

# Health analytics

## Análisis de la esperanza de vida mundial

Proyecto final de máster Data Analytics

Rosa Rubio  
Bamba Tandine  
Alicia Penalba  
Raúl Fernández  
Laia Campoy

---

Tutor **Pedro Costa**

**nuclio**<sup>®</sup>  
digital school

# Estructura de la presentación

---

- **Contexto y objetivo** del proyecto
- **Normalización** de datos
- Análisis de datos
  - **Hipótesis simples** y conclusiones
  - **Clusters** y conclusiones
- Modelos de **Machine Learning**
- **Recomendaciones** para la WHO + **Dashboard**

# Contexto

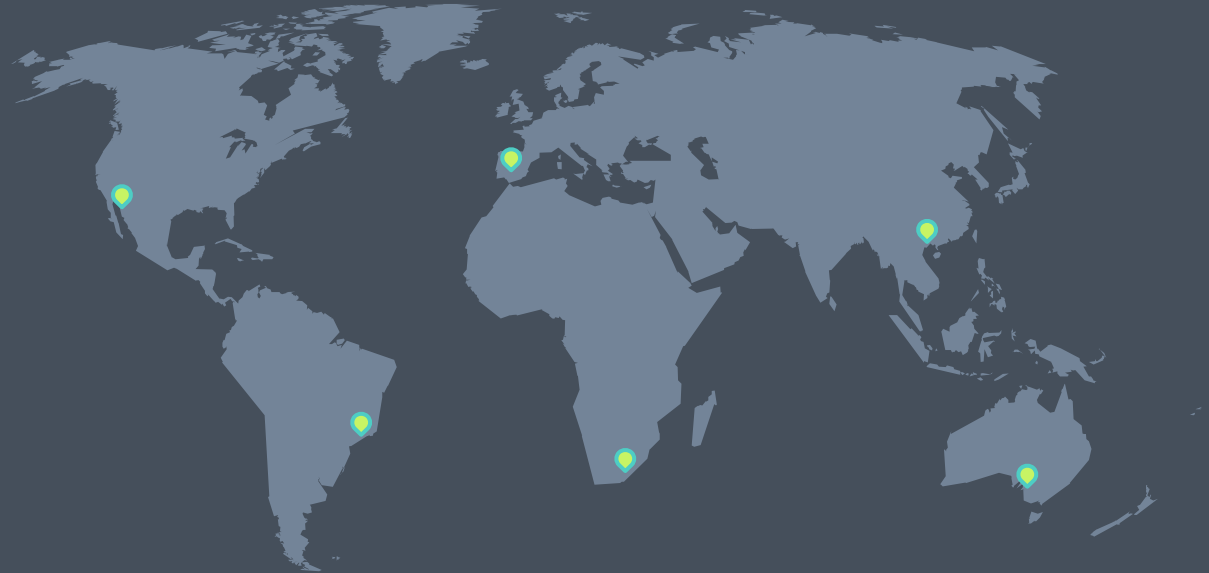
## ¿Qué es la esperanza de vida al nacer?

---

- Indicador demográfico que mide los **años promedio** que se espera que una persona viva desde su nacimiento
- Se basa en las **tasas de mortalidad** observadas en un año específico
- Los **avances sociales** han mejorado la **calidad de vida** y aumentado la **longevidad**
- Persisten **desigualdades** entre regiones: diferencias en **salud** y **condiciones socioeconómicas**

# Objetivo

- Identificar los factores que determinan la Esperanza de vida a nivel mundial
- Formular recomendaciones para mejorar las condiciones de vida mundial



# Normalización de datos



## Datos iniciales

8 tablas de Our World in data:

- Esperanza de vida
- PIB
- Gasto en Sanidad %PIB
- Causas de muerte varias
- Tasas de vacunación
- Pobreza/Desigualdad
- Suicidios
- Homicidios

Tabla adicional: **Población** por países (World Bank)



