



Facultad de Ciencias Económicas y Empresariales
ICADE

Indicador compuesto de innovación tecnológica por países

Autor: Carmen García Cucala
Director: Pablo Calvo Báscones

MADRID | Abril 2024

RESUMEN

La innovación tecnológica, fundamental en el desarrollo de nuevos productos y procesos, ha adquirido un papel crucial en la economía global, especialmente con la integración de tecnologías digitales y la inteligencia artificial.

Este trabajo desarrolla un indicador compuesto que mide la innovación tecnológica en diversos países, empleando la base de datos de los Indicadores Principales de Ciencia y Tecnología de la OCDE. La metodología adoptada combina técnicas de selección de variables, normalización, imputación de datos faltantes y análisis multivariable, utilizando modelos de *Machine Learning* para asignar pesos y agrupar los indicadores. Este proceso culmina con la creación de un indicador que permite comparar la innovación tecnológica entre naciones anualmente.

PALABRAS CLAVE

Indicador compuesto, indicador sintético, innovación tecnológica, comparativa países, normalización, imputación, correlación, extracción de pesos, agregación.

ABSTRACT

Technological innovation, fundamental in the development of new products and processes, has taken on a crucial role in the global economy, especially with the integration of digital technologies and artificial intelligence.

This project develops a composite indicator that measures technological innovation across various countries, using the OECD's Main Science and Technology Indicators database. The adopted methodology combines techniques for variable selection, normalization, imputation of missing data, and multivariable analysis, utilizing machine learning models to assign weights and aggregate the indicators. This process culminates in the creation of an indicator that allows for the annual comparison of technological innovation among nations.

KEYWORDS

Composite indicator, synthetic indicator, technological innovation, countries comparative, normalization, imputation, correlation, weight extraction, aggregation.

ÍNDICE DE CONTENIDOS

ÍNDICE DE SIGLAS	1
ÍNDICE DE TABLAS	2
ÍNDICE DE FIGURAS	3
ÍNDICE DE ECUACIONES	4
1. INTRODUCCIÓN.....	5
1.1 Interés por la cuestión.....	5
1.2 Objetivos del proyecto.....	6
1.3 Metodología.....	7
2. MARCO TEÓRICO	8
2.1. Innovación tecnológica.....	8
2.2. Indicadores compuestos: estado del arte	10
3. CONSTRUCCIÓN DEL INDICADOR COMPUESTO	11
3.1. Selección de variables	11
3.2. Categorías	12
3.3. Normalización	18
3.4. Imputación de datos faltantes	19
3.5. Correlaciones	22
3.6. Extracción de pesos factoriales	27
3.7. Agregación.....	31
3.8. Robustez y sensibilidad	33
4. CONCLUSIONES.....	40
5. DECLARACIÓN DE USO DE HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA EN TRABAJOS FIN DE GRADO	47
6. BIBLIOGRAFÍA.....	48
7. ANEXOS	52

ANEXO 1: Abreviación de variables	52
ANEXO 2: Pesos factoriales, pesos de variables y pesos dimensionales (AG 1) ..	58
ANEXO 3: Ejemplo de ecuación del indicador paso a paso para EE. UU en 2020	60
ANEXO 4: Pesos factoriales, pesos de variables y pesos dimensionales (AG 2)...	64

ÍNDICE DE SIGLAS

Terminología I+D:

BERD	<i>Business enterprise Expenditure on R&D (intramural):</i> Gastro de empresas en I+D (Intramuros)
FTE	<i>Full Time Equivalent (on R&D):</i> Equivalente a tiempo completo (en I+D)
GBARD	<i>Government Budget Allocations for R&D:</i> Asignaciones presupuestarias del gobierno para I+D
GERD	<i>Gross Domestic Expenditure on R&D (intramural):</i> Gasto Bruto Interno en I+D (intramuros)
GOVERD	<i>Government Expenditure on R&D (intramural):</i> Gasto del Gobierno en I+D (intramuros)
GUF	<i>General University Funds:</i> Fondos Generales Universitarios
HERD	<i>Higher Education Expenditure on R&D (intramural):</i> Gasto en I+D de la Educación Superior (intramuros)
NSE	<i>Natural Sciences and Engineering:</i> Ciencias Naturales e Ingeniería
PNP	<i>Private Non-Profit Institutions:</i> Instituciones Privadas sin fines de lucro
R&D	<i>Research and Development:</i> Investigación y desarrollo (I+D)
SSH	<i>Social Sciences and Humanities:</i> Ciencias Sociales y Humanidades

ÍNDICE DE TABLAS

Tabla 1. Resumen de la metodología.....	7
Tabla 2. Listado de países de la base de datos.....	11
Tabla 3. Indicadores del grupo <i>R&D General expenditure</i>	16
Tabla 4. Indicadores del <i>Research Personnel</i>	16
Tabla 5. Indicadores del grupo <i>HERD & BERD</i>	16
Tabla 6. Indicadores del grupo <i>Patents & Programmes</i>	17
Tabla 7. Indicadores del grupo <i>Trade Balance</i>	17
Tabla 8. Indicadores del grupo <i>Background Economic Indicators</i>	17
Tabla 9. Errores de imputación.....	22
Tabla 10. Indicadores del grupo <i>Input indicators</i>	34
Tabla 11. Indicadores del grupo <i>Impact indicators</i>	34
Tabla 12. Indicadores del grupo <i>Output indicators</i>	35
Tabla 13. Indicadores del grupo <i>Background Economic Indicators</i>	35
Tabla 14. Errores de imputación agrupación dos	35
Tabla 15. Factores más representativos por cada indicador en el grupo <i>Input indicators</i>	38
Tabla 16. Factores más representativos por cada indicador en el grupo <i>Impact indicators</i>	38
Tabla 17. Factores más representativos por cada indicador en el grupo <i>Output indicators</i>	39

ÍNDICE DE FIGURAS

Figura 1. Diagrama de flujo con metodología.....	7
Figura 2. Mapa de países de la base de datos	12
Figura 3. Correlaciones entre indicadores por grupo	23
Figura 4. Correlaciones entre indicadores por grupo tras aplicar la primera diferencia	25
Figura 5. Factores que explican el 80% de la varianza por cada grupo	29
Figura 6. Factores más representativos por cada indicador por grupo	30
Figura 7. Ranking de innovación tecnológica por países en 2020.....	32
Figura 8. Correlaciones por grupos	36
Figura 9. Correlaciones tras la primera diferencia por grupos	36
Figura 10. Factores que explican el 80% de la varianza por cada grupo en agrupación dos.....	37
Figura 11. Ranking de innovación tecnológica por países en 2020 con agrupación dos.....	39
Figura 12. Comparación del Indicador de Innovación Tecnológica (2020 vs. 2018 y 2015).....	41
Figura 13. Comparación del Indicador de Innovación Tecnológica para la dimensión HERD & BERD (2020 vs 2015 y 2019)	42
Figura 14. Comparación del Indicador de Innovación Tecnológica para la dimensión <i>Patents & Programmes</i> (2020 vs 2015 y 2019).....	43
Figura 15. Comparación del Indicador de Innovación Tecnológica para la dimensión <i>R&D General expenditure</i> (2020 vs 2015 y 2019).....	43
Figura 16. Comparación del Indicador de Innovación Tecnológica para la dimensión <i>Research Personnel</i> (2020 vs 2015 y 2019).....	44
Figura 17. Comparación del Indicador de Innovación Tecnológica para la dimensión <i>Trade Balance</i> (2020 vs 2015 y 2019).....	44
Figura 18. Evolución del Indicador de Innovación Tecnológica de España, Estados Unidos, Japón y Corea del Sur.....	46

ÍNDICE DE ECUACIONES

- (1) Min-Max scaling: $m = \frac{x-x_{min}}{x_{max}-x_{min}} \cdot 99 + 1$ 18
- (2) Imputación con la media: $Valor\ imputado = \mu = \frac{1}{n} \sum_{i=1}^n x_i$ 20
- (3) Imputación con KNN: $Valor\ imputado = \frac{1}{k} \sum_{i=1}^k x_{vecino\ i}$ 21
- (4) Error cuadrático medio: $ECM = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ 21
- (5) *Peso dimensional* = $P_d = \frac{1}{N}$ 31
- (6) *Peso factorial*_j = $PF_j = \frac{Varianza\ explicada_j}{\sum_j Varianza\ explicada_j}$ 31
- (7) *Peso variable*_{i,j} = $PV_{i,j} = \frac{(Carga\ factorial_{i,j})^2}{\sum_i (Carga\ factorial_{i,j})^2}$ 31
- (8) *Indicador Compuesto* = $\prod_d (\sum_j (\sum_i (V_i \cdot PV_{i,j}) \cdot PF_j))^{P_d}$ 32

1. INTRODUCCIÓN

1.1 Interés por la cuestión

La mayoría de las innovaciones que existen actualmente se manifiestan en forma de nuevos productos, procesos y modelos de negocio. Estas innovaciones suelen encontrar su fundamento en tecnologías digitales o incorporan datos o programas informáticos (OECD, 2019). Además, es importante destacar el emergente papel fundamental que juega la Inteligencia Artificial (IA) en este contexto; La aplicación de esta avanzada tecnología permite realizar investigaciones a gran escala, así como la adopción de nuevas técnicas de simulación y la creación de prototipos, elementos cruciales en el desarrollo de productos innovadores.

Así, la importancia de la innovación tecnológica se ve reflejada en varios ámbitos, como es en el mundo empresarial, ya que permite a las empresas ser más competitivas, más eficientes en los procesos de producción o la cadena de valor, y a estar más preparadas frente a cambios abruptos en el mercado; esta capacidad de adaptación atrae a clientes y potencia su fidelidad. Por otro lado, el uso de la tecnología también ayuda en la protección del medio ambiente: se están desarrollando tecnologías que ayudan a reducir la contaminación, a restaurar los ecosistemas o tecnologías que ayudan a proteger animales en peligro de extinción. Podemos ver la aplicación de la tecnología en el uso de drones y sensores que detectan la tala de árboles ilegal, o la caza en áreas protegidas. Además, con el uso de la IA, se pueden analizar imágenes de los satélites y detectar cambios en la calidad del agua (Rivera, 2023).

Comprendiendo esta situación en la que nos encontramos, existe un increíble número de indicadores que tratan de medir el nivel de innovación tecnológica en diferentes países, ya sea por ejemplo basándose en la inversión en I+D o en el número de investigadores dedicados a este ámbito (OECD, 2023). De esta manera, en el presente trabajo se agregarán los indicadores considerados más relevantes, empleando modelos de ML para el cálculo/asignación de pesos a dichas variables, con el fin de crear un indicador compuesto que nos permita comparar de manera sencilla la innovación tecnológica entre distintas naciones.

1.2 Objetivos del proyecto

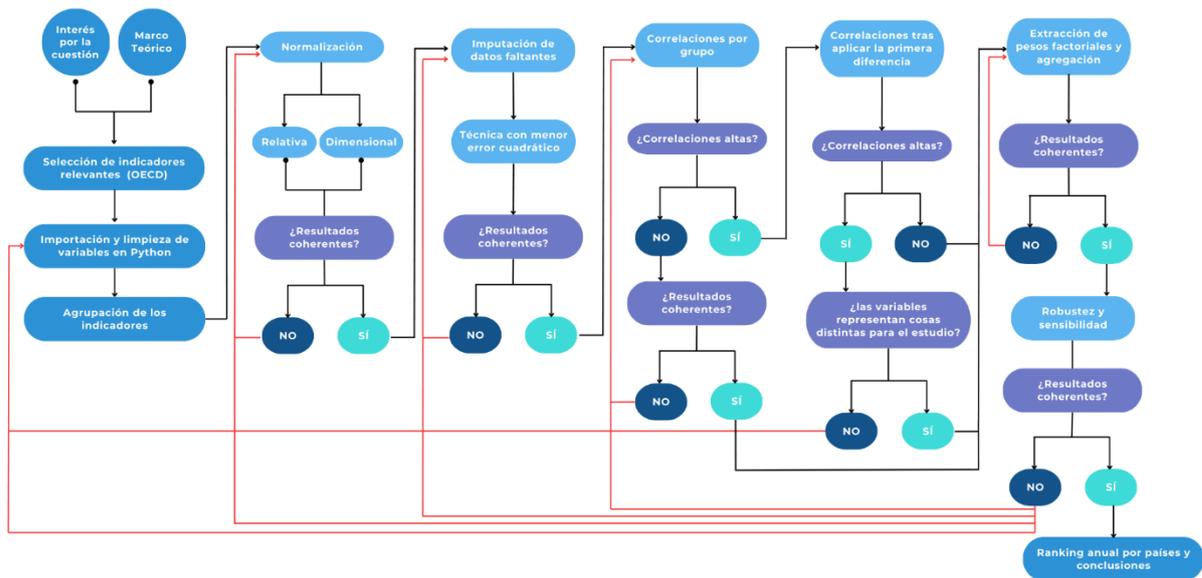
El principal objetivo de este proyecto es elaborar un indicador compuesto que mida la innovación tecnológica entre países. Para lograr este propósito, habrá que cumplir con los siguientes objetivos específicos, que constituyen los pasos para la creación del indicador:

- Selección de variables: elaborar un listado de variables potenciales a partir de una revisión bibliográfica, ver que variables son más representativas y por qué lo son.
- Imputación de los datos faltantes: en el caso de que haya datos vacíos, tratarlos con los 3 métodos generales que son eliminación de casos, imputación única, o imputación múltiple.
- Análisis multivariable: analizar las relaciones entre las variables del conjunto de datos y extraer información. Este paso ayudará a elegir las metodologías de, por ejemplo, la ponderación y agregación durante la fase de construcción del indicador.
- Normalización de los datos: normalización de los datos mediante distintos métodos, como pueden ser el ranking, la estandarización o el Min-Max. Es un paso necesario previo a la agregación ya que los indicadores en los data sets suelen tener distintas unidades de medida.
- Ponderación y agregación: mediante modelos de ML, adjudicar pesos a las distintas variables, basándonos en la importancia que tiene cada variable en la construcción del indicador. Para ello existen varios métodos como el Análisis de Componentes Principales (PCA), Modelos de Componentes no Observadas o el análisis conjunto.
- Robustez y sensibilidad: ayuda a calibrar la solidez del indicador. Lo ideal sería abordar todas las fuentes potenciales de incertidumbre, como son la selección de indicadores individuales, calidad de los datos, normalización, ponderación, método de agregación, etc.

1.3 Metodología

La metodología seguida para la elaboración del indicador compuesto ha sido resumida en el siguiente diagrama de flujo:

Figura 1. Diagrama de flujo con metodología



Fuente: Elaboración propia

Además, en la siguiente tabla he resumido la metodología empleada en las tres grandes secciones del trabajo:

Tabla 1. Resumen de la metodología

MARCO TEÓRICO	CONSTRUCCIÓN DEL INDICADOR	CONCLUSIONES
FUENTES ACADÉMICAS Y DOCUMENTALES: • GOOGLE SCHOLAR • DIALNET	PYTHON HANDBOOK ON CONSTRUCTING COMPOSITE INDICATORS	VISUALIZACIONES CON PYTHON

Fuente: Elaboración propia

2. MARCO TEÓRICO

2.1. Innovación tecnológica

La innovación se define como la implementación de productos nuevos o considerablemente mejorados, servicios, procesos, métodos de comercialización, o prácticas organizativas, ya sea dentro de las operaciones internas de una empresa, en la organización de su lugar de trabajo, o en sus relaciones con el exterior (OECD, 2007). Esto implica que, para que se considere una innovación, debe haber un cambio en el producto, proceso, método de comercialización, o estructura organizativa.

Además, se entiende por actividades innovadoras aquellas operaciones, ya sean científicas, tecnológicas, organizativas, financieras, o comerciales, destinadas o encaminadas hacia la innovación (OECD, 2007).

Por tanto, ¿qué constituye la innovación tecnológica? Este concepto comenzó a ganar relevancia en la década de 1960, marcando el inicio de un movimiento que posicionaba a la innovación tecnológica como un pilar clave para la prosperidad de las naciones. Durante este período, las naciones desarrolladas empezaron a vincular directamente la inversión en tecnología con el surgimiento de innovaciones, resultando en un aumento significativo en los gastos de investigación y desarrollo (I+D). Sin embargo, dichas inversiones no siempre se tradujeron en éxitos inmediatos, ya que las empresas a menudo operaban de manera poco sistemática y sin un control riguroso de sus procesos de innovación. Esta situación, sumada a las crisis económicas de los años 70, llevó a una reducción en los presupuestos de I+D, a pesar del reconocimiento de la importancia estratégica de la innovación tecnológica por parte de los líderes empresariales (Ruiz y Mandado, 1989).

Con el tiempo, se ha demostrado que, para aumentar la productividad en el departamento de I+D, es imprescindible adoptar una estrategia definida previamente, comprender a fondo el proceso de innovación tecnológica y considerar la tecnología como un elemento estratégico vital.

La teoría de la difusión de innovaciones, propuesta por (Rogers, 2003), proporciona un marco valioso para comprender cómo y por qué las nuevas tecnologías y prácticas se adoptan dentro de diferentes comunidades y sectores. Según Rogers, la adopción de innovaciones se ve influenciada por factores como la ventaja percibida, la compatibilidad, la complejidad, la posibilidad de experimentar y la observabilidad. Este

enfoque subraya la importancia de considerar no solo la naturaleza intrínseca de una innovación tecnológica sino también el contexto en el que se introduce.

Por otro lado, el concepto de innovación abierta, introducido por (Chesbrough, 2003), desafía el modelo tradicional de innovación cerrada, argumentando que las organizaciones pueden y deben utilizar ideas externas así como internas en sus procesos de innovación. Este enfoque refleja el creciente reconocimiento de que el conocimiento para impulsar la innovación tecnológica a menudo se encuentra más allá de los límites de la organización, y que la colaboración entre diferentes actores económicos puede acelerar el desarrollo y la aplicación de nuevas tecnologías.

La medición de la innovación tecnológica es un desafío complejo que ha sido abordado por instituciones como la OCDE y el Banco Mundial. Los indicadores utilizados para evaluar la innovación incluyen medidas de input, como el gasto en investigación y desarrollo (I+D), y de output, como el número de patentes registradas y artículos científicos publicados. Estos indicadores son fundamentales para comprender el dinamismo de innovación de los países y para establecer comparaciones internacionales sobre la capacidad y el desempeño innovador.

La relación entre innovación tecnológica y desarrollo económico ha sido ampliamente examinada en la literatura económica, destacando el trabajo de economistas como Schumpeter, (1943) y el desarrollo posterior de la teoría del crecimiento endógeno por Aghion & Howitt, (1997). Schumpeter articuló la idea de la "destrucción creativa", donde la innovación tecnológica actúa como una fuerza disruptiva que desplaza a las tecnologías, productos y servicios existentes, impulsando así el crecimiento económico a través de la generación de nuevas industrias y oportunidades de empleo. Aghion y Howitt profundizaron en cómo la innovación es impulsada por la expectativa de ganancias económicas, alimentando un ciclo de inversión en I+D y avances tecnológicos que sostiene el crecimiento económico a largo plazo.

Finalmente, la capacidad de un país para innovar no solo se mide por su producción tecnológica sino también por su capacidad para integrar y aplicar estas innovaciones de manera que fomente el desarrollo sostenible y la inclusión social. El Índice Global de Innovación y el Informe de Competitividad Global (Schwab, 2019) ofrecen perspectivas esenciales sobre cómo los países se desempeñan en estas

dimensiones, destacando la importancia de un enfoque holístico para evaluar el ecosistema de innovación.

2.2. Indicadores compuestos: estado del arte

Los indicadores compuestos se definen como medidas compuestas que integran y resumen varios indicadores individuales con el fin de representar conceptos complejos en un formato simplificado (Nardo et al., 2005). Estos indicadores son especialmente útiles para capturar las múltiples dimensiones de fenómenos tales como el desarrollo económico, la innovación tecnológica, la calidad de vida o la sostenibilidad ambiental. La capacidad de sintetizar información relevante en un solo valor facilita la comparación entre diferentes unidades de análisis, como pueden ser países, regiones o instituciones, y constituye una herramienta valiosa para la toma de decisiones basada en evidencias.

