

**Universidad Internacional de La Rioja**

**Escuela Superior de Ingeniería y  
Tecnología**

**Máster Universitario en Análisis y Visualización  
de Datos Masivos**

# Metodología para la visualización interactiva de datos de la Unión Europea

**Trabajo Fin de Máster**

**Tipo de trabajo:** Desarrollo de una metodología

**Presentado por:** Barroso García, Roberto

**Director/a:** Lancheros Cuesta, Diana

# Índice de contenidos

1	Introducción .....	7
1.1	Motivación.....	7
1.2	Planteamiento del trabajo.....	8
1.3	Estructura de la memoria .....	9
2	Contexto y estado del arte .....	10
3	Objetivos y metodología de trabajo .....	15
3.1	Objetivo general .....	16
3.2	Objetivos específicos .....	16
3.3	Metodología del trabajo.....	16
4	Desarrollo específico de la contribución .....	18
4.1	Identificación de requisitos .....	19
4.2	Descripción de la metodología .....	21
4.2.1	Bloque I: Proceso de ETL (extracción, transformación y carga de datos).....	21
4.2.1.1	Fuentes de los datos.....	22
4.2.1.2	Extracción y tratamiento de los datos.....	22
4.2.2	Bloque II: Desarrollo de las visualizaciones interactivas.....	35
4.2.2.1	Diseño, construcción y resultados .....	41
4.2.3	Bloque III: Despliegue de la aplicación web .....	75
4.3	Evaluación .....	77
5	Conclusiones y trabajo futuro .....	80
5.1	Conclusiones .....	80
5.2	Líneas de trabajo futuro .....	81
6	Bibliografía .....	82
7	Anexo.....	86
7.1	Script ETL – España: .....	86
7.2	Script ETL – Unión Europea.....	108

# Índice de figuras

<b>Figura 1.</b> Diagrama de flujo de tareas a alto nivel.....	9
<b>Figura 2.</b> Ejemplo de visualización de la esperanza de vida en Our World in Data .....	13
<b>Figura 3.</b> Ejemplo de la visualización de la esperanza de vida en Datos Macro .....	13
<b>Figura 4.</b> Ejemplo de visualización de la esperanza de vida en el INE.....	14
<b>Figura 5.</b> Ejemplo de visualización en El Orden Mundial .....	15
<b>Figura 6.</b> Diagrama de la metodología para la elaboración del estudio.....	16
<b>Figura 7.</b> Distribución porcentual de los grupos de edad que leyeron prensa en 2020 .....	19
<b>Figura 8.</b> Grupos de edad que han usado internet de forma frecuente en 2021 (en %) .....	20
<b>Figura 9.</b> Bloques en el desarrollo .....	21
<b>Figura 10.</b> Creación script en Jupyter Notebook .....	23
<b>Figura 11.</b> Carpetas creadas por los scripts.....	25
<b>Figura 12.</b> Importar datos en Tableau.....	36
<b>Figura 13.</b> Ejemplo de hoja en Tableau para el tratamiento de datos .....	37
<b>Figura 14.</b> Parámetros de una hoja de visualizaciones en Tableau .....	38
<b>Figura 15.</b> Ejemplo de construcción de una visualización simple.....	38
<b>Figura 16.</b> Formatear el panel en Tableau .....	39
<b>Figura 17.</b> Editar los ejes en Tableau .....	40
<b>Figura 18.</b> Opción para extraer los datos en Tableau .....	40
<b>Figura 19.</b> Opción para publicar en Tableau Public .....	41
<b>Figura 20.</b> Derechos humanos en la Unión Europea .....	42
<b>Figura 21.</b> Tendencia de los derechos humanos en la UE.....	43
<b>Figura 22.</b> Histórico de los derechos humanos en la UE.....	44
<b>Figura 23.</b> Evolución de los derechos humanos en España.....	45
<b>Figura 24.</b> Evolución del producto interior bruto en España.....	46
<b>Figura 25.</b> Evolución de la deuda total en España .....	46
<b>Figura 26.</b> Evolución del déficit total en España .....	47

<b>Figura 27.</b> Evolución del índice de precios al consumo en España.....	48
<b>Figura 28.</b> Balanza del comercio internacional en España.....	48
<b>Figura 29.</b> Evolución de los ingresos fiscales promedio en la Unión Europea.....	49
<b>Figura 30.</b> Evolución de los ingresos fiscales en España.....	50
<b>Figura 31.</b> Distribución de los impuestos por país de la UE .....	51
<b>Figura 32.</b> Evolución del gasto público promedio en la Unión Europea.....	51
<b>Figura 33.</b> Distribución del gasto público promedio en la Unión Europea .....	52
<b>Figura 34.</b> Gasto en Sanidad en España .....	53
<b>Figura 35.</b> Gasto en educación en España .....	53
<b>Figura 36.</b> Gasto en defensa en España .....	54
<b>Figura 37.</b> Índice de desigualdad GINI en la Unión Europea .....	55
<b>Figura 38.</b> Índice de desigualdad GINI en España.....	55
<b>Figura 39.</b> Tasa de desempleo en la Unión Europea .....	56
<b>Figura 40.</b> Tasa de desempleo en España .....	57
<b>Figura 41.</b> Tasa de desempleo en los países de la UE en el año 2020.....	57
<b>Figura 42.</b> Evolución del salario mínimo interprofesional en España .....	58
<b>Figura 43.</b> Esfuerzo vs Productividad en la Unión Europea .....	59
<b>Figura 44.</b> Puntuaciones en la relación esfuerzo y productividad en la UE .....	59
<b>Figura 45.</b> Esfuerzo vs Productividad en España .....	60
<b>Figura 46.</b> Percepción de la corrupción en la Unión Europea .....	61
<b>Figura 47.</b> Evolución del índice de corrupción en España.....	61
<b>Figura 48.</b> Emisiones de CO2 en la Unión Europea.....	62
<b>Figura 49.</b> Emisiones de CO2 en España.....	63
<b>Figura 50.</b> Consumo energético en España.....	63
<b>Figura 51.</b> Generación energética y uso de renovables en España .....	64
<b>Figura 52.</b> Territorios agrícolas en la Unión Europea.....	65
<b>Figura 53.</b> Territorios agrícolas en España .....	65
<b>Figura 54.</b> Número de muertes asociadas a factores de riesgo en la UE.....	66

<b>Figura 55.</b> Número de muertes asociadas a factores de riesgo en España .....	67
<b>Figura 56.</b> Suicidios por cada 100.000 habitantes en España .....	67
<b>Figura 57.</b> Componentes de la alimentación en la Unión Europea.....	68
<b>Figura 58.</b> Componentes de la alimentación en España .....	69
<b>Figura 59.</b> Problemas en la nutrición de los niños de la UE .....	70
<b>Figura 60.</b> Problemas en la nutrición de los niños de España.....	70
<b>Figura 61.</b> Población en España.....	71
<b>Figura 62.</b> Esperanza de vida en España .....	72
<b>Figura 63.</b> Balanza demográfica en España .....	73
<b>Figura 64.</b> Muertes a causa del terrorismo en la Unión Europea .....	74
<b>Figura 65.</b> Homicidios por cada 100.000 habitantes en España .....	74
<b>Figura 66.</b> Ejemplo básico para presentar las visualizaciones en la web .....	76
<b>Figura 67.</b> Cuestionario .....	78
<b>Figura 68.</b> Resultados del cuestionario de evaluación .....	79

## Resumen

El incesante incremento en el número de datos que se procesan actualmente de forma global hace obligatoria la utilización de metodologías de visualización de datos, que permitan al ser humano comprender una ingente cantidad de información que de otra forma sería imposible de manejar.

En el presente documento se propone el desarrollo de una nueva metodología para la visualización interactiva de grandes cantidades de datos referentes a un determinado país de la Unión Europea, mostrando diferentes materias relacionadas con la economía, sociedad, salud y bienestar del país en cuestión, en este caso España. Para ello se lleva a cabo un proceso de extracción, transformación y carga (ETL) por el que se recopilan grandes cantidades de datos mediante métodos de *web scraping*, y que son tratados y cargados haciendo uso de ciertas librerías de *Python*. Seguidamente se construyen las visualizaciones en la herramienta *Tableau*, que son actualizadas a tiempo real y mostradas en una aplicación web interactiva construida en *Brackets* con los lenguajes *HTML*, *CSS* y *Javascript*, y desplegada en *GithubPages*. De esta forma se pretende acercar al usuario una gran cantidad de información de forma más intuitiva y accesible.

**Palabras clave:** metodología, visualización, ETL, web scraping, Python, Tableau

## Abstract

The relentless increase in the amount of data currently being processed globally makes the use of interactive data visualisation methodologies mandatory, allowing humans to understand a vast amount of information that would otherwise be impossible to handle.

This paper proposes the development of a new methodology for the interactive visualisation of large amounts of data referring to a specific European Union country, showing different subjects related to the economy, society, health and welfare of the country in question. For this purpose, an extraction, transformation and loading (ETL) process is carried out, whereby large amounts of data are collected using web scraping methods, which are processed and loaded using certain Python libraries. The visualisations are then built in the Tableau tool, updated in real-time and displayed in an interactive web application built in Brackets with HTML, CSS and Javascript languages, and deployed in GitHub pages. In this way, the aim is to bring a large amount of information to the user in a more intuitive and accessible way.

**Keywords:** methodology, visualisation, ETL, web scraping, Python, Tableau

# 1 Introducción

## 1.1 Motivación

Uno de los motivos principales para la realización del trabajo es la singularidad del tema que se propone, y la forma de exponerlo, es decir, la forma de visualizar los datos que se ha escogido para mostrar la información dentro de un contexto. Existen varios ejemplos de aplicaciones web que exponen información sobre países, pero cometan algunos fallos, que son precisamente los que se proponen solucionar con esta metodología.

Estas webs recogen un tema concreto, por ejemplo, la cantidad de cerveza que se consume en la Unión Europea, y en un mapa coroplético simple muestran esos datos concretos con cifras y colores, sin aportar nada más con lo que poder interactuar (como ha variado en el tiempo, en qué año se consumió más o menos cantidad, comparado con los demás países, cuál es el que más consume por persona, etc.).

Un problema que se ha observado es la dificultad de encontrar toda la información sobre un país en concreto reunido en un solo sitio web, que muestre los datos en su contexto de forma sencilla, con comparaciones, históricas, dividida en categorías y sobre todo con visualizaciones que atrapen al lector por su estética, claridad y dinamismo, que le permitan interactuar con los datos.

Existen sitios web que reúnen los datos de un país y permiten hacer comparaciones o análisis, pero las visualizaciones son pobres, confusas, y a menudo incompletas, muchas veces no permiten realizar comparaciones intuitivas que de un vistazo permitan al usuario comprender lo que está viendo.

Se debe agregar que a menudo estos datos no se podrían considerar lo que denominamos *Big Data*, pues si se trata de una gran cantidad de datos, pero no los suficientes, suelen pertenecer a categorías específicas, o se refieren a años concretos. Esto puede ser causado por la necesidad actual de utilizar las visualizaciones para apoyar una noticia en concreto, y no para contar una “historia” como es el caso de este trabajo. Un ejemplo sería un artículo sobre la tasa de paro en cierto país donde mediante un extenso texto aportan toda la información y la visualización está ahí para apoyar la noticia, y mostrar por ejemplo solo su evolución en el tiempo, pero en este caso se pretende ir más allá, y en vez de mostrar un gráfico simple, se expone un dashboard interactivo con información de otros países, variaciones, rankings, etc. En definitiva, se pretende añadir valor a la visualización en decremento de largas explicaciones en texto.

Es relevante esta nueva metodología de visualización por dos razones importantes, en la actualidad es conocido que existen infinitas fuentes de información, si un usuario quiere enterarse sobre una determinada noticia, por ejemplo, la tasa de desempleo en España, solo tiene que realizar una búsqueda en internet y le aparecerán 10 o 20 noticias relacionadas, todas con un título llamativo para que pulse sobre el enlace, y finalmente en cada una de las noticias verá cierta información que se repite en todas las demás, la tasa de paro o quien tiene la culpa (dependiendo del medio serán unos u otros), todo ello resumido en numerosos párrafos de texto donde muchas veces se hace complicado encontrar lo que se quiere. Gorka R. Pérez (2022) escribe un artículo sobre el desempleo en España que sugiere unos resultados positivos en la tasa de desempleo del último mes, y sin embargo en esa misma fecha, Marta Ruiz (2022) publicaba un artículo con la misma noticia, pero en este caso las conclusiones extraídas eran más negativas, con los mismos datos.

Este sería el primer motivo por el que es relevante esta metodología, en este caso el usuario de un vistazo y en poco tiempo podría ver un histórico con la variación de la tasa de desempleo a lo largo de los años, cual ha sido el promedio de esta comparado con el promedio europeo, una comparación llamativa con los demás países, donde hay más, donde hay menos, etc... El usuario puede extraer sus propias conclusiones.

Otro motivo por el que es relevante este desarrollo es por la dificultad de encontrar algo similar, es decir, es complicado encontrar un sitio web o bibliografía que presente la información que se pretende representar con este trabajo de la misma forma y completitud, es decir, con estos datos y desde una perspectiva en este caso, europea.

## 1.2 Planteamiento del trabajo

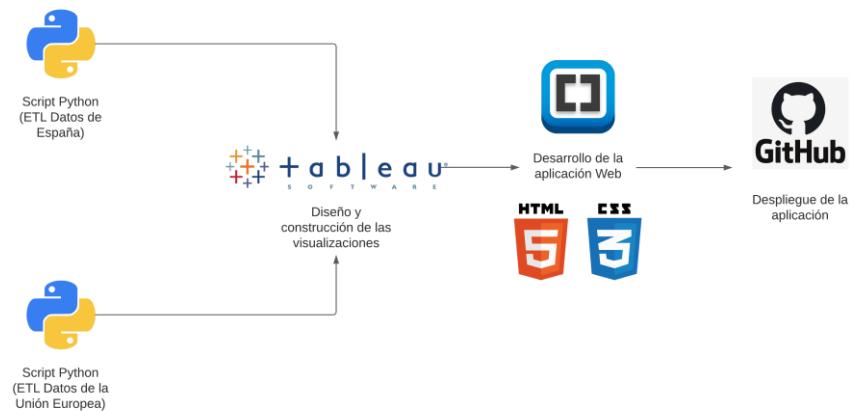
Se propone realizar una aplicación web que presente la información mediante visualizaciones interactivas en un formato más claro, motivados en solucionar los problemas comentados anteriormente, esto es, cada tema separado por categoría y exponiendo información de la Unión Europea y el país elegido de una forma más relacionada entre sí.

Mediante dos scripts de Python diferentes, uno para los datos de la UE y otro para los datos de España, se extraen y transforman de tal forma que se puedan asociar o vincular (una variable común, por ejemplo, la fecha o el país), y de esta manera presentarlos de forma homogénea en las visualizaciones interactivas que deben ser claras y sencillas para el lector.

Cada visualización se expone en la web con una breve descripción, corta y objetiva, que haga fácil al usuario su correcta lectura. De esta forma se consigue que una gran cantidad de

información dispersa en internet se reúna y se muestre de forma fácilmente interpretable, que permita de un vistazo percibir toda la información disponible dentro de un contexto. En la figura 1 se muestra un diagrama del flujo de tareas a alto nivel, de cómo se ha planteado el desarrollo del trabajo.

**Figura 1.** Diagrama de flujo de tareas a alto nivel



Fuente: Elaborada por el autor

Este planteamiento se enmarca en un único objetivo a nivel general, el de favorecer el aprendizaje y aumentar el conocimiento sobre un determinado país, en un contexto a nivel europeo.

### 1.3 Estructura de la memoria

Para entender el desarrollo del trabajo realizado se detallan los siguientes capítulos que definen la estructura del documento.

El primer capítulo comienza con la definición de un contexto donde se exponen los principales trabajos y autores relacionados con metodologías similares a la propuesta, mostrando así el conocimiento ya existente en este ámbito de estudio, junto con los problemas relacionados encontrados que se pretenden solucionar.

Seguidamente, en el segundo capítulo se exponen los objetivos y la metodología de trabajo, esto es, lo que se pretende conseguir, las soluciones a implementar, y a qué conclusiones se quiere llegar con la realización del trabajo. De este modo, se definen los objetivos generales del estudio y los objetivos específicos para llevar a cabo cada tarea del desarrollo. Por último, en otro apartado de este mismo capítulo, se muestra paso a paso la metodología seguida para la realización de estos objetivos específicos.

A continuación, en un total de tres capítulos, se describe el desarrollo completo del estudio realizado. El tercer capítulo presenta la identificación de requisitos donde se realiza una breve introducción a los contenidos implicados, los usuarios objetivo, la materia a enseñar y el contexto de uso habitual. Un cuarto capítulo más extenso con la descripción de la metodología, es decir, un manual detallado de como poder replicarla en distintos entornos, todo ello acompañado de los correspondientes diagramas y figuras que muestran cómo aplicar la metodología, esto es, el proceso de extracción y tratamiento de los datos, el proceso de construcción de las visualizaciones que ilustran los datos, y el proceso realizado para la construcción de la aplicación web donde se muestran los resultados obtenidos. El quinto capítulo y último de esta sección, trata sobre la evaluación que valida la metodología propuesta que soluciona los problemas encontrados, para ello se entregarán una serie de cuestionarios a los usuarios objetivo del estudio, para así con sus respuestas extraer las conclusiones pertinentes que aseguren el cumplimiento de los objetivos propuestos para la solución de los problemas en la metodología.

Finalmente, en el sexto capítulo se expone un breve resumen de las contribuciones del trabajo realizado, el alcance y las conclusiones extraídas de su realización, de igual forma se definen posibles líneas futuras que pueden aportar valor añadido a la metodología.

## 2 Contexto y estado del arte

Hoy en día, las grandes cantidades de datos que se manejan globalmente en todas las áreas han provocado que la forma de tratar y mostrar los datos para convertirlos en información y posteriormente en conocimiento, hayan adquirido un gran protagonismo.

Uno de los autores más relevantes sobre el tema de la visualización de datos es Alberto Cairo, autor de algunos de los libros más importantes en este ámbito, como son 'El Arte Funcional' (Cairo, 2011) donde expone cuales son las buenas prácticas para una correcta representación de la información, 'Nerd Journalism' (Cairo, 2017) donde realiza una visión retrospectiva de las dos últimas décadas y como ha cambiado todo el mundo de la visualización e infografías, desde diferentes puntos de vista y roles. Y, por último, 'How Charts Lie' (Cairo, 2019) donde expone como se pueden crear visualizaciones que aun conteniendo datos correctos pueden ser mostradas de determinada forma que conduzca a una extracción errónea de conclusiones.

En su trabajo 'Interactive Visualisations and Statistical Literacy', Jim Ridgway y Sinclair Sutherland (2017) indican que están surgiendo nuevas formas de relacionarse con los datos, y especialmente a través de las visualizaciones interactivas. Describen algunas de las habilidades necesarias para trabajar eficazmente con estas nuevas herramientas de

visualización. Afirman que las visualizaciones interactivas de datos tendrán un efecto tan profundo en lo que ellos llaman “la alfabetización estadística”, como la introducción de paquetes estadísticos en las estadísticas en las ciencias sociales en la década de 1960.

La idea principal del creciente protagonismo de la visualización de datos viene reforzada por el estudio llevado a cabo por José Luis Valero Sancho (2014), donde muestra la capacidad comunicativa de estas visualizaciones mediante comparativas y cuestionarios.

Al igual que el autor Carlos Olmeda-Gómez (2014) en su trabajo donde también hace énfasis en la importancia de los diferentes métodos de visualización de datos, dependiendo del ámbito y área relacionada, dando diferentes consejos y trucos para una representación adecuada.

En el ámbito del periodismo, Ana María Rodriguez Dominguez y Victor Federico Herrero Solana (2015) afirman en su estudio que, en ciertos periódicos españoles, se utilizan estos métodos de visualización, que sin embargo no han llegado a ese grado más avanzado de interactividad.

En el ámbito empresarial, el tratamiento y visualización de los datos ha pasado a tener una gran importancia sobre todo en la toma de decisiones, Fernando Jefferson Naranjo Solís (2020) lleva a cabo un estudio sobre una nueva metodología que indica como proceder en la construcción de estos elementos visuales que permitan aportar información de forma eficiente en esta área de conocimiento.

Existen algunos estudios como los llevados a cabo por Bart Rienties et al. (2018) donde se pone énfasis en el ámbito de la enseñanza y el protagonismo de las nuevas tecnologías que permiten la creación de visualizaciones útiles en esta área. Para ello en su estudio comprueban la importancia de apoyarse en este nuevo método para estimular a los alumnos de manera efectiva. Con la creciente disponibilidad de datos de análisis de aprendizaje, los profesores podrían utilizar paneles de análisis de los datos y visualización para ayudar a los alumnos con diferentes necesidades de aprendizaje. Sin embargo, la implementación de visualizaciones de análisis de aprendizaje por parte de los docentes también requiere de la aceptación de estos. Por eso en este estudio exploran la preparación de los profesores para el uso de estas visualizaciones. Los hallazgos indicaron que los participantes apreciaron el enfoque interactivo y práctico, pero al mismo tiempo se mostraron escépticos sobre la facilidad de uso percibida de las herramientas de análisis de aprendizaje que se les ofrecieron.

Otro ejemplo más actual que nos indica el creciente protagonismo que está adquiriendo la visualización de datos nos la muestra Roberto Therón Sánchez (2021) en su trabajo donde lleva a cabo una retrospectiva acerca de la importancia de este método para aportar información y como sobre todo en la última década ha cobrado gran importancia. Todo ello

como resultado del avance tecnológico y la capacidad de consumo de lo que se conoce como “Big Data”. Lleva a cabo un análisis de todas las disciplinas y roles implicados en la visualización de datos, y sus influencias tanto artísticas como científicas. Por último, aporta una reflexión sobre el futuro de la visualización de datos.

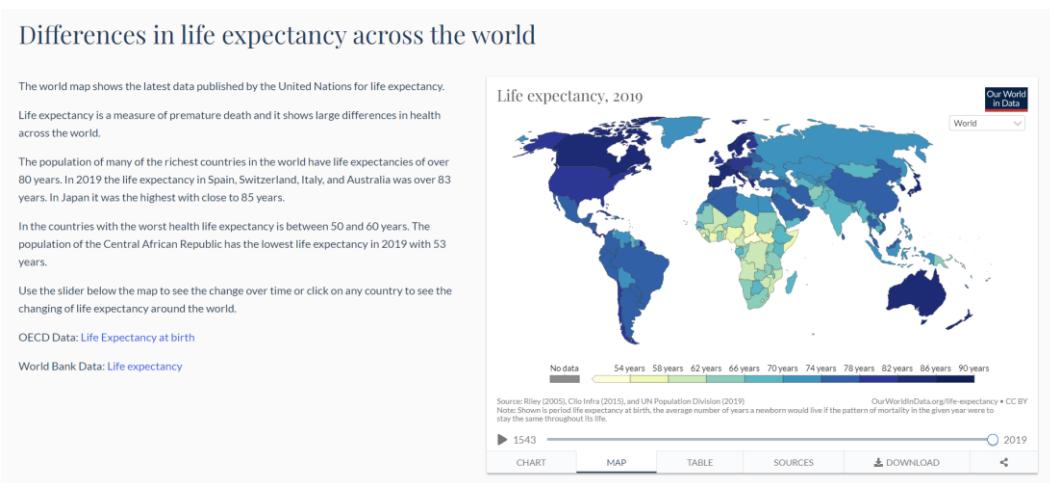
Enfocado a los ‘dashboards’ o paneles de información como método de visualización de datos existe un trabajo realizado con anterioridad por parte de Andrea Vázquez Ingelmo (2018) en el que estos paneles pueden ser personalizados por los usuarios cumpliendo las necesidades de cada uno a través de la parametrización de características. Se trata de un estudio interesante que puede aportar valor en un futuro para esta área.

En cuanto a metodologías de visualización, el libro de Isabel Meirelles (2013) expone los principios y metodologías de diseño, cognición, percepción e interacción humano-ordenador que se combinan con el conocimiento personal y las experiencias creativas del autor. Examina críticamente otras soluciones de diseño, actuales e históricas, para obtener una mayor comprensión de cómo resolver problemas específicos. Muestra una serie de casos de estudio de visualización actuales y las analiza por sus principios y métodos de diseño, aportando valiosas herramientas críticas y analíticas para mejorar en el proceso de diseño de las visualizaciones, indicando así las mejores prácticas.

Franklin Hernández Castro (2021) también expone otra metodología para el diseño de visualizaciones de datos y la toma de decisiones, en su trabajo indica la importancia de conocer los objetivos, y cumplir con la completitud y coherencia de los datos.

Algunas metodologías de visualización de datos que se pueden encontrar online son:

Una de las páginas web más conocidas actualmente en la visualización de datos es Our World in Data (<https://ourworldindata.org/>). Este sitio web al completo reúne todo tipo de artículos acerca de distintos temas muy variados sobre los países a nivel mundial, y los presenta mediante una visualización dinámica, junto con una descripción acerca de lo que se está mostrando. En la figura 2 podemos ver un ejemplo de visualización en la web mencionada.

**Figura 2.** Ejemplo de visualización de la esperanza de vida en Our World in Data

Nota. Muestra la esperanza de vida en distintos países mediante un gráfico coroplético junto con un breve artículo informativo. Tomado de *Life Expectancy* [Fotografía], de Roser, M. et al., 2019, OurWorldInData (<https://ourworldindata.org/life-expectancy>). CC BY 4.0

Un sitio web similar al anterior es Datos Macro (<https://datosmacro.expansion.com>) donde se reúnen los datos sobre un determinado país y permite hacer comparaciones con otros países, el inconveniente aquí es que se trata de una web con visualizaciones pobres y algo confusas, no aportan un contexto o perspectiva relevante, los datos en ocasiones se presentan de forma ambigua, se puede observar un ejemplo en la figura 3.

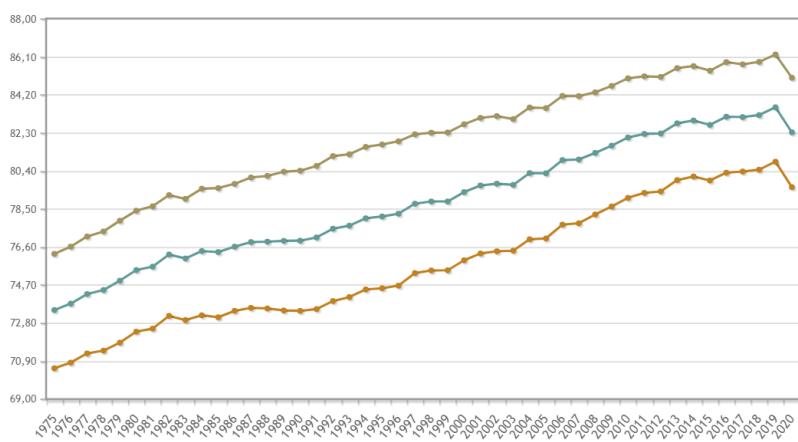
**Figura 3.** Ejemplo de la visualización de la esperanza de vida en Datos Macro

Nota. Muestra la esperanza de vida al nacer en España en 2021, dividido en hombres, mujeres y ambos. Tomado de *Esperanza de vida al nacimiento* [Fotografía], de DatosMacro, 2019, Expansión (<https://datosmacro.expansion.com/demografia/esperanza-vida/espana>). Obra de Dominio Público

Por otro lado, para el tema propuesto que se enmarca en España, se tienen datos oficiales y accesibles de forma online, que también se muestran por medio de visualizaciones, pero a menudo son complicados de encontrar, incompletos o se encuentran dispersos. Es el caso por ejemplo del Instituto Nacional de Estadística [INE]. (s.f.) donde se almacenan todo tipo de

estadísticas con datos de interés sobre España, pero las visualizaciones en ocasiones no son claras, están ordenadas de forma a veces aleatoria, son difíciles de encontrar y la mayoría cubren únicamente unos pocos años presentando valores actuales o de 10 años atrás, de esta forma no se percibe una evolución del todo real. A su vez son complicadas de extraer por el formato escogido, la información se encuentra en numerosos y extensos documentos que en definitiva no son manejables para el usuario medio, y por último muestran los datos aislados sin contexto o perspectiva frente a otros países o épocas. A continuación, se muestra un ejemplo de un gráfico en la figura 4.

**Figura 4.** Ejemplo de visualización de la esperanza de vida en el INE



Nota. Muestra la esperanza de vida en España entre el año 1975 y 2020. Tomado de *Esperanza de vida al nacimiento según sexo*, de El Instituto Nacional de Estadística, 2021a, Ministerio de Asuntos Económicos y Transformación Digital (<https://www.ine.es/jaxiT3/Datos.htm?t=1414#!tabs-grafico>).

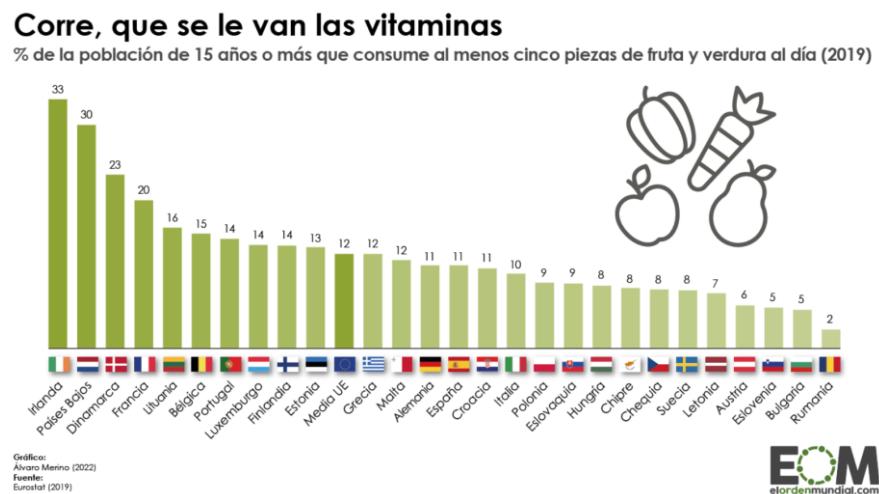
Obra de Dominio Público.

Esta vez referente en su mayoría a temas europeos sobresale un medio de análisis internacional denominado El Orden Mundial (<https://elordenmundial.com/>) que en su sitio web al completo, presenta mapas y gráficos de temas importantes, y otros más diversos acerca de la Unión Europea como por ejemplo los principales países productores de cerveza, o el color de cada pasaporte en el mundo.

Un punto fuerte de este sitio web es que las visualizaciones son llamativas y estéticas, además los temas tratados suelen ser curiosos e interesantes, pero el problema encontrado aquí es que no se tratan los temas más importantes referentes a un país en concreto, en el ámbito económico, de salud o de derechos humanos. Y los datos al igual que en las anteriores (muchas veces porque no es necesario para los temas que representan) no se muestran en un contexto de evolución en el tiempo sino en determinados momentos concretos.

De igual forma es un sitio web útil, que presenta visualizaciones llamativas y claras, que además resultan interesantes para el lector. Un ejemplo se puede observar en la figura 5.

**Figura 5.** Ejemplo de visualización en *El Orden Mundial*



Nota. Muestra el porcentaje de población mayor de 15 años que consume al menos 5 piezas de fruta y verdura al día. Tomado de *Corre, que se le van las vitamininas*, de Merino, A., 2022, El Orden Mundial (<https://elordenmundial.com/mapas-y-graficos/el-consumo-de-frutas-y-verduras-en-la-union-europea/>). CC BY-NC-ND 2.5 ES

En los sitios web comentados existe una gran cantidad de información que se encuentra dispersa, y aunque correcta, en ocasiones está mal presentada, resulta confusa o no aporta un contexto o perspectiva fácil de inferir. Sirven para encontrar una determinada información para un momento concreto, y a veces es complicada de encontrar. Es difícil encontrar la información concreta de un país en contraste con los datos de la Unión Europea y los países miembros.

Para el tema tratado resulta muy útil e interesante la experiencia visual que se propone con esta nueva metodología. Se pretende solucionar los problemas encontrados añadiendo valor que llame la atención sobre el usuario, mostrando la información dinámicamente en relación con la UE, de una manera más estética y ordenada de una forma concreta que facilite la comprensión de los datos. Además, al añadir interactividad el usuario se siente más protagonista y puede filtrar los datos para ver los detalles y relaciones más claramente.

### 3 Objetivos y metodología de trabajo

Se define el objetivo general y los objetivos específicos del desarrollo de la metodología propuesta.

### 3.1 Objetivo general

El objetivo principal es establecer una metodología de visualización interactiva de datos para conseguir que un usuario sea capaz de adquirir conocimiento sobre un país, de una forma sencilla e independiente, que le permita extraer sus propias conclusiones y realizar su propio análisis.

### 3.2 Objetivos específicos

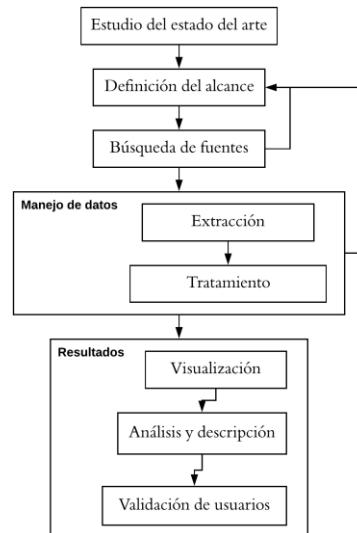
Para cumplir con el objetivo principal se plantean los siguientes subobjetivos:

- Establecer cuáles son los principales temas por tratar del país e identificar las posibilidades de captura de los datos.
- Determinar y describir todos los datos necesarios para desarrollar el estudio y los paneles interactivos.
- Desarrollar las visualizaciones interactivas con la claridad suficiente para que el usuario sea capaz de adquirir conocimiento con valor.
- Evaluar la metodología propuesta en un escenario real.

### 3.3 Metodología del trabajo

En orden de cumplimentar los objetivos indicados se ha configurado una metodología como la que se describe en la figura 6, y que se desarrolla más adelante.

**Figura 6.** Diagrama de la metodología para la elaboración del estudio



Fuente: Elaborada por el autor

**Estudio del estado del arte:** Es necesario conocer el contexto y cuáles son los problemas que existen en proyectos similares, que traten temas vinculados al que pretende resolver la metodología propuesta, y para ello se debe realizar una investigación previa. Esta parte es importante para poder hacer una contribución relevante y que el trabajo tenga un valor añadido.

**Definición del alcance:** Muchas veces los objetivos propuestos no se pueden cumplir por diversas razones, quizás no haya datos suficientes, sean incompletos o erróneos, y en otras ocasiones las fuentes encontradas no sean de confianza. Este paso se debe realizar de forma iterativa a lo largo del desarrollo del trabajo, hay que ir descubriendo que datos se tienen disponibles y como se pueden encajar dentro de un contexto, o que cumplan unos mínimos para los objetivos que se han propuesto.

**Búsqueda de fuentes:** En este paso se lleva a cabo un proceso de búsqueda de datos masivos acerca de la problemática a tratar en el trabajo, en este caso sobre España y la Unión Europea. Para ello se deben utilizar herramientas y recursos adecuados tanto de forma online en internet como de artículos académicos y científicos disponibles. Es muy importante contrastar que la información que se aporta sea veraz.

**Manejo de datos:** La parte más importante del proyecto, los datos, son los cimientos de todo el desarrollo y deben cumplir con una serie de requisitos primordiales que son, completitud, veracidad y relevancia. Para conseguir estos datos en la forma que se necesita se tienen 2 subprocessos:

- **Extracción:** Proceso por el cual se obtienen los datos, en este caso se han utilizado 2 formas diferentes, la primera que consiste en descargar los datasets directamente en crudo de la fuente, y la segunda mediante la construcción de un método de *web scraping* que recopila la información de forma automática de la web.
- **Tratamiento:** A parte de los requisitos mencionados anteriormente, los datos deben cumplir con un formato y forma adecuados que permita su correcto análisis y posterior visualización. Para ello mediante el lenguaje de programación Python (Versión 3.10) se lleva a cabo todo un proceso de limpieza y transformación de los datos, esto es, filtrar, cambiar formatos, eliminar vacíos o modificar tipos.

**Resultados:** Una vez se tienen los datos cumpliendo todos los requisitos es momento de mostrarlos de forma clara y objetiva al usuario en orden de aumentar sus conocimientos de forma sencilla, para ello se realiza lo siguiente:

- **Visualización:** Se construyen todas las visualizaciones necesarias que permitan conocer la información que se pretende aportar de forma completa, correcta, a la vez que estética para que el usuario reciba la información fácilmente sin esfuerzo y reduciendo al máximo las probabilidades de error al extraer conclusiones.

Este paso se realiza con la herramienta software Tableau (Stolte et al, 2003) que permite la construcción de visualizaciones con grandes volúmenes de datos de forma intuitiva y con numerosas opciones potentes de cálculo.

Finalmente se construye una página web donde se exponen todos los resultados del estudio haciendo uso de los lenguajes de programación y herramientas software siguientes: HTML5, CSS, Bootstrap (Otto, M. y Thornton, J. 2011), Brackets (Adobe, 2014) y GitHubPages (Preston et al. 2008).

- **Análisis y descripción:** Con las figuras ya construidas y ubicadas en un contexto, se pueden definir los principales cambios observables y cuantificables, de tal forma que se percibe como ha sido la evolución y la situación actual del país (en este caso España) desde una perspectiva que engloba a la Unión Europea y los países miembros.
- **Validación de usuarios:** Se trata de un paso necesario para comprobar que la metodología propuesta es válida para los lectores que al final son los usuarios finales del estudio y los que se van a apoyar en un futuro de este documento. Para ello se van a realizar una serie de entrevistas con el objetivo de sacar conclusiones y solucionar errores o ambigüedades.

