FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

El desarrollo autonómico de la industria del juego en España: construcción de un Índice de Desarrollo del Juego para la comparativa regional

Autor: Álvaro Gimeno Granado
5º, E3 Analytics

Madrid
Junio 2026
TRABAJO DE FIN DE GRADO

ÍNDICE

1. INTRODUCCIÓN 4

2. MARCO TEÓRICO 6

3. METODOLOGÍA 10

3.1 Variables, pirámide de indicadores y conversión de absolutos en ratios 10
3.1.1 Planteamiento y unidad de análisis 10
3.1.2 Cuatro dimensiones 12
3.1.3 Variables de contexto y denominadores 13
3.1.4 Dimensión económica 14
3.1.5 Dimensión intensidad de consumo 15
3.1.6 Dimensión estructural 16
3.1.7 Dimensión de dinámica 17
3.1.8 Los indicadores legislativos 18
3.1.9 Pirámide o árbol de indicadores 18
3.2 Procesamiento de los datos 19
3.2.1 Imputación de valores faltantes 19
3.2.2 Cálculo de tasas de variación para la dimensión de dinámica 21
3.3 Cálculo de correlaciones y eliminación de variables redundantes 22
3.3.1 Correlaciones encontradas 23
3.3.2 Medidas tomadas 24
3.4 Normalización 25
3.5 Ponderación 26
3.5.1 Ponderación mediante pesos iguales 26
3.5.2 Ponderación mediante criterio experto 27
3.5.3 Ponderación mediante Análisis de Componentes Principales (PCA) 28
3.5.4 Conclusiones sobre ponderación 28
3.6 Agregación 29
3.7 Análisis de sensibilidad 30
3.7.1 Técnicas empleadas 31
3.7.2 Robustez global entre escenarios 32
3.7.3 El peso de las decisiones metodológicas 33
3.7.4 Estabilidad por comunidad 34
3.7.5 Conclusión del análisis de sensibilidad 37

4. RESULTADOS 38
4.1 Elección del ranking de referencia 38
4.2 El ranking del IDJ en 2024 40
4.3 el perfil por dimensiones 42
4.4 Evolución temporal del ranking (2021-2024) 43
4.5 Síntesis 44
5. CONCLUSIÓN 45
*DECLARACIÓN USO DE INTELIGENCIA ARTIFICIAL 50
6. BIBLIOGRAFÍA 51
7. ANEXO 53
7.1 Tabla comparativa entre métodos de ponderación 53
7.2 Intervalos del 90% del ranking por simulación de Monte Carlo 53
7.3 Evolución del puesto en el ranking de referencia (2021-2024) 56
7.4 Ranking del IDJ y perfil por dimensión (2024, escala 0-100) 56
7.5 Variable 1: Número de locales de juegos de azar y apuestas 57
7.6 Variable 2: cifra de negocios 58
7.7 Variable 3: sueldos y salarios 58
7.8 Variable 4: Inversión en activos materiales 59
7.9 Variable 5: Personal ocupado en Juegos de azar y apuestas 60
7.10 Variable 6: Nº de empresas de JdA&A 60
7.11 Variable 7: Muertes de empresas de JdA&A 61
7.12 Variable 8: Locales de JdA&A 62
7.13 Variable 9: Cantidades jugadas en Apuestas 62
7.14 Variable 10: Cantidades jugadas en Loterías 63
7.15 Variables 11 y 12: cantidades totales jugadas 64
7.16 Magnitudes base 64
7.17 Variables legislativas recopiladas 66
7.18 Código 67

1. INTRODUCCIÓN

La industria del juego y las apuestas constituye una actividad económica con una presencia significativa dentro de la economía española. Durante las últimas décadas, el sector ha experimentado importantes transformaciones derivadas de los cambios tecnológicos, la expansión de las apuestas deportivas, la digitalización de numerosos productos de juego y la evolución de los hábitos de consumo. Sin embargo, aunque existen numerosas estadísticas sobre distintos aspectos de esta actividad, resulta mucho más difícil responder a una pregunta tan sencilla como: ¿qué regiones españolas presentan un mayor grado de desarrollo de la industria del juego?

La respuesta no es evidente porque el desarrollo de un sector económico es un fenómeno complejo y multidimensional. No puede explicarse únicamente mediante una sola variable, como la facturación, el número de empresas o la cantidad de dinero jugado. Cada una de estas magnitudes aporta información relevante, pero ninguna es capaz de describir por sí sola la realidad completa del sector. Una comunidad autónoma puede presentar una elevada actividad económica vinculada al juego, mientras que otra puede destacar por la intensidad de consumo de sus ciudadanos o por la densidad de su tejido empresarial especializado. Por ello, resulta necesario adoptar una perspectiva más amplia que integre simultáneamente distintas dimensiones del fenómeno.

En este contexto surge la principal pregunta de investigación que guía este trabajo: ¿Es posible construir un indicador compuesto que permita medir y comparar el grado de desarrollo de la industria del juego entre las distintas comunidades autónomas españolas?

Para responder a la cuestión, el trabajo consiste en desarrollar un Índice de Desarrollo del Juego (IDJ) construido a partir de información procedente exclusivamente de fuentes oficiales. Mediante técnicas propias del ámbito de Business Analytics y de la construcción de indicadores compuestos, se integran variables económicas, estructurales, de intensidad de consumo y de dinámica sectorial en una única medida sintética capaz de ordenar a las comunidades autónomas según el nivel de desarrollo relativo de su industria del juego.

La relevancia de esta investigación se puede analizar por partes. En primer lugar, permite transformar un conjunto amplio y heterogéneo de datos en una herramienta de análisis fácilmente interpretable. En segundo lugar, proporciona una metodología replicable que puede ser utilizada en futuras investigaciones sobre el sector. Finalmente, ofrece una visión comparativa del desarrollo autonómico de la actividad del juego, identificando similitudes y diferencias territoriales que serían difíciles de apreciar mediante el análisis aislado de variables individuales.

El objetivo se materializa con la obtención de un ranking autonómico contrastado, así como en la evaluación de su robustez a la hora de reflejar los resultados. Además, como parte de los objetivos también se incluye identificar hasta qué punto las conclusiones dependen de las decisiones metodológicas adoptadas durante la construcción del índice.

Conviene señalar que este estudio forma parte de una línea de investigación más amplia. La idea inicial que ha dado origen al proyecto estaba principalmente centrada en el análisis comparado de la regulación del juego en las distintas comunidades autónomas españolas. El propósito era estudiar cómo las diferencias normativas podían influir sobre la realidad del sector y elaborar un ranking autonómico basado en criterios legislativos. Sin embargo, durante la fase de diseño se consideró que resultaba metodológicamente más adecuado abordar primero la medición del propio desarrollo de la industria mediante indicadores objetivos y cuantificables. Por este motivo, la dimensión regulatoria ha quedado en un segundo plano, y se reserva como una futura línea de investigación complementaria. Los datos legislativos recopilados podrán servir posteriormente para construir un segundo indicador centrado en la intensidad regulatoria y analizar la relación existente entre regulación y el desarrollo del sector.

Además de su interés académico, la elección del tema responde también a una motivación personal. Como estudiante del doble grado en Business Analytics, siempre me ha resultado especialmente interesante la aplicación de técnicas cuantitativas para analizar fenómenos económicos complejos. El sector del juego presenta una combinación particularmente atractiva de elementos económicos, empresariales, sociales y regulatorios, convirtiéndolo en un ámbito interesante para aplicar herramientas avanzadas de análisis de datos. Asimismo, la existencia de diferencias territoriales importantes dentro de España ofrece una oportunidad interesante para estudiar cómo una misma actividad económica puede desarrollarse de forma distinta según el contexto regional.

En definitiva, este trabajo pretende contribuir al conocimiento de la realidad autonómica del sector del juego en España mediante la construcción de un indicador compuesto capaz de sintetizar en una única medida las múltiples dimensiones identificadas. A través de esta aproximación, se busca ofrecer una respuesta rigurosa y fundamentada a la pregunta de investigación planteada, aportando una herramienta útil para el análisis comparativo del desarrollo regional de esta industria.

2. MARCO TEÓRICO

Medir fenómenos económicos y sociales complejos es uno de los principales retos en Business Analytics y en las ciencias sociales aplicadas. A menudo, los conceptos que se quieren analizar no se pueden observar directamente con una única variable. En su lugar, se manifiestan a través de múltiples dimensiones interrelacionadas. Conceptos como el desarrollo económico, la competitividad regional, la innovación, la sostenibilidad y la calidad de vida son ejemplos comunes de realidades multidimensionales. Evaluarlas requiere combinar información de diversas fuentes e indicadores.

En este contexto, surgen los indicadores compuestos. Estos indicadores son herramientas estadísticas que combinan diferentes variables individuales en una única medida capaz de representar un fenómeno complejo. Los indicadores compuestos son una gran herramienta para resumir grandes volúmenes de información, facilitan comparaciones entre territorios o unidades de análisis y apoyan la toma de decisiones tanto públicas como privadas. Un referente importante para este trabajo ha sido la metodología desarrollada por la Organización para la Cooperación y el Desarrollo Económicos (OCDE) y el Joint Research Centre de la Comisión Europea. Esta metodología establece un marco sistemático para crear indicadores compuestos, basado en etapas sucesivas como la selección de variables, el tratamiento de datos, la normalización, la ponderación, la agregación y el análisis de sensibilidad, todas ellas presentes en la metodología desarrollada en este trabajo.[1]

Previo a este documento también la Comisión Europea encomendó una publicación en relación a la creación de indicadores compuestos titulada *Tools for Composite Indicators Building,* el cual ha sido de gran utilidad para asimilar muchos de los conocimientos

[1] OECD, *Handbook on Constructing Composite Indicators: Methodology and User Guide*, OECD Publishing, París, 2008, pp. 19-22

aplicados en los apartados donde se analiza la sensibilidad y robustez de dichos indicadores.[2] Por último, la obra *Methods for Constructing Composite Indices* también ha jugado un papel importante por su aporte crítico a conceptos como la distinción entre media aritmética y geométrica, así como otros conocimientos básicos para los métodos de agregación.[3]

El uso de indicadores compuestos ha crecido notablemente en las últimas décadas. Algunos ejemplos bien conocidos son el Índice de Desarrollo Humano (IDH) elaborado por el Programa de las Naciones Unidas para el Desarrollo, o el conocido sistema Nutri-Score, utilizado en diversos países europeos para resumir la calidad nutricional de los alimentos y sintetizar en una letra o color múltiples variables relacionadas con el contenido nutricional del producto. Estos índices y muchos otros surgen de una idea común: cuando un fenómeno es multidimensional, ninguna variable aislada es suficiente para capturar toda su complejidad.

El desarrollo del sector de los juegos de azar y las apuestas es un ejemplo de realidad multidimensional. Su análisis no puede limitarse a una única cifra económica, como la facturación o el volumen de juego. Varios factores relacionados con la actividad empresarial, la estructura productiva, el consumo y la evolución del mercado intervienen al mismo tiempo. Por ejemplo, una comunidad autónoma puede mostrar un gran volumen económico del sector, pero tener pocos negocios establecidos. Otra región puede contar con buena infraestructura, aunque su crecimiento reciente sea limitado. Por esta razón, construir un indicador compuesto es especialmente adecuado para sintetizar estas diferentes dimensiones en una sola medida, permitiendo comparaciones entre territorios. Precisamente, este trabajo propone la construcción de un Índice de Desarrollo del Juego (IDJ) que integre distintas dimensiones del sector para obtener una medida sintética comparable entre las comunidades autónomas españolas.

Desde el enfoque de Business Analytics, este tipo de métodos se incluyen en técnicas de análisis multicriterio y de integración de datos. Su objetivo no es solo describir la realidad

[2] NARDO, Michela; SAISANA, Michaela; SALTELLI, Andrea; TARANTOLA, Stefano, *Tools for Composite Indicators Building*, European Commission, Joint Research Centre, Ispra, 2005.
[3] MAZZIOTTA, Matteo; PARETO, Adriano, *Methods for Constructing Composite Indices: One for All or All for One?*, Rivista Italiana di Economia Demografia e Statistica, vol. 67, núm. 2, 2013, pp. 67-80.

observada. Buscan transformar grandes conjuntos de datos diversos en información que se pueda interpretar y usar para la toma de decisiones. Crear indicadores compuestos también comparte muchos elementos metodológicos con otras áreas del análisis de datos, como la reducción de dimensionalidad, la minería de datos y los sistemas de soporte a la decisión. En este sentido, el presente trabajo combina información procedente de distintas fuentes oficiales para transformarla en un indicador único capaz de sintetizar el grado de desarrollo del sector en cada territorio.

Un aspecto clave en la creación de indicadores compuestos es la elección de las variables que conforman el sistema. La literatura sugiere que las variables seleccionadas deben seguir una lógica conceptual y representar dimensiones claramente diferenciadas del fenómeno analizado. Por lo tanto, la construcción del Índice de Desarrollo del Juego se basa en una estructura jerárquica donde las variables individuales se agrupan en dimensiones homogéneas. En concreto, este trabajo organiza las variables en cuatro dimensiones principales: económica, estructural, intensidad de consumo y dinámica. Este enfoque, que se suele representar en forma de árbol o pirámide de indicadores, facilita la interpretación de los resultados y proporciona transparencia en el proceso metodológico. Otro elemento fundamental es la normalización de los datos. Las variables que componen un indicador compuesto generalmente se expresan en diferentes unidades, lo que dificulta su comparación directa. La literatura destaca varias técnicas de normalización, entre ellas la estandarización con puntuaciones Z y la normalización Min-Max. Esta última se usa a menudo en índices internacionales debido a su facilidad de interpretación, ya que transforma todas las variables a una escala común de entre 0 y 1. Esto permite combinar indicadores sin que las diferencias de escala afecten el resultado final. Siguiendo estas recomendaciones metodológicas, en este trabajo se empleará la normalización Min-Max para hacer comparables las variables seleccionadas antes de su agregación.[4]

La ponderación es otra cuestión muy debatida en la creación de indicadores compuestos. Los pesos determinan la importancia relativa de cada variable o dimensión dentro del índice final, lo que puede influir significativamente en los resultados. La literatura identifica tres enfoques principales. El primero consiste en asignar el mismo peso a todos los indicadores, basándose en la idea de que no hay suficiente evidencia para considerar

[4] OECD, *Handbook on Constructing Composite Indicators: Methodology and User Guide*, OECD Publishing, París, 2008, pp. 84-88.

unas variables más importantes que otras. El segundo enfoque se basa en criterios de expertos, donde los pesos reflejan una valoración conceptual del fenómeno. El tercer enfoque aplica métodos estadísticos para derivar los pesos a partir de los propios datos. Con el objetivo de evaluar la robustez de los resultados, este trabajo aplicará los tres enfoques y comparará posteriormente los rankings obtenidos.[5]

El Análisis de Componentes Principales (PCA) es uno de estos métodos. Esta técnica, que se usa ampliamente en Business Analytics y en la reducción de dimensionalidad, identifica patrones de variación en un conjunto de variables correlacionadas y sintetiza la información en un número reducido de componentes. Además de facilitar la comprensión de sistemas complejos, el PCA puede utilizarse para obtener pesos empíricos basados en la capacidad explicativa real de cada variable. Varios estudios sobre indicadores compuestos consideran el PCA una herramienta especialmente útil para complementar métodos tradicionales de ponderación y reducir el componente subjetivo en la asignación de pesos. Por ello, el PCA se incorpora en este trabajo como una alternativa a los esquemas de ponderación basados en pesos iguales y criterio experto.[6]

La etapa final del proceso es la agregación de los indicadores normalizados y ponderados. Los métodos más usados son la media aritmética y la media geométrica. La principal diferencia entre ambas es el grado de compensación permitido entre variables. La media aritmética permite una compensación total, de modo que valores altos en determinadas dimensiones pueden contrarrestar resultados bajos en otras. En cambio, la media geométrica penaliza más los desequilibrios y exige un desempeño más uniforme para alcanzar puntuaciones altas. La elección entre ambos métodos depende del fenómeno estudiado y es una de las decisiones metodológicas más importantes en la creación de cualquier indicador compuesto. En consecuencia, ambos métodos serán evaluados en este trabajo para analizar su impacto sobre el ranking autonómico resultante.[7]

---

[5] OECD, *Handbook on Constructing Composite Indicators: Methodology and User Guide*, OECD Publishing, París, 2008, pp. 93-104.

[6] OECD, *Handbook on Constructing Composite Indicators: Methodology and User Guide*, OECD Publishing, París, 2008, pp. 96-99; I. T. Jolliffe, *Principal Component Analysis*, 2nd ed., Springer, Nueva York, 2002, pp. 1-15.

[7] OECD, *Handbook on Constructing Composite Indicators: Methodology and User Guide*, OECD Publishing, París, 2008, pp. 118-125.

Por último, la literatura destaca la importancia de realizar un análisis de sensibilidad y robustez. Los resultados de un indicador compuesto pueden verse afectados por decisiones metodológicas como la selección de variables, los métodos de normalización, los esquemas de ponderación o las técnicas de agregación. Por ello, es necesario comprobar cuánto cambian las conclusiones obtenidas al modificar estas decisiones. Esta práctica se considera un estándar metodológico en la elaboración de indicadores compuestos y ayuda a reforzar la credibilidad y la validez de los resultados. Siguiendo esta recomendación, el trabajo incorpora un análisis de sensibilidad destinado a evaluar la estabilidad del Índice de Desarrollo del Juego frente a diferentes alternativas metodológicas.[8]

Así, este trabajo adopta el marco metodológico aceptado en la literatura sobre creación de indicadores compuestos para desarrollar un Índice de Desarrollo del Juego aplicable a las comunidades autónomas españolas. A partir de una selección de variables procedentes de fuentes oficiales, se construye una estructura multidimensional que integra aspectos económicos, estructurales, de intensidad de consumo y de dinámica del sector. La combinación de técnicas de normalización, ponderación, agregación y análisis de sensibilidad permite obtener una medida sintética del grado de desarrollo del sector y realizar comparaciones homogéneas entre territorios.

3. METODOLOGÍA

3.1 VARIABLES, PIRÁMIDE DE INDICADORES Y CONVERSIÓN DE ABSOLUTOS EN RATIOS

3.1.1 Planteamiento y unidad de análisis

El objetivo de este trabajo es construir un indicador compuesto que ordene a las comunidades autónomas según el grado de desarrollo de la realidad del juego de azar y apuestas (en las variables, abreviado a JdA&A, correspondiente al grupo 92 de la CNAE-2009). La unidad de análisis es la combinación *comunidad autónoma* × *año*, sobre un panel de las diecisiete comunidades autónomas. Inicialmente se contempló la inclusión

[8] Saisana, Michaela; Tarantola, Stefano, *State-of-the-Art Report on Current Methodologies and Practices for Composite Indicator Development*, Joint Research Centre, European Commission, Ispra, 2002, pp. 31-36

de las ciudades autónomas de Ceuta y Melilla, con el objetivo de ampliar el alcance comparativo del estudio. Sin embargo, la elevada ausencia de información y la falta de disponibilidad homogénea de datos para Ceuta y Melilla obligaron a excluirlas del análisis final.

El marco temporal común seleccionado comprende el período 2021-2024. La elección de este intervalo responde a criterios de disponibilidad, comparabilidad y consistencia de la información, ya que constituye el período más reciente para el que la mayoría de las variables analizadas presentan datos públicos accesibles, una cobertura territorial suficientemente completa y una metodología de recopilación relativamente homogénea entre comunidades autónomas. La adopción de un horizonte temporal común permite, además, garantizar la comparabilidad de los indicadores.

Las magnitudes de base proceden en su totalidad de la estadística estructural y de la contabilidad regional recogida en la base de datos del INE; tras la recopilación de estas magnitudes, mediante elaboración propia, se han calculado los ratios empleados para el estudio, que se describen a continuación.

Esta ha sido una decisión metodológica importante dentro del apartado, y que condiciona todo estudio: ninguna variable entra en el índice en términos absolutos. Una comunidad como Madrid o Cataluña tiene, por pura escala demográfica y económica, más locales, más empresas, más empleo y más euros jugados que La Rioja o Cantabria. Si compusiéramos el índice con valores brutos, no estaríamos midiendo *desarrollo del juego* sino *tamaño de la región*, y el ranking sería poco más que una reordenación del tamaño poblacional. Por eso toda variable absoluta se transforma en un ratio que neutraliza el efecto tamaño y la convierte en una medida de intensidad comparable entre territorios. La elección del denominador (población, PIB, total de ocupados, total de empresas, stock del año anterior) no es neutra: cada denominador responde a una pregunta distinta y se justifica según cada caso.

Además, conviene hacer la distinción desde el principio entre la orientación analítica de cada variable, es decir, si un valor alto indica más desarrollo del juego, y su lectura social, esto es, si más juego resulta deseable o no. El índice queremos que mida el desarrollo del sector, de modo que la orientación relevante para la agregación es la primera; la valoración social se reserva para la interpretación de resultados y para el apartado de

limitaciones. Esta distinción resulta especialmente importante en la dimensión de dinámica: la mortalidad empresarial se interpreta en sentido inverso, mientras que la variación neta de locales mantiene orientación directa, ya que un aumento del stock de locales indica expansión de la infraestructura física del sector.

3.1.2 Cuatro dimensiones

Las variables se agrupan en cuatro dimensiones que responden a preguntas complementarias sobre el fenómeno:

La dimensión económica responde a cuánto valor económico mueve el juego en la región: facturación, retribución del trabajo e inversión. Captura el peso monetario y de demanda del sector, es decir, su huella económica directa. Es la dimensión más cercana al concepto intuitivo de "mercado del juego" y, por su naturaleza monetaria, la que mejor se presta a la ratificación contra el PIB o la población.

La dimensión estructural responde a qué infraestructura y qué tejido empresarial sostienen el juego: cuántos locales, cuántas empresas y cuánto empleo especializado existen por habitante. Mientras la económica mide flujos (euros que circulan), la estructural mide el *stock* sobre el que esos flujos se apoyan. La separación de ambas evita confundir un mercado que mueve mucho dinero con pocos operadores grandes (alta económica, baja estructural) de otro muy dispersado en locales pequeños (alta estructural, económica más modesta): son dos formas distintas de desarrollo que el índice debe poder discriminar.

La dimensión de dinámica responde a en qué dirección se mueve el sector: si se expande o se contrae, y a qué ritmo. Las dos dimensiones anteriores son fotografías estáticas de un año; esta introduce la derivada temporal. Su pertinencia es doble: por un lado, un sector en crecimiento acelerado representa un grado de desarrollo distinto al de uno maduro y estancado, aunque ambos tengan hoy el mismo nivel; por otro, esta dimensión es la que materializa el dinamismo que el planteamiento del trabajo exige y la que justifica el cálculo posterior de tasas de variación.

La dimensión de intensidad de consumo responde a cuánto utilizan los ciudadanos los productos de juego más extendidos como son las apuestas y loterías. Mientras que la dimensión económica mide el valor monetario generado por el sector y la estructural su

capacidad instalada, esta dimensión se centra en el comportamiento de la demanda. A través de las cantidades jugadas en apuestas y loterías, ambas relativizadas por población, captura la frecuencia e intensidad con la que los residentes participan en las distintas modalidades de juego. Su inclusión resulta especialmente relevante porque permite distinguir entre comunidades donde existe una amplia infraestructura o un elevado volumen de negocio y aquellas donde el consumo efectivo por habitante es realmente elevado. De este modo, incorpora una perspectiva de participación ciudadana que complementa las restantes dimensiones y contribuye a una visión más completa del grado de desarrollo del sector.

3.1.3 Variables de contexto y denominadores

Cuatro magnitudes no entran directamente en el índice, pero son imprescindibles como denominadores para convertir las variables en ratios comparables homogeneizados.

(i) Población (personas).

Es el denominador *per cápita* por excelencia. Se utiliza para todas las variables cuya intensidad tiene sentido hacer el ratio por habitante. La razón es que el juego es un fenómeno de consumo de los residentes, de modo que normalizar por población aproxima la exposición media del ciudadano al sector. Su limitación es que ignora flujos turísticos: en comunidades con mucho juego de visitantes, como Baleares o Canarias, los ratios per cápita pueden infravalorar el tamaño real del mercado. Lo señalamos como sesgo conocido.

(ii) PIB regional (miles de €).

Es el denominador para variables de naturaleza monetaria y agregada, en particular la cifra de negocios. Dividir por el PIB responde a una pregunta distinta que hacerlo per cápita alejándonos de la cuestión sobre cuánto juego hay por persona para acercarnos a entender qué peso tiene el juego dentro de la economía regional. Es la ratificación estándar para comparar el tamaño económico de un sector entre territorios de muy distinta dimensión.

(iii) PIB per cápita (miles de €/persona).

No se usa como denominador sino como variable de contexto y posible control: aproxima la renta disponible regional, que la literatura vincula a los patrones de gasto en juego. Su utilidad principal aquí es interpretativa (ayuda a leer si el desarrollo del juego va de la mano de la riqueza o no) y, eventualmente, como variable a contrastar en correlaciones.

(iv) Total de ocupados (personas).

Resulta necesaria para la construcción de una de las métricas de intensidad del sector. En concreto, se utiliza como denominador para calcular la proporción de personal ocupado en las actividades de juegos de azar y apuestas sobre el empleo total de cada comunidad autónoma. El objetivo de esta ratio es medir el peso relativo del sector dentro del mercado laboral regional, permitiendo comparar territorios de distinto tamaño económico y demográfico. De este modo, no se evalúa únicamente el número absoluto de trabajadores vinculados al juego, sino la relevancia que dicha actividad tiene dentro de la estructura de empleo de cada comunidad autónoma. Como limitación, debe señalarse que esta medida no capta diferencias en productividad ni en la calidad del empleo generado.

3.1.4 Dimensión económica

Las siguientes variables han sido escogidas para formar la dimensión económica

(i) Cifra de negocios del sector / PIB regional *(ratio: € de facturación por € de PIB).*

Mide el peso de la facturación del juego dentro de la economía regional y es, probablemente, la variable más directa de «tamaño económico del sector». Aporta la magnitud central de actividad de mercado. Se normaliza respecto el PIB y no per cápita porque la cifra de negocios es un agregado de empresa, no un consumo individual: lo relevante no es cuánto factura el sector por habitante sino qué fracción de la economía regional representa. Entra en la dimensión económica por definición, al ser un flujo monetario de actividad. Orientación: a mayor valor, mayor desarrollo.

