



Escola Tècnica Superior d'Enginyeria Informàtica

Universitat Politècnica de València

ANÁLISIS SOBRE LA ESPERANZA DE VIDA DE LOS PAÍSES DE LA UNIÓN EUROPEA

Asignatura: PROYECTO I

Grado Ciencia de Datos, 1ºB2

Curso 2023-24

Grupo: Grupo Verde

Autores: Isabelle Archer

Nadal Bardisa Quintero

Fernanda de Paula Gonçalves

Carla Jiménez Argudo

Michele Romero Calero



ÍNDICE:

Introducción.....	1
Misión.....	1
Motivación.....	1
Objetivos.....	2
Metodología.....	2
Estructura.....	3
Estado del arte.....	4
Alcance del proyecto.....	6
Hipótesis.....	6
Requisitos.....	6
Restricciones.....	7
Entregables.....	7
Criterios de éxito.....	8
Alineación con los ODS.....	8
Calendario del proyecto.....	10
Diagrama de Gantt.....	10
Distribución de las tareas.....	11
Materiales y métodos.....	11
Origen de los datos.....	11
Análisis exploratorio y de calidad de los datos.....	12
Herramientas utilizadas.....	13
Resultados.....	13
Validación y comentarios de los resultados.....	13
Contribuciones y limitaciones del trabajo.....	17
Comparación con el estado del arte.....	19
Conclusiones.....	19
Bibliografía.....	21
Anexos.....	26
Tablas.....	26
Gráficos.....	28

ÍNDICE DE TABLAS:

Tabla 1. Restricciones del proyecto.....	7
Tabla 2. Diagrama de Gantt.....	10
Tabla 3. Lista de tareas del diagrama de Gantt.....	10
Tabla 4. Paquetes de trabajo.....	11
Tabla 5: Descripción de cada variable del dataset.....	26
Tabla 6: Resumen de los coeficientes de correlación.....	27

ÍNDICE DE GRÁFICOS:

Gráfico 1: Distribuciones de las variables del dataset.....	28
Gráfico 2: Regresión logarítmica de la variable adult mortality contra life expectancy.....	28
Gráfico 3: Regresión lineal de la variable income composition of resources contra life expectancy....	29
Gráfico 4: Regresión logarítmica de la variable total expenditure contra life expectancy.....	29
Gráfico 5: Esperanza de vida media por cada región de Europa.....	30

Introducción

La población del mundo está en continuo crecimiento. En 2023 hubo 134.279.612 nacimientos en todo el mundo (UNICEF, 2022). Con este dato, hay que preguntar, ¿Cuál sería la esperanza y la calidad de vida que van a tener estos bebés?

Según The World Factbook, se define la esperanza de vida como una comparación del número medio de años que van a vivir un grupo de personas que nacieron en el mismo año si la mortalidad en cada edad sigue constante y es también una medida de la calidad de vida (CIA.gov, 2024). Se calcula al nacimiento y a diferentes edades considerando que los patrones de mortalidad vigentes en los diferentes momentos del cálculo se mantienen a lo largo del horizonte de años de vida. (INE - Instituto Nacional de Estadística, 2024)

En 1900 la esperanza de vida fue de 32 años, y en 2021 se ha duplicado a 71 años (Dattani et al., 2023). Este aumento significativo se debe a la mejora de muchos factores, como por ejemplo, la tecnología, la medicina y la educación.

Misión

La misión de este proyecto es utilizar, analizar y comprender el nivel de esperanza de vida de los diferentes países de la Unión Europea entre los años 2000 y 2015. Esto se realiza para poder predecir y evaluar dentro de los próximos años la compleja interacción entre los diversos factores socioeconómicos y geográficos que impactan en la esperanza de vida de la población. A través de la recopilación, limpieza y análisis de datos, se busca identificar patrones, correlaciones y tendencias que contribuyan a una comprensión profunda de las variables que afectan a la esperanza de vida.

Motivación

Tras observar las variaciones de la esperanza de vida entre distintos países surgen diferentes preguntas. ¿Por qué la esperanza de vida varía entre países si las costumbres entre ellos son más o menos similares? ¿Existen factores que influyen directamente en la esperanza de vida como la alimentación, la educación o los estilos de vida? ¿Es posible predecir la esperanza de vida de un país determinado? Una de las propuestas de este proyecto es investigar e indagar sobre estos posibles factores que puedan repercutir sobre la esperanza de vida, para conseguir una predicción de la misma en cada país.

Asimismo, otra cuestión que ocupa la mente de la población en general es vivir el mayor tiempo posible y gozar de la mejor calidad de vida. Por ello, uno de los pilares fundamentales de la motivación de este proyecto es descubrir o adoptar decisiones más saludables y beneficiosas a nivel político, sanitario y social. De esta manera, se consigue mejorar nuestro presente y a su vez nuestro futuro. En definitiva, la motivación de este proyecto se basa en encontrar medidas que mejoren el bienestar común a largo plazo, mediante la comprensión de los diferentes factores que influyen en la esperanza de vida.

Objetivos

El objetivo general es analizar los factores geográficos, socioeconómicos, ambientales y sanitarios que influyen en la esperanza de vida de un país, para así poder proporcionar decisiones estratégicas, medioambientales y saludables.

OBJETIVOS ESPECÍFICOS:

1. Efectuar un análisis del estado del arte. Identificar las últimas investigaciones y tendencias relevantes relacionadas con la esperanza de vida y sus factores relacionados, situando el proyecto en un contexto actual y destacar su contribución diferenciada.
2. Adquirir conocimientos y habilidades prácticas para manipular y comprender conjuntos de datos mediante el uso de herramientas como librerías de Python (Pandas, NumPy, Matplotlib, Sklearn, BeautifulSoup).
3. Realizar un estudio comparativo de la esperanza de vida entre países de diferentes regiones identificando posibles patrones o diferencias entre los años 2000-2015.
4. Evaluar el posible impacto de diferentes variables en la esperanza de vida de un país mediante análisis estadísticos y gráficos descriptivos, usando librerías como XGBoost, Sklearn y Scipy.
5. Diseño y creación de una página web para presentar los datos recopilados de la esperanza de vida y sus factores relacionados, con el fin de ofrecer una visualización clara y accesible del análisis previamente realizado sobre factores socioeconómicos, geográficos, ambientales y de salud de los años 2000-2015.

Metodología

Se han seguido los siguientes pasos para la realización del trabajo:

1. Obtención de datos: se ha iniciado una búsqueda de datasets sobre la esperanza de vida y algunos factores relacionados. Los datasets incompletos o con variables irrelevantes han sido descartados, y se han seleccionado las más relevantes para el estudio. Además, se han añadido otras variables importantes, como el nivel de emisión de dióxido de carbono, el colesterol LDL en sangre y la temperatura media de cada país.

2. Estado del arte: se ha buscado información complementaria para asegurar la completitud del proyecto. Se recopilaron artículos y estudios relevantes para enriquecer el análisis.

3. Análisis exploratorio y calidad de los datos: se ha realizado un análisis exploratorio para entender la estructura de los datos y su calidad. También se ha estudiado la distribución de cada variable para evaluar la normalidad y luego, se ha aplicado unas transformaciones a las variables que presentan asimetría.

4. Análisis de los factores que influyen en la esperanza de vida: se ha analizado la relación entre la esperanza de vida y otras variables, y después país por país, en el rango de los 16 años investigados. Se ha estudiado el impacto de la posición geográfica en la esperanza de vida, así como el análisis específico de los países del sur de la Unión Europea.

5. Diseño e implementación de la web: se ha diseñado y desarrollado una página web para presentar los datos recopilados y los análisis realizados. La web ofrecerá una visualización clara y accesible de los resultados y una calculadora interactiva que permita a los usuarios introducir datos específicos y obtener una estimación de su propia esperanza de vida.

Estructura

El informe comienza con una introducción donde se define la esperanza de vida y sitúa el trabajo en su contexto, explicando los motivos y objetivos del estudio, así como la metodología utilizada. Se proporciona una visión clara de las metas a alcanzar y se describe el enfoque adoptado y los pasos seguidos en la elaboración del estudio. En la sección del estado del arte, se ofrece un panorama contextual del tema del proyecto.

Además, se detallan el alcance del proyecto, las restricciones, los requisitos, las hipótesis, y los criterios para evaluar su éxito, destacando la importancia de este apartado para medir la calidad y eficiencia del trabajo final.