El objetivo final como se comenta al principio del documento, son las personas que no tienen por qué disponer de un conocimiento extenso sobre el tema ni ser expertos en la problemática tratadas, el público objetivo es cualquier individuo que quiera informarse y conocer la situación real desde una perspectiva europea, este hecho se desarrolla más en detalle en la siguiente sección de identificación de requisitos.

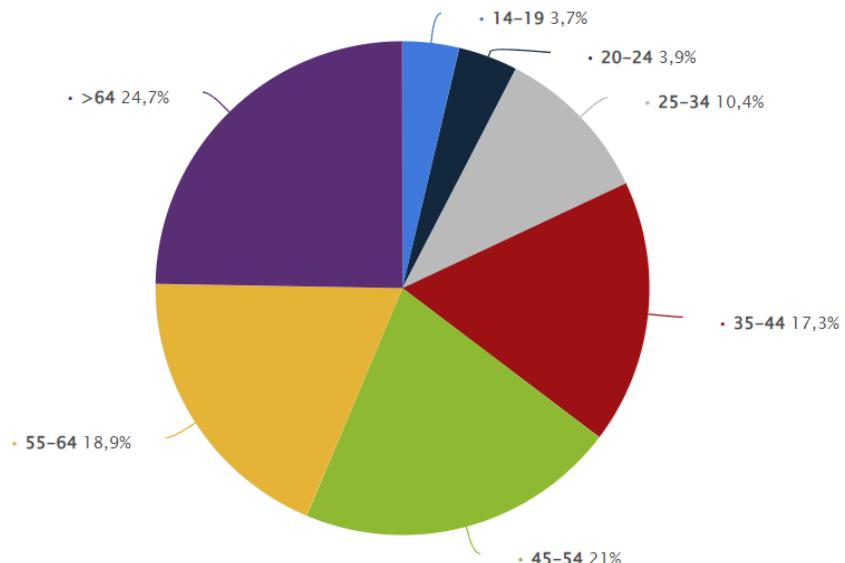
## 4 Desarrollo específico de la contribución

En este apartado se detalla el desarrollo del trabajo, pasando por la identificación de los requisitos, la descripción de la metodología y la evaluación que la valida.

## 4.1 Identificación de requisitos

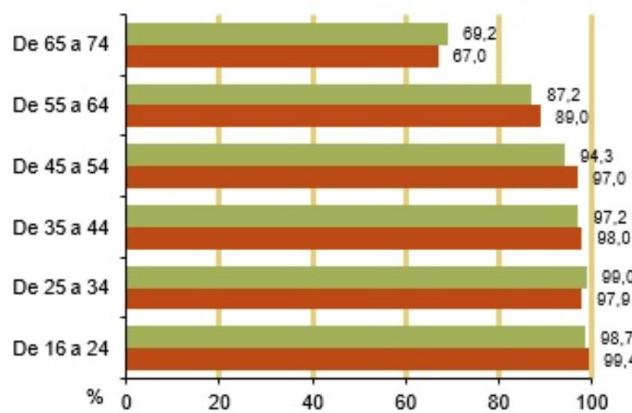
La idea principal y por la que surgió la temática escogida para la realización del proyecto es la de mostrar grandes cantidades de datos sobre un país concreto de tal forma que pueda ser comprendida por la gran mayoría del público, es decir, el usuario medio no debería tener un nivel muy alto de conocimientos sobre la Unión Europea o sobre conceptos como que mide el PIB de un país. Se entiende por usuario medio a una persona que suele estar interesada por este tipo de noticias, pero no es un experto en la materia, pueden ser lectores habituales de prensa que además están familiarizados con las nuevas tecnologías. Según la figura 7, el 67,6% de los individuos encuestados que consumen este tipo de medios están entre los 25 y 64 años, además los mayores de 64 años forman parte del 24,7%.

**Figura 7.** Distribución porcentual de los grupos de edad que leyeron prensa en 2020



Nota. El gráfico representa la distribución por edades del porcentaje de lectores de prensa en España en el año 2020. Tomado de *Distribución porcentual de individuos que leyeron periódicos de tirada diaria en España en 2021, por grupo de edad*, por Orús, A., 2021, Statista (<https://es.statista.com/estadisticas/476755/distribucion-de-lectores-de-periodicos-en-espana-por-edad/>). Obra de Dominio Público.

Según la figura 8 el público más aficionado al uso de internet comprende las edades entre los 16 años y los 54 años.

**Figura 8.** Grupos de edad que han usado internet de forma frecuente en 2021 (en %)

Nota. El gráfico representa la distribución por edades del porcentaje de personas que han utilizado regularmente internet en el año 2021. Tomado de *Población que ha usado Internet de manera frecuente en los últimos tres meses por grupos de edad*, por Instituto Nacional de Estadística, 2021b, INE ([https://www.ine.es/jaxi/Datos.htm?path=/t00/mujeres\\_hombres/tablas\\_2/l0/&file=C5G1.px#!tab grafico](https://www.ine.es/jaxi/Datos.htm?path=/t00/mujeres_hombres/tablas_2/l0/&file=C5G1.px#!tab grafico)). Obra de Dominio Público.

A la hora de evaluar la metodología propuesta, se van a tener en cuenta estos datos, y por tanto las visualizaciones y el cuestionario de evaluación se intentarán centrar en las personas que comprendan las edades entre 25 y 54 años que deberían poder extraer información de forma clara e independiente, extrayendo sus propias conclusiones. Los datos que se van a exponer en las visualizaciones interactivas se enumeran a continuación, junto con la razón de su elección.

- Unión Europea: Índice de los Derechos humanos, Producto interior bruto, Ingresos fiscales, Impuestos, Gasto público y distribución, Índice de desigualdad GINI, Tasa de desempleo, Esfuerzo vs productividad, Percepción de la corrupción, Emisiones de CO<sub>2</sub>, Territorio dedicado a fines agrícolas, Factores de riesgo para la salud, Alimentación y salud, Factores de riesgo en niños, Muertes por conflictos o terrorismo
- España: Índice de los derechos humanos, Producto interior bruto, Deuda (% del PIB), Déficit (% del PIB), Índice de precios al consumo, Exportaciones e Importaciones, Ingresos fiscales, Gasto público y distribución, Índice de desigualdad GINI, Tasa de desempleo, Salario mínimo interprofesional, Esfuerzo vs productividad, Percepción de la corrupción, Emisiones de CO<sub>2</sub> per cápita, Consumo y generación de energía, Energías renovables, Territorios con fines agrícolas, Factores de riesgo para la salud, Número de homicidios y suicidios, Alimentación y salud, Factores de riesgo en niños, Población y esperanza de vida, Balanza demográfica

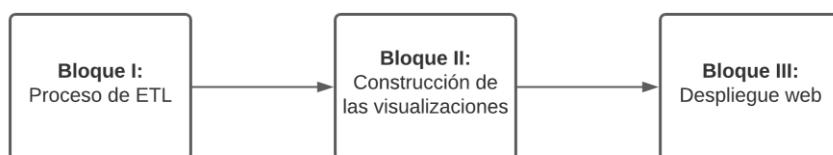
En primera instancia se ha decidido capturar esta información por pertenecer a las categorías que la Unión Europea denomina como “Las cuatro prioridades” y coloca como las más relevantes, donde hay que poner toda la atención, “proteger a los ciudadanos y las libertades, desarrollar una base económica sólida y dinámica, construir una Europa climáticamente neutra, ecológica, justa y social, y promover los intereses y valores europeos en el mundo” (Unión Europea, 2021).

## 4.2 Descripción de la metodología

Se trata de una nueva metodología para la exposición de grandes cantidades de datos referentes a un país de la Unión Europea, y, para su desarrollo, se ha decidido dividir en tres grandes bloques.

El primer bloque referente a la búsqueda y captura de datos, tratamiento y posterior carga de estos, el segundo bloque vinculado al diseño y construcción de las visualizaciones interactivas que le dan forma al trabajo, y un tercer y último bloque que permita el despliegue de todas las visualizaciones realizadas con el fin de presentarlas de forma ordenada, con el formato correcto y la claridad y completitud adecuadas. (Ver figura 9)

**Figura 9.** Bloques en el desarrollo



Fuente: Elaborada por el autor

A continuación, se exponen los pasos, herramientas y tecnologías necesarias para la correcta ejecución de cada uno de los bloques correspondientes a la metodología.

### 4.2.1 Bloque I: Proceso de ETL (extracción, transformación y carga de datos)

Para llevar a cabo este proceso es necesario realizar como primer paso una búsqueda exhaustiva de los datos disponibles referentes al tema sobre el que se va a realizar la exposición, y que se han relatado en el apartado anterior de identificación de requisitos, en

este caso, sobre los datos más relevantes de España como país, y los datos sobre la Unión Europea (el conjunto de países miembros) que aportarán contexto y perspectiva a los datos.

#### **4.2.1.1 Fuentes de los datos**

Referente a los datos europeos, el estudio se ha llevado a cabo a partir de los datos extraídos de Our World in Data por ser la fuente de datos más fiable y completa encontrada, además cuenta con la colaboración y apoyo de las instituciones académicas e informativas más influyentes y mejor valoradas del mundo.

En cuanto a los datos sobre España, se ha utilizado la aplicación web DatosMacro que se trata de una web muy completa y con una gran cantidad de información reunida sobre España y otros países. Utiliza fuentes fiables que son oficiales del país y reúne todos los datos en tablas, con las que se ha podido trabajar mediante un método de *Web Scraping* que se explica más adelante en detalle para la extracción de información.

Es importante en este punto escoger los datos correctos, descartando algunos más relevantes pero que no cumplen con los requisitos mínimos para ser una información veraz o completa.

Una vez se tienen las fuentes de los datos localizadas y se conocen los datos relevantes que se quieren extraer, es momento de construir los scripts que nos van a permitir realizar el proceso de ETL.

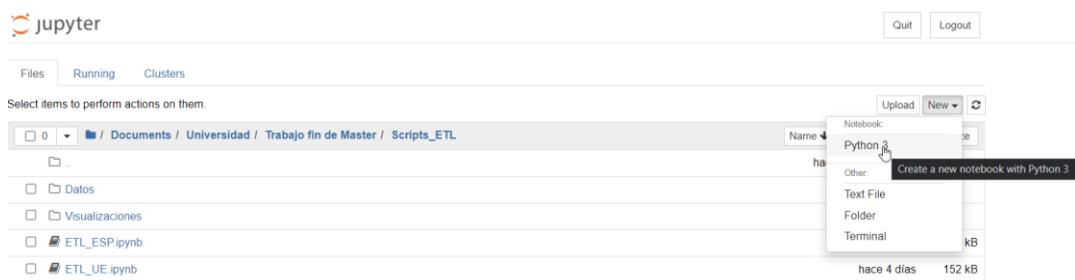
#### **4.2.1.2 Extracción y tratamiento de los datos**

En el proceso de captura y transformación de los datos se ha utilizado el lenguaje de programación Python (Versión 3.10) junto con las librerías: Pandas (McKinney, 2010), Matplotlib (Hunter, 2003), BeautifulSoup (Richardson, 2012) y Request (Reitz, 2011).

En la elaboración de los scripts de Python se ha utilizado la interfaz Jupyter Notebook (Pérez, 2014), se trata de un entorno de desarrollo para multitud de lenguajes, entre ellos, Python, y permite definir de una forma clara y sencilla las instrucciones como si se tratase de un cuaderno, de esta forma permite añadir títulos, anotaciones y comentarios de una manera que la pueda entender alguien que no se dedique de forma exclusiva a la programación. Esto hace más sencillo compartir el código en distintos formatos como .pdf por ejemplo, e incluir todo tipo de anotaciones que ayuden y guíen al usuario que pretenda implementar la nueva metodología.

Jupyter Notebook se puede instalar dentro de un sistema de gestión de entornos como es Conda (Anaconda Inc. 2017). Una vez instalado solo es necesario ejecutar *Jupyter Notebook* y comenzar con la creación de los scripts. Para ello accedes a la carpeta donde decidas posicionarlo, como se ve en la figura 10, y se pulsa sobre la pestaña ‘new’ que despliega varias opciones, en este caso la utilizada es Python 3.

**Figura 10.** Creación script en Jupyter Notebook



Fuente: Elaborada por el autor

Para la instalación de las librerías necesarias ya comentadas, solo es necesario ejecutar las siguientes instrucciones desde el prompt de Conda: ‘pip install pandas’, ‘pip install matplotlib’, ‘pip install request’ y ‘pip install beautifulsoup’.

Además de la descarga directa de los datasets como archivos ‘comma separated values’ (.csv), se ha utilizado un método de *web scraping* con el fin de leer los datos de la propia web, ya que, aun siendo datos abiertos de uso público, no están disponibles para su descarga. A continuación, se observa un ejemplo de código (por su extensión no es posible ponerlo al completo) con los comentarios correspondientes para la extracción del producto interior bruto en España.

Mediante este método como se muestra a continuación, se introduce la url del sitio web donde se encuentra la tabla con los datos, los métodos de la librería ‘BeautifulSoup’ permiten acceder al formato HTML y extraer de forma concreta los datos de la tabla, los cuales se van recorriendo con la instrucción ‘for’ y guardando uno a uno realizando los cambios en el formato que se desea en cada caso.

#### URL donde se encuentran los datos

```
url="https://datosmacro.expansion.com/pib/espana"
```

#### Petición GET con el "crudo" del contenido HTML

```
html_content = requests.get(url).text
```

#### Extracción del contenido HTML

```
soup = BeautifulSoup(html_content, "lxml")
```

## Extracción de tablas con los datos, del contenido HTML

```
tablas = soup.find_all('table')
```

## Extracción de los valores de cada tabla

```
tr = tablas[2].tbody.find_all("tr")
df_pib_esp = pd.DataFrame(columns = ['FECHA' , 'PIB',
'PIB_VAR_PORCENTAJE'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0' and values[3] != '0':
        df_pib_esp = df_pib_esp.append({ 'FECHA' : values[0] , 'PIB' :
values[1] , 'PIB_VAR_PORCENTAJE' : values[3]} , ignore_index=True)
    i = i + 1
```

## Modificación de tipos y guardado de datos

```
df_pib_esp[['PIB' , 'PIB_VAR_PORCENTAJE']].astype(float)
df_pib_esp['FECHA'] = pd.to_datetime(df_pib_esp.FECHA)
df_pib_esp['FECHA'] = df_pib_esp['FECHA'].dt.year
```

Al final del script todos los datos extraídos se reúnen en una sola tabla que será la utilizada en el momento de construir las visualizaciones. En el Anexo se dejan los scripts completos para su análisis.

Por otro lado, para el caso en el que se han descargado los datasets directamente en formato csv, se debe realizar también un tratamiento de los datos que nos permitan tener una información con el formato correcto, limpia y con la extensión correcta, un ejemplo de transformación sencillo sería el siguiente, con la cantidad de población total en la Unión Europea:

## Lectura del dataset descargado manualmente

```
df_population =
pd.read_csv("Datos/Generales_UE/Salud_alimentacion_contaminacion/population-
since-1800.csv")
```

## Filtrado de países de la Unión europea

```
df_population = df_population[df_population.Entity.isin(countries)]
```

## Renombrado de columnas

```
df_population_UE = df_population.rename(columns={'Entity': 'COUNTRY',
'Code': 'ISO_CODE', 'Year': 'YEAR', 'Population (historical estimates)': 'POPULATION'})
```

### Filtrado de años necesarios

```
df_population_UE = df_population_UE[df_population_UE['YEAR'] >= min_year]
```

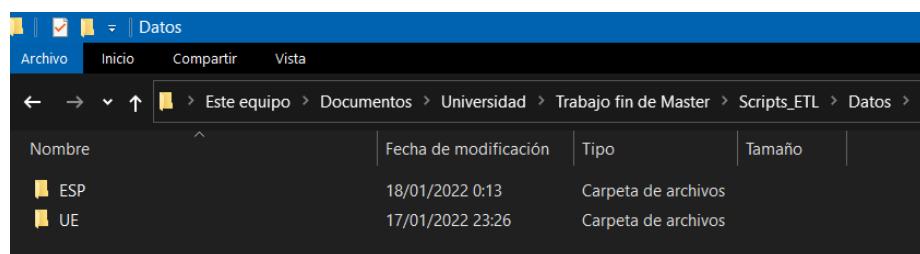
### Guardado de los datos

```
df_population_UE.to_excel('Datos/Generales_UE/UE/Paz_y_seguridad/population_UE.xlsx', index = False, sheet_name='population_UE')
```

En el Anexo se puede encontrar adjunto el código completo para su análisis.

Al ejecutar los dos scripts de Python quedará toda la información extraída de forma local en formato Excel (.xlsx), de la siguiente forma: (Ver figura 11)

**Figura 11.** Carpetas creadas por los scripts



Fuente: Elaborada por el autor

Dos directorios separados, uno para los datos referentes únicamente al país (España), y otro vinculada solo a los datos de la Unión Europea. La carpeta del país 'ESP' contiene el archivo 'TOTALES.csv' que contiene todos los datos ya tratados sobre España. Se trata de un archivo CSV con 56 columnas, una por cada atributo del dataset y 60 entradas, una por cada año del que se tienen datos que se describe a continuación:

- Fecha: Año en formato DATE
- PIB: Producto interior bruto en formato FLOAT
- PIB\_VAR\_PORCENTAJE: Producto interior bruto en formato FLOAT
- DEUDA\_TOTAL: Deuda total en formato FLOAT
- DEUDA\_PIB\_PORCENTAJE: Deuda en porcentaje respecto al PIB en formato FLOAT
- DEUDA\_PER\_CAPITA: Deuda por habitante en formato FLOAT

- DEFICIT: Déficit total en formato FLOAT
- DEFICIT\_PIB\_PORCENTAJE: Déficit en porcentaje respecto al PIB en formato FLOAT
- EDU\_GASTO\_TOTAL: Gasto total en educación en formato FLOAT
- EDU\_GASTO\_PIB\_PORCENTAJE: Gasto en educación como porcentaje del PIB en formato FLOAT
- SALUD\_GASTO\_TOTAL: Gasto total en salud en formato FLOAT
- SALUD\_GASTO\_PIB\_PORCENTAJE: Gasto en salud como porcentaje del PIB en formato FLOAT
- DEF\_GASTO\_TOTAL: Gasto total en defensa en formato FLOAT
- DEF\_GASTO\_PIB\_PORCENTAJE: Gasto en defensa como porcentaje del PIB en formato FLOAT
- SMI: Salario mínimo interprofesional en formato FLOAT
- SALARIO: Salario medio en formato FLOAT
- IPC: Índice de precios al consumo en formato FLOAT
- INDICE\_CORRUPCION: Índice de percepción de la corrupción en formato FLOAT
- TASA\_DESEMPLEO: Tasa de desempleo en formato FLOAT
- EXPORTACIONES: Ingresos en Exportaciones en formato FLOAT
- EXPORT\_PORCENTAJE\_PIB: Ingresos en exportaciones como porcentaje respecto al PIB en formato FLOAT
- IMPORTACIONES: Gastos en importaciones en formato FLOAT
- IMPORT\_PORCENTAJE\_PIB: Gastos en importaciones como porcentaje respecto al PIB en formato FLOAT
- BALANZA\_COMERCIO: Balanza de comercio en formato FLOAT
- BALANZA\_PIB\_PORCENTAJE: Balanza de comercio como porcentaje respecto al PIB en formato FLOAT

- NACIDOS: Número total de nacimientos en formato FLOAT
- NACIDOS\_HOMBRES: Número total de nacimientos hombres en formato FLOAT
- NACIDOS\_MUJERES: Número total de nacimientos mujeres en formato FLOAT
- IND\_FECUNDACION: Índice de fecundación en formato FLOAT
- MUERTES: Número total de muertes en formato FLOAT
- MUERTES\_HOMBRES: Número total de muertes en hombres en formato FLOAT
- MUERTES\_MUJERES: Número total de muertes en mujeres en formato FLOAT
- TASA\_MORTALIDAD: Tasa de mortalidad en formato FLOAT
- RIESGO\_POBREZA: Riesgo de pobreza en formato FLOAT
- ESP\_MUJERES: Esperanza de vida en mujeres en formato FLOAT
- ESP\_HOMBRES: Esperanza de vida en hombres en formato FLOAT
- ESPERANZA\_VIDA: Esperanza de vida en formato FLOAT
- IDH: Índice de derechos humanos en formato FLOAT
- S\_MUJERES: Número de suicidios en mujeres en formato FLOAT
- S\_HOMBRES: Número de suicidios en hombres en formato FLOAT
- TASA\_S\_MUJERES: Tasa de suicidios en mujeres en formato FLOAT
- TASA\_S\_HOMBRES: Tasa de suicidios en hombres en formato FLOAT
- SUICIDIOS\_POR\_100K: Suicidios por cada 100.000 habitantes en formato FLOAT
- HOMICIDIOS: Número total de homicidios en formato FLOAT
- HOMICIDIOS\_POR\_100K: Homicidios por cada 100.000 habitantes en formato FLOAT
- CO2\_TOTALES: Emisiones en toneladas de CO2 en formato FLOAT
- CO2\_PER\_CAPITA: Emisiones de CO2 por habitante en formato FLOAT
- GENERACION\_GWH: Generación energética en formato FLOAT

- CONSUMO\_GWH: Consumo energético en formato FLOAT
- CONSUMO\_PER\_CAPITA: Consumo energético por habitante en formato FLOAT
- GENERACION\_RENOVABLES: Generación de energías renovables en formato FLOAT
- DENSIDAD: Densidad de población en formato FLOAT
- HOMBRES\_POB: Número de habitantes hombres en formato FLOAT
- MUJERES\_POB: Número de habitantes mujer en formato FLOAT
- POBLACION\_TOTAL: Población total en formato FLOAT

Respecto a la Unión Europea, los sets de datos utilizados son:

Population\_UE: Dataset con la población total en cada país de la Unión Europea cada año desde 1946.

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- POPULATION: Total de población

deaths\_UE: Dataset con el número de muertes causadas por conflictos y terrorismos en los países de la Union europea desde 1990

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- CONFLICT\_TERRORISM\_DEATHS: Número total de muertes causadas por conflictos o terrorismos.

military\_expenditure: Dataset con el total de gasto militar por año y por persona en cada país de la Unión Europea desde 1960

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país

- YEAR: Año de la toma de datos
- MILITARY\_EXPENDITURE\_GDP\_PERCENTAGE: Gasto militar en porcentaje respecto al PIB
- MILITARY\_EXPENDITURE\_BILLION\_DOLLARS: Gasto militar en millones de dólares

agricultural\_area\_per\_capita\_UE: Dataset con el número total de hectáreas por persona dedicadas a fines agrícolas en los países de la Unión Europea desde el año 1961

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- AGRICULTURAL\_LAND\_PER\_CAP: Hectáreas dedicadas a la agricultura por persona

daily\_per\_capita\_fat\_suplly\_UE: Dataset con el consumo total de grasas por persona y día dentro de los países de la Unión Europea.

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- FAT\_QTY\_GR\_PER\_CAP\_DAY: Cantidad de grasa consumida por persona y día

daily\_per\_capita\_protein\_supply: Dataset con el consumo total de proteínas por persona y día dentro de los países de la Unión Europea.

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- PROTEIN\_QTY\_GR\_PER\_CAP\_DAY: Cantidad de proteínas consumida por persona y día

food\_supply\_ue: Dataset con información sobre la esperanza de vida y el consumo de calorías en los países pertenecientes a la Unión Europea desde el año 1950.

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- LIFE\_EXPECTANCY: Esperanza de vida
- CALORIC\_SUPPLY: Consumo de calorías
- POPULATION: Población
- CONTINENT: Continente

stunted\_wasted\_population: Construcción del dataset definitivo para la salud de los niños

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- CHILDREN\_STUNTED: Número de niños con retraso del crecimiento
- CHILDREN\_WASTED: Número de niños con debilidad
- POPULATION: Población
- STUNTED\_PERCENTAGE: Número de niños con retraso del crecimiento respecto a la población
- WASTED\_PERCENTAGE: Número de niños con debilidad respecto a la población
- SUM\_SW: Suma de porcentajes anteriores

deaths\_by\_risk\_factor: Dataset con la información referente a cuáles son los principales factores de riesgo de muerte y cuál es el impacto de cada una para la UE.

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- UNSAFE\_WATER: Agua no potable

- UNSAFE\_SANITATION: Sanidad pobre
- NO\_ACCESS\_HANDWASHING: Sin acceso al lavado de manos
- AIR POLLUTION\_SOLID\_FUEL: Contaminación debida a combustibles sólidos
- NON\_EXCLUSIVE\_BREASTFEEDING: Lactancia materna no exclusiva
- DISCONTINUED\_BREASTFEEDING: Lactancia materna discontinuada
- CHILD\_WASTING: Niños con debilidad
- CHILD\_STUNTING: Niños con retraso del crecimiento
- LOW\_BIRTH\_WEIGHT: Peso bajo al nacer
- SECONDHAND\_SMOKE: Fumador pasivo
- ALCOHOL\_USE: Abuso del alcohol
- DRUG\_USE: Consumo de drogas
- DIET\_LOW\_FRUITS: Dieta baja en frutas
- DIET\_LOW\_VEGETABLES: Dieta baja en verduras
- UNSAFE\_SEX: Sexo sin protección
- LOW\_PHYSICAL\_ACTIVITY: Actividad física baja
- FASTING\_PLASMA\_GLUCOSE: Glucosa plasmática en ayunas
- HIGH\_COLESTEROL: Colesterol alto
- HIGH\_BODYMASS\_INDEX: Índice de masa corporal alto
- HIGH\_BLOOD\_PRESSURE: Presión arterial alta
- SMOKING: Fumador
- IRON\_DEFICIENCY: Deficiencia de hierro
- VITAMIN\_A\_DEFICIENCY: Deficiencia de vitamina A
- LOW\_BONE\_MINERAL\_DENSITY: Baja densidad de minerales en los huesos
- AIR\_POLLUTION: Contaminación en el aire

- OUTDOOR\_AIR POLLUTION: Contaminación exterior en el aire
- DIET\_HIGH\_SODIUM: Dieta alta en sal
- DIET\_LOW\_GRAINS: Dieta baja en cereales
- DIET\_LOW\_NUTS\_SEEDS: Dieta baja en semillas y nueces

dietary\_composition: Dataset con la información de los componentes principales de la dieta en los países de la Unión Europea desde el año 1961

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- Other (FAO (2017)) (kilocalories per person per day): Otros
- Sugar (FAO (2017)) (kilocalories per person per day): azúcar
- Oils & Fats (FAO (2017)) (kilocalories per person per day): Aceites y grasas
- Meat (FAO (2017)) (kilocalories per person per day): Carne
- Dairy & Eggs (FAO (2017)) (kilocalories per person per day): Huevos y lácteos
- Fruit and Vegetables (FAO (2017)) (kilocalories per person per day): Frutas y verduras
- Starchy Roots (FAO (2017)) (kilocalories per person per day): Raíces almidonadas
- Pulses (FAO (2017)) (kilocalories per person per day): Legumbres
- Cereals and Grains (FAO (2017)) (kilocalories per person per day): Granos y cereales
- Alcoholic Beverages (FAO (2017)) (kilocalories per person per day): Bebidas alcohólicas

co2\_UE: Dataset con el total de emisiones de CO2 en toneladas y por persona para los países de Unión Europea desde el año 1946

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos

- CO2\_EMISSIONS\_TONNES: Emisiones de CO2 en toneladas
- CO2\_EMISSIONS\_TONNES\_PER\_CAPITA: Emisiones de CO2 en toneladas por persona
- POPULATION: Población

hr\_scores\_vs\_gdp\_UE: Dataset con la valoración en el cumplimiento de los derechos humanos dentro de los países de la Unión Europea desde el año 1990

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- GDP: Producto interior bruto (PIB)
- HR\_SCORE: Índice de cumplimiento de los derechos humanos

Gini\_index\_UE: Dataset con el índice Gini que sirve de referencia para indicar la desigualdad existente en cada país de la Unión Europea desde el año 1981

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- GINI\_INDEX: Índice Gini

expenditure\_education\_UE: Dataset con el gasto en educación en cada país de la Unión Europea desde el año 1971

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- EXPENDITURE\_EDUCATION: Gasto en educación

expenditure\_health\_UE: Dataset con el gasto en salud en cada país de la Unión Europea desde el año 1970

- COUNTRY: Países pertenecientes a la Unión Europea

- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- EXPENDITURE\_HEALTH: Gasto en salud

expenditure\_military\_UE: Dataset con el gasto militar en cada país de la Unión Europea desde el año 1970

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- EXPENDITURE\_MILITARY: Gasto militar

corruption\_perception: Dataset con la percepción de los ciudadanos sobre el nivel de corrupción en su país

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- CORRUPTION\_PERCEPTION: Percepción del nivel de corrupción

productivity\_vs\_hours\_worked\_UE: Dataset con el nivel de productividad y las horas trabajadas en la Unión Europea

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- HOURS\_WORKED\_ Horas trabajadas
- PRODUCTIVITY: Nivel de productividad

extreme\_poverty\_UE: Dataset con el porcentaje de población en riesgo de pobreza extrema

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país

- YEAR: Año de la toma de datos
- EXTREME\_POVERTY: Porcentaje de población en riesgo de pobreza extrema

taxes\_UE: Dataset con los datos sobre tasas de cada país de la Unión Europea

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- TAXES\_INCOME: Impuestos sobre la renta
- TAXES\_ON\_GOODS\_SERVICES: Impuestos sobre bienes y servicios

total\_expenditures\_UE: Dataset con la cantidad total de gastos

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- EXPENSES\_BY\_GDP: Total de gastos respecto al PIB

tax\_revenues: Dataset con la cantidad total de ingresos fiscales

- COUNTRY: Países pertenecientes a la Unión Europea
- ISO\_CODE: Código del país
- YEAR: Año de la toma de datos
- TAXES\_REVENUE\_BY\_GDP: Ingresos fiscales

#### **4.2.2 Bloque II: Desarrollo de las visualizaciones interactivas**

En la elaboración de las visualizaciones interactivas se ha utilizado el software de análisis e inteligencia de negocios Tableau que permite la construcción de análisis visuales para mostrar los datos, con un gran número de posibilidades que facilitan su construcción, y hace más fácil explorar los datos para descubrir tendencias, problemas, puntos de mejora, o simplemente una descripción de la información rápida que permita comprender los datos.

Por la imposibilidad de describir el desarrollo técnico de cada visualización en detalle (sería demasiado extenso y repetitivo), se ha optado por describir en detalle las opciones utilizadas que aporta el software y de qué forma se han utilizado, para más adelante describir por qué y cómo se ha construido cada una de las visualizaciones.

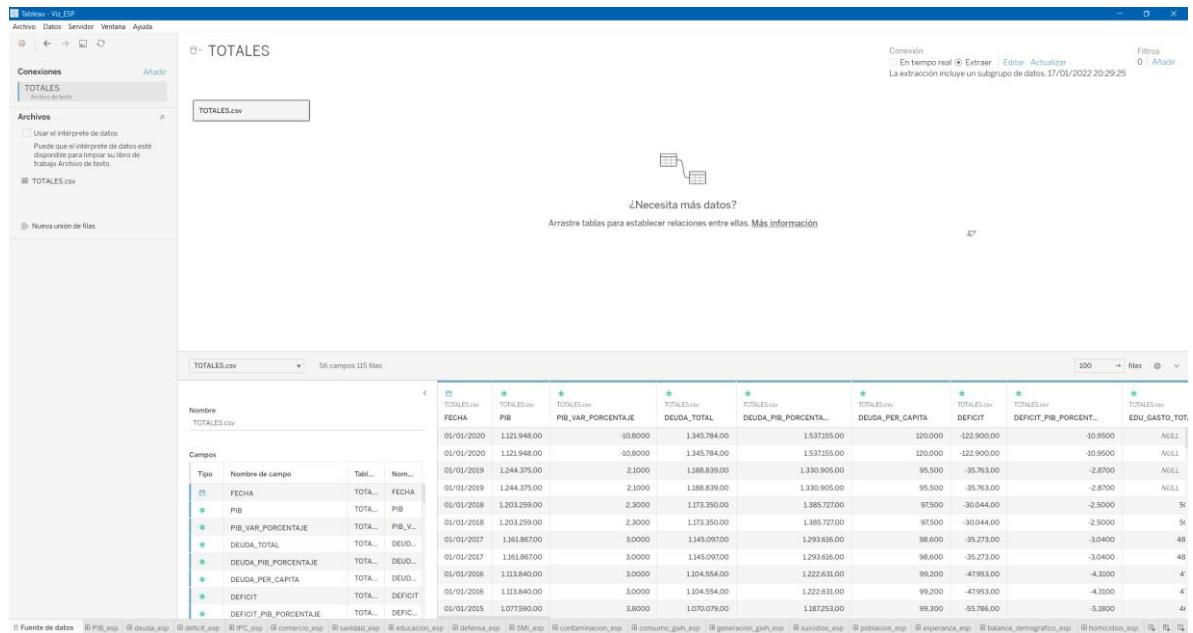
Se comienza con la importación de datos como se muestra en la figura 12.

**Figura 12.** Importar datos en Tableau



Fuente: Elaborada por el autor

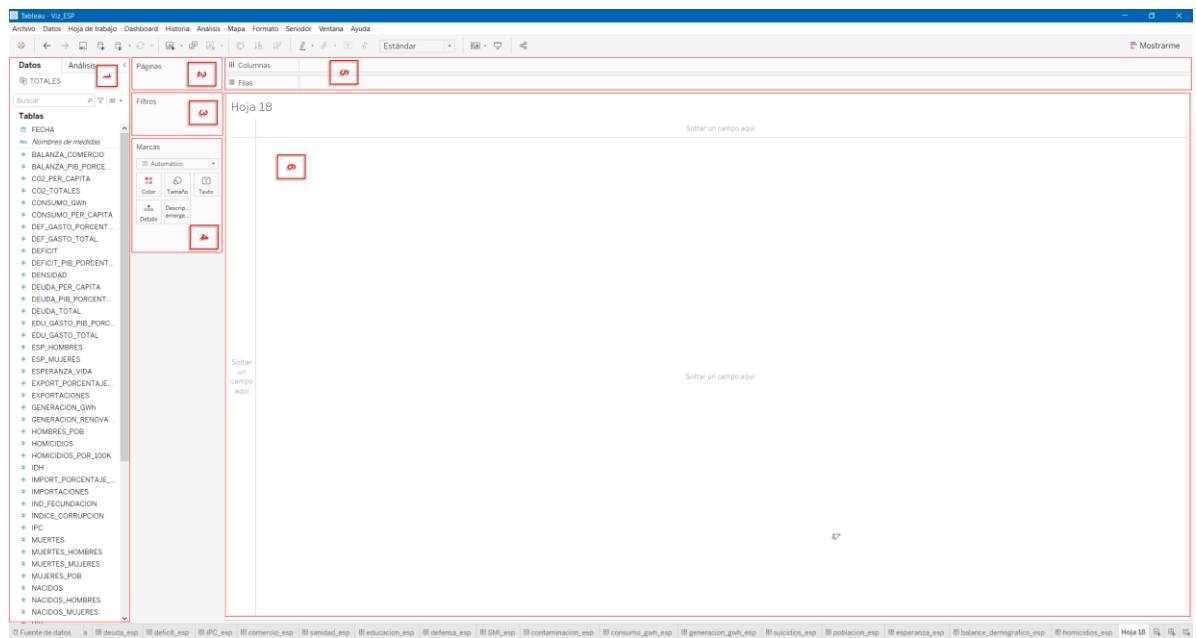
De esta forma se importan los datos y se puede comenzar a darles forma. Tableau siempre presenta una primera hoja referida a la fuente de datos, por si fuera necesario un tratamiento de estos, en este aspecto el software también es muy potente y nos muestra un gran número de opciones disponibles para el manejo de los datos, en la figura 13 se muestra un ejemplo de lo que podemos ver en esta hoja.

