	Datos disponibles desde 1990 para todos los países mencionados en Banco Mundial.	Descargar Datos
Tasa de desempleo	- "The Rise and Fall of Nations" (Mounk, Y. (2017)) - "The End of Poverty" (Sachs)	Banco Mundial: Unemployment rate	1991	2024	Datos disponibles desde 1991 para todos los países mencionados en Banco Mundial.	Descargar Datos
Tasa de inversión extranjera	- "Globalization and Its Discontents" (Stiglitz, J. E. (2002)) - "The Rise and Fall of Nations" (Mounk, Y. (2017))	Banco Mundial: Foreign direct investment	1990	2024	Datos disponibles desde 1990 para todos los países mencionados en Banco Mundial.	Descargar Datos
Tasa de deuda pública	- "The Debt Crisis" (Krugman, P. (1989)) - "Debt and Development" (Easterly, W. (2001))	Banco Mundial: Debt service	1990	2024	Datos disponibles desde 1990 para todos los países mencionados en Banco Mundial.	Descargar Datos
Tasa de innovación	- "The Innovator's Dilemma" (Christensen, C. M. (1997)) - "Innovation and Growth" (Aghion y Howitt)	Banco Mundial: Research and development expenditure	1996	2024	Datos disponibles desde 1996 para todos los países mencionados en Banco Mundial.	Descargar Datos
Diversificación económica	- "The Diversification Advantage" (Hausmann y Klinger) - "Economic Diversification and Development" (Lederman, D., & Maloney, W. F. (2007))	Banco Mundial: Export diversification	2000	2024	Datos disponibles desde 2000 para todos los países mencionados en Banco Mundial.	Descargar Datos
Tasa Arancelaria	-"The Globalization Paradox" - Dani Rodrik -"Free Trade Under Fire" - Douglas Irwin -"Why Nations Fail" - Acemoglu &	Banco Mundial	1990	2024	Datos disponibles desde 1990 para todos los países mencionados, en el Banco Mundial	Descargar Datos

Apéndice B: Ejemplo de Cálculo de Distancias y Representación JSON

Como ejemplo, supongamos un análisis en el que se evalúan distancias DTW para el atributo "tasa de crecimiento poblacional" (`Tasa_Crec_Pob_value`) entre segmentos. Los resultados se almacenaron en un formato similar al siguiente:


```
{
    "Tasa_Crec_Pob_value": {
        "(1, 2)": 4.541,
        "(1, 3)": 3.704,
        "(2, 3)": 4.715
    },
    "gini_value": {
        "(1, 2)": 6.123,
        "(1, 3)": 5.412,
        "(2, 3)": 7.345
    }
}
```


Apéndice C: Detalles de la metodología

a) Ejemplo resultado cluster entre Argentina y USA

La siguiente tabla consolida los resultados obtenidos a partir de las distintas metodologías aplicadas en el estudio, integrando tanto las clasificaciones basadas en los umbrales de la literatura como los agrupamientos generados mediante clustering simple (K-Medias y DBSCAN) y clustering dinámico (DTW). Esta consolidación permite comparar la coherencia entre enfoques y analizar la estabilidad de las clasificaciones, ofreciendo un panorama completo de la posición relativa de cada país en relación con la trampa de ingresos medios.***Tabla 1***

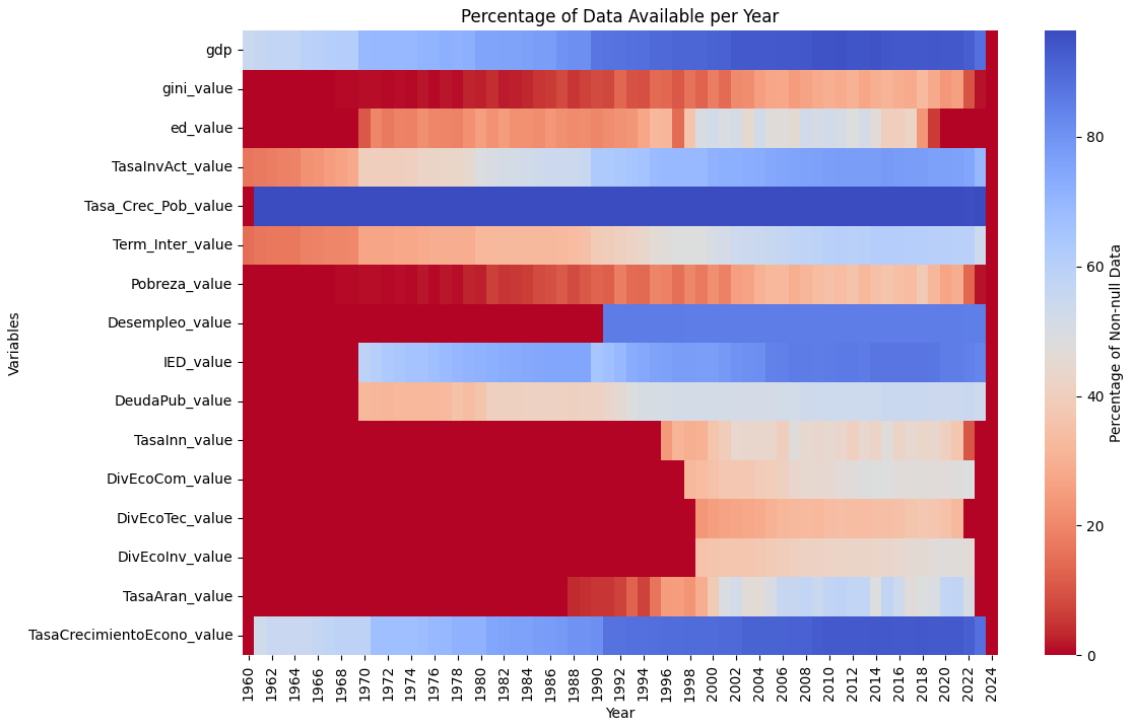
Comparación de Métodos de Clasificación MIT y Clustering DTW

CountryCode	MIT_Fixed	MIT_Hist	Cluster_PIB_Fixed	Cluster_PIB_Hist	Cluster_NoPIB_Fixed	Cluster_NoPIB_Hist
ARG	1	1	0	1	2	1
USA	0	0	1	1	1	1

Nota. ARG = Argentina; USA = Estados Unidos. MIT_Fixed representa el dato del último año con información disponible. MIT_Hist corresponde al promedio histórico de la serie. La tabla muestra la clasificación de los países según distintos métodos de clustering y segmentación MIT. Se ejemplifica la clasificación considerando todas las variables o excluyendo el PBIpc (PBIpc - No_PBIpc), lo que permite observar diferencias en los patrones de agrupamiento según el enfoque aplicado

b) Datos disponibles y nulos

Porcentaje de datos disponibles por año.

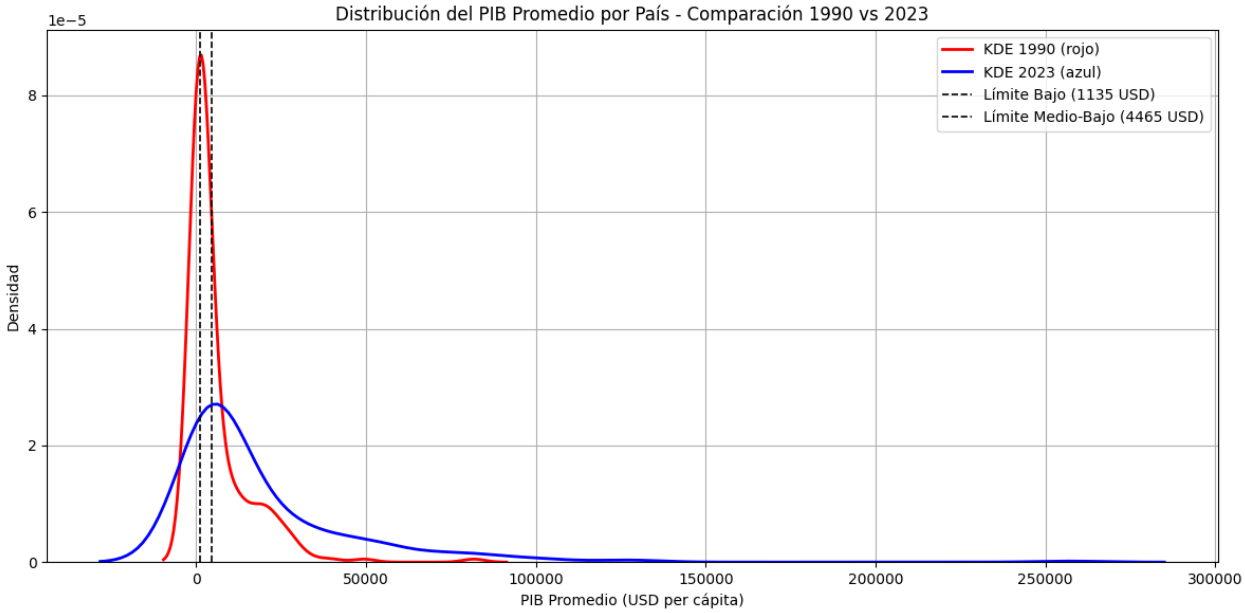


El mapa de calor destaca cómo la disponibilidad de datos mejora significativamente en las décadas más recientes, indicando una mayor calidad en los reportes contemporáneos. Estas disparidades también reflejan diferencias en la capacidad de recopilación de datos entre países, siendo más notorias en naciones en desarrollo.

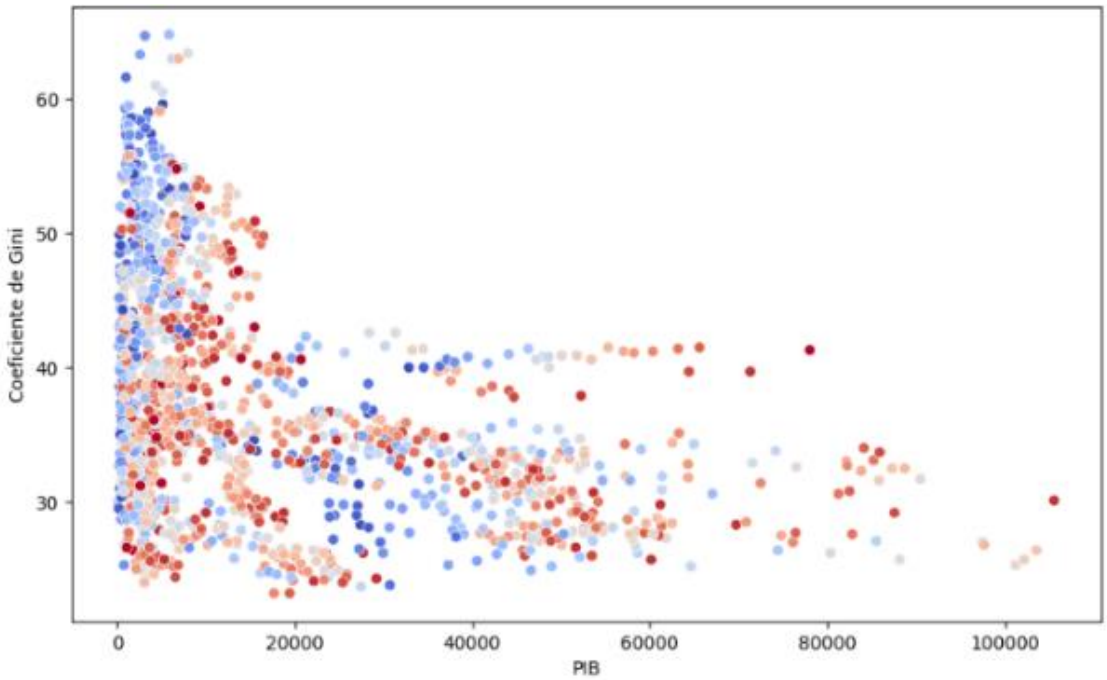
c) Relación entre PBIpc y Coeficiente de Gini

El diagrama de dispersión que relaciona el PBI per cápita con el coeficiente de Gini muestra una correlación negativa general, donde países con mayores niveles de PBIpc tienden a tener menores niveles de desigualdad. Sin embargo, los puntos dispersos indican excepciones notables, especialmente en niveles bajos de PBIpc, donde algunos países muestran desigualdades extremas. Esto refuerza la idea de que los niveles de desarrollo económico no siempre están alineados con una distribución equitativa de la riqueza, destacando la complejidad de las dinámicas económicas en países atrapados en la MIT.

