



Universidad Politécnica
de Madrid



**Escuela Técnica Superior de
Ingenieros Informáticos**

Grado en Administración y Dirección de Empresas

Trabajo Fin de Grado

**Análisis Socioeconómico de los Países
de la Unión Europea Usando Datos de
Eurostat a través del Lighthouse de
Economics @Intelligence**

Autora: Yana Nikolaeva Gizdova

Tutor: Raúl Gutiérrez Sanchis

Madrid, Junio 2024

Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

Trabajo Fin de Grado

Doble Grado en Ingeniería Informática y Administración y Dirección de Empresas

*Título: Análisis Socioeconómico de los Países de la Unión Europea Usando
Datos de Eurostat a través del Lighthouse de Economics @Intelligence
Junio 2024*

Autora: Yana Nikolaeva Gizdova

Tutor: Raúl Gutiérrez Sanchis

Departamento de Ingeniería de la Organización, Administración de
Empresas y Estadística
ETSI Informáticos
Universidad Politécnica de Madrid

Resumen

Este documento constituye el Trabajo Fin de Grado para el grado en Administración y Dirección de Empresas de título Análisis Socioeconómico de los Países de la Unión Europea Usando Datos de Eurostat a través del Lighthouse de Economics @Intelligence. Forma parte de un proyecto sinérgico con un trabajo complementario en Ingeniería Informática que desarrolla el dashboard del Lighthouse de Economics @Intelligence, herramienta empleada para el análisis en este estudio. Bajo la línea de investigación de "Inteligencia artificial para análisis económicos reales" del proyecto Economics @Intelligence, este trabajo aporta un análisis socioeconómico detallado de los países de la Unión Europea a través del estudio de las dimensiones de Calidad de Vida definidas por Eurostat. Cuenta, además, con un enfoque especial en la correlación de estos indicadores con los Objetivos de Desarrollo Sostenible de la Agenda 2030.

El objetivo principal es estudiar socioeconómicamente los distintos estados miembros tratando de buscar patrones y tendencias para poder evaluar su progreso hacia estos objetivos globales. A través de un análisis descriptivo, exploratorio, de series temporales, predictivo y de clustering, se revelan diferencias clave entre países, contribuyendo a un entendimiento más profundo de la calidad de vida y el bienestar económico y social en Europa. Este enfoque permite una comparación entre diferentes economías y trata de ofrecer bases para la orientación futura de políticas a nivel nacional y europeo. El trabajo destaca la necesidad de superar las métricas económicas tradicionales y adoptar un enfoque más integral que considere una variedad de indicadores de bienestar y sostenibilidad.

Abstract

This document constitutes the Final Degree Project for the degree in Business Administration titled "Socioeconomic Analysis of the Countries of the European Union Using Eurostat Data through the Lighthouse of Economics @Intelligence." It is part of a synergistic project with a complementary work in Computer Engineering that develops the dashboard of the Lighthouse of Economics @Intelligence, a tool used for analysis in this study. Under the research line "Artificial Intelligence for Real Economic Analysis" of the Economics @Intelligence project, this work provides a detailed socioeconomic analysis of the European Union countries through the study of the Quality of Life dimensions defined by Eurostat. It also focuses specifically on the correlation of these indicators with the Sustainable Development Goals of the 2030 Agenda.

The main objective is to socioeconomically study the different member states, trying to identify patterns and trends to evaluate their progress towards these global goals. Through descriptive, exploratory, time series, predictive, and clustering analysis, key differences between countries are revealed, contributing to a deeper understanding of the quality of life and economic and social well-being in Europe. This approach allows for comparisons between different economies and aims to provide a basis for future policy orientation at both national and European levels. The work emphasizes the need to move beyond traditional economic metrics and adopt a more comprehensive approach that considers a variety of well-being and sustainability indicators.

Tabla de contenidos

1	Introducción	1
1.1	Contexto	1
1.2	Objetivos del trabajo	2
1.3	Entorno de trabajo y tecnologías empleadas	2
2	Relación entre los indicadores de Calidad de Vida de Eurostat y la Agenda 2030	3
2.1	Dimensiones de Calidad de Vida de Eurostat	3
2.1.1	Descripción de cada dimensión	4
2.2	Contexto Objetivos de Desarrollo Sostenible – Agenda 2030	6
2.2.1	Introducción a la Agenda 2030 y sus objetivos	6
2.2.2	Relación entre los indicadores de Eurostat y los ODS	9
3	Desarrollo	11
3.1	Selección de variables relevantes	11
3.1.1	Estudio de la base de datos de Eurostat	11
3.1.2	Identificación de indicadores socioeconómicos clave	11
3.2	Metodología de análisis	13
3.2.1	Descripción de las técnicas de análisis	13
3.2.1.1	Metodología del análisis descriptivo y exploratorio	13
3.2.1.2	Metodología del análisis de series temporales	14
3.2.1.3	Metodología del análisis predictivo	15
3.2.1.4	Metodología del análisis de <i>clustering</i>	16
3.3	Análisis de la Dimensión 1: <i>Main GDP aggregates per capita</i>	17
3.3.1	Análisis descriptivo y exploratorio de la Dimensión 1	17
3.3.2	Análisis de series temporales de la Dimensión 1	18
3.3.3	Análisis predictivo de la Dimensión 1	19
3.3.4	Análisis de <i>clustering</i> de la Dimensión 1	20
3.4	Análisis de la Dimensión 2: <i>Employment rates by sex, age and educational attainment level</i>	21
3.4.1	Análisis descriptivo de la Dimensión 2	21
3.4.2	Análisis de series temporales de la Dimensión 2	22
3.4.3	Análisis predictivo de la Dimensión 2	23
3.4.4	Análisis de <i>clustering</i> de la Dimensión 2	23
3.5	Análisis de la Dimensión 3: <i>Population by educational attachment level, sex and age</i>	25
3.5.1	Análisis descriptivo de la Dimensión 3	25
3.5.2	Análisis de series temporales de la Dimensión 3	25
3.5.3	Análisis predictivo de la Dimensión 3	26
3.5.4	Análisis de <i>clustering</i> de la Dimensión 3	27

3.6	Análisis de la Dimensión 4: <i>Healthy life years by sex</i>	28
3.6.1	Análisis descriptivo de la Dimensión 4	28
3.6.2	Análisis de series temporales de la Dimensión 4	29
3.6.3	Análisis predictivo de la Dimensión 4	30
3.6.4	Análisis de <i>clustering</i> de la Dimensión 4	31
3.7	Análisis de la Dimensión 5: <i>Frequency of getting together with family and relatives or friends</i>	32
3.7.1	Análisis descriptivo y exploratorio de la Dimensión 5.....	32
3.7.2	Análisis de series temporales de la Dimensión 5	32
3.7.3	Análisis predictivo de la Dimensión 5	33
3.7.4	Análisis de <i>clustering</i> de la Dimensión 5	33
3.8	Análisis de la Dimensión 6: <i>Inability to face unexpected financial expenses</i>	34
3.8.1	Análisis descriptivo y exploratorio de la Dimensión 6.....	34
3.8.2	Análisis de series temporales de la Dimensión 6	35
3.8.3	Análisis predictivo de la Dimensión 6	36
3.8.4	Análisis de <i>clustering</i> de la Dimensión 6	36
3.9	Análisis de la Dimensión 7: <i>Gender employment gap</i>	37
3.9.1	Análisis descriptivo y exploratorio de la Dimensión 7.....	37
3.9.2	Análisis de series temporales de la Dimensión 7	37
3.9.3	Análisis predictivo de la Dimensión 7	39
3.9.4	Análisis de <i>clustering</i> de la Dimensión 7	39
3.10	Análisis de la dimensión 8: <i>Pollution, grime or other environmental problems</i>	40
3.10.1	Análisis descriptivo y exploratorio de la Dimensión 8	40
3.10.2	Análisis de series temporales de la Dimensión 8.....	41
3.10.3	Análisis predictivo de la Dimensión 8	42
3.10.4	Análisis de <i>clustering</i> de la Dimensión 8.....	42
3.11	Análisis de la Dimensión 9: <i>Average rating of satisfaction by domain, sex, age and educational attainment level</i>	43
3.11.1	Análisis descriptivo y exploratorio de la Dimensión 9	43
3.11.2	Análisis de series temporales de la Dimensión 9.....	44
3.11.3	Análisis predictivo de la Dimensión 9	45
3.11.4	Análisis de <i>clustering</i> de la Dimensión 9.....	45
4	Resultados y conclusiones	47
4.1	Resumen de los principales hallazgos.....	47
4.2	Limitaciones del estudio.....	48
4.3	Recomendaciones para futuras investigaciones	48
5	Análisis de Impacto	49
6	Bibliografía	50

1 Introducción

En la era digital, donde el volumen de información crece exponencialmente, es vital poder convertir datos en conocimiento aplicable para comprender las complejidades de las sociedades actuales. Este trabajo se centra en estudiar los indicadores socioeconómicos disponibles en la base estadística de Eurostat. Comenzando por un análisis exploratorio y descriptivo de los datos, identificando patrones, correlaciones y tendencias clave que emerjan de los indicadores seleccionados, su principal enfoque será en la técnica de *clustering*, un método estadístico que agrupa datos basándose en sus similitudes, revelando patrones ocultos y tendencias dentro de los grandes conjuntos de datos socioeconómicos de Eurostat.

El *clustering* es esencial en este análisis ya que permite la segmentación de los países de la UE en grupos homogéneos según indicadores socioeconómicos clave, lo que facilita la comparación y el contraste entre diversas economías y sociedades. Este enfoque es particularmente relevante para entender cómo cada país contribuye y se alinea con los objetivos de la Agenda 2030 para el Desarrollo Sostenible, permitiendo a los organismos europeos y a los estados miembros evaluar el progreso y orientar políticas de forma más informada y focalizada.

El nexo entre este análisis y la Agenda 2030 es de particular interés, ya que la segmentación de países según su desempeño socioeconómico ofrece una perspectiva crucial en la medición y avance hacia los Objetivos de Desarrollo Sostenible (ODS). Desde el punto de vista de la sostenibilidad y la equidad, el *clustering* de los indicadores socioeconómicos puede ayudar a identificar líderes y rezagados en áreas específicas, permitiendo dar una visión más profunda de la calidad de vida de los países estudiados. Estudios como este sobre dicha diferenciación podrían posibilitar la formulación de políticas más enfocadas y la asignación de recursos que puedan acelerar el progreso colectivo.

Este estudio pretende presentar una aplicación práctica y teóricamente fundada de técnicas estadísticas en el ámbito socioeconómico. Los datos pueden revelar tendencias y patrones, aunque el contexto y la complejidad real detrás de cada cifra requieren una interpretación cuidadosa y considerada. Por tanto, este trabajo busca analizar y tratar de relacionar los resultados obtenidos con el contexto socioeconómico europeo.

1.1 Contexto

Este trabajo de fin de grado en Administración y Dirección de Empresas forma parte de una colaboración sinérgica con otro complementario en Informática (con título *Diseño del Dashboard del Lighthouse Eurostat de Economics @Intelligence*), ambos centrados en el análisis de datos económicos de la Unión Europea utilizando la base de datos de Eurostat.

Se lleva bajo la tutela del del grupo de investigación Análisis de decisiones y Estadística (DASG) de la ETSIINF, UPM, aportando a su línea de investigación "Inteligencia artificial para análisis económicos reales". En esta se ha desarrollado Economics @Intelligence, un proyecto enfocado en la aplicación de la inteligencia artificial a la economía tratando de establecer una base de datos de acceso abierto para la consulta, análisis y descarga de datos económicos. Una parte fundamental de este proyecto es la plataforma Lighthouse que recoge datos económicos de distintas autoridades estadísticas mundiales y nacionales (como son el Banco Mundial y el Instituto Nacional de Estadística) y los ofrece

diversas visualizaciones gráficas de los mismos mediante técnicas de procesado y análisis de datos. El propósito de este trabajo de fin de grado es hacer uso del dashboard Lighthouse Eurostat, desarrollado en el trabajo complementario a este, para realizar un análisis exhaustivo de los indicadores socioeconómicos incluidos en la base de datos de Eurostat [1].

1.2 Objetivos del trabajo

El trabajo tiene definida la siguiente lista de objetivos:

- **O1:** Seleccionar variables relevantes para el análisis.
- **O2:** Realizar un análisis exploratorio y descriptivo de las variables.
- **O3:** Aplicar modelos predictivos y de *clustering*.

1.3 Entorno de trabajo y tecnologías empleadas

Este trabajo se lleva a cabo con el dashboard Lighthouse Eurostat diseñado y desarrollado en el trabajo complementario. Dado que se usará su versión en local, se usará el entorno de desarrollo de RStudio y el de Visual Studio Code además del lenguaje de programación R incluyendo una multitud de paquetes de R usados para el desarrollo de la plataforma.

2 Relación entre los indicadores de Calidad de Vida de Eurostat y la Agenda 2030

Este trabajo de investigación se adentra en el estudio de las dimensiones cubiertas por la temática de "Calidad de Vida" proporcionada por Eurostat, con el objetivo de comprender los indicadores que configuran el bienestar en los países de la Unión Europea. Se prestará especial atención a cómo estas métricas multifacéticas de calidad de vida, desde las condiciones materiales hasta la experiencia general de vida, no sólo reflejan la realidad socioeconómica de las naciones, sino que también revelan las interconexiones que contribuyen al tejido de las sociedades europeas.

2.1 Dimensiones de Calidad de Vida de Eurostat

La calidad de vida abarca numerosas dimensiones, incluyendo factores objetivos como la salud y el estado laboral, y la percepción subjetiva de estos, dependiendo de las prioridades de los ciudadanos. Medir la calidad de vida de manera comparable entre diferentes poblaciones y países es una tarea compleja que requiere un cuadro de mando de indicadores que cubra diversas dimensiones relevantes [2, 3].

El indicador más ampliamente utilizado para medir la actividad económica es el PIB, pero este no proporciona una imagen completa del bienestar de los ciudadanos. A pesar de su utilidad para evaluar la producción de mercado, el PIB se centra en aspectos cuantitativos de la economía y falla en capturar factores cruciales como la distribución de la riqueza, el impacto ambiental o la calidad de vida general. Por ejemplo, un incremento en el PIB puede reflejar un crecimiento económico que beneficia solo a una pequeña parte de la sociedad, sin traducirse en mejoras reales en los estándares de vida de la mayoría. Además, situaciones como el aumento de las exportaciones a costa de la explotación de recursos naturales demuestran cómo el PIB puede subir mientras se compromete la sostenibilidad a largo plazo. Por lo tanto, es esencial adoptar un enfoque más holístico y multidimensional, incorporando indicadores adicionales que reflejen de manera más precisa el bienestar económico, social y ambiental de una sociedad, para proporcionar una visión más precisa y equitativa del progreso y la calidad de vida en diversas regiones y comunidades [3].

El debate sobre cómo mejorar la medición del progreso y bienestar social ha impulsado iniciativas importantes, como el Informe de la Comisión Stiglitz/Sen/Fitoussi en 2009 [4, 5]. En respuesta a estos informes, el Sistema Estadístico Europeo ha establecido el Grupo de Patrocinio sobre Medición del Progreso, Bienestar y Desarrollo Sostenible. Este grupo ha producido un informe en 2011 que ha enfatizado la necesidad de adoptar un enfoque multidimensional para medir la calidad de vida y desarrollar indicadores de sostenibilidad. Siguiendo estas recomendaciones, Eurostat ha formado un grupo de expertos que ha trabajado entre 2012 y 2016 y ha entregado un informe final en 2017 con un cuadro de mando de indicadores de calidad de vida. Estos indicadores están disponibles en la sección de Calidad de Vida de Eurostat [3].

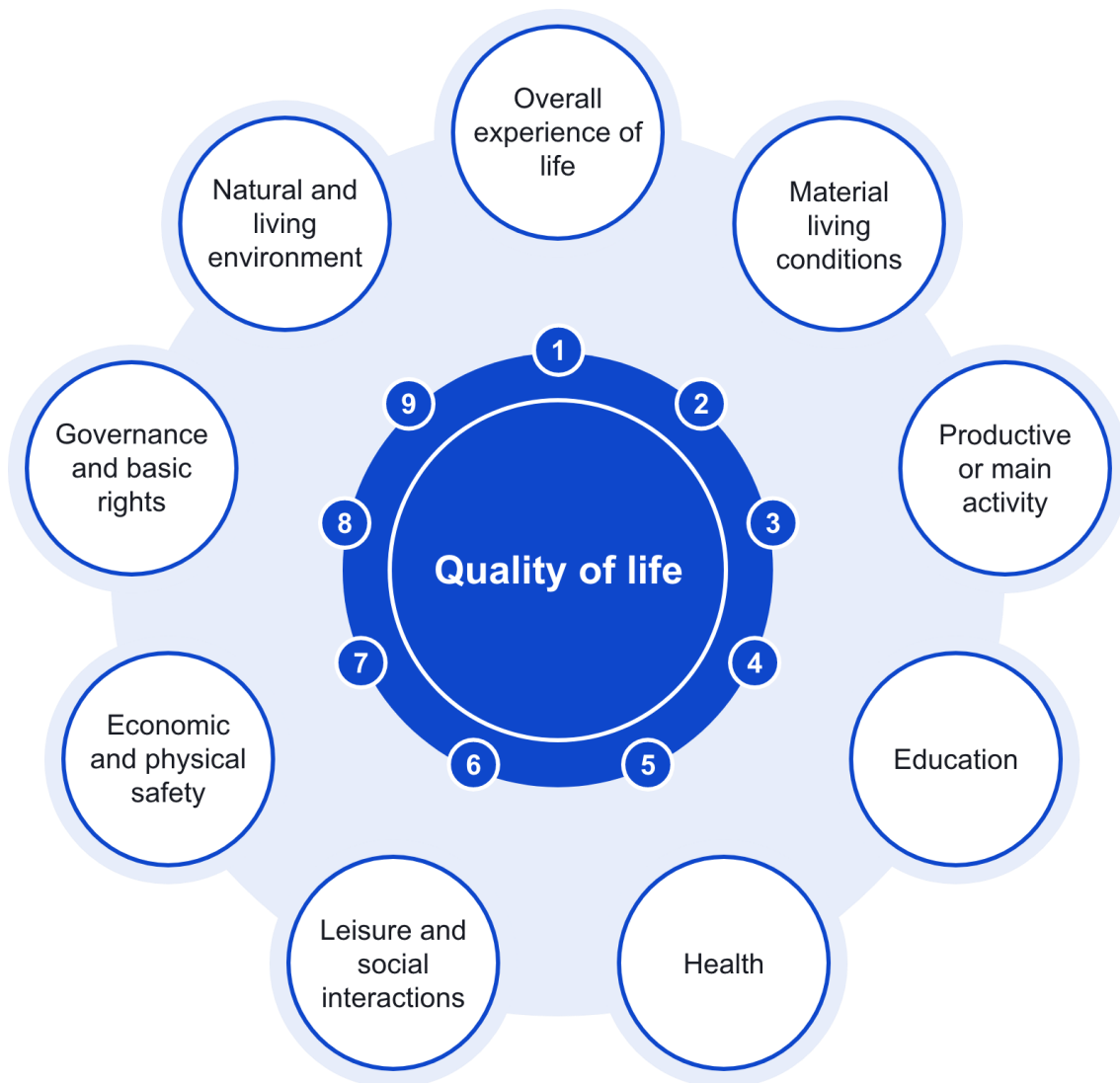


Figura 1 Dimensiones de Calidad de Vida definidas por Eurostat. Fuente: [6]

La Figura 1 es una representación gráfica de las nueve dimensiones que definen la calidad de vida según Eurostat. Se trata de dimensiones cuantificables estadísticamente que representan diversos aspectos que complementan la calidad de vida. Todas ellas conjuntamente tienen como objetivo ampliar el indicador tradicionalmente utilizado para medir el progreso económico y social, el Producto Interno Bruto (PIB). Ocho de estas dimensiones se centran en las capacidades funcionales que los individuos necesitan para perseguir efectivamente su bienestar de acuerdo con sus valores y prioridades personales. La última dimensión se ocupa del logro personal de la satisfacción y bienestar en la vida.

2.1.1 Descripción de cada dimensión

Dimensión 1 Condiciones materiales de vida:

La evaluación de las condiciones materiales de vida abarca tres subdimensiones: ingresos, consumo y bienestar material, que incluye privaciones y vivienda. Los ingresos, derivados de cuentas nacionales y encuestas a hogares como la EU-SILC, son fundamentales por su influencia en otros indicadores. El consumo se mide con indicadores nacionales y se prevé expandirlos mediante la Encuesta de Presupuestos Familiares. También se

están desarrollando indicadores integrados de ingresos, consumo y riqueza para una visión más completa de las condiciones de los hogares, complementados con datos adicionales de la EU-SILC sobre privaciones y vivienda [5].

Dimensión 2 Actividad productiva o principal:

Para evaluar la actividad productiva se consideran tres subdimensiones: calidad del empleo, cantidad de empleo y la inclusión de la población inactiva y el trabajo no remunerado. Los indicadores utilizados, provenientes de la EU-LFS y encuestas de estructura salarial, incluyen las horas de trabajo, el equilibrio entre la vida laboral y personal, y la seguridad laboral. Además, es fundamental incorporar datos sobre el trabajo no remunerado, crucial para la evaluación de la igualdad de género y la calidad de vida, aunque no todos los Estados miembros de la UE recopilan esta información de manera voluntaria [5].

Dimensión 3 Educación:

La educación es crucial para el progreso social y las oportunidades laborales en economías basadas en el conocimiento. En la UE, se mide a través de desempeño académico, habilidades, educación continua y acceso educativo, utilizando fuentes como la EU-LFS, la encuesta PIACC de la OCDE y estudios de TIC [5].

Dimensión 4 Salud:

La salud, vital para la calidad de vida y considerada capital humano, es crucial para el progreso social y el bienestar personal. En la UE, se evalúa a través de resultados de salud como la esperanza de vida y años de vida saludable, comportamientos de salud y acceso a servicios médicos, utilizando datos de la EU-SILC y la Encuesta Europea de Salud (EHIS) [5].

Dimensión 5 Ocio e interacciones sociales:

El ocio y las interacciones sociales, fundamentales para el bienestar, se miden en la UE en dos áreas: actividades de ocio y relaciones sociales. Estas evalúan la frecuencia y calidad de eventos sociales, problemas de acceso, frecuencia de contacto y satisfacción con las relaciones, además de voluntariado y cohesión social. Los datos provienen principalmente de módulos ad hoc de EU-SILC relacionados con la participación social y el bienestar subjetivo [5].

