

# INFORME EDA

## Empresa Farmacéutica

---

**PROYECTO:** Análisis de la Esperanza de Vida al Nacer para Factibilidad de lanzamiento de multivitamínico a nivel Global

---

## 1. Introducción

En el epicentro de la toma de decisiones informadas se encuentra el análisis exploratorio de datos (EDA), una práctica esencial que arroja luz sobre los matices y patrones ocultos dentro de conjuntos de datos vastos y complejos. Este informe se sumerge en un EDA exhaustivo, basado en ocho conjuntos de datos extraídos del Banco Mundial, cada uno de los cuales abarca factores cruciales que potencialmente inciden en la esperanza de vida.

Los datos, procedentes de una fuente confiable como el Banco Mundial, se han seleccionado meticulosamente para ofrecer una visión holística de los 44 países en estudio, distribuidos a lo largo de los 5 continentes. Estos conjuntos abarcan aspectos que van desde indicadores económicos y sociales hasta datos demográficos, permitiéndonos desentrañar las complejidades que subyacen a la variabilidad en la esperanza de vida.

A través de este EDA, no solo buscamos describir y visualizar la distribución de variables clave, sino también identificar relaciones y tendencias significativas entre los diversos factores. Al comprender la interconexión entre variables como el Producto Interno Bruto, la educación, la nutrición y otros indicadores relevantes, se espera revelar insights valiosos que orientarán la toma de decisiones estratégicas.

## Data sets

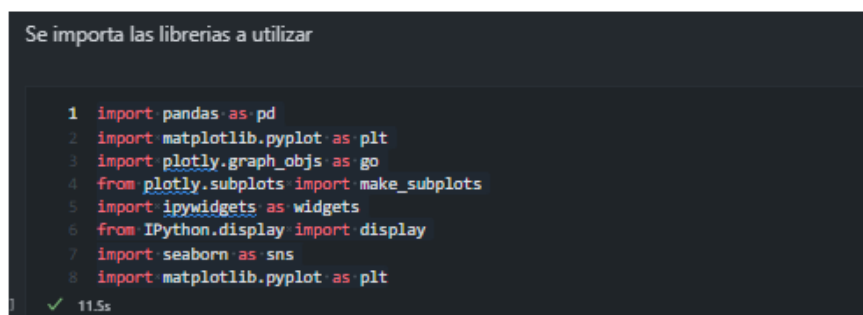
Los conjuntos de datos fueron descargados directamente desde la página oficial del Banco Mundial, accediendo específicamente al siguiente enlace:

<https://databank.worldbank.org/databases>

Este enlace proporciona acceso a una amplia variedad de series que se pueden filtrar según las necesidades específicas. Para el presente proyecto, se ha optado por descargar los factores más alineados con los objetivos establecidos, los cuales incluyen:

- Longevidad
- Población
- Economía
- Educación
- Inversión en Salud

Para analizar cada conjunto de datos, se importaron las siguientes librerías:



```
Se importa las librerías a utilizar

1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import plotly.graph_objs as go
4 from plotly.subplots import make_subplots
5 import ipywidgets as widgets
6 from IPython.display import display
7 import seaborn as sns
8 import matplotlib.pyplot as plt

✓ 11.5s
```

## Factor Educación

Este Dataset comprende un extenso período, desde 1987 hasta 2022, con un rango variable de observaciones, desde un mínimo de 3 a un máximo de 89. Las desviaciones estándar permanecen consistentes, entre 25 y 32, destacando la dispersión de los datos en cada año. La media de los indicadores se mantiene alrededor de 60-70, mostrando cierta estabilidad a lo largo de las décadas. Los valores mínimo y máximo varían notablemente, desde 0 hasta 12.60 como mínimo, y desde 99.86 hasta 129.95 como máximo, evidenciando la amplitud y diversidad de los datos a lo largo del tiempo. En resumen, el conjunto de datos proporciona un amplio espectro para analizar tendencias, cambios y distribución de indicadores específicos para múltiples países a lo largo de un amplio período temporal.

El conjunto de países en estudio refleja una diversidad geográfica significativa, representando un total de 44 naciones. Esta recopilación abarca una amplia distribución global, incluyendo países de distintos continentes, como Asia con naciones como China, India, y Japón; Europa con Alemania, España y Noruega; América con Estados Unidos, Argentina y Venezuela; África con Argelia, Nigeria y Rwanda; Oceanía con Australia, Fiji y Nueva Zelanda, entre otros. Esta diversidad geográfica sugiere una representación global en los datos, proporcionando la oportunidad de estudiar y comparar patrones, tendencias y características distintivas de cada continente en el análisis.

Un primer vistazo al conjunto de datos se obtuvo ejecutando `df.head()` en Pandas.

```
1 df.head()
```

	país	codigo_país	factor_(serie)	codigo_serie	1987	1988	1989	1990	1991	1992	...	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
0	Afganistán	AFG	Nivel de instrucción, al menos ciclo inferior ...	SESEC.CUAT.I.O.ZS	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	13.1258	NaN	NaN	NaN	NaN	NaN	11.6319	15.3429
1	Afganistán	AFG	Nivel de instrucción, al menos ciclo superior ...	SESEC.CUAT.IIP.ZS	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	9.7924	NaN	NaN	NaN	NaN	NaN	9.4613	12.5973
2	Afganistán	AFG	Nivel de instrucción, al menos escuela de nive...	SESEC.CUAT.PO.ZS	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.7049	NaN
3	Afganistán	AFG	Tasa de finalización del ciclo inferior de la ...	SESEC.CMPT.I.O.ZS	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	52.5213	54.6933	56.7369	59.7218	NaN	NaN	NaN	NaN
4	Alemania	DEU	Nivel de instrucción, al menos ciclo inferior ...	SESEC.CUAT.I.O.ZS	NaN	NaN	NaN	NaN	NaN	NaN	...	96.6718	96.7991	96.7161	96.5136	NaN	96.3323	NaN	95.6804	NaN	NaN

5 rows x 40 columns

Visualicemos los factores de la serie presentes en este conjunto de datos.

```
1 cantidad_series = len(df['factor_(serie)'].unique())
2 print('Cantidad de variables en estudio la categoría Educación:', cantidad_series)
3 serie_unicos = df['factor_(serie)'].unique()
4 print(serie_unicos)
```

Cantidad de variables en estudio la categoría Educación: 4

['Nivel de instrucción, al menos ciclo inferior de la escuela secundaria completo, población de más de 25 años, total (%) (acumulativo)'  
 'Nivel de instrucción, al menos ciclo superior de la escuela secundaria completo, población de más de 25 años, total (%) (acumulativo)'  
 'Nivel de instrucción, al menos escuela de nivel posterior a la secundaria completa, población de más de 25 años, total (%) (acumulativo)'  
 'Tasa de finalización del ciclo inferior de la educación secundaria, total (%) del grupo etario pertinente)']

Para facilitar la comprensión y análisis de los datos, se ha optado por reorganizar el dataframe de la siguiente manera:

	país	factor_(serie)	Año	Valor
0	Afganistán	Nivel de instrucción, al menos ciclo inferior ...	1987	NaN
1	Afganistán	Nivel de instrucción, al menos ciclo superior ...	1987	NaN
2	Afganistán	Nivel de instrucción, al menos escuela de nive...	1987	NaN
3	Afganistán	Tasa de finalización del ciclo inferior de la ...	1987	NaN
4	Alemania	Nivel de instrucción, al menos ciclo inferior ...	1987	NaN
...	...	...	...	...
6331	Kenya	Tasa de finalización del ciclo inferior de la ...	2022	NaN
6332	Noruega	Nivel de instrucción, al menos ciclo inferior ...	2022	NaN
6333	Noruega	Nivel de instrucción, al menos ciclo superior ...	2022	NaN
6334	Noruega	Nivel de instrucción, al menos escuela de nive...	2022	NaN
6335	Noruega	Tasa de finalización del ciclo inferior de la ...	2022	NaN

6336 rows x 4 columns

Aplicamos la función `.describe()` para obtener un resumen estadístico del conjunto de datos.

```
1 df_reorganizado['Valor'].describe()
```

count	1746.000000
mean	59.501718
std	29.032712
min	0.000000
25%	36.298250
50%	60.580000
75%	85.584600
max	129.953700
Name:	Valor, dtype: float64

Como se puede observar en el análisis anterior con un total de 1746 registros no nulos, se observa que la media de los valores es aproximadamente 59.50, con una dispersión significativa reflejada por una desviación estándar de alrededor de 29.03 unidades. El rango de valores es considerable, oscilando desde un mínimo de 0 hasta un máximo de 129.95. La mediana, situada en aproximadamente 60.58, indica que el 50% de los datos caen por debajo

de este valor, mientras que el otro 50% se encuentra por encima. Los percentiles muestran que el 25% de los datos son inferiores a 36.30 y el 75% son inferiores a 85.58, evidenciando la distribución de los valores en el conjunto de datos. Estos datos resumen la variabilidad, la centralidad y la amplitud de la información presente en la columna 'Valor'.

## Imputación

Ante la presencia significativa de valores faltantes (NaN) en nuestro conjunto de datos, se optó por aplicar una estrategia de imputación. Esta estrategia consiste en el reemplazo de los valores faltantes por la media de los datos, tomando en consideración la especificidad de cada país como también la del factor en estudio. Este enfoque nos permite abordar la ausencia de datos, asegurando una mayor integridad en nuestro conjunto de datos al evitar vacíos o valores faltantes, lo que resulta crucial para un análisis robusto y confiable.

```
1 def fill_missing_with_country_factor_mean(df):
2     means_by_country_factor = df.groupby(['país', 'factor_(serie)'])['Valor'].transform('mean')
3     df['Valor'] = df['Valor'].fillna(means_by_country_factor)
4     return df
5
6 # Llamada a la función para rellenar los valores faltantes con la media por país y factor
7 df_filled = fill_missing_with_country_factor_mean(df_reorganizado)
```

✓ 0.0s

1 df\_filled

	país	factor_(serie)	Año	Valor
0	Afganistán	Nivel de instrucción, al menos ciclo inferior ...	1987	13.366867
1	Afganistán	Nivel de instrucción, al menos ciclo superior ...	1987	10.617000
2	Afganistán	Nivel de instrucción, al menos escuela de nive...	1987	4.704900
3	Afganistán	Tasa de finalización del ciclo inferior de la ...	1987	43.790883
4	Alemania	Nivel de instrucción, al menos ciclo inferior ...	1987	96.720127
...	...	...	...	...
6331	Kenya	Tasa de finalización del ciclo inferior de la ...	2022	80.665067
6332	Noruega	Nivel de instrucción, al menos ciclo inferior ...	2022	99.546254
6333	Noruega	Nivel de instrucción, al menos ciclo superior ...	2022	74.675850
6334	Noruega	Nivel de instrucción, al menos escuela de nive...	2022	33.707636
6335	Noruega	Tasa de finalización del ciclo inferior de la ...	2022	99.388288

6336 rows x 4 columns

Volvemos aplicar la función `.describe()` para obtener un resumen estadístico del conjunto de datos, una vez realizado la imputación

```
1 df_filled['Valor'].describe()
```

✓ 0.0s

```
count    5868.000000
mean      52.076291
std       30.729631
min        0.000000
25%       25.728407
50%       50.626300
75%       78.337762
max      129.953700
Name: Valor, dtype: float64
```

Verificamos la presencia de datos Nan, resulta que hay países que no tienen datos en ciertos factores en ninguno de los años seleccionados.

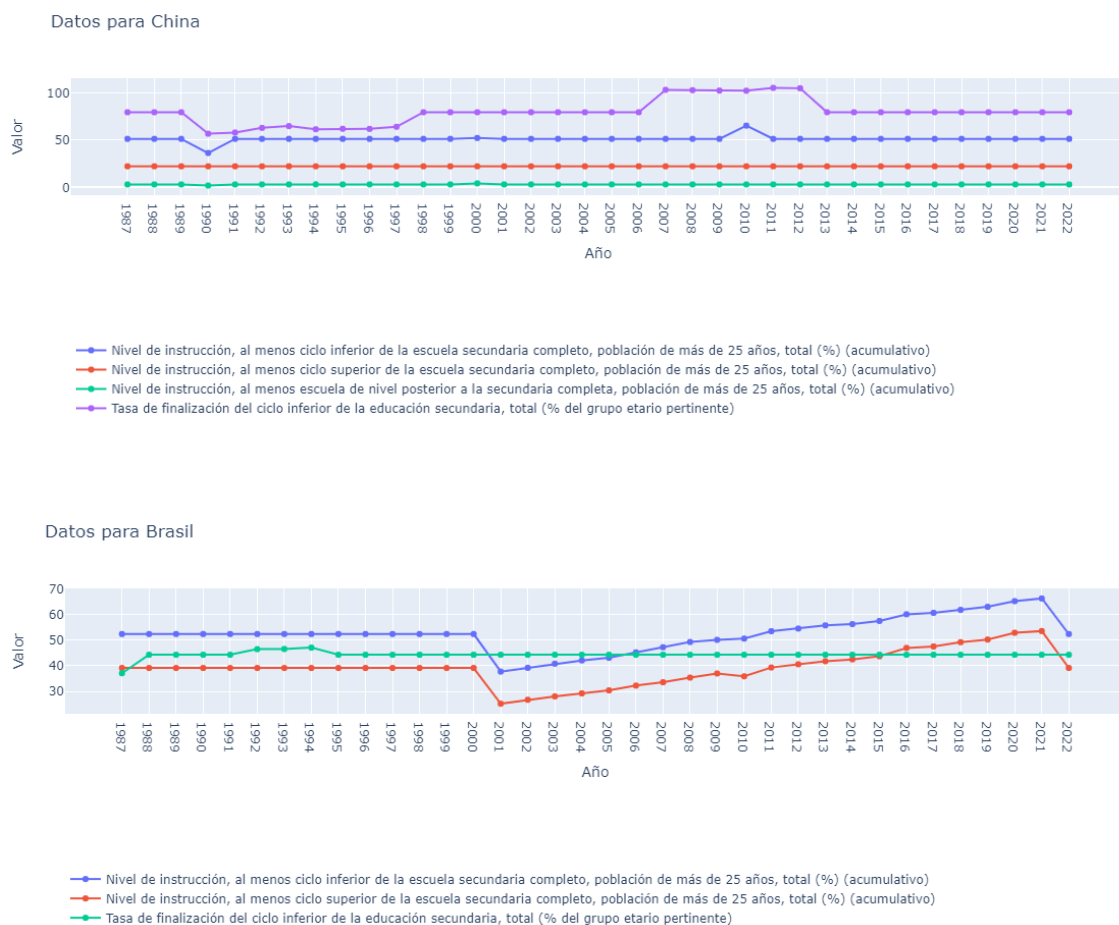
```
1 df_filled['Valor'].isna().sum()
✓ 0.0s
468
```

