



Indicador sintético sobre la actividad empresarial española

Facultad de Ciencias Económicas y Empresariales

Autora:
Raquel Del Río Torralbo
5º Doble grado de Derecho y Business Analytics
Director:
Prof. David Roch Dupré
Universidad Pontificia Comillas
Madrid, octubre de 2022

Dedicado a mi padre, gracias por ayudarme a encontrar mi camino y ser mi modelo a seguir

Resumen

Los indicadores sintéticos se han convertido en herramientas claves para poder analizar temas complejos que en el pasado requerían de múltiples recursos. La actividad económica, con sus múltiples fluctuaciones en estos últimos años, se convierte en un sector que puede beneficiarse en gran medida de un análisis capaz de extraer tendencias y puntos de inflexión. A través de este trabajo, se desarrollará un indicador compuesto en el que se analizará la actividad económica española en estos últimos años. El uso de herramientas como el SPSS y Matlab permitirán la creación de un indicador fiable y provechoso para aquellos que estudian la actividad económica y pretenden desarrollar legislaciones.

En este trabajo se ha seguido una metodología para la construcción del indicador compuesta, que consta de los siguientes pasos: la construcción de un marco teórico, un análisis de los datos, una imputación de la información ausente la normalización, ponderación y agregación de los datos y, finalmente un análisis de robustez. Todo ello ha permitido presentar un indicador sintético desde el que se han extraído tendencias y conclusiones.

Abstract

Synthetic indicators have become key tools for analyzing complex issues that in the past required multiple resources. Economic activity, with its multiple fluctuations in recent years, has become a sector that can benefit greatly from an analysis capable of extracting trends and turning points. Through this work, a composite indicator will be developed to analyze Spanish economic activity in recent years. The use of tools such as SPSS and MATLAB will allow the creation of a reliable and useful indicator for those who study economic activity and intend to develop legislation.

This project has followed a methodology for the construction of the composite indicator, which consists of the following steps: the construction of a theoretical framework, analysis of the data, imputation of the missing information, normalization, weighting and aggregation of the data and, finally, a robustness analysis. All this enabled the presentation of a synthetic indicator from which trends and conclusions were drawn.

Palabras Claves

Indicador sintético, imputación, actividad económica, análisis multivariante, normalización, ponderación, agregación.

Key Words

Synthetic indicator, imputation, economic activity, multivariate analysis, normalization, weighting, aggregation.

Índice

Índice de siglas.....	5
Índice de figuras.....	9
Índice de tablas	11
Índice de ecuaciones	12
1. Introducción	13
1.1. Introducción a la actividad empresarial española	13
1.2. Introducción a los indicadores sintéticos	16
1.3. Motivación	19
1.4. Objetivos	20
1.5. Recursos	21
1.6. Metodología	22
2. Estado del arte y construcción del indicador.....	23
2.1. Marco Teórico.....	23
2.2. Imputación de los datos.....	29
2.3. Análisis Multivariante.....	45
2.4. Normalización.....	73
2.5. Ponderación y agregación	81
2.6. Análisis de robustez	99
3. Conclusiones	103
4. Bibliografía	104
Anexos	111
Anexo 1.....	111
Anexo 2.....	111
Anexo 3.....	111
Anexo 4.....	112
Anexo 5.....	112
Anexo 6.....	112
Anexo 7.....	112
Anexo 8.....	112
Anexo 9.....	112
Anexo 10.....	113
Anexo 11.....	113

Índice de siglas

CCA	Consumo de cemento absoluto
CCB	Consumo de cemento_Base 100
CCVI	Consumo de cemento_Variación interanual
CLI	Composite leading indicator
CVEA	Crecimiento del volumen de las exportaciones_Absoluto
CVEB	Crecimiento del volumen de las exportaciones_Base 100
CVEVI	Crecimiento del volumen de las exportaciones_Variación Interanual
CVESA	Crecimiento del volumen de las exportaciones en España por Sector_Absoluto
CVEESB	Crecimiento del volumen de las exportaciones en España por Sector_Base 100
CVEESV	Crecimiento del volumen de las exportaciones en España por Sector_Variación interanual
CVEVI	Crecimiento del volumen de las exportaciones_Variación interanual
CVIA	Crecimiento del volumen de las importaciones_Absoluto
CVIB	Crecimiento del volumen de las importaciones_Base 100
CVIESA	Crecimiento del volumen de las importaciones en España por sector_Absoluto
CVIESB	Crecimiento del volumen de las importaciones en España por sector_Base 100
CVIESVI	Crecimiento del volumen de las importaciones en España por sector_Variación interanual
CVIVI	Crecimiento del volumen de las importaciones_Variación Interanual
DEEA	Demanda eléctrica en establecimientos de alojamiento
DEEAVI	Demanda eléctrica en establecimientos de alojamiento_Variación interanual

DEI	Demanda eléctrica en industria
DEIVI	Demanda eléctrica en industria_Variacion interanual
DES	Demanda eléctrica en servicios
DESVI	Demanda eléctrica en servicios_Variacion interanual
EISSA	Empresas inscritas en la Seguridad Social_Absoluto
EISSVI	Empresas inscritas en la Seguridad Social_Variacion interanual
ESCE	Evolución del stock de crédito – Empresas
ESCH	Evolución del stock de crédito – Hogares
FI	Facturación por industria
FICMMRVMM	Facturación por industria - Comercio al por mayor y al por menor; reparación de vehículos de motor y motocicletas
FIEM	Facturación por industria - Industrias extractivas y manufactureras
FISNM	Facturación por industria - Servicios no financieros de mercado
FISEEASGR	Facturación por industria - Suministro de energía eléctrica y agua, saneamiento y gestión de residuos
IAM	Indicador agregado de manufacturing
IASS	Indicadores de actividad del sector servicios
IASSAASA	Indicadores de actividad del sector servicios - Actividades administrativas y servicios auxiliares
IASSAPTC	Indicadores de actividad del sector servicios - Actividades profesionales técnicas y científicas
IASSCma	Indicadores de actividad del sector servicios - Comercio al por mayor
IASSCme	Indicadores de actividad del sector servicios - Comercio al por menor
IASSH	Indicadores de actividad del sector servicios - Hostelería
IASSIC	Indicadores de actividad del sector servicios - Información y comunicaciones

IASSRV	Indicadores de actividad del sector servicios - Reparación de vehículos
IASSTA	Indicadores de actividad del sector servicios - Transporte y almacenamiento
ICU	Índice de capacidad de uso
IGDE	Índice general de demanda eléctrica
IGDEVI	Índice general de demanda eléctrica_Variacion interanual
INE	Instituto Nacional de Estadística
IPI	Industrial Production Index
IPC_sb	IPC subyacente
IPRIM_VA	IPRIM_Variacion anual
IPRIM_VI	IPRIM_Variacion interanual
IPRIX_VA	IPRIX_Variacion anual
IPRIX_VI	IPRIX_Variacion interanual
ISE	Indicador de Sentimiento Económico
MAR	Missing at Random
MCAR	Missing Completely at Random
NMAR	Not Missing at Random
PCA	Análisis de los componentes principales
PIB_AA	PIB agregado_Absoluto
PIB_AB	PIB agregado_Base 100
PIB_AVI	PIB agregado_Variación interanual
PIB_BdE_Neutra	PIB previsión BdE-BCE neutra
PIB_BdE_Optim	PIB previsión BdE-BCE optimista
PIB_BdE_Pesim	PIB previsión BdE-BCE pesimista
PIB_VR	PIB_ValorReal
PIB_Neutro_Absol	Test_PIB previsión neutra_Absoluto

PIB_Neutro_100	Test_PIB previsión neutra_Base 100
PIB_Neutro_VI	Test_PIB previsión neutra_Variación Interanual
PIB_Optim_Absol	Test_PIB previsión optimista_Absoluto
PIB_Optim_100	Test_PIB previsión optimista_Base 100
PIB_Pesim_Absol	Test_PIB previsión pesimista_Absoluto
PIB_Pesim_100	Test_PIB previsión pesimista_Base 100
PIB_Pesim_VI	Test_PIB previsión pesimista_Variación Interanual
PTF	Productividad Total de los Factores
PMIA	Purchasing Manager's Index (PMI) of Aggregated
PMIM	Purchasing Manager's Index (PMI) of Manuf
PMISI	Purchasing Manager's Index (PMI) of Services Índice de gestores de compras
RNE	Registro de nuevas empresas
RNEVI	Registro de nuevas empresas Variación Interanual
RQE	Registro de quiebras empresas
RQEVI	Registro de quiebras empresas Variación Interanual
SA	Sector agrario
SAVI	Sector agrario - Variación interanual
SCL	Salarios y costes laborales
TMB	Tasa de morosidad bancaria
UEM	Unión Económica y Monetaria
VON	Visados de Obra Nueva
VONA	Visados de obra nueva_Absoluto
VONVI	Visados de obra nueva_Variación interanual

Índice de figuras

Figura 1: “Distribución de empresas por tamaño en España en 2022”	14
Figura 2: “Evolución del número de concursos, por forma jurídica 2012-2022”	14
Figura 3: “Metodología seguida en la construcción del indicador”	22
Figura 4: “Matriz de distancia de los datos de la categoría Actividad Económica”	50
Figura 5: “Matriz de distancia de los datos de la categoría Actividad Económica Sectorial”	51
Figura 6: “Matriz de distancia de los datos de la categoría Salud y Liquidez Financiera”	52
Figura 7: “Dendograma con los clústeres de la categoría Actividad Económica”	53
Figura 8: “Dendograma con los clústeres de la categoría Actividad Económica Sectorial”	53
Figura 9: “Dendograma con los clústeres de la categoría Actividad Económica Sectorial”	54
Figura 10: “Número óptimo de clústeres: Método del codo. Categoría: Actividad Económica”	55
Figura 11: “Número óptimo de clústeres: Método del codo. Categoría: Actividad Económica Sectorial”	55
Figura 12: “Número óptimo de clústeres: Método del codo. Categoría: Salud y Liquidez Financiera”	56
Figura 13: “Número óptimo de clústeres: Método de la silueta. Categoría: Actividad Económica Sectorial”	56
Figura 14: “Número óptimo de clústeres: Método de la silueta. Categoría: Actividad Económica Sectorial”	57
Figura 15: “Número óptimo de clústeres: Método de la silueta. Categoría: Salud y Liquidez Financiera”	57
Figura 16: “Representación gráfica de los clústeres de la categoría: Actividad Económica”	58
Figura 17: “Representación gráfica de las variables en los clústeres de la categoría: Actividad Económica”	59
Figura 18: “Representación gráfica de los clústeres de la categoría: Actividad Económica Sectorial”	60
Figura 19: “Representación gráfica de las variables en los clústeres de la categoría: Actividad Económica Sectorial”	60
Figura 20: “Representación gráfica de los clústeres de la categoría: Liquidez y Salud Financiera”	61
Figura 21: “Representación gráfica de las variables en los clústeres de la categoría: Liquidez y Salud Financiera”	62
Figura 22: “Muestra de la matriz de correlación de la categoría: Actividad Económica”	64
Figura 23: “Muestra de la matriz de correlación de la categoría: Actividad Económica Sectorial”	64
Figura 24: “Muestra de la matriz de correlación de la categoría: Liquidez y Salud Financiera”	65
Figura 25: “Análisis Multivariante categoría: Actividad Económica”	66
Figura 26: “Análisis Multivariante categoría: Actividad Económica Sectorial”	67
Figura 27: “Análisis Multivariante categoría: Actividad Económica Sectorial”	67
Figura 28: “Matriz Rotada: Actividad Económica”	68
Figura 29: “Matriz Rotada: Actividad Económica Sectorial”	69
Figura 30: “Matriz Rotada: Liquidez y Salud Financiera”	69
Figura 31: “Escala a la unidad: Actividad Económica”	70
Figura 32: “Escala a la unidad: Actividad Económica Sectorial”	70
Figura 33: “Escala a la unidad: Liquidez y Salud Financiera”	71
Figura 34: “Muestra de la Normalización: Actividad Económica”	81
Figura 35: “Ponderación: Actividad Económica”	90
Figura 36: “Ponderación: Actividad Económica Sectorial”	90
Figura 37: “Ponderación: Actividad Económica Sectorial”	91

Figura 38: “Agregación Aditiva: Actividad Económica”	95
Figura 39: “Agregación Aditiva: Actividad Económica Sectorial”	96
Figura 40: “Agregación Aditiva: Liquidez y Salud Financiera”	97
Figura 41: “Agregación Final”	98
Figura 42: “Análisis de Robustez”	100
Figura 43: “Comparativa entre las diferentes opciones del análisis de Robustez”	101
Figura 44: : “Comparativa entre las opciones (0100), (0110) y (0010)	102

Índice de tablas

Tabla 1: “Beneficios y desventajas propias del uso de indicadores sintéticos”	17
Tabla 2: “Beneficios y desventajas de las distintas técnicas de imputación”	33
Tabla 3: “Análisis de la imputación del indicador FICMMRVMM ”	39
Tabla 4 “Análisis de la imputación del indicador FIIEM”	40
Tabla 5: “Análisis de la imputación del indicador IASSCMe”	40
Tabla 6: “Análisis de la imputación del indicador IASSRV”	40
Tabla 7: “Análisis de la imputación del indicador CVEESB”	41
Tabla 8: “Análisis de la imputación del indicador CVEESV”	41
Tabla 9: “Análisis de la imputación del indicador CVIESB”	42
Tabla 10: “Análisis de la imputación del indicador CVIESVI”	42
Tabla 11: “Análisis de la imputación del indicador IPI”	43
Tabla 12: “Análisis de la imputación del indicador IPRIM_VA”	43
Tabla 13: “Análisis de la imputación del indicador IPRIX_VA”	44
Tabla 14: “Análisis de la imputación del indicador IAM”	44
Tabla 15: “Beneficios y desventajas de las distintas técnicas de análisis multivariante”	46
Tabla 16: “Número óptimo de Clústeres en la categoría: Actividad Económica”	58
Tabla 17: “Número óptimo de Clústeres en la categoría: Actividad Económica Sectorial”	58
Tabla 18: “Número óptimo de Clústeres en la categoría: Liquidez y Salud Financiera”	58
Tabla 19: “Código de colores para las matrices de correlación”	63
Tabla 20: “Indicadores Actividad Económica”	71
Tabla 21: “Indicadores Actividad Económica”	72
Tabla 22: “Indicadores Actividad Económica”	72
Tabla 23: “Beneficios y desventajas de las distintas técnicas de normalización”	76
Tabla 24: “Escalas de normalización Actividad Económica”	80
Tabla 25: “Escalas de normalización Liquidez y Salud Financiera”	80
Tabla 26: “Código de colores: normalizaciones”	81
Tabla 27: “Beneficios y desventajas de las distintas técnicas de ponderación”	85
Tabla 28: “Clasificación de los indicadores tras normalizar. Categoría: Actividad Económica”	88
Tabla 29: “Clasificación de los indicadores tras normalizar. Categoría: Actividad Económica”	89
Tabla 30: “Beneficios y desventajas de las distintas técnicas de agregación”	93

Índice de ecuaciones

(1)	$PIB = C + I + G + X - M$	15
(2)	$RMSE = \frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2$	36
(3)	$MAE = \frac{1}{N} \sum_{i=1}^N P_i - O_i $	36
(4)	$Z = x - \mu\theta$	73
(5)	$Z = x_i - \min(x) \text{max}x - \min(x)$	73
(6)	$y = x_i - x_{rxr}$	74
(7)	$I_{qct} = X_{qct} - E_t * (X_{qct}) E_t * (X_{qct} - E_t * X_{qct})$	74
(8)	$y = x_i - x_i - 1x_i - 1 * 100$	75
(9)	$Z = 1 - x_i - \min(x) \text{max}x - \min(x)$	79
(10)	$w_j = \phi_j \forall j \phi_j = \lambda_{ij}^2 \forall i (\lambda_{ij})^2$	89
(11)	$CIC = q = \frac{1}{Q} w_q * I_{qc}$	91
(12)	$CIC = q = \frac{1}{Q} x_{qc} w_q$	92

1. Introducción

1.1. Introducción a la actividad empresarial española

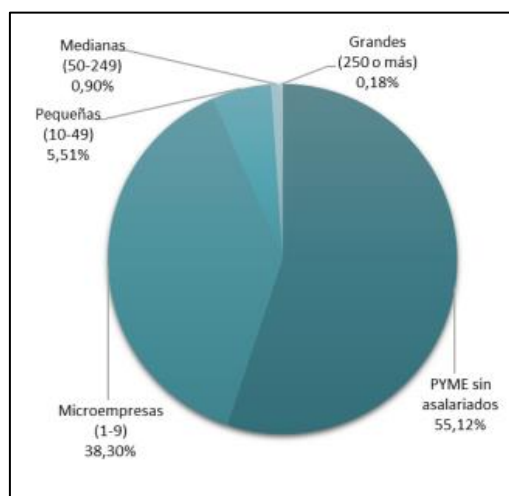
“La crisis derivada de la pandemia del Covid-19 ha generado una disrupción de la actividad económica de una magnitud elevadísima (...) con todas las previsiones disponibles, este año se producirán, probablemente, los retrocesos del PIB más abultados que se hayan registrado en tiempos de paz” (Hernández de Cos,2020), con esta declaración comenzaba el gobernador del Banco de España en su comparecencia ante la Comisión de Reconstrucción Social y Económica del Congreso de los Diputados de 2020.

Durante esta intervención el gobernador reflejó los principales retos a los que se iba a enfrentar España durante los años posteriores a la crisis del COVID 19. No obstante, la actividad económica ya venía lastrada por la etapa de recuperación previa a la pandemia, según los datos ofrecidos por la Comisión Europea, entre el 2013 y el 2019 el PTF español creció a un 0.4% frente al 0.5% de la UEM.

La situación económica ha complicado el estado de la actividad empresarial española, de media todavía no se ha alcanzado unos niveles de actividad empresarial similares a la época prepandemia. Esta situación se ve más acuciada en los sectores que más sufrieron el parón de la actividad, en concreto, el porcentaje de empresas con endeudamiento elevado se encuentra 7 puntos más alto que en 2019 (Banco de España, 2021). Del mismo modo, de acuerdo con los datos, más de la mitad de las empresas españolas son PYMES sin asalariados (Ministerio de Industria, Comercio y Turismo, 2022), lo cual, consecuentemente, es otro de los motivos por los que la actividad empresarial se ha visto tan afectada estos últimos años.

En la Figura 1: “Distribución de empresas por tamaño en España en 2022” se puede observar el resto de los actores que componen el panorama empresarial español.

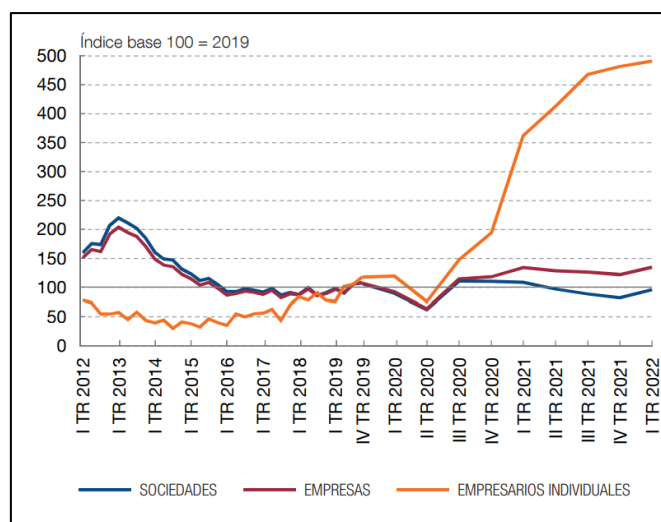
Figura 1: “Distribución de empresas por tamaño en España en 2022”



Fuente: Ministerio de Industria, Comercio y Turismo, 2022

Por otro lado, la mala situación económica ha provocado un aumento exponencial de los concursos de acreedores. Como se observa en la Figura 2: “Evolución del número de concursos, por forma jurídica 2012-2022”, la subida afecta especialmente a los empresarios individuales, los cuales han sido los más afectados por los concursos de acreedores desde principios del 2020.

Figura 2: “Evolución del número de concursos, por forma jurídica 2012-2022”



Fuente: Banco de España, 2022

Continuando con el análisis del contexto actual de la actividad empresarial, las conclusiones obtenidas por el Banco de España en su informe trimestral de la economía

española muestran que tanto la inflación como las consecuencias provocadas por la Guerra de Ucrania han afectado directamente a la actividad empresarial. Asimismo, la desaceleración producida a raíz de este contexto ha provocado una revisión a la baja del crecimiento del PIB español en 2022 (Banco de España, 2022). La actividad empresarial se encuentra íntimamente relacionado con este indicador.

$$(1) \text{ PIB} = C + I + G + X - M$$

Donde:

- Consumo (C)
- Inversión (I)
- Gasto público (G)
- Exportaciones (X)
- Importaciones (M).

La actividad empresarial afecta de forma transversal a todos estos factores: es una pieza clave para el desarrollo del consumo, tiene una importancia supina en la recepción de inversiones, el gasto público traducido en subvenciones puede ser decisivo para el desarrollo de la actividad empresarial, y las exportaciones e importaciones son motores del desarrollo de la actividad. El peso que tiene la actividad empresarial en el PIB denota la importancia de ésta y propulsa la necesidad de conocer su estado para entender en conjunto el desarrollo económico e industrial del país.

Comprendida la relevancia de la materia, en la actualidad surgen multitud de indicadores para medir la actividad empresarial, entre los más importantes se encuentran la Actividad Económica (como el consumo o el volumen de importaciones), la Actividad Económica Sectorial (por ejemplo, la actividad del sector servicios o los indicadores relativos a la gestión de compras) o un factor tan determinante como la Liquidez Financiera (Hipotecas o el número de empresas en quiebra entre otros). La combinación de todos estos factores permitiría crear un indicador capaz de monitorizar la actividad empresarial, lo cual puede suponer una gran ayuda para entender la situación económica.

En conclusión, la actividad empresarial se ve afectada por un gran número de factores, por ello, no se debe acudir a un único indicador tradicional para explicar el actual contexto empresarial de España.

1.2. Introducción a los indicadores sintéticos

En estos últimos años han proliferado el número de indicadores compuestos dados los claros beneficios que se desprenden del uso de los mismos. En concreto, su adopción por parte de las instituciones mundiales ha atraído la atención de los medios de comunicación y de los responsables políticos de todo el mundo, provocando directamente un aumento del número de aplicaciones (Salvatore Greco, 2017).

La base de cualquier indicador radica en su capacidad para reflejar comparaciones simples entre los distintos países, concretamente, asuntos que tradicionalmente son mostrados de una forma compleja (OECD, 2008). Por otro lado, los indicadores compuestos propiamente dichos, surgen de la combinación de una serie de indicadores individuales para producir una medida ordinal o cardinal agregada del rendimiento de cada país. Al tratarse de una unión de varios indicadores, las ponderaciones se erigen como piezas claves a la hora de construir el indicador sintético. En consecuencia, a la hora de establecer los pesos a los indicadores se debe tener en cuenta, la importancia relativa de cada indicador, la relación con los datos, los resultados de las combinaciones de indicadores y los potenciales efectos que puedan surgir del uso de matrices de puntuación en entornos multicriterio. (Saltelli, 2007). Dadas las dificultades que entraña combinar varios indicadores independientes, la construcción de un indicador compuesto es compleja y requiere del establecimiento de suposiciones que, por su propia naturaleza, deben ser analizadas debidamente para mantener unos resultados objetivos y analíticos.

Paralelamente, entre las ventajas propias del uso de este tipo de indicadores, encontramos su capacidad de integrar grandes cantidades de información en formatos sencillos y comprensibles. Ello favorece su uso como herramientas de comunicación y políticas y han propiciado que, entre los múltiples usos de los indicadores sintéticos, se haya sido capaz de aprovecharlos para comparar entre países su competitividad, su capacidad de innovación, su grado de globalización o su sostenibilidad medioambiental. (Freudenberg, M.,2003)

No obstante, no se debe olvidar que una construcción o interpretación errónea puede derivar en la emisión de mensajes políticos engañosos capaces de influir en conclusiones simplistas. Debido a ello, se debe comprender la función de “puente” que tienen los indicadores sintéticos, su utilidad no radica en crear dogmas, sino en inspirar el interés del público e iniciar debates (OECD,2008). Por ello, “la transparencia es esencial para construir indicadores creíbles” (OECD, 2008, p.22).

De esta manera, el incremento de indicadores compuestos en los diversos ámbitos políticos plantea dudas sobre la precisión y fiabilidad de los mismos. La capacidad de manipular los resultados y de distorsionar las conclusiones, provoca una potencial manipulación de las prescripciones políticas en base a ciertos indicadores sintéticos. (Freudenberg, M.,2003).

Igualmente, al tratarse de una combinación de indicadores, la ponderación de los distintos datos adquiere una importancia capital. Concretamente, los métodos de ponderación actuales deben ser constatemente analizados y mejorados para garantizar la objetividad y efectividad de los indicadores sintéticos.

Mostrados los riesgos y beneficios claves asociados a los indicadores sintéticos, listas como la elaborada por Saisana y Salteli en su informe de los indicadores sintéticos resultan esenciales a la hora de analizar la situación de los mismos (Saisana,M., 2004). Dichas indicaciones han sido reflejadas en la Tabla 1 mostrada a continuación.

Tabla 1: “Beneficios y desventajas propias del uso de indicadores sintéticos”

Pros	Contras
✓ Resumen cuestiones complejas o multidimensionales	✓ Pueden enviar mensajes políticos engañosos y no robustos si están mal contruidos o se interpretan erróneamente.
✓ Apoyan a la toma de decisiones	✓ Los resultados simples y generales pueden invitar a los políticos a sacar conclusiones políticas simplistas
✓ Ofrecen una visión global	✓ La construcción del indicador requiere la realización de juicios de valor en varias etapas de ésta
✓ Fáciles de interpretar	Requieren una mayor cantidad de datos para construir el indicador
✓ Encuentran tendencias	
✓ Clasifican los países en cuestiones complejas	
✓ Ayudan a atraer el interés del público	

✓ Reducen el tamaño de las listas de indicadores al incluir más información en el propio indicador	
--	--

Fuente: M. Saisana, 2004

Entendidos los potenciales riesgos asociados a este tipo de indicador, es razonable considerar que los beneficios obtenidos con el uso y la correcta monitorización de los indicadores sintéticos sobrepasan los puntos negativos. En consonancia, se debe notar que frases como: “ningún índice puede ser mejor que los datos que utiliza” (Naciones Unidas, 1992, p.56), no apelan al abandono de los indicadores sino a la mejora de los datos. Tanto la correcta selección de la información como el correcto tratamiento de los datos permiten la creación de mejores indicadores y, al mismo tiempo, son elementos claves a la hora de reducir la potencial manipulación de los resultados.

Para poder llevar a cabo este análisis los principales organismos, encabezados por la OCDE, han establecido una guía completa capaz de orientar a las personas decididas a construir un indicador sintético, mostrando los elementos esenciales con los que tiene que contar todo indicador que quiera considerarse confiable (Munda y Nardo, 2005). Dichos elementos son los siguientes y serán analizados en profundidad en los siguientes apartados:

1. Marco Teórico
2. Recolección de datos
3. Imputación de la información ausente
4. Análisis multivariante
5. Normalización
6. Ponderación y agregación
7. Análisis de robustez

Una vez establecido el indicador compuesto, los resultados proyectados a través del mismo pueden ser utilizados para apoyar la toma de decisiones más eficientes, óptimas y, sobre todo, basadas en datos que se alejan de sesgos e influencias subjetivas. Asimismo, la visión global que arroja este tipo de indicadores propicia la clasificación de los países en cuestiones complejas. (Saltelli, 2007).

Casos de éxito que demuestran el buen funcionamiento de los indicadores compuestos son múltiples y variados. Por ejemplo, indicadores como el Composite Leading Indicator (CLI) desarrollado por el OCDE, proporcionan a los gobiernos signos tempranos para detectar los

puntos de inflexión en los ciclos económicos, permitiendo conocer el futuro a largo plazo de la actividad económica en base al potencial mostrado por el indicador (OECD, 2022).

Por otro lado, indicadores como el Indicador de Sentimiento Económico (ISE) desarrollado por la Unión Europea se han demostrado útiles para medir la confianza en los sectores de la industria, los servicios, los consumidores, la construcción y el comercio minorista. A través del crecimiento del PIB el ISE puede analizar los movimientos de la economía en su conjunto. Otros indicadores económicos desarrollados por la Unión como el Indicador del Clima Empresarial, encargado de medir la actividad empresarial en la zona euro, o el Indicador de Incertidumbre Económica, capaz de medir la previsión de los consumidores y de las empresas respecto al futuro cercano, son pruebas latentes de la eficacia de los indicadores compuestos a la hora de analizar la economía europea (Comisión Europea, 2022).

Por todo lo anterior expuesto, se debe continuar apostando por la creación de estos mecanismos de estadística cuya utilidad como herramienta de comunicación y analítica ha quedado palpable.

1.3. Motivación

La actividad empresarial marca el desarrollo de la sociedad de un país, por ello, el análisis de la situación cobra especial relevancia. Dada la convulsa situación económica en la que el mundo se ha visto sumido estos últimos años, no es descabellado considerar que un buen análisis de las causas que han provocado las recesiones puede ser fundamental para paliar o frenar un futuro desaceleramiento económico.

Como se ha podido observar en la introducción a la realidad empresarial española, se trata de una materia multidisciplinar que no puede ser analizada a través de un solo indicador individual. Los múltiples factores que afectan a la actividad empresarial provocan que sea necesario relacionar todas aquellas variables que pueden ayudar a discernir la realidad del tema en cuestión.

Por otro lado, el análisis realizado sobre los indicadores compuestos muestra los beneficios de sintetizar la monitorización de una materia a través de un único indicador que englobe el conjunto de indicadores individuales que afectan a la misma. Los buenos resultados, reflejados por los indicadores económicos e industriales ya desarrollados, permiten analizar aspectos que afectan a la actividad empresarial, pero la dispersión de los indicadores impide un análisis

completo de la materia. La acumulación de los indicadores en un único indicador simplificaría la valoración de la actividad empresarial y permitiría comprender la situación de la misma.

Una vez comprendida la importancia de los indicadores sintéticos se puede comprender la oportunidad que se erige a la hora de desarrollar un indicador compuesto basado en la actividad empresarial. Sin duda, se muestra como una gran motivación proveer a la sociedad de una herramienta completa capaz de brindar una visión distinta a la proporcionada por los actuales indicadores existentes.

En base a lo anteriormente expuesto, la motivación de este trabajo se encuentra en la necesidad de desarrollar un indicador sintético capaz de ayudar a otorgar una visión coherente de la actividad empresarial española. En conclusión, este trabajo pretende servir como una herramienta capaz de analizar los factores claves que afectan a la actividad empresarial y, al mismo tiempo, encontrar patrones que puedan ayudar a desarrollar medidas para hacer frente a los cambios propios de los ciclos económicos. Todo ello, a través del desarrollo de un único indicador que simplifique el estudio de la materia.