La metodología para construir indicadores compuestos implica varios pasos críticos, que incluyen la selección de los indicadores individuales relevantes, la normalización de estos indicadores para hacerlos comparables, la asignación de pesos que reflejen su importancia relativa y la agregación de los valores normalizados en un único indicador compuesto (OECD, 2008). Este proceso metodológico debe ser riguroso y transparente para asegurar que el indicador compuesto resultante sea válido, fiable y relevante para los objetivos del estudio deseado.

La aplicación de indicadores compuestos es amplia y variada, extendiéndose a través de múltiples disciplinas y sectores. Por ejemplo, en el ámbito de la política, los indicadores compuestos se emplean para evaluar y monitorear el progreso hacia metas y objetivos estratégicos, como pueden ser los Objetivos de desarrollo sostenible elaborados por las Naciones Unidas (United Nations, 2015). Por otro lado, en el contexto empresarial, facilitan la evaluación de la innovación, la competitividad y el desempeño general de las organizaciones.

Sin embargo, la construcción y aplicación de indicadores compuestos no están exentas de desafíos y limitaciones. La selección de indicadores, la ponderación y la agregación pueden introducir sesgos y subjetividad en el análisis. Además, la simplificación que ofrecen puede ocultar variaciones importantes y detalles críticos de los fenómenos estudiados (Saltelli, 2007). Por lo tanto, es esencial abordar estos indicadores con un entendimiento crítico de sus fortalezas y debilidades, y considerarlos como una herramienta entre muchas en el arsenal de la investigación.

3. CONSTRUCCIÓN DEL INDICADOR COMPUESTO

3.1. Selección de variables

Las variables identificadas para la construcción del indicador compuesto de innovación tecnológica se derivan de los principales indicadores de Ciencia y Tecnología (MSTI) publicados por la Organización para la Cooperación y el Desarrollo Económicos (OECD) (OECD, 2023). La base de datos correspondiente a los MSTI es empleada por la OECD como un punto de referencia para las estadísticas internacionales sobre Investigación y Desarrollo (I+D).

Los MSTI son publicados de manera semestral, en marzo y septiembre: La edición de marzo implica una revisión de los datos anteriores y proporciona indicadores agregados más actualizados, aunque de naturaleza provisional. En contraste, los indicadores presentados en la edición de septiembre suelen basarse en datos confirmados por los países, considerándose como definitivos. Además, esta edición incluye indicadores más detallados que han sido revisados por la Secretaría a lo largo del año. Esta última edición se destaca por contar con información más completa sobre los indicadores de presupuestos públicos destinados a la I+D por parte de los países miembros de la OECD (OECD, 2023).

Por consiguiente, la base de datos empleada en el presente estudio corresponde a los MSTI de septiembre de 2023, la cual abarca información relativa a un conjunto de 45 países (véase tabla 1), habiéndose excluido de la base de datos para el análisis del presente trabajo los campos de ‘OECD - Total’ y ‘European Union – 27 countries’.

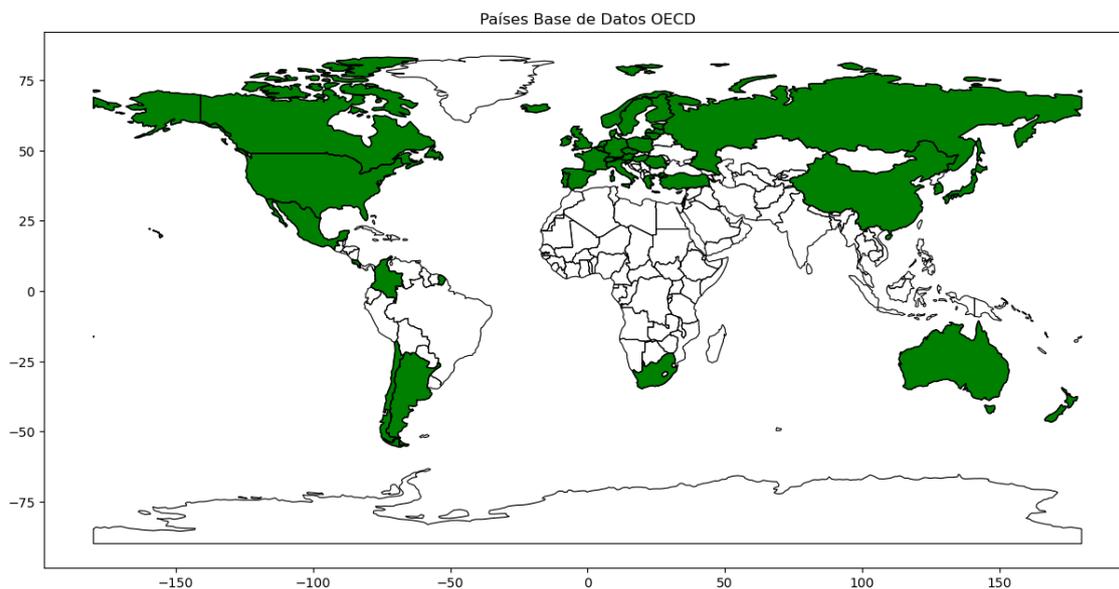
Tabla 2. Listado de países de la base de datos

Pais	Continente	Bandera	Pais	Continente	Bandera	Pais	Continente	Bandera	Pais	Continente	Bandera
South Africa	Africa	ZA	Finland	Europe	FI	Netherlands	Europe	NL	Russia	Europe/Asia	RU
China	Asia	CN	France	Europe	FR	Norway	Europe	NO	Turkey	Europe/Asia	TR
Israel	Asia	IL	Germany	Europe	DE	Poland	Europe	PL	Canada	North America	CA
Japan	Asia	JP	Greece	Europe	GR	Portugal	Europe	PT	Costa Rica	North America	CR
Singapore	Asia	SG	Hungary	Europe	HU	Romania	Europe	RO	Mexico	North America	MX
South Korea	Asia	KR	Iceland	Europe	IS	Slovakia	Europe	SK	USA	North America	US
Austria	Europe	AT	Ireland	Europe	IE	Slovenia	Europe	SI	Australia	Oceania	AU
Belgium	Europe	BE	Italy	Europe	IT	Spain	Europe	ES	New Zealand	Oceania	NZ
Czechia	Europe	CZ	Latvia	Europe	LV	Sweden	Europe	SE	Argentina	South America	AR
Denmark	Europe	DK	Lithuania	Europe	LT	Switzerland	Europe	CH	Chile	South America	CL
Estonia	Europe	EE	Luxembourg	Europe	LU	UK	Europe	GB	Colombia	South America	CO

Fuente: Elaboración propia

En la figura 1 podremos ver el listado de países representado en un mapamundi, observando que la mayor parte de los países estudiados son europeos.

Figura 2. Mapa de países de la base de datos



Fuente: Elaboración propia

En la presente base de datos se ha recopilado un conjunto de 133 indicadores que reflejan el nivel y la estructura de los esfuerzos realizados por los países miembros de la OECD y varias economías no miembros en el ámbito de la ciencia y la tecnología. Los indicadores abarcan los recursos dedicados a la investigación y el desarrollo, las familias de patentes, la balanza de pagos tecnológica y el comercio internacional en industrias intensivas en I+D. También se presentan las series económicas subyacentes utilizadas para calcular estos indicadores, como puede ser el PIB de cada país, o su población. Las series se presentan para un año de referencia y para años seleccionados de los que se dispone de datos (OECD, 2023).

3.2. Categorías

Previo a la categorización de las variables, se realiza una depuración de aquellas columnas que carecen de relevancia para el análisis, es decir, las que no contribuyen a la generación de información pertinente para la formación del indicador. En este contexto, dichas variables comprenden: "COU", "YEAR", "Unit Code", "Unit", "PowerCode Code", "PowerCode", "Reference Period Code", "Reference Period", "Flag Codes", "Flags".

Además, teniendo en cuenta que la base de datos tiene información de los años 2015 al 2023, se han eliminado los años 2022 y 2023 por falta de datos, y de la columna países se ha excluido la media europea y el total de la OECD.

Tras completar este proceso de depuración, la primera medida implementada ha consistido en llevar a cabo una agrupación de las variables con el propósito de clasificarlas de manera más accesible y facilitar su análisis. Dicha agrupación se ha realizado de manera manual, asignando cada indicador a su respectiva categoría o familia. A modo de ejemplo, todos los indicadores relacionados con investigadores han sido asignados al grupo tres denominado "*Research Personnel*".

En nuestra base de datos, se introduce una variable adicional denominada "*Groups*", donde se han establecido las siguientes categorías:

- Grupo 1, **HERD & BERD (HERD - *Higher Education Research & Development*, BERD - *Business Enterprise Research & Development*)**: Engloba todos los indicadores vinculados a la inversión en I+D en el ámbito de la enseñanza superior y el sector empresarial.
- Grupo 2, ***Trade Balance (Balanza Comercial)***: Incluye datos relativos a las importaciones, exportaciones y la balanza comercial de los países en cada periodo anual.
- Grupo 3, ***Research Personnel (Personal de investigación)***: Proporciona datos acerca del número de investigadores presentes en diversos sectores, como el universitario o el empresarial.
- Grupo 4, ***Patents & Programmes (Patentes y programas)***: Incorpora indicadores relacionados con las patentes emitidas por diferentes países, como el número de patentes triádicas o aquellas emitidas por sectores específicos. Además, incluye programas relacionados con la salud, medio ambiente, investigaciones sin orientación específica y programas espaciales.
- Grupo 5, ***R&D General expenditure (Gasto general en I+D)***: Este grupo engloba indicadores que no pueden asignarse a ninguno de los grupos anteriores y hace referencia a gastos generales en I+D, presupuestos destinados a la investigación y desarrollo, entre otros.
- Anexo, ***Background Economic Indicators (Indicadores económicos de contexto)***: Este último grupo recopila todos los indicadores económicos de

referencia necesarios, como la población de cada país o su Producto Interno Bruto (PIB). Este grupo se va a emplear únicamente para la normalización de los datos, posteriormente para la creación del indicador se eliminará, ya que no son indicadores que midan la innovación tecnológica de un país.

Una vez completada la categorización de los indicadores en grupos, procedemos a realizar una limpieza de las variables dentro de cada grupo, excluyendo indicadores que proporcionan información redundante, para así para evitar la duplicidad de información y asegurar la precisión y la claridad del análisis.

En el caso específico del grupo de *Trade Balance*, como se ha mencionado previamente, se incluyen datos relativos a importaciones, exportaciones, la balanza comercial de ciencia y tecnología y la cuota de mercado de las exportaciones. En este escenario, se ha optado por eliminar las importaciones y exportaciones, conservando la información de la balanza comercial de cada país y sus *exports market share*. La razón detrás de esta decisión radica en que, por un lado, la balanza comercial ya proporciona una visión clara del neto entre importaciones y exportaciones: esta medida refleja directamente el resultado final de la actividad comercial de un país en una industria específica, lo que es crucial para entender la posición competitiva del país en ese sector. Por otro lado, el "*export market share*" es un indicador particularmente útil cuando se evalúa la innovación tecnológica. Este indicador muestra qué tan significativas son las exportaciones de un país en el mercado global en comparación con otros países. Un alto *market share* puede indicar que el país es líder en esa tecnología o producto específico, lo cual es un resultado directo de innovaciones competitivas y efectivas.

Por otro lado, en cada uno de los grupos, con la excepción del conjunto de *Background Economic Indicators*, se ha observado la presencia de una variable común representada de diversas formas. Poniendo como ejemplo el grupo HERD y BERD, se observa que ambas variables se encuentran expresadas como porcentaje del PIB, en moneda nacional, como porcentaje del valor añadido en la industria y en precios constantes y PPP\$. Dada la redundancia, se ha tomado la decisión de elegir una única manera para representar la variable.

La representación seleccionada ha sido medir la inversión en I+D en precios constantes y ajustados por paridad de poder adquisitivo. Los expertos recomiendan emplear esta metodología cuando se busca realizar comparaciones entre países en

relación con variables como los salarios o el PIB, entre otros indicadores económicos (Leiva, 2023). La aplicación de esta metodología permite normalizar los datos, eliminando factores que podrían distorsionar las comparaciones directas, como las variaciones en las monedas nacionales, las diferencias en el poder adquisitivo, y otras consideraciones que podrían afectar la precisión de las comparaciones internacionales. Al centrarse en porcentajes, tasas o valores relativos, en lugar de cifras absolutas, se facilita una evaluación más precisa y equitativa de las condiciones económicas entre países, lo que contribuye a una interpretación más robusta de las disparidades y similitudes a nivel global (S, 2021).

Aunque los porcentajes del PIB pueden ofrecer una idea de la importancia relativa de la investigación y el desarrollo en la economía de un país, es posible que no tengan plenamente en cuenta las variaciones del coste de la vida y la inflación. Por lo tanto, para crear un indicador compuesto que permita comparaciones entre países, los valores a precios constantes y PPP\$ serían generalmente una opción más adecuada.

En cuanto al grupo de *Patents & Programmes*, se ha eliminado el indicador de "Número de familias de patentes triádicas (año de prioridad)" debido a su redundancia con el indicador de "Participación de países en familias de patentes triádicas (año de prioridad)". La información proporcionada por la participación de países ofrece una medida relativa y comparativa más relevante de la presencia global e impacto de la innovación de cada país, por este motivo se ha decidido mantener dicho indicador y eliminar el otro. Manteniendo indicadores que reflejan las solicitudes de patentes bajo el PCT y en sectores específicos, se mantiene una visión detallada de la actividad innovadora (WIPO, 2023).

Como último paso en esta limpieza de variables, en el grupo de *Research Personnel* se han eliminado variables relacionadas con la tasa anual compuesta de crecimiento, investigadores por cada mil de la fuerza laboral, investigadores como porcentaje del total nacional, e investigadores (en términos de conteo absoluto y desglose por género) en diferentes sectores (académico, gubernamental y empresarial), así como indicadores expresados en equivalentes a tiempo completo (FTE). Se ha llevado a cabo dicha eliminación ya que se ha decidido enfocar el análisis en indicadores que proporcionen una visión más directa y comparativa del personal dedicado a la

investigación y desarrollo (I+D), su crecimiento, y la representación de mujeres en estos roles.

Tras todos estos cambios, cada grupo ha quedado de la siguiente manera:

Tabla 3. Indicadores del grupo *R&D General expenditure*

Tabla para R&D General expenditure:

	Indicators
0	GERD at constant prices and PPP \$
1	Percentage of GERD financed by the business enterprise sector
2	Percentage of GERD financed by government
3	Percentage of GERD financed by the Higher Education and PNP sectors
4	Percentage of GERD financed by the rest of the world
5	Percentage of GERD performed by the Business Enterprise sector
6	Percentage of GERD performed by the Higher Education sector
7	Percentage of GERD performed by the Government sector
8	Percentage of GERD performed by the Private Non-Profit sector
9	GOVERD at constant prices and PPP \$
10	Percentage of GOVERD financed by the business sector
11	Defence Budget R&D as a percentage of Total GBARD
12	Civil Budget R&D as a percentage of Total GBARD
13	Civil GBARD for General University Funds (GUF) (current PPP\$)
14	General University Funds (GUF) as a percentage of Civil GBARD
15	Total GBARD at constant prices and PPP \$

Fuente: Elaboración propia

Tabla 4. Indicadores del *Research Personnel*

Tabla para Research Personnel:

	Indicators
0	Total researchers per thousand total employment
1	Total R&D personnel per thousand total employment
2	Total researchers (headcount)
3	Woman researchers (headcount)
4	Women researchers as a percentage of total researchers (headcount)
5	Business enterprise sector: Women researchers as a percentage of total researchers (headcount)
6	Government sector: Women researchers as a percentage of total researchers (headcount)
7	Higher education sector: Women researchers as a percentage of total researchers (headcount)
8	Business Enterprise researchers per thousand employment in industry
9	Total Business Enterprise personnel per thousand employment in industry

Fuente: Elaboración propia

Tabla 5. Indicadores del grupo *HERD & BERD*

Tabla para HERD & BERD:

	Indicators
0	BERD at constant prices and PPP \$
1	Business-financed BERD at constant prices and PPP \$
2	Percentage of BERD financed by the business sector
3	Percentage of BERD financed by government
4	Percentage of BERD financed by the Higher Education and PNP sectors
5	Percentage of BERD financed by the rest of the world
6	BERD performed in the aerospace industry (current PPP \$)
7	Percentage of BERD performed in the aerospace industry
8	BERD performed in the pharmaceutical industry (current PPP \$)
9	Percentage of BERD performed in the pharmaceutical industry
10	BERD performed in service industries (current PPP \$)
11	Percentage of BERD performed in service industries
12	HERD at constant prices and PPP \$
13	Percentage of HERD financed by the business sector
14	Percentage of BERD performed in the computer, electronic and optical industry
15	BERD performed in the computer, electronic and optical industry (current PPP \$)

Fuente: Elaboración propia

Tabla 6. Indicadores del grupo *Patents & Programmes*

Tabla para Patents & Programmes:

Indicators	
0	Civil GBARD for Economic Development programmes (current PPP\$)
1	Economic Development programmes as a percentage of Civil GBARD
2	Civil GBARD for Health and Environment programmes (current PPP\$)
3	Health and Environment programmes as a percentage of Civil GBARD
4	Civil GBARD for Space programmes (current PPP\$)
5	Space programmes as a percentage of Civil GBARD
6	Civil GBARD for Non-oriented Research programmes (current PPP\$)
7	Non-oriented Research programmes as a percentage of Civil GBARD
8	Number of "triadic" patent families (priority year)
9	Education and Social programmes as a percentage of Civil GBARD
10	Civil GBARD for Education and Social programmes (current PPP\$)
11	Number of patent applications filed under the PCT(priority year)
12	Number of patents in the biotechnology sector - applications filed under the PCT (priority year)
13	Number of patents in the ICT sector - applications filed under the PCT (priority year)

Fuente: Elaboración propia

Tabla 7. Indicadores del grupo *Trade Balance*

Tabla para Trade Balance:

Indicators	
0	Export market share: Aerospace industry
1	Trade Balance: Aerospace industry (current prices)
2	Export market share: Pharmaceutical industry
3	Trade Balance: Pharmaceutical industry (current prices)
4	Trade Balance: Computer, electronic and optical industry (current prices)
5	Export market share: Computer, electronic and optical industry

Fuente: Elaboración propia

Tabla 8. Indicadores del grupo *Background Economic Indicators*

Tabla para Background Economic Indicators:

Indicators	
0	Gross Domestic Product (current PPP\$)
1	Value Added of Industry (current PPP\$)
2	Total Employment
3	Population
4	Implicit GDP Price Indices
5	Industrial Employment
6	Purchasing Power Parity
7	Labour Force

Fuente: Elaboración propia

3.3. Normalización

La siguiente fase llevada a cabo ha consistido en la normalización de las variables. En el análisis, cada observación en el *DataFrame* representa un indicador relacionado con ciencia y tecnología de un país para un año específico. Dichos indicadores poseen valores expresados en diversas unidades, como pueden ser unidades monetarias, porcentajes o número de personas, entre otros.

En este enfoque de normalización de datos, se han diferenciado dos tipos principales: la normalización relativa y la normalización dimensional. Inicialmente, se ha aplicado la normalización relativa a dos grupos de datos: *Patents & Programmes* (únicamente a las variables de patentes) y el grupo de *Research Personnel*. Para estos grupos, se han ajustado los datos a patentes per cápita y personal de investigación per cápita, respectivamente. Esta metodología ofrece una visión más detallada y permite comparaciones más precisas entre los diferentes países (Banton, 2022).

Posteriormente, al preparar la base de datos para la agregación y creación del indicador, se ha implementado la normalización dimensional a través del método de *Min-Max Scaling*. Esta técnica tiene como objetivo principal estandarizar los datos, reduciendo de esta manera el impacto que las disparidades en la escala y la magnitud de los atributos podrían tener en el análisis (Blanco, 2023).