(ii) Sueldos y salarios del sector / población *(ratio: miles de € de masa salarial por habitante).*

Mide la retribución del trabajo que el sector genera por habitante. Aporta la faceta laboral (cuánto paga el sector) a la dimensión económica, complementaria a la variable sobre

empleo contemplada en la dimensión estructural. Se normaliza per cápita para hacerla comparable como intensidad regional. Entra en la dimensión económica porque mide un flujo monetario (rentas salariales), no un stock de empleo. Orientación: a mayor valor, mayor desarrollo. Su correlación con el empleo del sector deberá vigilarse en el paso de correlaciones. No obstante, no existe necesariamente una relación mecánica entre ambas, ya que la masa salarial del sector se relativiza mediante la población total de la comunidad autónoma, reflejando la renta laboral generada por habitante, y la variable de empleo se expresa como proporción del empleo total regional, captando el peso relativo del sector dentro de la estructura laboral de cada territorio. Al utilizar denominadores distintos, ambas variables miden dimensiones complementarias del fenómeno y pueden aportar información diferenciada sobre el grado de desarrollo del sector.

(iii) Inversión en activos materiales / población *(ratio: miles de € de inversión por habitante).*

Mide el esfuerzo inversor del sector (maquinaria, locales, equipamiento) por habitante. Aporta una señal de profundidad y de apuesta de capital, bajo la premisa de que un sector que invierte está consolidándose, no solo facturando. Se normaliza per cápita por la misma lógica de intensidad regional. Entra en la económica como flujo de capital. Orientación: a mayor valor, mayor desarrollo. Es una variable potencialmente volátil año a año (las inversiones son lumpy), lo que conviene tener presente de cara a la imputación.

3.1.5 Dimensión intensidad de consumo

(i) Cantidades jugadas en apuestas / población *(ratio: millones de € apostados por habitante).*

Mide el volumen de apuestas (deportivas y similares) por habitante, es decir, la intensidad de consumo como indica el nombre de la dimensión. Aporta el lado de la demanda: no lo que el sector factura o invierte, sino cuánto juega efectivamente la población. Per cápita es el denominador natural por ser consumo individual. Orientación: a mayor valor, mayor desarrollo del mercado.

(ii) Cantidades jugadas en loterías / población *(ratio: millones de € jugados por habitante).*

Análoga a la anterior, pero enfocada en loterías, modalidad de perfil distinto, más extendida, de menor riesgo individual y con fuerte componente público. Aporta la intensidad de consumo de un segmento aún más grande que el de apuestas. Mantenerla separada de las apuestas tiene valor analítico: permite distinguir comunidades donde el juego desarrollado es sobre todo lotería (consumo amplio y de bajo riesgo) de aquellas donde pesan más las apuestas. Se normaliza per cápita. Entra en la dimensión de intensidad de consumo al contemplar el volumen de demanda. Orientación: a mayor valor, mayor desarrollo.

3.1.6 Dimensión estructural

(i) Número de locales de JdA&A / población *(ratio: locales por habitante).*

Mide la densidad física de puntos de juego en el territorio. Aporta la variable de infraestructura más tangible: la presencia del juego en la calle. Per cápita captura la exposición física del ciudadano. Entra en la estructural por ser stock de infraestructura, no flujo monetario. Orientación: a mayor densidad, mayor desarrollo. Es también la variable con mayor carga interpretativa desde el punto de vista social, por su relación con la accesibilidad y el juego problemático.

(ii) Personal ocupado en JdA&A / total de ocupados *(ratio: proporción del empleo regional dedicada al sector).*

Mide qué fracción del empleo de la comunidad depende del juego. Aporta la magnitud laboral *en volumen* y, al normalizar por el total de ocupados en lugar de por población, mide específicamente la *especialización productiva* de la región en el sector. Entra en la dimensión estructural como componente del tejido productivo. Orientación: a mayor especialización, mayor desarrollo. Previsiblemente correlacionará con el número de empresas y con la masa salarial, lo que se evaluará antes de componer el índice.

(iii) Número de empresas de JdA&A / población *(ratio: empresas por habitante).*

Mide la densidad del tejido empresarial del sector. Aporta una visión del stock de operadores complementaria a la de locales: muchas empresas pequeñas frente a pocas grandes describen estructuras de mercado distintas. Per cápita la hace comparable como intensidad. Entra en la estructural. Orientación: a mayor densidad empresarial, mayor

desarrollo. Su solapamiento esperable con locales y empleo es el principal candidato a redundancia que el análisis de correlaciones deberá depurar.

3.1.7 Dimensión de dinámica

(i) Mortalidad empresarial: muertes de empresas de JdA&A / nº de empresas del sector *(ratio: tasa de desaparición de empresas).*

Mide la rotación destructiva del tejido empresarial. Aporta una señal de salud y estabilidad: una tasa alta indica contracción o inestabilidad del sector. Se normaliza sobre el stock de empresas del propio sector, no sobre población, porque mide una tasa *interna* de mortalidad. Entra en la dimensión de dinámica por ser una variable de cambio. Orientación analítica: una mortalidad alta apunta a un menor desarrollo/consolidación, de modo que entra con signo invertido en el índice (se reorientará en la normalización). Es uno de los casos donde la lectura de desarrollo y la lectura social divergen, y conviene explicitarlo.

(ii) Variación anual del número de locales: *(ratio: tasa de variación del stock de locales)*.

Mide la variación neta anual de la infraestructura física del sector. Esta variable no recoge únicamente los cierres de locales, sino el cambio neto del stock de establecimientos entre un año y el anterior. Por tanto, captura el resultado agregado de las altas y bajas producidas en cada ejercicio. Se calcula como la diferencia entre el número de locales del año actual y el número de locales del año anterior, dividida entre el stock de locales del año anterior, siguiendo el criterio habitual en el cálculo de tasas de variación. Una variación positiva indica crecimiento neto de la infraestructura y, por tanto, mayor dinamismo del sector; una variación negativa refleja contracción neta. Así pues, la orientación de la variable es directa y no requiere reorientación durante el proceso de normalización.

(iii) CAGR de las cantidades totales jugadas (2011 → año actual) *(tasa de crecimiento anual compuesta, largo plazo).*

Mide la tendencia estructural del volumen jugado a lo largo de más de una década. Aporta la trayectoria de fondo del mercado, filtrando el ruido de los años puntuales. No requiere

denominador adicional porque ya es una tasa. Entra en la dinámica como componente de largo plazo. Orientación: crecimiento alto implica sector en expansión y mayor desarrollo dinámico.

(iv) Variación anual de las cantidades totales jugadas (año anterior → año actual)

(tasa de variación interanual, corto plazo). Mide el impulso reciente del mercado. Aporta la lectura de coyuntura que el CAGR no capta, especialmente relevante en este marco temporal por el rebote del juego tras la pandemia. Es una tasa, sin denominador adicional. Entra en la dimensión de dinámica como componente de corto plazo (c.p.). Orientación: variación positiva implica expansión reciente. Combinar largo y corto plazo en la misma dimensión permite distinguir crecimiento sostenido de aceleraciones o frenazos puntuales, pero su correlación deberá vigilarse para no duplicar la misma señal de crecimiento.

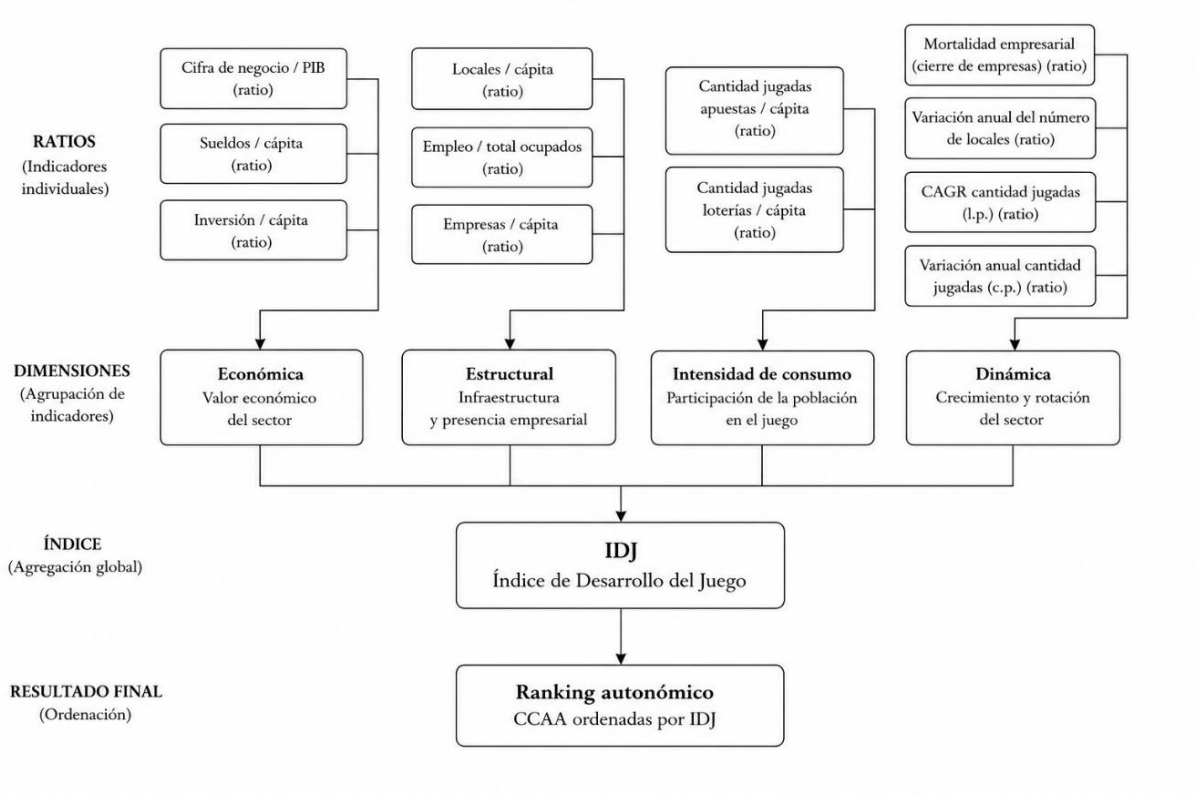
3.1.8 Los indicadores legislativos

Se ha recogido, además, un bloque de variables normativas (años desde la última modificación de la ley, número de modificaciones, año de la ley marco vigente, distancias mínimas de los locales a centros de enseñanza y a otros locales, tipo impositivo de las apuestas y tasa de la máquina B). Estas no se integran en el Índice de Desarrollo del Juego, ya que miden madurez y restricción regulatoria, no desarrollo del mercado, y constituyen la materia prima del indicador legal que el objetivo secundario contempla. Se documentan aquí y se incluirá alguna mención en el anexo, pero quedan fuera de la pirámide que sigue.

3.1.9 Pirámide o árbol de indicadores

La estructura del índice sigue el esquema clásico de cuatro niveles. En el Nivel 1 están las doce variables ya convertidas en ratios. En el Nivel 2 esas variables se agrupan en las cuatro dimensiones descritas. En el Nivel 3 las dimensiones se sintetizan, mediante normalización, ponderación y agregación, en un único IDJ. El Nivel 4 es el resultado final: el ranking de las comunidades autónomas ordenadas por su IDJ.

Esquema 3.1.9: Árbol con estructura jerárquica para la generación de indicadores compuestos



Una vez cerrado el inventario, definidas las dimensiones y fijados los ratios, las siguientes etapas consisten en completar e imputar los datos faltantes, construir los indicadores dinámicos, analizar posibles redundancias mediante correlaciones y normalizar las variables para hacerlas comparables. Posteriormente, se determinarán los pesos y el método de agregación más adecuados para obtener el Índice de Desarrollo del Juego (IDJ), finalizando con un análisis de sensibilidad que permita evaluar la robustez del ranking autonómico resultante frente a distintas decisiones metodológicas.

3.2 PROCESAMIENTO DE LOS DATOS

3.2.1 Imputación de valores faltantes

A diferencia de lo que ocurre en muchos ejercicios de construcción de indicadores compuestos, las bases de datos empleadas en este trabajo no presentan prácticamente valores faltantes en las variables finalmente seleccionadas. Las fuentes oficiales utilizadas, por lo general, proporcionan información precisa para todas las comunidades autónomas y para todo el período temporal analizado, lo que permite disponer de un panel completo y homogéneo.

Durante la fase exploratoria se identificaron algunas limitaciones de disponibilidad para determinados territorios (el caso de Ceuta y Melilla) y series estadísticas (años previos a 2021). No obstante, finalmente no fueron incorporados al análisis, por lo que dichas incidencias no implican tener que imputar valores ni afectan a la construcción del índice o al ranking autonómico resultante.

El único caso relevante de ausencia de datos se ha encontrado en la información sobre las muertes de empresas del sector del juego correspondientes al año 2024. En el momento de elaboración del trabajo, la última cifra disponible era la de 2023. Como consecuencia, la variable de mortalidad empresarial, definida como el cociente entre las empresas desaparecidas y el total de empresas del sector, quedaba sin valor para las diecisiete comunidades autónomas en 2024, mientras que el resto de variables del índice presentaban información completa para todo el período. Dado que el año de referencia para el cálculo del Índice de Desarrollo del Juego es precisamente 2024, resultaba imprescindible completar este único hueco mediante una técnica de imputación.

Se valoraron tres alternativas. La interpolación lineal se descartó porque está pensada para estimar valores situados entre dos observaciones conocidas, y en este caso 2024 constituye el extremo de la serie; aplicarla supondría en realidad una extrapolación, un uso para el que esta técnica no resulta adecuada. La regresión también se rechazó, ya que con solo tres observaciones por comunidad (2021, 2022 y 2023), una regresión sobre el tiempo equivaldría a prolongar una recta a partir de muy pocos puntos, sobreajustando un historial escaso. Plantear una regresión sobre el resto de datos de la propia variable también se desestimó por el riesgo de circularidad, al predecir una componente del indicador a partir de otras componentes del mismo, lo que restaría transparencia y replicabilidad al procedimiento.

Finalmente se optó por una media móvil. Esta opción es la más robusta frente a la volatilidad propia de una tasa de mortalidad empresarial, que depende de eventos discretos y puede oscilar de forma notable de un año a otro sin responder a una tendencia estable. La media móvil no asume ninguna dirección de cambio, mantiene el valor imputado dentro del rango efectivamente observado en cada territorio y exige supuestos mínimos, lo que la hace fácil de justificar y de reproducir. Implícitamente asume que el nivel de mortalidad de 2024 se aproxima al nivel típico reciente de cada comunidad, una

hipótesis razonable para una tasa de carácter estructural en un horizonte temporal de un solo año.

El procedimiento de imputación fue el siguiente. Para cada comunidad autónoma se calculó la media de su tasa de mortalidad empresarial en los años con dato disponible (2021 a 2023) y ese promedio se asignó como valor estimado de 2024. De este modo cada territorio recibe una imputación construida con su propia trayectoria, respetando las diferencias regionales que el índice pretende capturar. Cada valor imputado se marcó mediante una variable indicadora, con el fin de mantener la trazabilidad de las observaciones estimadas y poder evaluar después su influencia sobre los resultados. Conviene señalar, como limitación, que la imputación no captura un eventual repunte o caída atípica de la mortalidad en 2024; se trata de una estimación conservadora cuyo efecto sobre el ranking final se contrastará en el análisis de sensibilidad.

3.2.2 Cálculo de tasas de variación para la dimensión de dinámica

Las variables que integran la dimensión de dinámica no proceden directamente de las fuentes originales, sino que han sido elaboradas a partir de las series temporales recopiladas. Su objetivo es capturar la evolución reciente y la trayectoria de largo plazo del sector, incorporando una dimensión temporal que no está presente en las variables económicas, estructurales o de intensidad de consumo.

En concreto, se han calculado cuatro ratios:

a) La mortalidad empresarial. Se obtiene dividiendo el número de empresas desaparecidas en un año entre el total de empresas existentes en el sector.
b) La variación neta anual del número de locales. Se obtiene restando al número de locales del año actual el número de locales del año anterior, y dividiendo dicha diferencia entre el stock de locales del año anterior. De este modo, la variable refleja la tasa de crecimiento o contracción neta de la infraestructura física del sector

Para medir el crecimiento de la actividad del juego se utilizan dos aproximaciones complementarias.

c) Por un lado, la tasa de crecimiento anual compuesta (CAGR) de las cantidades jugadas entre 2011 y el año de referencia, que captura la tendencia estructural de largo plazo.
d) Por otro, la variación interanual de las cantidades jugadas, que refleja la evolución más reciente del mercado.

Una vez calculadas, estas variables pasan a formar parte de la base de datos final y son tratadas de la misma forma que el resto de los indicadores en las fases posteriores de análisis de correlaciones, normalización, ponderación y agregación.

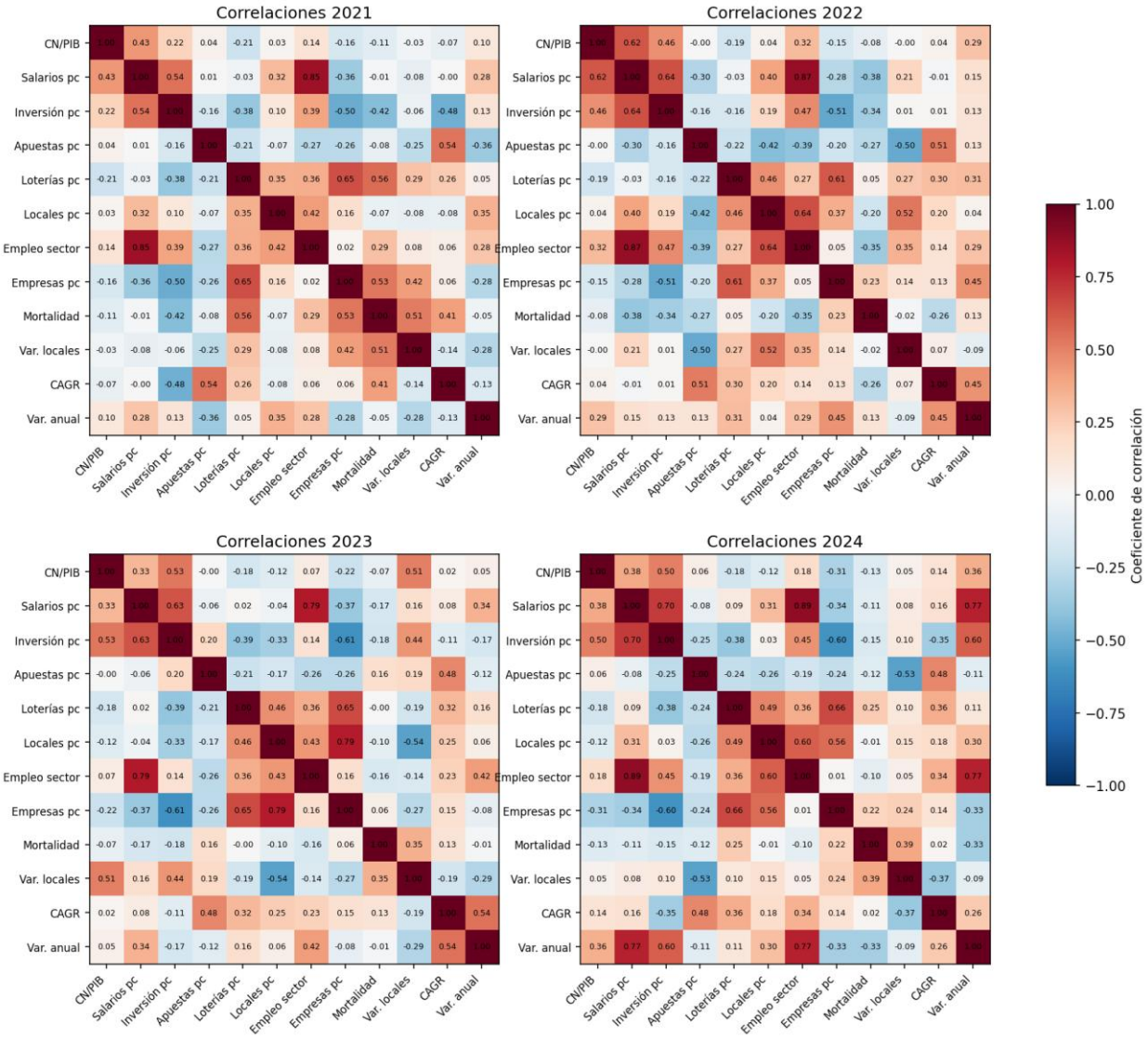
3.3 CÁLCULO DE CORRELACIONES Y ELIMINACIÓN DE VARIABLES REDUNDANTES

El paso previo antes de normalizar y ponderar las variables consiste en comprobar que ninguna de ellas duplica la información de otra. Cuando dos ratios están muy correlacionados, ambos empujan al índice en la misma dirección y, de hecho, ponderan por partida doble una misma faceta del fenómeno, sesgando el resultado hacia aquello que se mide repetido. Para evitarlo se calculó la matriz de correlaciones de Pearson entre los doce ratios, considerando redundante toda pareja cuyo coeficiente superase, en valor absoluto, el umbral de 0,7.

El análisis se realizó por separado para cada uno de los cuatro años del panel (2021 a 2024), en lugar de sobre los datos agrupados. Esta decisión responde al reducido tamaño muestral, diecisiete comunidades por año, que hace que correlaciones elevadas puedan surgir por azar en un ejercicio concreto. Calcular una matriz por año y exigir que la correlación se mantenga alta en la mayoría de ellos permite distinguir las redundancias estructurales, estables en el tiempo, de las puramente coyunturales.

Como control adicional de robustez se replicó el cálculo con el coeficiente de Spearman, menos sensible a valores extremos y a relaciones no lineales, para descartar que las correlaciones detectadas fueran un artificio del método de Pearson. El resultado es nítido: de las múltiples parejas posibles, una sola supera el umbral de forma sistemática. Las matrices de correlación se pueden observar a continuación.

Figura 3.3: *Matrices de correlación por año*



3.3.1 Correlaciones encontradas

Los salarios y sueldos per cápita y el peso del empleo del sector en el empleo regional presentan una correlación de Pearson de entre 0,79 y 0,89 en los cuatro años, confirmada además por Spearman (entre 0,71 y 0,90). Se trata, por tanto, de una redundancia robusta y no de un episodio aislado. Conviene señalar que esta era precisamente la relación que se había etiquetado como candidata a correlacionarse en la definición de las variables: ambas miden, desde ángulos distintos, la huella laboral del sector, y los datos confirman que esa información es, en la práctica, una sola.

El resto de los solapamientos que cabía esperar a priori no se materializaron. La densidad de locales y la de empresas solo superan el umbral en 2023 (0,79), permaneciendo en valores moderados el resto de años (entre 0,16 y 0,56), por lo que no constituyen una redundancia estable. Las dos variables de crecimiento del volumen jugado (la tasa compuesta de largo plazo y la variación interanual) en ningún momento alcanzan el umbral, lo que valida mantener separadas la tendencia de fondo y el impulso reciente. Y las cantidades jugadas en apuestas y en loterías presentan incluso una correlación ligeramente negativa (en torno a −0,22), señal de que responden a perfiles de consumo distintos: existen comunidades orientadas a la lotería y otras a la apuesta, y conviene que el índice las distinga. Únicamente en 2024 aparecieron dos correlaciones altas adicionales con la variación interanual, pero al no repetirse en ningún otro año se descartaron por coyunturales.

3.3.2 Medidas tomadas

La consecuencia práctica es la eliminación de una variable. De la pareja redundante se optó por descartar el empleo del sector y conservar los salarios per cápita, por dos motivos. El primero, empírico: el empleo del sector es la variable que más correlaciona, en promedio, con el conjunto del resto de indicadores (0,33 frente a 0,31), de modo que es la que más información comparte y la que menos pierde el índice al retirarse. El segundo, conceptual: la huella laboral del sector queda igualmente recogida por los salarios per cápita, mientras que la dimensión estructural conserva así un carácter homogéneo de stock (densidad de locales y de empresas), sin mezclar en ella una magnitud de naturaleza laboral.

Esta decisión obliga a matizar la justificación inicial, que presentaba ambas variables como complementarias por emplear denominadores distintos: el contraste empírico muestra que, pese a ello, capturan la misma señal.

Tras la depuración, el Índice de Desarrollo del Juego queda construido sobre once variables, repartidas en las cuatro dimensiones: tres económicas (cifra de negocios sobre PIB, salarios per cápita e inversión per cápita), dos de intensidad de consumo (apuestas y loterías per cápita), dos estructurales (locales y empresas per cápita) y cuatro de dinámica (mortalidad empresarial, variación de locales, CAGR y variación interanual del volumen

jugado). Sobre este conjunto depurado, y tomando 2024 como año de referencia, se procede en las etapas siguientes a la normalización de las variables, a la determinación de los pesos mediante tres esquemas alternativos (iguales, de experto y por componentes principales) y a su agregación, cerrando el proceso con el análisis de sensibilidad del ranking resultante.

3.4 NORMALIZACIÓN

Tras la depuración de variables redundantes, el siguiente paso consiste en normalizar los once indicadores que integran el Índice de Desarrollo del Juego (IDJ). La normalización resulta necesaria porque las variables se expresan en unidades y escalas muy diferentes: algunas son ratios económicos, otras representan densidades empresariales y otras son tasas de crecimiento o de mortalidad. Sin una transformación previa, las variables con valores numéricamente más elevados tendrían una influencia desproporcionada sobre el resultado final del índice.

Para evitar este problema se emplea la normalización Min-Max. Se ha escogido esta técnica de normalización por encima de otras como Z-score o transformaciones basadas en percentiles debido a su fácil interpretabilidad y a que constituye una de las metodologías más utilizadas en la elaboración de indicadores compuestos. Este procedimiento transforma cada variable a una escala común comprendida entre 0 y 1, donde 0 representa el peor resultado observado entre las comunidades autónomas y 1 el mejor. La transformación se realiza mediante la expresión:

$$X^* = \frac{X - X_{min}}{X_{max} - X_{min}}$$

De este modo, todas las variables pasan a ser directamente comparables y agregables en el índice, sin alterar las diferencias relativas existentes entre ellas.

La normalización se realiza de forma independiente para cada año del panel. En consecuencia, para cada uno de los ejercicios comprendidos entre 2021 y 2024, las once variables seleccionadas son reescaladas utilizando los valores mínimo y máximo observados entre las comunidades autónomas en ese mismo año. De este modo, cada indicador adopta siempre valores comprendidos entre 0 y 1 dentro de su correspondiente

ejercicio, donde 0 representa la comunidad autónoma con el peor resultado relativo y 1 la que presenta el mejor desempeño en esa variable

En aquellas variables cuya orientación es inversa al concepto de desarrollo (concretamente la mortalidad empresarial) la escala se reorienta tras la normalización para que valores más elevados representen siempre un mayor nivel de desarrollo. Con ello se garantiza que todas las variables contribuyan al índice en la misma dirección interpretativa.