1.4. Objetivos

El objetivo principal de este trabajo es construir un indicador compuesto capaz de medir la actividad económica española. En concreto, se van a seguir un conjunto de directrices técnicas a través de las cuales se logrará crear un indicador eficiente.

Los pasos que se van a seguir para poder construir el indicador se basan principalmente en la consecución de los siguientes objetivos:

- 1) Analizar los datos existentes en la base de datos, identificar el valor de los datos, la frecuencia, el número de datos y categorías... En definitiva, analizar de forma completa los datos disponibles.
- 2) Imputar los datos que faltan. Considerando las diferentes teorías existentes, eligiendo aquella o aquellas que mejor se adaptan a los datos actuales. Para ello se analizarán los valores faltantes y se completarán a través de las técnicas escogidas para ello.
- 3) Llevar a cabo un análisis multivariante. A través de este objetivo se logra analizar la correlación entre las distintas variables y la forma en la que se agrupan y se comportan los indicadores.
- 4) Normalizar los datos para poder compararlos. Con este fin, se valorarán las diferentes técnicas posibles y se escogerá aquella que mejor se adapte a la base de datos.

- 5) Ponderar y agregar los indicadores teniendo en cuenta la correlación y compensación entre ellos. Para ello se debe tener en cuenta el marco teórico escogido e integrar los resultados del análisis dentro del mismo.
- 6) Analizar la solidez y sensibilidad del indicador compuesto. A través de este análisis se podrán reformular las etapas previas en un proceso iterativo: determinar de qué indicadores se deben prescindir, el esquema de normalización que se debe seguir, la imputación de los datos que faltan, el tipo de ponderaciones que se deben realizar y, finalmente, escoger el método de ponderación y agregación necesario de acuerdo con las particularidades del indicador.
- 7) Relacionar el indicador compuesto con los indicadores publicados identificando los vínculos existentes entre ellos.
- 8) Presentar y visualizar el indicador compuesto. Con el objetivo de lograrlo, se debe establecer un mecanismo de visualización apto para el tipo de indicador.

A parte de estos objetivos, se sacarán conclusiones a través de los resultados reflejados por el indicador compuesto. El objetivo es lograr obtener tendencias y puntos de inflexión que nos permitan obtener justificaciones lógicas a la actividad económica actual.

1.5. Recursos

Los recursos que se emplearán en el desarrollo del indicador serán los siguientes:

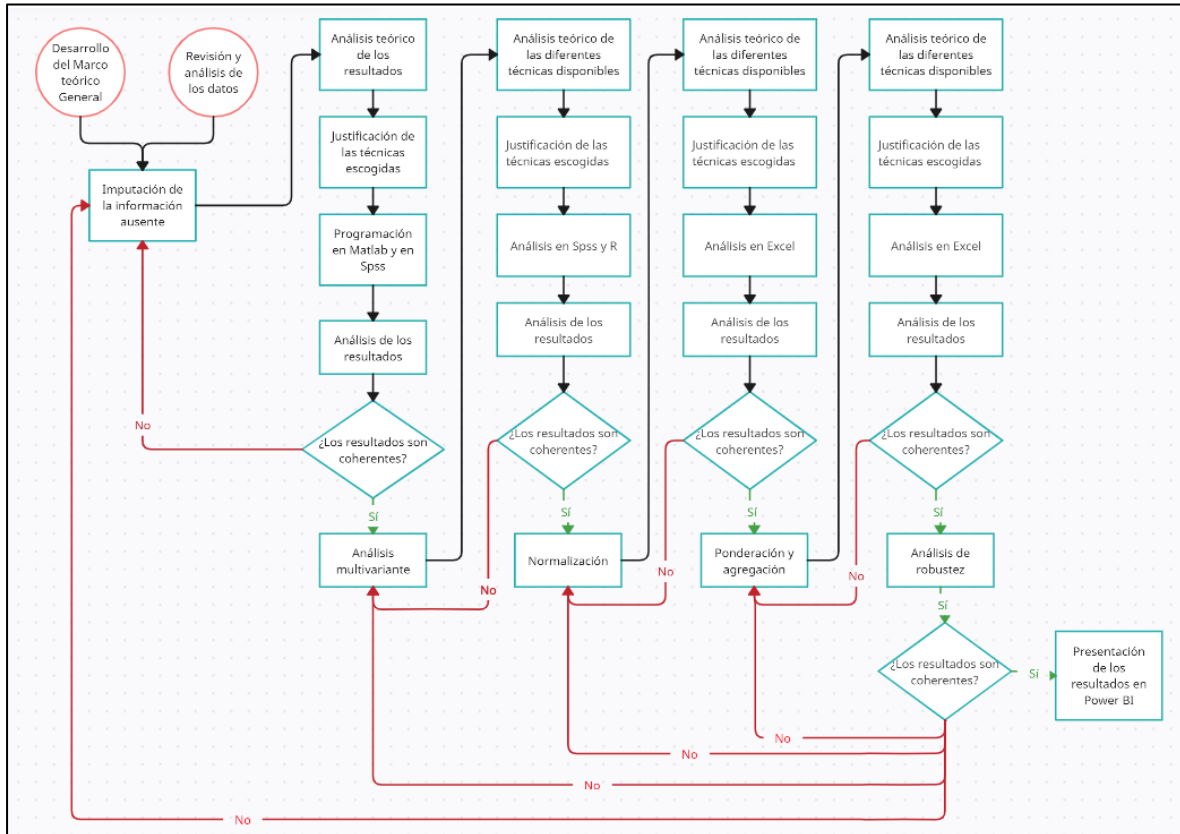
- 1) Una base de datos completa dividida en cinco categorías clave (actividad de la demanda y confianza de los consumidores, actividad económica, actividad económica nacional, actividad económica sectorial y liquidez y salud financiera) obtenida de la plataforma 360 Smart Vision
- 2) Uso de técnicas de SPSS
- 3) Uso de RStudio para analizar los datos
- 4) Python
- 5) Matlab para aplicar algunas técnicas de imputación de los datos

A través del uso de dichos recursos se construirá un indicador sintético capaz de medir la actividad económica española.

1.6. Metodología

La metodología escogida para llevar a cabo el TFG se ha ilustrado en un diagrama de flujo a fin de facilitar la visualización de los pasos a seguir.

Figura 3: “Metodología seguida en la construcción del indicador”



Fuente: Elaboración propia

2. Estado del arte y construcción del indicador

2.1. Marco Teórico

A la hora de construir un indicador sintético cobra especial relevancia la calidad de los datos y los procesos utilizados durante la construcción del indicador. La calidad de los datos radica tanto en la precisión de los resultados como en la utilidad que producen los mismos, es decir, deben ofrecer resultados en un marco temporal relevante, el acceso ha de ser libre y no han de provocar conflictos entre los distintos datos. Por todo ello, las características de calidad dependen directamente tanto de las perspectivas como de las necesidades y de las prioridades de los usuarios que van a hacer uso del indicador sintético. (OECD, 2008).

Para poder medir la aptitud de los indicadores, numerosas organizaciones han creado varias guías a través de las que analizar la calidad de los datos. En concreto, Eurostat y el FMI cuentan con el “Código de Buenas Prácticas de las Estadísticas Europeas” (Comisión Europea, 2018b) y el “Marco de Calidad de los Datos” (FMI,2012) respectivamente. Dichas guías han ofrecido a las instituciones un marco global sobre el que valorar la fuente de sus datos y las necesidades que presentan los mismos. Un buen marco teórico permite determinar la construcción de la estructura óptima para los sets de entrenamiento que serán utilizados en la posterior imputación.

Se debe recordar que un indicador sintético es una combinación integrada de indicadores individuales. A la hora de llevar a cabo este objetivo se debe tener en cuenta la dificultad que supone alcanzarlo, por ello, es necesario construir un marco teórico sólido que permita combinar los indicadores individuales en un compuesto fiable sobre el que obtener una base para seleccionar los componentes y ponderaciones necesarias en la elaboración del indicador. Adicionalmente, este marco debe ser capaz de reflejar las dimensiones o la estructura de los hechos que se pretenden medir (Comisión Europea, 2018).

El objetivo último es lograr que la información obtenida sea coherente con los datos originales y con los comportamientos analizados. En consecuencia, el desarrollo de esta parte del trabajo es esencial para seleccionar los indicadores que se deberán utilizar para la construcción del indicador sintético. Esta parte de la construcción del indicador marca el momento en el que se deberá definir la frecuencia de los datos que van a ser tratados y la forma en la que se va a imputar la información ausente (Comisión Europea, 2018).

Previo a futuros pasos, se debe hacer un análisis de los datos con los que se cuentan. Para hacerlo se puede acudir a códigos como el desarrollado por la Comisión Europea. Dicha fuente considera las siguientes seis dimensiones de calidad a la hora de analizar los datos estadísticos (Comisión Europea, 2017) :

- La pertinencia. Hace referencia al grado en el que los datos son capaces de cubrir las necesidades actuales y potenciales de los usuarios.
- La exactitud. Se deberá tener en cuenta la frecuencia con la que se han presentado los datos. Igualmente, cobrará especial relevancia el número de datos perdidos y el tipo de imputación resultante de la misma. El objetivo último es que el número de datos con los que se cuenta se encuentren lo más cercano posible a los valores reales.
- La oportunidad y puntualidad. Se ha de tener en cuenta la temporalidad de los datos, el marco temporal escogido y la distancia con el tiempo real en el que se van a presentar los datos. El objetivo es que los datos escogidos sean oportunos para medir futuras situaciones en base a los resultados que arroje el indicador.
- La accesibilidad y claridad. Se ha de valorar tanto la facilidad que tiene un usuario para acceder a los datos como la claridad de la información que se va a extraer de los mismos.
- La comparabilidad. La diferencia de los indicadores individuales provoca que los datos presenten diferencias (mayores o menores) que afectan al indicador sintético final. Por ello, en esta sección se debe valorar la comparabilidad de los datos y la pertinencia de estos.
- La coherencia. Se ha de seguir la línea de las anteriores dimensiones. En concreto, en esta sección se ha de valorar la coherencia que surge a la hora de combinar los datos de forma fiable atendiendo a las diferentes maneras y diversos usos.

Adicionalmente, conviene mencionar otras guías de calidad desarrolladas por organismos de la relevancia de la OECD. Desde esta institución los criterios de calidad de las estadísticas se centran en dos aspectos concretos: la calidad de las estadísticas nacionales y la calidad de los procesos internos de recogida, tratamiento, análisis y difusión de datos y metadatos (OECD, 2003).

De estos dos aspectos se extraen varias conclusiones. Primordialmente, se observa la importancia para la OECD tanto de la calidad solicitada para llevar a cabo los procedimientos de obtención de la información, como de la propia calidad de los datos empleados en la

construcción del indicador. Igualmente, al poner al mismo nivel ambos elementos se exige al desarrollador no caer en el error de centrar sus esfuerzos en obtener una excelencia de la calidad del dato, sin valorar los conceptos teóricos necesarios para extraer una información valiosa y medibles de los mismos.

El desarrollo de técnicas de extracción adecuadas ofrece al investigador un mayor aprovechamiento de los datos, mientras que una fuente de datos adecuada permite al indicador desplegar resultados coherentes capaces de ofrecer tendencias que deriven en explicaciones razonadas a asuntos complejos.

Dada la importancia de esta sección, lo primero que se debe hacer es categorizar los datos, agrupando los indicadores individuales en categorías que reflejen dinámicas y aspectos compartidos. El número de categorías dependerá tanto del número de indicadores individuales como del tipo de indicadores escogidos, cuánta mayor variedad de datos exista mayor número de categorías, y, por lo tanto, mayor enriquecimiento del futuro indicador sintético.

Una vez categorizados los datos, se debe escoger la frecuencia que se va a seguir a la hora de imputarlos. En este paso cobra especial relevancia tanto el número de datos perdidos como la forma en la que se comportan los datos con los que se va a construir el indicador. Asimismo, se ha de tener en cuenta el aspecto que se va a medir con el indicador y el ámbito del marco temporal y especial en el que va a encajar el mismo. Si se pretende obtener una herramienta capaz de encontrar tendencias, se debe acotar el ámbito de trabajo del indicador pues en caso contrario se corre el riesgo de que los datos no sean ni representativos ni fiables.

Una vez expuesta la parte teórica de esta sección de la construcción del indicador, procederemos a ponerla en práctica en nuestro indicador compuesto de la Actividad Empresarial Española creando el marco teórico.

Los datos con los que se van a trabajar provienen de la plataforma 360 Smart Vision. Desde esta plataforma se han escogido los datos relativos a la actividad empresarial, que se encuentran recopilados en un EXCEL.

Como se observan, los datos se encontraban en bruto y sin tratar, por ello, lo primero que se hizo fue recopilar la información en un EXCEL desde el que se categorizaron los datos y se identificaron la frecuencia de los distintos indicadores.

En el proceso de tratamiento se distinguieron cinco categorías diferenciadas (Activación de la demanda y confianza de los consumidores con tres indicadores, Actividad económica con

cincuenta y nueve indicadores, Actividad económica nacional con tres indicadores, Actividad económica sectorial con veintiocho indicadores y la Liquidez y salud financiera con veintisiete indicadores).

Tras categorizar los datos, analicé las frecuencias de los datos originales y la aparición de valores perdidos. Fijando un marco temporal centrado en los años 2016-2023.

En la primera categoría observé que la frecuencia era completamente mensual y con numerosos valores perdidos en todo el marco temporal, concentrándose el mayor número de datos disponibles entre los años 2019 y 2020.

En la segunda categoría encontré una mayor complejidad a la hora de analizar los datos, no solo las frecuencias variaron de mayor forma que en la primera categoría (existían frecuencias mensuales, trimestrales, anuales), sino que, además, existía un mayor número de valores perdidos y ningún año reunía datos de todos los indicadores. No obstante, al igual que ocurría en la primera categoría, los años con el mayor número de datos se situaban entre los años 2019 y 2021 (siendo el 2020 el año con el mayor número de datos). Asimismo, encontré datos duplicados en algunos indicadores, en este aspecto encontraba meses dónde se incluían datos europeos y datos nacionales, e incluso, datos de otros países europeos. El problema que encontré recaía en gran medida no tanto al hecho de que se duplicaran las observaciones, sino que no se seguía un criterio estable. De esta forma, encontraba meses dónde se intercalaban datos de diferentes orígenes sin seguir un criterio fijo.

En la tercera categoría, la frecuencia de los datos se dividía entre frecuencias mensuales y frecuencias trimestrales. Al igual que en el resto de las categorías, existía un gran número de valores perdidos y la información se concentraba entre el 2019 y el 2020.

En la cuarta categoría, encontré problemas similares a los presentes en la segunda categoría. Las frecuencias variaban entre los diferentes indicadores (mensuales y trimestrales), y aparecían datos duplicados de diferente origen sin un criterio estable. Asimismo, en esta categoría se concentraron un gran número de datos perdidos en casi todo el marco temporal, erigiéndose los años 2019 y 2020 como excepciones a la norma.

Finalmente, en la quinta y última categoría, los problemas tanto de la segunda como de la cuarta categoría se repitieron. Los indicadores presentaban frecuencias diferentes (trimestrales, anuales y mensuales) y algunos indicadores se mostraban con datos duplicados de diferentes

orígenes. Al mismo tiempo, observé un gran número de datos perdidos y una focalización de la información entre los años 2019 y 2021.

Vistos los resultados del primer análisis de los datos, observamos dos puntos clave.

En primer lugar, respecto a los datos duplicados se decidió prescindir de los datos nacionales puesto que el indicador sintético se centraba en la actividad empresarial española, por lo que, la incorporación de datos provenientes de otro origen no aportaba información relevante para el indicador.

En segundo lugar, la diferencia de frecuencias imposibilita la agregación de los datos por lo que previo a ello se debían escoger las categorías y las frecuencias pertinentes para que posteriormente la agregación pudiera desarrollarse correctamente. En este sentido, la frecuencia escogida marcaría el número de datos que posteriormente se imputarían, por lo que se debía escoger una frecuencia desde la que el número de datos que se imputaran fuera la menor posible.

Siguiendo esta línea de pensamiento, lo primero que se realizó fue definir el marco temporal de los datos que iba a tratar. Dada la falta de información opté por descartar los años 2016, 2017, 2018, 2021, 2022 y 2023 y me centré en los años 2019 y 2020.

Desde este marco temporal identifiqué dos escenarios posibles:

1. Escenario 1: Año 2020 y frecuencia mensual. Eliminando los datos que faltaban por su poca relevancia en el análisis.
2. Escenario 2: Año 2019 y 2020. Escogiendo los datos con frecuencia mensual.

Para poder decidir el marco teórico sobre el que se iba a hacer la imputación se analizaron los datos que conformaban los dos escenarios. En este sentido, se desarrolló el análisis siguiendo las diferentes categorías:

- 1) “Demanda y confianza del consumidor”. Tanto en el marco temporal del 2020 con en el del 2019-2020 no se contaba con suficientes datos, en ningún caso se llegaba al 70% de los datos no ausentes. Debido a ello, se decidió eliminar la categoría al no poder ofrecer información suficiente.
- 2) “Actividad económica”. En el caso del marco temporal del 2019-2020 solo se contaba con seis indicadores completos, en el resto de los indicadores no se alcanzaban el 70% de datos no ausentes.

En el caso del marco temporal del 2020, se encontraron 21 indicadores completos. Del resto de indicadores se decidió eliminar el “CVESA”, el “CVIESA”, el “ICU”, el “IPRIM_VI”, el “PIB_AA”, el “PIB_AVI” y el “PIB_VR”, todos estos indicadores no contaban con datos suficientes para alcanzar el 70% de información no ausente.

Por otro lado, se encontraron una serie de indicadores que, aunque en primera instancia no alcanzaban el 70% de datos no ausentes, uniendo los datos del final del 2019 se podía disminuir el número de información faltante por debajo del 30%. Este tipo de datos estaban conformados por los siguientes indicadores: “CVIESB”, “CVEESV”, “CVEESV”, “CVIESB”, “CVIESVI” y “IPRIX_VA”.

- 3) “Actividad Económica Nacional”. Tanto en el marco temporal del 2020 con en el del 2019-2020 no se contaba con suficientes datos, en ningún caso se llegaba al 70% de los datos no ausentes. Debido a ello, se decidió eliminar la categoría al no poder ofrecer información suficiente.
- 4) “Actividad Económica Sectorial”. En el caso del marco temporal del 2019-2020 solo se contaban con 6 indicadores completos, el resto no contaban con información suficiente para alcanzar el 70% de datos no ausentes.

Por su parte, en el marco temporal de 2020 se contaba con 20 indicadores completos. Del resto de indicadores, se decidió eliminar tres indicadores el “SA”, el “SA_VI” y “VON”.

Asimismo, existían cuatro indicadores cuyo número de datos no ausentes no alcanzaba el 30% por lo que podían ser imputados, estos eran: “FICMMRVMM”, “FIIEM”, “IASSCMe” y “IASSRV”.

Finalmente, respecto al “IAM” se apreciaba que no contaba con suficientes datos, no obstante, esta situación podía solventarse haciendo uso de los datos del final de 2019 debido a que en este caso los datos faltantes estarían por debajo del 30%.

- 5) “Liquidez y Salud financiera”. En el marco temporal del 2019-2020 no se contaba con ningún indicador que alcanzara el 70% de información no ausente.

En el caso del marco temporal del 2020 se contaba con 6 indicadores completos y se decidió eliminar el indicador “ICE” por falta de información.

Este análisis se puede observar en los dos EXCEL presentes en el Anexo III.

Una vez analizadas todas las categorías, se optó por escoger el marco teórico del escenario 1, es decir, el marco temporal del 2020 con una frecuencia mensual y cuyo análisis de las categorías se podría resumir de la siguiente manera:

- “Demanda y confianza del consumidor”. Categoría eliminada. No pudo imputar por no respetar la regla 70-30.
- “Actividad económica”: 336 observaciones
 - o 7 indicadores a imputar: 84 observaciones.
 - o 21 indicadores completos: 252 observaciones.
- “Actividad económica nacional” Eliminar categoría. No pudo imputar por no respetar la regla 70-30.
- “Actividad económica sectorial”: 300 observaciones.
 - o 5 indicadores a imputar: 60 observaciones.
 - o 20 indicadores completos: 240 observaciones.
- “Liquidez y salud financiera”: 72 observaciones.
 - o 0 indicadores a imputar
 - o 6 indicadores completos: 72 observaciones.

Observaciones totales: 708 observaciones en 3 categorías.

2.2. Imputación de los datos

Una vez categorizada la información y definido el marco teórico es necesario analizar los datos perdidos. Previo a este paso, en la parte práctica del indicador se creó un **EXCEL** donde se incorporaron aquellos indicadores completos, mientras paralelamente se procedió a realizar la imputación del resto de la base de datos.

El proceso de imputación es uno de los puntos clave de la construcción de un indicador compuesto. La elección de la técnica adecuada requiere de un repaso previo tanto de la definición y categorías de los datos ausentes, como de los métodos de imputación existentes para poder escoger aquel que mejor le convenga. Un dato ausente es aquel que no se encuentra disponible en la base de datos. Las causas que puede llevar a su desaparición pueden agruparse en tres principales categorías:

- Missing Completely At Random (MCAR, Ausentes completamente al azar): datos cuya ausencia es debida completamente al azar, es decir, que la probabilidad de tener valores faltantes no depende de los valores observados o no observados en otras variables. En otras palabras, la probabilidad de tener datos faltantes es la misma para todas las observaciones (OECD, 2008).
- Missing At Random (MAR, ausentes al azar): datos cuya ausencia depende del resto de datos observados en otras variables. De esta forma, si conocemos el valor de otras variables en el conjunto de datos, podemos predecir la probabilidad de que falte un valor en una variable determinada (OECD, 2008).
- Not Missing At Random (NMAR, no ausentes al azar): datos cuya ausencia depende de factores que no han sido observado y que son ajenos a las variables presentes en la base datos. En esta clase de datos ausente no es posible predecir sus valores a través del resto de variables observadas.

Los tipos de datos que se encontraron en nuestra base de datos fueron datos ausentes NMAR y datos ausentes MCAR. De esta forma, los tipos de datos se localizaron de la siguiente manera:

- La primera categoría “Actividad económica sectorial” presentó datos NMAR, en concreto, cuatro de los cinco indicadores que requerían ser imputados contenían datos ausentes en el mes de agosto. Al tratarse de datos ausentes localizados en un punto concreto no se puede considerar que dicha falta sea debida puramente al azar o que la misma pueda ser completada por el resto de las variables.
- La segunda categoría “Actividad económica” y el indicador faltante de la categoría “Actividad económica sectorial” presentaron datos MCAR.

Una vez expuestos los tipos de datos ausentes, se puede proceder a explicar las diferentes técnicas de imputación de los datos.

- Imputación simple

Esta técnica consiste en la imputación de medias o distribuciones predictivas de los valores perdidos. Dicha distribución se realiza empleando los datos observados, mediante una imputación implícita o explícita (OECD, 2008).

➤ Modelización Implícita

En este caso, la imputación se centra en un algoritmo construido con supuestos implícitos subyacentes, que deberán ser analizados para comprobar si son procedentes y ajustados a la materia. A la hora de emplear esta técnica no se ha de olvidar que se trata de una imputación, es decir, que el resultado de la misma no producirá datos reales por lo que la base original continuará estando incompleta (OECD, 2008). Para poder llevar a cabo la modelización implícita existen varios métodos para imputar datos.

- “Imputación en caliente”. Este paso se limita a completar los datos ausentes a través de la información extraída de usuarios similares.
- “Sustitución”. En este caso la imputación de los datos consiste en sustituir la información faltante por información no seleccionada inicialmente.
- “Imputación en frío”. En este caso, la sustitución es un poco más sofisticada que aquellas realizada en los pasos anteriores, la sustitución esta vez radica en fuentes externas.

Como se puede observar la modelización implícita tiende a la subjetividad, siendo poco objetiva y razonable y dependiendo enormemente de las equivalencias entre datos. Todas las opciones ofrecidas por estos modelos se alejan del modelo científico y no permiten la creación de un indicador sintético realmente relevante y valioso.

➤ Modelización Explícita

A diferencia de los anteriores métodos, en este caso las suposiciones se realizan explícitamente a través de un modelo estadístico formal. Los tipos de imputación dentro de esta modelización son variados por lo que para poder analizar mejor esta modelización a continuación se definen las principales técnicas.

➤ “Imputación a través de la media, la mediana o el modo”

Cómo su propio nombre indica, consiste en suplir los datos a través de la elección de un valor estadístico. El problema de este tipo de imputación es que el estimador escogido tiende a estar sesgado, la información escogida no va a poder ofrecer respuestas fieles capaces de suplir los datos ausentes. Debido a ello, esta imputación se presenta deficiente para poder ofrecer un indicador compuesto eficaz.

➤ “Imputación por regresión”

En este caso la técnica empleada es más compleja que la de los anteriores métodos analizados. El proceso de regresión se centra en eliminar del modelo las observaciones con datos faltantes sustituyéndolas por los resultados obtenidos de una ecuación de regresión. A través de dicha ecuación se logra sustituir los datos ausentes con el valor de una media condicionada por las covariables. (Medina & Galán, 2007).

➤ “Maximización de expectativas”

En este caso la imputación se centra en la interdependencia entre los parámetros del modelo y los valores que faltan. El objetivo es sustituir los datos ausentes por estimaciones extraídas a través de un proceso iterativo. Para llevar a cabo este proceso se tendrá que hacer uso de una función de probabilidad capaz de relacionar el conjunto de datos con el parámetro del modelo ausente, logrando de esta forma, encontrar la probabilidad de observar un conjunto de datos para un determinado valor (OECD, 2008).

➤ “Interpolación lineal”

La interpolación lineal es una técnica de imputación utilizada para estimar un valor desconocido en una serie de datos a partir de dos valores conocidos. En la interpolación lineal, se asume que los datos siguen una tendencia lineal que se utiliza para estimar un valor desconocido (MATLAB - MathWorks España, 2022).

▪ Imputación múltiple

En estos tipos de imputaciones no se utiliza un algoritmo de probabilidad parametrizada para todo el conjunto de datos. De esta forma, se realiza un proceso aleatorio que refleja la incertidumbre, es decir, consiste en ofrecer diferentes situaciones para tratar de encontrar los datos correctos. La imputación se realiza tantas veces como conjuntos de datos completos se creen. En cada conjunto de datos se calculan los valores ausentes junto con sus errores estándar y, una vez obtenido el número de conjuntos de datos completos, se calcula un valor estadístico (ya sea la media o la mediana) fruto de la combinación de todos los sets de datos. Tras ello, se calcula la varianza dentro de la imputación (OECD, 2008).

En este tipo de modelos se pueden utilizar múltiples métodos de imputación destacando entre todos ellos el método Monte Carlo. En este método se imputan los datos a través de una

distribución normal multivariante donde cada secuencia de variables aleatorias tiene una distribución dependiente de los elementos anteriores.

Todas las imputaciones anteriormente expuestas podrían resumirse en la siguiente tabla:

Tabla 2: “Beneficios y desventajas de las distintas técnicas de imputación”

Técnica	Pros	Contras
“Imputación en caliente” “Sustitución” “Imputación en frío”, “Métodos de imputación del valor anterior, siguiente o más cercano”	<ul style="list-style-type: none"> ✓ Sencillo y fácil de implementar ✓ Preciso cuando son datos MCAR 	<ul style="list-style-type: none"> ✓ Tendencia a la subjetividad (aparición de sesgos) ✓ A medida que aumentan los datos ausentes menos preciso ✓ No tiene en cuenta la variabilidad, subestimación de la varianza
“Interpolación lineal”	<ul style="list-style-type: none"> ✓ Sencillo y fácil de implementar ✓ Resultados precisos en conjuntos de datos MCAR ✓ Apropiado para variables continuas 	<ul style="list-style-type: none"> ✓ Poca precisión cuando hay un número grande de datos ausentes ✓ No tiene en cuenta las relaciones no lineales entre las variables ✓ Aparición de sesgos ✓ Puede generar valores imputados fuera del rango de los valores observados o no representativos
“Imputación a través de la media, la mediana o el modo”	<ul style="list-style-type: none"> ✓ Sencillo y fácil de implementar ✓ Robustos a los valores atípicos, no afectan la distribución general de los datos 	<ul style="list-style-type: none"> ✓ Imprecisos si los valores faltantes no son aleatorios o si hay muchos valores faltantes en el conjunto de datos ✓ Reducción de la variabilidad de los datos y subestimar la varianza

	<ul style="list-style-type: none"> ✓ Resultados precisos en conjuntos de datos MCAR 	<ul style="list-style-type: none"> ✓ No tiene en cuenta la relación entre las variables, lo que puede llevar a una imputación inexacta de los valores faltantes ✓ Aparición de sesgos ✓ No apropiados para conjuntos de datos con valores extremos o si las variables tienen una distribución no normal
“Imputación por regresión”	<ul style="list-style-type: none"> ✓ Más precisa que los métodos de imputación anteriores ✓ Conserva la relación entre las variables ✓ Sin sesgos en datos MCAR o MAR 	<ul style="list-style-type: none"> ✓ Computacionalmente costosa ✓ Poco precisa si los datos faltantes están relacionados con la variable objetivo de la regresión ✓ Dificultad para seleccionar las características relevantes cuando se tiene un conjunto de datos complejo o grande. ✓ Reducción de la variabilidad de los datos y subestimar la varianza
“Maximización de expectativas”	<ul style="list-style-type: none"> ✓ Evaluación objetiva de las opciones disponibles ✓ Considera la probabilidad: reduciendo el riesgo y la 	<ul style="list-style-type: none"> ✓ No tiene en cuenta la importancia relativa de los resultados ✓ Puede ser inadecuada para situaciones complejas ✓ Lentitud de procesamiento cuando la base de datos es grande

	incertidumbre asociados con la toma de decisiones. ✓ Adaptable ✓ Sin sesgos en datos MCAR o MAR	
Imputación múltiple	✓ Sin sesgos en datos MCAR o MAR ✓ Mejora la precisión de los datos ✓ Mantiene la distribución original de los de los datos	✓ Computacionalmente costoso ✓ Requiere supuestos ✓ No siempre garantiza la precisión

Fuente: Elaboración propia a partir de los datos de la OECD, 2008

De entre todos los métodos expuestos, se decidió prescindir de los multicriterio al depender de un gran número de supuestos (especialmente, al imponer que los datos se distribuyan de una forma normal) y de no garantizar la precisión de los resultados. Asimismo, se optó por no utilizar los métodos de imputación implícita (imputación en caliente, en frío o sustitución) y las técnicas basadas en la imputación a través de la media, la mediana o el modo puesto que producen valores imprecisos que subestiman la varianza y generan sesgos.