Dimensión 6 Seguridad económica y física:

La seguridad económica y física es esencial para que los ciudadanos gestionen cambios imprevistos en su entorno. Se evalúa mediante dos aspectos: la seguridad física, a través de tasas de homicidios y percepciones de crimen usando datos de la EU-SILC, y la seguridad económica, donde se usan variables como la capacidad de manejar gastos inesperados y el estado de los pagos atrasados. La falta de datos comparables en Europa obliga a usar estas variables proxy, que son cruciales para medir la resiliencia económica y la calidad de vida general [5].

Dimensión 7 Gobernanza y derechos básicos:

La gobernanza y los derechos básicos, esenciales para la calidad de vida, implican la capacidad de los ciudadanos para influir en el debate y las políticas públicas. Esta área se divide en confianza en instituciones, igualdad de oportunidades y participación en la sociedad, que incluye manifestaciones y peticiones. Se monitorea la confianza y satisfacción con servicios públicos, así como la discriminación y las disparidades laborales, utilizando datos de la EU-SILC y la LFS, con posibilidad de futuras expansiones de indicadores [5].

Dimensión 8 Entorno natural y de vida:

El entorno natural y de vida tiene gran influencia sobre el bienestar de los ciudadanos. Factores como la contaminación del aire, el agua y el ruido afectan a la salud y a la economía. Esta dimensión se mide con indicadores subjetivos (percepciones de ruido, contaminación y satisfacción con espacios verdes de la EU-SILC) y objetivos (niveles de contaminantes del aire según la Agencia Europea del Medio Ambiente), evaluando su impacto en la calidad de vida [5].

Dimensión 9 Experiencia general de vida:

La experiencia general de vida se mide en tres áreas: satisfacción vital, afecto (emociones positivas y negativas) y eudaimonía (propósito y bienestar psicológico). Estos indicadores, siguiendo las pautas de la OCDE, se recopilaron en los módulos Ad-Hoc EU-SILC de 2013 y 2018 [5].

2.2 Contexto Objetivos de Desarrollo Sostenible – Agenda 2030

2.2.1 Introducción a la Agenda 2030 y sus objetivos

La Agenda 2030 para el Desarrollo Sostenible es un plan de acción global adoptado por 193 Estados miembros de las Naciones Unidas en 2015. Su objetivo es promover el bienestar de las personas, la protección del planeta y la prosperidad económica, al tiempo que se fortalece la paz universal y se garantiza el acceso a la justicia. Reconociendo que la erradicación de la pobreza es el mayor desafío actual, la Agenda 2030 establece que sin lograrla no es posible alcanzar el desarrollo sostenible. Para su implementación, se plantea la movilización de recursos mediante alianzas globales, enfocándose especialmente en las necesidades de los más pobres y vulnerables [7, 8].

Los Objetivos de Desarrollo Sostenible (ODS) son un conjunto de 17 objetivos con 169 metas específicas que forman parte de la Agenda 2030. Estos objetivos constituyen un llamado universal a la acción para erradicar la pobreza, proteger el medio ambiente y mejorar la calidad de vida y las perspectivas de las personas en todo el mundo. Los ODS abarcan diversas esferas, incluyendo la económica, social y ambiental, y se diseñaron para ser integrados e indivisibles, reflejando la complejidad y la interconexión de los desafíos globales actuales. Al implementar los ODS, los países buscan crear un futuro más sostenible y equitativo para todos [7, 8].

En este apartado se presenta el estado de avance actual (2023) de los Objetivos de Desarrollo Sostenible de la Agenda 2030. En la Figura 2 se puede ver de forma gráfica dicho avance por indicador estando aquellos con mayor progreso en la parte superior derecha de la flecha.

Overview of EU progress towards the SDGs over the past 5 years, 2023 (Data mainly refer to 2016–2021 or 2017–2022)

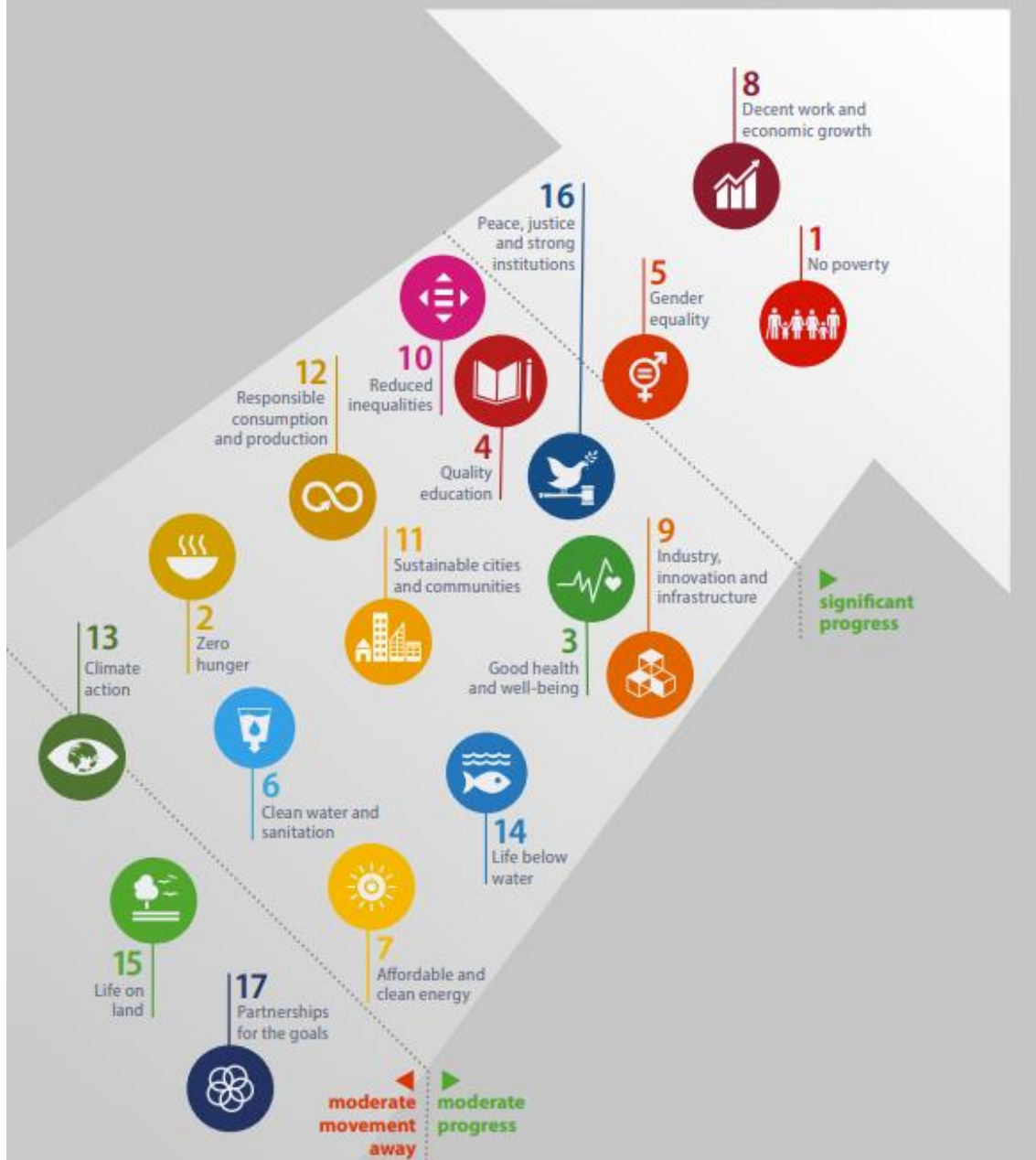


Figura 2 Vista general del progreso de la UE hacia los ODS de los 5 últimos años, 2023. Fuente: [9]

La Unión Europea, en su esfuerzo por avanzar hacia los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030, ha mostrado un progreso significativo en diversas áreas socioeconómicas, aunque enfrenta desafíos persistentes en el dominio ambiental. En términos de desarrollo económico y laboral (ODS 8), la UE ha experimentado una recuperación notable tras el impacto de la pandemia, con avances en el empleo y un aumento en el PIB per cápita. La lucha contra la pobreza (ODS 1) ha visto mejoras considerables, reflejadas en la disminución de la tasa de riesgo de pobreza o exclusión social y una mejor capacidad para cubrir necesidades básicas.

En cuanto a la igualdad de género (ODS 5), se ha observado un progreso alentador, con una disminución en la brecha salarial por hora entre hombres y

mujeres y un aumento en la representación femenina en puestos de liderazgo. En educación (ODS 4), la UE parece encaminada a cumplir sus metas para 2030, aunque persisten desafíos en la reducción de alumnos con bajo rendimiento académico.

La igualdad y reducción de desigualdades (ODS 10) también han mejorado, con una disminución en la desigualdad de ingresos dentro de los países y avances en la integración laboral de migrantes de fuera de la UE. Sin embargo, en temas ambientales como la acción por el clima (ODS 13), a pesar de reducir las emisiones de gases de efecto invernadero, la UE deberá intensificar sus esfuerzos para alcanzar las metas de 2030.

A nivel de salud y bienestar (ODS 3), aunque se ha progresado en la esperanza de vida saludable hasta 2019, la pandemia ha tenido impactos negativos, evidenciados en el aumento de la mortalidad evitable. En innovación e infraestructura (ODS 9), se han visto tendencias favorables en I+D e innovación, pero se requiere un progreso más significativo para alcanzar el objetivo del gasto del 3% del PIB en I+D.

La evaluación general revela que la UE ha avanzado firmemente hacia muchos de los objetivos socioeconómicos en el período más reciente de cinco años de datos disponibles, mientras que las tendencias en el dominio ambiental han sido menos favorables. Es fundamental que la UE continúe con políticas eficaces y adaptadas para abordar estos desafíos y mantener el impulso en la implementación de la Agenda 2030 [9].

Los objetivos directamente relacionados con las dimensiones señaladas de la temática de Calidad de Vida de Eurostat son los siguientes: ODS 1, ODS 3, ODS 4, ODS 8, ODS 10, ODS 11, ODS 13, ODS 16.

En el gráfico a continuación se puede ver la puntuación obtenida como media europea del progreso de los Objetivos de Desarrollo Sostenible de la última década. Cada círculo del gráfico corresponde a un país de los 34 estudiados en el Informe de Desarrollo Sostenible de Europa de 2023. Se puede ver una mejora progresiva, llegando a posicionarse la media en 72 en 2022.



Figura 3 Rendimiento sobre los ODS de los países europeos. Fuente: [10]

Por otro lado, en el siguiente mapa se puede observar la puntuación de desarrollo sostenible por país, marcando en verde los países que mejor avance estén teniendo y en rojo aquellos con métrica peor. De este ranking se puede observar que Finlandia se posiciona como el país con el mejor rendimiento en términos de desarrollo sostenible de 2022, mientras que Turquía se posiciona como la última de la lista de países.



SDG Index Rank	Country	SDG Index Score
1	Finland	80.6
2	Sweden	80.6
3	Denmark	80.0
4	Austria	77.7
5	Norway	77.0
6	Germany	75.4
7	Czechia	74.4
8	Slovenia	73.7
9	Iceland	73.7
10	Estonia	73.0
11	Switzerland	72.9
12	France	72.9
13	Poland	72.8
14	Ireland	71.4
15	Belgium	71.0
16	United Kingdom	70.7
17	Slovak Republic	70.1
18	Netherlands	70.1
19	Croatia	70.0
20	Portugal	70.0
21	Italy	69.9
22	Spain	69.6
23	Latvia	69.5
24	Hungary	69.0
25	Luxembourg	67.8
26	Lithuania	67.7
27	Malta	65.9
28	Greece	65.2
29	Romania	62.6
30	Serbia	62.5
31	North Macedonia	62.2
32	Cyprus	61.0
33	Bulgaria	59.5
34	Türkiye	57.1
European Union		72.0

Figura 4 Puntuación y ranking del rendimiento de los ODS de los países europeos. Fuente: [10]

2.2.2 Relación entre los indicadores de Eurostat y los ODS

Las dimensiones de calidad de vida, desglosadas en temas y subtemas con sus respectivos indicadores, reflejan de manera directa los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030. A continuación, se relacionan específicamente cada dimensión con los ODS correspondientes:

- Condiciones Materiales de Vida: Los temas de ingresos y consumo abordan el ODS 1 (Fin de la pobreza) y el ODS 10 (Reducción de las

desigualdades), evidenciando la necesidad de mejorar la distribución del ingreso y el acceso a servicios. La subtemática de privación material y las condiciones de vivienda se relacionan con el ODS 11 (Ciudades y comunidades sostenibles), reflejando la calidad de vida urbana y el acceso a viviendas asequibles y seguras.

- Productividad o Actividad Principal: El empleo y el desempleo (ODS 8 - Trabajo decente y crecimiento económico) y la calidad del empleo, que incluye la salud y la seguridad en el trabajo y el equilibrio entre trabajo y vida privada, son cruciales para asegurar trabajos justos y fomentar una economía próspera.
- Salud: La esperanza de vida y la salud (ODS 3 - Salud y bienestar), junto con el acceso a la atención sanitaria, resaltan la importancia de sistemas de salud sólidos y accesibles para todos.
- Educación: La consecución educativa y las habilidades (ODS 4 - Educación de calidad) son fundamentales para el desarrollo personal y profesional, permitiendo una sociedad más informada y competente.
- Ocio e Interacciones Sociales: La cantidad y calidad del ocio, así como las interacciones sociales (ODS 3 - Salud y bienestar, y ODS 11 - Ciudades y comunidades sostenibles), destacan la relevancia del bienestar social y mental y la importancia de las áreas verdes y espacios de ocio para la calidad de vida urbana.
- Seguridad Económica y Física: Los temas de seguridad económica y vulnerabilidad (ODS 1 - Fin de la pobreza y ODS 8 - Trabajo decente y crecimiento económico), así como la seguridad física y personal (ODS 11 - Ciudades y comunidades sostenibles y ODS 16 - Paz, justicia e instituciones sólidas) subrayan la importancia de una sociedad protegida y estable.
- Gobernanza y Derechos Básicos: La confianza en las instituciones y servicios públicos y la participación activa en la ciudadanía (ODS 16 - Paz, justicia e instituciones sólidas) son esenciales para garantizar la gobernanza efectiva y el respeto por los derechos humanos.
- Entorno Natural y Habitacional: La gestión de la contaminación, el acceso a espacios verdes y recreativos y el entorno construido (ODS 11 - Ciudades y comunidades sostenibles y ODS 13 - Acción por el clima) son vitales para la sostenibilidad ambiental y el bienestar humano.
- Experiencia General de la Vida: La satisfacción vital, los afectos y el sentido y propósito (ODS 3 - Salud y bienestar) enfatizan la importancia de la percepción subjetiva del bienestar como un aspecto integral del desarrollo humano [8].

3 Desarrollo

3.1 Selección de variables relevantes

3.1.1 Estudio de la base de datos de Eurostat

La base de datos de Eurostat está diseñada para ofrecer un acceso eficiente a una amplia gama de estadísticas que abarcan diferentes aspectos socioeconómicos de la Unión Europea. La estructura jerárquica de la base de datos comienza con categorías generales como estadísticas generales y regionales, economía y finanzas, población y condiciones sociales, industria, comercio y servicios, agricultura, comercio internacional, y ciencia y tecnología. Estas categorías generales se dividen en subcategorías más específicas, lo que permite a los usuarios explorar en profundidad temas particulares de interés.

Dentro de cada una de estas categorías, los datos se presentan en series temporales que pueden ser mensuales, trimestrales o anuales, proporcionando una visión detallada de las tendencias y cambios a lo largo del tiempo. Además, Eurostat ofrece indicadores especiales y vintages de datos, que facilitan el acceso a conjuntos de datos históricos. Esto es particularmente útil para realizar evaluaciones retrospectivas de políticas y analizar cambios a lo largo del tiempo.

Para garantizar una interpretación adecuada de los datos, cada conjunto de datos en Eurostat viene acompañado de metadatos exhaustivos. Estos metadatos incluyen detalles sobre las metodologías utilizadas, las definiciones aplicadas y las fuentes de los datos, lo cual es esencial para asegurar la precisión y la validez del análisis realizado con estos datos.

Todos los indicadores en la base de datos de Eurostat están identificados por un código único. Estos códigos no solo identifican los indicadores individuales, sino que también agrupan conjuntos de indicadores bajo una misma temática, facilitando el acceso temático a los datos. En las descargas masivas de datos desde la web de Eurostat, siempre se incluyen columnas con el código y el tipo de datos, diferenciando entre dataset, tabla o carpeta. Estas descargas están disponibles en formatos .tsv, .csv o sdmx, y también se pueden obtener a través de la API Statistics de Eurostat, que ofrece los datos en formato JSON-stat 2.0 [11, 12].

3.1.2 Identificación de indicadores socioeconómicos clave

En el estudio de la calidad de vida, es fundamental seleccionar indicadores que representen de manera integral y precisa las diferentes dimensiones que conforman este concepto. A continuación, se presenta una selección de indicadores clave para cada dimensión relevante junto con su código identificador.

Dimensión 1: Condiciones Materiales de Vida

Para evaluar las condiciones materiales de vida, se seleccionaron indicadores que reflejan los ingresos, la pobreza y la privación material. Los indicadores seleccionados incluyen [5]:

- *Main GDP aggregates per capita*: Este indicador proporciona una visión general de la distribución del ingreso entre diferentes grupos de edad y sexo, lo cual es crucial para entender la equidad económica [5].

- *Income quintile share ratio S80/S20*: Este ratio mide la desigualdad de ingresos, un factor importante para evaluar la cohesión social y económica [5].

Dimensión 2: Actividad Productiva o Principal

La evaluación de la actividad productiva se centra en el empleo y las condiciones laborales. Los indicadores seleccionados son:

- *Employment rates by sex, age and educational attainment level*: Este indicador mide la tasa de empleo, proporcionando información sobre la participación laboral [5].

Dimensión 3: Educación

Para la evaluación de la educación, se seleccionaron indicadores que reflejan el nivel educativo y las competencias de la población:

- *Population by educational attainment level, sex and age*: Este indicador mide el nivel educativo alcanzado por la población [5].

Dimensión 4: Salud

La salud es una dimensión crítica de la calidad de vida, y se seleccionaron indicadores que abarcan tanto la salud física como la mental:

- *Healthy life years by sex*: Mide el número de años de vida saludable, combinando la esperanza de vida con la calidad de salud [5].

Dimensión 5: Interacciones Sociales y Ocio

Para evaluar las interacciones sociales y el ocio, se seleccionaron indicadores que reflejan la participación en actividades y las relaciones sociales:

- *Frequency of getting together with family and relatives or friends*: Mide la frecuencia de reuniones con familiares y amigos [5].

Dimensión 6: Seguridad Económica y Física

La seguridad económica y física se evalúa a través de indicadores que reflejan la estabilidad financiera y la percepción de seguridad:

- *Inability to face unexpected financial expenses*: Mide la capacidad de los hogares para afrontar gastos imprevistos [5].

Dimensión 7: Gobernanza y Derechos Básicos

Para evaluar la gobernanza y los derechos básicos, se seleccionaron indicadores que reflejan la participación ciudadana y la confianza en las instituciones:

- *Gender employment gap*: Mide la disparidad salarial de género, un indicador clave de igualdad de oportunidades [5].

Dimensión 8: Entorno Natural y de Vida

Para evaluar el entorno natural y de vida, se seleccionaron indicadores que reflejan la percepción y el impacto ambiental:

- *Pollution, grime or other environmental problems*: Este indicador evalúa la percepción de problemas ambientales en el área local [5].

Dimensión 9: Experiencia General de Vida

La experiencia general de vida incluye indicadores que reflejan la satisfacción y el bienestar subjetivo:

- *Average rating of satisfaction by domain, sex, age and educational attainment level*: Mide la satisfacción general con la vida en diferentes dominios [5].

3.2 Metodología de análisis

3.2.1 Descripción de las técnicas de análisis

Dentro de la metodología de análisis que sigue este proyecto se utilizan técnicas de análisis descriptivo, análisis exploratorio, análisis de series temporales, análisis predictivo y análisis *clustering*.

3.2.1.1 Metodología del análisis descriptivo y exploratorio

En primer lugar, las estadísticas descriptivas proporcionan un resumen conciso y comprensible de las características principales de un conjunto de datos. Estas estadísticas ayudan a entender la distribución, la tendencia central y la variabilidad de los datos, proporcionando una base sólida para cualquier análisis posterior. A continuación, se describen las estadísticas descriptivas generadas por el código y su importancia en el análisis de calidad de vida.

La media es el promedio aritmético de un conjunto de valores [13].

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

La mediana es el valor central que divide el conjunto de datos en dos partes iguales.

Para un conjunto de datos ordenado x_1, x_2, \dots, x_n [14]:

- Si n es impar, la mediana es el valor en la posición $\left(\frac{n+1}{2}\right)$.
- Si n es par, la mediana es el promedio de los valores en las posiciones $\left(\frac{n}{2}\right)$ y $\left(\frac{n+1}{2} + 1\right)$.

La desviación estándar mide la dispersión de los datos respecto a la media [15].

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Los percentiles dividen el conjunto de datos en partes iguales. El percentil p es el valor debajo del cual se encuentra el $p\%$ de los datos.

Para el percentil p : $P_p = x_{\left(\frac{p(n+1)}{100}\right)}$

Donde x está ordenado [14].

El rango intercuartil es la diferencia entre el percentil 75 (Q3) y el percentil 25 (Q1) [14].

$$IQR = Q3 - Q1$$

El coeficiente de variación es la relación entre la desviación estándar y la media, expresada como un porcentaje [15].

$$CV = \left(\frac{s}{\bar{x}}\right) \times 100$$

El intervalo de confianza proporciona un rango dentro del cual se espera que se encuentre la media poblacional con un cierto nivel de confianza (normalmente 95%) [16].

$$\left(\bar{x} - t_{\alpha/2} \left(\frac{s}{\sqrt{n}} \right), \bar{x} + t_{\alpha/2} \left(\frac{s}{\sqrt{n}} \right) \right)$$

El código programado para generar las estadísticas descriptivas mencionadas anteriormente para un conjunto de datos específico funciona de la siguiente manera. Primero, se asegura que los datos han sido cargados y que contienen suficientes filas para el análisis. Acto seguido, se da lugar al cálculo de las estadísticas descriptivas por medio de los siguientes paquetes y funciones de R.