La normalización se realiza una vez completadas las fases de imputación, construcción de indicadores dinámicos y eliminación de redundancias, aplicándose sobre las once variables finalmente seleccionadas para la construcción del IDJ. Cabe mencionar que la correlación de Pearson es invariante ante cambios de escala, por lo que la normalización se ha realizado en un paso posterior al del cálculo de correlaciones para aplicarse únicamente a las variables que finalmente serán parte del indicador.

3.5 PONDERACIÓN

Una vez normalizadas las variables, el siguiente paso consiste en asignar pesos a cada una de ellas. La ponderación determina la contribución relativa de cada indicador al resultado final del (IDJ). En ausencia de ponderación, todas las variables tendrían la misma influencia sobre el índice, independientemente de su relevancia conceptual o de la información estadística que contienen. La asignación de pesos constituye, por tanto, una de las decisiones metodológicas más sensibles en la construcción de indicadores compuestos.

Con el fin de minimizar la arbitrariedad y evaluar la robustez de los resultados, se han construido tres esquemas alternativos de ponderación: pesos iguales, pesos de experto y pesos obtenidos mediante Análisis de Componentes Principales (PCA).

A continuación, se explica cada uno de los métodos de ponderación contemplados. Consultar el anexo para ver de forma conjunta la ponderación otorgada por cada método.

3.5.1 Ponderación mediante pesos iguales

El primer esquema asigna el mismo peso a cada una de las cuatro dimensiones del índice. De este modo, las dimensiones económica, estructural, intensidad de consumo y dinámica reciben cada una un peso del 25% sobre el resultado final. Una vez fijado el peso de cada dimensión, este se distribuye de forma equitativa entre las variables que la integran. Así, las tres variables de la dimensión económica reciben un peso individual del 8,33%; las dos variables de intensidad de consumo reciben un 12,5% cada una; las dos variables estructurales reciben igualmente un 12,5%; y las cuatro variables dinámicas reciben un 6,25% cada una. Este procedimiento tiene la ventaja de ser completamente transparente y reproducible, además de evitar juicios subjetivos sobre la importancia relativa de cada dimensión. Por ello constituye el punto de partida habitual en la literatura sobre indicadores compuestos cuando no existe un consenso claro sobre la superioridad de unas variables frente a otras.

3.5.2 Ponderación mediante criterio experto

El segundo esquema incorpora una valoración conceptual sobre la importancia relativa de las distintas dimensiones del desarrollo del juego. Bajo esta aproximación se considera que el peso económico del sector constituye la manifestación más directa de su nivel de desarrollo, ya que refleja la capacidad de generar actividad empresarial, rentas laborales e inversión. Por este motivo, la dimensión económica recibe un peso del 40% del índice total, distribuido entre la cifra de negocios sobre PIB (20%), los salarios per cápita (10%) y la inversión per cápita (10%). La dimensión de intensidad de consumo recibe un peso del 25%, repartido equitativamente entre apuestas y loterías. Del mismo modo, la dimensión estructural recibe un 25%, dividido entre densidad de locales y densidad empresarial. Por último, la dimensión dinámica recibe un peso conjunto del 10%, distribuido a partes iguales entre sus cuatro variables. La razón de esta menor ponderación en la dimensión dinámica es que sus variables capturan cambios de corto y medio plazo, mientras que el objetivo principal del índice es medir el grado de desarrollo alcanzado por el sector y no exclusivamente su ritmo reciente de crecimiento o contracción. Se considera, por tanto, que la dinámica complementa la medición del desarrollo, pero no debe determinarla en la misma medida que las variables económicas, estructurales o de consumo.

Este esquema refleja la hipótesis de que un sector desarrollado es, ante todo, un sector que genera actividad económica significativa, incluso aunque en un momento concreto experimente fluctuaciones temporales en sus tasas de crecimiento.

3.5.3 Ponderación mediante Análisis de Componentes Principales (PCA)

El tercer esquema utiliza una aproximación estadística basada en el Análisis de Componentes Principales (Principal Component Analysis, PCA). A diferencia de los dos métodos anteriores, los pesos no son fijados por el investigador, sino que se derivan de la estructura interna de los datos.

El PCA identifica combinaciones lineales de variables capaces de explicar la máxima variabilidad observada entre comunidades autónomas. A partir de la varianza que explica cada componente y de las cargas factoriales relacionadas con las variables, se pueden obtener un conjunto de pesos empíricos. Estos pesos reflejan la contribución estadística de cada variable a la variabilidad observada entre comunidades.

La aplicación del método retuvo cuatro componentes principales con autovalores superiores a la unidad, siguiendo el criterio de Kaiser. En conjunto, estos cuatro componentes explican el 70,6% de la varianza total del sistema de indicadores, lo que sugiere que capturan una parte sustancial de la información contenida en los datos originales.

Los pesos obtenidos mediante PCA presentan una distribución relativamente equilibrada, aunque otorgan una mayor importancia a variables como la inversión per cápita, la densidad empresarial y el crecimiento compuesto de las cantidades jugadas. Esto indica que dichos indicadores son especialmente relevantes para explicar las diferencias observadas entre comunidades autónomas. Por el contrario, variables como la cifra de negocios sobre PIB o la variación interanual de las cantidades jugadas reciben una ponderación menor, al aportar menos capacidad discriminatoria dentro de la muestra analizada.

3.5.4 Conclusiones sobre ponderación

Ninguno de los tres esquemas puede considerarse objetivamente superior en todos los contextos. Los pesos iguales maximizan la transparencia, los pesos de experto incorporan una interpretación conceptual del fenómeno y los pesos PCA aprovechan la información estadística contenida en los datos. Por este motivo, los tres esquemas se utilizarán en el análisis de sensibilidad posterior, permitiendo evaluar hasta qué punto las conclusiones del estudio dependen de la elección de los pesos y comprobar la robustez del ranking autonómico obtenido.

3.6 AGREGACIÓN

Una vez normalizadas las variables y fijados los pesos, la agregación es el paso que las combina en un único valor sintético, el IDJ, para cada comunidad y año. Dado que la elección del método de agregación no es neutral porque condiciona cómo se compensan entre sí las distintas variables, se plantearon dos alternativas, la media aritmética ponderada y la media geométrica ponderada, y se calcularon ambas para los tres esquemas de pesos definidos en el apartado anterior. De este modo se obtienen seis versiones del índice por comunidad y año, lo que permite valorar después hasta qué punto los resultados dependen de las decisiones metodológicas adoptadas.

La media aritmética ponderada consiste en multiplicar cada variable normalizada por su peso y sumar los productos. Es el método más intuitivo y se caracteriza por ser plenamente compensatorio: un valor bajo en una variable puede quedar enteramente contrarrestado por un valor alto en otra, ya que solo importa la suma final. Es decir, trata las variables como sustituibles entre sí y premia el nivel global de desarrollo con independencia de cómo se reparta entre dimensiones.

La media geométrica ponderada multiplica las variables elevadas cada una a su peso, lo que equivale a calcular la exponencial de la suma ponderada de sus logaritmos. Su comportamiento es distinto: es solo parcialmente compensatoria, de manera que un valor bajo en cualquier variable arrastra el índice con más fuerza y no puede compensarse del todo con un valor alto en otra. En la práctica, penaliza los perfiles desequilibrados y premia el desarrollo equilibrado entre las distintas facetas del sector; una comunidad que destaque mucho en una variable, pero flojee en otra, obtendrá un índice geométrico más bajo que el aritmético.

En el proceso se tomaron dos decisiones relevantes. La primera afecta a la media geométrica: como la normalización min-max deja en cero al peor valor de cada variable en cada año, y la media geométrica se anula en presencia de un solo cero, se desplazó la escala de las variables al intervalo [0,01; 1] únicamente para este método, evitando así los ceros sin alterar el orden de las observaciones; la media aritmética se calculó sobre los valores originales en [0, 1]. La segunda es que, al haberse normalizado las variables dentro de cada año, la comparación entre años se realiza sobre la posición relativa de cada comunidad (su ranking) y no sobre el nivel absoluto del índice, que no es directamente comparable de un año a otro.

Los resultados de esta fase, que se analizan en detalle en el apartado correspondiente, muestran que el índice es notablemente robusto. Las seis versiones producen ordenaciones muy parecidas entre sí, las correlaciones de rango de Spearman entre ellas oscilan entre 0,91 y 0,99, de modo que la posición de cada comunidad no parece depender demasiado del esquema de pesos o del método de agregación elegido. La ordenación también se mantiene estable a lo largo del tiempo, con correlaciones entre los rankings de años distintos de entre 0,81 y 0,92. A título ilustrativo, Cataluña ocupa la última posición en los cuatro años, mientras que el liderazgo se reparte entre la Comunidad de Madrid en los primeros ejercicios y Canarias en 2024. La diferencia entre métodos se aprecia bien en los desequilibrados: por ejemplo el Principado de Asturias, con un perfil irregular, cae bastante más con la media geométrica (de 0,43 a 0,23 con pesos de experto) que comunidades de perfil más homogéneo. Eso es justo lo que la geométrica está diseñada para captar.

3.7 ANÁLISIS DE SENSIBILIDAD

La construcción de un indicador compuesto exige tomar decisiones metodológicas que no son únicas ni neutrales, como la manera de ponderar las variables o el método empleado para agregarlas. El análisis de sensibilidad tiene por objeto comprobar, de forma cuantitativa y no a simple vista, en qué medida la ordenación de las comunidades depende de esas decisiones. Si vemos que el ranking se mantiene estable al cambiar de método, puede afirmarse con mayor certeza que refleja el fenómeno y no la elección concreta del analista.

El material de partida son las seis versiones del índice que resultan de combinar los tres esquemas de pesos (iguales, experto y componentes principales) con los dos métodos de agregación (media aritmética y media geométrica). El análisis se aplica a los cuatro años del panel (2021-2024) para verificar que la robustez no es un rasgo de un ejercicio puntual, sino que se sostiene también en el tiempo.

3.7.1 Técnicas empleadas

Se han empleado cuatro técnicas complementarias, ordenadas de lo más agregado a lo más granular:

1. Correlación de rangos de Spearman entre los seis escenarios: mide hasta qué punto las seis ordenaciones coinciden entre sí. Se resume en su valor medio (concordancia típica entre escenarios) y su valor mínimo (el caso de los dos escenarios más dispares).
2. Desplazamiento medio de rango (*Average Shift in Rank*, ASR): toma el rango mediano de cada comunidad entre los seis escenarios y calcula cuántas posiciones se aleja de él, en promedio, al cambiar de escenario. Es una medida directa, expresada en número de puestos.
3. Descomposición del efecto de cada decisión, que separa cuánto mueve el ranking cambiar solo los pesos (fijado el método de agregación) frente a cambiar solo el método (fijados los pesos), para identificar cuál de las dos decisiones tiene más fuerza.
4. Análisis de incertidumbre por simulación de Montecarlo: Se han generado 2.000 combinaciones de pesos aleatorios extraídos de una distribución de Dirichlet (lo que explora repartos de peso muy distintos a los tres propuestos), junto con una elección aleatoria del método de agregación. Para cada comunidad se obtuvo así la distribución completa de su posición, resumida en un rango mediano y un intervalo empírico del 90 % (percentiles 5 y 95).

Mientras las tres primeras técnicas examinan la sensibilidad ante las decisiones concretas que el estudio sí ha tomado, la cuarta somete el índice a un estrés mucho más severo, explorando miles de repartos de peso que nunca se plantearon.

3.7.2 Robustez global entre escenarios

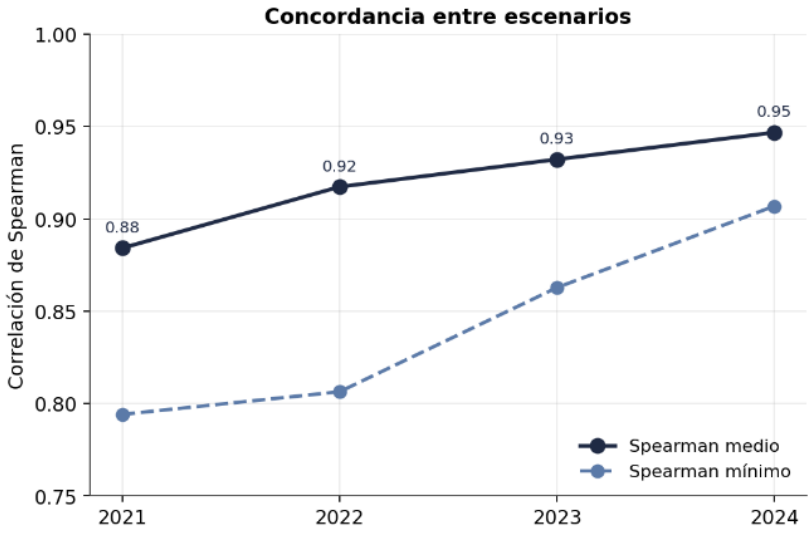
Los resultados confirman una robustez elevada (Tabla 3.7.2). La correlación de Spearman media entre los seis escenarios pasa de 0,88 en 2021 a 0,95 en 2024, y el mínimo o grado de acuerdo entre los dos escenarios más enfrentados nunca baja de 0,79, ascendiendo hasta 0,91 en 2024. Esto indica que incluso las dos versiones más dispares del índice ordenan a las comunidades de forma muy parecida.

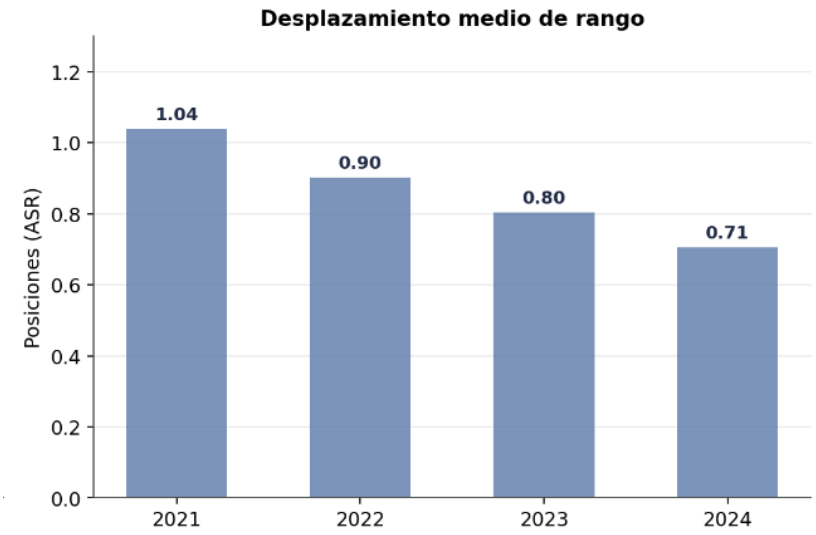
Tabla 3.7.2. *Indicadores de robustez del IDJ por año*

Año	Spearman medio	Spearman mínimo	ASR (posiciones)	Efecto ponderación	Efecto agregación
2021	0,884	0,794	1,04	1,29	1,84
2022	0,917	0,806	0,90	1,27	1,65
2023	0,932	0,863	0,80	1,59	1,25
2024	0,947	0,907	0,71	1,21	1,25

El desplazamiento medio de rango refuerza esta lectura: ronda una sola posición en todos los años y desciende de forma sostenida (1,04 en 2021; 0,71 en 2024). En términos prácticos, una comunidad autónoma típica no se mueve, de media, más allá de un puesto al cambiar de método. Los Gráficos 3.7.2 muestran ambas tendencias de forma conjunta: la concordancia entre escenarios aumenta y el desplazamiento de rango disminuye a lo largo del periodo 2021-2024.

Gráficos 3.7.2. *Evolución temporal de la robustez del IDJ, 2021-2024*





Conviene subrayar que la robustez de cada año se calcula de forma independiente (con su propia normalización y solo con los rankings de ese año), por lo que esta evolución no es que sea un arrastre metodológico, sino un reflejo de la estructura de los datos de cada ejercicio: en los años más recientes las comunidades aparecen algo más separadas entre sí y los indicadores algo más alineados, lo que hace el ranking menos sensible al método. La menor robustez de 2021 podría explicarse con la distorsión que el rebote posterior a 2020 introduce en las variables dinámicas de ese año, como posible hipótesis. En todo caso, con solo cuatro ejercicios esta mejora debemos leerla como una simple observación descriptiva, no como una tendencia consolidada ni extrapolable.

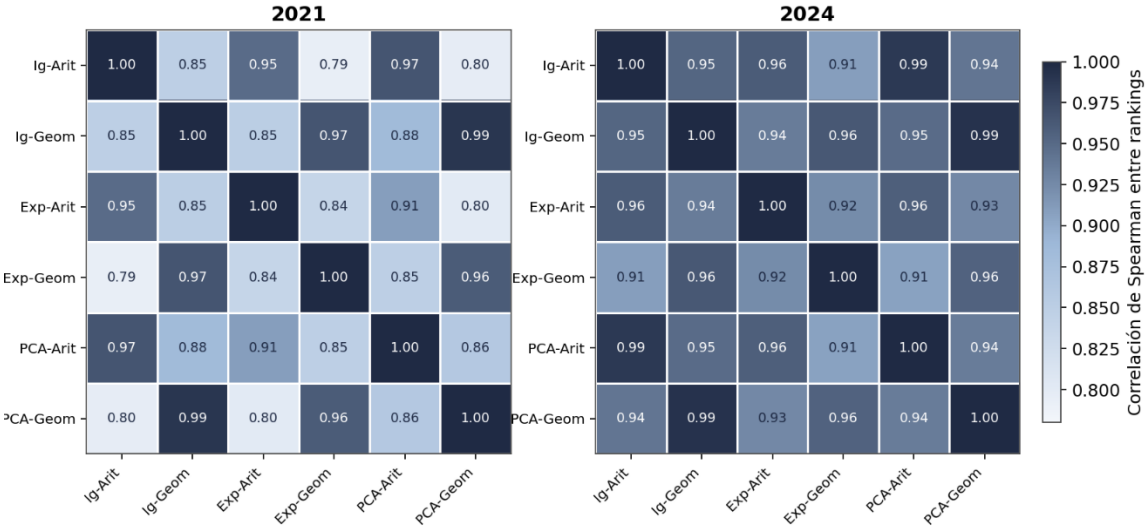
3.7.3 El peso de las decisiones metodológicas

La descomposición del efecto de cada decisión (últimas dos columnas de la Tabla 3.7.2) revela que ni la ponderación ni la agregación dominan de forma abrumadora: ambas mueven el ranking en torno a un par de posiciones. No obstante, sí se aprecia un matiz relevante. En los primeros años el método de agregación pesa algo más que la elección de pesos (1,84 frente a 1,29 en 2021); esa diferencia se estrecha aún más si nos fijamos en 2024, donde terminan prácticamente por igualarse en (1,25 frente a 1,21).

Este patrón se aprecia con nitidez en el mapa de calor de la Figura 3.7.3, que muestra la matriz de correlaciones de Spearman entre los seis escenarios. Se puede apreciar cómo los valores se organizan en bloques según el método de agregación: los dos escenarios aritméticos correlacionan entre sí de forma altísima (Ig-Arit ~ PCA-Arit = 0,99 en 2024),

y ocurre igual entre los geométricos (Ig-Geom ~ PCA-Geom = 0,99), mientras que las parejas que cruzan métodos distintos son las que registran los coeficientes más bajos (el mínimo de 2024, 0,91, corresponde a Exp-Geom ~ PCA-Arit). Dicho de otro modo, cambiar el esquema de pesos altera el ranking menos que cambiar de media aritmética a geométrica: el principal eje de divergencia metodológica es la agregación, no la ponderación. El contraste entre los dos paneles también ilustra la mejora de 2024 frente a 2021, pues las celdas más claras (correlaciones bajas) de 2021 prácticamente desaparecen en 2024.

Figura 3.7.3. *Concordancia entre los seis escenarios metodológicos, 2021 vs. 2024*



3.7.4 Estabilidad por comunidad

El análisis agregado certifica que el índice es robusto en su conjunto, pero la garantía no es uniforme a lo largo de la tabla. La Tabla 3.7.4(1) desciende al detalle de cada comunidad en 2024, contrastando dos intervalos de naturaleza distinta: el de los seis escenarios del estudio (mejor y peor posición efectivamente observada) y el intervalo del 90 % de Montecarlo (el rango que abarca el 90 % de las 2.000 simulaciones con pesos aleatorios). El primero mide la sensibilidad a las decisiones realmente adoptadas; el segundo, mucho más ancho por construcción, mide la resistencia del índice ante repartos de peso mucho más extremos.

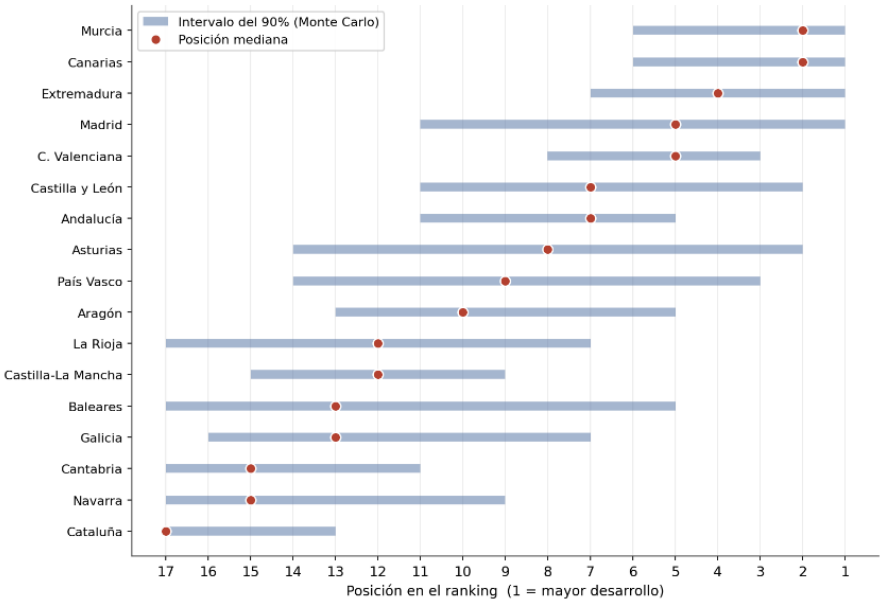
Tabla 3.7.4.(1) *Estabilidad de la posición de cada comunidad (2024)*

Comunidad	Rango mediano	Intervalo 6 escenarios	Amplitud	IC 90 % Montecarlo
Canarias	1,0	[1-2]	1	[1-6]
Murcia	2,0	[1-2]	1	[1-6]
Extremadura	3,0	[3-5]	2	[1-7]
C. Valenciana	4,5	[4-7]	3	[3-8]
Madrid	5,0	[3-7]	4	[1-11]
Castilla y León	5,5	[4-7]	3	[2-11]
Andalucía	7,5	[6-8]	2	[5-11]
Asturias	7,5	[5-10]	5	[2-14]
País Vasco	9,0	[8-10]	2	[3-14]
Aragón	10,0	[8-11]	3	[5-13]
Galicia	11,0	[9-12]	3	[7-16]
La Rioja	12,0	[9-13]	4	[7-17]
Castilla-La Mancha	13,0	[13-15]	2	[9-15]
Baleares	14,0	[12-16]	4	[5-17]
Navarra	14,5	[14-15]	1	[9-17]
Cantabria	16,0	[15-16]	1	[11-17]
Cataluña	17,0	[17-17]	0	[13-17]

La conclusión es clara: los extremos de la tabla son muy estables y la zona media es más volátil. Bajo las seis combinaciones metodológicas, Cataluña ocupa la última posición de forma invariable (amplitud 0, posición 17.ª en los cuatro años), y el bloque de cabeza de 2024 (Canarias, Murcia, Extremadura, la Comunidad Valenciana y la Comunidad de Madrid) se mantiene arriba con independencia del método: Canarias y Murcia oscilan apenas entre la 1.ª y la 2.ª posición. En cambio, comunidades de la zona intermedia como Asturias (banda [5-10]) o La Rioja ([9-13]) presentan amplitudes notablemente mayores: su posición exacta sí depende de la decisión metodológica.

La simulación de Montecarlo lo confirma de manera aún más exigente (Gráfico 3.7.4). Aun sometidas a 2.000 repartos de peso aleatorios, las comunidades de cabeza y de cola mantienen bandas comparativamente estrechas y ancladas en su extremo (Canarias y Murcia no caen del 6.º puesto; Cataluña no sube del 13.º), mientras que las comunidades centrales despliegan intervalos amplísimos (Asturias, por ejemplo, podría situarse entre la 2.ª y la 14.ª posición según cómo se reparta el peso). El contraste entre el intervalo estrecho de los seis escenarios y el ancho de Montecarlo para una misma comunidad mide, precisamente, cuánto margen de seguridad ofrece su posición frente a decisiones de ponderación arbitrarias.

Gráfico 3.7.4. *Incertidumbre de la posición de cada comunidad, 2024. Para la evolución de las bandas en los cuatro años, véase el anexo)*



La estabilidad de los extremos se reproduce en el tiempo, como se aprecia en la Tabla 3.7.4(2). Cataluña cierra el ranking los cuatro años, y el vértice superior lo ocupan de forma recurrente Canarias, Murcia, Madrid y la Comunidad Valenciana. Conviene señalar, no obstante, que algún movimiento sí refleja un cambio real del sector y no ruido metodológico: Extremadura escala desde un rango mediano de 10,5 en 2021 hasta el 3.º puesto en 2024, una trayectoria demasiado sostenida para atribuirla a la elección de método.

Tabla 3.7.4(2) *Rango mediano de cada comunidad por año (entre los seis escenarios)*

Comunidad	2021	2022	2023	2024
Canarias	1,5	2,0	2,5	1,0
Murcia	4,0	2,5	3,5	2,0
Extremadura	10,5	5,0	5,5	3,0
C. Valenciana	3,0	4,5	3,0	4,5
Madrid	3,0	2,0	1,5	5,0
Castilla y León	6,5	10,0	9,5	5,5
Andalucía	7,0	9,0	9,5	7,5
Asturias	9,5	5,5	8,0	7,5
País Vasco	8,5	9,0	5,5	9,0
Aragón	5,0	7,0	7,5	10,0
Galicia	13,0	12,0	13,0	11,0
La Rioja	14,0	14,5	11,0	12,0
Castilla-La Mancha	9,0	10,0	15,0	13,0
Baleares	15,0	16,0	16,0	14,0
Navarra	10,0	15,0	14,0	14,5
Cantabria	16,0	13,0	12,0	16,0
Cataluña	17,0	17,0	17,0	17,0

3.7.4 Conclusión del análisis de sensibilidad

En conjunto, el análisis de sensibilidad respalda la validez del Índice de Desarrollo del Juego como herramienta de ordenación. La imagen que ofrece del sector es estable frente a las decisiones metodológicas adoptadas, con correlaciones por encima de 0,88 de media y desplazamientos en torno a solamente una posición independientemente de la metodología. Esa estabilidad se reproduce, e incluso se refuerza, a lo largo de los cuatro años analizados. De ello se desprende una única cautela interpretativa, que conviene tener presente en el apartado de resultados: la pertenencia a los grupos de cabeza y de cola es

sólida y puede leerse literalmente, mientras que las posiciones intermedias deben interpretarse como bandas y no como puestos exactos. Con esta salvedad, el índice puede abordarse con confianza como base del análisis de resultados.