El calendario del proyecto delimita temporalmente las actividades y refleja las dificultades y facilidades encontradas. En "Materiales y métodos", se describen las fuentes de

datos y herramientas utilizadas, y el análisis exploratorio de los datos, incluyendo su calidad. Los resultados subrayan la importancia de la esperanza de vida como indicador de salud pública y las desigualdades en esta debido a factores sociales, económicos, políticos y de género. Las conclusiones presentan relaciones entre variables y predicciones sobre la influencia de diversos factores en la esperanza de vida. Finalmente, los anexos contienen gráficos y tablas detalladas, demasiado grandes para el cuerpo principal del informe.

Estado del arte

Durante los últimos 170 años, la esperanza de vida ha aumentado considerablemente, sin embargo siguen existiendo grandes discrepancias entre los países desarrollados y los países en desarrollo. Se cree que esta desigualdad en la esperanza de vida tiene sus raíces en los diferentes contextos socioeconómicos de los diferentes grupos sociales, como se puede ver según el estudio realizado por Richardson et al. (2013).

Uno de los principales objetivos de un gobierno es alargar la esperanza de vida de su población reduciendo su tasa de mortalidad al nivel mínimo posible, especialmente de acuerdo con los Objetivos de Desarrollo Sostenible de las Naciones Unidas (Naciones Unidas, 2024). El desarrollo económico determina mejoras en las condiciones sociales y un aumento en la esperanza de vida.

El estudio de Bilas et al. (2014) es de gran importancia para el desarrollo de este proyecto. Ellos estudiaron los factores determinantes de la esperanza de vida de 26 países de la Unión Europea de los años 2001-2011. Seleccionaron las variables: tasa de crecimiento del PIB, tasa de crecimiento de la población, nivel de educación, inscripción en educación, PIB per cápita y esperanza de vida. No pudieron encontrar los datos para los Países Bajos o para Polonia, así que no estudiaron todos los países de la Unión Europea (UE). Les interesaba la relación entre las variables y el sistema de salud en el país.

Aplicaron un modelo de regresión lineal y los resultados mostraron una significancia de 1% de las variables de PIB per cápita y de nivel de educación (Bilas et al., 2014). La relación entre la primera variable y la esperanza de vida está apoyada con estudios previos, que sigue una relación directa y positiva. Sin embargo, ellos encontraron que la relación entre el nivel de educación es distinta a la que surgieron los estudios previos. Este estudio sigue una relación negativa inversa y esto se debe al análisis a través del tiempo. En el contexto afirma que a mayor nivel de educación, más cantidad de estrés en el trabajo porque suelen trabajar más horas, y su ritmo de vida es mayor. Los residentes de un país con altos

estándares de vida viven más tiempo en promedio y tienen una mortalidad menor (Bilas et al., 2014). Por otra parte, se identificaron algunas limitaciones en el estudio, como la omisión de variables relacionadas con la dieta o la nutrición debido a la falta de datos comparables a nivel de la Unión Europea. Además, se observó la ausencia de dos países importantes dentro de la Unión Europea en el análisis.

Otro estudio que se puede destacar es un estudio portugués realizado en 2020. Esta exploración está enfocada en la supervivencia de una edad de 65 años. Poças et al. (2020) investigaron 20 países de la Unión Europea de 1990 hasta 2016. Las variables incluidas en este estudio están más enfocadas en la salud y el medio ambiente que el estudio de Bilas et al. (2014) incluyendo por ejemplo, fumadores, consumo de alcohol, y emisión de dióxido de carbono. Ellos tratan de crear una ecuación de salud estimada según las variables de nivel socioeconómico, demográfico, estilo de vida, sistema de salud y factores medioambientales. Aplicaron varios métodos de análisis. Luego, hicieron una prueba de Hausman y el modelo más apropiado es el modelo de efectos al azar (Poças et al., 2020). Los resultados encontrados son: un estilo de vida peligroso (fumar y beber alcohol) y la contaminación de dióxido de carbono, bajan la esperanza de vida en una edad avanzada (Poças et al., 2020). Es un estudio muy largo e interesante para ver cómo analizar los datos de forma adecuada. Sin embargo, sólo estudiaron 20 países y no incluyeron otras variables importantes como el índice del desarrollo humano, una variable importante, por su gran impacto socio-político en la sociedad.

Para estudiar la relación geográfica, se ha prestado especial atención al estudio de Richardson et al. (2013) sobre la diferencia de las regiones en la Unión Europea de 1991 hasta 2008. Para elegir las regiones, usaron la Nomenclatura Común de Unidades Territoriales Estadísticas (NUTS) para comparar entre países distintos a un nivel de 2. En adición de analizar la esperanza de vida geográficamente, también hicieron un análisis socio-espacial según los ingresos medios de la región. Finalmente, partieron los sexos para hacer un análisis más adecuado según el sexo. Los países del este de Europa tuvieron una tasa de crecimiento de la esperanza de vida más baja que la de Francia, España, Italia, y Suecia (Richardson et al., 2013). También, las regiones mantuvieron sus posiciones de esperanza de vida a lo largo del tiempo. Los resultados mostraron que la desigualdad geográfica se mantuvo para todas las regiones y los sexos, menos los hombres que viven en el este de Europa. La desigualdad socio-espacial tuvo un efecto sobre la desigualdad geográfica y la influía también. Una limitación del estudio fue la varianza en tamaño según la NUTS, aunque es un

método estandarizado. Este estudio muestra que el análisis geográfico que se quiere realizar en este proyecto es más difícil que el análisis general, pero no es imposible.

Si bien la esperanza de vida ha aumentado como resultado del progreso socioeconómico y médico, muchos factores de estilo de vida, como el abuso del alcohol, la obesidad y el tabaco, influyen de forma significativa en la esperanza de vida de las personas (Janssen et al., 2021).

Alcance del proyecto

Hipótesis

Se cree que pueden existir correlaciones entre la esperanza de vida y diversos factores socioeconómicos, geográficos y médicos en los distintos países de la Unión Europea. Es posible que estos no se hayan tenido en cuenta por la suficiente precisión obtenida en el estimador de esperanza de vida actual y, por tanto, la falta de necesidad de evaluar el efecto de los varios factores, quizá en menor medida, en el cálculo de la esperanza de vida. De este modo, se pretende aportar un modelo que, teniendo en cuenta estos nuevos factores, así como los anteriores, sea capaz de predecir de manera más precisa la cantidad de años que una persona vivirá según los estándares que caracterizan la zona donde vive y algunos datos añadidos de la propia persona (condiciones personales) como por ejemplo, si consume tabaco, alcohol, etc.

Requisitos

Para la realización del proyecto serán necesarias diversas cuestiones:

- El equipo adecuado para la realización de análisis descriptivos univariantes y multivariantes, que pueda extraer los datos de las variables obtenidas de un archivo de datos y que permita de manera visual reconocer patrones y distinguir comportamientos entre variables. Para lograrlo, se usan los distintos paquetes públicos preparados para Python3 con los que se puede analizar y representar los datos, como Pandas o matplotlib.
- Una o varias bases de datos con la suficiente cantidad limpia de información para permitir relacionar al menos una variable de dicho conjunto de datos con la esperanza de vida bruta obtenida de la publicación de la Unión Europea de esperanzas de vida de los países de la Unión en los diferentes años estudiados en este proyecto.

- Acceso a un servidor capaz de almacenar y mostrar los archivos de la página web, así como de ejecutar el código escrito para calcular la esperanza de vida de un usuario en base a sus datos.
- Acceso a equipos con las últimas versiones disponibles de programas de edición de vídeo de alta calidad, programas de edición y ejecución de código para creación de aplicaciones y programas de composición musical.

Restricciones

RESTRICCIÓN	DESCRIPCIÓN
Mantenimiento de los servidores.	La página web solo puede ser utilizada mientras los servidores donde se encuentra alojada la puedan cargar. Si son atacados o se encuentran en mantenimiento, la web quedará fuera del alcance de los usuarios.
Cambio de efecto de las variables.	El análisis, los datos de nuestra base y demás son estáticos, pero la realidad es continua y con el tiempo puede cambiar la afectación de una variable a la esperanza de vida, requiriendo la calculadora de constantes actualizaciones y revisiones.
Utilidad reducida.	La utilidad de la calculadora para la estimación de la esperanza de vida del usuario solo es válida para aquellos que vivan en alguno de los países de nuestro conjunto de datos.