Del resto de técnicas se decidió hacer uso de todas ellas para seleccionar aquella que ofreciese unos resultados más precisos. Para ello, se desarrolló un análisis más profundo de dichos métodos de imputación.

➤ “Imputación por regresión”

Este procedimiento se sustenta en la idea de que los indicadores tienen comportamientos parecidos. A partir de esta premisa, el modelo tiene en cuenta el conjunto de observaciones contenidas en los distintos indicadores, creando relaciones en función del

comportamiento del resto de datos. Si un indicador tiende a crecer o disminuir, el dato faltante en otro indicador tendrá un comportamiento similar dentro del modelo.

En resumidas cuentas, la imputación por regresión implica completar los datos faltantes a partir de la serie temporal tanto del indicador en cuestión como del resto de indicadores. Por ello, esta técnica requiere una alta relación entre los diferentes indicadores.

Antes de llevar a cabo la imputación propiamente dicha, se requiere analizar todos los subconjuntos de predictores mediante una regresión múltiple. Una vez escogidos los subconjuntos, se determina el conjunto óptimo haciendo uso de los siguientes criterios (OECD, 2008): el valor de la tasa de explicación del R², el valor del cuadrado medio residual RMSE, el valor del error absoluto medio MAE, el valor del C_p de Mallows y la regresión por pasos. El objetivo es que tras hacer uso de las técnicas anteriores se pueda escoger el subconjunto con el error más bajo. De entre todos estos métodos conviene explicar las fórmulas del RMSE y el MAE:

El cálculo del RMSE refleja la diferencia entre los valores reales y los valores predichos por el modelo de regresión. Cuanto más bajo sea el RMSE más ajustados serán los resultados predichos en referencia a los reales. La fórmula del RMSE es la siguiente:

$$(2) \quad RMSE = \frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2)^{1/2}$$

- N: Es el número de observaciones en el conjunto de datos.
- P: Es el valor real de la variable dependiente (la que queremos predecir) para la observación i.
- O: Es el valor predicho de la variable dependiente para la observación i.
- Σ: Es la suma de los valores para todas las observaciones.

El MAE, por otro lado, no toma en cuenta el cuadrado de las diferencias entre los valores reales y predichos, sino que utiliza la magnitud absoluta de la diferencia. El MAE se expresa en las mismas unidades que la variable dependiente y, al igual que el RMSE, cuanto menor sea su valor, mejor será el ajuste del modelo a los datos. La fórmula del MAE es la siguiente:

$$(3) \quad MAE = \frac{1}{N} \sum_{i=1}^N |P_i - O_i|$$

Donde:

- N: Es el número de observaciones en el conjunto de datos.
- P: Es el valor real de la variable dependiente (la que queremos predecir) para la observación i .
- O: Es el valor predicho de la variable dependiente para la observación i .
- Σ : Es la suma de los valores para todas las observaciones.
- $| |$: Es el valor absoluto de la diferencia entre los valores reales y los valores predichos.

El problema de la imputación por regresión es que depende en gran medida de la relación entre los indicadores. Por lo tanto, se deberá esperar a analizar en una matriz de correlación¹ los datos muestrales para ver si la regresión es la técnica más adecuada para imputar los datos ausentes.

➤ “Maximización de expectativas”

En este caso la imputación se centra en la interdependencia entre los parámetros del modelo y los valores que faltan. El objetivo es sustituir los datos ausentes por estimaciones extraídas a través de un proceso iterativo. Para llevar a cabo este proceso se tendrá que hacer uso de una función de probabilidad capaz de relacionar el conjunto de datos con el parámetro del modelo ausente, logrando de esta forma, encontrar la probabilidad de observar un conjunto de datos para un determinado valor (OECD, 2008).

En concreto, la maximización de expectativas cuenta con dos elementos principales: la estimación de expectativas y la maximización en sí misma. El primer elemento consiste en la estimación inicial de los datos ausentes (condicionada a los datos observados y a los parámetros actuales del modelo), mientras que el segundo elemento se basa en la maximización de la probabilidad de que los datos observados se correspondan con los valores estimados de las variables ausentes (para ello se ha de recalcular los parámetros de distribución con la anterior imputación). Estos dos pasos se repiten iterativamente hasta que se alcanza la convergencia y se obtienen los parámetros óptimos del modelo. Estos pasos se repiten hasta que no existan cambios entre las estimaciones y la matriz de varianza-covarianza. (OECD, 2008).

La principal aportación de este tipo de imputación son las múltiples opciones que ofrece, puede emplearse tanto para estimar la varianza de un dato como para el análisis factorial. La

¹ Véase la sección de análisis multivariante

facilidad para construir el proceso es un aspecto atractivo de este proceso, asimismo, la interpretación estadística que se puede extraer en cada caso y la fiable convergencia que ofrece son aspectos esenciales a la hora de valorar emplear este tipo de método explícito.

Su principal desventaja radica en la lentitud del proceso de convergencia cuando la cantidad de datos ausentes es muy elevada.

El fin último del proceso consiste en lograr unos resultados coherentes y lo más cercanos a los datos reales. Las iteraciones del proceso se irán repitiendo de manera continua el número de veces que sea necesario para completar toda la información faltante (OECD, 2008).

Una vez analizados los posibles métodos que se pueden emplear a la hora de imputar los datos, surge la cuestión sobre cuál de ellos escoger. Esta elección puede tomarse en base a varias reglas generales, principalmente estas indicaciones hacen referencia a la propia fuente de datos con la que se está trabajando, es decir, a las características principales de dichos conjuntos de datos (la escala en la que se han medido, el número de datos ausentes en proporción al conjunto o el tipo de indicador faltante) (OCED, 2008).

Un proceso más técnico para escoger el método de datos correctos recae en la evaluación de rendimiento de los diferentes procesos de imputación. Dicha evaluación puede llevarse a cabo a través del uso de dos algoritmos: el coeficiente de correlación y el coeficiente cuadrático de correlación. Asimismo, se puede medir el error medio de imputación del modelo a través del error cuadrático medio y el error absoluto medio (ambos explicados en (2) $RMSE = \frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2$) y (3) $MAE = \frac{1}{N} \sum_{i=1}^N |P_i - O_i|$)).

Finalmente, los diferentes métodos de imputación pueden medirse a través de métodos de bootstrapping para generar muestras de valores imputados. De esta manera se realiza un análisis de rendimiento de cada muestra y se calculan tanto los errores estándar como las desviaciones estándar de los análisis de rendimiento (OECD, 2008).

➤ “Interpolación lineal”

Este tipo de imputación toma como referencia dos puntos de la línea temporal y establece una línea recta entre ambos. A través de la unión de los puntos se puede obtener el valor de los datos ausentes (CK-12 Foundation, 2018). Otros métodos de interpolación lineal incluyen la técnica Spline cúbica a trozos, esta técnica se basa en el uso de polinomios cúbicos en cada uno de los intervalos de la función para estimar el valor de la función en puntos intermedios. Ello

garantiza que la función interpolada sea suave y continua (Navarro, 2013). Una variante de la técnica Spline es la interpolación PCHIP (Piecewise Cubic Hermite Interpolating Polynomial), en este caso, la interpolación utiliza polinomios cúbicos que mantienen la forma general de la función entre los puntos de control conocidos. Esto significa que los polinomios cúbicos se ajustan a los datos conocidos, pero también se adaptan a la forma general de la función, lo que reduce la oscilación y la propagación de errores de aproximación (MATLAB - MathWorks España, 2022).

Finalmente, la última variante de los métodos de interpolación que se prevén es la interpolación Makina (Modified Akima Cubic Interpolation). En este caso, la interpolación es polinómica y se basa en la creación de un conjunto de polinomios cúbicos en cada intervalo de los datos conocidos. En la interpolación Makima, los polinomios cúbicos se ajustan a los datos conocidos utilizando una fórmula que garantiza una mayor suavidad en los puntos de interconexión y una menor oscilación en los puntos intermedios. Esta fórmula utiliza una función de peso que pondera la contribución de cada punto de control a la estimación del valor de la función en puntos intermedios.

Una vez expuestos los tipos de datos y las técnicas que se van a emplear se acudió a SPSS IBM y Matlab para llevar a cabo la imputación propiamente dicha. Para ello, lo primero que se hizo fue una partición (70/30) de los datos reales originales, eliminando datos de forma aleatoria para poder imputar y comparar con los datos primigenios.

Tras la partición, se acudió al programa SPSS IBM para imputar todos los datos ausentes utilizando los métodos de maximización de expectativas y de regresión, y a Matlab para emplear los métodos de interpolación lineal.

Una vez realizada las imputaciones, los indicadores con datos NMAR presentaron las siguientes conclusiones:

Tabla 3: “Análisis de la imputación del indicador FICMMRVMM ”

FICMMRVMM

	RMSE	MAE
EM	0.011	0.011
Regresión	0.133	0.099
Linear	0.109	0.100

Spline	0.064	0.057
Pchip	0.118	0.098
Makima	0.109	0.094

Fuente: Elaboración propia

Tabla 4 “Análisis de la imputación del indicador FIIEM”

FIIEM

	RMSE	MAE
EM	0.042	0.039
Regresión	0.181	0.176
Linear	0.316	0.266
Spline	0.069	0.101
Pchip	0.122	0.063
Makima	0.065	0.051

Fuente: Elaboración propia

Tabla 5: “Análisis de la imputación del indicador IASSCMe”

IASSCMe

	RMSE	MAE
EM	0.034	0.033
Regresión	0.044	0.033
Linear	0.030	0.031
Spline	0.037	0.027
Pchip	0.029	0.027
Makima	0.027	0.020

Fuente: Elaboración propia

Tabla 6: “Análisis de la imputación del indicador IASSRV”

IASSRV

	RMSE	MAE
EM	0.127	0.123
Regresión	0.336	0.307
Linear	0.137	0.119

Spline	0.089	0.087
Pchip	0.097	0.088
Makima	0.080	0.075

Fuente: Elaboración propia

Por su lado, aquellos indicadores con datos MCAR reflejaron las siguientes conclusiones:

Tabla 7: “Análisis de la imputación del indicador CVEESB”

CVEESB

	RMSE	MAE
EM	0.011	0.011
Regresión	0.133	0.099
Linear	0.277	0.258
Spline	0.108	0.108
Pchip	0.099	0.097
Makima	0.104	0.102

Fuente: Elaboración propia

Tabla 8: “Análisis de la imputación del indicador CVEESV”

CVEESV

	RMSE	MAE
EM	0.042	0.039
Regresión	0.181	0.176
Linear	0.222	0.219
Spline	0.441	0.339
Pchip	0.086	0.086
Makima	0.130	0.123
Regresión	0.181	0.176

Fuente: Elaboración propia

Tabla 9: “Análisis de la imputación del indicador CVIESB”

CVIESB

	RMSE	MAE
EM	0.034	0.033
Regresión	0.044	0.033
Previous	0.201	0.179
Spline	0.195	0.195
Pchip	0.132	0.117
Makima	0.125	0.111

Fuente: Elaboración propia

Tabla 10: “Análisis de la imputación del indicador CVIESVI”

CVIESVI

	RMSE	MAE
EM	0.127	0.123
Regresión	0.336	0.307
Linear	0.022	0.016
Spline	0.068	0.065
Pchip	0.022	0.017
Makima	0.036	0.033

Fuente: Elaboración propia

Tabla 11: “Análisis de la imputación del indicador IPI”

IPI

	RMSE	MAE
EM	0.127	0.123
Regresión	0.336	0.307
Linear	0.106	0.105
Spline	0.255	0.209
Pchip	0.109	0.106
Makima	0.110	0.107

Fuente: Elaboración propia

Tabla 12: “Análisis de la imputación del indicador IPRIM_VA”

IPRIM_VA

	RMSE	MAE
EM	0.127	0.123
Regresión	0.336	0.307
Linear	0.020	0.017
Spline	0.002	0.002
Pchip	0.003	0.003
Makima	0.004	0.004

Fuente: Elaboración propia

Tabla 13: “Análisis de la imputación del indicador IPRIX_VA”

IPRIX_VA

	RMSE	MAE
EM	0.127	0.123
Regresión	0.336	0.307
Linear	0.017	0.017
Spline	0.008	0.007
Pchip	0.015	0.018
Makima	0.012	0.012

Fuente: Elaboración propia

Tabla 14: “Análisis de la imputación del indicador IAM”

IAM

	RMSE	MAE
EM	0.177	0.146
Regresión	0.841	0.839
Linear	0.920	0.914
Spline	1.053	1.023
Pchip	0.975	0.961
Makima	0.967	0.953

Fuente: Elaboración propia

Este análisis comparativo que acaba de ser expuesto ha sido reflejado en un EXCEL completo.

Finalmente, indicar que dado que la diferencia temporal podía afectar a la integridad del análisis se optó por eliminar aquellos indicadores que se imputaron con datos de finales de 2019 y que no lograron alcanzar las observaciones suficientes para cumplir la regla 70-30 de la imputación de datos ausentes en el periodo temporal de 2020. Estos indicadores fueron los siguientes: “CVEESB”, “CVEESV” y “CVIESB” y el “IAM”.

2.3. Análisis Multivariante

El análisis multivariante tiene como objetivo analizar conjuntos de datos en pos de encontrar patrones y relaciones entre las observaciones. Concretamente, las técnicas de análisis multivariante están orientadas a impulsar la creación del conocimiento a través de técnicas estadísticas que permiten el análisis simultáneo de múltiples mediciones (Hair et al., 1995).

Para poder exponer este apartado se van a enunciar los principales métodos utilizados en este tipo de análisis.

- Análisis de los componentes principales (PCA)

A través de esta técnica se pretende estudiar la varianza de los datos observados a través de la combinación lineal de los datos originales. El objetivo es reducir el número de variables a través de la identificación de las relaciones y correlaciones entre las muestras, convirtiendo en última estancia el conjunto de variables en un conjunto no correlacionado (conformado por los componentes principales) (OECD, 2008).

- Análisis factorial

Esta técnica es parecida al PCA, pero difiere en que, en vez de basarse en combinaciones lineales, los datos se fundan en los factores subyacentes del modelo (donde la varianza de los datos puede ser justificada por factores comunes y únicos) (OECD, 2008). A diferencia del PCA, más pensado para conjuntos de datos con menos variables, el análisis factorial suele ser más utilizado en la construcción de los indicadores compuestos. Específicamente, con el análisis factorial se busca reducir los datos para encontrar grupos homogéneos a partir del conjunto global de variables. Dichos subgrupos deberán estar conformados por variable correlacionados entre sí capaces de crear grupos independientes. En suma, el objetivo final radica en encontrar el mínimo número de dimensiones que permitan explicar de mejor manera la informa dispuesta en los datos (Pardo, 2019).

- Coeficiente Alfa de Cronbach

Esta opción es la estimación más común a la hora de analizar los datos de modelos o encuestas, y consiste en medir la variabilidad total de la muestra debida a la correlación de los indicadores. Este coeficiente permite conocer la medida en la que las variables miden la misma

habilidad o características, localizando la varianza total de los datos explicada por la varianza común (que es compartida por todos ellos) (OECD,2008).

- Análisis clúster

Esta técnica consiste en el uso de un conjunto de algoritmos con el propósito de clasificar los datos en base a características comunes (Massart and Kaufman 1983). Al igual que ocurría con el análisis factorial, el análisis clúster pretende reducir la dimensionalidad del conjunto, pero en este caso a través de similitudes/disimilitudes (OECD, 2008).

- Otros métodos de análisis

En último lugar, hay que mencionar que el análisis multivariante puede llevarse a cabo a través de un sinfín de métodos aparte de los ya mencionados. No obstante, se trata de métodos residuales y que no cuentan con un apoyo mayoritario.

Todos los métodos anteriormente expuestos podrían resumirse en la siguiente tabla:

Tabla 15: “Beneficios y desventajas de las distintas técnicas de análisis multivariante”

Técnica	Pros	Contras
Análisis de los componentes principales (PCA)	<ul style="list-style-type: none"> ✓ Reducción de dimensionalidad ✓ Identificación de patrones ✓ Mayor eficiencia ✓ Visualización de datos 	<ul style="list-style-type: none"> ✓ Pérdida de información ✓ Dificultad en la interpretación de los resultados ✓ Sensibilidad a los datos atípicos ✓ Dificultades en la elección del número de componentes ✓ Requiere que existe correlación entre los datos ✓ Funciona mejor con menos variables
Análisis factorial	<ul style="list-style-type: none"> ✓ Puede utilizarse con un gran número de variables 	<ul style="list-style-type: none"> ✓ Requiere que los datos estén correlacionados

	<ul style="list-style-type: none"> ✓ Reducción de dimensionalidad ✓ Identificación de las variables más importantes ✓ Combinación de variables ✓ Mayor eficiencia 	<ul style="list-style-type: none"> ✓ Dificultad en la interpretación de los resultados ✓ Sensibilidad a los datos atípicos ✓ Dificultades en la elección del número de factores ✓ Dependencia del modelo utilizado: los resultados pueden variar dependiendo del modelo escogido
Coeficiente Alfa de Cronbach	<ul style="list-style-type: none"> ✓ Evaluación de la confiabilidad ✓ Identificación de ítems problemáticos ✓ Facilidad de cálculo ✓ Amplia aplicabilidad 	<ul style="list-style-type: none"> ✓ Requiere que los datos estén correlacionados ✓ Dependencia de la longitud de la escala: sobreestimación de la confiabilidad para escalas cortas y subestimación para escalas largas. ✓ Sensibilidad a la homogeneidad de los ítems ✓ Sensibilidad al efecto de los errores de medición ✓ Limitaciones en la generalización de los resultados: se basa en una muestra específica, los resultados pueden no ser generalizables a otras poblaciones o muestras
Análisis clúster	<ul style="list-style-type: none"> ✓ No requiere de una correlación en los datos 	<ul style="list-style-type: none"> ✓ Sensibilidad a la selección de variables

	<ul style="list-style-type: none"> ✓ Identificación de patrones ✓ Identificación de grupos homogéneos ✓ Reducción de dimensionalidad ✓ Mayor eficiencia 	<ul style="list-style-type: none"> ✓ Sensibilidad al método de agrupamiento ✓ Dificultad para seleccionar el número de clústeres ✓ Dependencia del tipo de datos: puede no ser adecuado para datos categóricos o datos con valores atípicos
--	---	--

Fuente: Elaboración propia a partir de los datos de la OECD, 2008

Una vez analizados los posibles métodos que se pueden emplear a la hora de llevar a cabo el análisis multivariante, surge la cuestión sobre cuál de ellos escoger para imputar los valores ausentes.

Para ello, se decidió no hacer uso del análisis de los componentes principales dado que es una técnica que requiere que no existan un gran número de variables, en casos como el presentado en este trabajo el uso de este método requeriría tener que hacer significativas asunciones. Concretamente, esta técnica podría implicar que se excluyesen indicadores relevantes y que se incluyesen indicadores con una importancia no tan grande (OECD, 2008). Asimismo, se descartó el uso de la técnica Coeficiente Alfa de Cronbach al tratarse de un método orientado a datos de modelos y encuestas y, por consiguiente, poco apropiado para este tipo de trabajo (OECD, 2008).

Por otra parte, respecto a los análisis clúster y factorial se decidió llevar a cabo ambos análisis. De esta forma, se ha de comprender completamente la manera en la que funcionan antes de ponerlos en práctica.

- Análisis clúster

La agrupación de los datos en los conjuntos dependerá de las técnicas utilizadas para ello, ya sean jerárquicas o no jerárquicas (agrupación en árbol o k-means respectivamente). Asimismo, los grupos podrán delimitarse según el análisis de las distancias empleadas, esto determinará la valoración de la similitud o no de los casos del conjunto (cuanto más pequeña la distancia mayor la similitud) (OECD, 2008). En este sentido, el clustering jerárquico se caracteriza por comenzar el análisis de los datos con todas las observaciones que constituyen la base datos, configurando cada observación en el clúster que le corresponda por sus

características. Tras ello, el objetivo es que los clústeres converjan en grupos reducidos. Para ello, es necesario calcular las distancias entre los clústeres.

Entre las múltiples opciones de medida de distancia se encuentran (OECD, 2008):

- Euclídea. Se trata de una distancia geométrica en un espacio multidimensional y suele calcularse a partir de los datos originales. En esta medida si bien la incorporación de nuevos datos no afecta a la medición, la escala en la que se encuentren los datos puede interferir y dañar la obtención de una distancia representativa.
- Euclídea al cuadrado. Similar a la anterior medida, la diferencia radica en que se da mayor peso a las observaciones más alejadas.
- Manhattan. En este caso, la distancia se calcula a través de la media entre las distancias entre dimensiones.
- Chebychev. Desde dicha medida se tiene la posibilidad de definir los objetos como "diferentes" cuando presentan características diferentes en alguna dimensión.
- Potencia. A partir de esta distancia se puede atribuir un peso progresivo a aquellas dimensiones cuyas observaciones son muy diferentes.
- Porcentaje de desacuerdo. Es utilizada para los datos categóricos.

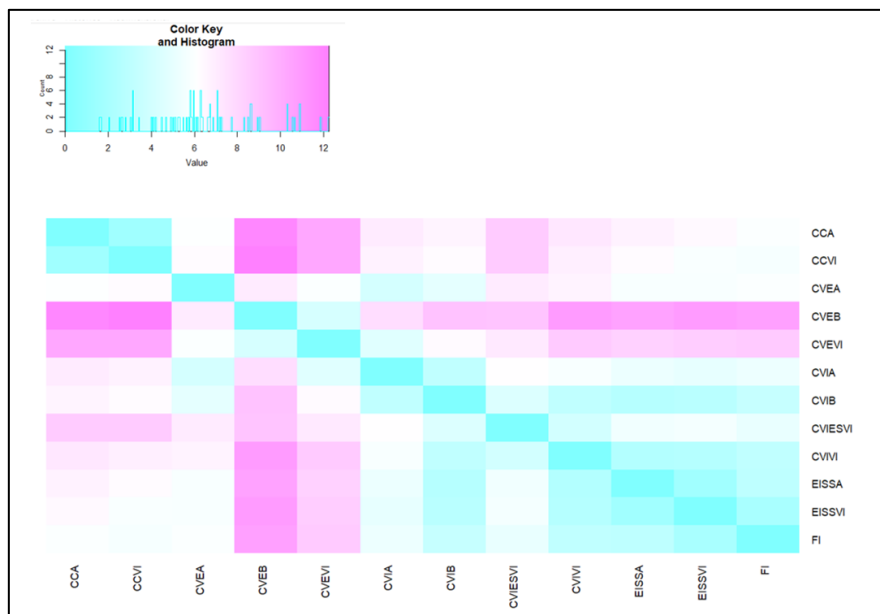
Tras haber mostrado las posibles distancias, se deben analizar las reglas de agrupación de las mismas entre los diferentes grupos. Entre las reglas más importantes de determinación se encuentran las siguientes (Spath, 1980):

- Vinculación simple (vecino más próximo). La distancia entre dos grupos viene definida por la distancia entre los dos elementos más cercanos de los distintos grupos.
- Vinculación completa (vecino más lejano). La distancia entre dos grupos viene marcada por la mayor distancia entre dos objetos pertenecientes a grupos diferentes.
- Media no ponderada de grupos de pares. La distancia entre dos grupos se calcula como la distancia media entre todos los pares de objetos de los dos grupos.
- Media ponderada de grupos de pares. Similar a la media no ponderada de grupos de pares, salvo que el tamaño del conglomerado, es decir, el número de objetos que contiene se utiliza como ponderación de la distancia media.
- Método de Ward (Ward, 1963). La pertenencia a un grupo se determina calculando la varianza de los elementos.

En la práctica, tomando los datos que fueron imputados en el anterior epígrafe se llevó a cabo esta evaluación haciendo uso de R y aplicando la distancia euclídea para concretar las distancias entre las categorías. La elección de esta distancia es debido al hecho de que son datos numéricos con pocos outliers, por ello, dada la simplicidad y la eficiencia de esta técnica se presenta como la más adecuada para estos datos.

Previo al estudio de los clústeres, se realizó un análisis preparatorio para tratar los datos. Ello implicó la estandarización previa de los datos para evitar que los resultados que se obtuviesen del análisis clúster presentaran sesgos. Asimismo, al tratarse de un análisis que reunía la información por clústeres, se presentaba imprescindible comprobar si los mismos presentaban características similares que permitiesen su agrupación posterior. Esta verificación se realizó a través de un resumen estadístico de los datos y de una visualización de la matriz de distancias.

Figura 4: “Matriz de distancia de los datos de la categoría Actividad Económica”

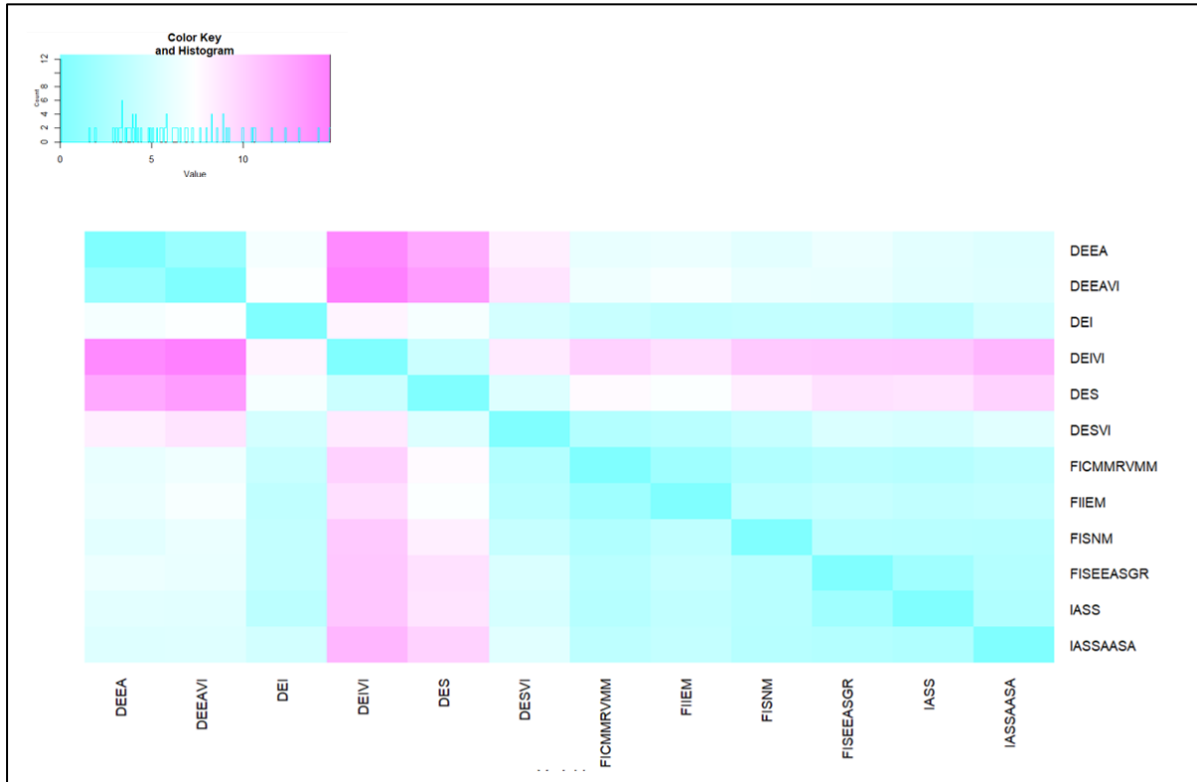


Fuente: Elaboración propia

Las escalas de colores mostraron las distancias entre las variables, de esta forma, los colores más morados denotaban una mayor distancia entre los datos, mientras que aquellos más azules mostraban una distancia menor. En la matriz de distancias se pudo observar las relaciones entre los datos y las potenciales agrupaciones que se podían crear a través de la similitud entre ellos. La distancia entre los datos y su correlación tiene una relación indirecta, pues a mayor distancia menor correlación y a menor distancia mayor correlación. Uniendo esta idea con los resultados reflejados por la Figura 4: “Matriz de distancia de los datos de la

“categoría Actividad Económica”, podemos concluir que existía una correlación alta entre los datos, lo cual podría permitir la agrupación de los mismos.

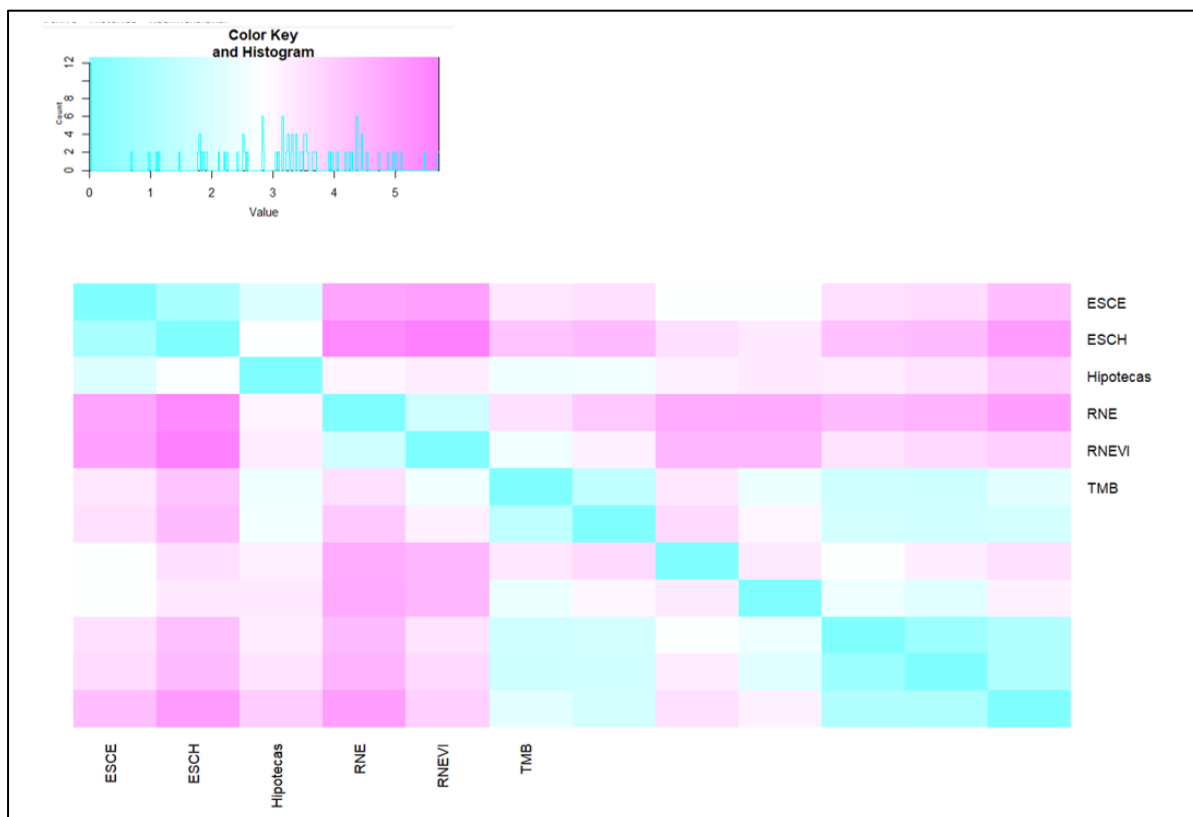
Figura 5: “Matriz de distancia de los datos de la categoría Actividad Económica Sectorial”



Fuente: Elaboración propia

Al igual que en la Figura 4: “Matriz de distancia de los datos de la categoría Actividad Económica”, en la Figura 5: “Matriz de distancia de los datos de la categoría Actividad Económica Sectorial” se apreciaba que los distintos colores permitían diferenciar las distancias entre los datos. Como ocurría con la primera categoría, los datos reflejaban una supremacía del color azul, y, por lo tanto, denotaban una distancia pequeña entre los mismos y una alta correlación.