4. RESULTADOS

El proceso metodológico descrito en los apartados anteriores no produce un único índice, sino seis versiones del Índice de Desarrollo del Juego (IDJ), fruto de combinar los tres esquemas de ponderación (iguales, experto y componentes principales) con los dos métodos de agregación (media aritmética y media geométrica).

Antes de exponer los resultados conviene, por tanto, fijar cuál de esas seis versiones se adopta como referencia. Esta decisión no es trivial ni completamente neutral: como ya puso de manifiesto el análisis de sensibilidad, no existe una elección correcta en sentido estricto, porque todo indicador compuesto incorpora supuestos discutibles en el momento de ponderar y agregar. Lo que sí puede hacerse es escoger la versión más defendible y comprobar que la imagen que ofrece no depende de esa elección.

4.1 ELECCIÓN DEL RANKING DE REFERENCIA

Se adopta como ranking de referencia la versión de pesos iguales con media geométrica, por tres motivos que se refuerzan entre sí:

1. Transparencia: Los pesos iguales son la opción más neutral y reproducible: reparten la importancia a partes iguales entre dimensiones sin introducir juicios subjetivos (como hace el esquema experto) ni hacer depender los pesos de la estructura de correlación de una muestra pequeña de 17 comunidades (como ocurre con el PCA). Es la elección por defecto recomendada cuando no existe un consenso teórico firme sobre la importancia relativa de cada componente.

2. Criterio de agregación: La media geométrica limita la *compensabilidad*: un valor alto en una dimensión no puede enjugar por completo un valor bajo en otra, de modo que penaliza los perfiles desequilibrados. Esto encaja con la naturaleza multidimensional del desarrollo del sector y sigue el precedente de indicadores

consolidados como el Índice de Desarrollo Humano, que en 2010 sustituyó la media aritmética por la geométrica por esta misma razón.[9]

3. Representatividad empírica: Es, además, la versión que mejor resume al resto. Si se toma como "consenso" la mediana de las seis ordenaciones, esta versión es la que más se le aproxima en los cuatro años, con una correlación de Spearman media de 0,980, la más alta de las seis, como se aprecia en la Tabla 4.1(1). Dicho de otro modo, es la versión más central, la que menos se aparta del acuerdo general.

Tabla 4.1.(1) *Proximidad de cada versión al ranking de consenso (correlación de Spearman con la mediana de las seis ordenaciones)*

Año	Ig-Arit	Ig-Geom	Exp-Arit	Exp-Geom	PCA-Arit	PCA-Geom
2021	0,931	**0,963**	0,927	0,950	0,948	0,933
2022	0,921	**0,988**	0,950	0,961	0,961	0,975
2023	0,955	**0,984**	0,948	0,946	0,988	0,984
2024	0,977	**0,986**	0,964	0,964	0,976	0,980
Media	0,946	**0,980**	0,947	0,955	0,968	0,968

Conviene insistir en que esta elección es cómoda precisamente porque el análisis de sensibilidad ya demostró que el ranking es robusto: como se aprecia en la Tabla 4.1(2), las seis versiones del año 2024 coinciden casi por completo, sobre todo en los extremos. La decisión de referencia no fuerza, por tanto, una imagen particular del sector, sino que selecciona la representación más transparente de un resultado que se mantiene estable desde múltiples perspectivas.

[9] Sudhir Anand, Recasting Human Development Measures, Human Development Report Office Discussion Paper, Programa de las Naciones Unidas para el Desarrollo (UNDP), Nueva York, marzo de 2018, p. 3.

Tabla 4.1(2). *Puesto de cada comunidad en las seis versiones del IDJ (2024)*

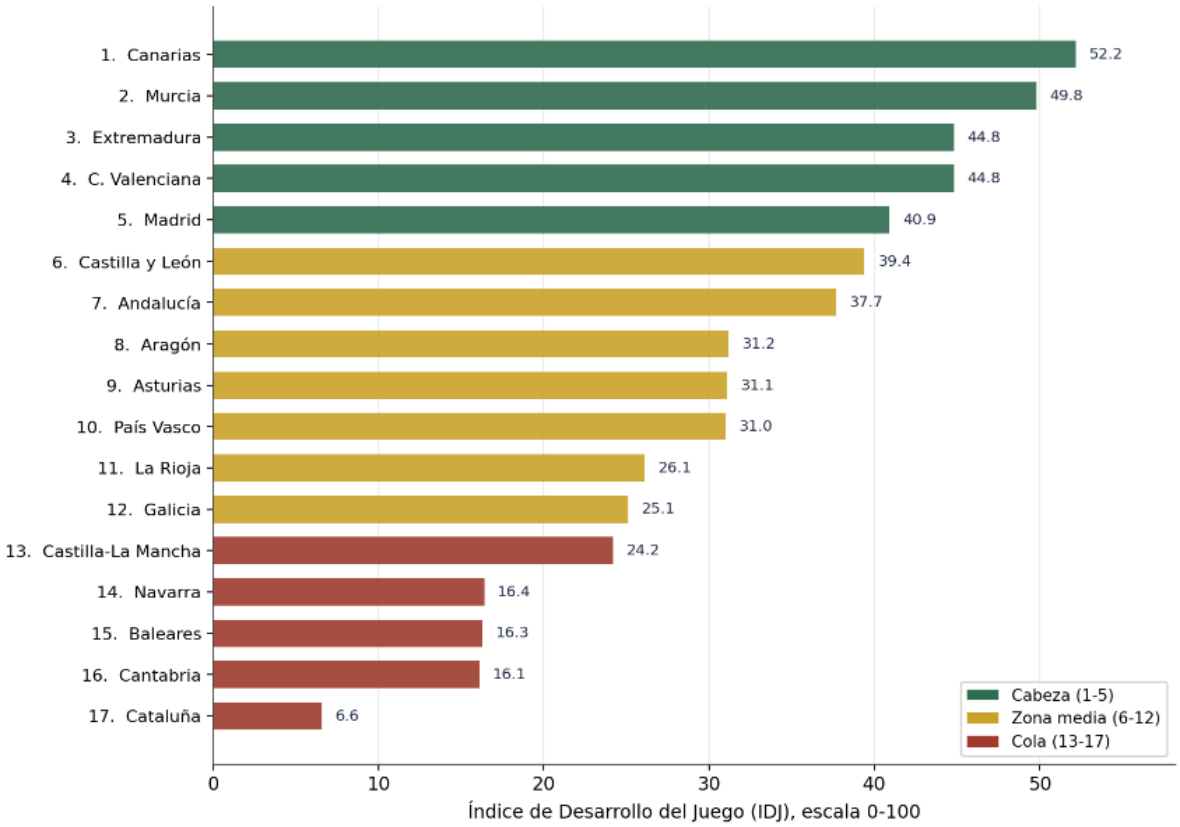
Comunidad	Ig-Arit	Ig-Geom	Exp-Arit	Exp-Geom	PCA-Arit	PCA-Geom
Canarias	1	1	1	1	2	1
Murcia	2	2	2	2	1	2
Extremadura	3	3	4	5	3	3
C. Valenciana	6	4	7	4	5	4
Madrid	7	5	3	3	7	5
Castilla y León	4	6	5	7	4	6
Andalucía	8	7	8	6	8	7
Aragón	10	8	11	11	10	8
Asturias	5	9	6	10	6	10
País Vasco	9	10	10	8	9	9
La Rioja	12	11	12	9	13	12
Galicia	11	12	9	12	11	11
Castilla-La Mancha	13	13	15	13	14	13
Navarra	15	14	14	15	15	14
Baleares	14	15	13	14	12	16
Cantabria	16	16	16	16	16	15
Cataluña	17	17	17	17	17	17

4.2 EL RANKING DEL IDJ EN 2024

El Gráfico 4.2 (ver también la tabla del Anexo 7.4) presenta el resultado principal del trabajo: la ordenación de las diecisiete comunidades autónomas según el IDJ de referencia en el último año del panel, 2024, sobre una escala 0-100. Encabezan la clasificación Canarias (52,2) y la Región de Murcia (49,8), seguidas de Extremadura y la Comunidad Valenciana (44,8 ambas) y de la Comunidad de Madrid (40,9), que cierran el grupo de cabeza. En el extremo opuesto, Cataluña ocupa la última posición con un valor muy bajo (6,6), precedida de Cantabria, Baleares y Navarra. Cabe mencionar que la escala 0-100 se utiliza únicamente para facilitar la lectura, sin implicar que el valor 100 sea alcanzable

en la práctica, ya que el índice de referencia se obtiene mediante agregación geométrica de variables normalizadas

Gráfico 4.2. *Ranking y puntuación del IDJ por comunidad autónoma, 2024*



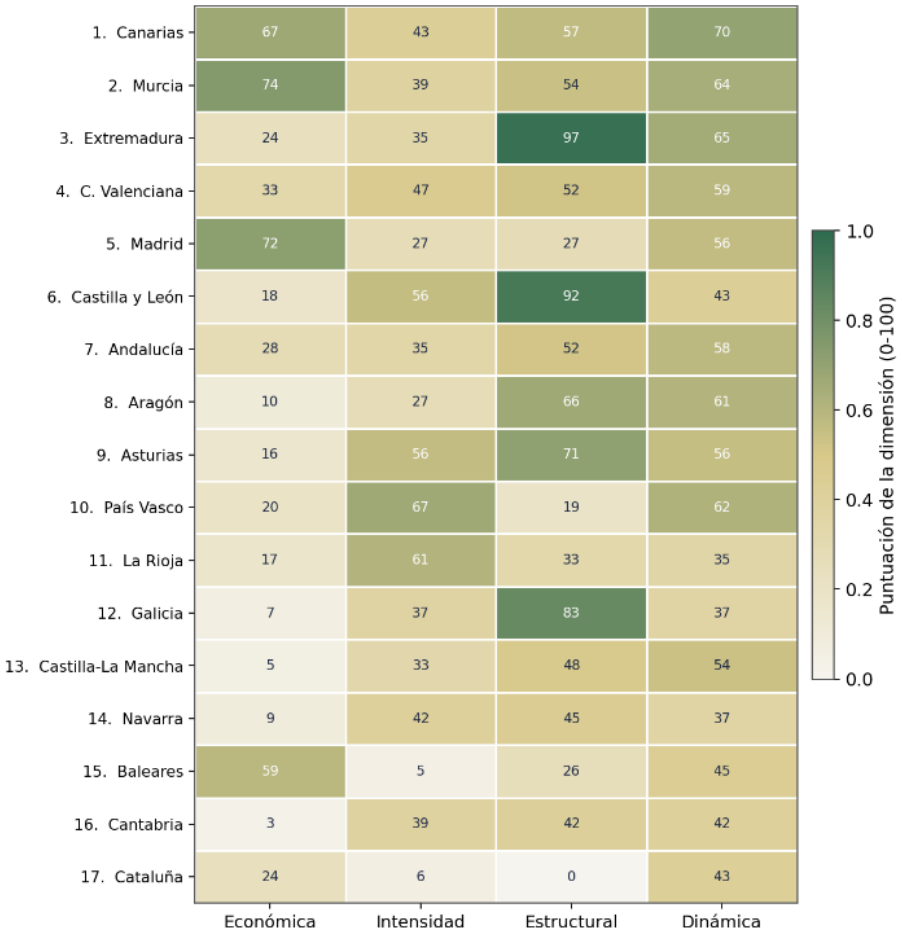
El resultado exige una lectura cuidadosa, porque puede resultar contraintuitivo a primera vista. El IDJ no mide el tamaño absoluto del sector del juego, sino su grado de desarrollo relativo dentro de cada territorio: intensidad de consumo por habitante, densidad de locales y empresas, peso económico per cápita y dinamismo. Por eso las dos comunidades con el sector más grande en términos absolutos, Cataluña y Madrid, que concentran las mayores cifras de negocio y de cantidades jugadas, no lideran la clasificación. Madrid se sostiene en el grupo de cabeza gracias a su músculo económico, pero Cataluña cae a la última posición porque, una vez relativizadas las magnitudes por población, su densidad estructural y su intensidad de consumo son las más bajas del panel. El índice premia, en suma, a las comunidades donde el juego está proporcionalmente más arraigado y desarrollado, no a las que simplemente mueven más dinero por su mayor tamaño.

La distribución de los valores dibuja con nitidez los tres tramos que ya anticipaba el análisis de sensibilidad: un grupo de cabeza compacto (puestos 1 a 5, con valores entre 40 y 52), una amplia zona media (puestos 6 a 12, entre 25 y 39, con diferencias pequeñas entre comunidades) y un grupo de cola (puestos 13 a 17), donde reaparece un escalón marcado hasta el valor mínimo de Cataluña. Coherentemente con lo visto en la sensibilidad, la pertenencia a la cabeza y a la cola puede leerse de forma literal, mientras que las posiciones dentro de la zona media, muy próximas entre sí, deben interpretarse con prudencia.

4.3 EL PERFIL POR DIMENSIONES

Para entender por qué cada comunidad ocupa su lugar, el Gráfico 4.3 descompone el índice en sus cuatro dimensiones (económica, intensidad de consumo, estructural y dinámica), puntuadas también de 0 a 100. El mapa de calor revela que las posiciones altas se alcanzan por caminos distintos, lo que confirma el carácter multidimensional del fenómeno. Ver Anexo 7.4 para comparar directamente dimensiones con IDJ.

Gráfico 4.3. *Perfil por dimensión de cada comunidad, 2024*



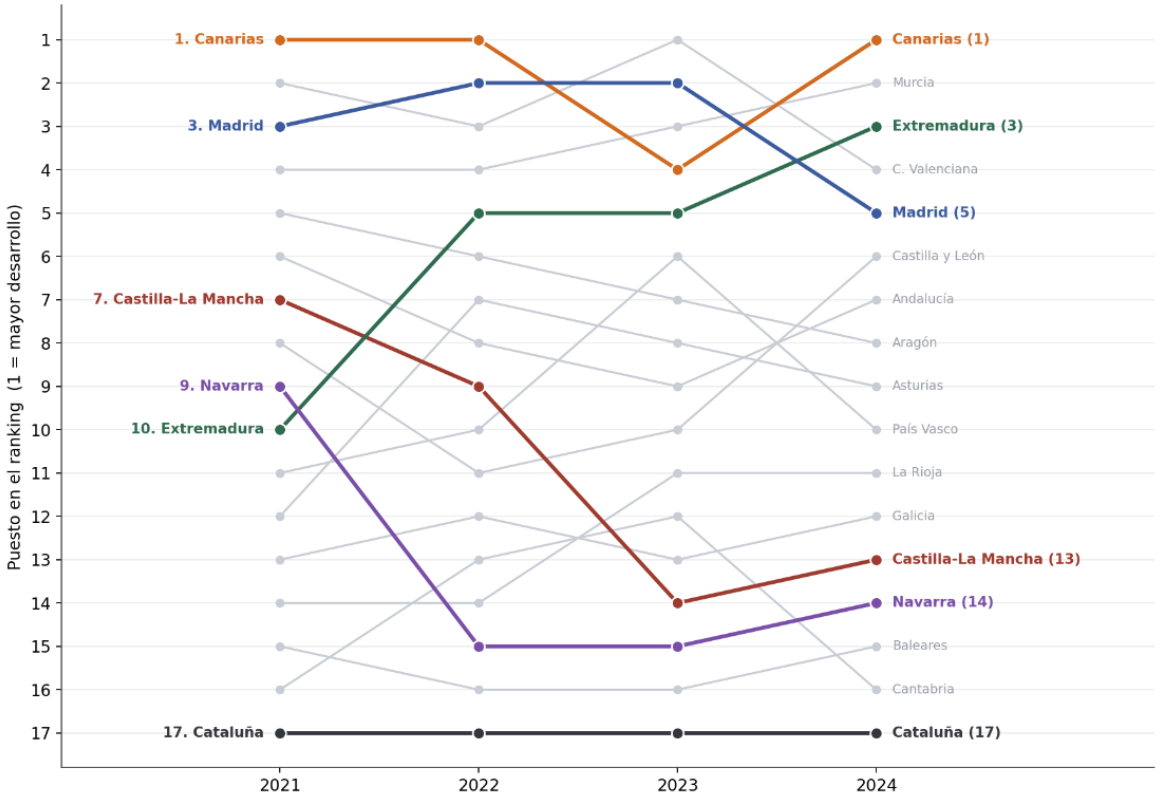
Los líderes combinan fortalezas en varias dimensiones. Canarias encabeza el índice con un perfil notablemente equilibrado y picos en la dimensión dinámica (70) y económica (67). Murcia se apoya en la mayor puntuación económica de todo el panel (74). El caso de Extremadura es el más singular: su tercer puesto se explica casi por completo por una densidad estructural extraordinaria (97), reflejo de un número muy elevado de locales y empresas por habitante en una comunidad poco poblada; lo mismo ocurre, en menor grado, con Castilla y León (92) y Galicia (83). En el polo contrario, Madrid exhibe un perfil opuesto al de Extremadura, con economía potente (72) pero densidad estructural baja (27), lo que ilustra cómo dos vías muy diferentes pueden conducir a posiciones cercanas.

La descomposición también explica los descensos. Baleares, pese a una dimensión económica alta (59), se hunde hasta el puesto 15 por la menor intensidad de consumo del panel (5). Y Cataluña queda última por acumular puntuaciones bajas en todas sus dimensiones; concretamente las mínimas en las dos dimensiones que la media geométrica más penaliza: estructural (0) e intensidad (6).

4.4 EVOLUCIÓN TEMPORAL DEL RANKING (2021-2024)

El IDJ se ha calculado para los cuatro años del panel, lo que permite analizar no solo una fotografía, sino el movimiento del sector (Gráfico 4.4 y Anexo 7.3 para tabla). Conviene precisar que, al normalizarse las variables año a año, la comparación temporal debe hacerse sobre los puestos del ranking más que sobre el valor absoluto del IDJ, que es una medida relativa dentro de cada año.

Gráfico 4.4. *Evolución del ranking del IDJ, 2021-2024*



El rasgo dominante es la estabilidad de los extremos: Cataluña cierra la tabla los cuatro años sin excepción, y la cabeza la ocupan de forma recurrente Canarias, Murcia, la Comunidad Valenciana y Madrid. Sobre ese fondo estable, destacan algunos movimientos de fondo que no son ruido metodológico, sino cambios reales del sector. El más llamativo es el ascenso sostenido de Extremadura, que escala del puesto 10 en 2021 al 3 en 2024 impulsada por su densidad estructural. En sentido contrario, Castilla-La Mancha (del 7 al 13) y Navarra (del 9 al 14) retroceden de manera clara a lo largo del periodo. La zona media, por su parte, muestra un intercambio de posiciones más intenso, reflejo de la proximidad de los valores que ya señalaba el análisis de sensibilidad.

4.5 SÍNTESIS

En conjunto, los resultados dibujan un mapa del desarrollo del juego en España que es estable en sus extremos y matizado en su zona central. La cabeza la forman comunidades donde el sector está proporcionalmente más arraigado, ya sea por su intensidad de consumo, su densidad empresarial o su peso económico relativo; la cola la integran territorios donde, una vez descontado el tamaño, el sector pesa comparativamente poco, con Cataluña **como** caso extremo. El análisis por dimensiones muestra además que no

hay un único patrón de desarrollo, sino varias vías que conducen a posiciones similares, y la perspectiva temporal confirma que esta estructura se mantiene e incluso parece que se consolida entre 2021 y 2024. Sobre esta base, y teniendo presente que las posiciones intermedias deben leerse como bandas y no como puestos exactos, se construyen las conclusiones del trabajo.

5. CONCLUSIÓN

El objetivo principal de este trabajo ha sido construir un indicador compuesto que permitiera medir el grado de desarrollo autonómico de la industria del juego en España. La pregunta de investigación planteada puede responderse afirmativamente: sí, hemos logrado sintetizar, mediante una metodología ordenada y replicable, distintas dimensiones del sector del juego en un Índice de Desarrollo del Juego (IDJ) robusto, capaz de mostrar una lectura territorial homogénea y comparable.

La principal aportación del trabajo no consiste exclusivamente en obtener un ranking autonómico, sino en demostrar que la realidad del juego no puede interpretarse correctamente a partir de una sola variable. La facturación, el número de locales, la cantidad jugada, el número de empresas o la evolución reciente del sector son piezas relevantes, pero ninguna de ellas era capaz de ofrecer por sí sola una imagen completa. El IDJ nos ha permitido integrar todas estas perspectivas en una medida común, que ha sido construida evitando que el análisis quede dominado por el tamaño absoluto de las comunidades autónomas. Esta decisión es fundamental, ya que, si se hubieran utilizado magnitudes brutas, comunidades como Madrid o Cataluña habrían aparecido siempre en las primeras posiciones por su peso demográfico y económico.

Uno de los resultados más importantes del trabajo es precisamente esa diferencia entre tamaño absoluto y desarrollo relativo. En 2024, Canarias y la Región de Murcia encabezan el ranking, no porque sean las economías más grandes, sino porque combinan de forma más intensa distintas manifestaciones del desarrollo del juego. Canarias presenta un perfil especialmente equilibrado, con puntuaciones elevadas en las dimensiones económica y dinámica. Murcia, por su parte, destaca por la dimensión económica, que es la más alta del panel. Estos resultados indican que el desarrollo del juego puede adoptar formas distintas: en unos casos se apoya en el peso económico del sector; en otros, en la

intensidad de consumo, en la densidad del tejido empresarial o en la evolución reciente de la actividad.

El caso de Extremadura es uno de los más ilustrativos. Su posición en el tercer lugar del ranking de 2024 no se explica por un gran tamaño económico, sino por una densidad estructural muy elevada. Es decir, el sector tiene una presencia proporcionalmente muy intensa en relación con el tamaño de la comunidad. Esta conclusión es relevante porque muestra que el índice también ha sido capaz de premiar a los territorios que pese a tener menor actividad monetaria cuentan con un juego más implantado en términos de infraestructura relativa. Algo similar, aunque con distinta intensidad, puede observarse en Castilla y León o Galicia, que han destacado en los análisis más por la dimensión estructural que por la económica.

Madrid representa el caso contrario. Su dimensión económica es muy elevada, como cabría esperar por el peso de su mercado, pero su densidad estructural es comparativamente más baja. Esto permite interpretar que Madrid tiene un sector potente en términos económicos, pero no necesariamente más extendido o transversal en relación con su población. De esta forma, el IDJ ayuda a distinguir entre concentración económica y desarrollo territorial del sector, una diferencia que habría quedado oculta en un análisis basado únicamente en cifras absolutas.

Cataluña constituye otro resultado especialmente significativo. A pesar de ser una de las comunidades con mayor volumen absoluto de actividad, ocupa la última posición del ranking de referencia en 2024 y se mantiene en esa posición durante todo el período analizado. Una vez relativizadas las variables por población, PIB u otros denominadores, Cataluña ha presentado valores bajos en muchas de las dimensiones, especialmente en la estructural y en la intensidad de consumo.

También aparecen casos que obligan a una lectura matizada. Baleares, por ejemplo, presenta una dimensión económica elevada, pero queda en la parte baja del ranking debido a su débil puntuación en intensidad de consumo. Este tipo de resultados demuestra la utilidad de emplear una media geométrica como ranking de referencia: el índice no permite que una fortaleza aislada compense por completo debilidades muy acusadas en otras dimensiones. En consecuencia, el IDJ no solo mide cuánto destaca una comunidad en algún aspecto concreto, sino hasta qué punto el desarrollo del sector es equilibrado.

En definitiva, la clasificación ha mostrado en sus múltiples variantes que sus extremos son bastante estables: Cataluña permanece en la última posición en todos los años del marco temporal, mientras que Canarias, Murcia, la Comunidad Valenciana y Madrid aparecen de forma recurrente en la zona alta. Esta estabilidad sugiere que el índice no está capturando únicamente fluctuaciones puntuales, sino patrones territoriales relativamente consistentes. Al mismo tiempo, algunos movimientos concretos, como el ascenso de Extremadura o el retroceso de Castilla-La Mancha y Navarra, pueden deberse tanto a que el IDJ es capaz de recoger cambios reales en la configuración del sector, como a la mayor sensibilidad del índice en las posiciones intermedias. Sobre este último apunte, cabe mencionar que esta zona del ranking ha reflejado diferencias entre comunidades más reducidas y, por tanto, pequeñas variaciones en alguna dimensión o en las decisiones metodológicas de ponderación y agregación han podido producir desplazamientos de varios puestos. Por ello, mientras que las posiciones de cabeza y cola pueden interpretarse con mayor seguridad, los puestos centrales deben entenderse más como bandas aproximadas de desarrollo relativo que como una ordenación exacta.

Una de las fortalezas y ramas que le han aportado solidez al estudio ha sido el trabajo realizado en el análisis de sensibilidad. La construcción de cualquier indicador compuesto implica decisiones metodológicas discutibles: qué variables incluir, cómo normalizarlas, qué pesos asignar y cómo agregarlas. Por ello, el análisis no se ha limitado a presentar un único resultado, sino que compara diferentes alternativas de ponderación y agregación. El hecho de que las seis versiones del índice produzcan clasificaciones similares, con correlaciones elevadas y desplazamientos medios de escasas posiciones, refuerza la validez del IDJ. Esto permite afirmar que las principales conclusiones no dependen de una elección metodológica concreta, sino que son robustas frente a cambios razonables en el diseño del indicador.

No obstante, esta robustez debe interpretarse correctamente. Que los rankings sean parecidos no significa que el índice esté libre de limitaciones ni que exista una única forma correcta de medir el fenómeno. En parte, la estabilidad de los resultados puede deberse a que las variables seleccionadas capturan una estructura territorial de fondo bastante consistente. Es decir, que determinadas comunidades pueden tender a situarse en posiciones altas o bajas con independencia del método porque presentan perfiles claramente diferenciados. Además, aunque se ha realizado un análisis de correlaciones y

se ha eliminado la única redundancia clara entre variables, no puede descartarse que varios indicadores reflejen parcialmente una misma realidad económica o territorial desde diferentes ángulos. Por tanto, la robustez del IDJ no debe entenderse como una prueba absoluta de exactitud, sino como una garantía razonable de que el ranking no es un resultado arbitrario producido por una única decisión metodológica.

