

Memoria Trabajo Grupal

Autores

Hevia González, Lucía
Crespo Rivas, Jorge
Pérez Alonso, Clara

Bases de Datos Analíticas
Grado en Ciencia e Ingeniería de Datos
Universidade da Coruña

Curso 2025-2026

Índice

1. Tableau	2
1.1. Introducción	2
1.2. Preparación de la base de datos	2
1.3. Descripción del dashborad	3
2. Power BI	5
2.1. Introducción	5
2.2. Preparación de la base de datos	5
2.3. Descripción del dashborad	6

1. Tableau

1.1. Introducción

La herramienta de visualización principal empleada para esta primera fase del proyecto ha sido Tableau Desktop, donde hemos ampliado el alcance del estudio dividiéndolo en dos vertientes complementarias: el análisis a gran escala de las tendencias de consumo geográfico-temporal y el análisis específico del perfil acústico ("ADN") de los géneros musicales en función de las características individuales de sus canciones.

Mientras que el primer enfoque busca responder a dónde y quién domina el mercado, analizando la distribución de streams por países y la hegemonía de ciertos artistas a lo largo de 2017, el segundo enfoque se centra en el cómo, desgranando las características técnicas (energía, positividad, tempo) que definen a los géneros más exitosos. La música no es un producto estático, ya que su consumo varía drásticamente según la región y la estacionalidad, y su composición interna sigue patrones cuantificables que determinan su viralidad.

El objetivo de esta sección es presentar una solución de inteligencia de negocios que permita, por un lado, monitorear la evolución de los grandes éxitos mundiales mediante mapas interactivos y rankings dinámicos, y por otro, explorar estadísticamente qué hace que una canción sea bailable o acústica mediante diagramas de dispersión y determinadas gráficas estadísticas. Gracias a la interactividad nativa de Tableau, el usuario puede transitar desde una visión global del mercado hasta el detalle técnico de una canción específica, facilitando la detección de patrones complejos que pasarían desapercibidos en tablas de datos tradicionales.

1.2. Preparación de la base de datos

Para garantizar la viabilidad y el rendimiento de los dashboards en Tableau, fue imprescindible ejecutar un proceso riguroso de transformación y limpieza utilizando Python (Pandas) sobre dos conjuntos de datos de naturaleza muy distinta.

El primer conjunto, denominado "Charts" (uno de los propuestos para la práctica), presentaba un reto de volumen gigante (aproximadamente 32GB de información con registros diarios). Cargar este archivo directamente en la herramienta de visualización era inviable ya que Tableau ni siquiera daba opción. Por tanto, la estrategia de limpieza se centró en la agregación y reducción de dimensionalidad.

En primer lugar, se eliminó el "ruido", descartando así variables irrelevantes para el análisis visual, tales como URLs, indicadores de tendencia diaria (trend) y tipos de ranking (rank), ya que el foco se puso en el volumen absoluto de reproducciones.

También se agruparon los datos temporal y geográficamente, pasando de un registro diario a una agregación mensual independiente en cada país (agrupando las reproducciones por Título, Artista, Región y Mes, sumando los streams totales). Esto redujo el dataset de 26 millones de filas a un archivo consolidado (charts_grouped.csv) de poco más de 1 millón de registros, optimizado para una respuesta ágil en Tableau.

Además, se ejecutó un tratamiento de fechas simple: se normalizó la variable temporal, convirtiendo las fechas completas en categorías mensuales (enero, febrero, etc.) para facilitar el filtrado estacional.

El segundo conjunto, "Attributes & Popularity" (obtenido de Kaggle), contenía metadatos técnicos de más de 114.000 pistas. Aquí, los problemas eran de consistencia y dispersión.