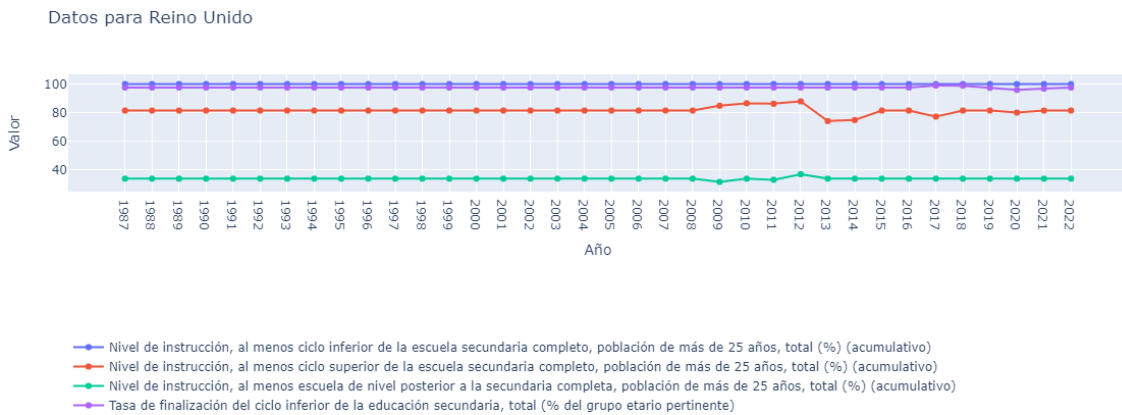
Por lo tanto, procedemos a eliminar los datos correspondientes a aquellos factores para los cuales no existen registros en ninguno de los años. Esta decisión se basa en evitar la generación de datos ficticios que podrían distorsionar la integridad del conjunto de datos, reconociendo las particularidades de cada país.

```
1 df_filled.dropna(inplace=True)
✓ 0.0s
```

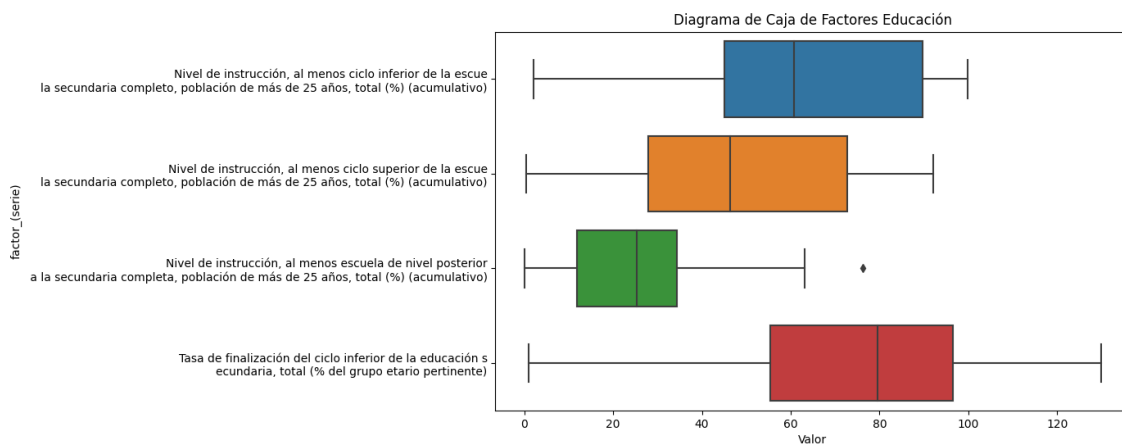
## Distribución

Ahora que contamos con la mayoría de los datos, procedemos a visualizar la distribución mediante gráficos. Se han seleccionado aleatoriamente algunos países para este análisis:





## Diagrama Boxplot del dataset de Educación



### 1.- Nivel de instrucción, al menos ciclo inferior de la escuela secundaria completo:

Este factor presenta un promedio del 60.47%, con una desviación estándar de 28.87%. El rango intercuartílico (IQR) abarca desde el 25% al 75%, con valores que oscilan entre el 45.16% y el 89.69%. El valor mínimo observado es 2.04%, mientras que el máximo alcanza el 99.86%.

### 2.- Nivel de instrucción, al menos ciclo superior de la escuela secundaria completo:

El nivel de instrucción, al menos ciclo superior de la escuela secundaria completo, tiene un promedio del 47.23%, con una desviación estándar de 26.09%. Su rango intercuartílico (IQR) va desde el 27.96% hasta el 72.74%. Los valores extremos registrados son 0.50% como mínimo y 92.18% como máximo.

### 3.- Nivel de instrucción, al menos escuela de nivel posterior a la secundaria completa:

Este factor presenta un promedio del 24.28%, con una desviación estándar de 15.85%. El rango intercuartílico (IQR) abarca desde el 25% al 75%, con valores que oscilan entre el 11.92% y el 34.38%. El valor mínimo registrado es 0%, mientras que el máximo alcanza el 76.29% siendo este último un outlier.

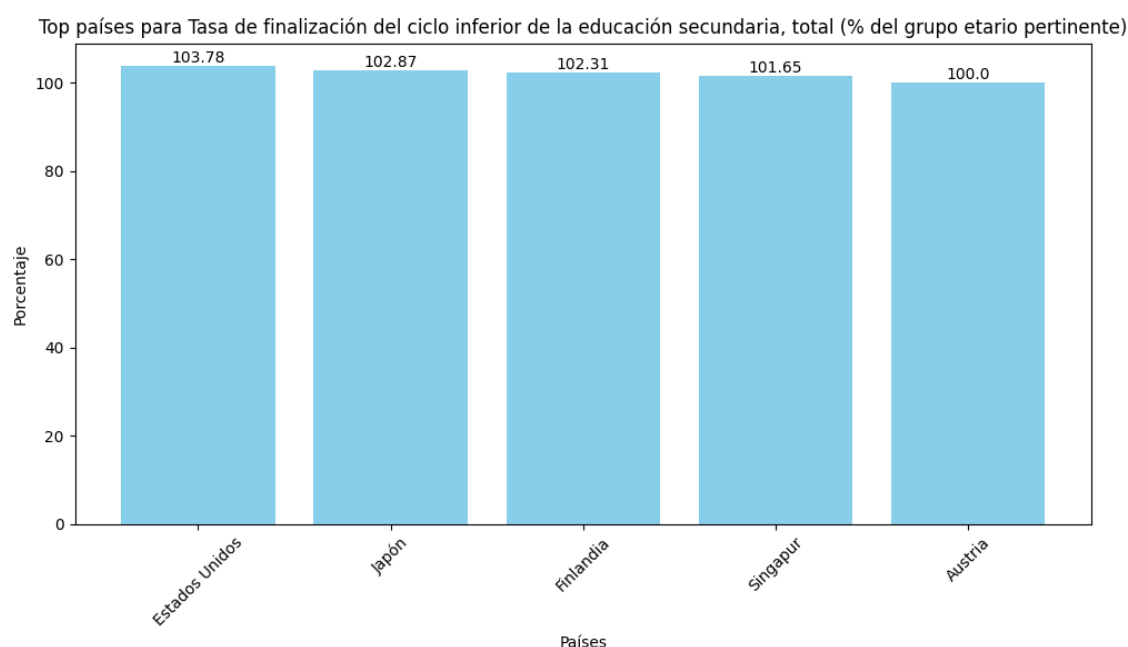
### 4.- Tasa de finalización del ciclo inferior de la educación secundaria:

La tasa de finalización del ciclo inferior de la educación secundaria exhibe un promedio del 72.75%, con una desviación estándar de 27.24%. El rango intercuartílico (IQR) comprende

desde el 25% al 75%, con valores que varían entre el 55.39% y el 96.55%. La cifra mínima registrada es 0.92%, mientras que la máxima alcanza el 129.95%.

## El mejor Factor de Educación y su top 5 países

El siguiente gráfico presenta los cinco países líderes en la Tasa de Finalización del Ciclo Inferior de la Educación Secundaria, expresada como un porcentaje del grupo etario correspondiente. Este indicador es crucial para evaluar el éxito del sistema educativo en proporcionar una educación completa a los estudiantes. En la cima de la lista se encuentra Estados Unidos con un impresionante 103.78%, seguido de cerca por Japón con 102.87%. Finlandia, reconocida por su excelencia educativa, ocupa el tercer lugar con un índice del 102.31%, seguida por Singapur con 101.65%. Cerrando la lista, Austria alcanza una tasa de finalización del 100.00%. Estos resultados destacan el compromiso y la eficacia de estos países en garantizar que sus jóvenes completen con éxito la etapa inicial de la educación secundaria.



## Factor Esperanza de vida

Al evaluar la esperanza de vida en el periodo continuo de 1987 a 2022, se observa una tendencia general al alza en los valores. Desde 1987, con un valor de aproximadamente 66.48, hasta el año 2022, donde se registra cerca de 75.11, se evidencia un crecimiento constante. La variabilidad, representada por la desviación estándar, fluctúa entre 7.98 y 13.06, sugiriendo una estabilidad relativa con ciertas variaciones a lo largo del tiempo. La presencia de valores extremos, con un rango desde 36.74 hasta 87.57, indica una amplia dispersión de datos.

Este análisis proporciona una visión general sobre la tendencia y dispersión de la serie temporal de la esperanza de vida en este extenso periodo. La identificación de valores extremos y la comprensión de la variabilidad ofrecen insights valiosos para comprender la evolución de esta variable a lo largo de los años. Este informe contribuye a la comprensión más



amplia de la dinámica de la esperanza de vida y su relevancia en el contexto de la salud pública y el bienestar social.

Un primer vistazo al conjunto de datos se obtuvo ejecutando `df.head()` en Pandas.

```
1 df.head()
```

	país	codigo_pais	factor_(serie)	codigo_serie	1987	1988	1989	1990	1991	1992	...	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
0	Afganistán	AFG	Esperanza de vida al nacer, total (años)	SP.DYN.LED0.IN	39.844	43.9580	45.1580	45.9670	46.6630	47.5960	...	62.4170	62.5450	62.6590	63.1360	63.0160	63.0810	63.5650	62.5750	61.982	NaN
1	Afganistán	AFG	Esperanza de vida al nacer, mujeres (años)	SP.DYN.LED0.FE.IN	43.453	46.5690	47.6100	48.3970	49.1440	50.3200	...	64.0270	64.2740	64.5760	65.0960	66.0990	66.4580	66.6770	65.4320	65.279	NaN
2	Afganistán	AFG	Esperanza de vida al nacer, varones (años)	SP.DYN.LED0.MA.IN	36.739	41.5720	42.8910	43.7090	44.3530	45.0700	...	60.7910	60.8120	60.7600	61.1930	60.1050	59.9230	60.6190	59.8660	58.915	NaN
3	Alemania	DEU	Esperanza de vida al nacer, total (años)	SP.DYN.LED0.IN	74.807	74.9987	75.1915	75.0856	75.3195	75.8195	...	80.4902	81.0902	80.6415	80.9902	80.9927	80.8927	81.2927	81.0415	80.901	NaN
4	Alemania	DEU	Esperanza de vida al nacer, mujeres (años)	SP.DYN.LED0.FE.IN	78.103	78.2860	78.4660	78.4190	78.7000	79.2000	...	83.0000	83.6000	83.1000	83.5000	83.4000	83.3000	83.7000	83.5000	83.380	NaN

5 rows × 40 columns

Visualicemos los factores de la serie presentes en este conjunto de datos.

```
1 cantidad_series = len(df['factor_(serie)'].unique())
2 print('Cantidad de variables estudio:', cantidad_series)
3 serie_unicos = df['factor_(serie)'].unique()
4 print(serie_unicos)

Cantidad de variables estudio: 3
['Esperanza de vida al nacer, total (años)'
 'Esperanza de vida al nacer, mujeres (años)'
 'Esperanza de vida al nacer, varones (años)']
```

Para facilitar la comprensión y análisis de los datos, se ha optado por reorganizar el dataframe de la siguiente manera:

	país	factor_(serie)	Año	Valor
0	Afganistán	Esperanza de vida al nacer, total (años)	1987	39.844
1	Afganistán	Esperanza de vida al nacer, mujeres (años)	1987	43.453
2	Afganistán	Esperanza de vida al nacer, varones (años)	1987	36.739
3	Alemania	Esperanza de vida al nacer, total (años)	1987	74.807
4	Alemania	Esperanza de vida al nacer, mujeres (años)	1987	78.103
...	...	...	...	...
4747	Kenya	Esperanza de vida al nacer, mujeres (años)	2022	NaN
4748	Kenya	Esperanza de vida al nacer, varones (años)	2022	NaN
4749	Noruega	Esperanza de vida al nacer, total (años)	2022	NaN
4750	Noruega	Esperanza de vida al nacer, mujeres (años)	2022	NaN
4751	Noruega	Esperanza de vida al nacer, varones (años)	2022	NaN

4752 rows × 4 columns

Aplicamos la función `.describe()` para obtener un resumen estadístico del conjunto de datos.

```
1 df_reorganizado['Valor'].describe()

count    4620.000000
mean      70.930790
std       10.244032
min       13.071000
25%       65.746750
50%       73.805500
75%       78.631700
max       87.710000
Name: Valor, dtype: float64
```

El resumen estadístico de la serie revela la presencia de 4620 observaciones. La media de los valores es de aproximadamente 70.93, con una desviación estándar de alrededor de 10.24, indicando una dispersión moderada de los datos respecto a la media. Los valores varían desde un mínimo de 13.07 hasta un máximo de 87.71. Los percentiles sugieren que el 25% de los valores son menores que 65.75, el 50% son inferiores a 73.81 y el 75% son menores que 78.63. Esto indica que la mayoría de los datos se concentran entre el percentil 25 y 75, con una dispersión más amplia hacia los valores máximos y mínimos.