Otro motivo por el que los resultados deben interpretarse con cautela recae en el concepto de que el IDJ mide desarrollo del sector, no deseabilidad social. Una puntuación alta no significa necesariamente que una comunidad se encuentre en una situación mejor desde el punto de vista social o económico en cuanto al juego. Además, algunas variables pueden verse afectadas por características territoriales particulares, como el turismo en comunidades como Canarias o Baleares, que puede alterar la lectura de los indicadores per cápita. Este trabajo no evalúa esos efectos sociales, sino el grado de implantación y desarrollo económico-sectorial de la actividad. Por ello, el ranking no debe leerse como una clasificación normativa de comunidades mejores o peores, sino como una herramienta analítica para describir una realidad económica y territorial.

Como posible ampliación del estudio, resulta especialmente interesante incorporar en el futuro la dimensión legislativa. El trabajo comenzó con la intención de analizar también las diferencias normativas entre comunidades autónomas, pero finalmente se optó por separar ambos planos. Esta decisión responde a una razón metodológica: la regulación no mide directamente el desarrollo del mercado, sino la intensidad, madurez o rigidez del marco normativo aplicable. Por ello, una continuación natural de este trabajo sería construir un segundo indicador compuesto de rigidez regulatoria y compararlo con el IDJ. De este modo, podría analizarse si las comunidades con mayor desarrollo del juego presentan marcos normativos más estrictos, más recientes o intervencionistas, o si, por el contrario, no existe una relación clara entre desarrollo sectorial y regulación autonómica.

En conclusión, el trabajo demuestra que los indicadores compuestos son una herramienta útil para analizar fenómenos económicos complejos y territorialmente heterogéneos. El Índice de Desarrollo del Juego permite transformar datos dispersos en una medida sintética, interpretable y robusta. Sus resultados muestran que el desarrollo autonómico de la industria del juego en España no responde únicamente al tamaño económico de cada territorio, sino a una combinación de factores económicos, estructurales, de consumo y de dinámica sectorial. El IDJ no agota el análisis del sector, pero sí ofrece una base sólida

para comprenderlo mejor y para desarrollar futuras investigaciones sobre su relación con la regulación, la política pública y sus efectos sociales.

***DECLARACIÓN USO DE INTELIGENCIA ARTIFICIAL**

Por la presente, yo, Álvaro Gimeno Granado, estudiante de E3 Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "El desarrollo autonómico de la industria del juego en España: construcción de un Índice de Desarrollo del Juego para la comparativa regional", 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:

1. Brainstorming de ideas de investigación: Utilizado para idear y esbozar posibles áreas de investigación.
2. Crítico: Para encontrar contra-argumentos a una tesis específica que pretendo defender.
3. Referencias: Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
4. Metodólogo: Para descubrir métodos aplicables a problemas específicos de investigación.
5. Interpretador de código: Para realizar análisis de datos preliminares.
6. Estudios multidisciplinares: Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
7. Constructor de plantillas: Para diseñar formatos específicos para secciones del trabajo.
8. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.
9. Generador previo de diagramas de flujo y contenido: Para esbozar diagramas iniciales.
10. Sintetizador y divulgador de libros complicados: Para resumir y comprender literatura compleja.
11. Generador de problemas de ejemplo: Para ilustrar conceptos y técnicas.
12. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
13. Traductor: Para traducir textos de un lenguaje a otro.

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: 17 de junio de 2026

Firma: Álvaro Gimeno Granado .

6. BIBLIOGRAFÍA

Anand, Sudhir (2018). *Recasting Human Development Measures*. Human Development Report Office Discussion Paper. Nueva York: United Nations Development Programme (UNDP).

Dirección General de Ordenación del Juego , *Detalle datos del mercado español del juego: premios y cantidades jugadas*, disponible en: *DG de Ordenación del Juego – Descarga de datos* (última consulta: 17 de junio de 2026).

Instituto Nacional de Estadística (INE), *Anuario Estadístico de España 2013 - 4. Cultura y ocio*, disponible en: INE – Anuario 2013 (última consulta: 17 de junio de 2026).

Instituto Nacional de Estadística (INE), *Empresas por CCAA, actividad principal (grupos CNAE 2009) y estrato de asalariados*, disponible en: INE - Tabla 39372 (última consulta: 17 de junio de 2026).

Instituto Nacional de Estadística (INE), *Locales por CCAA, actividad principal (grupos CNAE 2009) y estrato de asalariados*, disponible en: INE - Tabla 294 (última consulta: 17 de junio de 2026).

Instituto Nacional de Estadística (INE), *Magnitudes regionalizadas según comunidades y ciudades autónomas y actividad principal (CNAE-2009 a 1 y 2 dígitos)*, disponible en: INE - Tabla 36188 (última consulta: 17 de junio de 2026).

Instituto Nacional de Estadística (INE), *Ocupados por tipo de sector (público o privado), sexo y CCAA*, disponible en: INE - Tabla 65321 (última consulta: 17 de junio de 2026).

Instituto Nacional de Estadística (INE), *P.I.B. a precios de mercado y valor añadido bruto a precios básicos por ramas de actividad: Precios corrientes por comunidades y ciudades autónomas, magnitud y periodo*, disponible en: INE – Tabla 12955 (última consulta: 17 de junio de 2026).

Instituto Nacional de Estadística (INE), *Población según comunidad autónoma y provincia y sexo; Censo anual de poblacion*, disponible en: INE – Tabla 11807 (última consulta: 17 de junio de 2026).

JOLLIFFE, Ian T., *Principal Component Analysis*, 2nd ed., Springer, Nueva York, 2002.

MAZZIOTTA, Matteo; PARETO, Adriano, *Methods for Constructing Composite Indices: One for All or All for One?*, Rivista Italiana di Economia Demografia e Statistica, vol. 67, núm. 2, 2013, pp. 67-80.

NARDO, Michela; SAISANA, Michaela; SALTELLI, Andrea; TARANTOLA, Stefano, *Tools for Composite Indicators Building*, European Commission, Joint Research Centre, Ispra, 2005.

OECD; Joint Research Centre of the European Commission (JRC), *Handbook on Constructing Composite Indicators: Methodology and User Guide*, OECD Publishing, París, 2008.

SAISANA, Michaela; TARANTOLA, Stefano, *State-of-the-Art Report on Current Methodologies and Practices for Composite Indicator Development*, Joint Research Centre, European Commission, Ispra, 2002.

7. ANEXO

7.1 TABLA COMPARATIVA ENTRE MÉTODOS DE PONDERACIÓN

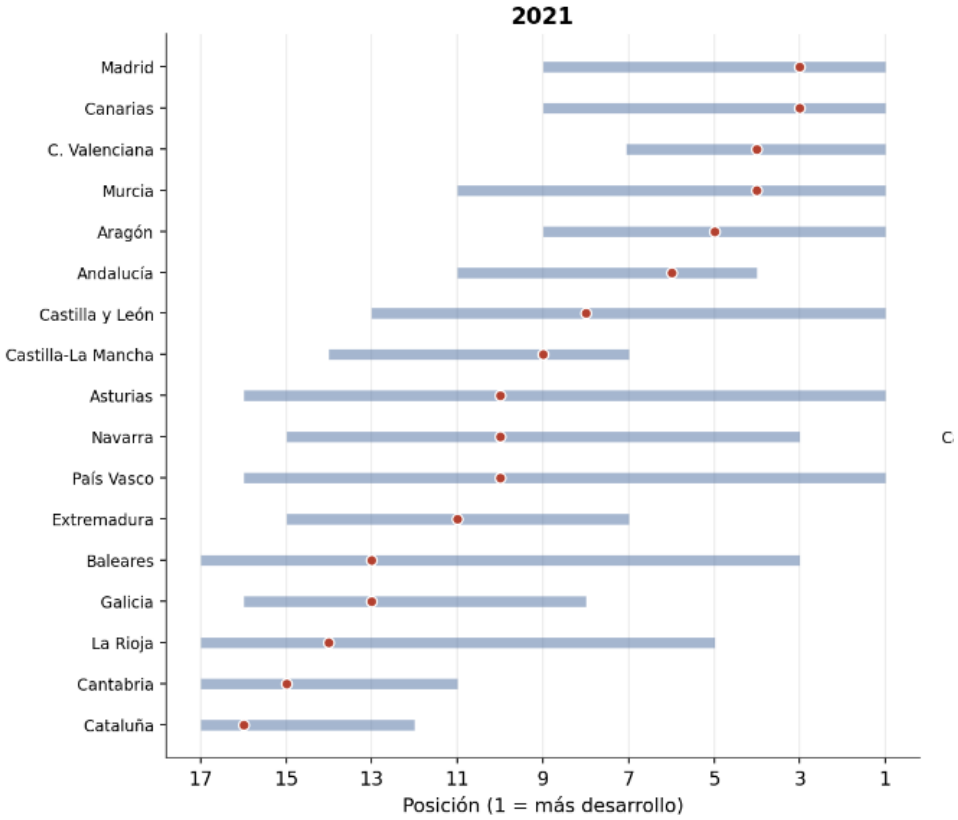
Variable	**Iguales**	**Experto**	**PCA**
cn_pib	8,33	20,0	5,36
salarios_pc	8,33	10,0	8,93
inversion_pc	8,33	10,0	12,34
apuestas_pc	12,50	12,5	9,14
loterias_pc	12,50	12,5	8,30
locales_pc	12,50	12,5	9,01
empresas_pc	12,50	12,5	13,41
mortalidad_emp	6,25	2,5	7,94
var_locales	6,25	2,5	7,31
cagr_jugado	6,25	2,5	12,48
variacion_jugado	6,25	2,5	5,78

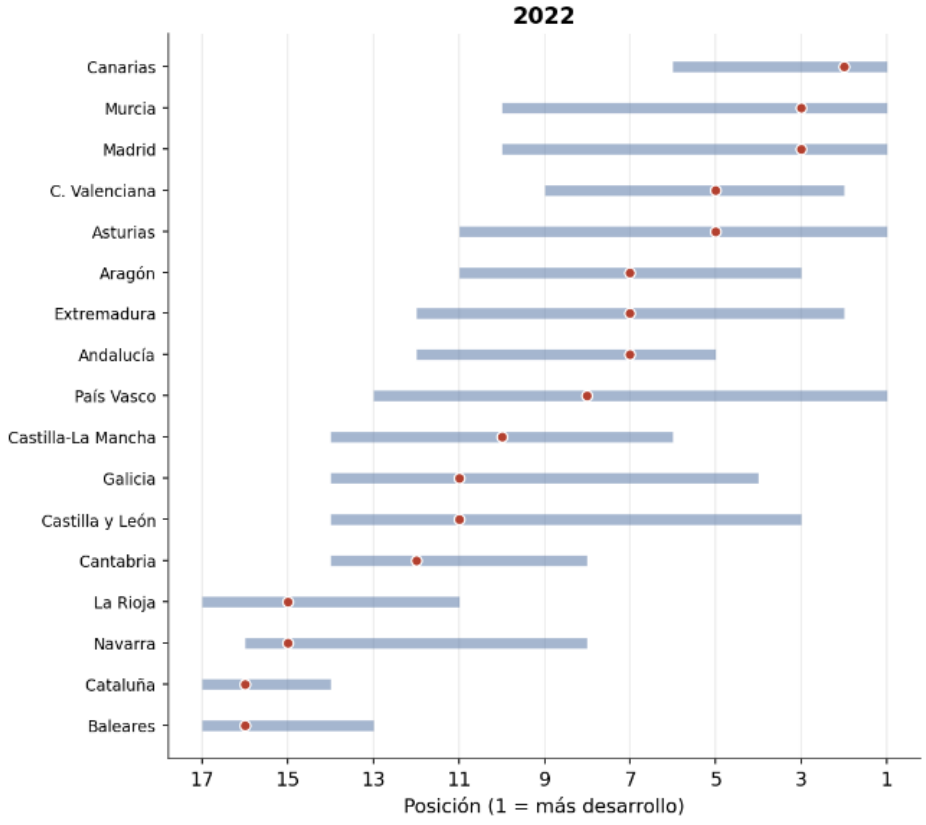
El valor de cada celda corresponde al % que recibe la variable en la ponderación

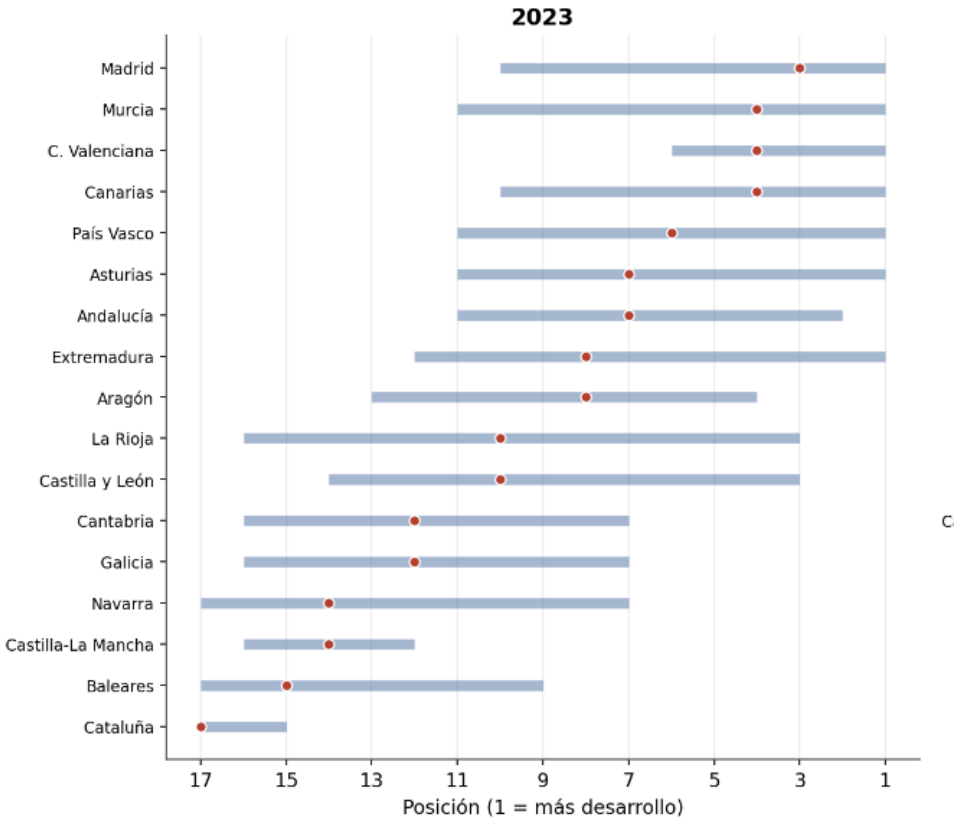
La tabla refleja lo explicado en el apartado 3.5 relativo a métodos de ponderación

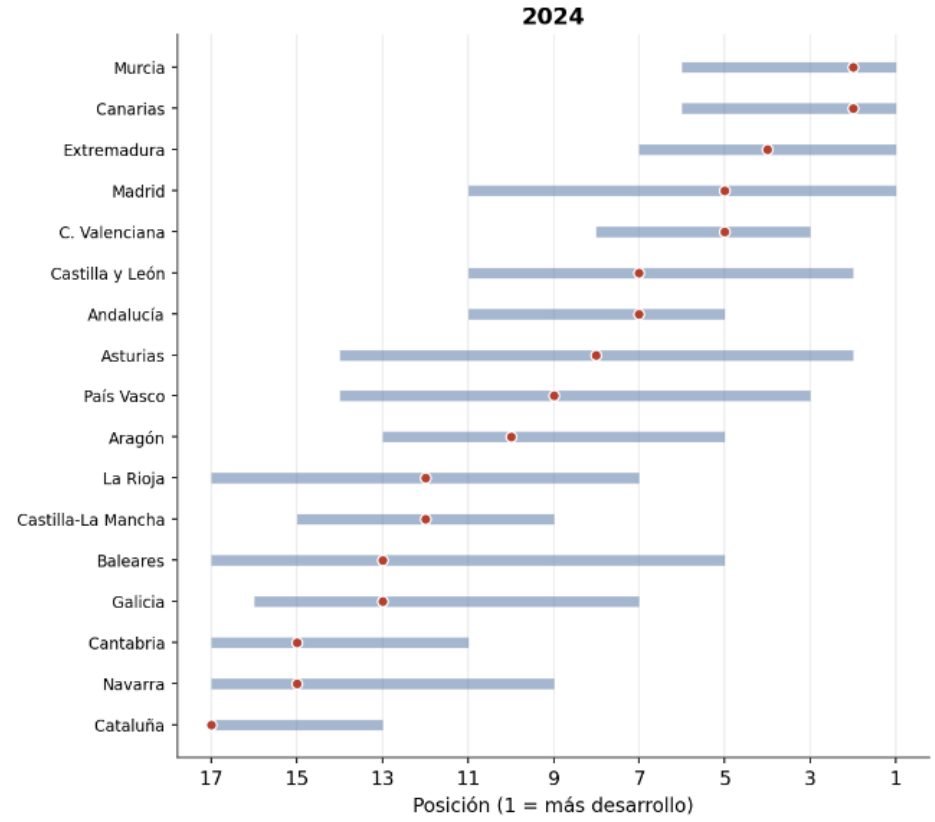
7.2 INTERVALOS DEL 90% DEL RANKING POR SIMULACIÓN DE MONTE CARLO, PARA 2021-2024

(siguiente página)


2021
Madrid
Canarias
C. Valenciana
Murcia
Aragón
Andalucía
Castilla y León
Castilla-La Mancha
Asturias
Navarra
País Vasco
Extremadura
Baleares
Galicia
La Rioja
Cantabria
Cataluña
17
15
13
11
9
7
5
3
1
Posición (1 = más desarrollo)


2022
Canarias
Murcia
Madrid
C. Valenciana
Asturias
Aragón
Extremadura
Andalucía
País Vasco
Castilla-La Mancha
Galicia
Castilla y León
Cantabria
La Rioja
Navarra
Cataluña
Baleares
17
15
13
11
9
7
5
3
1
Posición (1 = más desarrollo)





Los gráficos hacen referencia al apartado 3.7 relativo a la estabilidad de la posición de cada comunidad y concretamente a la mención del uso de la simulación de Monte Carlo.

7.3 EVOLUCIÓN DEL PUESTO EN EL RANKING DE REFERENCIA (2021-2024)

Comunidad	2021	2022	2023	2024
Canarias	1	1	4	1
Murcia	4	4	3	2
Extremadura	10	5	5	3
C. Valenciana	2	3	1	4
Madrid	3	2	2	5
Castilla y León	8	11	10	6
Andalucía	6	8	9	7
Aragón	5	6	7	8
Asturias	12	7	8	9
País Vasco	11	10	6	10
La Rioja	14	14	11	11
Galicia	13	12	13	12
Castilla-La Mancha	7	9	14	13
Navarra	9	15	15	14
Baleares	15	16	16	15
Cantabria	16	13	12	16
Cataluña	17	17	17	17

La tabla hace referencia al apartado 4.4 relativo a la evolución temporal del ranking

7.4 RANKING DEL IDJ Y PERFIL POR DIMENSIÓN (2024, ESCALA 0-100)

Puesto	Comunidad	IDJ	Económica	Intensidad	Estructural	Dinámica
1	Canarias	52,2	67	43	57	70
2	Murcia	49,8	74	39	54	64
3	Extremadura	44,8	24	35	97	65
4	C. Valenciana	44,8	33	47	52	59
5	Madrid	40,9	72	27	27	56
6	Castilla y León	39,4	18	56	92	43
7	Andalucía	37,7	28	35	52	58
8	Aragón	31,2	10	27	66	61
9	Asturias	31,1	16	56	71	56
10	País Vasco	31,0	20	67	19	62
11	La Rioja	26,1	17	61	33	35
12	Galicia	25,1	7	37	83	37

Puesto	Comunidad	IDJ	Económica	Intensidad	Estructural	Dinámica
13	Castilla-La Mancha	24,2	5	33	48	54
14	Navarra	16,4	9	42	45	37
15	Baleares	16,3	59	5	26	45
16	Cantabria	16,1	3	39	42	42
17	Cataluña	6,6	24	6	0	43

La tabla hace referencia al apartado 4.2 Y 4.3 relativos Ranking ofrecido por el IDJ para 2024

7.5 VARIABLE 1: NÚMERO DE LOCALES DE JUEGOS DE AZAR Y APUESTAS[10]

Dimensión: Estructural
Ratio en el que se usó: Nº de locales / cápita
Unidad: nº de locales

Comunidad Autónoma	2021	2022	2023	2024
Andalucía	2.789	2.703	2.689	2.781
Aragón	589	552	520	476
Asturias	295	465	314	331
Illes Balears	311	236	280	341
Canarias	780	751	614	840
Cantabria	129	91	209	141
Castilla y León	1.132	926	1.020	939
Castilla-La Mancha	716	896	664	644
Cataluña	1.657	1.855	1.587	1.574
C. Valenciana	1.876	1.666	1.623	1.720
Extremadura	268	423	563	402
Galicia	866	1.045	933	1.046
Madrid	2.052	1.884	1.943	1.777
Murcia	674	723	583	590
Navarra	237	84	180	199
País Vasco	586	583	689	527
La Rioja	66	46	88	81

[10] Disponible en: *INE - Tabla 36188*

7.6 VARIABLE 2: CIFRA DE NEGOCIOS[11]

Dimensión: Económica
Ratio en el que se usó: Cifra de negocios / PIB
Unidad: miles de € (mil€)

Comunidad Autónoma	**2021**	**2022**	**2023**	**2024**
Andalucía	1.466.786	1.761.429	1.818.930	2.116.344
Aragón	269.861	311.609	430.573	304.577
Asturias	130.174	173.605	205.364	214.168
Illes Balears	258.420	295.785	350.984	421.623
Canarias	568.162	688.553	796.530	934.597
Cantabria	63.975	88.145	136.421	92.521
Castilla y León	327.301	362.918	497.814	480.228
Castilla-La Mancha	316.102	326.557	322.601	349.092
Cataluña	1.308.690	1.596.407	2.015.982	1.933.255
C. Valenciana	1.022.669	1.280.669	1.262.441	1.456.595
Extremadura	157.186	200.993	203.625	222.542
Galicia	358.640	428.597	506.946	554.145
Madrid	11.020.004	11.439.580	11.585.632	11.966.129
Murcia	366.906	508.008	466.052	514.696
Navarra	131.731	142.592	158.690	158.352
País Vasco	595.292	701.108	729.321	1.101.529
La Rioja	52.898	57.865	95.093	80.898

7.7 VARIABLE 3: SUELDOS Y SALARIOS[12]

Dimensión: Económica
Ratio en el que se usó: Sueldos y salarios / cápita
Unidad: miles de € (mil€)

Comunidad Autónoma	**2021**	**2022**	**2023**	**2024**
Andalucía	261.697	301.201	321.791	353.748
Aragón	32.201	32.202	41.015	39.957
Asturias	27.481	28.083	35.167	36.987
Illes Balears	34.758	40.797	46.168	52.990
Canarias	75.390	93.863	101.658	110.163
Cantabria	11.545	13.591	15.876	15.404
Castilla y León	50.832	60.794	65.538	72.294

[11] Disponible en: *INE - Tabla 36188*
[12] Disponible en: *INE - Tabla 36188*

Comunidad Autónoma	2021	2022	2023	2024
Castilla-La Mancha	45.860	51.285	48.071	51.505
Cataluña	165.741	186.737	214.798	221.603
C. Valenciana	154.661	165.176	188.099	208.798
Extremadura	26.552	33.224	33.573	38.192
Galicia	52.345	68.665	67.412	71.057
Madrid	218.023	282.071	248.225	257.628
Murcia	64.186	69.821	79.849	85.613
Navarra	15.662	13.605	18.315	18.832
País Vasco	57.084	61.010	72.186	75.134
La Rioja	7.844	4.662	9.457	11.111

7.8 VARIABLE 4: INVERSIÓN EN ACTIVOS MATERIALES[13]

Dimensión: Económica
Ratio en el que se usó: Inversión en activos materiales / cápita
Unidad: miles de € (mil€)

Comunidad Autónoma	2021	2022	2023	2024
Andalucía	57.630	41.144	28.538	43.445
Aragón	10.187	9.953	10.656	6.267
Asturias	1.889	2.487	2.688	3.283
Illes Balears	25.603	5.833	8.379	19.094
Canarias	34.869	17.883	20.651	30.530
Cantabria	578	3.726	1.689	2.096
Castilla y León	4.043	5.793	7.859	16.222
Castilla-La Mancha	11.543	11.648	5.823	9.712
Cataluña	28.415	31.538	39.730	86.609
C. Valenciana	31.341	28.328	36.658	39.877
Extremadura	3.227	3.480	2.985	6.252
Galicia	5.430	5.247	4.939	11.323
Madrid	48.846	49.036	57.478	88.914
Murcia	12.613	15.124	11.195	24.738
Navarra	3.347	1.484	3.431	3.329
País Vasco	12.853	9.556	13.792	8.753
La Rioja	786	1.260	1.567	1.420

[13] Disponible en: *INE - Tabla 36188*

7.9 VARIABLE 5: PERSONAL OCUPADO EN JUEGOS DE AZAR Y APUESTAS[14]

Dimensión: Estructural
Ratio en el que se usó: Personal ocupado en JdA&A / total ocupados[15]
Unidad: personas

Comunidad Autónoma	**2021**	**2022**	**2023**	**2024**
Andalucía	12.989	13.176	14.004	14.320
Aragón	1.818	1.581	1.659	1.740
Asturias	1.589	1.483	1.483	1.602
Illes Balears	1.799	1.697	1.814	1.904
Canarias	4.285	4.478	4.458	4.849
Cantabria	691	612	755	707
Castilla y León	3.276	2.814	3.115	3.126
Castilla-La Mancha	2.608	2.671	2.368	2.367
Cataluña	8.576	7.766	8.186	8.577
C. Valenciana	8.360	8.072	8.452	8.811
Extremadura	1.403	1.695	1.888	1.655
Galicia	3.293	3.093	3.251	3.347
Madrid	10.452	11.024	9.712	10.072
Murcia	3.304	3.229	3.353	3.525
Navarra	751	629	715	780
País Vasco	2.756	2.495	2.812	2.727
La Rioja	425	228	395	424

7.10 VARIABLE 6: Nº DE EMPRESAS DE JDA&A[16]

Dimensión: Estructural
Ratio en el que se usó: Nº de empresas de JdA&A / cápita
Unidad: empresas

Comunidad Autónoma	**2021**	**2022**	**2023**	**2024**
Andalucía	1.994	1.973	1.836	1.823
Aragón	355	349	329	318
Asturias	331	305	284	274
Illes Balears	235	234	209	201

[14] Disponible en: *INE - Tabla 36188*
[15] Disponible en: *INE - Tabla 65321*
[16] Disponible en: *INE - Tabla 39372*

Comunidad Autónoma	2021	2022	2023	2024
Canarias	465	441	401	410
Cantabria	142	136	138	148
Castilla y León	739	727	704	689
Castilla-La Mancha	481	472	451	451
Cataluña	1.311	1.295	1.213	1.162
C. Valenciana	1.242	1.232	1.144	1.128
Extremadura	344	342	343	331
Galicia	775	728	706	708
Madrid	1.549	1.542	1.339	1.316
Murcia	290	280	269	273
Navarra	155	159	140	146
País Vasco	413	409	393	388
La Rioja	81	77	70	68

7.11 VARIABLE 7: MUERTES DE EMPRESAS DE JDA&A[17]

Dimensión: De dinámica
Ratio en el que se usó: Muertes de empresas de JdA&A / nº de empresas
Unidad: nº de bajas de empresas · (2024 sin dato en la BBDD)

Comunidad Autónoma	2021	2022	2023	2024
Andalucía	88	129	111	—
Aragón	13	23	23	—
Asturias	29	30	21	—
Illes Balears	6	30	11	—
Canarias	33	23	20	—
Cantabria	6	11	7	—
Castilla y León	51	53	41	—
Castilla-La Mancha	18	49	25	—
Cataluña	73	106	117	—
C. Valenciana	76	107	75	—
Extremadura	23	31	26	—
Galicia	64	78	69	—
Madrid	75	136	87	—
Murcia	17	14	22	—
Navarra	5	9	9	—
País Vasco	17	18	19	—

[17] Disponible en: *(2021)* Datos.gob, *(2022)* Datos.gob, *(2023)* Datos.gob, 2024 se imputó como se detalla en el contenido del trabajo

Comunidad Autónoma	2021	2022	2023	2024
La Rioja	7	11	10	—

7.12 VARIABLE 8: LOCALES DE JDA&A

Dimensión: De dinámica
Ratio en el que se usó: Variación de stock de locales = Diferencia de locales ese año / nº de locales del año anterior
Unidad: nº de locales

Comunidad Autónoma	2020	2021	2022	2023	2024	2025
Andalucía	2.479	2.341	2.352	2.475	2.492	2.321
Aragón	463	400	420	437	444	414
Asturias	406	324	337	357	386	311
Illes Balears	334	302	313	333	335	290
Canarias	627	562	549	590	610	566
Cantabria	181	192	183	177	181	192
Castilla y León	872	849	857	876	874	832
Castilla-La Mancha	554	535	534	546	561	537
Cataluña	1.587	1.456	1.492	1.566	1.590	1.446
C. Valenciana	1.622	1.559	1.553	1.636	1.626	1.533
Extremadura	426	406	431	437	431	412
Galicia	884	837	825	839	882	821
Madrid	1.979	1.714	1.745	1.950	1.972	1.698
Murcia	545	557	552	556	559	563
Navarra	219	221	212	228	223	211
País Vasco	526	569	552	559	556	572
La Rioja	117	101	103	115	110	104

Con este número de locales, el código calcula la variación y obtiene el ratio para usarse en el estudio.