Figura 6: “Matriz de distancia de los datos de la categoría Salud y Liquidez Financiera”



Fuente: Elaboración propia

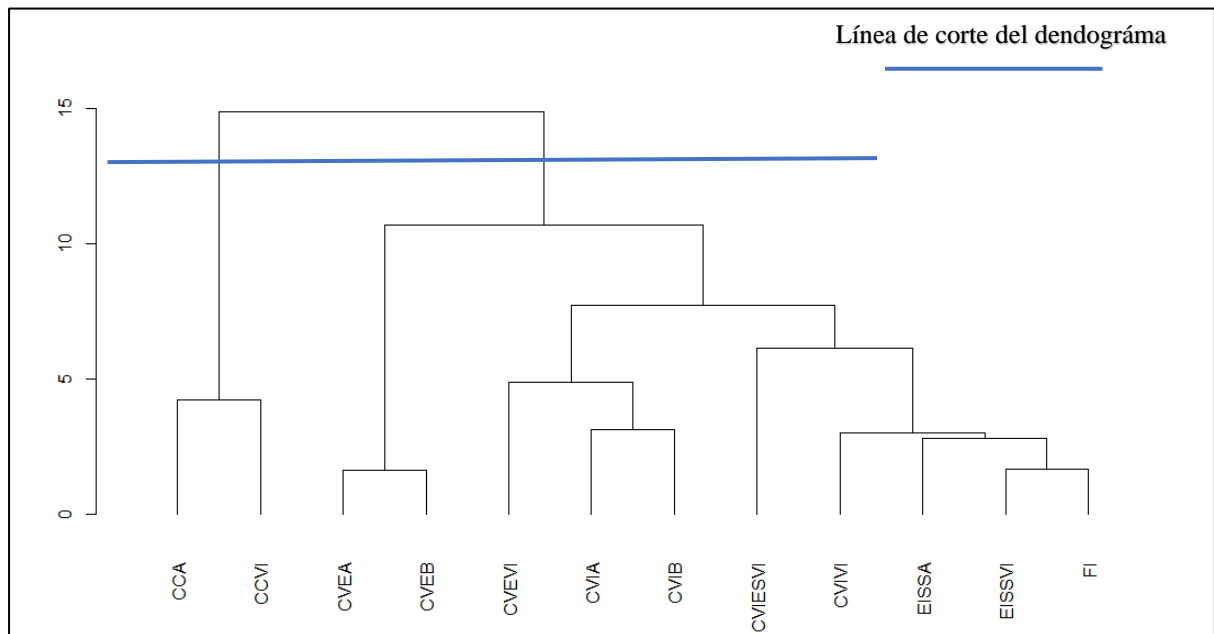
En esta última categoría, los resultados fueron distintos a los obtenidos en las categorías anteriores, con un fuerte predominio morado. Estos resultados reflejaron que los datos se encontraban a una distancia elevada y, por tanto, una correlación baja entre los mismos.

Tras haber tratado los datos, se llevó a cabo el análisis clúster. En consecuencia, se consideraron las dos posibles formas para identificar el número de clúster adecuados (el clustering jerárquico y el no jerárquico o k-means).

Se comenzó realizando el clustering jerárquico, para ello, se hizo uso de la matriz de distancias (antes expuesta) y se empleó el método de agrupación Ward. Esta técnica de agrupación fue escogida por su habilidad a la hora de minimizar la varianza total de cada clúster y, por su capacidad para crear grupos lo más homogéneos posibles. Una vez escogida la técnica, se visualizó el dendograma de cada categoría para poder determinar el número de clústeres óptimo en los que recoger la información. Para ello, se estudió la distancia entre los diferentes

niveles de decisión, es decir, aquellos momentos en los que se dividía la información en nuevos clústeres. Finalmente, el corte del dendograma se situó en la distancia entre niveles más significativa (el número de ramas cortadas sería el número óptimo de clústeres).

Figura 7: “Dendograma con los clústeres de la categoría Actividad Económica”

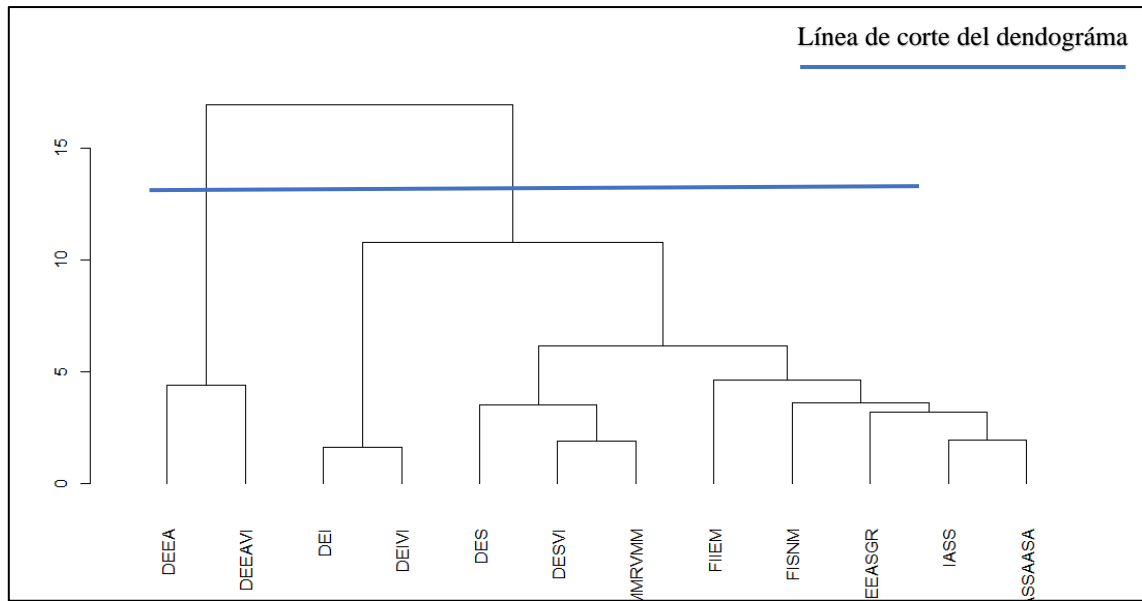


Fuente: Elaboración propia

En la primera categoría, como se observa en la Figura 7: “Dendograma con los clústeres de la categoría Actividad Económica” el número óptimo de clústeres a definir fueron dos. Esto es debido a que la distancia en el primer nivel de decisión es mayor que en el resto de las subdivisiones.

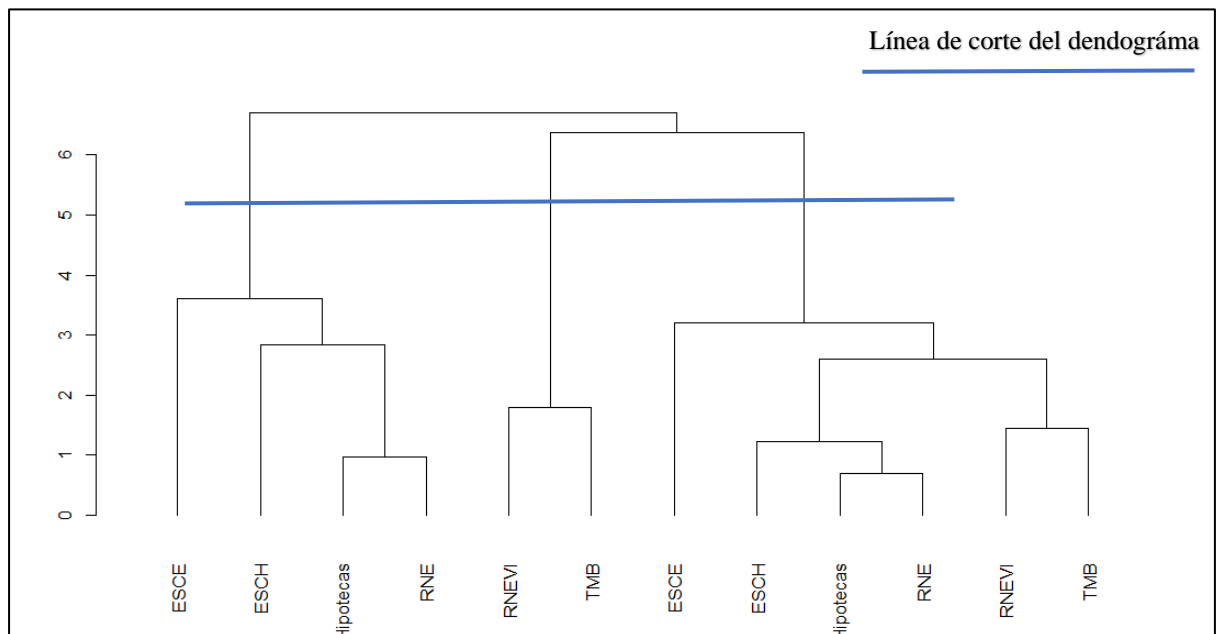
En la segunda y tercera categoría, como se muestra en la Figura 8: “Dendograma con los clústeres de la categoría Actividad Económica Sectorial” y en la Figura 9: “Dendograma con los clústeres de la categoría Actividad Económica Sectorial” el número óptimo de clústeres a definir se encontraba en 2 y 3 respectivamente.

Figura 8: “Dendograma con los clústeres de la categoría Actividad Económica Sectorial”



Fuente: Elaboración propia

Figura 9: “Dendograma con los clústeres de la categoría Actividad Económica Sectorial”



Fuente: Elaboración propia

Una vez finalizado el análisis jerárquico, se procedió a realizar al análisis clustering k-means para reevaluar el número de clústeres óptimos obtenidos con el anterior método.

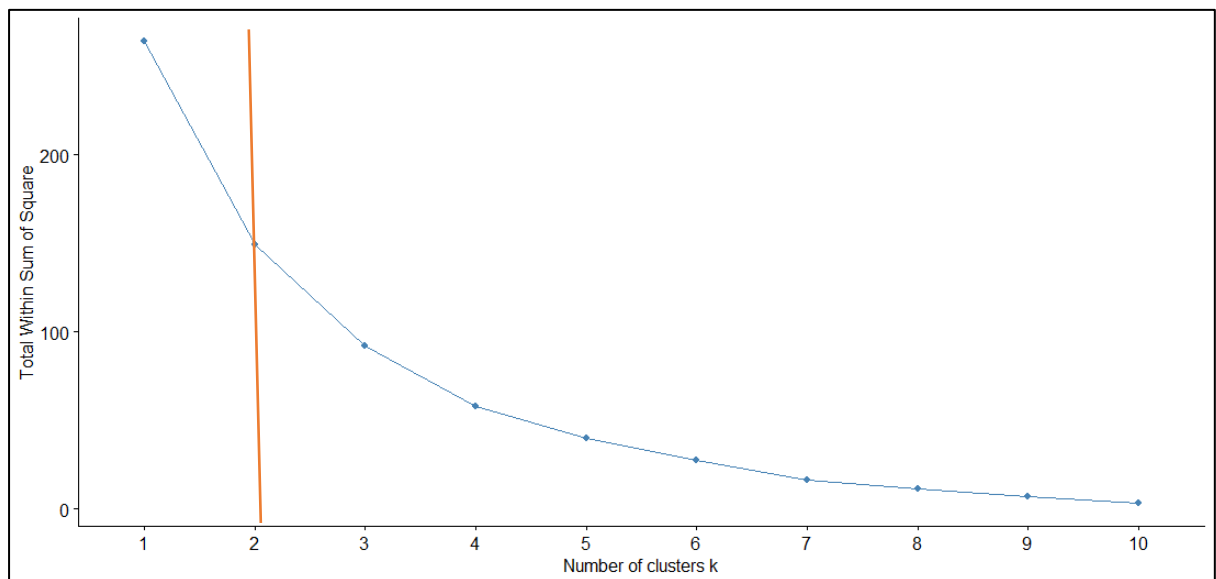
En este sentido, se incluyó inicialmente cada una de las observaciones en un clúster de forma aleatoria, con el objetivo de definir el valor del centroide de cada clúster. Dicho valor

estaba conformado por el vector de valores medios de cada variable y la suma de las distancias de cada observación en los centroides. En última instancia, las observaciones fueron reasignadas a aquellos clústeres con los que presentaba una mayor relación.

Concluidos los cálculos necesarios, se procedió a la definición del número de clústeres óptimo para clasificar los datos. Para poder llevar a cabo este paso, se valoraron dos técnicas visuales: el método del codo y el método de la silueta.

Por el método del codo se determinó el número de clúster en base al punto que presentaba un mayor cambio en el decrecimiento de la línea. En el caso de la primera categoría, como se observa en la Figura 10: “Número óptimo de clústeres: Método del codo. Categoría: Actividad Económica” el número óptimo estaría en el dos.

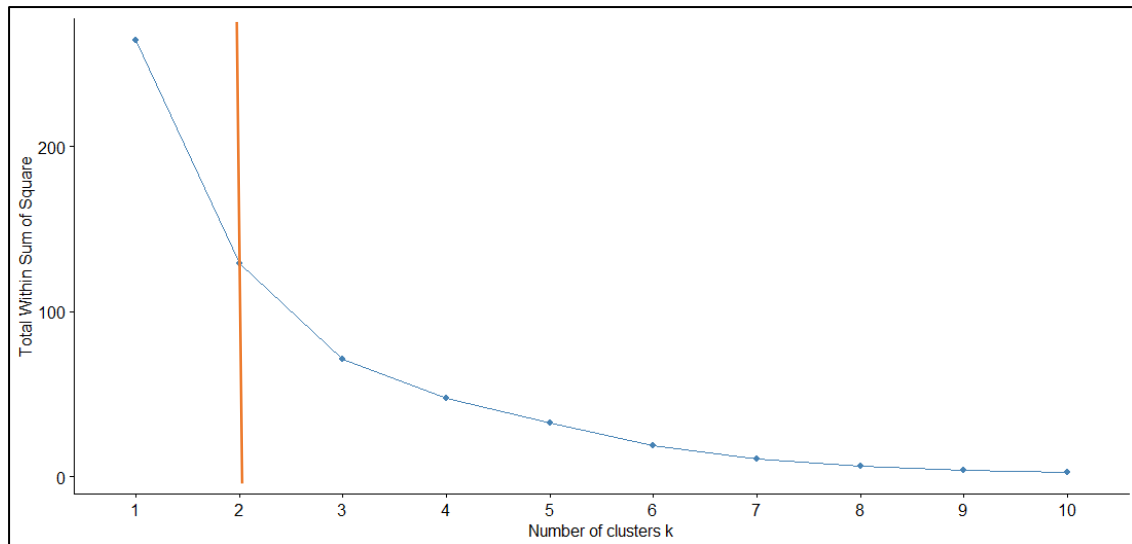
Figura 10: “Número óptimo de clústeres: Método del codo. Categoría: Actividad Económica”



Fuente: Elaboración propia

En la segunda categoría como se aprecia en la Figura 11: “Número óptimo de clústeres: Método del codo. Categoría: Actividad Económica Sectorial” el número óptimo se encontraba en el dos.

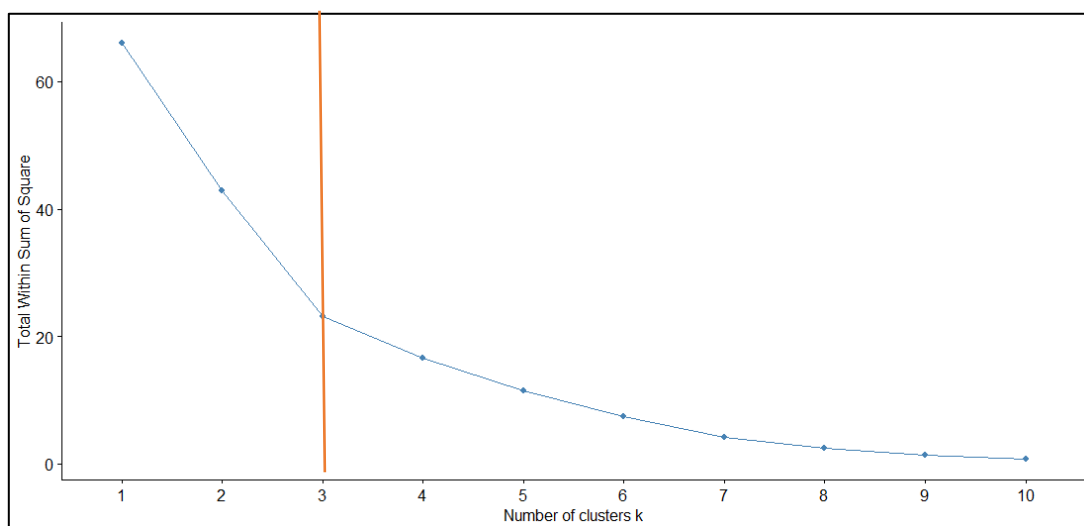
Figura 11: “Número óptimo de clústeres: Método del codo. Categoría: Actividad Económica Sectorial”



Fuente: Elaboración propia

Finalmente, en la última categoría como se desprende de la Figura 12: “Número óptimo de clústeres: Método del codo. Categoría: Salud y Liquidez Financiera” el número óptimo de clústeres estaba localizado en tres.

Figura 12: “Número óptimo de clústeres: Método del codo. Categoría: Salud y Liquidez Financiera”

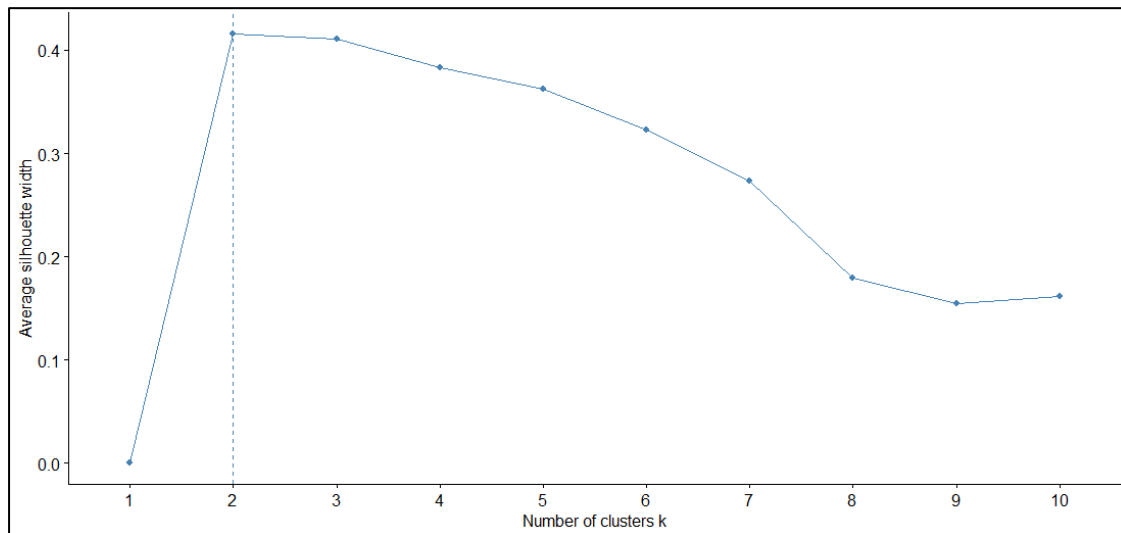


Fuente: Elaboración propia

Por su parte, por el método de la silueta se midió la distancia de las observaciones con su clúster, en comparación con la distancia entre los clústeres vecinos. En consecuencia, se buscó definir el acierto o el fallo de las asignaciones de cada una de las observaciones en cada uno de los clústeres. A través de este método, se logró definir el número de clústeres de una forma más óptima que en la lograda con la técnica del codo.

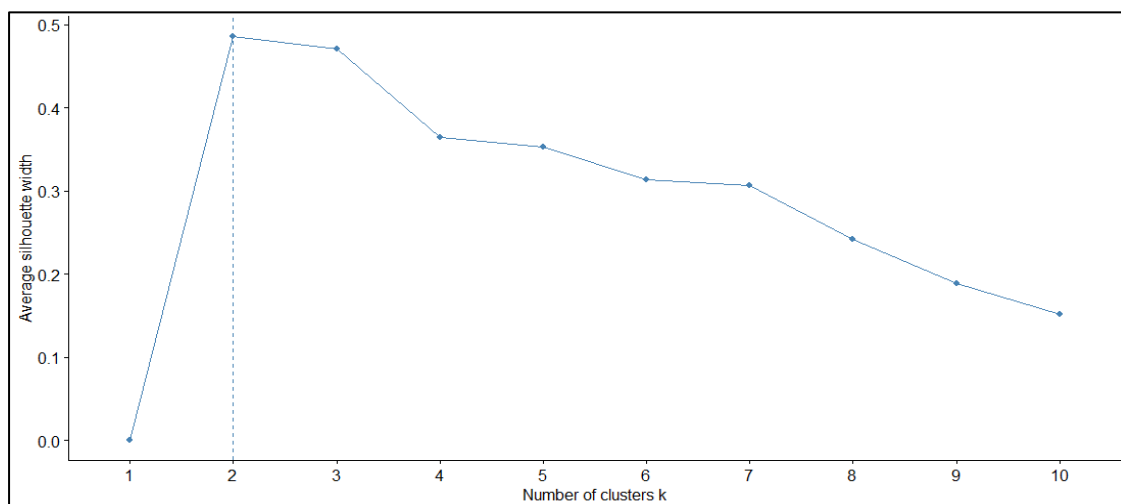
Así las cosas, se observó que en la primera y en la segunda categoría el número de clústeres óptimo eran dos. Por su parte, la tercera categoría situó en tres su número de clústeres óptimos.

Figura 13: “Número óptimo de clústeres: Método de la silueta. Categoría: Actividad Económica Sectorial”



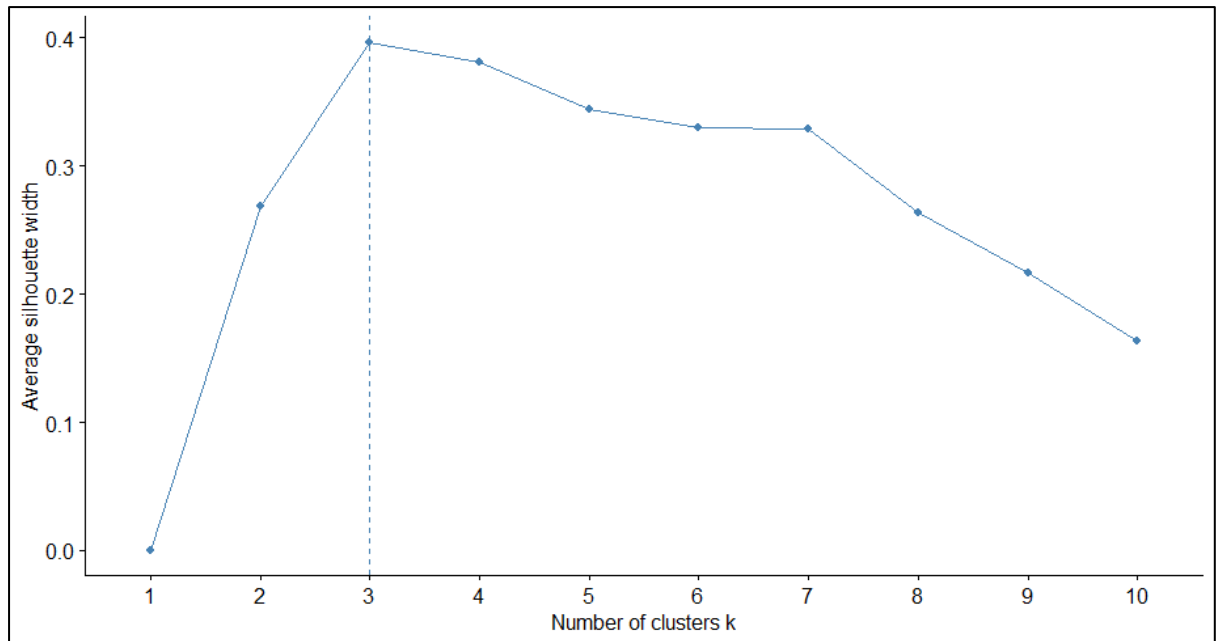
Fuente: Elaboración propia

Figura 14: “Número óptimo de clústeres: Método de la silueta. Categoría: Actividad Económica Sectorial”



Fuente: Elaboración propia

Figura 15: “Número óptimo de clústeres: Método de la silueta. Categoría: Salud y Liquidez Financiera”



Fuente: Elaboración propia

Una vez analizados ambos métodos, se observa que en este caso los resultados arrojados han sido los mismos (para las dos primeras categorías dos clústeres y para la tercera tres).

Localizado el número óptimo de clústeres, en la primera categoría la información se distribuyó de la siguiente forma:

Tabla 16: “Número óptimo de Clústeres en la categoría: Actividad Económica”

Primer Clúster	Segundo Clúster
2 variables	10 variables

Fuente: Elaboración propia

En la segunda categoría los clústeres se conformaron de esta manera:

Tabla 17: “Número óptimo de Clústeres en la categoría: Actividad Económica Sectorial”

Primer Clúster	Segundo Clúster
2 variables	10 variables

Fuente: Elaboración propia

Finalmente, en la tercera categoría los clústeres se desarrollaron de este modo:

Tabla 18: “Número óptimo de Clústeres en la categoría: Liquidez y Salud Financiera”

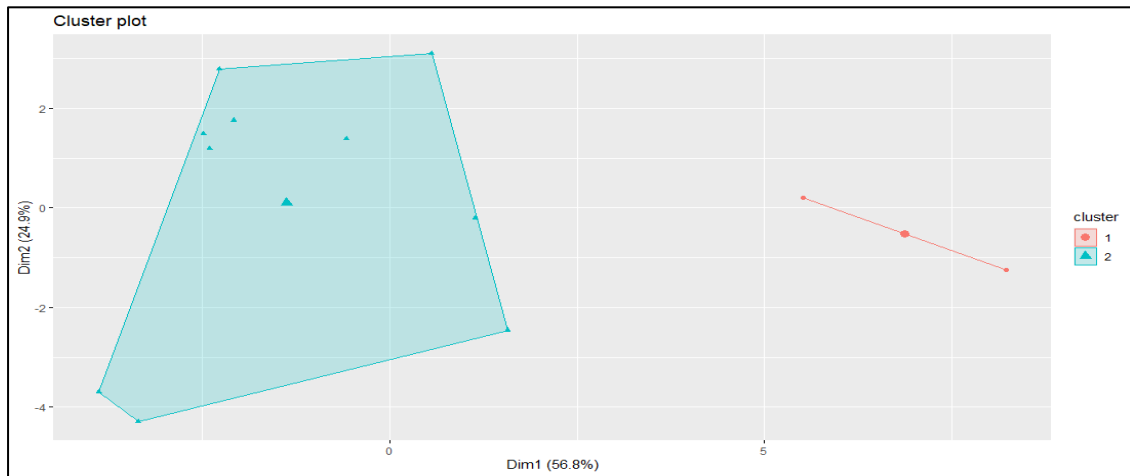
Primer Clúster	Segundo Clúster	Tercer Clúster

4 variables	2 variables	6 observaciones
-------------	-------------	-----------------

Fuente: Elaboración propia

Gráficamente y haciendo uso de técnicas PCA podemos observar que los clústeres de cada categoría quedaron separados de la siguiente manera. En la primera categoría, como se observa en la Figura 16: “Representación gráfica de los clústeres de la categoría: Actividad Económica”, los clústeres se mostraron alejados, con un primer grupo mucho mayor que el segundo.