**Figura 13.** Ejemplo de hoja en Tableau para el tratamiento de datos

Fuente: Elaborada por el autor

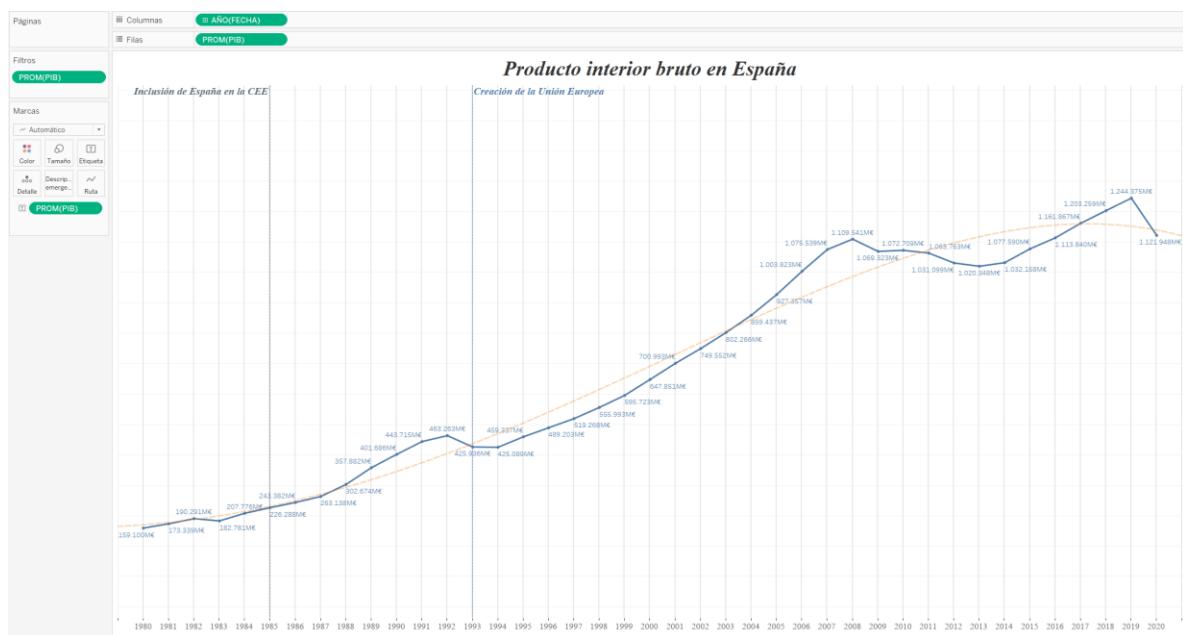
Cuando los datos ya se encuentran correctos, es momento de comenzar con la construcción de las visualizaciones. Para ello se crea una hoja nueva donde se tienen los siguientes parámetros que se muestran en la figura 14:

- Este recuadro muestra todas las columnas de las que se forma la tabla que se haya importado e indica también el tipo de dato del que se trata. En este caso las ya mencionadas anteriormente de TOTALES.csv
- En este apartado se incluyen las páginas que pueden formar la visualización, un ejemplo sería un año por página, y se puede utilizar para mostrar los datos para un año en concreto de la variación respecto al anterior. Se puede mover la barra interactiva para ir desplazándose entre los años.
- Se utiliza para añadir los filtros necesarios, por ejemplo, cuando se tienen datos de muchos países y solo quieres mostrar los de uno determinado.
- En este recuadro se colocan las variables de las que se quieren mostrar cierta información, es una propiedad fundamental para la visualización, añaden detalles y contexto a la vista.
- Como en cualquier gráfico normal, necesitas una variable en el eje X y otra variable en el eje Y, en Tableau este es el apartado donde se añaden esas variables.
- Vista donde se representa toda la visualización que se está construyendo.

**Figura 14.** Parámetros de una hoja de visualizaciones en Tableau

Fuente: Elaborada por el autor

A continuación, en la figura 15 se muestra un ejemplo de una de las visualizaciones del trabajo para comprobar cómo se construiría, el detalle se encuentra en el apartado de resultados más adelante.

**Figura 15.** Ejemplo de construcción de una visualización simple

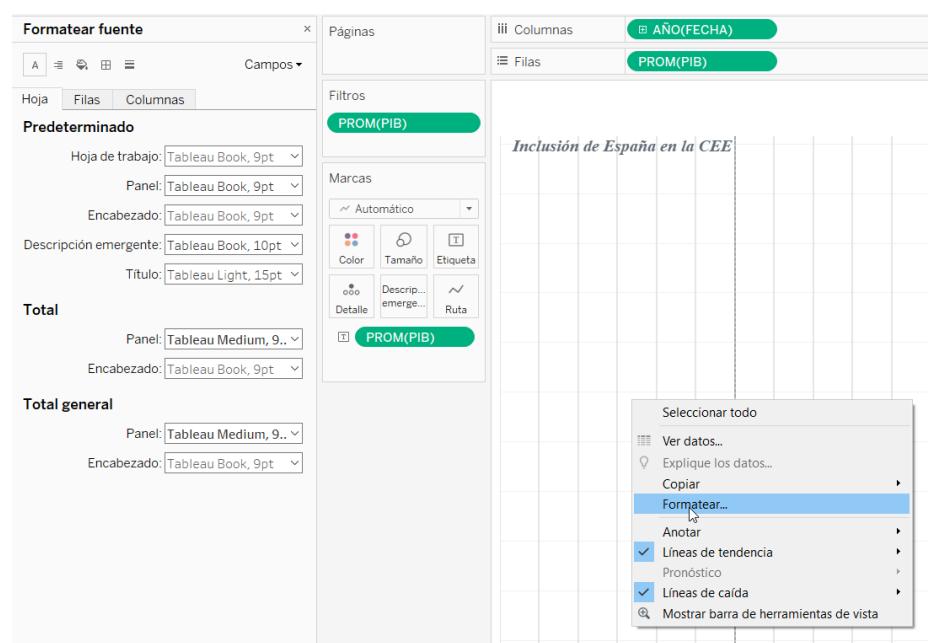
Fuente: Elaborada por el autor

Principalmente se puede comprobar que se tienen dos variables cuantitativas, la fecha en años, y el producto interior bruto de España en millones de euros. Se coloca la fecha en el eje X y el promedio del PIB en el eje Y. Tableau por defecto con estas variables mostrará un gráfico de líneas, pero esto se puede cambiar de forma sencilla en el apartado marcas que se vio anteriormente. Con este gráfico simple construido, se puede empezar a dar un formato concreto, esto es, colores, tamaños, etiquetas, añadir filtros, e incluso desde la pestaña ‘análisis’ se puede añadir líneas de tendencia o constantes.

También se pueden añadir lo que se conocen como ‘acciones’ muy útil a la hora de relacionar varios gráficos en un dashboard para que interactúen entre sus distintas hojas, de esta forma la información se vuelve más dinámica permitiendo filtrar ciertos datos de varias zonas distintas para ver los datos con mayor claridad.

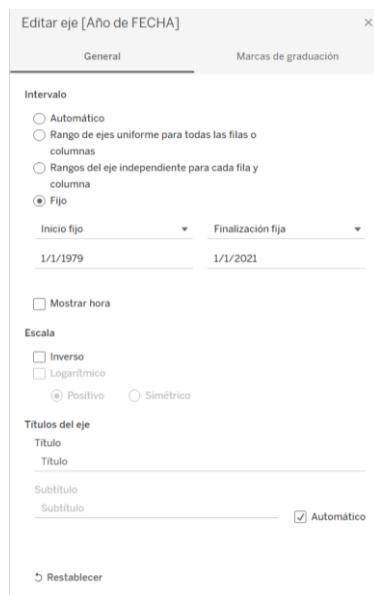
Para desarrollar más este ejemplo y poder explorar todas las opciones que ofrece Tableau, se tiene lo siguiente, desde el botón derecho del ratón sobre el panel principal de la vista, está la opción formatear, desde este menú que se abre a la izquierda se tienen todo tipo de propiedades que podemos aplicar sobre el panel, esto es, cada una de las fuentes de texto que forman la hoja, desde el título hasta las descripciones emergentes, concretando si se quiere modificar solo filas, columnas o todo. (Ver figura 16)

**Figura 16.** Formatear el panel en Tableau



Fuente: Elaborada por el autor

Al igual los ejes también son personalizables no solo en formato sino en intervalo y marcas de graduación. (Ver figura 17)

**Figura 17.** Editar los ejes en Tableau

Fuente: Elaborada por el autor

Una vez montado el dashboard definitivo es necesario guardarlo en Tableau Public, para ello primero se debe estar registrado y tener una cuenta disponible donde se van a subir de forma pública nuestras visualizaciones, esto se hace para después poder conseguir el script que nos permita acceder a ella desde el sitio web, de esta forma tener la representación actualizada en todo momento cada vez que se haga un cambio desde Tableau. Una vez se tenga la cuenta, desde la hoja de ‘fuente de datos’, se tienen que extraer los datos como se muestra en la figura 18, por tratarse de una cuenta gratuita. En caso contrario se puede utilizar Tableau Bridge que permite analizar la fuente de datos en tiempo real, esto permitiría actualizar las visualizaciones cada vez que se hace un cambio en los datos.

**Figura 18.** Opción para extraer los datos en Tableau

Conexión

En tiempo real  Extraer | [Editar](#) [Actualizar](#)

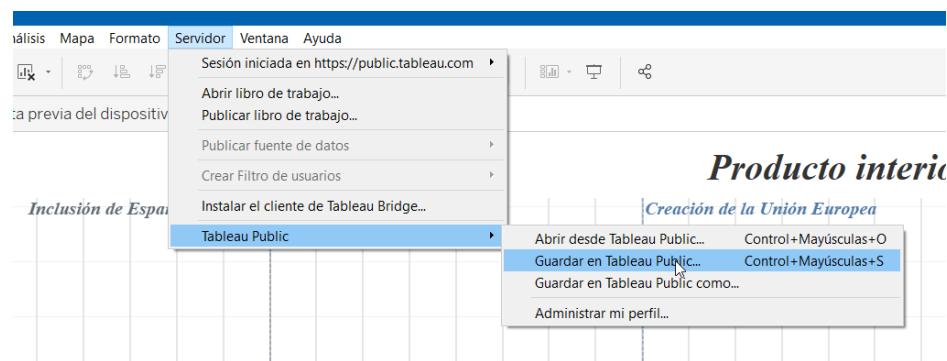
Filtros

0 | [Añadir](#)

La extracción incluye un subgrupo de datos. 17/01/2022 20:29:25

Fuente: Elaborada por el autor

Hecho esto se accede al Dashboard que se quiere publicar y se selecciona la siguiente opción, Servidor > Tableau Public > Guardar tal y como se muestra en la figura 19.

**Figura 19.** Opción para publicar en Tableau Public

Fuente: Elaborada por el autor

De esta forma ya se puede incrustar la visualización en el código HTML que se va a realizar en el siguiente bloque.

#### 4.2.2.1 Diseño, construcción y resultados

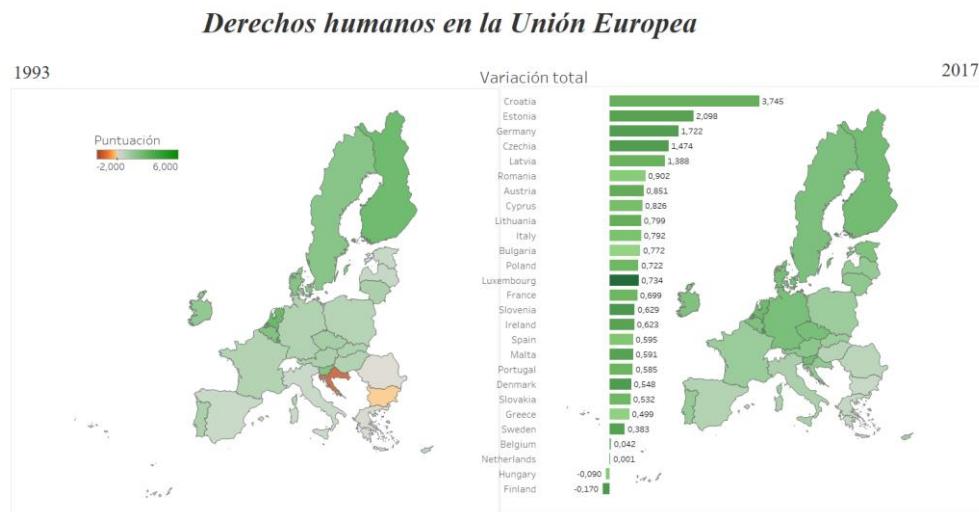
A continuación, se exponen los resultados del trabajo (de forma estática) sobre los temas escogidos para este estudio, para poder observar que visualizaciones se han decidido construir y poder replicarlas o mejorarlas en un futuro desde otro entorno, así poder extrapolarlo a otros países si fuera necesario. Se expone la figura junto con el proceso de diseño y construcción llevado a cabo.

Para la elaboración de las visualizaciones interactivas referentes a los derechos humanos se cuenta con el dato principal del índice de derechos humanos, el cual es una variable cuantitativa diferente en cada país y que cambia en el tiempo. Para su representación se ha decidido construir un total de 4 paneles informativos.

En la figura 20 se muestra el primer panel, cuyo objetivo es el de mostrar la posición física de cada país y la variación en el tiempo de este índice desde el año 1993 hasta el año 2017. Para ello se ha optado por un dashboard que reúna tres figuras, dos mapas coropléticos enfrentados el uno del otro, donde mediante el color (verde es un buen índice y rojo un mal índice) se representa la variable cuantitativa en cada año mostrado. Por último, un diagrama de barras horizontales central ordenado de mayor a menor que resume el cambio total sufrido por el índice para cada país entre los años mencionados. Para todos estos tipos de dashboards, ha sido muy importante la utilización de la opción 'acciones', por la que Tableau permite interactuar con cada uno de los gráficos que lo componen y filtrar por países o años pulsando sobre ellos, añadiendo de esta forma la interactividad que se propone. De esta forma

en este panel se puede consultar un país en concreto pulsando sobre él, y se visualizarán solo los datos referentes al mismo.

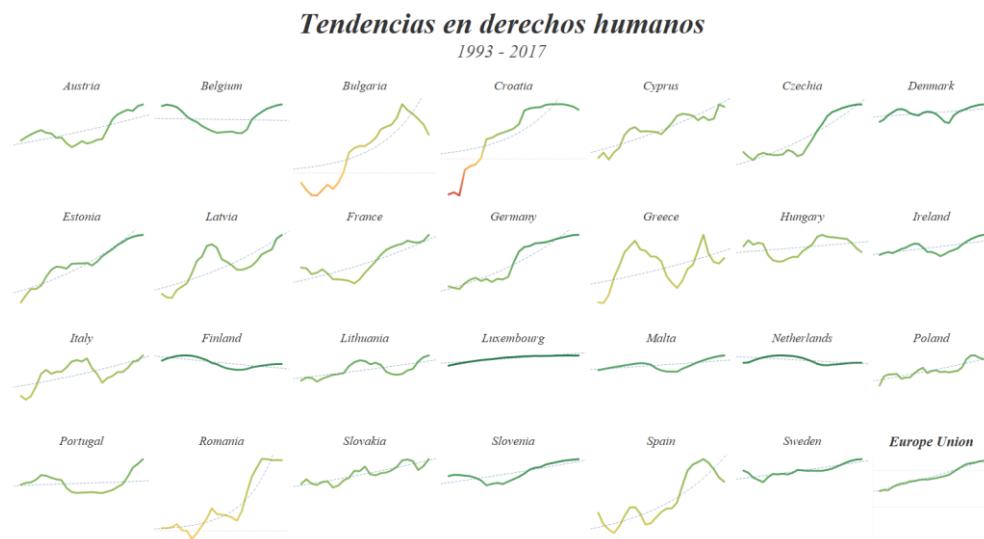
**Figura 20.** Derechos humanos en la Unión Europea



Fuente: Elaborada por el autor

Para observar mejor estas diferencias en el tiempo se propone el panel mostrado en la figura 21 como una infografía más intuitiva para el lector donde se perciben mejor las tendencias generales en la protección de los derechos humanos.

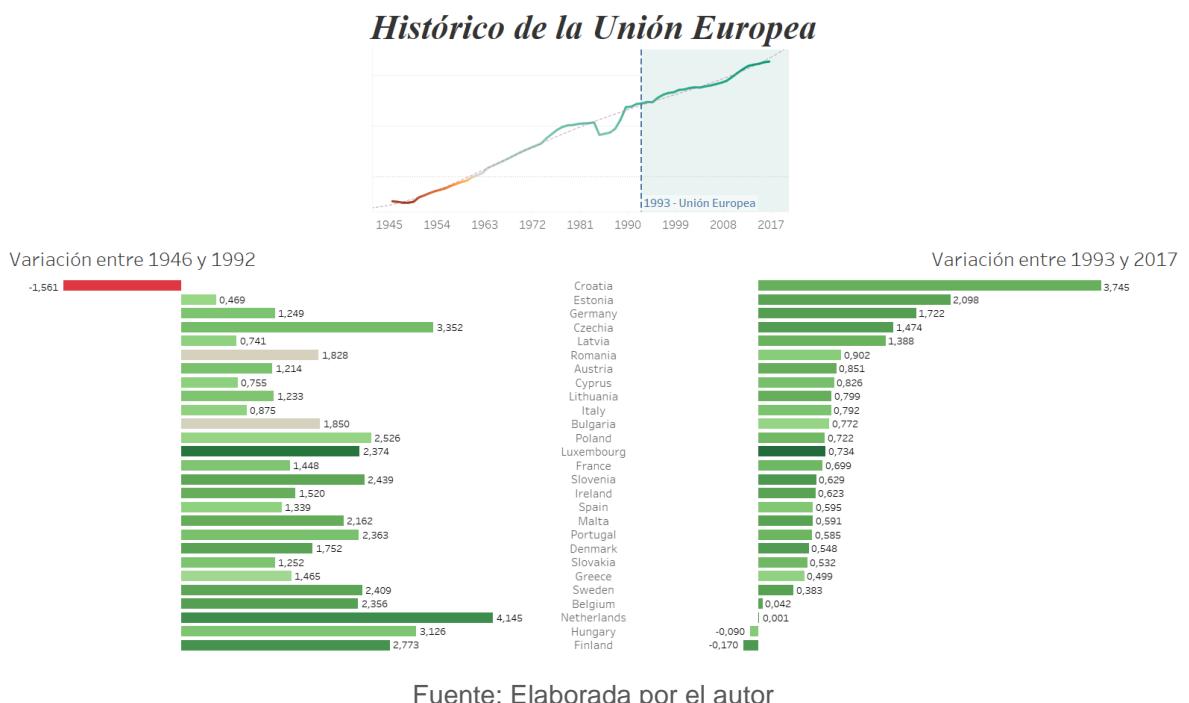
Se ha optado por un dashboard con un gráfico de líneas por cada país, en el que se aprecia el cambio del índice a lo largo de los años, desde 1993 hasta 2017, y al igual que en el anterior el color rojo indica peor valor, mientras que el verde muestra un mejor índice de los derechos humanos. Se ha decidido hacer así a parte de por estética, porque se aprecia mejor el cambio en el tiempo y permite una comparación mejor de la variación.

**Figura 21.** Tendencia de los derechos humanos en la UE

Fuente: Elaborada por el autor

En la figura 22 se muestra el tercer panel donde se ha decidido comprobar el histórico de este índice en la Unión Europea. Para ello se ha escogido posicionar 3 gráficos, un gráfico de líneas que es el que mejor muestra la evolución del índice de los derechos humanos desde el año 1946 hasta el año 2017, además se ha añadido la variable color que al igual que en los anteriores refleja de forma muy visual como es el impacto del valor del índice para esos años. Solo se deja el eje con los años ya que al ser interactivo el usuario se puede posicionar encima con el ratón y ver el valor concreto del dato, de esta forma se le quita carga visual al panel y lo hace más legible.

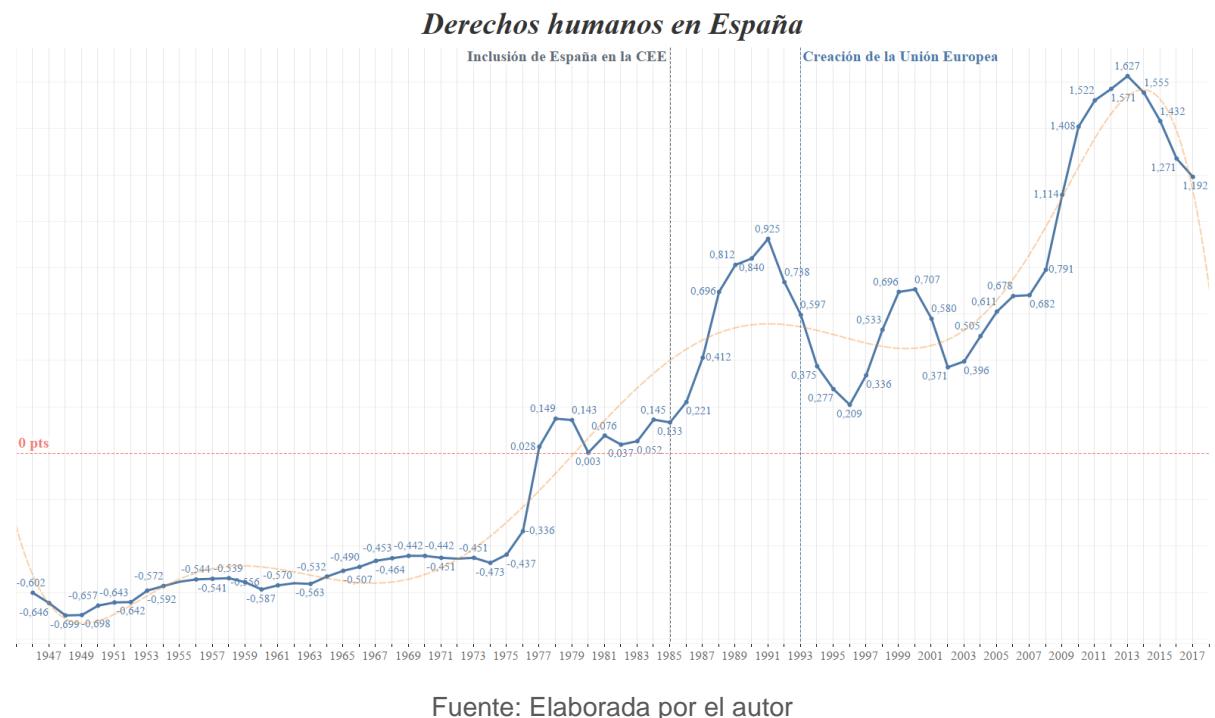
Seguidamente se añaden 2 diagramas de barras horizontales que muestran la misma información, pero para 2 intervalos de años diferentes. Esto es la variación total por país del índice entre los años indicados, junto con el color que indica el valor del índice al final del intervalo. De igual forma se puede pulsar sobre uno de los países para interactuar con los datos y visualizar únicamente su evolución y diferencias.

**Figura 22.** Histórico de los derechos humanos en la UE

Fuente: Elaborada por el autor

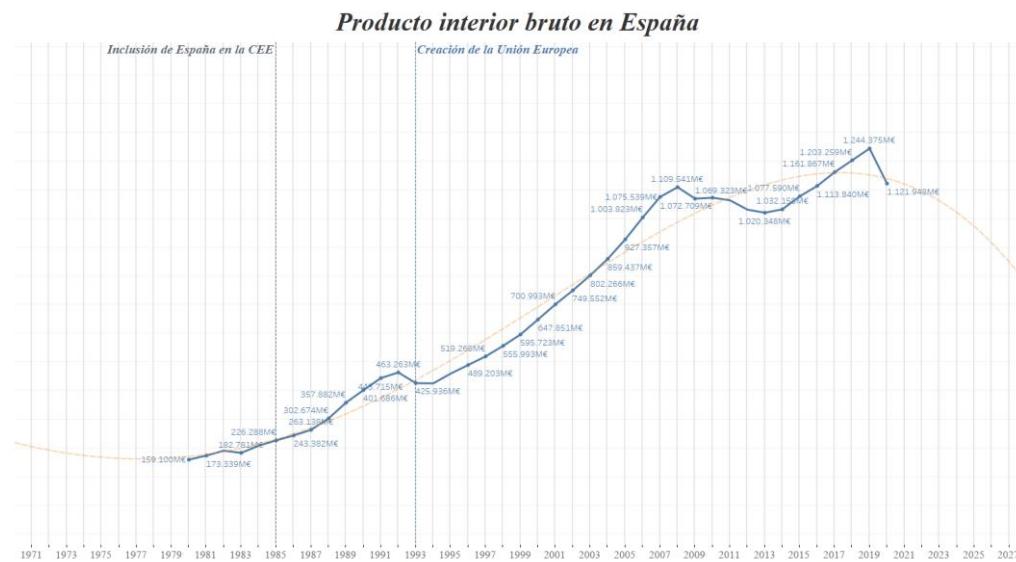
Por último, se expone en la figura 23 el dato para el país en concreto sobre el que se realiza el estudio, en este caso España. Para ello se opta por un formato, ya explicado anteriormente y que se va a repetir muy parecido para las gráficas similares en todo el trabajo como parte de la metodología cuando se pretende mostrar una variable cuantitativa para el país concreto y su evolución en el tiempo.

Consiste en un gráfico de líneas donde se aprecia la variación en el tiempo del índice de los derechos humanos. Únicamente se visualiza el eje de los años y serán las etiquetas las que indiquen el dato concreto que se coloque en el eje Y, en este caso el índice de derechos humanos. Además, se establecen 3 líneas constantes que marcan el año de la inclusión del país en la UE, la creación oficial de la UE y una línea horizontal marcando los 0 puntos en el índice que servirá de guía visual, junto con una línea de tendencia en color amarillo que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 23.** Evolución de los derechos humanos en España

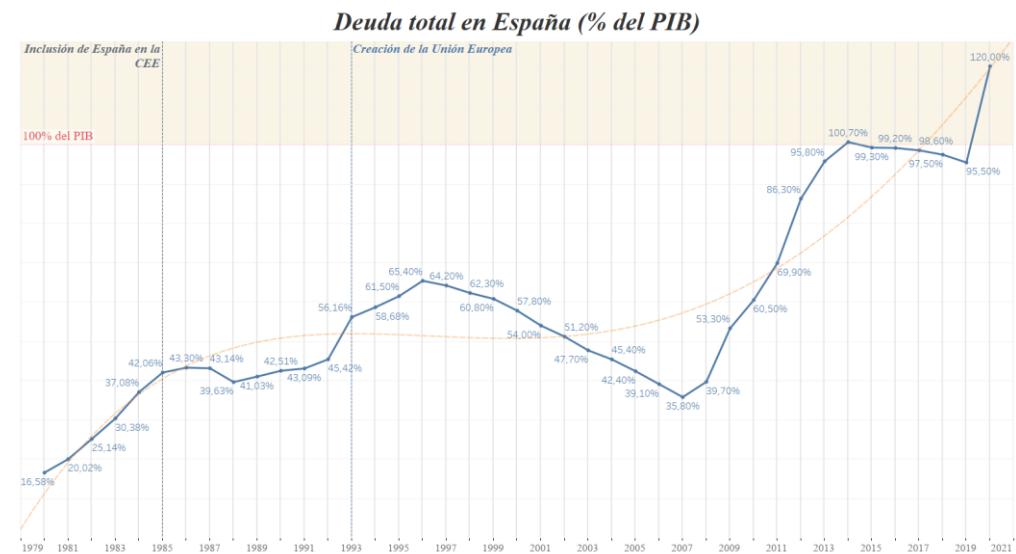
Para la siguiente sección sobre datos económicos del país se ha decidido comenzar con 5 paneles informativos.

En la figura 24 se muestra la evolución del producto interior bruto del país mediante el formato ya visto anteriormente por el que se ha optado utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con el valor concreto para cada año, además se añaden las líneas constantes que marcan el año de la inclusión del país en la UE y la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 24.** Evolución del producto interior bruto en España

Fuente: Elaborada por el autor

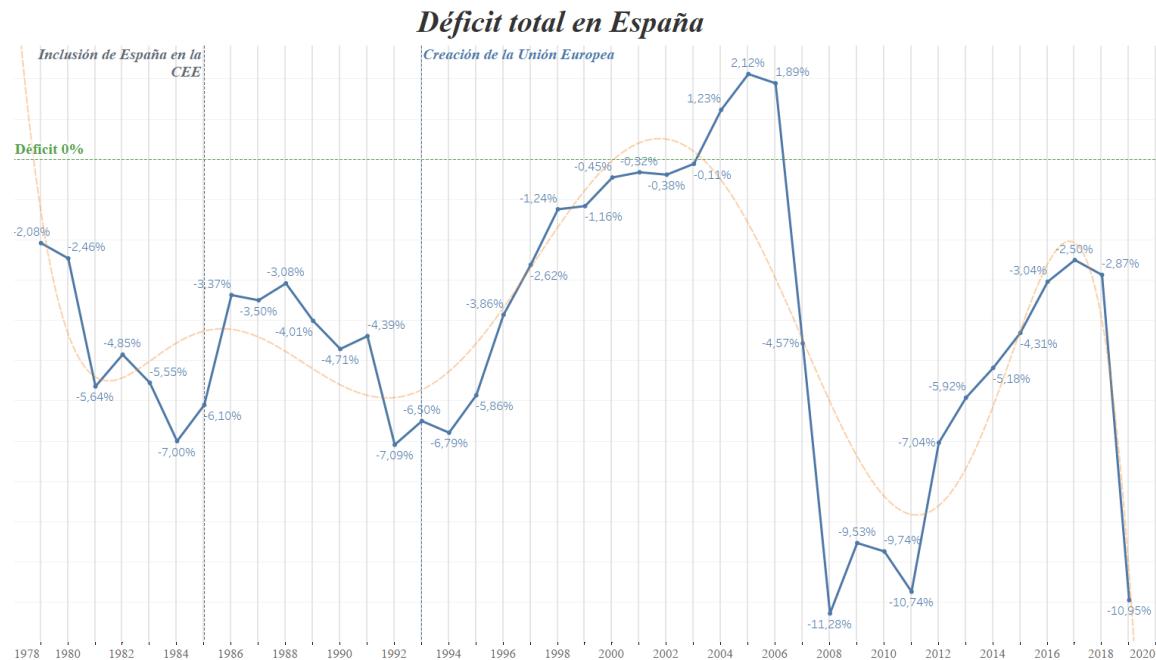
En la figura 25 se muestra la evolución de la deuda del país respecto al PIB en el que se ha optado utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con el valor concreto para cada año, y se añaden las líneas constantes que marcan el año de la inclusión del país en la UE y la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones y otra línea constante que marca el 100% de la deuda total respecto al PIB.

**Figura 25.** Evolución de la deuda total en España

Fuente: Elaborada por el autor

En la figura 26 se muestra la evolución del déficit del país respecto al PIB en el que se ha optado utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con el valor de déficit para cada año, y se añaden las líneas constantes que marcan el año de la inclusión del país en la UE y la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones y otra línea constante que marca el 0% de déficit respecto al PIB.

**Figura 26. Evolución del déficit total en España**



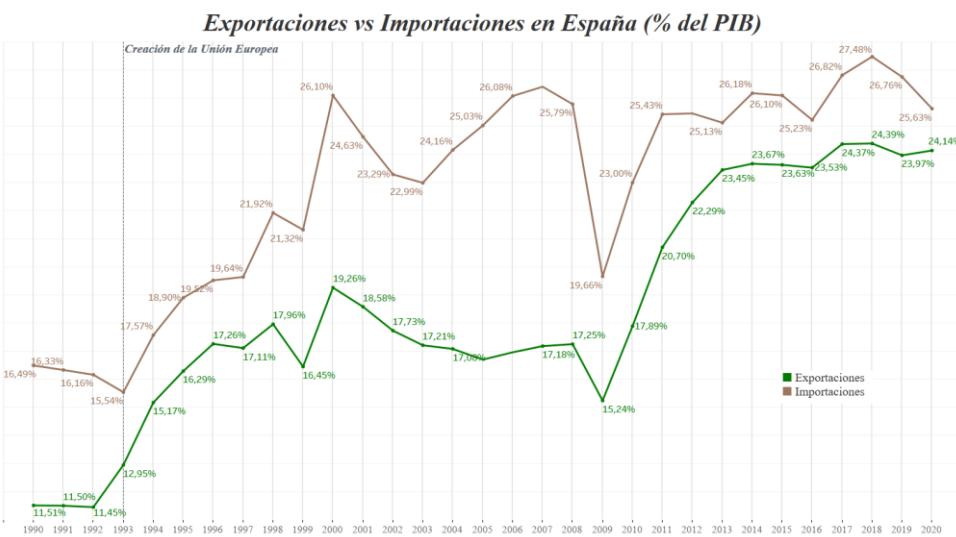
Fuente: Elaborada por el autor

En la figura 27 se muestra la evolución del índice de precios al consumo del país respecto al PIB en el que se ha optado utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con el valor concreto para cada año, y se añaden las líneas constantes que marcan el año de la inclusión del país en la UE y la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones y otra línea constante que marca el 0% de IPC respecto al PIB.

**Figura 27.** Evolución del índice de precios al consumo en España

Fuente: Elaborada por el autor

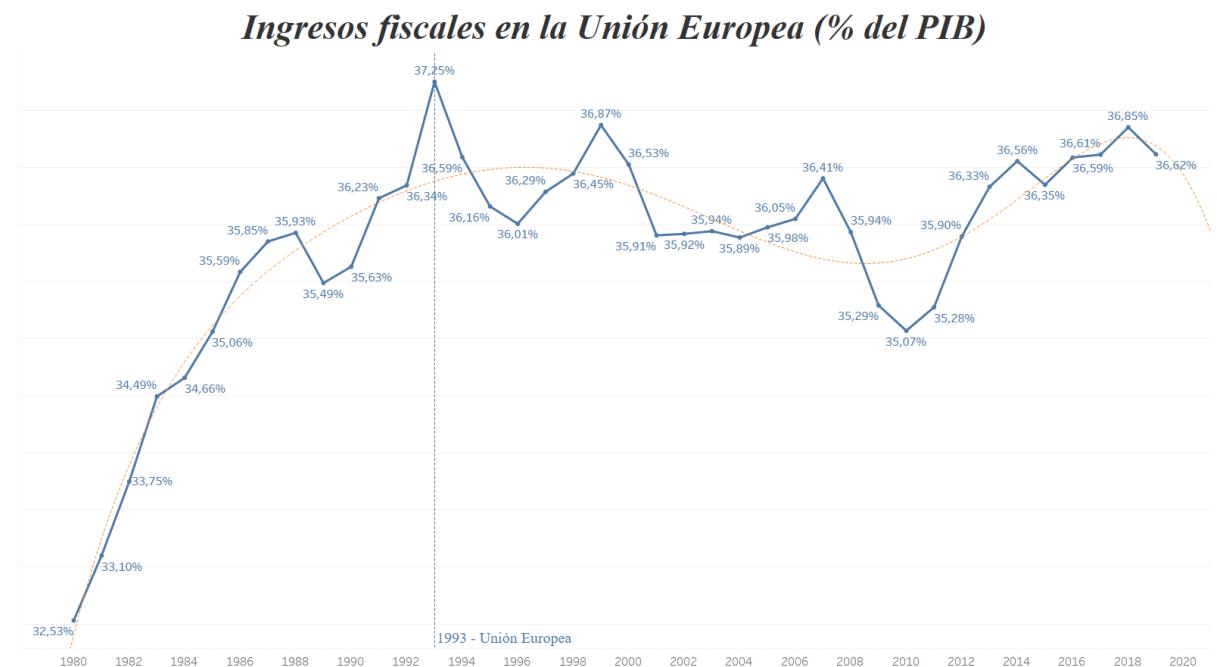
En la figura 28 se muestran las exportaciones e importaciones del país en porcentaje respecto al PIB. Se trata de 2 variables cuantitativas que se mueven en el mismo rango de valores y representan el mismo dominio, por tanto, se ha decidido mostrarlas juntas en un eje doble mediante un gráfico de líneas, de esta forma se puede apreciar fácilmente de forma visual cómo evolucionan y compararlas de forma sencilla. Los colores verde y marrón representan la categoría a la que pertenecen y se define en la leyenda. Las etiquetas indican el valor concreto de cada variable, para cada año.