Tabla 1. Restricciones del proyecto

Entregables

1. **La memoria del proyecto:** se ha detallado todo el proceso del trabajo e incluye la motivación, el estado del arte, el alcance del trabajo, el análisis de los datos, los resultados, las conclusiones y las limitaciones del proyecto. Se tiene valor para explicar lo que se ha hecho y eliminar dudas o preguntas sobre el trabajo.
2. **El dataset:** incluye varias variables que no han sido investigadas antes, juntas con variables que se sabe tienen influencia sobre la esperanza de vida. Abarca todos los países de la Unión Europea, menos Chipre, durante un largo período. Será de gran utilidad para cualquier investigador interesado en el estudio de la esperanza de vida, permitiéndole enfocar sus esfuerzos en un análisis más profundo en lugar de en la

búsqueda de datos. El dataset estará disponible en un repositorio de datos abiertos y accesibles.

3. **La página web y el modelo predictivo/calculadora:** se ha desarrollado una página web en paralelo al análisis de los datos, que sirve para mostrar visualmente los resultados del trabajo. También, es accesible a cualquier persona. El modelo predictivo también será útil para ahorrar tiempo en la búsqueda de datos actuales mediante un web scraper.

Criterios de éxito

Para considerar el proyecto un éxito, se deben cumplir los siguientes criterios medibles:

1. **Cumplimiento de objetivos específicos:** Se logrará el éxito si se cumplen todos los objetivos especificados en el proyecto dentro del tiempo y con los recursos asignados. Esto se evaluará mediante la revisión de hitos clave alcanzados.
2. **Impacto del modelo predictivo:** El modelo predictivo desarrollado debe demostrar una precisión de al menos 85% y un cuadrado medio del error de menos de 3 años en las predicciones sobre la esperanza de vida basadas en las variables introducidas. Esto se podría evaluar mediante pruebas de validación cruzada y comparación con datos históricos.
3. **Ampliación del conocimiento existente:** Otro criterio de éxito será la capacidad del proyecto para ampliar el conocimiento existente sobre la relación entre variables y la esperanza de vida. Encontrar al menos una relación nueva entre la esperanza de vida y una variable nueva se considera un éxito.

Alineación con los ODS

El objetivo de tres de los Objetivos de Desarrollo Sostenible de las Naciones Unidas es “garantizar una vida sana y promover el bienestar de todos a todas las edades” (Naciones Unidas, 2024). Este objetivo es muy importante para el proyecto pues trata de la salud de toda la población. Este trabajo puede destacar los factores influyentes que puedan afectar a una vida sana y prevenir a los gobiernos para que reflexionen donde es mejor invertir capital. Las tareas del objetivo tres son las vacunas y los servicios esenciales de salud.

Garantizar una educación inclusiva y equitativa de calidad y promover oportunidades de aprendizaje permanente para todos es el objetivo cuatro (Naciones Unidas,

2024). Los estudios previos, sobre todo Bilas et al. (2014), han mostrado que hay bastante relación entre nivel de educación y esperanza de vida. Si a algún país le falta un nivel de educación adecuado para el Objetivo Sostenible también le faltaría un nivel de educación adecuado para tener una esperanza de vida buena. Además, la desigualdad en el nivel de educación se puede ver según las diferencias en los factores socioeconómicos. Si se amplían los factores económicos, también se ampliará el nivel de educación y con ambos aumentos, la esperanza de vida también lo hará como consecuencia.

El objetivo ocho trata el crecimiento económico sostenido y empleo pleno. Es verdad que con un aumento de la situación económica, la esperanza y la calidad de vida serían mejores, hasta cierto punto. Bilas et al. (2014) han demostrado que si incrementa demasiado, las condiciones de vida serían más trabajo, más estrés, menos ejercicio, y menos tiempo libre.

Una de las variables estudiadas en este trabajo es la emisión de dióxido de carbono, que es un factor medioambiental. El objetivo 13 de los ODS es combatir el cambio climático y sus efectos (Naciones Unidas, 2024). Poças et al., encontraron en su estudio, que la relación entre la emisión de CO₂ y la esperanza de vida es una relación inversa. Con un aumento de las emisiones, la esperanza y la calidad de vida serían peor. Si se cambia a transporte más sostenible y a energía limpia, este factor mejoraría.

Calendario del proyecto

Diagrama de Gantt

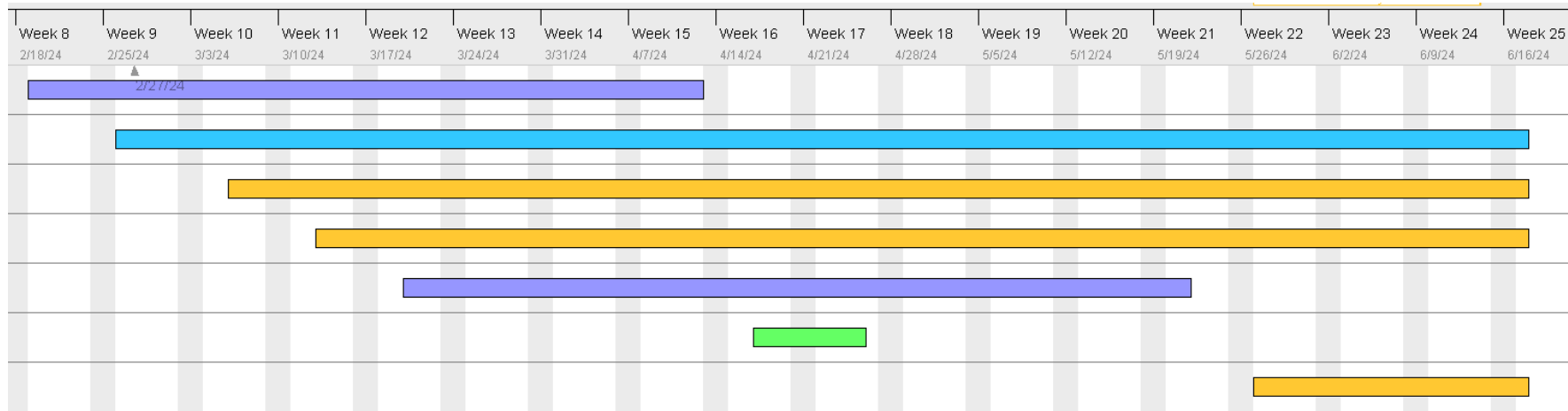


Tabla 2. Diagrama de Gantt.

Name	Begin ...▲	End date
Búsqueda y limpieza del dataset	2/19/24	4/12/24
Desarrollo de la página web	2/26/24	6/17/24
Desarrollo de la memoria	3/6/24	6/17/24
Desarrollo de la presentación	3/13/24	6/17/24
Análisis de los datos	3/20/24	5/21/24
División de trabajo y creación de los paquetes de trabajo	4/17/24	4/25/24
Sacar resultados y conclusiones	5/27/24	6/17/24

Tabla 3. Lista de tareas del diagrama de Gantt.

Distribución de las tareas

Responsable del Paquete	Paquete de Trabajo
Nadal	Desarrollo de página web <ul style="list-style-type: none">- Creación de una página web para publicar los resultados obtenidos y la calculadora.
Carla	Coordinación de trabajos <ul style="list-style-type: none">- Facilitar comunicación y organizar tareas entre los responsables de otros trabajos.
Michele	Creación del modelo predictivo y calculadora <ul style="list-style-type: none">- Desarrollar un modelo para predecir la esperanza de vida con datos del usuario.
Isabelle	Análisis de datos <ul style="list-style-type: none">- Creación y análisis del dataset para obtener resultados y conclusiones.
Fernanda	Difusión de resultados <ul style="list-style-type: none">- Publicar y difundir los resultados obtenidos del análisis de la esperanza de vida.

Tabla 4. Paquetes de trabajo.

Materiales y métodos

Origen de los datos

La primera fuente del dataset proviene de Kaggle. Se llama “Life Expectancy (WHO)” y fue subido al sitio web del usuario KumarRajarshi (2018). Incluye 193 países con 22 variables. Se han cogido los 27 países de la Unión Europea, a excepción de Chipre que no está incluido en el dataset original. Se eliminan además algunas variables que no aportan mucha información como hepatitis B, measles (sarampión) y polio.