## Selección países y período

Países de estudio:  
**60** de un total de 122

Representación equitativa:  
**Continentes** y **Regiones**

Período de estudio: **2000-2019**



## Ajuste de los datos

**Causas de muerte relativas:**  
Muertes por cada 100.000 habitantes

**Selección de columnas** de pobreza/desigualdad

**Creación de columnas** adicionales para el análisis

# Análisis de datos **HIPÓTESIS SIMPLES**

Agrupación de los datos en 6 Hipótesis:

Capacidad  
económica



Criminalidad y  
factores políticos /  
geográficos



Desarrollo en  
sanidad



Salud física



Salud mental



Nivel de accidentes



# Análisis de datos RESULTADOS

## Últimos 20 años

tendencia general al alza de la esperanza de vida

Mayor peso

**Capacidad económica** y **desarrollo en sanidad**

## Tasas de vacunación en menores de 1 año

Clave en África

**Enfermedades crónicas en aumento** en zonas envejecidas

## Resto de causas de muerte

peso menos relevante por sí solas



# Relación entre esperanza de vida, PIB y vacunación por continentes

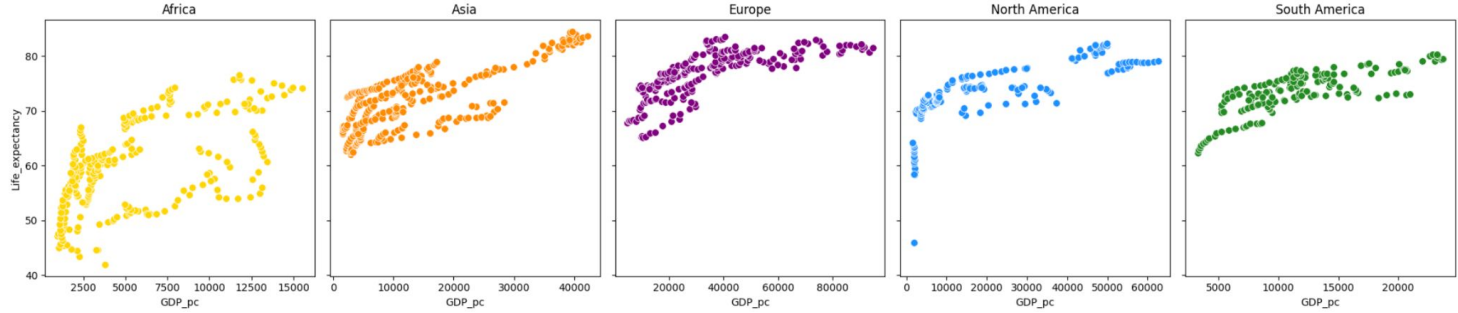
Relación positiva PIB  
y esperanza de vida,  
aunque **no siempre**  
se cumple esta  
relación

África:

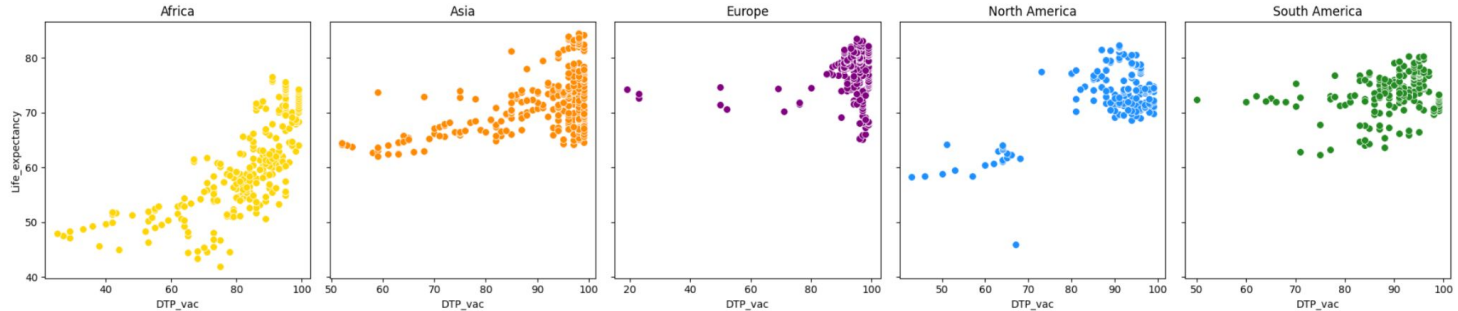
**Tasas de vacunación**  
impactan más debido  
a niveles iniciales  
bajos.

