# Batería de Ejercicios: Análisis con Pandas

A continuación, se presenta una batería de ejercicios relacionados con el dataset cargado. Estas actividades están diseñadas para poner en práctica conceptos como limpieza, manipulación, visualización y análisis avanzado de datos con Pandas.

## Exploración y limpieza de datos

* - Muestra las primeras 10 filas del dataset.
* - ¿Cuáles son las columnas categóricas y cuáles son numéricas?
* - Verifica si hay valores nulos en el dataset y calcula la proporción de valores nulos por columna.
* - Verifica si hay filas duplicadas en el dataset y elimínalas dejando solo la primera aparición.

## Transformaciones y manipulaciones

* - Convierte la columna `track\_album\_release\_date` al tipo de dato `datetime`.
* - Reduce el tamaño de las columnas numéricas (`energy`, `tempo`, `duration\_ms`) para optimizar el uso de memoria.
* - Extrae el nombre del primer artista de la columna `track\_artist`.
* - Crea una nueva columna que contenga únicamente el género (`playlist\_genre`) en mayúsculas.
* - Calcula el 'nivel de energía ajustado', definiéndolo como: `energy` × `valence`.
* - Convierte la duración de las canciones (`duration\_ms`) de milisegundos a minutos y guárdalo en una nueva columna.

## Manejo de datos faltantes

* - Rellena los valores faltantes en la columna `track\_album\_name` con el texto 'Desconocido'.
* - Imputa los valores faltantes en la columna `Precio` (si existiera) con la mediana de los precios.

## Visualización y análisis exploratorio

* - Crea un gráfico de barras para visualizar el promedio de popularidad (`track\_popularity`) por género (`playlist\_genre`).
* - Crea un histograma para la distribución de la columna `danceability`.
* - Crea un gráfico de dispersión que muestre la relación entre `tempo` y `energy`, diferenciando por género (`playlist\_genre`).
* - Representa un boxplot para analizar la distribución de `track\_popularity` por subgénero (`playlist\_subgenre`).

## Análisis de correlaciones

* - Calcula la matriz de correlación entre las columnas numéricas del dataset.
* - Representa un mapa de calor (heatmap) para visualizar las correlaciones.

## Análisis estadístico

* - Agrupa el dataset por `playlist\_genre` y calcula el promedio de la duración de las canciones y la popularidad por género.
* - ¿Cuál es el género más popular en promedio?
* - Analiza la tendencia de la popularidad (`track\_popularity`) en función de las fechas de lanzamiento (`track\_album\_release\_date`).

## Expresiones regulares

* - Extrae únicamente la parte numérica del `track\_id` (si contiene caracteres adicionales).
* - Verifica si las descripciones de las playlists (`playlist\_name`) contienen palabras clave como 'hits' o '2023'.

## Desafíos adicionales

* - Evalúa el uso de memoria del dataset antes y después de optimizar las columnas categóricas y numéricas.
* - Encuentra los 5 artistas con más canciones en el dataset.
* - Identifica el subgénero (`playlist\_subgenre`) con el mayor número de canciones instrumentales (`instrumentalness > 0.8`).
* - Genera un resumen con las 10 canciones más populares, incluyendo su artista, álbum y popularidad.