Adicionalmente, aparecen algunos datos faltantes o erróneos. Para rellenar la variable de consumo de alcohol, del data del WHO (WHO, 2019). De The World Bank (The World Bank, 2024), se han seleccionado las variables PIB, gastos totales y porcentaje de gastos. Para los datos faltantes de Chequia, se ha usado Statista (McEvoy, 2023) para el índice de desarrollo humano; UNData (UNData, 2015) para los años de la escolarización y The World Bank (The World Bank, 2022) para la población. Para los años de la escolarización de Alemania, utilizamos y GlobalDataLab (Lab, G. D., 2021). Para la

población de Eslovaquia, se han obtenido los datos de The World Bank (The World Bank, 2022).

Con esta revisión, faltan variables que se consideran significativas en relación a la esperanza de vida, así que se han añadido las variables: temperatura media, emisión de dióxido de carbono, colesterol LDL en sangre y porcentaje de adultos que fuman. Se han obtenido las temperaturas de Trading Economics (Trading Economics, 2024), el dióxido de carbono de The World Bank (The World Bank, 2020), el colesterol LDL de WHO (WHO, 2021) y el porcentaje de fumar de The World Bank (The World Bank, 2020).

Análisis exploratorio y de calidad de los datos

La calidad de los datos es buena. Es un dataset limpio y conciso con 20 variables de todos los países de la Unión Europea, a excepción de Chipre, de los años 2000 a 2015, ambos incluidos.

Hay dos variables medioambientales: *average temperature* y *CO2 emissions*; seis variables socio-económicas: *income composition of resources*, *population*, *GDP*, *percentage and total expenditure*, *schooling*; y nueve variables relacionadas con la salud: *smoking*, *alcohol*, *cholesterol*, *thinness 1-19 years and 5-9 years*, *under-five deaths*, *BMI*, y *adult mortality*. Una descripción de cada variable se puede ver en la *tabla 5* del anexo.

Todas las variables, con excepción del país y el estado del desarrollo, son variables cuantitativas. Se ha recodificado la variable *estado del desarrollo* como una variable cuantitativa, en la que con un 0 se representa un país desarrollado y con un 1 se representa un país en estado de desarrollo. La mayoría de las variables presentan asimetría por la izquierda o por la derecha. Las variables normales son: *life expectancy*, *alcohol*, *schooling*, *average temperature* y *cholesterol*, como se puede ver en el *gráfico 1* del anexo.

En principio, la variable *smoking* solo se dispone de un dato cada 5 años para cada país. Sin embargo, se considera que es una variable importante a estudiar y que se podría analizar usando una interpolación de los datos para obtener un valor para cada año. Para ello, se utilizó el método de interpolación de la librería *Scipy* de Python.

Se ha realizado un análisis univariable de la relación entre cada variable y la variable esperanza de vida. Se hizo una regresión lineal, también de la librería *Scipy*, para estudiar el efecto de las variables. Habían algunas variables que presentaban heterocedasticidad y para combatir los malos efectos de ésta, se hizo una transformación logarítmica de las mismas.

Asimismo, se hizo un árbol de decisión regresión usando XGBoost. Presenta un valor de R-cuadrado de 0.852, un valor bastante alto. El modelo puede predecir valores con un cuadrado medio del error (MSE) de 2.54.

Se han analizado las variables país por país. Se ha aplicado el modelo del árbol de decisión regresión a cada país para analizar la relación de cada variable con la esperanza de vida en este país.

Por último, se ha hecho un estudio geográfico agrupando los países en su región de la Unión Europea y se ha estudiado la esperanza de vida en estas regiones. La región norte incluye los países: Suecia, Finlandia y Dinamarca; la región sur incluye: España, Italia, Grecia, Portugal y Malta; la región este tiene los países: Bulgaria, Croacia, Chequia, Letonia, Lituania, Rumania, y Estonia; la región oeste incluye: Francia, Bélgica, Países Bajos, e Irlanda; y últimamente, la región central tiene los países: Austria, Luxemburgo, Alemania, Hungría, Polonia, Eslovaquia y Eslovenia.

Herramientas utilizadas

Se ha utilizado Jupyter Notebook con la librería Pandas para visualizar y analizar el dataset en formato CSV. También se ha usado la librería Scipy para hacer gráficos visuales y complejos dentro de Jupyter Notebook. Tanto la página web como la aplicación, se ha escrito en lenguaje HTML y CSS.

Para la compilación de todos los datos del dataset se ha usado Microsoft Excel. De este modo se consigue que todos los datos estén localizados en el mismo lugar, facilitando el análisis de las variables.

En cuanto a los análisis univariantes y multivariantes, estos se realizan gracias a los diversos paquetes de Python3 como Sklearn y XGBoost, para crear los modelos necesarios.

Resultados

Validación y comentarios de los resultados

Según las regresiones, y como se puede observar en la *tabla 6: resumen de los coeficientes de correlación* en el anexo, hay algunas variables con un valor alto de correlación. La variable *adult mortality* tiene el valor más alto con -0.83, una relación inversa, como se puede ver en el *gráfico 2: Regresión logarítmica de la variable adult mortality contra life expectancy*. Hay dos variables con un valor entre 0.7 y 0.8: *income*

composition of resources y total expenditure, mostrados en los *gráficos 3 y 4*. Las dos variables que miden la *delgadez* de la población hasta los 19 años también tienen altos coeficientes, de -0.66 y de -0.67.

Según el modelo del árbol de decisión regresión, la variable más influyente también es la variable *adult mortality* con un valor de 0.39, seguido por la variable *percentage expenditure* con un valor de 0.10.

A continuación se presenta el análisis país por país de las variables con las relaciones más significativas respecto a la esperanza de vida usando el modelo del árbol de decisión regresión:

Austria:

El modelo destaca en relación a la esperanza de vida, las variables: el índice de mortalidad en adultos (*adult mortality*) con un una relación de 0.23 sobre 1; para el producto interior bruto (*GDP*) un 0.68 y para *income composition* un 0.08. Además el error cuadrático es de tan sólo 1.122, el cual indica un bajo error respecto a las predicciones y el valor de R^2 es de 0.96128, lo que significa que el modelo explica el 96.128% de la variabilidad en la esperanza de vida basada en las variables utilizadas.

Bélgica:

Destaca en relación a la esperanza de vida el índice de masa corporal (*BMI*) con 0.07 sobre 1, también el índice de delgadez hasta los 19 años (*thinness 1-19*) con 0.01; la *temperatura media* con 0.21 y por último el porcentaje de personas fumadoras (*smoking*) con un valor de 0.71 sobre 1. El error cuadrático es de 3.502 y el valor de R^2 es de 0.8919.

Bulgaria:

Se distinguen en relación a la esperanza de vida, la tasa de mortalidad en adultos (*adult mortality*) con un 0.06 sobre 1; el *BMI* con un 0.01; índices de dióxido de carbono (*CO2*) con un 0.02; el colesterol (*cholesterol*) con un valor de 0.05; y el *income composition* con un 0.85. Por lo que respecta al error cuadrático, se observa un valor de 0.204, y el valor de R^2 es de 0.9384.

Croacia:

Cabe resaltar la *tasa de mortalidad adulta* con un 0.07; el *índice de delgadez hasta los 19 años* con un 0.85; y los años de escolarización obligatoria (*schooling*) con un 0.06. El error cuadrático es de 0.07 y el valor de R^2 es de 0.982.

Chequia:

Sobresalen el *producto interior bruto* con un valor de 0.02 sobre 1; la población (*population*) con 0.01; el *income composition* con un 0.86; los años de escolarización obligatoria con un 0.03; y la temperatura media (*average temperature*) con un 0.08. El error cuadrático es de 0.134 y el R^2 es de 0.974.

Dinamarca:

Destacan el índice de delgadez 5-9 años (*thinness 5-9*) con un 0.11 sobre 1; la *temperatura media* del país con un 0.02; y los niveles de *dióxido de carbono* con un 0.86. El error cuadrático es de 0.3 y el valor de R^2 es de 0.984615.

Estonia:

Se distinguen la *tasa de mortalidad en adultos* con un 0.35; el porcentaje de personas que consumen alcohol (*alcohol*) en relación a la esperanza de vida con un 0.05; el *índice de masa corporal* con un 0.04; el *porcentaje de personas fumadoras* con un 0.01; y la *población* con un 0.54. El error cuadrático es de 2.084 y el valor de R^2 es de 0.872459.