Del paquete **base** de R se obtienen las funciones que calculan las estadísticas: **mean()** calcula la media de los valores, excluyendo valores faltantes con el parámetro **na.rm = TRUE**; **median()** calcula la mediana de los valores, excluyendo valores faltantes con el parámetro **na.rm = TRUE**; **sd()** calcula la desviación estándar, excluyendo valores faltantes con el parámetro **na.rm = TRUE**; **quantile()** calcula los percentiles especificados, excluyendo valores faltantes con el parámetro **na.rm = TRUE**; **IQR()** calcula el rango intercuartil, excluyendo valores faltantes con el parámetro **na.rm = TRUE** [17].

Mientras, a través del paquete **DT** se hace uso de la función **datatable()** que genera una tabla interactiva para visualizar las estadísticas descriptivas calculadas [18].

Asimismo, se usa el paquete **htmltools** con la función **tags\$caption()** para añadir un título descriptivo a la tabla generada para contextualizar los datos presentados [19].

3.2.1.2 Metodología del análisis de series temporales

Dentro de la metodología de análisis de series temporales, se utilizan diversas técnicas para comprender y predecir tendencias en los datos a lo largo del tiempo. La primera etapa en este análisis es la visualización de la serie temporal, donde se grafican los valores de un indicador a lo largo de diferentes periodos. Esto se lleva a cabo mediante el uso de la librería **ggplot2** para crear gráficos de líneas que representan los datos temporales. Estos gráficos permiten identificar patrones como tendencias y estacionalidades que son cruciales para el análisis predictivo.

Para complementar el análisis de series temporales, se utilizan gráficos ACF (Función de Autocorrelación) y descomposición STL (Seasonal-Trend Decomposition using Loess). Estas herramientas se utilizan para diagnosticar la autocorrelación y descomponer la serie temporal en sus componentes de tendencia, estacionalidad y ruido.

El gráfico ACF muestra la correlación de la serie temporal consigo misma en diferentes retardos (lags). La fórmula para calcular la autocorrelación para un retardo k es:

$$p_k = \frac{\sum_{t=k+1}^N (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2}$$

donde x_t es el valor de la serie temporal en el tiempo t , \bar{x} es la media de la serie temporal, y N es el número total de observaciones. Esta fórmula permite identificar la correlación entre los valores de la serie temporal separados por k periodos.

STL descompone una serie temporal en tres componentes: tendencia, estacionalidad y residuales. La fórmula general para la descomposición STL es:

$$Y_t = T_t + S_t + R_t$$

donde Y_t es el valor observado de la serie temporal en el tiempo t , T_t es el componente de tendencia, S_t es el componente estacional, y R_t es el componente residual.

Esta descomposición se realiza utilizando el método LOESS (Locally Estimated Scatterplot Smoothing) para ajustar suavemente la tendencia y la estacionalidad.

Estas técnicas se implementan en el proyecto utilizando las funciones `plot_acf_diagnostics` y `plot_stl_diagnostics`, que generan gráficos detallados de ACF y STL respectivamente, proporcionando una visión clara de los patrones de autocorrelación y la estructura de la serie temporal.

3.2.1.3 Metodología del análisis predictivo

Por último, el análisis predictivo se lleva a cabo utilizando varios modelos de series temporales: ARIMA, ETS y Prophet.

El modelo ARIMA (Autoregressive Integrated Moving Average) es uno de los métodos más utilizados para el análisis de series temporales. Se compone de tres componentes: autoregresivo (AR), diferencia integrada (I) y media móvil (MA). Este modelo es efectivo para series temporales no estacionarias que pueden hacerse estacionarias mediante diferenciación. Su fórmula es:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_p \epsilon_{t-p} + \epsilon_t$$

donde y_t es el valor en el tiempo t , φ son los parámetros autoregresivos, θ son los parámetros de la media móvil, ϵ_t es el error en t , y c es una constante [20].

El ETS (Exponential Smoothing State Space Model) se basa en la descomposición de la serie temporal en tres componentes: error, tendencia y estacionalidad. Este modelo es adecuado para series temporales con patrones estacionales y de tendencia bien definidos. La fórmula básica del modelo ETS se presenta en la forma de ecuaciones de estado:

$$\begin{aligned} y_t &= l_{t-1} + b_{t-1} + s_{t-m} + \epsilon_t \\ l_t &= \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \\ b_t &= \beta(l_t + l_{t-1}) + (1 - \beta)b_{t-1} \\ s_t &= \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \end{aligned}$$

donde l_t es el nivel, b_t es la tendencia, s_t es la estacionalidad, ϵ_t es el error y α, β, γ son parámetros de suavización [21].

Prophet es un modelo desarrollado por Facebook que es especialmente útil para series temporales con datos faltantes y grandes fluctuaciones estacionales. Utiliza una combinación de regresión aditiva y multiplicativa para ajustar los componentes de la serie temporal. La fórmula básica de Prophet es:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

donde $g(t)$ es el componente de tendencia, $s(t)$ es el componente estacional, $h(t)$ es el componente de días festivos y ϵ_t es el error [22].

La evaluación de estos modelos se realiza mediante la división de los datos en conjuntos de entrenamiento y prueba utilizando técnicas de validación cruzada. Se ajustan los modelos a los datos de entrenamiento y se calibran usando los

datos de prueba. La precisión de los modelos se evalúa utilizando las métricas del Error Absoluto Medio (MAE), el Error Cuadrático Medio (MSE) y el Error Porcentual Absoluto Medio (MAPE).

El MAE mide la magnitud promedio de los errores en un conjunto de predicciones, sin considerar su dirección [20].

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$$

donde y_t es el valor observado en el tiempo t , \hat{y}_t es el valor predicho, y n es el número de observaciones.

El MSE es la media de los errores al cuadrado entre los valores observados y predichos. Al elevar al cuadrado las diferencias, esta métrica penaliza más fuertemente los errores grandes [20].

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2$$

El MAPE expresa la precisión como un porcentaje de los errores en relación con los valores observados. Es útil para comparar la precisión entre diferentes series temporales, ya que es una métrica relativa [20].

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

La librería **modeltime** de **R** se emplea para gestionar el flujo de trabajo de los modelos de series temporales, incluyendo la calibración y la evaluación de la precisión de los modelos.

El análisis predictivo se complementa con la generación de gráficos de pronóstico que muestran las predicciones futuras junto con los datos históricos. Estos gráficos se generan mediante **plot_modeltime_forecast** de **modeltime**, que facilita la visualización de las predicciones y su comparación con los valores reales.

3.2.1.4 Metodología del análisis de *clustering*

El análisis de *clustering* es una técnica de aprendizaje no supervisado cuyo objetivo es agrupar un conjunto de objetos de tal manera que los objetos en el mismo grupo (o clúster) sean más similares entre sí que los objetos en otros grupos. En este proyecto, se han utilizado varios algoritmos de *clustering* para analizar datos de indicadores de calidad de vida, implementados en R mediante diversas bibliotecas y funciones.

Primero, se lleva a cabo el preprocesamiento de datos. Este paso incluye la imputación de datos faltantes utilizando los métodos de *k-Nearest Neighbors* (knn), *bagging*, regresión lineal, media, mediana, moda y rol, implementados con la biblioteca *recipes* de R. Además, se normalizan los datos para que todas las variables tengan una escala comparable, utilizando la función `step_normalize`.

El siguiente paso es el Análisis de Componentes Principales (PCA), una técnica utilizada para reducir la dimensionalidad de los datos antes del *clustering*. El PCA descompone los datos en componentes principales que explican la mayor parte de la varianza en los datos originales. La fórmula básica del PCA es:

$$Z = XW,$$

donde Z es la matriz de componentes principales, X es la matriz de datos original y W es la matriz de pesos de los componentes principales. Este paso se realiza con la función `prcomp`.

Para el *clustering*, están implementados varios algoritmos. Sin embargo, se aplica al análisis únicamente el algoritmo de *K-Means*, el cual divide el conjunto de datos en k clústeres y minimiza la varianza dentro de cada clúster. La fórmula del objetivo de *K-Means* es [23]:

$$\sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

donde C_i es el i -ésimo clúster y μ_i es el centroide del i -ésimo clúster.

La validación del *clustering* se realiza principalmente mediante la métrica de Inercia, que es una medida de la suma de las distancias cuadradas dentro del clúster. En el caso de *K-Means*, una menor inercia indica un mejor ajuste de los datos dentro de los clústeres. La fórmula para calcular la inercia es [23]:

$$\sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

En la implementación del código, los datos se cargan y preprocesan filtrando períodos y agrupando por nombre del indicador y país. Los datos se organizan en un formato ancho utilizando `pivot_wider`. Se configura un pipeline de preprocesamiento utilizando `recipe`, que incluye la imputación de datos faltantes, la eliminación de predictores con baja varianza y alta correlación, y la normalización de los datos. Luego, se aplica PCA a los datos preprocesados para reducir la dimensionalidad antes del *clustering*. Se aplica el algoritmo de *clustering* seleccionado a los datos transformados por PCA, y los resultados se almacenan y visualizan mediante gráficos y tablas. La inercia se calcula y muestra para evaluar el rendimiento del *clustering*. También se generan gráficos de contribución de los componentes principales y correlaciones entre las variables.

Mediante esta metodología, se obtiene una visión detallada y estructurada de los patrones y grupos dentro de los datos, permitiendo una mejor interpretación y análisis de los indicadores de calidad de vida permitiendo, así, comprender cómo se agrupan los datos y qué características son más importantes para la formación de clústeres.

3.3 Análisis de la Dimensión 1: *Main GDP aggregates per capita*

3.3.1 Análisis descriptivo y exploratorio de la Dimensión 1

El análisis descriptivo del indicador de los agregados principales del PIB per cápita proporciona una visión completa de la distribución de los ingresos en términos de promedio, mediana, desviación estándar, percentiles y otros estadísticos clave.

Como se puede ver en la Figura 5 el valor promedio es 17124.068 euros, mientras que la mediana es de 16873.9 euros, lo que sugiere una distribución relativamente equilibrada de los ingresos. La desviación estándar, que es de 7462.881 euros, indica una variabilidad considerable en los ingresos per cápita entre los países. El intervalo intercuartílico (IQR) es de 10819.6 euros, reflejando la dispersión entre el 25% y el 75% de la población. Además, el coeficiente de

variación (0.4358) muestra la relación entre la desviación estándar y la media, proporcionando una medida relativa de la dispersión. Finalmente, el intervalo de confianza para la media, que va de 16661.193 a 17586.942 euros, proporciona una estimación del rango en el que se espera que caiga la verdadera media poblacional con un alto nivel de confianza.

Statistic	Value	Units
Mean	17124.0681318681	Chain linked volumes (2010), euro per capita
Median	16873.9	Chain linked volumes (2010), euro per capita
Standard Deviation	7462.88113571383	Chain linked volumes (2010), euro per capita
25th Percentile	11474	Chain linked volumes (2010), euro per capita
50th Percentile	16873.9	Chain linked volumes (2010), euro per capita
75th Percentile	22293.6	Chain linked volumes (2010), euro per capita
Interquartile Range	10819.6	Chain linked volumes (2010), euro per capita
Coefficient of Variation	0.435812394475662	Chain linked volumes (2010), euro per capita
Confidence Interval	16661.1933303262 - 17586.9429334101	Chain linked volumes (2010), euro per capita

Figura 5 Gráfico Descriptivo Main GDP aggregates per capita. Fuente: Elaboración propia

3.3.2 Análisis de series temporales de la Dimensión 1

La serie temporal del PIB per cápita muestra un crecimiento constante desde 1995 hasta 2020. El gráfico de ACF indica que hay una correlación positiva significativa en los primeros rezagos, lo que sugiere que los valores pasados tienen una fuerte influencia en los valores futuros. La función STL descompone la serie en sus componentes estacional, tendencia y ruido. La tendencia muestra un crecimiento estable, mientras que la componente estacional revela fluctuaciones anuales regulares.

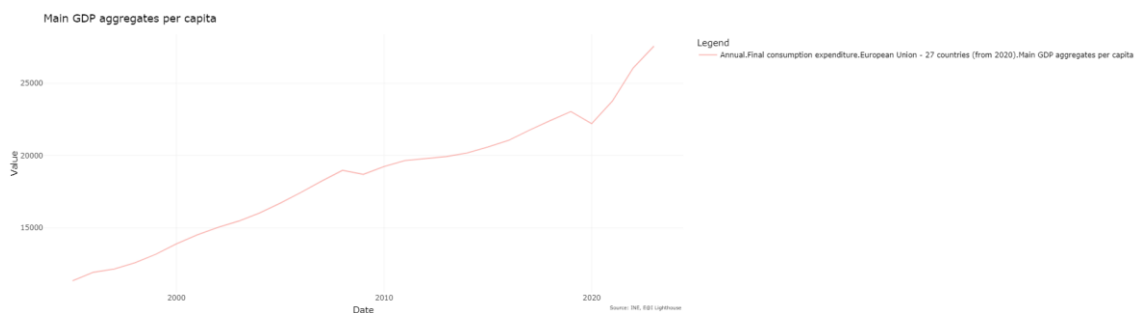


Figura 6 Serie temporal Main GDP aggregates per capita. Fuente: Elaboración Propia. Datos extraídos de: [12]

La ACF de la serie temporal del PIB per cápita muestra una disminución gradual, indicando que la serie tiene una correlación positiva significativa en los primeros rezagos. Esto sugiere que los valores del PIB per cápita en un año están influenciados por los valores de los años anteriores.

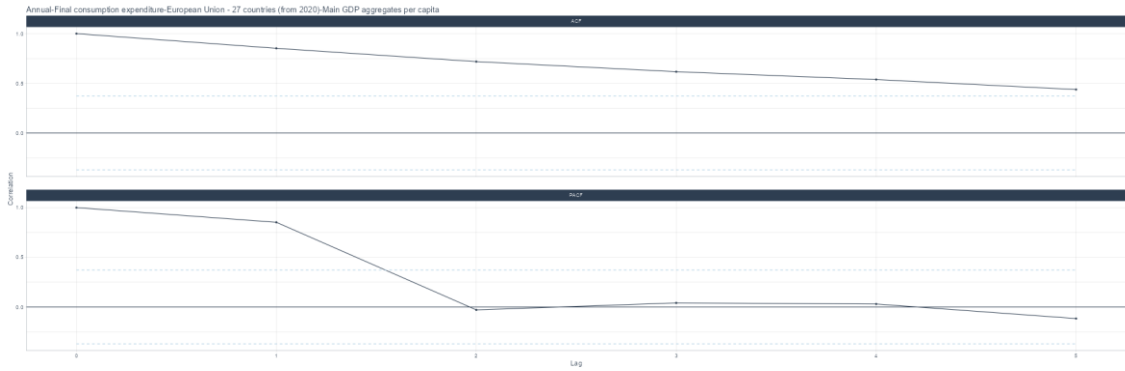


Figura 7 Análisis ACF Main GDP aggregates per capita. Fuente: Elaboración propia

El análisis STL descompone la serie en componentes de tendencia, estacionalidad y ruido. La tendencia muestra un crecimiento constante, mientras que la componente estacional revela fluctuaciones anuales regulares, probablemente relacionadas con ciclos económicos o eventos específicos.

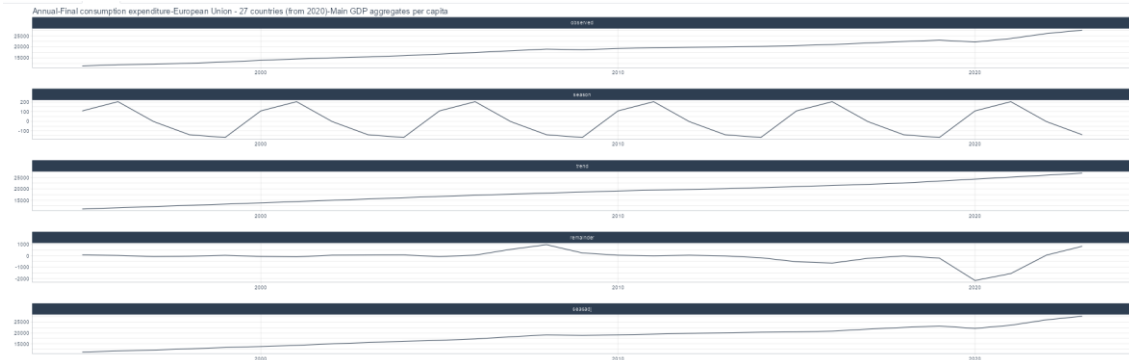


Figura 8 Análisis STL Main GDP aggregates per capita. Fuente: Elaboración propia. Datos extraídos de: [12]

3.3.3 Análisis predictivo de la Dimensión 1

El análisis predictivo a 5 años utiliza tres modelos: ARIMA(0,1,0) con deriva, ETS(M,A,N) y Prophet. Las métricas de precisión para el modelo ARIMA son MAE de 2724.18, RMSE de 2980.04 y R^2 de 0.99. Para el modelo ETS, las métricas son MAE de 2766.07, RMSE de 3025.21 y R^2 de 0.99. El modelo Prophet presenta un MAE de 2355.33, RMSE de 2648.43 y R^2 de 1.00. El gráfico de predicción muestra que los tres modelos siguen de cerca la tendencia histórica, con ligeras diferencias en las proyecciones futuras, siendo todas de tendencia de crecimiento positiva.

Accuracy for Annual-Final consumption expenditure-European Union - 27 countries (from 2020)-Main GDP aggregates per capita

.model_id	.model_desc	.type	mae	mape	mase	smape	rmse	rsq
1	ARIMA(0,1,0) WITH DRIFT	Test	2724.18	10.31	1.44	10.97	2980.04	0.99
2	ETS(M,A,N)	Test	2766.07	10.47	1.46	11.15	3025.21	0.99
3	PROPHET	Test	2355.33	8.87	1.24	9.39	2648.43	1.00

Figura 9 Métricas Modelos de Predicción Main GDP aggregates per capita. Fuente: Elaboración propia. Datos extraídos de: [12]

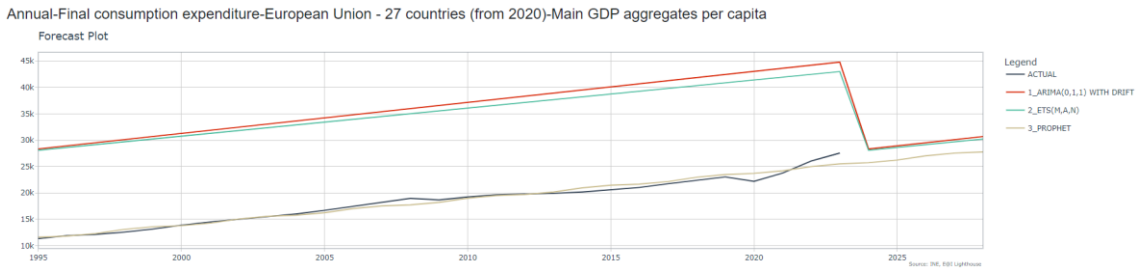


Figura 10 Gráficas Modelos de Predicción Main GDP aggregates per capita. Fuente: Elaboración propia. Datos extraídos de: [12]

3.3.4 Análisis de *clustering* de la Dimensión 1

El análisis de *clustering* aplicado al PIB per cápita en 2022 utiliza el algoritmo *K-means* con una configuración específica para obtener agrupaciones significativas entre los países europeos. Se permite un máximo de 5 valores faltantes por variable, se emplea el método de imputación *K-Nearest Neighbors* (KNN) y se aplica un umbral de correlación de 0.8 para evitar redundancias.

Para determinar el número óptimo de clústeres, se utiliza la inercia, una medida de la cohesión de los clústeres, que indica cuán cerca están los puntos dentro de un clúster. La inercia disminuye a medida que aumenta el número de clústeres, lo que sugiere una mejor separación entre los grupos. Sin embargo, esta disminución no es lineal, y en ciertos puntos, la reducción de inercia se vuelve menos significativa, lo que ayuda a identificar el número óptimo de clústeres.

En este caso, se observó que la inercia para 3 clústeres es 5.16, para 4 clústeres es 2.69, para 5 clústeres es 1.54, para 6 clústeres es 1.10 y para 7 clústeres es 0.72. La elección de cinco clústeres se justifica porque ofrece un equilibrio adecuado entre cohesión interna y separación significativa entre los clústeres. La reducción de inercia al pasar de 4 a 5 clústeres es significativa, mientras que las reducciones adicionales al aumentar el número de clústeres son menores, indicando que agregar más clústeres no proporcionaría una mejora sustancial en la segmentación.

El mapa de clústeres revela cómo los países con economías similares se agrupan juntos, mostrando patrones regionales. A continuación, se describen los clústeres identificados:

- Clúster 0 (Rojo: Irlanda y Reino Unido): Este clúster incluye a países como Irlanda y Reino Unido, caracterizados por tener un porcentaje bajo de PIB per cápita.
- Clúster 1 (Azul: España, Portugal y Grecia): Incluye países del sur de Europa como España, Portugal y Grecia, con niveles de PIB per cápita medio-bajos.
- Clúster 2 (Verde: Francia, Bélgica y Países Bajos): Este grupo incluye países de Europa occidental como Francia, Bélgica y Países Bajos, con niveles de PIB per cápita medios.
- Clúster 3 (Morado: Italia, Polonia y Hungría): Comprende países de Europa central y oriental como Italia, Polonia y Hungría, con niveles de PIB per cápita medios.
- Clúster 4 (Naranja: Suecia, Finlandia y Noruega): Incluye países nórdicos como Suecia, Finlandia y Noruega, con un alto porcentaje de PIB per cápita.
- Clúster 5 (Amarillo: Alemania y Austria): Agrupa a países del centro de Europa como Alemania y Austria, caracterizados por tener los niveles de PIB per cápita más altos del continente.

