INFORME MLOps

Empresa Farmacéutica

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

---

Índice

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.

2. 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 las siguientes categorías:

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

Selección de indicadores para el EDA

Después de hacer una exploración de los datos nos damos cuenta que algunos de los indicadores seleccionados presentan una cantidad considerable de valores nulos, por lo que decidimos quedarnos con aquellos que presentan por lo menos 1000 registros de un total de 1584 no nulos.

Indicador_name	Total
Population ages 0-14, male	1584
Population ages 0-14, female	1584
Urban population	1584
Rural population	1584
Population, total	1584
Population, male	1584
Population, female	1584
Population ages 65 and above, total	1584
Population ages 65 and above, male	1584
Population ages 65 and above, female	1584
Population ages 15-64, total	1584
Population ages 15-64, male	1584
Population ages 15-64, female	1584
Population ages 0-14, total	1584
Urban population growth (annual %)	1584
Population growth (annual %)	1583
GDP per capita (current US$)	1551

Inflation, GDP deflator (annual %)	1548
GDP per capita growth (annual %)	1548
Rural population growth (annual %)	1548
Life expectancy at birth, total (years)	1540
Life expectancy at birth, male (years)	1540
Life expectancy at birth, female (years)	1540
Inflation, consumer prices (annual %)	1405
Domestic private health expenditure per capita (current US$)	924
Out-of-pocket expenditure per capita (current US$)	924
Domestic general government health expenditure per capita (current US$)	924
Lower secondary completion rate, total (% of relevant age group)	739
Educational attainment, at least completed upper secondary, population 25+, total (%) (cumulative)	404
Educational attainment, at least completed lower secondary, population 25+, total (%) (cumulative)	370
Educational attainment, at least completed post-secondary, population 25+, total (%) (cumulative)	233

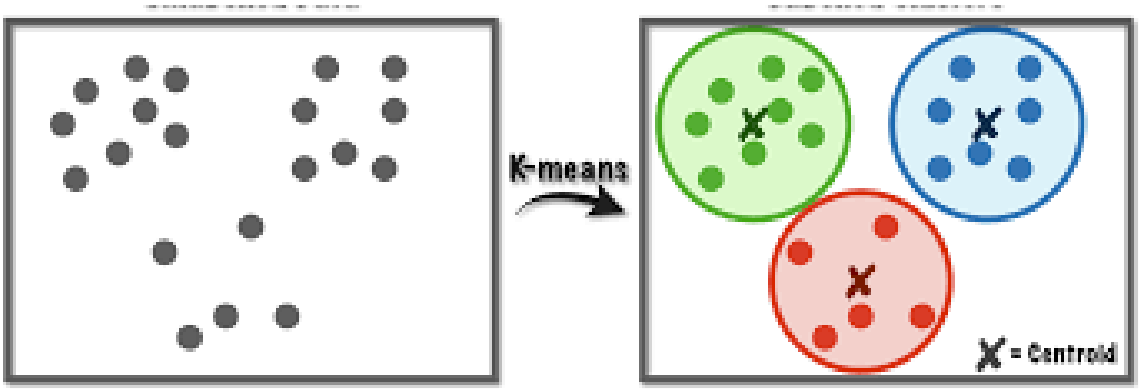
3. Selección de variables para el modelo

- País: Variable categórica que identifica la ubicación geográfica del registro.
- Año: Variable temporal que indica el año al que corresponde cada registro.
- GDP per capita (current US$): El Producto Interno Bruto (PIB) per cápita en dólares estadounidenses corrientes. Es un indicador económico que muestra la producción promedio por individuo en el país.
- Inflation, GDP deflator (annual %): Indica el cambio porcentual en el nivel general de precios en la economía.
- Inflation, consumer prices (annual %): Representa el cambio porcentual en el índice de precios al consumidor.
- Life expectancy at birth, total (years): La esperanza de vida al nacer en años.
- Population growth (annual %): La tasa de crecimiento poblacional anual, que indica la variación porcentual en la población en un año dado.
- Ratio_population ages 65 and above: La proporción de la población total que tiene 65 años o más.
- Ratio_urban population: La proporción de la población que vive en áreas urbanas en comparación con la población total.
- Urban population growth (annual %): La tasa de crecimiento anual de la población urbana.

4. Elección del modelo

Se ha decidido el uso de un modelo no supervisado con el objetivo de segmentar los países en base a sus indicadores y poder recomendar los que se encuentren en una mejor situación.

El algoritmo elegido es K-means.

