

# Segunda Entrega Proyecto Ing. Datos

## Base de datos de Spotify

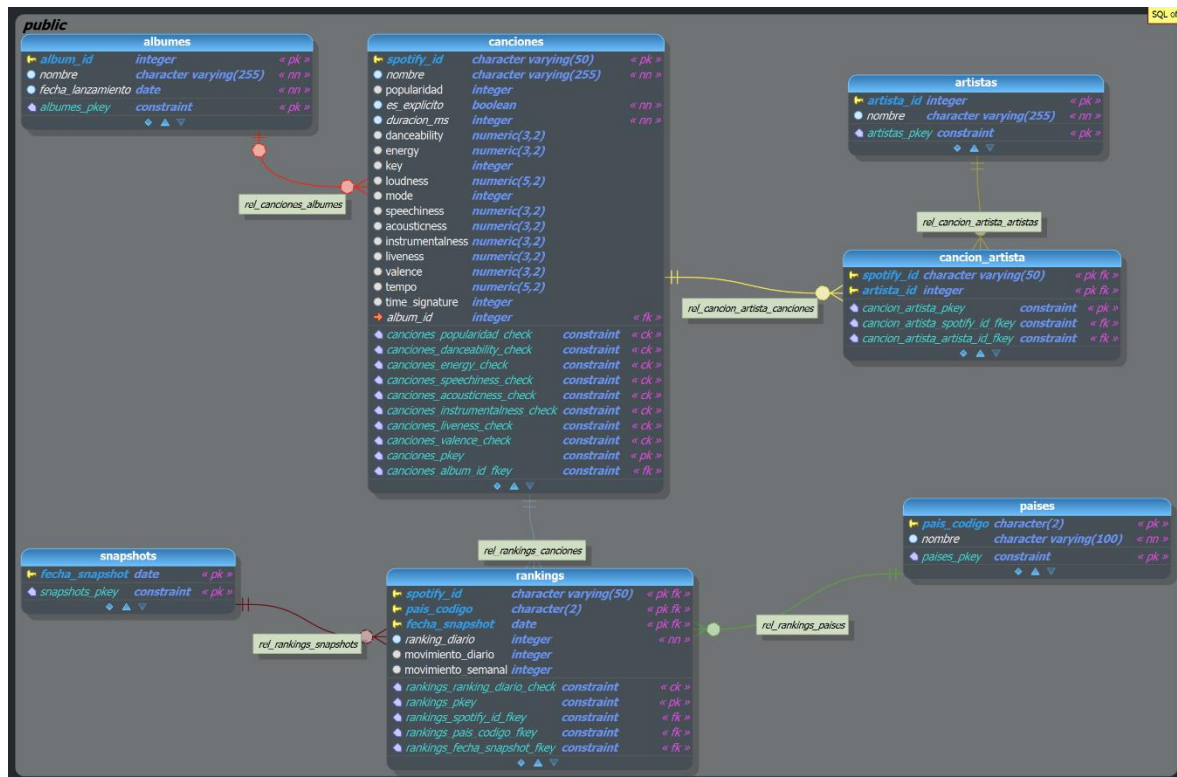
Carlos Julián Lucero Tobar, David Hernando Sanchez Lombana

### Planteamiento del Problema

Las organizaciones enfrentan varios desafíos en la gestión de estos datos:

1. Integridad y Consistencia: Garantizar que la información sobre álbumes, canciones y artistas sea precisa y coherente.
2. Complejidad de Relaciones: Manejar relaciones complejas, como la asociación muchos a muchos entre canciones y artistas, sin redundancias.
3. Acceso y Análisis Eficiente: Facilitar la generación de informes y el análisis de tendencias musicales de manera rápida y efectiva.
4. Escalabilidad: Adaptar la base de datos al crecimiento continuo de datos y nuevas métricas sin interrupciones.
5. Seguridad: Proteger la información sensible y controlar el acceso para mantener la confidencialidad y cumplir con normativas.