7.13 VARIABLE 9: CANTIDADES JUGADAS EN APUESTAS[18]

Dimensión: Intensidad de consumo
Ratio en el que se usó: Cantidades jugadas en Apuestas / cápita
Unidad: millones de € (mill€)

Comunidad Autónoma	2021	2022	2023	2024
Andalucía	187,25	220,63	240,19	267,12

[18] Disponible en: *DG de Ordenación del Juego – Descarga de datos*

Comunidad Autónoma	2021	2022	2023	2024
Aragón	55,87	53,92	58,14	60,65
Asturias	20,03	26,59	26,05	27,13
Illes Balears	4,06	3,59	4,05	3,72
Canarias	95,69	113,51	125,87	130,07
Cantabria	16,94	19,40	22,67	24,66
Castilla y León	63,24	83,37	88,74	95,87
Castilla-La Mancha	63,07	76,73	82,31	101,94
Cataluña	134,11	185,11	202,37	208,59
C. Valenciana	233,70	325,10	339,35	354,49
Extremadura	26,97	34,67	31,61	31,51
Galicia	120,74	142,29	148,73	157,60
Madrid	398,79	423,80	430,44	463,93
Murcia	92,14	94,35	115,96	125,35
Navarra	90,75	109,09	106,55	107,01
País Vasco	367,51	388,24	379,65	420,98
La Rioja	36,45	39,96	40,59	41,01

7.14 VARIABLE 10: CANTIDADES JUGADAS EN LOTERÍAS[19]

Dimensión: Intensidad de consumo
Ratio en el que se usó: Cantidades jugadas en Loterias / cápita
Unidad: millones de € (mill€)

Comunidad Autónoma	2021	2022	2023	2024
Andalucía	2.077,50	2.218,68	2.323,37	2.488,19
Aragón	310,46	316,97	327,42	338,18
Asturias	316,19	330,72	350,35	366,27
Illes Balears	230,13	243,58	252,06	264,76
Canarias	551,22	591,30	613,65	654,02
Cantabria	149,02	157,42	164,98	172,64
Castilla y León	770,01	799,38	821,31	837,68
Castilla-La Mancha	502,83	520,34	536,55	561,02
Cataluña	1.454,83	1.498,59	1.540,31	1.595,13
C. Valenciana	1.364,15	1.426,66	1.471,22	1.571,65
Extremadura	254,06	273,62	283,95	304,27
Galicia	633,92	660,11	700,93	737,21
Madrid	1.447,25	1.503,31	1.544,23	1.622,62
Murcia	362,01	377,63	387,44	406,67

[19] Disponible en: *DG de Ordenación del Juego – Descarga de datos*

Comunidad Autónoma	2021	2022	2023	2024
Navarra	126,90	129,46	132,48	135,04
País Vasco	523,93	530,42	546,41	564,08
La Rioja	78,10	86,88	90,90	93,67

7.15 VARIABLES 11 Y 12: CANTIDADES TOTALES JUGADAS[20]

Dimensión: De dinámica (largo plazo)
Ratio en el que se usó: CAGR (2011 → año actual) de las cantidades totales jugadas
Ratio en el que se usó: Variación anual de las cantidades totales jugadas
Unidad: millones de € (mill€)

Comunidad Autónoma	2011	2021	2022	2023	2024
Andalucía	3980	2.568,11	2.852,98	3.152,13	3.316,01
Aragón	719	432,10	473,63	485,82	503,14
Asturias	522	358,78	396,60	420,55	438,83
Illes Balears	606	234,18	247,16	256,11	268,48
Canarias	1175	813,44	904,30	929,20	1.062,63
Cantabria	283	183,21	205,74	219,33	226,70
Castilla y León	1655	909,22	948,94	966,97	994,19
Castilla-La Mancha	971	608,49	654,76	678,18	723,55
Cataluña	3596	2.079,61	1.925,21	1.992,55	2.056,67
C. Valenciana	2804	1.961,36	2.299,00	2.474,56	2.652,16
Extremadura	501	327,66	373,41	388,02	410,44
Galicia	1022	820,50	905,94	977,88	981,06
Madrid	4474	2.711,28	3.089,44	3.253,05	3.434,27
Murcia	759	517,56	557,78	605,87	670,94
Navarra	390	225,90	250,58	252,80	256,44
País Vasco	1070	1.045,91	1.094,06	1.164,60	1.220,98
La Rioja	171	119,72	135,36	139,94	142,98

7.16 MAGNITUDES BASE

Población[21]

Unidad: habitantes
Uso: denominador de todos los ratios «per cápita» (locales/cápita, sueldos/cápita, inversión/cápita, empresas/cápita, apuestas/cápita, loterías/cápita) y base del PIB per cápita.

[20] Disponible en: *INE – Anuario 2013 (datos 2011)* y en: *DG de Ordenación del Juego (datos 2021-24)*
[21] Disponible en: *INE – Tabla 11807*

Comunidad Autónoma	2021	2022	2023	2024
Andalucía	8.484.804	8.511.167	8.584.147	8.631.862
Aragón	1.331.938	1.328.215	1.341.289	1.351.591
Asturias	1.012.117	1.004.960	1.006.060	1.009.599
Illes Balears	1.183.415	1.187.043	1.209.906	1.231.768
Canarias	2.178.924	2.185.607	2.213.016	2.238.754
Cantabria	584.708	585.450	588.387	590.851
Castilla y León	2.385.223	2.375.583	2.383.703	2.391.682
Castilla-La Mancha	2.052.505	2.058.278	2.084.086	2.104.433
Cataluña	7.749.896	7.761.823	7.901.963	8.012.231
C. Valenciana	5.067.911	5.108.116	5.216.195	5.319.285
Extremadura	1.061.636	1.056.808	1.054.306	1.054.681
Galicia	2.698.177	2.692.825	2.699.424	2.705.833
Madrid	6.726.640	6.743.254	6.871.903	7.009.268
Murcia	1.518.279	1.529.658	1.551.692	1.568.492
Navarra	662.032	664.514	672.155	678.333
País Vasco	2.212.628	2.205.826	2.216.302	2.227.684
La Rioja	319.444	319.617	322.282	324.184

PIB[22]

Unidad: miles de € (mil€)
Uso: denominador del ratio Cifra de negocios/PIB y base del PIB per cápita.

Comunidad Autónoma	2021	2022	2023	2024
Andalucía	79.609.162	86.775.860	93.624.087	98.863.634
Aragón	17.728.783	19.127.404	20.676.129	22.292.111
Asturias	10.985.613	11.756.562	12.705.054	13.728.942
Illes Balears	13.787.462	15.968.531	17.706.873	19.107.809
Canarias	21.558.779	24.337.766	26.404.983	28.532.270
Cantabria	6.457.028	6.947.996	7.419.755	7.993.989
Castilla y León	26.325.332	28.159.905	30.406.557	32.273.730
Castilla-La Mancha	19.360.734	20.819.604	22.487.630	23.774.173
Cataluña	117.168.251	127.061.454	138.012.763	147.908.464
C. Valenciana	55.804.158	60.415.141	65.535.069	70.224.312
Extremadura	9.295.944	10.014.335	10.779.138	11.520.315
Galicia	29.061.032	31.165.065	33.481.458	35.979.262
Madrid	128.704.635	140.365.814	153.741.805	167.170.700
Murcia	16.378.824	17.507.223	18.698.044	20.435.873

[22] Disponible en: INE – Tabla 12955

Comunidad Autónoma	2021	2022	2023	2024
Navarra	9.894.417	10.646.972	11.554.307	12.398.161
País Vasco	35.735.847	38.461.409	41.285.223	43.784.988
La Rioja	3.794.039	4.065.668	4.434.348	4.736.506

Personal ocupado (total de ocupados)[23]

Unidad: miles de personas (datos T4 de cada año, INE)
Uso: denominador del ratio Personal ocupado en JdA&A / total ocupados.

Comunidad Autónoma	2021	2022	2023	2024
Andalucía	3.260,8	3.280,3	3.402,3	3.488,5
Aragón	600,3	600,2	620,5	624,5
Asturias	396,2	379,3	395,7	429,6
Illes Balears	520,6	554,3	564,2	596,9
Canarias	904,0	960,3	986,4	1.023,7
Cantabria	243,2	244,4	256,0	261,1
Castilla y León	989,3	1.019,1	1.013,2	1.041,5
Castilla-La Mancha	872,4	877,6	900,1	911,3
Cataluña	3.555,7	3.595,5	3.800,8	3.855,9
C. Valenciana	2.135,2	2.254,5	2.332,9	2.371,7
Extremadura	397,5	407,6	417,8	417,0
Galicia	1.093,1	1.104,8	1.135,9	1.153,3
Madrid	3.191,5	3.235,1	3.361,5	3.501,1
Murcia	650,8	660,5	678,5	671,9
Navarra	291,1	297,7	302,8	319,0
País Vasco	975,1	967,5	1.024,2	980,8
La Rioja	143,6	148,0	144,4	150,9

7.17 VARIABLES LEGISLATIVAS RECOPILADAS

Finalmente no fueron utilizadas para el estudio, pero se recogen en este anexo porque pueden ser de utilidad para un segundo estudio comparativo entre el IDJ obtenido y un futuro Índice de Presión Legislativa sobre el Juego, calculado con variables legislativas de este tipo. Todos los datos están a fecha de 2024 salvo mención expresa

[23] Disponible en: *INE - Tabla 65321*

Comunidad Autónoma	Años desde última modificación de la ley	Nº modificaciones de la ley hasta la fecha	Año de publicación de la ley marco vigente	Tipo impositivo ganancias apuestas	Tasa anual por operar máquina B (2025)
Andalucía	5	11	1986	10%	3.600,00 €
Aragón	1	11	2000	15%	3.290,00 €
Asturias	2	10	2014	14%	3.500,00 €
Illes Balears	1	3	2014	11%	3.640,00 €
Canarias	0	12	2010	12%	3.600,00 €
Cantabria	2	8	2022	12%	3.600,00 €
Castilla y León	0	14	1998	12%	3.600,00 €
Castilla-La Mancha	0	9	2021	12%	3.700,00 €
Cataluña	3	6	1984	15%	4.020,00 €
C. Valenciana	1	21	2020	20%	900,00 €
Extremadura	5	13	1998	10%	3.400,00 €
Galicia	0	15	2023	10%	3.740,00 €
Madrid	2	11	2001	10%	3.600,00 €
Murcia	6	12	1995	10%	3.620,00 €
Navarra	2	11	2006	20%	3.000,00 €
País Vasco	12	3	1991	10%	3.080,00 €
La Rioja	0	22	2022	10%	3.566,67 €
Fuente	Elaboración propia (consulta BOE)	Elaboración propia (consulta BOE)	BOE	*ANESAR*	*ANESAR*