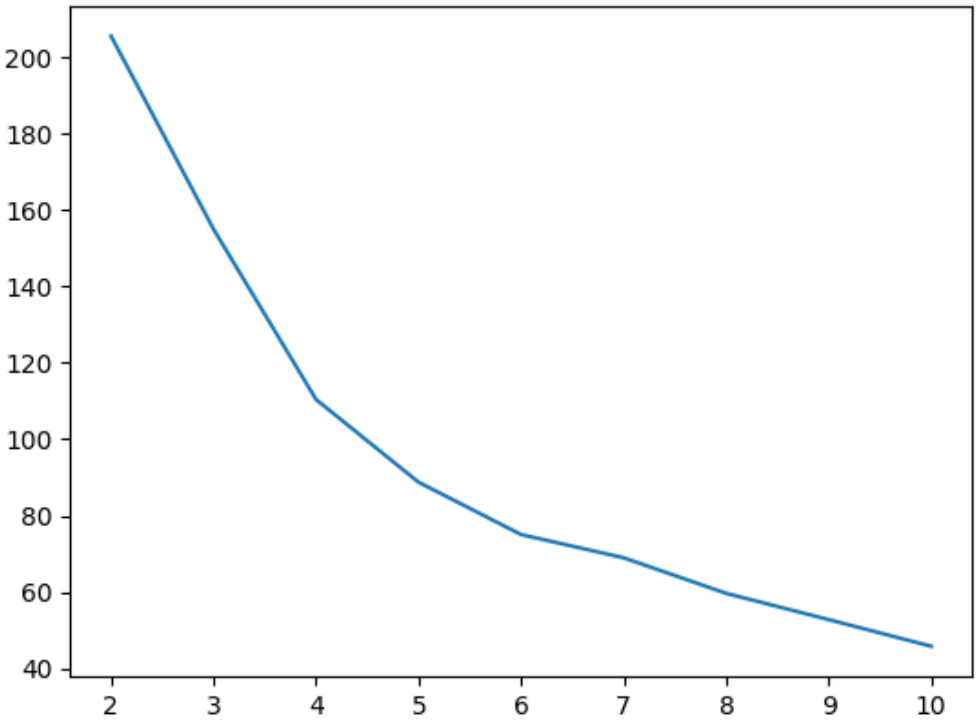


Se ha decidido tomar como referencia el promedio de los valores registrados en los últimos 5 años (2017-2021).

Pais	GDP per capita (current US$)	Inflation, GDP deflator (annual %)	Inflation, consumer prices (annual %)	Life expectancy at birth, total (years)	Population growth (annual %)	ratio_population ages 65 and above	ratio_urban population	Urban population growth (annual %)
AFG	0.000000	0.000000	0.000000	62.843800	2.929267	0.024091	0.257678	3.937778
ARE	43421.427078	2.395615	0.000000	79.302600	0.808158	0.014222	0.867812	1.116134
ARG	11100.883304	42.292984	0.000000	76.479600	0.992777	0.115880	0.919900	1.123750
AUS	55649.756424	2.702153	1.836326	82.929756	1.201004	0.159644	0.861286	1.331580
AUT	50282.610191	1.778456	1.951824	81.532683	0.495443	0.189736	0.585298	0.868419
BFA	797.775144	2.025153	1.148705	59.724600	2.735475	0.025614	0.299856	4.829883
BRA	8496.709453	6.049149	4.471526	74.406600	0.709181	0.090002	0.868184	1.003374
CAN	46751.935346	2.903484	1.985314	82.034541	1.139466	0.176072	0.814916	1.226117
CHE	85848.174807	0.122247	0.337790	83.612195	0.773234	0.185573	0.738636	0.842819
CHN	10378.507186	2.812982	1.993520	77.849600	0.350990	0.120393	0.602720	2.289982

