

VISUALIZACIÓN AVANZADA

PRÁCTICA 3.4: ANÁLISIS DE LOS DATOS SOBRE DISTINTOS

ASPECTOS DE LA POBLACIÓN MUNDIAL

USC, CURSO 2022-2023

DIANA MASCAREÑAS SANDE

ÍNDICE

ELECCIÓN DE UN DATASET	3
TEMA A TRATAR	5
PREPROCESADO DE LOS DATOS	5
ESTUDIO DE LOS DATOS.....	10
ANÁLISIS DE LOS RESULTADOS.....	13
BIBLIOGRAFÍA	17

ELECCIÓN DE UN DATASET

Para la elaboración de esta práctica, se ha elegido la base de datos *World Population Data*, que se puede encontrar en la fuente *kaggle.com*. La referencia exacta está disponible en el apartado de [bibliografía](#).

Este dataset cuenta con cinco documentos csv distintos (solo se emplearán cuatro). Se han cambiado el formato a xlsx ya que Excel advierte que se podría producir una pérdida de datos de no hacerlo. A continuación, se procede a explicar qué contiene cada uno de ellos:

<i>height_weight_data.xlsx</i>	
<i>country</i>	País
<i>male_height</i>	Altura de los hombres
<i>female_height</i>	Altura de las mujeres
<i>male_weight</i>	Peso de los hombres
<i>female_weight</i>	Peso de las mujeres
<i>male_bmi</i>	Índice de masa corporal de los hombres
<i>female_bmi</i>	Índice de masa corporal de las mujeres

<i>iq.xlsx</i>	
<i>country</i>	País
<i>iq</i>	Cociente intelectual medio
<i>education_expenditure_per_inhabitant</i>	Gasto en educación en dólares
<i>daily_max_temp</i>	Temperatura máxima al día en grados Celsius

<i>life_expectancy.xlsx</i>	
<i>country</i>	País
<i>male_life_expectancy</i>	Esperanza de vida de los hombres
<i>female_life_expectancy</i>	Esperanza de vida de las mujeres
<i>birth_rate</i>	Tasa de natalidad
<i>death_rate</i>	Índice de mortalidad

<i>quality_of_life.xlsx</i>	
<i>country</i>	País
<i>stability</i>	Estabilidad sobre 100
<i>rights</i>	Derechos sobre 100
<i>health</i>	Salud sobre 100
<i>safety</i>	Seguridad sobre 100
<i>climate</i>	Clima sobre 100
<i>costs</i>	Costes sobre 100
<i>popularity</i>	Popularidad sobre 100

TEMA A TRATAR

Es interesante conocer cómo se desenvuelve la vida en los diferentes países del mundo: ¿la calidad de vida de las personas en diferentes países del mundo puede alterar otros factores?

Los documentos que se han encontrado proporcionan información sobre el IMC, IQ, esperanza de vida, índice de natalidad y muertes. ¿Es posible que estos factores se vean afectados por la calidad de vida de las personas?

Se construirá una serie de representaciones estadísticas para responder a esta pregunta. El proceso de construcción de estos gráficos estará explicado en este documento pero, antes de comenzar a representar datos, es necesario observar cuáles van a ser realmente útiles. Por este motivo, el primer paso es realizar un preprocesado de datos.

PREPROCESADO DE LOS DATOS

En primer lugar, se seleccionará la primera columna de todos los ficheros, y en la pestaña de datos, se seleccionará la opción de *Texto en columnas*, eligiendo la coma como separador.

Ahora ya se tienen los datos por columnas en todos los ficheros. El siguiente paso es comprobar si existen errores de formato. Dado que la mayoría de datos son numéricos, no hay problemas de incompatibilidad en este aspecto.

A continuación, se pasarán todos los datos a tablas, para poder manejar la información más fácilmente.

El siguiente paso es comprobar documento a documento qué columnas van a interesar en el estudio:

<i>height_weight_data.xlsx</i>	
<i>country</i>	País
<i>male_height</i>	Altura de los hombres en cm
<i>female_height</i>	Altura de las mujeres en cm
<i>male_weight</i>	Peso de los hombres en kg
<i>female_weight</i>	Peso de las mujeres en kg
<i>male_bmi</i>	Índice de masa corporal de los hombres
<i>female_bmi</i>	Índice de masa corporal de las mujeres

En este caso, como el índice de masa corporal se define en función de la altura y peso, se podrían eliminar estas dos últimas columnas, tanto para hombres como para mujeres, quedando el índice de masa corporal de hombres y mujeres. Primero, se comprobará que los valores estén correctamente calculados: el IMC es el resultado de dividir el peso entre la altura en metros cuadrados. Los datos vienen dados con puntos para marcar los decimales, de modo que estos se sustituyen por comas mediante la opción de *Buscar y seleccionar*>*Reemplazar*.

Se añade una columna de comprobación que devuelve 0 si el IMC está bien calculado, y 1 si no:

	A	B	C	D	E	F
	country	male_height	female_height	male_weight	female_weight	
2	Netherlands	184,00	170,00	87,90	75	
3	Montenegro	183,00	170,00	90,40	75	
4	Estonia	182,00	168,00	89,90	75	
5	Denmark	182,00	169,00	86,80	70	
6	Bosnia and Herzegovina	182,00	167,00	87,10	70	
7	Iceland	181,00	168,00	89,20	72	
8	Czechia	181,00	168,00	91,90	74	
9	Slovenia	181,00	167,00	87,00	74	
10	Slovakia	181,00	167,00	89,50	71	
11	Croatia	181,00	167,00	91,30	74	
12	Serbia	180,00	168,00	86,90	72	
13	Sweden	180,00	167,00	87,10	70	

Ilustración 1: comprobación del cálculo del IMC

Todos los valores son 0, de modo que se puede eliminar el peso y altura de hombres y mujeres, además de esta última columna de comprobación.

Se añadirá una columna que medirá el nivel de peso en función de este valor. Los valores de esta nueva columna serán:

- Bajo peso: IMC por debajo de 18,5
- Normal: IMC entre 18,5 y 24,9
- Sobrepeso: IMC entre 25,0 y 29,9
- Obesidad: IMC de 30,0 o más

Excel spreadsheet showing a list of countries and their male BMI values. The formula bar displays a complex IF statement for categorizing BMI levels.

country	male_bmi
Netherlands	26,1
Montenegro	27,0
Estonia	27,0
Denmark	26,3
Bosnia and Herzegovina	26,4
Iceland	27,1
Czechia	28,1
Slovenia	26,6
Slovakia	27,4
Croatia	28,0
Serbia	26,7
Sweden	26,8
Norway	27,4
Lithuania	27,1

Ilustración 2: nivel de peso de los hombres en los países del mundo

Como se puede apreciar, no hay ningún país cuya media de IMC de los hombres dé como resultado un nivel de peso correspondiente a “Bajo peso”. Se procede a realizar el mismo procedimiento con las mujeres.

Excel spreadsheet showing a list of countries with both male and female BMI values. The formula bar displays a complex IF statement for categorizing BMI levels.

country	male_bmi	female_bmi	Nivel
Netherlands	26,10	25,30	Sobre
Montenegro	27,00	26,20	Sobre
Estonia	27,00	26,00	Sobre
Denmark	26,30	24,60	Sobre
Bosnia and Herzegovina	26,40	25,30	Sobre
Iceland	27,10	25,60	Sobre
Czechia	28,10	26,40	Sobre
Slovenia	26,60	26,60	Sobre
Slovakia	27,40	25,60	Sobre
Croatia	28,00	26,90	Sobre
Serbia	26,70	25,70	Sobre
Sweden	26,80	25,40	Sobre
Norway	27,40	26,20	Sobre
Lithuania	27,10	26,10	Sobre

Ilustración 3: nivel de peso de las mujeres en los países del mundo

El fichero resultante quedaría con la siguiente información:

<i>height_weight_data.xlsx</i>	
<i>country</i>	País
<i>male_bmi</i>	Índice de masa corporal de los hombres
<i>female_bmi</i>	Índice de masa corporal de las mujeres
<i>Nivel de peso de los hombres</i>	Nivel de peso en función del IMC de los hombres
<i>Nivel de peso de las mujeres</i>	Nivel de peso en función del IMC de las mujeres

Se mantendrá esta división entre hombres y mujeres, para después comprobar quiénes mantienen un mejor IMC en los distintos países del mundo.

Se continúa con el siguiente fichero:

<i>iq.xlsx</i>	
<i>country</i>	País
<i>iq</i>	Cociente intelectual medio
<i>education_expenditure_per_inhabitant</i>	Gasto en educación en dólares
<i>daily_max_temp</i>	Temperatura máxima al día en grados Celsius

En este caso, los únicos datos que van a interesar son los de la columna de *iq* (estará en formato numérico), de modo que se eliminan las anteriores y se añadirá una columna denominada categoría, que medirá el nivel de inteligencia de los habitantes en función del valor del cociente intelectual, siguiendo el siguiente criterio:

- Muy dotada: 130 o más
- Dotada: entre 121 y 130
- Inteligencia por encima de la media: entre 111 y 120
- Inteligencia media: 90-110
- Inteligencia por debajo de la media: 80-89
- Discapacidad intelectual: 79-70

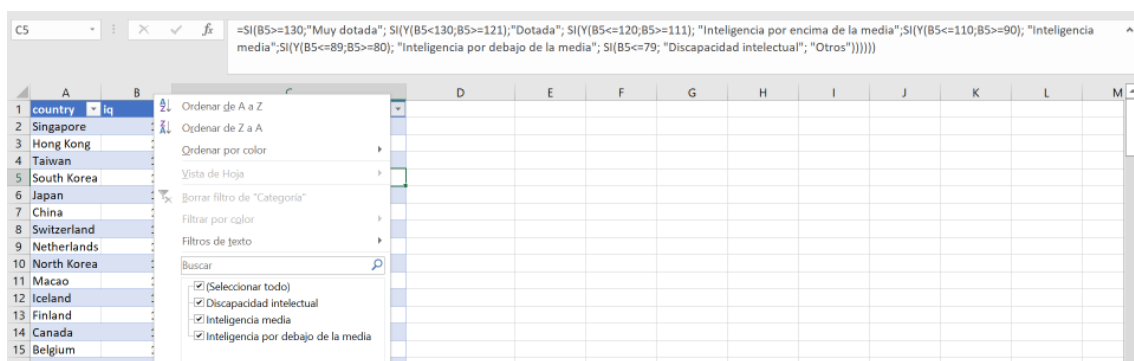


Ilustración 4: categoría por países en función de su iq

El fichero resultante almacena la siguiente información:

<i>iq.xlsx</i>	
<i>country</i>	País
<i>iq</i>	Cociente intelectual medio (en números)
<i>Categoría</i>	Categoría en función del iq

Se continúa con el siguiente fichero:

<i>life_expectancy.xlsx</i>	
<i>country</i>	País
<i>male_life_expectancy</i>	Esperanza de vida de los hombres (en años)
<i>female_life_expectancy</i>	Esperanza de vida de las mujeres (en años)
<i>birth_rate</i>	Tasa de natalidad (en tanto por mil)
<i>death_rate</i>	Índice de mortalidad (en tanto por mil)

De este fichero interesan todos los datos, de modo que, lo único que se hará será modificar el formato de los números (sustituir puntos por comas y poner en formato numérico, poner en formato tanto por mil las tasas).

Del último fichero:

<i>quality_of_life.xlsx</i>	
<i>country</i>	País
<i>stability</i>	Estabilidad sobre 100
<i>rights</i>	Derechos sobre 100
<i>health</i>	Salud sobre 100
<i>safety</i>	Seguridad sobre 100
<i>climate</i>	Clima sobre 100
<i>costs</i>	Costes sobre 100
<i>popularity</i>	Popularidad sobre 100
<i>calidad_media</i>	Media sobre 100

<i>tipo_vida</i>	Clasificación del tipo de vida en función de la media
------------------	---

También interesan todos los datos, de modo que lo único que se hace es pasar a porcentajes los valores, ya que están sobre 100. Se ha añade una nueva columna, producto de hacer la media del resto de los factores. Se trabajará con este valor para comprobar el nivel de felicidad o calidad de vida.

ESTUDIO DE LOS DATOS

Se creará un nuevo documento Excel: *P3.4-VA_DianaMascarenasSande*, que recogerá en cuatro hojas distintas el contenido de los cinco documentos tratados hasta ahora (se prescinde de la hoja. Por otra parte, la primera hoja de este documento será el diccionario de datos de cada hoja, explicando qué contiene cada una y en qué unidades, tal y como se ha explicado en este documento anteriormente.

Se eliminan de las tres tablas de análisis las filas que no estaban contenidas en *quality_of_life*.

A continuación, se añadirán todas las tablas al modelo de datos sobre el que se va a trabajar. Todas las tablas tienen un campo común: el país (*country*), de modo que se establecerán conexiones entre todas las tablas a través de este campo:

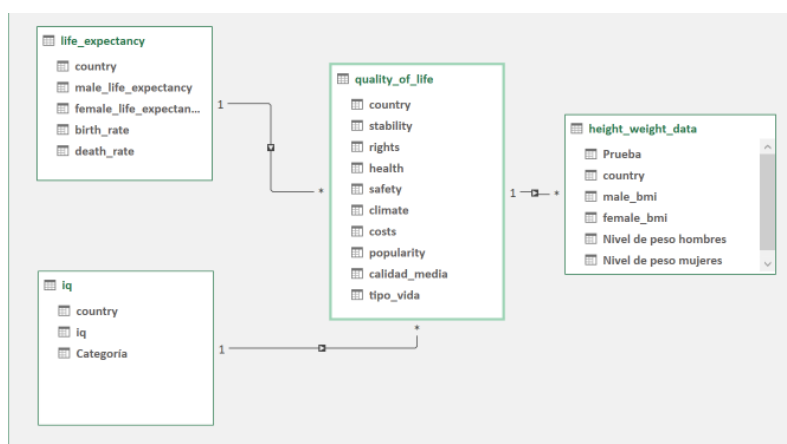


Ilustración 5: tablas relacionadas en el modelo de datos a través del campo de país

A partir de este modelo, se genera una tabla dinámica en una hoja nueva, llamada *Tabla Dinámica*. Se creará una hoja nueva llamada *Colores y Tipografía*, donde se especificarán

qué escala de colores y tipo de letra se emplearán en los dashboards. A partir de este momento, comienza el estudio.

A partir del modelo de datos se generan las tablas dinámicas sobre las que se construirán los siguientes gráficos (todos ellos ordenados en función del campo de la calidad de vida):

- Gráfico de tarjeta que representa el número total de países que se encuentran en la tabla de *quality_of_life*, la cual será la tabla principal del estudio. Son 137, de modo que no se puede visualizar todos los datos de estos 137 países en cada gráfica, puesto que no se acabaría entendiendo ningún dato. Es por esto que se selecciona el top 15 de países con mejor calidad de vida y el top 15 de países con peor calidad de vida. Estos estarán representados bajo el gráfico de tarjeta en forma de dos tablas, en las que se representará la calidad de vida mediante el formato condicional>barras de datos>relleno sólido. Será con este total de 30 países con los que se trabajará de ahora en adelante.
- Índice de Masa Corporal en los 15 países con mejor calidad de vida. Para ello se introducen los valores correspondientes en las filas de la tabla dinámica, a partir de la cual construirá un gráfico combinado que permitirá ver los valores IMC de hombres y mujeres, además del valor de la calidad de vida de forma clara y fácilmente visible. Se ha elegido este gráfico porque resulta adecuado para representar varios conjuntos de datos.
- Índice de Masa Corporal en los 15 países con peor calidad de vida. Se sigue el mismo procedimiento que en el gráfico anteriormente explicado.
- IQ de los países con mejor calidad de vida. Como solo se analiza un valor en función del país y calidad de vida, resulta más visual construir un gráfico de sectores. Aunque no se estén tratando porcentajes, ya que son relativamente pocos campos, se puede ver a qué país corresponde cada IP. Otra opción sería un gráfico de columnas o de barras, que se podría ordenar de mayor a menor. Sin embargo, como no interesa conocer esta ordenación en este caso concreto, si no los valores generales, no resulta vital elegir este gráfico de barras.
- IQ de los países con peor calidad de vida. Se sigue el mismo procedimiento que en el gráfico anteriormente explicado.

- Esperanza de vida de los 15 países con mejor calidad de vida. Se elige un gráfico de columnas en 2D, donde no se representa la calidad de vida porque ya se sabe que se encuentra en sus 15 valores más altos (ordenados de mayor a menor sobre el eje X), y se facilita la visualización de los datos referidos a la esperanza de vida de las personas de estos 15 países.
- Esperanza de vida de los 15 países con peor calidad de vida. Se sigue el mismo procedimiento que en el gráfico anteriormente explicado.
- Tasa de natalidad e índice de mortalidad de los 15 países con mejor calidad de vida. Se crea otro gráfico de barras en 2D. Para facilitar la visualización de los datos, se pone un límite en el eje Y de 30, ya que a pesar de que, los elementos que no aparezcan en la hoja de calidad de vida no aparecerán en el resto, sí puede ocurrir que haya países en esta hoja que no estén contemplados en el resto de hojas, de modo que los agrupa y suman un valor muy alto. Si se ajustara la escala de la gráfica a este valor, no se apreciarían los datos de estudio que realmente interesan.
- Tasa de natalidad e índice de mortalidad de los 15 países con peor calidad de vida. Se sigue el mismo procedimiento que en el gráfico anteriormente explicado pero, en este caso no se necesitaría hacer esa escala para limitar los valores del eje Y, ya que se está tratando con los valores más bajos de la calidad de vida.
- Se añade un gráfico de segmentación en función de *tipo_vida* que estará conectado a todas las gráficas salvo las de sectores, ya que en ese caso concreto interesa conocer su valor fijo.

Se podría haber dividido toda esta información en dos dashboards: uno para cada top 15 de países. Así, se dividirían las gráficas a la mitad en cada uno. Sin embargo, como es un estudio comparativo, ha resultado de más utilidad mantenerlos en el mismo dashboard.

ANÁLISIS DE LOS RESULTADOS

Se realiza un análisis del impacto de la calidad de vida en el índice de masa corporal. Inicialmente, se ha supuesto que la calidad de vida podría afectar a este valor porque se podía sospechar que los países con mayor calidad de vida iban a ser los más desarrollados, y viceversa. De este modo, podrían deducirse más casos de sobrepeso u obesidad en los de mejor calidad de vida que en el top 15 inferior. Sin embargo, al analizar las gráficas de IMC, se puede ver cómo los valores no distan mucho entre una gráfica y otra. Se concluye así que la calidad de vida no es una cuestión decisiva en este factor determinado.

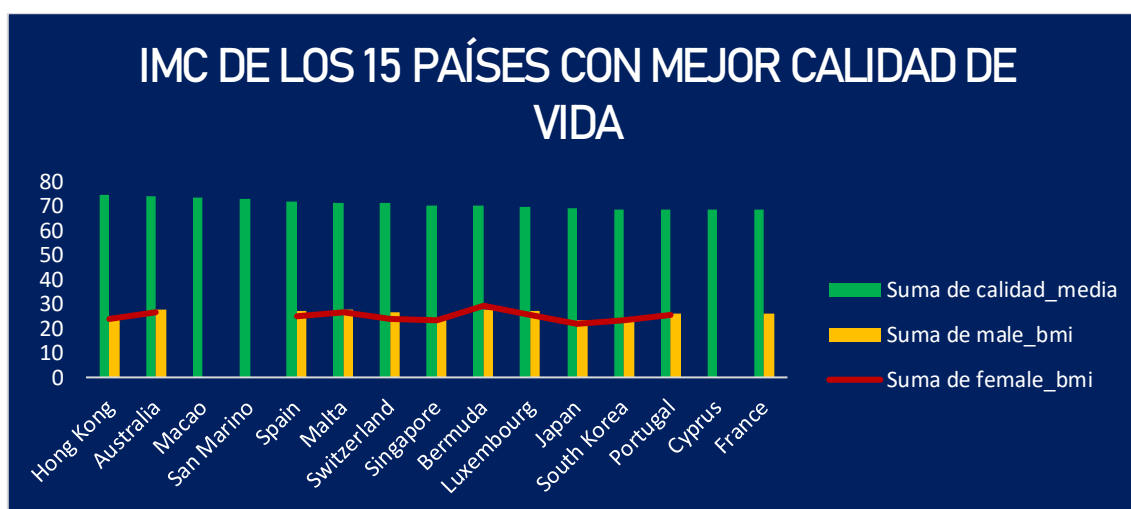


Ilustración 6: IMC top 15 países con mejor calidad de vida



Ilustración 7: IMC top 15 países con peor calidad de vida

Se realiza un análisis del impacto de la calidad de vida en el valor del IQ. En este caso, sí se aprecia una diferencia notable entre ambas gráficas. En el top 15 de mejor calidad de vida se aprecian valores superiores de IQ, mientras que en el top 15 inferior hay valores muy bajos de IQ. Sin embargo, al no especificarse en el dataset qué tipo de test se le ha realizado a la muestra de la población de la que se han obtenido los datos, quizás la prueba sea más bien subjetiva, y no se aplicara a diferentes tipos de inteligencia (emocional, musical... entre otros). No se puede concluir con seguridad que, para todos los tipos de inteligencia, la calidad de vida afecte. Sin embargo, se puede ver cómo este factor sí influye, aunque sea de forma parcial.

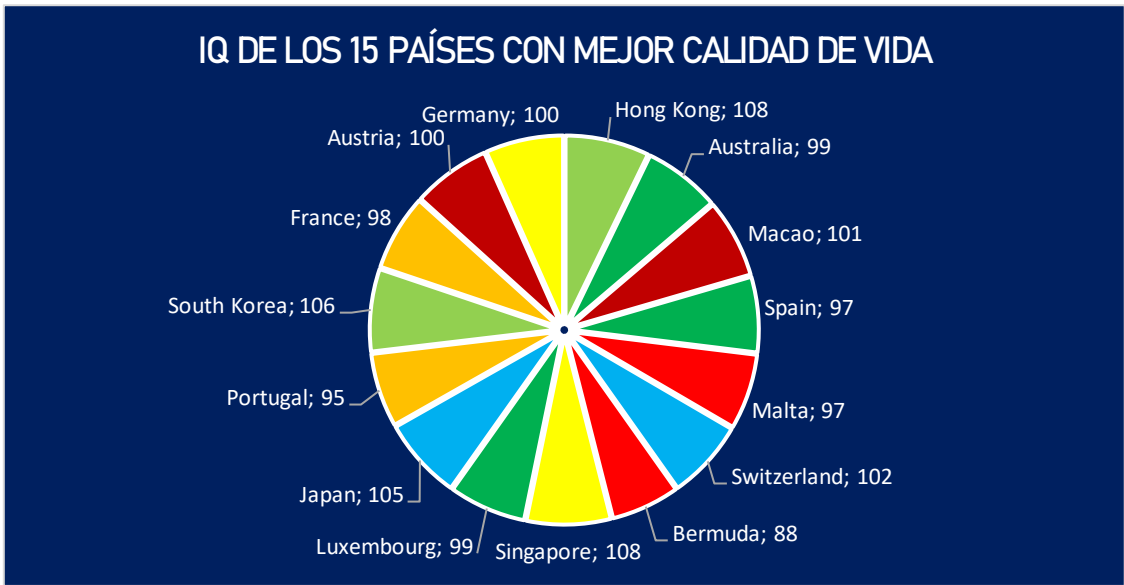


Ilustración 8: IQ top 15 países con mejor calidad de vida

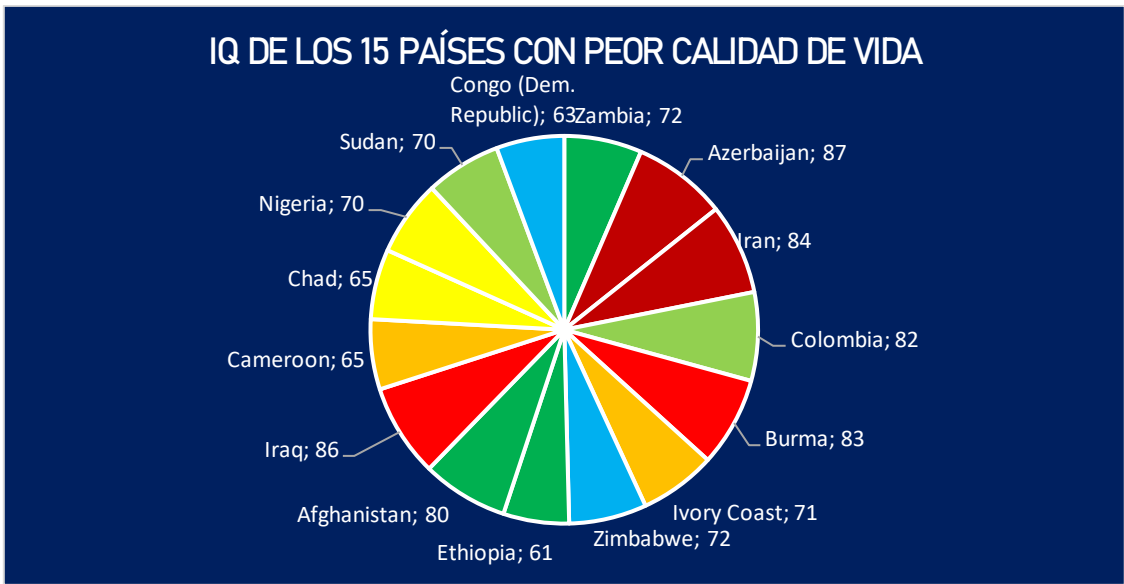


Ilustración 9: IQ top 15 países con peor calidad de vida

Se realiza un análisis del impacto de la calidad de vida en el valor de la esperanza de vida de hombres y mujeres. En este caso sí se puede apreciar una diferencia entre la esperanza de vida en las diferentes gráficas. En el top 15 superior, se alcanzan los 88 años de esperanza de vida para las mujeres, mientras que en el top 15 inferior llega a un máximo de 80. En el caso de los hombres, en el top 15 superior alcanzan los 83 años, mientras que en el inferior, apenas superan los 75. Se concluye así que la calidad de vida afecta a la esperanza de la misma.

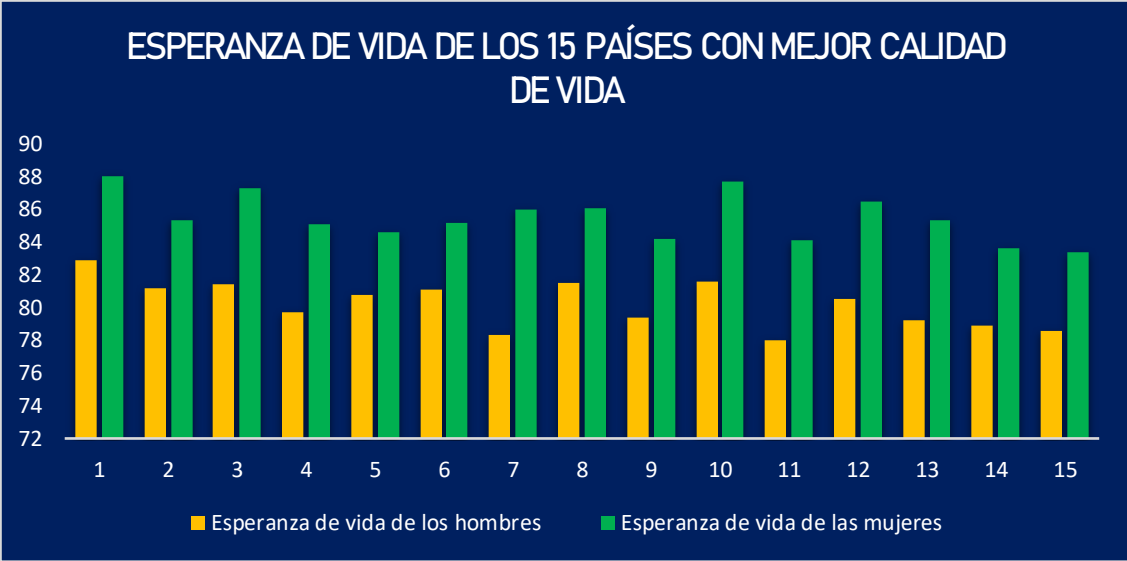


Ilustración 10: esperanza de vida top 15 países con mejor calidad de vida

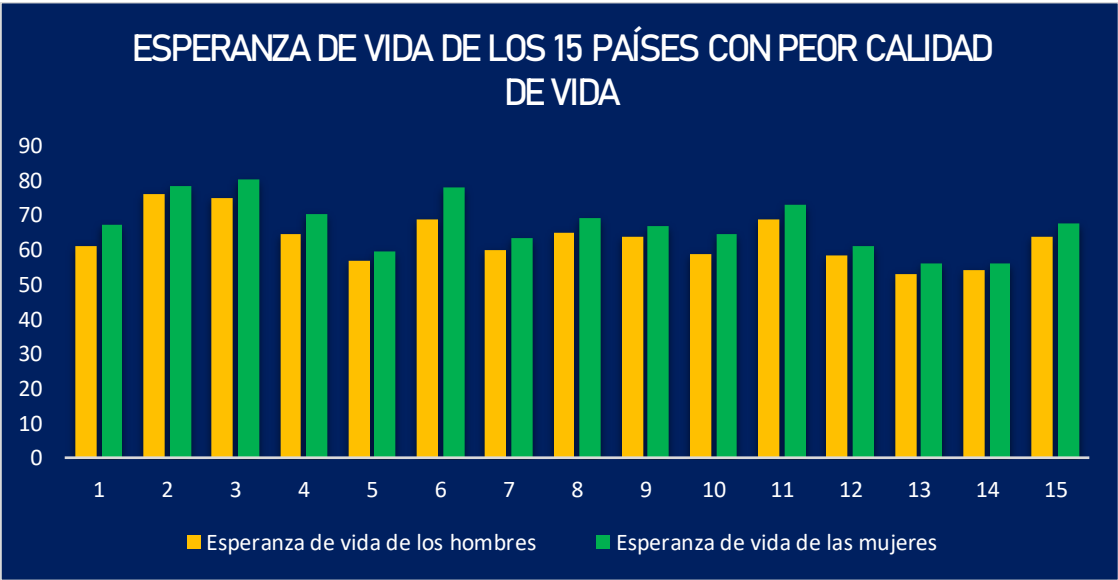


Ilustración 11: esperanza de vida top 15 países con peor calidad de vida

Finalmente, se analiza la tasa de natalidad e índice de mortalidad en función de la calidad de la vida de hombres y mujeres. Se puede apreciar que en el top 15 inferior hay una tasa de natalidad mayor que en el top 15 superior. Si se tiene en cuenta la falta de medios anticonceptivos (peor calidad de vida), este resultado tendría sentido. Por otra parte, fijándose en el valor del índice de mortalidad, en contra de lo que se pudiera creer en un primer momento, son valores bastante parecidos entre ambas gráficas, llegando a ser mayor el índice de mortalidad en el top 15 superior que el del inferior en algún caso. Sin embargo, sería interesante realizar un estudio en mayor profundidad sobre estos datos ya que, como se ha visto anteriormente, la esperanza de vida es menor en el top 15 inferior.

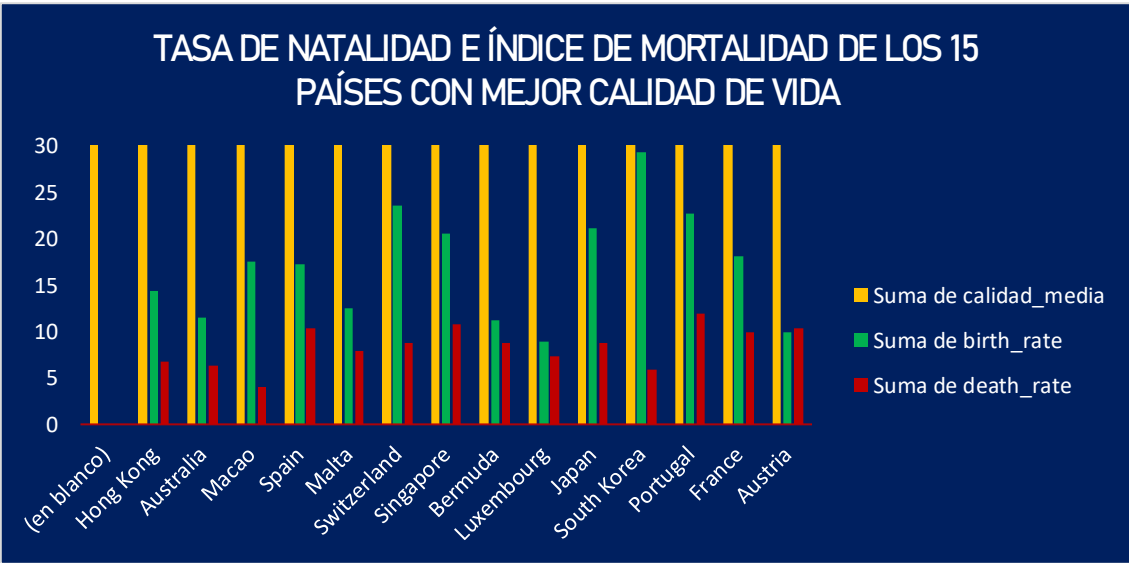


Ilustración 12: tasa de natalidad e índice de mortalidad top 15 países con mejor calidad de vida

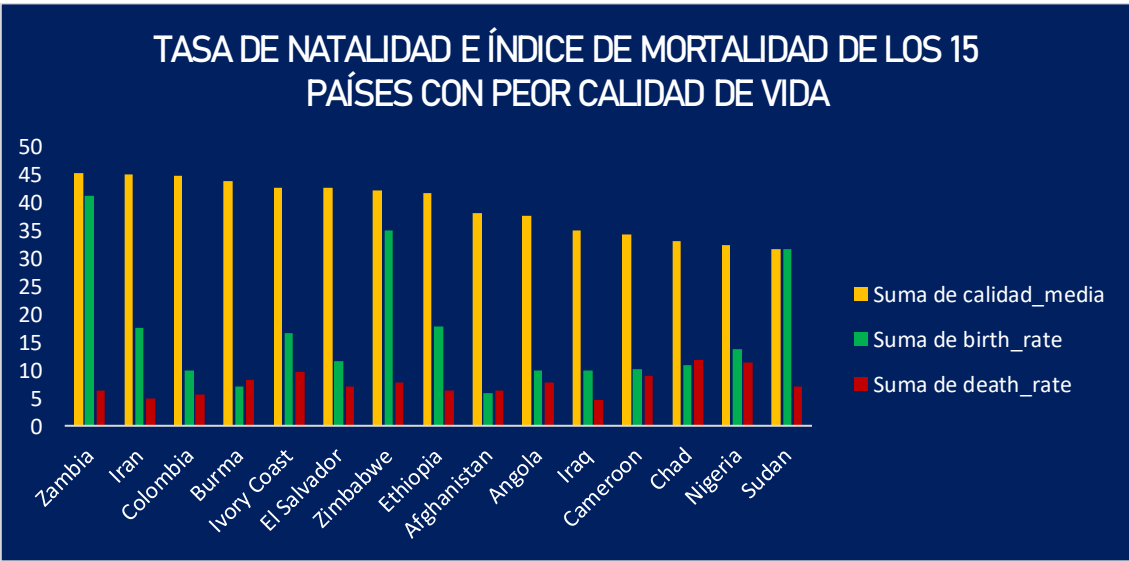


Ilustración 13: tasa de natalidad e índice de mortalidad top 15 países con peor calidad de vida

BIBLIOGRAFÍA

- [1] Fuente de datos. *World Population Data* (por Madhur Pant, actualizado el 23 de septiembre de 2022). <https://www.kaggle.com/datasets/madhurpant/world-population-data?select=iq.csv> (última visita realizada el 16/10/2022).
- [2] Criterio de clasificación de nivel de peso de adultos. *Acerca del índice corporal para adultos* (publicado por Centros para el Control y la Prevención de Enfermedades). https://www.cdc.gov/healthyweight/spanish/assessing/bmi/adult_bmi/index.html (última visita realizada el 17/10/2022).
- [3] Criterio de clasificación de categoría en función de iq. *¿Cuál es el coeficiente intelectual promedio de una persona?* (publicado por El Universo, El Mayor Diario Nacional). <https://www.eluniverso.com/larevista/sociedad/cual-es-el-coeficiente-intelectual-promedio-en-una-persona-nota/> (última visita realizada el 17/10/2022).