Europa:  
**Sistemas  
sanitarios  
fuertes** matizan  
este efecto.

Esperanza de Vida vs GDP\_pc



Esperanza de Vida vs DTP\_vac





# Análisis de datos **RESULTADOS**

## Últimos 20 años

tendencia general al alza de la esperanza de vida

Mayor peso

**Capacidad económica** y **desarrollo en sanidad**

## Tasas de vacunación en menores de 1 año

Clave en África

**Enfermedades crónicas en aumento** en zonas envejecidas

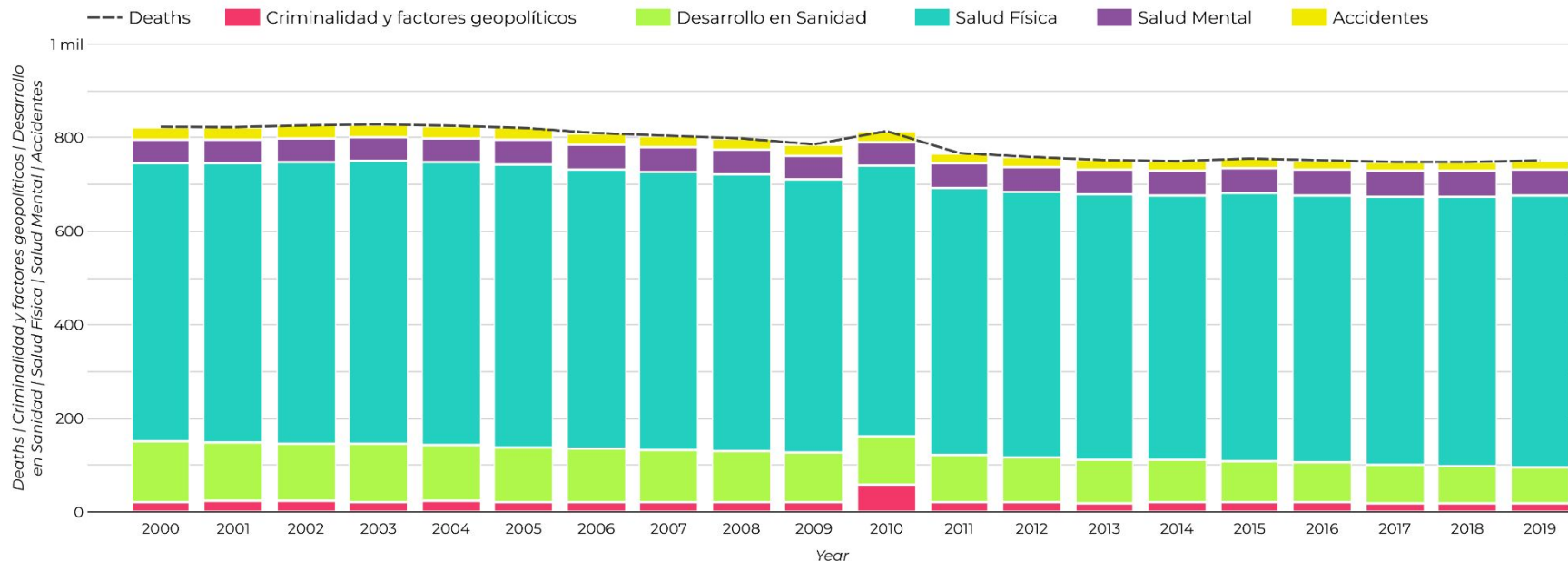
## Resto de causas de muerte

peso menos relevante por sí solas



# Causas de muerte por cada 100k, por hipótesis

Media de Muertes por Hipótesis



Las muertes por **accidentes**, **conflictos geopolíticos** y problemas de **salud mental** tienen un menor peso sobre el total.