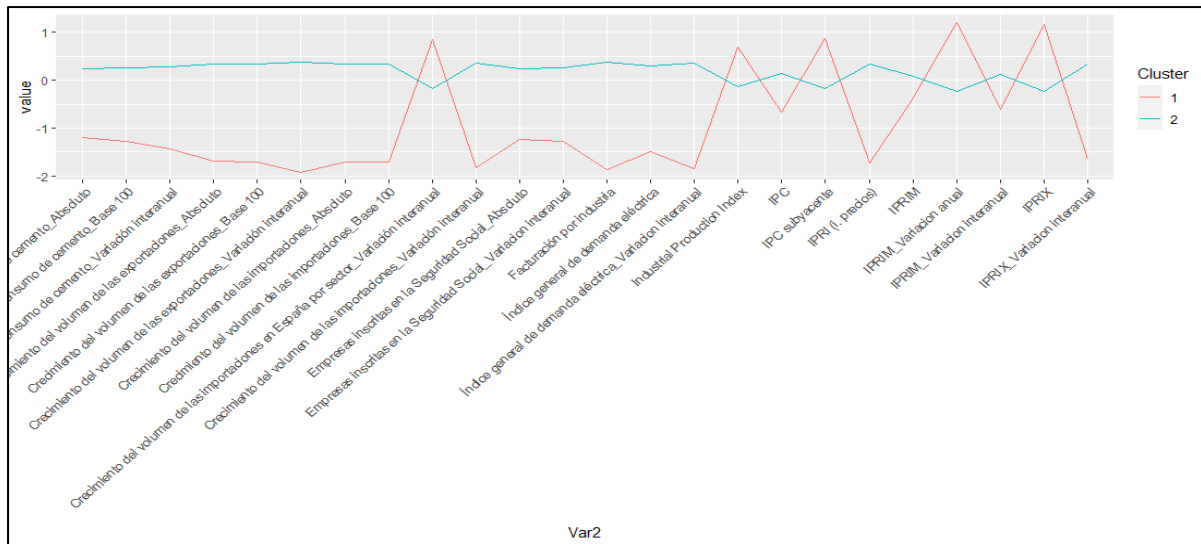
Figura 16: “Representación gráfica de los clústeres de la categoría: Actividad Económica”



Fuente: Elaboración propia

Por su parte, las variables dentro de los clústeres de esta categoría se distribuyeron de la siguiente forma:

Figura 17: “Representación gráfica de las variables en los clústeres de la categoría: Actividad Económica”



Fuente: Elaboración propia

En resumen, se observa que los Clústeres se relacionan de la siguiente manera:

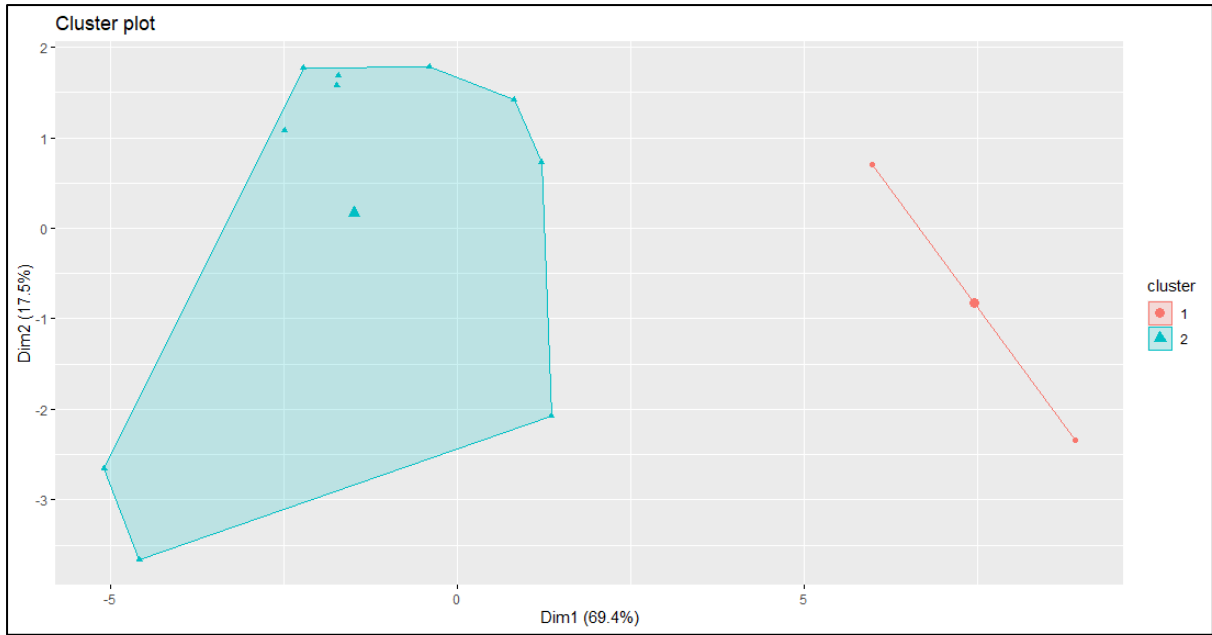
- El Clúster 1 se caracterizó por:
 - Valores bajos de los indicadores relativos: al cemento, las exportaciones, el número de empresas inscritas en la Seguridad Social, la FI, el IGDE, el IPRI y el IPRIX_VI.
 - Valores altos en el IMPRIM interanual, en el IPC_sb y en el IPRIM anual.

A destacar que, si bien se reflejaron niveles bajo generales respecto a las importaciones, se recogió un valor alto en el volumen de las importaciones de variación interanual.

- El Clúster 2 mostró valores más cercanos al centroide, lo que nos indica que los valores de los centroides coinciden con la media de las variables estandarizadas analizadas en el conjunto.

Por otro lado, la segunda categoría también reflejó dos clústeres muy separados, y al mismo tiempo, con un volumen de datos muy distinto entre ambos.

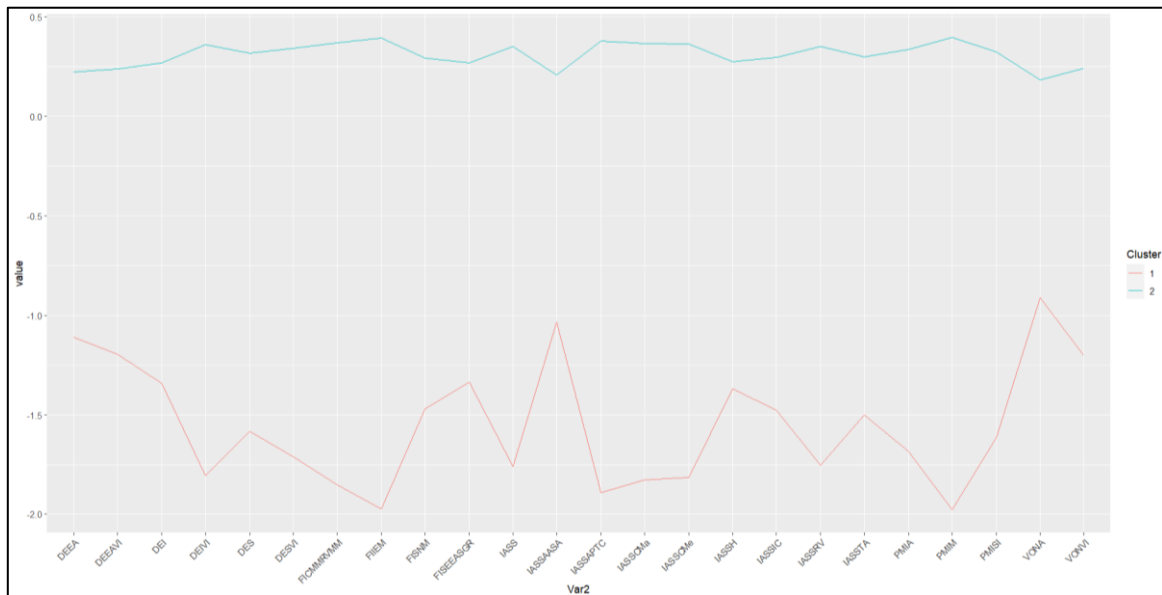
Figura 18: “Representación gráfica de los clústeres de la categoría: Actividad Económica Sectorial”



Fuente: Elaboración propia

Las variables dentro de los clústeres de esta categoría se agruparon de la siguiente manera:

Figura 19: “Representación gráfica de las variables en los clústeres de la categoría: Actividad Económica Sectorial”



Fuente: Elaboración propia

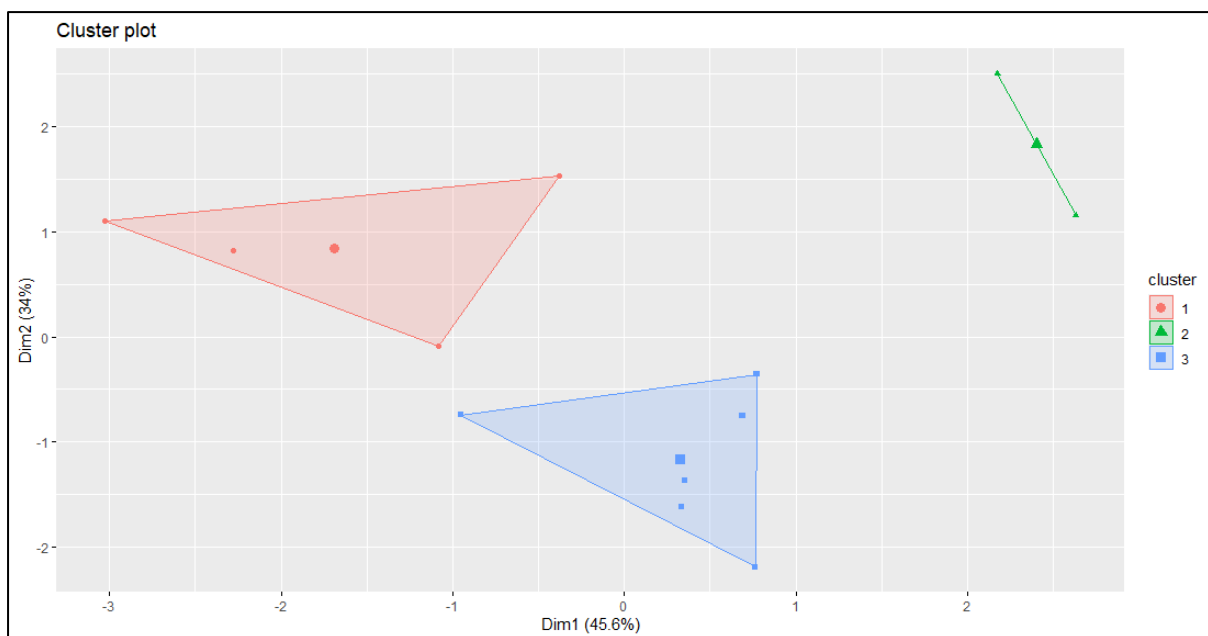
La segunda categoría reflejó los siguientes resultados:

- El Clúster 1 recogió valores bajos de todos los indicadores.

- El Clúster 2 contuvo los valores más cercanos al centroide, lo que nos indica que los valores de los centroides coinciden con la media de las variables estandarizadas analizadas en conjunto.

En la última categoría, como se observa en la Figura 20: “Representación gráfica de los clústeres de la categoría: Liquidez y Salud Financiera”, dos de los clústeres se encontraban a corta distancia mientras que el tercero se localizaba a una distancia mayor. En este caso, el volumen de las observaciones estaba más repartido que en las categorías anteriores lo que denota la idoneidad de la categoría a este tipo de análisis multivariante.

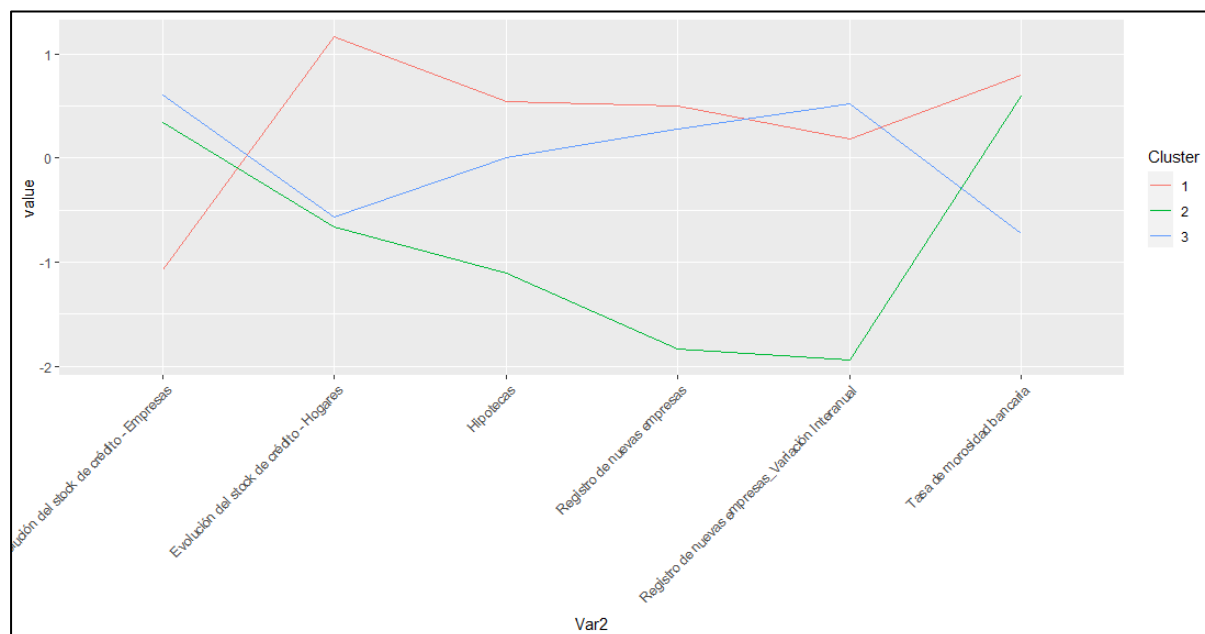
Figura 20: “Representación gráfica de los clústeres de la categoría: Liquidez y Salud Financiera”



Fuente: Elaboración propia

Las variables dentro de los clústeres de esta categoría se distribuyeron del siguiente modo:

Figura 21: “Representación gráfica de las variables en los clústeres de la categoría: Liquidez y Salud Financiera”



Fuente: Elaboración propia

Se pueden sacar las siguientes conclusiones de este gráfico:

- El Clúster 1 se caracteriza por una evolución alta del stock de créditos de los hogares por una baja evolución del stock de crédito de las empresas. Asimismo, el indicado TMB muestra valores altos.
- El Clúster 2, por su parte, se caracteriza por un valor bajo de los registros de nuevas empresas tanto de forma absoluta como en la variación interanual con una elevada TMB.
- El Clúster 3 contiene los valores más cercanos al centroide, lo que nos indica que los valores de los centroides coinciden con la media de las variables estandarizadas analizadas en conjunto.

Finalizado el análisis clúster² se procedió a realizar un análisis factorial.

- Análisis factorial

Este análisis requiere de una alta correlación entre los datos que vayan a ser tratados. Por ello, previo a la elección del número de factores se han de tratar los datos originales. Para ello, se ha llevado a cabo la matriz de correlación entre los distintos indicadores de cada una de las categorías.

Tanto la primera como la segunda categoría reflejaron una alta correlación entre los indicadores que las hacía óptimas para ser sujetos del análisis factorial. A continuación, se ofrece muestran las matrices de correlación de estas categorías. El código de colores que se va a observar en las Figura 22: “*Muestra de la matriz de correlación de la categoría: Actividad Económica*” Figura 23: “*Muestra de la matriz de correlación de la categoría: Actividad Económica Sectorial*” y Figura 24: “*Muestra de la matriz de correlación de la categoría: Liquidez y Salud Financiera*” sigue este esquema:

Tabla 19: “*Código de colores para las matrices de correlación*”

Rojo	Valores con una correlación mayor del 90%
Amarillo	Valores con una correlación mayor del 70%
Verde	Valores con una correlación mayor del 50%

Fuente: Elaboración propia

² Para ver el código empleado en este análisis clúster véase el Anexo 5.

Figura 22: “Muestra de la matriz de correlación de la categoría: Actividad Económica”

	CCA	CCB	CCVI	CVEA	CVLB	CVLEVI	CVESA	CVLESB	CVLESV	CVLEVI	EISSA	EISSV	FI	IGDE	IGDEVI	IPI	IPC	IPC_sb	IPRI	IPRIM	IPRIM_VA	IPRIM_VI	IPRIX	IPRIX_VI
CCA	1,00	0,97	0,82	0,64	0,62	0,53	0,43	0,43	-0,34	0,36	0,51	0,47	0,51	0,46	0,49	-0,38	-0,08	-0,36	0,28	-0,23	-0,37	-0,16	-0,38	0,27
CCB	0,97	1,00	0,91	0,67	0,67	0,62	0,45	0,45	-0,38	0,43	0,56	0,49	0,56	0,45	0,57	-0,36	-0,11	-0,46	0,37	-0,27	-0,45	-0,20	-0,42	0,34
CCVI	0,82	0,91	1,00	0,65	0,67	0,79	0,52	0,52	-0,42	0,63	0,72	0,62	0,73	0,51	0,70	-0,32	0,07	-0,45	0,63	-0,08	-0,51	-0,01	-0,46	0,58
CVEA	0,64	0,67	0,65	1,00	0,83	0,83	0,93	0,93	-0,17	0,81	0,65	0,64	0,80	0,89	0,83	0,03	0,24	-0,36	0,68	0,17	-0,44	0,24	-0,30	0,68
CVLB	0,62	0,67	0,67	1,00	1,00	0,85	0,93	0,93	-0,18	0,82	0,66	0,64	0,81	0,88	0,85	0,04	0,22	-0,38	0,70	0,15	-0,46	0,23	-0,31	0,70
CVLEVI	0,53	0,62	0,79	0,83	0,85	1,00	0,87	0,87	-0,40	0,96	0,83	0,81	0,97	0,78	0,95	-0,22	0,39	-0,42	0,95	0,29	-0,57	0,37	-0,52	0,92
CVESA	0,43	0,45	0,52	0,93	0,93	0,87	1,00	1,00	-0,19	0,91	0,72	0,75	0,89	0,94	0,90	0,02	0,46	-0,26	0,82	0,43	-0,41	0,49	-0,33	0,84
CVLESB	0,43	0,45	0,52	0,93	0,93	0,87	1,00	1,00	-0,19	0,91	0,72	0,75	0,89	0,94	0,90	0,02	0,46	-0,26	0,82	0,43	-0,41	0,49	-0,33	0,84
CVLESV	-0,34	-0,38	-0,42	-0,17	-0,18	-0,40	-0,19	-0,19	1,00	-0,28	-0,23	-0,34	-0,29	-0,10	-0,42	0,96	0,41	0,84	-0,41	0,45	0,91	0,42	0,98	-0,28
CVLEVI	0,36	0,43	0,63	0,81	0,82	0,96	0,91	0,91	-0,28	1,00	0,80	0,81	0,97	0,84	0,95	-0,08	0,52	-0,28	0,96	0,47	-0,47	0,54	-0,41	0,96
EISSA	0,51	0,56	0,72	0,65	0,66	0,83	0,72	0,72	-0,23	0,80	1,00	0,97	0,88	0,69	0,79	-0,14	0,51	-0,19	0,80	0,47	-0,29	0,49	-0,32	0,86
EISSV	0,47	0,49	0,62	0,64	0,64	0,81	0,75	0,75	-0,34	0,81	0,97	1,00	0,87	0,70	0,81	-0,26	0,48	-0,25	0,81	0,46	-0,38	0,49	-0,44	0,86
FI	0,51	0,56	0,73	0,80	0,81	0,97	0,89	0,89	-0,29	0,97	0,88	0,87	1,00	0,82	0,94	-0,17	0,56	-0,25	0,94	0,47	-0,43	0,54	-0,42	0,86
IGDE	0,46	0,45	0,51	0,89	0,88	0,78	0,94	0,94	-0,10	0,84	0,69	0,70	0,82	1,00	0,85	0,10	0,47	-0,13	0,72	0,46	-0,31	0,50	-0,21	0,78
IGDEVI	0,49	0,57	0,70	0,83	0,85	0,95	0,90	0,90	-0,42	0,95	0,79	0,81	0,94	0,85	1,00	-0,24	0,34	-0,44	0,92	0,30	-0,61	0,36	-0,54	0,92
IPI	-0,38	-0,36	-0,32	0,03	0,04	-0,22	0,02	0,02	0,90	-0,08	-0,14	-0,26	-0,17	0,10	-0,24	1,00	0,37	0,66	-0,22	0,44	0,71	0,39	0,89	-0,13
IPC	-0,08	-0,11	-0,01	0,24	0,23	0,37	0,49	0,49	0,42	0,52	0,51	0,49	0,56	0,47	0,34	0,37	1,00	0,62	0,48	0,97	0,42	0,98	0,32	0,61
IPC_sb	-0,36	-0,46	-0,45	-0,36	-0,38	-0,42	-0,26	-0,26	0,84	-0,28	-0,19	-0,25	-0,25	-0,13	-0,44	0,66	0,62	1,00	-0,37	0,63	0,93	0,58	0,85	-0,21
IPRI	0,28	0,37	0,63	0,68	0,70	0,95	0,82	0,82	-0,41	0,96	0,80	0,81	0,94	0,72	0,92	-0,22	0,48	-0,37	1,00	0,43	-0,53	0,50	-0,52	0,97
IPRIM	-0,23	-0,27	-0,08	0,17	0,15	0,29	0,43	0,43	0,45	0,47	0,47	0,46	0,47	0,46	0,30	0,44	0,97	0,63	0,43	1,00	0,45	0,99	0,37	0,58
IPRIM_VA	-0,37	-0,45	-0,51	-0,44	-0,46	-0,57	-0,41	-0,41	0,91	-0,47	-0,29	-0,38	-0,43	-0,31	-0,61	0,71	0,42	0,93	-0,53	0,45	1,00	0,41	0,93	-0,41
IPRIM_VI	-0,16	-0,20	-0,01	0,24	0,23	0,37	0,49	0,49	0,42	0,54	0,49	0,49	0,54	0,50	0,36	0,39	0,98	0,58	0,50	0,99	0,41	1,00	0,31	0,62
IPRIX	-0,38	-0,42	-0,46	-0,30	-0,31	-0,52	-0,33	-0,33	0,98	-0,41	-0,32	-0,44	-0,42	-0,21	-0,54	0,89	0,32	0,85	-0,52	0,37	0,93	0,31	1,00	-0,40
IPRIX_VI	0,27	0,34	0,58	0,68	0,70	0,92	0,84	0,84	-0,28	0,96	0,86	0,86	0,96	0,78	0,92	-0,13	0,61	-0,21	0,97	0,58	-0,41	0,62	-0,40	1,00

Fuente: Elaboración propia

Figura 23: “Muestra de la matriz de correlación de la categoría: Actividad Económica Sectorial”

	DEEA	DEEAVI	DEI	DEIVI	DES	DESVI	FICMMRVMM	FIEM	FISNM	FISEEASGR	IASS	IASSAASA	IASSAPTC	IASSCMa	IASSCMe	IASSH	IASSIC	IASSRV	IASSTA	PMIA	PMIM	PMISI	VONA	VONVI
DEEA	1,00	0,99	0,53	0,49	0,84	0,83	0,69	0,63	0,92	0,30	0,81	0,94	0,74	0,78	0,63	0,94	0,86	0,49	0,89	0,44	0,40	0,45	0,25	0,51
DEEAVI	0,99	1,00	0,49	0,51	0,84	0,85	0,74	0,66	0,92	0,30	0,84	0,92	0,76	0,81	0,67	0,96	0,88	0,55	0,89	0,53	0,46	0,53	0,25	0,47
DEI	0,53	0,49	1,00	0,77	0,79	0,67	0,70	0,76	0,68	0,39	0,70	0,62	0,67	0,73	0,69	0,57	0,63	0,63	0,70	0,48	0,63	0,45	0,31	0,67
DEIVI	0,49	0,51	0,77	1,00	0,82	0,84	0,85	0,93	0,73	0,74	0,83	0,55	0,91	0,83	0,85	0,60	0,74	0,82	0,76	0,74	0,89	0,70	0,52	0,69
DES	0,84	0,84	0,79	0,82	1,00	0,96	0,86	0,87	0,96	0,59	0,93	0,90	0,93	0,91	0,83	0,88	0,94	0,71	0,96	0,62	0,67	0,60	0,52	0,78
DESVI	0,83	0,85	0,67	0,84	0,96	1,00	0,90	0,90	0,95	0,62	0,95	0,83	0,97	0,93	0,87	0,90	0,94	0,77	0,94	0,72	0,75	0,71	0,52	0,72
FICMMRVMM	0,69	0,74	0,70	0,85	0,86	0,90	1,00	0,96	0,87	0,53	0,98	0,71	0,91	0,99	0,99	0,86	0,86	0,95	0,88	0,92	0,89	0,91	0,27	0,54
FIEM	0,63	0,66	0,76	0,93	0,87	0,90	0,96	1,00	0,84	0,67	0,94	0,66	0,95	0,95	0,95	0,76	0,84	0,91	0,85	0,84	0,90	0,82	0,37	0,63
FISNM	0,92	0,92	0,68	0,73	0,96	0,95	0,87	0,84	1,00	0,51	0,95	0,96	0,90	0,93	0,83	0,95	0,98	0,71	0,99	0,65	0,61	0,65	0,39	0,63
FISEEASGR	0,30	0,30	0,39	0,74	0,59	0,62	0,53	0,67	0,51	1,00	0,54	0,38	0,74	0,53	0,53	0,31	0,51	0,48	0,54	0,37	0,56	0,33	0,49	0,77
IASS	0,81	0,84	0,70	0,83	0,93	0,95	0,98	0,94	0,95	0,54	1,00	0,83	0,94	0,99	0,96	0,93	0,93	0,88	0,95	0,84	0,80	0,83	0,33	0,59
IASSAASA	0,94	0,92	0,62	0,55	0,90	0,83	0,71	0,66	0,96	0,38	0,83	1,00	0,75	0,81	0,65	0,90	0,91	0,49	0,95	0,43	0,37	0,44	0,32	0,56
IASSAPTC	0,74	0,76	0,67	0,91	0,93	0,97	0,91	0,95	0,90	0,74	0,94	0,75	1,00	0,93	0,88	0,81	0,91	0,82	0,90	0,75	0,83	0,72	0,53	0,73
IASSCMa	0,78	0,81	0,73	0,83	0,91	0,93	0,99	0,95	0,93	0,53	0,99	0,81	0,93	1,00	0,96	0,90	0,91	0,89	0,94	0,85	0,83	0,84	0,30	0,57
IASSCMe	0,63	0,67	0,69	0,85	0,83	0,87	0,99	0,95	0,83	0,53	0,96	0,65	0,88	0,96	1,00	0,82	0,81	0,96	0,83	0,94	0,90	0,93	0,26	0,55
IASSH	0,94	0,96	0,57	0,60	0,88	0,90	0,86	0,76	0,95	0,31	0,93	0,90	0,81	0,90	0,82	1,00	0,91	0,72	0,92	0,70	0,60	0,71	0,25	0,49
IASSIC	0,86	0,88	0,63	0,74	0,94	0,94	0,86	0,84	0,98	0,51	0,93	0,91	0,91	0,91	0,91	0,81	0,91	1,00	0,72	0,97	0,66	0,62	0,66	0,48
IASSRV	0,49	0,55	0,63	0,82	0,71	0,77	0,95	0,91	0,71	0,48	0,88	0,49	0,82	0,89	0,96	0,72	0,72	1,00	0,71	0,97	0,93	0,96	0,20	0,46
IASSTA	0,89	0,89	0,70	0,76	0,96	0,94	0,88	0,85	0,99	0,54	0,95	0,95	0,90	0,94	0,83	0,92	0,97	0,71	1,00	0,65	0,62	0,65	0,38	0,62
PMIA	0,44	0,53	0,48	0,74	0,62	0,72	0,92	0,84	0,65	0,37	0,84	0,43	0,75	0,85	0,94	0,70	0,66	0,97	0,65	1,00	0,90	1,00	0,17	0,29
PMIM	0,40	0,46	0,63	0,89	0,67	0,75	0,89	0,90	0,61	0,56	0,80	0,37	0,83	0,83	0,90	0,60	0,62	0,93	0,62	0,90	1,00	0,87	0,37	0,51
PMISI	0,45	0,53	0,45	0,70	0,60	0,71	0,91	0,82	0,65	0,33	0,83	0,44	0,72	0,84	0,93	0,71	0,66	0,96	0,65					

Por su parte, la tercera categoría reflejó una correlación considerablemente menor a la mostrada por las anteriores categorías.

Figura 24: “Muestra de la matriz de correlación de la categoría: Liquidez y Salud Financiera”

	Evolución.del.stock.de .crédito...Empresas	Evolución.del.stock.de.c rédito...Hogares	Hipotecas	Registro.de.nuevas.empr esas	Registro.de.nuevas.empresa s_Variación.Interanual	Tasa.de.morosidad.banc aria
Evolución.del.stock.de.crédito.. .Empresas	1,00	-0,52	-0,44	-0,34	0,07	-0,78
Evolución.del.stock.de.crédito.. .Hogares	-0,52	1,00	0,47	0,28	0,27	0,31
Hipotecas	-0,44	0,47	1,00	0,55	0,56	0,21
Registro.de.nuevas.empresas	-0,34	0,28	0,55	1,00	0,76	-0,12
Registro.de.nuevas.empresas_ Variación.Interanual	0,07	0,27	0,56	0,76	1,00	-0,44
Tasa.de.morosidad.bancaria	-0,78	0,31	0,21	-0,12	-0,44	1,00

Fuente: Elaboración propia

Una vez analizadas las matrices de correlación, de las dos primeras categorías se pudo concluir que ambas cuentan con indicadores con una correlación extremadamente alta (rondando el 1), y por ello, se consideró necesario eliminar aquellos KPI que no aportaban valor añadido al ofrecer información duplicada.

En la categoría “**Actividad Económica**” se decidió eliminar los siguientes indicadores:

- “**CCB**”. Debido a que contaba con una correlación de 0,97 y 0,91 con los indicadores “**CCA**” y “**CCVI**”. Asimismo, estos dos últimos indicadores reflejaron los mismos niveles de correlación con el resto de la base de datos que aquellos niveles mostrados por el “**CCB**”. Por todo ello, se pudo concluir que la información aportada por este indicador estaba duplicada en los otros dos KPIs.

En la categoría “**Actividad Económica Sectorial**” si bien existían indicadores con una correlación muy alta entre sí, la relación con el resto de la base de datos propiciaba que no fuera prudente eliminar ningún indicador puesto que podía afectar sensiblemente a la construcción del indicador compuesto.

Una vez tratados los datos y comprobada la correlación, se pasó al análisis factorial propiamente dicho. Para decidir qué factores se han de seleccionar a la hora de llevar a cabo el análisis multivariante existen una serie de criterios que se pueden escoger.

- Criterio de Kaiser. A través de esta regla se han de eliminar aquellas variables cuyos autovalores se encuentren por debajo de 1. La razón de eliminar dichas variables es

evitar la inclusión de un factor que explique menos varianza que aquella dispuesta en un indicador individual (OECD, 2008).

- Diagrama de Scree. En este caso, se han de mostrar los autovalores gráficamente para observar la bajada de los autovalores hasta su nivelización (comúnmente llamado regla del codo) (OECD, 2008).
- Criterios de varianza explicada. Desde dicho método, el objetivo es mantener suficientes factores explicar el 90% de la varianza (OECD, 2008).
- Criterio de Joliffe. Esta técnica es similar al Criterio de Kaiser, pero a diferencia del primero en vez de eliminar los autovalores menores de 1 se han de eliminar los autovalores inferiores a 0.70 (OECD, 2008).
- Comprensibilidad. Desde este método la reducción de factores consiste en limitarlos a aquellos cuya dimensión de significado sea correctamente entendible (OECD, 2008).

Entendidos los métodos se procedió a realizar el análisis en SPSS y una vez vistos los resultados se escogió el método más oportuno.