Finlandia:

Cabe resaltar el gasto total en salud (*total expenditure*) con un 0.75 y el *income composition* con un 0.17. El error cuadrático es de 3.13 y el valor de R^2 es de 0.907285.

Francia:

Destacan el *porcentaje de personas que consumen alcohol* con un valor de 0.25; y el *producto interior bruto* con un 0.75. El error cuadrático es de 12.5859 y el valor de R^2 es de 0.581705.

Alemania:

Sobresalen el *gasto total en salud* con un 0.12; el *índice de masa corporal* con un 0.01; y *fumadores* con un 0.86. El error cuadrático es de 0.254 y el valor de R^2 es de 0.99258.

Grecia:

Cabe destacar la *tasa de mortalidad* con un 0.16; el *índice de masa corporal* con un 0.01; el *porcentaje de gasto en salud* con un 0.07; el *income composition* con un 0.01; y las *emisiones de CO₂* con un 0.74. El error cuadrático es de 0.374 y el valor de R^2 es de 0.9862.

Hungría:

Se distinguen la *tasa de mortalidad en adultos* con un 0.06, el *índice de masa corporal* con un 0.05, el *índice de delgadez entre 5-9 años* con un valor de 0.01 y los *niveles de CO₂* con un 0.86. El error cuadrático es de 0.19 y el valor de R^2 es de 0.9623.

Italia:

Destacan el *gasto total en salud* con un 0.18; el *índice de masa corporal* con un 0.01; y los *niveles de CO₂* con un 0.8. El error cuadrático es de 0.47 y el valor de R^2 es de 0.978.

Irlanda:

Sobresalen la *tasa de mortalidad en adultos* con un 0.1 sobre 1; el porcentaje de gasto en salud (*percentage expenditure*) con un 0.04; el *income composition* con un 0.01; y los *niveles de CO₂* con un 0.79. El error cuadrático es de 7.09399 y el valor de R^2 es de 0.703725.

Letonia:

La *tasa de mortalidad en adultos* con un 0.25; el *índice de delgadez entre 1-19 años*, con un 0.38; el *número de personas que consumen alcohol* con un 0.01; el *índice de masa corporal* con un valor de 0.1; y el *porcentaje de gasto en salud* con un 0.01, son las variables con relación más significativa respecto a la esperanza de vida. El error cuadrático es de 3.292 y el valor de R^2 es de 0.7947.

Luxemburgo:

Resaltan la *tasa de mortalidad en adultos* con un 0.17; el *gasto total en salud* con un 0.8; el *porcentaje de gasto en salud* con un 0.01; y el *producto interior bruto* con un 0.02. El error cuadrático es de 14.718 y el valor de R^2 es de 0.41856.

Malta:

Se distinguen la *tasa de mortalidad en adultos* con un 0.05; el *índice de masa corporal* con un 0.01; el *income composition* 0.67; y los *años de escolarización obligatoria* con un 0.27. El error cuadrático es de 0.27 y el valor de R^2 es de 0.984.

Países Bajos:

Destacan la *tasa de mortalidad en adultos* con un 0.28; el *gasto total en salud* con un 0.01; el *porcentaje de gasto en salud* con un 0.7; y los *niveles de CO₂* con un 0.01. El error cuadrático es de 2.708 y el valor de R^2 es de 0.8824595.

Polonia:

La *tasa de mortalidad en adultos* con un 0.01; el *gasto total en salud* con un 0.06; el *porcentaje de gasto en salud* con un 0.01; el *income composition* con un 0.86; y los *niveles de CO₂* con un 0.06, son las variables que guardan mayor relación con la esperanza de vida. El error cuadrático es de 0.102 y el valor de R^2 es de 0.97519.

Portugal:

Resaltan la *tasa de mortalidad en adultos* con un 0.01; el *gasto total en salud* con un 0.04; la *temperatura media* con un 0.01; el *income composition* con un 0.67; y el *porcentaje*

de personas fumadoras con un 0.27. El error cuadrático es de 1.958 y el valor de R^2 es de 0.94632.

Rumanía:

Sobresalen la *población* con un 0.03; el *índice de delgadez entre 5-9 años*, con un 0.24; la *temperatura media* con un 0.01; el *income composition* con un 0.68; los *años de escolarización obligatoria* con un 0.01. El error cuadrático es de 0.218 y el valor de R^2 es de 0.98176.

Eslovaquia:

Se distinguen el *producto interior bruto* con un 0.06 sobre 1; la *población* con un 0.01; la *temperatura media* con un 0.08; y los *niveles de CO₂* con un 0.83. El error cuadrático es de 0.112 y el valor de R^2 es de 0.970276.

Eslovenia:

Destacan el *gasto total en salud* con un 0.1; la cantidad de *población* con un 0.01; y el *income composition* con un 0.8. El error cuadrático es de 8.14 y el valor de R^2 es de 0.8133.

España:

El *índice de masa corporal* con un 0.04; el *porcentaje del gasto en salud* con un 0.48; y el *income composition* con un 0.48, son las variables que presentan mayor relación con la esperanza de vida. El error cuadrático es de 0.38 y el valor de R^2 es de 0.98375.

Suecia:

Cabe resaltar el *número de personas que beben alcohol* con un 0.02; el *producto interior bruto* con un 0.64; el *income composition* con un 0.34; y el *número de personas fumadoras* con un 0.01. El error cuadrático es de 13.53199 y el valor de R^2 es de 0.3709.

No se ha podido hacer un modelo apto para el país Lituania.

En relación al análisis geográfico, que se puede ver en el *gráfico 5*, se ha analizado la esperanza de vida media de cada región de Europa.

Contribuciones y limitaciones del trabajo

Durante la elaboración de este proyecto se ha conseguido una serie de contribuciones al análisis de la esperanza de vida. Así, se ofrece una perspectiva detallada de los factores que influyen directamente en la Unión Europea. Al utilizarse datos actuales y métodos adecuados se han podido identificar tendencias para la toma de decisiones en políticas sociales, de salud y políticas. Además, una de las contribuciones más importantes ha sido brindar una

interpretación sencilla y detallada para mejorar la comprensión de las dinámicas estudiadas en cada región.

Se ha compilado un dataset muy grande con varias variables importantes relacionadas con la esperanza de vida, con todos los países de la Unión Europea, a excepción de Chipre. Resultaría útil para cualquier persona para estudiar y añadir al análisis ya hecho. También, se ha llevado a cabo un importante estudio de las distintas relaciones de factores multivariantes sobre la esperanza de vida.

Además, se ha creado una calculadora para que el usuario pueda introducir sus datos y la calculadora busque los datos más actuales del país, el cual introduce el propio usuario. Esto ahorra tiempo y recursos para el usuario e investigadores.

Las principales limitaciones para la realización de este proyecto son la falta de conocimiento sobre el tema, la escasez de tiempo para la realización del mismo, la falta de potencia de los equipos informáticos disponibles para la ejecución de programas pesados y la falta de recursos económicos que impide adquirir programas y equipos para facilitar la producción de las distintas partes del trabajo.

La funcionalidad de la página web depende de la estabilidad del servidor en el que residen los archivos que la hacen funcionar, igual que la calculadora.

Una de las principales limitaciones del estudio es el rango de años seleccionado, ya que es bastante reducido, abarcando solo 16 años. Que el intervalo sea tan corto limita la recogida de tendencias a largo plazo, haciendo que la predicción de la esperanza de vida no sea tan exacta. Esto se puede solucionar ampliando el rango, sin embargo los datos de la mayor parte de las variables no tienen un rango más amplio. Ampliarlo no es lo más factible para este estudio, ya que habría gran cantidad de datos faltantes, y para que no ocurra se debería buscar en multitud de otras bases de datos, lo cual es demasiado costoso. Por ello la mejor decisión es focalizar los recursos que se tienen en maximizar la interpretación de los datos con el rango obtenido.

Otra posible limitación es que la identificación de tendencias dentro de la Unión Europea no se pueden extrapolar para otros países que no pertenezcan a este grupo. Esto es debido a que los factores socioeconómicos, geográficos, ambientales y médicos varían considerablemente fuera de Europa. Se puede solucionar este problema realizando un análisis más exhaustivo, añadiendo datos de otras regiones del mundo al estudio. Al igual que con la limitación anterior, no se poseen los recursos necesarios para la realización de un análisis de tal magnitud y es más adecuado centrarse en la Unión Europea.