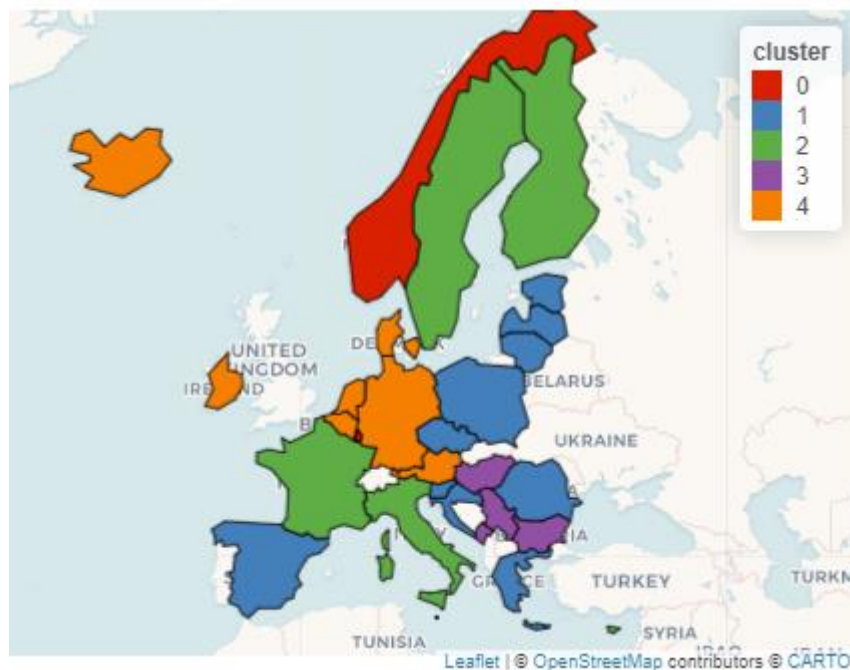


Figura 11 Análisis Clustering Main GDP aggregates per capita. Fuente: Elaboración propia. Datos extraídos de: [12]

3.4 Análisis de la Dimensión 2: *Employment rates by sex, age and educational attainment level*

3.4.1 Análisis descriptivo de la Dimensión 2

El análisis descriptivo de las tasas de empleo según el sexo, la edad y el nivel educativo se presenta en la tabla que muestra las estadísticas clave. El valor medio (Mean) de la tasa de empleo es de 65.52%, con una mediana (Median) de 68.4%. La desviación estándar (Standard Deviation) es de 14.79%, indicando variabilidad en las tasas de empleo entre los países. El percentil 25 (25th Percentile) es 57.5%, mientras que el percentil 75 (75th Percentile) es 76.7%, lo que da un rango intercuartílico (Interquartile Range) de 19.2%. El coeficiente de variación (Coefficient of Variation) es 0.23, y el intervalo de confianza (Confidence Interval) está entre 65.03% y 66.01%.

Employment rates by sex, age, educational attainment level, country of birth and degree of urbanisation		
Statistic	Value	Units
Mean	65.5207622298066	Percentage
Median	68.4	Percentage
Standard Deviation	14.7877890146672	Percentage
25th Percentile	57.5	Percentage
50th Percentile	68.4	Percentage
75th Percentile	76.7	Percentage
Interquartile Range	19.2	Percentage
Coefficient of Variation	0.22569622988818	Percentage
Confidence Interval	65.0317986924002 - 66.009725767213	Percentage

Figura 12 Descriptivo Employment rates by sex, age and educational attainment level. Fuente: Elaboración propia. Datos extraídos de: [12]

3.4.2 Análisis de series temporales de la Dimensión 2

El gráfico de series temporales muestra la evolución de las tasas de empleo desde 2005 hasta 2020 para diferentes grupos de edad y niveles educativos. Se observa un aumento gradual en las tasas de empleo a lo largo del tiempo, con algunas fluctuaciones en periodos específicos. Los grupos de edad más jóvenes y los que poseen niveles educativos más altos tienden a tener tasas de empleo más estables y elevadas.

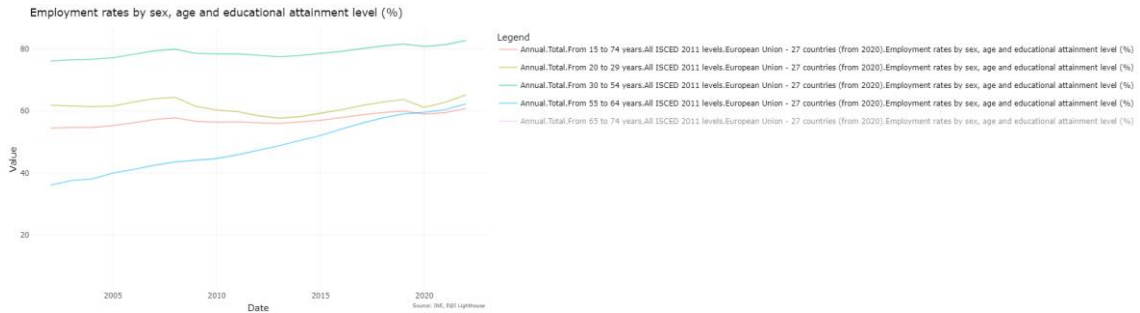


Figura 13 Serie Temporal Employment rates by sex, age and educational attainment level. Fuente: Elaboración propia. Datos extraídos de: [12]

El análisis de la ACF (Autocorrelation Function) y el STL (Seasonal-Trend Decomposition using LOESS) revela patrones importantes en las series temporales de las tasas de empleo. La ACF muestra la correlación de las tasas de empleo con sus propios valores en diferentes retardos (lags). Para las tasas de empleo, la ACF disminuye gradualmente, indicando una autocorrelación positiva moderada que se desvanece con el tiempo. La descomposición STL desglosa la serie temporal en componentes de tendencia, estacionalidad y residuo. Los resultados indican que la estacionalidad tiene un impacto significativo en las tasas de empleo, especialmente en los grupos de edad más jóvenes y mayores, mientras que la tendencia a largo plazo muestra un aumento constante en la participación laboral.

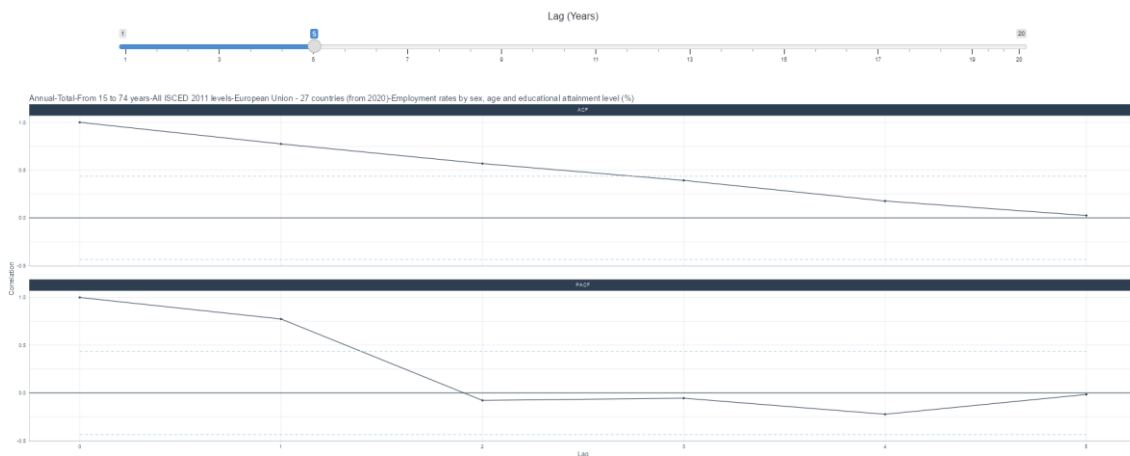


Figura 14 ACF Employment rates by sex, age and educational attainment level. Fuente: Elaboración propia. Datos extraídos de: [12]

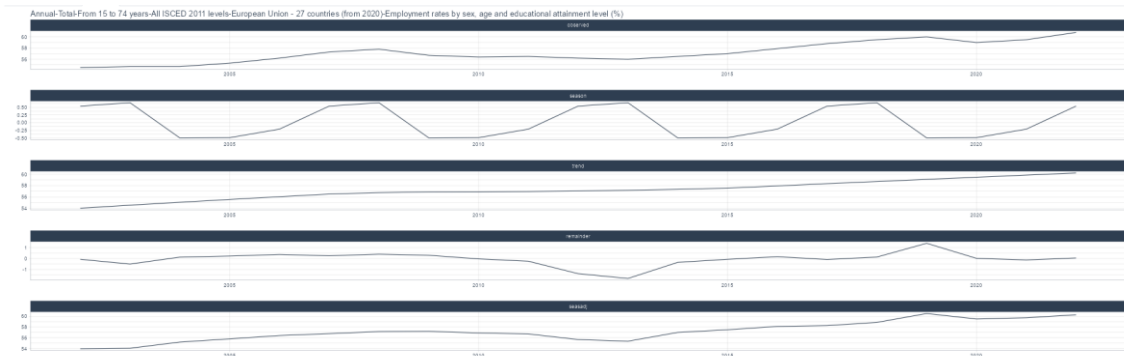


Figura 15 STL Employment rates by sex, age and educational attainment level. Fuente: Elaboración propia. Datos extraídos de: [12]

3.4.3 Análisis predictivo de la Dimensión 2

El análisis predictivo a 5 años utiliza tres modelos: ARIMA, ETS y Prophet. La tabla de precisión de los modelos muestra que el modelo ARIMA(1,1,1) con drift tiene un MAE de 0.89 y un RMSE de 1.03, mientras que el modelo ETS(M,N,N) presenta un MAE de 0.77 y un RMSE de 0.79. El modelo Prophet tiene un MAE de 1.92 y un RMSE de 1.94. El gráfico de predicción indica que el modelo ETS proporciona una buena predicción de las tasas de empleo, ajustándose mejor a los datos históricos comparado con los otros modelos, prediciendo, así una tendencia de crecimiento positiva para los próximos 5 años.

Accuracy for Annual-Total-From 15 to 74 years-All ISCED 2011 levels-European Union - 27 countries (from 2020)-Employment rates by sex, age and educational attainment level (%)

model_id	model_desc	type	mae	mape	mase	smape	rmse	rsq
1	ARIMA(0,1,1) WITH DRIFT	Test	0.89	1.50	0.99	1.49	1.03	0.94
2	ETS(M,N,N)	Test	0.77	1.28	0.85	1.28	0.79	NA
3	PROPHET	Test	1.92	3.22	2.14	3.17	1.94	0.94

Figura 16 Métricas Predicción Employment rates by sex, age and educational attainment level. Fuente: Elaboración propia. Datos extraídos de: [12]

Annual-Total-From 15 to 74 years-All ISCED 2011 levels-European Union - 27 countries (from 2020)-Employment rates by sex, age and educational attainment level (%)

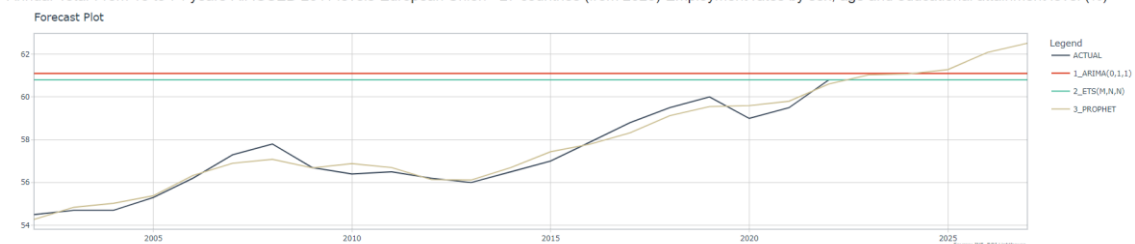


Figura 17 Gráficas Predicción Employment rates by sex, age and educational attainment level. Fuente: Elaboración propia. Datos extraídos de: [12]

3.4.4 Análisis de clustering de la Dimensión 2

El análisis de clustering aplicado a las tasas de empleo en 2022 utiliza el algoritmo K-means con una configuración específica para obtener agrupaciones significativas entre los países europeos. Se permite un máximo de 5 valores faltantes por variable, se emplea el método de imputación K-Nearest Neighbors (KNN) y se aplica un umbral de correlación de 0.8 para evitar redundancias.

Para determinar el número óptimo de clústeres, se utiliza la inercia, una medida de la cohesión de los clústeres, que indica cuán cerca están los puntos dentro de un clúster. La inercia disminuye a medida que aumenta el número de clústeres, lo que sugiere una mejor separación entre los grupos. Sin embargo, esta disminución no es lineal, y en ciertos puntos, la reducción de inercia se

vuelve menos significativa, lo que ayuda a identificar el número óptimo de clústeres.

En este caso, se observó que la inercia para 3 clústeres es 5.47, para 4 clústeres es 2.22, para 5 clústeres es 1.32 y para 6 clústeres es 0.93. La elección de cinco clústeres se justifica porque ofrece un equilibrio adecuado entre cohesión interna y separación significativa entre los clústeres. La reducción de inercia al pasar de 4 a 5 clústeres es significativa, mientras que las reducciones adicionales al aumentar el número de clústeres son menores, indicando que agregar más clústeres no proporcionaría una mejora sustancial en la segmentación.

El mapa de clústeres revela cómo los países con tasas de empleo similares se agrupan juntos, mostrando patrones regionales claros. A continuación, se describen los clústeres identificados:

- Clúster 0 (Rojo): Incluye países como Bélgica y Croacia, que tienen tasas de empleo más bajas en comparación con otros países europeos. Estos países se enfrentan a desafíos significativos en términos de empleo y mercado laboral.
- Clúster 1 (Azul): Este clúster agrupa a países como Alemania, que presentan tasas de empleo medias-altas. Estos países tienen economías estables y mercados laborales relativamente fuertes.
- Clúster 2 (Verde): Comprende países como Islandia y Noruega, que tienen las tasas de empleo más altas. Estos países se destacan por sus mercados laborales robustos y altos niveles de empleo.
- Clúster 3 (Morado): Agrupa a países como Bosnia y Herzegovina, que muestran tasas de empleo bajas y enfrentan desafíos económicos significativos que impactan en su mercado laboral.
- Clúster 4 (Naranja): Incluye países como Finlandia y Polonia, que tienen tasas de empleo medias. Estos países presentan mercados laborales desarrollados, pero con ciertos retos que limitan su crecimiento en términos de empleo.

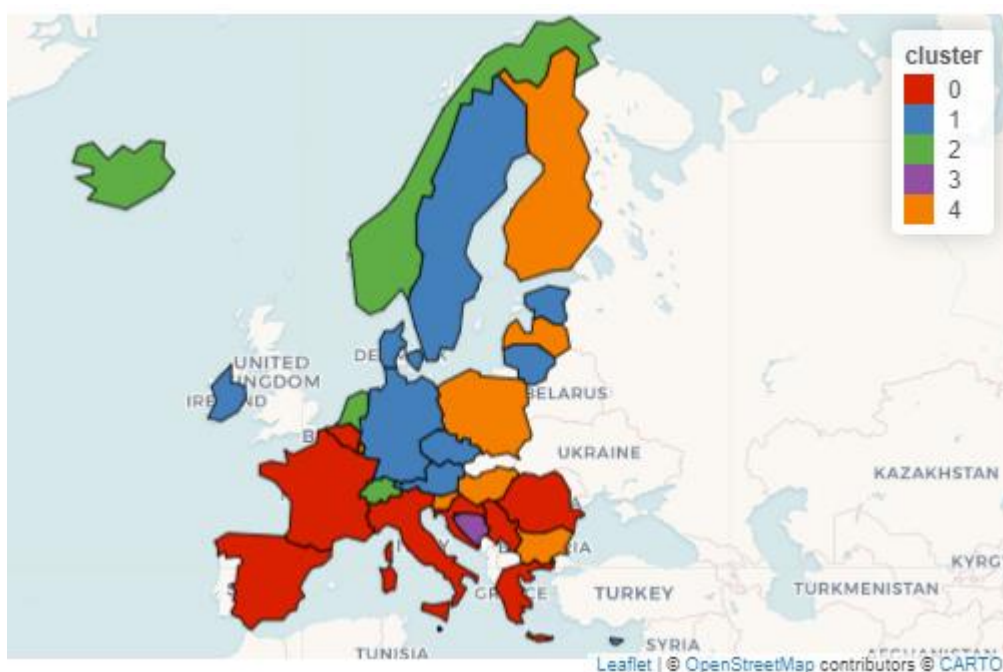


Figura 18 Clustering Employment rates by sex, age and educational attainment level. Fuente: Elaboración propia. Datos extraídos de: [12]

3.5 Análisis de la Dimensión 3: *Population by educational attachment level, sex and age*

3.5.1 Análisis descriptivo de la Dimensión 3

El análisis descriptivo de la dimensión de educación se basa en los estadísticos presentados. La media del porcentaje de la población por nivel de educación es del 41.01%, mientras que la mediana se sitúa en 31.8%. La desviación estándar es de 23.79%, lo que indica una variabilidad considerable en los niveles de educación entre los países. El rango intercuartílico es de 43.6%, con el 25° percentil en 21.6% y el 75° percentil en 65.2%. El coeficiente de variación es de 0.58, sugiriendo una dispersión moderada en los datos. El intervalo de confianza para la media es de 40.11% a 41.91%, lo que proporciona una estimación precisa del nivel medio de educación en la población.

Population by educational attainment level, sex and age (%) - main indicators		
Statistic	Value	Units
Mean	41.0098281016442	Percentage
Median	31.8	Percentage
Standard Deviation	23.7891707988644	Percentage
25th Percentile	21.6	Percentage
50th Percentile	31.8	Percentage
75th Percentile	65.2	Percentage
Interquartile Range	43.6	Percentage
Coefficient of Variation	0.580084626053592	Percentage
Confidence Interval	40.1101059054707 - 41.9095502978178	Percentage

Figura 19 Descriptivo *Population by educational attachment level, sex and age*. Fuente: Elaboración propia. Datos extraídos de: [12]

3.5.2 Análisis de series temporales de la Dimensión 3

El análisis de series temporales revela tendencias en los niveles educativos desde 2005 hasta 2020. Las series temporales indican un incremento gradual en los niveles de educación en la mayoría de los países europeos.



Figura 20 *Population by educational attachment level, sex and age*. Fuente: Elaboración propia. Serie Temporal Datos extraídos de: [12]

La ACF muestra la autocorrelación en diferentes retardos, indicando cómo los valores en diferentes momentos están correlacionados entre sí. En este caso, se observa una disminución gradual en la autocorrelación a medida que aumenta el lag, lo que sugiere que los niveles educativos tienen una dependencia temporal, pero esta disminuye con el tiempo. La descomposición STL separa la serie temporal en componentes de tendencia, estacionalidad y ruido. La tendencia muestra un incremento continuo en los niveles educativos a lo largo del tiempo, mientras que los componentes estacionales revelan patrones cíclicos

anuales.



Figura 21 ACF Population by educational attachment level, sex and age. Fuente: Elaboración propia. Datos extraídos de: [12]

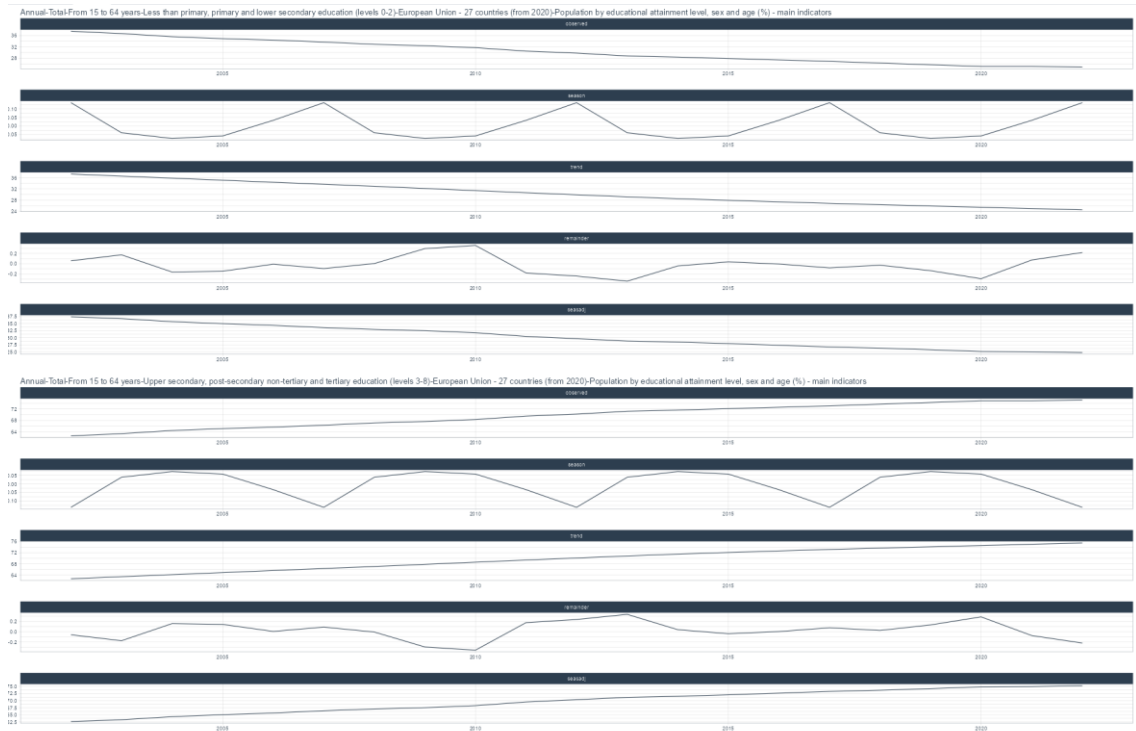


Figura 22 STL Population by educational attachment level, sex and age. Fuente: Elaboración propia. Datos extraídos de: [12]

3.5.3 Análisis predictivo de la Dimensión 3

El análisis predictivo utiliza tres modelos: ARIMA, ETS y Prophet. Los resultados de precisión indican que el modelo ARIMA(0,1,0) WITH DRIFT proporciona el ajuste más preciso con un RMSE de 1.54 y un R^2 de 0.75. El modelo ETS también muestra un buen desempeño con un RMSE de 1.54 y un R^2 de 0.76. El modelo Prophet, aunque menos preciso que los otros dos, aún ofrece una predicción razonable con un RMSE de 1.58 y un R^2 de 0.78. La tendencia futura predice un aumento continuo en los niveles educativos, alineado con la tendencia observada en el análisis de series temporales.