7.18 CÓDIGO

A continuación, se adjunta el código en lenguaje python detallado, cuya elaboración constituye parte fundamental de este trabajo y ha servido para obtener los resultados del estudio, así como la totalidad de gráficas, tablas y matrices que han servido para dar soporte a la explicación y visualización del trabajo. Además, a lo largo del código se realizan aclaraciones pertinentes que ayudan a su mayor entendimiento y facilitar seguir el transcurso del análisis.

```
"""
TFG - Indicador de Desarrollo del Juego (IDJ) por comunidades autónomas
Pasos 1 y 2: carga de los datos, construcción de los ratios e
imputación de la mortalidad empresarial de 2024.

se introducen a mano las magnitudes ABSOLUTAS de cada comunidad y año
(las que salen de las fuentes: INE y datos.gob.es). A partir de ellas se
calcularán más abajo los ratios que realmente entran en el índice, para
```

```
que el cálculo quede documentado y se pueda replicar sin depender de los
datos provenientes de los archivos Excel descargados de las bases de
datos oficiales.
"""

import pandas as pd
import numpy as np

# Cada fila es una comunidad-año con sus magnitudes absolutas.
# Orden de los campos: ver la lista "columnas" de debajo.
datos = [
    # -- 2021 ---
    ('Andalucía', 2021, 8484804, 79609162, 2789, 1466786, 261697, 57630,
12989, 3260.8, 1994, 88, 2479, 2492, 187.245616, 2077.502613, 3980,
2568.111747),
    ('Aragón', 2021, 1331938, 17728783, 589, 269861, 32201, 10187, 1818,
600.3, 355, 13, 463, 444, 55.871219, 310.464299, 719, 432.102916),
    ('Asturias, Principado de', 2021, 1012117, 10985613, 295, 130174,
27481, 1889, 1589, 396.2, 331, 29, 406, 386, 20.029137, 316.190243, 522,
358.780798),
    ('Balears, Illes', 2021, 1183415, 13787462, 311, 258420, 34758,
25603, 1799, 520.6, 235, 6, 334, 335, 4.0578, 230.125157, 606,
234.182956),
    ('Canarias', 2021, 2178924, 21558779, 780, 568162, 75390, 34869,
4285, 904, 465, 33, 627, 610, 95.689276, 551.22104, 1175, 813.439917),
    ('Cantabria', 2021, 584708, 6457028, 129, 63975, 11545, 578, 691,
243.2, 142, 6, 181, 181, 16.938291, 149.019385, 283, 183.213866),
    ('Castilla y León', 2021, 2385223, 26325332, 1132, 327301, 50832,
4043, 3276, 989.3, 739, 51, 872, 874, 63.241249, 770.010354, 1655,
909.222095),
    ('Castilla - La Mancha', 2021, 2052505, 19360734, 716, 316102, 45860,
11543, 2608, 872.4, 481, 18, 554, 561, 63.067313, 502.83329, 971,
608.494905),
    ('Cataluña', 2021, 7749896, 117168251, 1657, 1308690, 165741, 28415,
8576, 3555.7, 1311, 73, 1587, 1590, 134.114912, 1454.826529, 3596,
2079.605439),
    ('Comunitat Valenciana', 2021, 5067911, 55804158, 1876, 1022669,
154661, 31341, 8360, 2135.2, 1242, 76, 1622, 1626, 233.704348,
1364.149827, 2804, 1961.361926),
    ('Extremadura', 2021, 1061636, 9295944, 268, 157186, 26552, 3227,
1403, 397.5, 344, 23, 426, 431, 26.974077, 254.064423, 501, 327.66148),
    ('Galicia', 2021, 2698177, 29061032, 866, 358640, 52345, 5430, 3293,
1093.1, 775, 64, 884, 882, 120.73624, 633.918736, 1022, 820.49647),
    ('Madrid, Comunidad de', 2021, 6726640, 128704635, 2052, 11020004,
218023, 48846, 10452, 3191.5, 1549, 75, 1979, 1972, 398.788708,
1447.246821, 4474, 2711.281329),
    ('Murcia, Región de', 2021, 1518279, 16378824, 674, 366906, 64186,
12613, 3304, 650.8, 290, 17, 545, 559, 92.14408, 362.009252, 759,
517.559483),
    ('Navarra, Comunidad Foral de Navarra', 2021, 662032, 9894417, 237,
131731, 15662, 3347, 751, 291.1, 155, 5, 219, 223, 90.753967, 126.90137,
390, 225.901172),
    ('País Vasco', 2021, 2212628, 35735847, 586, 595292, 57084, 12853,
2756, 975.1, 413, 17, 526, 556, 367.513929, 523.932523, 1070,
1045.910762),
```

```
    ('Rioja, La', 2021, 319444, 3794039, 66, 52898, 7844, 786, 425,
143.6, 81, 7, 117, 110, 36.447151, 78.102979, 171, 119.716274),
    # --- 2022 ---
    ('Andalucía', 2022, 8511167, 86775860, 2703, 1761429, 301201, 41144,
13176, 3280.3, 1973, 129, 2492, 2475, 220.632365, 2218.677881, 3980,
2852.979791),
    ('Aragón', 2022, 1328215, 19127404, 552, 311609, 32202, 9953, 1581,
600.2, 349, 23, 444, 437, 53.923866, 316.972852, 719, 473.626735),
    ('Asturias, Principado de', 2022, 1004960, 11756562, 465, 173605,
28083, 2487, 1483, 379.3, 305, 30, 386, 357, 26.58616, 330.717033, 522,
396.602947),
    ('Balears, Illes', 2022, 1187043, 15968531, 236, 295785, 40797, 5833,
1697, 554.3, 234, 30, 335, 333, 3.588215, 243.576311, 606, 247.164525),
    ('Canarias', 2022, 2185607, 24337766, 751, 688553, 93863, 17883,
4478, 960.3, 441, 23, 610, 590, 113.512085, 591.297616, 1175, 904.2993),
    ('Cantabria', 2022, 585450, 6947996, 91, 88145, 13591, 3726, 612,
244.4, 136, 11, 181, 177, 19.404752, 157.420898, 283, 205.741467),
    ('Castilla y León', 2022, 2375583, 28159905, 926, 362918, 60794,
5793, 2814, 1019.1, 727, 53, 874, 876, 83.369933, 799.382425, 1655,
948.939798),
    ('Castilla - La Mancha', 2022, 2058278, 20819604, 896, 326557, 51285,
11648, 2671, 877.6, 472, 49, 561, 546, 76.726919, 520.335763, 971,
654.763723),
    ('Cataluña', 2022, 7761823, 127061454, 1855, 1596407, 186737, 31538,
7766, 3595.5, 1295, 106, 1590, 1566, 185.108086, 1498.593817, 3596,
1925.21236),
    ('Comunitat Valenciana', 2022, 5108116, 60415141, 1666, 1280669,
165176, 28328, 8072, 2254.5, 1232, 107, 1626, 1636, 325.103927,
1426.658313, 2804, 2298.99775),
    ('Extremadura', 2022, 1056808, 10014335, 423, 200993, 33224, 3480,
1695, 407.6, 342, 31, 431, 437, 34.666236, 273.622415, 501, 373.41426),
    ('Galicia', 2022, 2692825, 31165065, 1045, 428597, 68665, 5247, 3093,
1104.8, 728, 78, 882, 839, 142.290076, 660.11353, 1022, 905.938275),
    ('Madrid, Comunidad de', 2022, 6743254, 140365814, 1884, 11439580,
282071, 49036, 11024, 3235.1, 1542, 136, 1972, 1950, 423.800834,
1503.307819, 4474, 3089.438599),
    ('Murcia, Región de', 2022, 1529658, 17507223, 723, 508008, 69821,
15124, 3229, 660.5, 280, 14, 559, 556, 94.352002, 377.626178, 759,
557.780964),
    ('Navarra, Comunidad Foral de Navarra', 2022, 664514, 10646972, 84,
142592, 13605, 1484, 629, 297.7, 159, 9, 223, 228, 109.092408,
129.459879, 390, 250.580895),
    ('País Vasco', 2022, 2205826, 38461409, 583, 701108, 61010, 9556,
2495, 967.5, 409, 18, 556, 559, 388.236989, 530.416602, 1070,
1094.058595),
    ('Rioja, La', 2022, 319617, 4065668, 46, 57865, 4662, 1260, 228, 148,
77, 11, 110, 115, 39.962907, 86.875001, 171, 135.362453),
    # --- 2023 ---
    ('Andalucía', 2023, 8584147, 93624087, 2689, 1818930, 321791, 28538,
14004, 3402.3, 1836, 111, 2475, 2352, 240.191617, 2323.373998, 3980,
3152.134714),
    ('Aragón', 2023, 1341289, 20676129, 520, 430573, 41015, 10656, 1659,
620.5, 329, 23, 437, 420, 58.137401, 327.421703, 719, 485.817493),
    ('Asturias, Principado de', 2023, 1006060, 12705054, 314, 205364,
35167, 2688, 1483, 395.7, 284, 21, 357, 337, 26.053179, 350.351557, 522,
420.549296),
```

```
    ('Balears, Illes', 2023, 1209906, 17706873, 280, 350984, 46168, 8379,
1814, 564.2, 209, 11, 333, 313, 4.046977, 252.061806, 606, 256.108784),
    ('Canarias', 2023, 2213016, 26404983, 614, 796530, 101658, 20651,
4458, 986.4, 401, 20, 590, 549, 125.872044, 613.65273, 1175, 929.201548),
    ('Cantabria', 2023, 588387, 7419755, 209, 136421, 15876, 1689, 755,
256, 138, 7, 177, 183, 22.66643, 164.975325, 283, 219.330609),
    ('Castilla y León', 2023, 2383703, 30406557, 1020, 497814, 65538,
7859, 3115, 1013.2, 704, 41, 876, 857, 88.742196, 821.306562, 1655,
966.970541),
    ('Castilla - La Mancha', 2023, 2084086, 22487630, 664, 322601, 48071,
5823, 2368, 900.1, 451, 25, 546, 534, 82.312686, 536.545874, 971,
678.17712),
    ('Cataluña', 2023, 7901963, 138012763, 1587, 2015982, 214798, 39730,
8186, 3800.8, 1213, 117, 1566, 1492, 202.36539, 1540.307073, 3596,
1992.554474),
    ('Comunitat Valenciana', 2023, 5216195, 65535069, 1623, 1262441,
188099, 36658, 8452, 2332.9, 1144, 75, 1636, 1553, 339.350627,
1471.215347, 2804, 2474.560012),
    ('Extremadura', 2023, 1054306, 10779138, 563, 203625, 33573, 2985,
1888, 417.8, 343, 26, 437, 431, 31.614411, 283.951191, 501, 388.023197),
    ('Galicia', 2023, 2699424, 33481458, 933, 506946, 67412, 4939, 3251,
1135.9, 706, 69, 839, 825, 148.731577, 700.932005, 1022, 977.878232),
    ('Madrid, Comunidad de', 2023, 6871903, 153741805, 1943, 11585632,
248225, 57478, 9712, 3361.5, 1339, 87, 1950, 1745, 430.443423,
1544.229364, 4474, 3253.048676),
    ('Murcia, Región de', 2023, 1551692, 18698044, 583, 466052, 79849,
11195, 3353, 678.5, 269, 22, 556, 552, 115.964936, 387.435266, 759,
605.869327),
    ('Navarra, Comunidad Foral de Navarra', 2023, 672155, 11554307, 180,
158690, 18315, 3431, 715, 302.8, 140, 9, 228, 212, 106.553779,
132.481113, 390, 252.797217),
    ('País Vasco', 2023, 2216302, 41285223, 689, 729321, 72186, 13792,
2812, 1024.2, 393, 19, 559, 552, 379.65391, 546.40922, 1070,
1164.599241),
    ('Rioja, La', 2023, 322282, 4434348, 88, 95093, 9457, 1567, 395,
144.4, 70, 10, 115, 103, 40.5868, 90.896263, 171, 139.939122),
    # -- 2024 ---
    ('Andalucía', 2024, 8631862, 98863634, 2781, 2116344, 353748, 43445,
14320, 3488.5, 1823, np.nan, 2352, 2341, 267.117853, 2488.192518, 3980,
3316.00875),
    ('Aragón', 2024, 1351591, 22292111, 476, 304577, 39957, 6267, 1740,
624.5, 318, np.nan, 420, 400, 60.651491, 338.182725, 719, 503.138308),
    ('Asturias, Principado de', 2024, 1009599, 13728942, 331, 214168,
36987, 3283, 1602, 429.6, 274, np.nan, 337, 324, 27.129425, 366.272221,
522, 438.833949),
    ('Balears, Illes', 2024, 1231768, 19107809, 341, 421623, 52990,
19094, 1904, 596.9, 201, np.nan, 313, 302, 3.719617, 264.762965, 606,
268.482582),
    ('Canarias', 2024, 2238754, 28532270, 840, 934597, 110163, 30530,
4849, 1023.7, 410, np.nan, 549, 562, 130.068091, 654.020803, 1175,
1062.631238),
    ('Cantabria', 2024, 590851, 7993989, 141, 92521, 15404, 2096, 707,
261.1, 148, np.nan, 183, 192, 24.656418, 172.644478, 283, 226.697747),
    ('Castilla y León', 2024, 2391682, 32273730, 939, 480228, 72294,
16222, 3126, 1041.5, 689, np.nan, 857, 849, 95.866389, 837.681153, 1655,
994.191096),
```

```
    ('Castilla - La Mancha', 2024, 2104433, 23774173, 644, 349092, 51505,
9712, 2367, 911.3, 451, np.nan, 534, 535, 101.937904, 561.023651, 971,
723.553537),
    ('Cataluña', 2024, 8012231, 147908464, 1574, 1933255, 221603, 86609,
8577, 3855.9, 1162, np.nan, 1492, 1456, 208.587506, 1595.127017, 3596,
2056.672439),
    ('Comunitat Valenciana', 2024, 5319285, 70224312, 1720, 1456595,
208798, 39877, 8811, 2371.7, 1128, np.nan, 1553, 1559, 354.487615,
1571.646622, 2804, 2652.156634),
    ('Extremadura', 2024, 1054681, 11520315, 402, 222542, 38192, 6252,
1655, 417, 331, np.nan, 431, 406, 31.505884, 304.267086, 501,
410.443085),
    ('Galicia', 2024, 2705833, 35979262, 1046, 554145, 71057, 11323,
3347, 1153.3, 708, np.nan, 825, 837, 157.595555, 737.208556, 1022,
981.062199),
    ('Madrid, Comunidad de', 2024, 7009268, 167170700, 1777, 11966129,
257628, 88914, 10072, 3501.1, 1316, np.nan, 1745, 1714, 463.92535,
1622.624636, 4474, 3434.265375),
    ('Murcia, Región de', 2024, 1568492, 20435873, 590, 514696, 85613,
24738, 3525, 671.9, 273, np.nan, 552, 557, 125.350364, 406.670761, 759,
670.939563),
    ('Navarra, Comunidad Foral de Navarra', 2024, 678333, 12398161, 199,
158352, 18832, 3329, 780, 319, 146, np.nan, 212, 221, 107.012989,
135.042576, 390, 256.444609),
    ('País Vasco', 2024, 2227684, 43784988, 527, 1101529, 75134, 8753,
2727, 980.8, 388, np.nan, 552, 569, 420.983477, 564.081078, 1070,
1220.979685),
    ('Rioja, La', 2024, 324184, 4736506, 81, 80898, 11111, 1420, 424,
150.9, 68, np.nan, 103, 101, 41.005043, 93.674063, 171, 142.98094),
]

columnas = ['comunidad', 'anio', 'poblacion', 'pib', 'locales',
'cifra_negocios',
            'sueldos_salarios', 'inversion', 'ocupados_sector',
'ocupados_totales',
            'empresas', 'muertes_empresas', 'locales_serie_t',
'locales_serie_t1',
            'jugado_apuestas', 'jugado_loterias', 'jugado_total_2011',
'jugado_total']

df = pd.DataFrame(datos, columns=columnas)
df = df.sort_values(['comunidad', 'anio']).reset_index(drop=True)

# Total jugado en 2020 por comunidad. Se va a necesitar como año base
para la variación interanual de 2021 (en el panel 2021-2024 no tengo el
año previo).
jugado_total_2020 = {
    'Andalucía': 1961.415764,
    'Aragón': 349.481976,
    'Asturias, Principado de': 286.89814,
    'Balears, Illes': 187.174093,
    'Canarias': 619.018987,
    'Cantabria': 148.474547,
    'Castilla y León': 682.666218,
    'Castilla - La Mancha': 499.96288,
    'Cataluña': 1484.516766,
```

```
    'Comunitat Valenciana': 1392.516015,
    'Extremadura': 271.252198,
    'Galicia': 652.564445,
    'Madrid, Comunidad de': 2104.467537,
    'Murcia, Región de': 394.397485,
    'Navarra, Comunidad Foral de Navarra': 181.171176,
    'País Vasco': 849.820442,
    'Rioja, La': 107.889988,
}

# ------------------------------------------------------------------------
----
# Construcción de los ratios (Nivel 1 de la pirámide de indicadores)
# ------------------------------------------------------------------------
----

# Dimensión económica
df['cn_pib'] = df['cifra_negocios'] / df['pib']              # cifra de
negocios / PIB
df['salarios_pc'] = df['sueldos_salarios'] / df['poblacion']  # sueldos y
salarios per cápita
df['inversion_pc'] = df['inversion'] / df['poblacion']       # inversión
per cápita

# Dimensión de intensidad de consumo
df['apuestas_pc'] = df['jugado_apuestas'] / df['poblacion']   # jugado en
apuestas per cápita
df['loterias_pc'] = df['jugado_loterias'] / df['poblacion']     # jugado
en loterías per cápita

# Dimensión estructural
df['locales_pc'] = df['locales'] / df['poblacion']       # locales per
cápita
df['empleo_sector'] = df['ocupados_sector'] / (df['ocupados_totales'] *
1000)  # ocupados pasados a personas
df['empresas_pc'] = df['empresas'] / df['poblacion']        # empresas
per cápita

# Dimensión de dinámica
# Mortalidad empresarial: muertes / nº de empresas del sector (NaN en
2024, sin datos -> se imputará)
df['mortalidad_emp'] = df['muertes_empresas'] / df['empresas']

# Variación interanual del nº de locales (dimensión de dinámica).
# Diferencia entre el stock de locales del año actual y el del añoa
nterior, dividida entre el stock del año anterior. Valor positivo:
aumento neto de locales; valor negativo: reducción neta de locales. La
orientación es directa porque valores más altos indican mayor
dinamismo/expansión del sector.
df['var_locales'] = (df['locales_serie_t'] - df['locales_serie_t1']) /
df['locales_serie_t1']

# CAGR del total jugado entre 2011 y el año de referencia
df['cagr_jugado'] = (df['jugado_total'] / df['jugado_total_2011']) ** (1
/ (df['anio'] - 2011)) - 1
```

```
# Variación interanual del total jugado. Para 2021 el año previo es 2020;
# para el resto cojo el año anterior dentro del propio panel.
df['jugado_total_prev'] =
df.groupby('comunidad')['jugado_total'].shift(1)
es2021 = df['anio'] == 2021
df.loc[es2021, 'jugado_total_prev'] = df.loc[es2021,
'comunidad'].map(jugado_total_2020)
df['variacion_jugado'] = (df['jugado_total'] - df['jugado_total_prev']) /
df['jugado_total_prev']

# Variable de contexto (no entra en el índice, sirve para interpretar)
df['pib_pc'] = df['pib'] / df['poblacion']

# Lista de los 12 ratios que forman el índice, por si la necesito después
ratios = ['cn_pib', 'salarios_pc', 'inversion_pc', 'apuestas_pc',
'loterias_pc', 'locales_pc', 'empleo_sector', 'empresas_pc',
'mortalidad_emp',
          'var_locales', 'cagr_jugado', 'variacion_jugado']


# ----------------------------------------------------------------------
# Imputación de la mortalidad empresarial de 2024
# ----------------------------------------------------------------------
--
# El INE no publica todavia las muertes de empresas del sector en 2024
(la serie
# llega a 2023), por lo que 'mortalidad_emp' está vacía en 2024. La serie
es corta
# (solo 2021-2023) y la tasa de mortalidad es volátil y sin tendencia
clara, por
# lo que voy a imputar 2024 con la media móvil de los años disponibles de
cada comunidad.

# Marco qué valores son imputados para poder trazarlos después
df['mortalidad_imputada'] = df['mortalidad_emp'].isna()

# Media de la mortalidad de cada comunidad con los años que sí tienen
dato
media_mortalidad = (df[df['anio'] < 2024]
                    .groupby('comunidad')['mortalidad_emp']
                    .mean())

# Relleno los huecos de 2024 con esa media
falta = df['mortalidad_imputada']
df.loc[falta, 'mortalidad_emp'] = df.loc[falta,
'comunidad'].map(media_mortalidad)


#
======================================================================
# Paso 3: analisis de correlaciones y eliminacion de variables
redundantes
# ======================================================================
# Para detectar variables que aportan practicmente la misma informacion
calculo
# la matriz de correlaciones de Pearson entre los 12 ratios, una por anio
```

```
# (2021-2024), y considero redundante toda pareja con |r| > 0,7. Hacerlo
anio a
# anio permite quedarse solo con las redundancias que se repiten y
descartar las que aparecen en un anio suelto (que suelen ser ruido por el
tamanio muestral,
# n = 17 comunidades).

UMBRAL = 0.7
anios = [2021, 2022, 2023, 2024]

# Una matriz de correlaciones por anio
matrices_corr = {a: df[df['anio'] == a][ratios].corr() for a in anios}

# Parejas que superan el umbral, contando en cuantos anios se repite cada
una
def parejas_altas(matriz):
    pares = []
    for i, x in enumerate(ratios):
        for y in ratios[i + 1:]:
            r = matriz.loc[x, y]
            if abs(r) > UMBRAL:
                pares.append((x, y, round(r, 3)))
    return pares

veces_redundante = {}
for a in anios:
    for x, y, r in parejas_altas(matrices_corr[a]):
        veces_redundante[(x, y)] = veces_redundante.get((x, y), 0) + 1

# Considero redundancia "real" la que se repite en la mayoria de los
anios (>=3)
redundancias = {par: n for par, n in veces_redundante.items() if n >= 3}

# Control de robustz con Spearman (menos sensible a valores extremos y a
la
# falta de linealidad), por si Pearson estuviera inflando la correlacion
matrices_spearman = {a: df[df['anio'] == a][ratios].corr('spearman') for
a in anios}

# De la pareja redundante (salarios per capita ~ empleo del sector)
elimino el
# empleo del sector: es la que mas correlaciona de mediaa con el resto de
# variables, y su informacion (la huella laboral del sector) ya queda
recogida
# por los salarios per capita.
variables_descartadas = ['empleo_sector']
variables_indice = [v for v in ratios if v not in variables_descartadas]



#
================================================================================
# Paso 4: normlizacion min-max de las variables del indice
#
================================================================================
==
```

```
# Me llevo las 11 variables a una escala comun [0, 1] con min-max. La
normalizacion
# se hace dentro de cada anio por separado: para cada variable y cada
anio uso el
# minimo y el maximo de las 17 comunidades de ese anio. Asi cada anio
queda
# repartido de 0 a 1 y el ranking de un anio no depende de los niveles de
otros.
#
# El min-max no conoce la orientacion de cada variable. Casi todas van en
el
# sentido "a mas valor, mas desarrollo"; la unica excepcion es la
mortalidad
# empresarial, donde un valor alto significa MENOS desarrollo, asi que la
# normalizo invertida para que, ya normalizada, un 1 indique siempre mas
# desarrollo en todas las variables.

# Variables en las que un valor alto significa MENOS desarrollo (se
invierten)
variables_invertidas = ['mortalidad_emp']

for v in variables_indice:
    grupo = df.groupby('anio')[v]
    vmin = grupo.transform('min')
    vmax = grupo.transform('max')
    if v in variables_invertidas:
        df[v + '_norm'] = (vmax - df[v]) / (vmax - vmin)
    else:
        df[v + '_norm'] = (df[v] - vmin) / (vmax - vmin)

# Lista de las variables ya normalizadas
variables_norm = [v + '_norm' for v in variables_indice]



#
==========================================================================
==
# Paso 5: ponderacion de las variables (tres esquemas)
#
==========================================================================
==
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA

# Variables agrupadas por dimension (ya sin el empleo del sector,
descartado)
dimensiones = {
    'economica':   ['cn_pib', 'salarios_pc', 'inversion_pc'],
    'intensidad':  ['apuestas_pc', 'loterias_pc'],
    'estructural': ['locales_pc', 'empresas_pc'],
    'dinamica':    ['mortalidad_emp', 'var_locales', 'cagr_jugado',
'variacion_jugado'],
}
```

```
# --- 5.1 Pesos iguales ------------------------------------------------
----
# Cada dimension recibe el 25% y ese 25% se reparte a partes iguales
entre las
# variables que contiene (p. ej. una dimension con 2 variables -> 12,5%
cada una).
pesos_iguales = {}
for dim, vars_dim in dimensiones.items():
    peso_var = 0.25 / len(vars_dim)
    for v in vars_dim:
        pesos_iguales[v] = peso_var
pesos_iguales = pd.Series(pesos_iguales).reindex(variables_indice)

# --- 5.2 Pesos de experto ---------------------------------------------
----
# Pesos asignados segun criterio experto. Por dimmension: economica 40%,
# intensidad 25%, estructural 25% y dinamica 10%.
pesos_experto = pd.Series({
    'cn_pib': 0.20, 'salarios_pc': 0.10, 'inversion_pc': 0.10,
    'apuestas_pc': 0.125, 'loterias_pc': 0.125,
    'locales_pc': 0.125, 'empresas_pc': 0.125,
    'mortalidad_emp': 0.025, 'var_locales': 0.025,
    'cagr_jugado': 0.025, 'variacion_jugado': 0.025,
}).reindex(variables_indice)

# --- 5.3 Pesos por PCA ------------------------------------------------
# Pesos derivados de un analisis de componentes principales sobre las
variables
# ya normalizadas (todo el panel, 68observaciones. Por ultimo se
# normalizarán los pesos para que sumen 1.
X = StandardScaler().fit_transform(df[variables_norm].values)
pca = PCA().fit(X)
autovalores = pca.explained_variance_
var_explicada = pca.explained_variance_ratio_

cargas = pca.components_.T * np.sqrt(autovalores)   # cargas factoriales
(loadings)
retenidos = autovalores > 1                          # componentes a
retener
var_ret = var_explicada[retenidos]

cargas2 = cargas[:, retenidos] ** 2                # cargas al cuadrado
comp_dominante = cargas2.argmax(axis=1)              # componente
principal de cada variable
cargas2_norm = cargas2 / cargas2.sum(axis=0)        # reescaladas dentro
de cada componente

pesos_pca = np.array([cargas2_norm[j, comp_dominante[j]] *
var_ret[comp_dominante[j]]
                      for j in range(len(variables_indice))])
pesos_pca = pd.Series(pesos_pca / pesos_pca.sum(),
index=variables_indice)

# Los tres esquemas juntos, para usarlos luego en la agregacion
ponderaciones = {'iguales': pesos_iguales, 'experto': pesos_experto,
'pca': pesos_pca}
```

```
#
================================================================================
==
# Paso 6: agregacion y calculo del IDJ
#
================================================================================
==
# Combino las variables normalizadas con cada vector de pesos mediante
dos
# metodos de agregacion: media aritmetica ponderada y media geometrica
# ponderada. Con 3 esquemas de pesos x 2 metodos salen 6 versiones del
IDJ por
# comunidad y anio.
#
# La media aritmetica trabaja sobre los valores normalizados [0, 1]. La
media
# geometrica necesita valores estrictmente positivos, asi que para ella
# desplazo la escala a [EPS, 1] (con EPS = 0,01) y evito asi los ceros,
que la
# anularian. El desplazamiento solo afecta a la geometrica.

EPS = 0.01

# Matriz de variables normalizadas, con los nombres base para alinear los
pesos
Xnorm = df[variables_norm].copy()
Xnorm.columns = variables_indice
Xshift = EPS + (1 - EPS) * Xnorm        # escala [EPS, 1] para la
geometrica

for nombre, w in ponderaciones.items():
    w = w.reindex(variables_indice)
    # Aritmetica: suma ponderada de las variables normalizadas
    df['IDJ_' + nombre + '_arit'] = (Xnorm * w).sum(axis=1)
    # Geometrica: producto ponderado = exp(suma de w * ln(x))
    df['IDJ_' + nombre + '_geom'] = np.exp((np.log(Xshift) *
w).sum(axis=1))

columnas_idj = ['IDJ_' + n + '_' + m
                for n in ponderaciones for m in ('arit', 'geom')]

# Ranking por anio de cada version del IDJ (1 = mayor desarrollo)
for c in columnas_idj:
    df[c + '_rank'] =
df.groupby('anio')[c].rank(ascending=False).astype(int)



#
================================================================================
==
# Paso 7: analisis de sensibilidad
```

```
#
==========================================================================
==
# Tenemos 6 versiones del IDJ (3 esquemas de pesos x 2 metodos de
agregacion).
# El analisis de sensibilidad mide hasta que punto la posicion de cada
comunidad
# depende de esas decisiones metodologicas, en lugar de juzgarlo a ojo.
Se aplica
# a los cuatro anios para comprobar que la robustez se mantiene en el
tiempo.

import itertools

rank_cols = [c + '_rank' for c in columnas_idj]   # los 6 rankings por
anio

# --- 7.1 Correlacion de rangos (Spearman) entre los 6 escenarios, por
anio ---
# Si los 6 escenarios ordenan a las comunidades de forma parecida, la
# correlacion de Spearman entre sus rankings sera alta y el indice sera
robusto.
robustez_spearman = []
for a in anios:
    rk = df[df['anio'] == a][rank_cols]
    mat = rk.corr('spearman').values
    fuera_diag = mat[np.triu_indices(len(columnas_idj), 1)]
    robustez_spearman.append({'anio': a,
                              'spearman_medio': fuera_diag.mean(),
                              'spearman_min': fuera_diag.min()})
robustez_spearman = pd.DataFrame(robustez_spearman)

# --- 7.2 Desplazamiento medio de rango (ASR) por anio ------------------
------
# Para cada comunidad se toma su rango mediano entre los 6 escenarios y
se mide
# cuanto se aleja de el cada escenario. El promedio (Average Shift in
Rank) dice,
# en numero de posiciones, cuanto se mueve de media una comunidad al
cambiar de
# escenario. Cuanto mas bajo, mas robusto.
asr = []
for a in anios:
    R = df[df['anio'] == a][rank_cols].values
    mediana = np.median(R, axis=1, keepdims=True)
    asr.append({'anio': a, 'asr': np.abs(R - mediana).mean()})
asr = pd.DataFrame(asr)

# --- 7.3 Intervalo de rango por comunidad y anio -----------------------
------
# Mejor y peor posicion que alcanza cada comunidad entre los 6
escenarios.
estabilidad = []
for a in anios:
    sub = df[df['anio'] == a]
    R = sub[rank_cols]
```

```
    tmp = pd.DataFrame({'anio': a, 'comunidad': sub['comunidad'].values,
                        'rank_mediano': R.median(axis=1).values,
                        'mejor_rank': R.min(axis=1).values.astype(int),
                        'peor_rank': R.max(axis=1).values.astype(int)})
    tmp['amplitud'] = tmp['peor_rank'] - tmp['mejor_rank']
    estabilidad.append(tmp)
estabilidad = pd.concat(estabilidad, ignore_index=True)

# --- 7.4 Descomposicion: efecto de la ponderacion vs de la agregacion --
------
# Cuanto mueve el ranking cambiar SOLO los pesos (fijado el metodo)
frente a
# cambiar SOLO el metodo (fijados los pesos).
esquemas_peso = list(ponderaciones)          # iguales, experto, pca
metodos = ['arit', 'geom']
efecto = []
for a in anios:
    sub = df[df['anio'] == a]
    # efecto de los pesos: amplitud de rango entre los 3 esquemas, dentro
de cada metodo
    ef_peso = []
    for m in metodos:
        cols = ['IDJ_' + e + '_' + m + '_rank' for e in esquemas_peso]
        R = sub[cols].values
        ef_peso.append((R.max(1) - R.min(1)).mean())
    # efecto de la agregacion: diferencia de rango entre arit y geom,
dentro de cada esquema
    ef_metodo = []
    for e in esquemas_peso:
        cols = ['IDJ_' + e + '_' + m + '_rank' for m in metodos]
        R = sub[cols].values
        ef_metodo.append(np.abs(R[:, 0] - R[:, 1]).mean())
    efecto.append({'anio': a,
                   'efecto_ponderacion': np.mean(ef_peso),
                   'efecto_agregacion': np.mean(ef_metodo)})
efecto = pd.DataFrame(efecto)


def monte_carlo_incertidumbre(anio, n_iter=2000, semilla=7):
    """Analisis de incertidumbre: pesos aleatorios + metodo de agregacion
al
    azar. Devuelve, por comunidad, el rango mediano y el intervalo de
    confianza del 90% (percentiles 5 y 95) de su posicion."""
    sub = df[df['anio'] == anio]
    Xn = sub[variables_norm].values                # 17 x 11
(normalizadas)
    Xs = EPS + (1 - EPS) * Xn                       # desplazadas para
la geometrica
    rng = np.random.default_rng(semilla)
    W = rng.dirichlet(np.ones(len(variables_indice)), size=n_iter)   #
pesos aleatorios
    usa_geom = rng.integers(0, 2, size=n_iter)       # 0 = aritmetica, 1
= geometrica
    idj_arit = Xn @ W.T                              # 17 x n
    idj_geom = np.exp(np.log(Xs) @ W.T)              # 17 x n
    idj = np.where(usa_geom[None, :] == 1, idj_geom, idj_arit)
```

```
    rangos = (-idj).argsort(0).argsort(0) + 1        # 1 = mayor IDJ
    return pd.DataFrame({'comunidad': sub['comunidad'].values,
                         'rank_mediano': np.median(rangos, axis=1),
                         'p5': np.percentile(rangos, 5, axis=1),
                         'p95': np.percentile(rangos, 95, axis=1)})


if __name__ == '__main__':
    pd.set_option('display.width', 200)
    pd.set_option('display.max_columns', 30)

    # Resumen de redundancias detectadas
    print('Parejas con |r| > 0,7 y nº de años en que se repiten:')
    for (x, y), n in sorted(veces_redundante.items(), key=lambda t: -
t[1]):
        rs = [round(matrices_corr[a].loc[x, y], 2) for a in anios]
        print('  {} ~ {}: {}/4 años  (Pearson por año: {})'.format(x, y,
n, rs))
    print()
    print('Variables descartadas por redundancia:',
variables_descartadas)
    print('Variables finales del índice
({}):'.format(len(variables_indice)))
    for v in variables_indice:
        print('   -', v)

    # Comprobación de la normalización (cada variable, en cada año, debe
ir de 0 a 1)
    print()
    print('Rango de las variables normalizadas por año (min / max):')
    print(df.groupby('anio')[variables_norm].agg(['min',
'max']).round(2).T.to_string())
    print()
    print('Ejemplo: valores normalizados de 2024')
    cols24 = ['comunidad'] + variables_norm
    print(df[df['anio'] == 2024][cols24].round(3).to_string(index=False))

    #Comparación de los tres esquemas de ponderación
    print()
    tabla_pesos = pd.DataFrame(ponderaciones).reindex(variables_indice)
    print('Pesos de cada variable por esquema (%):')
    print((tabla_pesos * 100).round(2).to_string())
    print('Suma de pesos por esquema (%):', (tabla_pesos.sum() *
100).round(1).to_dict())

    # IDJ de 2024 con los 6 esquemas (ordenado por el experto aritmético)
    print()
    t24 = df[df['anio'] == 2024][['comunidad'] + columnas_idj].copy()
    t24 = t24.sort_values('IDJ_experto_arit', ascending=False)
    print('IDJ 2024 (6 versiones):')
    print(t24.round(3).to_string(index=False))

    # ----Resultados del analisis de sensibilidad (todos los años) ----
    print()
    print('7.1 Robustez entre los 6 escenarios (Spearman de rankings) por
año:')
```

```
    print(robustez_spearman.round(3).to_string(index=False))
    print()
    print('7.2 Desplazamiento medio de rango (ASR), en posiciones:')
    print(asr.round(2).to_string(index=False))
    print()
    print('7.4 Efecto de cada decisión (posiciones de rango que mueve):')
    print(efecto.round(2).to_string(index=False))

    # Tabla de rango mediano por comunidad y año (estabilidad temporal))
    print()
    piv = estabilidad.pivot(index='comunidad', columns='anio',
values='rank_mediano')
    piv = piv.sort_values(2024)
    print('Rango mediano por comunidad y año (entre los 6 escenarios):')
    print(piv.to_string())

    # Monte Carlo por año
    mc = {a: monte_carlo_incertidumbre(a) for a in anios}
    print()
    print('Monte Carlo 2024 (rango mediano e intervalo 90% con pesos
aleatorios):')
    m24 = mc[2024].sort_values('rank_mediano')
    print(m24.to_string(index=False))

    # Figura: intervalos de rango Monte Carlo, un panel por año
    import matplotlib.pyplot as plt
    fig, ejes = plt.subplots(2, 2, figsize=(15, 14))
    for ax, a in zip(ejes.flatten(), anios):
        d = mc[a].sort_values('rank_mediano',
ascending=False).reset_index(drop=True)
        y = range(len(d))
        ax.hlines(y, d['p5'], d['p95'], color='steelblue', alpha=0.6,
linewidth=4)
        ax.plot(d['rank_mediano'], y, 'o', color='darkred')
        ax.set_yticks(list(y))
        ax.set_yticklabels(d['comunidad'], fontsize=9)
        ax.set_xlabel('Posición en el ranking (1 = más desarrollo)')
        ax.set_title('Incertidumbre del ranking ' + str(a), fontsize=12)
        ax.invert_xaxis()
        ax.grid(axis='x', alpha=0.3)
    fig.suptitle('Análisis de sensibilidad: intervalos del 90% del
ranking (Monte Carlo)', fontsize=14)
    plt.tight_layout(rect=[0, 0, 1, 0.98])
    plt.savefig('sensibilidad_rangos.png', dpi=150, bbox_inches='tight')
    print('\nGuardado: sensibilidad_rangos.png')

    # Exporto la tabla de rangos medianos por año para el TFG
    piv.to_csv('sensibilidad_rangos_por_anio.csv')
    print('Guardado: sensibilidad_rangos_por_anio.csv')

    # Mapa de calor de la matriz de correlaciones de cada año
    import matplotlib.pyplot as plt

    etiquetas = {'cn_pib': 'CN/PIB', 'salarios_pc': 'Salarios pc',
'inversion_pc': 'Inversión pc',
```

```
                'apuestas_pc': 'Apuestas pc', 'loterias_pc': 'Loterías
pc', 'locales_pc': 'Locales pc',
                'empleo_sector': 'Empleo sector', 'empresas_pc':
'Empresas pc', 'mortalidad_emp': 'Mortalidad',
                'var_locales': 'Var. locales', 'cagr_jugado': 'CAGR',
'variacion_jugado': 'Var. anual'}
    nombres = [etiquetas[v] for v in ratios]

    fig, ejes = plt.subplots(2, 2, figsize=(15, 13))
    for ax, a in zip(ejes.flatten(), anios):
        m = matrices_corr[a].values
        im = ax.imshow(m, cmap='RdBu_r', vmin=-1, vmax=1)
        ax.set_xticks(range(len(ratios)))
        ax.set_yticks(range(len(ratios)))
        ax.set_xticklabels(nombres, rotation=45, ha='right', fontsize=8)
        ax.set_yticklabels(nombres, fontsize=8)
        ax.set_title('Correlaciones {}'.format(a), fontsize=12)
        for i in range(len(ratios)):
            for j in range(len(ratios)):
                ax.text(j, i, '{:.2f}'.format(m[i, j]), ha='center',
va='center',
                        fontsize=6, color='black')
    fig.colorbar(im, ax=ejes, shrink=0.6, label='Coeficiente de
correlación')
    fig.suptitle('Matrices de correlación de los ratios por año',
fontsize=14)
    plt.savefig('matrices_correlacion.png', dpi=150, bbox_inches='tight')
    print('\nGuardado: matrices_correlacion.png')
```