Es relevante destacar que existen diversos métodos para llevar a cabo la normalización de datos, y la elección de una metodología específica depende del objetivo del análisis y la naturaleza intrínseca de los datos en consideración (OECD, 2008). Al emplear *Min-Max Scaling* no solo se estandarizan los datos a una escala común, sino que también se evita la pérdida de impacto de los valores bajos en la suma geométrica de variables, lo cual es crucial para la creación de un indicador compuesto más representativo. La técnica consiste en sustituir cada valor de una columna por un nuevo valor empleando la siguiente fórmula:

$$m = \frac{x - x_{min}}{x_{max} - x_{min}} \cdot 99 + 1$$

Donde

- m es el nuevo valor
- x es el valor original
- x_{min} es el valor mínimo de la columna

- x_{max} es el valor máximo de la columna

Al aplicar esta técnica de normalización, los valores de cada columna serán de entre 0 y 1, pero, para este trabajo se han modificado los límites deseados, cambiándolos de 0-1 a 1-100, ¿el motivo? es una estrategia para evitar la pérdida de información cuando se agregan variables geoméricamente, paso que será explicado más adelante, permitiendo una representación más precisa de los datos en el indicador compuesto.

Si se empleara la normalización estándar de 0-1 y alguna de las variables tuviera un valor de 0, al agregar geoméricamente las variables, el producto de esa variable con cualquier otra sería 0, lo que resultaría en que el indicador compuesto completo también fuera 0, incluso si las otras variables tienen valores significativos. De esta manera, al ampliar el rango de normalización a 1-100, se asegura que incluso si una variable tiene un valor muy bajo, aún contribuirá en cierta medida al indicador compuesto final, permitiendo así una representación más precisa de la información.

3.4. Imputación de datos faltantes

El siguiente paso en el tratamiento de los datos es la imputación de datos faltantes, ya que la falta de datos suele dificultar la elaboración de indicadores compuestos sólidos. Esta falta de datos puede de forma aleatoria o no aleatoria (OECD, 2008). Existen tres procedimientos generales para imputar los datos faltantes:

- Eliminación
- Imputación única
- Imputación múltiple

La eliminación consiste en trabajar únicamente con las observaciones que contienen información completa, es decir, eliminar todas aquellas observaciones que tienen valores nulos. Esta metodología asume que los datos faltantes tienen las mismas características que los datos que están completos y que la ausencia de datos se generó de manera aleatoria. Si embargo, para la elaboración de este indicador se ha decidido emplear técnicas de imputación única e imputación múltiple.

Imputación única:

Consiste en considerar los datos que faltan como parte del análisis e intentar imputar los valores mediante imputación única (mismo valor para todos los datos faltantes)

- Imputación con la media: se asume que los datos faltantes suceden de manera aleatoria. Consiste en imputar el valor promedio de las observaciones que disponen de información. Con esta metodología el valor medio de las variables se conserva, pero otros estadísticos como la varianza o el sesgo pueden ser afectados.

$$\text{Valor imputado} = \mu = \frac{1}{n} \sum_{i=1}^n x_i$$

Donde

- μ es la media de los valores no nulos
 - x_i son los valores observados
 - n es el número de observaciones no nulas
- Imputación con la mediana: Implica reemplazar los datos faltantes con el valor de la mediana de las observaciones de las que se dispone información, que se encuentra en el centro de un conjunto de datos ordenados, dividiéndolo en dos partes iguales. La mediana no se influencia por valores extremos (*outliers*), por lo que es útil en la presencia de *outliers* (Urrego, 2023).

Imputación múltiple:

Se realizan varias imputaciones de las observaciones faltantes, para posteriormente analizar los conjuntos de datos completados y combinar los resultados obtenidos, para obtener una estimación final (Román Radío, 2017).

- Imputación con valores aleatorios: El método busca completar registros vacíos usando datos completos de otros campos mediante una selección aleatoria, evitando sesgar la variabilidad del estimador al introducir valores de forma impredecible. Esto asegura una imputación equitativa sin patrones predecibles.

- Imputación KNN: consiste en asignar a cada dato faltante un valor obtenido a partir de la información disponible de los k vecinos más cercanos. Esta técnica ayuda a conservar la estructura original de los datos y evita distorsionar la distribución de la variable imputada (Chiapella, 2020).

$$\text{Valor imputado} = \frac{1}{k} \sum_{i=1}^k x_{\text{vecino } i}$$

Donde

- k es el número de vecinos más cercanos a considerar
- $x_{\text{vecino } i}$ son los valores de los k vecinos más cercanos

Para implementar estas estrategias en nuestra base de datos en Python, inicialmente se ha generado una réplica del *DataFrame*, eliminando los valores nulos de la columna denominada 'Value'. En este nuevo *DataFrame* se han suprimido un 10% de los valores en la columna 'Value' y se le ha sometido a las técnicas de imputación de dichos datos faltantes. Posteriormente, se ha llevado a cabo una comparación entre los valores asignados y los originales, permitiéndonos calcular el error cuadrático medio (ECM) asociado a cada técnica utilizada.

El ECM es una métrica fundamental en el análisis estadístico y de datos, utilizada para cuantificar la diferencia entre valores predichos y los valores reales u observados. Es especialmente útil en contextos donde es crucial medir la precisión de modelos predictivos o técnicas de imputación.

La ecuación del ECM se define como la media de los cuadrados de las diferencias entre los valores observados (y_i) y los valores imputados o predichos (\hat{y}_i), expresada como:

$$ECM = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Donde

- n es la media de los valores no nulos

En este análisis, el ECM proporciona una medida cuantitativa del error introducido por diferentes técnicas de imputación al reemplazar los valores nulos. Al minimizar el ECM, se busca la técnica de imputación que más se acerque a los valores reales.

Una vez que hayamos calculado el ECM para cada técnica de imputación, la estrategia seleccionada para este estudio será aquella que presente el menor ECM (véase tabla 9). En este contexto, la imputación mediante la media ha demostrado ser la más eficaz, al registrar el menor valor de error de imputación.

Tabla 9. Errores de imputación

Técnica de Imputación	Error (MSE)
Imputación Media	0.791703032
Imputación Mediana	0.793086364
Imputación Aleatoria	1.265112063
Imputación kNN	0.791730007

Fuente: Elaboración propia

Tras identificar que la imputación con la media es la técnica con el menor ECM, dicha técnica ha sido la empleada para imputar los datos faltantes de toda la base de datos.

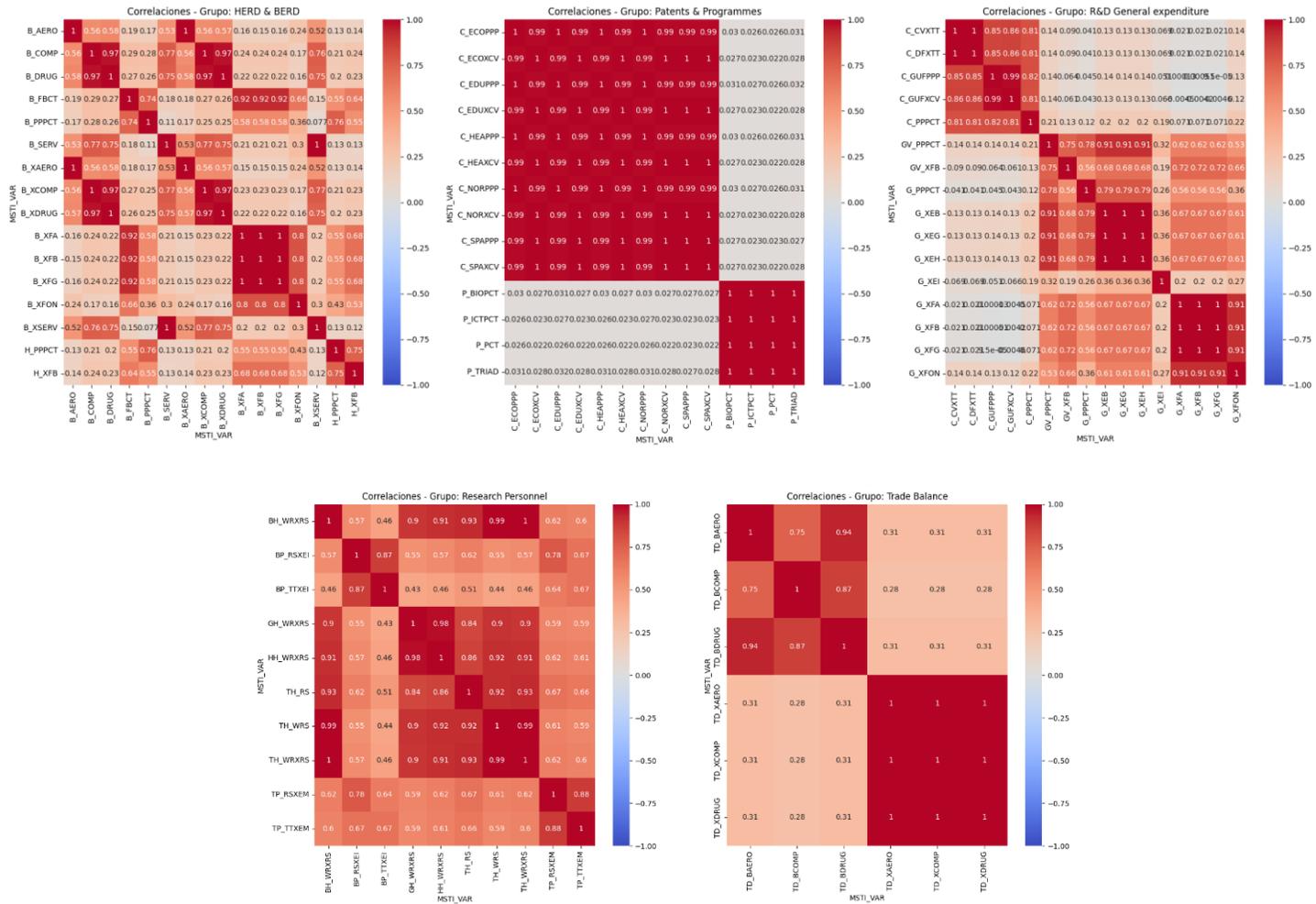
3.5. Correlaciones

Una vez hemos completado la normalización e imputación de nuestros datos, procedemos a calcular las correlaciones entre los indicadores dentro de cada grupo. Este paso es crucial para la posterior extracción de pesos factoriales a través de *Factor Analysis* (FA), ya que aquellos indicadores que tengan una correlación alta serán agrupados en el FA, repartiendo el peso de cada factor entre los indicadores más representativos dentro del mismo.

Para realizar el cálculo de la matriz de correlación por grupos, hemos creado un diccionario que almacena un *DataFrame* por cada grupo. Esto nos permite iterar y calcular las correlaciones entre cada par de indicadores dentro de su grupo respectivo.

Como podemos observar en la Figura 3, hay en varios grupos donde existen correlaciones muy altas entre indicadores: *Research Personnel, Patents & Programmes* y *Trade Balance*. Otros grupos como el de *HERD & BERD* también cuentan con correlaciones altas en algunas de sus variables, pero no son tan significativas.

Figura 3. Correlaciones entre indicadores por grupo



Fuente: Elaboración propia

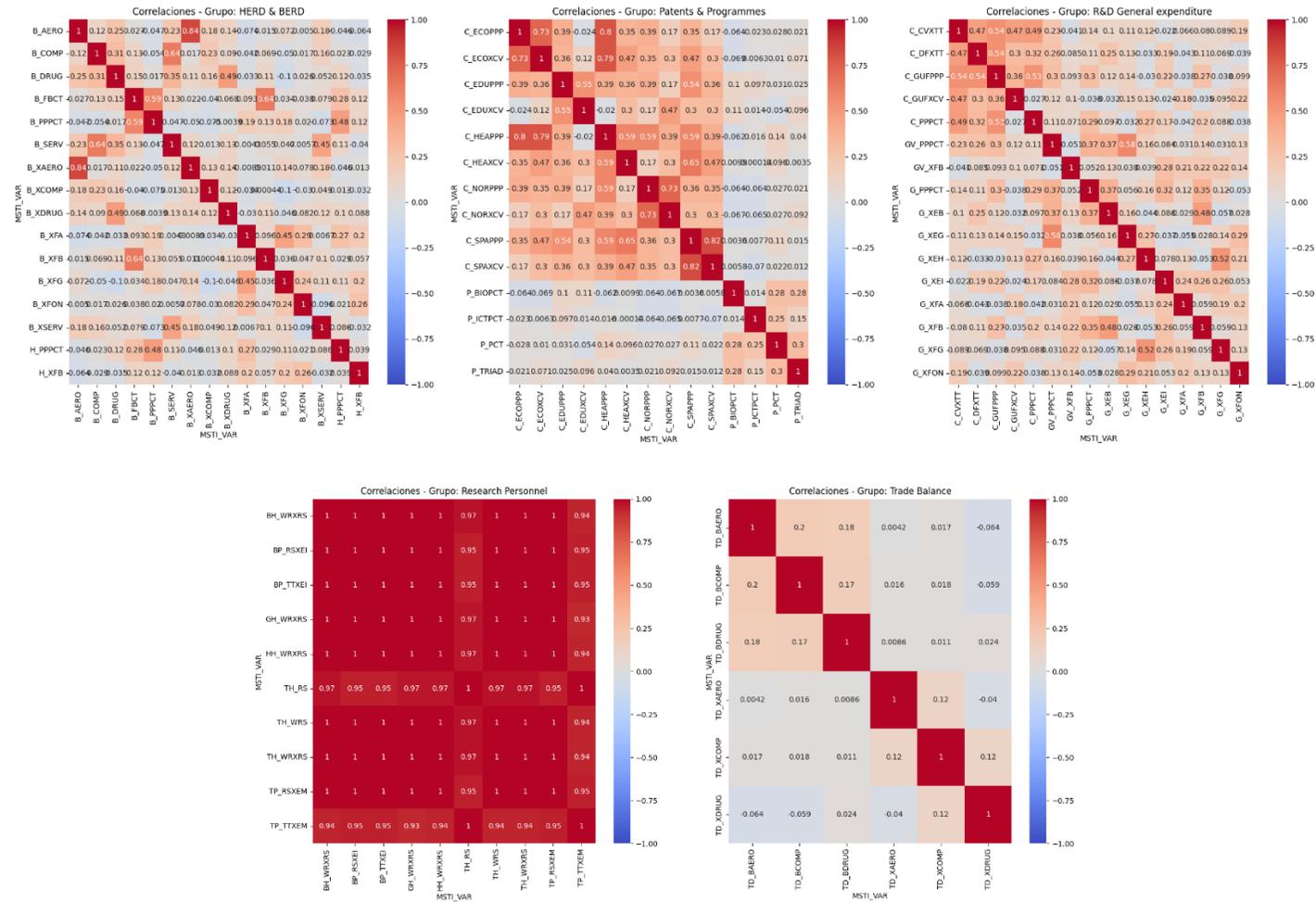
Al existir correlaciones altas entre variables en varios grupos, se ha querido comprobar si estas correlaciones son reales o si se tratan de correlaciones espurias. Una correlación espuria es “una correlación entre dos o más variables que no implica la determinación causal de una sobre la otra, sino el resultado de la covariación con terceras variables que no consideramos en el análisis” (de los Campos, 2021). De esta manera, una correlación espuria ocurre cuando dos variables parecen estar correlacionadas (positiva o negativamente) pero no hay una relación causal o lógica que explique esta asociación, la correlación es más bien una coincidencia o está influenciada por otra variable no considerada en el análisis.

Se ha realizado este análisis ya que, en el contexto de la creación de nuestro indicador compuesto en el que se va a emplear FA, tener un grupo de variables con correlaciones muy altas puede indicar multicolinealidad, lo cual significa que las variables están tan relacionadas entre sí que es difícil distinguir sus efectos individuales. Esto puede complicar la interpretación de los factores resultantes, ya que varios indicadores estarían midiendo esencialmente lo mismo.

Además, es recomendable que ambas series temporales sean "estacionarias". Esto quiere decir que, en términos generales, sean constantes en media y varianza. Dicho de un modo más técnico: que su media y varianza sean independientes respecto al tiempo.

Tras aplicar la primera diferencia (véase figura 4), las correlaciones altas de los grupos *Research Personnel* y *HERD & BERD* han mayoritariamente desaparecido, lo que indica que, aunque en su estado original estos grupos parecían estar altamente correlacionados, dicha correlación era en gran medida el resultado de tendencias comunes a lo largo del tiempo más que de relaciones intrínsecas estables. Sin embargo, las correlaciones altas en los grupos de *Patents & Programmes* y *Trade Balance* siguen existiendo, lo que indica que en este grupo las variables sí que están probablemente influenciadas por factores intrínsecos o relaciones directas entre las variables dentro del propio grupo de programas.

Figura 4. Correlaciones entre indicadores por grupo tras aplicar la primera diferencia



Fuente: Elaboración propia

Para entender esto mejor, profundicemos en el grupo de *Trade Balance y Patents & Programmes*. La balanza comercial incluye variables sobre las importaciones y exportaciones de la industria aeroespacial, farmacéutica e informática, electrónica y óptica. La presencia de estas correlaciones altas puede interpretarse como evidencia de que existen tendencias compartidas en el tiempo o de movimientos cíclicos conjuntos entre estos sectores, los cuales podrían estar influenciados por factores económicos globales, políticas comerciales, o incluso innovaciones tecnológicas que afectan a múltiples industrias a la vez.

En cuanto al grupo *Patents & Programmes*, que incluye variables tales como la asignación gubernamental de presupuesto a programas de desarrollo económico, salud y medio ambiente, espacio, investigación no orientada, y programas de educación y sociales, ha mostrado correlaciones elevadas tanto en el análisis inicial como tras aplicar la técnica de la primera diferencia.

Esto indica que las fuertes relaciones entre las cantidades asignadas a diferentes programas y su proporción relativa al GBARD civil son estables a través del tiempo, sugiriendo una coherencia en las políticas de financiamiento gubernamental. El hecho de que siga habiendo correlaciones altas tras aplicar la primera diferencia refleja que las variaciones anuales en la asignación de fondos mantienen una proporción similar año tras año, lo que puede interpretarse como un reflejo de una estrategia política a largo plazo y un compromiso sostenido con estas áreas de inversión.

Otra posible causa es que exista una integración estructural en el modo en que se distribuyen los fondos para los programas. Por ejemplo, un aumento en los fondos para el desarrollo económico podría estar sistemáticamente ligado por incrementos proporcionales en programas de salud y medio ambiente, espacio y otros ámbitos, lo cual podría ser indicativo de una política de financiamiento integrada y holística que busca un equilibrio entre diversos sectores de desarrollo e investigación.

Básicamente, los resultados nos sugieren que los cambios en las asignaciones presupuestarias no son arbitrarios o a modo reacción, sino que obedecen a un plan de financiamiento gubernamental bien establecido.

Centrándonos en el resto de los grupos, al aplicar la técnica de la primera diferencia la mayoría de las correlaciones anteriormente identificadas han disminuido o han desaparecido. Esto sugiere que, como se ha mencionado anteriormente, las

correlaciones iniciales altas posiblemente eran resultado de tendencias temporales comunes más que de relaciones propias entre las variables. La aplicación de la primera diferencia tiene el efecto de controlar estas tendencias, permitiéndonos observar las variaciones año a año y proporcionando así un análisis más ajustado a las fluctuaciones a corto plazo y a los cambios inmediatos en el comportamiento de la balanza comercial. Esta disminución también podría estar indicando que los sectores estudiando en el grupo de la balanza comercial operan con una mayor independencia relativa de la que los valores originales sugerían.