# Análisis de datos **CLUSTERING k-means**

¿Qué **factores** influyen más en la **esperanza de vida** de diferentes grupos de **países** y cómo podemos **simplificar esta complejidad** a través de una metodología adecuada?

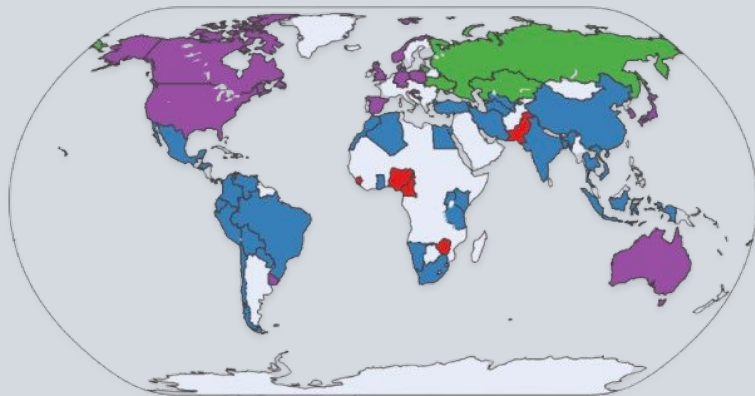
**K-means:** técnica que **agrupa datos** en **grupos** de manera que los elementos dentro de un mismo grupo **son más similares entre sí**.



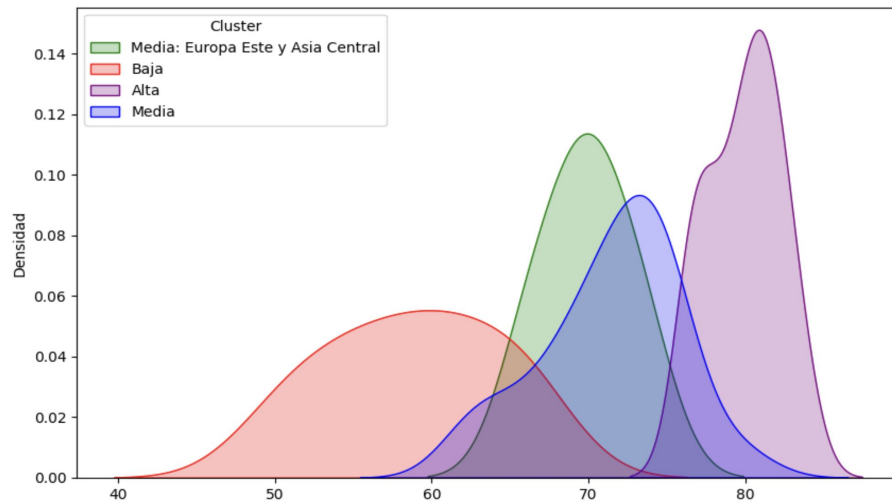
Esto nos permitió **analizar la variación en la esperanza de vida** dentro de cada grupo.

El modelo creó **cuatro grupos**, que luego renombramos según su esperanza de vida media.