Accuracy for Annual-Total-From 15 to 64 years-Less than primary, primary and lower secondary education (levels 0-2)-European Union - 27 countries (from 2020)-Population by educational attainment level, sex and age (%) - main indicators

_model_id	_model_desc	type	mae	mape	mape	smape	rmse	rsq
1	ARIMA(1,0) WITH DRIFT	Test	0.72	2.69	7.22	2.95	0.87	0.75
2	ETS(A,A,N)	Test	0.72	2.67	7.16	2.94	0.87	0.75
3	PROPHET	Test	0.42	1.68	4.21	1.71	0.51	0.78

Accuracy for Annual-Total-From 15 to 64 years-Tertiary education (levels 5-8)-European Union - 27 countries (from 2020)-Population by educational attainment level, sex and age (%) - main indicators

_model_id	_model_desc	type	mae	mape	mape	smape	rmse	rsq
1	ARIMA(1,0) WITH DRIFT	Test	0.34	1.13	0.52	1.14	0.34	0.98
2	ETS(A,A,N)	Test	0.33	1.12	0.51	1.12	0.34	0.98
3	PROPHET	Test	0.44	1.48	0.67	1.49	0.44	0.98

Accuracy for Annual-Total-From 15 to 64 years-Upper secondary, post-secondary non-tertiary and tertiary education (levels 3-8)-European Union - 27 countries (from 2020)-Population by educational attainment level, sex and age (%) - main indicators

_model_id	_model_desc	type	mae	mape	mape	smape	rmse	rsq
1	ARIMA(1,0) WITH DRIFT	Test	0.72	0.95	7.22	0.95	0.87	0.75
2	ETS(A,A,N)	Test	0.72	0.95	7.24	0.95	0.87	0.75
3	PROPHET	Test	0.43	0.58	4.33	0.58	0.52	0.79

Figura 23 Métricas Predicción Population by educational attachment level, sex and age. Fuente: Elaboración propia. Datos extraídos de: [12]

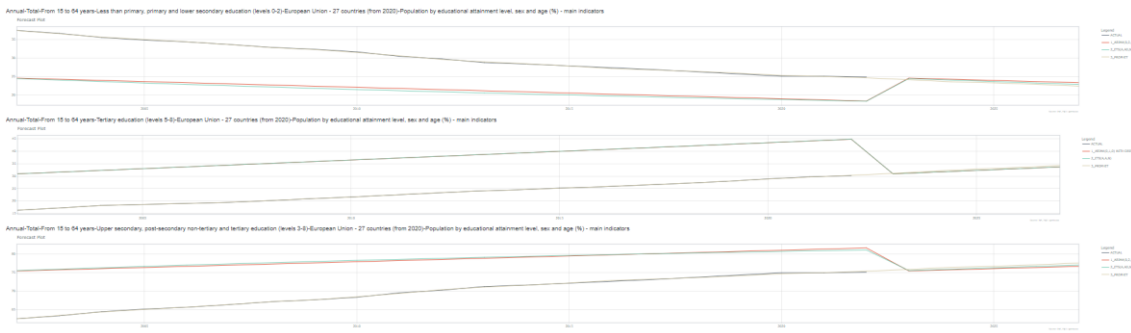


Figura 24 Gráficas Predicción Population by educational attachment level, sex and age. Fuente: Elaboración propia. Datos extraídos de: [12]

3.5.4 Análisis de clustering de la Dimensión 3

El análisis de *clustering* para 2022 se realiza utilizando el algoritmo *K-means* con una configuración específica: un máximo de 5 valores faltantes por variable, imputación mediante *K-Nearest Neighbors* y un umbral de correlación de 0.8. Se identifican seis clústeres. La inercia, que mide la cohesión de los clústeres, es de 2.00 para 4 clústeres, 1.32 para 5 clústeres, 0.76 para 6 clústeres y 0.50 para 7 clústeres. La elección de seis clústeres se justifica por la significativa reducción de la inercia al pasar de 5 a 6 clústeres, lo que indica una mejor separación de los grupos. El mapa de clústeres muestra cómo los países con niveles educativos similares se agrupan, destacando patrones regionales.

Los clústeres identificados son los siguientes:

- Clúster 0 (Rojo: Irlanda, Chipre y Luxemburgo): Este clúster incluye a países como Irlanda, Chipre y Luxemburgo, caracterizados por tener un porcentaje bajo de población con educación avanzada.
- Clúster 1 (Azul: Alemania, Grecia, Portugal y otros): Incluye países del sur y centro de Europa como Alemania, Grecia y Portugal, con niveles educativos medio-bajos.
- Clúster 2 (Verde: Italia, Rumania y otros): Este grupo incluye países de Europa del sur y este como Italia y Rumania, con niveles educativos medios.
- Clúster 3 (Morado: Bulgaria, Croacia, Hungría y otros): Comprende países de Europa central y oriental como Bulgaria, Croacia y Hungría, con niveles educativos medios.
- Clúster 4 (Naranja: Bélgica, Lituania, Noruega y otros): Incluye países de Europa occidental y nórdica como Bélgica, Lituania y Noruega, con un alto porcentaje de población con educación avanzada.

- Clúster 5 (Amarillo: España, Francia, Finlandia y otros): Agrupa a países del centro y norte de Europa como España, Francia y Finlandia, caracterizados por tener los niveles educativos más altos del continente.

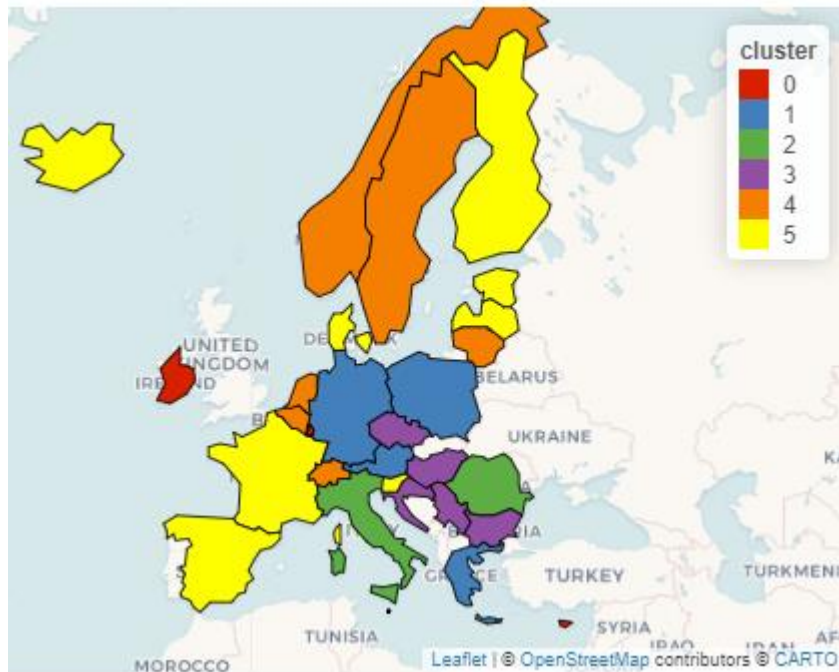


Figura 25 Clustering Population by educational attachment level, sex and age. Fuente: Elaboración propia. Datos extraídos de: [12]

3.6 Análisis de la Dimensión 4: *Healthy life years by sex*

3.6.1 Análisis descriptivo de la Dimensión 4

El análisis descriptivo de los años de vida saludable por sexo se basa en las estadísticas presentadas. La media es de 62.04%, mientras que la mediana se sitúa en 62.4%. La desviación estándar es de 4.85%, indicando una variabilidad moderada entre los países. El rango intercuartílico es de 6.7%, con el 25º percentil en 58.5% y el 75º percentil en 65.2%. El coeficiente de variación es de 0.078, lo que sugiere una dispersión baja en los datos. El intervalo de confianza para la media es de 61.62% a 62.46%, proporcionando una estimación precisa de los años de vida saludable promedio en la población.

Healthy life years by sex (from 2004 onwards)		
Statistic	Value	Units
Mean	62.04444444444444	Percentage
Median	62.4	Percentage
Standard Deviation	4.84715498978404	Percentage
25th Percentile	58.5	Percentage
50th Percentile	62.4	Percentage
75th Percentile	65.2	Percentage
Interquartile Range	6.7	Percentage
Coefficient of Variation	0.0781239163826224	Percentage
Confidence Interval	61.6276616204608 - 62.4612272684281	Percentage

Figura 26 Descriptivo Healthy life years by sex. Fuente: Elaboración propia. Datos extraídos de: [12]

3.6.2 Análisis de series temporales de la Dimensión 4

El análisis de series temporales incluye tanto la función de autocorrelación (ACF) como la descomposición STL (Seasonal-Trend decomposition using Loess). La ACF muestra la autocorrelación en diferentes retardos, indicando cómo los valores en diferentes momentos están correlacionados entre sí. En este caso, se observa una disminución gradual en la autocorrelación a medida que aumenta el lag, lo que sugiere que los años de vida saludable tienen una dependencia temporal que disminuye con el tiempo. La descomposición STL separa la serie temporal en componentes de tendencia, estacionalidad y ruido. La tendencia muestra un aumento general en los años de vida saludable a lo largo del tiempo, mientras que los componentes estacionales revelan patrones cíclicos anuales.

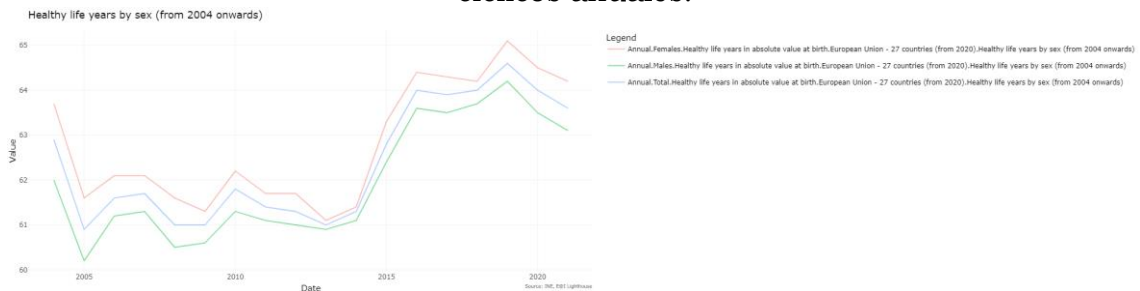


Figura 27 Serie Temporal Healthy life years by sex. Fuente: Elaboración propia. Datos extraídos de: [12]

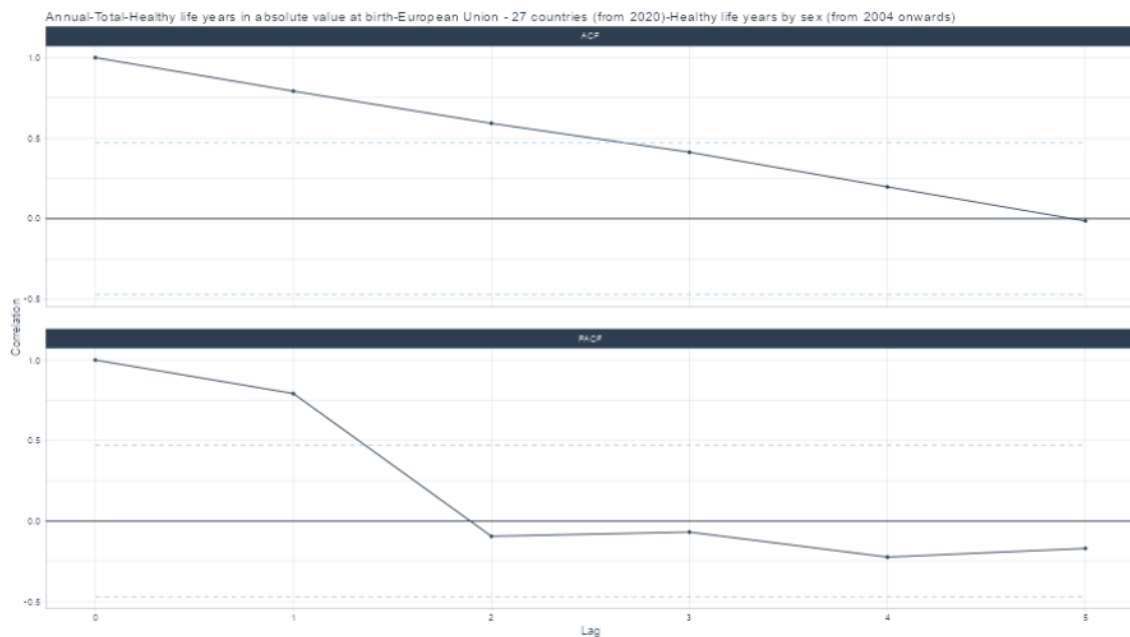


Figura 28 ACF Healthy life years by sex. Fuente: Elaboración propia. Datos extraídos de: [12]

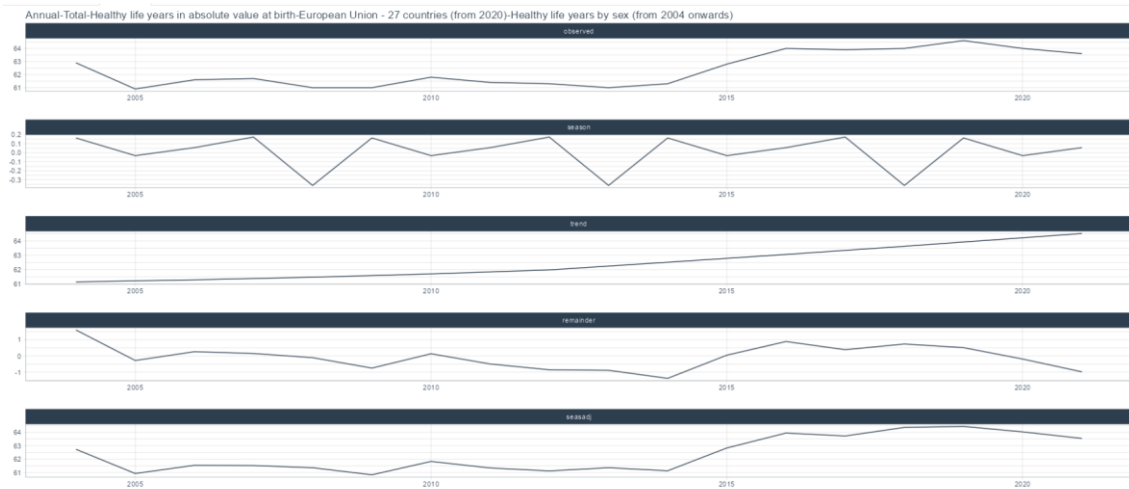


Figura 29 STL Healthy life years by sex. Fuente: Elaboración propia. Datos extraídos de: [12]

3.6.3 Análisis predictivo de la Dimensión 4

El análisis predictivo utiliza tres modelos: ARIMA, ETS y Prophet. Los resultados de precisión indican que el modelo ARIMA(0,1,0) WITH DRIFT proporciona el ajuste más preciso para los datos de mujeres con un RMSE de 1.16 y un R^2 de 0.98. El modelo ETS también muestra un buen desempeño con un RMSE de 1.70 y un R^2 de 0.82. Para los hombres, el modelo ARIMA(0,1,0) proporciona un RMSE de 1.25 y un R^2 de 0.94, mientras que ETS y Prophet tienen un RMSE de 1.95 y 2.85, respectivamente. En términos de predicción total, ARIMA sigue siendo el modelo más preciso con un RMSE de 1.25 y un R^2 de 0.94. La tendencia futura predice un aumento continuo en los años de vida saludable, alineado con la tendencia observada en el análisis de series temporales.

Accuracy for Annual-Females-Healthy life years in absolute value at birth-European Union - 27 countries (from 2020)-Healthy life years by sex (from 2004 onwards)

model_id	model_desc	type	mae	mape	mase	smape	rmse	rsq
1	ARIMA(0,1,0)	Test	0.75	1.17	2.5	1.16	0.76	NA
2	ETS(M,N,N)	Test	0.75	1.17	2.5	1.16	0.76	NA
3	PROPHET	Test	1.86	2.89	6.2	2.85	1.86	1

Accuracy for Annual-Males-Healthy life years in absolute value at birth-European Union - 27 countries (from 2020)-Healthy life years by sex (from 2004 onwards)

model_id	model_desc	type	mae	mape	mase	smape	rmse	rsq
1	ARIMA(0,1,0)	Test	0.90	1.42	2.25	1.41	0.92	NA
2	ETS(M,N,N)	Test	0.86	1.36	2.16	1.35	0.89	NA
3	PROPHET	Test	1.84	2.90	4.59	2.86	1.84	1

Accuracy for Annual-Total-Healthy life years in absolute value at birth-European Union - 27 countries (from 2020)-Healthy life years by sex (from 2004 onwards)

model_id	model_desc	type	mae	mape	mase	smape	rmse	rsq
1	ARIMA(0,1,0)	Test	0.80	1.25	2.00	1.25	0.82	NA
2	ETS(M,N,N)	Test	0.80	1.25	2.00	1.25	0.82	NA
3	PROPHET	Test	1.86	2.91	4.64	2.87	1.86	1

Figura 30 Métricas Predicción Healthy life years by sex. Fuente: Elaboración propia. Datos extraídos de: [12]

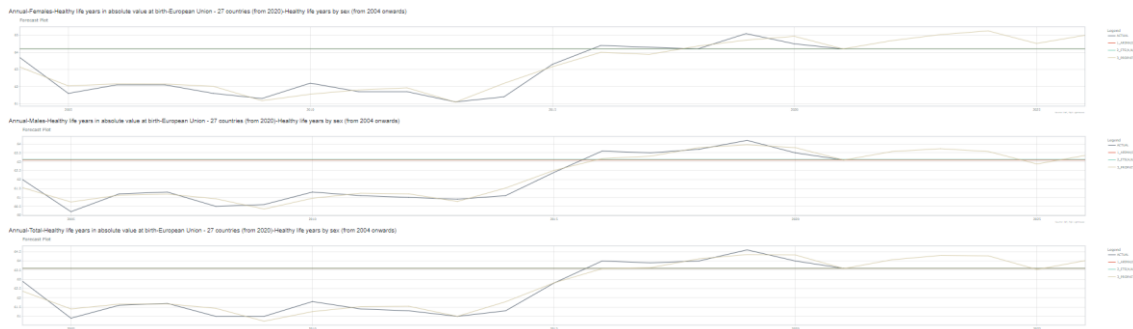


Figura 31 Gráficas Predicción Healthy life years by sex. Fuente: Elaboración propia. Datos extraídos de: [12]

3.6.4 Análisis de *clustering* de la Dimensión 4

El análisis de *clustering* aplicado a los años de vida saludable en 2022 utiliza el algoritmo *K-means* con una configuración específica para obtener agrupaciones significativas entre los países europeos. Se permite un máximo de 5 valores faltantes por variable, se emplea el método de imputación *K-Nearest Neighbors* (KNN) y se aplica un umbral de correlación de 0.8 para evitar redundancias. Para determinar el número óptimo de clústeres, se utiliza la inercia, una medida de la cohesión de los clústeres, que indica cuán cerca están los puntos dentro de un clúster. La inercia disminuye a medida que aumenta el número de clústeres, lo que sugiere una mejor separación entre los grupos. Sin embargo, esta disminución no es lineal, y en ciertos puntos, la reducción de inercia se vuelve menos significativa, lo que ayuda a identificar el número óptimo de clústeres. En este caso, se observó que la inercia para 3 clústeres es 2.47, para 4 clústeres es 1.50, para 5 clústeres es 0.79 y para 6 clústeres es 0.56. La elección de cinco clústeres se justifica porque ofrece un equilibrio adecuado entre cohesión interna y separación significativa entre los clústeres. La reducción de inercia al pasar de 4 a 5 clústeres es significativa, mientras que las reducciones adicionales al aumentar el número de clústeres son menores, indicando que agregar más clústeres no proporcionaría una mejora sustancial en la segmentación:

- Clúster 0 (Rojo: Italia y Malta): Este clúster incluye a países como Italia y Malta, caracterizados por tener los valores más bajos de años de vida saludable.
- Clúster 1 (Azul: Irlanda y Suecia): Incluye países como Irlanda y Suecia, con valores medios de años de vida saludable.
- Clúster 2 (Verde: Austria y Finlandia): Este grupo incluye países como Austria y Finlandia, que tienen valores medios-altos de años de vida saludable.
- Clúster 3 (Morado: Bélgica y Alemania): Comprende países como Bélgica y Alemania, con niveles altos de años de vida saludable.
- Clúster 4 (Naranja: Croacia y Rumanía): Incluye países como Croacia y Rumanía, que tienen un porcentaje bajo-medio de años de vida saludable.

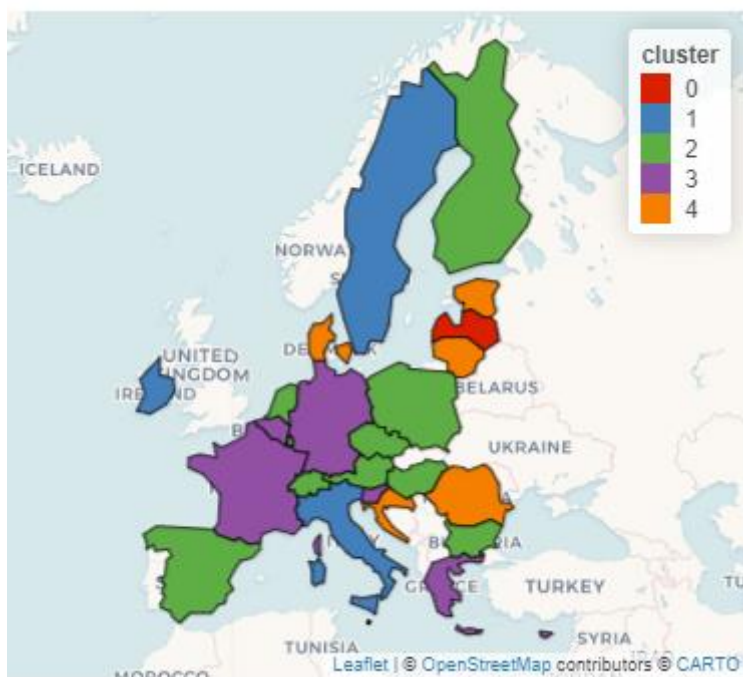


Figura 32 Clustering Healthy life years by sex. Fuente: Elaboración propia. Datos extraídos de: [12]

3.7 Análisis de la Dimensión 5: *Frequency of getting together with family and relatives or friends*

3.7.1 Análisis descriptivo y exploratorio de la Dimensión 5

El análisis descriptivo de la frecuencia de reuniones con familiares y amigos se basa en los estadísticos presentados. La media es del 13.52%, mientras que la mediana es de 13.45%. La desviación estándar es de 4.88%, lo que indica una variabilidad moderada en la frecuencia de las reuniones entre los países europeos. El rango intercuartílico es de 6.8%, con el 25° percentil en 10% y el 75° percentil en 16.8%. El coeficiente de variación es de 0.36, sugiriendo una dispersión moderada en los datos. El intervalo de confianza para la media es de 12.68% a 14.36%, proporcionando una estimación precisa de la frecuencia media de reuniones.

Frequency of getting together with family and relatives or friends by sex, age and educational attainment level		
Statistic	Value	Units
Mean	13.52	Percentage
Median	13.45	Percentage
Standard Deviation	4.88091362651065	Percentage
25th Percentile	10	Percentage
50th Percentile	13.45	Percentage
75th Percentile	16.8	Percentage
Interquartile Range	6.8	Percentage
Coefficient of Variation	0.361014321487474	Percentage
Confidence Interval	12.679586896625 - 14.360413103375	Percentage

Figura 33 Descriptivo Frequency of getting together with family and relatives or friends. Fuente: Elaboración propia. Datos extraídos de: [12]

3.7.2 Análisis de series temporales de la Dimensión 5

El análisis de series temporales incluye la función de autocorrelación (ACF) y la descomposición STL (Seasonal-Trend decomposition using Loess). La ACF

muestra una disminución gradual en la autocorrelación a medida que aumenta el lag, lo que sugiere que la frecuencia de reuniones tiene una dependencia temporal que disminuye con el tiempo. No cuenta con representación STL. La tendencia muestra un aumento continuo en la frecuencia de reuniones a lo largo del tiempo.

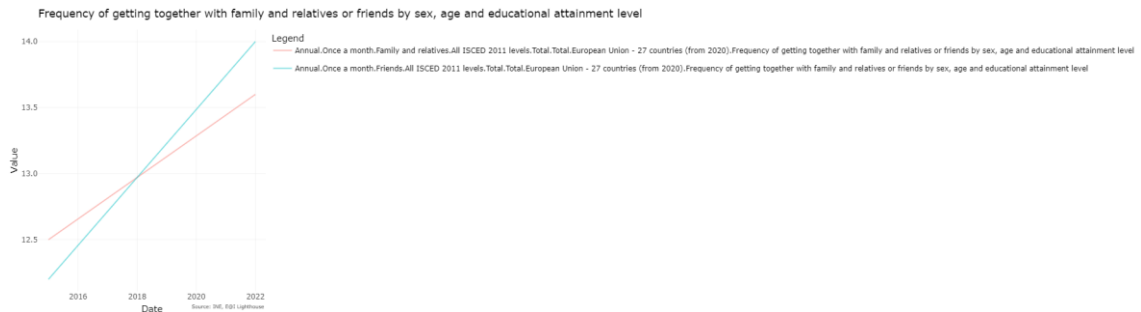


Figura 34 Serie Temporal Frequency of getting together with family and relatives or friends. Fuente: Elaboración propia. Datos extraídos de: [12]

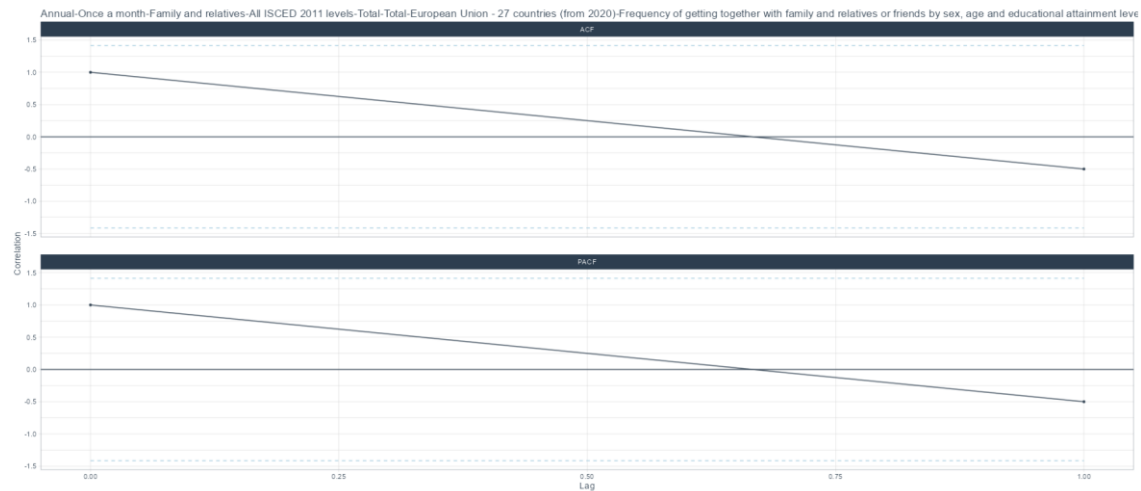


Figura 35 ACF Frequency of getting together with family and relatives or friends. Fuente: Elaboración propia. Datos extraídos de: [12]

3.7.3 Análisis predictivo de la Dimensión 5

El análisis predictivo utiliza tres modelos: ARIMA, ETS y Prophet. La frecuencia de los datos resulta ser insuficiente para obtener predicciones precisas.

3.7.4 Análisis de clustering de la Dimensión 5

El análisis de *clustering* para 2022 se realiza utilizando el algoritmo *K-means* con una configuración específica: un máximo de 5 valores faltantes por variable, imputación mediante *K-Nearest Neighbors* y un umbral de correlación de 0.8. Se identifican seis clústeres. La inercia, que mide la cohesión de los clústeres, es de 2.00 para 4 clústeres, 1.32 para 5 clústeres, 0.76 para 6 clústeres y 0.50 para 7 clústeres. La elección de seis clústeres se justifica por la significativa reducción de la inercia al pasar de 5 a 6 clústeres, lo que indica una mejor separación de los grupos. El mapa de clústeres muestra cómo los países con niveles similares de frecuencia de reuniones se agrupan juntos, destacando patrones regionales. Se identifican:

- Clúster 0 (Rojo): Incluye países como Croacia, Francia, Irlanda, Luxemburgo, Países Bajos y Eslovaquia, caracterizados por una frecuencia media de reuniones con familiares y amigos.
- Clúster 1 (Azul): Comprende países como Bulgaria, Estonia, Letonia, Lituania y Polonia, que tienen una frecuencia de reuniones relativamente baja en comparación con otros países.
- Clúster 2 (Verde): Agrupa a países como Bélgica, Grecia, Italia, Malta, Portugal y España, que presentan niveles medios de frecuencia de reuniones.
- Clúster 3 (Morado): Incluye a países como Chequia, Alemania, Islandia, Montenegro, Rumanía, Serbia, Suecia, Suiza y Reino Unido, caracterizados por una frecuencia alta de reuniones.
- Clúster 4 (Naranja): Comprende a Chipre, Macedonia del Norte, que tienen una frecuencia significativamente alta de reuniones con familiares y amigos.
- Clúster 5 (Amarillo): Incluye a Austria, Dinamarca, Hungría, Eslovenia, que presentan niveles muy altos de frecuencia de reuniones.

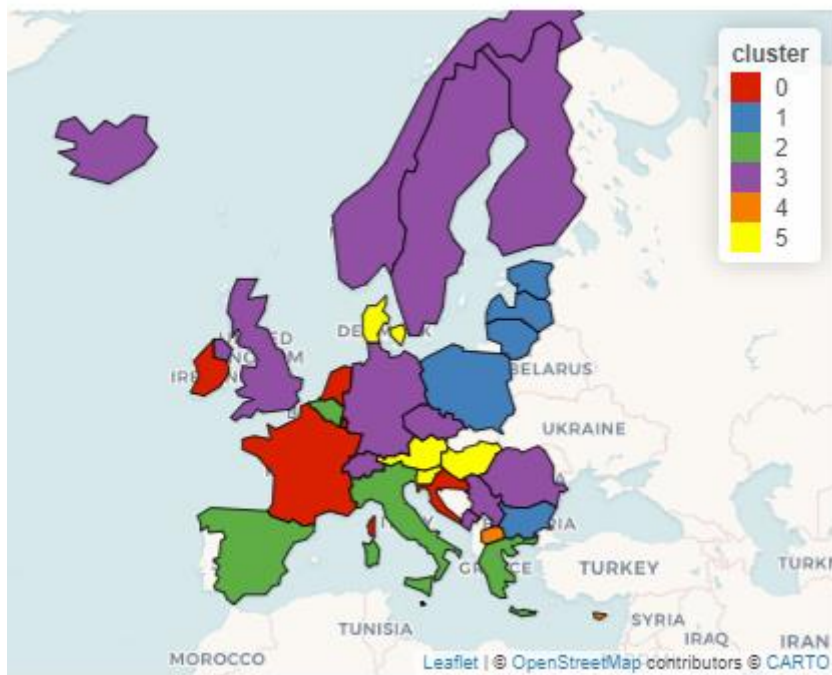


Figura 36 Clustering Frequency of getting together with family and relatives or friends. Fuente: Elaboración propia. Datos extraídos de: [12]

3.8 Análisis de la Dimensión 6: *Inability to face unexpected financial expenses*

3.8.1 Análisis descriptivo y exploratorio de la Dimensión 6

El análisis descriptivo de la incapacidad para enfrentar gastos financieros imprevistos revela que la media es del 36.37%, con una mediana ligeramente inferior del 33.7%. La desviación estándar es de 15.04%, indicando una variabilidad considerable entre los países. El rango intercuartílico es de 20.28%, con el 25° percentil en 24.53% y el 75° percentil en 44.8%. El coeficiente de variación es de 0.41, sugiriendo una dispersión moderada. El intervalo de confianza para la media está entre 35.19% y 37.54%, proporcionando una estimación precisa del nivel medio de incapacidad para enfrentar gastos imprevistos en la población.

Inability to face unexpected financial expenses - EU-SILC survey		
Statistic	Value	Units
Mean	36.365873015873	Percentage
Median	33.7	Percentage
Standard Deviation	15.0417583172576	Percentage
25th Percentile	24.525	Percentage
50th Percentile	33.7	Percentage
75th Percentile	44.8	Percentage
Interquartile Range	20.275	Percentage
Coefficient of Variation	0.413622912632735	Percentage
Confidence Interval	35.1890452710785 - 37.5427007606675	Percentage

Figura 37 Descriptivo Inability to face unexpected financial expenses. Fuente: Elaboración propia. Datos extraídos de: [12]

3.8.2 Análisis de series temporales de la Dimensión 6

El análisis temporal muestra que la incapacidad para enfrentar gastos financieros imprevistos ha disminuido con el tiempo, especialmente en los segmentos por debajo del 60% de la renta media equivalente. La función de autocorrelación (ACF) muestra una disminución gradual en la autocorrelación a medida que aumenta el lag, lo que sugiere que los valores están correlacionados con los valores previos pero esta correlación disminuye con el tiempo. La función parcial de autocorrelación (PACF) confirma esta observación, mostrando una disminución similar en los primeros lags.

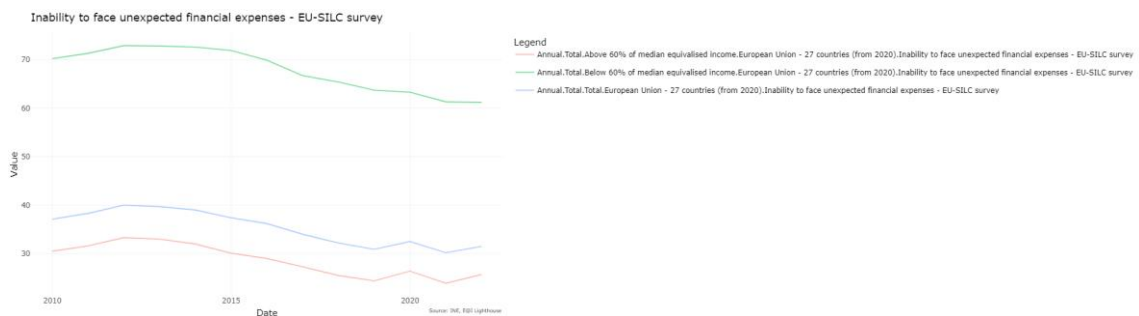


Figura 38 Serie Temporal Inability to face unexpected financial expenses. Fuente: Elaboración propia. Datos extraídos de: [12]

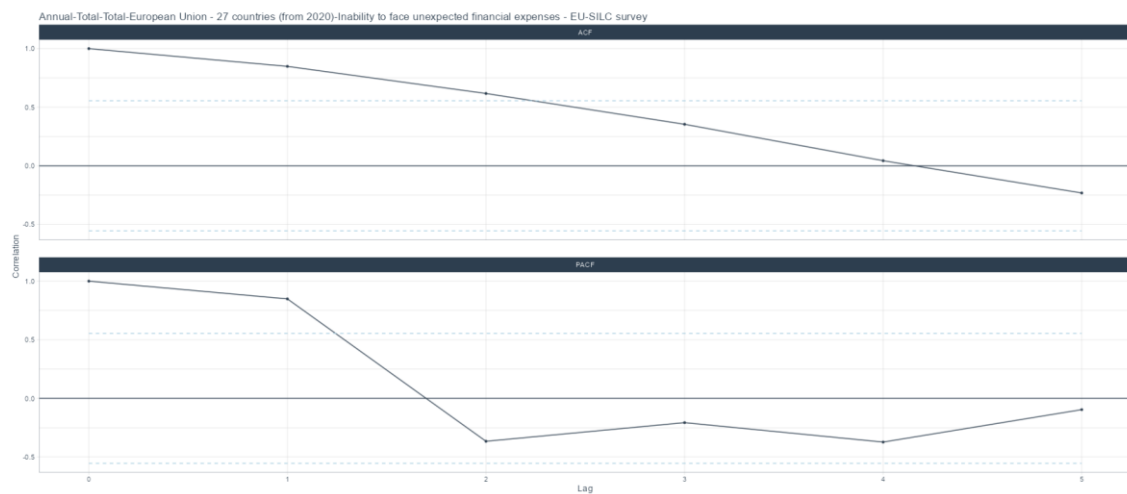


Figura 39 ACF Inability to face unexpected financial expenses. Fuente: Elaboración propia. Datos extraídos de: [12]

3.8.3 Análisis predictivo de la Dimensión 6

El análisis predictivo utiliza tres modelos: ARIMA, ETS y Prophet. Los resultados de precisión indican que el modelo ARIMA(0,1,0) es el que proporciona el ajuste más preciso para la incapacidad total para enfrentar gastos imprevistos, con un RMSE de 5.23 y un R^2 de 1.0. El modelo ETS(A,N,N) también muestra un buen desempeño con un RMSE de 5.27 y un R^2 de 1.0. El modelo Prophet, aunque menos preciso que los otros dos, aún ofrece una predicción razonable con un RMSE de 5.32 y un R^2 de 1.0. La tendencia futura predice una continuación de la disminución en la incapacidad para enfrentar gastos imprevistos.

Accuracy for Annual-Total-Above 60% of median equivalised income-European Union - 27 countries (from 2020)-Inability to face unexpected financial expenses - EU-SILC survey

_model_id	_model_desc	.type	mae	mape	mase	smape	rmse	rsq
1	ARIMA(0,1,0)	Test	1.60	6.59	0.89	6.31	1.84	NA
2	ETS(A,N,N)	Test	1.60	6.59	0.89	6.31	1.84	NA
3	PROPHET	Test	1.52	6.00	0.84	6.21	1.76	1

Accuracy for Annual-Total-Below 60% of median equivalised income-European Union - 27 countries (from 2020)-Inability to face unexpected financial expenses - EU-SILC survey

_model_id	_model_desc	.type	mae	mape	mase	smape	rmse	rsq
1	ARIMA(0,2,0)	Test	1.45	2.37	14.50	2.34	1.46	1
2	ETS(A,N,N)	Test	2.05	3.35	20.50	3.29	2.05	NA
3	PROPHET	Test	1.26	2.07	12.64	2.09	1.47	1

Accuracy for Annual-Total-Total-European Union - 27 countries (from 2020)-Inability to face unexpected financial expenses - EU-SILC survey

_model_id	_model_desc	.type	mae	mape	mase	smape	rmse	rsq
1	ARIMA(0,1,0)	Test	1.65	5.40	1.27	5.23	1.77	NA
2	ETS(A,N,N)	Test	1.65	5.39	1.27	5.23	1.77	NA
3	PROPHET	Test	1.31	4.21	1.00	4.26	1.36	1

Figura 40 Métricas Predicción Inability to face unexpected financial expenses. Fuente: Elaboración propia. Datos extraídos de: [12]

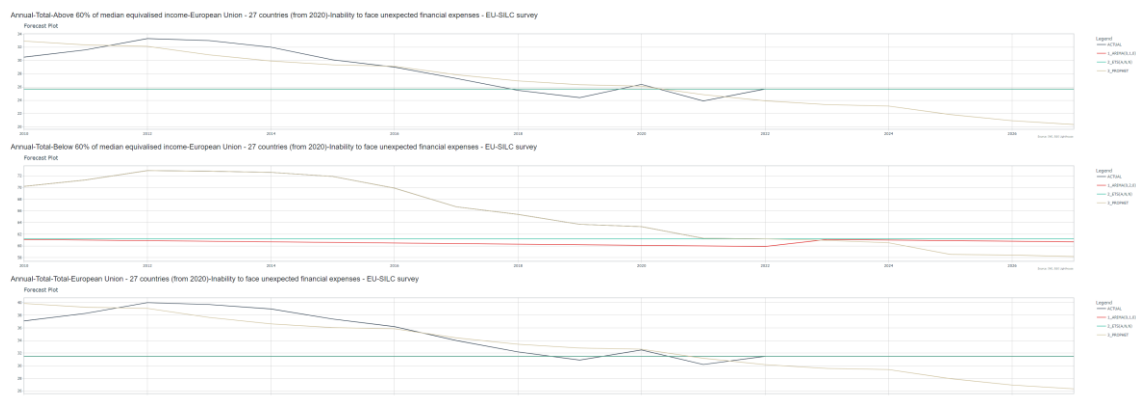


Figura 41 Gráficas Predicción Inability to face unexpected financial expenses. Fuente: Elaboración propia. Datos extraídos de: [12]

3.8.4 Análisis de clustering de la Dimensión 6

El análisis de *clustering* para 2022 se realiza utilizando el algoritmo *K-means* con una configuración específica: un máximo de 5 valores faltantes por variable, imputación mediante *K-Nearest Neighbors* y un umbral de correlación de 0.8. Se identifican cinco clústeres. La inercia, que mide la cohesión de los clústeres, es de 4.00 para 3 clústeres, 2.01 para 4 clústeres, 1.14 para 5 clústeres y 0.84 para 6 clústeres. La elección de cinco clústeres se justifica por la significativa reducción de la inercia al pasar de 4 a 5 clústeres, lo que indica una mejor separación de los grupos. Los clústeres identificados son los siguientes:

- Clúster 0 (Rojo: Bélgica, Finlandia, Noruega, Polonia, Suecia, Eslovenia)
- Clúster 1 (Azul: Bulgaria, Croacia, Chipre, Grecia, Letonia, Lituania, Rumania)
- Clúster 2 (Verde: Estonia, Francia, Alemania, Hungría, Irlanda, Italia, Portugal, Serbia, España)

- Clúster 3 (Morado: Montenegro)
- Clúster 4 (Naranja: Austria, República Checa, Dinamarca, Luxemburgo, Malta, Países Bajos, Suiza)

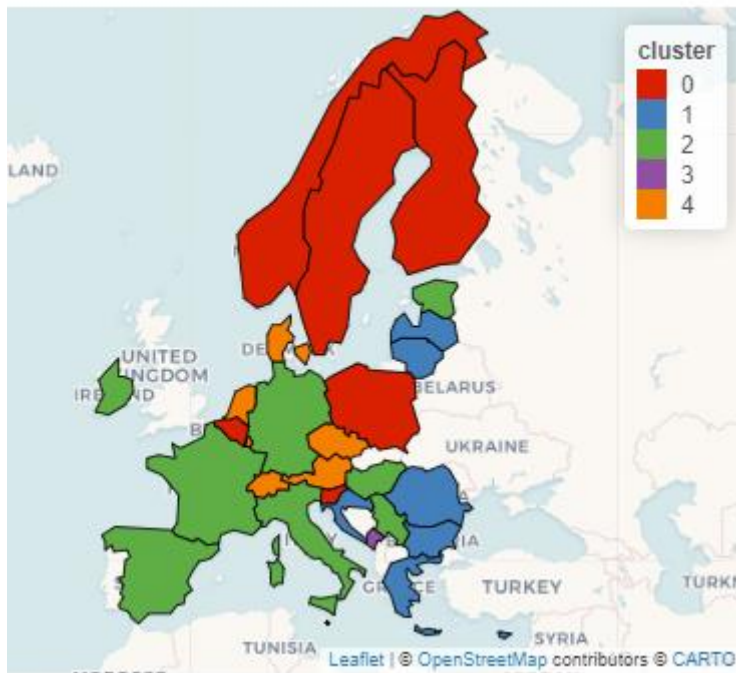


Figura 42 Clustering Inability to face unexpected financial expenses. Fuente: Elaboración propia. Datos extraídos de: [12]

3.9 Análisis de la Dimensión 7: Gender employment gap

3.9.1 Análisis descriptivo y exploratorio de la Dimensión 7

El análisis descriptivo del Gender Employment Gap (diferencia de empleo de género) muestra unas estadísticas que indican que la brecha de empleo de género tiene una media de 11.15%, con una mediana de 9.4%, lo que sugiere que la mayoría de los países tienen una brecha de empleo de género alrededor de estos valores. La desviación estándar de 7.74% muestra una variabilidad considerable en la brecha de empleo de género entre los países.

Statistic	Value	Units
Mean	11.151689860835	Percentage of total population
Median	9.4	Percentage of total population
Standard Deviation	7.7386333639648	Percentage of total population
25th Percentile	6.4	Percentage of total population
50th Percentile	9.4	Percentage of total population
75th Percentile	13.8	Percentage of total population
Interquartile Range	7.4	Percentage of total population
Coefficient of Variation	0.693942663447185	Percentage of total population
Confidence Interval	10.473772570446 - 11.829607151224	Percentage of total population

Figura 43 Descriptivo Gender employment gap Fuente: Elaboración propia. Datos extraídos de: [12]

3.9.2 Análisis de series temporales de la Dimensión 7

La serie temporal muestra una tendencia decreciente en los últimos 20 años, significando esto que se está reduciendo la brecha de género en los empleos europeos de forma general. La ACF muestra la autocorrelación en diferentes retardos, indicando cómo los valores en diferentes momentos están

correlacionados entre sí. En este caso, se observa una disminución gradual en la autocorrelación a medida que aumenta el lag, lo que sugiere que la brecha de empleo de género tiene una dependencia temporal, pero esta disminuye con el tiempo.