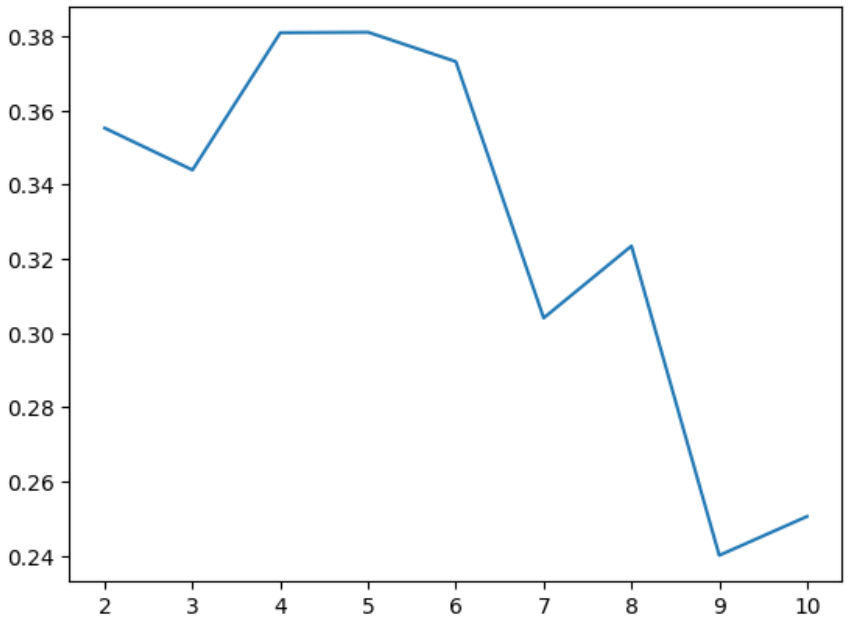
5. Selección del número óptimo de clusters

5.1 Método del codo



En el eje x tenemos el número de clusters y en el eje y la suma de las distancias al cuadrado. El "codo" en el gráfico es el punto donde la disminución en la suma de las distancias se hace más gradual después de haber descendido abruptamente. Este punto sugiere el número óptimo de clusters, ya que agregar más clusters a partir de ese punto no proporcionará una mejora significativa en la reducción de la suma de distancias.

5.2 Coeficiente de silhoutte

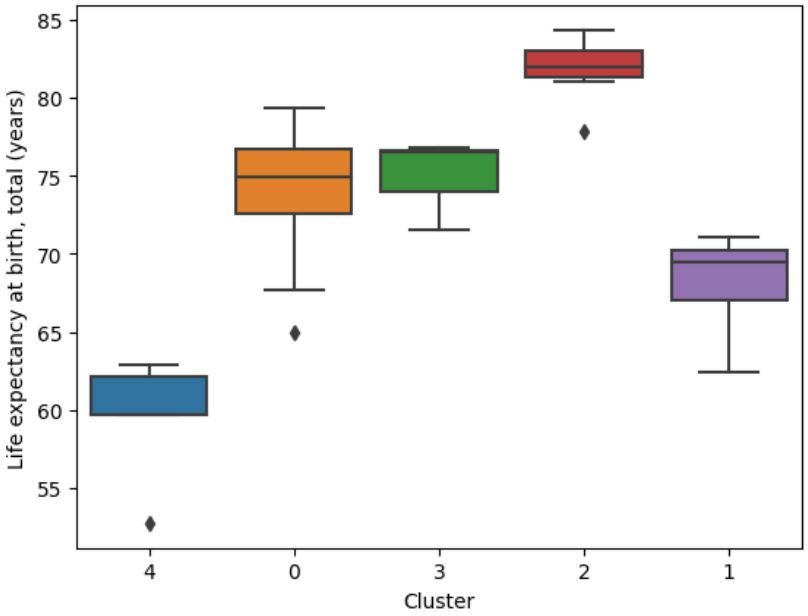


El coeficiente de silueta oscila entre -1 y 1. Un valor alto indica que la muestra está bien ubicada en su cluster y que está separada de los clusters vecinos, mientras que un valor bajo indica que la muestra podría estar asignada al cluster incorrecto o que está muy cerca de los límites del cluster.

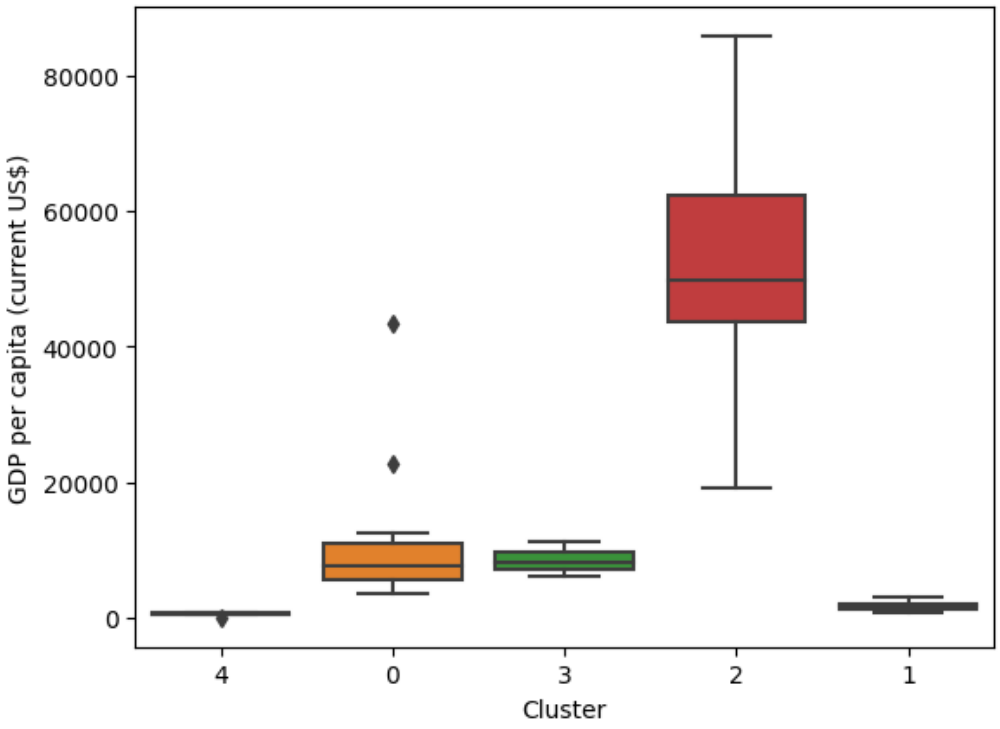
En base a los gráficos anteriores el número de clusters elegido es 5