# Análisis de datos CLUSTERING grupos



Los 60 países en el mapa por cluster



Distribución de la esperanza de vida por cluster

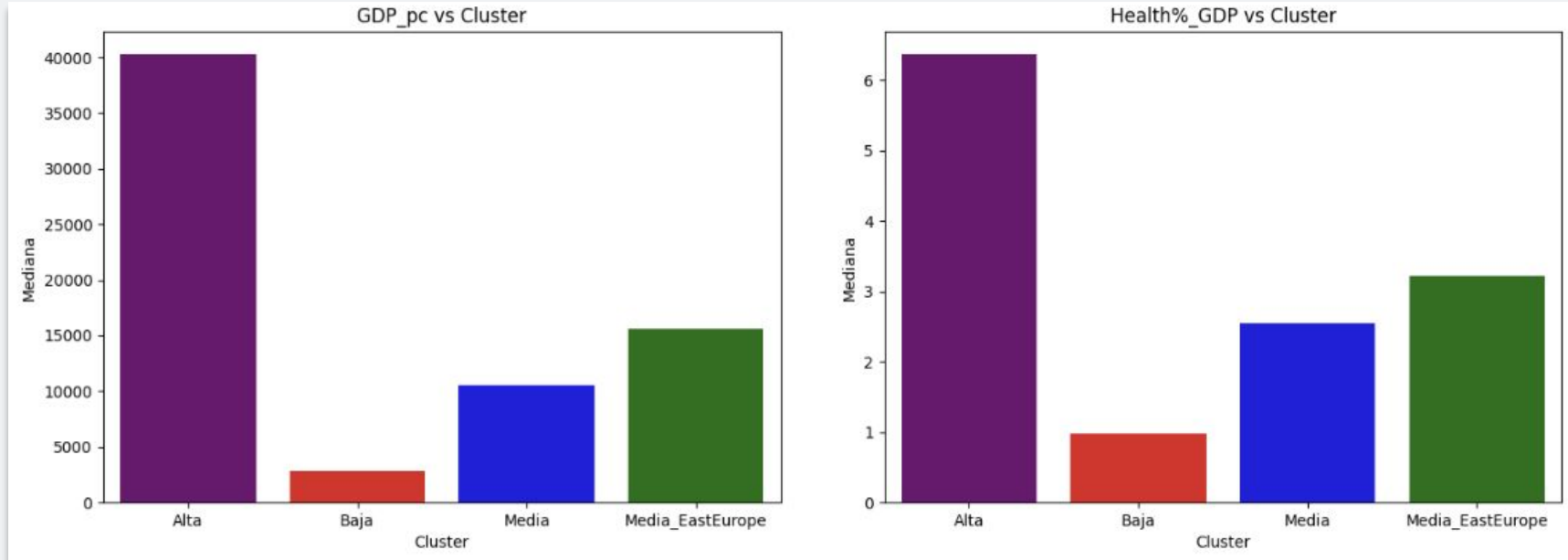
# Análisis de datos Características de los clusters

Cluster esperanza de vida **Alta** (79,8 años)

- Población reducida, **menor extensión territorial**
- **Desarrollo económico, limitaciones PIB** como medidor
- **Sanidad eficiente, estilo de vida y factores culturales** pueden influir más
- Desafíos: **enfermedades crónicas y salud mental**



## Cluster esperanza de vida **Alta** (79,8 años)



Su PIB per cápita y gasto en sanidad es muy superior al resto.

## Cluster esperanza de vida **Alta** (79,8 años)



Country	Life_expectancy	GDP_pc	Health%_GDP
Japan	82,8	39.409,61	7,35
Spain	81,83	34.432,67	6,32
United Kingdom	80,28	39.574,59	7,84
United States	78,67	54.983,23	7,89

Llegados a cierto punto, más crecimiento económico o mayor gasto en sanidad no siempre impactan positivamente a la esperanza de vida

# Análisis de datos Características de los clusters

Cluster esperanza de vida **Alta** (79,8 años)

- Población reducida, **menor extensión territorial**
- **Desarrollo económico, limitaciones PIB** como medidor
- **Sanidad eficiente, estilo de vida y factores culturales** pueden influir más
- Desafíos: **enfermedades crónicas y salud mental**