En resumen, aunque el estudio tiene limitaciones, estas también abren puertas a futuras investigaciones que podrían superar estos desafíos y proporcionar una comprensión aún más robusta de la esperanza de vida a nivel global.

Comparación con el estado del arte

Si se compara el trabajo con el de Bilas et al., se puede ver que el análisis del trabajo sigue con el estudio de ellos. Se ha encontrado una relación directa y positiva con la esperanza de vida y el PIB per cápita, que tiene sentido actual. Cuánto más dinero un país tiene, más años podría esperar la población de este país. En relación con la segunda variable, el nivel de escolarización, este trabajo no tiene una variable en sí para medir esta relación, pero sí tiene la variable *schooling*, que mide el número de años de escolarización obligatorio de cada país. Según el análisis realizado, la variable tiene una relación directa y positiva, que significa con más años de escolarización obligatorio, más años la gente puede esperar vivir.

Con respecto al estudio de Poças et al., las variables de *smoking* y *alcohol* siguen los resultados de este estudio. Tienen una relación inversa y negativa con la esperanza de vida, que tiene sentido en el mundo real. Es un estilo de vida peligroso, según el estudio (Poças et al., 2020), y baja la esperanza de vida. En contra del estudio, se ha encontrado en el trabajo una relación directa y positiva entre la variable *CO2 emissions* y la esperanza de vida. No tiene sentido que con mayor emisiones de CO2 aumente la esperanza de vida. Se puede decir que los países con mayor esperanza de vida tienen otros factores que influyen más en la esperanza de vida y ocultan el efecto de la variable de CO2 emisiones, pero esto necesita más investigación.

Según el análisis geográfico del trabajo, siguen los mismos resultados del estudio realizado por Richardson et al., donde se ha encontrado que los países de la región este tiene menor esperanza de vida que los países de otras regiones.

Conclusiones

Se ha cumplido el objetivo general del proyecto de analizar los distintos factores que influyen en la esperanza de vida de un país, para así poder proporcionar decisiones estratégicas, medioambientales y saludables. Se han encontrado varias relaciones entre las variables del dataset y la esperanza de vida dentro de la Unión Europea entre los años 2000 y 2015.

OBJETIVOS ESPECÍFICOS:

1. Se ha cumplido con el primer objetivo específico de efectuar un análisis del estado del arte. Se han identificado algunas de las últimas investigaciones relacionadas con la esperanza de vida y sus factores relacionados, situando el proyecto en un contexto actual. Se han utilizado algunos estudios relevantes para poder continuar y plantear análisis, adquiriendo así nuevos conocimientos.
2. En cuanto al segundo objetivo específico, este se ha cumplido, ya que se han utilizado numerosas librerías de Python para cargar el conjunto de datos y realizar el análisis, lo que ha permitido adquirir nuevos conocimientos.
3. El objetivo específico número tres se considera un éxito, pues se ha realizado un estudio comparativo entre países de diferentes regiones, como se puede observar en resultados.
4. Se ha cumplido con el cuarto objetivo específico de evaluar los impactos de varias variables en la esperanza de vida mediante distintas librerías de Python. Se han mostrado al menos tres relaciones entre las variables y la esperanza de vida.
5. El quinto objetivo que trata de diseñar y crear una página web para presentar todo el trabajo, también ha sido alcanzado con éxito. Además, se ha vinculado la calculadora con la página web para que el usuario pueda introducir sus datos y obtener su esperanza de vida.

Se ha identificado que la relación más importante entre la esperanza de vida y los países de la Unión Europea es la variable *adult mortality*. Cada país tiene sus factores distintos que afectan a la esperanza de vida, pero hay factores en común que afectan a la mayoría de los países.

Una conclusión del análisis geográfico es que la esperanza de vida media de la región este y la región central, en los años 2000-2015, siempre es menor que las otras regiones. Mientras que, en las regiones del norte, sur, y oeste varían entre ellas cuál tiene la esperanza de vida media más alta. Otra conclusión interesante es que todas las regiones tienen una esperanza de vida media mayor en 2015 en comparación con la de 2000.

Este trabajo sirve para ver cómo los gobiernos de los países de la Unión Europea pueden alargar la esperanza de vida de su población y mejorar la calidad de su vida. Cada gobierno puede saber dónde invertir su presupuesto para tener los mejores resultados posibles.

Bibliografía

Bilas, V., Franc, S., & Bošnjak, M. (2014). Determinant factors of life expectancy at birth in the European union countries. PubMed, 38(1), 1-9. Recuperado el 11 de marzo de 2024, de <https://pubmed.ncbi.nlm.nih.gov/2485159>

CIA.gov. (2024). Central Intelligence Agency. Recuperado el 19 de febrero de 2024, de <https://www.cia.gov/the-world-factbook/field/life-expectancy-at-birth/country-comparison/>

Dattani, S., Rodés-Guirao, L., Ritchie, H., Ortiz-Ospina, E., & Roser, M. (2023). Life expectancy. Our World in Data. Recuperado el 26 de febrero de 2024, de <https://ourworldindata.org/life-expectancy>

INE - Instituto Nacional de Estadística. (2024). Productos y Servicios / Publicaciones / Publicaciones de descarga gratuita. Recuperado el 13 de marzo de 2024, de https://www.ine.es/ss/Satellite?L=es_ES&c=INESeccion_C&cid=1259926380048&p=1254735110672&pagename=ProductosYServicios/PYSLayout

Janssen, F., Bardoutsos, A., Gewily, S. E., & De Beer, J. (2021). Future life expectancy in Europe taking into account the impact of smoking, obesity, and alcohol. eLife, 10. Recuperado el 9 de marzo de 2024, de <https://doi.org/10.7554/elife.66590>

KumarRajarshi. (2018). Life expectancy (WHO). Kaggle. Recuperado el 12 de febrero de 2024, de <https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who>

Lab, G. D. (2021). Subnational HDI (v7.0). Global Data Lab. Recuperado el 25 de febrero de 2024, de

<https://globaldatalab.org/shdi/table/esch/CZE+DEU/?levels=1%2B4&years=2015%2B2014%2B2008%2B2000&interpolation=0&extrapolation=0>

McEvoy, O. (2023). Czechia HDI score 2021. Statista. Recuperado el 23 de marzo de 2024, de <https://www.statista.com/statistics/880414/human-development-index-of-czechia/>

Naciones Unidas. (2024). The 17 goals | sustainable development. United Nations. Recuperado el 20 de marzo de 2024, de <https://sdgs.un.org/goals>

Poças, A., Soukiazis, E. & Antunes, M. (2020). Factors Explaining Life Expectancy at Age 65: A Panel Data Approach Applied to European Union Countries. Soc Indic Res 150, 265–288 . Recuperado el 10 de abril de 2024, de <https://doi.org/10.1007/s11205-020-02290-2>

Richardson, E., Pearce, J., Mitchell, R., Shortt, N. & Tunstall, H. (2013). Have regional inequalities in life expectancy widened within the European Union between 1991 and 2008?, European Journal of Public Health, Volume 24, Issue 3, June 2014, Pages 357–363, Recuperado el 6 de abril de 2024, de <https://doi.org/10.1093/eurpub/ckt084>

The World Bank. (2020). CO2 emissions (metric tons per capita) - european union. World Bank Open Data. Recuperado el 8 de abril de 2024, de <https://data.worldbank.org/indicator/EN.ATM.CO2E.PC?end=2015&locations=EU&start=1999>

The World Bank. (2024). Current health expenditure (% of GDP) - european union. World Bank Open Data. Recuperado el 6 de marzo de 2024, de <https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS?end=2015&locations=EU&start=2000>

The World Bank. (2024). Current health expenditure per capita (current US\$) - european union. World Bank Open Data. Recuperado el 6 de marzo de 2024, de <https://data.worldbank.org/indicator/SH.XPD.CHEX.PC.CD?end=2014&locations=EU&start=2000>

The World Bank. (2022). GDP per capita (current US\$) - slovak republic. World Bank Open Data. Recuperado el 25 de marzo de 2024, de <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2015&locations=SK&start=1998>

The World Bank. (2022). Población, total - czechia. World Bank Open Data. Recuperado el 25 de marzo de 2024, de <https://datos.bancomundial.org/indicador/SP.POP.TOTL?end=2015&locations=CZ&start=2000>

The World Bank. (2022). Población, total - slovak republic. World Bank Open Data. Recuperado el 25 de marzo de 2024, de <https://datos.bancomundial.org/indicador/SP.POP.TOTL?end=2015&locations=SK&start=2000>

The World Bank. (2020). Prevalence of current tobacco use (% of adults) - european union.

