1. Actividad Práctica: **Explorando Patrones Ocultos en Datos Reales** Tema: **Reducción de dimensionalidad + Clustering no supervisado** Fecha de entrega: **Miercoles 23/07/2025**
   * Objetivo: Aplicar técnicas de reducción de dimensionalidad y clustering para explorar, visualizar y descubrir patrones ocultos en un conjunto de datos realista.
   * Dataset: **Spotify Tracks Dataset (**[**https://www.kaggle.com/datasets/zaheenhamidani/ultimate-spotify-tracks-db**](https://www.kaggle.com/datasets/zaheenhamidani/ultimate-spotify-tracks-db)**)**

(Contiene características musicales como energy, danceability, acousticness, tempo, valence, etc.)

1. Paso a Paso para el desarrollo de la actividad:
   1. Exploración Inicial de Datos (EDA)
   2. Inspeccionar columnas: ¿cuáles son numéricas y útiles para clustering?
   3. Normalizar/estandarizar las variables numéricas.
   4. **OPCIONAL** Usar mapas de correlación para eliminar variables redundantes.
   5. Reducción de Dimensionalidad
   * Aplicar PCA para reducir la dimensionalidad, visualizando varianza explicada.
   * Aplicar t-SNE para explorar agrupamientos no lineales (diferente forma de visualizar).
   * Graficar los resultados en 2D: ¿hay grupos naturales visibles?
   1. Clustering No Supervisado

Agrupar instancias similares **sin usar etiquetas**, y explorar qué representan esos grupos.

* + Aplicar clustering (con KMeans o con DBSCAN) sobre los datos reducidos con PCA.
  + Visualizar los clusters en 2D con color.
  + Etiquetar los clusters con algún nombre provisional
  1. Interpretación y Evaluación

Validar si los clusters tienen sentido, y qué representan en términos reales.

* + Calcular métricas como Silhouette Score (-1: mal agrupado) o +1 :excelente agrupamiento) para ver que tan bien estan armados los clusters.
  + Analizar características promedio por cluster. (En este punto será necesario unir los resultados de la reduccion de dimensionalidad con los datos originales para hacer el análisis basado en las columnas iniciales) --> Esta union se puede realizar con un merge, los datos originales y los datos despues de la reducción de dimensionalidad conservan el orden, y pueden unirse por el indice.
  + Usar la columna género para determinar ¿Qué género predomina en cada cluster? ¿Estos generos son movidos, energéticos, pasivos?

**Realizar todos los puntos en un Jupyter Notebook y subirlo al repositorio con todas las celdas ejecutadas para su posterior revision**

1. Si tienen dudas o preguntas, las pueden hacer aqui mismo en el canal general, para que todos puedan aportar a la solución