Figura 25: “Análisis Multivariante categoría: Actividad Económica”

Componente	Varianza total explicada											
	Autovalores iniciales			Sumas de cargas al cuadrado de la extracción			Sumas de cargas al cuadrado de la rotación					
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado			
1	13,308	55,452	55,452	13,308	55,452	55,452	11,011	45,880	45,880			
2	6,599	27,495	82,947	6,599	27,495	82,947	5,270	21,958	67,839			
3	1,793	7,472	90,419	1,793	7,472	90,419	4,433	18,473	86,311			
4	1,142	4,758	95,177	1,142	4,758	95,177	2,128	8,866	95,177			
5	0,561	2,337	97,514									
6	0,337	1,404	98,917									
7	0,127	0,527	99,445									
8	0,064	0,268	99,713									
9	0,040	0,165	99,878									
10	0,019	0,080	99,958									
11	0,010	0,042	100,000									
12	8,469E-16	3,529E-15	100,000									
13	7,628E-16	3,178E-15	100,000									
14	5,695E-16	2,373E-15	100,000									
15	4,352E-16	1,813E-15	100,000									
16	2,735E-16	1,140E-15	100,000									
17	6,902E-17	2,876E-16	100,000									
18	-3,766E-17	-1,569E-16	100,000									
19	-1,233E-16	-5,137E-16	100,000									
20	-2,667E-16	-1,111E-15	100,000									
21	-3,694E-16	-1,539E-15	100,000									
22	-6,018E-16	-2,507E-15	100,000									
23	-6,731E-16	-2,804E-15	100,000									
24	-1,546E-15	-6,442E-15	100,000									

Fuente: Elaboración propia

Figura 26: “Análisis Multivariante categoría: Actividad Económica Sectorial”

Componente	Varianza total explicada								
	Autovalores iniciales			Sumas de cargas al cuadrado de la extracción			Sumas de cargas al cuadrado de la rotación		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	18,015	72,059	72,059	18,015	72,059	72,059	9,327	37,310	37,310
2	3,096	12,382	84,441	3,096	12,382	84,441	8,963	35,851	73,161
3	2,065	8,261	92,702	2,065	8,261	92,702	4,885	19,541	92,702
4	0,677	2,707	95,409						
5	0,524	2,097	97,505						
6	0,258	1,032	98,537						
7	0,164	0,655	99,192						
8	0,105	0,422	99,613						
9	0,052	0,207	99,821						
10	0,034	0,135	99,956						
11	0,011	0,044	100,000						
12	9,838E-16	3,934E-15	100,000						
13	8,137E-16	3,255E-15	100,000						
14	5,350E-16	2,140E-15	100,000						
15	4,283E-16	1,713E-15	100,000						
16	3,608E-16	1,443E-15	100,000						
17	7,983E-17	3,193E-16	100,000						
18	-2,804E-17	-1,122E-16	100,000						
19	-5,494E-17	-2,598E-16	100,000						
20	-9,888E-17	-3,955E-16	100,000						
21	-2,118E-16	-8,471E-16	100,000						
22	-3,033E-16	-1,213E-15	100,000						
23	-3,318E-16	-1,327E-15	100,000						
24	-6,046E-16	-2,418E-15	100,000						
25	-1,063E-15	-4,253E-15	100,000						

Fuente: Elaboración propia

Figura 27: “Análisis Multivariante categoría: Actividad Económica Sectorial”

Componente	Varianza total explicada								
	Autovalores iniciales			Sumas de cargas al cuadrado de la extracción			Sumas de cargas al cuadrado de la rotación		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	3,128	44,680	44,680	3,128	44,680	44,680	3,066	43,794	43,794
2	2,549	36,413	81,094	2,549	36,413	81,094	2,611	37,300	81,094
3	0,636	9,085	90,179						
4	0,445	6,354	96,533						
5	0,136	1,947	98,480						
6	0,092	1,310	99,790						
7	0,015	0,210	100,000						

Fuente: Elaboración propia

A la luz de los resultados obtenidos, se discutió el método más adecuado para elegir los factores. De esta forma, tanto el diagrama de Scree como el método de la Comprensibilidad fueron descartados por su carácter subjetivo y por ser menos efectivos en comparación con el resto de los métodos.

El criterio de Joliffe fue descartado debido a que escogía los factores en base al número capaz de explicar el 70 % de la varianza explicada. De esta forma, se corría el riesgo de perder información relevante en el proceso del análisis.

Descartados el resto de criterios, se decidió proceder a determinar el número de los factores a través de los criterios de varianza explicada (más del 90%) y del criterio de Kaiser.

En la primera categoría, por el Criterio de Kaiser se obtuvieron cuatro factores cuyos autovalores se encontraban por encima de 1. Por su parte, con los Criterios de varianza

explicada se obtuvieron tres factores con los que se explicaba el 90% de la varianza. Entre ambos criterios se decidió proceder con el primero puesto que a través de la configuración de cuatro factores se podía explicar de forma más razonada la categoría (dicha subdivisión reflejaba una mayor relación entre los indicadores).

En la segunda categoría, por ambos criterios se obtuvieron tres factores por lo que no es necesario elegir un método en concreto.

En la tercera categoría, por el Criterio de Kaiser se obtuvieron dos factores cuyos autovalores se encontraban por encima de 1. Por su parte, con los Criterios de varianza explicada se obtuvieron tres factores con los que se explicaba el 90% de la varianza. Entre ambos criterios se decidió proceder con el primero pues ofrecía una separación más relevante en cada uno de los factores.

Una vez escogido el número de factores, se requiere elegir una rotación que favorezca la explicabilidad de los resultados. Para llevar a cabo la rotación existen múltiples opciones, siendo la más común la “rotación varimax” que será utilizada para conocer la forma en la que se habían agrupado los indicadores por factor. Los resultados de este proceso se pueden ver a continuación (resaltado en color amarillo: el factor al que se ha asignado cada indicador):

Figura 28: “Matriz Rotada: Actividad Económica”

	Matriz de componente rotado*									
	1	2	3	4						
Consumo de cemento_Absoluto	-0,172	0,936	0,168	0,073	Índice general de demanda eléctrica	-0,042	0,989	0,105	-0,069	
Consumo de cemento_Variación interanual	0,702	0,042	0,216	0,413	Índice general de demanda eléctrica_Variación interanual	0,941	-0,082	0,239	0,103	
Crecimiento del volumen de las exportaciones_Absoluto	0,112	0,974	0,179	0,040	Industrial Production Index	-0,053	-0,211	-0,943	0,020	
Crecimiento del volumen de las exportaciones_Base 100	0,121	0,972	0,183	0,058	IPC	-0,603	-0,040	0,503	0,601	
Crecimiento del volumen de las exportaciones_Variación interanual	0,950	0,060	0,242	0,100	IPC subyacente	0,297	-0,371	0,712	0,478	
Crecimiento del volumen de las importaciones_Absoluto	0,160	0,982	0,170	-0,089	IPRI (i. precios)	-0,941	0,241	-0,168	0,029	
Crecimiento del volumen de las importaciones_Base 100	0,920	0,160	-0,121	0,132	IPRIM	0,200	-0,965	-0,008	0,123	
Crecimiento del volumen de las importaciones en España por sector_Variación interanual	0,169	0,221	0,949	0,092	IPRIM_Variación anual	0,265	0,380	0,811	0,298	
Crecimiento del volumen de las importaciones_Variación interanual	0,977	-0,063	0,130	-0,022	IPRIM_Variación interanual	-0,628	0,389	0,448	0,500	
Empresas inscritas en la Seguridad Social_Absoluto	-0,195	0,970	0,071	-0,059	IPRIX	-0,287	-0,348	-0,886	-0,069	
Empresas inscritas en la Seguridad Social_Variación interanual	0,860	0,134	0,153	-0,149	IPRIX_Variación interanual	0,969	-0,168	0,087	-0,132	
Facturación por industria	0,984	-0,067	0,102	-0,051						

Fuente: Elaboración propia

Figura 29: “Matriz Rotada: Actividad Económica Sectorial”

	Matriz componente rotado				Componente		
	1	2	3		1	2	3
Demanda eléctrica en establecimientos de alojamiento	0.005	0.105	0.005	Indicadores de actividad del sector servicios - Actividades profesionales técnicas y científicas	0.040	0.033	0.070
Demanda eléctrica en establecimientos de alojamiento_Variación interanual	0.010	0.099	0.003	Indicadores de actividad del sector servicios - Comercio al por mayor	0.067	0.046	0.017
Demanda eléctrica en industria	0.020	0.022	0.044	Indicadores de actividad del sector servicios - Comercio al por menor	0.079	0.022	0.015
Demanda eléctrica en industria_Variación interanual	0.051	0.008	0.090	Indicadores de actividad del sector servicios - Hostelería	0.028	0.084	0.002
Demanda eléctrica en servicios	0.021	0.062	0.056	Indicadores de actividad del sector servicios - Información y comunicaciones	0.024	0.068	0.030
Demanda eléctrica en servicios_Variación interanual	0.032	0.051	0.051	Indicadores de actividad del sector servicios - Reparación de vehículos	0.094	0.006	0.009
Facturación por industria - Comercio al por mayor y al por menor; reparación de vehículos de motor y motocicletas	0.073	0.030	0.015	Indicadores de actividad del sector servicios - Transporte y almacenamiento	0.024	0.074	0.027
Facturación por industria - Industrias extractivas y manufactureras	0.064	0.020	0.044	Purchasing Manager's Index (PMI) of Aggregated	0.100	0.006	0.001
Facturación por industria - Servicios no financieros de mercado	0.022	0.080	0.024	Purchasing Manager's Index (PMI) of Manuf	0.084	0.002	0.031
Facturación por industria - Suministro de energía eléctrica y agua, saneamiento y gestión de residuos	0.011	0.001	0.148	Purchasing Manager's Index (PMI) of Services Índice de gestores de compras	0.008	0.007	0.000
Indicadores de actividad del sector servicios	0.053	0.050	0.019	Visados de obra nueva_Absoluto	0.000	0.003	0.134
Indicadores de actividad del sector servicios - Actividades administrativas y servicios auxiliares	0.005	0.102	0.013	Visados de obra nueva_Variación interanual	0.003	0.016	0.150

Fuente: Elaboración propia

Figura 30: “Matriz Rotada: Liquidez y Salud Financiera”

	Matriz de componente rotado ^a	
	1	2
Evolución del stock de crédito - Empresas	0,883	-0,348
Evolución del stock de crédito - Hogares	-0,452	0,555
Hipotecas	-0,234	0,811
Registro de nuevas empresas	0,023	0,867
Registro de nuevas empresas_Variación Interanual	0,419	0,868
Tasa de morosidad bancaria	-0,961	-0,088

Fuente: Elaboración propia

Una vez hecha la matriz rotada, se ha de encontrar el porcentaje de varianza explicada escalado a la unidad para evitar que los indicadores tiendan a incluirse en el primer factor (al tener éste la varianza explicada más elevada). Los resultados de este proceso se pueden ver a continuación (resaltado en color amarillo: el factor al que se ha asignado cada indicador):

En la primera categoría una serie de indicadores han cambiado de factor.

Figura 31: “Escala a la unidad: Actividad Económica”

	Escala a la unidad					Escala a la unidad			
	1	2	3	4		1	2	3	4
Consumo de cemento_Absoluto	0.003	0.137	0.006	0.004	Índice general de demanda eléctrica	0.000	0.152	0.002	0.004
Consumo de cemento_Variación interanual	0.056	0.000	0.010	0.138	Índice general de demanda eléctrica_Variación interanual	0.101	0.001	0.012	0.009
Crecimiento del volumen de las exportaciones_Absoluto	0.001	0.148	0.007	0.001	Industrial Production Index	0.000	0.007	0.194	0.000
Crecimiento del volumen de las exportaciones_Base 100	0.002	0.147	0.007	0.003	IPC	0.041	0.000	0.055	0.292
Crecimiento del volumen de las exportaciones_Variación interanual	0.103	0.001	0.013	0.008	IPC subyacente	0.010	0.021	0.110	0.185
Crecimiento del volumen de las importaciones_Absoluto	0.003	0.000	0.006	0.006	IPRI (i. precios)	0.101	0.009	0.006	0.001
Crecimiento del volumen de las importaciones_Base 100	0.095	0.004	0.003	0.014	IPRIM	0.005	0.145	0.000	0.012
Crecimiento del volumen de las importaciones en España por sector_Variación interanual	0.003	0.008	0.196	0.007	IPRIM_Variación anual	0.008	0.022	0.143	0.072
Crecimiento del volumen de las importaciones_Variación interanual	0.108	0.001	0.004	0.000	IPRIM_Variación interanual	0.045	0.024	0.044	0.202
Empresas inscritas en la Seguridad Social_Absoluto	0.004	0.147	0.001	0.003	IPRIX	0.009	0.019	0.171	0.004
Empresas inscritas en la Seguridad Social_Variación interanual	0.084	0.003	0.005	0.018	IPRIX_Variación interanual	0.107	0.004	0.002	0.014
Facturación por industria	0.110	0.001	0.002	0.002					0.000

Fuente: Elaboración propia

En la segunda categoría, los cambios entre los indicadores y los factores han sido mayores, cambiando hasta seis indicadores con este método.

Figura 32: “Escala a la unidad: Actividad Económica Sectorial”

	Componente				Componente		
	1	2	3		1	2	3
Demanda eléctrica en establecimientos de alojamiento	0.005	0.107	0.005	Indicadores de actividad del sector servicios - Actividades profesionales técnicas y científicas	0.041	0.034	0.082
Demanda eléctrica en establecimientos de alojamiento_Variación interanual	0.010	0.100	0.003	Indicadores de actividad del sector servicios - Comercio al por mayor	0.057	0.047	0.020
Demanda eléctrica en industria	0.020	0.022	0.052	Indicadores de actividad del sector servicios - Comercio al por menor	0.079	0.022	0.018
Demanda eléctrica en industria_Variación interanual	0.051	0.008	0.105	Indicadores de actividad del sector servicios - Hostelería	0.028	0.085	0.003
Demanda eléctrica en servicios	0.021	0.063	0.069	Indicadores de actividad del sector servicios - Información y comunicaciones	0.024	0.069	0.035
Demanda eléctrica en servicios_Variación interanual	0.032	0.052	0.061	Indicadores de actividad del sector servicios - Reparación de vehículos	0.094	0.008	0.011
Facturación por industria - Comercio al por mayor y al por menor; reparación de vehículos de motor y motocicletas	0.073	0.031	0.018	Indicadores de actividad del sector servicios - Transporte y almacenamiento	0.024	0.076	0.032
Facturación por industria - Industrias extractivas y manufactureras	0.065	0.021	0.052	Purchasing Manager's Index (PMI) of Aggregated	0.100	0.006	0.001
Facturación por industria - Servicios no financieros de mercado	0.022	0.081	0.028	Purchasing Manager's Index (PMI) of Manufacturing	0.085	0.002	0.036
Facturación por industria - Suministro de energía eléctrica y agua, saneamiento y gestión de residuos	0.011	0.001	0.174	Purchasing Manager's Index (PMI) of Services Índice de gestores de compras	0.090	0.008	0.000
Indicadores de actividad del sector servicios	0.053	0.051	0.023	Visados de obra nueva_Absoluto	0.000	0.003	0.157
Indicadores de actividad del sector servicios - Actividades administrativas y servicios auxiliares	0.005	0.103	0.016	Visados de obra nueva_Variación interanual			

Fuente: Elaboración propia

En esta última categoría, la escala a la unidad no ha cambiado la asociación de los indicadores, ello puede deberse a que no es una categoría correlacionada y, por tanto, no apta para este tipo de análisis.

Figura 33: “Escala a la unidad: Liquidez y Salud Financiera”

	1	2
Evolución del stock de crédito - Empresas	0,364	0,047
Evolución del stock de crédito - Hogares	0,096	0,119
Hipotecas	0,026	0,253
Registro de nuevas empresas	0,000	0,289
Registro de nuevas empresas_Variación Interanual	0,082	0,290
Tasa de morosidad bancaria	0,432	0,003

Fuente: Elaboración propia

Así las cosas, en la primera categoría se puede observar que los indicadores se han agrupado de la siguiente forma:

Tabla 20: “Indicadores Actividad Económica”

Primer factor (importaciones, exportaciones, cemento y seguridad social)	Segundo factor (relación del volumen de las importaciones con las principales métricas industriales)	Tercer factor (métricas económicas más relacionadas con el poder económico de los ciudadanos)	Cuarto factor (métricas relacionadas con el consumo de cemento, las importaciones y métricas relacionadas con el IPC)
“CVEVI”	“CCA”	“CVIESVI”	“CCVI”
“CVIB”	“CVEA”	“IPI”	“CVIA”
“FI”	“CVEB”	“IPRIM_VI”	“IPC”
“IGDEVI”	“EISSA”	“IPRIX”	“IPC_sb”
“IPRI”	“IGDE”		“IPRIM_VA”
“IPRIX_VA”			
“EISSVI”			

Fuente: Elaboración propia

En la segunda categoría, los indicadores se han agrupado de esta forma:

Tabla 21: “Indicadores Actividad Económica”

Primer factor (facturaciones por industria e indicadores del sector servicios)	Segundo factor (demanda eléctrica de sectores más estáticos en interiores y sector servicios)	Tercer factor (demanda eléctrica y suministros)
“FICMMRVMM”	“DEEA”	“DEI”
“FIEM”	“DEEAVI”	“DEIVI”
“IASS”	“DES”	“ FISEEASGR”
“IASSCMa”	“FISNM”	“IASSAPTC”
“IASSCMe”	“IASSAASA”	“VONA”
“IASSRV”	“IASSH”	“VONVI”
“PMIA”	“IASSIC”	
“PMIM”	“IASSTA”	
“PMISI”		
“DESVI”		

Fuente: Elaboración propia

Por último, la tercera categoría se ha agrupado de la siguiente forma:

Tabla 22: “Indicadores Actividad Económica”

Primer factor	Segundo factor
“ESCE”	“ESCH”
“TMB”	“Hipotecas”
	“RNE”
	“TMB”

Fuente: Elaboración propia

En conclusión, tras haber analizado las tres categorías se demuestra que aquellas que contaban con una alta correlación entre sus datos eran aptas para este tipo de análisis, mientras que aquellas con una correlación baja funcionaban mejor con un análisis clúster. De esta forma, las dos primeras categorías funcionaban con el análisis factorial y la tercera con el análisis clúster. Para ver los análisis factoriales completos véase Anexo 7.

2.4. Normalización

La normalización es el proceso por el cual se transforman las observaciones de los indicadores, que suelen encontrarse en unidades distintas, a una misma escala que permita la comparación entre los datos (Pollesch & Dale, 2016).

Para llevar a cabo este proceso existen varios métodos, el objetivo radica en elegir aquel que, valorando las propiedades de las unidades de medidas, garantice la solidez de la comparativa frente a posibles valores atípicos (Ebert & Welsch, 2004).

- 1) Estandarización (o z-scores). Esta técnica se enfoca en la media y en la desviación típica. El fin último de este método es encontrar una escala común en la que la media es cero y la desviación típica es uno. A partir de ello, las desviaciones típicas podrán estar por encima o por debajo de la media (Gupta, S., 2021). La fórmula para esta técnica es la siguiente:

$$(4) \quad Z = \frac{x-\mu}{\sigma}$$

Donde:

- Z es el valor estandarizado de la variable X
 - x es el valor original de la variable
 - μ es la media de la variable
 - σ es la desviación estándar de la variable
- 2) Min Max. En este caso se busca desplazar las observaciones hacia la media del indicador. En concreto, se resta el valor de la observación menos el valor mínimo del indicador entre la resta del valor máximo y mínimo del indicador (Rojas, K., 2022). La fórmula para esta técnica es la siguiente:

$$(5) \quad Z = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

Donde:

- x_i es el valor original de la variable
- $\min(x)$ es el valor mínimo de la variable
- $\max(x)$ es el valor máximo de la variable
- Z es el valor normalizado

- 3) Distancia a una referencia. En este caso se divide el valor de la observación menos el valor de la referencia en un momento inicial por el valor de la referencia. A partir de esta técnica, la normalización tiene en cuenta la evolución de los indicadores en el tiempo (Comisión Europea, 2020). La fórmula para esta técnica es la siguiente:

$$(6) \quad y = \frac{x_i - x_r}{x_r}$$

Donde:

- x_i es el valor original de la variable
 - x_r es el valor de referencia utilizado para normalizar los datos
 - y es el valor normalizado de la variable
- 4) Normalización de valores por encima o por debajo de la media. En este caso se distribuyen los valores que están por encima, cerca o por debajo de un rango porcentual definido arbitrariamente en torno a la media. De esta forma, el valor normalizado será uno si el valor del indicador está por encima del rango de la media, menos uno si está por debajo y cero en caso de estar en el rango marcado (Comisión Europea, 2020).
- 5) Normalización de indicadores cíclicos. El objetivo es reducir el riesgo de señales falsas para prever mejor los ciclos de las actividades económicas. Específicamente, la normalización se realiza restando la media y dividiéndola por la media de los valores absolutos de la resta entre la observación y la media (Comisión Europea, 2020). La fórmula para esta técnica es la siguiente:

$$(7) \quad Iqc^t = \frac{Xqc^t - E_t * (Xqc^t)}{E_t * (|Xqc^t - E_t * (Xqc^t)|)}$$

Donde:

- Xqc^t es el valor del indicador en el tiempo t
- $E_t * (Xqc^t)$ es la media del indicador en el tiempo t
- $|Xqc^t - E_t * (Xqc^t)|$ es el valor absoluto de la diferencia entre el valor del indicador en el tiempo t y su media en el tiempo t

- Iqc^t es el valor normalizado del indicador en el tiempo t utilizando la fórmula que proporcionaste
- 6) Porcentaje de diferencias en puntos temporales consecutivos. Esta técnica se basa en la resta del valor de la observación por el valor de la observación por el momento temporal anterior dividido por éste por 100. Al final la normalización representa el porcentaje de crecimiento con respecto al punto anterior en lugar del nivel absoluto (Comisión Europea, 2020). La fórmula para esta técnica es la siguiente:

$$(8) \quad y = \frac{x_i - x_{i-1}}{x_{i-1}} * 100$$

Donde:

- x_i es el valor de la serie en el punto de tiempo i
 - $x_{(i-1)}$ es el valor de la serie en el punto de tiempo i-1
 - y es el porcentaje de cambio entre x_i y x_{i-1}
- 7) Clasificación de los indicadores. Este método requiere seguir el rendimiento de los indicadores en clasificaciones relativas. Sin embargo, no es posible seguir el rendimiento absoluto de cada país a lo largo del tiempo pues siempre se va a encontrar clasificado en función del resto de indicadores sin importar si mejora personalmente (Comisión Europea, 2020).
- 8) Escala categórica. Este procedimiento está pensado para valores categóricos a los que se les asigna una puntuación. Al no contar con ningún indicador categórico se descarta este método (OECD, 2008).

En la siguiente Tabla se puede observar las ventajas e inconvenientes de cada una de las técnicas.

Tabla 23: “Beneficios y desventajas de las distintas técnicas de normalización”

Técnica	Pros	Contras
Estandarización	<ul style="list-style-type: none"> ✓ Facilita la comparación y el análisis de diferentes variables ✓ Ayuda a identificar valores atípicos ✓ Facilita la visualización de datos 	<ul style="list-style-type: none"> ✓ Reduce la interpretación intuitiva de los datos (los valores originales pueden perder su significado en la nueva escala) ✓ Puede eliminar información importante ✓ No recomendable cuando los datos originales no están distribuidos normalmente ✓ Puede generar valores negativos
Min Max	<ul style="list-style-type: none"> ✓ Fácil de entender e implementar ✓ Mantiene la forma de la distribución original ✓ Mejora el rendimiento de algunos algoritmos de aprendizaje automático ✓ Útil en situaciones en las que se necesitan datos en una escala específica 	<ul style="list-style-type: none"> ✓ No adecuada para datos que tienen valores extremos ✓ No es efectiva en datos con distribuciones muy sesgadas ✓ Penaliza los datos negativos
Distancia a una referencia	<ul style="list-style-type: none"> ✓ Compara valores con respecto a una referencia común. ✓ No le afectan los outliers 	<ul style="list-style-type: none"> ✓ Imprecisa, dependiente de la referencia escogida ✓ Poco representativa cuando la base de datos es grande

	<ul style="list-style-type: none"> ✓ Capaz de comparar valores que tienen diferentes unidades de medida ✓ Ayuda a resaltar las diferencias relativas entre los valores 	
Porcentaje de diferencias en puntos temporales consecutivos	<ul style="list-style-type: none"> ✓ Identificar patrones de cambio en los datos a lo largo del tiempo. ✓ Fácil de calcular 	<ul style="list-style-type: none"> ✓ No tiene en cuenta la magnitud absoluta de los cambios. ✓ No es adecuada para datos que son muy variables ✓ Sensible a la elección del intervalo de tiempo
Normalización de valores por encima o por debajo de la media	<ul style="list-style-type: none"> ✓ Sencilla y fácil de implementar. ✓ Igualar la escala de los datos y facilitar la comparación 	<ul style="list-style-type: none"> ✓ Sensible a valores atípicos ✓ No apropiada para datos que no tienen una distribución normal o simétrica
Normalización de indicadores cíclicos	<ul style="list-style-type: none"> ✓ Identifica patrones comunes en diferentes ciclos ✓ Permite la comparación de datos entre diferentes momentos y lugares. ✓ Identifica señales tempranas de cambios económicos y de negocios. 	<ul style="list-style-type: none"> ✓ No es apta para datos que no tienen una estructura cíclica clara ✓ Sensible a los cambios en la estructura de los ciclos
Clasificación por indicadores	<ul style="list-style-type: none"> ✓ Sencilla de aplicar y entender 	<ul style="list-style-type: none"> ✓ No tiene en cuenta el rendimiento absoluto de

	<ul style="list-style-type: none"> ✓ Identificar países que están por encima o por debajo de la media. ✓ Útil para realizar análisis comparativos y detectar patrones o tendencias 	<p>cada país o entidad en el tiempo</p> <ul style="list-style-type: none"> ✓ Se ve afectada en gran medida por los sesgos de los datos ✓ No tiene en cuenta la variabilidad entre los países o entidades
Escala categórica	<ul style="list-style-type: none"> ✓ Fácil de entender y aplicar ✓ Útil para la presentación de datos a audiencias no técnicas 	<ul style="list-style-type: none"> ✓ Pierde información sobre la distribución original de los datos. ✓ No es adecuado para datos continuos o variables que no se pueden agrupar en categorías. ✓ La conversión de categorías a valores numéricos puede ser subjetiva y no siempre representa la información de manera precisa. ✓ Difícil establecer la escala adecuada para la normalización

Fuente: Elaboración propia a partir de los datos de la OECD, 2008

Tras haber analizado los diferentes métodos se deberán valorar las características principales de los datos del trabajo y escoger una técnica en base a las mismas. De esta forma, se descartaron tanto la normalización categórica como los métodos cíclicos por no tratarse ni de datos categóricos ni de ciclos temporales. Asimismo, se descartó el método referente al porcentaje de diferencias en puntos temporales consecutivos al tratarse de una técnica enfocada más a valorar observaciones a lo largo de varios años (no aplicable a estos datos al ser el marco de referencia doce meses) (Comisión Europea, 2020).

Por su parte, se descartó tanto el método de referencia como el método de clasificación al tratarse de métodos que requieren de un enfoque subjetivo para marcar la normalización. Ninguno de los métodos se apoya en criterios objetivos y medibles por lo que, y dado la existencia de otras técnicas, es preferible no hacer uso de estos procedimientos.

Adicionalmente, se decidió no optar por el método de normalización basado en los valores superiores, inferiores e iguales a la media de las observaciones de cada indicador. La razón de esta decisión estriba en la arbitrariedad para seleccionar el rango de la media y la pérdida de información a nivel absoluto (Comisión Europea, 2020).

Tras descartar los anteriores métodos, la elección se situó entre la estandarización y la normalización Min Max. Entre ambos métodos, se escogió el último mencionado puesto que manejaba de mejor forma los valores atípicos y agrupaba los datos en una misma escala (Code Academy, s.f.). Para ello, se hizo uso de la ecuación (5).

La técnica min max requiere ponderar el papel de los valores atípicos en los indicadores. De esta forma, se podrá escoger los valores máximos y mínimos absolutos (sin analizar los valores atípicos) o los relativos (una vez descartados los valores extremos). Aunque la base de datos no contaba con un gran número de atípicos se optó por filtrarlos para que la escala de normalización fuera más realista.

En la primera categoría solo se consideraron dos valores atípicos en los indicadores: “CCA” y el “IPC_sb” por contar con un valor atípico muy bajo. Por otro lado, en la segunda y en la tercera categoría no se consideró ningún valor atípico.

Tras elegir el método y proceder a su normalización fue necesario discriminar entre aquellos cuyos valores elevados indicaban algo positivo (cuanto mayor era el valor del indicador más positivo era su efecto), de aquellos que mostraban lo contrario (y que requerían de un ajuste para emplazarlos en la misma escala del resto de datos).

Este ajuste se llevó a cabo haciendo uso de la siguiente fórmula:

$$(9) \quad Z = 1 - \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

Donde:

- x_i es el valor original de la variable
- $\min(x)$ es el valor mínimo de la variable
- $\max(x)$ es el valor máximo de la variable

- Z es el valor normalizado

En la primera categoría, la normalización se llevó a cabo de acuerdo con las siguientes escalas (las negativas tuvieron que ser invertidas (haciendo uso de la ecuación 9) para estar en la misma escala que las positivas):

Tabla 24: “Escala de normalización Actividad Económica”

Escala positiva	Escala negativa
“CCA”	“CVIA”
“CCVI”	“CVIB”
“CVEA”	“CVIESVI”
“CVEB”	“CVIESVI”
“CVEVI”	“IPC”
“EISSA”	“IPC_sb”
“EISSVI”	“IPRI”
“FI”, “IPI”, “IPRIX”, “IPRIX_VI”	“IPRIM”
“IGDE”	“IPRIM_VA”
“IGDEVI”	“IPRIM_VI”

Fuente: Elaboración propia

En la segunda categoría, la normalización se realizó en su totalidad en una escala positiva. Finalmente, en la tercera categoría se valoraron las siguientes escalas:

Tabla 25: “Escala de normalización Liquidez y Salud Financiera”

Escala positiva	Escala negativa
ESCE	TMB
ESCH	
Hipotecas	
RNE	
RNEVI	

Fuente: Elaboración propia

El objetivo último de la ponderación es definir la importancia relativa de cada una de las variables, con la esperanza de obtener un indicador objetivo y completo capaz de reconocer de forma efectiva los KPI que más afectan a los resultados. Para llevar a cabo esta parte de la construcción, la ponderación puede realizarse atendiendo a varias técnicas.

- Técnicas estadísticas

- “PCA y Análisis factorial”

En este método de ponderación se utilizan los pesos de los componentes principales como ponderaciones para cada variable. No obstante, estos primeros resultados pueden ser difíciles de interpretar en relación con los datos originales, ya que pueden ser fruto de combinaciones lineales de varias variables. Por ello, se ve necesario rotar la matriz de componentes principales para identificar los componentes que explican la mayor cantidad de varianza en los datos y que están más correlacionados con las variables de interés. Ello puede permitir la asignación de pesos específicos a aquellos indicadores que cuenta con una mayor importancia relativa en la construcción de los componentes principales (OECD, 2008).

A partir de estos indicadores, se pueden crear indicadores compuestos intermedios sumando las puntuaciones estandarizadas de cada indicador, ponderando cada una de ellas por la carga factorial correspondiente. Dichos componentes se deberán agregar asignado un peso a cada uno de ellos igual a la proporción de la varianza explicada en el conjunto de datos (Nicoletti et al, 2000).

- “Análisis envolvente de los datos”

En este caso la ponderación hace uso de herramientas de programación lineal para crear una frontera de eficiencia capaz de servir como referencia para medir el rendimiento de los indicadores. De esta forma, se deberá establecer la frontera en base a una serie de suposiciones (OECD, 2008):

- Ponderaciones positivas, mayor valor de la observación mayor rendimiento del indicador.
- Se deberá clasificar todos los indicadores de forma estandarizada (en una misma escala y dimensión)
- Se podrá prever una combinación lineal de los mejores indicadores de la frontera.