World Bank Open Data. Recuperado el 19 de marzo de 2024, de <https://data.worldbank.org/indicator/SH.PRV.SMOK?end=2001&locations=EU>

Trading Economics. (2024). Trading economics: 20 million indicators from 196 countries.

TRADING ECONOMICS | 20 million INDICATORS FROM 196 COUNTRIES.
Recuperado el 25 de marzo de 2024, de <https://tradingeconomics.com/>

UNData. (2015). Undata | record view | school life expectancy (years). Primary to tertiary education (ISCED 1 to 8). United Nations. Recuperado el 19 de marzo de 2024, de https://data.un.org/Data.aspx?d=UNESCO&f=series%3ASLE_1T6

UNICEF. (2022). How many babies are born a year?. UNICEF DATA. Recuperado el 19 de marzo de 2024, de <https://data.unicef.org/how-many/how-many-babies-are-born-a-year/>

WHO. (2019). Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol). World Health Organization. Recuperado el 10 de abril de 2024, de [https://www.who.int/data/gho/data/indicators/indicator-details/GHO/alcohol-recorded-per-capita-\(15-\)-consumption-\(in-litres-of-pure-alcohol\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/alcohol-recorded-per-capita-(15-)-consumption-(in-litres-of-pure-alcohol))

WHO. (2021). Mean non-HDL cholesterol (age-standardized estimate). World Health Organization. Recuperado el 4 de marzo de 2024, de https://www.who.int/data/gho/data/indicators/indicator-details/GHO/chol_05

Anexos

Tablas

Variable	Descripción
Country	Nombre del país.
Year	Año.
Status	País desarrollado o en estado de desarrollo.
Life expectancy	Esperanza de vida en años.
Adult mortality	Probabilidad de morir entre las edades 15 y 60 años cada 1000 de población.
Alcohol	Consumición de alcohol en litros de las personas mayores que 15 años per cápita.
Total expenditure	Gastos en el sector de salud per cápita en dólares.
BMI	Índice de masa corporal de la población.
Under-five deaths	Números de muertes de niños menores de cinco años cada 1000 personas.
Percentage expenditure	Gastos en el sector de salud como un porcentaje del PIB.
GDP	Producto interior bruto (PIB) per cápita.
Population	Población del país.
Thinness 1-19 years	Porcentaje de la prevalencia de delgadez de las personas de las edades de 1 hasta 19 años.
Thinness 5-9 years	Porcentaje de la prevalencia de delgadez de las personas de las edades de 5 hasta 9 años.
Income composition of resources	Índice de desarrollo humano (IDH) de un rango de 0 a 1.
Schooling	Número de años obligatorios de escolarización.
Average temperature	Media de la temperatura del país en celsio.
CO2 emissions	Emisiones de dióxido de carbono en toneladas métricas per cápita.
Cholesterol	Media de colesterol no HDL.
Smoking	Porcentaje de la prevalencia de consumición de tabaco en adultos.

Tabla 5: Descripción de cada variable del dataset.

	Regresión lineal	Regresión Logarítmica
Variable	R value	R value
Smoking	-0.46	-0.47
Schooling	0.42	0.42
Average temperature	0.14	0.05
Status	0.13	-
Under-five deaths	0.05	0.09
Adult Mortality	-0.79	-0.83
BMI	0.26	0.26
Population	0.16	0.15
Thinness 5-9 years	-0.66	-0.62
Thinness 1-19 years	-0.67	-0.63
Total expenditure	0.7	0.77
Alcohol	-0.27	-0.27
Year	0.36	0.35
CO2 emissions	0.18	0.26
Cholesterol	-0.57	-0.57
Percent expenditure	0.15	-0.02
GDP	0.37	0.28
Income composition of resources	0.7	0.7

Tabla 6: Resumen de los coeficientes de correlación.

Gráficos

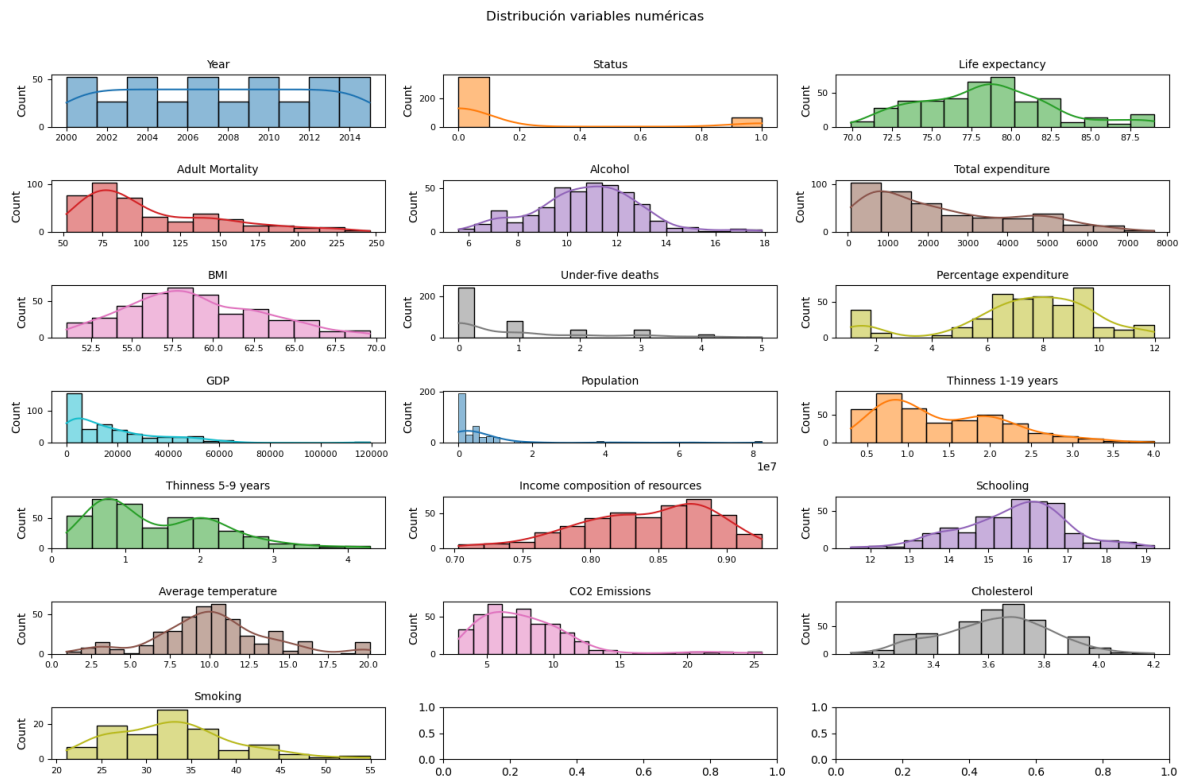


Gráfico 1: Distribuciones de las variables del dataset.

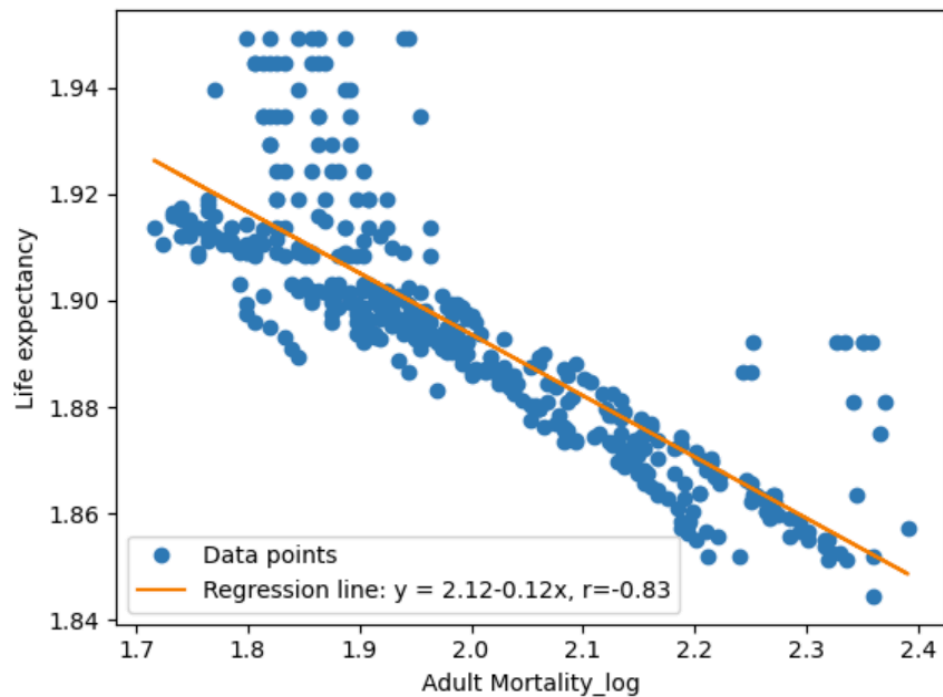


Gráfico 2: Regresión logarítmica de la variable adult mortality contra life expectancy.