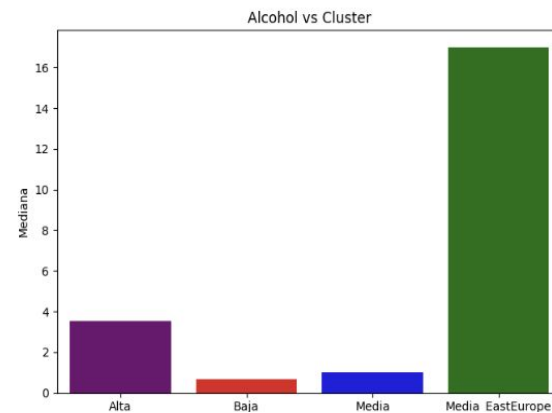
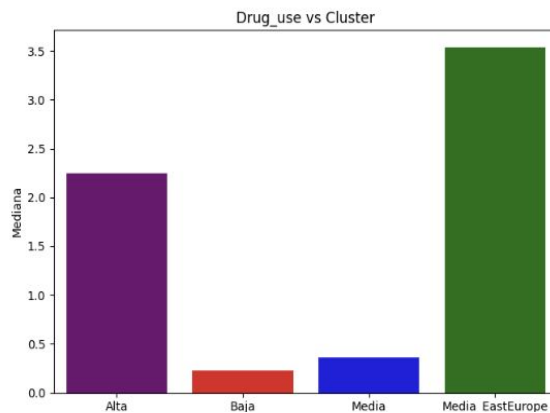
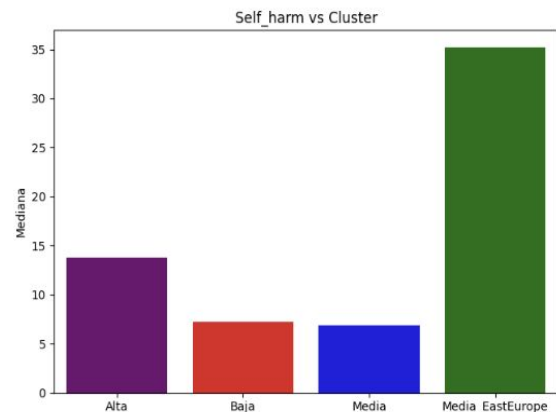
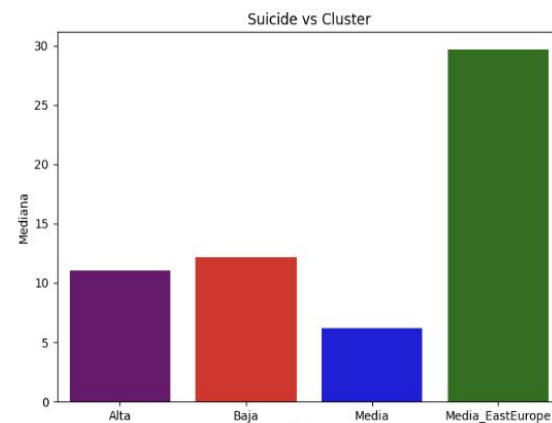
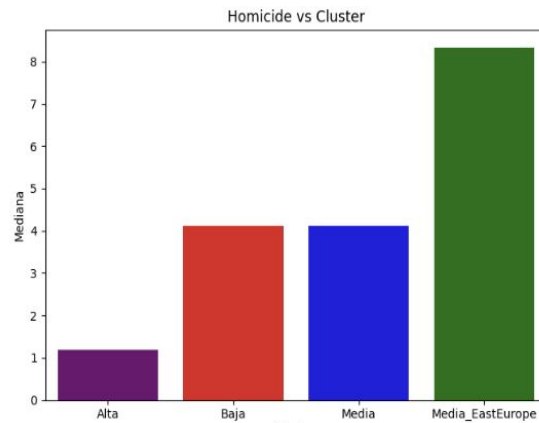
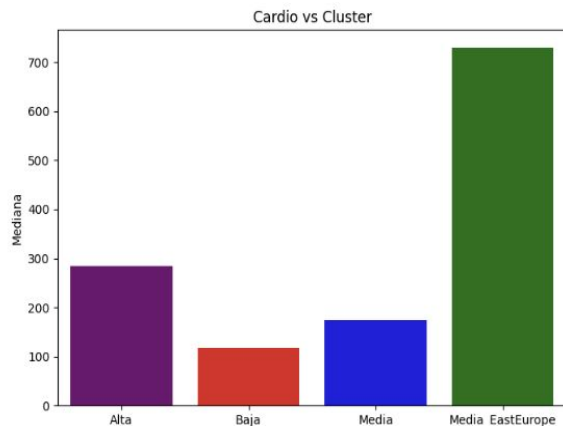
# Análisis de datos

## Características de los clusters

Cluster esperanza de vida **Media** (72,5 años) y **Media Europa del Este / Asia Central** (70,7 años)



