

# MEMORIA DEL PROYECTO: ANÁLISIS Y STORYTELLING DE DATOS DE SALUD EN APPLE Y MÚSICA EN SPOTIFY

---

## 1. Descripción de los Datos y Objetivo del Estudio

Este estudio analiza datos personales de salud obtenidos de Apple Health y datos musicales extraídos de Spotify. El objetivo principal fue explorar posibles tendencias en medidas de salud como actividad física, frecuencia cardíaca, tendencias musicales en la evolución de los géneros, artistas y canciones, además de posibles correlaciones entre la exposición al ruido y las preferencias musicales, empleando técnicas de limpieza, transformación y visualización de datos en **Power BI**. Se buscó responder preguntas clave como:

- ¿Cómo varían las métricas de salud según el paso del tiempo?
- ¿Cómo varían las métricas musicales según el paso del tiempo?
- ¿Qué dispositivos capturaron mayor cantidad de registros y cómo afecta la calidad de los datos?
- ¿Existe una relación entre la exposición al ruido y la música escuchada?

Los datos se estructuraron en múltiples dimensiones que incluyen **frecuencia cardíaca, pasos diarios, energía quemada, exposición sonora, géneros musicales y rankings de canciones**, abarcando los años **2022, 2023 y 2024**.

---

## 2. Metodología de Trabajo

El trabajo se desarrolló en varias fases:

### 2.1. Procesamiento y Limpieza de Datos

Se utilizó **Python (Jupyter Notebook)** para transformar los datos antes de importarlos a Power BI. Algunas acciones clave fueron:

- **Tratar variables o eliminar con valores nulos** dependiendo el tipo de variable
- **Uniformización de géneros musicales** para consolidar categorías específicas.
- **Creación de métricas derivadas** como día de la semana y tipo de actividad.
- **Filtrado de valores extremos o inconsistentes** en métricas de salud.

- **Uniformización de actividades** para categorizarlas por tipos.
- **Conversión de fechas** a un formato estándar.

Problema encontrado: **Diferencias en los nombres de géneros musicales entre años.**

- **Solución:** Se implementó un **diccionario de mapeo** para agrupar categorías afines (ej. "Trap Latino" dentro de "Trap").

Problema encontrado: **Las variables de salud no especifican a que categoría corresponden.**

- **Solución:** Se creó una **columna de clasificación de tipo de actividad** basada en el significado de cada variable, agrupando los datos en categorías como "Gasto Energético", "Frecuencia Cardíaca", "Actividad Física", "Exposición al Ruido" y "Sueño". Esto permitió diferenciar métricas como **frecuencia cardíaca, calorías quemadas y pasos diarios**, facilitando su análisis comparativo en Power BI.

- 

## 2.2. Modelado y Visualización en Power BI

Se construyó un **dashboard interactivo** con visualizaciones clave:

- **Comparación de exposición al ruido vs. porcentaje acústico de canciones** (gráfico de líneas con colores contrastantes: azul #1E90FF para ruido y verde #46DE7D para acústica).
- **Distribución de géneros musicales a lo largo de los años** (gráfico de cintas para mostrar la evolución de preferencias).
- **Tendencias de actividad física y frecuencia cardíaca** (gráfico de área y KPIs de salud).
- **Proporción de registros capturados por iPhone vs. Apple Watch** (gráfico de anillo con colores diferenciados).

---

## 3. Conclusiones y Reflexiones

El análisis permitió obtener **hallazgos relevantes** sobre la relación entre la música y la salud personal: **✓ El 88.69% de los registros fueron capturados por el iPhone**, lo que sugiere que los datos de salud podrían no ser tan precisos como los obtenidos

con el Apple Watch. ✓ **Los picos de exposición al ruido ocurrieron en días laborales**, mientras que los fines de semana hubo una reducción. ✓ **Se observó una relación evolución de géneros musicales:** años más recientes presentaron incrementos notables en el género Tropical Pop, reducciones en el género Pop Latino y fluctuaciones en géneros como Trap.

### **Futuras Mejoras**

Podríamos enriquecer los análisis con datos adicionales:

- Incorporar **datos de reproducción musical con marca de tiempo** para correlacionar canciones específicas con actividad física.
- Recopilar datos de calidad del sueño para buscar **patrones de sueño** y su relación con la música escuchada antes de dormir.
- Aplicar **modelos predictivos** para estimar el tipo de género a escuchar en función de patrones de sueño, exposición al ruido y actividad física realizada.