### Diagrama Relacional en tercera forma normal:



## **Implementación en PostgreSQL:**

La base de datos implementada en PostgreSQL se encuentra en este repositorio de GitHub: <https://github.com/CvrlosLucero/SpotifyDB/> (En la rama llamada "David" y nombre del archivo es ddl.)

## **Carga masiva de datos:**

Para cargar la información en nuestra base de datos PostgreSQL utilizando archivos CSV, primero organizamos todos los datos que queríamos insertar. Comenzamos seleccionando y extrayendo la información desde el archivo base descargado de [Kaggle](#) en CSV, asegurándonos de que todo estuviera en el formato correcto para evitar errores durante la importación.

Una vez que los datos estaban listos, los separamos en varios archivos CSV, uno para cada entidad que necesitábamos insertar en PostgreSQL. Por ejemplo, un archivo para "Álbumes", otro para "Canciones", y así sucesivamente. Esto nos facilitó el proceso de carga, ya que pudimos tratar cada tabla de forma independiente.

Después de preparar los archivos, nos conectamos a la base de datos PostgreSQL utilizando la terminal con PgAdmin. Ahí empleamos el comando COPY, que es una forma rápida y eficiente de cargar datos en PostgreSQL directamente desde los archivos CSV. Por ejemplo, para cargar los datos de los álbumes, utilizamos un comando como `copy Albumes FROM './albumes.csv' DELIMITER ',' CSV HEADER`. Repetimos este proceso para cada archivo que habíamos preparado, cargándolos uno por uno en sus respectivas tablas.

Una vez que terminamos de cargar todos los datos, verificamos que se hubieran importado correctamente. Para esto, ejecutamos algunas consultas simples como `SELECT COUNT(*)` en cada tabla y nos aseguramos de que los números coincidieran con lo que esperábamos.

## **Conexión de Base de Datos PostgreSQL a Python:**

La conexión a la base de datos en Python se encuentra en este repositorio de GitHub: <https://github.com/CvrlosLucero/SpotifyDB/> (En la rama llamada "Carlos" y nombre del archivo es connection.py)

## **Escenarios Posibles de análisis con datos de la DB:**

### **1) Análisis de Tendencias de Popularidad de Canciones por País.**

Evaluar cómo varía la popularidad de las canciones en diferentes países y detectar patrones de aceptación regional.

Usando la tabla Rankings junto con Canciones, Paises, y Snapshots, se puede determinar cómo cambian los rankings de canciones en un país específico. Se podrían identificar patrones estacionales o eventos que favorezcan a ciertas canciones.

Esto ayudaría a discográficas y plataformas a comprender la distribución de popularidad por región y a enfocar estrategias de promoción en países donde el interés por una canción va en aumento.

### **2) Características Musicales Relacionadas con el Éxito.**

Identificar qué características musicales son más comunes entre las canciones exitosas.

Usando los datos de las columnas como danceability, energy, valence, tempo, entre otras, junto con Rankings, para encontrar patrones en las características que tienen las canciones mejor posicionadas.

Esto permite a los productores musicales podrían utilizar estos datos para crear canciones que tengan un mayor potencial de éxito, basándose en las características que son más populares.

### **3) Duración y Éxito de las Canciones en las Listas de Popularidad**

Se analizará la relación entre la duración de una canción (duracion\_ms) y su rendimiento en los rankings de popularidad. El objetivo es identificar si las canciones más cortas o más largas tienden a obtener mejores posiciones y si ciertos atributos (valence, acousticness, speechiness) influyen en la permanencia en los rankings. Además, se estudiará si esta relación varía según el país, ayudando a comprender cómo las preferencias de duración impactan en el éxito en diferentes mercados.

#### 4) Análisis del Impacto del Movimiento en el Ranking en la Popularidad de Canciones.

Se evaluará cómo los cambios en el ranking (movimiento\_diario y movimiento\_semanal) influyen en la popularidad de canciones. El objetivo es identificar si ciertos atributos (energy, tempo, danceability) están relacionados con ascensos en el ranking y cómo estas fluctuaciones afectan la longevidad de las canciones en diferentes países. Esto permitirá entender qué factores impulsan una canción a mantenerse o subir en las listas.