En un entorno diferente y partiendo del código anterior, se ha desarrollado el siguiente código expresamente para la realización de tablas y gráficas del apartado 3.7, Análisis de Sensibilidad.

```
"""
Figuras y tablas de calidad para el apartado de ANÁLISIS DE SENSIBILIDAD.
Aquí se obtienen lad siguientes :
  - Tabla resumen de robustez por año (Spearman medio/mín, ASR,
descomposición)
  - Mapa de calor de la matriz de Spearman 6x6 (2021 vs 2024)
  - Tabla por comunidad 2024 (rango mediano + intervalo de los 6
escenarios + IC90 Monte Carlo)
  - Figura de intervalos Monte Carlo 2024 (principal) y panel 4 años
  - Figura de evolución temporal de la robustez
"""
import os
import numpy as np
import pandas as pd
import matplotlib
matplotlib.use('Agg')
import matplotlib.pyplot as plt
from matplotlib.colors import LinearSegmentedColormap

import sensibilidad_datos_1 as s

OUT = 'salidas_tfg'
```

```
os.makedirs(OUT, exist_ok=True)

# --- estilo general ---
plt.rcParams.update({
    'font.family': 'DejaVu Sans',
    'font.size': 11,
    'axes.titlesize': 13,
    'axes.titleweight': 'bold',
    'axes.edgecolor': '#444444',
    'axes.linewidth': 0.8,
    'figure.facecolor': 'white',
    'axes.facecolor': 'white',
})

TINTA = '#1f2a44'      # azul tinta
ACENTO = '#b3402f'     # rojo teja (puntos)
BANDA = '#5b7aa8'      # azul banda

df = s.df
anios = s.anios
columnas_idj = s.columnas_idj
rank_cols = [c + '_rank' for c in columnas_idj]
estab = s.estabilidad

ESCEN = ['Ig-Arit', 'Ig-Geom', 'Exp-Arit', 'Exp-Geom', 'PCA-Arit', 'PCA-
Geom']

# nombres cortos de comunidad para los gráficos
CORTO = {
    'Andalucía': 'Andalucía', 'Aragón': 'Aragón',
    'Asturias, Principado de': 'Asturias', 'Balears, Illes': 'Baleares',
    'Canarias': 'Canarias', 'Cantabria': 'Cantabria',
    'Castilla y León': 'Castilla y León', 'Castilla - La Mancha':
'Castilla-La Mancha',
    'Cataluña': 'Cataluña', 'Comunitat Valenciana': 'C. Valenciana',
    'Extremadura': 'Extremadura', 'Galicia': 'Galicia',
    'Madrid, Comunidad de': 'Madrid', 'Murcia, Región de': 'Murcia',
    'Navarra, Comunidad Foral de Navarra': 'Navarra', 'País Vasco': 'País
Vasco',
    'Rioja, La': 'La Rioja',
}

# ================================================================
# TABLA 1 - Resumen de robustez por año
#
================================================================================
==
t1 = s.robustez_spearman.merge(s.asr, on='anio').merge(s.efecto,
on='anio')
t1 = t1.rename(columns={
    'spearman_medio': 'Spearman medio',
    'spearman_min': 'Spearman mínimo',
    'asr': 'ASR (posiciones)',
    'efecto_ponderacion': 'Efecto ponderación',
    'efecto_agregacion': 'Efecto agregación',
})
```

```
t1['anio'] = t1['anio'].astype(int)
t1 = t1.set_index('anio').round(3)
t1.to_csv(os.path.join(OUT, 'tabla1_robustez_por_anio.csv'))
print('TABLA 1 - Robustez por año')
print(t1.to_string())
print()

# ================================================================
# TABLA 2 - Por comunidad 2024 (6 escenarios + Monte Carlo)
#
================================================================================
==
mc24 = s.monte_carlo_incertidumbre(2024)
e24 = estab[estab['anio'] == 2024].copy()
t2 = e24.merge(mc24, on='comunidad')
t2['intervalo_6esc'] = t2.apply(lambda r: f"[{int(r['mejor_rank'])}-
{int(r['peor_rank'])}]", axis=1)
t2['IC90_montecarlo'] = t2.apply(lambda r: f"[{int(r['p5'])}-
{int(r['p95'])}]", axis=1)
t2 = t2.sort_values('rank_mediano_x')[
    ['comunidad', 'rank_mediano_x', 'intervalo_6esc', 'amplitud',
'IC90_montecarlo']]
t2 = t2.rename(columns={'rank_mediano_x': 'rango_mediano',
                        'amplitud': 'amplitud_6esc'})
t2.to_csv(os.path.join(OUT, 'tabla2_comunidad_2024.csv'), index=False)
print('TABLA 2 - Por comunidad 2024')
print(t2.to_string(index=False))
print()

#================================================================================
# TABLA 3 - Rango mediano por comunidad y año
# ================================================================
piv = estab.pivot(index='comunidad', columns='anio',
values='rank_mediano').sort_values(2024)
piv.to_csv(os.path.join(OUT, 'tabla3_rango_mediano_por_anio.csv'))
print('TABLA 3 - Rango mediano por comunidad y año')
print(piv.to_string())
print()


# ================================================================
# FIGURA 1 - Mapa de calor de la matriz de Spearman 6x6 (2021 vs 2024)
#
================================================================================
==
def matriz_spearman(anio):
    rk = df[df['anio'] == anio][rank_cols]
    m = rk.corr('spearman').values
    return m

cmap = LinearSegmentedColormap.from_list('robustez', ['#f3f6fb',
'#9db8d8', '#1f2a44'])
fig, axes = plt.subplots(1, 2, figsize=(13, 5.6))
for ax, anio in zip(axes, [2021, 2024]):
    m = matriz_spearman(anio)
    im = ax.imshow(m, cmap=cmap, vmin=0.78, vmax=1.0)
```

```
    ax.set_xticks(range(6)); ax.set_yticks(range(6))
    ax.set_xticklabels(ESCEN, rotation=45, ha='right', fontsize=9)
    ax.set_yticklabels(ESCEN, fontsize=9)
    ax.set_title(f'{anio}', fontsize=13)
    for i in range(6):
        for j in range(6):
            val = m[i, j]
            color = 'white' if val > 0.93 else TINTA
            ax.text(j, i, f'{val:.2f}', ha='center', va='center',
                    fontsize=8.5, color=color)
    ax.set_xticks(np.arange(-.5, 6, 1), minor=True)
    ax.set_yticks(np.arange(-.5, 6, 1), minor=True)
    ax.grid(which='minor', color='white', linewidth=1.4)
    ax.tick_params(which='minor', length=0)
fig.colorbar(im, ax=axes, shrink=0.75, label='Correlación de Spearman
entre rankings', pad=0.02)
fig.suptitle('Concordancia entre los seis escenarios metodológicos del
IDJ',
             fontsize=14, fontweight='bold', y=1.02)
fig.savefig(os.path.join(OUT, 'fig1_spearman_6x6.png'), dpi=160,
bbox_inches='tight')
plt.close(fig)
print('Guardado: fig1_spearman_6x6.png')


#
==========================================================================
==
# FIGURA 2 -Intervalos Monte Carlo 2024
# ============================================================
mc = {a: s.monte_carlo_incertidumbre(a) for a in anios}

d = mc[2024].copy()
d['corto'] = d['comunidad'].map(CORTO)
d = d.sort_values('rank_mediano', ascending=False).reset_index(drop=True)
y = np.arange(len(d))

fig, ax = plt.subplots(figsize=(10, 7.2))
ax.hlines(y, d['p5'], d['p95'], color=BANDA, alpha=0.55, linewidth=7,
          label='Intervalo del 90% (Monte Carlo)')
ax.plot(d['rank_mediano'], y, 'o', color=ACENTO, markersize=8,
        markeredgecolor='white', markeredgewidth=1.2, zorder=3,
        label='Posición mediana')
ax.set_yticks(y)
ax.set_yticklabels(d['corto'], fontsize=10)
ax.set_xlabel('Posición en el ranking  (1 = mayor desarrollo)',
fontsize=11)
ax.set_xticks(range(1, 18))
ax.invert_xaxis()
ax.grid(axis='x', alpha=0.25)
ax.set_axisbelow(True)
ax.set_title('Incertidumbre de la posición de cada comunidad (2024)',
fontsize=14, pad=12)
ax.legend(loc='upper left', frameon=True, framealpha=0.9,
edgecolor='#cccccc', fontsize=10)
for spine in ['top', 'right']:
```

```
    ax.spines[spine].set_visible(False)
fig.tight_layout()
fig.savefig(os.path.join(OUT, 'fig2_montecarlo_2024.png'), dpi=160,
bbox_inches='tight')
plt.close(fig)
print('Guardado: fig2_montecarlo_2024.png')


# ============================================================
# FIGURA 3 - Panel Monte Carlo de los 4 años)
#
============================================================================
==
fig, ejes = plt.subplots(2, 2, figsize=(15, 14))
for ax, a in zip(ejes.flatten(), anios):
    dd = mc[a].copy()
    dd['corto'] = dd['comunidad'].map(CORTO)
    dd = dd.sort_values('rank_mediano',
ascending=False).reset_index(drop=True)
    yy = np.arange(len(dd))
    ax.hlines(yy, dd['p5'], dd['p95'], color=BANDA, alpha=0.55,
linewidth=6)
    ax.plot(dd['rank_mediano'], yy, 'o', color=ACENTO, markersize=6,
            markeredgecolor='white', markeredgewidth=1, zorder=3)
    ax.set_yticks(yy); ax.set_yticklabels(dd['corto'], fontsize=9)
    ax.set_xlabel('Posición (1 = más desarrollo)')
    ax.set_title(f'{a}', fontsize=13)
    ax.set_xticks(range(1, 18, 2))
    ax.invert_xaxis()
    ax.grid(axis='x', alpha=0.25); ax.set_axisbelow(True)
    for spine in ['top', 'right']:
        ax.spines[spine].set_visible(False)
fig.suptitle('Intervalos del 90% del ranking por simulación de Monte
Carlo (2021-2024)',
             fontsize=15, fontweight='bold')
fig.tight_layout(rect=[0, 0, 1, 0.98])
fig.savefig(os.path.join(OUT, 'fig3_montecarlo_panel.png'), dpi=150,
bbox_inches='tight')
plt.close(fig)
print('Guardado: fig3_montecarlo_panel.png')


# ====================================================================
# FIGURA 4- Evolución temporal de la robustez
#
============================================================================
==
fig, (axA, axB) = plt.subplots(1, 2, figsize=(13, 5))

# Panel A:Spearman medio y mínimo
rs = s.robustez_spearman
axA.plot(rs['anio'], rs['spearman_medio'], '-o', color=TINTA,
linewidth=2.2,
         markersize=8, label='Spearman medio')
axA.plot(rs['anio'], rs['spearman_min'], '--o', color=BANDA, linewidth=2,
         markersize=7, label='Spearman mínimo')
axA.set_ylim(0.75, 1.0)
```

```
axA.set_xticks(anios)
axA.set_ylabel('Correlación de Spearman')
axA.set_title('Concordancia entre escenarios', fontsize=12)
axA.legend(frameon=False, fontsize=10, loc='lower right')
axA.grid(alpha=0.25); axA.set_axisbelow(True)
for v, a in zip(rs['spearman_medio'], rs['anio']):
    axA.annotate(f'{v:.2f}', (a, v), textcoords='offset points',
xytext=(0, 9),
                 ha='center', fontsize=9, color=TINTA)

# Panel B: ASR
asr = s.asr
axB.bar(asr['anio'], asr['asr'], color=BANDA, alpha=0.8, width=0.55)
axB.set_xticks(anios)
axB.set_ylabel('Posiciones (ASR)')
axB.set_title('Desplazamiento medio de rango', fontsize=12)
axB.set_ylim(0, 1.3)
for v, a in zip(asr['asr'], asr['anio']):
    axB.annotate(f'{v:.2f}', (a, v), textcoords='offset points',
xytext=(0, 5),
                 ha='center', fontsize=10, color=TINTA,
fontweight='bold')
axB.grid(axis='y', alpha=0.25); axB.set_axisbelow(True)
for ax in (axA, axB):
    for spine in ['top', 'right']:
        ax.spines[spine].set_visible(False)
fig.suptitle('Evolución temporal de la robustez del IDJ (2021-2024)',
             fontsize=14, fontweight='bold')
fig.tight_layout(rect=[0, 0, 1, 0.96])
fig.savefig(os.path.join(OUT, 'fig4_evolucion_robustez.png'), dpi=160,
bbox_inches='tight')
plt.close(fig)
print('Guardado: fig4_evolucion_robustez.png')

print('\nTodo generado en', OUT)
```

Adicionalmente, el código desarrollado para las figuras, tablas y gráficas del apartado 3.7 Resultados, se muestra a continuación:

```
"""
Figuras y tablas del apartado de RESULTADOS

Tomando como ranking de referencia la version PESOS IGUALES + MEDIA
GEOMETRICA
(IDJ_iguales_geom)

Importa el código generado hasta el momento y produce:
  - Tabla A: IDJ de referencia 2024 (puesto, IDJ 0-100, perfil por
dimension)
  - Tabla B: evolucion del ranking 2021-2024 (referencia)
  - Tabla C: las 6 versiones del IDJ en 2024 (transparencia)
  - Tabla D: cercania de cada version al consenso (justificacion de la
eleccion)
```

```
  - Figura 1: ranking del IDJ 2024 (barras horizontales por tramos)
  - Figura 2: perfil por dimension de cada comunidad 2024 (mapa de calor)
  - Figura 3: evolucion del ranking 2021-2024 (bump chart)
"""
import os
import numpy as np
import pandas as pd
import matplotlib
matplotlib.use('Agg')
import matplotlib.pyplot as plt
from matplotlib.colors import LinearSegmentedColormap

import sensibilidad_datos_1 as s

OUT = 'salidas_tfg2'
os.makedirs(OUT, exist_ok=True)

plt.rcParams.update({
    'font.family': 'DejaVu Sans', 'font.size': 11,
    'axes.titlesize': 13, 'axes.titleweight': 'bold',
    'axes.edgecolor': '#444444', 'axes.linewidth': 0.8,
    'figure.facecolor': 'white', 'axes.facecolor': 'white',
})

TINTA = '#1f2a44'
CABEZA = '#2e6b4f'
MEDIA = '#c9a227'
COLA = '#9e3b2e'
BANDA = '#5b7aa8'

df = s.df.copy()
anios = s.anios
dimensiones = s.dimensiones

REF = 'IDJ_iguales_geom'
RANKREF = REF + '_rank'

CORTO = {
    'Andalucía': 'Andalucía', 'Aragón': 'Aragón',
    'Asturias, Principado de': 'Asturias', 'Balears, Illes': 'Baleares',
    'Canarias': 'Canarias', 'Cantabria': 'Cantabria',
    'Castilla y León': 'Castilla y León', 'Castilla - La Mancha':
'Castilla-La Mancha',
    'Cataluña': 'Cataluña', 'Comunitat Valenciana': 'C. Valenciana',
    'Extremadura': 'Extremadura', 'Galicia': 'Galicia',
    'Madrid, Comunidad de': 'Madrid', 'Murcia, Región de': 'Murcia',
    'Navarra, Comunidad Foral de Navarra': 'Navarra', 'País Vasco': 'País
Vasco',
    'Rioja, La': 'La Rioja',
}
df['corto'] = df['comunidad'].map(CORTO)

# Escala 0-100 del IDJ de referencia (solo cambia la escala, no el orden)
df['IDJ100'] = (df[REF] * 100).round(1)
```

```
# Perfil por dimension: media de las variables normalizadas de cada
dimension
DIM_NOMBRE = {'economica': 'Económica', 'intensidad': 'Intensidad',
              'estructural': 'Estructural', 'dinamica': 'Dinámica'}
for dim, vars_dim in dimensiones.items():
    cols = [v + '_norm' for v in vars_dim]
    df['dim_' + dim] = df[cols].mean(axis=1)

dim_cols = ['dim_' + d for d in dimensiones]


# ============================================================
# TABLA A - IDJ de referencia 2024 con perfil por dimension
# ============================================================================
a24 = df[df['anio'] == 2024].sort_values(REF, ascending=False).copy()
tA = a24[['corto', RANKREF, 'IDJ100'] + dim_cols].copy()
tA[dim_cols] = (tA[dim_cols] * 100).round(0).astype(int)
tA = tA.rename(columns={'corto': 'Comunidad', RANKREF: 'Puesto',
'IDJ100': 'IDJ (0-100)',
                        'dim_economica': 'Económica', 'dim_intensidad':
'Intensidad',
                        'dim_estructural': 'Estructural', 'dim_dinamica':
'Dinámica'})
tA.to_csv(os.path.join(OUT, 'tablaA_idj_referencia_2024.csv'),
index=False)
print('TABLA A - IDJ de referencia 2024 (perfil por dimensión, 0-100)')
print(tA.to_string(index=False))
print()

#
============================================================================
# TABLA B - Evolucion del ranking 2021-2024 (referencia)
# ============================================================
piv_rank = df.pivot(index='corto', columns='anio',
values=RANKREF).astype(int)
piv_rank = piv_rank.sort_values(2024)
piv_idj = df.pivot(index='corto', columns='anio', values='IDJ100')
tB = piv_rank.copy()
tB.columns = [f'Puesto {a}' for a in tB.columns]
for a in anios:
    tB[f'IDJ {a}'] = piv_idj[a].round(1)
tB = tB[[f'Puesto {a}' for a in anios] + [f'IDJ {a}' for a in anios]]
tB.index.name = 'Comunidad'
tB.to_csv(os.path.join(OUT, 'tablaB_ranking_evolucion.csv'))
print('TABLA B - Evolución del ranking y del IDJ 2021-2024')
print(tB.to_string())
print()

#
============================================================================
==
# TABLA C - Las 6 versiones del IDJ en 2024 (rank de cada una)
# ================================================================
etiqueta = {'IDJ_iguales_arit': 'Ig-Arit', 'IDJ_iguales_geom': 'Ig-Geom',
            'IDJ_experto_arit': 'Exp-Arit', 'IDJ_experto_geom': 'Exp-
Geom',
```

```
            'IDJ_pca_arit': 'PCA-Arit', 'IDJ_pca_geom': 'PCA-Geom'}
cols_rank = [c + '_rank' for c in s.columnas_idj]
tC = df[df['anio'] == 2024][['corto'] + cols_rank].copy()
tC.columns = ['Comunidad'] + [etiqueta[c.replace('_rank', '')] for c in
cols_rank]
tC = tC.sort_values('Ig-Geom')
tC.to_csv(os.path.join(OUT, 'tablaC_seis_versiones_2024.csv'),
index=False)
print('TABLA C - Puesto de cada comunidad en las 6 versiones (2024)')
print(tC.to_string(index=False))
print()

#
================================================================================
==
# TABLA D - Cercania de cada version al consenso (justificacion)
#
================================================================================
==
filas = []
for a in anios:
    sub = df[df['anio'] ==a]
    R = sub[cols_rank].values
    consenso = pd.Series(np.median(R,
axis=1)).rank(method='average').values
    fila = {'Año': a}
    for c in s.columnas_idj:
        r = sub[c + '_rank'].values
        fila[etiqueta[c]] = round(pd.Series(r).corr(pd.Series(consenso),
method='spearman'), 3)
    filas.append(fila)
tD = pd.DataFrame(filas).set_index('Año')
tD.loc['Media'] = tD.mean().round(3)
tD.to_csv(os.path.join(OUT, 'tablaD_cercania_consenso.csv'))
print('TABLA D - Cercanía de cada versión al ranking de consenso
(Spearman)')
print(tD.to_string())
print()


# ================================================================
# FIGURA 1 Ranking del IDJ 2024 (figura d barras horizontales por tramos)
#
================================================================================
==
d = a24.copy().reset_index(drop=True)
# tramos: cabeza (1-5), media (6-12), cola (13-17)
def color_tramo(p):
    if p <= 5: return CABEZA
    if p <= 12: return MEDIA
    return COLA
colores = [color_tramo(p) for p in d[RANKREF]]

fig, ax = plt.subplots(figsize=(10, 7.4))
y = np.arange(len(d))[::-1]
ax.barh(y, d['IDJ100'], color=colores, alpha=0.9, height=0.66)
```

```
for yi, idj, com in zip(y, d['IDJ100'], d['corto']):
    ax.text(idj + 0.8, yi, f'{idj:.1f}', va='center', ha='left',
fontsize=9, color=TINTA)
ax.set_yticks(y)
ax.set_yticklabels([f'{p}.  {c}' for p, c in zip(d[RANKREF],
d['corto'])], fontsize=10)
ax.set_xlabel('Índice de Desarrollo del Juego (IDJ), escala 0-100',
fontsize=11)
ax.set_xlim(0, max(d['IDJ100']) + 6)
ax.set_title('Ranking del IDJ por comunidad autónoma (2024)',
fontsize=14, pad=12)
ax.grid(axis='x', alpha=0.25); ax.set_axisbelow(True)
for spine in ['top', 'right']:
    ax.spines[spine].set_visible(False)
# leyenda de tramos
from matplotlib.patches import Patch
leg = [Patch(facecolor=CABEZA, label='Cabeza (1-5)'),
       Patch(facecolor=MEDIA, label='Zona media (6-12)'),
       Patch(facecolor=COLA, label='Cola (13-17)')]
ax.legend(handles=leg, loc='lower right', frameon=True, framealpha=0.9,
          edgecolor='#cccccc', fontsize=9.5)
fig.tight_layout()
fig.savefig(os.path.join(OUT, 'res_fig1_ranking_2024.png'), dpi=160,
bbox_inches='tight')
plt.close(fig)
print('Guardado: res_fig1_ranking_2024.png')


#
==========================================================================
==
# FIGURA 2 -perfil por dimension 2024 (mapa de calor)
# ==========================================================================
prof = a24[['corto'] + dim_cols].set_index('corto')
prof.columns = [DIM_NOMBRE[d.replace('dim_', '')] for d in dim_cols]
cmap = LinearSegmentedColormap.from_list('perfil', ['#f6f4ee', '#d8c98a',
'#2e6b4f'])

fig, ax = plt.subplots(figsize=(8.2, 8.6))
im = ax.imshow(prof.values, cmap=cmap, vmin=0, vmax=1, aspect='auto')
ax.set_xticks(range(prof.shape[1])); ax.set_xticklabels(prof.columns,
fontsize=10.5)
ax.set_yticks(range(prof.shape[0]))
ax.set_yticklabels([f'{p}.  {c}' for p, c in zip(a24[RANKREF],
prof.index)], fontsize=9.5)
for i in range(prof.shape[0]):
    for j in range(prof.shape[1]):
        val = prof.values[i, j]
        ax.text(j, i, f'{val*100:.0f}', ha='center', va='center',
fontsize=8.5,
                color='white' if val > 0.55 else TINTA)
ax.set_xticks(np.arange(-.5, prof.shape[1], 1), minor=True)
ax.set_yticks(np.arange(-.5, prof.shape[0], 1), minor=True)
ax.grid(which='minor', color='white', linewidth=1.4)
ax.tick_params(which='minor', length=0)
```

```
fig.colorbar(im, ax=ax, shrink=0.5, label='Puntuación de la dimensión (0-
100)', pad=0.02)
ax.set_title('Perfil por dimensión de cada comunidad (2024)',
fontsize=13.5, pad=12)
fig.tight_layout()
fig.savefig(os.path.join(OUT, 'res_fig2_perfil_dimensiones.png'),
dpi=160, bbox_inches='tight')
plt.close(fig)
print('Guardado: res_fig2_perfil_dimensiones.png')


# ================================================================
# FIGURA 3 - Evolucion del ranking 2021-2024
# ================================================================
rk = df.pivot(index='corto', columns='anio', values=RANKREF)
# comunidades a destacar por su movimiento
destacar = {'Extremadura': '#2e6b4f', 'Castilla-La Mancha': '#9e3b2e',
            'Navarra': '#7a4ea8', 'Madrid': '#3b5b9e', 'Canarias':
'#d2691e',
            'Cataluña': '#33373d'}
fig, ax = plt.subplots(figsize=(11, 8))
for com in rk.index:
    ys = rk.loc[com, anios].values
    if com in destacar:
        ax.plot(anios, ys, '-o', color=destacar[com], linewidth=2.6,
markersize=8,
                zorder=3, markeredgecolor='white', markeredgewidth=1.2)
        ax.text(anios[0] - 0.08, ys[0], f'{int(ys[0])}. {com}',
ha='right', va='center',
                fontsize=9.5, color=destacar[com], fontweight='bold')
        ax.text(anios[-1] + 0.08, ys[-1], f'{com} ({int(ys[-1])})',
ha='left', va='center',
                fontsize=9.5, color=destacar[com], fontweight='bold')
    else:
        ax.plot(anios, ys, '-o', color='#c8ccd4', linewidth=1.4,
markersize=5, zorder=1)
        ax.text(anios[-1] + 0.08, ys[-1], com, ha='left', va='center',
                fontsize=8.5, color='#9aa0aa')
ax.set_yticks(range(1, 18))
ax.invert_yaxis()
ax.set_xticks(anios)
ax.set_xlim(anios[0] - 1.1, anios[-1] + 1.4)
ax.set_ylabel('Puesto en el ranking  (1 = mayor desarrollo)')
ax.set_title('Evolución del ranking del IDJ (2021-2024)', fontsize=14,
pad=12)
ax.grid(axis='y', alpha=0.25); ax.set_axisbelow(True)
for spine in ['top', 'right']:
    ax.spines[spine].set_visible(False)
fig.tight_layout()
fig.savefig(os.path.join(OUT, 'res_fig3_evolucion_ranking.png'), dpi=160,
bbox_inches='tight')
plt.close(fig)
print('Guardado: res_fig3_evolucion_ranking.png')
```