1. Ejercicio Práctico 2: Agrupación de países

Contexto: Te encuentras como analista de datos para la ONG "Ayuda Internacional", organización comprometida con la lucha contra la pobreza y con proporcionar a las personas de países subdesarrollados servicios básicos y ayuda en tiempos de desastres y calamidades naturales. Esta ONG ha logrado recaudar aproximadamente 10 millones de dólares. Asimismo, tu equipo ha recopilado datos socioeconómicos actuales sobre todos los países del mundo.

El desafío: Utilizar este conjunto de datos para diseñar un método de agrupación que permita distribuir este dinero de manera eficaz según los países más necesitados y que se tomen decisiones estratégicas basadas en datos.

Objetivo: Basándote en el conjunto de datos proporcionado, tu tarea es agrupar los países en clústeres significativos utilizando factores socioeconómicos y de salud. La categorización deberá estar fundamentada en un análisis exploratorio de los datos, la extracción de las características y la correcta elección del método a emplear.

2. Instrucciones:

- **Análisis Exploratorio de los datos (2 puntos):**
    - **Ejercicio 1**- Llevar a cabo un procedimiento de EDA para obtener información acerca de los datos:
        - Declara el tamaño del dataset y muestra algunos registros por pantalla. Muestra los estadísticos (media, desviación típica, min, max...) relevantes de cada una de las variables. Observando el tipo de características, ¿podemos usarlas todas en un algoritmo de agrupación?
        - Analiza la distribución de cada una de las variables mediante gráficas de densidad y boxplots (se valora utilizar el mínimo código posible para mostrar todas las gráficas). ¿Están bien distribuidas las variables? ¿Qué podemos comentar de este análisis? Responde razonadamente.
- **Extracción de las características (2 puntos)**
    - **Ejercicio 2**: Realizaremos transformaciones en nuestros datos en crudo para obtener las variables con las que realizar el clustering:
        - Muestra la matriz de correlaciones por pantalla y comenta resumidamente sus conclusiones.
        - Vamos a **agrupar todas nuestras nueve variables en tres grandes indicadores** diferentes: **Salud** (4 variables), **Comercio** (2 variables) y **Finanzas** (3 variables). Definir las agrupaciones en base a lo que significa cada variable. Para construir cada uno de los indicadores, sumaremos cada una de las variables agrupadas en ese indicador dividida por su media (Por ejemplo: Comercio = (variable1/media_variable1) + (variable2/media_variable2)
        - Una vez construidos los indicadores, tenemos que asegurarnos que los tres estén a la **misma escala**. Para ello tenemos que decidirnos por la **estandarización** o la normalización, ambas funciones nativas de SKLearn. Visualiza la distribución de

los 3 grandes indicadores, decide que método emplear (estandarización/normalización) y aplícalo a nuestros datos.

- **Entrenamiento y evaluación del modelo (3 puntos)**
  - **Ejercicio 3:** Una vez tenemos los datos estandarizados, procedemos a aplicar nuestro método de agrupación.

    - **Selección del algoritmo de clustering**: Argumentar de manera razonada, teniendo en cuenta nuestro problema concreto, que método de agrupación es mejor aplicar en este caso. Existe uno claramente diferenciado.

    - **Ajuste de hiperparámetro/s**: Una vez seleccionado el modelo, ajusta los hiperparámetro/s para decidir el o los mejores valores del mismo. Recuerda utilizar todas las técnicas vistas en clase y hacer un análisis completo de la situación para poder tomar la decisión con toda la información disponible. La elección de los valor/es de los hiperparámetros deben estar justificados con gran profundidad y evidencias.

    - **Evaluación de la agrupación**: Utiliza métricas vistas en clase y la representación visual de la agrupación final realizada para valorar la calidad de la misma. Usa las librerías 2D y 3D vistas en las prácticas.

- **Evaluación del modelo (3 puntos)**
  - **Ejercicio 4**: Interpretar y explicar los resultados obtenidos en base a nuestro problema concreto. Se inlcuye:
    - Desde la ONG se tiene conocimiento de que las variables más representativas para decidir si un país necesita o no ayuda son el ingreso neto (income) y la mortalidad infantil (child_mort). Dibuja un boxplot de ambas variables respecto a los clústeres etiquetados para determinar el nivel de ayuda para cada uno de los diferentes grupos. **Nota**: Si hemos obtenido dos clústers finales, los niveles de ayuda serán (necesita ayuda/no necesita ayuda), si son 3 (no necesita ayuda / necesita ayuda moderada/ necesita mucha ayuda) y así sucesivamente.

    - Utilizando la librería *kaileido* y *plotly.express* dibuja un mapa mundi dónde se refleje el nivel de ayuda necesario por país en función de los clústeres realizados. Que cada nivel de ayuda (clúster) tenga un color diferente.

    - **Conclusiones**: En función de nuestro análisis, ¿qué países debería priorizar Ayuda Internacional para depositar su ayuda y sus recursos? ¿Qué mejoras o implementaciones podríamos hacer a este análisis para mejorar sus resultados? Razona las respuestas de manera argumentada.

3. Entrega:

- Se espera que completes el NoteBook adjunto al presente documento (Feedback_Clustering.ipynb) abarcando todas las fases mencionadas anteriormente. El informe (Notebook) debe incluir el código en Python empleado, gráficos, resultados y conclusiones.
- Solo debe ser entregado el documento 'Feedback_Clustering .ipynb' con la extensión de nombre y apellidos. Ejemplo: Feedback_Clustering_Fernando_Alonso.ipynb.

4. Instrucciones y anotaciones extra:

- En el notebook proporcionado (Feedback_Clustering.ipynb), se detallan de manera precisa las instrucciones y requerimientos para cada ejercicio.
- Es fundamental que todas las respuestas, comentarios y análisis, especialmente aquellos relacionados con los gráficos y métricas, estén adecuadamente justificados.
- Esta práctica es de carácter **individual**, por lo que no se permite el trabajo en grupo.
- La fecha límite para la entrega de la práctica es el **15 de febrero**, incluido este día.