Para tratar estos datos se normalizaron los géneros, ya que el dataset original contenía una fragmentación excesiva de estos que haría inviable la comprensión visual en la mayoría de los gráficos (manteniendo solo el top25 de géneros con más presencia). Además, denotamos una incongruencia técnica no relevante, Tableau identificaba las métricas o atributos de las canciones como un número entre 0 y 1000 en vez de entre 0 y 1 como se especificaba en el dataset, pero

no se tomaron medidas al respecto ya que, al ser para todas las variables de este tipo el mismo caso, no se presentaban variaciones en las escalas comparativas (los gráficos y tablas contenidos en el dashboard son exactamente iguales independientemente de que la escala sea 0-1 o 0-1000). Por último, se transformaron las unidades de la duración de las canciones (ya que nadie se guía por milisegundos en cuanto a duración de una canción se refiere).

1.3. Descripción del dashborad

El proyecto culmina en la creación de dos dashboards interactivos en Tableau, cada uno diseñado para abordar una dimensión distinta del negocio musical.

Dashboard 1: Tendencias Globales y Ranking de Artistas

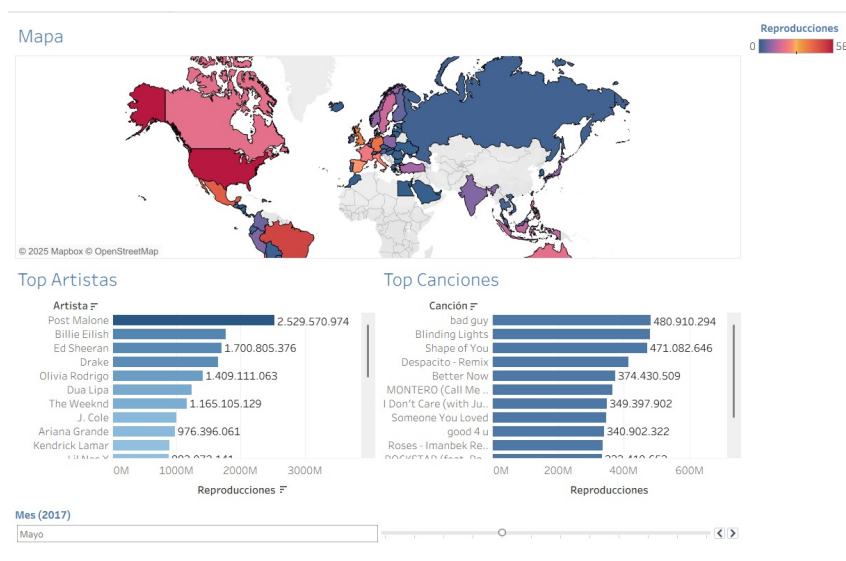


Figura 1: Dashboard 2 de Tableau

Este panel ofrece una visión ejecutiva del consumo musical. Su elemento central es un mapa de calor geográfico que representa la intensidad de reproducciones por país. Esta visualización permite identificar de un vistazo los mercados clave (como EE. UU. o países de Latinoamérica) frente a regiones con menor penetración.

Junto al mapa, se presenta un ranking de artistas dinámico mediante un gráfico de barras horizontales. Esta gráfica no es estática; incluye una jerarquía interactiva que permite al usuario hacer desglose desde el nombre del artista hasta sus canciones individuales, revelando qué temas específicos impulsaron su éxito y son la razón de su popularidad en ese territorio. La interactividad se gestiona mediante una Barra Deslizante Temporal situada en la zona inferior. Al desplazar el control a través de los meses del año, tanto el mapa como el ranking se actualizan en tiempo real, permitiendo visualizar la "película" del año musical: cómo ciertos artistas dominan el verano en x territorio y otros surgen hacia final de año.

Por último, se incorpora un gráfico de barras con el Top de canciones, donde se presentan los temas más reproducidos durante 2017. Esta visualización complementa el ranking de artistas y permite observar qué canciones alcanzaron mayor impacto en ese año, ofreciendo una perspectiva completa del comportamiento del consumo musical en Spotify.