- **Diversidad geográfica / socioeconómica:** países extensos y poblados (México, China, India, Rusia)
- Correlación positiva PIB y gasto en sanidad, pero con menor dispersión
- **Tasas de vacunación** altas, pero con variabilidad. **Enfermedades crónicas** con niveles altos.
- **Europa del Este:**
  - Enfermedades **cardiovasculares**, problemas graves de **salud mental** (alcoholismo, suicidio, autolesiones)
  - **Contexto geopolítico** como factor clave



## Europa del Este:

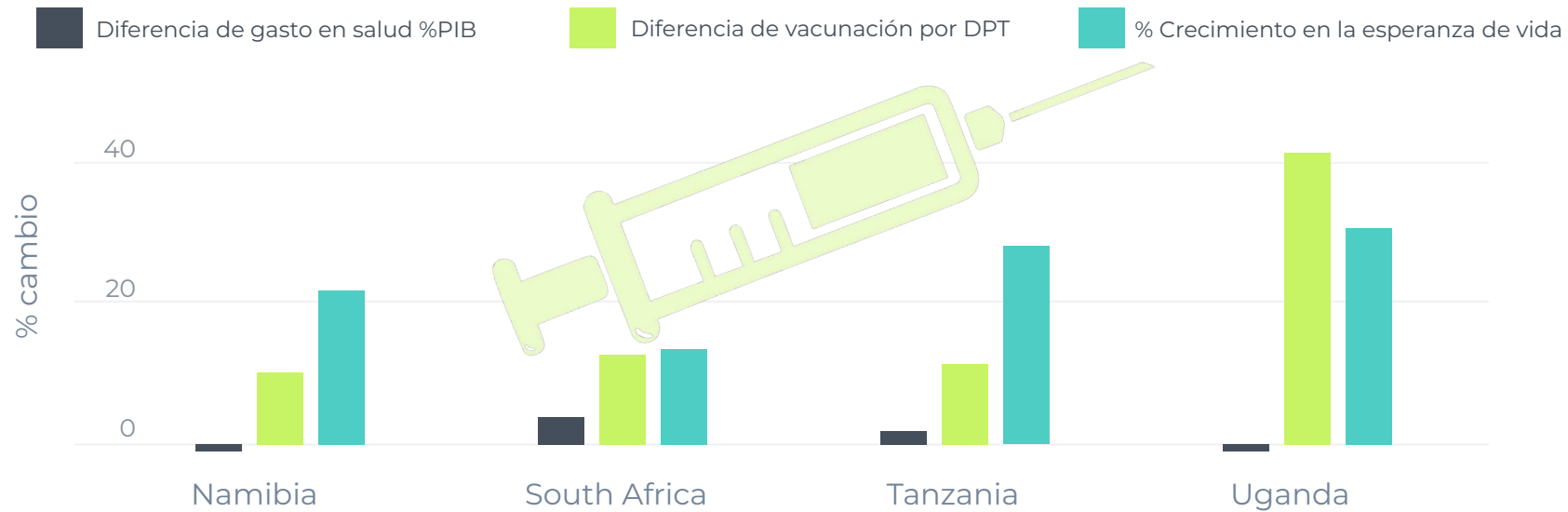
Las cardiopatías, homicidios y variables relacionadas con la salud mental definen este cluster

# Análisis de datos Características de los clusters

Cluster esperanza de vida **Baja** (58,5 años)

- Países en **África** y **Asia** ( exceptuando **Haití** )
- **Alta mortalidad por enfermedades infecciosas** ( VIH / SIDA, malaria )
- **Impactos**: Inversión en salud desigual, altos niveles de pobreza y desigualdad
- Progresos países con *cambio de cluster* : **Mejoras** en **tasas** de **vacunación**, reducción mortalidad neonatal y enfermedades infecciosas.





Para países con menor esperanza de vida, el incremento de la vacunación es crucial

# Conclusión

## En respuesta a nuestra pregunta

La esperanza de vida es un **indicador multifacético** que refleja la **interacción** de factores económicos, sanitarios y sociales

Claras **diferencias** de esperanza de vida **entre países más desarrollados** y ricos y **países pobres**.