Una vez hemos entendido lo que pasa en cada grupo, con el fin de crear el indicador compuesto, en el grupo *Patents & Programmes*, donde sigue habiendo correlaciones muy altas tras aplicar la primera diferencia, habría que eliminar aquellas variables que siguen presentando una correlación alta tras la primera diferencia, para evitar así la multicolinealidad y la información redundante. Se marcaría el *threshold* en 0,8 y -0,8, es decir, aquellas variables que tienen una correlación alta tanto negativa como positiva. Sin embargo, como las variables que pertenecen a este grupo son distintas conceptualmente, es decir, cada patente y programa pertenece a un sector y representa una inversión diferente, se ha decidido mantener todos los indicadores y no proceder a la eliminación. Posteriormente, a la hora de asignar pesos en el grupo de *Patents & Programmes*, el peso asignado a estas variables muy parecidas se va a repartir entre ellas.

3.6. Extracción de pesos factoriales

El siguiente paso para crear el indicador ha sido el cálculo de los pesos factoriales. Para la consecución de este objetivo, se han determinado los pesos factoriales mediante el FA. FA es una técnica estadística utilizada para identificar patrones subyacentes en un conjunto de variables, reduciendo así la dimensionalidad de los datos mientras se retiene la mayor parte de la información original. Se basa en el principio de que las observaciones realizadas en múltiples variables pueden explicarse mediante un número menor de factores latentes o no observados, que representan las dimensiones subyacentes compartidas por las variables (Fabrigar et al., 1999).

Esta técnica es especialmente útil en áreas de investigación donde se pretende explorar conceptos complejos que no pueden medirse directamente, como la innovación tecnológica, permitiendo a los investigadores inferir la presencia de factores comunes a partir de patrones de correlación entre variables.

Se ha seleccionado el FA para este estudio por varias razones. En primer lugar, el FA es apropiado cuando el objetivo es simplificar y reducir la complejidad de los datos, identificando variables subyacentes comunes que explican la correlación entre múltiples variables observadas (Costello & Osborne, 2019). Esto es particularmente relevante en el análisis de la innovación tecnológica, donde múltiples indicadores interrelacionados pueden reflejar dimensiones fundamentales de la innovación.

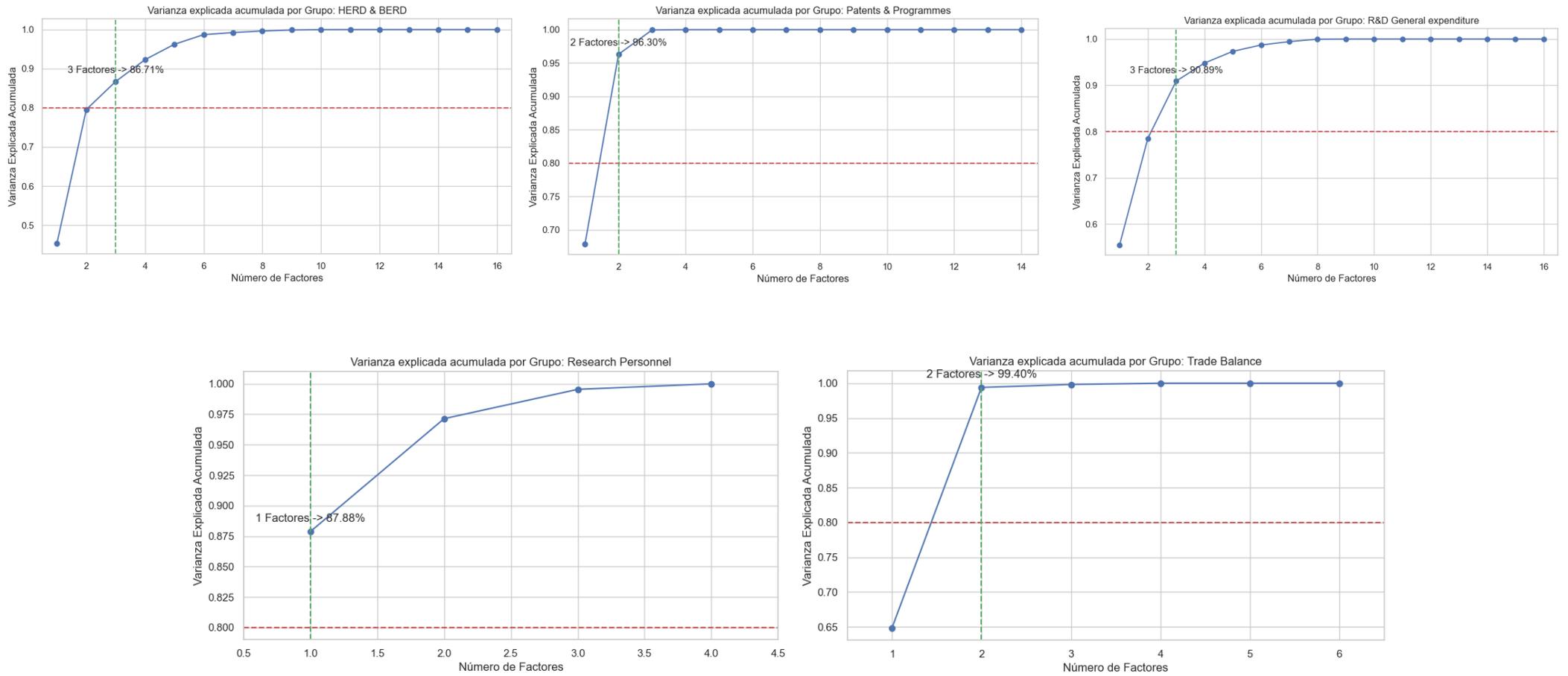
En segundo lugar, el FA permite una mejor interpretación y comprensión de los datos al agrupar variables correlacionadas bajo factores comunes, facilitando la identificación de áreas clave de innovación tecnológica.

Por último, este método es útil para identificar y eliminar redundancias en el set de datos, ya que las variables altamente correlacionadas se combinan en un solo factor, mejorando así la eficiencia del análisis.

Por otro lado, se ha preferido el FA sobre otras técnicas de reducción de dimensiones, como el Análisis de Componentes Principales (PCA), porque nuestro objetivo es explorar la estructura subyacente de los datos y entender cómo las variables originales se relacionan con los factores latentes (Thompson, 2004). Mientras que el PCA se enfoca en maximizar la varianza explicada, el FA busca modelar las correlaciones entre variables, haciendo suposiciones sobre los factores comunes y los errores específicos de cada variable, lo que lo hace más adecuado para estudios exploratorios en ciencias sociales y de comportamiento.

El análisis comienza calculando los autovalores y los vectores propios para cada factor, utilizando métodos de extracción como el de componentes principales o el análisis de ejes principales. Los *loadings* factoriales, que representan las correlaciones entre las variables originales y los factores extraídos, se calculan, normalizan y ordenan. A continuación, con el fin de determinar el número de factores necesarios para cada grupo de indicadores, se ha aplicado el criterio de que el número óptimo de factores debe explicar al menos el 80% de la varianza total (véase figura 5). Posteriormente, se aplica una rotación (con promax) para maximizar la varianza de los *loadings* factoriales, reduciendo la multipertenencia de una variable a varios factores, mejorando así la interpretación de la información contenida en cada factor.

Figura 5. Factores que explican el 80% de la varianza por cada grupo



Fuente: Elaboración propia

Las variables se asocian preferentemente con el factor al que más contribuyen, lo que permite identificar los pesos factoriales. Estos pesos se utilizan para interpretar los factores en términos de las variables originales, ofreciendo *insights* significativos sobre las dimensiones subyacentes que estructuran el *dataset*.

En este análisis, tras identificar los factores significativos mediante el criterio del 80% de la varianza explicada y realizar la rotación adecuada a través de promax, se recalculan los *loadings* para estos factores significativos (véase figura 6)

Figura 6. Factores más representativos por cada indicador por grupo

Grupo: HERD & BERD				Grupo: Patents & Programmes			Grupo: R&D General expenditure			
Variable	Factor 1	Factor 2	Factor 3	Variable	Factor 1	Factor 2	Variable	Factor 1	Factor 2	Factor 3
B_AERO	0.771219	-0.121651	0.065202	C_ECOPPP	0.957668	0.036087	C_CVXTT	-0.051762	-0.015231	1.009487
B_COMP	0.970525	-0.017594	0.021054	C_ECOCVC	0.981053	-0.023804	C_DFXTT	-0.051764	-0.015232	1.009474
B_DRUG	0.974073	-0.031027	0.037095	C_EDUPPP	0.958056	0.033328	C_GUFPPP	0.004361	-0.002453	0.927287
B_FBCT	-0.003170	1.026424	-0.047589	C_EDUCVC	0.981051	-0.023806	C_GUFXCV	-0.051760	-0.015237	1.009488
B_PPPCT	-0.062344	0.362843	0.490326	C_HEAPPP	0.943350	0.034759	C_PPPCT	0.119706	0.023720	0.706245
B_SERV	0.859257	0.188922	-0.136446	C_HEAXCV	0.981052	-0.023796	GV_PPPCT	0.702189	0.246254	0.048310
B_XAERO	0.769753	-0.121147	0.065331	C_NORPPP	0.958024	0.032963	GV_XFB	0.097907	0.595091	0.057750
B_XCOMP	0.966767	-0.014086	0.019781	C_NORXCV	0.981050	-0.023816	G_PPPCT	1.010707	-0.010623	-0.019718
B_XDRUG	0.970608	-0.028276	0.036217	C_SPAPPP	0.981111	-0.022953	G_XEB	1.006815	-0.000902	-0.019063
B_XFA	-0.001939	1.025159	-0.045309	C_SPAXCV	0.981052	-0.023804	G_XEG	1.006820	-0.000902	-0.019078
B_XFB	-0.002953	1.026083	-0.046554	P_BIOPCT	-0.000082	0.999259	G_XEH	1.006827	-0.000908	-0.019083
B_XFG	-0.001945	1.025151	-0.045295	P_ICTPCT	-0.000082	0.999259	G_XEI	0.891551	-0.089832	0.005704
B_XFON	-0.008568	0.901718	0.071927	P_PCT	-0.000082	0.999259	G_XFA	-0.023601	1.016451	-0.022030
B_XSERV	0.849247	0.200758	-0.141828	P_TRIAD	-0.000082	0.999259	G_XFB	-0.027002	1.018885	-0.022341
H_PPPCT	-0.014084	-0.098031	1.072468				G_XFG	-0.023566	1.016432	-0.022037
H_XFB	0.076785	0.247035	0.506936				G_XFON	-0.026974	1.018873	-0.022348

Grupo: Research Personnel		Grupo: Trade Balance		
Variable	Factor 1	Variable	Factor 1	Factor 2
BP_RSXEI	-0.928606	TD_BAERO	0.468311	0.861852
BP_TTXEI	-0.903906	TD_BCOMP	0.469157	0.847841
TP_RSXEM	-0.926785	TD_BDRUG	0.465987	0.861037
TP_TTXEM	-0.903205	TD_XAERO	0.983558	-0.214754
		TD_XCOMP	0.983558	-0.214753
		TD_XDRUG	0.983558	-0.214753

Fuente: Elaboración propia

Este enfoque metódico asegura que los factores seleccionados para el modelo final son aquellos que proporcionan la representación más fiel y simplificada del set de datos original, permitiendo así una interpretación precisa de las principales dimensiones de variabilidad en el contexto de la innovación tecnológica entre países.

3.7. Agregación

Una vez identificados los *loadings* para cada factor más representativos para cada indicador mediante el FA, el siguiente paso crucial en la construcción del indicador es identificar, por cada dimensión, los pesos de las variables dentro de cada factor, los pesos factoriales y los pesos dimensionales.

Para ello, inicialmente asignamos igual importancia a cada dimensión, donde el peso de cada una es el recíproco del número total de dimensiones consideradas.

$$\text{Peso dimensional} = P_d = \frac{1}{N}$$

Donde:

- N es el número total de dimensiones consideradas

A nivel factorial, los pesos se derivan de la proporción de la varianza explicada asociada al autovalor del factor, calculada como la varianza explicada acumulada ajustada por la varianza explicada acumulada de los factores anteriores.

$$\text{Peso factorial}_j = PF_j = \frac{\text{Varianza explicada}_j}{\sum_j \text{Varianza explicada}_j}$$

Donde:

- $\text{Varianza explicada}_j$ es la varianza explicada por el factor j

Por último, dentro de cada factor, asignamos pesos a las variables basándonos en el cuadrado de las cargas factoriales, normalizado por la suma de los cuadrados de todas las cargas dentro de ese factor.

$$\text{Peso variable}_{i,j} = PV_{i,j} = \frac{(\text{Carga factorial}_{i,j})^2}{\sum_i (\text{Carga factorial}_{i,j})^2}$$

Donde:

- $\text{Carga factorial}_{i,j}$ es la carga del factor j para la variable i

Finalmente, el indicador compuesto se calcula, a nivel de factor mediante una agregación aritmética y a nivel de dimensión una agregación geométrica: Cada contribución de cada variable es ponderada por su correspondiente peso de variable y luego multiplicada por el peso factorial del factor asociado. La contribución total de cada grupo se eleva a la potencia del peso dimensional asignado a esa dimensión en particular y el resultado de cada dimensión se multiplica.

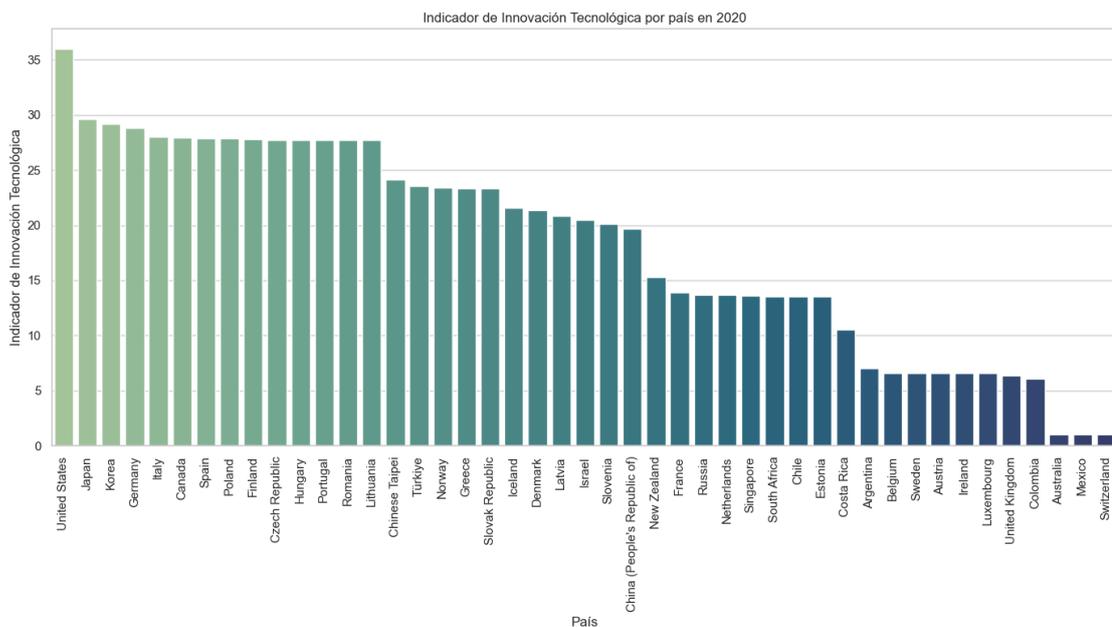
$$\text{Indicador Compuesto} = \prod_d \left(\sum_j \left(\sum_i (V_{i,t_i} \cdot PV_{i,j}) \cdot PF_j \right) \right)^{P_d}$$

Donde:

- V_i es el valor de la variable i en el año j
- $PV_{i,j}$ es el peso de la variable i para el factor j
- PF_j es el peso factorial del factor j
- P_d es el peso de la dimensión

Finalmente, se desarrolla un ranking anual para cada país, ordenado de acuerdo con el valor del indicador normalizado (véase figura 7), permitiendo una comparativa clara de la capacidad de innovación tecnológica entre diferentes naciones.

Figura 7. Ranking de innovación tecnológica por países en 2020



Fuente: Elaboración propia

3.8. Robustez y sensibilidad

Una vez ya tenemos el indicador compuesto creado, toca estudiar la sensibilidad y robustez de dicho indicador. Este proceso consiste en estudiar la solidez del indicador compuesto en términos de, por ejemplo, el mecanismo para incluir o excluir indicadores individuales, el esquema de normalización, la imputación de los datos que faltan, la elección de las ponderaciones y el método de agregación (OECD, 2008).

En este caso, como previamente se han aplicado diferentes métricas para normalizar y para imputar datos faltantes, se ha decidido emplear un nuevo método de clasificación de las variables con el fin de buscar nuevos *insights* sobre la influencia de diferentes tipos de datos en el indicador compuesto.

En la construcción del indicador se realizó inicialmente una agrupación de indicadores basada en sus características temáticas. Por ejemplo, todos los datos relacionados con patentes y programas fueron agrupados bajo el título "*Patents & Programmes*", mientras que los relacionados con el personal de investigación se agruparon en "*Research Personnel*".

Para explorar la robustez y la sensibilidad del indicador compuesto resultante, se ha implementado una nueva metodología para agrupar las variables. Los datos se han reorganizado en cuatro nuevas categorías: *Input Indicators*, *Output Indicators*, *Impact Indicators*, y *Background Economic Indicators*. Este enfoque se alinea con la literatura que sugiere que una reevaluación de las agrupaciones de indicadores puede proporcionar *insights* significativos sobre la influencia de distintos tipos de datos en el indicador compuesto final (Saisana et al., 2005).

La reasignación de los indicadores se ha realizado de la siguiente manera:

- ***Background Economic Indicators***: Incluye indicadores que proporcionan un contexto económico general, no directamente vinculado con actividades de innovación pero que establecen un marco para el mismo (este grupo se ha mantenido de la agrupación original)
- ***Input Indicators***: Consisten en recursos invertidos en innovación, como gastos en I+D y personal de investigación.
- ***Output Indicators***: Reflejan los resultados directos de las actividades de innovación, incluyendo patentes y publicaciones.

- **Impact Indicators:** Evalúan el efecto de la innovación en la sociedad y la economía, como los cambios en el mercado laboral tecnológico y las contribuciones económicas.