La forma de interpretar este método de ponderación consiste en otorgar la mejor puntuación a aquellos indicadores localizados en la frontera, mientras que los indicadores de peor puntuación serán aquellas fuera de la frontera. En concreto, se le otorgará la puntuación de 1 a los indicadores en la frontera y los peores indicadores se les otorgará la puntuación de -1 (OECD, 2008).

➤ “Beneficio de la duda”

La ponderación a través del beneficio de la duda es un enfoque utilizado en la toma de decisiones en el que se asigna un mayor peso a la opción más favorable en función de la incertidumbre asociada con la elección. Esto se hace para dar lugar a la posibilidad de que se produzcan resultados positivos incluso si las pruebas disponibles no son concluyentes (OECD, 2008). De esta forma, el índice compuesto resultante de la ponderación oscilará entre 0 (el peor rendimiento) y el 1 (el mejor rendimiento) (Melyn & Moesen, 1991). Este método es especialmente útil en el ámbito político, pudiendo generar incentivos y ayudando a definir las compensaciones (OECD, 2008)

Sin embargo, un riesgo que se deriva de estas ponderaciones es que no se fijan de una única manera, en consecuencia, se pueden dar múltiples soluciones dependiendo del conjunto de restricciones impuestas a las ponderaciones. Así, cuando mayor sea el rango de variación, menor será la posibilidad de obtener una única solución (Cherchye et al., 2008).

▪ Técnicas participatorias

➤ “Proceso de asignación presupuestaria (BAL)”

El proceso de asignación presupuestaria (también conocido como criterio experto) se basa en la asignación de un presupuesto entre los diferentes indicadores, a cada uno se le asigna una ponderación subjetiva en base a la importancia relativa de los mismos. En este caso las ponderaciones se calculan como presupuestos medios, y sus principales beneficios radican en la transparencia y sencillez del método. Asimismo, hay que mencionar que este método consta de cuatro fases: la selección de expertos para la valoración, la asignación del presupuesto a los indicadores, el cálculo de las ponderaciones y, opcionalmente, la continua asignación presupuestaria hasta alcanzar la convergencia (OECD, 2008).

➤ “Ponderación a través de la opinión pública”

Como ocurre en el anterior método es subjetivo y presenta dificultades para la población para asignar a cada indicador un grado de preocupación (OECD, 2008).

➤ “Proceso de jerarquía analítica (AHP)”

Esta ponderación permite la descomposición de un problema en una estructura jerárquica, ponderando tanto los aspectos cualitativos como los cuantitativos agregándolos al proceso de evaluación y valorando las observaciones por pares de variables (así se tendrá que expresar la preferencia entre un indicador u otro a través de una escala de 1 a 9, no por importancia si no por contribución al indicador) (OECD, 2008). En consecuencia, por esta técnica la evaluación por pares permite compensar los rendimientos de los indicadores con otros que hayan ofrecido mejores resultados (Forman, 1983). La ponderación hace referencia directa a dicha compensación, midiendo la posibilidad de intercambiar un indicador por otro (OECD, 2008).

➤ “Ponderación equiponderante”

La ponderación en este caso se limita a otorgar un mismo peso a cada uno de los indicadores que conforman el indicador compuesto. Su simplicidad ha difundido su uso, pero no se ha de olvidar que el empleo de este método puede provocar que se diluya la importancia de aquellos indicadores que realmente están afectando a la materia analizada.

➤ “Análisis Conjuntos”

Este método consiste en la evaluación de los diferentes escenarios, cada uno de ellos conformado por una serie de indicadores individuales, a través de la preferencia de cada componente. Esta ponderación se nutre de la opinión de personas que deberán ir escogiendo una combinación de indicadores. A partir de las decisiones individuales se estimará la función de preferencia en función de la frecuencia de dichas elecciones (OECD, 2008).

En la siguiente Tabla se puede observar las ventajas e inconvenientes de cada uno de estos métodos de ponderación.

Tabla 27: “Beneficios y desventajas de las distintas técnicas de ponderación”

Técnica	Pros	Contras
“PCA y Análisis factorial”	<ul style="list-style-type: none"> ✓ Asigna pesos en función de la importancia relativa de cada variable (objetiva) ✓ Permite identificar las variables que más afectan al indicador final ✓ Ponderación precisa 	<ul style="list-style-type: none"> ✓ Requiere que los datos estén correlacionados
“Análisis envolvente de los datos”	<ul style="list-style-type: none"> ✓ Asigna pesos a las variables en función de su importancia relativa en la evaluación de la eficiencia ✓ Ayuda a reducir el efecto de las variables que pueden no ser relevantes para la evaluación de la eficiencia 	<ul style="list-style-type: none"> ✓ Difícil determinar los pesos óptimos para cada variable ✓ Se asume que las variables tienen una distribución normal
“Beneficio de la duda”	<ul style="list-style-type: none"> ✓ Ayuda a evitar decisiones precipitadas y a considerar múltiples posibilidades ✓ Permite la inclusión de información que puede ser relevante, pero que no es concluyente o está incompleta ✓ Fomenta la imparcialidad y la justicia al considerar todas las posibles opciones 	<ul style="list-style-type: none"> ✓ Falta de confianza en las decisiones tomadas ✓ Puede llevar a la procrastinación ✓ Puede ser utilizado para justificar decisiones arbitrarias o prejuiciosas.
“Proceso de asignación	<ul style="list-style-type: none"> ✓ Cuando se cuenta con los expertos adecuados realiza 	<ul style="list-style-type: none"> ✓ Depende de los expertos escogidos, gran

presupuestaria (BAL)”	una asignación más justa de los pesos	componente de subjetividad <ul style="list-style-type: none"> ✓ Requiere de una inversión de tiempo y recursos para obtener las opiniones adecuadas
“Ponderación a través de la opinión pública”	<ul style="list-style-type: none"> ✓ Permite que la ponderación refleje el sentir general de la sociedad ✓ Puede mejorar la aceptación y la legitimidad de las ponderaciones 	<ul style="list-style-type: none"> ✓ Difícil de aplicar adecuadamente debido a la complejidad y la falta de conocimiento en algunos temas. ✓ Influenciado por factores como la emoción ✓ Difícil llegar a un consenso o equilibrio entre las diferentes opiniones y percepciones de la comunidad.
“Proceso de jerarquía analítica (AHP)”	<ul style="list-style-type: none"> ✓ Permite considerar múltiples criterios y alternativas ✓ Puede ayudar a identificar y evaluar las prioridades 	<ul style="list-style-type: none"> ✓ Compensa resultados ✓ Ponderaciones subjetivas basadas en la asignación personal de los valores
“Ponderación equiponderante”	<ul style="list-style-type: none"> ✓ Simple y fácil de aplicar ✓ Puede ser adecuado para situaciones en las que los criterios son igualmente importantes o en situaciones donde no se dispone de información suficiente para asignar pesos ✓ Proporciona una forma equitativa de evaluar 	<ul style="list-style-type: none"> ✓ No considera las diferencias en la importancia relativa de los indicadores ✓ No tiene en cuenta las preferencias o juicios de los expertos en la asignación de pesos

	diferentes criterios y alternativas	
“Análisis Conjuntos”	<ul style="list-style-type: none"> ✓ Permite identificar la importancia relativa de cada variable y cómo afecta la elección del consumidor 	<ul style="list-style-type: none"> ✓ La asignación subjetiva ✓ La asignación de pesos puede cambiar dependiendo de cómo se presenta agrupación de indicadores ✓ La ponderación puede ser afectada por sesgos

Fuente: Elaboración propia a partir de los datos de la OECD, 2008

La elección de la técnica de la ponderación adecuada deberá basarse en función de la naturaleza de los datos, aunque por su objetividad las técnicas estadísticas han de primar sobre las participatorias, aquellas bases de datos que cuenten con datos poco correlacionados (y por tanto inadecuadas para técnicas como el PCA/Análisis Factor) pueden ser candidatas para hacer uso de métodos como el criterio experto.

Por ello, tanto para la primera categoría como para la segunda se optó por la ponderación basada en el análisis de los componentes principales. Dado que ambas categorías presentaban una alta correlación y eran aptas para la realización de un análisis factorial (que fue llevado a cabo en el análisis multivariante). En este caso, se ha de comentar que se vio necesario rehacer el análisis multivariante con los datos ya normalizados. De este análisis, se desprende un cambio en las agrupaciones y en los valores de las observaciones de los indicadores.

En la primera categoría, que fue la única en la que se consideraron dos valores atípicos, sufrió un cambio en la distribución de dos indicadores, lo cual tiene sentido dado los cambios que sufrieron los datos al ser adaptados a la escala positiva. La clasificación se modificó de esta forma:

Tabla 28: “Clasificación de los indicadores tras normalizar. Categoría: Actividad Económica”

Primer factor (importaciones, exportaciones, cemento y seguridad social)	Segundo factor (relación del volumen de las importaciones con las principales métricas industriales)	Tercer factor (métricas económicas más relacionadas con el poder económico de los ciudadanos)	Cuarto factor (métricas relacionadas con el consumo de cemento, las importaciones y métricas relacionadas con el IPC)
“CVEVI”	“CCA”	“CVIVI”	“CCVI”
“CVIB”	“CVEA”	“IPI”	“CVIA”
“FI”	“CVEB”	“IPRIM_VI”	“IPC”
“IGDEVI”	“EISSA”	“IPRIX”	“IPC_sb”
“IPRI”	“IGDE”		“IPRIM_VA”
“IPRIX_VA”			
“EISSVI”			

Fuente: Elaboración propia

En la segunda categoría, lo que varió respecto al primer análisis fue un intercambio entre el primer y el segundo factor, si bien los indicadores se agruparon de la misma forma, el orden de los factores se modificó. Esto se debió a que el valor de la varianza explicada del primer factor y del segundo factor era muy parecida (37,454 y 36,306 respectivamente), esto provoca que las pequeñas variaciones provocadas por la normalización pueden alternar los grupos una vez hecha la misma. En consecuencia, la nueva clasificación fue la siguiente:

Tabla 29: “Clasificación de los indicadores tras normalizar. Categoría: Actividad Económica”

Primer factor (demanda eléctrica de sectores más estáticos en interiores y sector servicios)	Segundo factor (facturaciones por industria e indicadores del sector servicios)	Tercer factor (demanda eléctrica y suministros)
“DEEA”	“FICMMRVMM”	“DEI”
“DEIVI”	“FIIEM”	“ FISEEASGR”
“DES”	“IASS”	“IASSAPTC”
“FISNM”	“IASSCMA”	“VONA”
“IASSAASA”	“IASSCMe”	“VONVI”
“IASSH”	“IASSRV”	
“IASSIC”	“PMIA”	
“IASSTA”	“PMIM”	
“DESVI”	“PMISI”	

Fuente: Elaboración propia

Tras repetir el análisis factorial y haber analizado los resultados, se procedió a otorgar el peso que le correspondía por su importancia a cada indicador. Antes de continuar con la ponderación, se calculó la varianza explicada por los indicadores que conformaban cada uno de los factores. Tras ello, se elevó al cuadrado la misma y se dividió entre la suma total de las varianzas explicadas. Con esta referencia se multiplicó por cada peso relativo de cada indicador sobre el componente.

$$(10) \quad w_j = \frac{\sum \phi_j}{\sum \phi_j} = \frac{\sum \lambda_{ij}^2}{\sum \forall i (\lambda_{ij})^2}$$

Donde:

- w_j : Coeficiente que representa la proporción de la varianza total explicada por el indicador intermedio j , en relación con la varianza total explicada por todos los indicadores juntos.
- ϕ_j : Carga del indicador intermedio j en el primer componente principal, es decir, el valor de la contribución del indicador j al primer componente principal.
- λ_{ij} : Coeficiente de correlación entre el indicador i y el componente principal j .

- i: Índice que indica los indicadores utilizados en el análisis de componentes principales
- j: Índice que indica el componente principal que se está considerando.

Una vez hecho esto, los resultados de la ponderación de cada categoría fueron los siguientes (resaltado en color amarillo: el factor al que se ha asignado cada indicador):

Figura 35: “Ponderación: Actividad Económica”

	1	2	3	4		1	2	3	4
Consumo de cemento_Absoluto		0.023508827			Índice general de demanda eléctrica		0.026204245		
Consumo de cemento_Variación interanual				0.005	Índice general de demanda eléctrica_Variación interanual	0.024			
Crecimiento del volumen de las exportaciones_Absoluto		0.025			Industrial Production Index			0.024	
Crecimiento del volumen de las exportaciones_Base 100		0.025			IPC				0.010
Crecimiento del volumen de las exportaciones_Variación interanual	0.024				IPC subyacente				0.006
Crecimiento del volumen de las importaciones_Absoluto				0.0002139	IPRI (i precios)	0.024			
Crecimiento del volumen de las importaciones_Base 100	0.023				IPRIM		0.025		
Crecimiento del volumen de las importaciones en España por sector_Variación interanual			0.024		IPRIM_Variación anual			0.018	
Crecimiento del volumen de las importaciones_Variación interanual	0.026				IPRIM_Variación interanual				0.006697359
Empresas inscritas en la Seguridad Social_Absoluto		0.53091451			IPRIX			0.021034945	
Empresas inscritas en la Seguridad Social_Variación interanual					IPRIX_Variación interanual	0.025			
Facturación por industria	0.058693406								
	0.026								

Fuente: Elaboración propia

Figura 36: “Ponderación: Actividad Económica Sectorial”

	1	Componentes				1	Componentes	
		2	3				2	3
Demanda eléctrica en establecimientos de alojamiento	0.058				Indicadores de actividad del sector servicios - Actividades profesionales técnicas y científicas			0.020
Demanda eléctrica en establecimientos de alojamiento_Variación interanual	0.056				Indicadores de actividad del sector servicios - Comercio al por mayor			0.032
Demanda eléctrica en industria			0.015		Indicadores de actividad del sector servicios - Comercio al por menor			0.045
Demanda eléctrica en industria_Variación interanual			0.026		Indicadores de actividad del sector servicios - Hostelería	0.048		
Demanda eléctrica en servicios	0.035				Indicadores de actividad del sector servicios - Información y comunicaciones	0.040		
Demanda eléctrica en servicios_Variación interanual	0.030				Indicadores de actividad del sector servicios - Reparación de vehículos			0.054
Facturación por industria - Comercio al por mayor y al por menor, reparación de vehículos de motor y motocicletas		0.041			Indicadores de actividad del sector servicios - Transporte y almacenamiento	0.043		
Facturación por industria - Industrias extractivas y Facturación por industria - Servicios no financieros de mercado	0.046		0.037		Purchasing Manager's Index (PMI) of Aggregated			0.058
Facturación por industria - Suministro de energía eléctrica y agua, saneamiento y gestión de residuos			0.041		Purchasing Manager's Index (PMI) of Manuf			0.050
Indicadores de actividad del sector servicios			0.029		Purchasing Manager's Index (PMI) of Services I			0.057
Indicadores de actividad del sector servicios - Actividades	0.056				Índice de gestores de compras			0.040
					Visados de obra nueva_Absoluto			0.045
					Visados de obra nueva_Variación interanual			0.045

Fuente: Elaboración propia

Respecto a la última categoría, al no contar con datos correlacionados el análisis factor no era apto para llevar a cabo la agrupación de las observaciones. Por su parte, el análisis clúster había permitido aglutinar los indicadores en tres clústeres diferenciados que reducían los datos y la convertían en una categoría apta para aplicar el BAP (criterio experto).

De esta forma, se consideró dar más valor a la primera subcategoría al contar con los indicadores más influyentes para la liquidez y la salud financiera (stock del crédito en empresas y hogares y la tasa de morosidad) (Sánchez et al., 2020). En consecuencia, la primera

subcategoría recibió el 50% del peso total de la categoría, y las otras dos subcategorías se repartieron el 50% restante. Los indicadores que conformaban las subcategorías, por su parte, recibieron el mismo peso al tener una importancia similar en cada una de ellas.

Figura 37: “Ponderación: Actividad Económica Sectorial”

	Componente		
	1	2	3
Evolución del stock de crédito - Empresas	0,160		
Evolución del stock de crédito - Hogares	0,160		
Hipotecas			0,260
Registro de nuevas empresas		0,130	
Registro de nuevas empresas_Variación Interanual		0,130	
Tasa de morosidad bancaria	0,160		

Fuente: Elaboración propia

Todas las ponderaciones pueden verse explicadas en detalle en el Anexo 9.

Tras la ponderación, se procedió a realizar la agregación de los indicadores. En este paso, se combinan múltiples indicadores en un solo indicador (OECD,2008). Por lo tanto, se está llegando al final de la construcción del indicador compuesto que deberá englobar todos los indicadores individuales. Para llevar a cabo esta etapa, existen varias técnicas que deben ser expuestas para poder, en la parte práctica, escoger aquella que más se adapta al conjunto de datos.

- Agregación aditiva

Esta técnica se basa en la suma ponderada de los valores normalizados y el peso asignado a cada uno de los indicadores. En suma, las técnicas de agregación aditiva permiten ponderar la contribución de cada variable por separado, dichas contribuciones se relacionan directamente con una “tasa de sustitución” por la que se relacionan los indicadores entre ellos “sustituyendo” los resultados más altos con los más bajos (Saisana et al., 2005). Esta situación provoca una potencial pérdida de información (al considerar cada indicador por separado sin valorar las relaciones entre ellos) y puede producir un indicador compuesto sesgado (OECD, 2008). Para evitarlo, será esencial que los indicadores no se correlacionen entre sí (Nardo et al., 2005, p. 75). La fórmula para llevar a cabo la agregación aditiva es la siguiente:

$$(11) \quad CI_C = \sum_{q=1}^Q w_q * I_{qc}$$

Donde:

- CI_C es puntuación agregada.
- w_q son los pesos asignados a cada factor, indican la importancia relativa de cada indicador en la medida global. Los pesos deben sumar 1.
- I_{qc} son los valores normalizados.

▪ Agregación geométrica

Se trata de un método intermedio entre la compensación y la no compensación, limitando la compensación entre variables, y más concretamente penalizando la misma cuando una observación tiene un valor muy bajo. En este caso, a diferencia del anterior, la fórmula no es una simple suma ponderada, sino que, por este método, el sumatorio se sustituye por la elevación de los valores normalizados a la potencia de los valores ponderados, para finalmente multiplicar los resultados. Por ello, en el caso de que el valor del indicador sea 0 o cercano al mismo el valor agregado será cercano a ese valor (OECD, 2008). La fórmula empleada para llevar a cabo la ponderación es la siguiente:

$$(12) \quad CI_C = \prod_{q=1}^Q x_{qc}^{w_q}$$

Donde:

- CI_C : es el valor agregado para la alternativa c (por ejemplo, la agregación de todos los indicadores en enero)
- x_{qc} : es el valor del indicador q para la alternativa c
- w_q : es el peso del indicador q en el factor en el que está agrupado
- Q : es el número total de indicadores considerados en el análisis

A diferencia de la agregación aditiva, la agregación geométrica tiene una propiedad llamada “efecto de amortiguación” o “efecto de atenuación”, que permite reducir el impacto de los valores extremos en el resultado final en comparación con la agregación aditiva. No obstante, la agregación sigue viéndose afectada por los resultados negativos (OECD, 2008).

▪ Métodos multicriterio

Si bien son técnicas que aceptan que no existe una regla de ponderación perfecta, buscan la elaboración de métodos de clasificación razonables que se acerquen a dicha exactitud. Asimismo, este tipo de técnicas están conformadas por dos grandes grupos: aquellos que hacen uso de compensaciones y aquellos que no lo hacen (OECD, 2008). Como se ha observado el uso de compensaciones puede desvirtuar el indicador provocando la aparición de sesgos e inexactitudes en los resultados. Por ello, se va a tomar como referencia el método multicriterio no compensatorio.

En el caso de la no compensación, la agregación se compone de dos pasos: la comparación por pares de las observaciones según el conjunto de indicadores individuales utilizados, y, la clasificación de las observaciones en un orden previo (Munda y Nardo, 2007). De esta manera, en lugar de asignar a cada factor un valor número y sumarlos, en este caso la ponderación por el método de no compensación establece un nivel mínimo de aceptabilidad. Tras ello, se evalúa el cumplimiento de dicho nivel descartando aquellos que no alcanzan el nivel mínimo (OECD, 2008).

En conclusión, tras haber analizado la ponderación y la agregación, el indicador sintético estará construido, solo quedando la realización de un análisis de robustez que pondere la eficiencia del mismo.

Tabla 30: “Beneficios y desventajas de las distintas técnicas de agregación”

Técnica	Pros	Contras
“Agregación Aditiva”	<ul style="list-style-type: none"> ✓ Fácil de entender y aplicar. ✓ Es útil cuando los factores no están correlacionados ✓ Permite comparar alternativas en términos de sus valores totales. 	<ul style="list-style-type: none"> ✓ No tiene en cuenta las interacciones entre elementos ✓ Sensible a la elección de la escala utilizada para medir los factores ✓ Le afectan en gran medida los datos atípicos o extremos ✓ Compensa los valores más altos y bajos (falla en la representativa)

		<ul style="list-style-type: none"> ✓ No tiene en cuenta la importancia relativa de los diferentes elementos
“Agregación Geométrica”	<ul style="list-style-type: none"> ✓ Tiene en cuenta la interacción entre los elementos ✓ Menos sensible a la elección de la escala o la unidad de medida ✓ Reduce el impacto de los valores atípicos (gracias al efecto de amortiguación) 	<ul style="list-style-type: none"> ✓ Menos intuitiva y puede ser difícil de entender ✓ Si un elemento tiene un valor de cero, el resultado de la agregación geométrica será siempre cero ✓ No tiene en cuenta la importancia relativa de los diferentes elementos
“Métodos multicriterio compensatorios”	<ul style="list-style-type: none"> ✓ Flexibilidad ✓ Interpretación intuitiva ✓ Adecuados para decisiones complejas ✓ Reducción de incertidumbre 	<ul style="list-style-type: none"> ✓ Dificultad para establecer ponderaciones ✓ Falta de transparencia ✓ Ignorancia de la interdependencia de los criterios ✓ Sensibilidad a la elección del método de agregación ✓ Falta de precisión
“Métodos multicriterio no compensatorios”	<ul style="list-style-type: none"> ✓ Capacidad para tratar la interdependencia ✓ Fácil de entender ✓ Adecuados para decisiones complejas ✓ Mayor precisión 	<ul style="list-style-type: none"> ✓ Dificultad para establecer pesos ✓ Menos flexible que los métodos compensatorios ✓ Requiere más información ✓ Resultados más complejos ✓ Sensibilidad a la elección del método de agregación

Fuente: Elaboración propia a partir de los datos de la OECD, 2008

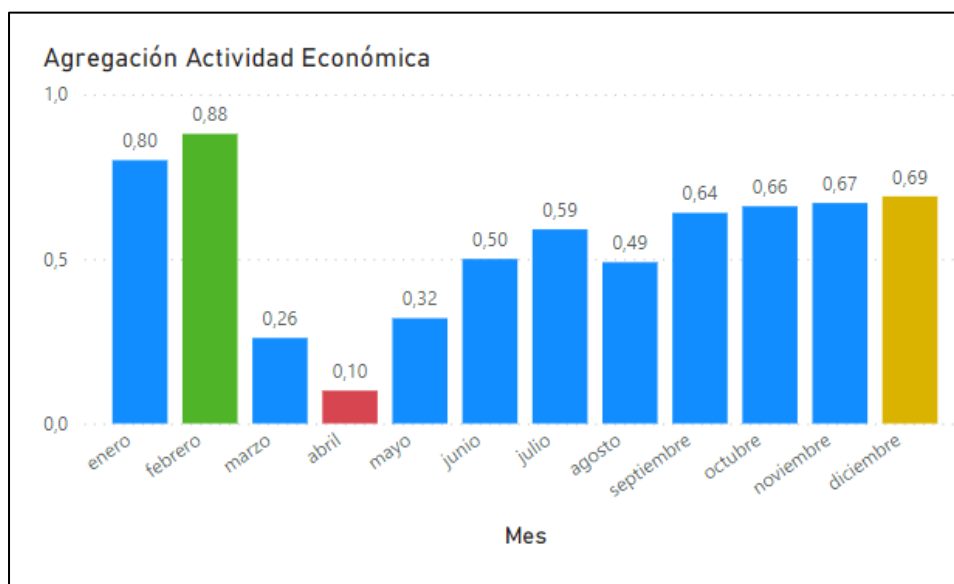
Entre todos los métodos anteriormente expuestos, se optó por una agregación aditiva puesto que todas las categorías estaban libres de correlaciones, en el caso de las dos primeras gracias

al análisis factor, y, en el caso de la tercera por la propia naturaleza de sus datos. La razón por la que se decidió no optar por la agregación geométrica en este punto fue que la ingente cantidad de datos podía provocar penalizaciones demasiado grandes. Por su parte, la agregación por métodos multicriterio fue descartada al tratarse de un tipo de agregación enfocada a clasificaciones, lo cual no casa con los datos tratados en este trabajo.

De esta forma, se multiplicaron los datos normalizados por el peso asignado a cada indicador, para después, sumar cada observación mensual por cada uno de los indicadores, para finalmente obtener una sola serie de datos para cada una de las categorías. Los resultados se pueden analizar en profundidad en el Anexo 10.

Como se puede observar en Figura 38: “Agregación Aditiva: Actividad Económica”, la actividad económica española se encontraba en un proceso de crecimiento a principios del 2020, con la pandemia sufrió una bajada importante alcanzando el mínimo en abril. A partir del levantamiento de las restricciones la actividad económica fue mejorando, aunque no alcanzó los niveles prepandemia.

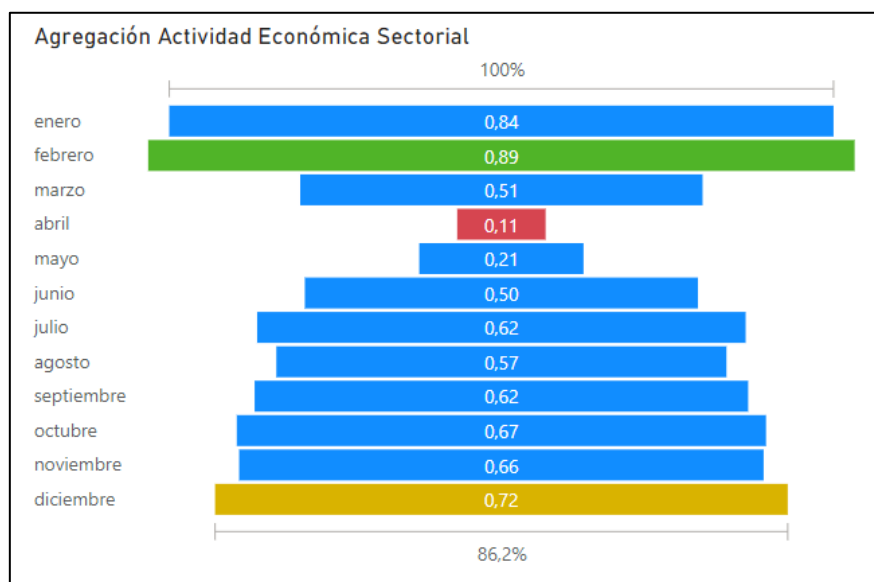
Figura 38: “Agregación Aditiva: Actividad Económica”



Fuente: Elaboración propia

En la segunda categoría, los resultados como pueden verse en la **¡Error! La autoreferencia al marcador no es válida.** fueron parecidos a los de la Actividad Económica general, pero ligeramente superiores en este caso. Por ello, podemos observar que la pandemia lastró el desarrollo económica de la actividad económica en la mayor parte de los sectores (siendo sus efectos transversales a toda la economía nacional).

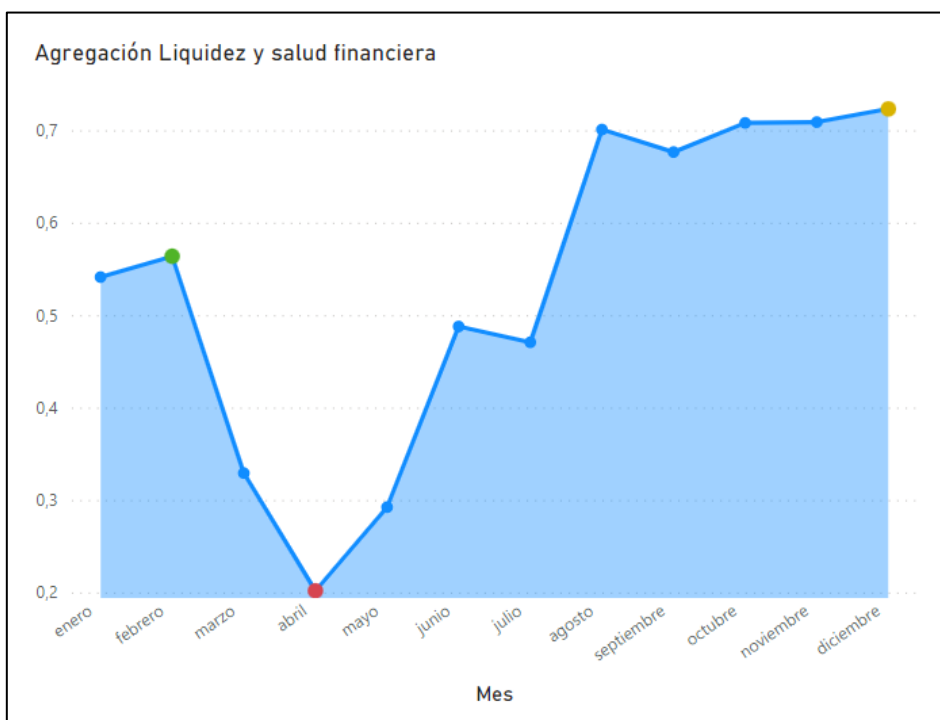
Figura 39: “Agregación Aditiva: Actividad Económica Sectorial”



Fuente: Elaboración propia

Por último, la última categoría como se puede apreciar en la Figura 40: “Agregación Aditiva: *Liquidez y Salud Financiera*” fue la única que fue capaz de recuperar los niveles pre-pandémicos, ello pudo deberse gracias en gran medida a que la tasa de morosidad bajo significativamente comparada con la del 2019 (Público, 2019). Por lo demás, el comportamiento de la categoría fue similar a las anteriores, comenzó subiendo y se desplomó con la llegada del COVID 19 en marzo alcanzando el pico más bajo en abril.

Figura 40: “Agregación Aditiva: Liquidez y Salud Financiera”



Tercera categoría: Liquidez y Salud Financiera

Tras ello, para hacer la agregación final entre las tres categorías lo primero que se hizo fue hacer una ponderación por el criterio experto entre las mismas. De esta forma, recordando que el indicador compuesto que se pretendía construir era relativo a la “actividad empresarial”, se consideró que las tres categorías tenían una importancia similar, y que, por ello, otorgar más peso a una categoría podía desvirtuar el indicador final.