Influencia de estos factores: varía según el **contexto socioeconómico** y **geopolítico**.

Importancia de un enfoque **contextualizado** en las políticas de salud pública.

# Modelos de Machine Learning

## REGRESIÓN LOGÍSTICA

Categorización de la variable:  
umbral en 72.65 años

## ÁRBOL DE DECISIÓN

Jerarquización de los datos:  
preguntas simples

¿Cuáles son los factores que permiten **predecir** con mayor precisión la esperanza de vida en distintos países?

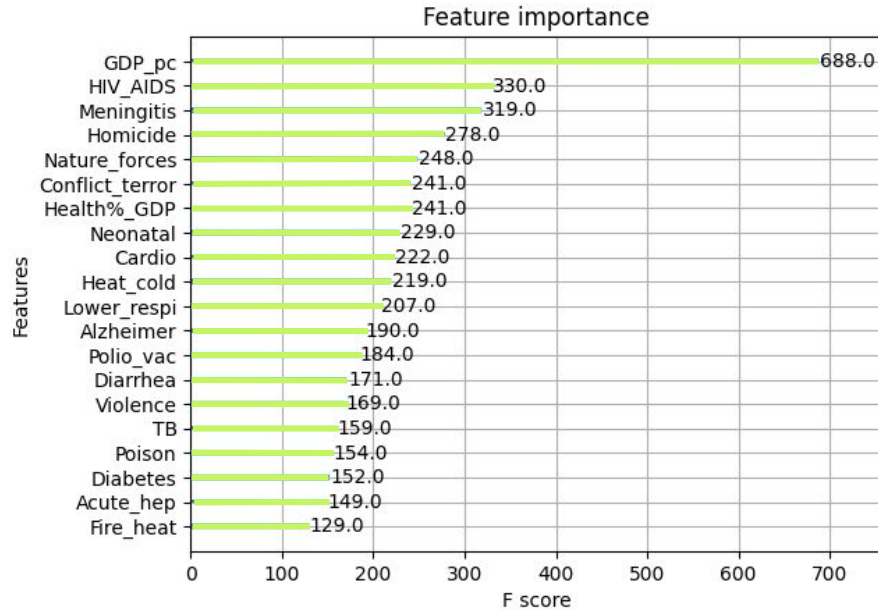
- Conjunto de árboles de decisión
- Variables con más relevancia: muertes por meningitis y neonatales

## RANDOM FOREST

- Modelo altamente eficiente y preciso
- Variables con más relevancia: PIB, VIH, meningitis

## XGBOOST

# Modelos de Machine Learning



Influencia de variables en el modelo XGBoost

# Modelos de Machine Learning



- En general, los modelos hacen mejores predicciones con los países con mayor esperanza de vida
- XGBoost es el modelo que realiza mejores predicciones
- Con el modelo de regresión logística se pierde información (granularidad de los datos)



País	Árbol decisión	Random Forest	XGBoost	R. log	Valor real
Japan	82.38	83.51	83.74	1	83.65
Spain	82.38	82.12	82.82	1	82.86
México	74.19	74.33	74.52	1	74.80
Zimbabwe	60.31	57.56	59.14	0	58.85



# Recomendaciones



## PAÍSES EN DESARROLLO y más pobres

### IMPORTANCIA DEL CONTEXTO



Fortalecer sistemas de salud y seguridad social y garantizar el acceso universal a la atención médica.



Implementar sistemas de atención primaria de calidad aumentando la inversión en infraestructura especialmente en áreas desfavorecidas.



Reducir las desigualdades sociales y las barreras económicas.

## PAÍSES MÁS RICOS y con mayor esperanza de vida



Centrarse en la prevención de enfermedades crónicas y en mejorar la calidad de vida de las poblaciones envejecidas.



Promover estilos de vida saludables, destacando el ejercicio regular, para la salud física y mental.

# ¡Gracias!

## ¿Alguna pregunta?

Link al [DASHBOARD](#)