Tras realizar la limpieza de variables, de la misma manera que en la agrupación 1, cada grupo contiene las siguientes variables:

Tabla 10. Indicadores del grupo *Input indicators*

Tabla para Input indicators:

	Indicators
0	BERD performed in the aerospace industry (current PPP \$)
1	BERD performed in the computer, electronic and optical industry (current PPP \$)
2	BERD performed in the pharmaceutical industry (current PPP \$)
3	Business-financed BERD at constant prices and PPP \$
4	BERD at constant prices and PPP \$
5	BERD performed in service industries (current PPP \$)
6	Percentage of BERD performed in the aerospace industry
7	Percentage of BERD performed in the computer, electronic and optical industry
8	Percentage of BERD performed in the pharmaceutical industry
9	Percentage of BERD financed by the rest of the world
10	Percentage of BERD financed by the business sector
11	Percentage of BERD financed by government
12	Percentage of BERD financed by the Higher Education and PNP sectors
13	Percentage of BERD performed in service industries
14	Civil Budget R&D as a percentage of Total GBARD
15	Defence Budget R&D as a percentage of Total GBARD
16	Civil GBARD for General University Funds (GUF) (current PPP\$)
17	General University Funds (GUF) as a percentage of Civil GBARD
18	Total GBARD at constant prices and PPP \$
19	GOVERD at constant prices and PPP \$
20	Percentage of GOVERD financed by the business sector
21	GERD at constant prices and PPP \$
22	Percentage of GERD performed by the Business Enterprise sector
23	Percentage of GERD performed by the Government sector
24	Percentage of GERD performed by the Higher Education sector
25	Percentage of GERD performed by the Private Non-Profit sector
26	Percentage of GERD financed by the rest of the world
27	Percentage of GERD financed by the business enterprise sector
28	Percentage of GERD financed by government
29	Percentage of GERD financed by the Higher Education and PNP sectors
30	HERD at constant prices and PPP \$
31	Percentage of HERD financed by the business sector

Fuente: Elaboración propia

Tabla 11. Indicadores del grupo *Impact indicators*

Tabla para Impact indicators:

	Indicators
0	Business enterprise sector: Women researchers as a percentage of total researchers (headcount)
1	Business Enterprise researchers per thousand employment in industry
2	Total Business Enterprise personnel per thousand employment in industry
3	Government sector: Women researchers as a percentage of total researchers (headcount)
4	Higher education sector: Women researchers as a percentage of total researchers (headcount)
5	Trade Balance: Aerospace industry (current prices)
6	Trade Balance: Computer, electronic and optical industry (current prices)
7	Trade Balance: Pharmaceutical industry (current prices)
8	Export market share: Aerospace industry
9	Export market share: Computer, electronic and optical industry
10	Export market share: Pharmaceutical industry
11	Total researchers (headcount)
12	Woman researchers (headcount)
13	Women researchers as a percentage of total researchers (headcount)
14	Total researchers per thousand total employment
15	Total R&D personnel per thousand total employment

Fuente: Elaboración propia

Tabla 12. Indicadores del grupo *Output indicators*

Tabla para Output indicators:

	Indicators
0	Civil GBARD for Economic Development programmes (current PPP\$)
1	Economic Development programmes as a percentage of Civil GBARD
2	Civil GBARD for Education and Social programmes (current PPP\$)
3	Education and Social programmes as a percentage of Civil GBARD
4	Civil GBARD for Health and Environment programmes (current PPP\$)
5	Health and Environment programmes as a percentage of Civil GBARD
6	Civil GBARD for Non-oriented Research programmes (current PPP\$)
7	Non-oriented Research programmes as a percentage of Civil GBARD
8	Civil GBARD for Space programmes (current PPP\$)
9	Space programmes as a percentage of Civil GBARD
10	Number of patents in the biotechnology sector - applications filed under the PCT (priority year)
11	Number of patents in the ICT sector - applications filed under the PCT (priority year)
12	Number of patent applications filed under the PCT(priority year)
13	Number of "triadic" patent families (priority year)

Fuente: Elaboración propia

Tabla 13. Indicadores del grupo *Background Economic Indicators*

Tabla para Background Economic Indicators:

	Indicators
0	Labour Force
1	Gross Domestic Product (current PPP\$)
2	Industrial Employment
3	Implicit GDP Price Indices
4	Purchasing Power Parity
5	Total Employment
6	Population
7	Value Added of Industry (current PPP\$)

Fuente: Elaboración propia

Con esta nueva agrupación, se realiza la misma normalización que en la agrupación uno y para la imputación de datos faltantes la técnica seleccionada también ha sido imputación con la media, siendo los errores de la tabla 14 los ECM de imputación.

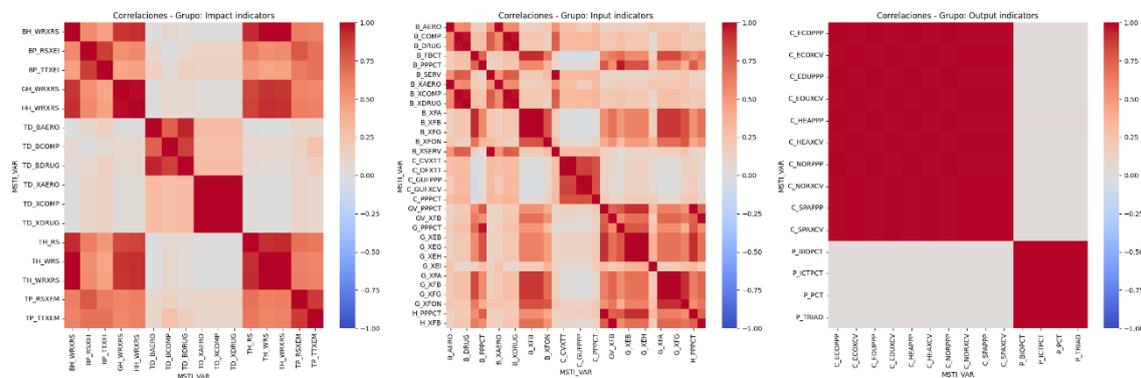
Tabla 14. Errores de imputación agrupación dos

Técnica de Imputación	Error (MSE)
Imputación Media	0.058046334
Imputación Mediana	0.070313702
Imputación Aleatoria	0.11975464
Imputación kNN	0.058047388

Fuente: Elaboración propia

En cuanto a las correlaciones (véase figura 8) observamos que también hay correlaciones altas en los distintos grupos, destacando el de *Output Indicators* (el grupo de *Background Economic Indicators* ha sido eliminado tras la normalización de las variables, ya que deja de ser necesario para el resto del análisis), por lo que aplicamos de la misma manera la primera diferencia para comprobar si estas correlaciones son reales o si se tratan de correlaciones espurias, es decir, que dos variables parezcan estar correlacionadas (positiva o negativamente) pero no hay una relación causal o lógica que explique esta asociación.

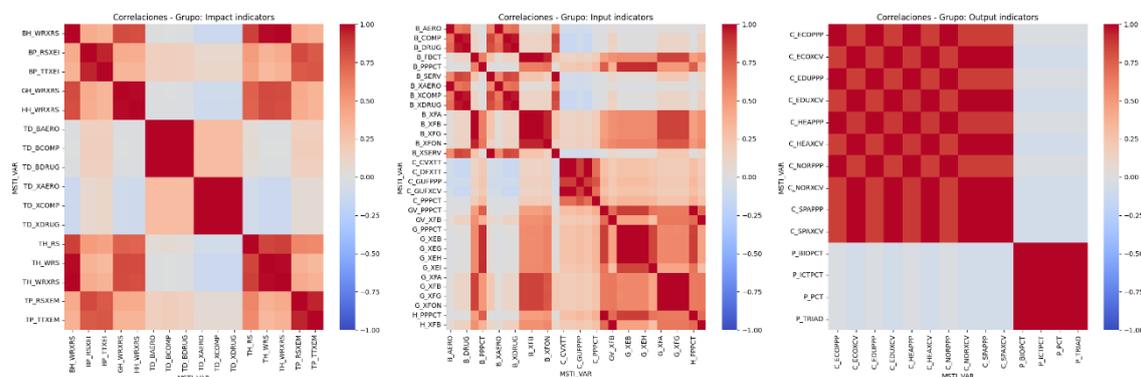
Figura 8. Correlaciones por grupos



Fuente: Elaboración propia

Tras este análisis sigue algunas variables que antes parecían estar correlacionadas ya no lo están, y hay otras que mantienen las correlaciones altas (véase figura 9). Aun así, para el análisis se decide no eliminar ninguna variable, ya que, aunque presenten correlaciones altas, con el propósito de la creación de este indicador se consideran que son igual de relevantes al proporcionar información distinta.

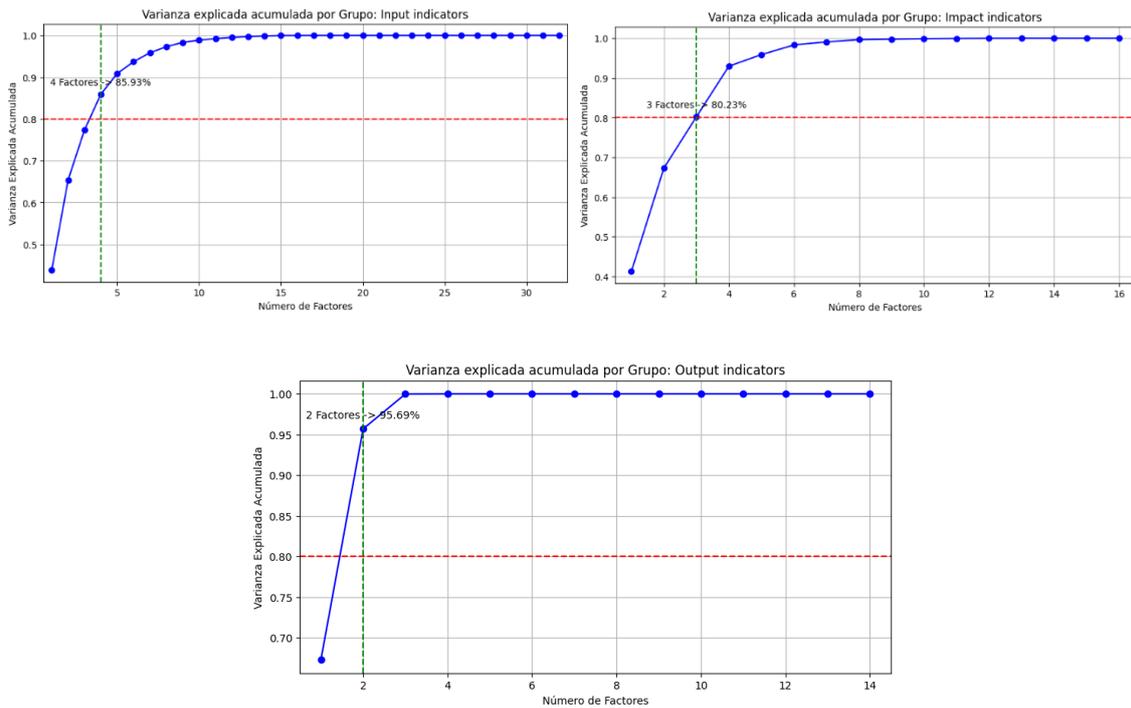
Figura 9. Correlaciones tras la primera diferencia por grupos



Fuente: Elaboración propia

Posteriormente empleamos las correlaciones tras la primera diferencia para realizar un FA. En este análisis se indica que los factores seleccionados para cada grupo (véase figura 10) serán los suficientes para explicar un 80% de la varianza.

Figura 10. Factores que explican el 80% de la varianza por cada grupo en agrupación dos



Fuente: Elaboración propia

Una vez se tienen esos factores, ya se realiza el FA con los factores necesarios y realizando una rotación promax para maximizar la varianza de los *loadings* factoriales, reduciendo la multipertenencia de una variable a varios factores.

De esta manera, obtendremos el peso repartido de cada factor por cada dimensión en cada variable que compone dicha dimensión (véase tablas 15, 16 y 17). Cada variable pertenecerá al factor en el que tenga más importancia, es decir, en el que el valor del factor en dicha variable en valor absoluto sea más grande.

Tabla 15. Factores más representativos por cada indicador en el grupo *Input indicators*

Grupo: Input indicators					Grupo: Input indicators				
Variable	Factor 1	Factor 2	Factor 3	Factor 4	Variable	Factor 1	Factor 2	Factor 3	Factor 4
B_AERO	-0.122531	0.041688	0.776698	0.019119	G_XEB	-0.010822	1.008083	-0.002595	-0.024136
B_COMP	-0.010838	0.008631	0.968675	-0.019804	G_XEG	-0.010822	1.008089	-0.002602	-0.024152
B_DRUG	-0.015270	0.013032	0.971263	-0.020218	G_XEH	-0.010822	1.008094	-0.002599	-0.024159
B_FBCT	0.982927	-0.018209	-0.012004	-0.067824	G_XEI	0.017112	0.842364	-0.000426	-0.026531
B_PPPCT	0.041378	0.917775	-0.015133	-0.072627	G_XFA	0.962305	-0.022065	-0.018337	0.029613
B_SERV	0.155274	-0.088025	0.857803	0.004699	G_XFB	0.965578	-0.026034	-0.019836	0.029242
B_XAERO	-0.121853	0.041783	0.775060	0.018394	G_XFG	0.962280	-0.022025	-0.018343	0.029605
B_XCOMP	-0.005917	0.007092	0.964130	-0.023720	G_XFON	0.965558	-0.026000	-0.019840	0.029233
B_XDRUG	-0.011396	0.011813	0.967161	-0.023312	H_PPPCT	0.090677	0.814684	-0.009292	0.071400
B_XFA	0.983512	-0.016798	-0.011009	-0.068178	H_XFB	0.547324	0.134529	0.055894	0.107435
B_XFB	0.986837	-0.020565	-0.012464	-0.068269					
B_XFG	0.983499	-0.016774	-0.011014	-0.068183					
B_XFON	0.847621	0.098222	-0.012059	-0.074469					
B_XSERV	0.169470	-0.093024	0.846286	-0.002124					
C_CVXTT	-0.006883	-0.066464	-0.029219	1.005609					
C_DFXTT	-0.006885	-0.066466	-0.029211	1.005658					
C_GUFPPP	-0.013411	0.000092	-0.004343	0.929748					
C_GUFXCV	-0.006891	-0.066461	-0.029207	1.005672					
C_PPPCT	-0.015101	0.134060	0.054304	0.718496					
GV_PPPCT	0.089210	0.816377	-0.009585	0.070589					

Fuente: Elaboración propia

Tabla 16. Factores más representativos por cada indicador en el grupo *Impact indicators*

Grupo: Impact indicators			
Variable	Factor 1	Factor 2	Factor 3
BH_WRXRS	0.925066	-0.161224	-0.069279
BP_RSXEI	0.627469	0.195524	0.078955
BP_TTXEI	0.584151	0.194910	0.103368
GH_WRXRS	0.868310	-0.158116	-0.039917
HH_WRXRS	0.859119	-0.156178	-0.043391
TD_BAERO	-0.059726	0.978476	-0.057716
TD_BCOMP	-0.044439	0.977890	-0.051804
TD_BDRUG	-0.071449	0.985425	-0.061846
TD_XAERO	-0.026561	-0.064640	1.017894
TD_XCOMP	-0.026562	-0.064639	1.017893
TD_XDRUG	-0.026565	-0.064639	1.017893
TH_RS	0.903830	-0.061511	-0.042642
TH_WRS	0.902229	-0.168623	-0.073234
TH_WRXRS	0.925066	-0.161224	-0.069279
TP_RSXEM	0.647520	0.207619	0.049457
TP_TTXEM	0.606410	0.209389	0.054426

Fuente: Elaboración propia

Tabla 17. Factores más representativos por cada indicador en el grupo *Output indicators*

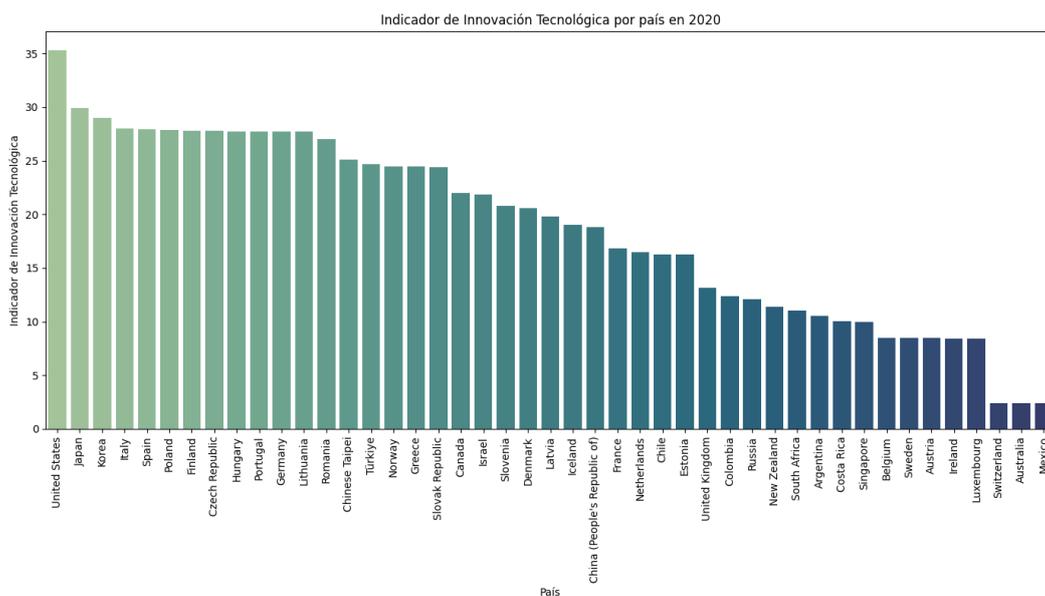
Grupo: Output indicators

Variable	Factor 1	Factor 2
C_ECOPPP	0.957993	0.036188
C_ECOXCV	0.981145	-0.023861
C_EDUPPP	0.958417	0.033325
C_EDUXCV	0.981142	-0.023863
C_HEAPPP	0.942889	0.034801
C_HEAXCV	0.981143	-0.023854
C_NORPPP	0.958384	0.032947
C_NORXCV	0.981141	-0.023874
C_SPAPPP	0.981204	-0.022978
C_SPAXCV	0.981143	-0.023862
P_BIOPCT	0.000272	0.999281
P_ICTPCT	0.000272	0.999281
P_PCT	0.000272	0.999281
P_TRIAD	0.000272	0.999281

Fuente: Elaboración propia

Por último, se calculan de la misma manera que con la agrupación uno los pesos de las variables, pesos factoriales y pesos dimensionales. Una vez obtenidos, se emplea la ecuación para crear el indicador compuesto y realizamos un ranking anual por países como podemos ver en la figura 11:

Figura 11. Ranking de innovación tecnológica por países en 2020 con agrupación 2



Fuente: Elaboración propia

4. CONCLUSIONES

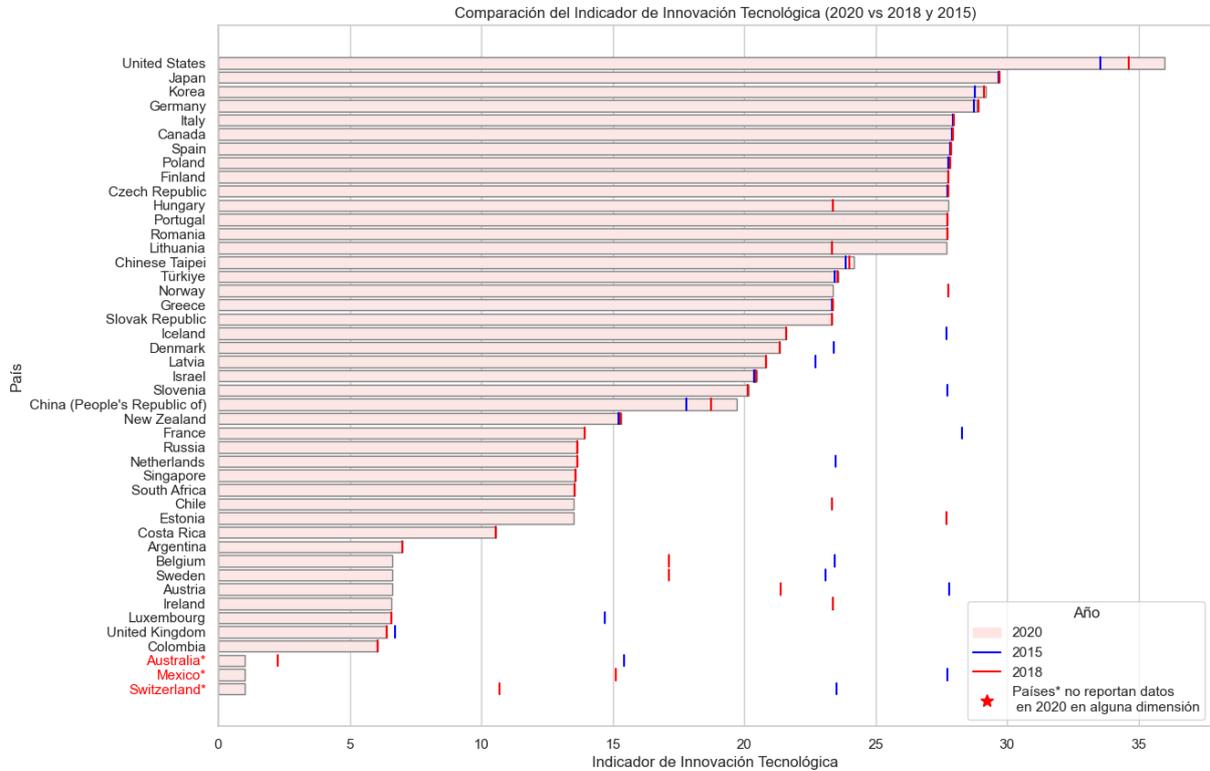
La importancia de desarrollar un indicador compuesto que mida la innovación tecnológica reside en la capacidad de integrar múltiples variables que reflejen los distintos aspectos de la innovación en un único indicador simplificado. Este indicador permite una comparación más efectiva y directa entre países, facilitando el entendimiento de las capacidades de innovación tecnológica y su impacto en la economía global.