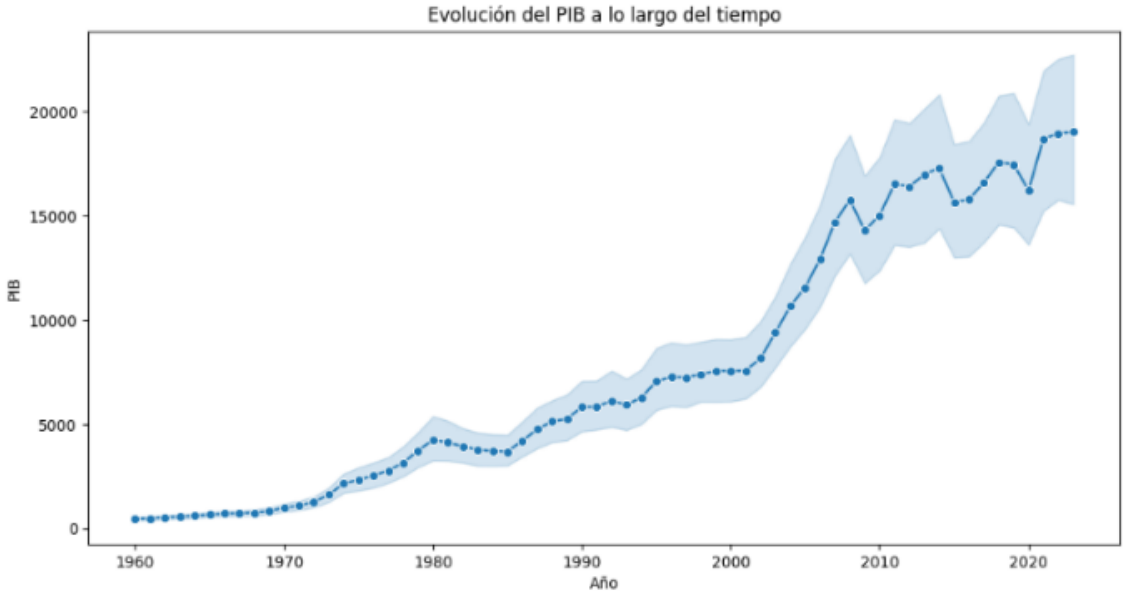
Distribución del PBIpc



Relación entre variable PBIpc y Coeficiente de Gini


Coeficiente de Gini
60
50
40
30
0
20000
40000
60000
80000
100000
PIB

Evolución del PBIpc a lo largo del tiempo



d) Herramientas y Software Utilizados

Para el desarrollo del presente estudio, se emplearon herramientas y software especializados que facilitaron tanto el procesamiento de datos como el análisis estadístico y la visualización de resultados. En particular, se utilizaron **Stata** y **Python** para la gestión y análisis de datos en panel, así como para la implementación de técnicas de **clustering** y análisis dinámico. Estas herramientas permitieron manipular bases de datos extensas y aplicar metodologías avanzadas de segmentación y análisis multivariado. La visualización de resultados se realizó mediante Python, utilizando las bibliotecas **Matplotlib** y **Seaborn**, lo que permitió generar gráficos claros, personalizados y ajustados a las necesidades específicas del análisis, facilitando la interpretación y comunicación de los hallazgos principales.

e) Proceso de normalización

El proceso de normalización aplicado tuvo como objetivo preparar las series temporales para su análisis sin excluir variables ni realizar imputaciones de datos faltantes. En primer lugar, se segmentaron las series temporales en bloques definidos por país y variable, como se mostró en el ejemplo gráfico anterior. Esta segmentación permitió organizar los datos de forma estructurada y facilitó su posterior procesamiento.

La normalización se llevó a cabo utilizando un enfoque basado en el cálculo del *z-score*. Este procedimiento ajusta cada serie de datos restando su media y dividiéndola entre su desviación estándar, asegurando que todas las variables sean comparables al estar en la misma escala. La fórmula utilizada fue:

$$z_i = \frac{x_i - \mu}{\sigma}$$

Donde:

• z_i es el valor normalizado de la observación x_i

• x_i es el valor original de la observación,

• μ es la media de la serie temporal correspondiente,

• σ es la desviación estándar de la serie temporal.

Este paso fue esencial para mitigar el impacto de las disparidades en las escalas de las diferentes variables, como el PBIpc (gdp) frente al índice de Gini ($gini_value$), y garantizar que todas las variables contribuyeran equitativamente al análisis de clustering.

La decisión de no descartar variables ni imputar valores faltantes permitió preservar la integridad de los datos originales y centrarse exclusivamente en los patrones existentes. De esta manera, el proceso de normalización facilitó la identificación de similitudes entre segmentos de series temporales sin alterar su estructura fundamental, maximizando el valor analítico del conjunto de datos.

f) Enfoque principal MIT_Hist

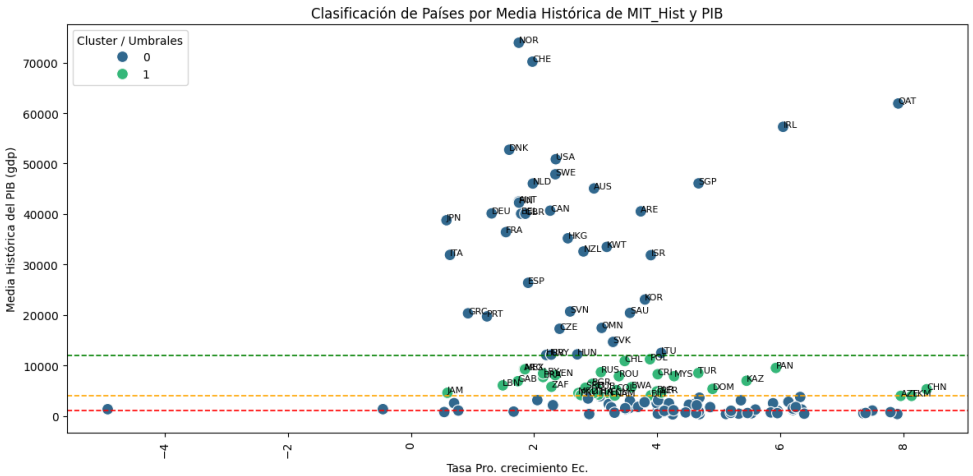
Se aplicó un modelo de *clustering K-Means* con dos grupos, utilizando variables socioeconómicas normalizadas como PBI per cápita histórico, inversión, pobreza, desigualdad, desempleo y diversificación económica. El modelo identificó un primer grupo compuesto por países con ingresos medios y desafíos estructurales persistentes

(como Argentina, Brasil o Sudáfrica), y un segundo grupo conformado por economías más desarrolladas o con mejores indicadores estructurales. El PBI per cápita histórico resultó ser la variable con mayor poder discriminante entre ambos clusters, consolidándose como proxy robusto del nivel de desarrollo.

A partir de esta segmentación general, se identificó un subconjunto de interés: países del primer cluster y países del segundo con PBIpc histórico superior a USD 12.000. Esta condición permitió acotar la muestra a países desarrollados y de ingreso medio alto, para profundizar en la identificación de casos en situación de MIT. Dentro de este grupo, algunos países como Corea del Sur, Israel o Chile muestran trayectorias destacadas, mientras que otros, pese a altos ingresos, mantienen debilidades estructurales, lo que refuerza la necesidad de incorporar múltiples dimensiones —productivas, institucionales y sociales— para evaluar adecuadamente el desarrollo económico.

Figura 4

Agrupación de países por media historia de MIT_Hist y PBIpc



Para variable MIT_Fixed: Análisis complementario

El análisis de los resultados del *clustering* basado en la clasificación de las variables MIT_Fixed y el PBIpc**,** según los umbrales de la literatura, revela una segmentación significativa en los datos. El gráfico segmenta a los países en dos *clusters* principales que reflejan patrones claros en términos de desarrollo económico.

En primer lugar, el *Cluster* 1 agrupa principalmente a países de ingreso medio, formando una franja bien definida en la distribución de los datos. Este grupo se distingue por características económicas y sociales que lo separan tanto de las economías de ingreso bajo como de las de ingreso alto. La existencia de esta franja intermedia resalta cómo los países de ingreso medio constituyen un subconjunto relativamente homogéneo y diferenciado del resto del mundo.

Por otro lado, el *Cluster* 0 es más heterogéneo, ya que incluye tanto a países de ingreso bajo como a países de ingreso alto. Los países de ingreso bajo se ubican en la parte inferior de la distribución, con valores reducidos tanto en PBIpc como en MIT_Fixed, mientras que los países de ingreso alto ocupan la parte superior, con valores elevados en ambas dimensiones. Este patrón indica que el *Cluster* 0 refleja la diversidad del resto del mundo, agrupando economías con características significativamente diferentes.

Desde una perspectiva visual, el *scatterplot* confirma esta segmentación. El *Cluster* 1 aparece como una banda intermedia en los valores del PBIpc, lo que refuerza la idea de que estos países representan un grupo transicional con características que los distinguen claramente. Por su parte, el *Cluster* 0 se divide visualmente en dos subgrupos: economías avanzadas en la parte superior y países en desarrollo en la inferior.

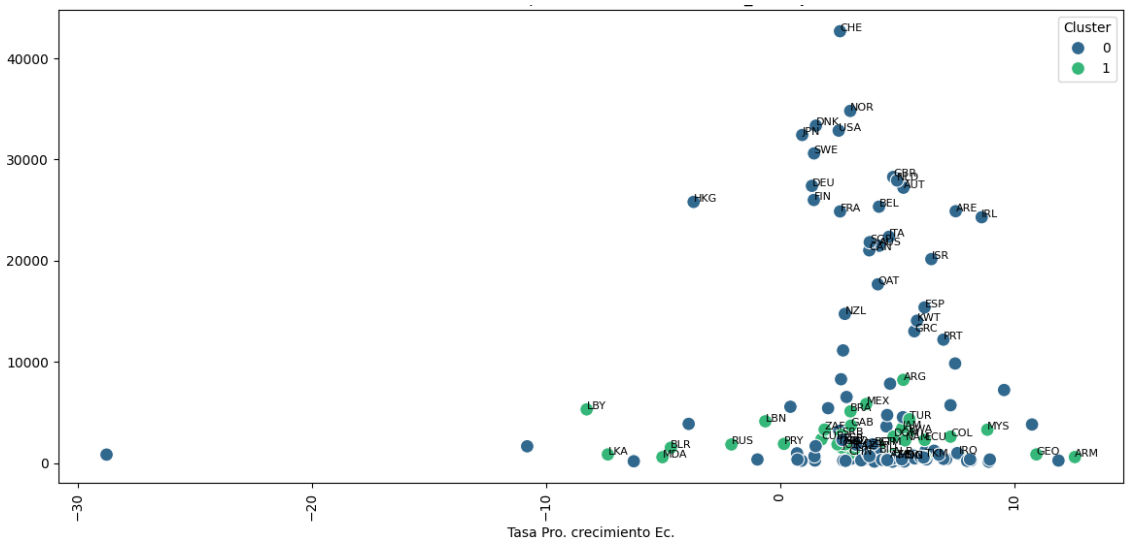
Este patrón de *clustering* destaca cómo los países de ingreso medio tienen una identidad económica única que los separa tanto de las economías avanzadas como de aquellas en desarrollo. En el contexto de estudios de desarrollo económico, esto sugiere que los países de ingreso medio pueden requerir políticas específicas para abordar desafíos y oportunidades diferentes a las de los otros dos grupos.