Finalmente, respecto a la agregación final, se optó por la agregación geométrica. Si bien en el paso anterior al tratar con un gran número de datos no era la mejor opción, en este caso, al tratar con solo tres series temporales este tipo de agregación es óptima al tener un menor grado de compensación, equilibrando la “tasa de sustitución” que pudiese afectar a la objetividad del indicador.

Figura 41: “Agregación Final”



Gráfico de elaboración propia que refleja la evolución de indicador a lo largo del 2020

Tras haber terminado todos los pasos el indicador estaría completo, con ello, a través de la Figura 41: “Agregación Final” podemos ver la evolución de la Actividad Empresarial durante el 2020. Como se puede observar en la gráfica, la actividad empresarial española creció en febrero (dato respaldado por el número de empresas inscritas en febrero y enero (Ministerio de Industria, Comercio y Turismo, 2020)), y con la llegada del COVID, sufrió un retroceso considerable, especialmente en abril con la cuarentena nacional y el parón de la industria (Chanel Partner, 2020). Con las medidas económicas del gobierno (la ampliación de los ERTES y los subsidios) y el levantamiento de las restricciones, la actividad empresarial pudo recuperarse a partir del mes de mayo (Ministerio de Industria, Comercio y Turismo, 2020). Finalmente, hacia diciembre de 2020 se lograron resultados parecidos a los de enero de 2020 aunque sin alcanzar el nivel previo (INE, 2023).

2.6. Análisis de robustez

Una vez construido el indicador se ha de llevar a cabo un análisis de robustez que permita garantizar el buen funcionamiento del mismo (Kennedy, 2003). Durante la construcción del indicador se han de realizar una serie de elecciones subjetivas que van a afectar al resultado final del indicador (el método de normalización adecuado, el número de categorías en la imputación o los métodos de ponderación y agregación) (OECD, 2008). Dicho efecto será mayor o menor dependiendo del punto de la construcción del indicador en el que se tome la decisión, y podrá ser medido por análisis de incertidumbre y de la sensibilidad.

El análisis de incertidumbre requiere el establecimiento de factores para cada una de las decisiones tomadas. En concreto, los factores han de hacer referencia a la imputación, la normalización, las exclusiones de indicadores, la agregación y la ponderación (OECD, 2008). Por otro lado, el análisis de sensibilidad busca evaluar el cambio de la varianza en función de los cambios en los factores (OECD, 2008). Estos métodos permiten localizar aquellos factores que más han afectado a la construcción del indicador.

En este sentido, a cada factor le correspondería un número (por ejemplo: el uso de la normalización por min max sería $X_1=1$, mientras que la normalización por estandarización se asociaría a $X_2=2$).

Dichos factores podrían ponderar tanto la elección de técnicas, como dentro de una técnica el tratamiento subjetivo de los datos (por ejemplo: en el caso del análisis multivariante a la hora de analizar la correlación de los datos se pudo optar por eliminar aquellos indicadores que presentaban una correlación cercana al 90%).

A través de este análisis se pueden localizar aquellas partes de la construcción del indicador que han podido afectar en mayor medida al resultado final, y, con ello, poder replantarse el uso de ciertos métodos y decisiones.

De esta forma, se han analizado las fases del indicador donde fue necesario tomar una decisión y se han establecido una serie de escenarios alternativos:

- 1) Normalización: En nuestro indicador se optó por el método Min Max. Para el análisis de robustez se valorará el método de Estandarización.

- 2) Análisis multivariante: En nuestro indicador en la primera categoría, se optó por el criterio de Kaiser (cuatro factores) en vez de hacer uso de los criterios de varianza explicada (tres factores). Para el análisis de robustez se valorará la inclusión de tres factores en la primera categoría.
- 3) Ponderación: En nuestro indicador en las dos primeras categorías se optó por la ponderación en base al análisis de componentes principales y en la tercera categoría por el criterio experto. En el análisis de robustez se valorará la ponderación equiponderada.
- 4) Agregación: En nuestro indicador se hizo uso de la agregación rítmica y geométrica a la hora de agregar las categorías (actividad económica, actividad económica sectorial y liquidez y salud financiera). En el análisis de robustez se valorará la agregación geométrica en todo el proceso.

Figura 42: “Análisis de Robustez”

Normalización X_1	Análisis multivariante X_2	Ponderación X_3	Agregación X_4	X			
<ul style="list-style-type: none"> • Normalización Min Max • Normalización Estandarizada 	<ul style="list-style-type: none"> • Categoría: Actividad Económica (3 factores) • Categoría: Actividad Económica (4 factores) 	<ul style="list-style-type: none"> • Ponderación basada en el análisis de los componentes principales y criterio experto • Ponderación equiponderante 	<ul style="list-style-type: none"> • Aditiva • Geométrica 	<table border="1"> <tr><td>(1,1,1,1)</td></tr> <tr><td>(1,1,1,0)</td></tr> </table>	(1,1,1,1)	(1,1,1,0)	
			(1,1,1,1)				
		(1,1,1,0)					
		<ul style="list-style-type: none"> • Aditiva • Geométrica 	<table border="1"> <tr><td>(1,1,0,1)</td></tr> <tr><td>(1,1,0,0)</td></tr> </table>	(1,1,0,1)	(1,1,0,0)		
		(1,1,0,1)					
		(1,1,0,0)					
	<ul style="list-style-type: none"> • Ponderación basada en el análisis de los componentes principales y criterio experto • Ponderación equiponderante 	<ul style="list-style-type: none"> • Categoría: Actividad Económica (3 factores) • Categoría: Actividad Económica (4 factores) 	<ul style="list-style-type: none"> • Ponderación basada en el análisis de los componentes principales y criterio experto • Ponderación equiponderante 	<ul style="list-style-type: none"> • Aditiva • Geométrica 	<table border="1"> <tr><td>(1,0,1,1)</td></tr> <tr><td>(1,0,1,0)</td></tr> </table>	(1,0,1,1)	(1,0,1,0)
				(1,0,1,1)			
			(1,0,1,0)				
			<ul style="list-style-type: none"> • Aditiva • Geométrica 	<table border="1"> <tr><td>(1,0,0,1)</td></tr> <tr><td>(1,0,0,0)</td></tr> </table>	(1,0,0,1)	(1,0,0,0)	
	(1,0,0,1)						
	(1,0,0,0)						
<ul style="list-style-type: none"> • Ponderación basada en el análisis de los componentes principales y criterio experto • Ponderación equiponderante 	<ul style="list-style-type: none"> • Categoría: Actividad Económica (3 factores) • Categoría: Actividad Económica (4 factores) 	<ul style="list-style-type: none"> • Ponderación basada en el análisis de los componentes principales y criterio experto • Ponderación equiponderante 	<ul style="list-style-type: none"> • Aditiva • Geométrica 	<table border="1"> <tr><td>(0,1,1,1)</td></tr> <tr><td>(0,1,1,0)</td></tr> </table>	(0,1,1,1)	(0,1,1,0)	
			(0,1,1,1)				
(0,1,1,0)							
<ul style="list-style-type: none"> • Aditiva • Geométrica 	<table border="1"> <tr><td>(0,1,0,1)</td></tr> <tr><td>(0,1,0,0)</td></tr> </table>	(0,1,0,1)	(0,1,0,0)				
(0,1,0,1)							
(0,1,0,0)							
<ul style="list-style-type: none"> • Ponderación basada en el análisis de los componentes principales y criterio experto • Ponderación equiponderante 	<ul style="list-style-type: none"> • Categoría: Actividad Económica (3 factores) • Categoría: Actividad Económica (4 factores) 	<ul style="list-style-type: none"> • Ponderación basada en el análisis de los componentes principales y criterio experto • Ponderación equiponderante 	<ul style="list-style-type: none"> • Aditiva • Geométrica 	<table border="1"> <tr><td>(0,0,1,1)</td></tr> <tr><td>(0,0,1,0)</td></tr> </table>	(0,0,1,1)	(0,0,1,0)	
			(0,0,1,1)				
(0,0,1,0)							
<ul style="list-style-type: none"> • Aditiva • Geométrica 	<table border="1"> <tr><td>(0,0,0,1)</td></tr> <tr><td>(0,0,0,0)</td></tr> </table>	(0,0,0,1)	(0,0,0,0)				
(0,0,0,1)							
(0,0,0,0)							

En esta figura se puede apreciar un diagrama con las diferentes opciones posibles fruto del análisis de robustez, en los recuadros naranjas se pueden ver las opciones escogidas durante la elaboración de este indicador.

De estas opciones surgen estos distintos indicadores compuestos (en rojo se puede apreciar el indicador construido en este trabajo):

Figura 43: “Comparativa entre las diferentes opciones del análisis de Robustez”

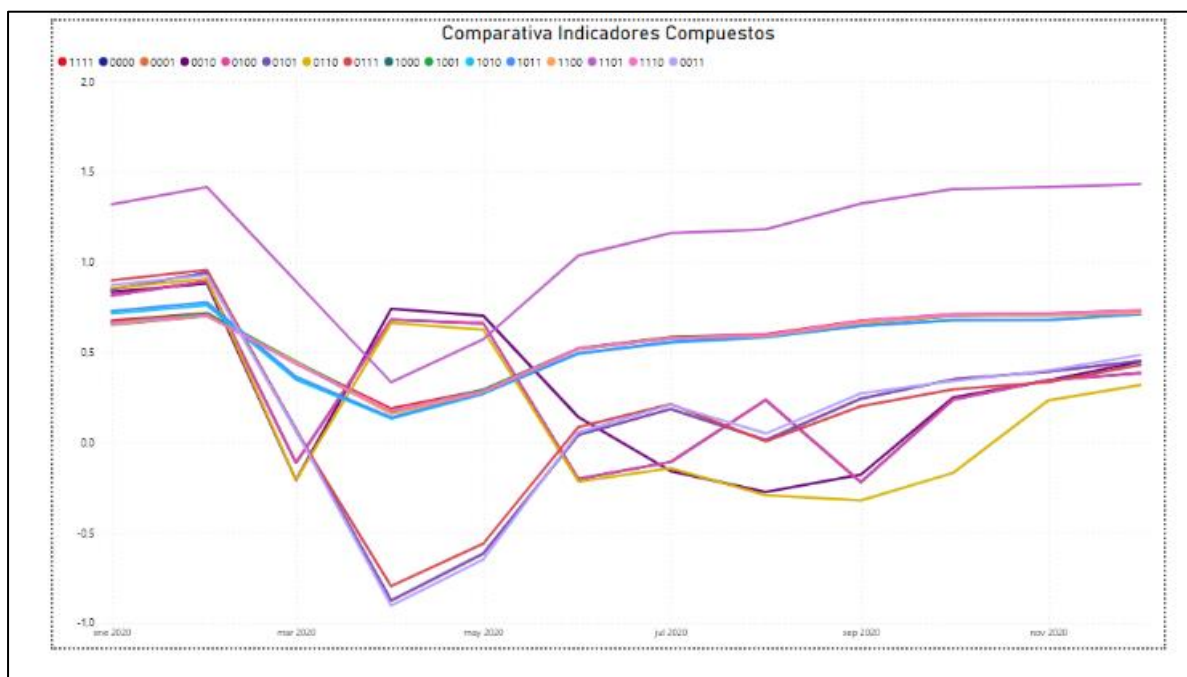


Gráfico de elaboración propia con la comparativa entre los potenciales indicadores compuestos

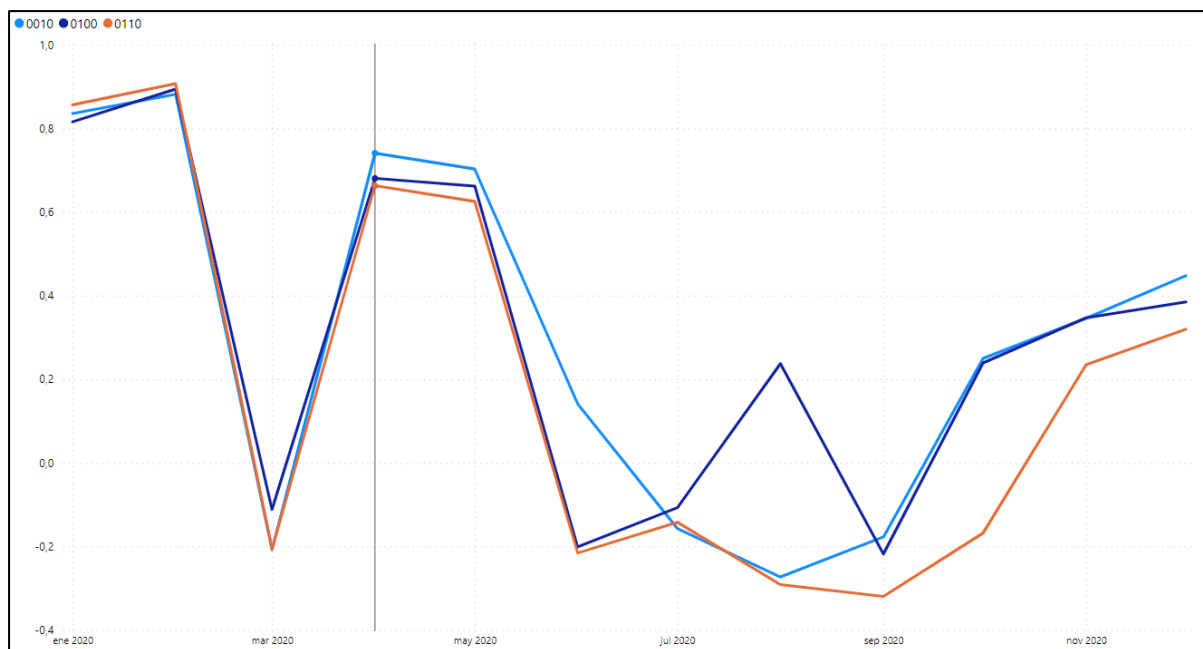
Los resultados de cada uno de los análisis pueden verse en profundidad en el Anexo 11.

De este proceso se pueden extraer diversas conclusiones. En primer lugar, podemos observar que independientemente de la técnica escogida en marzo se produce un descenso de la actividad empresarial. Esto es coherente con la declaración del Estado de Alarma, la cuarenta nacional y el consecuente parón de la actividad económica empresarial.

En segundo lugar, se observa que cuando los valores se normalizan a través de la técnica de la estandarización la tendencia sufre pequeños cambios. De esta forma, en los escenarios (0100), (0110) y (0010) la recuperación se comienza a notar en abril para volver a caer en mayo, esto es debido tanto por haber optado por una normalización estandarizada (no incluye los datos en una misma escala) y por la agregación geométrica que castiga los valores negativos y los valores nulos. Todo ello provoca que la tendencia viré y se aleje del resto de posibilidades. Ello puede verse en la

Figura 44: : “Comparativa entre las opciones (0100), (0110) y (0010).

Figura 44: : “Comparativa entre las opciones (0100), (0110) y (0010)



Por su parte, los escenarios (0001), (0101) y (0011) aunque hacen uso de valores estandarizados, la aplicación de agregación aditiva no trastoca tanto la línea de tendencia (al hacer uso de compensaciones), pero al tratarse de valores estandarizados el descenso en marzo y abril es más acentuado (al poder tomar valores negativos).

En tercer lugar, se puede observar que en todos los indicadores a partir de septiembre comienzan la recuperación de la actividad empresarial. Esto es coherente con la situación del país en esos meses, el levantamiento de las restricciones y la recuperación paulatina del consumo permitieron relanzar la actividad empresarial.

En conclusión, se puede observar la importancia de cada uno de los pasos realizados durante la construcción del indicador y la relevancia de tomar determinadas decisiones. La dificultad de predecir los efectos del análisis hace esencial mantener una coherencia y una explicación razonada de cada una de las elecciones que se vayan a tomar. Vistas todas las opciones, considero que las decisiones tomadas durante la construcción del indicador fueron las adecuadas para los tipos de datos con los que contaba al arrojar los resultados más parecidos a la situación que se dio en la realidad.

3. Conclusiones

Las principales conclusiones que pueden extraerse de este trabajo son las siguientes:

Primera. El uso de indicadores compuestos permite obtener una herramienta única y sencilla capaz de agrupar múltiples indicadores individuales, los cuales pueden provenir de sectores de actividad muy diversos. Gracias a ello, se puede monitorizar materias como la actividad empresarial de forma sencilla, obteniendo una visión general de la situación del asunto analizado.

Segunda. Durante la construcción del indicador se han de tomar múltiples decisiones subjetivas. El peso que tienen estas disyuntivas provoca que sea necesario llevar a cabo una explicación exhaustiva sobre las técnicas y métodos obtenidos. De esta forma, se debe primar la transparencia y la coherencia para que los usuarios que hagan uso de estos indicadores conozcan en todo momento la forma en la que han sido construidos. Ello permitirá un aumento de la confianza en los indicadores compuestos.

Tercera. Cada una de las partes de construcción del indicador puede ser llevada a cabo a través de diversos métodos. La elección de un método u otro ha de ser tomada en base a las características propias de las bases de datos. El análisis crítico de estas técnicas muestra que el uso de métodos más complejos puede aportar una mayor objetividad al proceso de construcción y, por consiguiente, un indicador compuesto más sólido.

Cuarta. La necesidad de explicar las decisiones durante la construcción del indicador no solo responde a motivos de transparencia, sino que, además, permite al lector analizar las diversas técnicas y cambiar el indicador en base a sus conocimientos propios de la materia. Todo ello, facilitará la mejora continua del indicador compuesto al enriquecerlo con su propio análisis.

Quinta. El COVID frenó el avance de la actividad empresarial que se estaba produciendo en España, la pandemia sin duda afectó a los resultados que se pudiesen haber obtenido durante los meses de marzo y abril. No obstante, las medidas de apoyo promovidas desde Europa y la relajación de las restricciones permitieron a la actividad empresarial recuperarse y acercarse a los valores obtenidos a principio de año.

4. Bibliografía

Amat, R. (2017). *Clustering y Heatmaps: agrupando datos con visualización*. Ciencia de datos. Recuperado el 28 de febrero de 2023 en:

https://www.cienciadedatos.net/documentos/37_clustering_y_heatmaps

Banco de España (2021). *Informe anual 2021*. Recuperado el 21 de octubre de 2022 en:

https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesAnuales/InformesAnuales/21/Fich/InfAnual_2021.pdf

Banco de España (2022). *Informe trimestral de la economía española*. Recuperado el 19 de octubre de 2022 en:

<https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/InformesBoletinesRevistas/BoletinEconomico/22/T2/Fich/be2202-it.pdf>

Banco de España (2022b). *La evolución de la solvencia y de la demografía empresarial en España desde el inicio de la pandemia*. Recuperado el 15 de octubre de 2022 en: <https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/InformesBoletinesRevistas/ArticulosAnaliticos/22/T3/Fich/be2203-art13.pdf>

Becker, W. (2022b, julio 7). *Composite Indicator Development and Analysis in R with COINr*. Recuperado el 9 de abril de 2023 en: <https://bluefoxr.github.io/COINrDoc/>

Channel Partner. (2020). Parón drástico, disolución y creación de empresas abril de 2020. Recuperado el 23 de abril de 2023 en: <https://www.channelpartner.es/negocios/noticias/1118528002202/paron-drastico-disolucion-y-creacion-de-empresas-abril-de-2020.1.html>

Cherchye, L., Moesen, W., Rogge, N., van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R., Tarantola, S. (2008). *Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index*, Journal of the Operational Research Society, 59, 239-251.

Codecademy. (s.f.). *Normalization*. Recuperado el 6 de marzo de 2023 en: <https://www.codecademy.com/article/normalization>

Comisión Europea, Eurostat. (2017) “*Código de buenas prácticas de las estadísticas europeas para los servicios estadísticos nacionales y comunitarios*”, Oficina de Publicaciones. Recuperado el 27 de noviembre en: <https://data.europa.eu/doi/10.2785/2990>

Comisión Europea. (2018) “*Competence Centre on Composite Indicators and Scoreboards. Step 1: Theoretical Framework / Knowledge for Policy*”, Recuperado el 27 de noviembre de 2022 en: https://knowledge4policy.ec.europa.eu/composite-indicators/10-step-guide/step-1-theoretical-framework_en.

Comisión Europea. (2018b). *European statistics code of practice: for the national statistical authorities and Eurostat (EU statistical authority)*, Publications Office. Recuperado el 26 de noviembre de 2022 en:

<https://data.europa.eu/doi/10.2785/798269>

Comisión Europea (2018c). *Informe sobre España 2018*, SWD(2018) 207 final.

Comisión Europea. (2020). *Step 5: Normalisation*. Knowledge for Policy. Recuperado el 6 de marzo en:

https://knowledge4policy.ec.europa.eu/composite-indicators/10-step-guide/step-5-normalisation_en#distance-to-a-reference-country

Comisión Europea (2022). *The Joint Harmonised EU Programme of Business and Consumer Surveys*.

Cox, D.R. et al. (1992) “*Quality-of-life assessment: Can we keep it simple?*” *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 155(3), p. 353.

CK-12 Foundation. (2018). *Interpolación y extrapolación lineal*. Recuperado el 26 de abril de 2023 en:

<https://www.ck12.org/algebra/linear-interpolation-and-extrapolation/lesson/Interpolaci%C3%B3n-y-extrapolaci%C3%B3n-lineal-BSC-ALG-SPN/>

Ebert U. and Welsch H. (2004), *Meaningful environmental indices: a social choice approach*, *Journal of Environmental Economics and Management*, Vol. 47: 270-283.

FMI (2012) *Data Quality Assessment Framework (DQAF) for National Accounts Statistics*. Recuperado el 23 de abril de 2023 en:

https://dsbb.imf.org/content/pdfs/dqrs_nag.pdf

Freudenberg, M. (2003). “*Composite Indicators of Country Performance*.” OECD Science, Technology and Industry Working Papers

Gan, X., Fernandez, I. C., Guo, J., Wilson, M., Zhao, Y., Zhou, B. & Wu, J. (2017). *When to use what: Methods for weighting and aggregating sustainability indicators. Ecological Indicators, 81*, 491-502.

Gobierno de España. (2021). *Informe de situación de la economía*. Recuperado el 11 de octubre de 2022 en:

<https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/asuntos-economicos/Documents/2021/290721-Informe-de-Situacion-Economia-espanola-2021.pdf>

Greco, Salvatore, et al. (2018) “*On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness.*” *Social Indicators Research*, vol. 141, no. 1, pp. 61–94.

Gupta, S. (2021). Z-Score Normalization. *ML Concepts*. Recuperado el 5 de marzo de 2023 en:

<https://ml-concepts.com/2021/10/08/z-score-normalization/>

Hair J.F., Black W.C., B.J., Babin, Anderson R.E. and R.L.,Tatham (2006), *Multivariate data analysis*, sexta edición, Pearson Prentice Hall, Upper Saddle River, NJ, p. 6.

Hernández de Cos, P. (2020). *Los principales retos de la economía española tras el Covid-19*. Comparecencia en la Comisión para la Reconstrucción Social y Económica de España tras el Covid-19. Congreso de los Diputados, el 23 de junio de 2020. *Research Papers in Economics*.

Instituto Nacional de Estadística (INE). (2021). *Directorio Central de Empresas (DIRCE)*.

Recuperado el 23 de abril de 2023 en:

https://www.ine.es/prensa/dirce_2021.pdf

Kaufmann, D., Kraay, A. y Mastruzzi, M. (2006): *Governance Matters V: Aggregate and Individual Governance Indicators for 1996-2005*. The World Bank.

Kennedy P. (2003), *A guide to Econometrics*, 5th edition, MIT Press

MATLAB - MathWorks España. (2022). *Fill missing values*. Recuperado 31 de diciembre de 2022 en:

<https://es.mathworks.com/help/matlab/ref/fillmissing.html>

Massart D.L. and Kaufman L. (1983), *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*, John Wiley & Sons, Inc.

Ministerio de Industria (2022). *Cifras y PyME*. Gobierno de España. Recuperado el 11 de octubre de 2022 en:

<http://www.ipyme.org/Publicaciones/CifrasPYME-enero2022.pdf>

Ministerio de Industria, Comercio y Turismo. (2020). *Nota explicativa medidas adoptadas para sectores*. Recuperado el 23 de abril de 2023 en:

https://www.mincotur.gob.es/es-es/COVID-19/Documents/Nota_explicativa_medidas_adoptadas_para_sectores_MINCOTUR.pdf

Medina, F., & Galán, M. (2007). *Imputación de Datos: Teoría y práctica - CEPAL*. Estudios Estadísticos y Prospectivos, 54. Recuperado el 30 de noviembre del 2022 en

https://repositorio.cepal.org/bitstream/handle/11362/4755/1/S0700590_es.pdf

Melyn W. and Moesen W.W. (1991). *Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available*. Public Economic research Paper 17, CES, KU Leuven.

Ministerio de Industria, Comercio y Turismo. (2021, 19 febrero). *El déficit comercial de 2020 se sitúa en 16.755 millones de euros, el 60,6% menos [Comunicado de prensa]*. Recuperado el 23 de abril de 2023 en:

[https://www.mincotur.gob.es/es-es-GabinetePrensa/NotasPrensa/2021/Paginas/20210219Np_comex_enero_diciembre.aspx](https://www.mincotur.gob.es/es-es/GabinetePrensa/NotasPrensa/2021/Paginas/20210219Np_comex_enero_diciembre.aspx)

Munda G. and Nardo M. (2005). *Constructing Consistent Composite Indicators: The Issue of Weights*, EUR 21834 EN, Joint Research Centre, Ispra.

Munda G. and Nardo M. (2007). *Non-compensatory/Non-Linear composite indicators for ranking countries: a defensible setting*, Forthcoming in Applied Economics.

Naciones Unidas (1992). *Human Development Report*. Oxford.

Navarro, R. (2013). Splines Cúbicos. Recuperado el 26 de abril de 2023 en:

<https://sites.google.com/site/analisisnumerico2013/interpolacion/splines-cubicos>

Nicoletti G., Scarpetta S. and Boylaud O. (2000). *Summary indicators of product market regulation with an extension to employment protection legislation*, OECD, Economics department working papers No. 226, ECO/WKP(99)18. Recuperado el 27 de marzo de 2023 en:

<http://www.oecd.org/eco/eco>.

OECD. (2008). *Handbook on Constructing Composite Indicators: METHODOLOGY AND USER GUIDE*.

OECD. (2022). *Composite leading indicator (CLI)*. Disponible en:

<https://www.oecd.org/newsroom/composite-leading-indicators-cli-oecd-october-2022.htm>

OECD. (2003). *Quality Framework and Guidelines for OECD Statistical Activities*, Paris. Recuperado el 28 de noviembre de 2023 en:

www.oecd.org/statistics

Pardo, Antonio (2019). “*Guía para el análisis de datos SPSS 11.0.*”, Mc Graw Hill, Capítulo 20, 420-456. Recuperado el 13 de febrero de 2023 en:

<https://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/GuiaSPSS/20factor.pdf>

Pollesch, N., & Dale, V. (2016). *Normalization in sustainability assessment: Methods and implications*. *Ecological Economics*, 130, 195-208. Recuperado el 5 de marzo de 2023 en:

<https://www.sciencedirect.com/science/article/abs/pii/S0921800915305899>

Público. (2021, 5 de febrero). La banca cierra 2020 con una tasa de morosidad del 4,51%. Público. Recuperado el 27 de abril de 2023 en:

<https://www.publico.es/economia/banca-cierra-2020-tasa-morosidad-4-51.html>

Rojas, K. (2022). *Transformación, estandarización e imputación de datos*. En *Ciencia de Datos y Estadística Aplicada*. Recuperado el 6 de marzo en:

https://bookdown.org/keilor_rojas/CienciaDatos/transformación-estandarización-e-imputación-de-datos.html

Saisana, M., Saltelli, A. & Tarantola, S. (2005). *Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators*. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(2), 307-323.

Saisana, M. & Tarantola, S. (2002). *State-of-the-art report on current methodologies and practices for composite indicator development*. EUR 20408 EN, Joint Research Centre.

Saltelli, A (2006). “*Composite Indicators between Analysis and Advocacy.*” *Social Indicators Research*, vol. 81, no. 1, pp. 65–77.

Saltelli, Andrea (2002). “*Making Best Use of Model Evaluations to Compute Sensitivity Indices.*” *Computer Physics Communications*, vol. 145, no. 2, pp. 280–297.

Sánchez, S. S., Condori, W. J. A., Huamaní, O. G., Alcántara, R. L., Garcés, J. V., & Monrroy, K. M. B. (2020). La morosidad: factores motivacionales de los responsables financieros, en el contexto del COVID – 19. *Revista de Investigación Valor Contable*, 7(1), 37-49.

Sharpe, A. (2004). *Literature Review of Frameworks for Macro-Indicators*. Centre for the Study of Living Standards. Ottawa.

Spath H. (1980), *Cluster Analysis Algorithms*, Chichester, England: Ellis Horwood.

Zimmermann H.J. y Zysno P. (1983). *Decisions and evaluations by hierarchical aggregation of information*, *Fuzzy Sets and Systems*, 10: 243-260

Anexos

Anexo 1



Datos Actividad
Empresarial.xlsx

Anexo 2



Actividad_Empresaria
_Resumen.xlsx

Anexo 3



Datos_No_Imputar.xls
x

Anexo 4



DataSet_Imputacione
s_Actividad_económic



DataSet_Imputacione
s_Actividad_Económics



DataSet_imputacione
Agregador_Manufac

Anexo 5



Script_Análisis_Clúster
.docx



Análisis_multivariante
_Actividad_Económica



Análisis_Multivariante
_Actividad_Económica



Análisis_multivariante
_Liquidez_Salud_Finan

Anexo 6



Guión_Matriz de
Correlación.docx



Matriz_Correlacion_A
ctividad_Economica_Si



Matriz_Correlacion_Li
quidez_y_salud_financi



Matriz_Correlacion_A
ctividad_Economica.xl:

Anexo 7



Análisis_factorial_Liqu
idez_Salud_Financiera.



Análisis_factor_Activid
ad_Económica_Sectori



Análisis_factorial_Acti
vidad_Económica.xlsx

Anexo 8



Normalización_Activi
dad_Económica_Secto



Normalización_Activi
dad_Económica_Min_!



Normalización_Liquid
ez_Salud_Financiera.xl:

Anexo 9



Ponderación_Liquidez_Salud_Financiera.xlsx



Ponderación_Activida
d_Económica_Sectoria



Ponderación_Activida
d_Económica_4_min_rr

Anexo 10



Agregación_Liquidez_y_Salud_financiera.xlsx



Agregación_Actividad_Económica_Sectori



Agregación_Actividad_Económica_4_min_

Anexo 11



Indicadores_Compu
estos.pbix



Todos los posibles
indicadores.xlsx