6. Visualizaciones de los diferentes indicadores

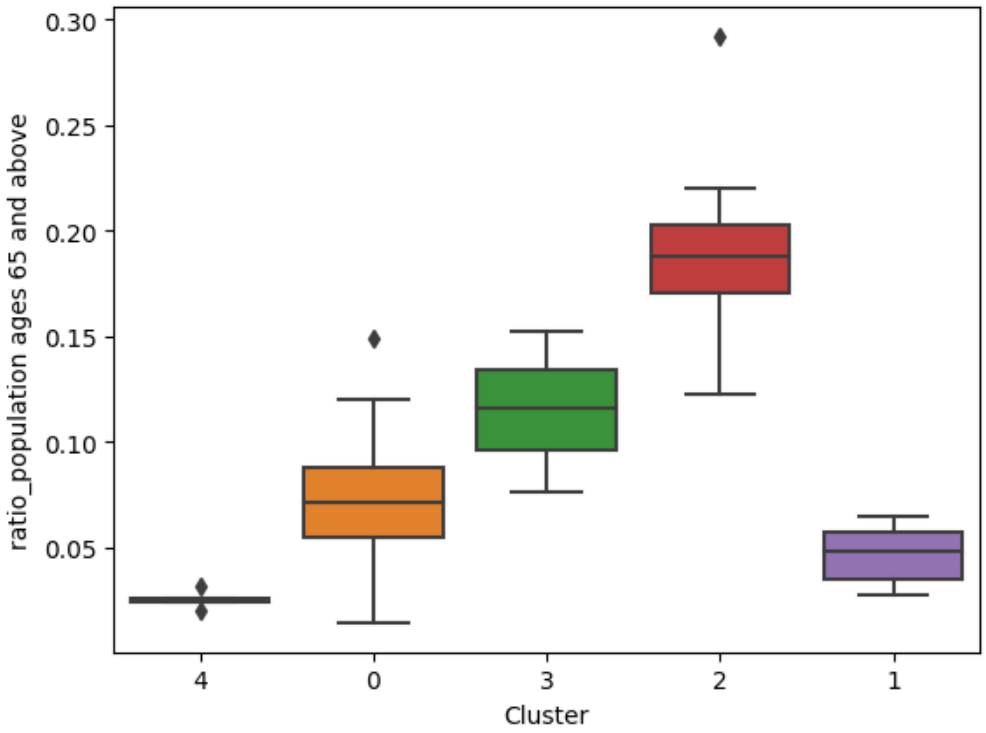
Observamos que el cluster 2 agrupa los valores mayores para el indicador de “Life expectancy at birth, total (years)”:



Observamos que el cluster 2 agrupa los valores mayores para el indicador de “GDP per capita (US$)”:



Observamos que el cluster 2 agrupa los valores mayores para el indicador de "Ratio_population ages 65 and above"



7. Elección del cluster para la recomendación de países

En base a los gráficos anteriores observamos que el cluster número 2 resalta sobre los demás en cada uno de los indicadores por lo que es el elegido.

Países recomendados

Pais	Continente
Australia	Oceania
Austria	Europa
Canada	América del Norte
Denmark	Europa
Finland	Europa
Germany	Europa
Greece	Europa
Japan	Asia
Netherlands	Europa
New Zealand	Oceania
Norway	Europa
Singapore	Asia
Spain	Europa
Switzerland	Europa
United Kingdom	Europa
United States	América del Norte