Para mejorar el análisis, es útil agregar variables adicionales que capturen aspectos más específicos del desarrollo económico y social, así como aplicar técnicas de *subclustering* dentro del *Cluster* 0 para diferenciar mejor los países de ingreso alto y bajo. Asimismo, índices de validación como el de Calinski-Harabasz o el coeficiente de Silueta podrían reforzar la calidad de la segmentación realizada.

En conclusión, el *clustering* realizado identifica una franja intermedia de economías de ingreso medio que actúa como un grupo transicional entre las economías de ingreso bajo y alto. Esto no solo resalta la heterogeneidad global en términos de desarrollo económico,

sino también la importancia de estudiar los factores que definen a los países de ingreso medio en un contexto más amplio.

Clasificación de países por media historia de MIT_Fixed y PBIpc



g) Para observar la determinación del Número Óptimo de Clusters basado en MIT_Hist, con y sin PBIpc.

El análisis del *clustering* basado en la media histórica de las variables de desarrollo, incluyendo el PBIpc (MIT_Hist), se confirma con los resultados del método *Elbow* y los índices de Silueta y Calinski-Harabasz. Según el método *Elbow*, el gráfico muestra una disminución significativa en la pendiente al pasar de 1 a 2 *clusters*, lo que indica una mejora sustancial en la reducción de la suma de los errores cuadráticos (SSE). Sin embargo, a partir de 3 *clusters*, la pendiente se vuelve menos pronunciada, sugiriendo que la ganancia adicional al aumentar el número de *clusters* es cada vez menor. Esto respalda la elección de 2 *clusters* como una solución óptima desde el punto de vista de simplicidad y compactación.

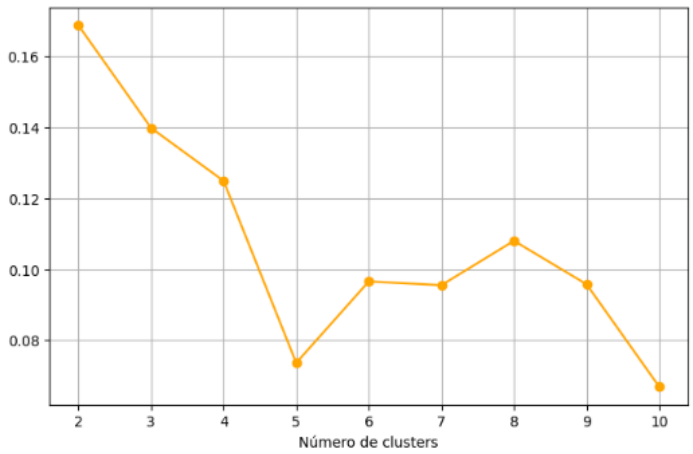
El coeficiente de Silueta también identifica 2 *clusters* como la solución más adecuada, con un puntaje de 0.169, que es el más alto en comparación con las demás configuraciones. Este índice mide la cohesión y separación entre los *clusters*, y aunque el puntaje no es particularmente alto, 2 *clusters* resultan ser la mejor opción dentro de los

valores evaluados. Por su parte, el índice de Calinski-Harabasz, que mide la separación entre *clusters* y la compacidad dentro de ellos, también señala 2 *clusters* como la configuración más favorable, con un puntaje de 4.62, el más alto en la serie analizada.

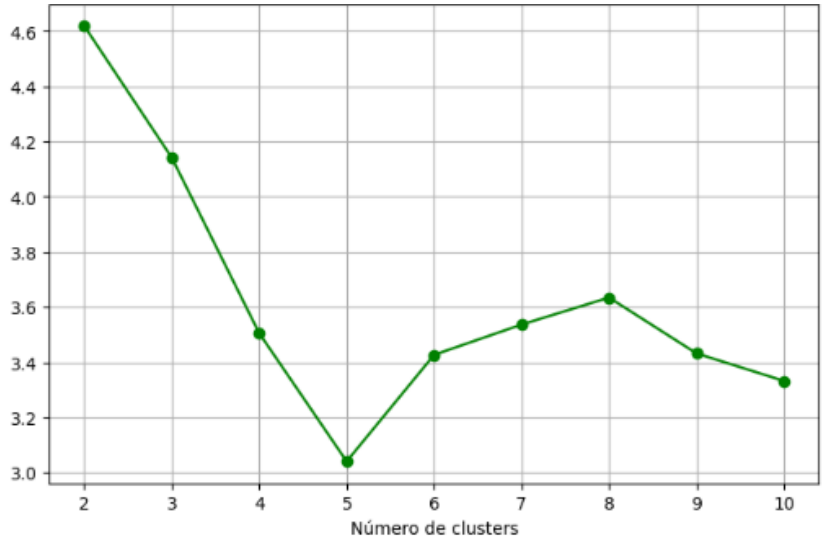
El gráfico de *Elbow* confirma estas observaciones. Mientras que la disminución inicial de la SSE es evidente entre 1 y 2 *clusters*, el punto de inflexión no es completamente claro en los valores posteriores, aunque la pendiente entre 4 y 5 *clusters* podría interpretarse como una alternativa razonable si se busca mayor granularidad en los resultados.

En resumen, el análisis sugiere que 2 *clusters* son la solución más simple y efectiva, alineándose con los resultados de los tres métodos evaluados. Sin embargo, si el objetivo del análisis es explorar una segmentación más detallada de los datos, se podría considerar un mayor número de *clusters*, como 4 o 5, donde aún hay cierta coherencia en los índices y el gráfico de *Elbow*. Esto dependerá del nivel de detalle requerido para los objetivos específicos del estudio.

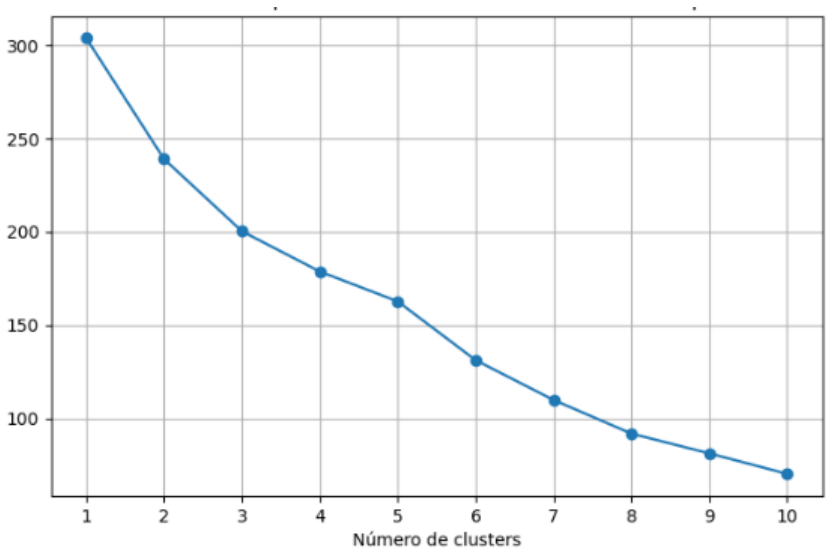
Número de clusters óptimo para la variable MIT_Hist (con PBIpc) evaluado en Coeficiente de Silueta



Número de clusters óptimo para la variable MIT_Hist (con PBIpc) evaluado en Índice de Calinski-Harabasz,



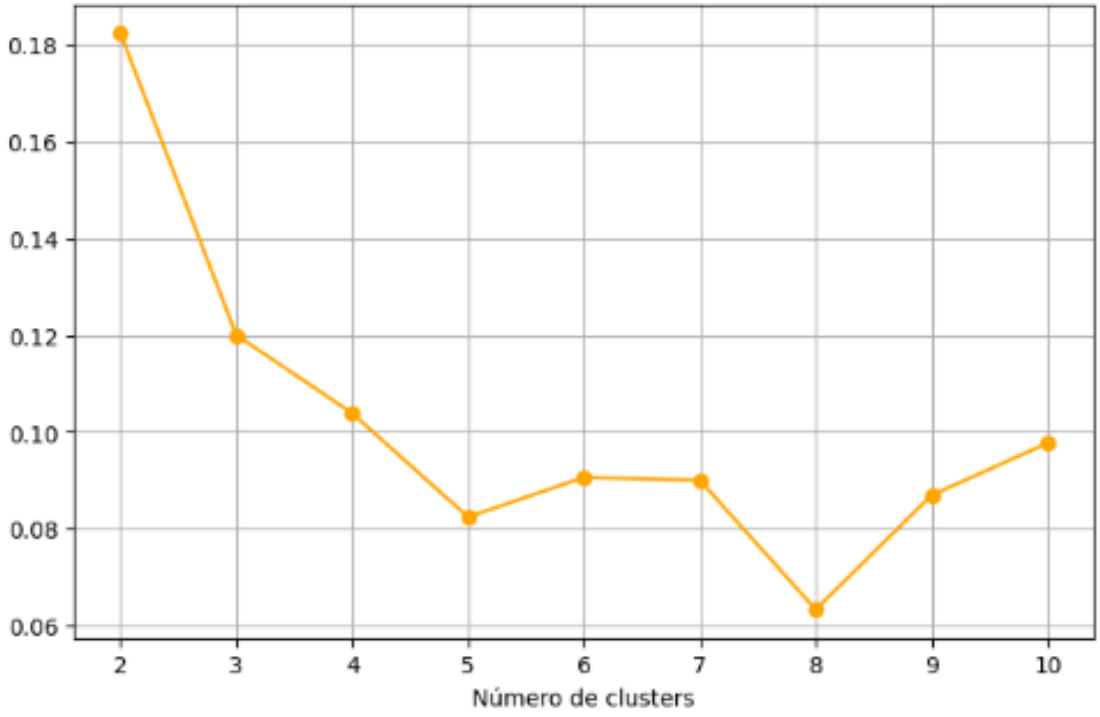
Número de clusters óptimo para la variable MIT_Hist (con PBIpc) evaluado en Método Elbow



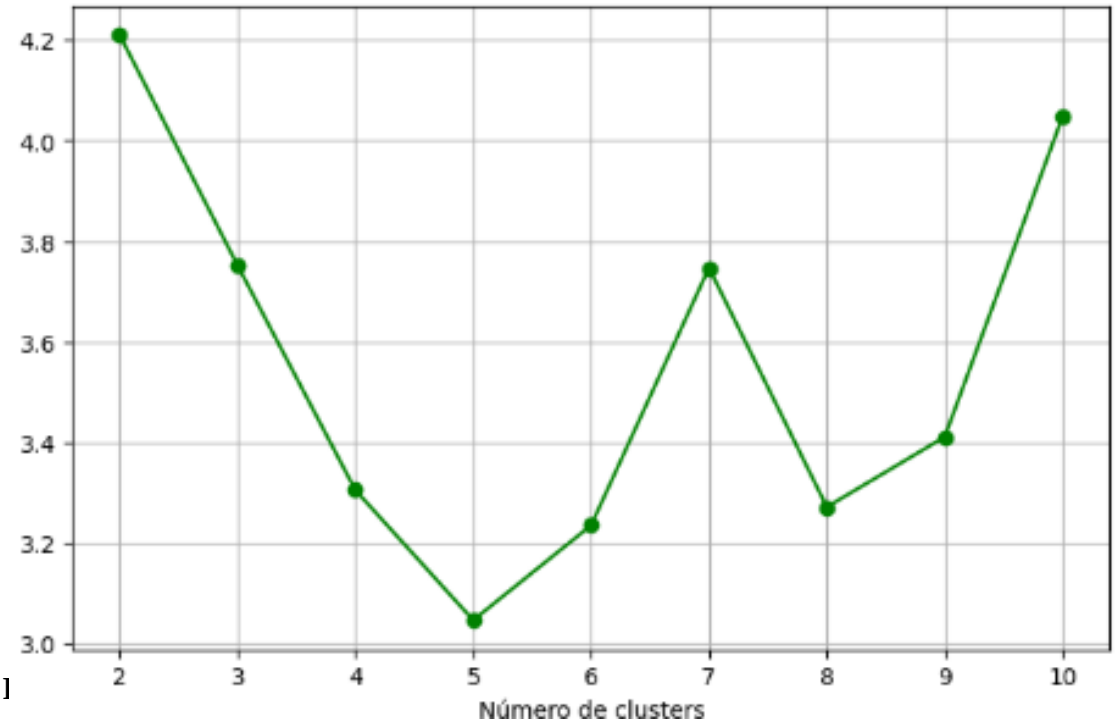
El análisis de *clustering* sin incluir la variable PBIpc (gdp) se centró en evaluar cómo las demás variables de interés relacionadas con el desarrollo de los países forman agrupaciones.

Los resultados obtenidos mediante el método *Elbow*, el coeficiente de Silueta y el índice de Calinski-Harabasz revelan algunos patrones, aunque con ciertas limitaciones en la claridad de los *clusters*. En el método *Elbow*, se observa una disminución progresiva en la suma de errores cuadráticos (SSE) a medida que aumenta el número de *clusters*. Si bien no hay un punto de codo definido, la reducción en la SSE se desacelera notablemente a partir de 6 *clusters*, sugiriendo que esta cantidad podría ser adecuada según esta métrica. En cuanto al coeficiente de Silueta, el valor más alto se alcanza con 2 *clusters* (0.182), aunque los valores en general son bajos, lo que indica que la separación entre *clusters* no es fuerte. Sin embargo, 6 *clusters* logran un coeficiente moderado (0.090), lo que puede justificar su elección considerando el contexto. Por otro lado, el índice de Calinski-Harabasz, que mide la compacidad interna y la separación entre *clusters*, tiene su valor máximo con 10 *clusters* (4.048), pero se observa un incremento moderado en los valores a partir de 6 *clusters*, reforzando nuevamente esta cantidad como una opción razonable. En general, la ausencia de la variable PBIpc parece limitar la claridad y definición de los *clusters*, dado que las variables restantes presentan una estructura menos diferenciada. A pesar de ello, la elección de 6 *clusters* emerge como una solución balanceada, apoyada por las tres métricas analizadas. Esta agrupación permite identificar patrones basados exclusivamente en las demás variables, y resulta pertinente profundizar en la interpretación de las características de cada grupo para comprender cómo estas variables interactúan en ausencia del PBIpc.

Silueta vs Número de Clusters para variable MIT_Hist sin PBIpc

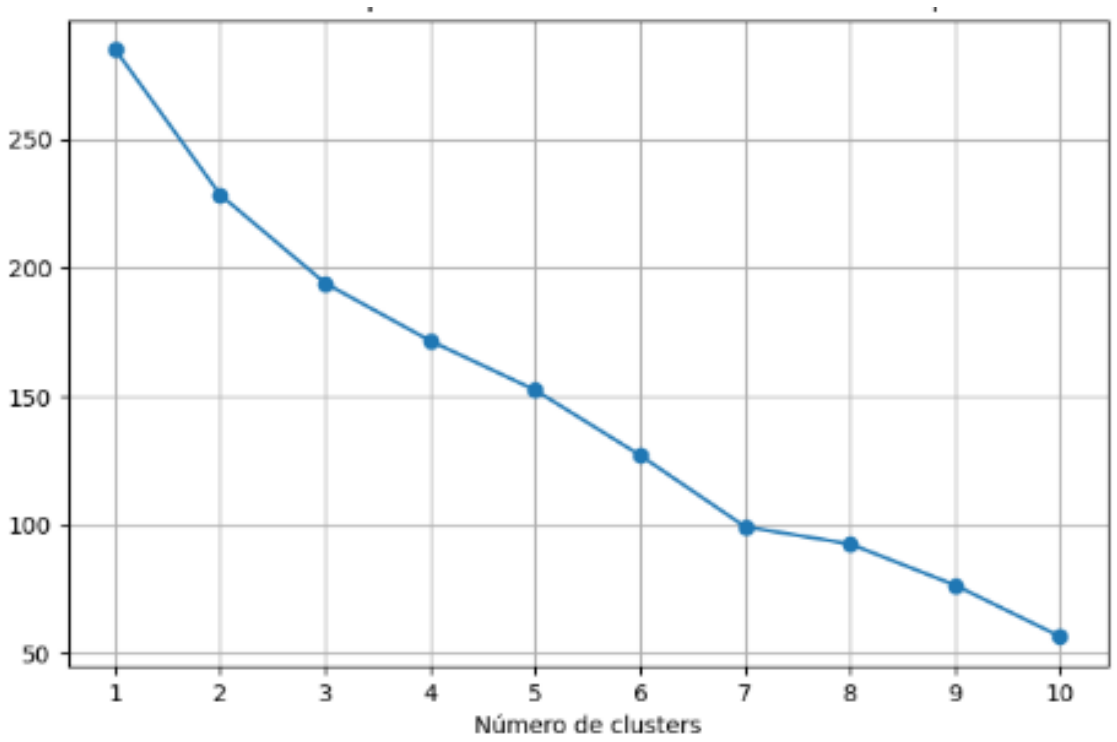


Índice de Calinski-Harabasz vs Número de Clusters para variable MIT_Hist sin PBIpc



]

Método Elbow para determinar el número de clusters óptimo para variable MIT_Hist sin PBIpc



h) Evaluación de Clustering con Cuatro Grupos Basado en Variables de Desarrollo y PBIpc para variable MIT_Fixed

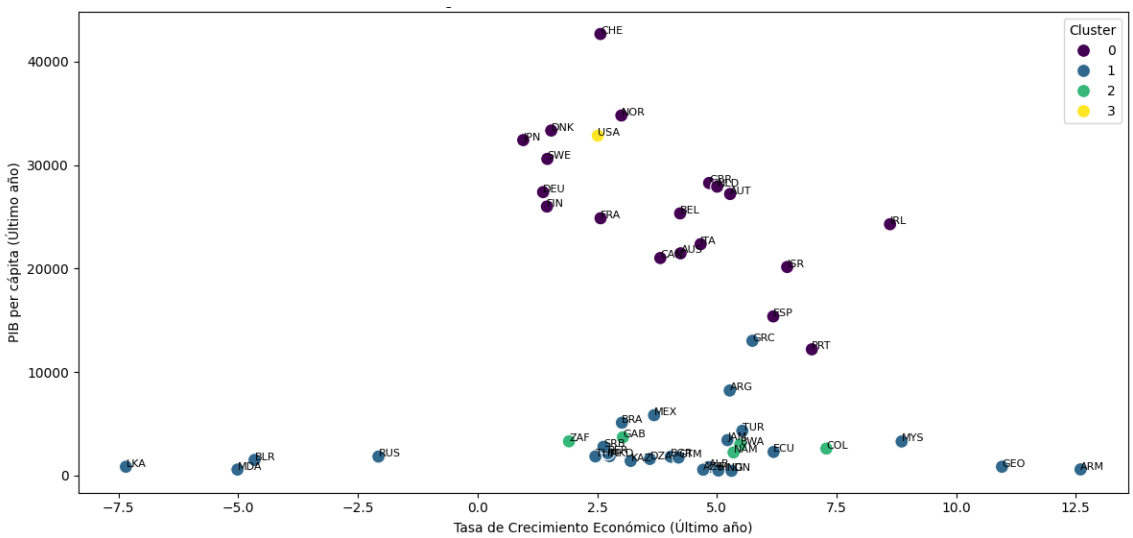
El análisis de *clustering* para MIT_Fixed con PBIpc identifica patrones entre países basados en variables socioeconómicas clave. Se generaron cuatro clusters considerando criterios combinados de *Elbow*, Silhouette y Calinsky-Harabasz. El *Cluster* 0 agrupa a países desarrollados con altos niveles de PBI per cápita, como Suiza, Japón y Noruega, caracterizados por baja desigualdad, diversificación económica elevada e inversión extranjera significativa. El *Cluster* 1 incluye economías emergentes y en desarrollo, como Argentina, Brasil y México, con niveles de desigualdad y pobreza moderados, un crecimiento económico sostenido, y una menor diversificación en comparación con el *Cluster* 0. El *Cluster* 2 agrupa a países con alta desigualdad y desafíos estructurales, como Sudáfrica y Namibia, que presentan bajos niveles de PBI per cápita, altas tasas de desempleo y pobreza, y un crecimiento económico desigual. Por último, el *Cluster* 3 se distingue como un grupo de *outliers* con características diferenciadas, como Estados Unidos, que combina un alto PBIpc total con características estructurales que lo diferencian de los países del *Cluster* 0.

El PBIpc emerge como una variable clave para diferenciar los *clusters*, especialmente entre países desarrollados y en desarrollo. Sin embargo, la asignación de Estados Unidos en un *cluster* distinto al de otras economías desarrolladas sugiere que existen otras dimensiones que pueden estar influyendo en la clasificación. Estas diferencias pueden estar relacionadas con factores como la composición sectorial del PBIpc, la estructura del mercado laboral, la distribución de ingresos y la dinámica de crecimiento económico.

De manera general, los *clusters* reflejan diferencias significativas en términos de nivel de desarrollo económico, desigualdad y estructura productiva. La clasificación obtenida muestra que países con niveles similares de PBI per cápita pueden pertenecer a distintos *clusters*, lo que sugiere que la estructura económica y social de cada país desempeña un papel importante en la segmentación observada.

Este análisis muestra que, si bien el PBIpc es un diferenciador central en la agrupación de países, otras variables estructurales pueden influir en la asignación de *clusters*, lo que indica la necesidad de considerar múltiples dimensiones para una caracterización más completa del desarrollo económico.

Clustering basado en variables de desarrollo con PBIpc (MIT_Fixed)



i) Clasificación de Países Mediante Clustering K-*Means*: Análisis de Desarrollo Socioeconómico (Sin PBIpc)

El gráfico refleja los resultados de un análisis de *clustering* utilizando el método de K-*means* excluyendo el PBIpc como variable en el conjunto de datos históricos (MIT_Hist).

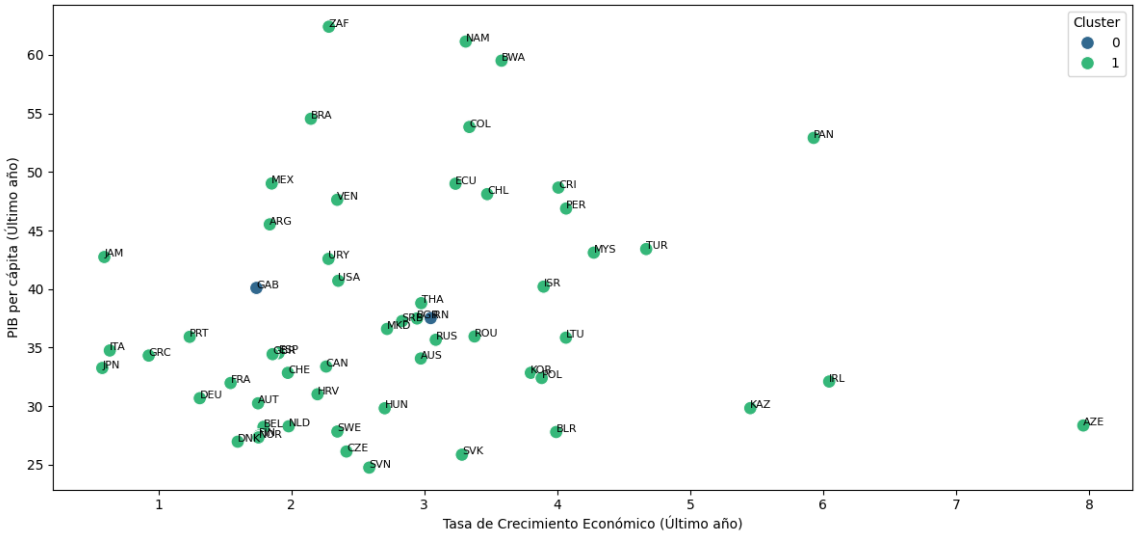
A primera vista, la distribución de los puntos en el gráfico pierde la estructura observada en análisis previos que incluían el PBIpc. Esto sugiere que el PBIpc desempeña un papel fundamental en la configuración de los *clusters*, lo que no solo refleja su relevancia económica, sino también su capacidad de diferenciar países según su nivel de desarrollo y características macroeconómicas.

El hecho de que los puntos aparezcan dispersos y sin un patrón claramente distinguible indica que las demás variables no logran proporcionar la misma capacidad explicativa o diferenciadora que el PBIpc. Es notable cómo, en ausencia de esta variable clave, el modelo tiene dificultades para segmentar a los países en grupos significativamente diferenciados. Esto también podría estar relacionado con el peso que el PBIpc tiene en la economía global como un indicador integrador de diversos aspectos del desarrollo, como la inversión, la productividad y la capacidad de generar riqueza.

En términos de interpretación, la exclusión del PBIpc resalta la necesidad de considerar variables complementarias que puedan compensar su influencia y que permitan capturar dimensiones alternativas del desarrollo. Sin embargo, en este caso, la dispersión sugiere que el resto de las variables no logra captar una estructura de *clusters* clara y robusta.

Figura 6

Clustering basado en variables de desarrollo (media histórica)



Nota: No hay umbrales ya que se quitó variable de PBI percápita

(Para más detalles, ver **Apéndice C, sección H**: "Evaluación de Clustering con cuatro grupos basado en variables de desarrollo y PBIpc para variable MIT_Fixed", y **Apéndice C, sección G**: "Evaluación de Clustering con cuatro grupos basado en variables de desarrollo y PBIpc para variable MIT_Hist")

j) Evaluación de Clustering con Cuatro Grupos Basado en Variables de Desarrollo y sin PBIpc para variable MIT_Fixed

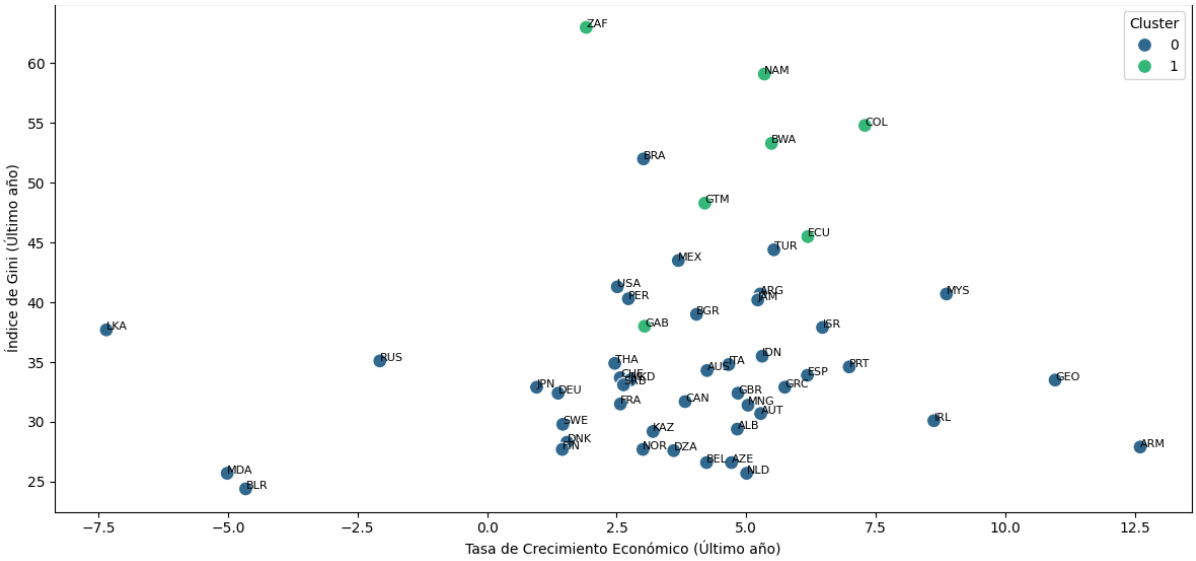
El análisis de *clustering* para MIT_Fixed sin el PBIpc evidencia una mayor dispersión de los valores, lo que sugiere que la exclusión de esta variable reduce la capacidad de cohesión dentro de los grupos y expone diferencias más marcadas entre los países. Sin embargo, incluso con esta dispersión, se mantienen patrones claros de desigualdad en los países asignados al *Cluster* 1, como Botsuana, Sudáfrica y Colombia.

Estos países muestran índices de Gini significativamente más altos, lo que refleja una mayor desigualdad económica y social. Además, suelen estar asociados con tasas de desempleo elevadas y desafíos estructurales en sus economías, como una mayor dependencia de sectores específicos. Por otro lado, el *Cluster* 0, compuesto por países como Austria, Finlandia y Japón, continúa reflejando menor desigualdad, estabilidad económica y una mayor diversificación tecnológica y económica, aunque su dispersión también es notable.

La mayor dispersión al excluir el PBIpc como variable sugiere que esta magnitud económica actuaba como un elemento unificador dentro de los grupos, simplificando la identificación de patrones. Sin embargo, la eliminación del PBIpc también permite destacar con mayor claridad las diferencias en términos de desigualdad y otros indicadores, reforzando la percepción de que los países del *Cluster* 1 enfrentan mayores desafíos en la distribución equitativa de los recursos.

En conclusión, aunque la dispersión de los datos es mayor al quitar el PBIpc, los patrones de desigualdad siguen siendo evidentes, con una división clara entre países más igualitarios y aquellos que enfrentan mayores barreras estructurales.

Clustering basado en variables de desarrollo sin PBIpc (MIT_Fixed)



Nota. La *clusterización* se realizó sin incluir el PBIpc como variable, utilizando en su lugar el coeficiente de Gini como criterio de segmentación. Se observa una dispersión significativa de los puntos, lo que sugiere que la clasificación sin la variable PBIpc genera grupos con menor cohesión estructural. Esto indica que, si bien el coeficiente de Gini, u otra variable, podrían aportar información relevante, su capacidad para segmentar países de manera clara es menor cuando se resta el PBIpc al modelo.

k) Análisis de Correlación: Exploración de Relaciones Entre Variables Económicas

El análisis de correlación es una herramienta clave para comprender las interrelaciones entre las variables económicas consideradas en este estudio. La matriz de correlación presentada revela una combinación de asociaciones positivas y negativas que destacan aspectos críticos de las dinámicas económicas globales. Este análisis no solo permite identificar patrones de comportamiento conjunto entre variables, sino que también cumple un rol fundamental como insumo para la posterior combinación de atributos en análisis multivariados, tales como técnicas de *clustering* multidimensional, donde se integran series temporales de múltiples indicadores económicos para capturar trayectorias estructurales más complejas.

6.1.1.1. Correlaciones Destacadas

Para evitar sesgos del panel completo, el análisis se centra en dos matrices diagnósticas: (i) correlaciones within, que eliminan los promedios país y capturan co-

movimientos estructurales; y (ii) correlaciones en primeras diferencias (Δ), que miden co-variación en cambios anuales y reducen el riesgo de correlaciones espurias por niveles o tendencias. Los resultados muestran redundancias claras y estables entre TasaInvAct_value e IED_value (ρ≈0,84 within; ρ≈0,65 Δ), lo que justifica tratarlas como un bloque de inversión (índice o selección de una de ellas según marco teórico); consistencia también entre gini, pobreza y desempleo, que avala un bloque de desigualdad socioeconómica; y un vínculo relevante entre ed y TasaInn (ρ≈0,58 within; ρ≈0,35 Δ), que respalda un bloque de capacidades estructurales (capital humano/innovación). La relación entre DivEcoCom y DivEcoTec es moderada (ρ≈0,36 within; ρ≈0,09 Δ): pueden mantenerse separadas para preservar matices (composición comercial vs. tecnológica) o consolidarse en un indicador de diversificación cuando el argumento empírico requiera parsimonia. Dos contrastes dinámicos orientan la interpretación: ed con TasaCrecimientoEcono es positiva en within pero negativa en Δ (≈−0,35), sugiriendo efectos rezagados de la educación que no se reflejan en variaciones anuales; y Term_Inter con Pobreza es positivo en Δ pero neutro en within, coherente con shocks externos de corto plazo más que con estructura. En síntesis, las matrices within y Δ se emplean solo con fines diagnósticos para reducir colinealidad y clarificar dimensiones antes del agrupamiento: en el multidimensional, se privilegia una representación parsimoniosa por bloques (inversión, desigualdad, capacidades, diversificación) para mejorar estabilidad e interpretabilidad; en el unidimensional, cada variable se analiza por separado con DTW y la agregación posterior por moda ponderada (pesos de Random Forest) deja explícita la contribución de cada dimensión. La matriz del panel completo se relega al apéndice con una advertencia metodológica sobre sus limitaciones.