**Figura 28.** Balanza del comercio internacional en España

Fuente: Elaborada por el autor

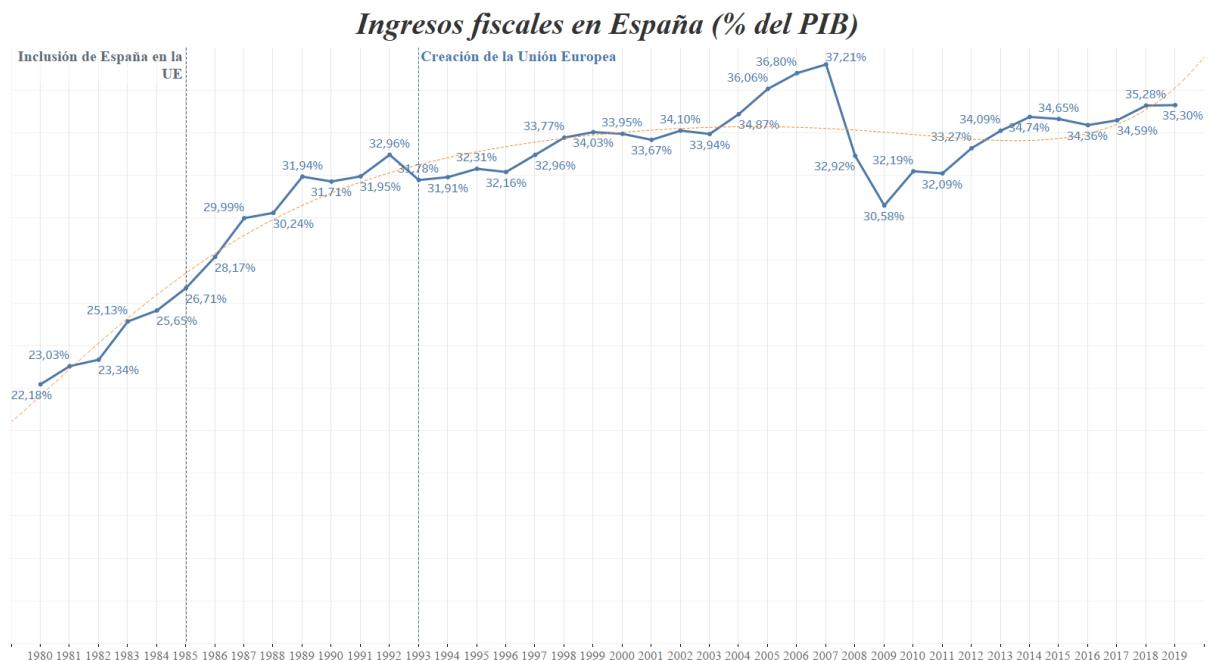
En la figura 29 se muestra la evolución de los ingresos fiscales en la Unión Europea respecto al PIB total en el que se ha optado utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con el valor concreto para cada año, y se añaden las líneas constantes que marcan el año de la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 29.** Evolución de los ingresos fiscales promedio en la Unión Europea



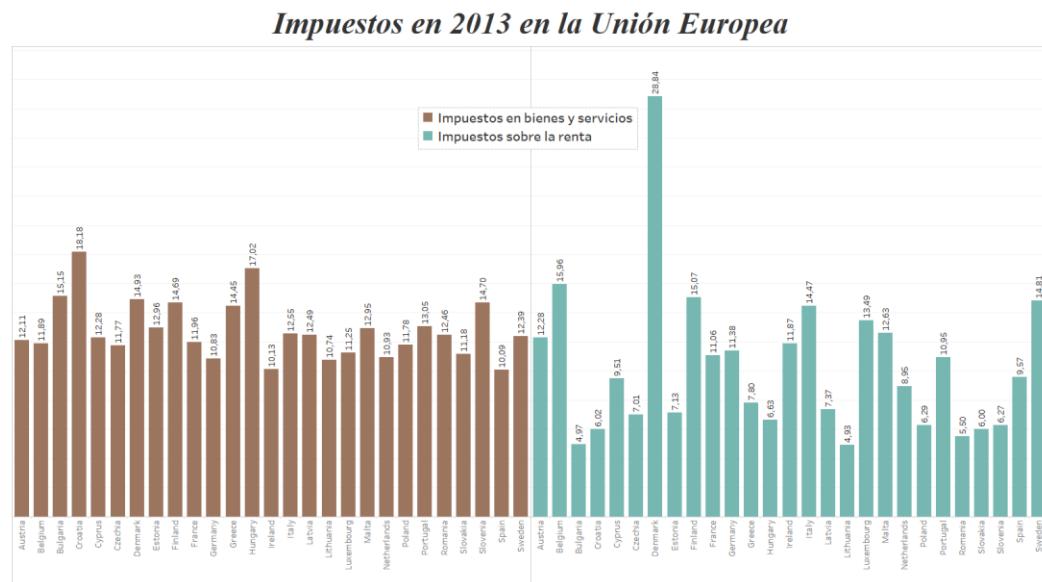
Fuente: Elaborada por el autor

En la figura 30 se muestra la evolución de los ingresos fiscales del país respecto al PIB en el que se ha optado utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con el valor de los ingresos como porcentaje del PIB en cada año para España, y se añaden las líneas constantes que marcan el año de la inclusión del país en la UE y la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 30.** Evolución de los ingresos fiscales en España

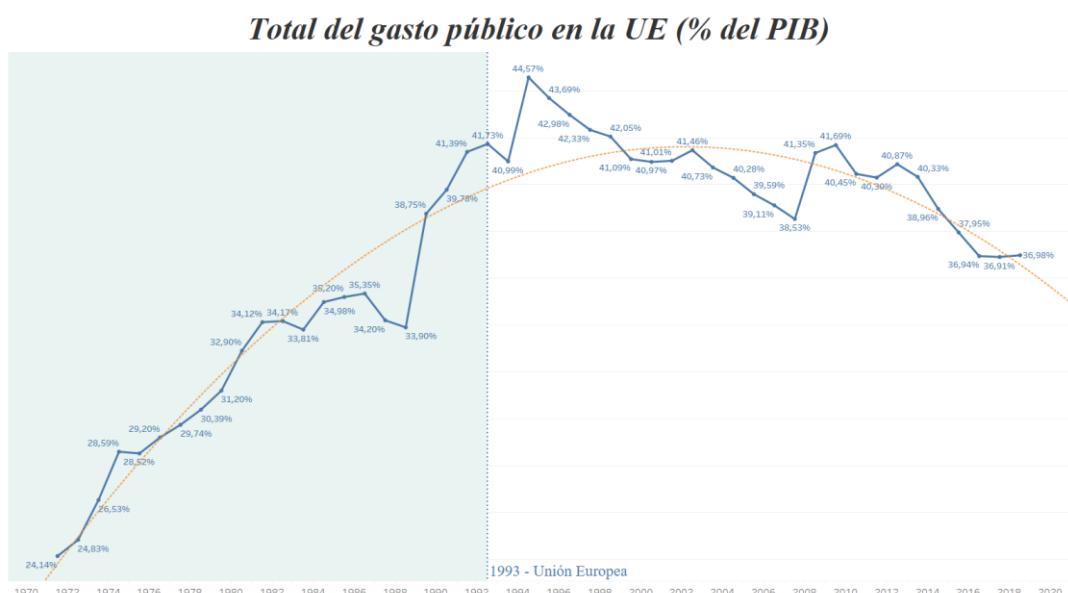
Fuente: Elaborada por el autor

Para representar en la figura 31 los impuestos de cada país de la Unión Europea en el año del que se tienen datos se ha decidido realizar un diagrama de barras verticales, que permite realizar una comparación rápida e inequívoca del valor de las variables cuantitativas para cada país. Se representan 2 categorías, los impuestos de bienes y servicios, y los impuestos sobre la renta, se tiene un color para cada una que hace más sencillo distinguirlas y añade algo de estética a la visualización, lo que la hace más interesante. Además, pulsando sobre cualquiera de las barras el usuario puede interactuar con el panel y filtrar los datos por país para facilitar las comparaciones entre ellos.

**Figura 31.** Distribución de los impuestos por país de la UE

Fuente: Elaborada por el autor

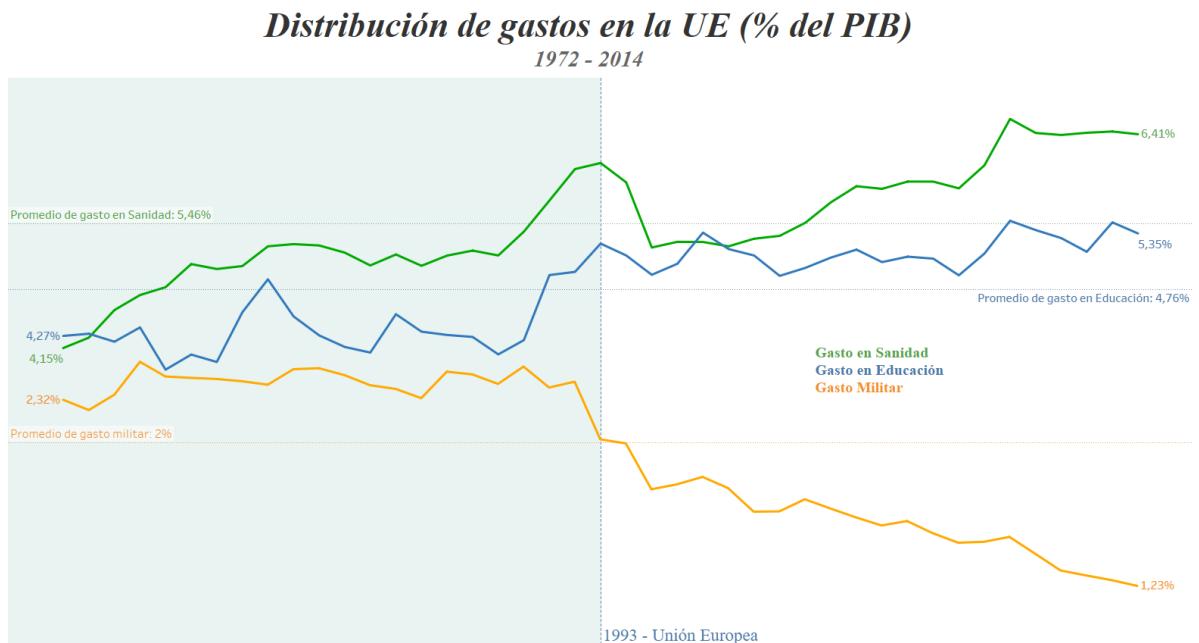
En la figura 32 se muestra la evolución del gasto público en la Unión Europea respecto al PIB total en el que se ha optado utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con su valor concreto, y se añaden las líneas constantes que marcan el año de la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 32.** Evolución del gasto público promedio en la Unión Europea

Fuente: Elaborada por el autor

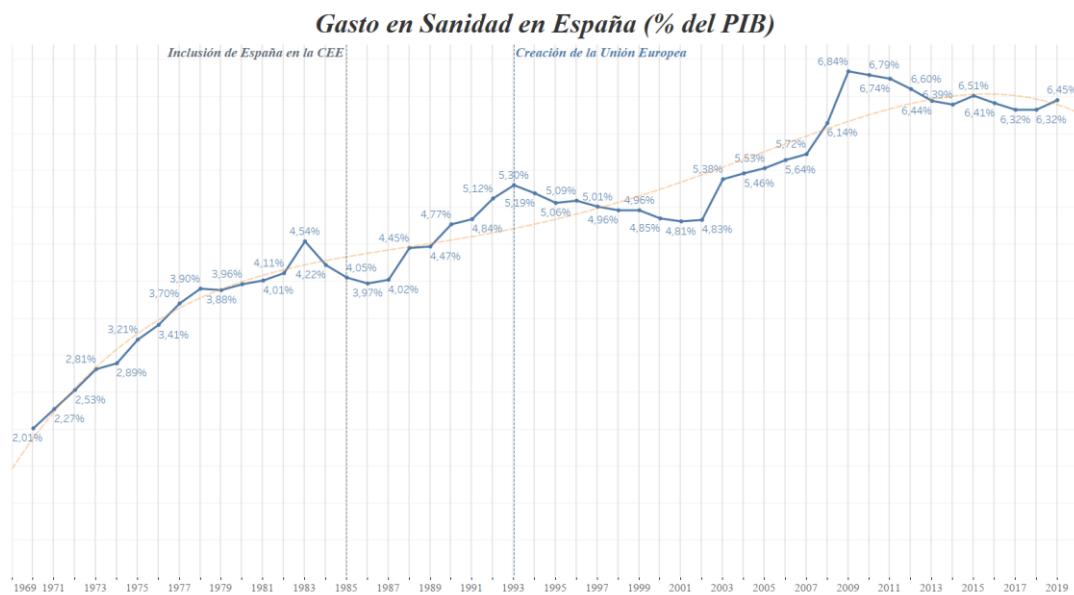
En la figura 33 se representa la distribución de gasto en los países de la UE, se ha decidido establecer un mismo panel con las 3 variables cuantitativas que son el gasto en sanidad, gasto en educación y gasto militar, de tal forma que mediante un gráfico de líneas se pueda comparar de forma ágil y sencilla como ha sido la distribución de estos gastos y su evolución en el tiempo. Para favorecer la visualización cada una de las categorías se representa de un color diferente y con contraste, que se refleja en la leyenda.

**Figura 33.** Distribución del gasto público promedio en la Unión Europea



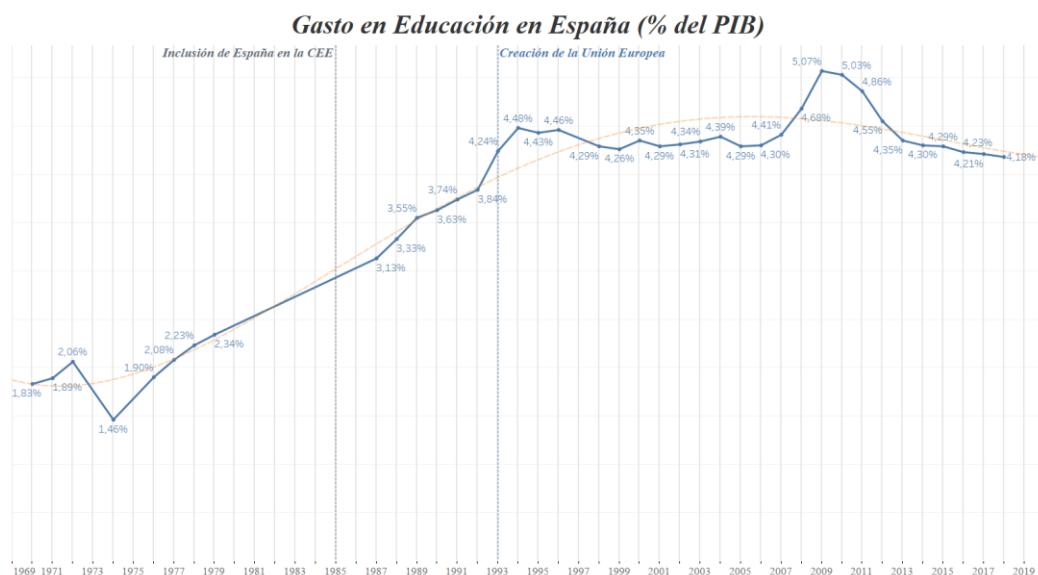
Fuente: Elaborada por el autor

En la figura 34 se muestra la evolución del gasto en sanidad del país respecto al PIB en el que se ha optado utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con su valor concreto por año, y se añaden las líneas constantes que marcan el año de la inclusión del país en la UE y la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 34.** Gasto en Sanidad en España

Fuente: Elaborada por el autor

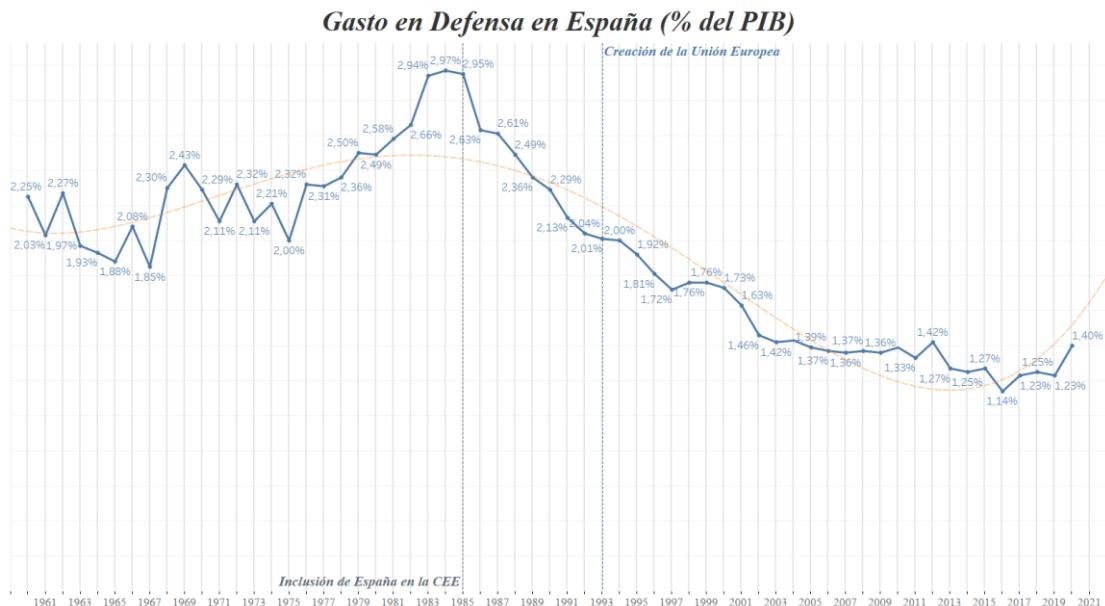
En la figura 35 se muestra la evolución del gasto en educación del país respecto al PIB en el que se ha optado utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con su valor concreto, y se añaden las líneas constantes que marcan el año de la inclusión del país en la UE y la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 35.** Gasto en educación en España

Fuente: Elaborada por el autor

En la figura 36 se muestra la evolución del gasto militar del país respecto al PIB en el que se ha optado utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con su valor concreto, y se añaden las líneas constantes que marcan el año de la inclusión del país en la UE y la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 36.** Gasto en defensa en España

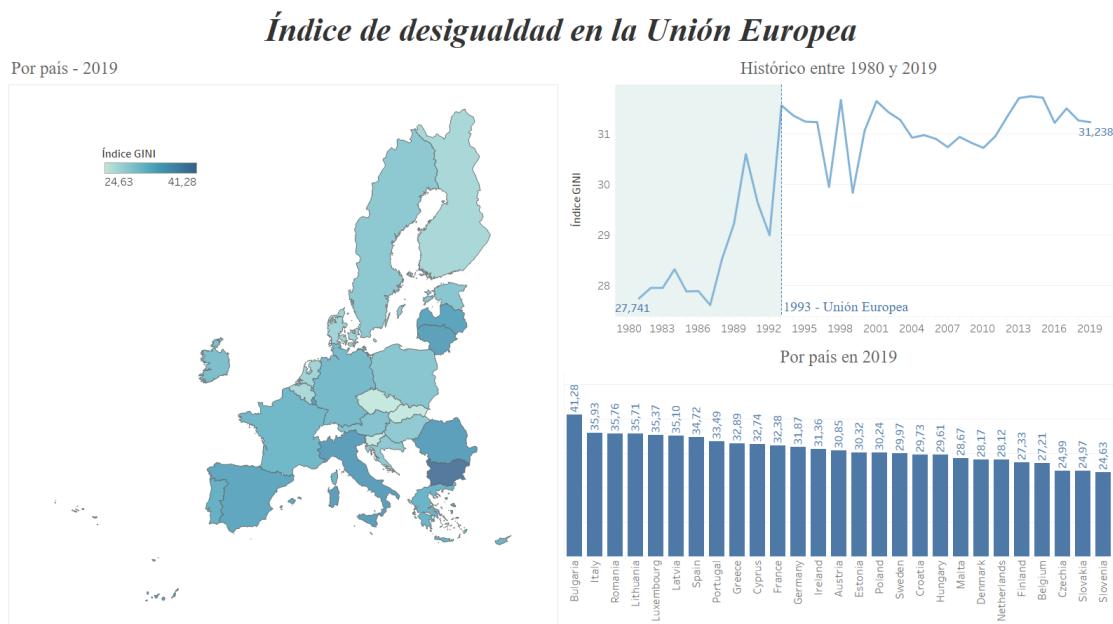


Fuente: Elaborada por el autor

En la figura 37 se muestran los datos del índice de desigualdad GINI en la Unión Europea, para ello se ha decidido mostrar un panel con 3 gráficos diferentes. En uno de ellos se ha optado por un mapa coroplético que ayude al usuario a posicionar el país que este visualizando de forma interactiva y así sepa situarlo en el globo, en color azul se representa el valor de la variable cuantitativa, en este caso el índice GINI, en el que más oscuro significa mayor desigualdad.

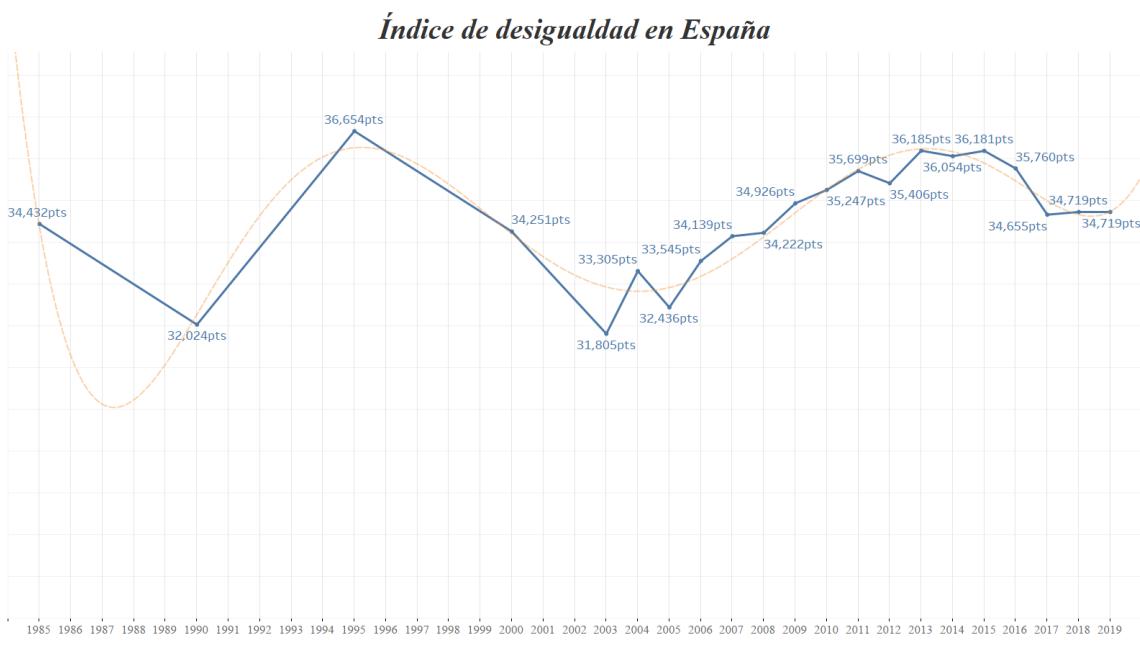
Por otro lado, se ha construido un gráfico de líneas por ser el que mejor representa la evolución del índice a lo largo del tiempo. Cada uno de los gráficos es interactivo con los demás y pulsando sobre cualquiera de ellos se obtienen opciones de filtrado, por ejemplo, por país o por año, todos los demás cambian dinámicamente.

Por último, un diagrama de barras verticales que permite realizar una rápida comparación de este valor de desigualdad entre los países de la UE.

**Figura 37.** Índice de desigualdad GINI en la Unión Europea

Fuente: Elaborada por el autor

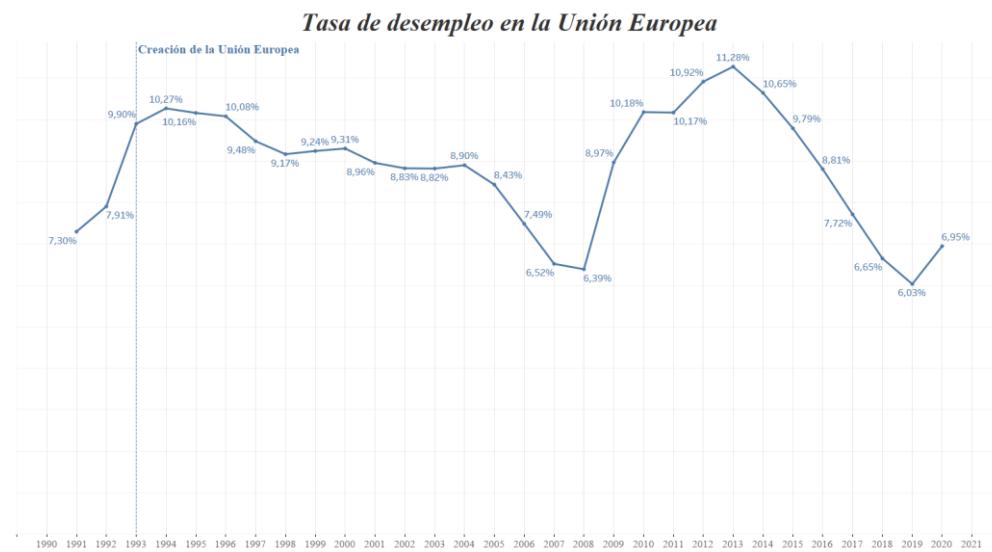
En la figura 38 se muestra la evolución del índice GINI del país en el que se ha optado por utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con su valor concreto, junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 38.** Índice de desigualdad GINI en España

Fuente: Elaborada por el autor

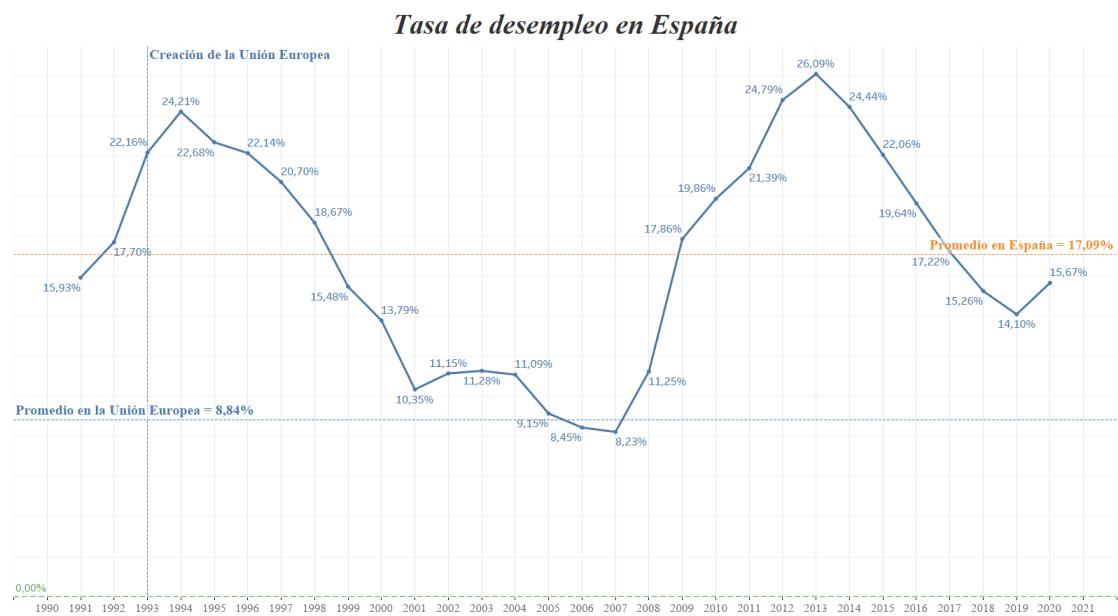
En la figura 39 se muestra la evolución de la tasa de desempleo en la Unión Europea en el que se ha optado utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con su valor concreto, y se añade la línea constante que marca el año de la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 39.** Tasa de desempleo en la Unión Europea



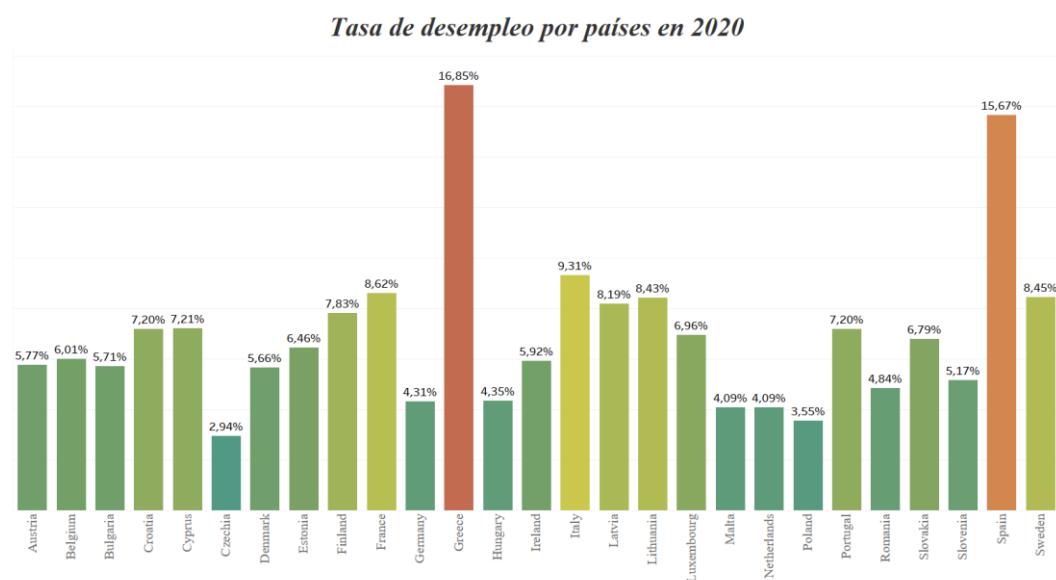
Fuente: Elaborada por el autor

En la figura 40 se muestra la evolución de la tasa de desempleo del país en el que se ha optado utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con su valor concreto, y se añaden las líneas constantes que marcan el año de la creación oficial de UE junto con una línea constante horizontal que marca el promedio del país y otra línea constante horizontal que marca el promedio europeo, esto permite realizar una mejor valoración de la situación real del país respecto a sus vecinos europeos, además una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 40.** Tasa de desempleo en España

Fuente: Elaborada por el autor

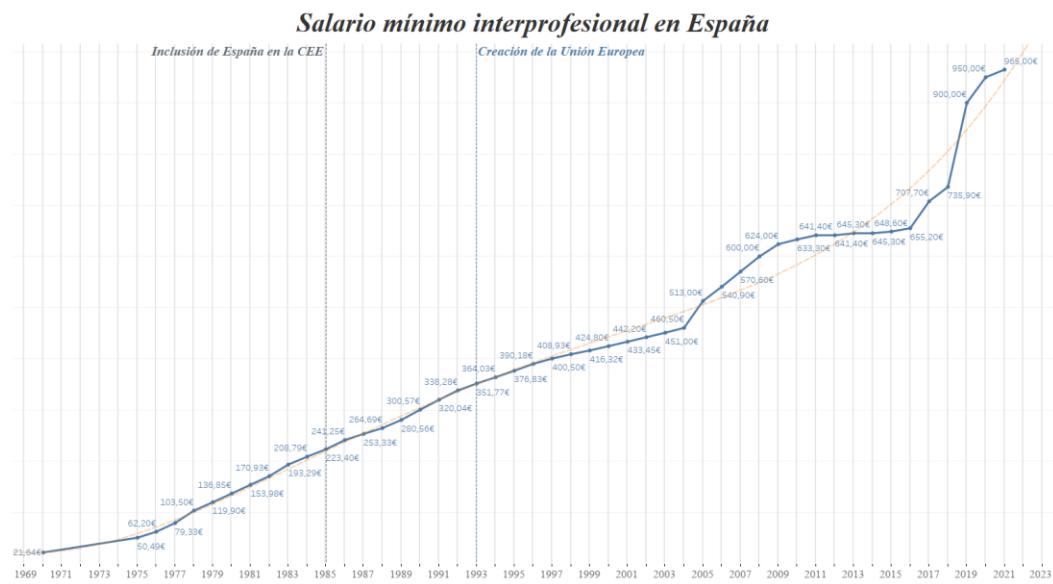
En la figura 41 se representa la tasa de desempleo por país más actual para un año concreto, se ha escogido un diagrama de barras verticales por ser el gráfico más adecuado para poder comparar la variable cuantitativa en cada país de forma clara y sencilla. El color que va en gradiente desde el verde al rojo representa la misma variable de forma más visual y clara al lector (verde es una buena tasa y rojo una mala tasa de desempleo frente a la media).

**Figura 41.** Tasa de desempleo en los países de la UE en el año 2020

Fuente: Elaborada por el autor

En la figura 42 se muestra la evolución del salario mínimo interprofesional del país en el que se ha optado utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con su valor concreto, y se añaden las líneas constantes verticales que marcan el año de la inclusión del país en la UE y la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

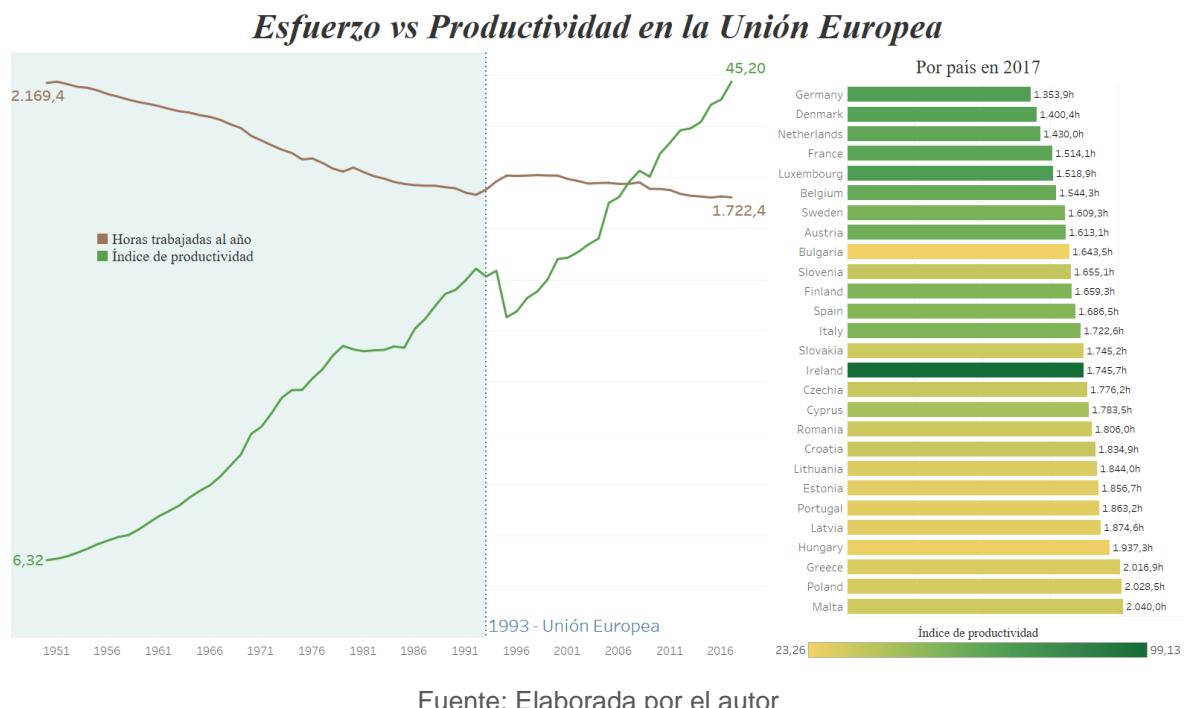
**Figura 42.** Evolución del salario mínimo interprofesional en España



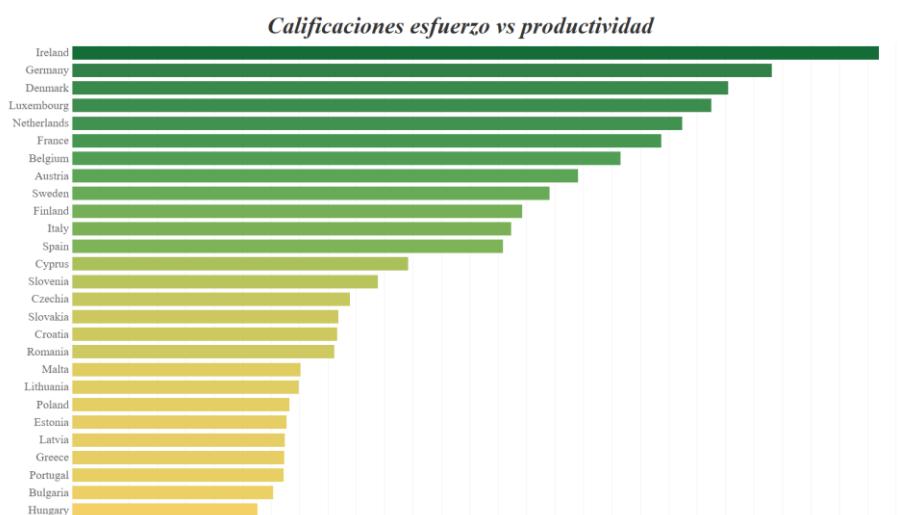
Fuente: Elaborada por el autor

En la figura 43 se muestra mediante 2 gráficas una comparativa del esfuerzo contra la productividad de los países de la Unión Europea. Para el primero se ha decidido realizar un gráfico de líneas de eje doble ya que se representan dos variables cuantitativas, horas trabajadas al año, e índice de productividad. De esta forma se puede apreciar la evolución de las variables a lo largo del tiempo, para distinguirla correctamente se han escogido 2 colores con un buen contraste entre sí.

Seguidamente se ha añadido otro gráfico esta vez de barras horizontales ordenado de menor a mayor en cuanto a horas trabajadas, de esta forma es más visual la comparación entre los países, el color representa el índice de productividad (amarillo significa poca productividad, mientras que verde oscuro es una buena productividad). Este panel permite la interactividad mediante la cual pulsando sobre uno de los países, el resto de gráficos se filtran mostrando únicamente los resultados concretos referentes al mismo.

**Figura 43.** Esfuerzo vs Productividad en la Unión Europea

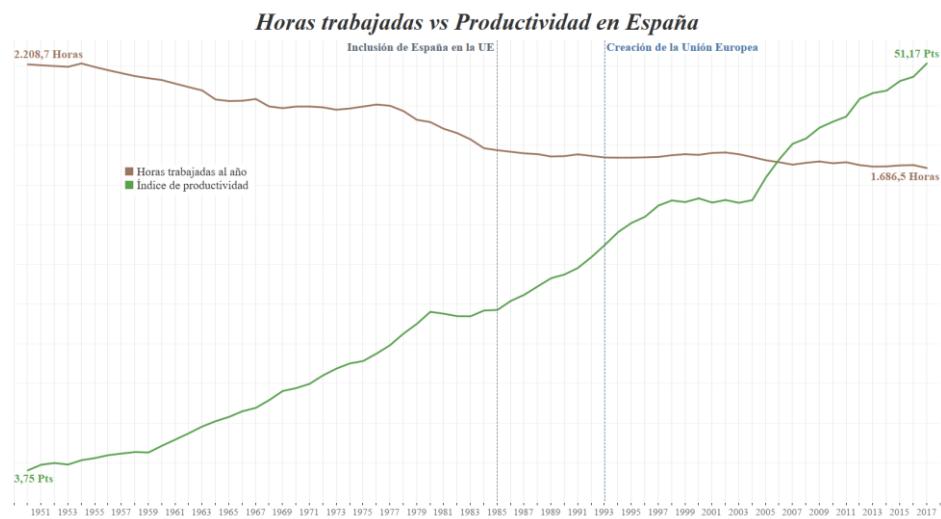
En la figura 44 se representa una calificación del nivel de productividad en función de su nivel de esfuerzo, por país. Para reforzar la extracción de conclusiones del anterior panel. Y para ello se ha optado por construir un diagrama de barras horizontales que permite realizar una comparación sencilla y visual de esta variable cuantitativa calculada. Además, el color representa la misma variable y acompaña a la vista del usuario para una mejor lectura.