Los indicadores compuestos son esenciales para abordar la complejidad de la innovación tecnológica, que incluye desde la inversión en I+D hasta los resultados como patentes y otros outputs de innovación. Por lo tanto, crear un indicador que consolide esta información en una medida coherente es crucial para analizar tendencias, hacer comparaciones internacionales y apoyar la toma de decisiones en política pública y estrategia empresarial.

El este trabajo se ha desarrollado un indicador compuesto robusto, utilizando una metodología que integra técnicas como la normalización de datos, imputación de datos faltantes, y análisis factorial para la asignación de pesos y agregación de indicadores. Este enfoque no solo garantiza que el indicador sea representativo y equitativo, sino que también mejora la precisión en la comparación de la innovación tecnológica entre diferentes países. Gracias a los dos tipos de clasificación de las variables, obteniendo resultados muy similares, el indicador ha demostrado ser robusto y sensible a las variaciones en los datos, garantizando que el ser representativo y equitativo para la comparación de la innovación tecnológica entre diferentes países.

Con los resultados se ha creado un ranking anual que refleja las capacidades de innovación tecnológica de los países. Con la información proporcionada por este ranking se ha elaborado un gráfico de barras (véase figura 13) comparando el valor del indicador en los años 2015, 2018 y 2020 entre los distintos países estudiados. Por tanto, el gráfico ofrece una representación visual de la posición relativa de cada país en términos de su desempeño en innovación tecnológica.

Figura 12. Comparación del Indicador de Innovación Tecnológica (2020 vs. 2018 y 2015)



Fuente: Elaboración propia

Observamos una consistencia en la innovación, ya que algunas naciones muestran una posición sólida y consistente en el indicador a través de los años evaluados, especialmente destacando que EE. UU, Japón y Corea son siempre el top 3. Esta constancia puede sugerir una estrategia efectiva y sostenida de inversión en investigación y desarrollo (I+D), así como un ecosistema robusto que favorece la innovación continua.

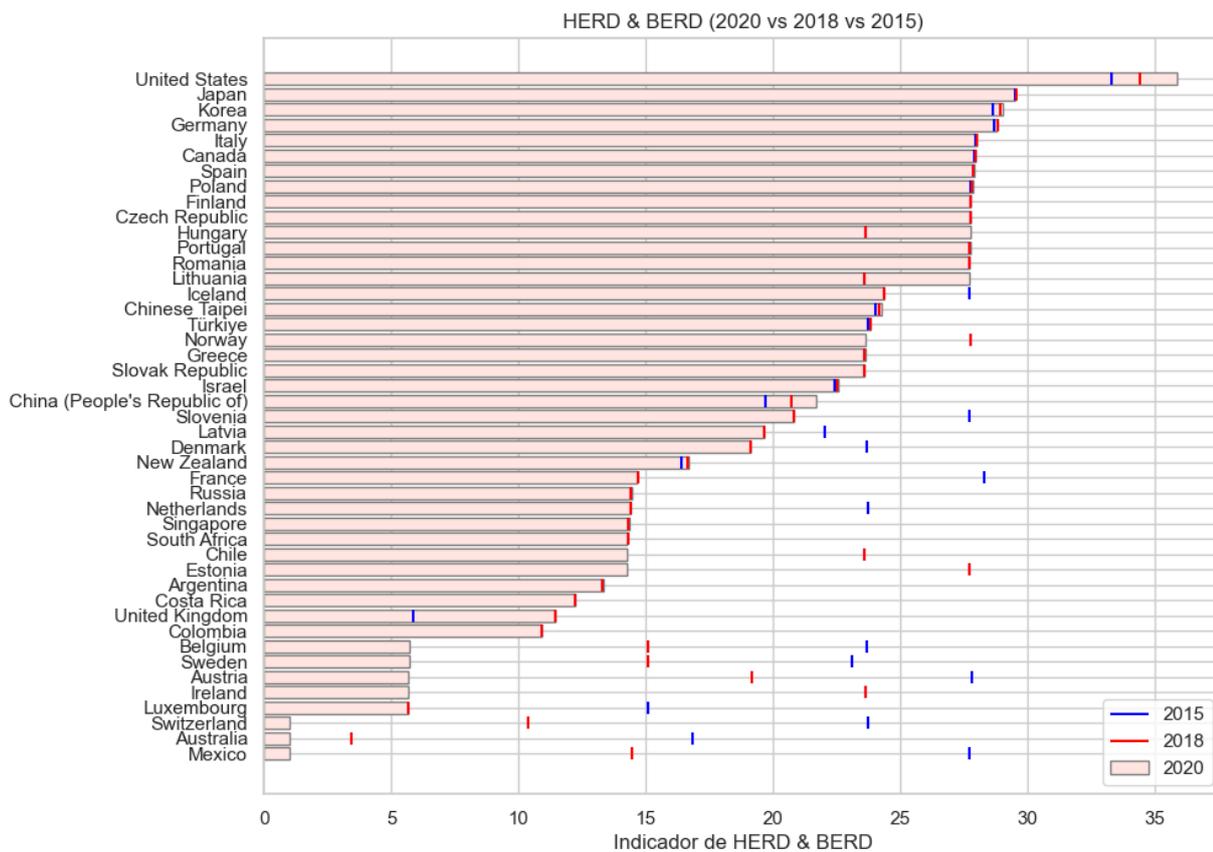
Además, cabe mencionar que para el año 2020, países como Australia, México y Suiza reportaron un volumen de datos menor, lo que se refleja en un valor bastante bajo del indicador. Por tanto, que tenga un valor tan bajo en el indicador ese año no necesariamente refleja una disminución real en las actividades de innovación, sino que radica en la falta de disponibilidad de datos.

Por otro lado, la constancia observada en el top 3 del ranking global compuesto, destacando a EE. UU., Japón y Corea, bastante seguidos de Alemania, indica una robustez en sus ecosistemas de innovación.

A continuación, se ha decidido llevar a cabo un análisis del indicador de cada país por cada dimensión para los años 2015, 2018 y 2020, con el objetivo de estudiar si en todos los grupos siguen estando los mismos países en cabeza del ranking, o si hay países que destacan en una dimensión u otra. Por tanto, este análisis por dimensiones (véase figuras de la 13 a la 17) individuales del indicador nos ofrece una perspectiva más detallada. Permite identificar fortalezas y debilidades específicas dentro de los distintos grupos estudiados que componen la innovación tecnológica.

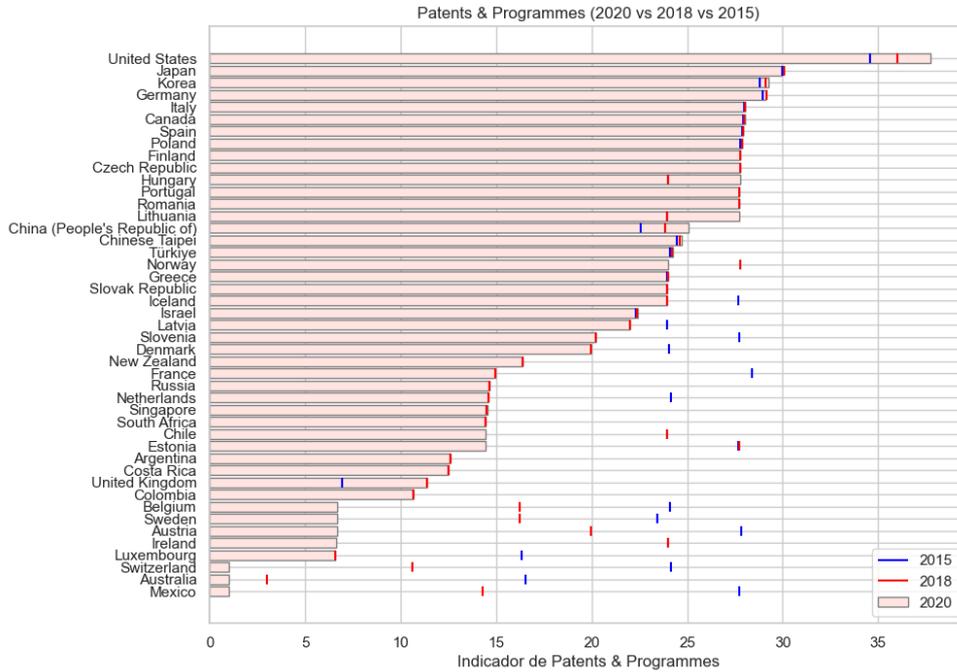
Por ejemplo, si el top 3 global no coincide con el top 3 en ciertas dimensiones como HERD & BERD, esto puede señalar que hay países que, si bien no lideran el indicador compuesto, operan muy bien en áreas específicas. Estas áreas podrían ser la clave de su potencial competitivo y de su capacidad de influir en la economía global de formas únicas.

Figura 13. Comparación del Indicador de Innovación Tecnológica para la dimensión HERD & BERD (2020 vs 2015 y 2018)



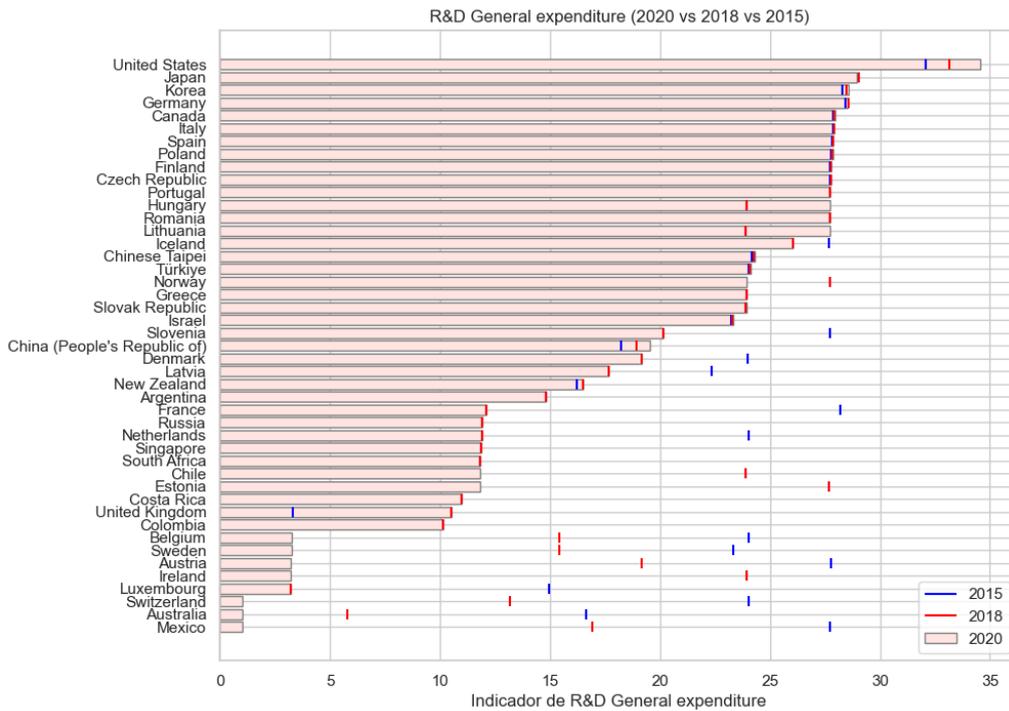
Fuente: Elaboración propia

Figura 14. Comparación del Indicador de Innovación Tecnológica para la dimensión *Patents & Programmes* (2020 vs 2015 y 2018)



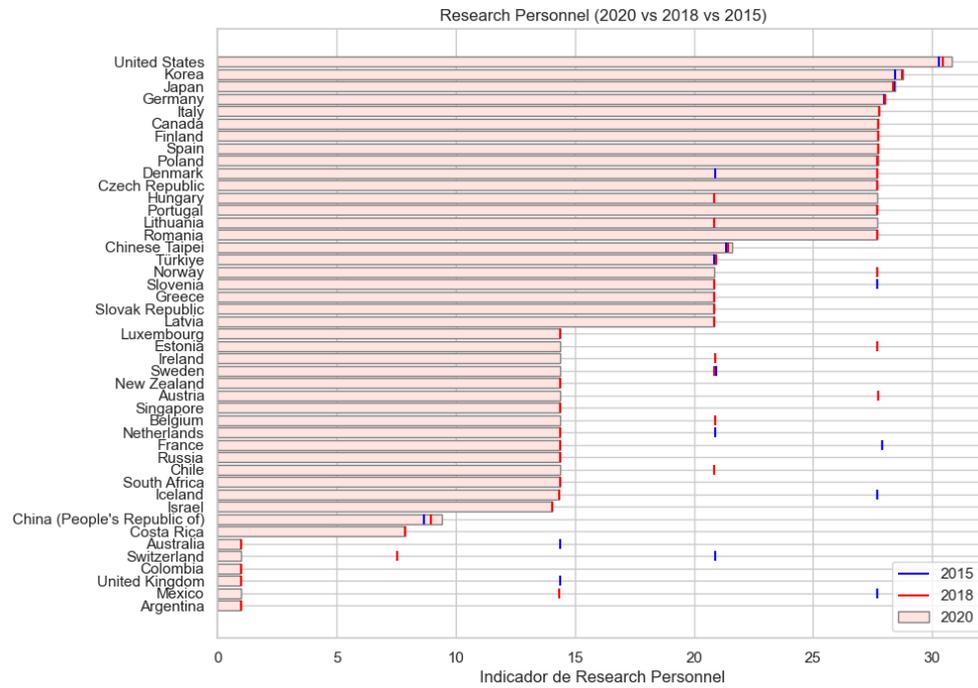
Fuente: Elaboración propia

Figura 15. Comparación del Indicador de Innovación Tecnológica para la dimensión *R&D General expenditure* (2020 vs 2015 y 2018)



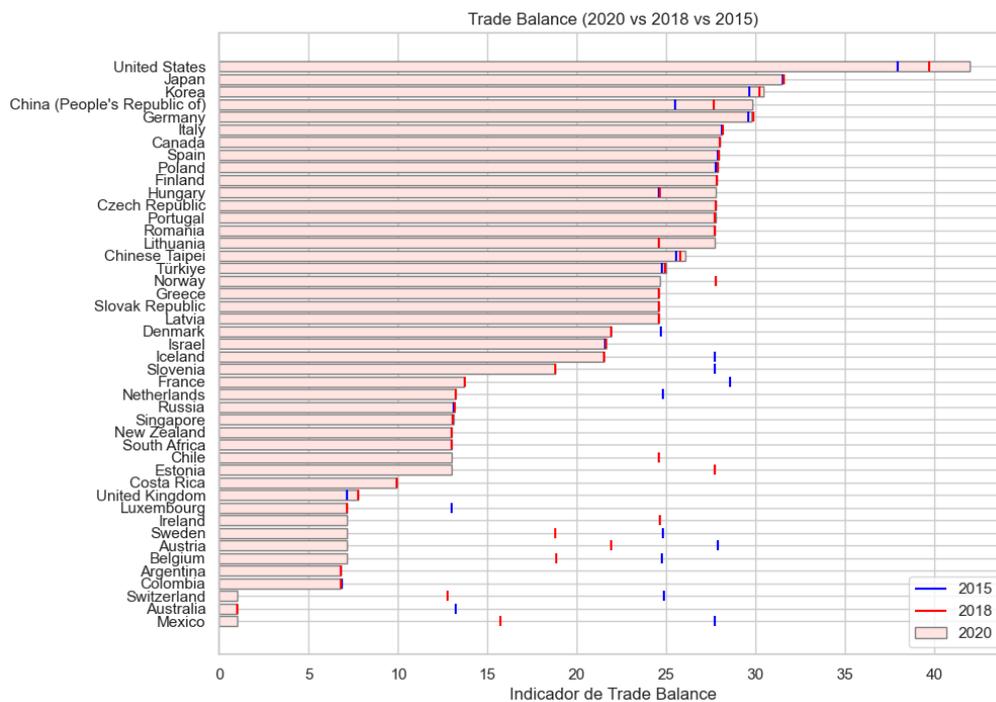
Fuente: Elaboración propia

Figura 16. Comparación del Indicador de Innovación Tecnológica para la dimensión *Research Personnel* (2020 vs 2015 y 2018)



Fuente: Elaboración propia

Figura 17. Comparación del Indicador de Innovación Tecnológica para la dimensión *Trade Balance* (2020 vs 2015 y 2018)



Fuente: Elaboración propia

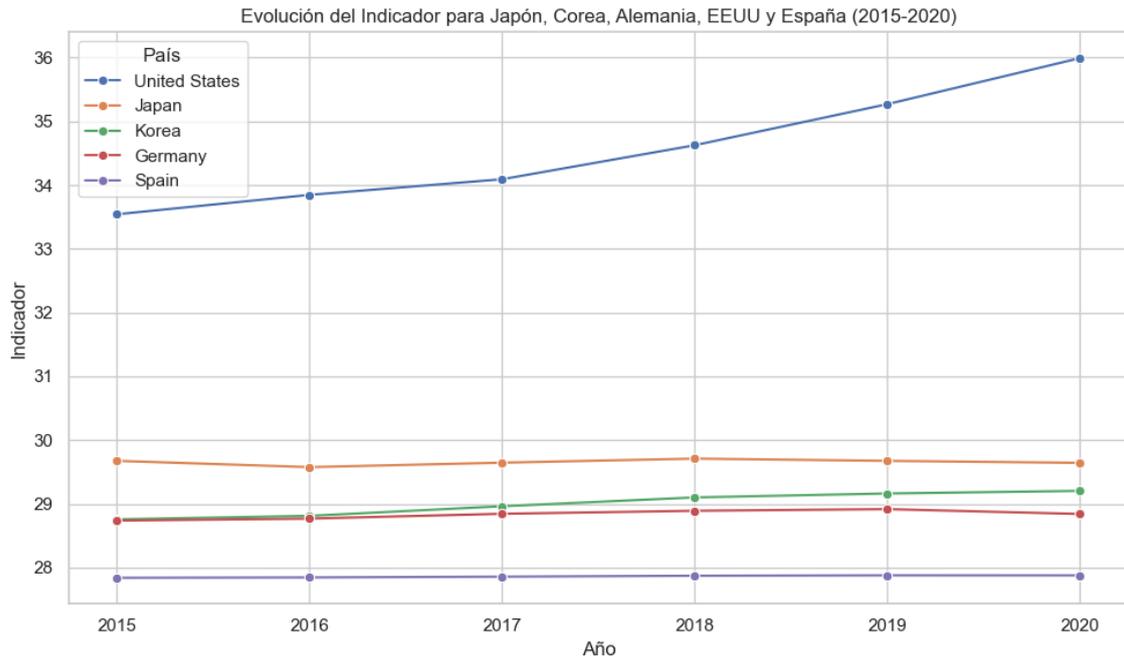
La variación en los rankings por dimensión revela la especialización y las estrategias únicas de cada país en su enfoque hacia la innovación. Mientras la gran mayoría mantienen una consistencia en todas las dimensiones, otros presentan picos de excelencia en áreas concretas, reflejando estrategias nacionales que podrían estar orientadas a fortalecer ciertas industrias o sectores de I+D. Específicamente destacamos que EE. UU se mantiene como top 1 en todas las dimensiones, mientras que Corea, Japón y Alemania van intercálándose, dependiendo de la dimensión. Además, observamos que China destaca especialmente en la dimensión de *Trade Balance*, indicando que su mayor potencial de innovación tecnológica reside en las importaciones y exportaciones del país.