## Imputación

Ante la presencia significativa de valores faltantes (NaN) en nuestro conjunto de datos, se optó por aplicar una estrategia de imputación. Esta estrategia consiste en el reemplazo de los valores faltantes por la media de los datos, tomando en consideración la especificidad de cada país como también la del factor en estudio. Este enfoque nos permite abordar la ausencia de datos, asegurando una mayor integridad en nuestro conjunto de datos al evitar vacíos o valores faltantes, lo que resulta crucial para un análisis robusto y confiable.

```
1 def fill_missing_with_country_factor_mean(df):
2     means_by_country_factor = df.groupby(['país', 'factor_(serie)'])['Valor'].transform('mean')
3     df['Valor'] = df['Valor'].fillna(means_by_country_factor)
4     return df
5
6 # Llamada a la función para rellenar los valores faltantes con la media por país y factor
7 df_filled = fill_missing_with_country_factor_mean(df_reorganizado)

✓ 0.0s

1 df_filled
✓ 0.0s

   país      factor_(serie)  Año  Valor
0  Afganistán  Esperanza de vida al nacer, total (años)  1987  39.844000
1  Afganistán  Esperanza de vida al nacer, mujeres (años)  1987  43.453000
2  Afganistán  Esperanza de vida al nacer, varones (años)  1987  36.739000
3  Alemania    Esperanza de vida al nacer, total (años)  1987  74.807000
4  Alemania    Esperanza de vida al nacer, mujeres (años)  1987  78.103000
...
4747  Kenya  Esperanza de vida al nacer, mujeres (años)  2022  60.417771
4748  Kenya  Esperanza de vida al nacer, varones (años)  2022  56.859286
4749  Noruega    Esperanza de vida al nacer, total (años)  2022  79.708663
4750  Noruega    Esperanza de vida al nacer, mujeres (años)  2022  82.224857
4751  Noruega    Esperanza de vida al nacer, varones (años)  2022  77.312286

4752 rows x 4 columns
```

Volvemos aplicar la función `.describe()` para obtener un resumen estadístico del conjunto de datos, una vez realizado la imputación

```
1 df_filled['Valor'].describe()

count    4752.000000
mean      70.930790
std       10.226149
min       13.071000
25%       65.736750
50%       73.809000
75%       78.600000
max       87.710000
Name: Valor, dtype: float64
```

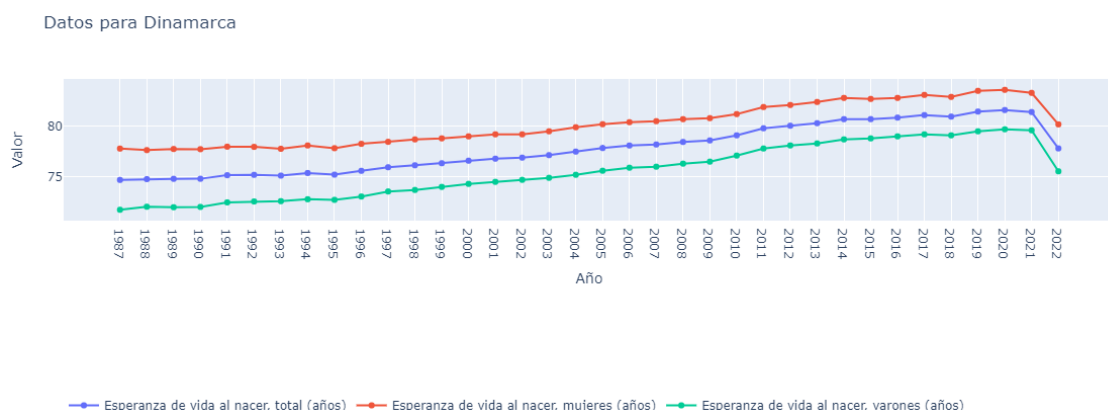
El resumen estadístico de la serie comprende un total de 4752 observaciones. Los valores muestran una media de aproximadamente 70.93 con una desviación estándar de alrededor de 10.22, lo que sugiere una dispersión moderada en relación a la media. Los valores oscilan entre un mínimo de 13.07 y un máximo de 87.71. Los percentiles muestran que el 25% de los datos son menores que 65.75, el 50% son inferiores a 73.8, y el 75% son menores que 78.64. Esto indica que la mayor parte de los datos se concentran entre los percentiles 25 y 75, mientras que hay una dispersión más amplia hacia los valores máximos y mínimos.

Verificamos la presencia de datos Nan, resulta que hay países que no tienen datos en ciertos factores en ninguno de los años seleccionados.

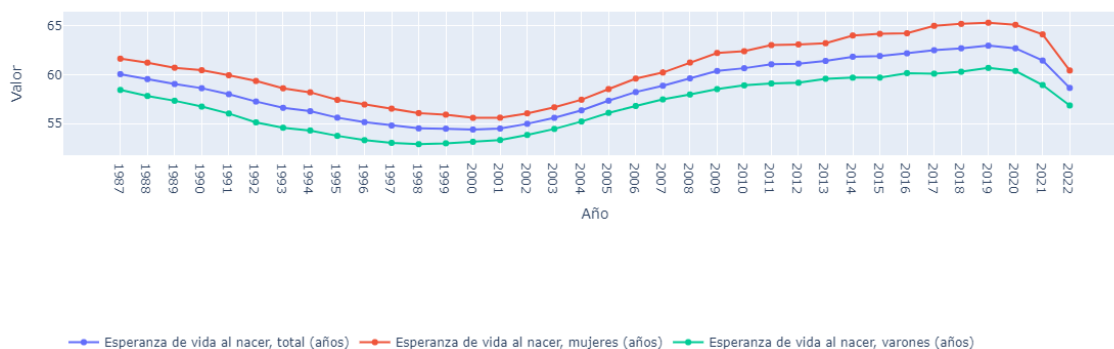
```
1 df_filled['Valor'].isna().sum()
✓ 0.0s
0
```

Se puede observar que no hay valores NaN en ninguno de los factores de este dataset de la esperanza de vida. Todos los datos asociados a cada factor están completos y disponibles para el análisis.

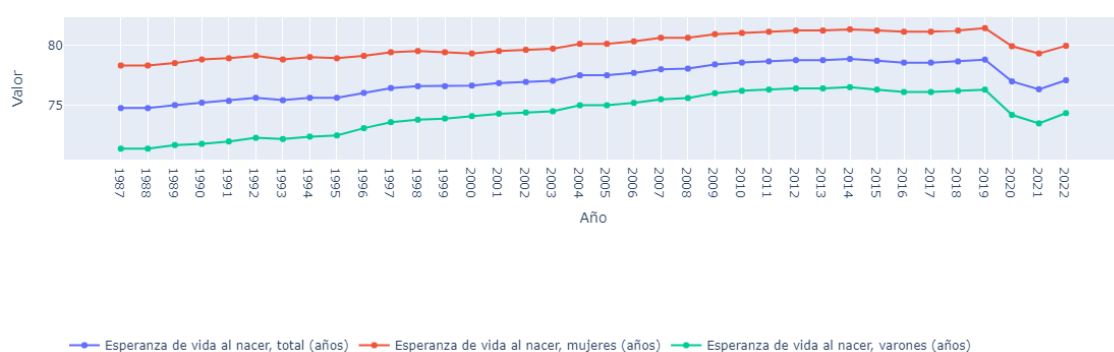
## Distribución



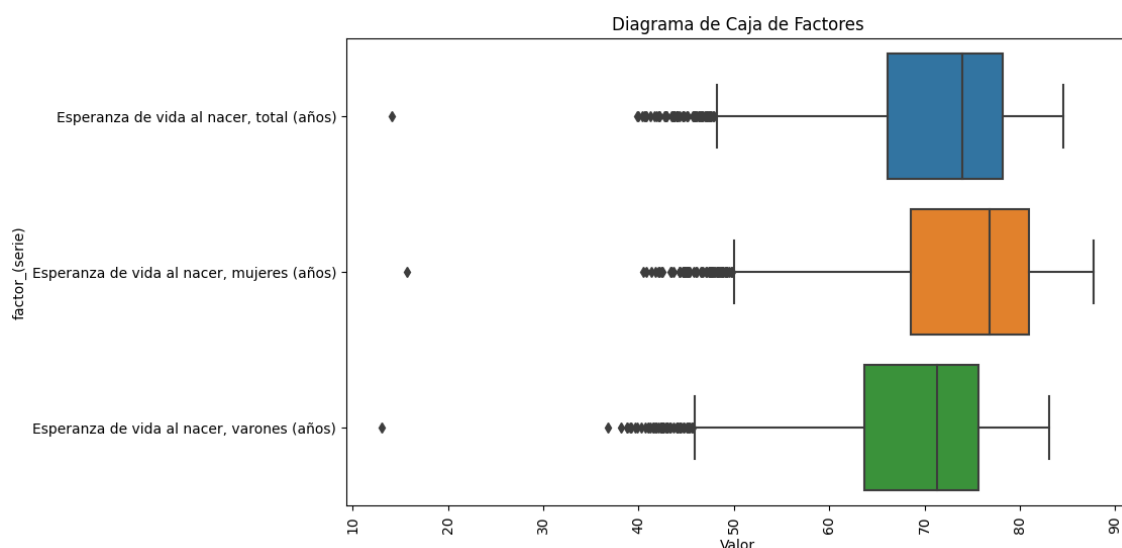
Datos para Kenya



Datos para Estados Unidos



## Diagrama Boxplot del dataset de Esperanza de vida



### 1.- Esperanza de vida al nacer, total:

Al observar el gráfico de caja para 'Esperanza de vida al nacer, total', se destacan características significativas en su distribución. El rango intercuartílico (IQR) abarca desde 66.18 hasta 78.20 años, con una mediana de 73.98 años. Además, se registran valores extremos que van desde

14.10 hasta 84.56 años. Esta variabilidad refleja diferencias sustanciales en la duración de la vida para este factor.

## 2.- Esperanza de vida al nacer, mujeres:

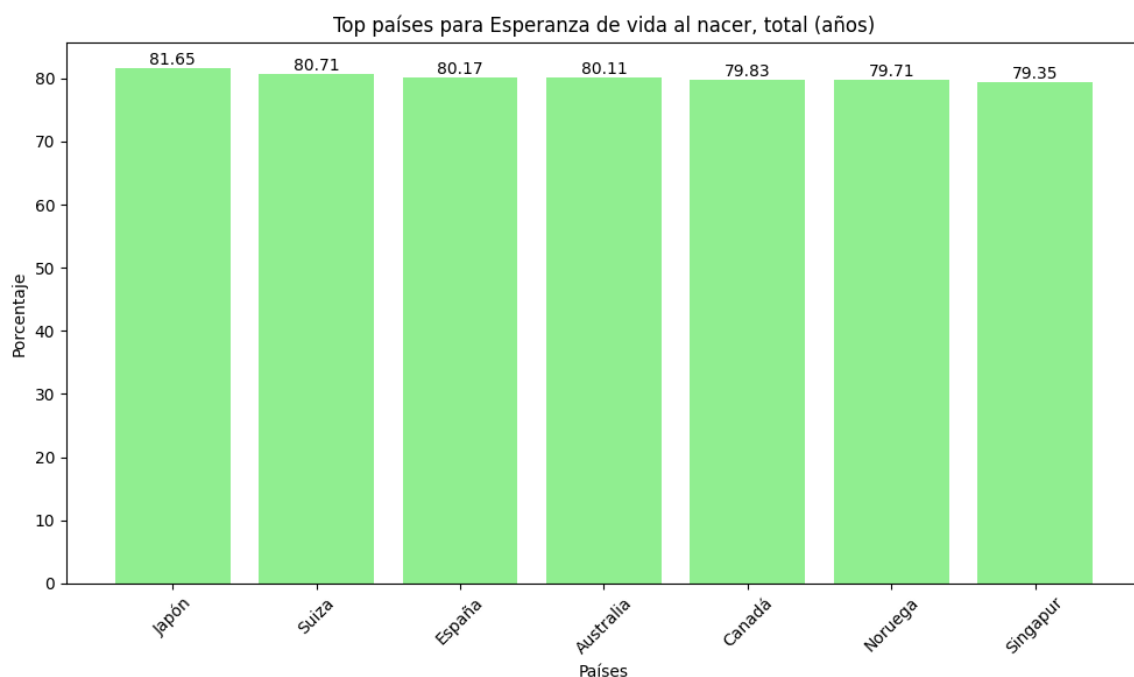
En cuanto a 'Esperanza de vida al nacer, mujeres', el IQR se extiende desde 68.57 hasta 81.00 años, con una mediana de 76.83 años. Se observan valores extremos entre 15.66 y 87.71 años. Estas diferencias en la distribución indican variaciones notables en la esperanza de vida al nacer, específicamente en el contexto femenino.

## 3.- Esperanza de vida al nacer, varones:

Por último, para 'Esperanza de vida al nacer, varones', el IQR abarca desde 63.73 hasta 75.63 años, con una mediana de 71.31 años. Los valores extremos se encuentran en el rango de 13.07 a 83.10 años. Estas particularidades revelan discrepancias en la distribución de la esperanza de vida al nacer entre los distintos grupos demográficos masculinos.

En el análisis de la esperanza de vida al nacer, es crucial destacar que el valor mínimo identificado como un outlier pertenece a Rwanda en el año 1994. Este año en particular fue testigo de un genocidio que tuvo consecuencias humanitarias devastadoras, con un impacto significativo en la población y las condiciones de vida en el país. Durante aproximadamente 100 días, se llevó a cabo un intento sistemático de exterminio que resultó en la pérdida masiva de vidas, estimada entre 800,000 y 1 millón de personas. Las consecuencias de este trágico evento se reflejan en la esperanza de vida registrada, subrayando la importancia de comprender el contexto histórico al analizar los datos demográficos.

## Top 7 países con el mejor índice de esperanza de vida al nacer.



Al analizar los datos del dataset de esperanza de vida, se revela un grupo de siete países que se destacan por tener los mejores valores en este indicador. Japón lidera la lista con una notable esperanza de vida de 81.65 años, seguido de cerca por Suiza con 80.71 años y España con 80.17 años. Australia y Canadá también muestran sólidos resultados, registrando cifras de 80.11 y 79.83 años respectivamente. Noruega y Singapur completan este grupo selecto con esperanzas de vida de 79.71 y 79.35 años, destacando una tendencia positiva en la longevidad de la población en estos países. Este análisis resalta la importancia de factores como el sistema de salud, condiciones socioeconómicas y calidad de vida en la determinación de la esperanza de vida, evidenciando patrones notables en estas naciones líderes.

## Factor Salud

El análisis integral de la serie temporal abarcando desde 1987 hasta 2022 presenta diversas tendencias y patrones en los valores de un factor específico. Antes del año 2000, la falta de datos limita la visión retrospectiva, pero a partir de 1999 hasta 2002, se observa una variabilidad significativa con una desviación estándar que fluctúa entre 568.99 y 588.21. Es crucial tener en cuenta que la ausencia de datos previos al 2000 se debe a la falta de estudios en ese período.

Posteriormente, entre 2004 y 2022, la serie temporal revela un aumento constante en la cantidad de observaciones hasta 2018, seguido de una abrupta disminución y la falta de datos en 2022. La media de los valores muestra una tendencia ascendente de 524.95 en 2004 a 914.20 en 2021, señalando un progreso en los valores a lo largo de los años. Sin embargo, la amplia variabilidad, expresada por una desviación estándar que varía entre 870.50 y 1508.19, indica fluctuaciones notables. El rango extenso, de 1.19 a 7717.88, subraya la diversidad de los datos, y los percentiles detallan la distribución. A pesar de la complejidad y la presencia de valores extremos, este análisis proporciona una visión completa de la evolución de la serie temporal desde 1987 hasta 2022, reconociendo las distintas tendencias y la falta de datos más recientes en 2022.

