



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.....	5
Misión.....	5
Motivación.....	5
Objetivos.....	6
Metodología.....	7
Estructura.....	8
Estado del arte.....	8
Alcance del proyecto.....	11
Hipótesis.....	11
Requisitos.....	11
Restricciones.....	12
Entregables.....	13
Criterios de éxito.....	13
Alineación con los ODS.....	14
Calendario del proyecto.....	15
Diagrama de Gantt.....	15
Distribución de las tareas.....	15
Materiales y métodos.....	16
Origen de los datos.....	16
Análisis exploratorio y de calidad de los datos.....	16
Herramientas utilizadas.....	17
Resultados.....	18
Validación y comentarios de los resultados.....	18
Contribuciones y limitaciones del trabajo.....	18
Comparación con el estado del arte.....	20
Conclusiones.....	20
Bibliografía.....	21
Anexos.....	26
Tablas.....	26
Gráficos.....	27

ÍNDICE DE TABLAS:

Tabla 1. Restricciones del proyecto.....	12
Tabla 2. Lista de tareas del diagrama de Gantt.....	15
Tabla 3. Diagrama de Gantt.....	15
Tabla 4. Paquetes de trabajo.....	15
Tabla 5: resumen de los coeficientes de correlación.....	26

ÍNDICE DE GRÁFICOS:

Gráfico 1: Regresión logarítmica de la variable adult mortality contra life expectancy.....	27
Gráfico 2: Regresión lineal de la variable income composition of resources contra life expectancy....	27
Gráfico 3: Regresión logarítmica de la variable total expenditure contra life expectancy.....	28

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, s. f.). Por otro lado, la Organización Mundial de la Salud (OMS) define la esperanza de vida como el número promedio de años que una persona puede esperar vivir desde el nacimiento (Esperanza de Vida, s. f.). 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, s. f.)

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 han estudiado las relaciones entre la esperanza de vida y otras variables para cada país. Y por último, se ha llevado a cabo un análisis adicional considerando la posición geográfica de los países, así como un estudio específico de los países del sur de la Unión Europea.

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, 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 causa-efecto 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, s. f.). El desarrollo económico determina mejoras en las condiciones sociales y un aumento en la esperanza de vida.

Hay muchos estudios realizados sobre este tema, pero la mayoría de ellos son en inglés. Por lo que, la base de datos utilizada en este proyecto estará basada en las diversas investigaciones de estudios globales publicados en inglés.

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.

Además, 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, pues 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, consumición 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, incluyendo un modelo de efectos al azar, modelo de efectos fijos, y un modelo de Least Squares Dummy Variables. 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 polución 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 influyó 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.

Los estudios pasados usan análisis de datos de paneles para extraer conclusiones sobre el dataset en general. Para hacer el análisis geográfico, se puede referenciar y ampliar el estudio de Richardson et al. (2013).

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 relaciones entre la esperanza de vida y otros 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 son enfermedades padecidas o drogas consumidas con frecuencia.

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.
Acceso restringido.	De momento no hay un sistema de creación de cuenta para la utilización de la calculadora desde la web, sin embargo, los que tengan claves de acceso podrán hacer uso libre de ella tanto en la web como en la aplicación.

Tabla 1. Restricciones del proyecto

Entregables

Un entregable que es la memoria del proyecto. En la que se ha detallado todo el proceso: la motivación, el estado del arte, el alcance del trabajo, el análisis de los datos, los resultados, las conclusiones, y las limitaciones del proyecto.

Otro entregable es el propio dataset. Ha sido costoso conseguir un dataset tan grande, compilando tantas variables distintas en un rango largo del tiempo para tantos países. Este dataset puede servir para cualquier otra persona que quiere estudiar el tema de esperanza de vida, y puede enfocar y usar el tiempo para hacer un mejor análisis en vez de buscar datos. Estará subido a un repositorio de datasets libres y accesibles.

La página web, la app, y el modelo predictivo/la calculadora es otro entregable. A lo largo del proyecto se ha creado la página web en paralelo al análisis de los datos. Sirve para mostrar visualmente los resultados del trabajo. También, es accesible para cualquier persona para ver y usar. El modelo predictivo también servirá para un futuro análisis para un investigador para entretener y conformar a lo que necesite.

Criterios de éxito

Si se cumplen todos los objetivos especificados, el trabajo se consideraría un éxito. Se prevé que los objetivos sean alcanzables dentro del tiempo y con los recursos asignados para el proyecto.

Se puede considerar como otra condición de éxito: ampliar una relación existente entre la esperanza de vida y una variable. Por ejemplo, se sabe que hay una relación entre un factor geográfico y la esperanza de vida. Con la variable de temperatura media, se podría proporcionar más contexto a esta relación y afirmar que existe una relación directa o inversa

con la esperanza de vida. Aún no se sabe si esto es cierto, pero sería considerado un éxito si se pudiera lograr.

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, s. f.). 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, s. f.). Los estudios previos, sobre todo Bilas et al., 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, s. f.). 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

Name	Begin date	End date	Notes
Propuestas y selección del tema	2/19/24	2/21/24	
Elección dataset	3/8/24	3/8/24	
Definición de misión y objetivos	3/6/24	3/25/24	
Sustitución de variables y selección definitiva	3/8/24	3/26/24	
Comprobación de viabilidad de uso de página web y app móvil	2/26/24	3/29/24	
Búsqueda de literatura y trabajos relacionados	3/8/24	3/18/24	
Confección del cronograma	3/13/24	3/19/24	
Análisis de datos repetidos y/o ausentes	3/11/24	3/22/24	
Búsqueda y comparación de diferentes datasets	2/21/24	3/8/24	
Desarrollo de la aplicación móvil y la página web	3/18/24	4/12/24	
Estudio unidimensional de países	3/20/24	4/12/24	
Análisis de factores geográficos	3/20/24	4/12/24	
Elaboración de conclusiones	4/12/24	4/24/24	
Realización de presentación del proyecto	4/22/24	4/24/24	
Desarrollo de la presentación	3/13/24	4/23/24	
Análisis de factores socioeconómicos	3/20/24	4/12/24	
Evaluación de la operatividad	4/22/24	5/17/24	
Efectividad de la página web y app móvil	4/22/24	5/17/24	
Presentación final	5/17/24	6/13/24	

Tabla 2. Lista de tareas del diagrama de Gantt.

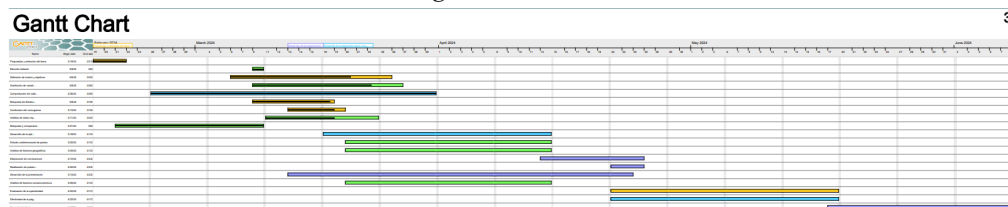


Tabla 3. Diagrama de Gantt.

Distribución de las tareas

Encargados	Paquete de Trabajo
Nadal	Desarrollo de app y página web
Carla	Coordinación de trabajos
Michele	Creación del modelo predictivo y calculadora
Isabelle	Análisis de datos
Fernanda	Difusión de resultados

Tabla 4. Paquetes de trabajo.

Materiales y métodos

Origen de los datos

La primera fuente del dataset es de Kaggle. Se llama “Life Expectancy (WHO)” y fue

subido al sitio web del usuario KumarRajarshi (2018). Incluye 193 países con 22 variables. Hemos cogido los 27 países de la Unión Europea, menos Chipre que no está incluido en el dataset original. Se eliminan además algunas variables que pensamos que no aportan mucha información como hepatitis B, measles (sarampión) y polio.

Adicionalmente, aparecen algunos datos faltantes o erróneos. Para rellenar la variable de consumición de alcohol, sacamos el data del WHO (WHO, s. f.). De The World Bank (The World Bank, s. f.), sacamos las variables PIB per cápita, gastos totales y porcentaje de gastos. Para los datos faltantes de Chequia, usamos Statista (McEvoy, 2023) para el índice de desarrollo humano, UNData (UNData, s. f.) para los años de la escolarización y The World Bank (The World Bank, s. f.) para la población. Para los años de la escolarización de Alemania, utilizamos y GlobalDataLab (Lab, G. D., s. f.). Para la población de Eslovaquia, sacamos los datos de The World Bank (The World Bank, s. f.).

Con esta revisión, faltan variables que consideremos importantes a estudiar, así que añadimos las variables temperatura media, emisión de dióxido de carbono, colesterol LDL en sangre y porcentaje de adultos que fuman. Sacamos las temperaturas de Trading Economics (Trading Economics, s. f.), el dióxido de carbono de The World Bank (The World Bank, s. f.), el colesterol LDL de WHO (WHO, s. f.) y el porcentaje de fumar de The World Bank (The World Bank, s. f.).

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. La única variable con una salvedad: el porcentaje de la población que fuma. Solo se dispone de un dato cada 5 años para esta variable, sin embargo se considera que es una variable importante a estudiar y que se podría analizar usando una interpolación de los datos. Para ello, se utilizó el método de interpolación de la librería *Scipy* de Python. Todas las variables, con excepción del país y el estado del desarrollo, son variables cualitativas.

En primer lugar, 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 tomó una transformación logarítmica de las variables con heterocedasticidad. Según las regresiones, y como se puede observar en la *tabla 5: resumen de los coeficientes de correlación* en el anexo, hay algunas

variables con un valor alto de correlación. La variable *adult mortality* tiene lo más alto con -0.83, una relación inversa, como se puede ver en el *gráfico 1: 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 2 y 3*. Las variables que miden la delgadez de la población hasta los 19 años también tienen altos coeficientes, de -0.62. Se puede afirmar que, para toda Europa, la variable más importante para controlar y medir la esperanza de vida es la variable *adult mortality*, pues esta refleja la probabilidad de morir entre las edades de 15 y 60 años por cada 1000 personas. Otras variables influyentes incluyen el gasto de capital per cápita en el sector de salud y el Índice de Desarrollo Humano.

Asimismo, se hizo un árbol de decisión regresión usando XGBoost. Tiene un valor de R-cuadrado de 0.85, un valor bastante alto. El modelo puede predecir valores con un cuadrado medio del error (MSE) de 2.54. Según este modelo, la variable más influyente también es la variable *adult mortality*, seguido por la variable *percentage expenditure*. La cual es parecida a la de *total expenditure*, porque es mostrado como un porcentaje del presupuesto.

Se han analizado las variables país por país.

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 SS.

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 Pandas y matplotlib, que a su vez permiten representar los datos de manera visual.

Resultados

Validación y comentarios de los resultados

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 cruciales 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 enorme 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 y a los resultados escritos aquí. También, se ha llevado a cabo un importante estudio de las distintas relaciones de factores multivariantes sobre la esperanza de vida.

Las principales limitaciones para la realización de este proyecto son la falta de conocimiento sobre el tema, la falta de datos fiables, 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 por supuesto la falta de recursos económicos que impide adquirir programas y equipos para facilitar la producción de las distintas partes del trabajo. Programas de pago como Adobe Photoshop, Filmora o Final Cut son algunos ejemplos, de los cuales no se ha podido hacer uso; otros como Android Studio o Blender son ejemplos de programas que requieren de recursos informáticos superiores a los utilizados.

La funcionalidad de la página web queda totalmente en manos de la estabilidad del servidor en el que residen los archivos que la hacen funcionar, igual que la calculadora. El vídeo presenta una calidad de audio y resolución variable también por la falta de equipo profesional específico, fruto de la falta general de capital invertido en el proyecto.

Una de las principales limitaciones del trabajo es el rango de años que se ha escogido. Esto es debido a que el rango es muy pequeño, de tan 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 mucho 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 de salud 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

Conclusiones

Se ha cumplido con el objetivo específico número cinco 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 conocimientos nuevos.

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. (s. f.). 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>

Esperanza de vida. (s. f.). Datosmacro.com. Recuperado el 26 de febrero de 2024, de <https://datosmacro.expansion.com/diccionario/esperanza-de-vida#:~:text=La%20esperanza%20de%20vida%20es%20un>

INE - Instituto Nacional de Estadística. (s. f.). Productos y Servicios / Publicaciones / Publicaciones de descarga gratuita. Recuperado el 13 de marzo de 2024, de https://ine.es/ss/Satellite?c=INESeccion_C&cid=1259944484459&p=1254735110672&pagename=ProductosYServicios%2FPYSLayout¶m1=PYSDetalleFichaIndicador¶m3=1259947308577

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. (s.f.). 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. (s. f.). 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. (s. f.). 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. (s. f.). 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. (s. f.). 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. (s. f.). 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. (s. f.). 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. (s. f.). 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. (s. f.). 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.PRVS.MOK?end=2001&locations=EU>

Trading Economics. (s. f.). 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. (n.d.). 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. (s. f.). 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. (s. f.). 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

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 5: resumen de los coeficientes de correlación.

Gráficos

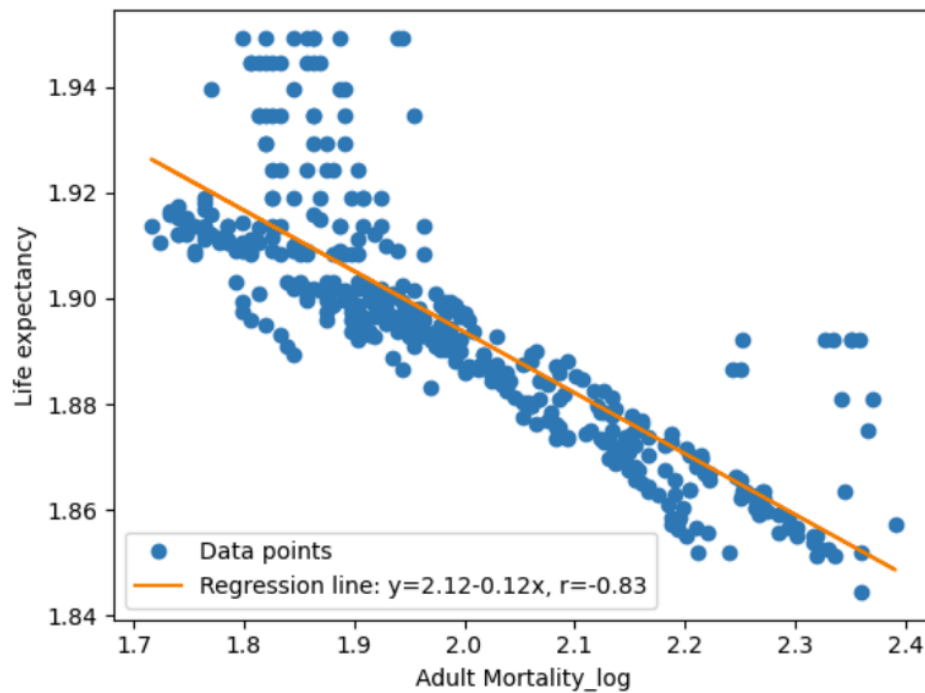


Gráfico 1: Regresión logarítmica de la variable adult mortality contra life expectancy.

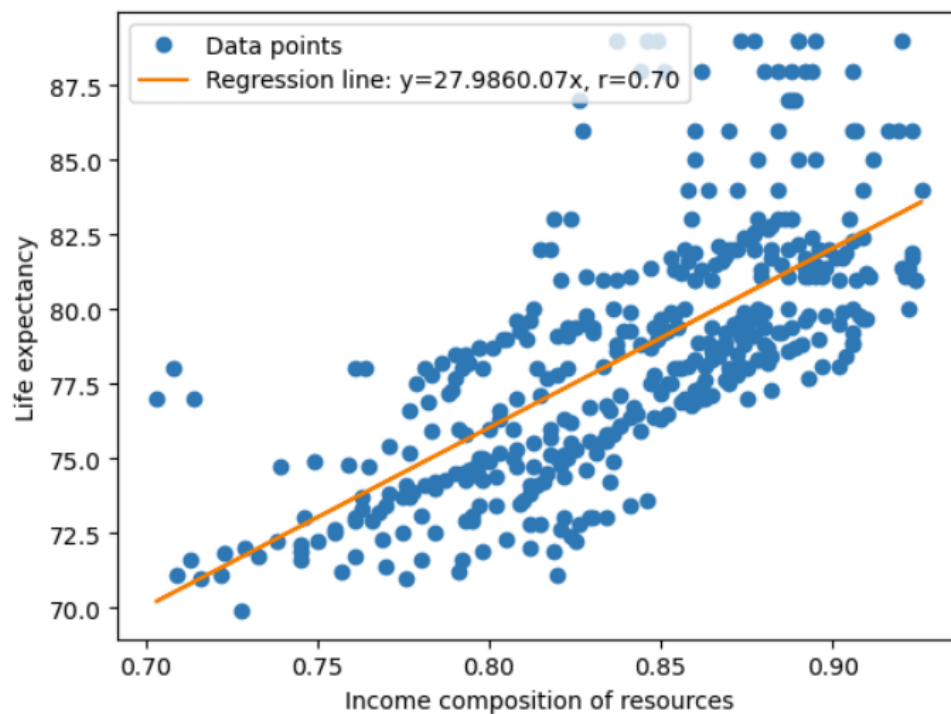


Gráfico 2: Regresión lineal de la variable income composition of resources contra life expectancy.

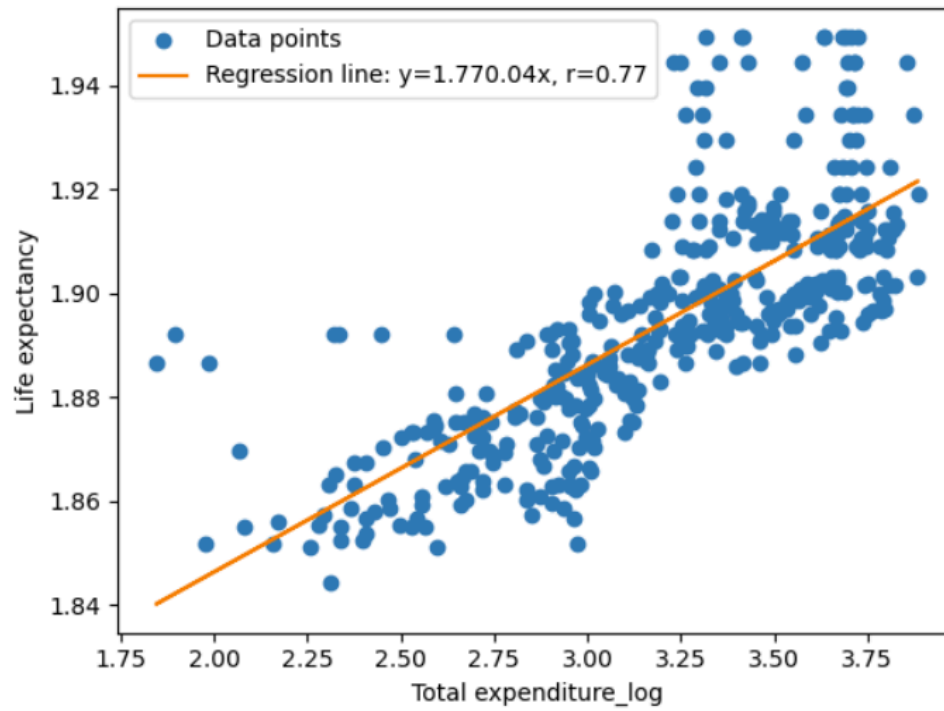


Gráfico 3: Regresión logarítmica de la variable total expenditure contra life expectancy.