Otro *insight* obtenido de este análisis es que Finlandia, además de estar siempre en el top 10 países más innovadores tecnológicamente hablando, destaca en la dimensión de *Research Personnel*. Este dato cobra importancia al saber que este país ha sido en los últimos años uno de los editores de investigación científica mayores del mundo, por delante de países con una intensidad en ciencia tradicionalmente sólida como Reino Unido y Alemania. Según resultados de investigación de la UE “en los últimos 20 años el número de publicaciones de investigadores finlandeses en las revistas científicas de reputación internacional había aumentado dos veces y medio hasta alcanzar las 8.300 publicaciones en 2005” (Comisión Europea, 2023).

La información de los rankings dimensionales se vuelve aún más valiosa al considerar la variabilidad en la recolección de datos. Por ejemplo, las posiciones cambiantes de ciertos países podrían ser un indicativo de cambios en su entorno de innovación o, como en el caso de Suiza, Australia y México en 2020, podrían deberse a irregularidades en la disponibilidad de datos. Estas fluctuaciones nos recuerdan la necesidad de realizar interpretaciones cautelosas y consideradas de los indicadores, teniendo en cuenta los contextos económicos y políticos subyacentes que pueden influir en los datos que han sido reportados por cada país.

Como último análisis, se ha querido realizar una comparación entre los países que han sido los top 3 todos los años estudiados: Estados Unidos, Japón, Corea del sur y Alemania, con España (véase figura 18).

Figura 18. Evolución del Indicador de Innovación Tecnológica de España, Estados Unidos, Japón, Corea del Sur y Alemania



Fuente: Elaboración propia

Se observa que EE. UU. tiene una tendencia ascendente sostenida, lo que se puede interpretar como un compromiso de este país con la innovación y el desarrollo tecnológico. Por otro lado, Japón, Corea, Alemania y España muestran una relativa estabilidad en sus valores de indicador, con un leve crecimiento por parte de Corea del sur en el año 2016, adelantando a la potencia alemana. Estos resultados sugieren una tasa de crecimiento más moderada en la capacidad de innovación tecnológica de estos países durante los años estudiados.

Como conclusión de este análisis, la elaboración de este indicador ha permitido realizar comparaciones significativas entre países, evaluando su progreso y compromiso con la innovación tecnológica a lo largo del tiempo. La utilidad de este indicador es notable, ya que proporciona una métrica cuantitativa que puede ser empleada para informar políticas públicas, dirigir inversiones en I+D y fomentar estrategias competitivas en el ámbito global. Por tanto, la elaboración de este indicador compuesto es un claro ejemplo de cómo la investigación aplicada puede generar herramientas analíticas que ayuden a comprender mejor el panorama tecnológico mundial, contribuyendo así al debate académico y a la toma de decisiones basada en evidencia.

5. DECLARACIÓN DE USO DE HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA EN TRABAJOS FIN DE GRADO

Por la presente, yo, María del Carmen García Cucala, estudiante de ADE y Business Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado “Indicador compuesto de innovación tecnológica por países”, declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación [el alumno debe mantener solo aquellas en las que se ha usado ChatGPT o similares y borrar el resto. Si no se ha usado ninguna, borrar todas y escribir “no he usado ninguna”]:

1. Interpretador de código: Para realizar análisis de datos preliminares.

2. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.

3. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: marzo de 2024

Firma: _____



6. BIBLIOGRAFÍA

Aghion, P., & Howitt, P. W. (1997). *Endogenous Growth Theory*. The MIT Press.

<https://mitpress.mit.edu/9780262528467/endogenous-growth-theory/>

Banton, C. (2022). Per Capita: What It Means, How It's Determined, Uses, and

Examples. *Investopedia*. <https://www.investopedia.com/terms/p/percapita.asp>

Blanco, J. I. (2023). “¿Por qué la normalización es clave e importante en Machine

Learning y Ciencia de Datos? *Medium*. [https://jorgeiblanco.medium.com/por-](https://jorgeiblanco.medium.com/por-qu%C3%A9-la-normalizaci%C3%B3n-es-clave-e-importante-en-machine-learning-y-ciencia-de-datos-4595f15d5be0)

[qu%C3%A9-la-normalizaci%C3%B3n-es-clave-e-importante-en-machine-](https://jorgeiblanco.medium.com/por-qu%C3%A9-la-normalizaci%C3%B3n-es-clave-e-importante-en-machine-learning-y-ciencia-de-datos-4595f15d5be0)

[learning-y-ciencia-de-datos-4595f15d5be0](https://jorgeiblanco.medium.com/por-qu%C3%A9-la-normalizaci%C3%B3n-es-clave-e-importante-en-machine-learning-y-ciencia-de-datos-4595f15d5be0)

Chesbrough, H. W. (2003). *Open Innovation: The New Imperative for Creating and*

Profiting from Technology. Harvard Business Press.

Chiapella, L. (2020). *Impacto de estrategias para el tratamiento de información faltante*

sobre la estimación de modelos de regresión de Cox.

<https://rehip.unr.edu.ar/items/bb3f2db8-32a9-4a61-b669-ebd6c52f2f74>

Comisión Europea. (2023). Finlandia encabeza la clasificación de publicaciones

científicas. *CORDIS | European Commission*.

[https://cordis.europa.eu/article/id/26851-finland-tops-scientific-publishing-](https://cordis.europa.eu/article/id/26851-finland-tops-scientific-publishing-table/es)

[table/es](https://cordis.europa.eu/article/id/26851-finland-tops-scientific-publishing-table/es)

Costello, A. B., & Osborne, J. (2019). *Best practices in exploratory factor analysis:*

Four recommendations for getting the most from your analysis.

<https://doi.org/10.7275/JYJ1-4868>

de los Campos, H. (2021). Relaciones espurias. *la diaria*.

<https://ladiaria.com.uy/ciencia/articulo/2021/8/relaciones-espurias/>

- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299. <https://doi.org/10.1037/1082-989X.4.3.272>
- Fern, o M. y M. G. (2007). *Imputación de datos: Teoría y práctica*. 54.
- Ferrando, P. J., & Anguiano-Carrasco, C. (2010). El Análisis Factorial Como Técnica De Investigación En Psicología. *Papeles del Psicólogo*, 31(1), 18-33.
- Godin, B. (s. f.). *DOCUMENT DE RECHERCHE*.
- Gubert, X. A. (2019). La industria 4.0, el nuevo motor de la innovación industrial. *Dirección y Organización*, 69, Article 69. <https://doi.org/10.37610/dyo.v0i69.563>
- IEB. (2019, febrero 12). ¿Es el PIB el indicador apropiado para conocer el crecimiento de un país? *IEB*. <https://www.ieb.es/es-el-pib-el-indicador-apropiado-para-conocer-el-crecimiento-de-un-pais/>
- Leiva, M. (2023). ¿Qué es y para qué se usa la medición de Paridad de Poder Adquisitivo? <https://chequeado.com/el-explicador/que-es-y-para-que-se-usa-la-medicion-de-paridad-de-poder-adquisitivo/>
- Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). Tools for Composite Indicators Building. *JRC Publications Repository*. <https://publications.jrc.ec.europa.eu/repository/handle/JRC31473>
- OECD. (2007). *Oslo Manual: Guía para la recogida e interpretación de datos sobre innovación, 3ª edición*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/science-and-technology/manual-de-oslo_9789264065659-es

- OECD. (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. *OECD Publishing, Paris*. <https://doi.org/10.1787/9789264043466-en>
- OECD. (2019a). Digital Innovation: Seizing Policy Opportunities. *OECD Publishing, Paris*. <https://doi.org/10.1787/a298dc87-en>
- OECD. (2019b). Measuring the Digital Transformation: A Roadmap for the Future. *OECD Publishing, Paris*. . <https://doi.org/10.1787/a298dc87-en>
- OECD. (2023). Main Science and Technology Indicators, Volume 2022 Issue 2. *OECD Publishing, Paris*. <https://doi.org/10.1787/1cdcb031-en>
- OECD Data. (s. f.). Conversion rates—Purchasing power parities (PPP). *theOECD*. Recuperado 11 de noviembre de 2023, de <http://data.oecd.org/conversion/purchasing-power-parities-ppp.htm>
- Rivera, Ó. (2023). *Cómo la tecnología ayuda a proteger el medio ambiente* | LinkedIn. <https://www.linkedin.com/pulse/c%C3%B3mo-la-tecnolog%C3%ADa-ayuda-protoger-el-medio-ambiente-oscar-rivera-/?originalSubdomain=es>
- Rogers, E. M. (2003). *Diffusion of Innovations, 5th Edition*. Simon and Schuster.
- Román Radío, G. (2017). *Los valores perdidos en el muestreo de poblaciones finitas. Técnicas de imputación*.
- Ruiz González, M., & Mandado Pérez, E. (1989). *La innovación tecnológica y su gestión*. Marcombo.
- S, J. (2021). Paridad de poder adquisitivo: ¿Qué es y cómo se hace el cálculo? *Economia3*. <https://economia3.com/paridad-poder-adquisitivo-que-es-funcion/>
- Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and Sensitivity Techniques as Tools for the Analysis and Validation of Composite Indicators. *Journal of the*

- Royal Statistical Society Series A*, 168, 307-323. <https://doi.org/10.1111/j.1467-985X.2005.00350.x>
- Saltelli, A. (2007). Composite Indicators Between Analysis and Advocacy. *Social Indicators Research*, 81, 65-77. <https://doi.org/10.1007/s11205-006-0024-9>
- Schumpeter, J. A. (1943). *Capitalism, Socialism and Democracy*. Routledge.
- Schwab, K. (2019). Global Competitiveness Report 2019. *Foro Económico Mundial*. <https://es.weforum.org/publications/global-competitiveness-report-2019/>
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications* (pp. x, 195). American Psychological Association. <https://doi.org/10.1037/10694-000>
- Toth, G. (2020). Factor Analysis with Python. *DataSklr*. <https://www.datasklr.com/principal-component-analysis-and-factor-analysis/factor-analysis>
- United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. <https://wedocs.unep.org/xmlui/handle/20.500.11822/9814>
- University, C., Insead, & Wipo. (2020). *Global Innovation Index 2020: Who Will Finance Innovation?*
- Urrego, N. (2023). Tratamiento de valores vacíos II: Estrategias de imputación estadística (moda, mediana y media). *Medium*. <https://medium.com/@nicolasurrego/tratamiento-de-valores-vac%C3%ADos-ii-estrategias-de-imputaci%C3%B3n-estad%C3%ADstica-moda-mediana-y-media-2665b0f53a4c>
- WIPO (Ed.). (2023). *World Intellectual Property Indicators 2023*. World Intellectual Property Organization. <https://doi.org/10.34667/tind.48541>

7. ANEXOS

ANEXO 1: Abreviación de variables

G_NC	GERD en moneda nacional (para el área euro: pre-EMU euro o EUR)
G_PPPCT	GERD a precios constantes o PPA\$ (Paridad de Poder Adquisitivo)
G_GRO	GERD, tasa de crecimiento anual compuesta (a precios constantes)
TP_RS	Investigadores totales (FTE)
TP_RSGRO	Investigadores totales, tasa de crecimiento anual compuesta
TP_RSXLF	Investigadores totales por cada mil trabajadores en la fuerza laboral
TP_RSXEM	Investigadores totales por cada mil empleos totales
TP_TT	Personal I+D total (FTE)
TP_TTGRO	Personal I+D total, tasa de crecimiento anual compuesta
TP_TTXLF	Personal I+D total por cada mil trabajadores en la fuerza laboral
TP_TTXEM	Personal I+D total por cada mil empleos totales
G_XFB	Porcentaje de GERD financiado por el sector empresarial
G_XFG	Porcentaje de GERD financiado por el gobierno
G_XFON	Porcentaje de GERD financiado por sectores de Educación Superior e Instituciones Privadas Sin Fines de Lucro
G_XFA	Porcentaje de GERD financiado por el resto del mundo
G_XEB	Porcentaje de GERD realizado por el sector empresarial
G_XEH	Porcentaje de GERD realizado por el sector de la Educación Superior
G_XEG	Porcentaje de GERD realizado por el sector del gobierno
G_XEI	Porcentaje de GERD realizado por el sector privado sin fines de lucro
TH_RS	Investigadores Totales (número de personas)
TH_WRS	Investigadores mujeres (número de personas)

TH_WRXRS	Investigadores mujeres como porcentaje del total de investigadores (número de personas)
BH_RS	Sector empresarial: Total de investigadores (número de personas)
BH_WRS	Sector empresarial: Investigadores mujeres (número de personas)
BH_WRXRS	Sector empresarial: Investigadores mujeres como porcentaje del total de investigadores (número de personas)
GH_RS	Sector gubernamental: Total de investigadores (número de personas)
GH_WRS	Sector gubernamental: Investigadores mujeres (Nº de personas)
GH_WRXRS	Sector gubernamental: Investigadores mujeres como porcentaje del total de investigadores (Nº de personas)
HH_RS	Sector de educación superior: Total de investigadores (Nº de personas)
HH_WRS	Sector de educación superior: Investigadores mujeres (Nº de personas)
HH_WRXRS	Sector de educación superior: Investigadores mujeres como porcentaje del total de investigadores (número de personas)
B_NC	BERD en moneda nacional (para la zona euro: euro pre-EMU o EUR)
B_PPPCT	BERD a precios constantes y PPA\$ (Paridad de Poder Adquisitivo)
B_GRO	BERD, tasa de crecimiento anual compuesta (a precios constantes)
BP_RS	Investigadores de empresas (FTE)
BP_RSGRO	Investigadores de empresas, tasa de crecimiento anual compuesta
BP_RSXRS	Investigadores de empresas como porcentaje del total nacional
BP_RSXEI	Investigadores de empresas por cada mil empleos en la industria
BP_TT	Personal total de I+D de empresas (FTE)
BP_TTGRO	Personal total de I+D de empresas, tasa de crecimiento anual compuesta
BP_TTXTT	Personal total de I+D de empresas como porcentaje del total nacional
BP_TTXEI	Personal total de I+D de empresas por cada mil empleos en la industria
B_FBCT	BERD financiado por empresas a precios constantes y PPA\$

B_FBGRO	BERD financiado por empresas, tasa de crecimiento anual compuesta (precios constantes)
B_XFB	Porcentaje de BERD financiado por el sector empresarial
B_XFG	Porcentaje de BERD financiado por el gobierno
B_XFON	Porcentaje de BERD financiado por los sectores de Educación Superior y PNP
B_XFA	Porcentaje de BERD financiado por el resto del mundo
B_AERO	BERD realizado en la industria aeroespacial (PPP\$ actuales)
B_XAERO	Porcentaje de BERD realizado en la industria aeroespacial
B_DRUG	BERD realizado en la industria farmacéutica (PPP\$ actuales)
B_XDRUG	Porcentaje de BERD realizado en la industria farmacéutica
B_SERV	BERD realizado en industrias de servicios (PPP\$ actuales)
B_XSERV	Porcentaje de BERD realizado en industrias de servicios
H_NC	HERD en moneda nacional (para la zona euro: euro pre-EMU o EUR)
H_PPPCT	HERD a precios constantes y PPA\$ (Paridad de Poder Adquisitivo)
H_GRO	HERD, tasa de crecimiento anual compuesta (a precios constantes)
H_XFB	Porcentaje de HERD financiado por el sector empresarial
HP_RS	Investigadores de Educación Superior (FTE)
HP_RSGRO	Investigadores de Educación Superior, tasa de crecimiento anual compuesta
HP_RSXRS	Investigadores de Educación Superior como porcentaje del nacional total
HP_TT	Personal total de I+D de Educación Superior (FTE)
HP_TTGRO	Personal total de I+D de Educación Superior, tasa de crecimiento anual compuesta

GV_NC	GOVERD en moneda nacional (para la zona euro: euro pre-EMU o EUR)
GV_PPPCT	GOVERD total a precios constantes y PPA\$
GV_GRO	GOVERD, tasa de crecimiento anual compuesta (precios constantes)
GV_XFB	Porcentaje de GOVERD financiado por el sector empresarial
GP_RS	Investigadores del gobierno (FTE)
GP_RSGRO	Investigadores del gobierno, tasa de crecimiento anual compuesta
GP_RSXRS	Investigadores del gobierno como porcentaje del nacional total
GP_TT	Personal total de I+D del gobierno (FTE)
GP_TTGRO	Personal total de I+D del gobierno, tasa de crecimiento anual compuesta
C_NC	GBARD en moneda nacional (para la zona euro: euro pre-EMU o EUR)
C_DFXTT	Presupuesto de Defensa en I+D como porcentaje del total de GBARD
C_CVXTT	Presupuesto Civil en I+D como porcentaje del total de GBARD
C_ECOPPP	GBARD civil para programas de desarrollo económico (PPP\$ actuales)
C_ECOXCV	Programas de desarrollo económico como porcentaje de GBARD civil
C_HEAPPP	GBARD civil para programas de Salud y Medio Ambiente (PPP\$ actuales)
C_HEAXCV	Programas de Salud y Medio Ambiente como porcentaje de GBARD civil
C_SPAPPP	GBARD civil para programas espaciales (PPP\$ actuales)
C_SPAXCV	Programas espaciales como porcentaje de GBARD civil
C_NORPPP	GBARD civil para programas de Investigación no orientada (PPP\$ actuales)
C_NORXCV	Programas de Investigación no orientada como porcentaje de GBARD civil

C_GUFPPP	GBARD civil para Fondos Generales Universitarios (FGU) (PPP\$ actuales)
C_GUFXCV	Fondos Generales Universitarios (FGU) como porcentaje de GBARD civil
P_TRIAD	Número de familias de patentes 'triádicas' (año de prioridad)
P_XTRIAD	Participación de países en familias de patentes 'triádicas' (año de prioridad)
TD_XAERO	Cuota de mercado de exportación: Industria aeroespacial
TD_IAERO	Importaciones totales: Industria aeroespacial (precios actuales)
TD_EAERO	Exportaciones totales: Industria aeroespacial (precios actuales)
TD_XDRUG	Cuota de mercado de exportación: Industria farmacéutica
TD_IDRUG	Importaciones totales: Industria farmacéutica (precios actuales)
TD_EDRUG	Exportaciones totales: Industria farmacéutica (precios actuales)
C_EDUXCV	Programas de Educación y Sociales como porcentaje de GBARD civil
C_EDUPPP	GBARD civil para programas de Educación y Sociales (PPP\$ actuales)
P_PCT	Número de solicitudes de patentes presentadas bajo el Tratado de Cooperación en Materia de Patentes (PCT) (año de prioridad)
P_BIOPCT	Número de patentes en el sector de la biotecnología, solicitudes presentadas bajo el PCT (año de prioridad)
P_ICTPCT	Número de patentes en el sector de las TIC - solicitudes presentadas bajo el PCT (año de prioridad)
B_XCOMP	Porcentaje de BERD realizada en la industria informática, electrónica y óptica
TD_ECOMP	Exportaciones totales: industria informática, electrónica y óptica (precios actuales)
TD_XCOMP	Cuota de mercado de exportación: Industria informática, electrónica y óptica

TD_ICOMP	Importaciones totales: Industria informática, electrónica y óptica (precios actuales)
B_COMP	BERD realizada en la industria informática, electrónica y óptica (PPP\$ actuales)
C_PPPCT	GBARD total a precios constantes y PPA\$

Otros:

GDP	<i>Gross Domestic Product</i> : Producto Interior Bruto
ISIC	<i>International Standard Industrial Classification</i> : Clasificación Internacional Industrial Uniforme
PCT	<i>Patent Co-operation Treaty</i> : Tratado de Cooperación en materia de Patentes
PPP	<i>Purchasing Power Parity</i> : Paridad de Poder Adquisitivo (PPA)

ANEXO 2: Pesos factoriales, pesos de variables y pesos dimensionales (AG 1)

Pesos factoriales

	Factor 1	Factor 2	Factor 3
HERD & BERD	0.5231	0.3944	0.0826
Patents & Programmes	0.7048	0.2952	nan
R&D General expenditure	0.6105	0.2536	0.1360
Research Personnel	1.0000	nan	nan
Trade Balance	0.6523	0.3477	nan