La descomposición STL separa la serie temporal en componentes de tendencia, estacionalidad y ruido. La tendencia muestra un decremento continuo en la brecha de empleo de género a lo largo del tiempo, mientras que los componentes estacionales revelan patrones cíclicos anuales.

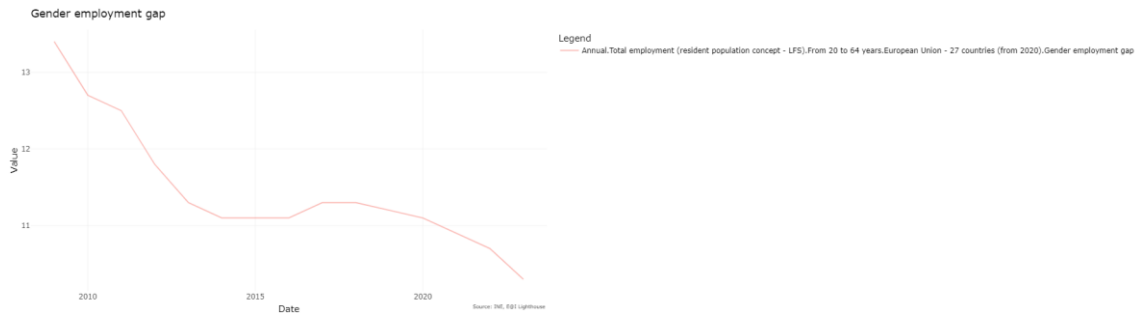


Figura 44 Serie Temporal Gender employment gap Fuente: Elaboración propia. Datos extraídos de: [12]

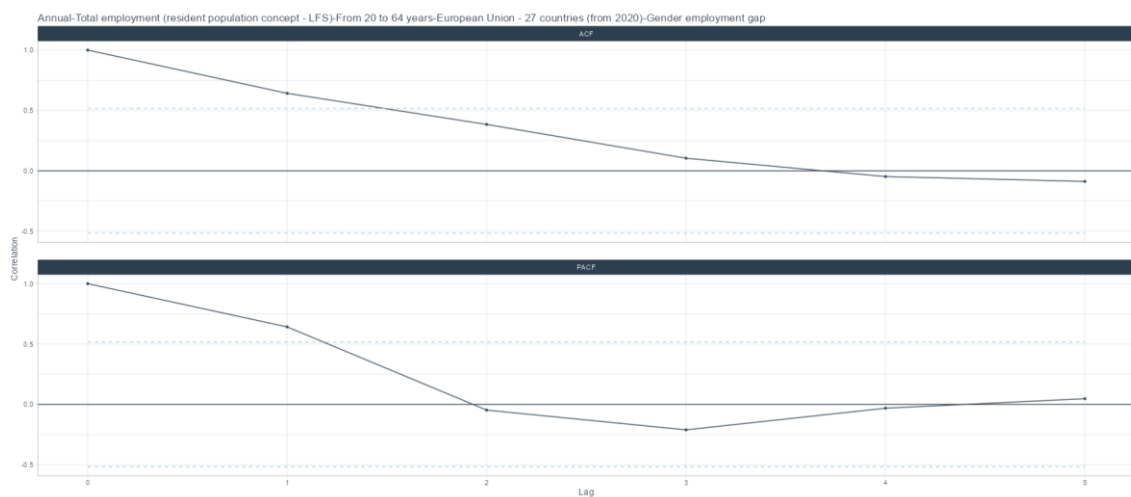


Figura 45 ACF Gender employment gap Fuente: Elaboración propia. Datos extraídos de: [12]

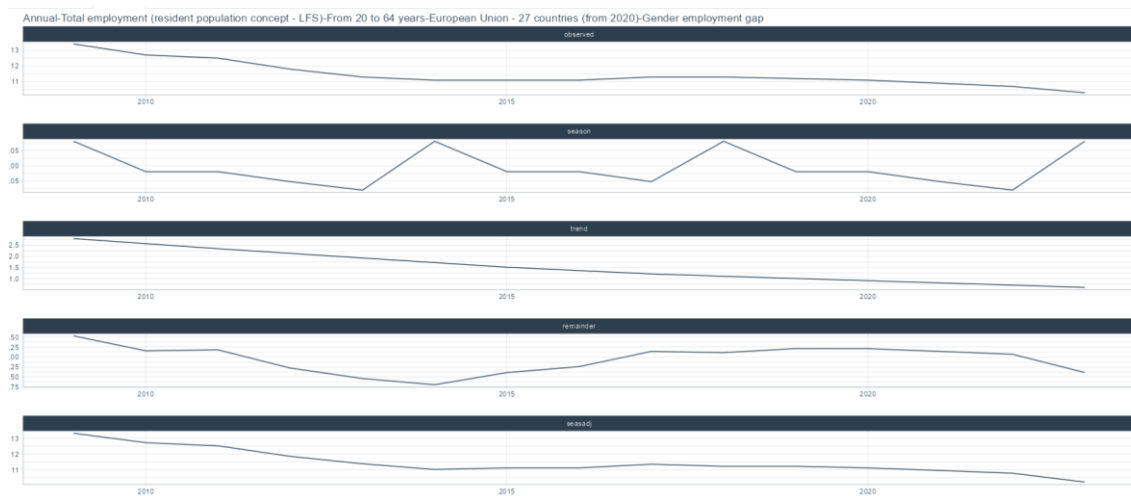


Figura 46 STL Gender employment gap Fuente: Elaboración propia. Datos extraídos de: [12]

3.9.3 Análisis predictivo de la Dimensión 7

El análisis predictivo utiliza tres modelos: ARIMA, ETS y Prophet. Los resultados de precisión indican que el modelo ARIMA(0,2,0) proporciona el ajuste más preciso con un RMSE de 0.14 y un R² de 1.0. El modelo ETS también muestra un buen desempeño con un RMSE de 0.45. El modelo Prophet, aunque menos preciso que los otros dos, aún ofrece una predicción razonable con un RMSE de 0.38. La tendencia futura predice una disminución continua en la brecha de empleo de género, alineado con la tendencia observada en el análisis de series temporales.

Accuracy for Annual-Total employment (resident population concept - LFS)-From 20 to 64 years-European Union - 27 countries (from 2020)-Gender employment gap

.model_id	.model_desc	.type	mae	mape	mase	smape	rmse	rsq
1	ARIMA(0,2,0)	Test	0.10	0.97	0.25	0.96	0.14	1
2	ETS(M,N,N)	Test	0.40	3.85	1.00	3.76	0.45	NA
3	PROPHET	Test	0.35	3.35	0.87	3.29	0.38	1

Figura 47 Métricas Predicción Gender employment gap Fuente: Elaboración propia. Datos extraídos de: [12]

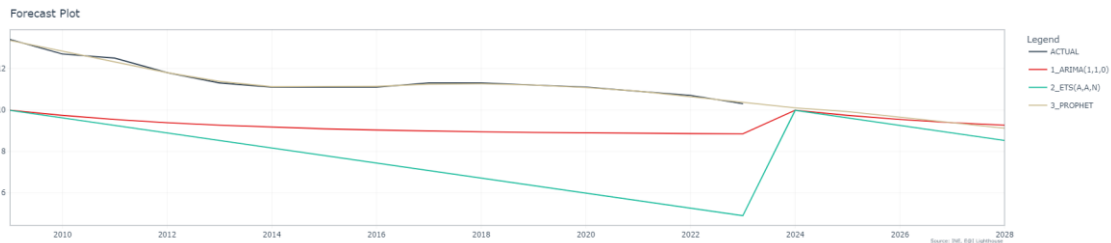


Figura 48 Gráficas Predicción Gender employment gap Fuente: Elaboración propia. Datos extraídos de: [12]

3.9.4 Análisis de clustering de la Dimensión 7

El análisis de *clustering* para la brecha de empleo de género en 2022 se realiza utilizando el algoritmo *K-means* con una configuración específica para obtener agrupaciones significativas entre los países europeos. Se permite un máximo de 5 valores faltantes por variable, se emplea el método de imputación *K-Nearest Neighbors* (KNN) y se aplica un umbral de correlación de 0.8 para evitar redundancias. Para determinar el número óptimo de clústeres, se utiliza la inercia, una medida de la cohesión de los clústeres, que indica cuán cerca están los puntos dentro de un clúster. La inercia disminuye a medida que aumenta el número de clústeres, lo que sugiere una mejor separación entre los grupos. Sin embargo, esta disminución no es lineal, y en ciertos puntos, la reducción de inercia se vuelve menos significativa, lo que ayuda a identificar el número óptimo de clústeres. En este caso, se observó que la inercia para 4 clústeres es 2.75, para 5 clústeres es 1.41, para 6 clústeres es 0.76, para 7 clústeres es 0.44 y para 8 clústeres es 0.33. La elección de siete clústeres se justifica porque ofrece un equilibrio adecuado entre cohesión interna y separación significativa entre los clústeres. La reducción de inercia al pasar de 6 a 7 clústeres es significativa, mientras que las reducciones adicionales al aumentar el número de clústeres son menores, indicando que agregar más clústeres no proporcionaría una mejora sustancial en la segmentación. Clústeres Identificados:

- Clúster 0 (Rojo): Incluye países como Estonia, Finlandia y Lituania, con una brecha de empleo de género relativamente baja.

- Clúster 1 (Azul): Incluye países como Chipre y Malta, con una brecha de empleo de género más alta.
- Clúster 2 (Verde): Incluye Bosnia y Herzegovina, con la brecha de empleo de género más alta.
- Clúster 3 (Morado): Incluye países como Dinamarca, Francia, y Noruega, con una brecha de empleo de género media.
- Clúster 4 (Naranja): Incluye países como Grecia e Italia, con una brecha de empleo de género alta.
- Clúster 5 (Amarillo): Incluye países como Croacia, Hungría, Irlanda y España, con una brecha de empleo de género moderada.
- Clúster 6 (Marrón): Incluye países como Austria, Bélgica, Bulgaria, y Alemania, con una brecha de empleo de género baja.

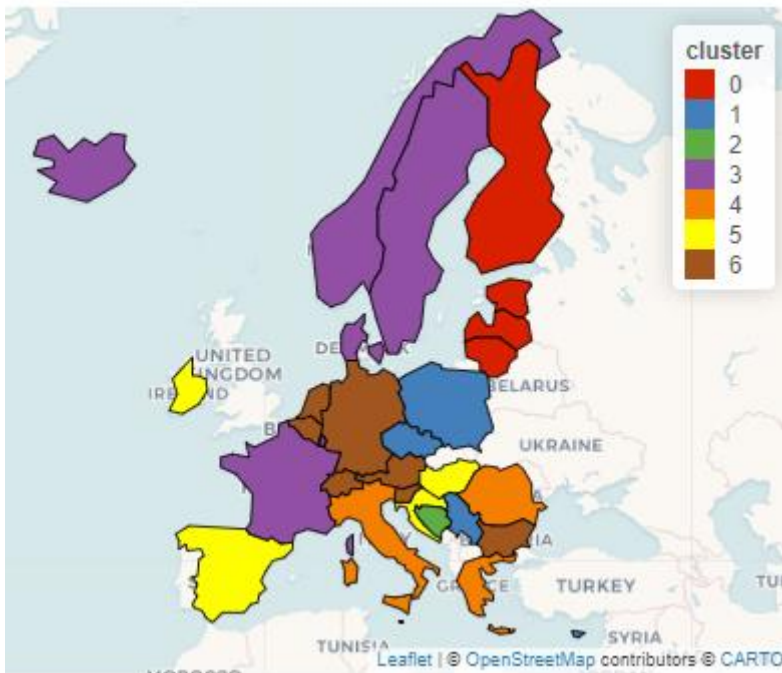


Figura 49 Clustering Gender employment gap Fuente: Elaboración propia. Datos extraídos de: [12]

3.10 Análisis de la dimensión 8: *Pollution, grime or other environmental problems*

3.10.1 Análisis descriptivo y exploratorio de la Dimensión 8

La media de la proporción de personas que experimentan problemas de contaminación y otros problemas ambientales en 2020 es del 14.46%. La mediana se sitúa en 13.6%, con una desviación estándar de 6.68%, indicando una variabilidad considerable en los datos. Los percentiles 25 y 75 son 9.3% y 17.4%, respectivamente, y el rango intercuartílico es de 8.1%. El coeficiente de variación es de 0.4623, sugiriendo una dispersión moderada. El intervalo de confianza para la media es de 13.90% a 15.01%.

Pollution, grime or other environmental problems - EU-SILC survey		
Statistic	Value	Units
Mean	14.4557522123894	Percentage
Median	13.6	Percentage
Standard Deviation	6.6823645596978	Percentage
25th Percentile	9.3	Percentage
50th Percentile	13.6	Percentage
75th Percentile	17.4	Percentage
Interquartile Range	8.1	Percentage
Coefficient of Variation	0.462263357971137	Percentage
Confidence Interval	13.9035641537704 - 15.0079402710084	Percentage

Figura 50 Descriptivo Pollution, grime or other environmental problems Fuente: Elaboración propia. Datos extraídos de: [12]

3.10.2 Análisis de series temporales de la Dimensión 8

El gráfico de series temporales muestra tres series diferenciadas para la contaminación y problemas ambientales en la UE: total por encima del 60% del ingreso mediano, total por debajo del 60% del ingreso mediano, y total de la Unión Europea. Las tres exponen tendencias decrecientes, aunque se tiene una clara diferencia en términos totales entre la población por encima del ingreso mediano que experimenta muchos menos problemas de contaminación que la población por debajo de este 60%. La función de autocorrelación (ACF) indica una disminución gradual en la autocorrelación a medida que aumenta el lag, lo que sugiere una dependencia temporal que disminuye con el tiempo.

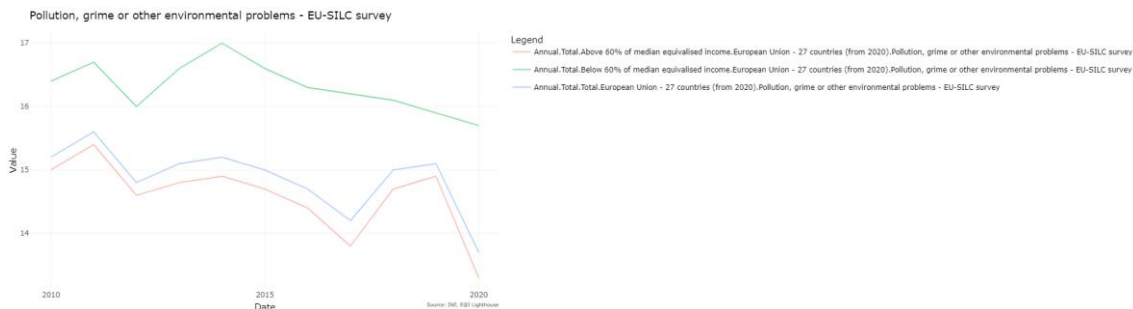


Figura 51 Serie Temporal Pollution, grime or other environmental problems Fuente: Elaboración propia. Datos extraídos de: [12]

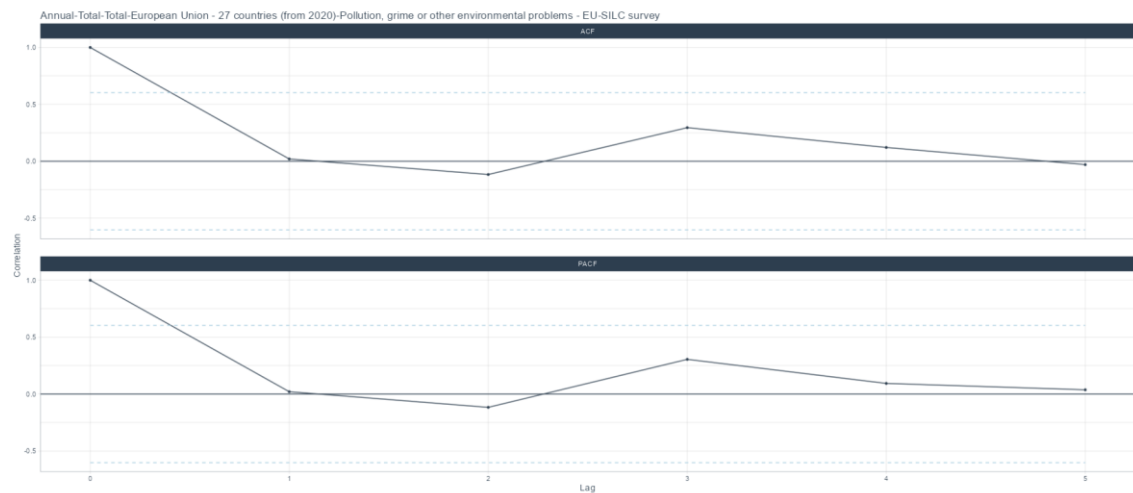


Figura 52 ACF Pollution, grime or other environmental problems Fuente: Elaboración propia. Datos extraídos de: [12]

3.10.3 Análisis predictivo de la Dimensión 8

Se utilizan tres modelos para el análisis predictivo: ARIMA, ETS y Prophet. Para la serie por encima del 60% del ingreso mediano, el modelo ARIMA(0,1,0) con media no nula presenta un MAE de 0.58 y un RMSE de 0.68. El modelo ETS(A,N,N) tiene un MAE de 5.84 y un RMSE de 6.12, mientras que Prophet muestra un MAE de 0.58 y un RMSE de 0.68. Para la serie por debajo del 60% del ingreso mediano, ARIMA presenta un MAE de 0.37 y un RMSE de 0.64, ETS tiene un MAE de 0.51 y un RMSE de 0.64, y Prophet muestra un MAE de 0.39 y un RMSE de 0.61. En la serie total, ARIMA tiene un MAE de 0.70 y un RMSE de 0.51, ETS presenta un MAE de 0.70 y un RMSE de 0.51, y Prophet muestra un MAE de 0.49 y un RMSE de 0.51. Los modelos predicen una ligera disminución en los problemas de contaminación en los próximos años.

Accuracy for Annual-Total-Above 60% of median equivalised income-European Union - 27 countries (from 2020)-Pollution, grime or other environmental problems - EU-SILC survey

.model_id	.model_desc	.type	mae	mape	mase	smape	rmse	rsq
1	ARIMA(0,1,0)	Test	0.80	5.93	0.50	5.68	1.00	NA
2	ETS(A,N,N)	Test	0.80	5.84	0.50	5.68	0.88	NA
3	PROPHET	Test	0.58	4.20	0.37	4.16	0.60	1

Accuracy for Annual-Total-Below 60% of median equivalised income-European Union - 27 countries (from 2020)-Pollution, grime or other environmental problems - EU-SILC survey

.model_id	.model_desc	.type	mae	mape	mase	smape	rmse	rsq
1	ARIMA(0,0,0) WITH NON-ZERO MEAN	Test	0.63	4.01	3.17	3.93	0.64	NA
2	ETS(A,N,N)	Test	0.63	4.01	3.17	3.93	0.64	NA
3	PROPHET	Test	0.39	2.46	1.94	2.51	0.46	1

Accuracy for Annual-Total-Total-European Union - 27 countries (from 2020)-Pollution, grime or other environmental problems - EU-SILC survey

.model_id	.model_desc	.type	mae	mape	mase	smape	rmse	rsq
1	ARIMA(0,0,0) WITH NON-ZERO MEAN	Test	0.70	5.07	0.50	4.86	0.91	NA
2	ETS(A,N,N)	Test	0.70	5.07	0.50	4.86	0.91	NA
3	PROPHET	Test	0.49	3.45	0.35	3.41	0.51	1

Figura 53 Métricas Predicción Pollution, grime or other environmental problems Fuente: Elaboración propia. Datos extraídos de: [12]

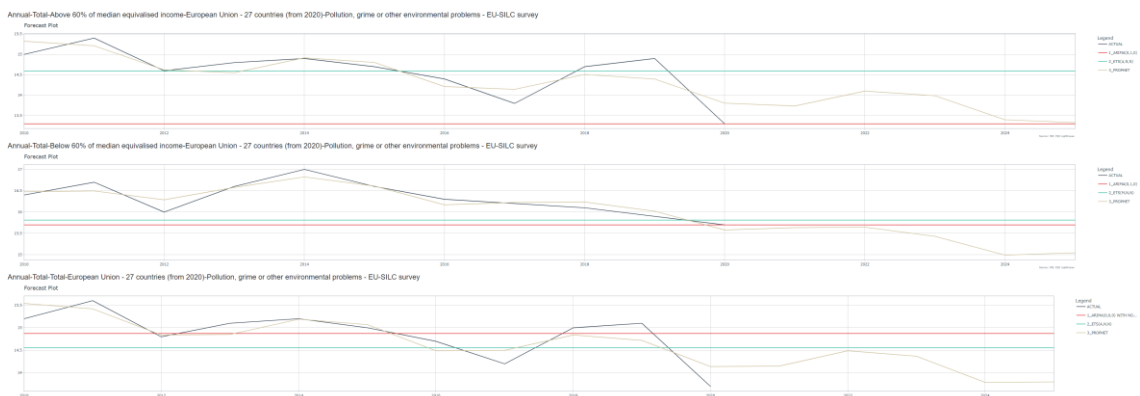


Figura 54 Gráficas Predicción Pollution, grime or other environmental problems Fuente: Elaboración propia. Datos extraídos de: [12]

3.10.4 Análisis de clustering de la Dimensión 8

El análisis de *clustering* identifica seis clústeres, utilizando un máximo de 5 valores faltantes por variable y la imputación mediante *K-Nearest Neighbors*. La inercia disminuye a medida que aumenta el número de clústeres, justificando la elección de seis clústeres con una inercia de 0.89. A continuación, se detallan los clústeres y algunos países representativos de cada uno:

- Clúster 0 (Rojo): Incluye a Italia, Letonia, Lituania, Luxemburgo, Países Bajos, Serbia, y Eslovenia. Estos países tienen niveles intermedios de problemas de contaminación.
- Clúster 1 (Azul): Malta. Caracterizado por niveles muy altos de problemas de contaminación en comparación con otros países europeos.

