

*Health analytics: Análisis de la esperanza de vida
a nivel mundial
(2000-2019)*

Memoria del proyecto final del máster en *Data Analytics*

Nuclio Digital School

Tutor: Pedro Costa

Rosa Rubio
Laia Campoy
Raúl Fernández
Bamba Tandine
Alicia Penalba

Septiembre de 2024

0. ÍNDICE

1. INTRODUCCIÓN.....	2
2. NORMALIZACIÓN DE DATOS.....	4
2.1 Descripción de los datos iniciales.....	4
2.2 Países y marco temporal.....	4
2.3 Elección y creación de columnas.....	5
3. ANÁLISIS DE DATOS.....	6
3.1 Esperanza de vida: hipótesis simples.....	6
3.1.1 Insights primer análisis.....	10
3.2 Análisis multivariable: Clustering K-means.....	11
2.2 Insights análisis de clusters.....	15
4. MODELOS DE MACHINE LEARNING.....	18
4.1 Regresión Logística.....	18
4.2 Árboles de decisión.....	20
4.3 <i>Random forest</i>	20
4.4 <i>XGBoost</i>	21
4.5 Comparativa y uso de los modelos.....	21
5. CONCLUSIONES FINALES.....	23
5. 1 Recomendaciones para la <i>World Health Organization</i>	25
5.2 Link al Dashboard.....	26
6. BIBLIOGRAFÍA.....	27
7. ANEXOS.....	29

1. INTRODUCCIÓN

El desarrollo y crecimiento de nuestra sociedad ha sido posible gracias a los avances y descubrimientos que hemos ido acumulando a lo largo de la historia. Muchos de ellos han sido clave para la mejora de las condiciones sanitarias, que han tenido un impacto claramente visible en nuestras perspectivas de futuro. Cada vez vivimos más tiempo y disfrutamos de una mejor calidad de vida, pero las diferencias en la longevidad entre distintas regiones del mundo persisten, reflejando desigualdades en el acceso a servicios de salud y en las condiciones socioeconómicas.

Para entender y medir estas diferencias, se utiliza la esperanza de vida al nacer, un indicador demográfico ampliamente usado por entidades gubernamentales y organizaciones internacionales. Este indicador mide la probabilidad de que los habitantes de un país vivan hasta cierta edad, basándose en la duración de vida promedio de una cohorte de población hipotética expuesta a las tasas de mortalidad observadas en un año específico, sin tener en cuenta cambios futuros (Esteban Ortiz-Ospina, 2017). Este promedio, junto con otros datos de carácter socioeconómico y de salud pública puede darnos una imagen general de los países y de su situación en momentos determinados, así como también ayudarnos a establecer comparaciones entre ellos.

El interés por la demografía y sus efectos directos en la economía y en la vida de las personas se ha visto representado en los trabajos de múltiples autores que han moldeado nuestra comprensión actual de la esperanza de vida. Uno de los pioneros fue John Graunt en el siglo XVII, quien sentó las bases de esta disciplina mediante el análisis de registros de mortalidad en el Reino Unido. Posteriormente, Edmond Halley desarrolló una de las primeras tablas de vida basadas en datos reales, lo que permitió estimar la esperanza de vida de manera más precisa. En el siglo XIX, William Farr avanzó en la estadística sanitaria, mejorando la metodología para calcular la esperanza de vida, relacionándola con otros factores. En el siglo XX, Alfred Lotka y Benjamin Gompertz hicieron contribuciones significativas a la modelización matemática de la mortalidad, con la ley de Gompertz siendo especialmente influyente en la comprensión de cómo la mortalidad aumenta con la edad. Finalmente, en tiempos más recientes, Samuel H. Preston es ampliamente reconocido por su trabajo sobre la relación entre el desarrollo económico y la esperanza de vida, especialmente a través de la “curva de Preston”, que ha sido fundamental en el estudio de la longevidad a nivel global. De igual manera, el demógrafo Massimo Livi Bacci, en su obra *Historia mínima de la población mundial* (1990), subraya cómo los factores económicos, sociales y ambientales han influido profundamente en la evolución de las poblaciones a lo largo del tiempo. Su análisis pone énfasis en la importancia de comprender estas dinámicas

para enfrentar desafíos contemporáneos, como el envejecimiento poblacional y las migraciones masivas. Este enfoque complementa el estudio de la esperanza de vida al resaltar la interconexión entre estos factores y la longevidad de las poblaciones.

En resumen, el análisis de la esperanza de vida ha abarcado una variedad de caminos que reflejan la complejidad de este tema, pero todos coinciden en la relevancia de variables económicas, sociales y ambientales como determinantes clave (Monsef, A. et al., 2015).

Basándonos en este hecho, el objetivo de nuestro trabajo es identificar qué variables de estudio son determinantes para la esperanza de vida y formular recomendaciones para políticas públicas, poniendo el énfasis en la mejora de las condiciones de vida de las distintas regiones mundiales. Se estudiará la evolución de varios factores para 60 países entre los años 2000 y 2019. En nuestra muestra se incluye una representación de todos los continentes y de sus respectivas regiones. Los datos para la realización del análisis se han extraído de *Our World in Data*, una base de datos pública que se apoya de las estadísticas de otras organizaciones como el Banco Mundial, la Organización Mundial de la Salud o la Universidad de Oxford.

La estructura del trabajo se divide en tres partes principales. Primero, expondremos el proceso de limpieza y normalización de los datos, así como la elección de los países de estudio y las variables. En segundo lugar, pasaremos al análisis principal, dividido a su vez en dos partes: la primera, donde establecemos grupos de hipótesis simples y realizamos un análisis de correlaciones temporal y por continentes; y la segunda, en la que presentaremos un modelo de *Clustering k-means* que hará agrupaciones por países teniendo en cuenta múltiples variables a la vez. Después de este apartado, presentaremos modelos de *machine learning* con el objetivo de hacer predicciones sobre la esperanza de vida. Finalmente, expondremos las conclusiones del análisis con recomendaciones para la *World Health Organization* (WHO), que vendrán acompañadas por un *dashboard* diseñado en *Looker Studio*.

2. NORMALIZACIÓN DE DATOS

2.1 Descripción de los datos iniciales

El primer paso que hemos dado ha sido tomar contacto con los datos que recibimos de *Our World in Data*, distribuidos en 8 tablas distintas. En todas ellas había un marco temporal y un número de países diferente. En algunos casos, se incluían agrupaciones de países comúnmente usadas para realizar estudios estadísticos, o bien territorios que no están constituidos como países. La distribución de los datos¹ era la siguiente:

- **Causas de muerte:** número de muertes por causa en números absolutos. En este conjunto había muertes por 30 causas distintas.
- **Suicidios:** número de suicidios por cada 100.000 habitantes
- **PIB per cápita:** valor del PIB per cápita en \$ internacionales en precios de 2017²
- **Vacunación:** % de menores de 1 año vacunados para 11 tipos de vacunas distintas.
- **Homicidios:** número de homicidios por cada 100.000 habitantes
- **Esperanza de vida:** años de esperanza de vida al nacer.
- **Indicadores de pobreza:** tabla con distintos indicadores de pobreza y desigualdad.
- **Gasto público en sanidad:** % de gasto público en sanidad respecto al PIB del país.

2.2 Países y marco temporal

Después de entender el contexto de los datos, tomamos varias decisiones. La primera fue seleccionar únicamente aquellos países que fueran comunes en todos los datasets. Para ello nos apoyamos de los códigos de país proporcionados, que comparamos con los códigos ISO, para buscar inconsistencias o países que pudieran tener nombres diferentes en los distintos conjuntos de datos. El listado obtenido fue de 122 países, de los cuales elaboramos una lista de 60 como muestra para nuestro estudio. Para elaborarla nos basamos en un criterio geográfico, seleccionando países de todos los continentes con representación de varias regiones. Además, incluimos diferentes tamaños territoriales. La lista puede observarse de manera detallada en la Tabla 2 de los anexos.

Posteriormente, tomamos la decisión de limitar el marco de estudio entre 2000 y 2019. Pese a disponer de datos para años anteriores en algunas variables, estos no eran siempre

¹ Para información más detallada, consultar la tabla 1 del Anexo.

² Los dólares internacionales son una moneda hipotética usada para hacer comparaciones monetarias sobre los estándares de vida. Los datos expresados en esta moneda tienen en cuenta la inflación existente entre países en el tiempo, así como también las diferencias en el coste de vida entre países (Fuente: Traducción de *Our World in Data*)

anuales y tampoco había consistencia entre los países. También decidimos excluir el año 2020, ya que no disponíamos de datos para la mayoría de causas de muerte. Además, consideramos que la pandemia COVID19 fue un factor excepcional que podría tener efectos en nuestro análisis, y que merecería un estudio aislado.

2.3 Elección y creación de columnas

El siguiente paso fue buscar la población de cada uno de los países para poder convertir el número de muertes por cada causa de valores absolutos a relativos. Una vez aplicada la fórmula, las causas de muerte pasaron a estar en la misma unidad que los suicidios u homicidios, es decir, en número de muertes por cada 100.000 habitantes. Cabe aclarar que el análisis se ha realizado sin diferenciar por grupos de edad ni por sexo. Aunque consideramos que sería interesante, nuestros datos no incluían ninguno de estos componentes. Debido a la magnitud del problema al que nos enfrentamos, decidimos aplicar un enfoque generalizado. Somos conscientes de esta limitación en nuestro estudio y consideramos que las diferencias por sexo y edad podrían ser una ampliación relevante en el futuro.

Durante el proceso de normalización, diseñamos un conjunto de hipótesis iniciales (que será detallado más adelante) para agrupar las causas de muerte en categorías específicas, facilitando así el análisis. Estas hipótesis sirvieron como base para la creación de nuevas columnas que reflejan la suma de muertes por cada grupo y su representación porcentual respecto al total, además de guiar la decisión en el orden y selección específica de columnas en el conjunto de datos.

Por ejemplo, los datos de pobreza y desigualdad incluían más de 100 columnas distintas, muchas de las cuales representaban los mismos valores en diferentes formatos (absoluto y relativo) o proporcionaban información sobre distintos percentiles poblacionales. Además, el número de valores nulos era muy elevado. Por todo ello, decidimos simplificar el uso de este dataset a cuatro columnas principales: el Índice de Gini y el Índice de Palma, comúnmente usados en estudios económicos sobre desigualdad, y el número total y el porcentaje de población que vive con menos de 2.15\$ al día, que es el umbral de pobreza reconocido internacionalmente y ajustado por paridad de poder adquisitivo, utilizado por el Banco Mundial. Del resto de los conjuntos de datos, decidimos utilizar toda la información disponible, aunque en el caso de las tasas de vacunación, nuestro enfoque se centró en las vacunas con mayor relevancia dentro de la comunidad sanitaria internacional, que coincidían con aquellas con menos datos faltantes³. El tratamiento de estos valores nulos se comentará en el apartado de análisis posterior. El último paso realizado fue simplificar los

³ Las vacunas en las que nos hemos centrado han sido: DTP, Tuberculosis, HepatitisB3, Sarampión y Polio.

nombres de las columnas y unir todos los datos en un único dataset mediante los países como clave, obteniendo así un conjunto único para proceder con el análisis.

3. ANÁLISIS DE DATOS

3.1 Esperanza de vida: hipótesis simples

Para empezar a desglosar el problema, observamos la evolución de la esperanza de vida por zonas y planteamos una serie de hipótesis simples con el objetivo de ver su relación con cada una de las variables por país.

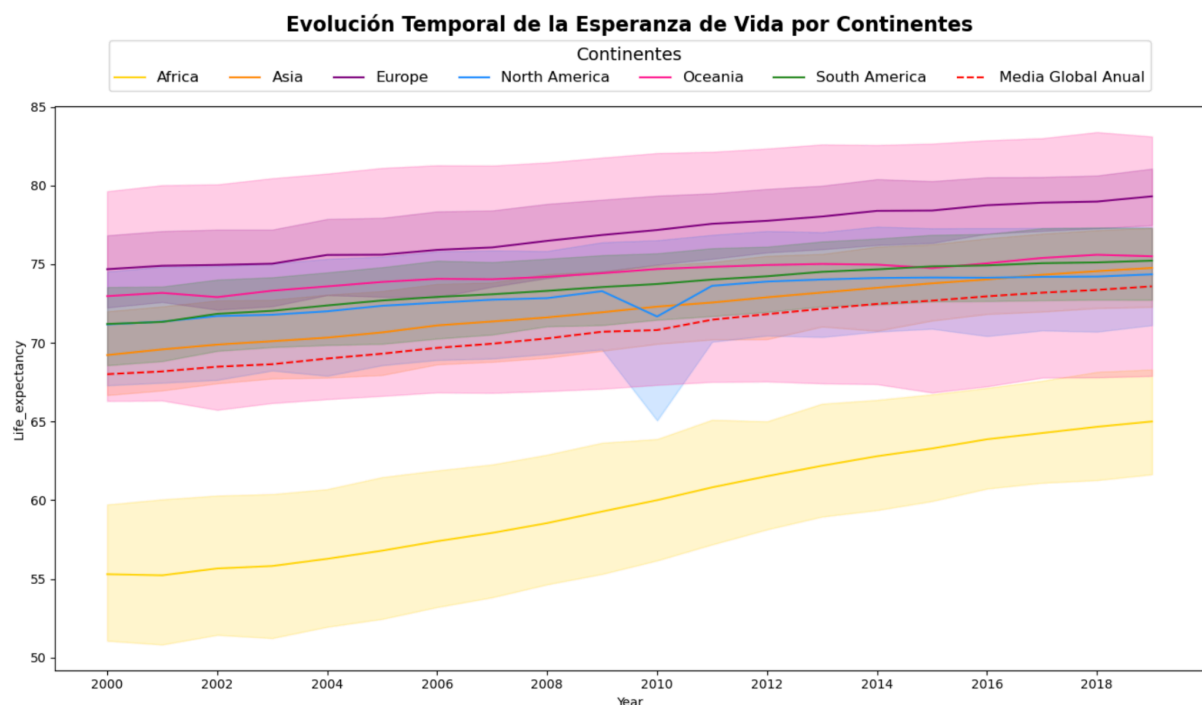


Gráfico 1: Evolución temporal de la EV por Continentes y media global. Fuente: elaboración propia con los datos de *Our World in Data*. Para el resto de gráficos, consultar el apartado 4 del Notebook *Estudio de la Esperanza de vida*.

El Gráfico 1 representa la evolución temporal de la esperanza de vida por continentes, así como la media global anual. Este gráfico destaca la notable disparidad entre África y las demás regiones, con Europa liderando en términos de esperanza de vida. Aunque la tendencia global es crecientemente positiva, África ha experimentado el mayor aumento, dado que partía de valores considerablemente más bajos. Por otro lado, se observa una disminución abrupta en Norteamérica debido al devastador terremoto de Haití en 2010. Hacia 2019, la esperanza de vida global se acercaba a los 70 años.

Con esta perspectiva, planteamos las siguientes hipótesis simples, que nos han permitido explorar cómo diferentes factores podrían estar influyendo en la esperanza de vida.

La esperanza de vida depende de:

Hipótesis 1: La capacidad económica del país

- Desarrollo económico: PIB per cápita y % de gasto en sanidad respecto al PIB
- Pobreza y desigualdad: índices de Gini, Palma y % por debajo de 2,15\$/día.

Hipótesis 2: Del grado de criminalidad y de factores políticos y geográficos

- Homicidios, guerras
- Causas de muerte geográficas: temperaturas extremas y fuerzas de la naturaleza

Hipótesis 3: Del desarrollo en sanidad

- Tasas de vacunación a menores de 1 año: DTP, Sarampión, Hepatitis, Tuberculosis y Polio.
- Muertes por enfermedades infecciosas, neonatales y maternas, diarrea, crónicas de riñón

Hipótesis 4: De la salud física

- Muertes por enfermedades infecciosas: HIV/AIDS, Infecciones respiratorias
- Muertes por enfermedades crónicas no transmisibles: cardiovasculares, respiratorias crónicas y diabetes.
- Muertes por cáncer
- Muertes por enfermedades somáticas relacionadas con el estilo de vida y la nutrición: deficiencia nutricional, de proteína, cirrosis y enfermedades digestivas.

Hipótesis 5: De la salud mental

- Muertes por enfermedades degenerativas: Parkinson, Alzheimer
- Muertes por suicidios, alcoholismo, consumo de drogas y auto lesiones

Hipótesis 6: Del nivel de accidentes

- Muertes por accidentes de tráfico, incendios, ahogamientos y envenenamientos

Una vez definido el *roadmap*, analizamos cada una de estas hipótesis en los distintos continentes, para tratar de identificar patrones por zonas. Para ello, realizamos tres tipos de representaciones, todas ellas con un código de color particular para cada continente. En primer lugar, usamos gráficos de series temporales para ver la evolución durante el período estudiado por países. En segundo lugar, nos centramos en la relación entre variables,

mediante diagramas de dispersión y la representación de las correlaciones de Spearman en mapas de calor. Para las hipótesis 2, 4, 5 y 6, totalmente relacionadas con el número de muertes también añadimos otras representaciones adicionales.

Debemos aclarar que en este proceso tuvimos siempre presente que estábamos asumiendo que la esperanza de vida depende de una única variable. Esta simplificación nos permitió conocer mejor el contexto de los datos antes de pasar a la siguiente fase del análisis, obteniendo una fotografía general por zonas geográficas. Aquí mostramos un ejemplo de las visualizaciones para el caso de África en la hipótesis de capacidad económica⁴:

Evolución de GDP_pc en Africa

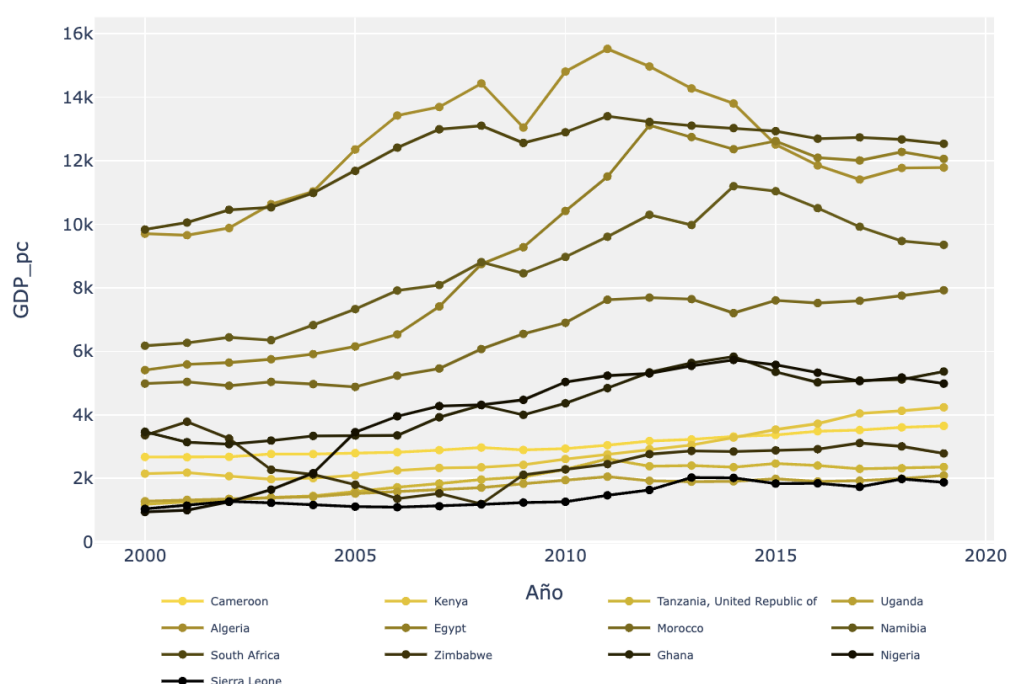


Gráfico 2: Evolución temporal del PIB per cápita en países Africanos. Fuente: elaboración propia con los datos de *Our World in Data*.

⁴ Para el resto de las visualizaciones, consultar el apartado 5. "Continentes" del Notebook *Estudio de la Esperanza de vida*

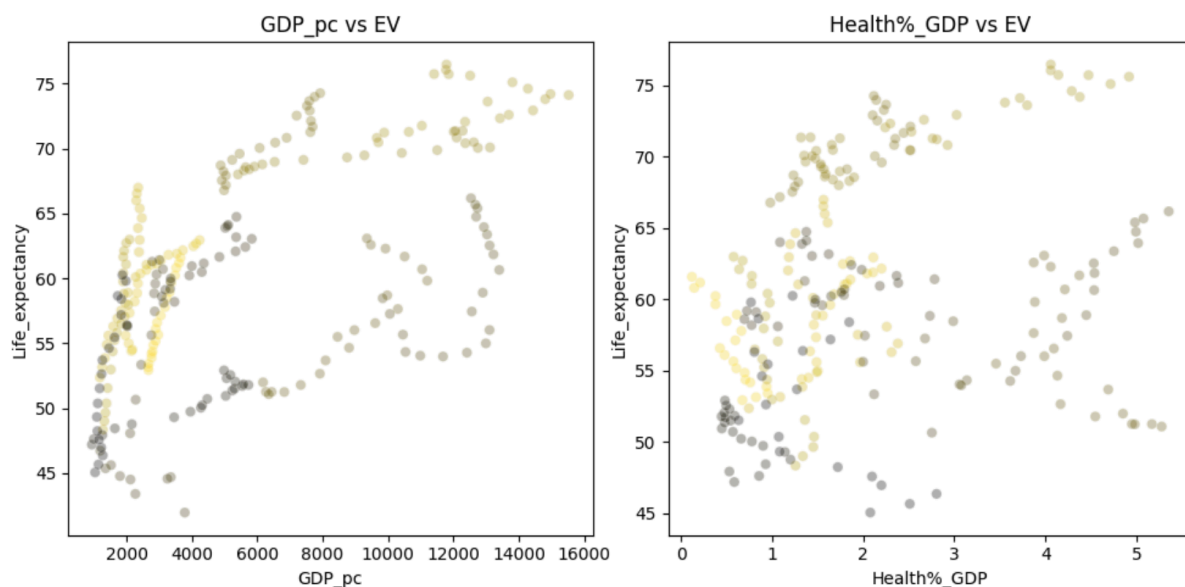


Gráfico 3: Diagrama de dispersión: PIB per cápita y % de gasto en sanidad con esperanza de vida. Datos de países Africanos para el período estudiado. Fuente: elaboración propia con los datos de *Our World in Data*.

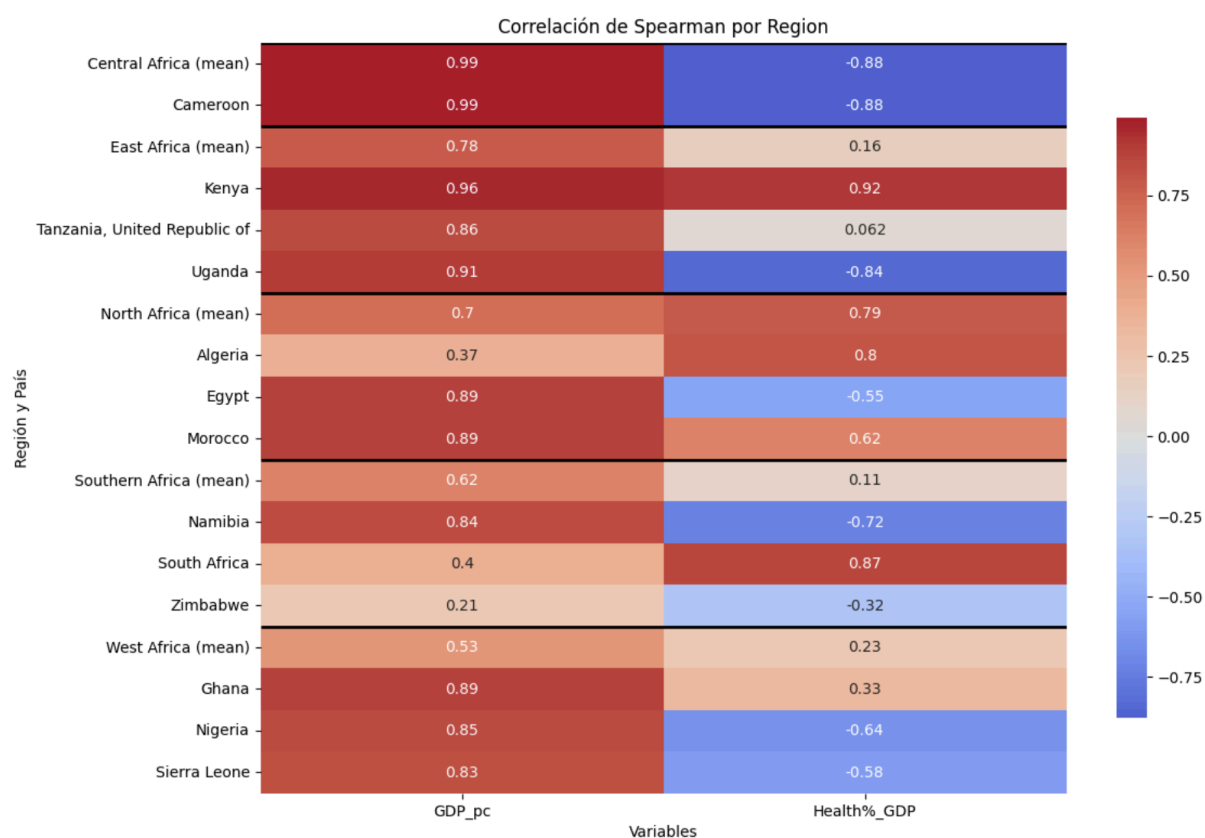


Gráfico 4: Mapa de calor de las correlaciones de Spearman: PIB per cápita y % de gasto en sanidad con esperanza de vida. Datos de países Africanos para el período estudiado. Fuente: elaboración propia con los datos de *Our World in Data*.

3.1.1 Insights primer análisis

Durante los últimos 20 años, la tendencia general de la esperanza de vida ha sido positiva en la mayoría de las regiones del mundo. En África, aunque el crecimiento ha sido desigual, muchas regiones han visto incrementos en la esperanza de vida gracias a mejoras en el acceso a la atención médica y campañas de vacunación más efectivas. En Asia y América Latina, también se ha observado un aumento sostenido.

Las causas de muerte han seguido una tendencia general a la baja en muchas regiones, con una notable disminución de enfermedades infecciosas y neonatales, particularmente en áreas donde estas eran más prevalentes. En África, la reducción de muertes por Malaria y VIH/SIDA en algunas regiones ha contribuido al aumento de la esperanza de vida, aunque estas enfermedades siguen teniendo un impacto significativo en ciertos lugares. En Asia y América Latina, la disminución de la mortalidad infantil y por enfermedades infecciosas ha sido clave para mejorar la longevidad. Sin embargo, las enfermedades crónicas, como las cardiopatías y el cáncer, han mostrado una tendencia al alza, especialmente en regiones con poblaciones envejecidas como Europa, Oceanía y América del Norte. Estas enfermedades han comenzado a predominar como las principales causas de muerte, reflejando los desafíos asociados con el envejecimiento de la población.

Las diferencias regionales en África son particularmente marcadas, con variaciones significativas en la esperanza de vida y las causas de muerte. Estas diferencias tienden a disminuir en regiones con esperanzas de vida más altas. En Asia y América Latina, aunque también existen variaciones regionales, estas son menos pronunciadas y tienen un impacto menor en la esperanza de vida general. En Europa y América del Norte, las diferencias regionales son mínimas, lo que refleja una mayor homogeneidad en términos de desarrollo económico y acceso a servicios de salud.

Aunque las hipótesis planteadas no pueden considerarse de manera aislada, hemos observado que las relacionadas con la capacidad económica y el desarrollo sanitario son las que muestran una relación más fuerte con la esperanza de vida, especialmente en las zonas que partían de niveles más bajos. El resto de factores, como las muertes por causas físicas, accidentes, conflictos geopolíticos y problemas de salud mental, no parecen tener un peso tan relevante por sí solas y no presentan patrones tan marcados. Sin embargo, combinadas con factores como la inestabilidad política y la debilidad de los sistemas sanitarios, pueden tener efectos significativos en la esperanza de vida. Un claro ejemplo es el caso Haitiano.

En resumen, la tendencia general en la esperanza de vida ha sido positiva en las últimas dos décadas, con muchas causas de muerte disminuyendo, especialmente aquellas relacionadas con enfermedades infecciosas y neonatales. Sin embargo, las enfermedades

crónicas, que afectan principalmente a las poblaciones envejecidas, han mostrado una tendencia al alza, lo que plantea nuevos desafíos en la gestión de la salud a nivel global.

3.2 Análisis multivariable: Clustering K-means

El análisis exploratorio previo nos permitió identificar cómo las diferentes variables se relacionaban individualmente con la esperanza de vida en distintos continentes y regiones. En esta segunda fase del proyecto, nuestro objetivo fue agrupar los países en función de un conjunto de factores clave. Utilizamos un modelo de *clustering k-means* para identificar nuevos patrones entre los países, con el propósito de obtener un análisis más profundo y complejo que complementara la primera parte del estudio. Decidimos excluir la esperanza de vida del proceso de agrupamiento, ya que nuestro enfoque era agrupar los países basándonos en variables independientes relacionadas con niveles de riqueza, tasas de vacunación y causas de muerte, para después analizar cómo la esperanza de vida variaba dentro de cada cluster.

Aunque muchas causas de muerte no parecían tener un impacto relevante en análisis individuales, observamos que en algunos contextos regionales, como en África, las causas relacionadas con enfermedades infecciosas o muertes prematuras tenían un peso significativo. Por otro lado, muchas causas de muerte, como las cardiopatías y el cáncer, están directamente relacionadas con la longevidad. Por estos motivos, decidimos incluir estas variables en el modelo.

Sabemos que este tipo de técnicas suelen comenzar con enfoques más sencillos, ya que incluir múltiples variables puede generar ruido. No obstante, decidimos seguir con nuestra metodología inicial, mitigando la complejidad mediante el Análisis de Componentes Principales⁵ (PCA), que nos permitió reducir la dimensionalidad y eliminar redundancias entre variables.

Metodología adoptada:

1. Eliminación de variables irrelevantes o con muchos nulos: Eliminamos aquellas variables con un porcentaje alto de datos faltantes, como las columnas de pobreza y desigualdad. También eliminamos la columna “año” y todas aquellas con valores no numéricos.

⁵ El PCA es una técnica de reducción de dimensionalidad que transforma un conjunto de variables posiblemente correlacionadas en un conjunto más pequeño de variables no correlacionadas, llamadas componentes principales. Estos componentes capturan la mayor parte de la variabilidad del conjunto de datos original, lo que permite simplificar el análisis manteniendo la información más relevante. En un modelo con 39 variables, esta era una parte crucial del proceso.

2. Imputación de valores faltantes: Las columnas restantes con valores nulos fueron imputadas utilizando la mediana, estadístico empleado en fases previas del análisis.
3. Análisis de multicolinealidad: Para asegurar la independencia de las variables seleccionadas, calculamos el VIF⁶ (*Variance Inflation Factor*). Variables con un VIF elevado pueden distorsionar los resultados debido a su alta correlación con otras⁷. En algunos casos, optamos por mantener variables a pesar de presentar VIFs superiores a 10, ya que representaban aspectos importantes que no queríamos excluir.

Fases del modelo:

1. Creación de la variable X y selección del número de clusters mediante el “Método del Codo”. En este caso el modelo indicó que 4 era el número adecuado.
2. Reducción de dimensionalidad con PCA, lo que nos permitió manejar la complejidad por el elevado número de variables.
3. Aplicación del modelo k-means, un algoritmo que asigna cada país al clúster más cercano basado en la media de las variables en ese clúster (el centroide).
4. Representación de los grupos en un mapa, permitiendo visualizar la distribución geográfica de los clusters.

El resultado obtenido fue el siguiente:

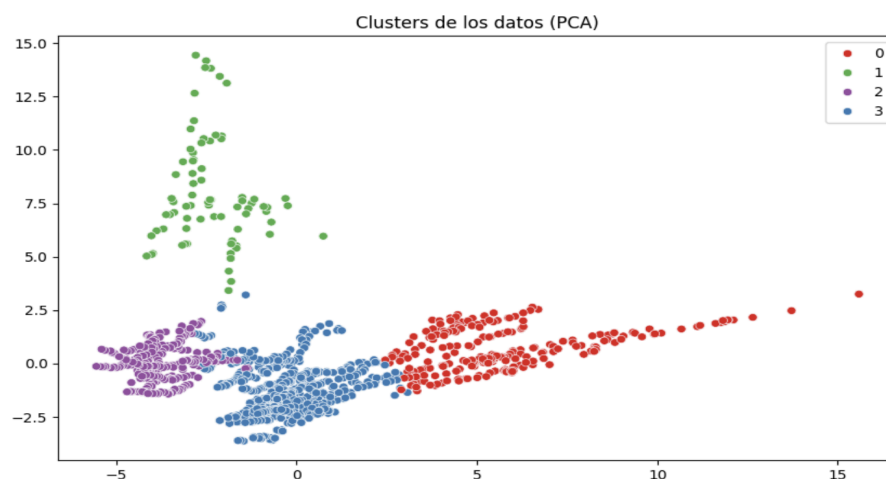


Gráfico 5 : Clusters del modelo k-means. Fuente: elaboración propia.

⁶ Un VIF de 1: Indica que las variables son independientes entre sí. Uno de entre 1 y 5 indica que hay una relación moderada, pero no problemática. Si es mayor de 5 o 10, puede haber redundancia entre esa variable y el resto, causando distorsión en los resultados. La esperanza de vida mostraba un VIF de 39,73. Para más coeficientes, consultar el apartado 6.1.3 del Notebook.

⁷ En este momento nos dimos cuenta de la relación entre las columnas de deficiencia nutricional y de proteína. Ya habíamos percibido que sus tendencias eran muy similares, pero observando los valores en las tablas vimos que sus valores eran prácticamente idénticos. Por ello, excluimos la columna de proteína.

Representados en el mapa para el final del período estudiado:

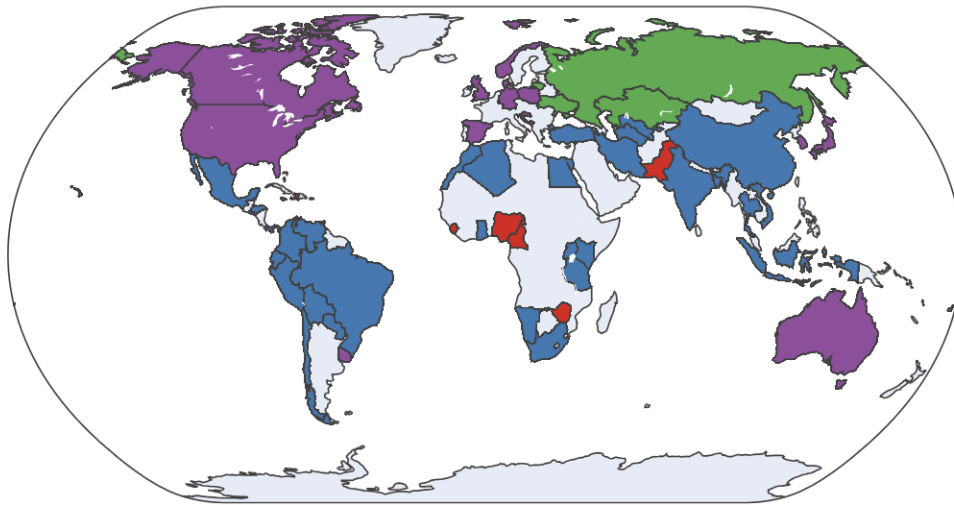


Imagen 1: Agrupación de países por clusters en el mapamundi. Fuente: Elaboración propia.

Después de obtener los grupos, los nombramos en base a su esperanza de vida media en el período 2000-2019. El grupo morado está formado por los países con mayor esperanza, el azul y el verde con un nivel medio y por último el grupo rojo, en el que están las naciones con un nivel más bajo.

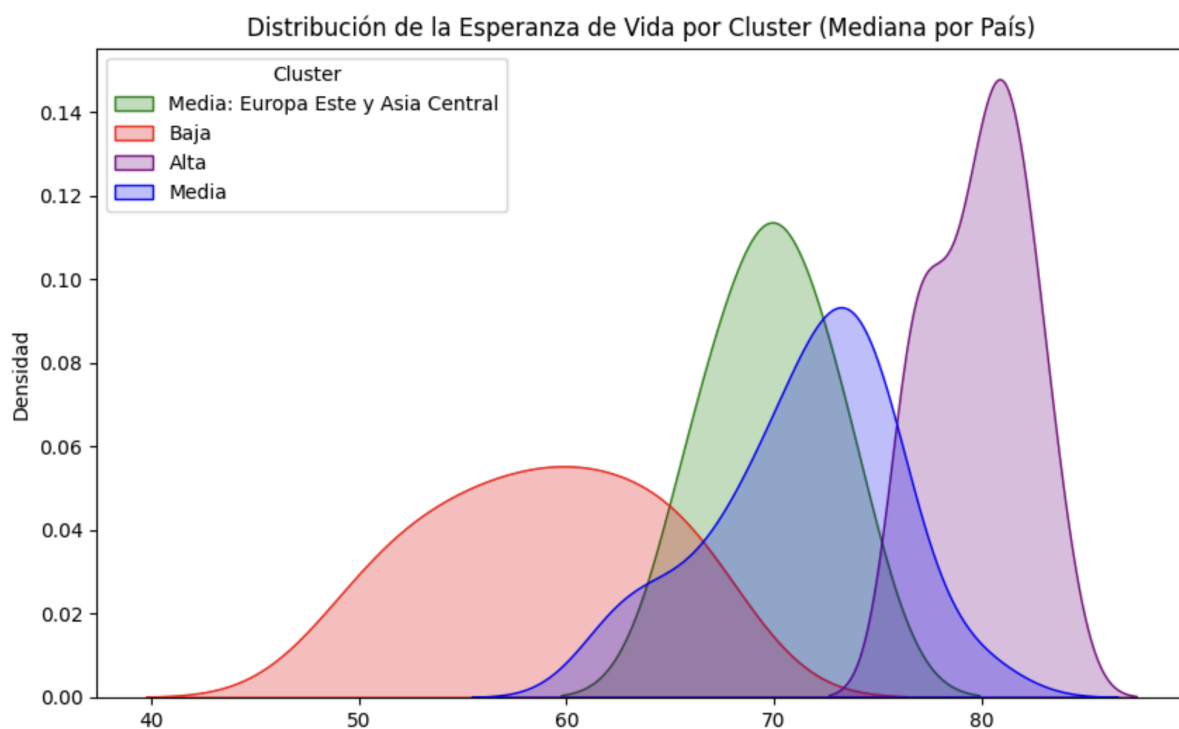


Gráfico 6: Densidad de la esperanza de vida en los distintos clusters. Fuente: elaboración propia.

La variabilidad de los datos entre países es más notable en el grupo rojo, tal y como muestra la curva más ancha. En el resto de casos, los países convergen más. Es decir, a medida que la esperanza de vida aumenta, la densidad de datos es mayor. Las diferencias entre los clusters de media esperanza de vida son menos notables, lo que sugiere que comparten unos niveles de longevidad similares pero características propias en lo que respecta al resto de variables.

Posteriormente, indagando más en cada uno de los grupos observamos cómo algunos países se encuentran en más de uno, ya que los datos tenían un factor temporal. Es decir, un total de 15 países⁸ ha sufrido una evolución en sus datos que ha provocado que el modelo los agrupara de manera distinta al final del período estudiado. Todos ellos muestran una mejora considerable en muchas de sus variables.

Al analizar los clusters, nos enfocamos en estudiar sus características principales en materia de economía, salud y determinadas causas de muerte o agrupaciones de estas. Para ello usamos los valores medianos del período, representados en *boxplots* y gráficos de barras, así como también el cálculo de las correlaciones. En el caso de los países con cambio de cluster, nos centramos en el cálculo de las diferencias entre el año 2019 y 2000, para medir su evolución de manera sencilla. Estos son algunos de los ejemplos gráficos:⁹

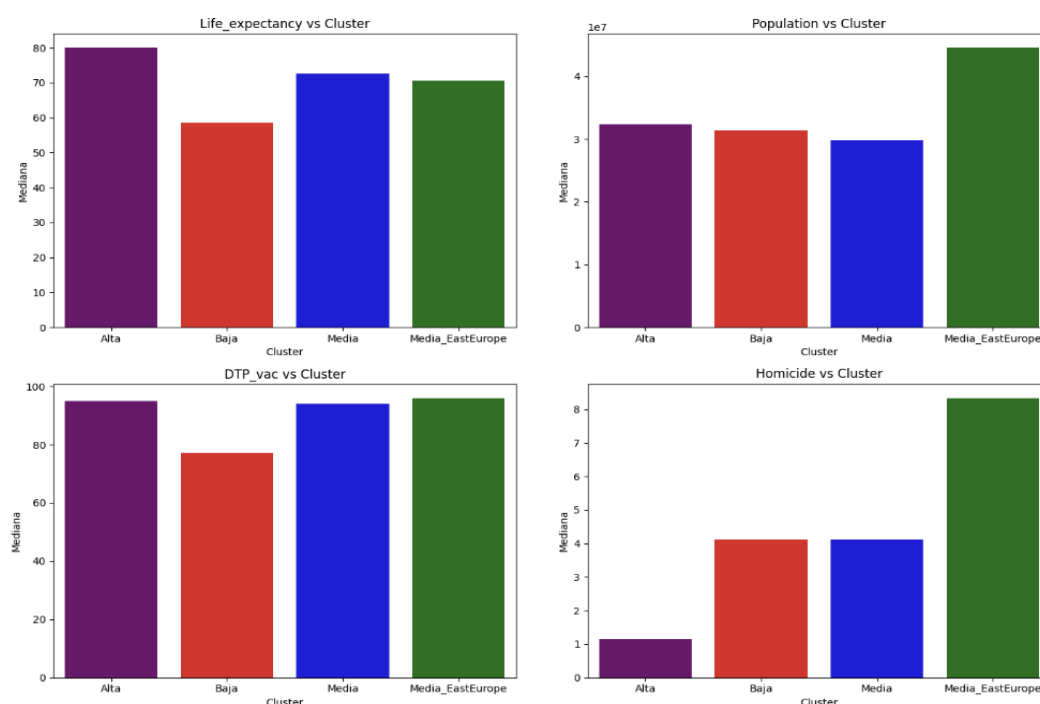


Gráfico 7: Valores medianos de distintas variables por cluster. Fuente: elaboración propia.

⁸ Estos países son: Bangladesh, Bolivia, Ghana, India, Indonesia, Kazakhstan, Kenya, Korea, Montenegro, Namibia, South Africa, Tanzania y Uganda.

⁹ Para ver todas las visualizaciones, consultar el apartado 6.3 del Notebook.

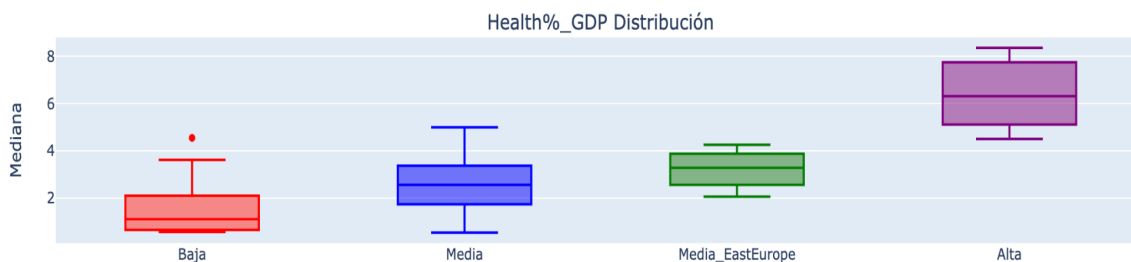


Gráfico 8: Distribución de los valores medianos del % de inversión en sanidad pública respecto al PIB. Fuente: elaboración propia.

2.2 Insights análisis de clusters

El análisis de los diferentes clusters de esperanza de vida revela patrones distintivos que varían según la región geográfica y el nivel de desarrollo socioeconómico de los países.

En el Cluster de Alta Esperanza de Vida (Alta), se observa que, si bien es necesario alcanzar ciertos niveles de riqueza para entrar en este grupo, un mayor PIB o inversión en sanidad no siempre se traduce en un incremento adicional de la esperanza de vida. Esta conclusión puede verse reforzada por el trabajo de Joseph Stiglitz, Amartya Sen y Jean-Paul Fitoussi en su libro *“Medir nuestras vidas: Las limitaciones del PIB como indicador del progreso”*, en el que plantean que el PIB se trata de una medida insuficiente del bienestar general y del desarrollo social y proponen un enfoque más amplio y multifacético que incluya indicadores de sostenibilidad, equidad, salud, educación, y bienestar subjetivo, entre otros. Ejemplos claros de esta tendencia son España y Japón, donde, pese a encontrarse en el grupo de países ricos no destacan por tener los niveles más altos de PIB, pero su esperanza de vida es superior a muchos otros¹⁰. Este cluster también se caracteriza por una alta prevalencia de enfermedades crónicas, como las cardiovasculares y el cáncer, que son típicas de poblaciones envejecidas. La mortalidad por enfermedades infecciosas es casi inexistente, reflejando un control efectivo de estas patologías. No obstante, se presentan desafíos importantes en el ámbito de la salud mental, como lo demuestran las altas tasas de suicidio en países como Japón y Corea del Sur. Además, la incidencia de muertes por accidentes es considerablemente baja, lo que podría estar vinculado a mejores normativas de seguridad y servicios de emergencia en estos países.

El Cluster de Media Esperanza de Vida (Media) es notable por su diversidad geográfica¹¹ y socioeconómica, lo que se traduce en una alta dispersión en variables como el PIB y la

¹⁰ Se puede observar la tabla comparativa de países de alta esperanza de vida en el apartado de anexos. Para más tablas comparativas, ver el apartado 6.3.2 del Notebook

¹¹ Muchos países que pertenecen a los grupos de media esperanza de vida son muy extensos y tienen una mayor cantidad de población, factores que aumentan las dificultades de gestión. En los anexos se puede observar un mapa que refleja esta realidad.

inversión en sanidad. De todos modos, estas variables presentan las correlaciones más altas y positivas del conjunto de grupos, indicando que su relación con la esperanza de vida es más estrecha que en otros clusters. Este grupo muestra una mayor dispersión en muertes causadas por conflictos y factores geográficos, siendo los datos de países latinoamericanos, como Colombia y El Salvador, los principales *outliers*. Las tasas de vacunación, aunque elevadas, presentan más variabilidad que en el cluster de alta esperanza de vida. La mortalidad por enfermedades crónicas sigue siendo significativa, aunque menos predominante en comparación con los países de alta esperanza de vida. A su vez, la incidencia de enfermedades infecciosas es más baja que en el cluster de baja esperanza de vida, pero sigue siendo relevante en algunos países.

En el Cluster de Media Esperanza de Vida (Europa del Este y Asia Central), se observa una alta prevalencia de enfermedades cardiovasculares, con países como Lituania y Ucrania registrando tasas muy elevadas. Este cluster presenta menor dispersión socioeconómica en comparación con otros clusters intermedios, reflejando una relativa homogeneidad en PIB e inversión en sanidad. La correlación de estas variables también es elevada y positiva, al igual que ocurría en el cluster anterior. Las enfermedades infecciosas y neonatales están mejor controladas en esta región, aunque los desafíos en cuanto a enfermedades crónicas y mentales son similares a los del cluster de alta esperanza de vida. Sin embargo, aquí destacan los niveles más altos de suicidios, autolesiones, alcoholismo y consumo de drogas, especialmente en países como Rusia, donde también se registran altas muertes por conflictos y factores geográficos. En resumen, este cluster parece estar muy definido por determinadas causas de muerte, pero a parte de esto, sus diferencias en esperanza de vida respecto al grupo mencionado anteriormente no son muy significativas.

El Cluster de Baja Esperanza de Vida (Baja) está marcado por altas tasas de mortalidad debido a enfermedades infecciosas y neonatales, particularmente en países africanos como Tanzania y Uganda. La malaria es una causa de muerte significativa en esta región, con tasas elevadas en países como Nigeria. Las enfermedades crónicas son menos comunes, en parte debido a la baja longevidad y la falta de diagnósticos precisos. Los valores del PIB per cápita son muy bajos en comparación con el resto. La inversión en salud es extremadamente desigual, pero esta dispersión no dista mucho de la observada en los clusters de esperanza de vida media. Esto puede dar indicios de que el dinero invertido en estas zonas no se traduzca en mejoras para la población. De hecho, su correlación con la esperanza de vida muestra un valor negativo¹². Los datos de este grupo, refuerzan las ideas compartidas en la obra de Michel Chossudovsky; "La globalización de la pobreza y el nuevo orden mundial", libro en el que expone cómo la globalización ha profundizado las

¹² Con esta conclusión no queremos indicar que invertir en sanidad sea algo negativo para estos países, sino poner énfasis en el hecho de que probablemente haya otros factores como la inestabilidad política o la corrupción, que no están permitiendo que se produzcan mejoras reales.

desigualdades, debilitado a los estados y empobrecido a grandes sectores de la población mundial. Chossudovsky llama la atención sobre la necesidad de repensar el modelo de desarrollo global, priorizando las necesidades humanas y la justicia social por encima de los intereses corporativos y financieros. Consideramos que, aunque los datos de pobreza no hayan sido introducidos en el modelo por la gran cantidad de valores nulos, se acaban reflejando en el resto de variables y se hacen notables en este cluster.

Por lo que respecta a los cambios entre grupos¹³, los países que han pasado del clúster de esperanza de vida media al alta, han experimentado un aumento moderado en la esperanza de vida y en el PIB per cápita. Corea del Sur ha destacado por su crecimiento en innovación tecnológica, mientras que el crecimiento en Montenegro se puede relacionar con el proceso de acercamiento a la UE. Además, la inversión en sanidad ha aumentado ligeramente. Aunque las tasas de vacunación no han incrementado, estos países ya contaban con sistemas sanitarios bien establecidos. No quiere decir que esta variable no sea importante, pero su efecto en la esperanza de vida no es el mismo que en el caso de países con una alta prevalencia de determinadas enfermedades. En cuanto a las causas de muerte, no se han observado grandes cambios, excepto en Corea, donde ha habido un aumento en las muertes relacionadas con enfermedades mentales, posiblemente ligado a un mayor número de suicidios y de casos de Alzheimer.

Por otro lado, los países que han pasado del clúster de esperanza de vida baja al de media, en su mayoría africanos y asiáticos, como Uganda, Tanzania e India, han experimentado un crecimiento poblacional significativo y un notable aumento en la esperanza de vida, con mejoras en las tasas de vacunación, especialmente en DTP¹⁴ y sarampión. El crecimiento del PIB ha sido dispar. En India, Bangladesh, Indonesia y Kazajistán ha sido muy notable, a causa de la apuesta por sectores manufactureros ligados a la deslocalización de fábricas en países más ricos. Pero en cambio, los países que han aumentado más su esperanza de vida no son los que más han crecido económicamente (como Namibia o Uganda). Las muertes neonatales y por enfermedades infecciosas han disminuido considerablemente, particularmente en Bangladesh e India, lo que parece ser un factor clave en el cambio de clúster.

¹³ Para observar estas diferencias, usar los mapas interactivos Plotly del apartado 6.3.3 del Notebook

¹⁴ En Uganda, el incremento de las vacunaciones de DTP a menores de 1 año ha sido de un 40% en 20 años.

4. MODELOS DE *MACHINE LEARNING*

Después de realizar el análisis previo, procedimos a entrenar diversos modelos de *machine learning* con el objetivo de predecir la esperanza de vida de diferentes países. Para este propósito:

- Utilizamos el conjunto de datos completo, con información de 122 países.
- Eliminamos las columnas con valores nulos y aquellas que agrupaban causas de muerte.
- Imputamos el resto de los valores nulos utilizando la mediana.

Iniciamos con modelos sencillos y aumentamos progresivamente su complejidad. El primer modelo fue de categorización (1 para esperanza de vida alta y 0 para los valores que estuvieran por debajo del umbral). Los modelos siguientes se enfocaron en predecir valores exactos de esperanza de vida en años. Finalmente, realizamos una comparativa entre los diferentes modelos para evaluar su rendimiento.

Para reducir el riesgo de *overfitting*, optamos por reducir las variables temporales, seleccionando solo años representativos: 2000, 2006, 2010, 2016 y 2019. Las razones para esta selección fueron:

1. **Reducción del ruido y complejidad:** Al utilizar solo ciertos años, evitamos que el modelo aprenda patrones específicos de cada año, lo que promueve una mejor generalización.
2. **Evitar el sobredimensionamiento:** Disminuir la cantidad de años evita un número innecesario de parámetros, ayudando a que el modelo se enfoque en las relaciones fundamentales entre las variables y la esperanza de vida.
3. **Evitar tendencias cíclicas de corto plazo:** Seleccionar todos los años podría inducir al modelo a detectar patrones cíclicos que no generalizan bien fuera del periodo de entrenamiento.

4.1 Regresión Logística

Uno de los modelos más utilizados para problemas de clasificación es la regresión logística. Se trata de un modelo que destaca por ser fácilmente interpretable. De manera general, la regresión logística es un modelo estadístico que estima la probabilidad de que un determinado evento ocurra, basándose en las variables independientes. El modelo utiliza

una transformación logística (sigmoide) que toma los valores de las variables independientes y los convierte en probabilidades entre 0 y 1. Posteriormente, asigna una categoría según un umbral predefinido, lo que permite clasificar los resultados. En nuestro caso, hemos convertido en categórica la esperanza de vida (Alta o Baja). Seleccionamos la mediana global de la esperanza de vida (72.65 años) como umbral porque este valor divide los países en dos grupos relativamente equitativos, asegurando que un grupo refleje países con esperanza de vida alta y el otro con esperanza de vida más baja. Este indicador proporciona un umbral representativo que minimiza el impacto de valores extremos, y refleja de manera más robusta la tendencia central de la distribución de la esperanza de vida entre los países. Por ello decidimos usarla como valor para categorizar la variable. Este tipo de modelo podría ser útil para decidir si aplicar o no ciertas políticas públicas.

Para evaluar el modelo, visualizamos e interpretamos la matriz de confusión y extrajimos las métricas más usadas: *accuracy*, *precision* y *recall*¹⁵. En general, el modelo consiguió buenos resultados, alcanzando una precisión del 93%. Predijo 77 verdaderos positivos y 93 falsos positivos.

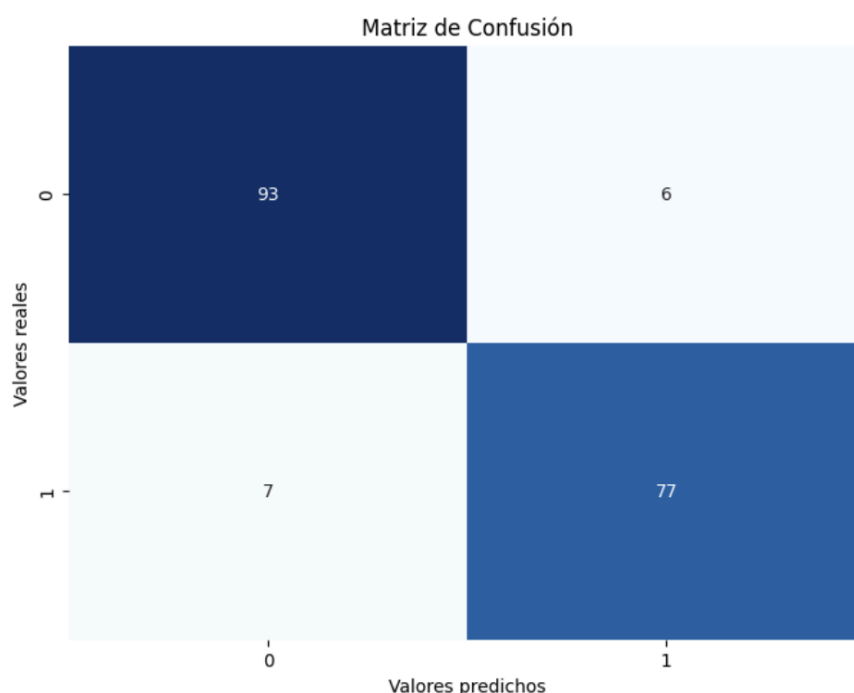


Gráfico 10: Matriz de confusión del modelo de regresión logística. Fuente: elaboración propia.

¹⁵ *Accuracy* (Precisión global): es el % de predicciones correctas sobre el total de predicciones realizadas. Mide qué tan bien el modelo clasifica en general. *Precision* (Precisión): Es la proporción de verdaderos positivos sobre el total de predicciones positivas. Indica cuántas de las predicciones positivas del modelo fueron correctas. *Recall* (Sensibilidad o Tasa de verdaderos positivos): Es la proporción de verdaderos positivos sobre el total de ejemplos reales positivos. Mide la capacidad del modelo para encontrar todos los casos positivos. Para observar los valores resultantes consultar el apartado 7.1.1 del Notebook.

4.2 Árboles de decisión

El primer modelo de regresión que entrenamos fue un árbol de decisión. Los árboles de decisión se usan tanto para modelos de regresión como de clasificación. Su funcionamiento imita el proceso de tomar decisiones, dividiendo los datos en subconjuntos más pequeños de manera jerárquica en base a preguntas simples sobre las características de los datos. El árbol está compuesto por nodos que representan decisiones basadas en características de los datos. El proceso comienza desde un nodo raíz y luego se bifurca en ramas, donde cada bifurcación corresponde a una decisión tomada en función de un umbral de una característica. Al final de las bifurcaciones, se encuentran hojas que contienen las predicciones. La evaluación de este modelo y los siguientes la hicimos a través de las métricas RMSE (*Root Mean Squared Error*), MAE (*Mean Absolute Error*) y R2 (*Coefficient of Determination*)¹⁶, que podremos ver en una tabla comparativa después de presentarlos todos.

4.3 Random forest

Para el modelo de complejidad media-alta, decidimos usar el *Random Forest*. Se trata de un conjunto de árboles de decisión combinados para mejorar la precisión del modelo y controlar el posible sobreajuste. A grandes rasgos, se trata de un conjunto de múltiples árboles que se entrenan con distintas muestras aleatorias. Cuando se construye cada árbol se emplea un conjunto aleatorio de variables para crear diversidad entre ellos y mejorar la generalización. Por último, predice el valor medio de las predicciones de todos los árboles. Una vez entrenado el modelo, no sólo usamos las métricas mencionadas anteriormente, sino que también observamos la importancia que este le había dado a cada variable¹⁷, siendo la causa de muerte por meningitis la que ocupó el primer lugar, seguida por las muertes neonatales y el PIB per cápita.

¹⁶ RMSE: Mide la raíz cuadrada del error cuadrático medio, conforme más bajo es este indicador, las predicciones de este modelo están más cerca de los valores reales. MAE: error medio entre las predicciones y los valores reales. Su valor muestra la media de los errores en el conjunto de predicciones. R2: Mide la proporción de la varianza de la variable dependiente que es explicada por el modelo. Las cifras varían entre 0 y 1 siendo 1 un indicador de que el modelo explica perfectamente los datos.

¹⁷ Los gráficos de importancia de las características se encuentran en el apartado de Anexos.

4.4 XGBoost

A grandes rasgos, el *XGBoost* es un algoritmo de *machine learning* basado en árboles de decisión que optimiza el rendimiento a través de una técnica llamada boosting, la cual mejora los errores gradualmente al entrenar varios modelos secuenciales. Es uno de los mejores modelos para predecir la esperanza de vida ya que permite manejar grandes conjuntos de datos con gran diversidad, proporciona un excelente rendimiento en muchos problemas de regresión y es óptimo a nivel de velocidad y uso de memoria. Pero, ¿cómo funciona? El modelo comienza entrenando un árbol de decisión básico. Luego, construye otros árboles que intentan corregir los errores del primero, ajustando continuamente las predicciones mediante el uso del gradiente. Este proceso se repite hasta que los errores se minimizan. Además, es conocido por su eficiencia y precisión, ya que incluye técnicas de regularización que evitan el sobreajuste y permiten que el modelo generalice mejor en datos nuevos. La importancia que el modelo asignó a las variables fue distinta a la del *Random Forest*. En este caso, la variable PIB per cápita apareció como la más relevante, lo que valida la hipótesis de que el crecimiento económico puede tener un impacto en la mejora de la calidad de vida y el acceso a servicios de salud (aunque a partir de cierto nivel tenga sus limitaciones). Además, el modelo también dió importancia a las muertes por enfermedades como el VIH/SIDA y la meningitis, junto con el gasto en salud pública. En general, mostró una versión más balanceada que el *Random Forest*, dando mayor importancia a distintas causas de muerte.

4.5 Comparativa y uso de los modelos

	RMSE	MAE	R2
Árbol de decisión	2.36	1.74	0.94
Random Forest	1.68	1.31	0.97
XGBoost	1.53	1.18	0.98

Tabla 1: Comparativa de métricas de evaluación de los modelos de regresión. Fuente: elaboración propia.

Como se observa en la tabla, el modelo que presenta unos mejores resultados es el *XGBoost*, consiguiendo los niveles más bajos de errores medios y un coeficiente de determinación de 0.98. Por lo tanto, es el que debería ofrecer predicciones más precisas.

Para comprobarlo, creamos una función a la que le dimos un año y listado de países para los que queríamos hacer la predicción, para obtener los resultados por modelos. Para este

ejemplo, usamos países de todas las regiones, con distintos niveles de esperanza de vida y características. Las predicciones para el año 2014 fueron las siguientes:

País	Árbol decisión	Random Forest	XGBoost	R. log	Valor real
Japan	82.38	83.51	83.74	1	83.65
Spain	82.38	82.12	82.82	1	82.86
Norway	78.85	80.86	81.52	1	82.09
United States	79.14	78.84	79.28	1	79.02
Uruguay	76.86	76.92	77.11	1	77.37
Mexico	74.19	74.33	74.52	1	74.80
Russian Federation	75.53	72.53	71.11	1	71.57
Egypt	76.75	71.33	71.48	0	70.42
India	68.43	67.42	67.52	0	69.07
Bolivia, Plurinational State of	67.63	68.33	67.68	0	67.16
Haiti	63.39	62.53	62.69	0	62.99
Uganda	62.35	61.25	60.83	0	60.41
Zimbabwe	60.31	57.56	59.14	0	58.85

Tabla 2: Comparativa de las predicciones de la esperanza de vida para varios países en el año 2014. Fuente: elaboración propia.

Este ejemplo confirma los resultados de las métricas anteriores, en las que el modelo *XGBoost* destacaba por ser el más preciso, con predicciones más cercanas al valor real en la mayoría de los casos. Los países con menor esperanza de vida, como Zimbabwe, Uganda o Bolivia, y aquellos con contextos más conflictivos o desafiantes, como Egipto tras la Primavera Árabe o India con su elevada población, muestran mayor variabilidad entre los modelos, sugiriendo la influencia de factores externos no capturados adecuadamente. En contraste, los países con mayor esperanza de vida, como Japón y España, presentan predicciones más consistentes entre los modelos, lo que indica que los algoritmos capturaron mejor los factores que influyen en estos países. Aunque la regresión logística predice bien en casi todos los casos (salvo en Rusia, debido a su umbral de 62.65 años), consideramos que la complejidad aportada por otros modelos es más relevante para hacer predicciones generales, y categorizarlas puede tener sentido solo en casos muy específicos.

5. CONCLUSIONES FINALES

La esperanza de vida es un indicador multifacético que refleja la interacción de factores económicos, sanitarios y sociales. A lo largo de este análisis, hemos identificado variables clave como el PIB per cápita, el gasto en sanidad, las tasas de vacunación y la prevalencia de enfermedades crónicas como determinantes cruciales. Sin embargo, estos factores no operan de manera aislada. Su influencia varía considerablemente según el contexto socioeconómico y geopolítico de cada región, lo que subraya la necesidad de adoptar un enfoque contextualizado en las políticas de salud pública.

En los continentes con altos niveles de desarrollo económico, como Europa y América del Norte, la relación positiva entre el PIB per cápita y la esperanza de vida es evidente. Países como Noruega y Luxemburgo, con ingresos elevados, disfrutan de una esperanza de vida superior a los 80 años. Sin embargo, a medida que se alcanza un cierto nivel de desarrollo, el impacto adicional del PIB o del gasto en sanidad sobre la longevidad parece ser limitado. Japón, por ejemplo, aunque no tiene el PIB per cápita más alto, presenta una de las mayores esperanzas de vida a nivel mundial, superando a economías más grandes. De manera similar, España muestra que un mayor PIB no siempre se traduce en un incremento proporcional de la longevidad. Además, si observamos la evolución de países que han cambiado del cluster de baja esperanza de vida al de media, vemos cómo los que han aumentado en mayor medida este indicador no han sido los que han incrementado más su PIB. Estos casos revelan las limitaciones de los indicadores económicos tradicionales como medidas de éxito y sugieren que otros factores, como las costumbres culturales, el estilo de vida y el entorno social, pueden ser igualmente significativos en estados que ya disfrutaban de garantías sanitarias y de ciertos niveles de riqueza.

El gasto en sanidad emerge como un factor determinante, pero su impacto depende en gran medida de la eficiencia con la que se gestionen los recursos. Evidencia de ello son las correlaciones tan dispares de esta variable con la esperanza de vida para las distintas agrupaciones de países. En América Latina, Uruguay o Chile han logrado mejoras notables en la esperanza de vida, superando los 77 años, gracias a sistemas de salud eficaces. En contraste, Bolivia, con un gasto sanitario inferior pero no muy distante¹⁸, muestra una esperanza de vida significativamente menor, lo que evidencia la necesidad de optimizar la administración de los recursos sanitarios. Este contraste es aún más pronunciado en Estados Unidos¹⁹, donde una alta inversión en sanidad no se traduce en una esperanza de

¹⁸ El % de inversión para el período estudiado de Uruguay, Chile y Bolivia ha sido de 4,82, 3,19 y 3,04 respectivamente.

¹⁹ La privatización del sistema, junto con la crisis de los opiáceos que ha ido en aumento en los últimos años son factores que no se reflejan en el % de gasto en sanidad pública de Estados Unidos (Para más información, ver recursos audio y web de El Orden Mundial).

vida comparable a la de otros países desarrollados. Las ineficiencias y la falta de cobertura sanitaria universal para amplios sectores de la población, en gran parte debido a la privatización del sistema, parecen ser factores determinantes. Esto subraya la necesidad de un enfoque más integral que garantice no solo la cantidad, sino también la equidad y eficiencia en el acceso a los servicios de salud.

Las tasas de vacunación a menores de un año resultan ser un aspecto clave, especialmente en regiones con altos índices de mortalidad infantil, como África subsahariana y Asia del Sur. En Uganda, por ejemplo, el incremento en las tasas de vacunación infantil ha contribuido directamente a la reducción de la mortalidad por enfermedades infecciosas, derivando en un aumento de 15 años en la esperanza de vida desde el año 2000. Sin embargo, en otras regiones como la zona del Amazonas en América Latina, las tasas de vacunación han sido más inconsistentes, factor que podría estar dificultando el manejo total de algunas enfermedades que ya son inexistentes en partes más desarrolladas. Estos ejemplos destacan la importancia de campañas de vacunación sostenibles y bien organizadas para mejorar la esperanza de vida en las áreas más vulnerables. Con esto no queremos decir que esta variable no sea fundamental en los países que gozan de unos niveles de esperanza de vida más altos, pero es evidente que las prioridades o preocupaciones de estas naciones en este aspecto son muy diferentes a los de los países más pobres y con una estructura sanitaria más débil.

A medida que las sociedades envejecen, como es el caso en Europa y América del Norte, las enfermedades crónicas, como las cardiovasculares y el cáncer, se han convertido en las principales causas de muerte. En países como Alemania y el Reino Unido, más del 30% de las muertes están relacionadas con enfermedades cardiovasculares. Este cambio en el perfil epidemiológico resalta la necesidad de adaptar los sistemas de salud para gestionar a largo plazo las enfermedades crónicas, en lugar de enfocarse solo en el tratamiento de enfermedades agudas. La prevención se perfila como una estrategia crucial para afrontar estos desafíos, lo que permitirá mitigar el impacto del envejecimiento poblacional si se promueve la calidad de vida en la vejez.

Aunque la falta de datos específicos sobre pobreza y desigualdad no nos ha permitido un análisis más exhaustivo, estas variables no deben ser ignoradas. En este sentido, nos apoyamos en el trabajo de Esther Duflo y Abhijit Banerjee, quienes subrayan que la pobreza y la desigualdad impactan negativamente la salud de millones de personas. En Nigeria, donde más del 40% de la población vive por debajo de la línea de pobreza, la esperanza de vida es de apenas 54 años en 2019. En contraste, los países con políticas inclusivas y sistemas de salud robustos disfrutan de expectativas de vida mucho más elevadas. Esto confirma la importancia de reducir las desigualdades económicas y sociales para mejorar la salud pública. Un enfoque contextualizado, que tenga en cuenta las realidades locales y no

trate la pobreza como un fenómeno endémico, es clave para lograr mejoras sostenibles en la esperanza de vida.

El uso de modelos de machine learning como el XGBoost ha reforzado estas conclusiones. La importancia de las variables económicas, reflejadas en el gráfico de características del modelo, confirma su relación con la esperanza de vida, mientras que la relevancia de factores como el VIH/SIDA, la mortalidad neonatal y las enfermedades crónicas destaca el papel central de las políticas de salud pública. De todos modos, comparando los modelos hemos visto cómo la predicción de la esperanza de vida es compleja, y los resultados pueden ser bastante distintos cuanto más desafiante sea el contexto del país.

Por último, los conflictos y desastres naturales pueden tener un impacto devastador en la esperanza de vida, especialmente en países con infraestructuras frágiles. El terremoto de Haití en 2010, que resultó en más de 200,000 muertes, redujo drásticamente la esperanza de vida del país. Este evento pone de relieve cómo las externalidades, agravadas por la falta de recursos, pueden alterar significativamente la longevidad. De manera similar, en países como Nigeria y Camerún, los conflictos internos y el terrorismo han disminuido la esperanza de vida, evidenciando la fragilidad de los sistemas de salud en zonas de inestabilidad.

5. 1 Recomendaciones para la *World Health Organization*

Después de revisar todas las conclusiones, estas son nuestras recomendaciones específicas para la WHO:

Para los países en desarrollo y más pobres, la prioridad debe ser fortalecer los sistemas de salud y garantizar el acceso universal a la atención médica. Esto implica implementar sistemas de atención primaria de calidad que proporcionen prevención, diagnóstico temprano y tratamiento accesible para todos los ciudadanos, especialmente en las zonas rurales y para los grupos vulnerables. Además, es esencial expandir la cobertura de salud para asegurar que los servicios médicos, medicamentos y vacunas sean accesibles sin barreras económicas, lo que ayudará a reducir las tasas de mortalidad y a mejorar la esperanza de vida. Al mismo tiempo, es necesario invertir en infraestructura sanitaria para garantizar que los sistemas de salud tengan la capacidad de responder adecuadamente, incluso en áreas desfavorecidas.

Además, es crucial reducir las desigualdades sociales y mejorar las condiciones de vida de la población, invirtiendo en educación y desarrollo comunitario para empoderar a las personas con el conocimiento y los recursos para mejorar su salud. Como señalan Duflo y Banerjee en *Repensar la pobreza*, estas intervenciones deben adaptarse al contexto local y

estar respaldadas por instituciones sólidas, de manera que su impacto sea sostenible. Asimismo, fortalecer los sistemas de seguridad social y mejorar el entorno urbano y rural son pasos fundamentales para crear condiciones más saludables y equitativas, lo que permitirá mejorar el bienestar general y la longevidad de la población.

Para los países más ricos, donde la esperanza de vida es alta y los cambios en las dinámicas sociales son cada vez mayores²⁰, el enfoque debe centrarse en prevenir enfermedades crónicas y mejorar la calidad de vida, particularmente en la vejez. En este contexto, la promoción de estilos de vida saludables es clave, no solo para prevenir enfermedades físicas, sino también para mejorar la salud mental. Estudios recientes han demostrado que el ejercicio regular tiene un efecto comparable al de algunos medicamentos en la reducción de los síntomas de la depresión leve a moderada y el manejo de la ansiedad (Pearce, M., et al. (2022)). Estos hallazgos subrayan la importancia de fomentar el ejercicio no solo para prevenir enfermedades físicas, sino también para prevenir trastornos psicológicos. Además, el ejercicio regular también juega un papel importante en la prevención del declive cognitivo. El trabajo de Schlegel, P., et al. (2019), sugirió que la actividad física puede reducir el riesgo de desarrollar demencia y mejorar la función cognitiva en personas mayores, mientras que una revisión de 2020 en la revista *Neuropsychobiology* concluyó que este supuesto parece ser cierto. Al orientar las inversiones en sanidad hacia la prevención de enfermedades y la promoción de una vida activa y saludable, la OMS puede contribuir a mejorar no solo la esperanza de vida, sino también la calidad de vida de las personas mayores.

5.2 Link al Dashboard

[Dashboard](#)

²⁰ Estos cambios se exponen en "Una sociedad de hijos" de Manuel Castells. El autor ofrece una visión profunda de cómo las transformaciones en la estructura familiar y los cambios en las dinámicas sociales están afectando a las familias en la sociedad contemporánea. El libro explora la influencia de la modernidad líquida, las desigualdades en el cuidado infantil, y el impacto de la tecnología y la economía en la vida familiar

6. BIBLIOGRAFÍA

Graunt, J. (1662). *Natural and Political Observations Made upon the Bills of Mortality*. London: Roycroft.

Halley, E. (1693). An Estimate of the Degrees of the Mortality of Mankind, Drawn from Curious Tables of the Births and Funerals at the City of Breslaw. *Philosophical Transactions of the Royal Society of London*, 17(196), 596-610.

Farr, W. (1852). *Report on the Mortality of Cholera in England, 1848-49*. London: W. Clowes & Sons.

Gompertz, B. (1825). On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London*, 115, 513-585.

Lotka, A. J. (1939). *Analytical Theory of Biological Populations*. New York: Plenum Press.

Preston, S. H. (1975). The Changing Relation between Mortality and Level of Economic Development. *Population Studies*, 29(2), 231-248.

Livi Bacci, M. (1990). *Historia mínima de la población mundial*. Editorial Crítica.

Monsef, A. et. al (2015). *Determinants of life expectancy: a panel data approach*. Asian Economic and Financial Review. 5(11): 1251-1257.

Stiglitz, J., Sen, A., & Fitoussi, J.-P. (2010). *Medir nuestras vidas: Las limitaciones del PIB como indicador del progreso*. Taurus.

Chossudovsky, M. (1997). *La globalización de la pobreza y el nuevo orden mundial*. Ediciones Ecoamérica.

Duflo, E., & Banerjee, A. (2012). *Repensar la pobreza: Un giro radical en la lucha contra la desigualdad global*. Editorial Taurus.

Pearce, M., et al. (2022). Association between physical activity and risk of depression: A systematic review and meta-analysis. *JAMA Psychiatry*, 79(6), 550-559.

López-Torres Hidalgo, J., et al. (2019). Effectiveness of physical exercise in the treatment of depression in older adults as an alternative to antidepressant drugs in primary care. *BMC Psychiatry*, 19(1), 21.

Schlegel, P., et al. (2019). "Muscle-gut-brain axis": Can physical activity help patients with Alzheimer's disease due to microbiome modulation? *Journal of Alzheimer's Disease*, 71(3), 861-878.

Recursos audio y web:

Esteban Ortiz-Ospina (2017) - *"Life Expectancy" – What does this actually mean?"*
Publicado online en OurWorldInData.org. Más información:
'<https://ourworldindata.org/life-expectancy-how-is-it-calculated-and-how-should-it-be-interpreted>' [Recurso Online]

Hernandez, C(2023) - *"Así han aumentado las muertes por sobredosis en Estados Unidos: la epidemia de fentanilo."* Publicado online en: El Orden Mundial.com Mas información:
'<https://elordenmundial.com/mapas-y-graficos/asi-han-aumentado-muertes-sobredosis-estados-unidos-epidemia-fentanilo/>' [Recurso Online]

No es el fin del mundo, por el Orden Mundial (2024). *El problema de la sanidad en Estados Unidos*. Spotify. <https://open.spotify.com/show/4uP7uqC9AGzwZfJp1LLmAC>
[Audio podcast]

No es el fin del mundo, por el Orden Mundial (2024). *El problema del envejecimiento*. Spotify.
<https://open.spotify.com/episode/6VOduwGtVsDf8ggKgs9OPw?si=2c57e52d24fc41c2>
[Audio Podcast]

7. ANEXOS

Tabla 3: *check* inicial de los datos.

	Causas de muerte	Suicidios	PIB per cápita	Vacunación	Homicidios	Esperanza de vida	Indicadores de pobreza	Gasto público en sanidad
Shape	6840, 34	3880,4	10108,4	7897, 14	4204, 4	20755,4	2602, 109	4014, 4
Países	205	184	182	196	209	238	192	191
Período	1990-2019 (30)	2000-2019 (20)	1950-2019 (70, pero sólo de 51)	1980-2021(42)	1990-2021 (32), aunque no para todos los países	1950-2021 (Para años anteriores hay datos, pero pocos) Min: 1543	60 (1963-2022) Aunque ningún país tiene datos para todos.	20 (2000-2019) Hay 20 datos para la mayoría de países. Sólo 10 disponen de menos años.
Desc. Columnas	Distintos tipos de muerte y su cuenta	Muertes por cada 100.000 habitantes	PIB per cápita	% de menores vacunados para 11 enfermedades	Homicidios por cada 100.000 habitantes	Esperanza de vida	Indicadores de pobreza varios (107 columnas)	Gasto público en sanidad en % del PIB
Valores nulos	Sólo en 'Code'	Sólo en 'Code'	-	Gran cantidad de nulos (> 80% en algunas)	Sólo en 'Code'	Sólo en 'Code'	Sí, en múltiples columnas. (> 40% en algunas)	Sólo en 'Code'

Fuente: elaboración propia

Tabla 4: Países de estudio

África				
Región	Tamaño	País 1	País 2	País 3
North Africa	Medium	Algeria	Egypt	Morocco
West Africa	Large	Ghana	Nigeria	Sierra Leone
Central Africa	Small	Cameroon		
East Africa	Large	Kenya	Tanzania, United Republic of	Uganda
Southern Africa	Large	South Africa	Zimbabwe	Namibia

Asia				
Región	Tamaño	País 1	País 2	País 3
East Asia	Medium	China	Japan	Korea, Republic of
South Asia	Medium	India	Pakistan	Bangladesh
Southeast Asia	Large	Indonesia	Thailand	Viet Nam
Central Asia	Small	Kazakhstan	Uzbekistan	Turkmenistan
West Asia	Large	Iran, Islamic Republic of	Armenia	Türkiye

América del Norte				
Región	Tamaño	País 1	País 2	País 3
North America	Medium	Canada	United States	Mexico
Central America	Medium	El Salvador	Panama	Honduras
Caribbean	Small	Haiti	Jamaica	Trinidad and Tobago

América del Sur				
Región	Tamaño	País 1	País 2	País 3
Southern Cone	Medium	Chile		
Andean Region	Medium	Peru	Bolivia, Plurinational State of	Ecuador
Amazon Region	Medium	Brazil	Colombia	
Atlantic Region	Small	Venezuela, Bolivarian Republic of	Paraguay	Uruguay

Europa				
Región	Tamaño	País 1	País 2	País 3
Western Europe	Medium	United Kingdom	Germany	Luxembourg
Eastern Europe	Large	Poland	Ukraine	Russian Federation
Northern Europe	Medium	Norway	Lithuania	Denmark
Southern Europe	Large	Spain	Croatia	Montenegro

Oceanía				
Región	Tamaño	País 1	País 2	País 3
Australasia	Small	Australia		
Melanesia	Small	Fiji		
Micronesia	Small			
Polynesia	Small			

Fuente: elaboración propia

Tabla 5: Datos para los países del cluster “Alta” ordenados por esperanza de vida

Country	Life_expectancy	Population	GDP_pc	Health%_GDP	DTP_vac
Japan	82.8	127586820.0	39409.61	7.35	97.0
Korea, Republic of	82.47	50776220.0	37245.48	3.76	98.0
Australia	81.91	21840032.0	48166.91	6.04	92.0
Spain	83.59	46399558.0	34432.67	6.32	96.5
Canada	81.21	33778668.0	47053.13	7.27	91.0
Norway	80.95	4859471.0	75764.23	7.5	94.0
Luxembourg	80.59	502475.5	66694.7	5.24	99.0
United Kingdom	80.28	62501717.0	39574.59	7.84	93.0
Germany	80.01	81478995.0	45850.93	8.16	94.0
Denmark	79.1	5538652.5	47677.08	8.36	94.0
United States	78.67	309847440.0	54983.22	7.89	95.0
Uruguay	76.8	3348407.0	16593.28	4.82	94.5
Montenegro	76.68	632123.0	17391.38	4.88	91.5
Croatia	76.6	4375411.0	21593.8	5.78	95.5
Poland	76.05	38565600.0	22383.44	4.43	99.0

Fuente: elaboración propia

Imagen 1: Mapamundi con ajuste al tamaño real de los países.



Fuente: @Neilkaye

Gráfico 11: ancia de las características del modelo *Random Forest*:

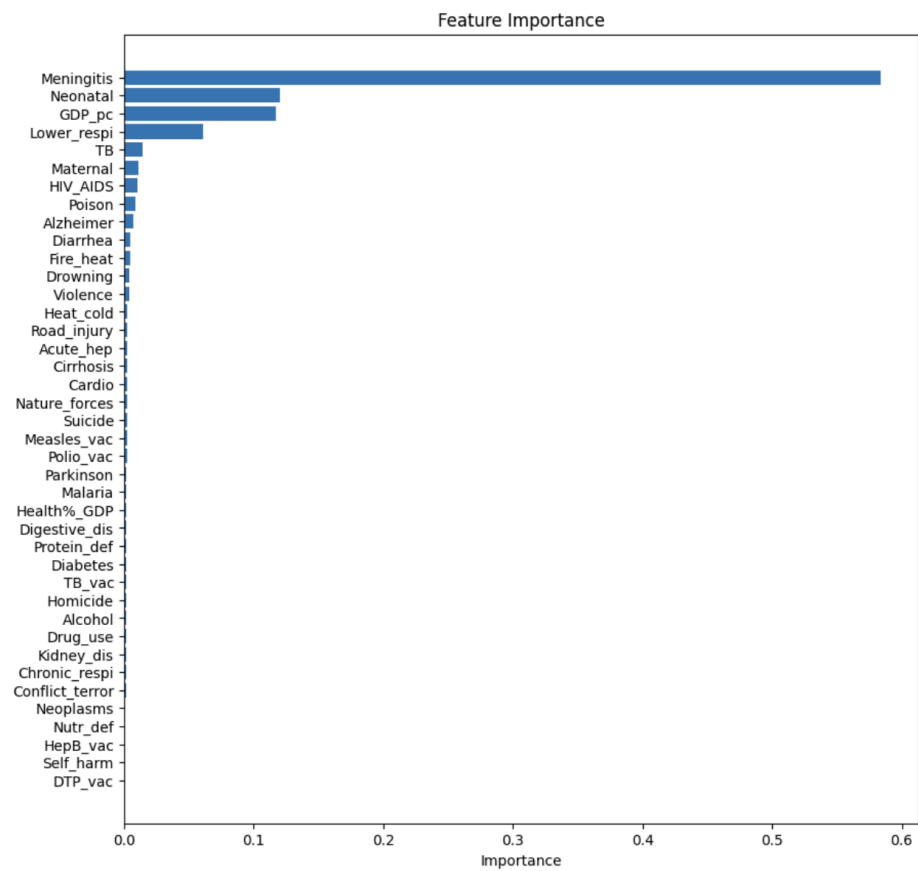


Gráfico 12: Importancia de las características del modelo *XGBoost*:

