



PRACTICA DE LABORATORIO

Unidad 2: introducción a la limpieza  
de datos.

**MATERIA:**

**Introducción a la Ciencia de Datos**

***PRACTICA:***

***Limpieza de una base de datos  
ensuciada***

**Alumnos:**

- ° ***Medina Sanchez Sugey***
- ° ***Hernandez Muñoz Karol***

**Profesor:**

***Jaime Alejandro Romero Sierra***

# **ANÁLISIS Y VISUALIZACIÓN DE TENDENCIAS MUSICALES EN SPOTIFY**

## ***Objetivo General:***

Evaluar como los algoritmos de recomendación en las listas de reproducción y las interacciones de los usuarios en Spotify afectan la popularidad mediante lo explicito, la duración o listas de reproducción de las canciones, esto para desarrollar estrategias comerciales que maximicen el crecimiento de los artistas y optimicen el retorno de inversión en el marketing musical.

Esto permitirá analizar no solo la popularidad, sino también las implicaciones económicas y las oportunidades para monetizar esa popularidad.

## ***Descripción del problema:***

A medida que Spotify incrementa como una de las plataformas más influyentes en la industria musical, un problema crítico que suele tener:

### ***La Desigualdad en la visibilidad de canciones***

La plataforma tiene millones de artistas, pero la realidad es que una gran cantidad de ellos, especialmente los menos conocidos, enfrentan una lucha por que su música pueda ser escuchada. Este problema suele ser por los algoritmos de recomendación que, si bien son sofisticados, tienden a favorecer a artistas que ya tienen de popularidad. Esto crea un vínculo adictivo en el que sólo pocas canciones se destacan, mientras que muchas otras, que podrían resonar profundamente con los oyentes, quedan en el olvido.

Esto no solo afecta a los artistas, si no que también limita las experiencias musicales de los oyentes. Muchos usuarios de Spotify se ven atrapados donde lo “popular” destaca, y se pierden la oportunidad de descubrir nuevos géneros y artistas. Además, esto pone un desafío para la diversidad musical, ya que los oyentes se ven expuestos a un repertorio limitado que puede ayudar a las

tendencias en lugar de apoyar a la diversidad de música que se encuentra en ella.

Esta desigualdad no solo impresiona a los artistas y oyentes, sino que también tiene implicaciones para la propia plataforma. Ya que al no promover la diversidad, corre el riesgo de perder una base de usuarios de nuevas experiencias musicales. En este sentido, este problema se convierte en una necesidad para asegurar no solo la salud de la plataforma, sino también la vida útil musical como forma de arte.

## ***RECURSOS DISPONIBLES:***

### **° TECNOLOGIAS Y HERRAMIENTAS:**

En este informe se utilizaran las herramientas antes vistas, como: Python además de la importación de librerías, como es el caso de, pandas para la manipulación de datos , matplotlib para la creación de graficos en dos dimensiones como lo pudimos apreciar en nuestra base de datos, y plotly para la eliminación de los datos atípicos de la base malita.

### **° DATOS :**

- Id: es el identificador único con el que se puede encontrar una canción en esta plataforma de Spotify.
- Name: es la nombración de caracteres para encontrar con facilidad una canción
- Genre: es la clasificación de la diversidad de las canciones musicales
- Artist: Es el nombre de los autores de las propias pistas.
- Album: colección de un reportorio de dichas canciones
- Popularity: es la puntuación de la visualización de rango estadístico que puede llegar a tener una canción en la plataforma, esto se puede calificar entre 0 – 100, donde 100 es el rango mas alto.
- Duration\_ms: es el periodo de tiempo en milisegundos que suelen tener la música promedio.
- Explicit: es el que indica si la canción o pista contiene dichos sinónimos de ellos.

## **Hipótesis:**

### **↳ Si la música tiene contenido explícito que relación puede tener con la popularidad:**

Las letras explícitas en las canciones aumenta su atractivo ya que estas tienden a resonar con audiencias jóvenes que buscan autenticidad y un sentido de rebeldía en su música.

### **↳ Que relación existe entre género y popularidad:**

El género puede afectar la popularidad, dado que ciertos de ellos , como el reguetón o el pop, se alinean más con las tendencias culturales y preferencias de las audiencias actuales, resultando en un mayor número de reproducciones y reconocimiento en listas de éxitos.

### **↳ Duración y la popularidad:**

La duración de las canciones influye en la popularidad; ya que aquellas que tiene un formato más corto, alrededor de 3 minutos, son más propensas a volverse virales, ya que se adaptan mejor a los hábitos de escucha rápida de las plataformas digitales como Spotify.

## **DEFINICION DE STAKEHOLDERS CLAVE:**

### **↳ Marketing:**

Hacer estrategias que puedan desarrollar mas audiencia para los artistas menos escuchados hacia los oyentes, mediante la popularidad y el genero.

### **↳ Oyentes:**

El impacto que pueden llegar a tener las influencias del marketing para que pueda influenciar la popularidad y genero para aumentar la cantidad de oyentes en Spotify.

## **PREGUNTAS CLAVE:**

- ▼ ¿Cuáles son los factores que determinan la popularidad de una canción en Spotify?
- ▼ ¿Cómo afecta la visibilidad de canciones menos conocidas en el crecimiento de nuevos artistas?
- ▼ ¿Qué papel juegan los algoritmos de recomendación por parte del marketing en este proceso?
- ▼ ¿Qué diferencias existen en la promoción de artistas menos escuchados a artistas ya populares en la plataforma?
- ▼ ¿Qué estrategias de marketing podrían implementarse para aumentar la visibilidad de canciones menos populares?
- ▼ ¿Qué impacto tiene la falta de diversidad musical en la experiencia de los oyentes en Spotify?
- ▼ ¿Cómo se podrían evaluar los cambios en la visibilidad y el alcance de los artistas menos escuchados?
- ▼ ¿Qué papel juegan las listas de reproducción por los oyentes en la promoción de nueva música?
- ▼ ¿Qué papel juegan las tendencias culturales y sociales en la validacion de nuevos géneros y
- ▼ ¿Cómo influyen las interacciones sociales en la popularidad de una canción en Spotify?

## **FUENTES DE DATOS IDENTIFICADAS**

- ➊ Los datos de Reproducción en Spotify
- ➋ Estadísticas de Listas de Reproducción
- ➌ Análisis de Contenido de Letras
- ➍ Datos Demográficos de los oyentes
- ➎ Comentarios y Reseñas de los oyentes
- ➏ Análisis de las Redes Sociales
- ➐ Encuestas de Preferencias Musicales
- ➑ Informes de Tendencias de la Industria
- ➒ Datos de Críticas Musicales

# **JUSTIFICACIÓN DEL PROYECTO**

El análisis de la popularidad de las canciones en plataformas de streaming como Spotify es crucial en el contexto actual de la industria musical, donde la competencia es muy grande y las dinámicas de consumo están en constante evolución. A medida que los oyentes se vuelven más selectivos y buscan experiencias personalizadas como comprender los factores que influyen en la popularidad de una canción que se convierte en una necesidad estratégica tanto para artistas como para sellos discográficos y plataformas de streaming.

Este proyecto se centra en examinar elementos como el contenido explícito, el género musical y la duración de las canciones, como aspectos que han demostrado tener un impacto significativo en la opinión del público. Al identificar cómo estas variables interactúan con las preferencias del oyente, se pueden desarrollar estrategias más efectivas para promocionar nuevos lanzamientos y aumentar la visibilidad de artistas menos escuchados.

Además, este análisis no solo beneficia a los artistas y las plataformas, sino que también ayuda con la experiencia del consumidor. Al proporcionar a los oyentes acceso a una gama más amplia de música, se aumenta la diversidad cultural y se les permite descubrir nuevos talentos que, de otro modo, podrían pasar desapercibidos.

En resumen, este proyecto no solo busca entender la dinámica de la popularidad musical en un entorno digital, sino que también tiene el potencial de transformar la forma en que se produce, promociona y consume la música, beneficiando a todos los autores involucrados en el desarrollo musical.

## ***CUANTOS DATOS Y QUE TIPO SON:***

*El data frame cuenta con 7303 filas y 8 columnas.*

- ﴿ ID cuenta con 6187 valores únicos que son tipo carácter
- ﴿ NAME cuenta con 4518 valores únicos y es tipo carácter
- ﴿ GERNER cuenta con 126 valores únicos y son tipo carácter
- ﴿ ARTIST cuenta con 4651 valores únicos y son tipo carácter
- ﴿ ALBUM cuenta con 4949 valores únicos y son tipo carácter
- ﴿ POPULARITY cuenta con valores que van de 0 a 90 y su tipo es decimal
- ﴿ DURATION cuenta con datos que van de los 30,1 k a 3,60 m y son tipo decimal
- ﴿ EXPLICIT cuenta con 1194 en true y 5106 el false y es tipo carácter