Un primer vistazo al conjunto de datos se obtuvo ejecutando `df.head()` en Pandas.

```
1 df.head()
```

	país	codigo_país	factor_(serie)	codigo_serie	1987	1988	1989	1990	1991	1992	...	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
0	Afganistán	AFG	Gasto nacional en salud del gobierno general p...	SH.XPD.GHED.PC.CD	NaN	NaN	NaN	NaN	NaN	NaN	...	2.8348	2.9856	3.1064	3.1222	3.4096	2.7733	2.4879	6.1311	2.6858	NaN
1	Afganistán	AFG	Gasto de bolsillo per cápita (US\$ actuales)	SH.XPD.OOPC.PC.CD	NaN	NaN	NaN	NaN	NaN	NaN	...	40.5669	43.9721	47.0753	46.7139	50.5048	54.3732	57.0254	60.0667	62.7903	NaN
2	Afganistán	AFG	Gasto interno privado en salud per cápita (US\$...	SH.XPD.PVTD.PC.CD	NaN	NaN	NaN	NaN	NaN	NaN	...	40.5669	43.9721	47.0753	46.7139	50.5058	54.4835	57.0254	61.2048	62.9002	NaN
3	Alemania	DEU	Gasto nacional en salud del gobierno general p...	SH.XPD.GHED.PC.CD	NaN	NaN	NaN	NaN	NaN	NaN	...	3897.2982	4087.5269	3559.0362	3665.8242	3929.0932	4251.8416	4227.0487	4651.7547	NaN	NaN
4	Alemania	DEU	Gasto de bolsillo per cápita (US\$ actuales)	SH.XPD.OOPC.PC.CD	NaN	NaN	NaN	NaN	NaN	NaN	...	679.1858	690.4245	605.2361	614.9713	655.0565	726.2255	738.2104	743.7962	NaN	NaN

5 rows x 40 columns

Visualicemos los factores de la serie presentes en este conjunto de datos.

```
1 cantidad_series = len(df['factor_(serie)'].unique())
2 print('Cantidad de variables estudio:', cantidad_series)
3 serie_unicos = df['factor_(serie)'].unique()
4 print(serie_unicos)

Cantidad de variables estudio: 3
['Gasto nacional en salud del gobierno general per cápita (US$ a precios actuales)'
'Gasto de bolsillo per cápita (US$ actuales)'
'Gasto interno privado en salud per cápita (US$ a precios actuales)']
```

Para facilitar la comprensión y análisis de los datos, se ha optado por reorganizar el dataframe de la siguiente manera:

	país	factor_(serie)	Año	Valor
0	Afganistán	Gasto nacional en salud del gobierno general p...	1987	NaN
1	Afganistán	Gasto de bolsillo per cápita (US\$ actuales)	1987	NaN
2	Afganistán	Gasto interno privado en salud per cápita (US\$...	1987	NaN
3	Alemania	Gasto nacional en salud del gobierno general p...	1987	NaN
4	Alemania	Gasto de bolsillo per cápita (US\$ actuales)	1987	NaN
...	...	...	...	...
4747	Kenya	Gasto de bolsillo per cápita (US\$ actuales)	2022	NaN
4748	Kenya	Gasto interno privado en salud per cápita (US\$...	2022	NaN
4749	Noruega	Gasto nacional en salud del gobierno general p...	2022	NaN
4750	Noruega	Gasto de bolsillo per cápita (US\$ actuales)	2022	NaN
4751	Noruega	Gasto interno privado en salud per cápita (US\$...	2022	NaN

4752 rows x 4 columns

## Imputación

Ante la presencia significativa de valores faltantes (NaN) en nuestro conjunto de datos, se optó por aplicar una estrategia de imputación. Esta estrategia consiste en el reemplazo de los valores faltantes por la media de los datos, tomando en consideración la especificidad de cada país como también la del factor en estudio. Este enfoque nos permite abordar la ausencia de datos, asegurando una mayor integridad en nuestro conjunto de datos al evitar vacíos o valores faltantes, lo que resulta crucial para un análisis robusto y confiable.

```
1 def fill_missing_with_country_factor_mean(df):
2     means_by_country_factor = df.groupby(['país', 'factor_(serie)'])['Valor'].transform('mean')
3     df['Valor'] = df['Valor'].fillna(means_by_country_factor)
4     return df
5
6 # Llamada a la función para rellenar los valores faltantes con la media por país y factor
7 df_filled = fill_missing_with_country_factor_mean(df_reorganizado)

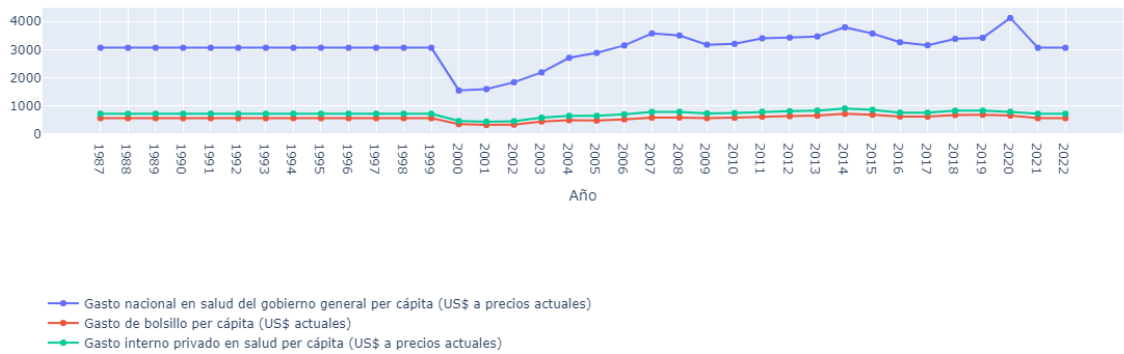
1 df_filled

   país  factor_(serie)  Año  Valor
0  Afganistán  Gasto nacional en salud del gobierno general p...  1987    2.463610
1  Afganistán      Gasto de bolsillo per cápita (US$ actuales)  1987   37.732205
2  Afganistán  Gasto interno privado en salud per cápita (US$...  1987   37.800170
3  Alemania    Gasto nacional en salud del gobierno general p...  1987  3312.234576
4  Alemania      Gasto de bolsillo per cápita (US$ actuales)  1987   581.818381
...    ...          ...    ...    ...
4747  Kenya      Gasto de bolsillo per cápita (US$ actuales)  2022   16.611381
4748  Kenya  Gasto interno privado en salud per cápita (US$...  2022   22.982057
4749  Noruega    Gasto nacional en salud del gobierno general p...  2022  5886.812005
4750  Noruega      Gasto de bolsillo per cápita (US$ actuales)  2022  1049.097077
4751  Noruega  Gasto interno privado en salud per cápita (US$...  2022  1070.962068

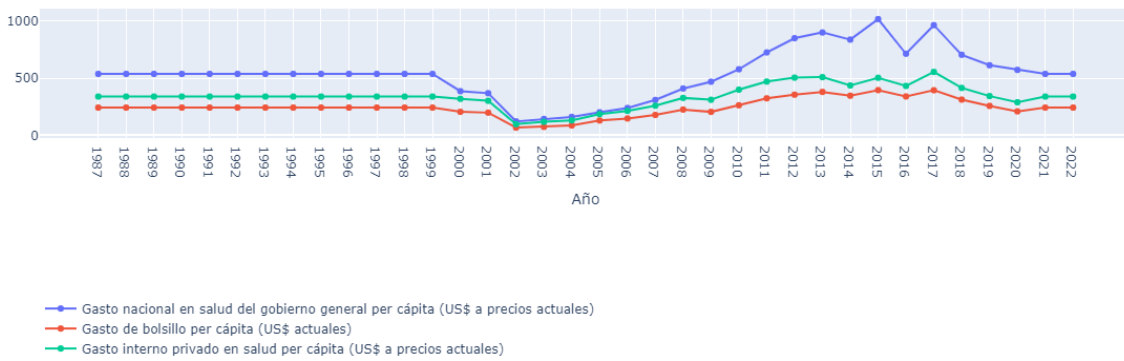
4752 rows x 4 columns
```

# Distribución

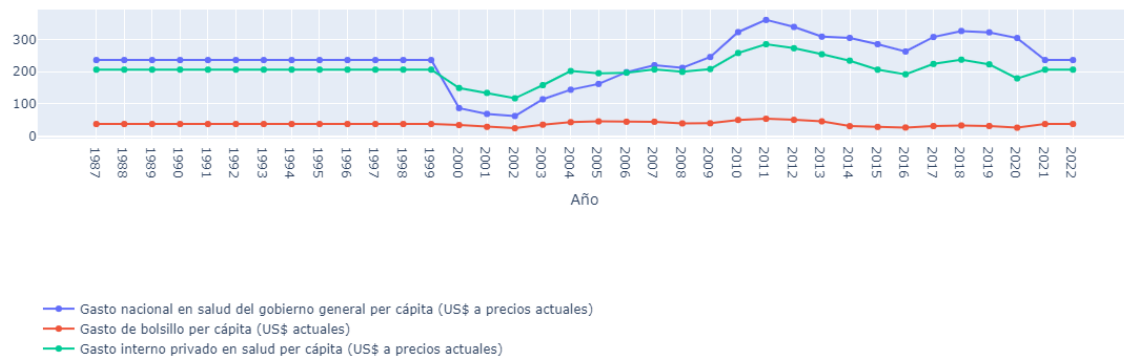
Datos para Reino Unido



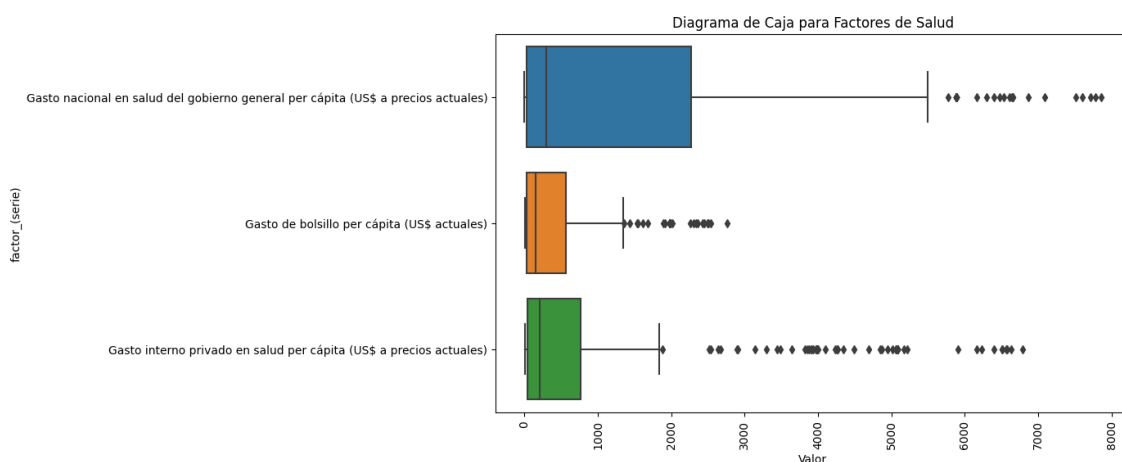
Datos para Argentina



Datos para Sudáfrica







### 1. Gasto nacional en salud del gobierno general per cápita (US\$ a precios actuales):

El diagrama de caja revela una amplia dispersión en los valores, destacando la presencia de outliers tanto en el extremo inferior como en el superior. El 25% de los datos se concentran en el rango de 33.51 a 2274.02, mientras que el 75% restante muestra una variación significativa, alcanzando hasta 7857.19. La presencia de valores extremadamente bajos y altos indica una marcada heterogeneidad en el gasto gubernamental en salud per cápita.

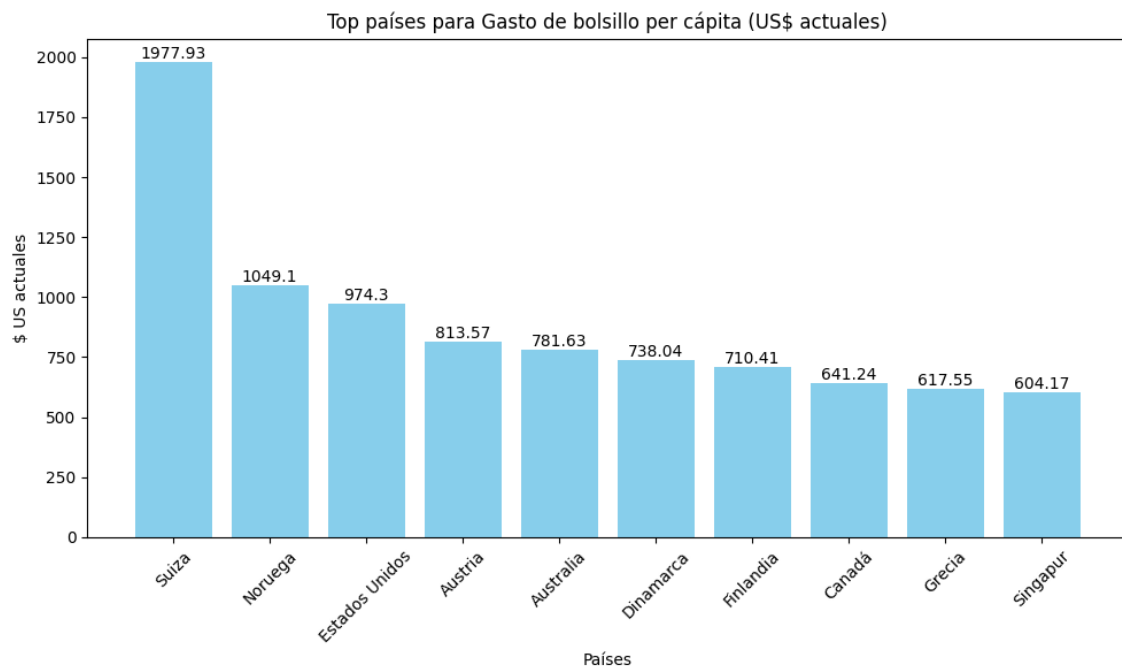
### 2. Gasto de bolsillo per cápita (US\$ actuales):

El diagrama de caja revela una distribución similarmente amplia en los datos. Se observa la presencia de outliers tanto en la parte inferior como en la superior del rango intercuartílico. El 25% de los datos se encuentra entre 33.92 y 560.93, mientras que el 75% restante muestra una variabilidad considerable, alcanzando hasta 2761.98. Los valores extremos sugieren una gran disparidad en los gastos de bolsillo per cápita.

### 3. Gasto interno privado en salud per cápita (US\$ a precios actuales):

En este caso, el diagrama de caja indica una distribución con una amplia variabilidad. La presencia de outliers tanto en la parte inferior como en la superior del rango intercuartílico destaca la heterogeneidad en los datos. El 25% de los valores se concentra entre 39.01 y 765.65, mientras que el 75% restante muestra una dispersión significativa, llegando hasta 6788.90. La existencia de valores atípicos resalta las diferencias notables en el gasto interno privado en salud per cápita.

En general, estos diagramas de caja sugieren que cada factor presenta una distribución muy dispersa, con la presencia de valores atípicos que indican heterogeneidad en los patrones de gasto en salud. Estos outliers deben considerarse al interpretar los datos, ya que representan situaciones excepcionales que pueden influir en el análisis general de cada factor.

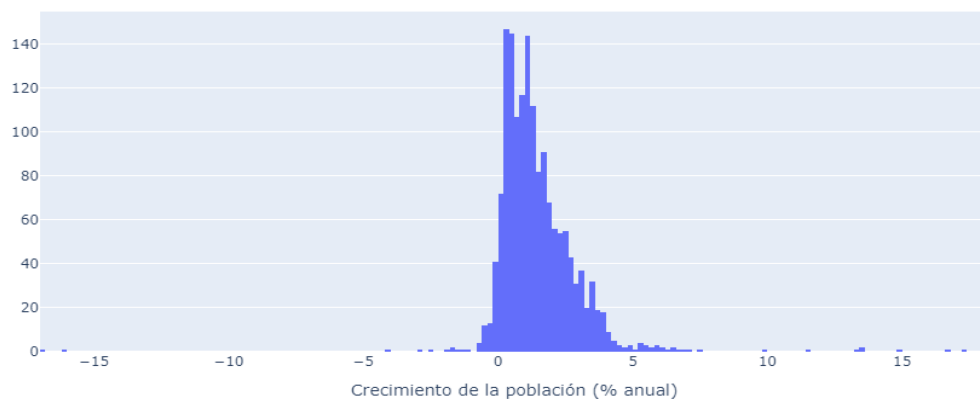


Al explorar los datos sobre el gasto de bolsillo en salud per cápita, se identifican diez países que sobresalen por sus notables cifras en este indicador. Suiza encabeza la lista con un marcado gasto de 1977.93 dólares estadounidenses por persona, seguida por Noruega con 1049.1 y Estados Unidos con 974.3. Austria, Dinamarca, Finlandia, Canadá, Grecia y Singapur también muestran niveles considerables de gasto individual en salud, reflejando diferentes enfoques en la financiación de la atención médica. Este análisis subraya la variabilidad en los patrones de gasto de bolsillo entre estos países, destacando la influencia de las políticas de salud y las preferencias individuales en la inversión personal en el bienestar médico.

## FACTOR POBLACIONAL

Dataset que cuenta con los siguientes indicadores:

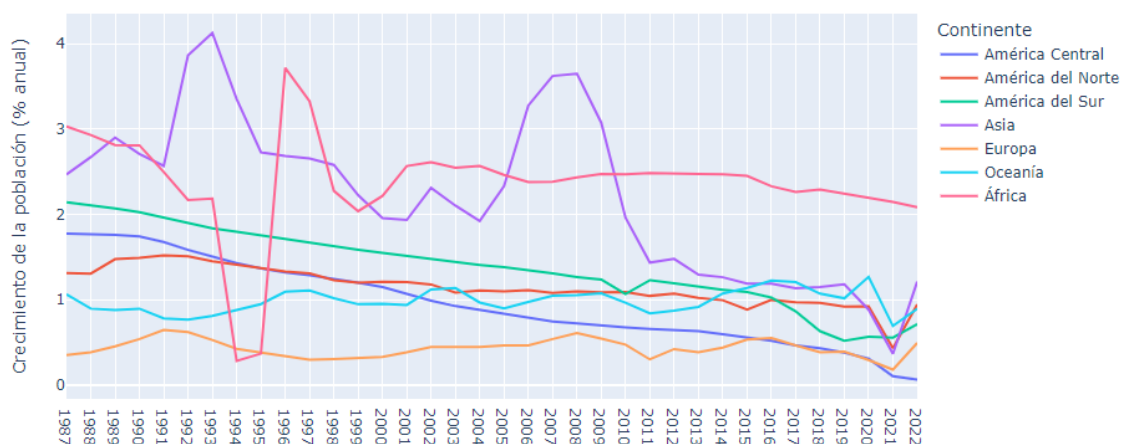
- **Crecimiento de la población**



Con respecto al indicador “Crecimiento de la población (% anual)” podemos observar que los valores se concentran entre 0 - 4%, también observamos valores muy alejados que van hasta -15%/15%”

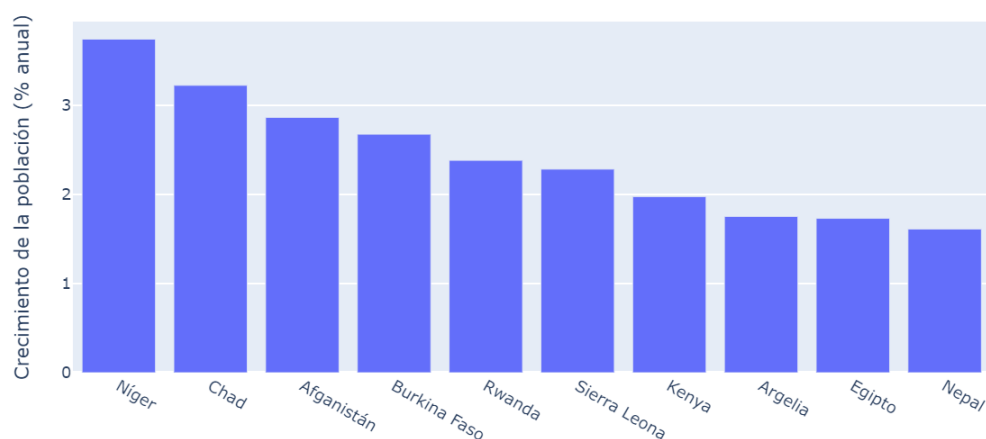
país Name	Año	Crecimiento de la población (% anual)
Rwanda	1994	-16.049153
Rwanda	1995	-16.880630

Rwanda sufrió entre los años 1994 - 1995 una disminución en su población de hasta 16% debido al genocidio ocurrido en el que se calcula murieron entre 500 000 - 1 000 000 personas.



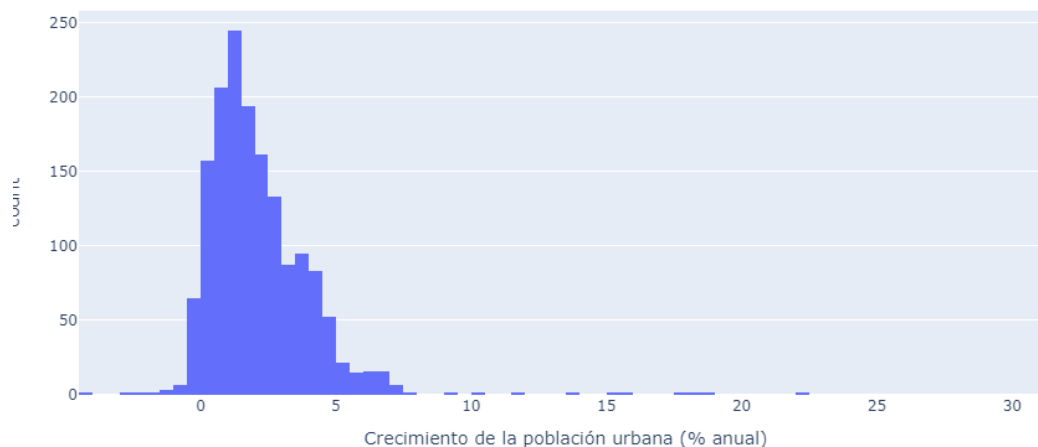
En promedio podemos ver que es el continente africano el que tiene un mayor crecimiento de la población (% anual)

TOP 10 países con mayor Crecimiento de la población (% anual) en promedio (2018-2022)



Países con mayor crecimiento promedio de la población (% anual), todos de África

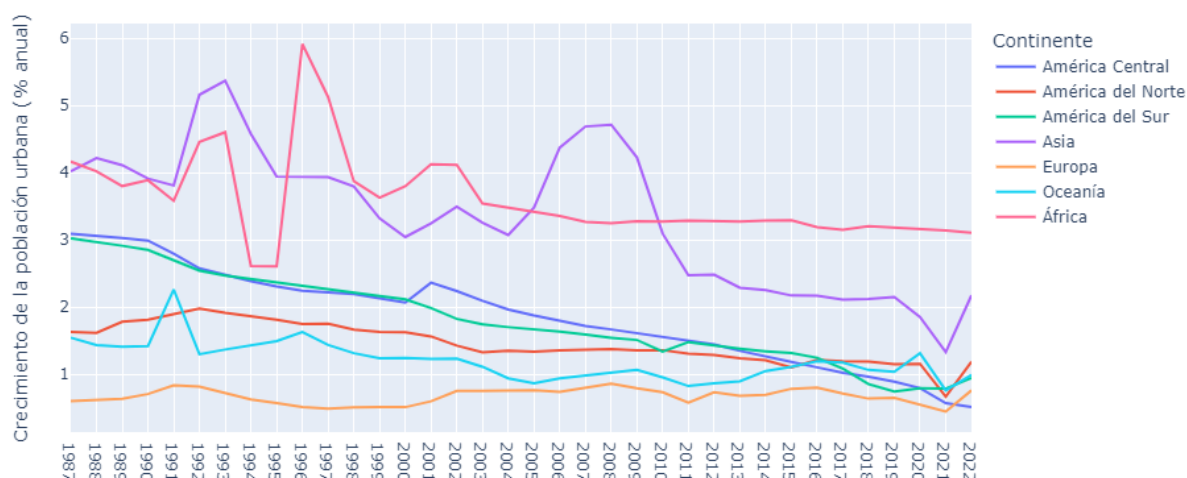
- **Crecimiento de la población urbana**



Con respecto al indicador “Crecimiento de la población urbana (% anual)” podemos observar que los valores se concentran entre 0 - 5%, podemos observar también que hay algunos datos con más del 20%

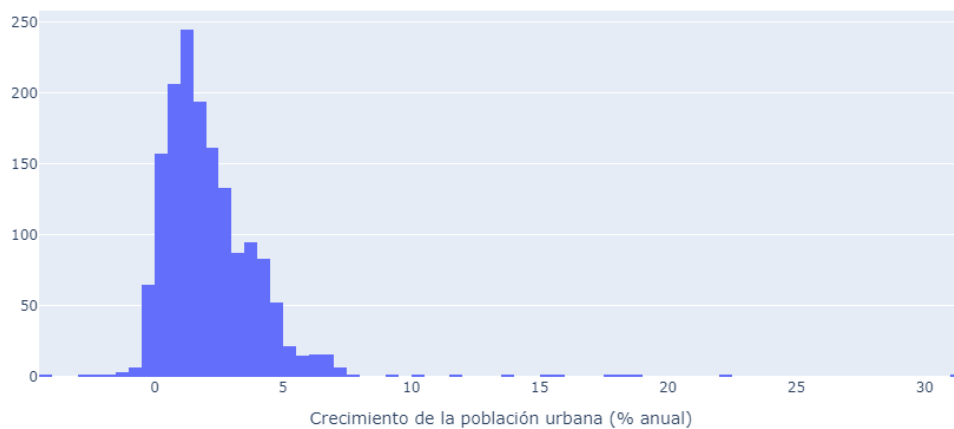
país Name	Año	Crecimiento de la población urbana (% anual)
Afganistán	1993	15.380336
Emiratos Árabes Unidos	2007	18.580685
Emiratos Árabes Unidos	2008	17.836557
Rwanda	1992	15.818698
Rwanda	1993	18.282681
Rwanda	1996	31.143425
Rwanda	1997	22.414590

Países con mayores registros de crecimiento de población urbana



En promedio podemos ver que es el continente africano el que tiene un mayor crecimiento de la población urbana (% anual)

- **Crecimiento de la población rural (% anual)**

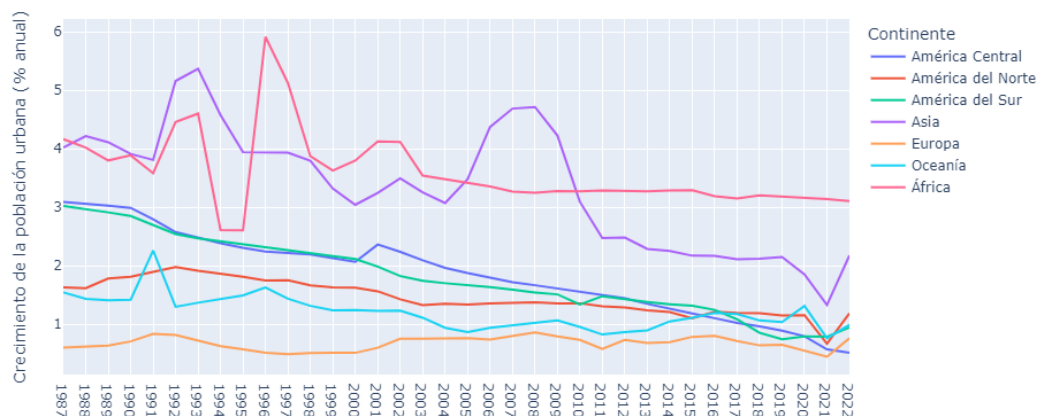


Con respecto al indicador “Crecimiento de la población rural (% anual)” podemos observar que los valores se concentran entre 0 - 4%, también observamos valores muy alejados mayores a 10%”

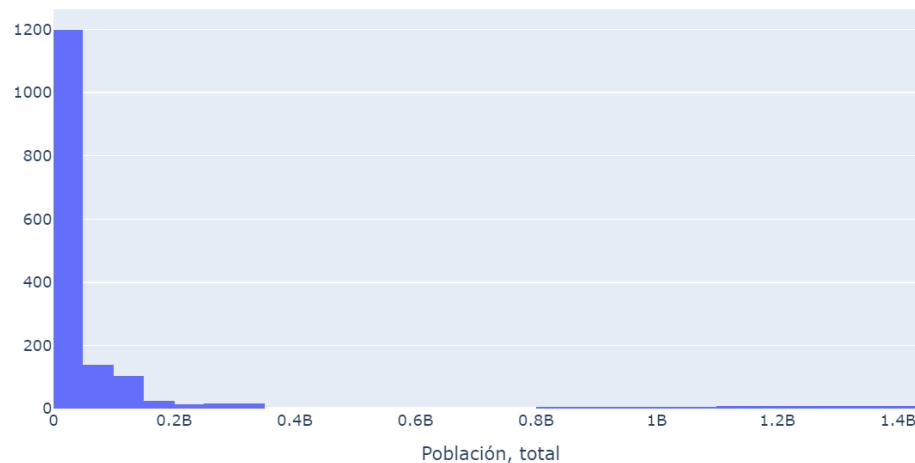
Países con mayores registros de crecimiento de población rural (% anual)

país Name	Año	Crecimiento de la población rural (% anual)
Afganistán	1993	14.851229
Emiratos Árabes Unidos	2007	15.943176
Emiratos Árabes Unidos	2008	15.231707
Rwanda	1996	14.906115

En promedio podemos ver que es el continente africano el que tiene un mayor crecimiento de la población rural (% anual).



- **Población total**

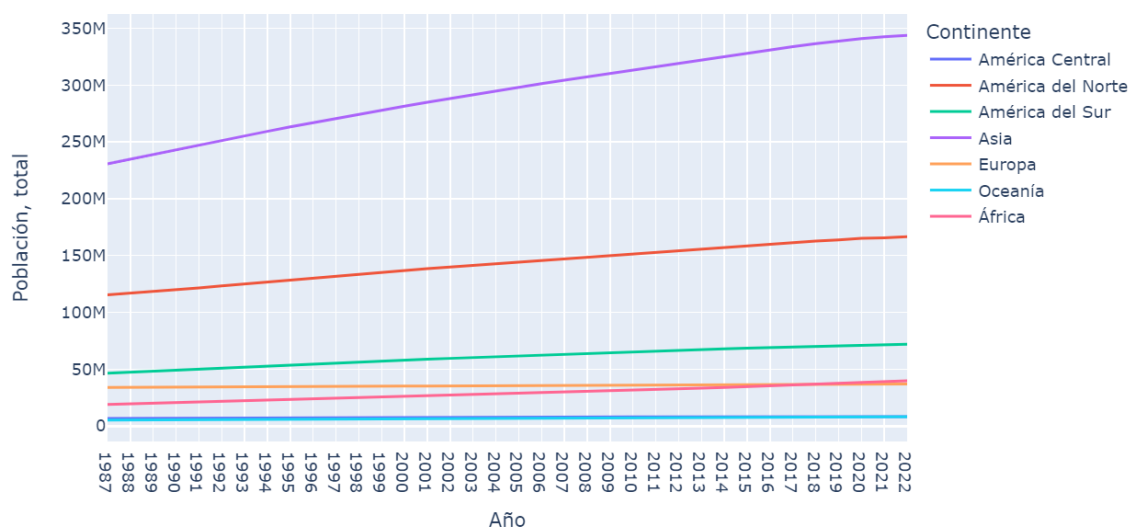


Podemos observar que la mayor parte de registros de “Población total” se concentran entre los valores de 0 - 50 M y podemos observar que hay registros muy alejados que superan los 800 M

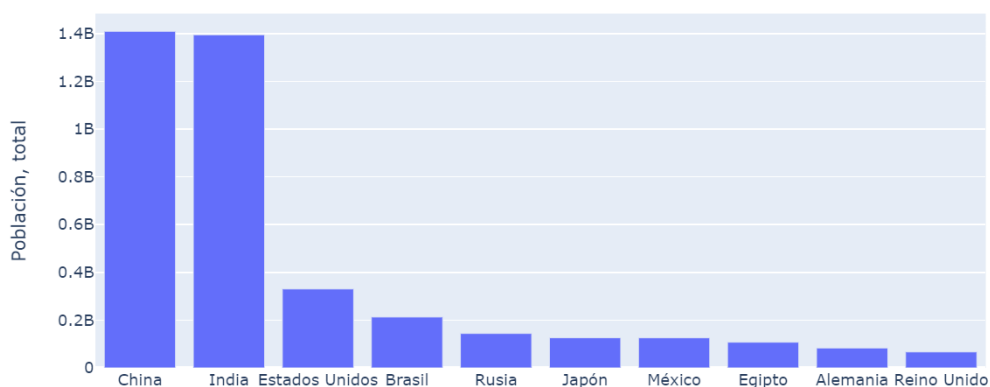
Países cuyos registros superan los 1000 M

país Name	Año	Población, total
China	2020	1.411100e+09
China	2021	1.412360e+09
China	2022	1.412175e+09
India	2020	1.396387e+09
India	2021	1.407564e+09
India	2022	1.417173e+09

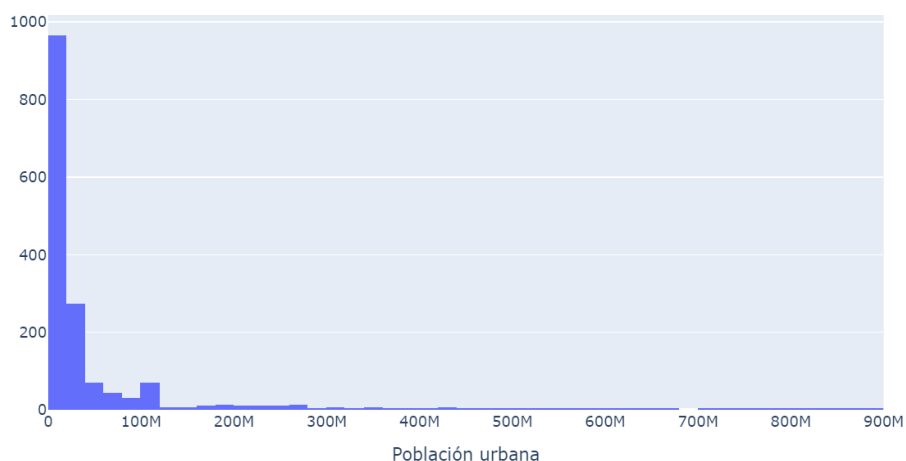
Podemos observar que el continente asiático es el que presenta mayor población seguido del continente de América del Norte



TOP 10 países con mayor Población en promedio (2018-2022)



- **Población urbana total**

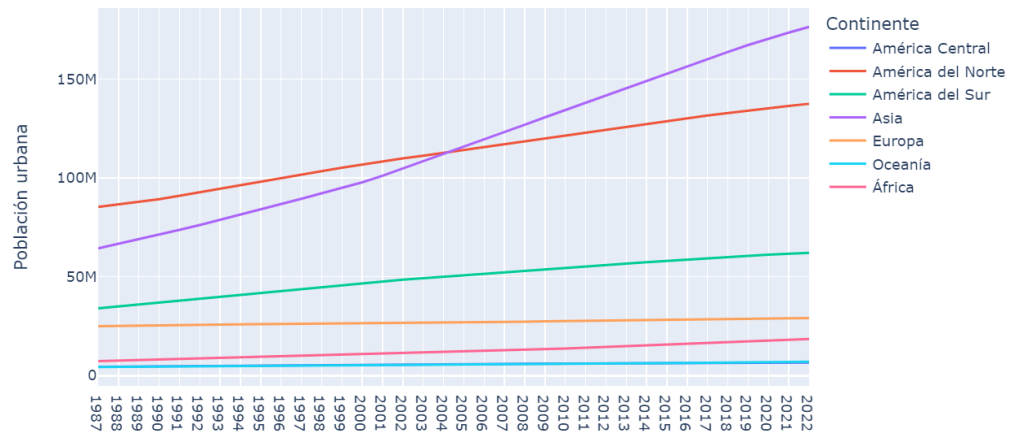


Podemos observar que la mayor parte de registros de “Población total urbana” se concentran en mayor medida entre los valores de 0 - 20 M y podemos observar que hay registros muy alejados que superan los 500 M

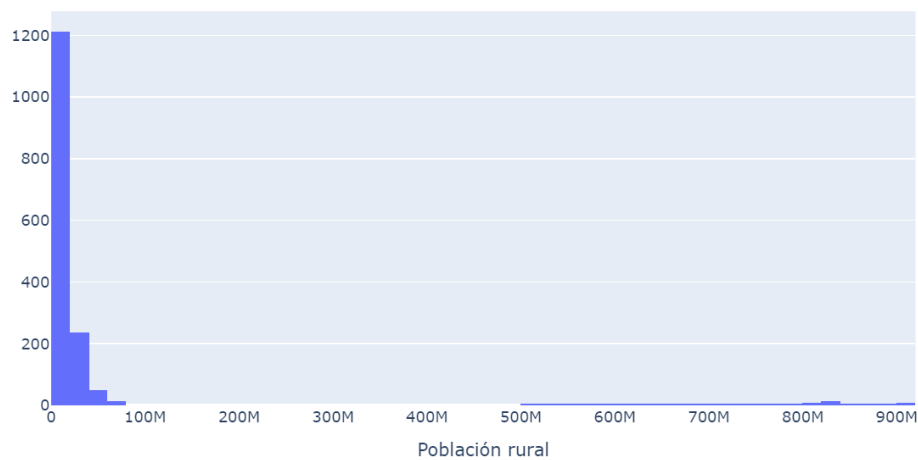
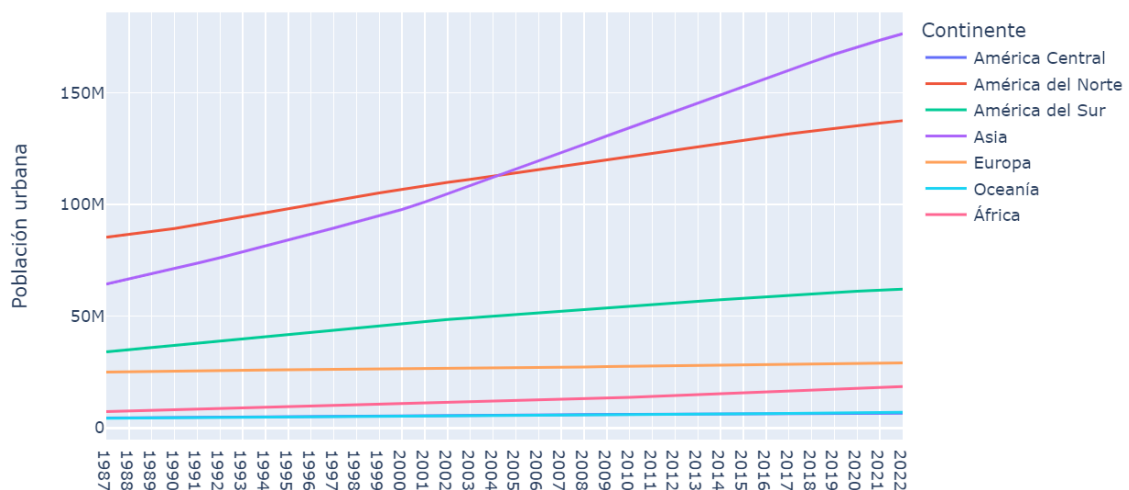
Registros de países que superan los 400 M

país Name	Año	Población urbana
China	2020	866810508.0
China	2021	882894483.0
China	2022	897578430.0
India	2020	487702168.0
India	2021	498179071.0
India	2022	508368361.0

Podemos observar que el continente asiático es el que presenta mayor población urbana seguido del continente de América del Norte



- **Población rural total**



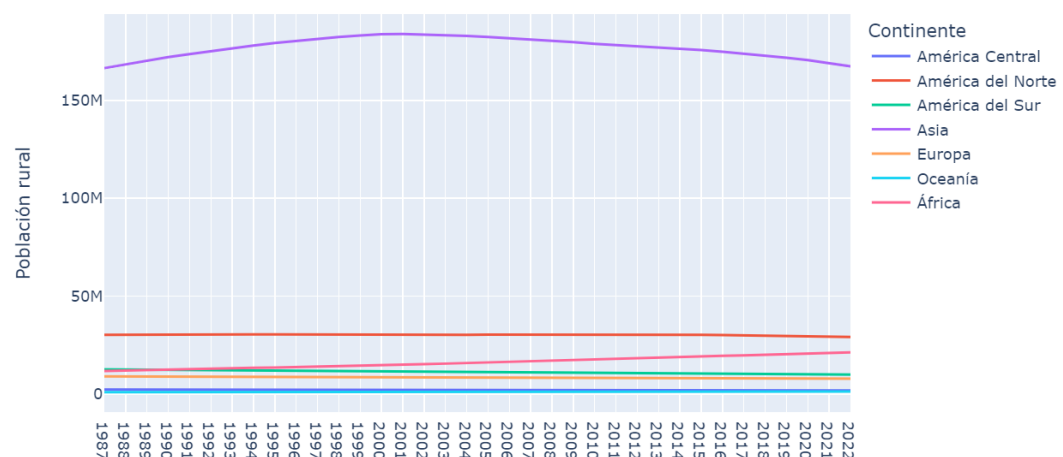


Podemos observar que la mayor parte de registros de “Población rural total” se concentran entre los valores de 0 - 20 M y podemos observar que hay registros muy alejados que superan los 500 M

Países con registros mayores a 500 M

país Name	Año	Población rural
China	2020	544289492.0
China	2021	529465517.0
China	2022	514596570.0
India	2022	908804812.0

Podemos observar que el continente asiático es el que presenta mayor población rural seguido del continente de América del Norte



## FACTOR LONGEVIDAD

Dataset que cuenta con los siguientes indicadores:

- **Población de 65 años de edad y más, total**

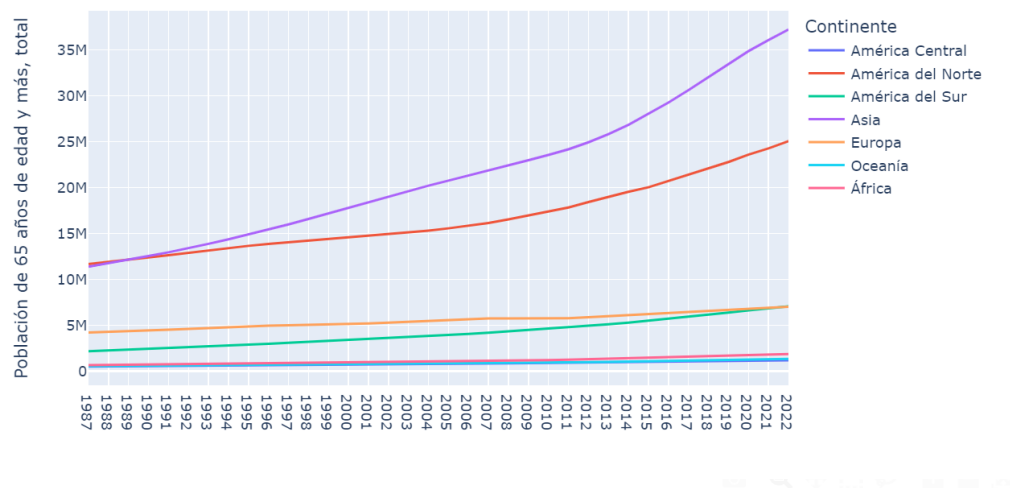


Con respecto al indicador “Población de 65 años de edad y más, total” podemos observar que los valores se concentran entre 0 - 5 M, podemos observar también que hay algunos datos muy alejados de más de 100 M.

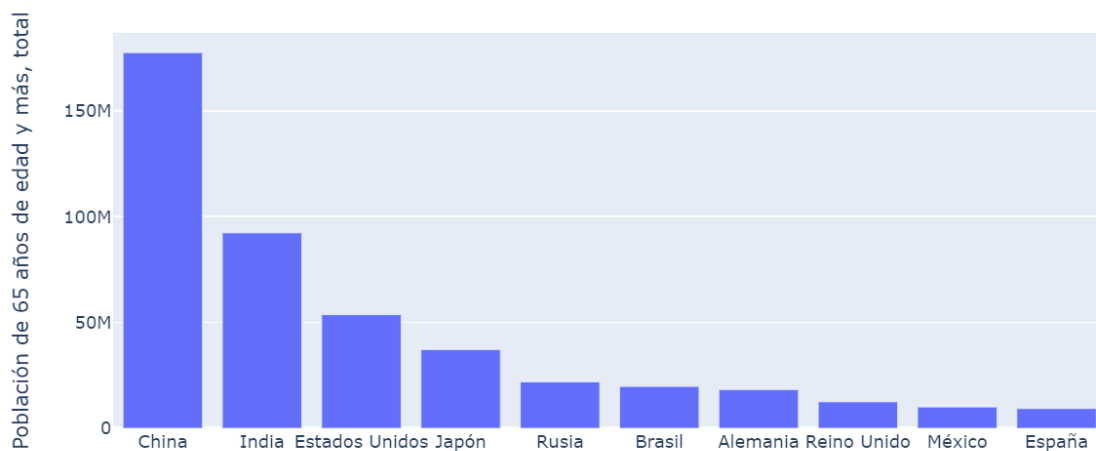
Registro de países con una población mayor a 150 M, podemos observar que China es el único país con esos registros

país Name	Año	Población de 65 años de edad y más, total
China	2017	152834660.0
China	2018	161037596.0
China	2019	169240734.0
China	2020	177774074.0
China	2021	185721855.0
China	2022	193783391.0

En promedio podemos ver que es el continente asiático el que tiene una mayor Población de 65 años de edad a más, seguido de América del Norte

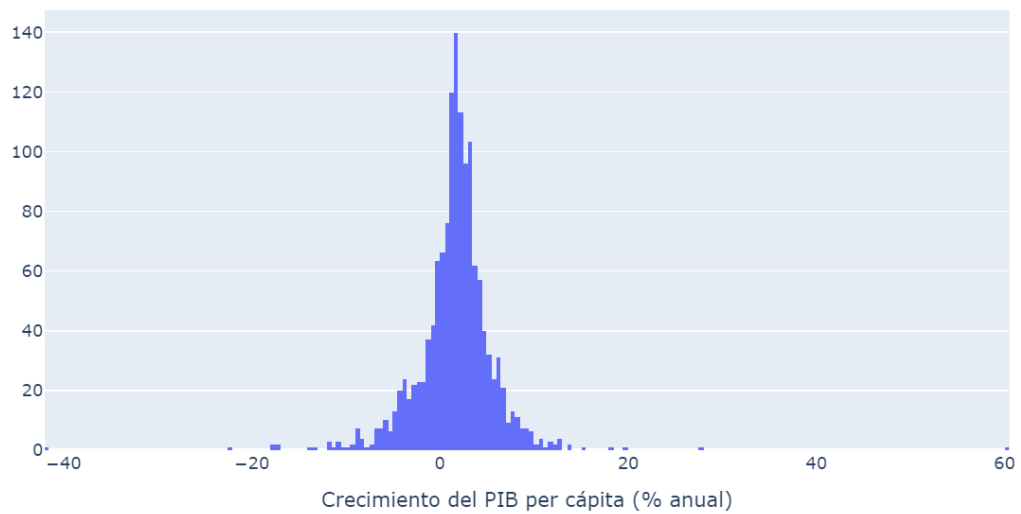


TOP 10 países con mayor Población de 65 años de edad a más (2018-2022)



## FACTOR ECONÓMICO

- Crecimiento del PIB per cápita (% anual)



Con respecto al indicador “Crecimiento del PIB per cápita (% anual)” podemos observar que los valores se concentran entre -2 - 7 %, podemos observar también que hay algunos datos muy alejados tanto negativos como positivos, en los negativos tenemos registros menores a -20% y en los positivos tenemos valores mayores a 20%.

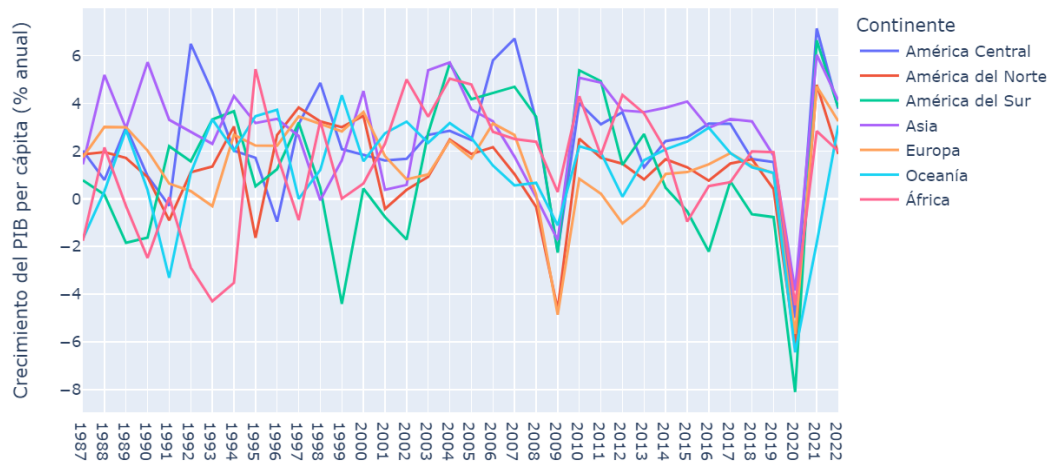
país Name	Año	Crecimiento del PIB per cápita (% anual)
Chad	2004	27.831095
Rwanda	1995	60.090541

Registros cuyos valores superan el 20%

país Name	Año	Crecimiento del PIB per cápita (% anual)
Rwanda	1994	-41.586861
Sierra Leona	2015	-22.383315

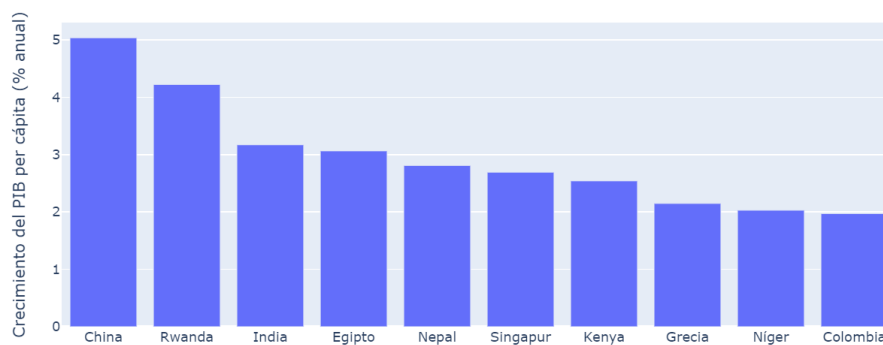
Registros que están por debajo del -20%

A lo largo de los años, en promedio el crecimiento de pbi per cápita de cada continente oscila entre -4% a 6% anual, vemos una caída en el año 2020 en todos los continentes producto de la pandemia COVID 19, pero se muestra una recuperación en los años posteriores.

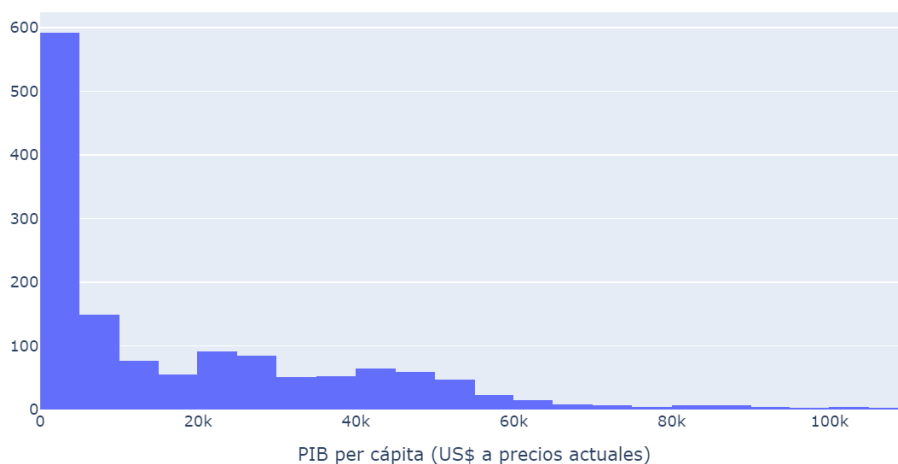


Países con mayor crecimiento de pbi per cápita en promedio en los últimos 5 años, china lidera el ranking seguido de Rwanda ewe India

TOP 10 países con mayor Crecimiento del PIB per cápita (% anual) en promedio (2018-2022)



- **PIB per cápita (US\$ a precios actuales)**

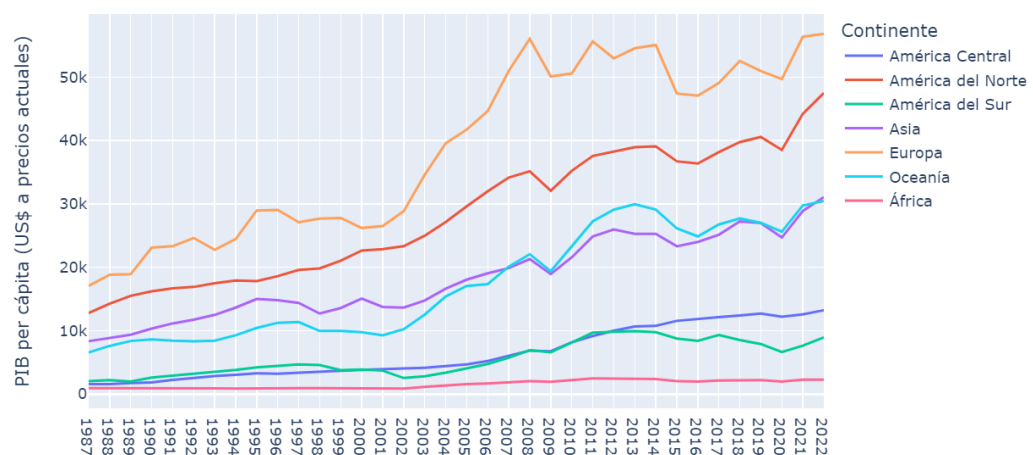


Con respecto al indicador “PIB per cápita (US\$ a precios actuales)” podemos observar que los valores se concentran entre 0-10K, podemos observar también que hay algunos datos muy alejados que superan los 90K.

Países con registros que superan los 90 K de PIB per cápita

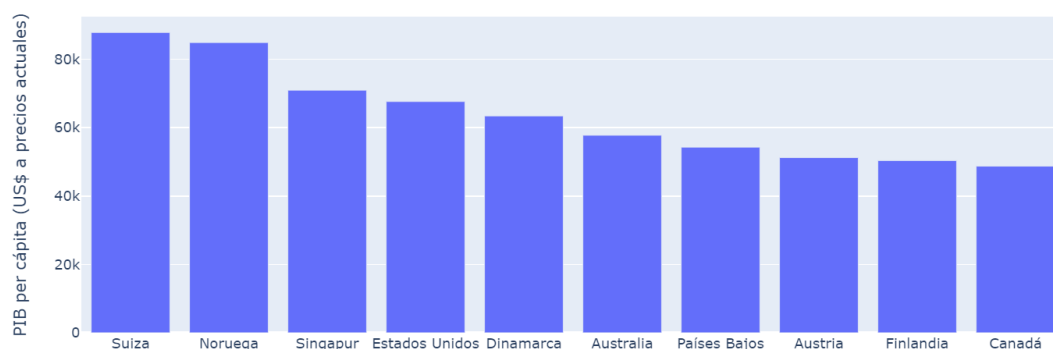
país Name	Año	PIB per cápita (US\$ a precios actuales)
Noruega	2008	97503.540780
Noruega	2011	101221.813477
Noruega	2012	102175.919298
Noruega	2013	103553.840134
Noruega	2014	97666.695184
Noruega	2021	90655.391023
Noruega	2022	106148.778628
Suiza	2011	90476.758965
Suiza	2021	91991.600458
Suiza	2022	92101.469741

A lo largo de los años, en promedio el “PIB per cápita” de cada continente ha ido en incremento, África parece ser el único continente que no ha crecido significativamente.



Tomando en cuenta los últimos 5 años, Suiza lidera el ranking con mayor PIB per cápita en promedio, seguido de Noruega y Singapur.

TOP 10 países con mayor PIB per cápita (US\$ a precios actuales) en promedio (2018-2022)



## ● INFLACIÓN

El análisis abarcando desde 1987 hasta 2022 revela la evolución compleja y variable de la serie temporal. La cantidad de observaciones varía entre 77 y 86 a lo largo de los años. La media exhibe fluctuaciones notables, desde valores bajos como 3.96 en 2003 hasta picos significativos como 110.48 en 1990. La desviación estándar, que varía de 5.50 a 44.09, sugiere una dispersión considerable en los datos. Se observan valores extremos, con el mínimo de -26.30 en 1999 y el máximo de 401.59 en 2021. Los percentiles detallan la distribución de los datos, resaltando, por ejemplo, que el 25% más bajo de los valores está por debajo de 2.14 y el 50% está por debajo de 3.26. Este análisis proporciona una comprensión integral de la complejidad y variabilidad de la serie temporal a lo largo de los años, destacando años específicos con valores extremos.

Un primer vistazo al conjunto de datos se obtuvo ejecutando `df.head()` en Pandas.

```
1 df.head()
```

	país	codigo_pais	factor_(serie)	codigo_serie	1987	1988	1989	1990	1991	1992	...	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
0	Afganistán	AFG	Inflación, índice de deflación del PIB (% anual)	NY.GDP.DEFL.KD.ZG	NaN	NaN	NaN	NaN	NaN	NaN	...	4.8228	0.5669	2.4476	-2.1975	2.4037	2.0713	6.5215	7.8217	0.4774	NaN
1	Afganistán	AFG	Inflación, precios al consumidor (% anual)	FP.CPI.TOTL.ZG	NaN	NaN	NaN	NaN	NaN	NaN	...	7.3858	4.6740	-0.6617	4.3839	4.9760	0.6261	2.3024	NaN	NaN	NaN
2	Alemania	DEU	Inflación, índice de deflación del PIB (% anual)	NY.GDP.DEFL.KD.ZG	1.2797	1.6905	2.8791	3.3966	3.0851	5.3062	...	1.9594	1.8779	1.8537	1.3278	1.5037	2.0075	2.1244	1.8108	3.0576	5.4828
3	Alemania	DEU	Inflación, precios al consumidor (% anual)	FP.CPI.TOTL.ZG	0.2499	1.2741	2.7806	2.6965	4.0470	5.0570	...	1.5047	0.9068	0.5144	0.4917	1.5095	1.7322	1.4457	0.1449	3.0667	6.8726
4	Argelia	DZA	Inflación, índice de deflación del PIB (% anual)	NY.GDP.DEFL.KD.ZG	8.8420	9.0610	16.0114	30.2596	53.7886	21.9261	...	-0.0933	-0.3006	-6.4557	1.5489	6.3907	6.7573	-0.4722	-5.0259	15.5672	19.7109

5 rows × 40 columns

Visualicemos los factores de la serie presentes en este conjunto de datos.

```
1 cantidad_series = len(df['factor_(serie)'].unique())
2 print('Cantidad de variables estudio:', cantidad_series)
3 serie_unicos = df['factor_(serie)'].unique()
4 print(serie_unicos)
```

✓ 0.0s

Cantidad de variables estudio: 2  
['Inflación, índice de deflación del PIB (% anual)'  
'Inflación, precios al consumidor (% anual)']

Para facilitar la comprensión y análisis de los datos, se ha optado por reorganizar el dataframe de la siguiente manera:

	país	factor_(serie)	Año	Valor
0	Afganistán	Inflación, índice de deflación del PIB (% anual)	1987	NaN
1	Afganistán	Inflación, precios al consumidor (% anual)	1987	NaN
2	Alemania	Inflación, índice de deflación del PIB (% anual)	1987	1.2797
3	Alemania	Inflación, precios al consumidor (% anual)	1987	0.2499
4	Argelia	Inflación, índice de deflación del PIB (% anual)	1987	8.8420
...	...	...	...	...
3163	Japón	Inflación, precios al consumidor (% anual)	2022	2.4977
3164	Kenya	Inflación, índice de deflación del PIB (% anual)	2022	6.0088
3165	Kenya	Inflación, precios al consumidor (% anual)	2022	7.6569
3166	Noruega	Inflación, índice de deflación del PIB (% anual)	2022	28.0333
3167	Noruega	Inflación, precios al consumidor (% anual)	2022	5.7641

3168 rows × 4 columns

Aplicamos la función `.describe()` para obtener un resumen estadístico del conjunto de datos.

```
Descripción para el factor 'Inflación, índice de deflación del PIB (% anual)':  
count    1548.000000  
mean      19.449905  
std       153.708483  
min       -26.300000  
25%        1.445925  
50%        3.622000  
75%        8.145500  
max       3046.091200  
Name: Valor, dtype: float64  
  
Descripción para el factor 'Inflación, precios al consumidor (% anual)':  
count    1405.000000  
mean      14.418154  
std       122.747322  
min       -8.974700  
25%        1.520400  
50%        3.151200  
75%        6.802800  
max       2947.732800  
Name: Valor, dtype: float64
```

#### Inflación, índice de deflación del PIB (% anual):

El factor "Inflación, índice de deflación del PIB (% anual)" abarca un total de 1548 observaciones. La media de este indicador se sitúa en un 19.45%, destacando una marcada variabilidad, indicada por la desviación estándar de 153.71. La distribución muestra un valor mínimo de -26.30% y un valor máximo significativamente alto de 3046.09%. Este amplio rango, evidenciado por los percentiles que van desde 1.45% hasta 8.15%, revela la presencia de valores extremos que afectan notablemente la media. La interpretación de este factor requeriría una atención especial a los valores atípicos que podrían influir en la percepción general de la inflación.

#### Inflación, precios al consumidor (% anual):

En cuanto al factor "Inflación, precios al consumidor (% anual)", se registran 1405 observaciones. La media para este indicador es del 14.42%, con una desviación estándar de 122.75, indicando nuevamente una variabilidad considerable. El valor mínimo es de -8.97%, y el valor máximo alcanza los 2947.73%. Los percentiles revelan que la mayoría de las observaciones se sitúan entre el 1.52% y el 6.80%. Al igual que en el factor anterior, se observa una influencia significativa de valores extremos, lo que enfatiza la necesidad de un análisis más profundo para comprender la verdadera dinámica de la inflación en este contexto.

## Imputación

Ante la presencia significativa de valores faltantes (NaN) en nuestro conjunto de datos, se optó por aplicar una estrategia de imputación. Esta estrategia consiste en el reemplazo de los valores faltantes por la media de los datos, tomando en consideración la especificidad de cada país como también la del factor en estudio. Este enfoque nos permite abordar la ausencia de datos, asegurando una mayor integridad en nuestro conjunto de datos al evitar vacíos o valores faltantes, lo que resulta crucial para un análisis robusto y confiable.

```

1 def fill_missing_with_country_factor_mean(df):
2     means_by_country_factor = df.groupby(['país', 'factor_(serie)'])['Valor'].transform('mean')
3     df['Valor'] = df['Valor'].fillna(means_by_country_factor)
4     return df
5
6 # Llamada a la función para rellenar los valores faltantes con la media por país y factor
7 df_filled = fill_missing_with_country_factor_mean(df_reorganizado)

```

✓ 0.0s

```

1 df_filled

```

✓ 0.0s

	país	factor_(serie)	Año	Valor
0	Afganistán	Inflación, índice de deflación del PIB (% anual)	1987	6.112895
1	Afganistán	Inflación, precios al consumidor (% anual)	1987	6.124627
2	Alemania	Inflación, índice de deflación del PIB (% anual)	1987	1.279700
3	Alemania	Inflación, precios al consumidor (% anual)	1987	0.249900
4	Argelia	Inflación, índice de deflación del PIB (% anual)	1987	8.842000
...	...	...	...	...
3163	Japón	Inflación, precios al consumidor (% anual)	2022	2.497700
3164	Kenya	Inflación, índice de deflación del PIB (% anual)	2022	6.008800
3165	Kenya	Inflación, precios al consumidor (% anual)	2022	7.656900
3166	Noruega	Inflación, índice de deflación del PIB (% anual)	2022	28.033300
3167	Noruega	Inflación, precios al consumidor (% anual)	2022	5.764100

3168 rows x 4 columns

Verificamos la presencia de datos Nan, resulta que hay países que no tienen datos en ciertos factores en ninguno de los años seleccionados.

```

1 df_filled['Valor'].isna().sum()

```

✓ 0.0s

72

Por lo tanto, procedemos a eliminar los datos correspondientes a aquellos factores para los cuales no existen registros en ninguno de los años. Esta decisión se basa en evitar la generación de datos ficticios que podrían distorsionar la integridad del conjunto de datos, reconociendo las particularidades de cada país.

```

1 df_filled[df_filled['Valor'].isna()]['país'].unique()

```

✓ 0.0s

array(['Argentina', 'Cuba'], dtype=object)

```

1 df_filled.dropna(inplace=True)

```

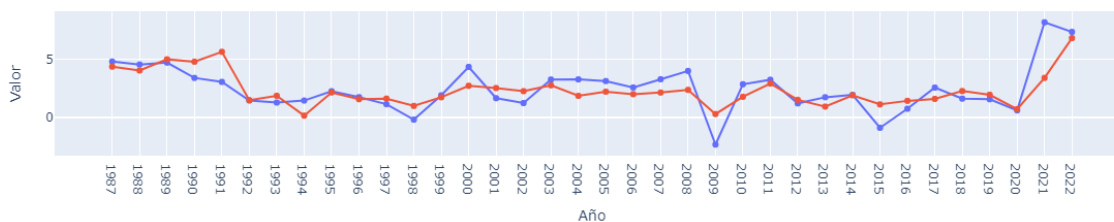
✓ 0.0s



## Distribución

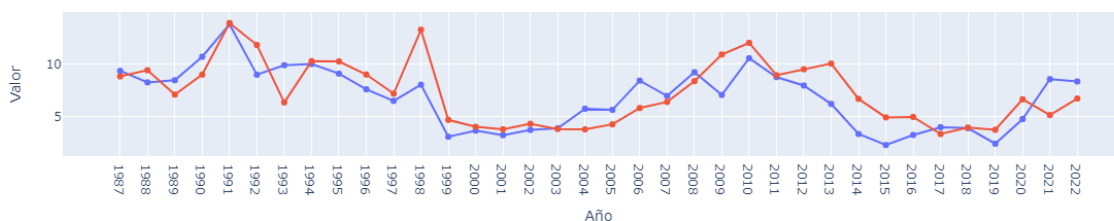
Ahora que contamos con la mayoría de los datos, procedemos a visualizar la distribución mediante gráficos. Se han seleccionado aleatoriamente algunos países para este análisis:

Datos para Canadá



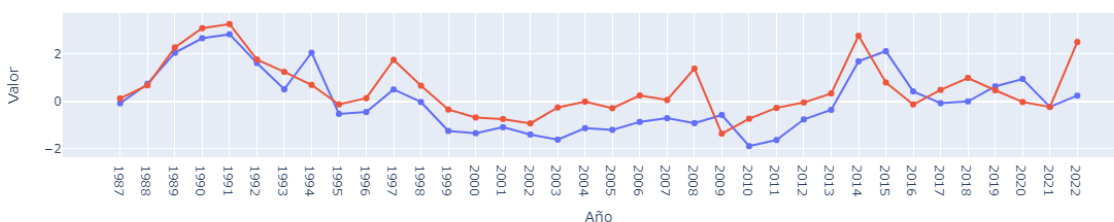
— Inflación, índice de deflación del PIB (% anual) — Inflación, precios al consumidor (% anual)

Datos para India



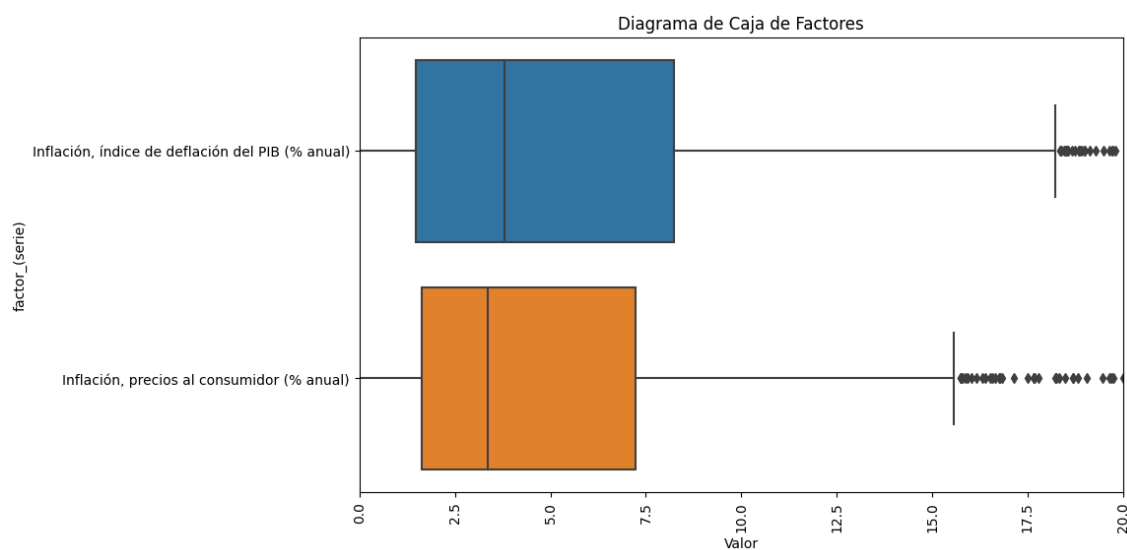
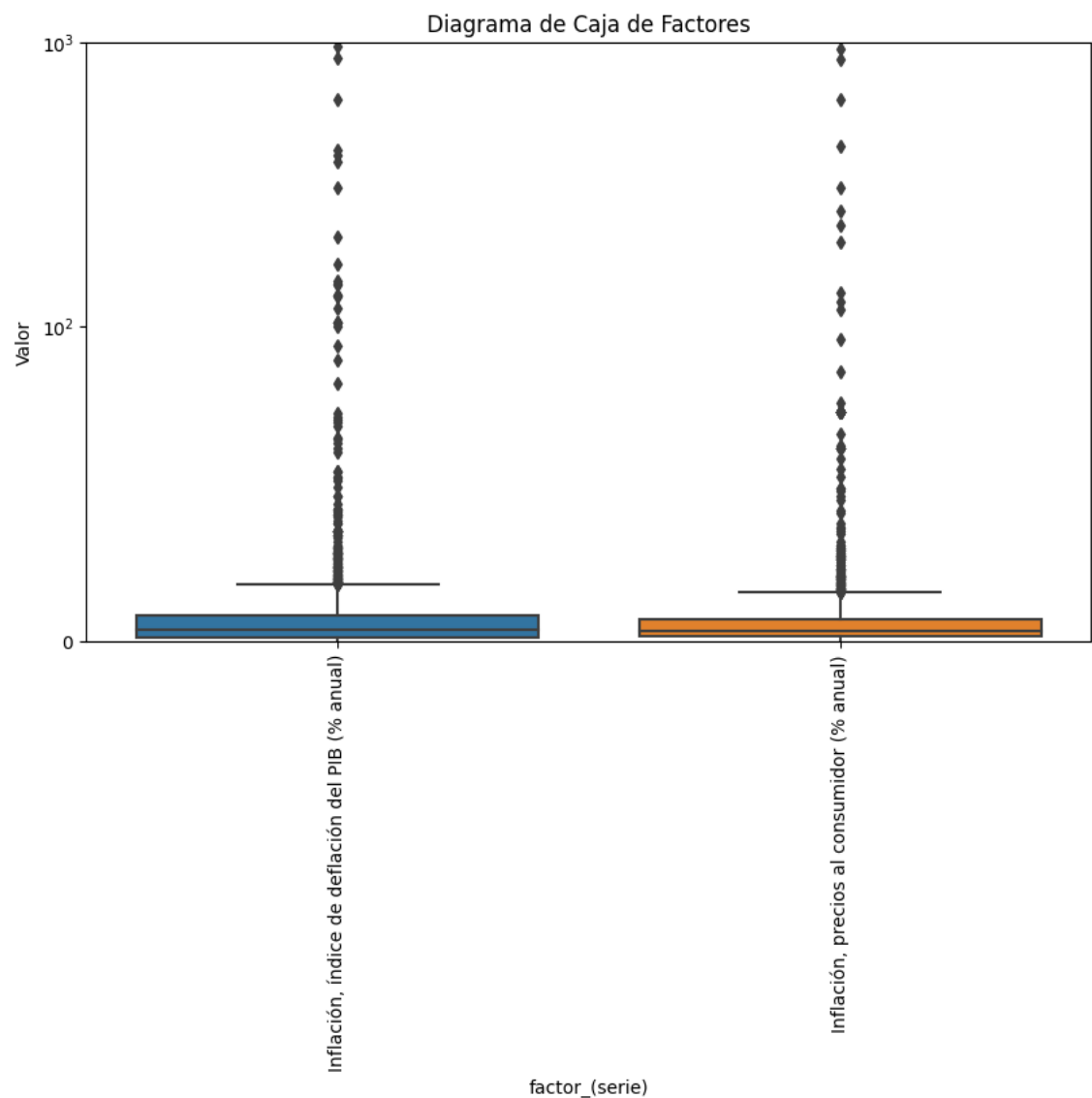
— Inflación, índice de deflación del PIB (% anual) — Inflación, precios al consumidor (% anual)

Datos para Japón



— Inflación, índice de deflación del PIB (% anual) — Inflación, precios al consumidor (% anual)

Diagrama Boxplot del dataset de inflación



Top 5 de países con mayor índice de inflación en el tiempo estudiado