**Figura 44.** Puntuaciones en la relación esfuerzo y productividad en la UE

Fuente: Elaborada por el autor

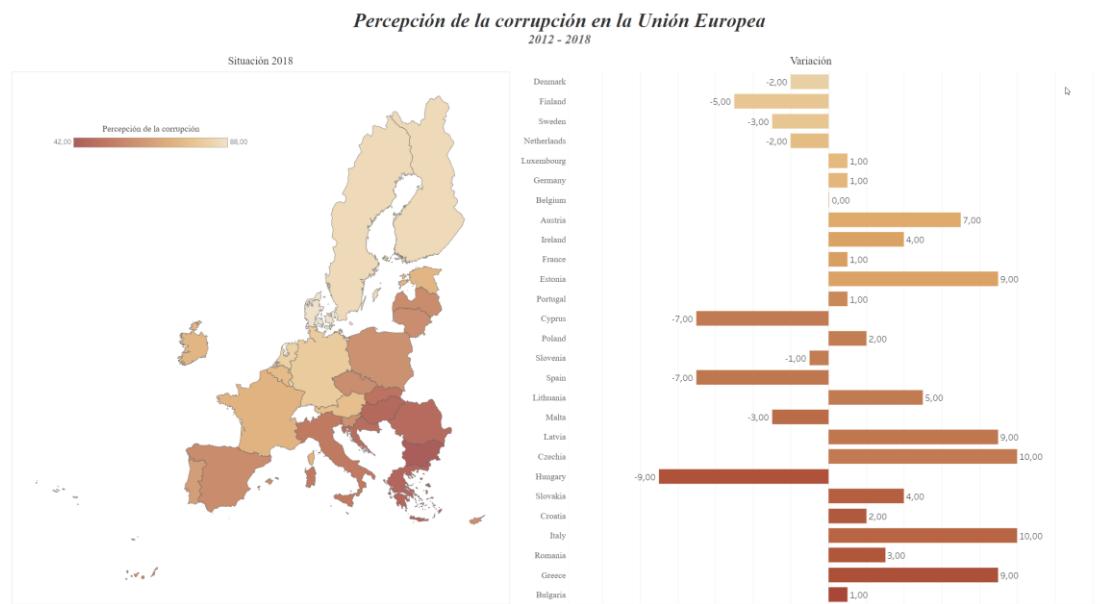
Al igual que en el anterior caso, en la figura 45 se representa una comparativa del esfuerzo contra la productividad del caso concreto del país, se utiliza un gráfico de líneas de eje doble ya que se representan dos variables cuantitativas, horas trabajadas al año, e índice de productividad. De esta forma se puede apreciar la evolución de las variables a lo largo del tiempo, para distinguirla correctamente se han escogido 2 colores con un buen contraste entre sí. Además, se añaden las líneas constantes verticales que marcan el año de la inclusión del país en la UE y la creación oficial de UE.

**Figura 45. Esfuerzo vs Productividad en España**



Fuente: Elaborada por el autor

Para la figura 46 que muestra la percepción de la corrupción en la Unión Europea se ha decidido construir un panel informativo con 2 gráficos diferentes. Un mapa coroplético para que el usuario pueda situar físicamente el país de estudio, y cada país tiene un color que representa la variable cuantitativa, en este caso, cuanto más oscuro, peor es la percepción de la corrupción. El otro gráfico, esta vez de barras horizontales, representa la variación por país de este índice entre los años indicados, ordenado de menor a mayor variación. Al igual que en el anterior, el color representa también la variable de percepción de la corrupción. Este panel permite la interacción del usuario, pulsando sobre cualquiera de los países y filtrando los datos concretos referentes a cada uno.

**Figura 46.** Percepción de la corrupción en la Unión Europea

Fuente: Elaborada por el autor

En la figura 47 se muestra la evolución del índice de percepción de la corrupción del país, en el que se ha optado utilizar un gráfico de líneas que representa de la mejor forma la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con su valor concreto, y se añade una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 47.** Evolución del índice de corrupción en España

Fuente: Elaborada por el autor

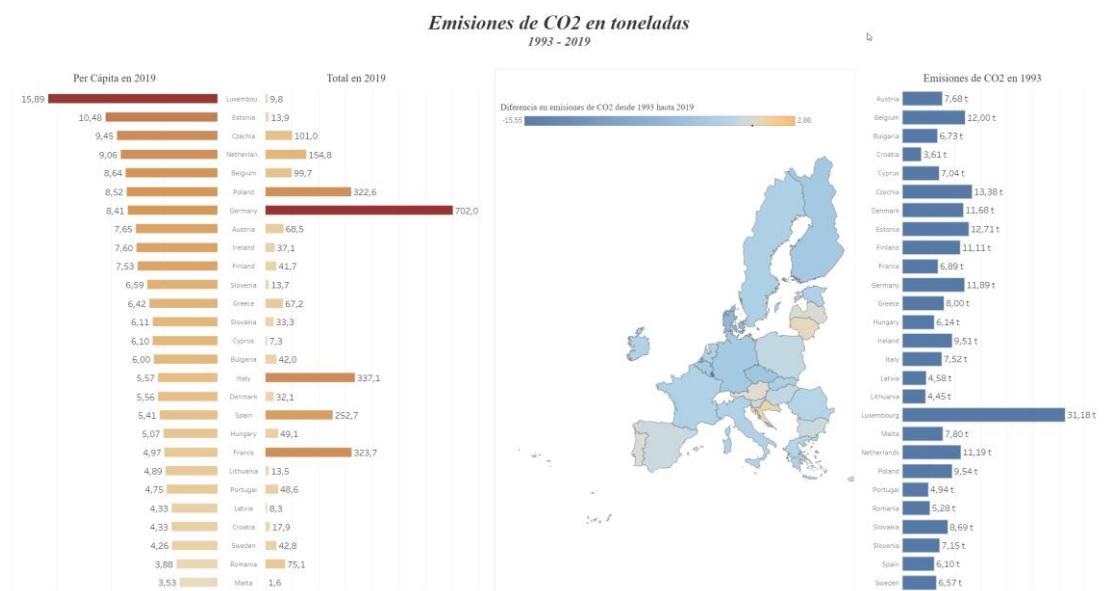
En la figura 48 se muestra un panel informativo con los datos de emisiones de CO<sub>2</sub> en la Unión Europea, la variable cuantitativa son las toneladas de CO<sub>2</sub> emitidas desde el año 1993 hasta el 2019. Para representar estos datos se han decidido construir 4 gráficos en un mismo panel.

Los 2 primeros gráficos de la izquierda son diagramas de barras horizontales por ser los que mejor permiten comparar las 2 variables cuantitativas para cada país, las emisiones per cápita y las emisiones totales. Además, colocándose enfrentados en contraposición, el usuario puede extraer conclusiones más realistas de los datos.

Por otro lado, se muestra un mapa coroplético para que el usuario pueda situar físicamente el país objetivo en el mapa, junto con colores que indican de forma visual y fácil como ha sido su evolución en emisión de gases.

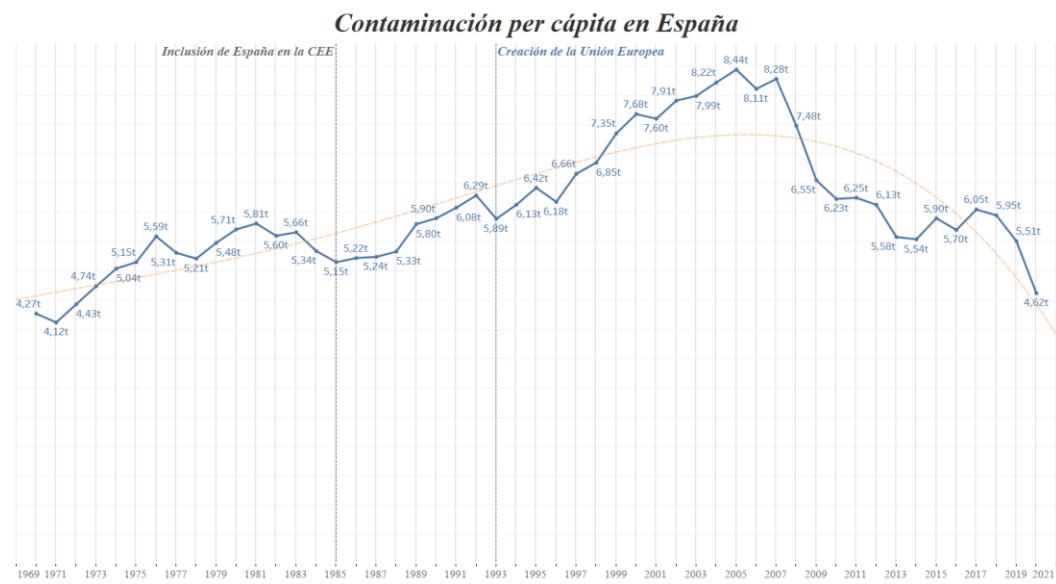
Por último, se ha elegido otro diagrama de barras horizontal por las mismas razones que el anterior, es el más adecuado para poder realizar comparaciones entre países de forma rápida e inequívoca, esta vez con los datos de emisiones en 1993 para cada país.

**Figura 48.** Emisiones de CO<sub>2</sub> en la Unión Europea



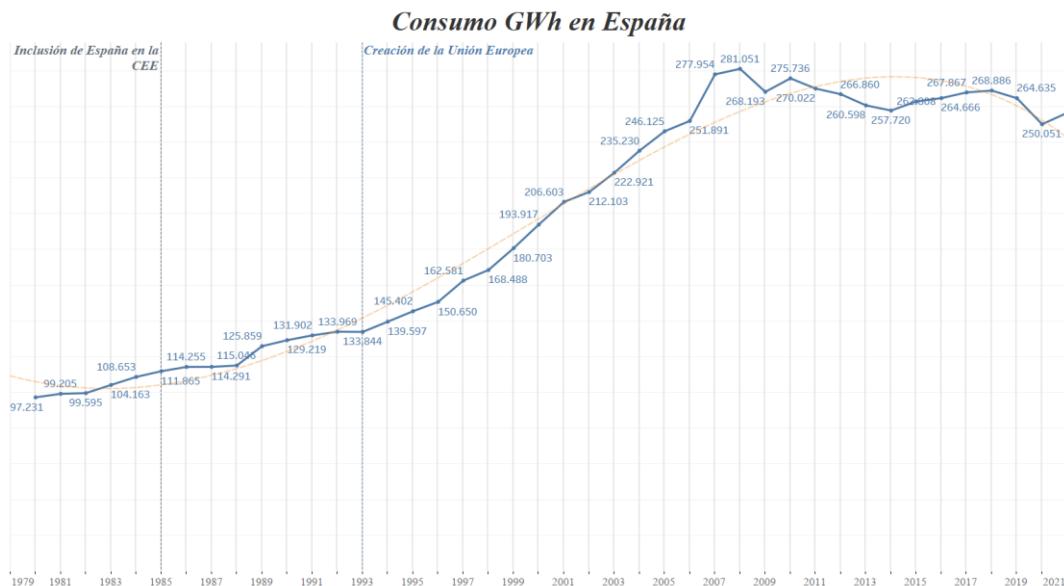
Fuente: Elaborada por el autor

En la figura 49 se muestra la evolución de las emisiones de CO<sub>2</sub> del país en el que se ha optado por utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con su valor concreto, y se añaden las líneas constantes verticales que marcan el año de la inclusión del país en la UE y la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 49.** Emisiones de CO<sub>2</sub> en España

Fuente: Elaborada por el autor

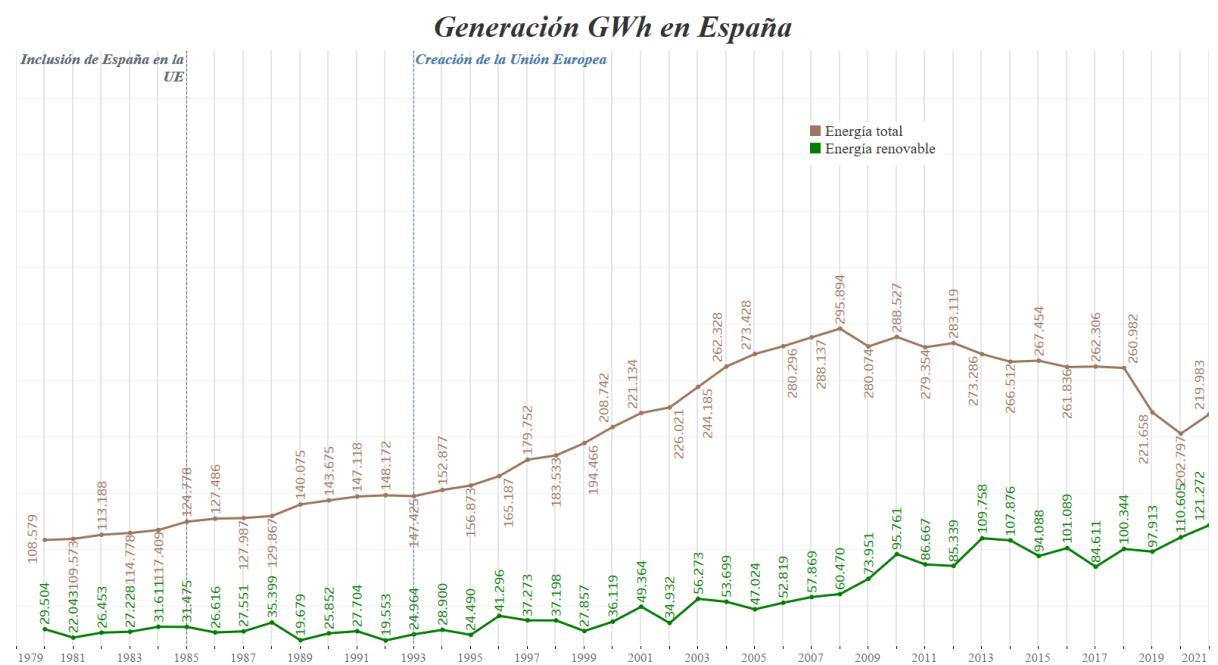
En la figura 50 se muestra la evolución del consumo energético del país en el que se ha optado por utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con su valor concreto, y se añaden las líneas constantes verticales que marcan el año de la inclusión del país en la UE y la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 50.** Consumo energético en España

Fuente: Elaborada por el autor

En la figura 51 se muestra la energía total generada por el país a lo largo del tiempo y de esa energía cuanta proviene de fuentes renovables. Se trata de 2 variables cuantitativas que se mueven en el mismo rango de valores y representan el mismo dominio, por tanto, se ha decidido mostrarlas juntas en un eje doble mediante un gráfico de líneas, y de esta forma se puede apreciar fácilmente de forma visual cómo evolucionan y compararlas fácilmente. Los colores verde y marrón representan la categoría a la que pertenecen y se define en la leyenda, las etiquetas indican el valor concreto de la variable para cada año.

**Figura 51.** Generación energética y uso de renovables en España



Fuente: Elaborada por el autor

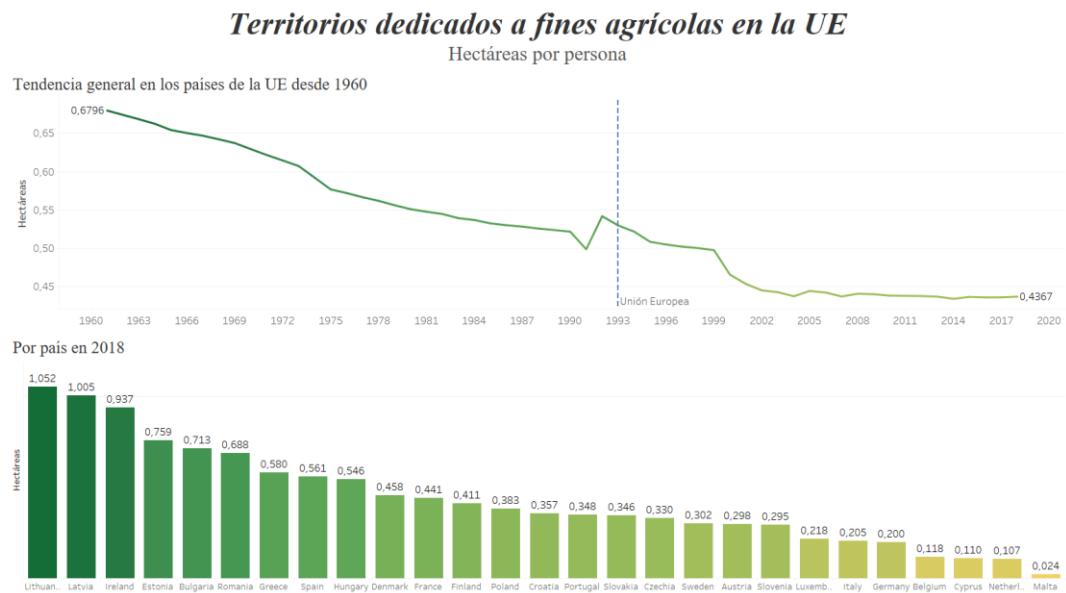
En la figura 52 se representa la extensión de los territorios agrícolas en los países de la Unión Europea. Para ello se ha decidido construir un dashboard con 2 gráficas.

La primera es una gráfica de líneas ya que se pretende representar la evolución en el tiempo de una variable cuantitativa, en este caso la extensión media en hectáreas de los territorios agrícolas, desde el año 1960 hasta el 2018. De esta forma junto con el color que también describe esta variable, es más sencillo para el usuario comprobar la tendencia general.

A continuación, se ha elegido un diagrama de barras verticales ordenado de mayor a menor para representar este mismo dato en un año concreto, ya que es la forma más clara e inequívoca para el usuario de comprobar que países cuentan con mayor o menor cantidad de territorio dedicado actualmente. En este panel el usuario puede interactuar pulsando sobre

cualquiera de los países y filtrar los datos referentes a cada país en concreto, cambiando los gráficos dinámicamente.

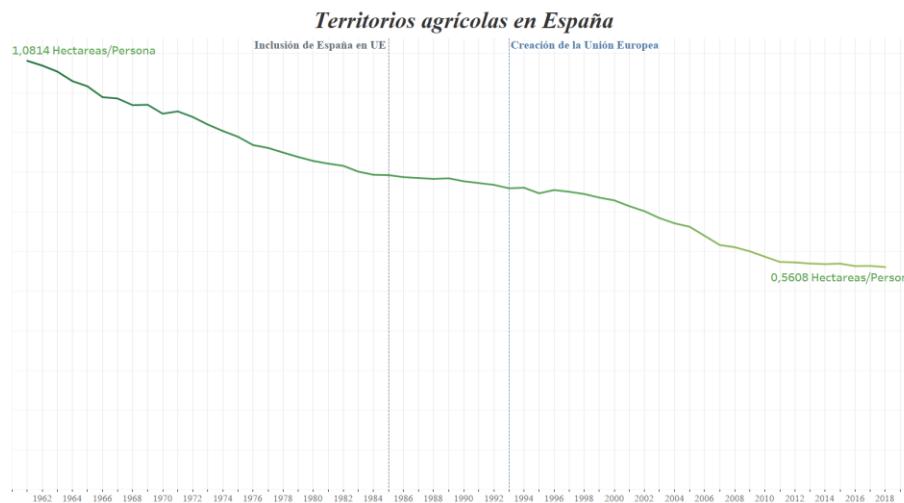
**Figura 52.** Territorios agrícolas en la Unión Europea



Fuente: Elaborada por el autor

En la figura 53 se muestra la evolución de la extensión de territorio agrícola del país, y se ha optado por utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con su valor concreto al principio y al final del intervalo. Además, se añaden las líneas constantes verticales que marcan el año de la inclusión del país en la UE y la creación oficial de la UE.

**Figura 53.** Territorios agrícolas en España



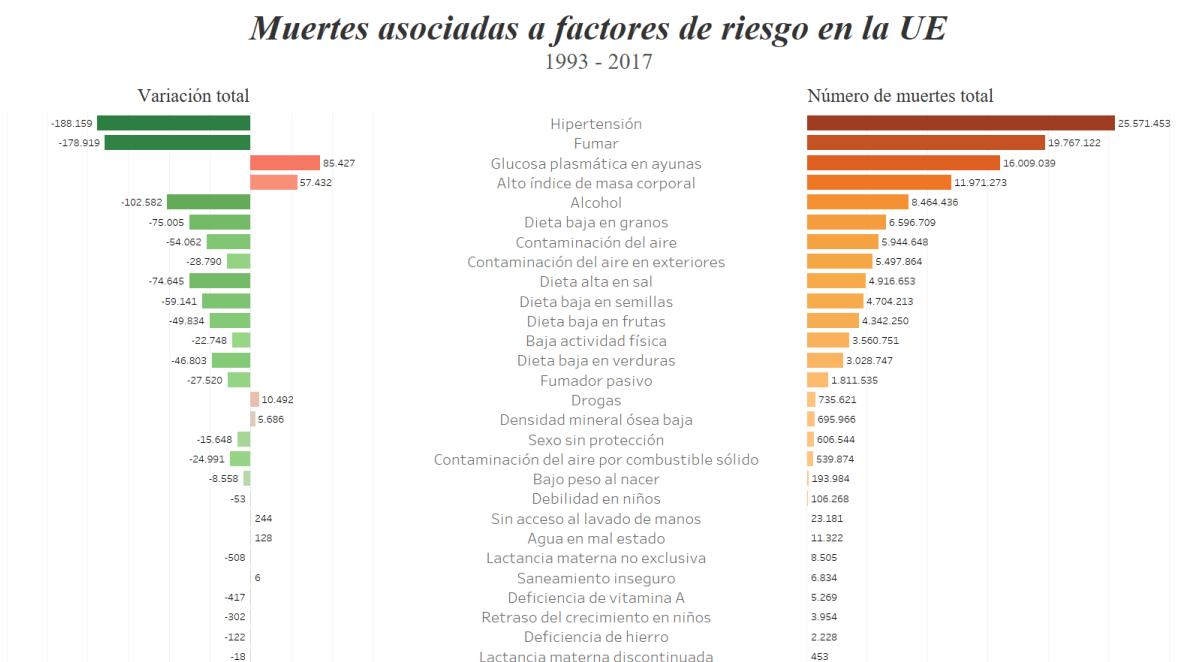
Fuente: Elaborada por el autor

En la figura 54 se exponen los principales factores de riesgo para la salud en la Unión Europea. Se ha decidido crear un dashboard con 2 gráficos diferentes.

El primer gráfico de barras horizontales representa la variación en el número de muertes al año de cada factor desde 1993 a 2017, se ha escogido este tipo de gráfico por ser el más claro e inequívoco para el lector a la hora de sacar conclusiones. El color representa con un verde si se ha mejorado en cifras y con rojo si se ha empeorado, por ser lo más intuitivo para el lector antes que una cifra o una longitud.

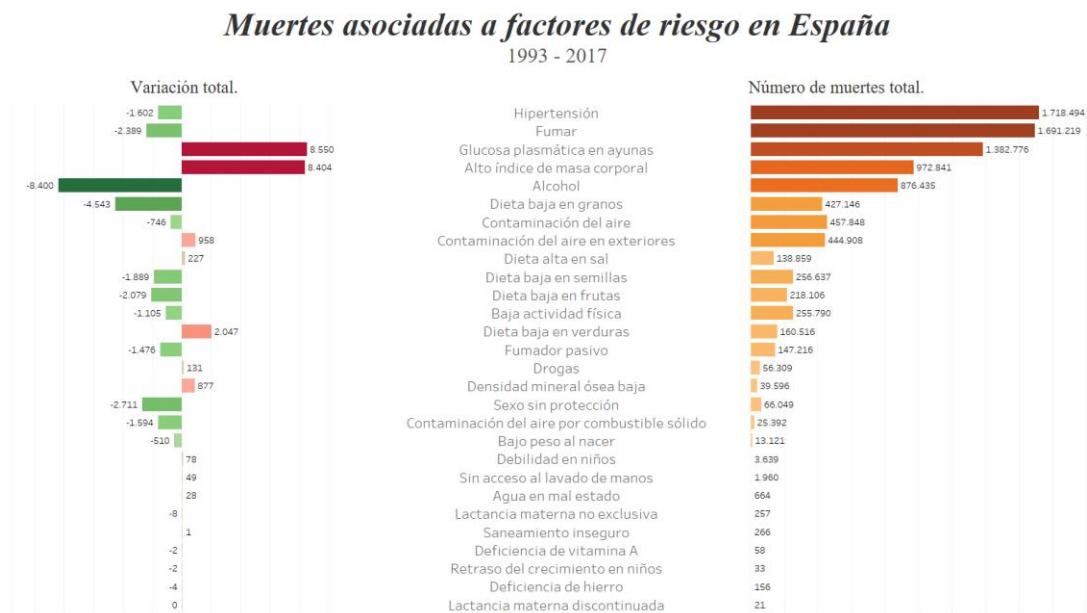
De igual forma se ha escogido para el segundo gráfico el mismo formato, pero esta vez para representar el número de muertes total provocado por el factor de riesgo.

**Figura 54.** Número de muertes asociadas a factores de riesgo en la UE



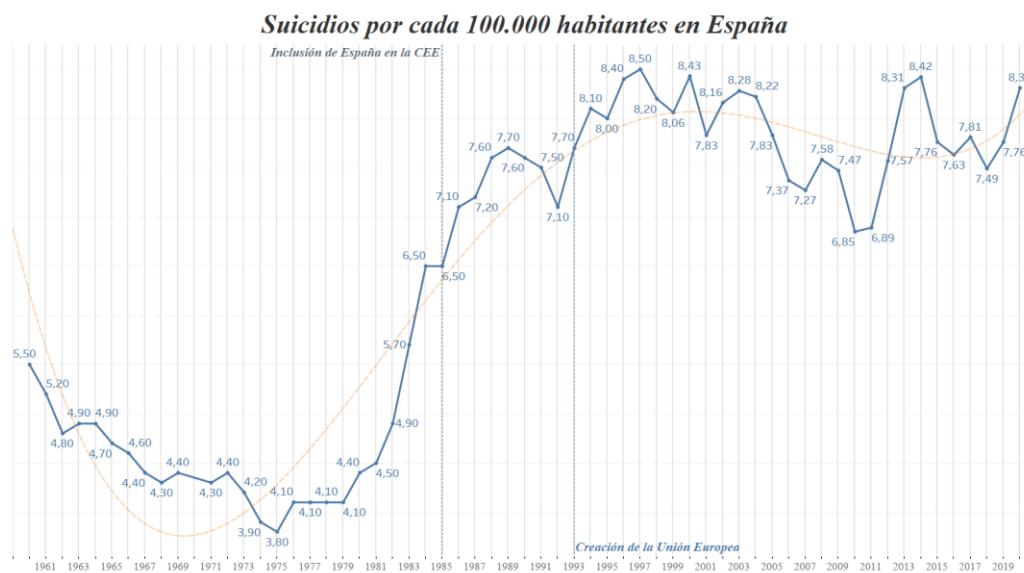
Fuente: Elaborada por el autor

La figura 55 se ha diseñado y pensado de la misma forma que la figura 56 pero esta vez representa los mismos datos para el caso de España en concreto.

**Figura 55.** Número de muertes asociadas a factores de riesgo en España

Fuente: Elaborada por el autor

En la figura 56 se muestra la evolución de la tasa de suicidio por cada 100.000 habitantes del país en el que se ha optado por utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con su valor concreto, y se añaden las líneas constantes verticales que marcan el año de la inclusión del país en la UE y la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

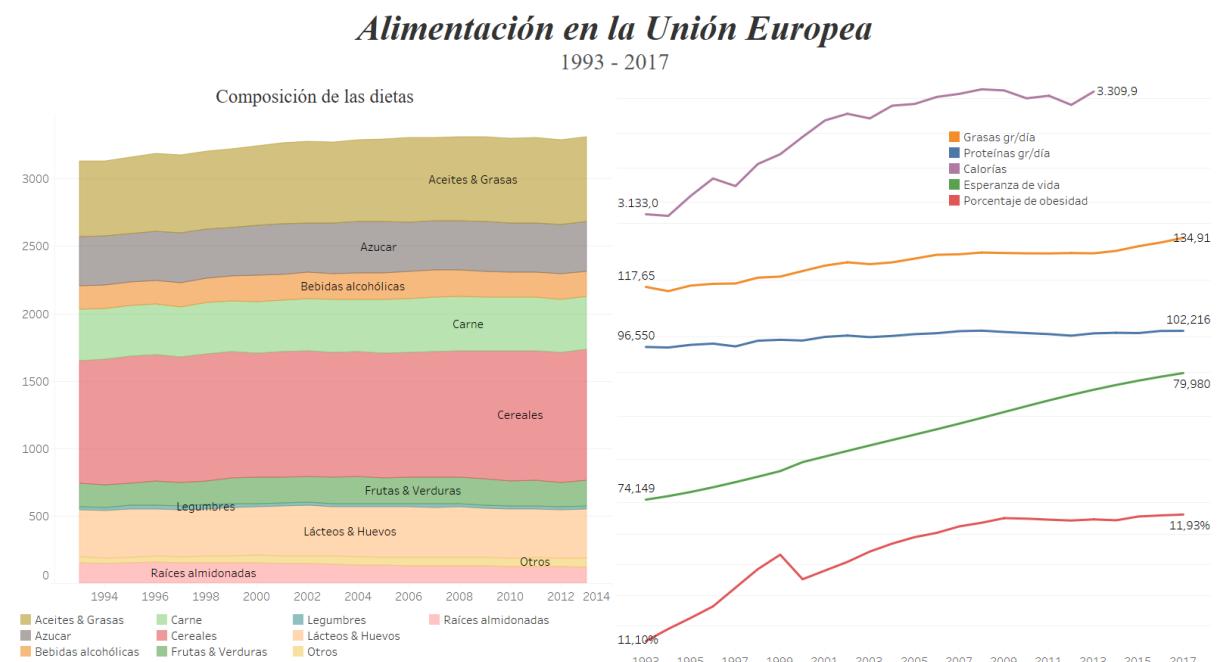
**Figura 56.** Suicidios por cada 100.000 habitantes en España

Fuente: Elaborada por el autor

En la figura 57 referente a la alimentación en la Unión Europea se ha decidido construir 2 gráficos. Para el primero que muestra la composición de las dietas europeas se ha escogido un diagrama de flujo acumulado, de esta forma el usuario puede comprobar a lo largo del tiempo, los diferentes compuestos que forman una dieta, reuniendo todos los ingredientes, uno por cada área hasta hacer el total de calorías consumidas. Se ha utilizado colores diferentes con contrastes para que la visualización sea más clara.

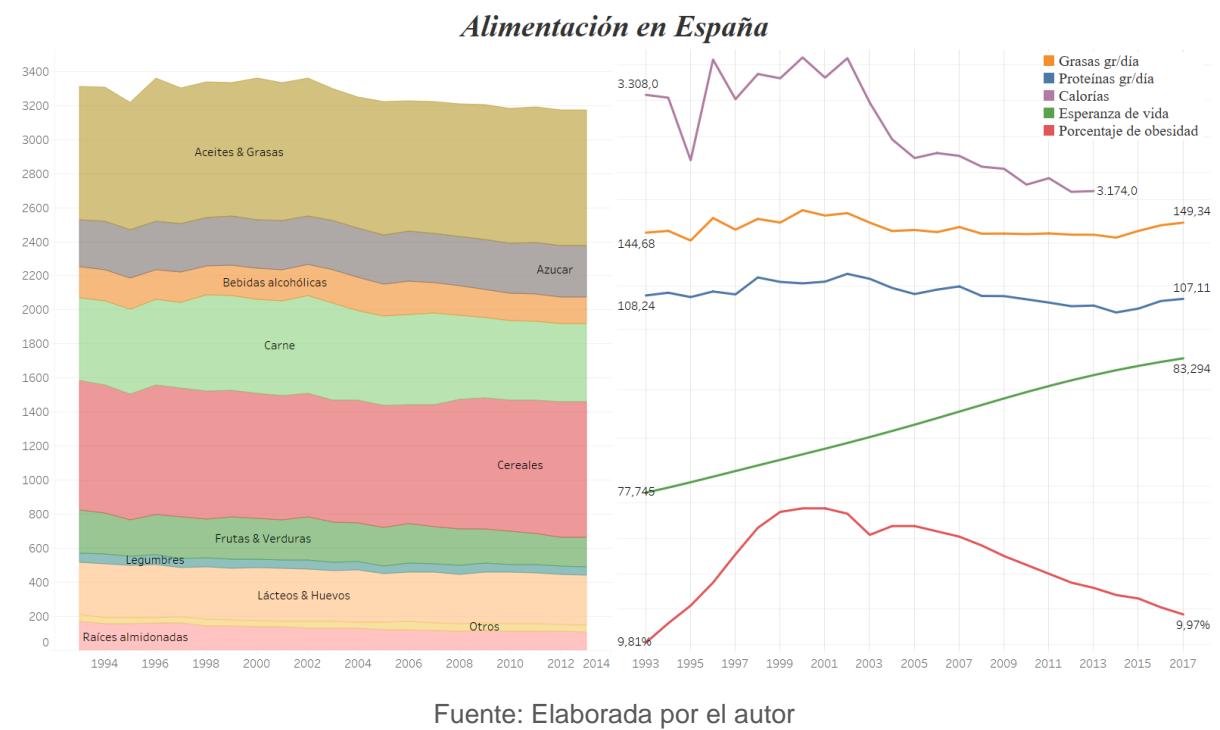
Seguidamente se escoge para el siguiente un gráfico de líneas ya que es el que mejor transmite las tendencias, en este caso del consumo de grasas, hidratos, proteínas, calorías, esperanza de vida y porcentaje de obesidad.

**Figura 57.** Componentes de la alimentación en la Unión Europea

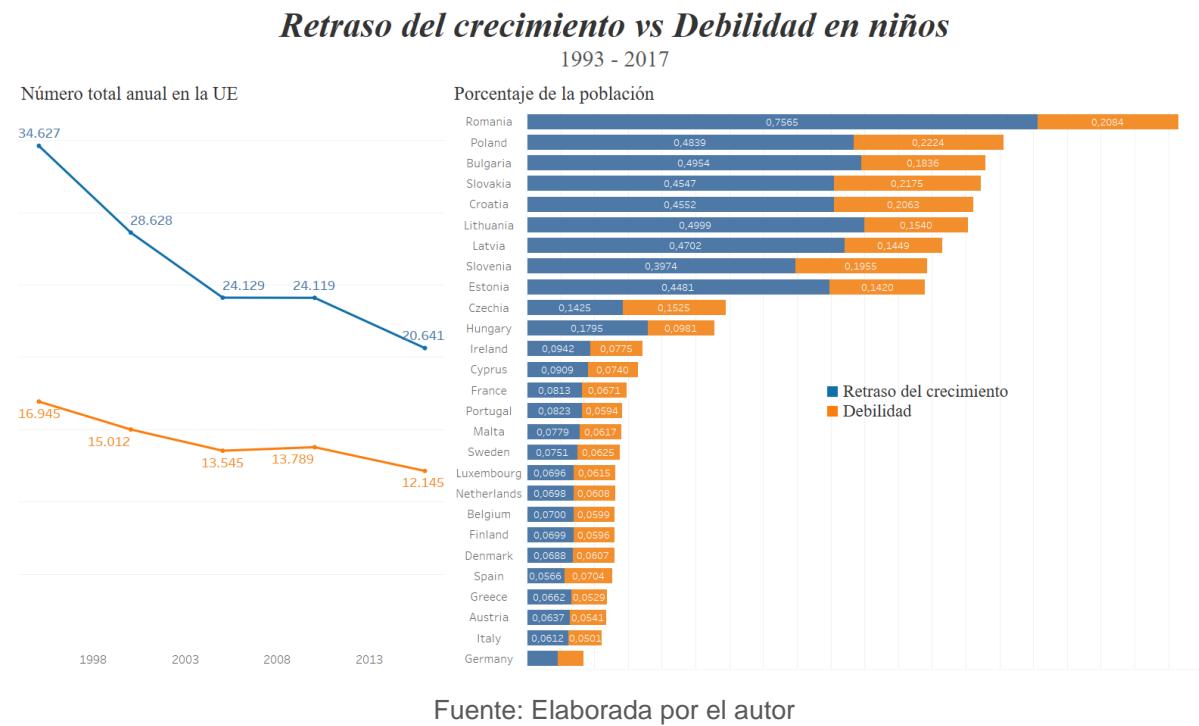


Fuente: Elaborada por el autor

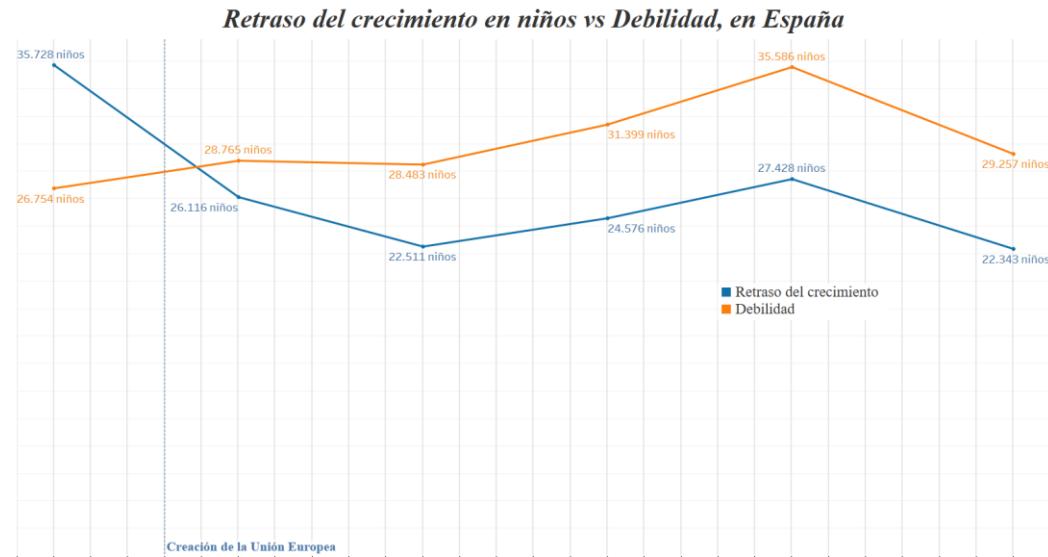
En la figura 58 es el mismo caso anterior, pero con los datos concretos de España, se ha construido con la misma forma para facilitar al lector la comparación entre los datos.

**Figura 58.** Componentes de la alimentación en España

En la figura 59 para representar los problemas de nutrición de los niños se han utilizado 2 gráficas diferentes sobre las mismas 2 categorías, el retraso del crecimiento y la debilidad en niños. La primera es una gráfica de líneas de eje doble que permite comparar y a la vez observar las tendencias de las 2 variables fácilmente, además se ha utilizado un color diferente para cada una con un contraste marcado. Para la segunda se ha optado por un diagrama de barras laterales acumulado con el objetivo de observar cada país en concreto y en detalle y de esta forma comparar que porcentaje es mayor para cada categoría, además al estar ordenado de mayor a menor se tiene una visión global de los países que mejor y peor posicionados están.

**Figura 59.** Problemas en la nutrición de los niños de la UE

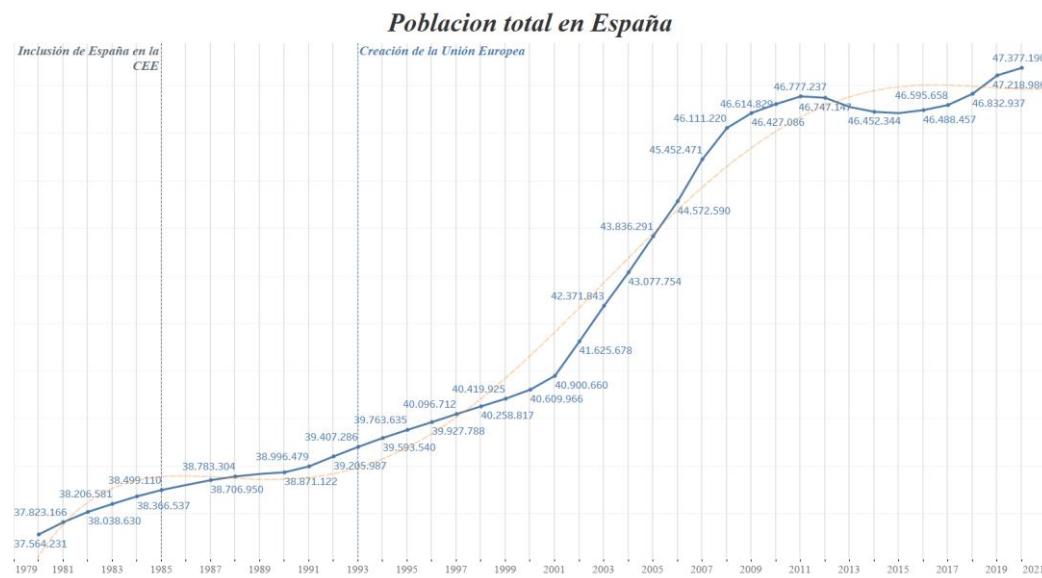
La figura 60 se compone de los mismos elementos que la anterior ya que representa la misma información, pero referente a los datos del país únicamente.

**Figura 60.** Problemas en la nutrición de los niños de España

En la figura 61 se muestra la evolución del número de habitante del país en el que se ha optado por utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución

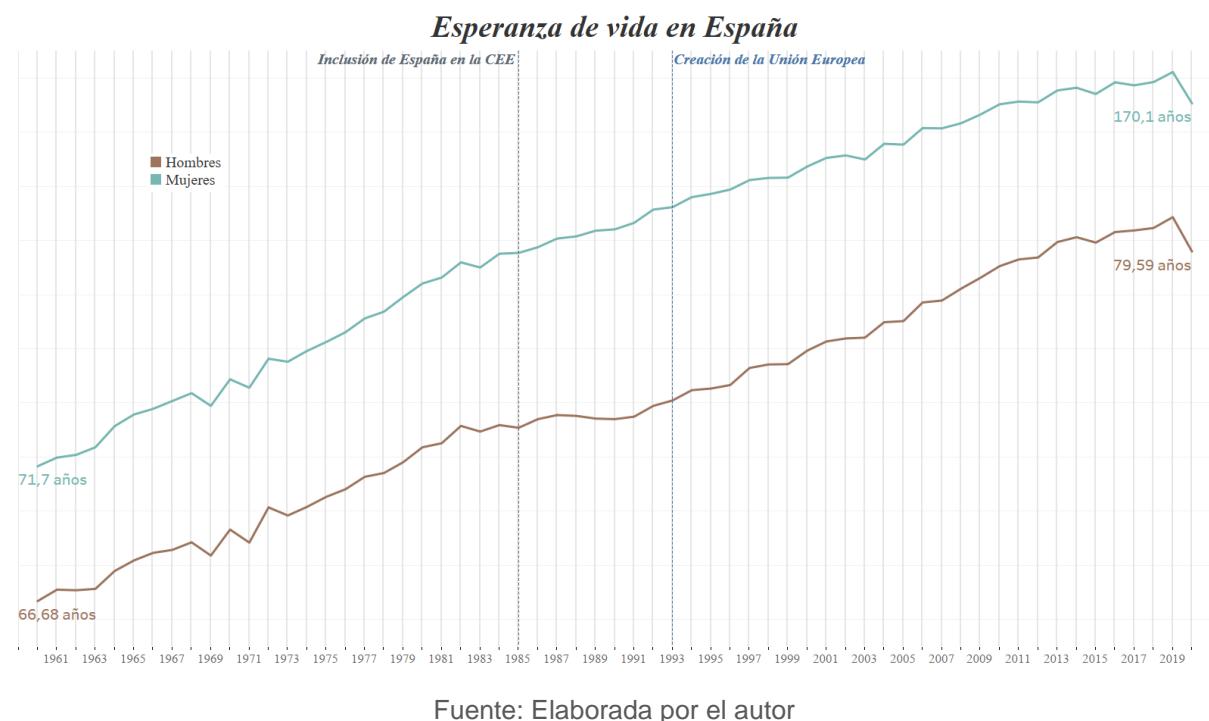
en el tiempo, acompañado de etiquetas con su valor concreto, y se añaden las líneas constantes verticales que marcan el año de la inclusión del país en la UE y la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 61.** Población en España

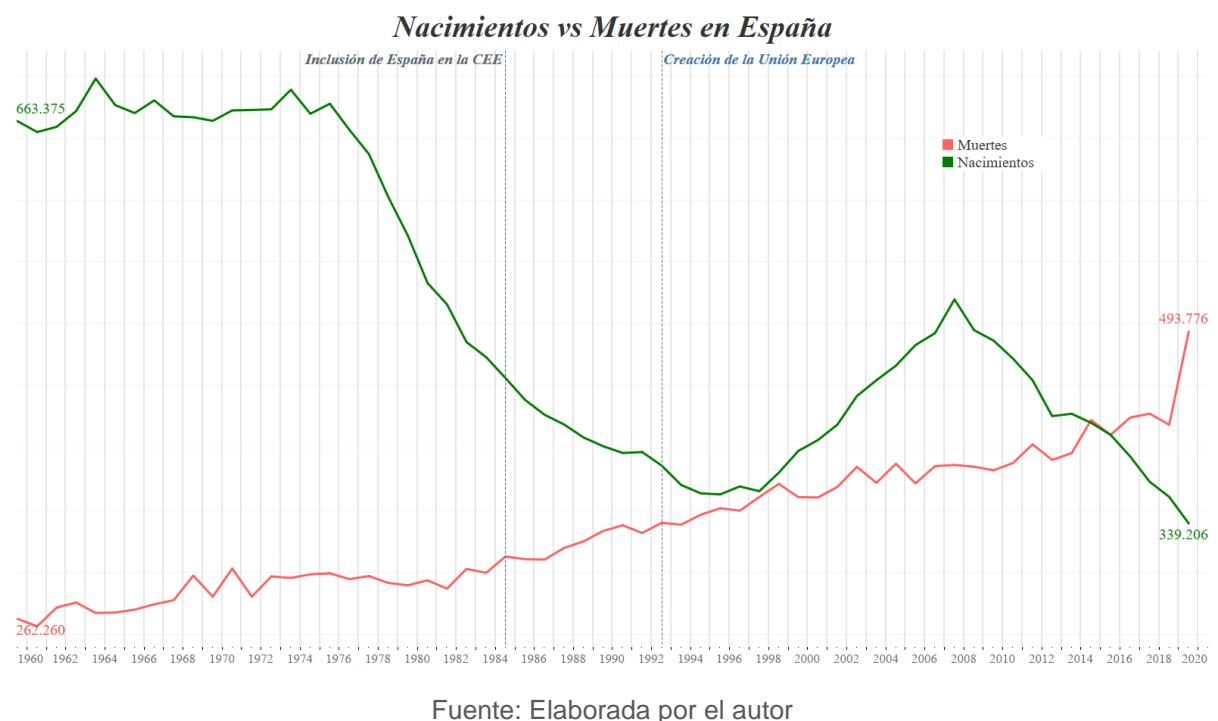


Fuente: Elaborada por el autor

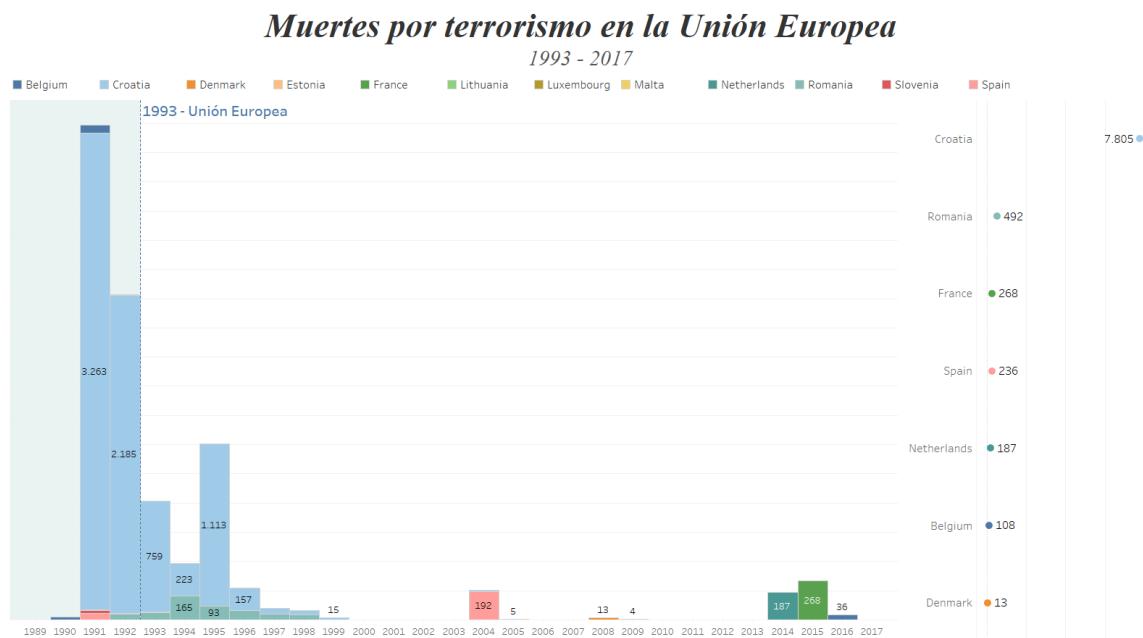
En la figura 62 se muestra la evolución de la esperanza de vida en hombres y mujeres del país a lo largo del tiempo. Se trata de 2 variables cuantitativas que se mueven en el mismo rango de valores y representan el mismo dominio, por tanto, se ha decidido mostrarlas juntas en un eje doble mediante un gráfico de líneas, de esta forma se puede apreciar fácilmente de forma visual cómo evolucionan y compararlas de forma sencilla. Los colores azul y marrón representan la categoría a la que pertenecen y se define en la leyenda. Las etiquetas indican el valor concreto de la variable al principio y al final del intervalo. Además, se añaden las líneas constantes verticales que marcan el año de la inclusión del país en la UE y la creación oficial de UE que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 62.** Esperanza de vida en España

En la figura 63 se muestra el número de nacimientos y muertes en el país a lo largo del tiempo, se trata de 2 variables cuantitativas que se mueven en el mismo rango de valores y representan el mismo dominio, por tanto, se ha decidido mostrarlas juntas en un eje doble mediante un gráfico de líneas, de esta forma se puede apreciar fácilmente de forma visual cómo evolucionan y compararlas de forma sencilla. Los colores verde y rojo representan la categoría a la que pertenecen y se define en la leyenda. Las etiquetas indican el valor concreto de la variable al principio y al final del intervalo. Además, se añaden las líneas constantes verticales que marcan el año de la inclusión del país en la UE y la creación oficial de la UE que apoya de forma visual al usuario en la extracción de conclusiones.

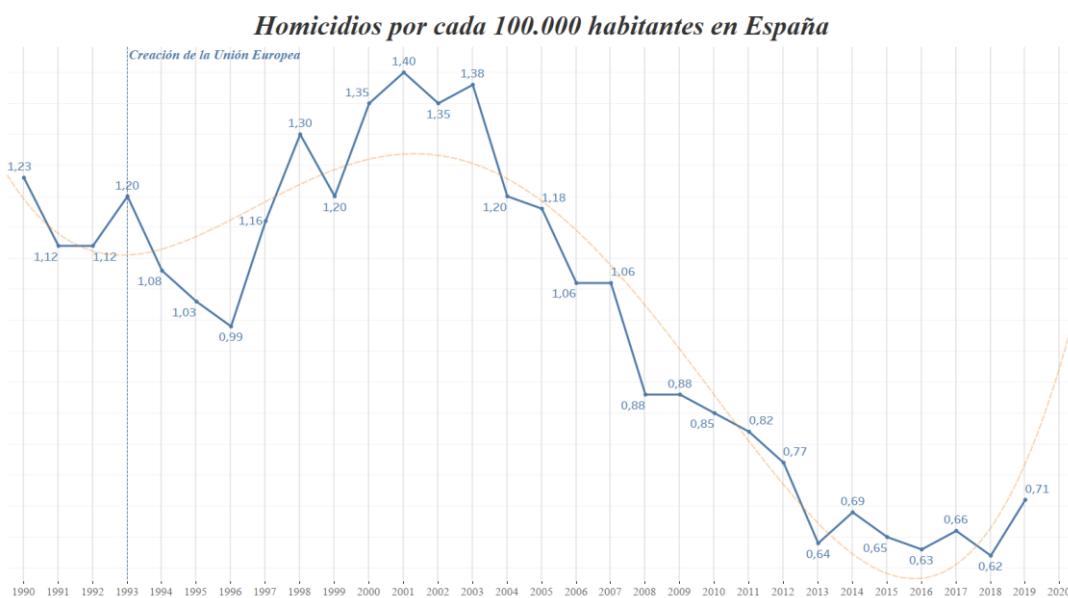
**Figura 63.** Balanza demográfica en España

En la figura 64 se muestra un panel informativo que representa en número de muertes por conflictos o terrorismo en la Unión Europea. Para visualizarlo se ha decidido utilizar un diagrama de barras verticales acumulado para que el usuario pueda comprobar de forma sencilla como evoluciona en el tiempo el número de muertes en cada país, que vienen distinguidos por diferentes colores. A su vez se ha construido otro gráfico a la derecha donde se han colocado los países que sobresalen en esta métrica de mayor a menor junto con el número total de muertes. Se ha hecho de esta forma porque permite al usuario comprobar la diferencia que existe entre países, y en general el lector puede comprobar tanto los países que más y menos muertes han tenido, como su evolución en el tiempo y cuando se han producido. Además permite interactuar al usuario pulsando sobre los países filtrando la información referente a cada uno.

**Figura 64.** Muertes a causa del terrorismo en la Unión Europea

Fuente: Elaborada por el autor

En la figura 65 se muestra la evolución del número de homicidios por cada 100.000 habitantes del país, en el que se ha optado por utilizar un gráfico de líneas que representa la variable cuantitativa y su evolución en el tiempo, acompañado de etiquetas con su valor concreto, y se añade la línea constante vertical que marca el año de la creación oficial de la UE junto con una línea de tendencia que apoya de forma visual al usuario en la extracción de conclusiones.

**Figura 65.** Homicidios por cada 100.000 habitantes en España

Fuente: Elaborada por el autor

Para la exposición de todas estas visualizaciones y descripciones es necesario el despliegue en un sitio web que las reúna en un formato legible y atractivo para el lector, la forma de hacerlo se describe a continuación.

#### **4.2.3 Bloque III: Despliegue de la aplicación web**

Una vez construidas las visualizaciones, estas se exponen en una aplicación web construida con lenguaje HTML y CSS (Bootstrap), en la herramienta Brackets. Esta aplicación web se compone de pestañas con los diferentes temas tratados donde se representan las visualizaciones con unas breves descripciones de lo que se observa.

El primer paso es tener instalada la herramienta que permite la edición de archivos HTML, en este caso se ha utilizado Brackets, que es de libre uso y distribución.

En la realización del proyecto se ha decidido establecer un archivo HTML por cada categoría estudiada, por tanto, una pestaña para cada tema tratado. En este caso se tienen los siguientes archivos:

- index.html – Pantalla de inicio a la página
- derechos.html – Visualizaciones de los derechos humanos
- economía\_cuentas.html – Visualizaciones de los datos económicos nacionales
- economía\_ingresos.html - Visualizaciones de los datos económicos en ingresos
- economía\_desigualdad.html - Visualizaciones de los datos económicos de desigualdad
- economía\_empleo.html - Visualizaciones de los datos de empleo
- economía\_corrupcion.html - Visualizaciones de los datos de corrupción
- salud\_contaminacion.html – Visualizaciones de los datos sobre contaminación
- salud\_riesgo.html - Visualizaciones de los datos sobre factores de riesgo
- salud\_alimentacion.html - Visualizaciones de los datos sobre alimentación
- salud\_demografia - Visualizaciones de los datos demográficos
- paz.html – Visualizaciones sobre la seguridad
- cuestionario.html – Cuestionario para evaluar la metodología

La estructura seguida para incrustar los dashboards en el código HTML del sitio web, ha sido la de copiar y pegar el trozo de código que indica Tableau para estos casos. Tal y como se muestra en la figura 66.

**Figura 66.** Ejemplo básico para presentar las visualizaciones en la web

```

159 <div class="w-100">
160 </div>
161 ▼ <div class="col-lg-8">
162   <div class="tableauPlaceholder" id="viz1642458069274" style="position:
163     relative"><object class="tableauViz" style="display:none;"><param
164       name="host_url" value="https%3A%2F%2Fpublic.tableau.com%2F" /> <param
165       name="embed_code_version" value="3" /> <param name="site_root" value=''/><param
166       name="name" value="Viz_ESP&#47;balance_demoografico_esp" /><param
167       name="tabs" value="no" /><param name="toolbar" value="yes" /><param
168       name="animate_transition" value="yes" /><param
169       name="display_static_image" value="yes" /><param name="display_spinner"
170       value="yes" /><param name="display_overlay" value="yes" /><param
171       name="display_count" value="yes" /><param name="language" value="es-ES"
172       /><param name="filter" value="publish:yes" /></object></div>
173   <script type="text/javascript"> var divElement =
174     document.getElementById('viz1642458069274'); var
175     vizElement = divElement.getElementsByTagName('object')[0];
176     if ( divElement.offsetWidth > 800 ) {
177       vizElement.style.width='100%';vizElement.style.height=
178       (divElement.offsetWidth*0.75)+'px';} else if ( divElement.offsetWidth >
179       500 ) { vizElement.style.width='100%';vizElement.style.height=
180       (divElement.offsetWidth*0.75)+'px';} else {
181       vizElement.style.width='100%';vizElement.style.height='727px';
182       var scriptElement = document.createElement('script');
183       scriptElement.src =
184       'https://public.tableau.com/javascripts/api/viz_v1.js';
185       vizElement.parentNode.insertBefore(scriptElement, vizElement);
186     </script>
187   </div>
188   <div class="col-lg-4 align-self-center">
189     <p class="lead">En la figura se indica una comparativa entre el número de
190     nacimientos y el número de muertes en España cada año, donde es notable
191     el sorpaso que ha habido en los últimos años, esto se debe sobre todo a
192     la disminución de nacimientos desde el año 1976 que se ha reducido a la
193     mitad hasta los años 90, cuando hubo un rebote hasta la crisis de 2009,
194     cuando ya se desplomó hasta los 339,206 nacimientos. El número de muertes
195     se ha incrementado de forma constante desde 1960 y la población ha
196     aumentado en 13M de habitantes desde el año 1970, esto junto a la elevada
197     esperanza de vida ha provocado que la pirámide poblacional se encuentre
198     descompensada y haya más gente de avanzada edad y menos jóvenes.</p>
199   </div>
200 
```

Fuente: Elaborada por el autor

El primer recuadro representa el trozo de código que indica Tableau para compartir la visualización que se ha publicado con anterioridad en Tableau Public. De esta forma se consigue que una vez desplegado el sitio web, esta se actualice automáticamente al realizar cambios sobre ella. De esta forma cuando se guarde algún cambio nuevo en Tableau, se refleje directamente sobre la web ya desplegada.

Una vez se tienen los archivos formados, se puede desplegar la aplicación web, utilizando la funcionalidad *Github Pages*. Para ello necesitas una cuenta de Github y crear un repositorio que contenga el proyecto con las características que se indican, e ir actualizándolo con los archivos que se han construido, es decir, los ficheros HTML.

En este paso es muy recomendable utilizar GitHub Desktop que permite trabajar con los archivos de forma local e ir guardando los cambios para posteriormente hacer un *commit* y *push* que actualiza el repositorio, de esta forma se facilita bastante el hacer cambios.

Los resultados finales del trabajo se pueden comprobar en el sitio web creado al completo (<https://robertobarrosogarcia.github.io/>).

## 4.3 Evaluación

Como se comenta en apartados anteriores, el objetivo del trabajo es que el público general, familiarizado con este tipo de información pero que no tiene que ser experto, pueda aumentar sus conocimientos sobre el tema tratado, en este caso España y la Unión Europea. En el apartado de identificación de requisitos se llega a la conclusión que las personas que comprendan las edades entre 25 y 54 años deberían poder extraer información de forma sencilla y clara, y ser capaces de poder sacar sus propias conclusiones, además pueden ser los más interesados en este tipo de informaciones.

Para la evaluación del comportamiento de los usuarios con las visualizaciones y las descripciones se ha creado un cuestionario anónimo. (Ver figura 67)

**Figura 67. Cuestionario**

**Cuestionario para validar la nueva metodología de visualización interactiva propuesta**

Aviso: Este formulario se realiza en 5 minutos, y no guarda nombres, correos electrónicos, o cualquier otra información personal, es totalmente anónimo y solo se guardan las respuestas indicadas más adelante.

El objetivo es medir el nivel de comprensión de las visualizaciones construidas para un trabajo de fin de Master. De esta forma, solucionar posibles errores, y entender si las representaciones son útiles para exponer la información que se pretende mostrar.

El sitio web donde se pueden encontrar es el siguiente:  
<https://robertobarrosogarcia.github.io/>

 rbgeshd@gmail.com (no compartidos) [Cambiar de cuenta](#) 

\*Obligatorio

Introduce tu fecha de nacimiento \*

Fecha  
dd/mm/aaaa

Profesión \*

Tu respuesta

¿El sitio web le ha resultado sencillo y claro en la exposición de la información?

Sí  
 No

¿Las visualizaciones le han resultado fáciles para interactuar?

Sí  
 No

¿Cuál cree que ha sido su nivel de comprensión de la información que ha visto? \*

0    1    2    3    4    5  
No se entiende                     Se entiende perfectamente

¿Le han resultado de utilidad las visualizaciones para aumentar sus conocimientos sobre el tema tratado? \*

No  
 Sí

A continuación, puedes añadir cualquier comentario o propuesta que creas relevante

Tu respuesta

**Enviar** **Borrar formulario**

Nunca envíes contraseñas a través de Formularios de Google.  
Google Formularios Este contenido no ha sido creado ni aprobado por Google.

Fuente: Elaborada por el autor

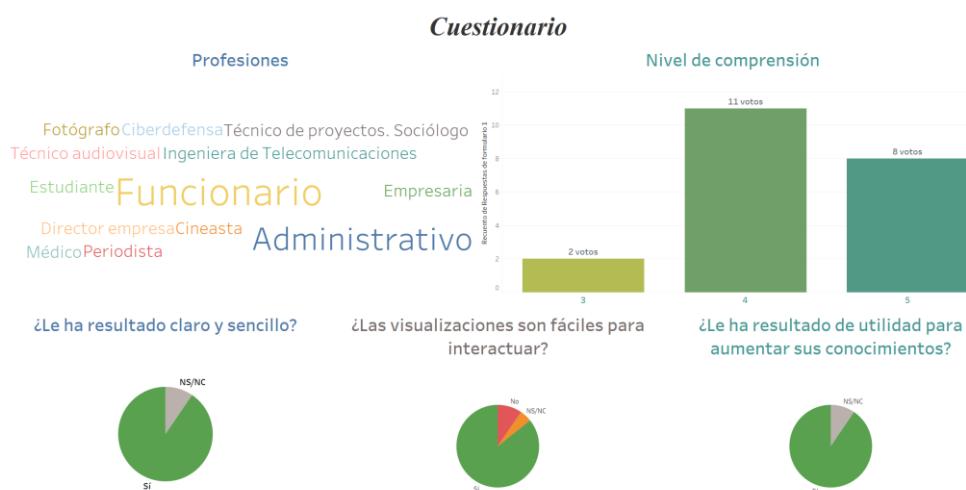
El cuestionario tiene el objetivo de evaluar la metodología propuesta, analizar el nivel de usabilidad y conocer las opiniones del usuario de primera mano, donde pueden dejar sus comentarios y puntuaciones sobre diferentes aspectos, después de interactuar con el sitio web al menos 10min. Se ha tenido en cuenta que el cuestionario se tarde como máximo en realizar alrededor de 5min, si es más extenso el público tiende a cansarse y dar respuestas que no van acorde con su pensamiento y de esta forma acabar más rápido, pero así no se sacan las conclusiones correctas que aporten alguna mejora a la metodología.

En definitiva, este cuestionario se compone de lo siguiente: Breve introducción y descripción del proyecto (1min) y unas preguntas (4min):

- Edad y profesión
- ¿El sitio web le ha resultado sencillo y claro en la exposición de la información? (Sí o no)
- ¿Las visualizaciones le han resultado fáciles para interactuar? (Sí o no)
- ¿Cuál cree que ha sido su nivel de comprensión de la información que ha visto? (0=mala a 5=perfecto)
- ¿Le han resultado de utilidad las visualizaciones para aumentar sus conocimientos? (si o no).
- Comentarios y propuestas de mejora.

Los resultados de los cuestionarios los podemos observar en el dashboard mostrado en la figura 68. Donde se reflejan las respuestas de un total de 22 personas que han realizado el cuestionario, con edades comprendidas entre los 24 y 49 años.

**Figura 68.** Resultados del cuestionario de evaluación



Fuente: Elaborada por el autor

## 5 Conclusiones y trabajo futuro

A partir del desarrollo de la metodología propuesta y la validación de los usuarios se pueden extraer conclusiones y posibles líneas de trabajo futuras.

### 5.1 Conclusiones

Respecto a los objetivos específicos indicados, el primer subobjetivo es el de establecer los temas que se van a tratar e identificar las posibilidades de captura de los datos. La información que se puede recoger sobre un país puede ser infinita, pero en este caso, la propuesta es relacionarlo con su pertenencia a la Unión Europea, y por lo tanto se han decidido establecer unos temas en los que la organización ha puesto énfasis desde su creación, y denomina como “Las cuatro prioridades”, estas serían “proteger a los ciudadanos y las libertades, desarrollar una base económica sólida y dinámica, construir una Europa climáticamente neutra, ecológica, justa y social, y promover los intereses y valores europeos en el mundo” (Unión Europea, 2021). Una vez se tienen acotados los temas, es muy importante no solo encontrar los datos relacionados, sino que además estos tengan la veracidad necesaria en este tipo de estudios. Después de un gran proceso de búsqueda, se han encontrado fuentes que cumplían con estos requisitos, y algunos datos que se han tenido que descartar por falta de veracidad o por falta de completitud, en definitiva, el resultado de la investigación ha sido correcto con una definición clara de los temas tratados y una buena selección de las fuentes de las que se extraen los datos.

El segundo objetivo específico de determinar y describir los datos que se van a recoger y tratar se ha cumplido mediante la elaboración de scripts donde se han extraído, transformado y unido de forma homogénea todos los datos, de esta forma se tiene reunida toda la información que se tiene descrita en la memoria y que sirve para alimentar las visualizaciones interactivas.

Para el siguiente subobjetivo de desarrollar las visualizaciones interactivas se han seguido las indicaciones descritas en las diferentes publicaciones indicadas en el contexto del arte, con el objetivo de construir unos paneles que permitan al usuario adquirir conocimientos sobre el tema tratado de forma clara y sencilla, con el valor añadido de la interactividad. Se ha conseguido con la herramienta software utilizada, Tableau, y los resultados como se verá más adelante, han sido los adecuados, por las positivas evaluaciones recibidas.

Por último, relacionado con el objetivo anterior, se ha propuesto otro subjetivo en orden de comprobar si para el usuario ha tenido un impacto significativo el desarrollo de la metodología

propuesta, y considera que se ha cumplido con los objetivos. Por ello se ha repartido un cuestionario a modo de evaluación en un escenario real, que permita extraer resultados y conclusiones del público, los resultados han sido buenos como se indica en detalle más adelante.

La metodología de visualización de datos propuesta tenía como objetivo principal el mostrar de forma sencilla al usuario una gran cantidad de información acerca de un país en relación con la Unión Europea, de tal forma que el usuario se interesara, y fuera capaz de adquirir conocimiento de forma fácil e independiente. Según los resultados de la prueba de evaluación y los comentarios sugeridos se puede concluir que se ha conseguido. De las 22 personas encuestadas, el 90.9% ha confirmado que le ha resultado un sitio web sencillo y claro en la exposición de la información. El 86% ha contestado que las visualizaciones le han resultado intuitivas y por tanto fáciles para interactuar con ellas. Además, el 90.9% afirma que ha comprendido la información mostrada y ha conseguido aumentar sus conocimientos acerca de temas que desconocía. Según algunos comentarios recibidos, la estética e interactividad ha añadido interés en ciertos usuarios, sobre temas, que de otra forma no les habrían llamado la atención.

El público ha concluido con sus comentarios que se trata de una buena forma de observar los resultados y es de gran utilidad para la extracción de información y conclusiones. Una de las motivaciones era la de dar protagonismo a las figuras y menos a los extensos textos explicativos, sin perder información, y por el 'feedback' recibido se puede deducir que se ha conseguido, dando además así mayor visibilidad a los temas tratados.

## 5.2 Líneas de trabajo futuro

Existen multitud de caminos futuros y estudios muy interesantes que se pueden realizar, técnicamente se puede desarrollar una nueva funcionalidad para optimizar la recolección de los datos, ya sea un script más avanzado y parametrizable que permita todo el proceso de ETL de una sola vez y más rápido. Se pueden realizar ligeros cambios en los scripts que extraen información del país y poder ejecutarlos para otros países, por ejemplo, se podría realizar el mismo estudio, pero enfocado en cualquier otro miembro de la Unión Europea. De igual forma, automatizar algunas construcciones de Tableau para recoger los datos en streaming y realizar las actualizaciones automáticamente.

El despliegue web se puede hacer de muy diversas formas y con un formato totalmente diferente que pueda facilitar aún más la comprensión de la información. Otra posible mejora

a nivel técnico sería la de añadir un seguimiento o ‘tracking’ de la web para comprobar como es el comportamiento de los usuarios y su usabilidad respecto al sitio, e incluso añadir un motor de búsqueda que facilite aún más encontrar la información. En cuanto a los temas escogidos, no solo se puede aumentar la información sobre lo que ya está publicado en este documento, sino que se pueden añadir más puntos interesantes que no eran posibles de abarcar en este estudio por falta de tiempos o falta de información. Ejemplos como las migraciones, las relaciones internacionales, tanto en campos económicos como militares o políticos, asuntos demográficos como la densidad de población y su distribución, el protagonismo de las religiones y los distintos idiomas que se hablan en todo el territorio, incluso los deportes y su importancia a nivel global. Por último, un tema muy importante que no se ha podido tratar por su amplitud es la inversión en ciencia y tecnología que ha permitido avances en todos los demás sectores tratados aquí y por lo tanto está muy relacionado con todos los cambios observados.

En definitiva, se pueden ampliar o añadir nuevos temas y con otros puntos de vista desde otras fuentes de información, contrastando con lo que ya se tiene. Es destacable la aportación que puede tener este tipo de metodología en el ámbito de la divulgación informativa, histórica o comunicativa, sobre todo en el caso de que se pueda desarrollar más, añadiendo todo lo indicado en este apartado.

## 6 Bibliografía

- Adobe (2014). *Brackets* (Versión 2.0.1) [Software de computador]. Adobe. <https://brackets.io/>
- Alonso, N. (2021 octubre). *España se mantiene como el séptimo país europeo con el salario mínimo más alto tras la última subida*. Newtral. <https://www.newtral.es/espana-salario-minimo-europa-ultima-subida/20211009/>
- Anaconda Inc. (2017). *Conda* (Versión 4.11) [Software de computador]. Anaconda Inc. <https://docs.conda.io/en/latest/index.html>
- Cairo, A. (2011). *El arte funcional*. Alamut.
- Cairo, A. (2017). *Nerd Journalism*. Universidad Oberta de Catalunya.
- Cairo, A. (2019). *How Charts Lie*. Norton & Company.
- DatosMacro. (2019). *Esperanza de vida al nacimiento* [Fotografía]. Expansión. <https://datosmacro.expansion.com/demografia/esperanza-vida/españa>

Hernández-Castro, F. (2021). *Dashboard design cookbook, metodología para el diseño de visualizaciones de datos*. Instituto Tecnológico de Costa Rica.

<https://hdl.handle.net/2238/13281>

Herrero Solana, V. F., Rodríguez Domínguez, A. M. (2015). *Periodismo de datos, infografía y visualización de la información*. Universidad de Granada.

<https://dx.doi.org/10.1344/BiD2015.34.4>

Hunter, J. (2003). *Matplotlib* (Versión 1.5.1) [Software de computador]. The Matplotlib Development team. <https://matplotlib.org/>

Instituto Nacional de Estadística. (s.f.). Ministerio de Asuntos Económicos y Transformación Digital. Recuperado el 14 de diciembre de 2021 de <https://www.ine.es/>

Instituto Nacional de Estadística (2021a). *Esperanza de vida al Nacimiento según sexo* [Fotografía]. INE. <https://www.ine.es/jaxiT3/Datos.htm?t=1414#!tabs-grafico>

Instituto Nacional de Estadística, (2021b). *Población que ha usado Internet de manera frecuente en los últimos tres meses por grupos de edad* [Fotografía]. INE.

[https://www.ine.es/jaxi/Datos.htm?path=/t00/mujeres\\_hombres/tablas\\_2/I0/&file=C5G1.px#!abs-grafico](https://www.ine.es/jaxi/Datos.htm?path=/t00/mujeres_hombres/tablas_2/I0/&file=C5G1.px#!abs-grafico)

McKinney, W. (2010). *Pandas* (Versión 1.3.5) [Software de computador]. Pandas Development Team. <https://pandas.pydata.org/>

Meirelles, I. (2013 octubre). *Design for information*. Rockport Publishers.

Merino, A. (2022). *Corre, que se van las vitaminas*. elordenmundial.

<https://elordenmundial.com/mapas-y-graficos/el-consumo-de-frutas-y-verduras-en-la-union-europea/>

Naranjo Solís, J.F. (2020). *Visualizaciones de datos: una propuesta para facilitar el análisis de información empresarial*. Universidad de Sevilla. <https://hdl.handle.net/11441/114805>

Olmeda-Gómez, C. (2014). *Visualización de la Información*. Profesional de la Información. <https://doi.org/10.3145/epi.2014.may.01>

Orús, A. (2022 2 febrero). *Distribución porcentual de individuos que leyeron periódicos de tirada diaria en España en 2021, por grupo de edad* [Fotografía]. Statista.

<https://es.statista.com/estadisticas/476755/distribucion-de-lectores-de-periodicos-en-espana-por-edad/>

Otto, M. y Thornton, J. (2011). *Bootstrap* (Versión 5.1.3) [Software de computador]. Twitter.  
<https://getbootstrap.com/>

Pellicer, L. (2018 diciembre 7). *La corrupción le cuesta a la UE más de 900.000 millones al año, según los Verdes*. El País.

[https://elpais.com/economia/2018/12/06/actualidad/1544123736\\_288597.html](https://elpais.com/economia/2018/12/06/actualidad/1544123736_288597.html)

Pérez, Fernando. (2014). *Jupyter Notebook* [Software de computador]. Proyecto Jupyter.  
<https://jupyter.org/>

Preston Werner, T. Wanstrath, C. Hyett, P. y Chacon S. (2010). *GitHubPages* [Software de computador]. GitHub. <https://pages.github.com/>

R. Pérez, G. (2022 febrero 2). *El paro crece en enero en 17.173 personas, el menor ritmo en este mes en 24 años*. El País. <https://elpais.com/economia/2022-02-02/el-paro-crece-en-enero-en-17173-personas-el-menor-ritmo-en-este-mes-en-24-anos.html>

Reitz, K. (2011). *Request* (Versión 2.26) [Software de computador]. Apache.  
<https://docs.python-requests.org/en/latest/>

Richardson, L. (2012). *BeautifulSoup* (Versión 4.6.0) [Software de computador]. Crummy.  
<https://www.crummy.com/software/BeautifulSoup/>

Rienties, B., Herodotou, C., Olney, T., Schencks, M., Boroowa, A. (2018 noviembre). *Making Sense of Learning Analytics Dashboards*. International Review of Research in Open and Distributed Learning. <https://files.eric.ed.gov/fulltext/EJ1198041.pdf>

Romero Caro, I. (2021 diciembre 14). *Suicidios y salud mental en España, al descubierto: cifras y evolución de una epidemia tabú*. Onda Cero.  
[https://www.ondacero.es/noticias/sociedad/suicidios-espana-descubierto-datos-cifras-evolucion\\_2021121461b88871454389000144c314.html#:~:text=En%20total%2C%203.941%20personas%20se,primera%20vez%20los%201.000%20suicidios](https://www.ondacero.es/noticias/sociedad/suicidios-espana-descubierto-datos-cifras-evolucion_2021121461b88871454389000144c314.html#:~:text=En%20total%2C%203.941%20personas%20se,primera%20vez%20los%201.000%20suicidios)

Roser, M., Ortiz-Espina, E. y Ritchie, H. (2019). *Life Expectancy* [Fotografía]. OurWorldInData. <https://ourworldindata.org/life-expectancy>

Ruiz, M. (2022 enero 27). *Las claves del paro: España sigue liderando el desempleo en Europa a pesar de la bajada histórica de 2021*. Cope.  
[https://www cope.es/actualidad/economia/noticias/las-claves-del-paro-espana-sigue-liderando-desempleo-europa-pesar-bajada-historica-2021-20220127\\_1751981](https://www cope.es/actualidad/economia/noticias/las-claves-del-paro-espana-sigue-liderando-desempleo-europa-pesar-bajada-historica-2021-20220127_1751981)

Stolte, C. Hanrahan, P. Chabot, C. (2003) *Tableau* (Versión 2021.4) [Software de computador]. Salesforce. <https://www.tableau.com/>

Sutherland, S., Ridgway, J. (2017 mayo). *Interactive Visualisations and Statistical Literacy*. Statistics Education Research Journal. [http://iase-web.org/documents/SERJ/SERJ16\(1\)\\_Sutherland.pdf](http://iase-web.org/documents/SERJ/SERJ16(1)_Sutherland.pdf)

Therón Sanchez, R. (2021 octubre). *Visualización de datos: caminos de ida y vuelta entre arte y ciencia en la producción y consumo de imágenes*. Universidad de Salamanca. <http://hdl.handle.net/10366/147775>

Unión Europea. (2021). *Prioridades de la Unión*. Web oficial de la Unión Europea. [https://european-union.europa.eu/priorities-and-actions/eu-priorities\\_es](https://european-union.europa.eu/priorities-and-actions/eu-priorities_es)

Valero Sancho, José Luis. (2014). *La visualización de datos*. Ámbitos: Revista Internacional de Comunicación. <http://hdl.handle.net/11441/66886>

Vázquez Ingelmo, Andrea. (2018 julio). *Aplicación de ingeniería de dominio para la generación de dashboards personalizados*. Universidad de Salamanca. <http://repositorio.grial.eu/handle/grial/1300>

## 7 Anexo

A continuación, se exponen los archivos que se han construido para la elaboración del trabajo.

### 7.1 Script ETL – España:

Trabajo de Fin de Máster: Metodología para la visualización interactiva de datos sobre un país de la Unión Europea.

Este documento forma parte del desarrollo de un trabajo de fin de grado realizado en la Universidad Internacional de La Rioja (UNIR).

```
# importación de librerías
from bs4 import BeautifulSoup
import requests
import pandas as pd

España

PIB Anual

url="https://datosmacro.expansion.com/pib/españa"

# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[2].tbody.find_all("tr")
df_pib_esp = pd.DataFrame(columns = ['FECHA' , 'PIB',
'PIB_VAR_PORCENTAJE'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
```

```

if values[0] != '0' and values[1] != '0' and values[3] != '0':
    df_pib_esp = df_pib_esp.append({'FECHA' : values[0], 'PIB' :
values[1], 'PIB_VAR_PORCENTAJE' : values[3]}, ignore_index=True)
    i = i + 1

df_pib_esp[['PIB', 'PIB_VAR_PORCENTAJE']] = df_pib_esp[['PIB',
'PIB_VAR_PORCENTAJE']].astype('float')
df_pib_esp['FECHA'] = pd.to_datetime(df_pib_esp.FECHA)
df_pib_esp['FECHA'] = df_pib_esp['FECHA'].dt.year
# df_pib_esp.to_excel('Datos/ESP/pib.xlsx', index=False)
df_pib_esp

df_pib_esp.info()

```

### Evolución de la Deuda

```
url="https://datosmacro.expansion.com/deuda/espana"
```

```

# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[1].tbody.find_all("tr")
df_deuda_esp = pd.DataFrame(columns = ['FECHA', 'DEUDA_TOTAL',
'DEUDA_PIB_PORCENTAJE', 'DEUDA_PER_CAPITA'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0' and values[2] != '0' and
values[3] != '0':
        df_deuda_esp = df_deuda_esp.append({'FECHA' : values[0],
'DEUDA_TOTAL' : values[1], 'DEUDA_PIB_PORCENTAJE' : values[2],

```

```
'DEUDA_PER_CAPITA' : values[3]} , ignore_index=True)
i = i + 1

df_deuda_esp[['DEUDA_TOTAL', 'DEUDA_PIB_PORCENTAJE', 'DEUDA_PER_CAPITA']]
= df_deuda_esp[['DEUDA_TOTAL', 'DEUDA_PIB_PORCENTAJE',
'DEUDA_PER_CAPITA']].astype(float)
df_deuda_esp['FECHA'] = pd.to_datetime(df_deuda_esp.FECHA)
df_deuda_esp['FECHA'] = df_deuda_esp['FECHA'].dt.year
# df_deuda_esp.to_excel('Datos/ESP/deuda.xlsx', index=False)
df_deuda_esp
```

Evolución del déficit

```
url="https://datosmacro.expansion.com/deficit/espana"
```

```
# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[1].tbody.find_all("tr")
df_deficit_esp = pd.DataFrame(columns = ['FECHA', 'DEFICIT',
'DEFICIT_PIB_PORCENTAJE'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0' and values[3] != '0':
        df_deficit_esp = df_deficit_esp.append({'FECHA' : values[0],
'DEFICIT' : values[1], 'DEFICIT_PIB_PORCENTAJE' : values[3]}, ignore_index=True)
    i = i + 1

df_deficit_esp[['DEFICIT', 'DEFICIT_PIB_PORCENTAJE']] =
df_deficit_esp[['DEFICIT', 'DEFICIT_PIB_PORCENTAJE']].astype(float)
```

```

df_deficit_esp['FECHA'] = pd.to_datetime(df_deficit_esp.FECHA)
df_deficit_esp['FECHA'] = df_deficit_esp['FECHA'].dt.year
# df_deficit_esp.to_excel('Datos/ESP/deficit.xlsx', index=False)
df_deficit_esp

Gasto público en educación

url="https://datosmacro.expansion.com/estado/gasto/educacion/espana"

# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")
df_gasto_edu_esp = pd.DataFrame(columns = ['FECHA' , 'EDU_GASTO_TOTAL',
'EDU_GASTO_PIB_PORCENTAJE'])
i = 0

for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0' and values[4] != '0':
        df_gasto_edu_esp = df_gasto_edu_esp.append({'FECHA' : values[0] ,
'EDU_GASTO_TOTAL' : values[1], 'EDU_GASTO_PIB_PORCENTAJE' : values[4]} ,
ignore_index=True)
    i = i + 1

df_gasto_edu_esp[['EDU_GASTO_TOTAL', 'EDU_GASTO_PIB_PORCENTAJE']] =
df_gasto_edu_esp[['EDU_GASTO_TOTAL',
'EDU_GASTO_PIB_PORCENTAJE']].astype(float)
df_gasto_edu_esp['FECHA'] = pd.to_datetime(df_gasto_edu_esp.FECHA)
df_gasto_edu_esp['FECHA'] = df_gasto_edu_esp['FECHA'].dt.year
# df_gasto_edu_esp.to_excel('Datos/ESP/gasto_educacion.xlsx', index=False)
df_gasto_edu_esp

```

## Gasto público en salud

```
url="https://datosmacro.expansion.com/estado/gasto/salud/espana"

# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")
df_gasto_salud_esp = pd.DataFrame(columns = ['FECHA' ,
'SALUD_GASTO_TOTAL', 'SALUD_GASTO_PORCENTAJE_PIB'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[2] != '0' and values[5] != '0':
        df_gasto_salud_esp = df_gasto_salud_esp.append({'FECHA' :
values[0], 'SALUD_GASTO_TOTAL' : values[2], 'SALUD_GASTO_PORCENTAJE_PIB' :
values[5]}, ignore_index=True)
    i = i + 1

df_gasto_salud_esp[['SALUD_GASTO_TOTAL', 'SALUD_GASTO_PORCENTAJE_PIB']] =
df_gasto_salud_esp[['SALUD_GASTO_TOTAL',
'SALUD_GASTO_PORCENTAJE_PIB']].astype(float)
df_gasto_salud_esp['FECHA'] = pd.to_datetime(df_gasto_salud_esp.FECHA)
df_gasto_salud_esp['FECHA'] = df_gasto_salud_esp['FECHA'].dt.year
# df_gasto_salud_esp.to_excel('Datos/ESP/gasto_salud.xlsx', index=False)
df_gasto_salud_esp
```

## Gasto público en defensa

```
url="https://datosmacro.expansion.com/estado/gasto/defensa/espana"

# Petición GET del documento en "crudo" en formato HTML
```

```

html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")
df_gasto_def_esp = pd.DataFrame(columns = [ 'FECHA' , 'DEF_GASTO_TOTAL',
'DEF_GASTO_PORCENTAJE_PIB'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0' and values[4] != '0':
        df_gasto_def_esp = df_gasto_def_esp.append({ 'FECHA' : values[0] ,
'DEF_GASTO_TOTAL' : values[1], 'DEF_GASTO_PORCENTAJE_PIB' : values[4]} ,
ignore_index=True)
    i = i + 1

df_gasto_def_esp[['DEF_GASTO_TOTAL', 'DEF_GASTO_PORCENTAJE_PIB']] =
df_gasto_def_esp[['DEF_GASTO_TOTAL',
'DEF_GASTO_PORCENTAJE_PIB']].astype(float)
df_gasto_def_esp['FECHA'] = pd.to_datetime(df_gasto_def_esp.FECHA)
df_gasto_def_esp['FECHA'] = df_gasto_def_esp['FECHA'].dt.year
# df_gasto_def_esp.to_excel('Datos/ESP/gasto_defensa.xlsx', index=False)
df_gasto_def_esp

SMI

df_smi_esp = pd.read_csv('Datos/ESP/ext_manual/smi.csv')
df_smi_esp[['SMI']] = df_smi_esp[['SMI']].astype(float)
df_smi_esp

# df_smi_esp.to_excel('Datos/ESP/smi_esp.xlsx', index=False)

IPC

```

```
df_ipc = pd.read_csv('Datos/ESP/ext_manual/ipc.csv')
df_ipc[['IPC']] = df_ipc[['IPC']].astype(float)
df_ipc

# df_ipc.to_excel('Datos/ESP/ipc.xlsx', index=False)
```

Tasa de desempleo

```
df_desempleo = pd.read_csv('Datos/ESP/ext_manual/desempleo.csv')
df_desempleo[['TASA_DESEMPLEO']] =
df_desempleo[['TASA_DESEMPLEO']].astype(float)
df_desempleo

# df_desempleo.to_excel('Datos/ESP/desempleo.xlsx', index=False)
```

Salario medio

```
url="https://datosmacro.expansion.com/mercado-laboral/salario-
medio/espana"
```

```
# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")
df_salario_medio_esp = pd.DataFrame(columns = ['FECHA' , 'SALARIO'])
i = 0

for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0':
        df_salario_medio_esp = df_salario_medio_esp.append({'FECHA' :
values[0] , 'SALARIO' : values[1]} , ignore_index=True)
    i = i + 1
```

```
df_salario_medio_esp[['SALARIO']] =
df_salario_medio_esp[['SALARIO']].astype(float)
df_salario_medio_esp['FECHA'] = pd.to_datetime(df_salario_medio_esp.FECHA)
df_salario_medio_esp['FECHA'] = df_salario_medio_esp['FECHA'].dt.year
# df_salario_medio_esp.to_excel('Datos/ESP/salario_medio.xlsx',
index=False)
df_salario_medio_esp
```

Índice corrupción

```
url="https://datosmacro.expansion.com/estado/indice-percepcion-
corrupcion/espana"
```

*# Petición GET del documento en "crudo" en formato HTML*

```
html_content = requests.get(url).text
```

*# Análisis por etiqueta del contenido html*

```
soup = BeautifulSoup(html_content, "lxml")
```

```
tablas = soup.find_all('table')
```

```
tr = tablas[0].tbody.find_all("tr")
```

```
df_corrupcion_esp = pd.DataFrame(columns = ['FECHA' ,
'INDICE_CORRUPCION'])
```

```
i = 0
```

```
for i in range(len(tr)):
```

```
    values = []
```

```
    for td in tr[i].find_all("td"):
```

```
        values.append(td.get("data-value"))
```

```
    if values[0] != '0' and values[1] != '0':
```

```
        df_corrupcion_esp = df_corrupcion_esp.append({'FECHA' : values[0]
, 'INDICE_CORRUPCION' : values[1]} , ignore_index=True)
```

```
    i = i + 1
```

```
df_corrupcion_esp[['INDICE_CORRUPCION']] =
```

```
df_corrupcion_esp[['INDICE_CORRUPCION']].astype(float)
```

```
df_corrupcion_esp['FECHA'] = pd.to_datetime(df_corrupcion_esp.FECHA)
```

```
df_corrupcion_esp['FECHA'] = df_corrupcion_esp['FECHA'].dt.year
```

```
# df_corrupcion_esp.to_excel('Datos/ESP/corrucion.xlsx', index=False)
df_corrucion_esp

Comercio

url="https://datosmacro.expansion.com/comercio/exportaciones/espana"

# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")
df_export_esp = pd.DataFrame(columns = ['FECHA' , 'EXPORTACIONES',
'EXPORT_PORCENTAJE_PIB'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0' and values[3] != '0':
        df_export_esp = df_export_esp.append({'FECHA' : values[0] ,
'EXPORTACIONES' : values[1], 'EXPORT_PORCENTAJE_PIB' : values[3]} ,
ignore_index=True)
    i = i + 1

df_export_esp[['EXPORTACIONES','EXPORT_PORCENTAJE_PIB']] =
df_export_esp[['EXPORTACIONES','EXPORT_PORCENTAJE_PIB']].astype(float)
df_export_esp['FECHA'] = pd.to_datetime(df_export_esp.FECHA)
df_export_esp['FECHA'] = df_export_esp['FECHA'].dt.year
# df_export_esp.to_excel('Datos/ESP/exportaciones.xlsx', index=False)
df_export_esp

url="https://datosmacro.expansion.com/comercio/importaciones/espana"

# Petición GET del documento en "crudo" en formato HTML
```

```
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")
df_import_esp = pd.DataFrame(columns = ['FECHA' , 'IMPORTACIONES',
'IMPORT_PORCENTAJE_PIB'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0' and values[3] != '0':
        df_import_esp = df_import_esp.append({'FECHA' : values[0] ,
'IMPORTACIONES' : values[1], 'IMPORT_PORCENTAJE_PIB' : values[3]} ,
ignore_index=True)
    i = i + 1

df_import_esp[['IMPORTACIONES','IMPORT_PORCENTAJE_PIB']] =
df_import_esp[['IMPORTACIONES','IMPORT_PORCENTAJE_PIB']].astype(float)
df_import_esp['FECHA'] = pd.to_datetime(df_import_esp.FECHA)
df_import_esp['FECHA'] = df_import_esp['FECHA'].dt.year
# df_import_esp.to_excel('Datos/ESP/importaciones.xlsx', index=False)
df_import_esp

url="https://datosmacro.expansion.com/comercio/balanza/espana"

# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')
```

```

tr = tablas[0].tbody.find_all("tr")
df_balanza_esp = pd.DataFrame(columns = [ 'FECHA' , 'BALANZA_COMERCIO',
'BALANZA_PIB_PORCENTAJE'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0' and values[4] != '0':
        df_balanza_esp = df_balanza_esp.append({ 'FECHA' : values[0] ,
'BALANZA_COMERCIO' : values[1], 'BALANZA_PIB_PORCENTAJE' : values[4]} ,
ignore_index=True)
    i = i + 1

df_balanza_esp[['BALANZA_COMERCIO','BALANZA_PIB_PORCENTAJE']] =
df_balanza_esp[['BALANZA_COMERCIO','BALANZA_PIB_PORCENTAJE']].astype(float)
df_balanza_esp['FECHA'] = pd.to_datetime(df_balanza_esp.FECHA)
df_balanza_esp['FECHA'] = df_balanza_esp['FECHA'].dt.year
# df_balanza_esp.to_excel('Datos/ESP/balanza.xlsx', index=False)
df_balanza_esp

```

Natalidad

```
url="https://datosmacro.expansion.com/demografia/natalidad/espana"
```

```
# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text
```

```
# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")
```

```
tablas = soup.find_all('table')
```

```

tr = tablas[0].tbody.find_all("tr")
df_natalidad_esp = pd.DataFrame(columns = [ 'FECHA' , 'NACIDOS',
'NACIDOS_HOMBRES','NACIDOS_MUJERES','IND_FECUNDACION'])
i = 0

```

```

for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0' and values[2] != '0' and
values[3] != '0' and values[4] != '0':
        df_natalidad_esp = df_natalidad_esp.append({'FECHA' : values[0] ,
'NACIDOS' : values[1], 'NACIDOS_HOMBRES' : values[2], 'NACIDOS_MUJERES' :
values[3], 'IND_FECUNDACION' : values[4]}, ignore_index=True)
    i = i + 1

df_natalidad_esp[['NACIDOS','NACIDOS_HOMBRES','NACIDOS_MUJERES','IND_FECUN
DACION']] =
df_natalidad_esp[['NACIDOS','NACIDOS_HOMBRES','NACIDOS_MUJERES','IND_FECUN
DACION']].astype(float)
df_natalidad_esp['FECHA'] = pd.to_datetime(df_natalidad_esp.FECHA)
df_natalidad_esp['FECHA'] = df_natalidad_esp['FECHA'].dt.year
# df_natalidad_esp.to_excel('Datos/ESP/natalidad.xlsx', index=False)
df_natalidad_esp

```

Mortalidad

```
url="https://datosmacro.expansion.com/demografia/mortalidad/espana"
```

```

# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")
df_mortalidad_esp = pd.DataFrame(columns = ['FECHA' , 'MUERTES',
'MUERTES_HOMBRES','MUERTES_MUJERES','TASA_MORTALIDAD'])
i = 0
for i in range(len(tr)):
    values = []

```

```

for td in tr[i].find_all("td"):
    values.append(td.get("data-value"))
if values[0] != '0' and values[1] != '0' and values[2] != '0' and
values[3] != '0' and values[4] != '0':
    df_mortalidad_esp = df_mortalidad_esp.append({'FECHA' : values[0]
, 'MUERTES' : values[1], 'MUERTES_HOMBRES' : values[2], 'MUERTES_MUJERES'
: values[3], 'TASA_MORTALIDAD' : values[4]} , ignore_index=True)
    i = i + 1

df_mortalidad_esp[['MUERTES',
'MUERTES_HOMBRES','MUERTES_MUJERES','TASA_MORTALIDAD']] =
df_mortalidad_esp[['MUERTES',
'MUERTES_HOMBRES','MUERTES_MUJERES','TASA_MORTALIDAD']].astype(float)
df_mortalidad_esp['FECHA'] = pd.to_datetime(df_mortalidad_esp.FECHA)
df_mortalidad_esp['FECHA'] = df_mortalidad_esp['FECHA'].dt.year
# df_mortalidad_esp.to_excel('Datos/ESP/mortalidad.xlsx', index=False)
df_mortalidad_esp

```

Población

```
url="https://datosmacro.expansion.com/demografia/poblacion/espana"
```

```
# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text
```

```
# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")
```

```
tablas = soup.find_all('table')
```

```

tr = tablas[0].tbody.find_all("tr")
df_poblacion_esp = pd.DataFrame(columns = ['FECHA' , 'DENSIDAD',
'HOMBRES_POB','MUJERES_POB','POBLACION_TOTAL'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))

```

```

    if values[0] != '0' and values[1] != '0' and values[2] != '0' and
values[3] != '0' and values[4] != '0':
    df_poblacion_esp = df_poblacion_esp.append({'FECHA' : values[0] ,
'DENSIDAD' : values[1], 'HOMBRES_POB' : values[2], 'MUJERES_POB' :
values[3], 'POBLACION_TOTAL' : values[4]} , ignore_index=True)
    i = i + 1

df_poblacion_esp[['DENSIDAD',
'HOMBRES_POB','MUJERES_POB','POBLACION_TOTAL']] =
df_poblacion_esp[['DENSIDAD',
'HOMBRES_POB','MUJERES_POB','POBLACION_TOTAL']].astype(float)
df_poblacion_esp['FECHA'] = pd.to_datetime(df_poblacion_esp.FECHA)
df_poblacion_esp['FECHA'] = df_poblacion_esp['FECHA'].dt.year
# df_poblacion_esp.to_excel('Datos/ESP/poblacion.xlsx', index=False)
df_poblacion_esp

```

Riesgo de pobreza

```
url="https://datosmacro.expansion.com/demografia/riesgo-pobreza/espana"
```

```

# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")
df_pobreza_esp = pd.DataFrame(columns = ['FECHA' , 'RIESGO_POBREZA'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[6] != '0':
        df_pobreza_esp = df_pobreza_esp.append({'FECHA' : values[0] ,

```

```
'RIESGO_POBREZA' : values[6]} , ignore_index=True)
    i = i + 1

df_pobreza_esp[['RIESGO_POBREZA']] =
df_pobreza_esp[['RIESGO_POBREZA']].astype(float)
df_pobreza_esp['FECHA'] = pd.to_datetime(df_pobreza_esp.FECHA)
df_pobreza_esp['FECHA'] = df_pobreza_esp['FECHA'].dt.year
# df_pobreza_esp.to_excel('Datos/ESP/pobreza.xlsx', index=False)
df_pobreza_esp

Esperanza de vida

url="https://datosmacro.expansion.com/demografia/esperanza-vida/espana"

# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")
df_vida_esp = pd.DataFrame(columns = ['FECHA' , 'ESP_MUJERES',
'ESP_HOMBRES' , 'ESPERANZA_VIDA'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0' and values[2] != '0' and
values[3] != '0':
        df_vida_esp = df_vida_esp.append({'FECHA' : values[0] ,
'ESP_MUJERES' : values[1], 'ESP_HOMBRES' : values[2], 'ESPERANZA_VIDA' :
values[3]} , ignore_index=True)
    i = i + 1

df_vida_esp[['ESP_MUJERES', 'ESP_HOMBRES', 'ESPERANZA_VIDA']] =
df_vida_esp[['ESP_MUJERES', 'ESP_HOMBRES',
```

```
'ESPERANZA_VIDA']].astype(float)
df_vida_esp['FECHA'] = pd.to_datetime(df_vida_esp.FECHA)
df_vida_esp['FECHA'] = df_vida_esp['FECHA'].dt.year
# df_vida_esp.to_excel('Datos/ESP/esperanza_vida.xlsx', index=False)
df_vida_esp

Índice desarrollo humano

url="https://datosmacro.expansion.com/idh/espana"

# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")
df_idh_esp = pd.DataFrame(columns = ['FECHA' , 'IDH'])
i = 0

for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0':
        df_idh_esp = df_idh_esp.append({'FECHA' : values[0] , 'IDH' : values[1]} , ignore_index=True)
    i = i + 1

df_idh_esp[['IDH']] = df_idh_esp[['IDH']].astype(float)
df_idh_esp['FECHA'] = pd.to_datetime(df_idh_esp.FECHA)
df_idh_esp['FECHA'] = df_idh_esp['FECHA'].dt.year
# df_idh_esp.to_excel('Datos/ESP/idh.xlsx', index=False)
df_idh_esp
```

Suicidios

```
url="https://datosmacro.expansion.com/demografia/mortalidad/causas-  
muerte/suicidio/espana"

# Petición GET del documento en "crudo" en formato HTML  
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html  
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")
df_suicidio_esp = pd.DataFrame(columns = ['FECHA' ,  
'S_MUJERES','S_HOMBRES','SUICIDIOS','TASA_S_MUJERES','TASA_S_HOMBRES','SUI  
CIDIOS_POR_100K'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0' and values[2] != '0' and  
values[3] != '0' and values[4] != '0' and values[5] != '0' and values[6]  
!= '0':
        df_suicidio_esp = df_suicidio_esp.append({'FECHA' : values[0] ,  
'S_MUJERES' : values[1], 'S_HOMBRES' : values[2], 'SUICIDIOS' : values[3],  
'TASA_S_MUJERES' : values[4], 'TASA_S_HOMBRES' : values[5],  
'SUICIDIOS_POR_100K' : values[6]} , ignore_index=True)
    i = i + 1

df_suicidio_esp[['S_MUJERES','S_HOMBRES','SUICIDIOS','TASA_S_MUJERES','TAS  
A_S_HOMBRES','SUICIDIOS_POR_100K']] =
df_suicidio_esp[['S_MUJERES','S_HOMBRES','SUICIDIOS','TASA_S_MUJERES','TAS  
A_S_HOMBRES','SUICIDIOS_POR_100K']].astype(float)
df_suicidio_esp['FECHA'] = pd.to_datetime(df_suicidio_esp.FECHA)
df_suicidio_esp['FECHA'] = df_suicidio_esp['FECHA'].dt.year
# df_suicidio_esp.to_excel('Datos/ESP/suicidio.xlsx', index=False)
df_suicidio_esp
```

## Homicidios

```
url="https://datosmacro.expansion.com/demografia/homicidios/espana"

# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")
df_homicidio_esp = pd.DataFrame(columns =
['FECHA','HOMICIDIOS','HOMICIDIOS_POR_100K'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0' and values[6] != '0':
        df_homicidio_esp = df_homicidio_esp.append({'FECHA' : values[0] ,
'HOMICIDIOS' : values[1], 'HOMICIDIOS_POR_100K' : values[6]} ,
ignore_index=True)
    i = i + 1

df_homicidio_esp[['HOMICIDIOS','HOMICIDIOS_POR_100K']] =
df_homicidio_esp[['HOMICIDIOS','HOMICIDIOS_POR_100K']].astype(float)
df_homicidio_esp['FECHA'] = pd.to_datetime(df_homicidio_esp.FECHA)
df_homicidio_esp['FECHA'] = df_homicidio_esp['FECHA'].dt.year
# df_homicidio_esp.to_excel('Datos/ESP/homicidio.xlsx', index=False)
df_homicidio_esp
```

## Emisiones CO2

```
url="https://datosmacro.expansion.com/energia-y-medio-ambiente/emisiones-
co2/espana"
```

```
# Petición GET del documento en "crudo" en formato HTML
```

```

html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")
df_emisiones_esp = pd.DataFrame(columns =
['FECHA','CO2_TOTALES','CO2_PER_CAPITA'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0' and values[3] != '0':
        df_emisiones_esp = df_emisiones_esp.append({'FECHA' : values[0] ,
'CO2_TOTALES' : values[1], 'CO2_PER_CAPITA' : values[3]} , ignore_index=True)
    i = i + 1

df_emisiones_esp[['CO2_TOTALES','CO2_PER_CAPITA']] =
df_emisiones_esp[['CO2_TOTALES','CO2_PER_CAPITA']].astype(float)
df_emisiones_esp['FECHA'] = pd.to_datetime(df_emisiones_esp.FECHA)
df_emisiones_esp['FECHA'] = df_emisiones_esp['FECHA'].dt.year
# df_emisiones_esp.to_excel('Datos/ESP/emisiones.xlsx', index=False)
df_emisiones_esp

```

Consumo electricidad

```
url="https://datosmacro.expansion.com/energia-y-medio-ambiente/electricidad-consumo/espana"
```

```

# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

```

```

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")
df_consumo_gwh_esp = pd.DataFrame(columns =
['FECHA', 'GENERACION_GWh', 'CONSUMO_GWh', 'CONSUMO_PER_CAPITA'])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[1] != '0' and values[2] != '0' and
values[3] != '0':
        df_consumo_gwh_esp = df_consumo_gwh_esp.append({'FECHA' :
values[0], 'GENERACION_GWh' : values[1], 'CONSUMO_GWh' : values[2],
'CONSUMO_PER_CAPITA' : values[3]}, ignore_index=True)
    i = i + 1

df_consumo_gwh_esp[['GENERACION_GWh', 'CONSUMO_GWh', 'CONSUMO_PER_CAPITA']]
=
df_consumo_gwh_esp[['GENERACION_GWh', 'CONSUMO_GWh', 'CONSUMO_PER_CAPITA']].
astype(float)
df_consumo_gwh_esp['FECHA'] = pd.to_datetime(df_consumo_gwh_esp.FECHA)
df_consumo_gwh_esp['FECHA'] = df_consumo_gwh_esp['FECHA'].dt.year
# df_consumo_gwh_esp.to_excel('Datos/ESP/consumo_gwh.xlsx', index=False)
df_consumo_gwh_esp

url="https://datosmacro.expansion.com/energia-y-medio-
ambiente/electricidad-generacion/espana"

# Petición GET del documento en "crudo" en formato HTML
html_content = requests.get(url).text

# Análisis por etiqueta del contenido html
soup = BeautifulSoup(html_content, "lxml")

tablas = soup.find_all('table')

tr = tablas[0].tbody.find_all("tr")

```

```

df_renovables_esp = pd.DataFrame(columns =
[ 'FECHA' , 'GENERACION_RENOVABLES' ])
i = 0
for i in range(len(tr)):
    values = []
    for td in tr[i].find_all("td"):
        values.append(td.get("data-value"))
    if values[0] != '0' and values[4] != '0':
        df_renovables_esp = df_renovables_esp.append({ 'FECHA' : values[0]
, 'GENERACION_RENOVABLES' : values[4]} , ignore_index=True)
    i = i + 1

df_renovables_esp[['GENERACION_RENOVABLES']] =
df_renovables_esp[['GENERACION_RENOVABLES']].astype(float)
df_renovables_esp['FECHA'] = pd.to_datetime(df_renovables_esp.FECHA)
df_renovables_esp['FECHA'] = df_renovables_esp['FECHA'].dt.year
# df_renovables_esp.to_excel('Datos/ESP/renovables.xlsx', index=False)
df_renovables_esp

```

## Merge

```

join_type = "outer"
inner_merged = pd.merge(df_pib_esp, df_deuda_esp, how=join_type,
on=["FECHA"])
inner_merged1 = pd.merge(inner_merged, df_deficit_esp, how=join_type,
on=["FECHA"])
inner_merged2 = pd.merge(inner_merged1, df_gasto_edu_esp, how=join_type,
on=["FECHA"])
inner_merged3 = pd.merge(inner_merged2, df_gasto_salud_esp, how=join_type,
on=["FECHA"])
inner_merged4 = pd.merge(inner_merged3, df_gasto_def_esp, how=join_type,
on=["FECHA"])
inner_merged5 = pd.merge(inner_merged4, df_smi_esp, how=join_type,
on=["FECHA"])
inner_merged6 = pd.merge(inner_merged5, df_salario_medio_esp,
how=join_type, on=["FECHA"])
inner_merged7 = pd.merge(inner_merged6, df_ipc, how=join_type,

```

```
on=["FECHA"])
inner_merged8 = pd.merge(inner_merged7, df_corrupcion_esp, how=join_type,
on=["FECHA"])
inner_merged9 = pd.merge(inner_merged8, df_desempleo, how=join_type,
on=["FECHA"])
inner_merged10 = pd.merge(inner_merged9, df_export_esp, how=join_type,
on=["FECHA"])
inner_merged11 = pd.merge(inner_merged10, df_import_esp, how=join_type,
on=["FECHA"])
inner_merged12 = pd.merge(inner_merged11, df_balanza_esp, how=join_type,
on=["FECHA"])
inner_merged13 = pd.merge(inner_merged12, df_natalidad_esp, how=join_type,
on=["FECHA"])
inner_merged14 = pd.merge(inner_merged13, df_mortalidad_esp,
how=join_type, on=["FECHA"])
inner_merged15 = pd.merge(inner_merged14, df_pobreza_esp, how=join_type,
on=["FECHA"])
inner_merged16 = pd.merge(inner_merged15, df_vida_esp, how=join_type,
on=["FECHA"])
inner_merged17 = pd.merge(inner_merged16, df_idh_esp, how=join_type,
on=["FECHA"])
inner_merged18 = pd.merge(inner_merged17, df_suicidio_esp, how=join_type,
on=["FECHA"])
inner_merged19 = pd.merge(inner_merged18, df_homicidio_esp, how=join_type,
on=["FECHA"])
inner_merged20 = pd.merge(inner_merged19, df_emisiones_esp, how=join_type,
on=["FECHA"])
inner_merged21 = pd.merge(inner_merged20, df_consumo_gwh_esp,
how=join_type, on=["FECHA"])
inner_merged22 = pd.merge(inner_merged21, df_renovables_esp,
how=join_type, on=["FECHA"])
inner_merged23 = pd.merge(inner_merged22, df_poblacion_esp, how=join_type,
on=["FECHA"])
inner_merged23.to_csv('Datos/ESP/TOTALES.csv', index=False, sep=';',
decimal=',')
inner_merged23.info()
```

## 7.2 Script ETL – Unión Europea

Trabajo de Fin de Máster: Metodología para la visualización interactiva de datos sobre un país de la Unión Europea.

Este documento forma parte del desarrollo de un trabajo de fin de grado realizado en la Universidad Internacional de La Rioja (UNIR).