- Clúster 2 (Verde): Austria, Chipre, Chequia, Dinamarca, Estonia, Finlandia, Irlanda, y Suiza. Estos países presentan niveles bajos de problemas de contaminación.
- Clúster 3 (Morado): Francia, Grecia, y Turquía. Presentan niveles variados de problemas de contaminación, con algunas regiones más afectadas que otras.
- Clúster 4 (Naranja): Alemania, Bélgica, Bulgaria, Hungría, Montenegro, Macedonia del Norte, Portugal, Rumanía, y España. Estos países tienen niveles moderadamente altos de problemas de contaminación.
- Clúster 5 (Amarillo): Albania, Croacia, y Suecia. Presentan niveles altos de problemas de contaminación en comparación con otros clústeres.

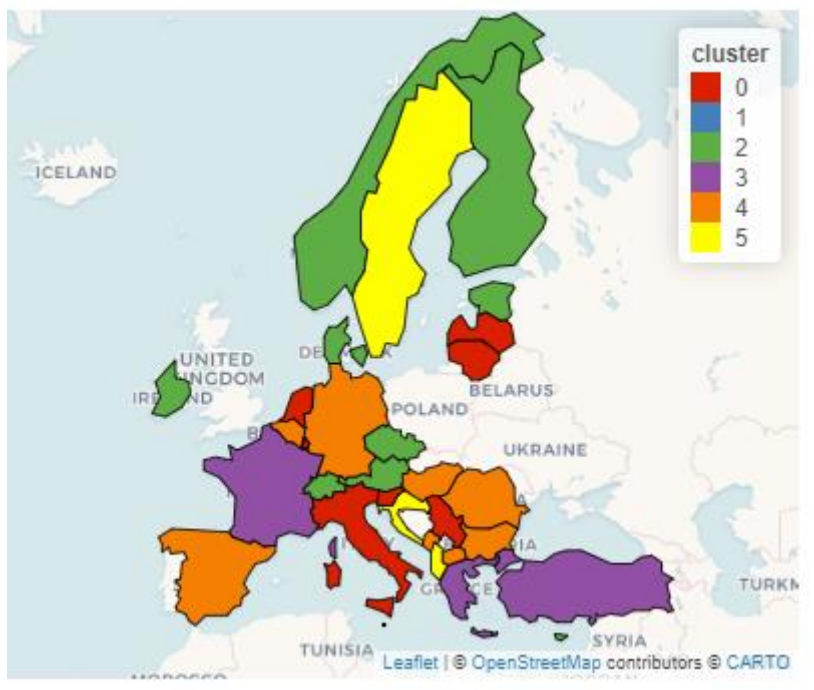


Figura 55 Clustering Pollution, grime or other environmental problems Fuente: Elaboración propia. Datos extraídos de: [12]

3.11 Análisis de la Dimensión 9: Average rating of satisfaction by domain, sex, age and educational attainment level

3.11.1 Análisis descriptivo y exploratorio de la Dimensión 9

El análisis descriptivo muestra los siguientes valores estadísticos para la satisfacción promedio: Media: 7.047, Mediana: 7.2, Desviación estándar: 0.707, Percentil 25: 6.7, Percentil 50: 7.2, Percentil 75: 7.575, Rango intercuartílico: 0.875, Coeficiente de variación: 0.100, Intervalo de confianza: 6.933 - 7.161. Estos datos indican que la satisfacción promedio se encuentra en un rango moderadamente alto, con una variabilidad relativamente baja, lo que sugiere una percepción consistente de satisfacción entre los encuestados.

Average rating of satisfaction by domain, sex, age and educational attainment level		
Statistic	Value	Units
Mean	7.04733333333333	Rating (0-10)
Median	7.2	Rating (0-10)
Standard Deviation	0.706983066224653	Rating (0-10)
25th Percentile	6.7	Rating (0-10)
50th Percentile	7.2	Rating (0-10)
75th Percentile	7.575	Rating (0-10)
Interquartile Range	0.874999999999999	Rating (0-10)
Coefficient of Variation	0.10031923179803	Rating (0-10)
Confidence Interval	6.93326811959819 - 7.16139854706848	Rating (0-10)

Figura 56 Descriptivo Average rating of satisfaction by domain, sex, age and educational attainment level Fuente: Elaboración propia. Datos extraídos de: [12]

3.11.2 Análisis de series temporales de la Dimensión 9

El análisis temporal, representado en el gráfico, muestra varias series de tiempo para diferentes dominios de satisfacción: Satisfacción laboral, Sentido de la vida, Satisfacción general, Satisfacción con la vivienda, Satisfacción con el tiempo de desplazamiento, Satisfacción con la situación financiera, Satisfacción con el entorno vital, Satisfacción con las relaciones personales, Satisfacción con el tiempo libre y las áreas verdes. A pesar de la falta de suficientes datos para un análisis predictivo detallado, las tendencias generales muestran que la mayoría de los dominios han mantenido un nivel de satisfacción relativamente estable a lo largo del tiempo, con ligeras variaciones, obteniendo mejores resultados la satisfacción en las relaciones personales en primer lugar. El análisis de la Función de Autocorrelación (ACF) del total muestra que existe una autocorrelación significativa en los datos, lo que indica que los valores de satisfacción en un año están correlacionados con los de los años anteriores. Sin embargo, la falta de datos suficientes impide realizar un análisis predictivo robusto.



Figura 57 Serie Temporal Average rating of satisfaction by domain, sex, age and educational attainment level Fuente: Elaboración propia. Datos extraídos de: [12]

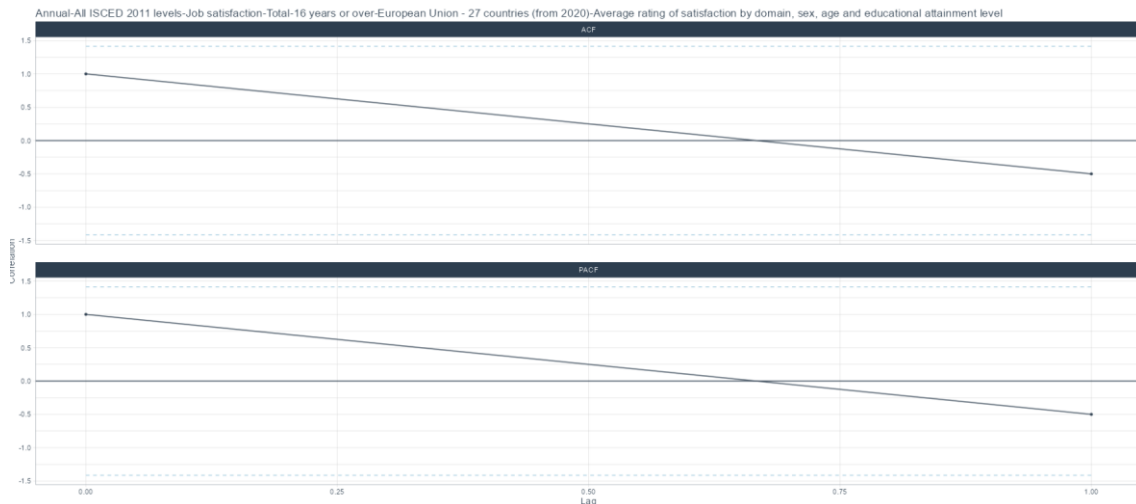


Figura 58 ACF Average rating of satisfaction by domain, sex, age and educational attainment level
Fuente: Elaboración propia. Datos extraídos de: [12]

3.11.3 Análisis predictivo de la Dimensión 9

No se dispone de suficientes datos para realizar análisis predictivo.

3.11.4 Análisis de clustering de la Dimensión 9

Para determinar el número óptimo de clústeres, se utilizaron las inercias con los siguientes valores: para 4 clústeres es 2.32, para 5 clústeres es 1.35, para 6 clústeres es 0.99, y para 7 clústeres es 0.69 determinando óptima la inercia para 6 clústeres. Se identificaron seis clústeres con la siguiente composición:

- Clúster 0 (Rojo) incluye a países como Chipre, Estonia, Francia, Italia, Letonia, Lituania, Luxemburgo, España.
- Clúster 1 (Azul) incluye a países como Bulgaria, Montenegro, Serbia.
- Clúster 2 (Verde) incluye a países como Bélgica, Finlandia, Países Bajos, Noruega, Polonia, Rumania.
- Clúster 3 (Morado) incluye a países como Croacia, Alemania, Grecia, Hungría, Letonia.
- Clúster 4 (Naranja) incluye a países como Austria, Suiza.
- Clúster 5 (Amarillo) incluye a países como Chequia, Dinamarca, Irlanda, Malta, Suecia.

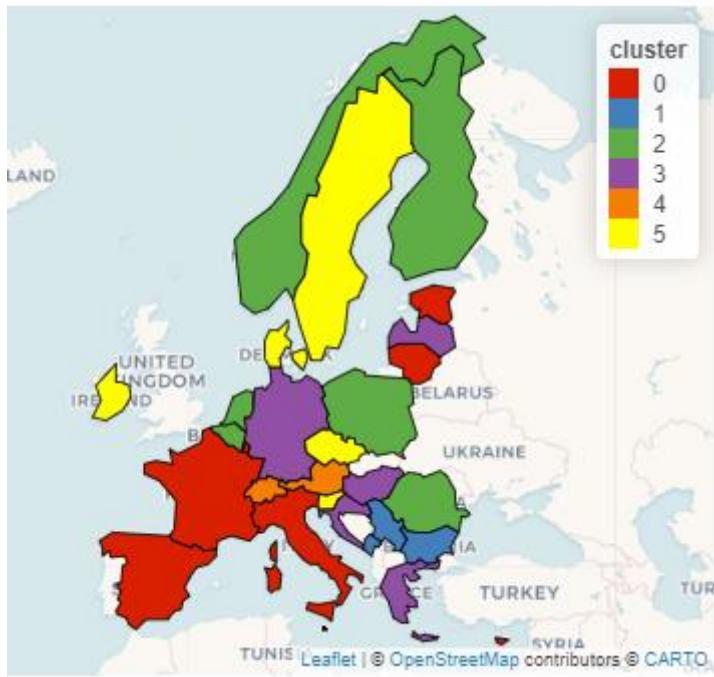


Figura 59 Clustering Average rating of satisfaction by domain, sex, age and educational attainment level Fuente: Elaboración propia. Datos extraídos de: [12]

4 Resultados y conclusiones

Resumen de resultados obtenidos en el TFG. Y conclusiones personales del estudiante sobre el trabajo realizado.

4.1 Resumen de los principales hallazgos

El análisis de cada dimensión de calidad de vida, utilizando los indicadores de Eurostat, ha proporcionado resultados detallados y significativos que permiten entender las variadas facetas del bienestar socioeconómico en la Unión Europea. A continuación, se resume cada dimensión con sus respectivos resultados y se extraen conclusiones clave.

Dimensión 1: Condiciones Materiales de Vida

Se ha analizado la distribución de ingresos, consumo y privaciones materiales, destacando la disparidad económica entre los países miembros. Los indicadores han mostrado que, aunque algunos países muestran un alto nivel de ingresos medios, otros sufren de tasas elevadas de riesgo de pobreza y privación material. Este contraste pone de relieve la necesidad de políticas dirigidas a reducir la desigualdad y mejorar las condiciones materiales en toda la Unión.

Dimensión 2: Empleo

El análisis ha revelado diferencias significativas en las tasas de empleo y las condiciones laborales entre los países. Mientras que algunos países exhiben altas tasas de empleo y condiciones laborales favorables, otros enfrentan desafíos como el alto desempleo y la inseguridad laboral. Estos resultados sugieren la importancia de fortalecer las políticas de empleo y capacitación para fomentar un mercado laboral más inclusivo y resiliente.

Dimensión 3: Educación

Se ha observado que el acceso a la educación y la calidad de la misma varían considerablemente. Los países con sistemas educativos fuertes muestran mejores resultados en términos de desempeño académico y oportunidades laborales. Este hallazgo subraya la necesidad de invertir en educación para garantizar que todos los ciudadanos de la UE tengan las mismas oportunidades de desarrollo personal y profesional.

Dimensión 4: Salud

El análisis ha indicado que la salud general, medida a través de expectativas de vida saludable y acceso a servicios médicos, es mejor en algunos países que en otros. Los desafíos incluyen disparidades en la atención médica y en los resultados de salud, lo que requiere un enfoque coordinado para mejorar los sistemas de salud en toda Europa.

Dimensión 5: Ocio e Interacciones Sociales

Los resultados han demostrado que el ocio y las interacciones sociales juegan un papel crucial en la calidad de vida. Países con mayores oportunidades de ocio y fuertes redes sociales tienden a tener ciudadanos más satisfechos y saludables. Promover la cohesión social y el acceso a actividades de ocio puede ser una estrategia efectiva para mejorar el bienestar general.

Conclusiones Relevantes

El análisis multidimensional ha permitido identificar áreas clave donde los países de la Unión Europea muestran tanto fortalezas como debilidades. Los resultados sugieren que, mientras algunos países avanzan bien en aspectos

como empleo y salud, otros se quedan atrás, especialmente en condiciones materiales de vida y acceso equitativo a servicios esenciales como la educación y la salud. Las políticas futuras deberán enfocarse en cerrar estas brechas, promoviendo un desarrollo más equitativo y sostenible que beneficie a todos los ciudadanos de la UE. La inversión en educación y salud, junto con la mejora de las condiciones laborales y la reducción de la desigualdad económica, son esenciales para alcanzar los Objetivos de Desarrollo Sostenible y asegurar una alta calidad de vida para todos los europeos.

4.2 Limitaciones del estudio

El estudio, a pesar de su alcance y profundidad, enfrenta varias limitaciones que deben ser consideradas al interpretar los resultados. Una de las principales restricciones ha sido la dependencia de los datos disponibles a través de Eurostat, que, aunque extensos y detallados, pueden presentar desfases temporales en la actualización y diferencias en la metodología de recolección entre países. Esto podría introducir ciertos sesgos en el análisis, especialmente al comparar indicadores entre países con sistemas estadísticos de distinto rigor o capacidad.

Además, el estudio no ha explorado el uso de diversos métodos de *clustering* más allá de los implementados, limitándose a técnicas específicas que, aunque efectivas, podrían no capturar todas las posibles estructuras de agrupación o patrones subyacentes en los datos. La elección de variables, aunque exhaustiva, siempre está sujeta a la relevancia y precisión de lo que los indicadores pueden realmente medir, y algunas dimensiones importantes de la calidad de vida podrían haberse subrepresentado o interpretado de manera superficial.

4.3 Recomendaciones para futuras investigaciones

Dadas las limitaciones mencionadas, se recomienda que futuras investigaciones amplíen el rango de métodos de análisis, especialmente explorando diferentes técnicas de *clustering* que puedan ofrecer nuevas perspectivas sobre la agrupación de países según sus características socioeconómicas. Además, sería beneficioso incorporar datos de fuentes adicionales que complementen y corroboren los obtenidos de Eurostat, para mitigar el riesgo de sesgos debido a las limitaciones de una sola fuente.

Se sugiere también realizar estudios longitudinales que permitan seguir la evolución de los indicadores de calidad de vida a lo largo del tiempo, facilitando una mejor comprensión de las tendencias y la efectividad de las políticas implementadas en respuesta a los hallazgos de estudios como este. Por último, se recomienda llevar a cabo análisis comparativos más detallados que puedan identificar con precisión las causas de las disparidades observadas entre países y regiones, lo que a su vez puede orientar de manera más efectiva las decisiones políticas y económicas a nivel de la Unión Europea.

5 Análisis de Impacto

Este capítulo se centra en evaluar el impacto potencial de los resultados obtenidos durante la realización de este Trabajo Fin de Grado en diversos contextos.

Para empezar, este estudio analiza detalladamente las dimensiones de calidad de vida en la Unión Europea utilizando datos de Eurostat, proporcionando una evaluación profunda de cómo diversos factores socioeconómicos afectan el bienestar general. Estos análisis y sus resultados tienen una relación directa y beneficiosa con los Objetivos de Desarrollo Sostenible (ODS) propuestos por las Naciones Unidas, debido a que cada dimensión explorada en este estudio se alinea con objetivos específicos diseñados para promover un desarrollo global sostenible y equitativo.

Por otro lado, estos resultados también tienen el potencial de contribuir a informar políticas públicas y programas sociales que buscan mejorar la calidad de vida en la Unión Europea. Al entender las disparidades y necesidades específicas de diferentes grupos, los responsables políticos pueden diseñar intervenciones más efectivas que apunten a reducir la desigualdad y promover una mayor cohesión social y, así, favorecer la calidad de vida de los ciudadanos.

Desde un punto de vista personal, los resultados del estudio han tenido un impacto directo y positivo en términos de desarrollo académico y profesional, proporcionando una base sólida en análisis de datos y comprensión de dinámicas socioeconómicas complejas. Llevando todo esto a construir y ampliar las habilidades analíticas y críticas, esenciales para futuras investigaciones o aplicaciones prácticas.


Asimismo, cómo apoya la línea de investigación del grupo de Análisis de decisiones y Estadística de la ETSIINF, UPM., este análisis contribuye a una mejor comprensión de los factores económicos, sociales y culturales que influyen en la calidad de vida y el desarrollo sostenible. Al proporcionar una visión más clara de las interrelaciones entre diversos indicadores económicos y su impacto en la vida de los ciudadanos, se pueden fomentar políticas más robustas y fundamentadas que apunten al crecimiento económico inclusivo.

6 Bibliografía

- [1] Decision Analysis and Statistics Group Universidad Politécnica de Madrid, «Economics @Intelligence,» 2024. [En línea]. Available: <http://vps159.cesvima.upm.es/>.
- [2] European Commission, «Quality of life indicators,» 2023. [En línea]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Quality_of_life_indicators.
- [3] European Commission, «Quality of life indicators - overall experience of life,» 2023. [En línea]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Quality_of_life_indicators_-_overall_experience_of_life.
- [4] J. E. Stiglitz, A. Sen y J.-P. Fitoussi, «Report by the Commission on the Measurement of Economic Performance and Social,» 2009.
- [5] European Commission, «Quality of life indicators - measuring quality of life,» 2023. [En línea]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Quality_of_life_indicators_-_measuring_quality_of_life.
- [6] Eurostat, «Quality of Life,» 2024. [En línea]. Available: <https://ec.europa.eu/eurostat/web/quality-of-life/information-data>.
- [7] Ministerio de Derechos Sociales, Consumo y Agenda 2030, «Conoce la Agenda 2030,» Ministerio de Derechos Sociales, Consumo y Agenda 2030, 2024. [En línea]. Available: https://www.mdsocialesa2030.gob.es/agenda2030/conoce_la_agenda.htm.
- [8] Naciones Unidas, «Objetivos de Desarrollo Sostenible,» 2015. [En línea]. Available: <https://www.un.org/sustainabledevelopment/es/2015/09/la-asamblea-general-adopta-la-agenda-2030-para-el-desarrollo-sostenible/>.
- [9] Eurostat, «Sustainable development in the European Union,» Luxembourg: Publications Office of the European Union, Luxembourg, 2023.
- [1] Sustainable Development Solutions Network, «Europe Sustainable Development Report 2023/24,» Dublin University Press, Dublin, 2024.
- [1] Eurostat, «API Statistics - data query,» 2024. [En línea]. Available: <https://wikis.ec.europa.eu/display/EUROSTATHELP/API+Statistics+-+data+query>.
- [1] Eurostat, «Eurostat Data Browser,» 2024. [En línea]. Available: https://ec.europa.eu/eurostat/databrowser/explore/all/all_themes?lang=en&display=list&sort=category.
- [1] P. Newbold, Statistics for Business and Economics, 2012.
3]
- [1] D. Moore, Introduction to the Practice of Statistics, 2017.
4]

- [1 Snedecor, Statistical Methods, 1989.
5]
- [1 Tibshirani, The Elements of Statistical Learning, 2009.
6]
- [1 «Documentation for package 'base' version 4.5.0,» [En línea]. Available:
7] <https://stat.ethz.ch/R-manual/R-devel/library/base/html/00Index.html> .
- [1 Y. Xie, «DT: A Wrapper of the JavaScript Library 'DataTables',» 2024. [En
8] línea]. Available: <https://cran.r-project.org/web/packages/DT/index.html> .
- [1 J. Cheng, «htmltools: Tools for HTML,» 2024, [En línea]. Available:
9] <https://cran.r-project.org/web/packages/htmltools/index.html> .
- [2 G. Box, Time Series Analysis: Forecasting and Control, 2008.
0]
- [2 R. J. Hyndman, Forecasting: principles and practice. OTexts, 2018.
1]
- [2 S. J. Taylor, Forecasting at scale. The American Statistician, 2018.
2]
- [2 Y. Li, A Clustering Method Based on K-Means Algorithm, 2012.
3]
- [2 Agencia Española de Protección de Datos, «Aproximación a los espacios de
4] datos desde la perspectiva del RGPD,» 2023. [En línea]. Available:
<https://www.aepd.es/guias/aproximacion-espacios-datos-rgpd.pdf>.
- [2 Eurostat, «Código de Buenas Práctias de las Estadísticas Europeas,»
5] Luxemburgo: Oficina de Publicaciones de la Unión, Luxemburgo, 2017.
- [2 Eurostat, «Data validation,» 2024. [En línea]. Available:
6] <https://ec.europa.eu/eurostat/web/main/data/data-validation>.
- [2 ROpenGov, «eurostat R package,» 2023. [En línea]. Available:
7] <https://ropengov.github.io/eurostat/>.
- [2 C. Chatfield, The Analysis of Time Series: An Introduction, 2004.
8]
- [2 R. B. Cleveland, STL: A seasonal-trend decomposition procedure based on
9] loess. Journal of Official Statistics, 1990.

Este documento esta firmado por



Firmante	CN=tfgm.fi.upm.es, OU=CCFI, O=ETS Ingenieros Informaticos - UPM, C=ES
Fecha/Hora	Sun Jun 30 23:43:16 CEST 2024
Emisor del Certificado	EMAILADDRESS=camanager@etsiinf.upm.es, CN=CA ETS Ingenieros Informaticos, O=ETS Ingenieros Informaticos - UPM, C=ES
Numero de Serie	561
Metodo	urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature)