

World Happiness Report :)

Exploratory Data Analysis (EDA)



Realizado por:

Raquel Hernández Lozano
2025

01. INTRODUCCIÓN	2
02. HIPÓTESIS	3
03. EXPLORACIÓN DE LOS DATOS.....	4
04. ANÁLISIS GLOBAL	11
05. ANÁLISIS CON DATOS LIMPIOS Y NORMALIZADOS	12
06. CONCLUSIONES	15
07. STORYTELLING FINAL.....	16

01. INTRODUCCIÓN

El World Happiness Report es un informe anual que mide y compara la felicidad en más de 150 países, usando datos sobre bienestar subjetivo, salud, apoyo social, libertad y más. Nació en 2012 como parte de un movimiento global para incluir la felicidad como objetivo de políticas públicas.

A continuación, vamos a analizar los datos de los que disponemos para profundizar en el 'Happiness Score' 2023 a nivel mundial. Como fuentes de datos hemos tomado 2 datasets de Kaggle.

El primero world-data-2023.csv con muchos datos cada fila representa un país, y cada columna una variable. Hay más de 60 columnas, muchas de ellas numéricas, categóricas o porcentuales. Algunas están en formato texto con símbolos (\$, %, comas)

Incluye variables sobre geografía como *country*, *population* de salud como *life expectancy* o *Physicians per thousand*, economía *Unemployment rate* y sociedad como *Birth Rate*

El segundo whr2023.csv contiene menos columnas, pero tiene la principal que queremos analizar Happiness score, una variable que nos indica del 0 al 10 la puntuación con que califican los habitantes de cada país su nivel de felicidad. Además, aquí tenemos columnas que queremos analizar para ver su correlación o no con el nivel de felicidad tales como, Social support, Freedom to make choices o Healthy life expectancy.

La idea de unir 2 datasets fue por intentar enriquecer un poco el encontrado que incluía variables de felicidad con otros datos por país más relativos a salud, economía y sociedad.

En el documento notebook se detalla sobre la limpieza de los csv's y la unión de ellos con un merge para poder cruzar y analizar variables.

Tomando estos datos de base se ha hecho un profundo *exploratory data analysis*, EDA para poder sacar conclusiones sobre los factores que afectan positiva o negativamente a la felicidad de los diferentes habitantes del mundo por país.

02. HIPÓTESIS

Con tantísimos datos y variables acerca de tantos países al inicio surgieron muchísimas hipótesis como estas:

- ¿Los países con menor desempleo son más felices?
- ¿Qué es lo que más influye en la felicidad de un país?
- ¿Los países con mayor esperanza de vida tienden a ser más felices?
- ¿La percepción de corrupción está inversamente relacionada con la felicidad?
- ¿El acceso a médicos influyen positivamente en el nivel de felicidad?
- ¿La libertad para tomar decisiones personales predice niveles más altos de felicidad?

Incluso algunas más alguna más surgieron tras en el análisis y se contemplaron.

Finalmente llegamos a rechazar varias hipótesis nulas.

1. *"La percepción de corrupción no influye en la felicidad"*
2. *"El apoyo social no está relacionado con la felicidad"*
3. *"La región geográfica no influye en el nivel de felicidad"*
4. "El desempleo no afecta a la felicidad"
5. "La libertad para tomar decisiones no afecta al bienestar"
6. "La tasa de fertilidad no está relacionada con la felicidad"
7. "La esperanza de vida saludable no influye en la felicidad"

03. EXPLORACIÓN DE LOS DATOS

Al inicio cada csv se analizó por separado y se limpió poco a poco hasta quedarse con las columnas apropiadas, ya que de primeras teníamos demasiada información imposible de analizar y gestionar toda.

Whr2023.csv --> 137 rows × 21 columns

world-data-2023 --> 195 rows × 35 columns

Finalmente, el documento analizado y unido tiene 124 rows × 19 columns.

Ambos tenían la columna de país, hubo que renombrar para poder igualar "Country name": "Country"}, inplace=True y así poner hacer el merge inner.

Columnas interesantes se quedaron fuera del análisis ya que el tiempo y los recursos era limitados.

Fueron detectados datos nulos en varios países y fueron eliminados ya que no suponían países relevantes en el contexto que queríamos analizar. Se quedan fuera del EDA "North Macedonia", "Comoros" y "Uzbekistan".

Se procede a realizar el EDA sobre 124 países de todos los continentes.

Quitamos nulos en países no relevantes para poder usar las columnas numéricas North Macedonia", "Comoros" y "Uzbekistan

Hacemos zoom en las variables numéricas y sacamos las medidas de tendencia central del set de datos.

Centrándonos sobre todo en Happiness score donde observamos una media de 5.55, con valor máximo de 7.8 donde son más felices y 1.86 donde son más infelices

El mayor % de países se sitúan entre 4.79 y 6.33 puntos de felicidad.

Hay desigualdad global, algunos países están por debajo de 2, otros por encima de 7.

	count	mean	std	min	25%	50%	75%	max
Happiness score	127.0	5.55	1.14	1.86	4.79	5.7	6.32	7.8

Aquí se deja detallado el resto de las columnas con su definición.

Variables Definición

Country	Nombre del país
Land Area Km2	Superficie terrestre en kilómetros cuadrados
Co2-Emissions	Emisiones de dióxido de carbono (toneladas métricas)
CPI	Índice de precios al consumidor
Fertility Rate	Número promedio de hijos por mujer
Forested Area %	Porcentaje del territorio cubierto por bosques
Gasoline Price	Precio del litro de gasolina en dólares
Life expectancy	Esperanza de vida promedio en años
Physicians per thousar	Número de médicos por cada mil habitantes
Population	Población total del país
Total tax rate	Tasa total de impuestos (% sobre beneficios comerciales)
Unemployment rate	Tasa de desempleo (%)
Regional indicator	Región geográfica según el World Happiness Report
Happiness score	Puntuación media de felicidad (0-10)
Social support	Índice de apoyo social percibido (0-1)
Healty life expentancy	Años de vida saludable estimados (0-1)
Freedom to make choices	Índice de libertad para tomar decisiones personales
Generosity	Índice de generosidad (donaciones, ayuda a otros)
Perception of corruption	Índice de percepción de corrupción (valores altos = más corrupción percibida) (0-1)

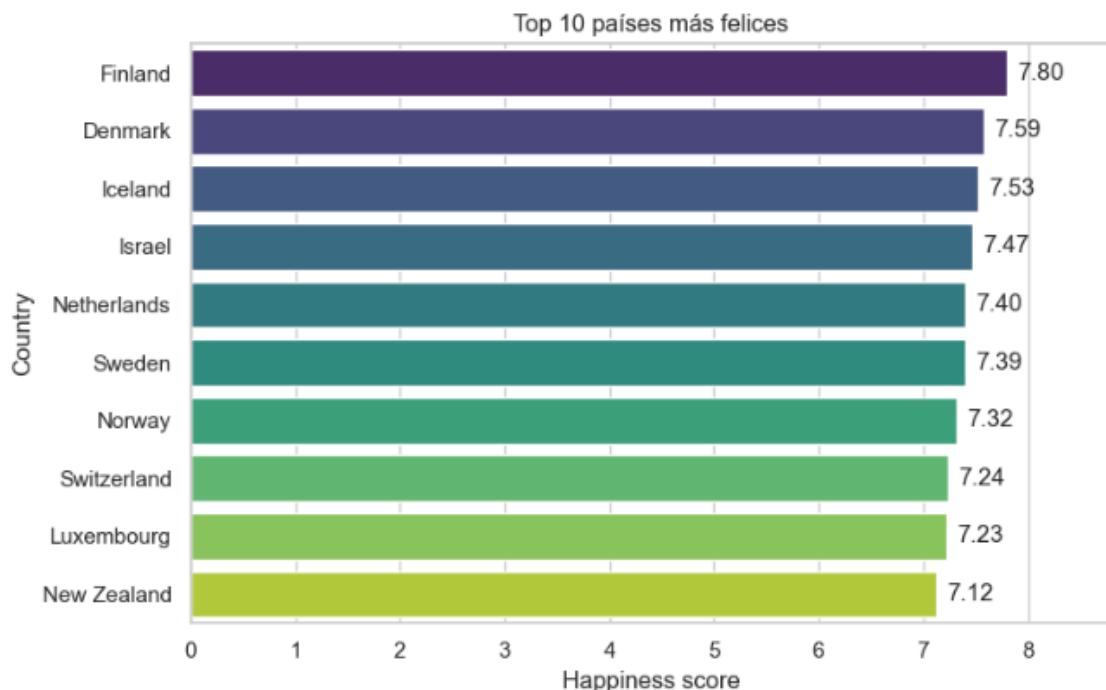
Se realizó una limpieza completa de los datos, eliminando símbolos como %, \$, y convirtiendo todas las variables numéricas. Se identificaron variables categóricas (región, país) y numéricas (PIB, felicidad, salud, corrupción). Posteriormente se calcularon estadísticas descriptivas básicas (media, mediana, moda) y se detectaron valores atípicos.

Inicialmente no se observan casi outliers significativos en los datos. Se detectaron valores atípicos en pequeña proporción sobre todo en Co2-Emissions con 14, Population con 14 también, Land Area(Km2) con 13 pero ninguno de estos es correlativo al happiness score como luego veremos.

Perceptions of corruption, tiene 9 outliers junto con Unemployment rate con 7. Ambas variables si están correlacionadas con el nivel de felicidad.

Analizando en profundidad se observa que los países con un bajo Perceptions of corruption son países con Happiness score muy alto (Ej: Finland) . Al igual que si nos fijamos en los outliers de la tasa de desempleo vemos que se asocian a niveles de felicidad significativamente bajos (Ej: Afghanistan). Esto sugiere que el desempleo podría estar relacionado con el bienestar percibido en estos países.

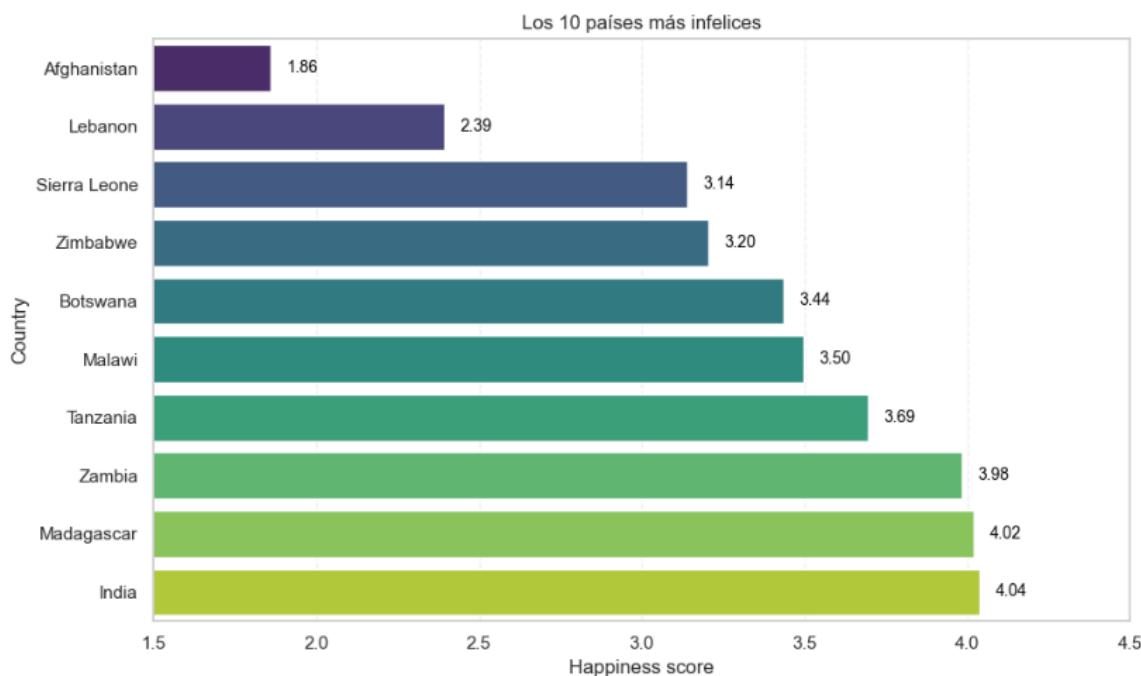
Análisis univariante y sus principales graficas



¿Qué tienen en común los países más felices?

- Sistema sanitario sólido, todos tienen altos valores en Physicians per thousand, lo que indica buen acceso a atención médica. Esto se refleja también en su alta Healthy life expectancy.
- Apoyo social fuerte, la mayoría tienen Social support alto y están en regiones con buena cohesión comunitaria.
- Libertad personal elevada, los valores de Freedom to make life choices son altos, lo que indica que las personas sienten autonomía y control sobre sus vidas.
- Baja corrupción percibida, la categoría Corrupcion_cat es baja en todos los casos, lo que sugiere confianza en las instituciones.

- Generosidad y cohesión social, aunque varía, la mayoría tienen valores positivos en Generosity, lo que refleja una cultura de colaboración y ayuda mutua.
- Ubicación geográfica y modelo social, todos menos 1 pertenecen a Europa, especialmente al norte y centro, donde predominan modelos de bienestar social, educación pública fuerte y políticas inclusivas.

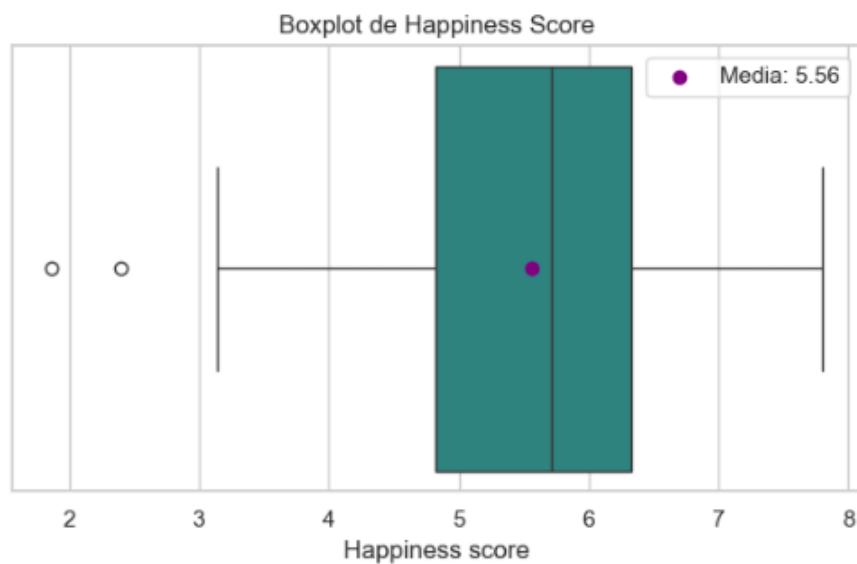
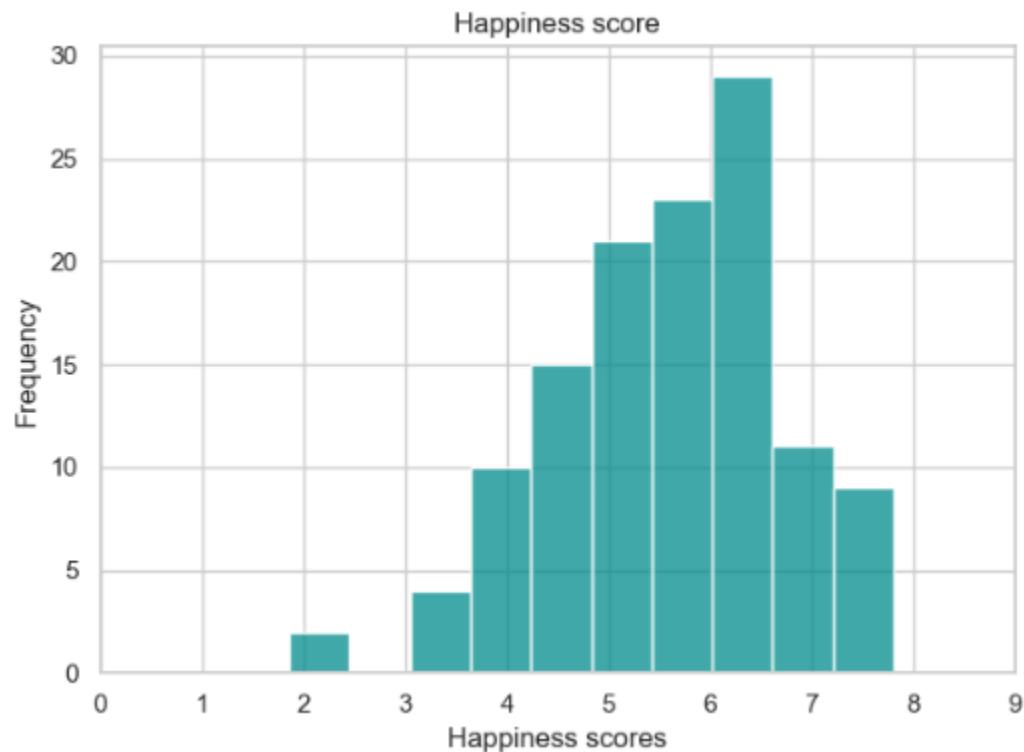


¿Qué tienen en común los países menos felices?

- Bajo apoyo social, la mayoría tienen valores bajos en lo que indica que las personas no sienten que pueden contar con otros en momentos difíciles.
- Libertad personal limitada, los valores suelen ser bajos, lo que refleja restricciones sociales, políticas o económicas que limitan la autonomía individual.
- Alta percepción de corrupción, muchos están en la categoría alta, lo que indica desconfianza en las instituciones y sensación de injusticia.
- Esperanza de vida reducida Los valores son bajos, lo que puede reflejar problemas en el sistema de salud, nutrición o seguridad.
- Recursos sanitarios escasos, tienen pocos médicos por cada mil habitantes, lo que limita el acceso a atención médica básica.
- Inestabilidad económica y social, algunos presentan altos niveles de desempleo, pobreza o conflictos internos, lo que afecta directamente el bienestar.

- Ubicación geográfica y contexto histórico, muchos pertenecen a regiones con desafíos estructurales como África Subsahariana o zonas afectadas por conflictos prolongados.

El nivel medio de felicidad se encuentra en la media 5.56

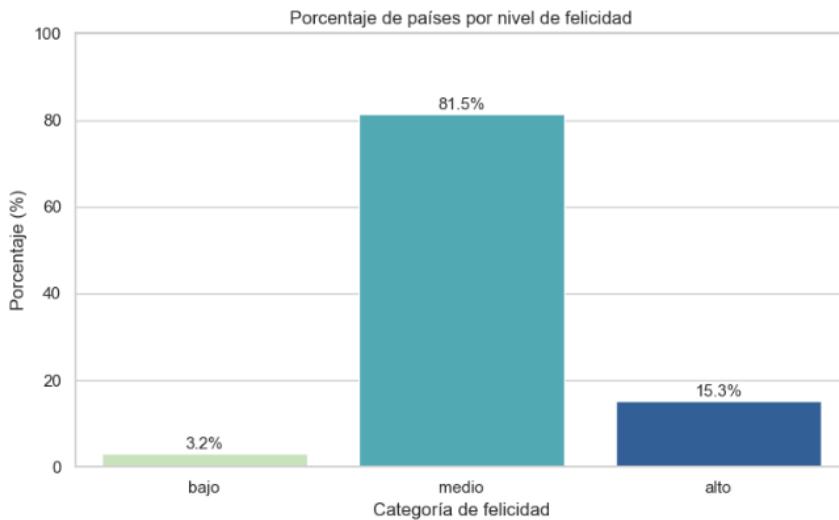
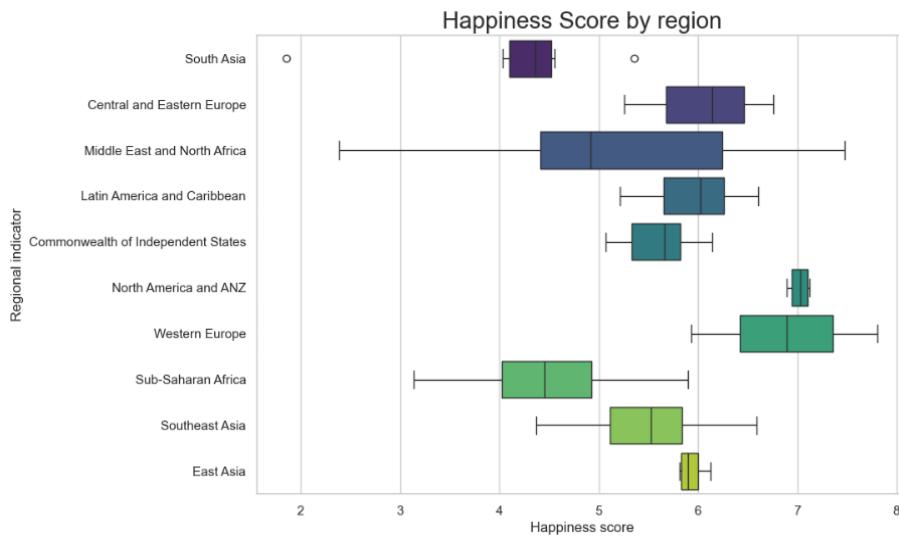


Para poder ver mejor la distribución de niveles de felicidad por países se ha convertido la variable numérica de Happiness Score en una variable categórica donde:

- Felicidad baja: 0.00 – 3.33
- Felicidad media: 3.34 – 6.66
- Felicidad alta: 6.67 – 10.00

Otras 2 variables numéricas fueron convertidas en categoricas para poder entender mejor los grafico con bajo, medio y alto, Social_cat, y Corrupcion_cat

Happiness score by region nos demuestra que la región si influye en el bienestar.



De los 124 países analizados, solo 19 presentan un nivel de felicidad alto (entre 6.67 y 10.00), lo que representa apenas el 15 % del total. Este grupo reducido será objeto de un análisis más profundo para identificar qué variables podrían estar influyendo en su bienestar.

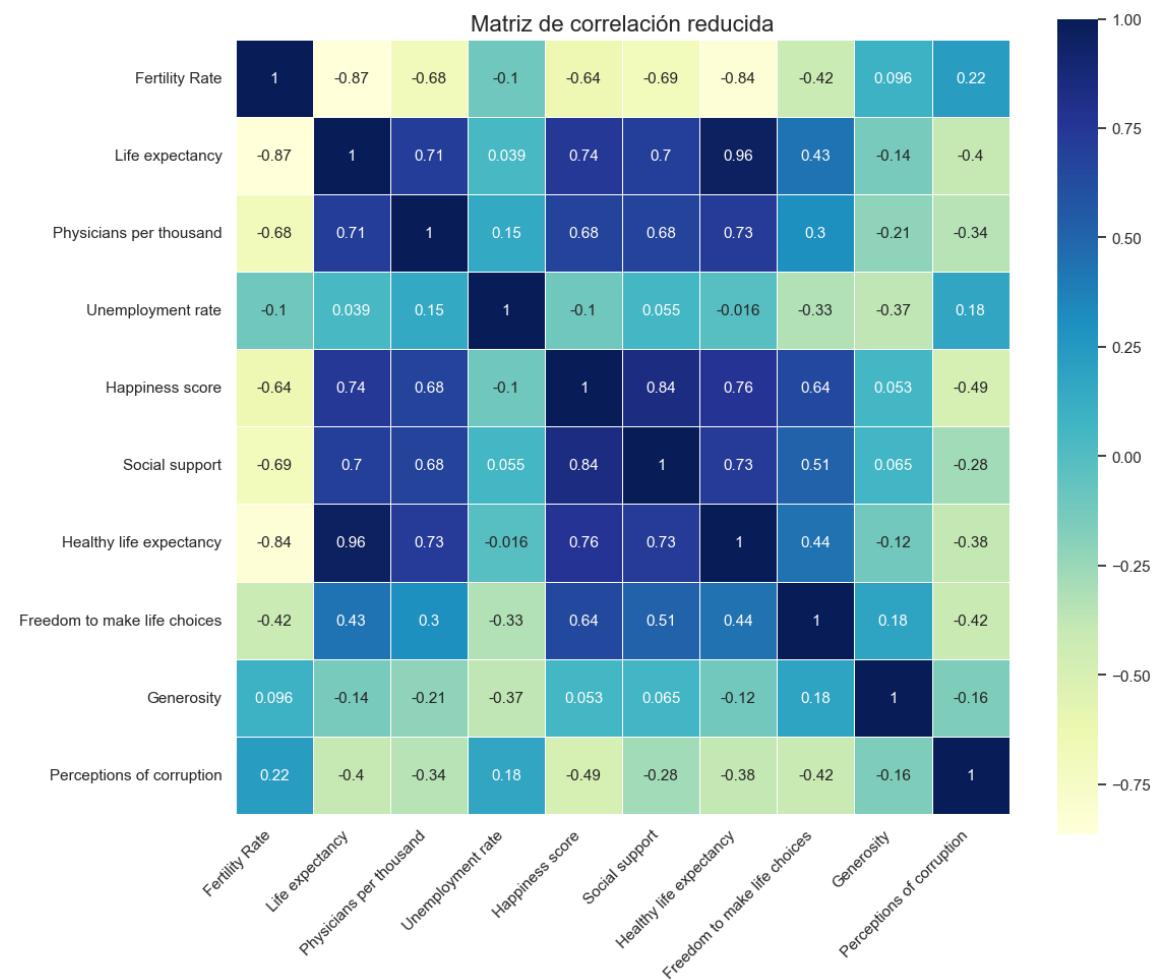
Como dato interesante para hacer zoom, descubrimos que nuestro país, **España**, se encuentra en la posición 29 de 124 en el ranking happiness score analizado.

Si nos comparamos con Finlandia, el país más feliz, vemos lo siguiente:

	Dimensión	España	Finlandia
0	Felicidad	medio	alto
1	Apoyo social	alto	alto
2	Corrupción percibida	alta	bajo

04. ANÁLISIS GLOBAL

Para poder observar las correlaciones entre las variables que influyen en la felicidad de manera global, mostramos la matriz de correlación reducida. Puesto que la 1 tenían demasiados valores que no mostraban correlación y se decidió limpiar para poder ver con más claridad



05. ANÁLISIS CON DATOS LIMPIOS Y NORMALIZADOS

Se aplicó una normalización para permitir comparaciones entre variables en distintas escalas. Las estadísticas descriptivas del Happiness Score tras la limpieza son las siguientes:

Media	5.54
Mediana	5.48
Desviación estándar	1.12
Mínimo	2.40
Máximo	7.80
Países analizados	124

Las correlaciones más destacadas en el análisis revelan que la felicidad está fuertemente asociada con la esperanza de vida saludable y el apoyo social, dos factores que muestran una relación positiva clara y consistente.

Por otro lado, variables como la tasa de fertilidad y el número de médicos por mil habitantes presentan relaciones más débiles, aunque siguen siendo relevantes para comprender el bienestar general. Estos hallazgos permiten enfocar el análisis en los factores con mayor impacto sobre la felicidad global.

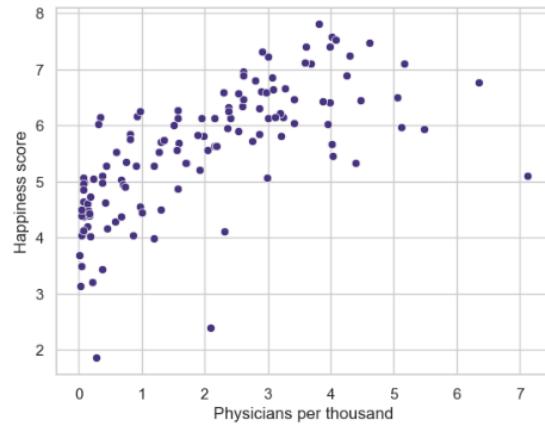
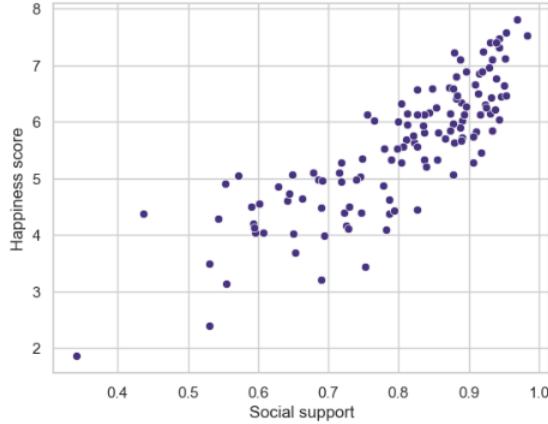
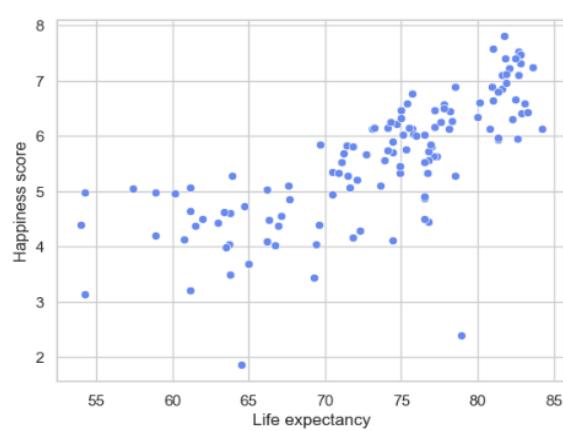
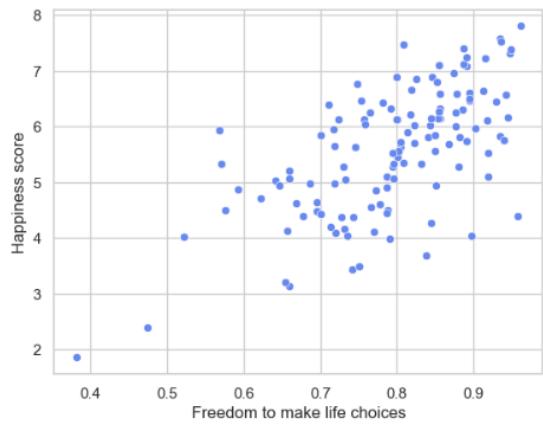
En contraste, la percepción de corrupción mantiene una correlación negativa moderada, lo que sugiere que en contextos donde se percibe mayor corrupción, los niveles de felicidad tienden a ser más bajos.

Correlaciones positivas y negativas contrastadas con chi-cuadrado y Pearson

Buscando p-valor cercanos a 0 para poder demostrar que son estadísticamente significativas.

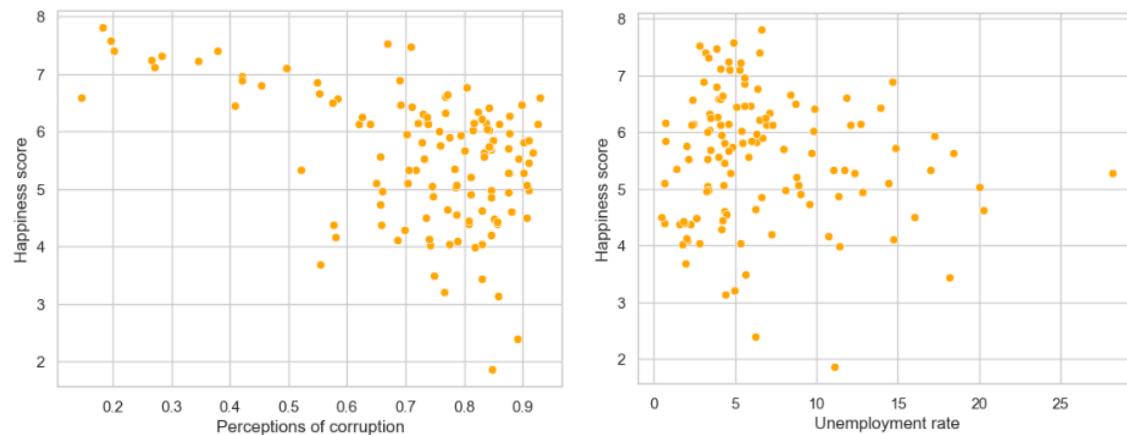
VARIABLES MÁS CORRELACIONADAS CON LA FELICIDAD (POSITIVAS)

Esperanza de vida saludable	+0.78
Apoyo social	+0.75
Libertad para decidir	+0.69
Médicos por mil habitantes	+0.52



VARIABLES MÁS CORRELACIONADAS CON LA FELICIDAD (NEGATIVAS)

Percepcion de la corrupcion	-0.51
Tasa de desempleo	-0.44
Tasa de fertilidad	-0.39



06. CONCLUSIONES

Los resultados confirman que el bienestar no depende únicamente de la riqueza material, sino de un equilibrio entre salud, apoyo social e instituciones confiables. La percepción de corrupción, por su parte, reduce significativamente los niveles de felicidad.

- +Más años de vida saludable se asocian con mayor felicidad.
 - +Sentir apoyo social está fuertemente ligado al bienestar.
 - +La autonomía personal contribuye significativamente a la felicidad.
 - +Más médicos por habitante refleja mejor acceso a salud.
-
- A mayor percepción de corrupción, menor felicidad. La confianza institucional es clave.
 - El desempleo afecta negativamente al bienestar general.
 - En este contexto global, los países con mayor fertilidad tienden a tener menor felicidad.
- Por otro lado, dejamos reflejados los puntos que no guardan correlación significativa con el happiness score analizado.
- El tamaño del país no guarda relación con la felicidad.
- Las emisiones totales no se asocian directamente con el bienestar.
- La proporción de área forestal no muestra impacto claro en la felicidad.
- La carga fiscal no parece influir directamente en el nivel de felicidad.
- El número de habitantes no está vinculado al Happiness score.

07. STORYTELLING FINAL

Si queremos construir sociedades más felices, debemos mirar más allá. Este análisis revela que la longevidad, el apoyo social y la confianza en las instituciones son pilares del bienestar. Países con acceso médico y baja corrupción tienden a ser más felices, independientemente de su riqueza. La felicidad no es un lujo, es un indicador de progreso.

En conjunto, estos resultados apoyan la idea de que la felicidad de un país no depende solo del dinero, sino también de la salud pública, el apoyo social y la confianza institucional.

El World Happiness Report, de donde parte toda la exploración de datos es mucho más que un ranking: es una herramienta global que revela cómo factores como la salud, el apoyo social y la libertad personal influyen en el bienestar. Desde 2012, ha guiado a gobiernos y ciudadanos hacia una visión más humana del progreso