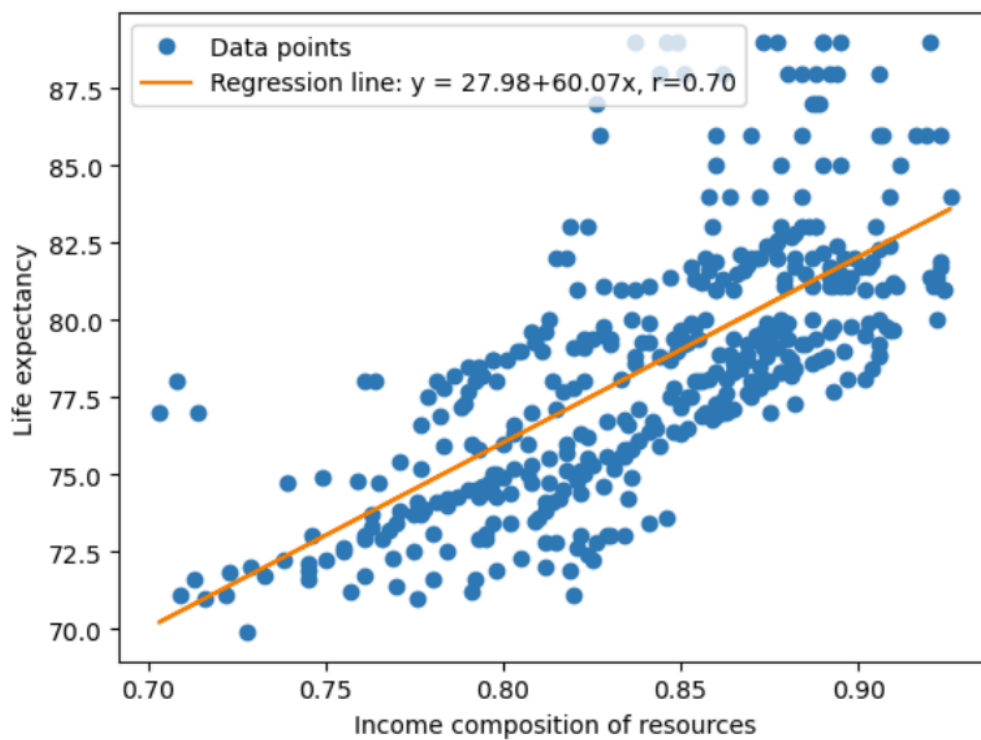


Gráfico 3: Regresión lineal de la variable income composition of resources contra life expectancy.

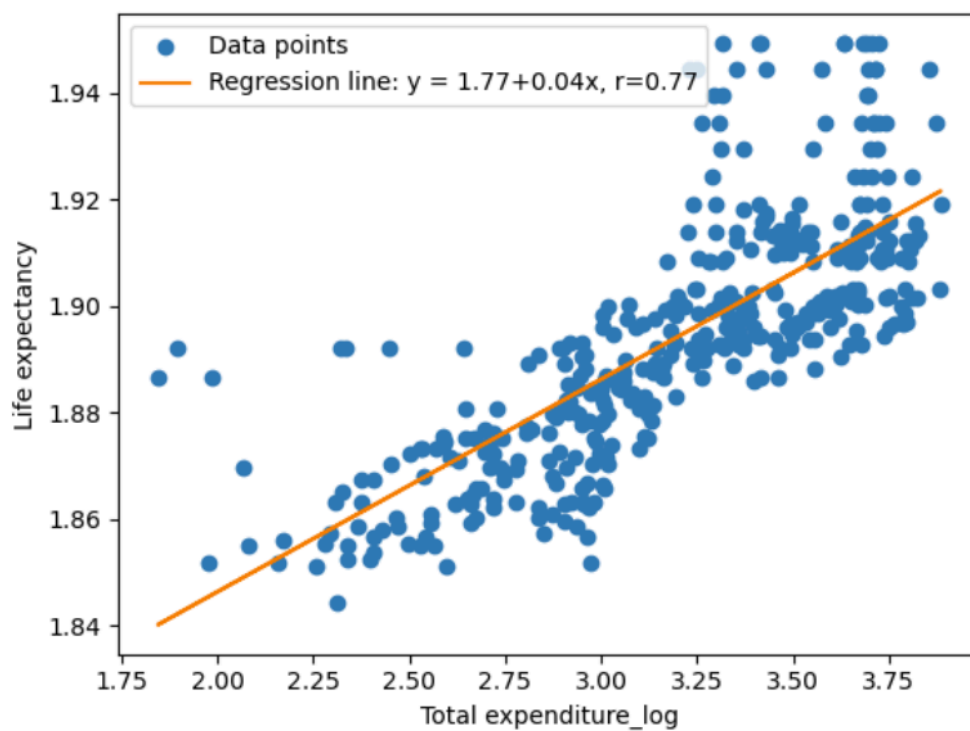


Gráfico 4: Regresión logarítmica de la variable total expenditure contra life expectancy.

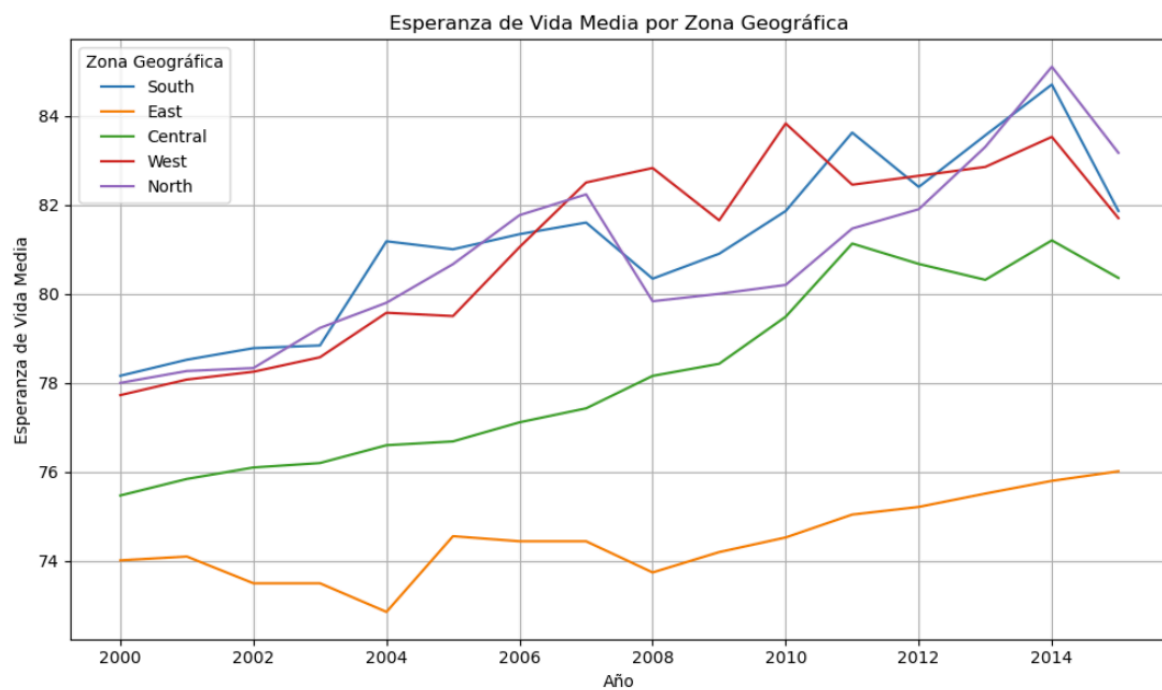


Gráfico 5: Esperanza de vida media por cada región de Europa.