



Calidad de Vida y Gasto en Salud en Europa

Proyecto Final de Visualización de datos

Diego Huerta Ocaña

170212

2 de marzo del 2024

Índice

Introduccion	3
Contexto del problema	3
Delimitación del problema	3
Objetivo	4
Metodologia	4
Datos.....	4
Herramientas de ciencia de datos	4
Herramientas de visualización	5
Resultados	5
Conclusión y recomendaciones.....	8
Síntesis del proyecto.....	14
Recomendaciones basadas en los resultados.....	14
Limites del proyecto y direcciones futura	15
Referencias.....	15

Introduccion

Contexto del problema

La calidad de vida ha sido un objetivo de estudio importante para la sociedad a lo largo de la historia. Este concepto manifiesta la capacidad que tiene una sociedad para prolongar la vida de sus ciudadanos, también refleja la efectividad de políticas sociales en la salud, la calidad del ambiente, la eficacia de los servicios ofrecidos por la sociedad, y otros factores que impacten directa o indirectamente en la salud o el bienestar de la población.

La calidad de vida es un concepto cualitativo que considera diversos aspectos. Por lo tanto, es un desafío definirlo concretamente, así como poder medirlo en una sociedad. Debido a esto, a menudo se utilizan conceptos relacionados para estudiar la calidad de vida.

Una de las métricas más utilizadas para medir la calidad de vida es la esperanza de vida, esta es una medida cuantitativa de la salud de los individuos de una sociedad. Formalmente, la esperanza de vida se define como el número medio de años que se espera que un individuo viva, al momento que nace. Para calcular esta métrica se hace uso de análisis estadísticos de mortalidad y otros factores demográficos.

Determinar los factores influyen o contribuyen al aumento de la esperanza de vida es una tarea desafiante debido a la complejidad de las interacciones entre diferentes variables. Usualmente se asume que variables como gasto en salud mejoran directamente la esperanza de vida. Es decir, los gobiernos invierten en infraestructura médica pública para así poder tener mejor calidad de vida en sus ciudadanos. Sin embargo, existiendo tantos factores que influyen en la esperanza de vida, intentar solucionar el problema de mala esperanza de vida únicamente invirtiendo en infraestructura medica podría ser una solución ingenua. Además, se necesita un entendimiento profundo de la distribución y causas de la esperanza de vida.

Es de interés identificar y determinar factores que impulsan la esperanza de vida. De este modo, distintos gobiernos alrededor del mundo podrán poner más énfasis en esos aspectos, para así mejorar la esperanza de vida, y por tanto calidad de vida, de sus ciudadanos. Contribuyendo a una mejor sociedad.

Delimitación del problema

Debido a que la historia y localización de los países puede afectar en su esperanza de vida, es conveniente limitar el análisis a una región geográfica específica. Para el estudio únicamente se consideran países de Europa. Esto se hace para delimitar el análisis a una región concreta, con un conjunto razonable de países que comparten características geográficas similares. Para analizar países que tengan información disponible de fácil acceso, se analizan 46 países de Europa. Sin embargo, algunas de las visualizaciones consideran un número ligeramente menor de países, cuando no todos los países tengan información disponible. Por lo tanto, no todas las visualizaciones consideran los mismos países para el análisis.

Con el propósito de obtener conclusiones contemporáneas, únicamente se analizan datos del siglo XXI. Además, debido a que los últimos años no tienen aún información tan completa, se delimita el análisis temporal para solo considerar años entre el 2000 y 2021.

Es imposible considerar todos los factores que pueden estar correlacionados con la esperanza de vida. Por lo tanto, se delimita el estudio para considerar únicamente variables geográficas y temporales, gasto en salud per cápita, gasto externo en salud per cápita, gasto de los países en sectores de hospitales, farmacias, clínicas, dentistas y doctores. Únicamente para la última visualización se considera un número mayor de variables de salud en los países.

Objetivo

Se busca analizar el fenómeno de la esperanza de vida para los países de Europa. Describir su distribución e identificar potenciales variables influyentes para este indicador. Se buscan describir algunos factores correlacionados con la esperanza de vida para estos países. De este modo, se busca tener mejor conocimiento sobre cuáles son los agentes que parecen impulsar la esperanza de vida en los países de Europa.

Metodología

Datos

Las fuentes de datos usadas para la obtención de información son:

- Country Comparisons – Life Expectancy at Birth (CIA)
- Global Health Expenditure Database (OMS)
- Global Healthsites Mapping Project (open source)
- The World Bank Group
- Natural Earth Data

Los datos históricos de esperanza de vida de los países, gasto en salud per cápita y gasto externo en salud per cápita se obtienen de “The World Bank Group”. Para tener datos más completos actuales sobre la esperanza de vida de todos los países se usa la base de datos de “Country Comparisons”. La información sobre los gastos de cada país en diversos factores (hospitales, clínicas, farmacias, etc.) se obtiene de “Global Healthsites Mapping Project”. Usando “Global Health Expenditure Database” se obtiene información general de cada país en aspectos de salud. Finalmente, los datos proporcionados por “Natural Earth” otorgan información sobre la población y localización de los países. De este modo, se obtienen 7 bases de datos usando estas fuentes. El análisis y la visualización se limitan a considerar solo estos datos.

Recolección y limpieza de datos

El manejo de datos de diversas fuentes y diversos formatos requiere de un manejo muy minucioso de la información. Esto se lleva a cabo usando el lenguaje de programación python, el cual cuenta

con librerías especializadas para el manejo y unión de conjuntos de datos, tal como pandas y numpy.

La información de varios países se unifica de las distintas fuentes de datos. Para unir los datos se usa el nombre de los países como su identificador. Antes de hacer la unión de los datos, se tiene especial cuidado con resolución de entidades, pues un mismo país puede estar representado con diferentes nombres. Las variables y países con gran número de valores nulos no se consideran para el análisis. Cuando se tienen pocos valores faltantes, se usan técnicas de especializadas para llenarlos. El código del análisis y los datos se pueden encontrar en el siguiente repositorio: <https://github.com/DiegoHuerta1/Calidad-de-Vida-y-Gasto-en-Salud-en-Europa>.

Herramientas de ciencia de datos

Para el análisis de los datos se hace uso mayormente de estadística descriptiva y de inferencia estadística, se usan técnicas tales como Kernel Density Estimation para la estimación de funciones de densidad. Para identificar que países que tienen alta o baja esperanza de vida se utiliza K-means, que es una técnica de agrupamiento de datos. Para poder visualizar datos con varias variables de salud, se hace uso de PCA, la cual es una técnica de reducción de dimensión.

Finalmente, para determinar la eficacia de variables para explicar la esperanza de vida de los países se hace uso de regresión logística. Específicamente, se hace uso de la prueba de hipótesis basada en deviance de la regresión logística para validar el modelo.

Herramientas de visualización

Usando el lenguaje python de programación, se hace uso de las librerías matplotlib, seaborn y plotly para poder llevar a cabo las visualizaciones. Usando estas librerías especializadas, se realizan gráficas de barras, series de tiempo, gráficas de dispersión, entre otras. Cada una con las personalizaciones necesarias para poder obtener la mayor cantidad de información de las gráficas. Usando estas herramientas, se buscan crear visualizaciones, tanto estáticas como dinámicas. Para las gráficas dinámicas se hace uso de la librería plotly.

Al estar trabajando con datos de países, se aprovecha la información geoespacial para crear visualizaciones en formato de mapa. Esto se lleva a cabo con librerías especializadas, tal como es geopandas.

Resultados

Se observa la esperanza de vida de los distintos países de Europa y se visualiza la distribución. Esta visualización se muestra en la Figura 1. Al hacer uso de un histograma se observan dos grandes agrupaciones. Mas aun, al estimar la densidad de la esperanza de vida con KDE se reafirma la presencia de dos grupos, pues se ven dos acumulaciones de densidad.

Ya que se identifican dos grupos, se utiliza la técnica de agrupación K-means para dividir los países en ambos grupos. Dado que la diferencia en los grupos es la esperanza de vida, y un grupo tiene

mayor esperanza de vida que el otro, estos grupos se refieren por los nombres “Esperanza de vida baja” y “Esperanza de vida alta”. Observar características diferentes en países de estos grupos permite hacer inferencia sobre las características que permiten tener esperanza de vida alta, o baja. En el gráfico de barras se puede observar la esperanza de vida de cada país, así como ver qué país pertenece a qué grupo. Este gráfico deja ver el salto que hay entre los países con esperanza de vida baja, a los que tienen esperanza de vida alta. Se puede observar que se identificaron más países en la categoría de esperanza de vida alta.

Más aun, usando el gráfico de violín se observa la distribución de esperanza de vida de cada grupo. Se puede notar como el grupo de esperanza de vida alta tiene valores sin mucha varianza, a excepción de un outlier muy separado del resto (en el gráfico de barras se puede ver que este es Mónaco). Para los países con esperanza de vida baja, se tiene más variabilidad, y se tienen unos pocos países (Kosovo, Rusia, Ucrania y Moldova) con esperanza de vida ligeramente menor que el resto, aunque estos no se consideran outliers.

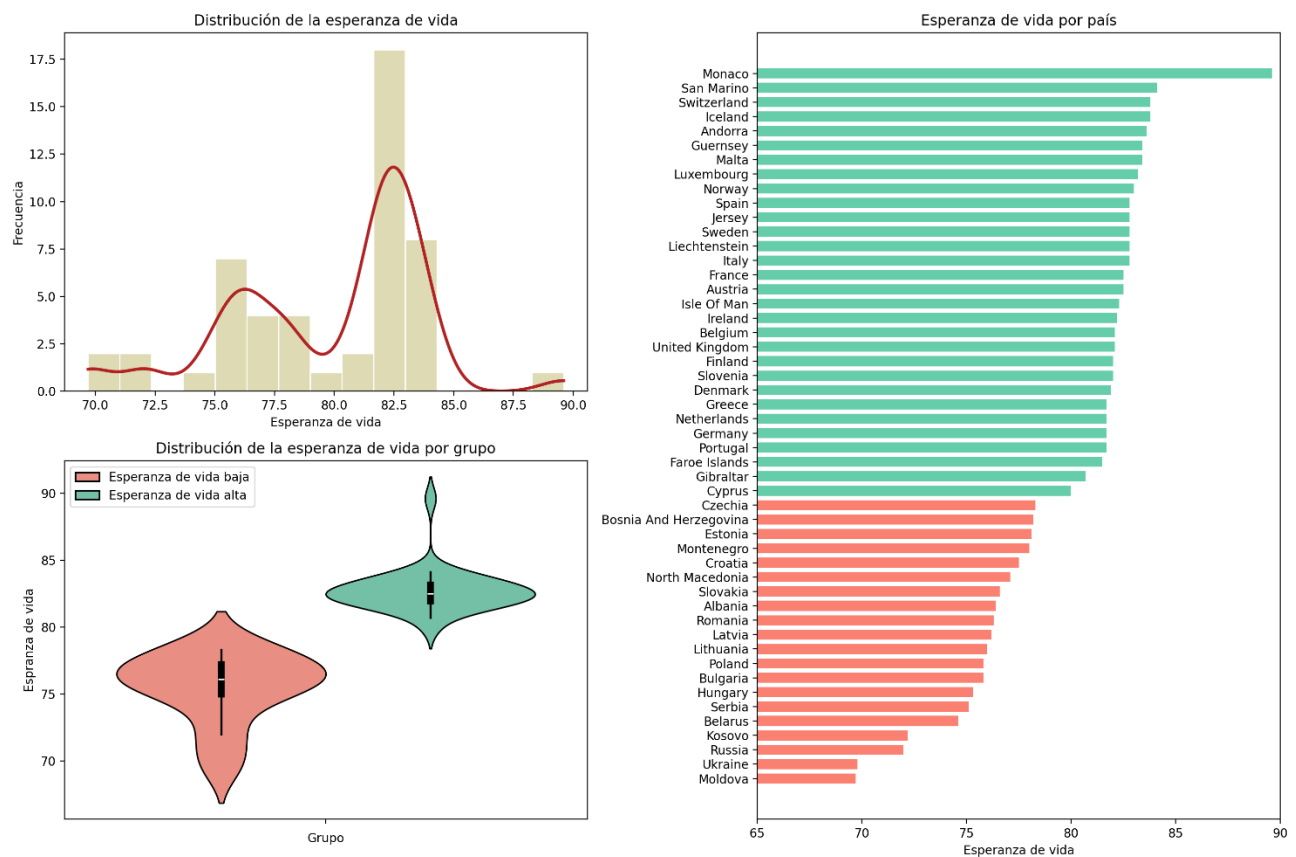


Figura 1

Ya que se identificó la distribución de la esperanza de vida, y los grupos de países, la primera variable a explorar es la localización geográfica. En la Figura 2 se muestran dos mapas que muestran la esperanza de vida de cada país de Europa, así como la categoría a la que pertenecen.

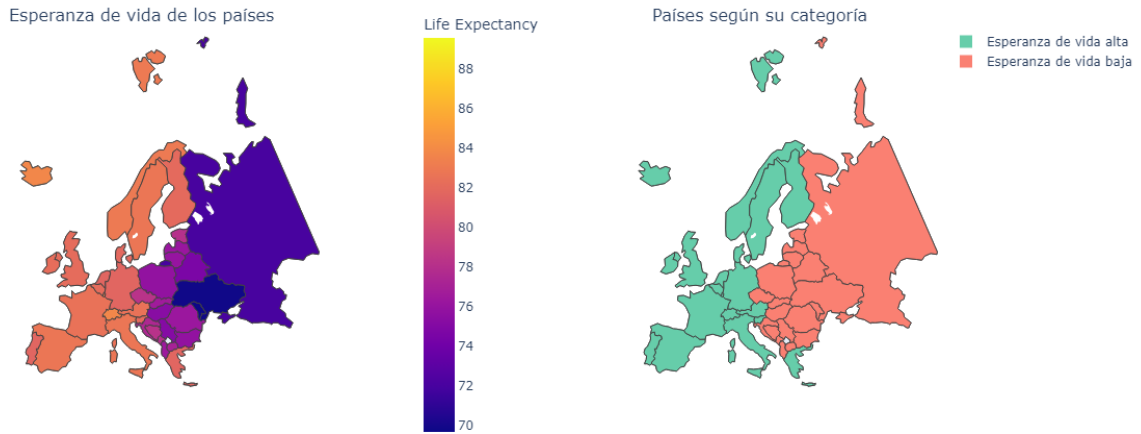


Figura 2

Se puede inferir que la posición geográfica de los países tiene una fuerte influencia en su esperanza de vida. Se observa que la categoría de países con esperanza de vida alta se carga al este, mientras en el oeste se encuentran países con esperanza de vida más baja. Esta división es casi perfecta, por lo que se puede pensar que la esperanza de vida de un país es fuertemente influenciada por situaciones políticas y económicas derivadas de la posición geográfica del país.

Posteriormente se analiza el factor temporal en la esperanza de vida. En la Figura 3 se presenta una visualización que muestra datos históricos de la esperanza de vida de estos países. Cada línea representa un país, y el color de la línea representa su categoría. En esta grafica interactiva se pueden explorar los datos de los ultimos 20 años usando el deslizador de abajo y los botones del costado.

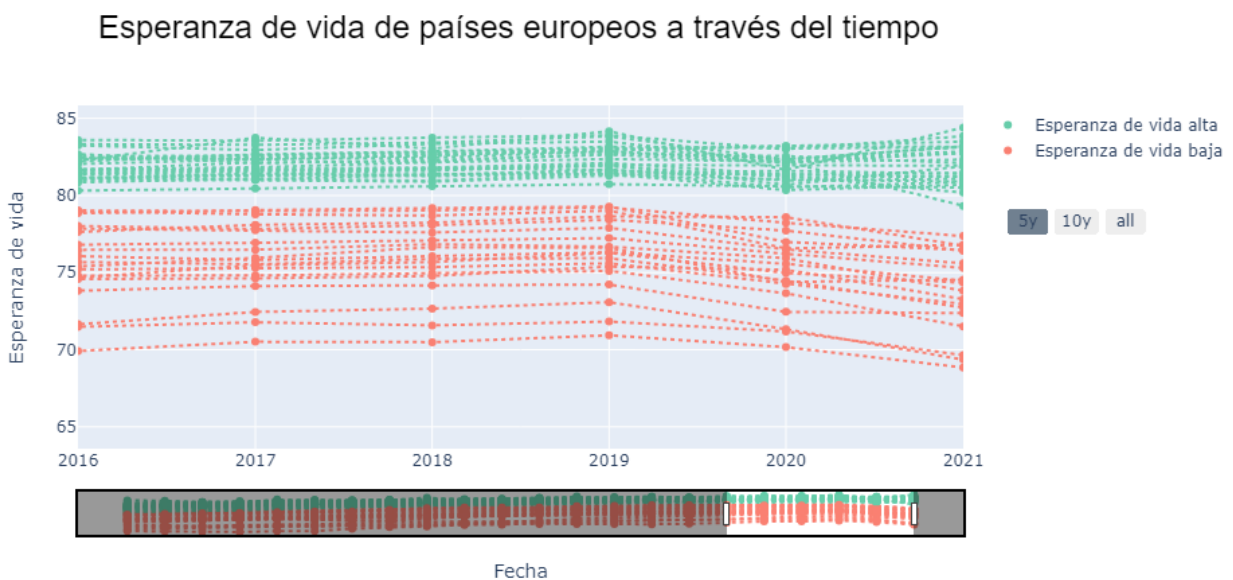


Figura 3

Se puede apreciar que la división entre países con esperanza de vida alta y baja lleva años manteniéndose. Es decir, los países que tienen actualmente esperanza de vida alta, llevan años teniéndola. Aunque llega a cambiar el orden de países entre países del mismo grupo, no cambia el orden entre países de distintos grupos. Esto infiere que la división de países según su esperanza de vida es bastante fuerte, estos grupos están bien delimitados desde hace ya varios años.

Este análisis deja ver que, sin importar los cambios implementados por países con esperanza de vida baja, alcanzar la esperanza de vida del grupo más alto es poco probable, o incluso imposible. Por lo tanto, se infiere que tener cambios significativos en la esperanza de vida de un país es un desafío muy difícil, por la cantidad de factores que influyen, los cuales se ven afectados por la historia de cada país.

Posteriormente, se enfoca el análisis en analizar la correlación de la esperanza de vida con el gasto en salud per cápita, y el gasto externo en salud per cápita. Dado que usualmente se asume que estas variables impulsan la esperanza de vida, es de interés explorar y describir este impacto. En la Figura 4 se presenta una evolución temporal de estas variables, donde se toma el promedio de todos los países en Europa. Dado que las 3 variables de interés están en diferente escala, y únicamente es de interés comparar comportamiento relativo entre ellas, se normalizan las variables para que tomen valores únicamente en el intervalo $[0,1]$. Se usa una normalización Min-Max para lograr esto. Ya que se tienen las variables en la serie de tiempo, se pueden observar patrones en comportamiento, es decir, se puede observar si el crecimiento en una variable se relaciona con el crecimiento de alguna otra.

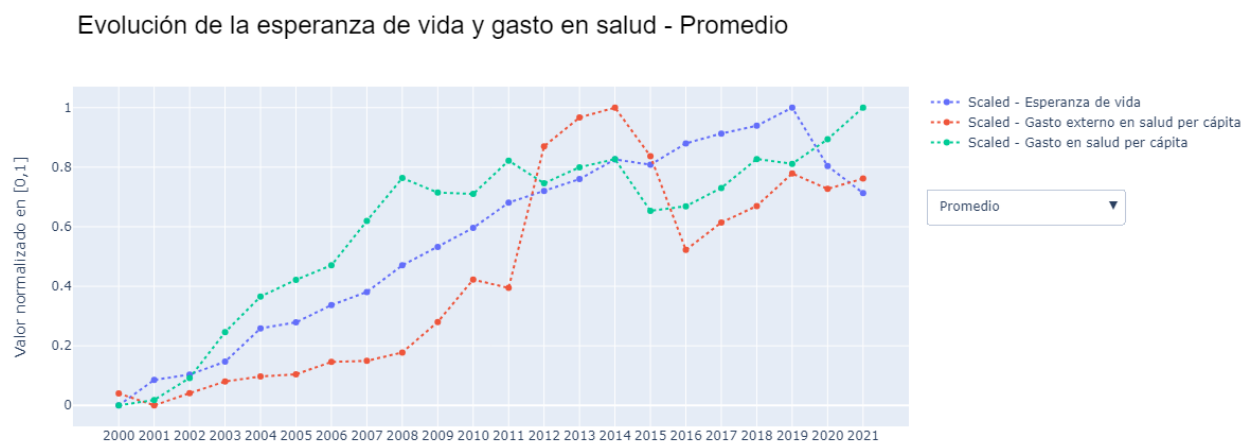


Figura 4

Para el promedio de países en Europa se observa una tendencia creciente parecida para la esperanza de vida y el gasto en salud per cápita, infiriendo una correlación entre estas variables. Se puede ver que el gasto externo en salud no sigue un comportamiento tan similar a las otras dos variables. Aunque también tiene tendencia creciente, se puede ver una diferencia en comportamiento, denotando una correlación menor con esperanza de vida.

Usando la misma visualización se puede ver la evolución temporal de estas variables para otros países. Esto permite encontrar tendencias particulares de cada país. Por ejemplo, en la Figura 5 se observan los ejemplos de Estonia y United Kingdom. Se muestra que para United Kingdom las 3

variables muestran una tendencia similar. Sin embargo, para Estonia el gasto externo en salud tiene una tendencia muy diferente a las otras dos variables. Por lo tanto, se observa que el impacto que tenga el gasto en salud, o el gasto externo en salud en la esperanza de vida depende de cada país, pues puede estar influenciado por muchos otros factores.



Figura 5

Después de analizar la correlación de estas variables temporalmente, es de interés visualizar esta correlación a nivel país. Tomando los datos del 2021, en la Figura 6 se presentan dos mapas de dispersión de la esperanza de vida contra el gasto en salud per cápita, y el gasto externo en salud per cápita. Cada punto representa un país, donde el tamaño del punto representa la cantidad de población y el color del punto denota el grupo al que pertenece el país. Esta gráfica, a diferencia de la Figura 5, permite realizar un análisis a nivel país, con valores absolutos de las variables, en lugar de un análisis a nivel temporal, con valores relativos de las variables.

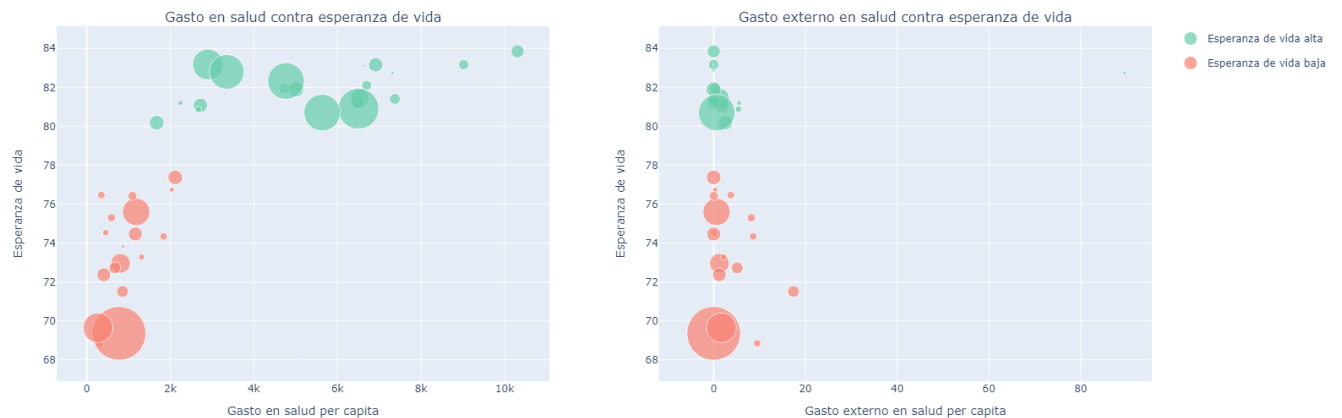


Figura 6

Como algunas de las observaciones anteriores inferían, se puede ver que el gasto externo en salud per cápita y la esperanza de vida no parecen seguir un patrón específico. Es decir, sin importar la cantidad que se invierta externamente en un país para gasto en salud, esto parece no tener ningún efecto en la esperanza de vida del país. Es interesante ver que se tiene un outlier (Luxemburgo) para el cual se tiene muchísimo gasto externo en salud per cápita, pero ni siquiera es el país con más esperanza de vida.

Por otro lado, si se puede observar correlación entre el gasto en salud per cápita y la esperanza de vida. Esta no es una correlación lineal tan clara, pero se puede apreciar que los países con esperanza de vida alta (verdes) tienen más gasto en salud per cápita que los países con baja esperanza de vida (rojos).

Es interesante ver que, aunque se tenga una correlación general entre estas variables, este patrón se rompe al observar algún grupo en específico. Es decir, si solo se observan los países con esperanza de vida alta (verdes) o solo se observan los países con esperanza de vida baja (rojos) entonces no es claro que exista una correlación tan clara entre el gasto en salud per cápita y la esperanza de vida. De nuevo, esto indica una fuerte separación entre estos grupos. Pues el gasto en salud es útil para diferenciarlos, pero no para observar patrones dentro de los países de un mismo grupo. Estas observaciones se presentan en la Figura 7, donde se observan capturas de pantallas provenientes de la visualización de la Figura 6.

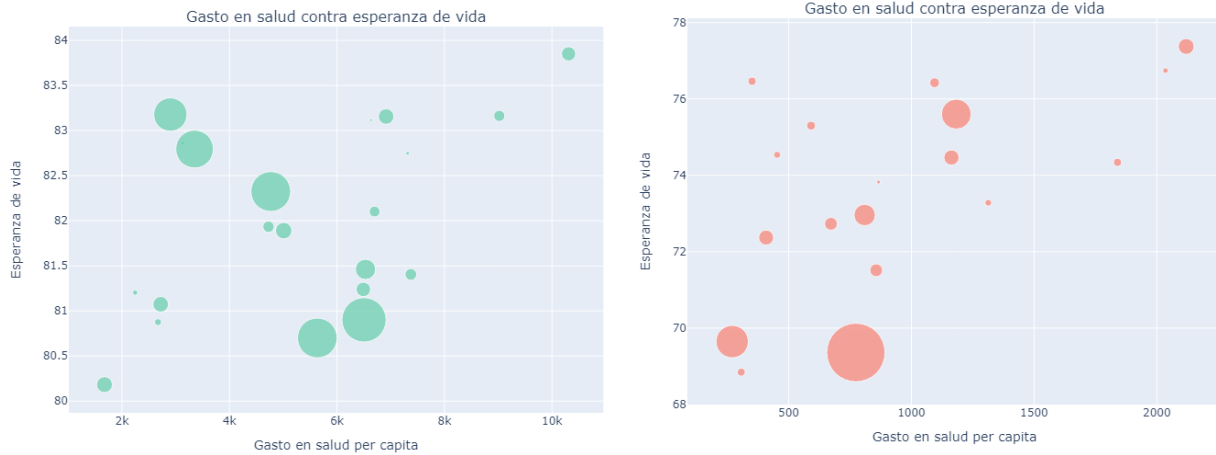


Figura 7

Ya que se identificó una correlación entre el gasto en salud per cápita y la esperanza de vida. Es de interés observar la evolución de esta correlación. La Figura 8 muestra una captura de pantalla de una visualización dinámica que muestra la gráfica de dispersión de estas variables a través del tiempo. Al igual que con la visualización pasada, el color de los puntos denota su grupo, mientras que el tamaño representa la población de los países.

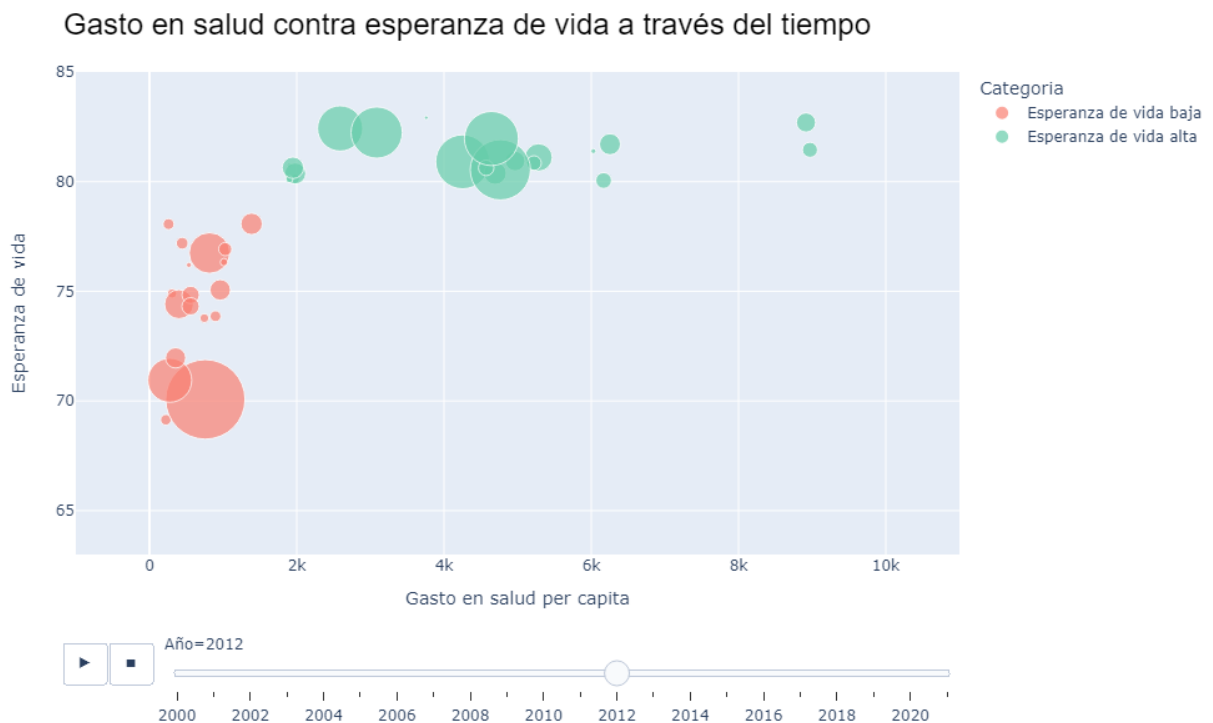


Figura 8

Usando esta visualización se puede observar como cambia el gasto en salud de los países de ambos grupos. Es interesante notar que, durante mas de 10 años, todos los países con esperanza de vida alta tenían mas gasto en salud per cápita que todos los países con esperanza de vida baja, recalando que esta variable también presenta valores muy diferentes para países de ambos grupos. Además, se pueden observar casos particulares, como el de Albania, el cual pertenece al grupo de países con esperanza de vida baja, este país por algunos años fue el país con mayor esperanza de vida de su grupo; sin embargo, es de los que presentan menor gasto en salud per cápita.

Analizar la variable de gasto en salud es analizar “cuanto” se gasta en salud, ahora se enfoca en analisis en “que” se gasta este dinero. Se obtiene el gasto de cada país en los siguientes cinco rubros: Hospital, Farmacia, Clínica, Dentista, Doctor. Por lo que se puede normalizar para saber las proporciones de gasto de cada rubro. Posteriormente se agrupan los países según su esperanza de vida, para analizar como ambos grupos invierten dinero en estos rubros. La Figura 9 presenta esta informacion. La línea vertical colocada en cada barra indica un intervalo al 95% de confianza para el gasto de los países de ese grupo en el dominio correspondiente.

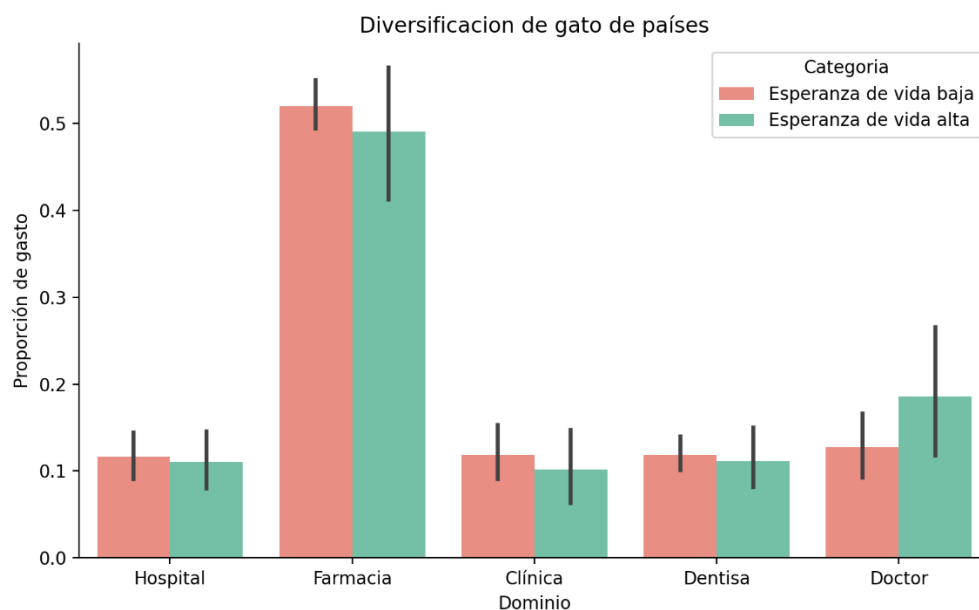


Figura 9

Se puede observar ligeras diferencias en la diversificación de gastos para ambos grupos; por ejemplo, los países con esperanza de vida alta invierten mas en factores relacionados con Doctores que los países con esperanza de vida baja. Sin embargo, estas diferencias no son grandes. Al considerar los intervalos de confianza, se observa que, para cada rubro, los intervalos de confianza de ambos grupos se intersecan. Esto quiere decir que ambos grupos invierten aproximadamente las mismas proporciones en los cinco rubros. Por lo tanto, la esperanza de vida de un país no parece estar influenciada por la diversificación del gasto en salud.

Finalmente se busca detectar la influencia de otras variables sociales de salud en la esperanza de vida. Se recolectan 44 indicadores para cada país, los cuales estén relacionados con la sociedad del país, específicamente su salud. Algunas de las variables consideradas son: tasa de fertilidad adolescente, porcentajes de niños de un año vacunados contra el HepB3, tasa de mortalidad por suicidio, porcentaje de hombres desempleados e incidencia de tuberculosis por cada 100,000 personas.

El objetivo es determinar si estas variables están correlacionadas o no con la esperanza de vida de los países. Para cada país se considera un vector de características asociado con los valores de las 44 variables consideradas. Posteriormente se usa PCA para reducción de estas variables a dos dimensiones con el objetivo de visualizarla. La Figura 10 muestra las variables después de la reducción de dimensionalidad. Cada punto representa un país, su color representa el grupo, y el tamaño representa la esperanza de vida, donde un tamaño mayor denota una mayor esperanza de vida.

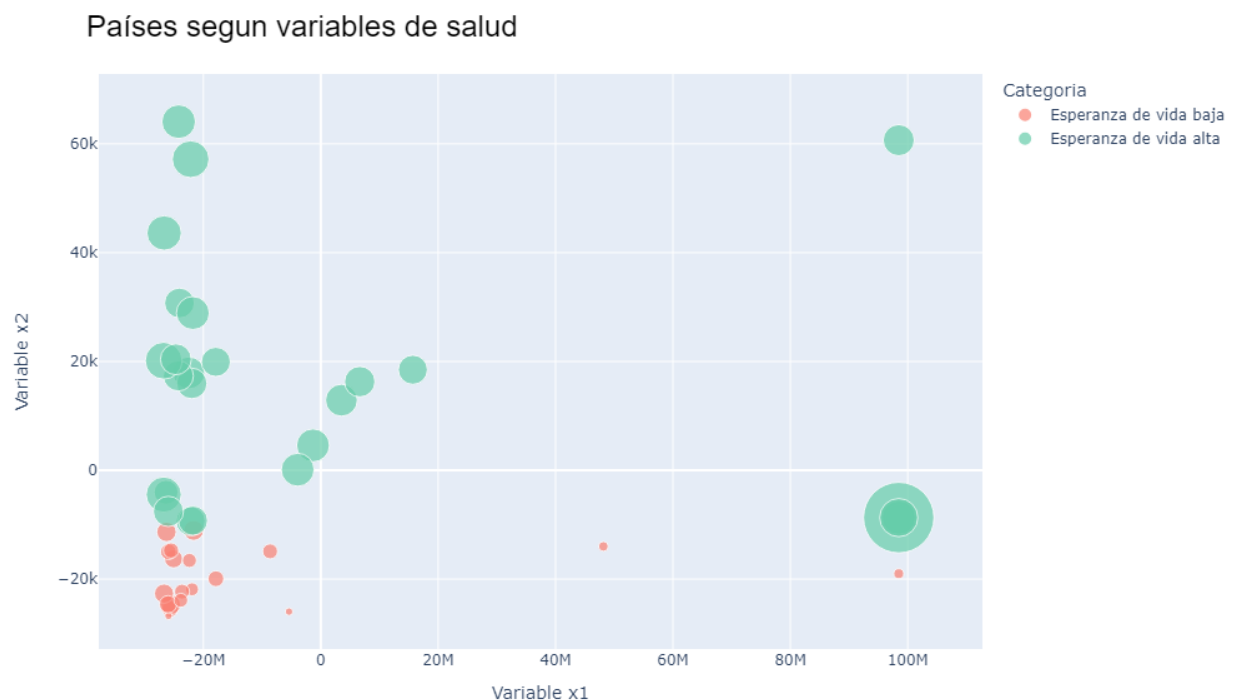


Figura 10

Parece que los países con esperanza de vida baja se agrupan sin mezclarse con los países con esperanza de vida alta. Hay países de distintos grupos que están cercanos, por ejemplo, en la esquina inferior derecha se observa un país con esperanza de vida muy baja (Kosovo) cercano a dos países con esperanza de vida muy alta (Gibraltar y Mónaco). Sin embargo, parece ser que las variables resultado del PCA discriminan de buena manera a los países de acuerdo con su grupo. Es decir, se puede inferir que las 44 variables consideradas tienen un impacto real en la esperanza de vida de los países.

Para reforzar la idea de que las variables consideradas separan de buena manera a los países de acuerdo con su grupo se realiza un modelo de regresión logística. Se utilizan las 44 variables consideradas para el PCA para predecir el grupo de un país, es decir, predecir si tiene alta o baja esperanza de vida. Después de crear el modelo, se calcula su deviance residual, y se resta con el deviance del modelo nulo. Esto permite hacer una prueba de hipótesis donde en la hipótesis nula el modelo propuesto no ayuda a discriminar los grupos. Al calcular el p-value de la prueba se obtiene un valor de 0.03. Esto quiere decir que se rechaza la hipótesis nula. Por lo tanto, hay evidencia estadística de que las variables consideradas por el modelo son de ayuda para clasificar los países de acuerdo con su grupo.

Este hallazgo recalca el hecho de que la esperanza de vida no depende únicamente del gasto en salud, pues hay muchos factores sociales que pueden influenciarlo.

Conclusión y recomendaciones

Síntesis del proyecto

En este proyecto se analizó la esperanza de vida para los países de Europa. Se identificaron dos grupos de acuerdo con su esperanza de vida, y se encontró un fuerte impacto de la localización geográfica de los países para determinar su esperanza de vida. Se observó la fuerte tendencia histórica que separa estos grupos, creando una brecha entre estos países, la cual puede ser un desafío cerrar.

En cuanto a la cantidad de dinero invertida, se encontró que el gasto externo en salud no tiene un impacto claro en la esperanza de vida, mientras que el gasto en salud tiene mayor impacto. Sin embargo, la correlación entre gasto en salud y esperanza de vida no es tan clara al analizar países del mismo grupo, únicamente es perceptible al considerar y comparar ambos grupos. Además, se encontró que ambos grupos diversifican los gastos de la misma manera.

Finalmente, se analizó como un gran número de variables sociales determinan el grupo al que pertenece cada país. Recalcando que este indicador toma en cuenta varios aspectos de la sociedad, no únicamente variables que se impulsan subiendo el gasto en salud.

Recomendaciones basadas en los resultados

La preocupación principal es la brecha de esperanza de vida que existe entre los países de los dos grupos identificados. Se deben de buscar cerrar esta brecha para ayudar a los países con esperanza de vida baja alcanzar un mayor valor de este indicador.

Subir drásticamente el gasto en salud puede ser una estrategia para que los países con baja esperanza de vida puedan obtener mayor esperanza de vida. Al cerrar la brecha entre grupos del gasto en salud se pudiera cerrar la brecha entre grupos de esperanza de vida. No es necesario hacer cambios en gasto externo de salud, ni en cómo se diversifica el gasto. Finalmente mejorar políticas

sociales puede mejorar a subir la esperanza de vida, pues se encontró que varios indicadores de la sociedad pueden tener impacto.

Límites del proyecto y direcciones futura

Es importante reconocer que este proyecto, al ser un proyecto escolar con tiempo limitado de realización, considera un número muy limitado de variables en el análisis, y no se consideran todas las interacciones posibles entre ellas, además de que la delimitación temporal se enfoca únicamente en los últimos años. Un análisis más exhaustivo puede revelar conclusiones más contundentes e interesantes sobre este fenómeno.

Además, como se menciona en algunas de las visualizaciones, cada país puede tener comportamientos específicos y no seguir el patrón general. Por lo tanto, las conclusiones obtenidas del análisis general de todos los países deben tomarse con cuidado antes de aplicarlas a un país específico, pues puede que este país no siga este patrón.

Finalmente, este proyecto, aunque usa técnicas estadísticas, es mayormente de visualización. Un enfoque más estadístico puede dar resultados más confiables, pues la significancia estadística ayuda a respaldar las conclusiones obtenidas.

Como trabajo futuro se puede extender este análisis para considerar más variables. Se pueden usar técnicas econométricas para el análisis de las series de tiempo y técnicas estadísticas y de aprendizaje de máquina para encontrar más correlaciones entre las variables. Finalmente, este análisis puede aumentar el enfoque para estudiar países de otra región, o incluso de todo el mundo.

Referencias

UN-Habitat. (2024, February). Quality of Life Initiative: Synthesis paper 1 - A brief history of quality of life [PDF]. <https://unhabitat.org/>

/2023). Life Expectancy – What does this actually mean?. Our World in Data. Retrieved from <https://ourworldindata.org/grapher/life-expectancy>