Figura 7

Matriz de correlación entre atributos (Primeras diferencias)

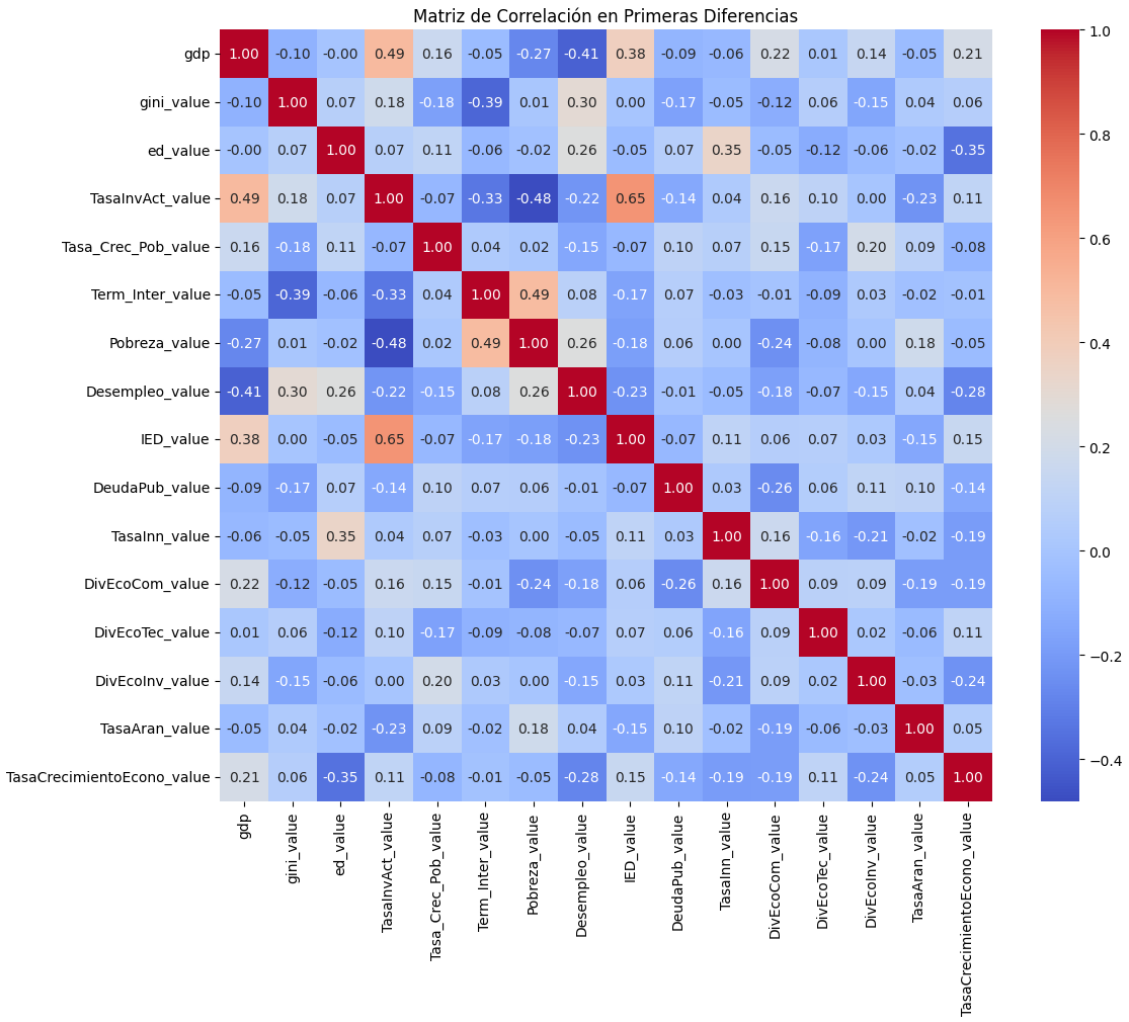
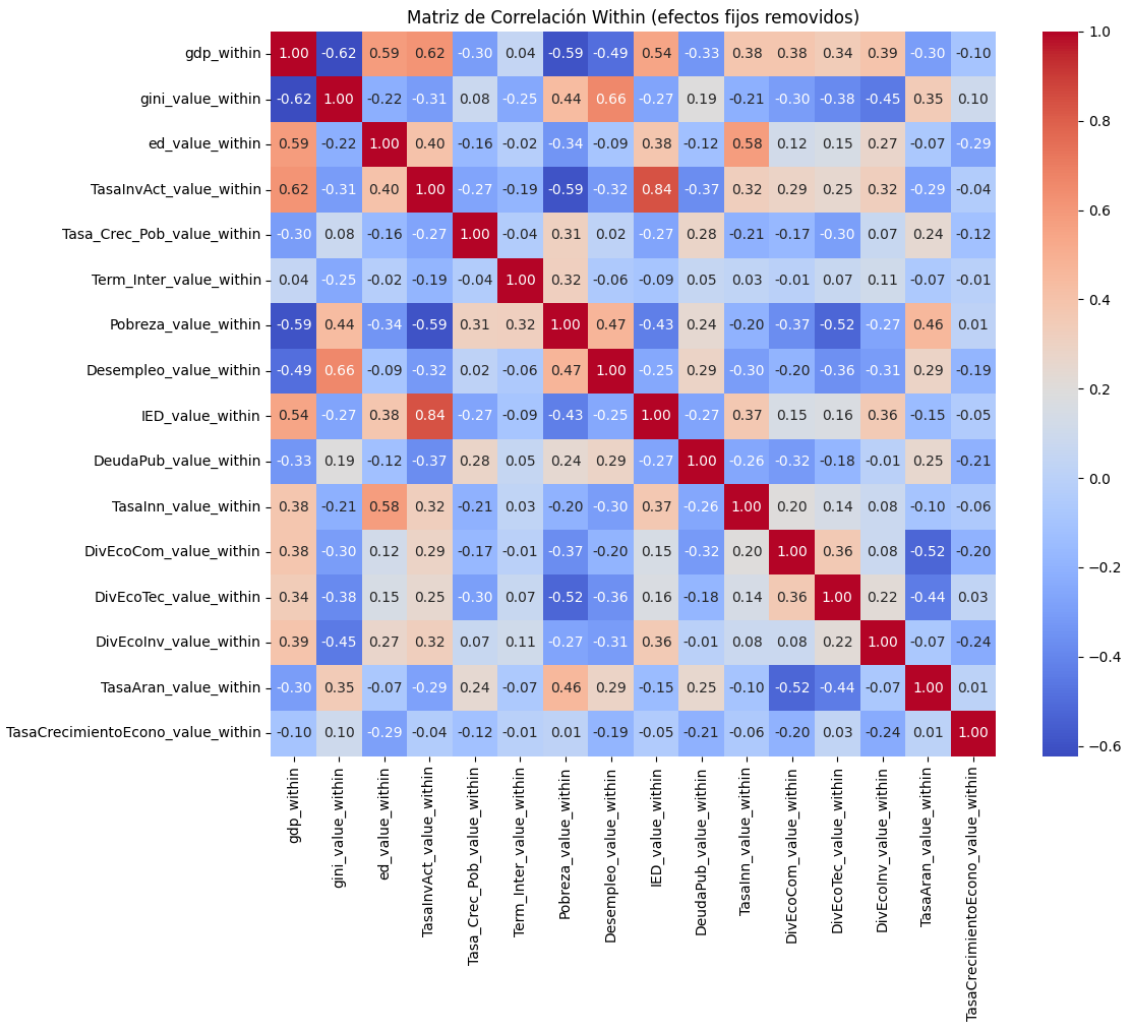


Figura 7

Matriz de correlación entre atributos (Primeras diferencias)



l) DTW Multidimensional en Segundo Nivel: Exclusión del Cluster Verde y Sin Variables Combinadas

El presente análisis de *clustering* jerárquico basado en DTW Multidimensional se enfoca en los países MIT lejos del desarrollo, excluyendo aquellos que forman parte del *Cluster* Verde, lo que permite estudiar economías en diferentes fases de la trampa de ingresos medios sin la influencia de países más desarrollados. A diferencia del análisis previo, en esta iteración se han eliminado las variables combinadas para evitar redundancias generadas por correlaciones altas entre atributos.

i. Atributos Considerados. Para esta fase del análisis, se seleccionaron 16 variables económicas clave con el objetivo de capturar el comportamiento estructural de las economías en términos de crecimiento, desigualdad, inversión y flujo de capital. Con el fin de garantizar un análisis más preciso y evitar posibles distorsiones, no se incluyeron variables sintéticas o combinadas, permitiendo así que el *clustering* se base en patrones subyacentes más específicos.

Las variables seleccionadas abarcan distintas dimensiones económicas fundamentales. En primer lugar, el crecimiento y desarrollo económico se representan a través del PBIpc (gdp) y la tasa de crecimiento económico (TasaCrecimientoEcono_value). En cuanto a la distribución del ingreso y la desigualdad, se incluyen el coeficiente de Gini (gini_value) y la tasa de pobreza (Pobreza_value), proporcionando una visión sobre la equidad económica en cada país.

Por otro lado, el análisis de inversión y mercado laboral considera indicadores como la tasa de inversión activa (TasaInvAct_value), la inversión extranjera directa (IED_value), la tasa de desempleo (Desempleo_value) y la deuda pública (DeudaPub_value), los cuales permiten evaluar la estabilidad y sostenibilidad del crecimiento económico. Asimismo, la diversificación económica se mide mediante tres variables específicas: diversificación comercial (DivEcoCom_value), diversificación tecnológica (DivEcoTec_value) y diversificación en inversión (DivEcoInv_value), elementos clave para entender la resiliencia de las economías ante cambios globales.

La apertura comercial y la competitividad se reflejan en la tasa de innovación (TasaInn_value) y la tasa arancelaria (TasaAran_value), indicadores que permiten evaluar el grado de integración de las economías en los mercados internacionales. Finalmente, se incluyen factores institucionales y demográficos, representados por la duración de los mandatos políticos (Term_Inter_value) y la tasa de crecimiento poblacional (Tasa_Crec_Pob_value), variables que influyen en la estabilidad económica y en la capacidad de los países para implementar políticas de desarrollo.

La selección de estos atributos se realizó a partir de un riguroso análisis de correlación, asegurando que cada variable aporte valor al proceso de *clustering* y que no exista redundancia con otras dimensiones del estudio. Con este enfoque, se busca una

segmentación más precisa y representativa de las diferencias estructurales entre las economías analizadas.

ii. **Países Incluidos y Exclusiones.** Se excluyeron países fuera del MIT, territorios insulares o dependientes de servicios financieros, así como economías con estructuras económicas extremadamente particulares. La siguiente tabla presenta los países considerados y su asignación a *clusters*:

Tabla 8

Clasificación Cluster segundo nivel sin variables combinadas

Cluster	Países
C1	ARG, BRA, CHL, CHN, COL, CRI, CUB, DOM, EAS, MEX, MYS, PAN, PER, THA, URY, VCT.
C2	BEL, ESP, FRA, GRC, IRL, ITA, NLD, PRT, SVK.
C3	AUT, CHE, FIN, HKG, ISL, JPN, KOR, LUX, MDV, SWE, TUR, ZAF.
C4	BGR, BLR, CEB, CZE, EST, HRV, HUN, KAZ, LTU, LVA, MKD, POL, ROU, RUS, SRB.
C5	AUS, CAN, CYP, DNK, GBR, ISR, JAM, MLT, NOR, NZL, SAU, SGP, USA, VEN.

Interoretación del Clustering y Agrupación de Países

Cluster	Interpretación
Cluster 1 (C1): Mercados Emergentes y en Desarrollo	Agrupa a economías emergentes de América Latina, Asia y el Caribe, caracterizadas por un crecimiento dinámico, pero con desafíos estructurales en productividad, innovación y estabilidad institucional. Países como China y Tailandia han logrado avances significativos en manufactura y comercio, mientras que economías latinoamericanas como Brasil, Argentina y México aún enfrentan volatilidad económica y dependencia de materias primas. Este grupo refleja la heterogeneidad de los mercados en desarrollo, con países en distintas etapas de consolidación económica.
Cluster 2 (C2): Europa Occidental con Economías Diversificadas	Compuesto por economías desarrolladas de Europa Occidental, este grupo se caracteriza por altos ingresos per cápita, integración en la Unión Europea y estructuras económicas diversificadas. España, Italia y Grecia han enfrentado crisis fiscales en el pasado, mientras que Países Bajos e Irlanda se destacan por su dinamismo financiero y comercial. Eslovaquia aparece en este grupo, posiblemente por su integración en la UE y su crecimiento sostenido en manufactura e industria.
Cluster 3 (C3): Economías Avanzadas y Especializadas	Incluye economías altamente desarrolladas con modelos productivos específicos. Suiza, Japón y Corea del Sur destacan por su liderazgo en tecnología y manufactura avanzada, mientras que Suecia y Finlandia se caracterizan por políticas de bienestar y alto nivel de innovación. Turquía y Sudáfrica aparecen en este grupo, sugiriendo que ciertos factores económicos los acercan a países con alta especialización, aunque sus desafíos estructurales pueden diferenciarse del resto.
Cluster 4 (C4): Europa del Este y Asia Central en Transición	Agrupa a países de Europa del Este y Asia Central con economías en transición. Son países con un pasado de planificación centralizada y que han adoptado modelos de mercado con diferentes grados de éxito. Rusia y Kazajistán mantienen estructuras económicas dependientes de recursos naturales, mientras que Polonia y la República Checa han logrado consolidarse como economías de manufactura y servicios. El desarrollo institucional y la estabilidad política varían significativamente dentro del grupo.
Cluster 5 (C5): Economías Altamente Desarrolladas y Globalizadas	Compuesto por algunas de las economías más avanzadas del mundo, este grupo se caracteriza por altos niveles de ingreso, mercados financieros sólidos y liderazgo en comercio global. Estados Unidos, Reino Unido, Canadá y Australia dominan la escena económica global con industrias diversificadas. Singapur y Noruega destacan por su estabilidad macroeconómica y liderazgo en comercio y recursos naturales. Arabia Saudita y Jamaica son casos atípicos en el grupo, lo que podría reflejar la necesidad de ajustes en la segmentación. Venezuela también se encuentra en este cluster, lo que sugiere que ciertas variables utilizadas podrían estar generando anomalías en la clasificación.

iii. **Resumen del Clustering.** El análisis comparativo entre los modelos de clustering con y sin variables combinadas ha revelado diferencias significativas en la asignación de los países a los distintos grupos, especialmente en el caso de las economías emergentes. En el modelo que incluye variables combinadas, estos países se agrupan en C4, manteniendo cierta coherencia tanto a nivel regional como estructural. Sin embargo, cuando se eliminan estas variables, los mismos países son trasladados a C1, lo que sugiere que, en ausencia de combinaciones de indicadores, el modelo tiende a clasificar a los

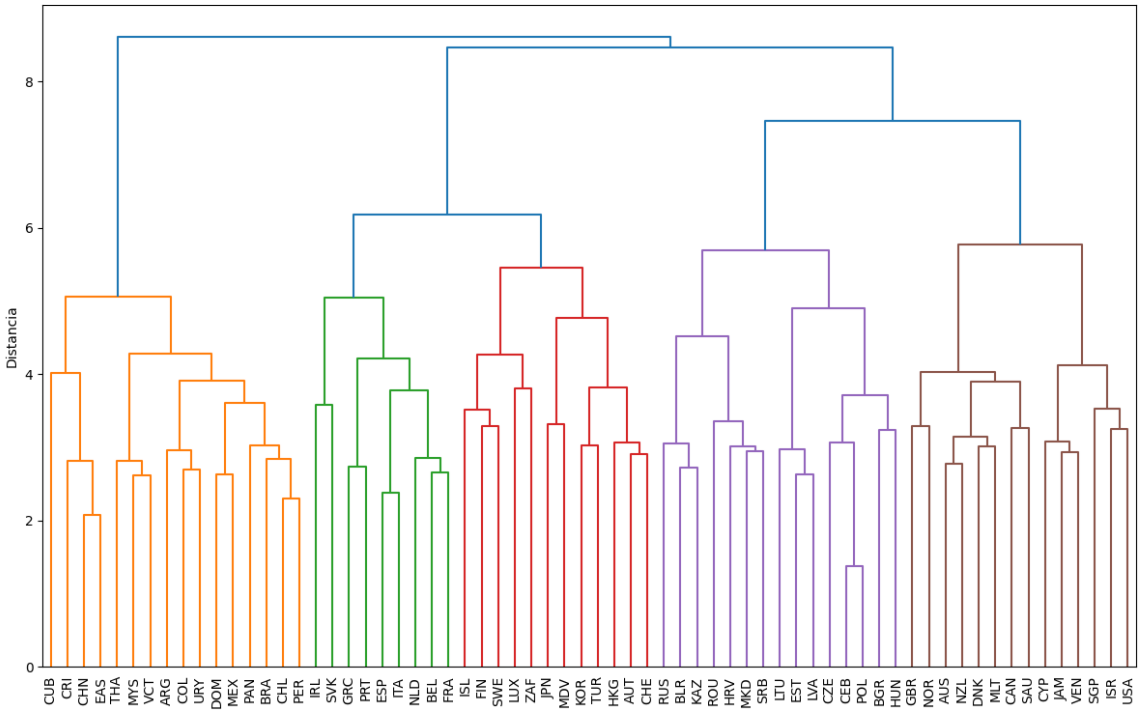
países únicamente en función de valores absolutos, sin captar relaciones económicas más complejas.

En contraste, la segmentación de las economías avanzadas muestra una mayor estabilidad en ambos modelos. En ambos casos, C5 se mantiene como el grupo que agrupa a las economías más desarrolladas, incluyendo países como Estados Unidos, Canadá, Australia y el Reino Unido. No obstante, en el modelo sin variables combinadas, C5 también incorpora a Arabia Saudita y Jamaica, lo que genera ciertas inconsistencias en la clasificación y plantea dudas sobre la precisión de este enfoque.

Otro hallazgo relevante es la persistencia de anomalías en la clasificación de ciertos países. Venezuela continúa apareciendo en C5 en ambos modelos, lo que indica que algunos indicadores podrían estar sobrevalorando su similitud con economías avanzadas. Turquía, por su parte, permanece en C3 en ambas metodologías, lo que sugiere que su economía posee características híbridas que dificultan su clasificación en un solo grupo. Un caso particular es Sudáfrica, cuya asignación varía dependiendo del enfoque utilizado: con variables combinadas aparece en C5, mientras que sin ellas se traslada a C3, lo que evidencia que la metodología de clustering impacta significativamente su clasificación.

En términos generales, el clustering sin variables combinadas genera una segmentación más dependiente de valores absolutos, lo que limita su capacidad para capturar dinámicas económicas complejas. Por el contrario, el uso de variables combinadas permite reflejar con mayor precisión las relaciones estructurales y económicas entre los países, proporcionando una visión más realista del desarrollo económico. La eliminación de estas variables da lugar a clusters que siguen patrones geográficos más estrictos, pero con una menor capacidad explicativa en lo que respecta a la especialización productiva y la transición económica de los países.

Dendrograma de Clustering Basado en DTW segundo nivel sin variables combinadas



m) Random Forest con SHAP

Para evaluar la importancia de las variables en la segmentación de países obtenida mediante *clustering*, se ha utilizado un modelo de *Random Forest* complementado con el análisis de valores SHAP (*SHapley Additive exPlanations). Random Forest* es un algoritmo basado en múltiples árboles de decisión que permite determinar el impacto de diversas variables en la clasificación de observaciones, en este caso, la asignación de países a distintos *clusters* (Breiman, 2001). Sin embargo, a diferencia del análisis de importancia de variables basado en reducción de impureza o permutación, SHAP permite una interpretación más detallada al descomponer el impacto individual de cada variable en la predicción del modelo (Lundberg & Lee, 2017).
El enfoque de SHAP se basa en la teoría de juegos, asignando un valor de contribución a cada característica en función de cómo cambia la predicción cuando la variable es considerada en distintos escenarios (Lundberg & Lee, 2017). Esto permite entender no solo la magnitud de la importancia de una variable, sino también la dirección de su efecto, es decir, si un valor alto de una variable aumenta o disminuye la probabilidad de asignación a un cluster determinado. En este análisis, se ha empleado un conjunto de árboles con parámetros ajustados para evitar sobreajuste, asegurando una interpretación estable de los valores SHAP.

i. Interpretación de la Importancia de Variables.

El gráfico de SHAP muestra el impacto de cada variable en la asignación de *clusters*. Las variables están ordenadas según su importancia, y el color representa el valor de la característica: valores altos en rojo y valores bajos en azul. Se observa que Pobreza_value es la variable con mayor influencia en la segmentación. La distribución de los puntos sugiere que valores altos de pobreza (rojo) tienden a tener un impacto positivo en la probabilidad de asignación a ciertos *clusters*, mientras que valores bajos (azul) reducen esta probabilidad, lo que indica que el *clustering* está capturando diferencias estructurales en términos de desarrollo socioeconómico.

TasaInn_value y Indicador_Crecimiento aparecen como las siguientes variables más relevantes. La innovación es un factor diferenciador clave, lo que sugiere que los países con mayores tasas de innovación tienden a agruparse en clusters distintos de aquellos con baja capacidad innovadora. De manera similar, el crecimiento económico tiene un efecto

significativo en la clasificación de países, indicando que las trayectorias de expansión juegan un papel determinante en la diferenciación de grupos.

Otra variable con alto impacto es el indicador de Flujo de Capital (Indicador_Flujo_Capital), lo que sugiere que la integración al comercio global y la facilidad para realizar transacciones internacionales han influido en la segmentación. Sorprendentemente, el coeficiente de Gini **(**gini_value**)** muestra una importancia moderada, indicando que la desigualdad de ingresos no ha sido un factor predominante en la diferenciación de clusters, aunque sigue teniendo un efecto observable. La educación **(**ed_value**)** también juega un papel relevante, reflejando la relación entre inversión en capital humano y el desarrollo económico de los países.

Las variables con menor impacto incluyen Tasa_Crec_Pob_value**,** Term_Inter_value y DeudaPub_value, lo que sugiere que factores como la tasa de crecimiento poblacional y la deuda pública no han sido determinantes en la diferenciación de los clusters, al menos en la estructura de datos utilizada. No obstante, esto no implica que estas variables sean irrelevantes para el desarrollo económico, sino que, dentro del modelo actual, otras características han sido más explicativas.

ii. **Implicaciones para el Análisis Unidimensional**.

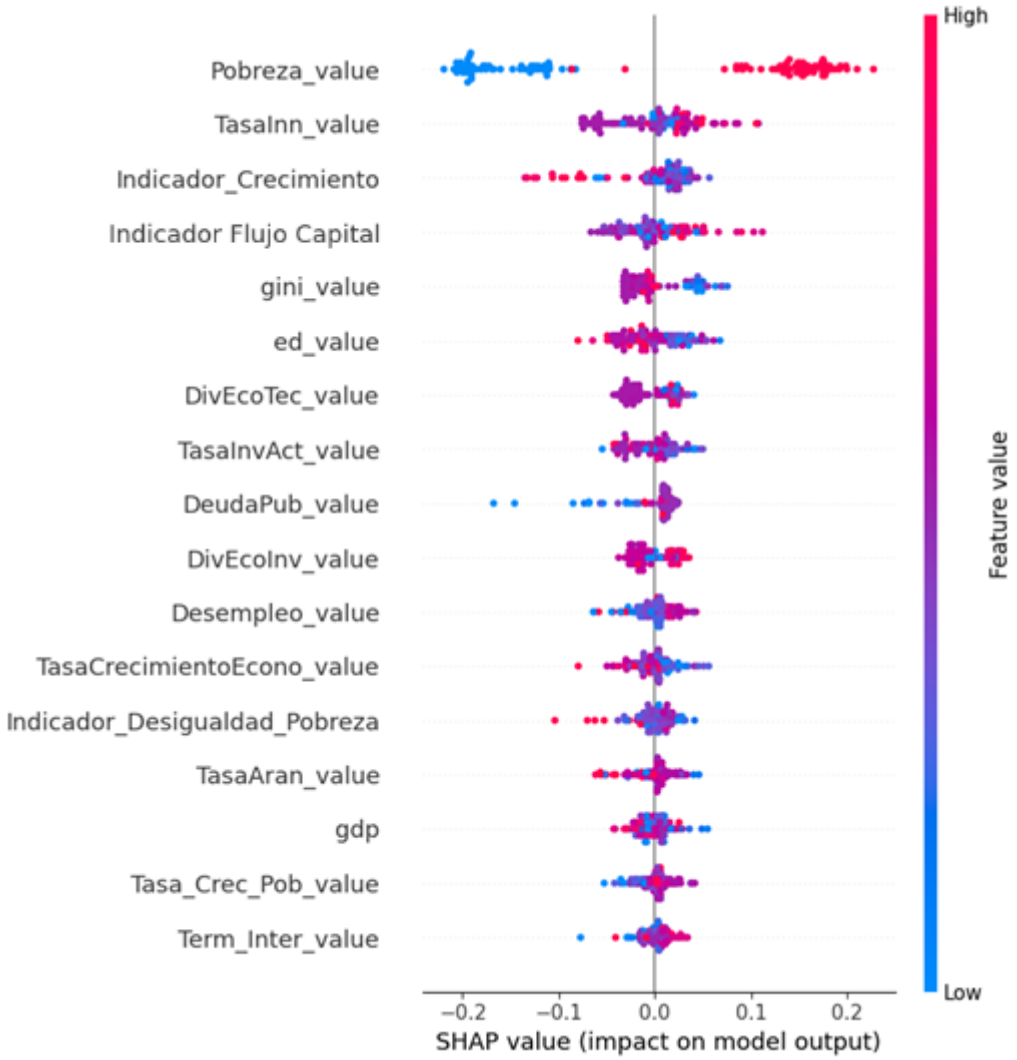
Dado que la pobreza, la innovación, el crecimiento económico y la apertura comercial han sido los principales determinantes en la segmentación de los países, el siguiente paso en el análisis debe centrarse en estas variables clave. Un análisis de *clustering* unidimensional sobre estas dimensiones permitirá evaluar con mayor precisión si ciertos países muestran patrones específicos de estancamiento económico, como la presencia de la *Middle-Income Trap* (MIT).

Es recomendable iniciar con una segmentación basada en Pobreza_value y TasaInn_value, dado su impacto significativo en la asignación de clusters. Posteriormente, variables como Indicador_Crecimiento y Indicador_Flujo_Capital pueden ser utilizadas para evaluar diferencias en dinámicas de desarrollo y especialización productiva.

En conclusión, el uso de SHAP en combinación con *Random Forest* ha permitido una interpretación más detallada del impacto de cada variable en la segmentación de países,

confirmando que las diferencias en pobreza, innovación, crecimiento y flujo de capital han sido los principales factores en la diferenciación de *clusters*. A partir de estos hallazgos, el siguiente paso del estudio debe enfocarse en un análisis unidimensional sobre estas variables, con el objetivo de detectar patrones específicos en la dinámica de desarrollo económico y refinar la identificación de países potencialmente atrapados en la Middle-Income Trap.

Impacto de cada variable en la asignación de Clusters según Random Forest con SHAP



l) Cluster de K-Medias

El algoritmo K-Means (K-Medias), propuesto por MacQueen (1967), constituye el primer enfoque de agrupamiento aplicado en este estudio, con el objetivo de segmentar a los países en conglomerados homogéneos en función de variables estructurales y

socioeconómicas clave. Este análisis se complementa con el algoritmo Density-Based Spatial Clustering of Applications with Noise (DBSCAN), desarrollado por Ester, Kriegel, Sander y Xu (1996), que permite identificar agrupaciones de países con formas irregulares y la detección de observaciones atípicas. El proceso inicia con un análisis exploratorio de las principales tendencias de variables relevantes como el PBI per cápita, la desigualdad (medida a través del índice de Gini), la inversión en investigación y desarrollo (I+D), la diversificación productiva y el nivel educativo. Este análisis preliminar permite identificar patrones generales y correlaciones significativas entre las variables y la condición de estar o no atrapado en la trampa de ingresos medios (MIT).

Inicialmente, la formación de grupos sigue un enfoque que utiliza los umbrales de ingreso per cápita definidos en la literatura, como punto de partida para clasificar a los países. Sin embargo, para superar las limitaciones de esta aproximación estática, se adopta un esquema complementario que prescinde de dichos umbrales y construye clusters basados exclusivamente en características estructurales, lo que permite captar la heterogeneidad subyacente entre economías. En línea con este objetivo, se contempla además una subselección metodológica posterior centrada en países desarrollados y aquellos casos que, si bien presentan niveles de ingreso per cápita relativamente elevados (superiores a los 12.000 USD en promedio histórico), comparten simultáneamente rasgos estructurales similares a los observados en economías en desarrollo. Esta estrategia analítica permitirá examinar con mayor detalle las trayectorias de desarrollo de estos países, evaluando si el crecimiento en los niveles de ingreso ha estado acompañado por transformaciones estructurales significativas o si, por el contrario, persisten condiciones típicas de la trampa de ingresos medios. Esta decisión metodológica responde a la necesidad de ampliar el enfoque más allá del ingreso y considerar configuraciones complejas de desarrollo.

El procedimiento de K-Means y DBSCAN se aplica bajo dos perspectivas temporales: un enfoque de estado fijo, que utiliza los valores de un año específico (2024), proporcionando una imagen instantánea de la situación de cada país; y un enfoque de media histórica, que considera el promedio de las variables clave a lo largo de toda la serie temporal disponible, reflejando la trayectoria estructural de cada economía. En ambas perspectivas, se generan clasificaciones tanto incluyendo el PBI per cápita como variable central, como excluyéndolo para concentrar el análisis en factores estructurales como la desigualdad, la educación y la innovación. Esta combinación de enfoques da

lugar a cuatro clasificaciones complementarias, que permiten observar cómo varían los patrones de agrupamiento cuando se priorizan diferentes conjuntos de variables.

El número óptimo de clusters en cada configuración se determina mediante el método del codo (Elbow Method) y el coeficiente de silueta, garantizando así la solidez de las segmentaciones obtenidas. En términos generales, los resultados del clustering reflejan la existencia de un primer grupo que podría estar compuesto por países atrapados estructuralmente en la MIT, y economías en transición con cierto dinamismo económico, y un segundo grupo conformado por economías consolidadas que han superado la MIT, con altos niveles de diversificación productiva e innovación. Esta aproximación integral, al combinar distintos conjuntos de variables y perspectivas temporales, proporciona un marco analítico robusto para comprender las distintas trayectorias de los países y evaluar cómo las condiciones estructurales y las dinámicas de largo plazo inciden en la permanencia o superación de la trampa de ingresos medios.

n) **Enfoque de *cluster* con variables combinadas y muestra 1990 – 2024**

Una vez definidas las variables clave mediante el proceso de selección alternativo, se aplican nuevamente las técnicas de *clustering* para evaluar si el uso de variables seleccionadas mejora la calidad y coherencia de las agrupaciones. Este proceso se realiza tanto con el algoritmo K-Medias como con el Clustering Dinámico basado en DTW, lo que permite observar el comportamiento de las economías bajo diferentes metodologías y con un conjunto depurado de variables. La combinación de variables seleccionadas permite identificar patrones más claros y diferenciados entre los *clusters*, evidenciando que los factores estructurales asociados a la innovación, la inversión productiva y la desigualdad son determinantes clave para explicar la permanencia o salida de la MIT. Este enfoque destaca la importancia de adaptar los criterios de segmentación a la realidad estructural de cada país, evitando la dependencia exclusiva de umbrales estáticos que, si bien son utilizados ampliamente en la literatura, pueden no capturar adecuadamente la complejidad de las trayectorias de desarrollo económico.

Complementariamente, se lleva a cabo un análisis comparativo detallado entre las distintas clasificaciones generadas, con el fin de evaluar la consistencia y estabilidad de los resultados obtenidos bajo distintos enfoques. Para ello, la base de datos en panel fue transformada a un formato de corte transversal, lo que permitió condensar la información

temporal de cada país en un único vector representativo. Esta transformación respondió tanto a la necesidad analítica de sintetizar las trayectorias económicas como a las exigencias técnicas de los algoritmos utilizados, particularmente K-Medias, que requiere datos transversales. Se adoptaron dos estrategias: una basada en el valor del último año disponible (Estado Fijo) y otra en el promedio histórico de las variables (Media Histórica). Ambas versiones fueron contrastadas con las agrupaciones obtenidas mediante clustering dinámico (DTW), lo que permitió explorar el grado de coincidencia entre metodologías y detectar posibles discrepancias asociadas a factores coyunturales o estructurales.

y las agrupaciones derivadas del *clustering* dinámico DTW. Este análisis permitió identificar cuántos países clasificados en la MIT según la media histórica coincidían con los mismos *clusters* generados con el enfoque de estado fijo, proporcionando información sobre la estabilidad de las categorías definidas y la pertinencia de los umbrales tradicionales. Adicionalmente, se contrastan las trayectorias dinámicas identificadas mediante DTW con el estado actual de los países, lo que permitió detectar casos en los que países con trayectorias históricamente similares presentaban posiciones actuales divergentes. Este tipo de análisis comparativo no solo permite validar parcialmente los resultados obtenidos con variables seleccionadas, sino que también contribuye a identificar posibles factores transitorios o estructurales que expliquen las discrepancias entre métodos, reforzando la necesidad de un enfoque flexible y adaptativo para comprender las condiciones que determinan la permanencia o superación de la MIT.

El análisis se centra en un conjunto de series temporales que representan variables económicas clave de 276 países en el período comprendido entre 1960 y 2024. Estas variables incluyen indicadores como el PBIpc (gdp), el índice de Gini (gini_value), la tasa de inversión activa (TasaInvAct_value), la pobreza (Pobreza_value), la inversión extranjera directa (IED_value), entre otras, que son esenciales para evaluar las dinámicas económicas y sociales de los países. Después de un proceso de integración y limpieza de datos, se seleccionaron 138 países que cuentan con datos suficientemente completos para el análisis. Las series temporales fueron normalizadas para garantizar la comparabilidad entre países y variables, teniendo en cuenta la heterogeneidad en las economías analizadas.

Sin embargo, tras realizar pruebas de *clusterización* con distintos períodos de datos, se observó que los resultados mejoran significativamente al considerar únicamente la información a partir de 1990. Cuando se emplea la serie completa desde 1960, la segmentación tiende a generar solo dos grandes grupos, perdiendo granularidad y capacidad explicativa. Por otro lado, al realizar la *clusterización* únicamente con los datos del último año disponible, se obtuvo un número elevado de *clusters*, lo que sugiere que la estructura de los datos cambia considerablemente en función del período analizado. En cambio, al limitar el análisis al intervalo 1990-2024, la *clusterización* produce una segmentación más equilibrada, capturando tanto la evolución histórica reciente como la diversidad entre países de manera coherente con la realidad económica contemporánea. Por este motivo, la presente investigación se centra en el período 1990-2024.