Pesos de las variables por factor y por grupo

Grupo: HERD & BERD

	Factor 1	Factor 2	Factor 3
0	0.0928	nan	nan
1	0.1469	nan	nan
2	0.1479	nan	nan
3	nan	0.2098	nan
4	nan	nan	0.1457
5	0.1151	nan	nan
6	0.0924	nan	nan
7	0.1457	nan	nan
8	0.1468	nan	nan
9	nan	0.2093	nan
10	nan	0.2097	nan
11	nan	0.2093	nan
12	nan	0.1620	nan
13	0.1124	nan	nan
14	nan	nan	0.6983
15	nan	nan	0.1560

Grupo: Patents & Programmes

	Factor 1	Factor 2
0	0.0974	nan
1	0.1022	nan
2	0.0975	nan
3	0.1022	nan
4	0.0945	nan
5	0.1022	nan
6	0.0975	nan
7	0.1022	nan
8	0.1022	nan
9	0.1022	nan
10	nan	0.2500
11	nan	0.2500
12	nan	0.2500
13	nan	0.2500

Grupo: R&D General expenditure

	Factor 1	Factor 2	Factor 3
0	nan	nan	0.2308
1	nan	nan	0.2308
2	nan	nan	0.1947
3	nan	nan	0.2308
4	nan	nan	0.1128
5	0.0921	nan	nan
6	nan	0.0788	nan
7	0.1909	nan	nan
8	0.1894	nan	nan
9	0.1894	nan	nan
10	0.1894	nan	nan
11	0.1486	nan	nan
12	nan	0.2298	nan
13	nan	0.2309	nan
14	nan	0.2298	nan
15	nan	0.2309	nan

Grupo: Research Personnel

	Factor 1
0	0.2571
1	0.2436
2	0.2561
3	0.2432

Grupo: Trade Balance

	Factor 1	Factor 2
0	nan	0.3374
1	nan	0.3259
2	nan	0.3367
3	0.3333	nan
4	0.3333	nan
5	0.3333	nan

ANEXO 3: Ejemplo de ecuación del indicador paso a paso para EE. UU en 2020

- Grupo HERD & BERD

--- Grupo: HERD & BERD ---

- Factor 1 -

Contribución de BP_RSXEI: 27.69833451804058 * 0.09276048518183153 = 2.5693109486221157
Contribución de BP_TTXEI: 27.69833451972528 * 0.14685763687027106 = 4.067711952809109
Contribución de B_AERO: 30.05020982722361 * 0.1479359605453727 = 4.445506655380323
Contribución de B_COMP: 37.98562417414141 * 0.0 = 0.0
Contribución de B_DRUG: 37.09218457901894 * 0.0 = 0.0
Contribución de B_FBCT: 72.38039440727728 * 0.11513310806999315 = 8.333379771441782
Contribución de B_PPPCT: 79.69597507931122 * 0.092389489169437 = 7.363070426437745
Contribución de B_SERV: 51.37336316540973 * 0.1456761853802418 = 7.483875576090713
Contribución de B_XAERO: 27.6987627970459 * 0.1468380332723055 = 4.067231853194324
Contribución de B_XCOMP: 27.70020785657467 * 0.0 = 0.0
Contribución de B_XDRUG: 27.70004515890775 * 0.0 = 0.0
Contribución de B_XFA: 27.6992017872351 * 0.0 = 0.0
Contribución de B_XFB: 27.70708868797629 * 0.0 = 0.0
Contribución de B_XFG: 27.69888568241157 * 0.11240910151054737 = 3.113606852403249
Contribución de B_XFON: 27.69834933444921 * 0.0 = 0.0
Contribución de B_XSERV: 27.70264579855884 * 0.0 = 0.0

Peso del Factor 1: 0.523064988645809

Contribución del Factor 1: 21.677745350579148

- Factor 2 -

Contribución de BP_RSXEI: 27.69833451804058 * 0.0 = 0.0
Contribución de BP_TTXEI: 27.69833451972528 * 0.0 = 0.0
Contribución de B_AERO: 30.05020982722361 * 0.0 = 0.0
Contribución de B_COMP: 37.98562417414141 * 0.2097914127147094 = 7.969057758343142
Contribución de B_DRUG: 37.09218457901894 * 0.0 = 0.0
Contribución de B_FBCT: 72.38039440727728 * 0.0 = 0.0
Contribución de B_PPPCT: 79.69597507931122 * 0.0 = 0.0
Contribución de B_SERV: 51.37336316540973 * 0.0 = 0.0
Contribución de B_XAERO: 27.6987627970459 * 0.0 = 0.0
Contribución de B_XCOMP: 27.70020785657467 * 0.20929419033113678 = 5.79749257534599
Contribución de B_XDRUG: 27.70004515890775 * 0.20967105065841685 = 5.807897571753781
Contribución de B_XFA: 27.6992017872351 * 0.20929112277645012 = 5.797197042061888
Contribución de B_XFB: 27.70708868797629 * 0.16195222351928676 = 4.487224620263838
Contribución de B_XFG: 27.69888568241157 * 0.0 = 0.0
Contribución de B_XFON: 27.69834933444921 * 0.0 = 0.0
Contribución de B_XSERV: 27.70264579855884 * 0.0 = 0.0

Peso del Factor 2: 0.39438120508301117

Contribución del Factor 2: 11.775776962553044

- Factor 3 -

Contribución de BP_RSXEI: 27.69833451804058 * 0.0 = 0.0
Contribución de BP_TTXEI: 27.69833451972528 * 0.0 = 0.0
Contribución de B_AERO: 30.05020982722361 * 0.0 = 0.0
Contribución de B_COMP: 37.98562417414141 * 0.0 = 0.0
Contribución de B_DRUG: 37.09218457901894 * 0.14574280333432585 = 5.405918962340471
Contribución de B_FBCT: 72.38039440727728 * 0.0 = 0.0
Contribución de B_PPPCT: 79.69597507931122 * 0.0 = 0.0
Contribución de B_SERV: 51.37336316540973 * 0.0 = 0.0
Contribución de B_XAERO: 27.6987627970459 * 0.0 = 0.0
Contribución de B_XCOMP: 27.70020785657467 * 0.0 = 0.0
Contribución de B_XDRUG: 27.70004515890775 * 0.0 = 0.0
Contribución de B_XFA: 27.6992017872351 * 0.0 = 0.0
Contribución de B_XFB: 27.70708868797629 * 0.0 = 0.0
Contribución de B_XFG: 27.69888568241157 * 0.0 = 0.0
Contribución de B_XFON: 27.69834933444921 * 0.6982623695672133 = 19.340715039372952
Contribución de B_XSERV: 27.70264579855884 * 0.15599482709846088 = 4.32146944151609

Peso del Factor 3: 0.08255380627117986

Contribución del Factor 3: 2.3996825803229855

Suma de las contribuciones de todos los factores en el grupo HERD & BERD: 35.85320489345518

Contribución del Grupo HERD & BERD elevada al peso dimensional: (35.85320489345518) ** 0.2 = 2.0459998456128288

- Grupo Patents & Programmes

--- Grupo: Patents & Programmes ---

- Factor 1 -

Contribución de BP_RSXEI: 27.69833451804058 * 0.09737568261470736 = 2.697144230984713
Contribución de BP_TTXEI: 27.69833451972528 * 0.10219961742983807 = 2.830759191359601
Contribución de B_AERO: 30.05020982722361 * 0.09746197099709628 = 2.928752678637525
Contribución de B_COMP: 37.98562417414141 * 0.10219908212953334 = 3.8820959247146645
Contribución de B_DRUG: 37.09218457901894 * 0.0944989801253029 = 3.505173613336778
Contribución de B_FBCT: 72.38039440727728 * 0.10219937388027565 = 7.397230989631144
Contribución de B_PPPCT: 79.69597507931122 * 0.09745533226724906 = 7.766797731716676
Contribución de B_SERV: 51.37336316540973 * 0.1021987876357341 = 5.250295432275154
Contribución de B_XAERO: 27.6987627970459 * 0.10221193258983505 = 2.8311440758334867
Contribución de B_XCOMP: 27.70020785657467 * 0.10219924033042817 = 2.8309401999368893
Contribución de B_XDRUG: 27.70004515890775 * 0.0 = 0.0
Contribución de B_XFA: 27.6992017872351 * 0.0 = 0.0
Contribución de B_XFB: 27.70708868797629 * 0.0 = 0.0
Contribución de B_XFG: 27.69888568241157 * 0.0 = 0.0

Peso del Factor 1: 0.7048413535702586

Contribución del Factor 1: 29.547185006907252

- Factor 2 -

Contribución de BP_RSXEI: 27.69833451804058 * 0.0 = 0.0
Contribución de BP_TTXEI: 27.69833451972528 * 0.0 = 0.0
Contribución de B_AERO: 30.05020982722361 * 0.0 = 0.0
Contribución de B_COMP: 37.98562417414141 * 0.0 = 0.0
Contribución de B_DRUG: 37.09218457901894 * 0.0 = 0.0
Contribución de B_FBCT: 72.38039440727728 * 0.0 = 0.0
Contribución de B_PPPCT: 79.69597507931122 * 0.0 = 0.0
Contribución de B_SERV: 51.37336316540973 * 0.0 = 0.0
Contribución de B_XAERO: 27.6987627970459 * 0.0 = 0.0
Contribución de B_XCOMP: 27.70020785657467 * 0.0 = 0.0
Contribución de B_XDRUG: 27.70004515890775 * 0.25000000005437467 = 6.925011291233118
Contribución de B_XFA: 27.6992017872351 * 0.2499999999083879 = 6.924800444271193
Contribución de B_XFB: 27.70708868797629 * 0.25000000000843553 = 6.926772172227797
Contribución de B_XFG: 27.69888568241157 * 0.25000000002880207 = 6.924721421400678

Peso del Factor 2: 0.29515864642974143

Contribución del Factor 2: 8.176279785283816

Suma de las contribuciones de todos los factores en el grupo Patents & Programmes: 37.723464792191066

Contribución del Grupo Patents & Programmes elevada al peso dimensional: (37.723464792191066) ** 0.2 = 2.06691356219111

63

- Grupo R&D General expenditure

--- Grupo: R&D General expenditure ---

- Factor 1 -

Contribución de BP_RSXEI: 27.69833451804058 * 0.0 = 0.0
Contribución de BP_TTXEI: 27.69833451972528 * 0.0 = 0.0
Contribución de B_AERO: 30.05020982722361 * 0.0 = 0.0
Contribución de B_COMP: 37.98562417414141 * 0.0 = 0.0
Contribución de B_DRUG: 37.09218457901894 * 0.0 = 0.0
Contribución de B_FBCT: 72.38039440727728 * 0.09213304750942862 = 6.66862631667686
Contribución de B_PPPCT: 79.69597507931122 * 0.0 = 0.0
Contribución de B_SERV: 51.37336316540973 * 0.19093728846982086 = 9.809090662378706
Contribución de B_XAERO: 27.6987627970459 * 0.18944139767565016 = 5.247292338158677
Contribución de B_XCOMP: 27.70020785657467 * 0.18944324949890579 = 5.247617388144626
Contribución de B_XDRUG: 27.70004515890775 * 0.1894459920987277 = 5.247662536308837
Contribución de B_XFA: 27.6992017872351 * 0.14859902474746692 = 4.1160743718664285
Contribución de B_XFB: 27.70708868797629 * 0.0 = 0.0
Contribución de B_XFG: 27.69888568241157 * 0.0 = 0.0
Contribución de B_XFON: 27.69834933444921 * 0.0 = 0.0
Contribución de B_XSERV: 27.70264579855884 * 0.0 = 0.0

Peso del Factor 1: 0.6104715238235386

Contribución del Factor 1: 22.182315265360366

- Factor 2 -

Contribución de BP_RSXEI: 27.69833451804058 * 0.0 = 0.0
Contribución de BP_TTXEI: 27.69833451972528 * 0.0 = 0.0
Contribución de B_AERO: 30.05020982722361 * 0.0 = 0.0
Contribución de B_COMP: 37.98562417414141 * 0.0 = 0.0
Contribución de B_DRUG: 37.09218457901894 * 0.0 = 0.0
Contribución de B_FBCT: 72.38039440727728 * 0.0 = 0.0
Contribución de B_PPPCT: 79.69597507931122 * 0.07877145421752077 = 6.277767852280641
Contribución de B_SERV: 51.37336316540973 * 0.0 = 0.0
Contribución de B_XAERO: 27.6987627970459 * 0.0 = 0.0
Contribución de B_XCOMP: 27.70020785657467 * 0.0 = 0.0
Contribución de B_XDRUG: 27.70004515890775 * 0.0 = 0.0
Contribución de B_XFA: 27.6992017872351 * 0.0 = 0.0
Contribución de B_XFB: 27.70708868797629 * 0.22976113319680008 = 6.366012094533674
Contribución de B_XFG: 27.69888568241157 * 0.23086049366152883 = 6.394578422515789
Contribución de B_XFON: 27.69834933444921 * 0.2297523093197718 = 6.363759723935471
Contribución de B_XSERV: 27.70264579855884 * 0.2308546096043785 = 6.395283480834678

Peso del Factor 2: 0.25356884799069795
Contribución del Factor 2: 8.062830486242207

- Factor 3 -

Contribución de BP_RSXEI: 27.69833451804058 * 0.2308224985414517 = 6.393398778891063
Contribución de BP_TTXEI: 27.69833451972528 * 0.2308165202097869 = 6.393233189449608
Contribución de B_AERO: 30.05020982722361 * 0.1947287178331331 = 5.85163883027187
Contribución de B_COMP: 37.98562417414141 * 0.23082324626857773 = 8.76796508341348
Contribución de B_DRUG: 37.09218457901894 * 0.11280901714705062 = 4.184332886196114
Contribución de B_FBCT: 72.38039440727728 * 0.0 = 0.0
Contribución de B_PPPCT: 79.69597507931122 * 0.0 = 0.0
Contribución de B_SERV: 51.37336316540973 * 0.0 = 0.0
Contribución de B_XAERO: 27.6987627970459 * 0.0 = 0.0
Contribución de B_XCOMP: 27.70020785657467 * 0.0 = 0.0
Contribución de B_XDRUG: 27.70004515890775 * 0.0 = 0.0
Contribución de B_XFA: 27.6992017872351 * 0.0 = 0.0
Contribución de B_XFB: 27.70708868797629 * 0.0 = 0.0
Contribución de B_XFG: 27.69888568241157 * 0.0 = 0.0
Contribución de B_XFON: 27.69834933444921 * 0.0 = 0.0
Contribución de B_XSERV: 27.70264579855884 * 0.0 = 0.0

Peso del Factor 3: 0.13595962818576338
Contribución del Factor 3: 4.29504198390427

Suma de las contribuciones de todos los factores en el grupo R&D General expenditure: 34.54018773550684
Contribución del Grupo R&D General expenditure elevada al peso dimensional: (34.54018773550684) ** 0.2 = 2.030789637891
913

- **Grupo Research Personnel**

--- Grupo: Research Personnel ---

- Factor 1 -

Contribución de BP_RSXEI: 27.69833451804058 * 0.257093890656433 = 7.121072585946429
Contribución de BP_TTXEI: 27.69833451972528 * 0.2435986658988382 = 6.747277336624815
Contribución de B_AERO: 30.05020982722361 * 0.2560864010620998 = 7.695450085814638
Contribución de B_COMP: 37.98562417414141 * 0.24322104238262898 = 9.238903107189463
Peso del Factor 1: 1.0
Contribución del Factor 1: 30.802703115575348

Suma de las contribuciones de todos los factores en el grupo Research Personnel: 30.802703115575348
Contribución del Grupo Research Personnel elevada al peso dimensional: (30.802703115575348) ** 0.2 = 1.98480463754777

- Grupo Trade Balance

--- Grupo: Trade Balance ---

- Factor 1 -

Contribución de BP_RSXEI: 27.69833451804058 * 0.0 = 0.0
Contribución de BP_TTXEI: 27.69833451972528 * 0.0 = 0.0
Contribución de B_AERO: 30.05020982722361 * 0.0 = 0.0
Contribución de B_COMP: 37.98562417414141 * 0.33333326968545873 = 12.661872307009558
Contribución de B_DRUG: 37.09218457901894 * 0.3333333542900946 = 12.364062303671703
Contribución de B_FBCT: 72.38039440727728 * 0.33333337602444657 = 24.126801225758708

Peso del Factor 1: 0.6522660530720015

Contribución del Factor 1: 32.06066100172542

- Factor 2 -

Contribución de BP_RSXEI: 27.69833451804058 * 0.3373871980190264 = 9.34506347283539
Contribución de BP_TTXEI: 27.69833451972528 * 0.3259356205152062 = 9.027873848924415
Contribución de B_AERO: 30.05020982722361 * 0.33667718146576736 = 10.11721994708455
Contribución de B_COMP: 37.98562417414141 * 0.0 = 0.0
Contribución de B_DRUG: 37.09218457901894 * 0.0 = 0.0
Contribución de B_FBCT: 72.38039440727728 * 0.0 = 0.0

Peso del Factor 2: 0.3477339469279985

Contribución del Factor 2: 9.906994835694656

Suma de las contribuciones de todos los factores en el grupo Trade Balance: 41.967655837420075

Contribución del Grupo Trade Balance elevada al peso dimensional: (41.967655837420075) ** 0.2 = 2.1114604078612054

- Multiplicación de dimensiones

Multiplicación de las contribuciones de todos los grupos al indicador compuesto: 35.99097058528304

ANEXO 4: Pesos factoriales, pesos de variables y pesos dimensionales (AG 2)

Pesos factoriales

	Factor 1	Factor 2	Factor 3	Factor 4
Impact indicators	0.5154	0.3233	0.1613	nan
Input indicators	0.5094	0.2518	0.1397	0.0991
Output indicators	0.7049	0.2951	nan	nan

Pesos de las variables por factor y por grupo

Grupo: Impact indicators

	Factor 1	Factor 2	Factor 3
0	0.1346	nan	nan
1	0.0619	nan	nan
2	0.0537	nan	nan
3	0.1186	nan	nan
4	0.1161	nan	nan
5	nan	0.3319	nan
6	nan	0.3315	nan
7	nan	0.3366	nan
8	nan	nan	0.3333
9	nan	nan	0.3333
10	nan	nan	0.3333
11	0.1285	nan	nan
12	0.1281	nan	nan
13	0.1346	nan	nan
14	0.0660	nan	nan
15	0.0579	nan	nan

Grupo: Input indicators

	Factor 1	Factor 2	Factor 3	Factor 4
0	nan	nan	0.0942	nan
1	nan	nan	0.1466	nan
2	nan	nan	0.1473	nan
3	0.1085	nan	nan	nan
4	nan	0.1213	nan	nan
5	nan	nan	0.1149	nan
6	nan	nan	0.0938	nan
7	nan	nan	0.1452	nan
8	nan	nan	0.1461	nan
9	0.1086	nan	nan	nan
10	0.1093	nan	nan	nan
11	0.1086	nan	nan	nan
12	0.0807	nan	nan	nan
13	nan	nan	0.1119	nan
14	nan	nan	nan	0.2291
15	nan	nan	nan	0.2291
16	nan	nan	nan	0.1958
17	nan	nan	nan	0.2291
18	nan	nan	nan	0.1169
19	nan	0.0960	nan	nan
20	0.0335	nan	nan	nan
21	nan	0.1458	nan	nan
22	nan	0.1464	nan	nan
23	nan	0.1464	nan	nan
24	nan	0.1464	nan	nan
25	nan	0.1022	nan	nan
26	0.1040	nan	nan	nan
27	0.1047	nan	nan	nan
28	0.1040	nan	nan	nan
29	0.1047	nan	nan	nan
30	nan	0.0956	nan	nan
31	0.0336	nan	nan	nan

Grupo: Output indicators

	Factor 1	Factor 2
0	0.0974	nan
1	0.1022	nan
2	0.0975	nan
3	0.1022	nan
4	0.0944	nan
5	0.1022	nan
6	0.0975	nan
7	0.1022	nan
8	0.1022	nan
9	0.1022	nan
10	nan	0.2500
11	nan	0.2500
12	nan	0.2500
13	nan	0.2500