### *Importación de librerías*

```
import pandas as pd
import matplotlib.pyplot as plt
```

### *Variables estáticas*

```
countries = ['Germany', 'Austria', 'Belgium', 'Bulgaria', 'Cyprus',
'Croatia', 'Denmark', 'Slovakia', 'Slovenia', 'Spain', 'Estonia',
'Finland', 'France', 'Greece', 'Hungary', 'Ireland', 'Italy', 'Latvia',
'Lithuania', 'Luxembourg', 'Malta', 'Netherlands', 'Poland', 'Portugal',
'Czechia', 'Romania', 'Sweden']
min_year = 1946
```

### *General*

#### *Población*

Dataset con la población total en cada país de la Unión Europea cada año desde 1946 -  
COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país -  
YEAR: Año de la toma de datos - POPULATION: Total de población

```
df_population = pd.read_csv("Datos/UE/ext_manual/population-since-
1800.csv")

df_population = df_population[df_population.Entity.isin(countries)]
df_population_UE = df_population.rename(columns={'Entity': 'COUNTRY',
'Code': 'ISO_CODE', 'Year': 'YEAR', 'Population (historical estimates)': 'POPULATION'})

df_population_UE = df_population_UE[df_population_UE['YEAR'] >= min_year]
df_population_UE.to_excel('Datos/UE/population_UE.xlsx', index = False,
sheet_name='population_UE')
```

```
df_population_UE.head()
```

### *PIB - Producto interior bruto*

Dataset con el producto interior bruto por persona en cada país de la Unión Europea cada año desde 1946 - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - PIB\_PER\_CAPITA: Producto interior bruto por persona

```
df_GDP_per_capita = pd.read_csv("Datos/UE/ext_manual/gdp-per-capita-worldbank.csv")
```

```
df_GDP_per_capita_UE =
df_GDP_per_capita[df_GDP_per_capita.Entity.isin(countries)]
df_GDP_per_capita_UE = df_GDP_per_capita_UE.rename(columns={'Code':
'ISO_CODE', 'Entity': 'COUNTRY', 'Year': 'YEAR', 'GDP per capita, PPP
(constant 2017 international $)': 'PIB_PER_CAPITA'})
df_GDP_per_capita_UE = df_GDP_per_capita_UE[df_GDP_per_capita_UE['YEAR']
>= min_year]
df_GDP_per_capita_UE.to_excel('Datos/UE/GDP_per_capita_UE.xlsx', index =
False, sheet_name='GDP_per_capita_UE')
```

```
df_GDP_per_capita_UE.head()
```

### Paz y seguridad

#### *Muertes por conflictos y terrorismos*

Dataset con el número de muertes causadas por conflictos y terrorismos en los países de la Unión Europea desde 1990 - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - CONFLICT\_TERRORISM\_DEATHS: Número total de muertes causadas por conflictos o terrorismos.

```
df_death_by_war = pd.read_csv("Datos/UE/ext_manual/conflict-and-terrorism-
deaths-ihme-2017.csv")
```

```
df_deaths_UE = df_death_by_war[df_death_by_war.Entity.isin(countries)]
df_deaths_UE = df_deaths_UE.rename(columns={'Entity': 'COUNTRY', 'Code':
'ISO_CODE', 'Year': 'YEAR', 'Conflict and terrorism deaths (IHME 2017)':
```

```
'CONFLICT_TERRORISM_DEATHS'})  
df_deaths_UE = df_deaths_UE[df_deaths_UE['YEAR'] >= min_year]  
df_deaths_UE.to_excel('Datos/UE/df_deaths_UE.xlsx', index = False,  
sheet_name='df_deaths_UE')  
  
df_deaths_UE.head()
```

### *Gasto militar*

Dataset con el total de gasto militar por año y por persona en cada país de la Unión Europea desde 1960 - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos -  
 MILITARY\_EXPENDITURE\_GDP\_PERCENTAGE: Gasto militar en porcentaje respecto al PIB - MILITARY\_EXPENDITURE\_BILLION\_DOLLARS: Gasto militar en millones de dólares

```
df_military_expenditure = pd.read_csv("Datos/UE/ext_manual/military-  
expenditure-share-gdp-sipri.csv")  
df_military_expenditure2 = pd.read_csv("Datos/UE/ext_manual/military-  
expenditure-by-country-in-thousands-of-2000-us-dollars.csv")  
military_merged = pd.merge(df_military_expenditure,  
df_military_expenditure2, left_on=['Entity', 'Code', 'Year'],  
right_on=['Entity', 'Code', 'Year'])  
  
df_military_expenditure_UE =  
military_merged[military_merged.Entity.isin(countries)]  
df_military_expenditure_UE =  
df_military_expenditure_UE.rename(columns={'Entity': 'COUNTRY', 'Code':  
'ISO_CODE', 'Year': 'YEAR', 'Military expenditure (% of GDP)':  
'MILITARY_EXPENDITURE_GDP_PERCENTAGE', 'Military expenditure (1914-2007,  
real prices) (Correlates of War: National Material Capabilities (v4.0))':  
'MILITARY_EXPENDITURE_BILLION_DOLLARS'})  
df_military_expenditure_UE =  
df_military_expenditure_UE[df_military_expenditure_UE['YEAR'] >= min_year]  
df_military_expenditure_UE.to_excel('Datos/UE/military_expenditure_UE.xlsx  
, index = False, sheet_name='military_expenditure_UE')  
  
df_military_expenditure_UE.head()
```

### Salud, Alimentación y contaminación

## Territorios agrícolas

Dataset con el número total de hectáreas por persona dedicadas a fines agrícolas en los países de la Unión Europea desde el año 1961 - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - AGRICULTURAL\_LAND\_PER\_CAP: Hectáreas dedicadas a la agricultura por persona

```
df_agricultural_area_per_capita =
pd.read_csv("Datos/UE/ext_manual/agricultural-area-per-capita.csv")

df_agricultural_area_per_capita =
df_agricultural_area_per_capita[df_agricultural_area_per_capita.Entity.isin(countries)]
df_agricultural_area_per_capita_UE =
df_agricultural_area_per_capita.rename(columns={'Entity': 'COUNTRY',
'Code': 'ISO_CODE', 'Year': 'YEAR', 'agricultural_land_per_capita':
'AGRICULTURAL_LAND_PER_CAP'})
df_agricultural_area_per_capita_UE =
df_agricultural_area_per_capita_UE[df_agricultural_area_per_capita_UE['YEAR'] >= min_year]
df_agricultural_area_per_capita_UE.to_excel('Datos/UE/agricultural_area_per_capita_UE.xlsx', index = False,
sheet_name='agricultural_area_capita_UE')

df_agricultural_area_per_capita_UE.head()
```

## Consumo de grasas

Dataset con el consumo total de grasas por persona y día dentro de los países de la Unión Europea. - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - FAT\_QTY\_GR\_PER\_CAP\_DAY: Cantidad de grasa consumida por persona y día

```
df_daily_per_capita_fat_supply = pd.read_csv("Datos/UE/ext_manual/daily-
per-capita-fat-supply.csv")

df_daily_per_capita_fat_supply =
df_daily_per_capita_fat_supply[df_daily_per_capita_fat_supply.Entity.isin(
countries)]
```

```

df_daily_per_capita_fat_supply_UE =
df_daily_per_capita_fat_supply.rename(columns={'Entity': 'COUNTRY',
'Code': 'ISO_CODE', 'Year': 'YEAR', 'Fat supply quantity (g/capita/day)
(FAO, 2020)': 'FAT_QTY_GR_PER_CAP_DAY'})
df_daily_per_capita_fat_supply_UE =
df_daily_per_capita_fat_supply_UE[df_daily_per_capita_fat_supply_UE[ 'YEAR'
] >= min_year]
df_daily_per_capita_fat_supply_UE.to_excel('Datos/UE/daily_per_capita_fat_
supply_UE.xlsx', index = False, sheet_name='fat_supply_UE')

df_daily_per_capita_fat_supply_UE.head()

```

### *Consumo de proteínas*

Dataset con el consumo total de proteínas por persona y día dentro de los países de la Unión Europea. - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - PROTEIN\_QTY\_GR\_PER\_CAP\_DAY: Cantidad de proteínas consumida por persona y día

```

df_daily_per_capita_protein_supply =
pd.read_csv("Datos/UE/ext_manual/daily-per-capita-protein-supply.csv")

df_daily_per_capita_protein_supply =
df_daily_per_capita_protein_supply[df_daily_per_capita_protein_supply.Entity.isin(countries)]
df_daily_per_capita_protein_supply_UE =
df_daily_per_capita_protein_supply.rename(columns={'Entity': 'COUNTRY',
'Code': 'ISO_CODE', 'Year': 'YEAR', 'Protein supply quantity
(g/capita/day) (FAO, 2020)': 'PROTEIN_QTY_GR_PER_CAP_DAY'})
df_daily_per_capita_protein_supply_UE =
df_daily_per_capita_protein_supply_UE[df_daily_per_capita_protein_supply_U
E['YEAR'] >= min_year]
df_daily_per_capita_protein_supply_UE.to_excel('Datos/UE/daily_per_capita_
protein_supply_UE.xlsx', index = False, sheet_name='protein_supply_UE')

df_daily_per_capita_protein_supply_UE.head()

```

### *Esperanza de vida y consumo de calorías*

Dataset con información sobre la esperanza de vida y el consumo de calorías en los países pertenecientes a la Unión Europea desde el año 1950. - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - LIFE\_EXPECTANCY: Esperanza de vida - CALORIC\_SUPPLY: Consumo de calorías - POPULATION: Población - CONTINENT: Continente

```
df_food_supply = pd.read_csv("Datos/UE/ext_manual/food-supply-vs-life-expectancy.csv")

df_food_supply = df_food_supply[df_food_supply.Entity.isin(countries)]
df_food_supply_UE = df_food_supply.rename(columns={'Entity': 'COUNTRY',
'Code': 'ISO_CODE', 'Year': 'YEAR', 'Estimates, 1950 - 2020: Annually interpolated demographic indicators - Life expectancy at birth, both sexes combined (years)': 'LIFE_EXPECTANCY', 'Daily caloric supply (FAO (2017) & Various historical sources)': 'CALORIC_SUPPLY', 'Population (historical estimates)': 'POPULATION', 'Continent': 'CONTINENT'})

df_food_supply_UE = df_food_supply_UE[df_food_supply_UE['YEAR'] >= min_year]
df_food_supply_UE.to_excel('Datos/UE/df_food_supply_UE.xlsx', index = False, sheet_name='food_supply_UE')

df_food_supply_UE.head()
```

### *Retraso en el crecimiento*

Dataset con la cantidad de niños que sufren de retraso en el crecimiento en los países de la Unión Europea desde el año 1990 - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - CHILDREN\_STUNTED: Número de niños con retraso del crecimiento

```
df_children_stunted = pd.read_csv("Datos/UE/ext_manual/number-children-stunted-ihme.csv")

df_children_stunted =
df_children_stunted[df_children_stunted.Entity.isin(countries)]
df_children_stunted_UE = df_children_stunted.rename(columns={'Entity': 'COUNTRY', 'Code': 'ISO_CODE', 'Year': 'YEAR', 'Number of children who are stunted (IHME)': 'CHILDREN_STUNTED'})

df_children_stunted_UE =
```

```
df_children_stunted_UE[df_children_stunted_UE['YEAR'] >= min_year]
df_children_stunted_UE.to_excel('Datos/UE/children_stunted_UE.xlsx', index = False, sheet_name='children_stunted_UE')

df_children_stunted_UE.head()
```

### *Debilidad*

Dataset con la cantidad de niños que sufren de debilidad en los países de la Unión Europea desde el año 1990 - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - CHILDREN\_WASTED: Número de niños con debilidad

```
df_children_wasted = pd.read_csv("Datos/UE/ext_manual/number-children-wasted-ihme.csv")

df_children_wasted =
df_children_wasted[df_children_wasted.Entity.isin(countries)]
df_children_wasted_UE = df_children_wasted.rename(columns={'Entity': 'COUNTRY', 'Code': 'ISO_CODE', 'Year': 'YEAR', 'Number of children who are wasted (IHME)': 'CHILDREN_WASTED'})

df_children_wasted_UE =
df_children_wasted_UE[df_children_wasted_UE['YEAR'] >= min_year]
df_children_wasted_UE.to_excel('Datos/UE/children_wasted_UE.xlsx', index = False, sheet_name='children_wasted_UE')

df_children_wasted_UE.head()
```

### *Merge de salud en niños*

Construcción del dataset definitivo para la salud de los niños - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - CHILDREN\_STUNTED: Número de niños con retraso del crecimiento - CHILDREN\_WASTED: Número de niños con debilidad - POPULATION: Población - STUNTED\_PERCENTAGE: Número de niños con retraso del crecimiento respecto a la población - WASTED\_PERCENTAGE: Número de niños con debilidad respecto a la población - SUM\_SW: Suma de porcentajes anteriores

```

df_children_stunted_wasted = pd.merge(df_children_stunted_UE,
df_children_wasted_UE, how="inner", on=[ "COUNTRY", "ISO_CODE", "YEAR"])
df_stunted_wasted_population = pd.merge(df_children_stunted_wasted,
df_population_UE, how="inner", on=[ "COUNTRY", "ISO_CODE", "YEAR"])
df_stunted_wasted_population[ 'STUNTED_PERCENTAGE' ] =
df_stunted_wasted_population[ 'CHILDREN_STUNTED' ] /
df_stunted_wasted_population[ 'POPULATION' ] * 100
df_stunted_wasted_population[ 'WASTED_PERCENTAGE' ] =
df_stunted_wasted_population[ 'CHILDREN_WASTED' ] /
df_stunted_wasted_population[ 'POPULATION' ] * 100
df_stunted_wasted_population[ 'SUM_SW' ] =
df_stunted_wasted_population[ 'STUNTED_PERCENTAGE' ] +
df_stunted_wasted_population[ 'WASTED_PERCENTAGE' ]
df_stunted_wasted_population.to_excel('Datos/UE/stunted_wasted_population.
xlsx', index = False, sheet_name='df_stunted_wasted_population')
df_stunted_wasted_population.head()

```

### *Principales riesgos de muerte*

Dataset con la información referente a cuales son los principales factores de riesgo de muerte y cual es el impacto de cada una para los países de dentro de la Unión Europea - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - UNSAFE\_WATER: Agua no potable - UNSAFE\_SANITATION: Sanidad pobre - NO\_ACCESS\_HANDWASHING: Sin acceso al lavado de manos - AIR POLLUTION\_SOLID\_FUEL: Contaminación debida a combustibles sólidos - NON\_EXCLUSIVE\_BREASTFEEDING: Lactancia materna no exclusiva - DISCONTINUED\_BREASTFEEDING: Lactancia materna discontinuada - CHILD\_WASTING: Niños con debilidad - CHILD\_STUNTING: Niños con retraso del crecimiento - LOW\_BIRTH\_WEIGHT: Peso bajo al nacer - SECONDHAND\_SMOKE: Fumador pasivo - ALCOHOL\_USE: Abuso del alcohol - DRUG\_USE: Consumo de drogas - DIET\_LOW\_FRUITS: Dieta baja en frutas - DIET\_LOW\_VEGETABLES: Dieta baja en verduras - UNSAFE\_SEX: Sexo sin protección - LOW\_PHYSICAL\_ACTIVITY: Actividad física baja - FASTING\_PLASMA\_GLUCOSE: Glucosa plasmática en ayunas - HIGH\_COLESTEROL: Colesterol alto - HIGH\_BODYMASS\_INDEX: Índice de masa corporal alto - HIGH\_BLOOD\_PRESSURE: Presión arterial alta - SMOKING: Fumador -

IRON\_DEFICIENCY: Deficiencia de hierro - VITAMIN\_A\_DEFICIENCY: Deficiencia de vitamina A - LOW\_BONE\_MINERAL\_DENSITY: Baja densidad de minerales en los huesos - AIR POLLUTION: Contaminación en el aire - OUTDOOR\_AIR\_POLLUTION: Contaminación exterior en el aire - DIET\_HIGH\_SODIUM: Dieta alta en sal - DIET\_LOW\_GRAINS: Dieta baja en cereales - DIET\_LOW\_NUTS\_SEEDS: Dieta baja en semillas y nueces

```
df_deaths_by_risk_factor = pd.read_csv("Datos/UE/ext_manual/number-of-deaths-by-risk-factor.csv")

df_deaths_by_risk_factor =
df_deaths_by_risk_factor[df_deaths_by_risk_factor.Entity.isin(countries)]
df_deaths_by_risk_factor_UE =
df_deaths_by_risk_factor.rename(columns={'Entity': 'COUNTRY', 'Code':
'ISO_CODE', 'Year': 'YEAR', 'Deaths - Unsafe water source - Sex: Both -
Age: All Ages (Number)': 'UNSAFE_WATER', 'Deaths - Unsafe sanitation -
Sex: Both - Age: All Ages (Number)': 'UNSAFE_SANITATION', 'Deaths - No
access to handwashing facility - Sex: Both - Age: All Ages (Number)':
'NO_ACCESS_HANDWASHING', 'Deaths - Household air pollution from solid
fuels - Sex: Both - Age: All Ages (Number)': 'AIR_POLLUTION_SOLID_FUEL',
'Deaths - Non-exclusive breastfeeding - Sex: Both - Age: All Ages
(Number)': 'NON_EXCLUSIVE_BREASTFEEDING', 'Deaths - Discontinued
breastfeeding - Sex: Both - Age: All Ages (Number)':
'DISCONTINUED_BREASTFEEDING', 'Deaths - Child wasting - Sex: Both - Age:
All Ages (Number)': 'CHILD_WASTING', 'Deaths - Child stunting - Sex: Both
- Age: All Ages (Number)': 'CHILD_STUNTING', 'Deaths - Low birth weight
for gestation - Sex: Both - Age: All Ages (Number)': 'LOW_BIRTH_WEIGHT',
'Deaths - Secondhand smoke - Sex: Both - Age: All Ages (Number)':
'SECONDHAND_SMOKE', 'Deaths - Alcohol use - Sex: Both - Age: All Ages
(Number)': 'ALCOHOL_USE', 'Deaths - Drug use - Sex: Both - Age: All Ages
(Number)': 'DRUG_USE', 'Deaths - Diet low in fruits - Sex: Both - Age: All
Ages (Number)': 'DIET_LOW_FRUITS', 'Deaths - Diet low in vegetables - Sex:
Both - Age: All Ages (Number)': 'DIET_LOW_VEGETABLES', 'Deaths - Unsafe
sex - Sex: Both - Age: All Ages (Number)': 'UNSAFE_SEX', 'Deaths - Low
physical activity - Sex: Both - Age: All Ages (Number)':
'LOW_PHYSICAL_ACTIVITY', 'Deaths - High fasting plasma glucose - Sex: Both
```

```

- Age: All Ages (Number)': 'FASTING_PLASMA_GLUCOSE', 'Deaths - High total
cholesterol - Sex: Both - Age: All Ages (Number)': 'HIGH_COLESTEROL',
'Deaths - High body-mass index - Sex: Both - Age: All Ages (Number)': 
'HIGH_BODYMASS_INDEX', 'Deaths - High systolic blood pressure - Sex: Both
- Age: All Ages (Number)': 'HIGH_BLOOD_PRESSURE', 'Deaths - Smoking - Sex:
Both - Age: All Ages (Number)': 'SMOKING', 'Deaths - Iron deficiency -
Sex: Both - Age: All Ages (Number)': 'IRON_DEFICIENCY', 'Deaths - Vitamin
A deficiency - Sex: Both - Age: All Ages (Number)': 
'VITAMIN_A_DEFICIENCY', 'Deaths - Low bone mineral density - Sex: Both -
Age: All Ages (Number)': 'LOW_BONE_MINERAL_DENSITY', 'Deaths - Air
pollution - Sex: Both - Age: All Ages (Number)': 'AIR POLLUTION', 'Deaths
- Outdoor air pollution (all ages) (IHME)': 'OUTDOOR_AIR_POLLUTION',
'Deaths - Diet high in sodium - Sex: Both - Age: All Ages (Number)': 
'DIET_HIGH_SODIUM', 'Deaths - Diet low in whole grains - Sex: Both - Age:
All Ages (Number)': 'DIET_LOW_GRAINS', 'Deaths - Diet low in nuts and
seeds - Sex: Both - Age: All Ages (Number)': 'DIET_LOW_NUTS_SEEDS'})
df_deaths_by_risk_factor_UE =
df_deaths_by_risk_factor_UE[df_deaths_by_risk_factor_UE['YEAR'] >=
min_year]
df_deaths_by_risk_factor_UE.to_excel('Datos/UE/deaths_by_risk_factor_UE.xls',
index = False, sheet_name='deaths_by_risk_factor_UE')

df_deaths_by_risk_factor_UE.head()

```

### *Muertes por obesidad*

Dataset con la información referente al porcentaje de muertos a causa de la obesidad para países de la Unión Europea desde el año 1990 - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - OBESITY\_PERCENTAGE: Porcentaje de obesidad

```

df_share_of_deaths_obesity = pd.read_csv("Datos/UE/ext_manual/share-of-
deaths-obesity.csv")

df_share_of_deaths_obesity =
df_share_of_deaths_obesity[df_share_of_deaths_obesity.Entity.isin(countrie
s)]
df_share_of_deaths_obesity_UE =

```

```
df_share_of_deaths_obesity.rename(columns={'Entity': 'COUNTRY', 'Code': 'ISO_CODE', 'Year': 'YEAR', 'Obesity (IHME, 2019)': 'OBESITY_PERCENTAGE'})
df_share_of_deaths_obesity_UE =
df_share_of_deaths_obesity_UE[df_share_of_deaths_obesity_UE['YEAR'] >=
min_year]
df_share_of_deaths_obesity_UE.to_excel('Datos/UE/share_of_deaths_obesity_U
E.xlsx', index = False, sheet_name='share_of_deaths_obesity_UE')

df_share_of_deaths_obesity_UE.head()
```

### *Principales componentes de la dieta*

Dataset con la información de los componentes principales de la dieta en los países de la Unión Europea desde el año 1961 - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - Other (FAO (2017)) (kilocalories per person per day): Otros - Sugar (FAO (2017)) (kilocalories per person per day): Azucar - Oils & Fats (FAO (2017)) (kilocalories per person per day): Aceites y grasas - Meat (FAO (2017)) (kilocalories per person per day): Carne - Dairy & Eggs (FAO (2017)) (kilocalories per person per day): Huevos y lácteos - Fruit and Vegetables (FAO (2017)) (kilocalories per person per day): Frutas y verduras - Starchy Roots (FAO (2017)) (kilocalories per person per day): Raíces almidonadas - Pulses (FAO (2017)) (kilocalories per person per day): Legumbres - Cereals and Grains (FAO (2017)) (kilocalories per person per day): Granos y cereales - Alcoholic Beverages (FAO (2017)) (kilocalories per person per day): Bebidas alcohólicas

```
df_dietary_composition = pd.read_csv("Datos/UE/ext_manual/dietary-
compositions-by-commodity-group.csv")

df_dietary_composition =
df_dietary_composition[df_dietary_composition.Entity.isin(countries)]
df_dietary_composition_UE =
df_dietary_composition.rename(columns={'Entity': 'COUNTRY', 'Code': 'ISO_CODE', 'Year': 'YEAR'})
df_dietary_composition_UE =
df_dietary_composition_UE[df_dietary_composition_UE['YEAR'] >= min_year]
df_dietary_composition_UE.to_excel('Datos/UE/dietary_composition_UE.xlsx',
index = False, sheet_name='dietary_composition_UE')
```

```
df_dietary_composition_UE.head()
```

### *Emisiones de CO2*

Dataset con el total de emisiones de CO2 en toneladas y por persona para los países de Unión Europea desde el año 1946 - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - CO2\_EMISSIONS\_TONNES: Emisiones de CO2 en toneladas - CO2\_EMISSIONS\_TONNES\_PER\_CAPITA: Emisiones de CO2 en toneladas por persona - POPULATION: Población

```
df_co2 = pd.read_csv("Datos/UE/ext_manual/owid-co2-data.csv")

df_co2 = df_co2[['iso_code', 'country', 'year', 'co2', 'co2_per_capita',
'population']]

df_co2_UE = df_co2[df_co2.country.isin(countries)]
df_co2_UE = df_co2_UE.rename(columns={'iso_code': 'ISO_CODE', 'country':
'COUNTRY', 'year': 'YEAR', 'co2': 'CO2_EMISSIONS_TONNES',
'co2_per_capita': 'CO2_EMISSIONS_TONNES_PER_CAPITA',
'population': 'POPULATION'})

df_co2_UE = df_co2_UE[df_co2_UE['YEAR'] >= min_year]
df_co2_UE.to_excel('Datos/UE/co2_UE.xlsx', index = False,
sheet_name='co2_UE')

df_co2_UE.head()
```

### Derechos humanos

#### *Índice de derechos humanos*

Dataset con la valoración en el cumplimiento de los derechos humanos dentro de los países de la Unión Europea desde el año 1990 - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - GDP: Producto interior bruto (PIB) - HR\_SCORE: Índice de cumplimiento de los derechos humanos

```
df_hr_scores_VS_GDP = pd.read_csv("Datos/UE/ext_manual/human-rights-score-
vs-gdp-per-capita.csv")

df_hr_scores_VS_GDP = df_hr_scores_VS_GDP[['Entity', 'Code', 'Year', 'GDP
per capita, PPP (constant 2017 international $)', 'Human Rights Score']]
```

```
(Schnakenberg & Fariss, 2014; Fariss, 2019)']]  
df_hr_scores_VS_GDP_UE =  
df_hr_scores_VS_GDP[df_hr_scores_VS_GDP.Entity.isin(countries)]  
df_hr_scores_VS_GDP_UE = df_hr_scores_VS_GDP_UE.rename(columns={'Code':  
'ISO_CODE', 'Entity': 'COUNTRY', 'Year': 'YEAR', 'GDP per capita, PPP  
(constant 2017 international $)': 'GDP', 'Human Rights Score (Schnakenberg  
& Fariss, 2014; Fariss, 2019)': 'HR_SCORE'})  
df_hr_scores_VS_GDP_UE =  
df_hr_scores_VS_GDP_UE[df_hr_scores_VS_GDP_UE['YEAR'] >= min_year]  
df_hr_scores_VS_GDP_UE.to_excel('Datos/UE/hr_scores_vs_gdp_UE.xlsx', index  
= False, sheet_name='hr_scores_vs_gdp_UE')  
  
df_hr_scores_VS_GDP_UE.head()
```

## Economía y bienestar

### *Desigualdad*

Dataset con el indice Gini que sirve de referencia para indicar la desigualdad existente en cada país de la Unión Europea desde el año 1981 - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - GINI\_INDEX: Índice Gini

```
df_gini_index = pd.read_csv("Datos/UE/ext_manual/economic-inequality-gini-  
index.csv")  
  
df_gini_index_UE = df_gini_index[df_gini_index.Entity.isin(countries)]  
df_gini_index_UE = df_gini_index_UE.rename(columns={'Code': 'ISO_CODE',  
'Entity': 'COUNTRY', 'Year': 'YEAR', 'Gini index': 'GINI_INDEX'})  
df_gini_index_UE = df_gini_index_UE[df_gini_index_UE['YEAR'] >= min_year]  
df_gini_index_UE.to_excel('Datos/UE/gini_index_UE.xlsx', index = False,  
sheet_name='gini_index_UE')  
  
df_gini_index_UE.head()
```

### *Gasto en educación*

Dataset con el gasto en educación en cada país de la Unión Europea desde el año 1971 - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - EXPENDITURE\_EDUCATION: Gasto en educación

```

df_expenditure_education = pd.read_csv("Datos/UE/ext_manual/total-
government-expenditure-on-education-gdpslope.csv")
df_expenditure_education.head()

df_expenditure_education_UE =
df_expenditure_education[df_expenditure_education.Entity.isin(countries)]
df_expenditure_education_UE =
df_expenditure_education_UE.rename(columns={'Code': 'ISO_CODE', 'Entity':
'COUNTRY', 'Year': 'YEAR', 'Government expenditure on education, total (%
of GDP)': 'EXPENDITURE_EDUCATION'})
df_expenditure_education_UE =
df_expenditure_education_UE[df_expenditure_education_UE['YEAR'] >=
min_year]
df_expenditure_education_UE.to_excel('Datos/UE/expenditure_education_UE.xls
sx', index = False, sheet_name='expenditure_education_UE')

df_expenditure_education_UE.head()

```

### *Gasto en salud*

Dataset con el gasto en salud en cada país de la Unión Europea desde el año 1970 -  
COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país -  
YEAR: Año de la toma de datos - EXPENDITURE\_HEALTH: Gasto en salud

```

df_expenditure_health = pd.read_csv("Datos/UE/ext_manual/public-health-
expenditure-share-GDP-OWID.csv")

df_expenditure_health_UE =
df_expenditure_health[df_expenditure_health.Entity.isin(countries)]
df_expenditure_health_UE =
df_expenditure_health_UE.rename(columns={'Code': 'ISO_CODE', 'Entity':
'COUNTRY', 'Year': 'YEAR', 'Public expenditure on health %GDP (OWID
extrapolated series)': 'EXPENDITURE_HEALTH'})
df_expenditure_health_UE =
df_expenditure_health_UE[df_expenditure_health_UE['YEAR'] >= min_year]
df_expenditure_health_UE.to_excel('Datos/UE/expenditure_health_UE.xlsx',
index = False, sheet_name='expenditure_health_UE')

df_expenditure_health_UE.head()

```

### *Gasto militar*

Dataset con el gasto militar en cada país de la Unión Europea desde el año 1970 -  
 COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país -  
 YEAR: Año de la toma de datos - EXPENDITURE\_MILITARY: Gasto militar

```
df_expenditure_military = pd.read_csv("Datos/UE/ext_manual/military-
expenditure-share-gdp-sipri.csv")
df_expenditure_military.head()

df_expenditure_military_UE =
df_expenditure_military[df_expenditure_military.Entity.isin(countries)]
df_expenditure_military_UE =
df_expenditure_military_UE.rename(columns={'Code': 'ISO_CODE', 'Entity':
'COUNTRY', 'Year': 'YEAR', 'Military expenditure (% of GDP)': 'EXPENDITURE_MILITARY'})
df_expenditure_military_UE =
df_expenditure_military_UE[df_expenditure_military_UE['YEAR'] >= min_year]
df_expenditure_military_UE.to_excel('Datos/UE/expenditure_military_UE.xlsx',
', index = False, sheet_name='expenditure_military_UE')

df_expenditure_military_UE.head()
```

### *Percepción de la corrupción*

Dataset con la percepción de los ciudadanos sobre el nivel de corrupción en su país -  
 COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país -  
 YEAR: Año de la toma de datos - CORRUPTION\_PERCEPTION: Percepción del nivel de  
 corrupción

```
df_corruption_perception = pd.read_csv("Datos/UE/ext_manual/human-
development-index-vs-corruption-perception-index.csv")

df_corruption_perception =
df_corruption_perception[['Entity', 'Code', 'Year', 'Human Development Index
(UNDP)', 'Corruption Perception Index - Transparency International
(2018)']]
df_corruption_perception_UE =
df_corruption_perception[df_corruption_perception.Entity.isin(countries)]
```

```

df_corruption_perception_UE =
df_corruption_perception_UE.rename(columns={'Code': 'ISO_CODE', 'Entity':
'COUNTRY', 'Year': 'YEAR', 'Human Development Index (UNDP)': 'HUMAN_DEVELOPMENT_INDEX', 'Corruption Perception Index - Transparency International (2018)': 'CORRUPTION_PERCEPTION'})
df_corruption_perception_UE =
df_corruption_perception_UE[df_corruption_perception_UE[ 'YEAR' ] >=
min_year]
df_corruption_perception_UE.to_excel('Datos/UE/corruption_perception_UE.xlsx', index = False, sheet_name='corruption_perception_UE')

df_corruption_perception_UE.head()

```

### *Percepción de la corrupción por institución*

Dataset con la percepción de los ciudadanos sobre el nivel de corrupción en su país por institución - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - Police: Percepción del nivel de corrupción en la Policía - Legislature: Percepción del nivel de corrupción en la legislatura - Government officials: Percepción del nivel de corrupción en el gobierno - Local government councillors: Percepción del nivel de corrupción en el gobierno local - Prime Minister / President: Percepción del nivel de corrupción por parte del presidente o primer ministro - Tax officials: Percepción del nivel de corrupción en los Funcionarios tributarios - Business executives: Percepción del nivel de corrupción en las empresas - Judges and magistrates: Percepción del nivel de corrupción en jueces y magistrados - Religious leaders: Percepción del nivel de corrupción en la iglesia - Traditional leaders: Percepción del nivel de corrupción en los líderes

```

df_corruption_by_institution =
pd.read_csv("Datos/UE/ext_manual/perception-of-corruption-by-
institution.csv")

df_corruption_by_institution_UE =
df_corruption_by_institution[df_corruption_by_institution.Entity.isin(coun
tries)]
df_corruption_by_institution_UE =
df_corruption_by_institution_UE.rename(columns={'Code': 'ISO_CODE',

```

```
'Entity': 'COUNTRY', 'Year': 'YEAR'})  
df_corruption_by_institution_UE =  
df_corruption_by_institution_UE[df_corruption_by_institution_UE['YEAR'] >=  
min_year]  
df_corruption_by_institution_UE.to_excel('Datos/UE/corruption_by_institution_UE.xlsx', index = False, sheet_name='corruption_by_institution_UE')  
  
df_corruption_by_institution_UE.head()
```

### *Nivel de productividad por horas trabajadas*

Dataset con el nivel de productividad y las horas trabajadas en la Unión Europea - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - HOURS\_WORKED\_ Horas trabajadas - PRODUCTIVITY: Nivel de productividad

```
df_productivity_vs_hours_worked =  
pd.read_csv("Datos/UE/ext_manual/productivity-vs-annual-hours-worked.csv")  
  
df_productivity_vs_hours_worked =  
df_productivity_vs_hours_worked[['Entity', 'Code', 'Year', 'Average annual  
hours worked by persons engaged (avh) (PWT 9.1 (2019))', 'Productivity (PWT  
9.1 (2019))']]  
df_productivity_vs_hours_worked_UE =  
df_productivity_vs_hours_worked[df_productivity_vs_hours_worked.Entity.isin(countries)]  
df_productivity_vs_hours_worked_UE =  
df_productivity_vs_hours_worked_UE.rename(columns={'Code': 'ISO_CODE',  
'Entity': 'COUNTRY', 'Year': 'YEAR', 'Average annual hours worked by  
persons engaged (avh) (PWT 9.1 (2019))': 'HOURS_WORKED', 'Productivity  
(PWT 9.1 (2019))': 'PRODUCTIVITY'})  
df_productivity_vs_hours_worked_UE =  
df_productivity_vs_hours_worked_UE[df_productivity_vs_hours_worked_UE['YEAR'] >= min_year]  
df_productivity_vs_hours_worked_UE.to_excel('Datos/UE/productivity_vs_hours_worked_UE.xlsx', index = False,  
sheet_name='productivity_vs_hours_worked_UE')  
  
df_productivity_vs_hours_worked_UE.head()
```

### *Porcentaje de compensaciones a empleados en el servicio público*

Dataset con el porcentaje del gasto reservado a compensaciones en el servicio público -  
 COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país -  
 YEAR: Año de la toma de datos - COMPENSATION\_SPENDING: Porcentaje del gasto  
 reservado a compensaciones en el servicio público

```
df_employee_compensation_spending =
pd.read_csv("Datos/UE/ext_manual/share-of-employee-compensation-in-public-
spending.csv")

df_employee_compensation_spending_UE =
df_employee_compensation_spending[df_employee_compensation_spending.Entity
.isin(countries)]

df_employee_compensation_spending_UE =
df_employee_compensation_spending_UE.rename(columns={'Code': 'ISO_CODE',
'Entity': 'COUNTRY', 'Year': 'YEAR', 'Compensation of employees (% of
expense)': 'COMPENSATION_SPENDING'})

df_employee_compensation_spending_UE =
df_employee_compensation_spending_UE[df_employee_compensation_spending_UE[
'YEAR'] >= min_year]
df_employee_compensation_spending_UE.to_excel('Datos/UE/compensation_spend
ing_UE.xlsx', index = False, sheet_name='compensation_spending_UE')

df_employee_compensation_spending_UE.head()
```

### *Porcentaje de la población en riesgo de pobreza extrema*

Dataset con el porcentaje de población en riesgo de pobreza extrema - COUNTRY: Países  
 pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma  
 de datos - EXTREME\_POVERTY: Porcentaje de población en riesgo de pobreza extrema

```
df_extreme_poverty = pd.read_csv("Datos/UE/ext_manual/share-of-population-
in-extreme-poverty.csv")

df_extreme_poverty_UE =
df_extreme_poverty[df_extreme_poverty.Entity.isin(countries)]
df_extreme_poverty_UE = df_extreme_poverty_UE.rename(columns={'Code':
'ISO_CODE', 'Entity': 'COUNTRY', 'Year': 'YEAR', '$1.90 per day - share of
```

```
population below poverty line': 'EXTREME_POVERTY'})  
df_extreme_poverty_UE =  
df_extreme_poverty_UE[df_extreme_poverty_UE['YEAR'] >= min_year]  
df_extreme_poverty_UE.to_excel('Datos/UE/extreme_poverty_UE.xlsx', index =  
False, sheet_name='extreme_poverty_UE')  
  
df_extreme_poverty_UE.head()
```

### Tasas

Dataset con los datos sobre tasas de cada país de la Unión Europea - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - TAXES\_INCOME: Impuestos sobre la renta - TAXES\_ON\_GOODS\_SERVICES: Impuestos sobre bienes y servicios

```
df_taxes = pd.read_csv("Datos/UE/ext_manual/taxes-on-income-vs-taxes-on-  
goods-and-services.csv")  
  
df_taxes = df_taxes[['Entity', 'Code', 'Year', 'Total taxes on income,  
profits, and capital gains (% of GDP) (ICTD (2021))', 'Total taxes on goods  
and services (% of GDP) (ICTD (2021))']]  
df_taxes_UE = df_taxes[df_taxes.Entity.isin(countries)]  
df_taxes_UE = df_taxes_UE.rename(columns={'Code': 'ISO_CODE', 'Entity':  
'COUNTRY', 'Year': 'YEAR', 'Total taxes on income, profits, and capital  
gains (% of GDP) (ICTD (2021))': 'TAXES_INCOME', 'Total taxes on goods and  
services (% of GDP) (ICTD (2021))': 'TAXES_ON_GOODS_SERVICES'})  
df_taxes_UE = df_taxes_UE[df_taxes_UE['YEAR'] >= min_year]  
df_taxes_UE.to_excel('Datos/UE/taxes_UE.xlsx', index = False,  
sheet_name='taxes_UE')  
  
df_taxes_UE.head()
```

### Total de gastos respecto al PIB

Dataset con la cantidad total de gastos - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - EXPENSES\_BY\_GDP: Total de gastos respecto al PIB

```
df_total_expenditures = pd.read_csv("Datos/UE/ext_manual/total-gov-  
expenditure-gdp-wdi.csv")
```

```

df_total_expenditures_UE =
df_total_expenditures[df_total_expenditures.Entity.isin(countries)]
df_total_expenditures_UE =
df_total_expenditures_UE.rename(columns={'Code': 'ISO_CODE', 'Entity':
'COUNTRY', 'Year': 'YEAR', 'Expense (% of GDP)': 'EXPENSES_BY_GDP'})
df_total_expenditures_UE =
df_total_expenditures_UE[df_total_expenditures_UE['YEAR'] >= min_year]
df_total_expenditures_UE.to_excel('Datos/UE/total_expenditures_UE.xlsx',
index = False, sheet_name='total_expenditures_UE')

df_total_expenditures_UE.head()

```

### *Total de ingresos fiscales*

Dataset con la cantidad total de ingresos fiscales - COUNTRY: Países pertenecientes a la Unión Europea - ISO\_CODE: Código del país - YEAR: Año de la toma de datos - TAXES\_REVENUE\_BY\_GDP: Ingresos fiscales

```

df_tax_revenues = pd.read_csv("Datos/UE/ext_manual/total-tax-revenues-
gdp.csv")

df_tax_revenues_UE =
df_tax_revenues[df_tax_revenues.Entity.isin(countries)]
df_tax_revenues_UE = df_tax_revenues_UE.rename(columns={'Code':
'ISO_CODE', 'Entity': 'COUNTRY', 'Year': 'YEAR', 'Total tax revenue (% of
GDP) (ICTD (2021))': 'TAXES_REVENUE_BY_GDP'})
df_tax_revenues_UE = df_tax_revenues_UE[df_tax_revenues_UE['YEAR'] >=
min_year]
df_tax_revenues_UE.to_excel('Datos/UE/tax_revenues_UE.xlsx', index =
False, sheet_name='tax_revenues_UE')

df_tax_revenues_UE.head()

```