Dashboard 2: ADN Musical y Análisis de Géneros

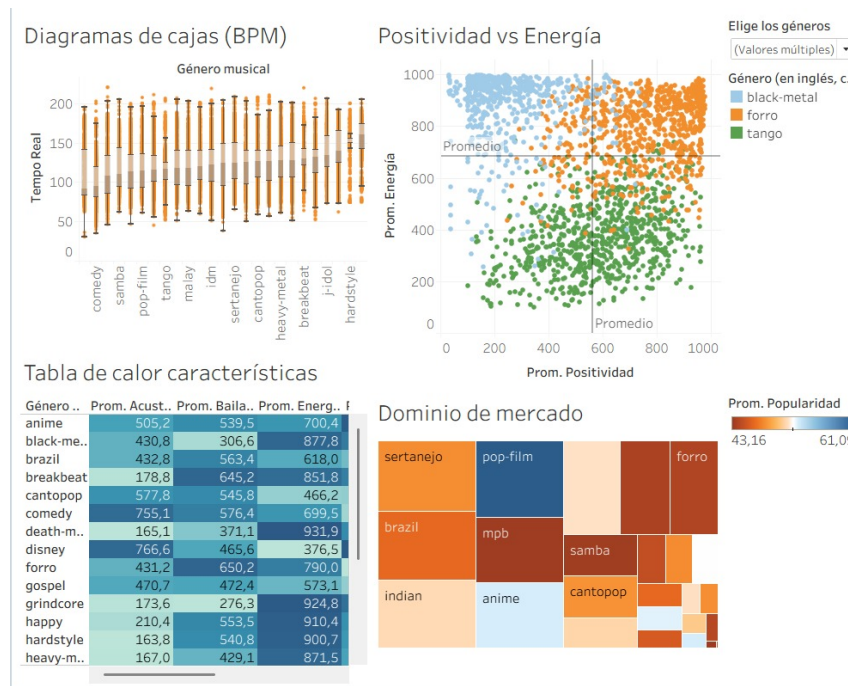


Figura 2: Dashboard 1 de Tableau

Este segundo panel presenta un enfoque analítico-descriptivo compuesto por cuatro visualizaciones avanzadas e interconectadas que permiten profundizar en la estructura técnica de la música. La narrativa visual se articula entorno al gráfico de dominio de mercado, un Tree Map que utiliza rectángulos anidados para ilustrar qué géneros producen una mayor cantidad de éxitos. En esta representación, el tamaño de cada elemento indica el volumen de canciones con una popularidad superior a 40, mientras que el color refleja su popularidad media, funcionando a su vez como el filtro principal que condiciona el resto del dashboard. Complementando esta visión, el Mapa de Emociones despliega un gráfico de dispersión que cruza las variables de Positividad (Valence) y Energía para dividir el espectro musical en cuadrantes emocionales, facilitando el posicionamiento psicológico de cada género entre estados opuestos como la tristeza o la euforia. Este último cuenta con un menú desplegable en el que se puede seleccionar interactivamente qué géneros comparar en el gráfico.

Para profundizar en los aspectos rítmicos y estructurales, se incluyen Diagramas de Caja que examinan la distribución del Tempo (BPM) ordenados de menor a mayor velocidad, lo que permite comprender la variabilidad rítmica y evidenciar contrastes notables, como el rango compacto y rápido de la música black metal frente a la tranquilidad de, por ejemplo, el tango. Finalmente, se integra una tabla de calor de características, diseñada como una matriz que compara los promedios de atributos clave como acusticidad, bailabilidad e instrumentalidad por género, utilizando leyendas de color independientes por columna para agilizar la detección visual de los rasgos distintivos que definen la identidad de cada estilo musical.

La integración de estas cuatro gráficas permite realizar tanto un análisis profundo sobre un género concreto como la comparación de varios géneros y sus respectivas tendencias características a la vez.

2. Power BI

2.1. Introducción

La segunda herramienta de visualización empleada en el proyecto es Power BI, donde centramos nuestro estudio en el tema de la estacionalidad de los streams musicales, analizando cómo varía el volumen de reproducciones a lo largo de los meses y los años, y cómo estos cambios se relacionan con los diferentes géneros musicales y con los artistas más dominantes en el período comprendido entre 2017 y 2021. La estacionalidad es un fenómeno especialmente relevante en el ámbito musical, ya que el consumo puede cambiar en función de diversos factores externos, como los períodos de vacaciones, los lanzamientos de nuevos álbumes, las tendencias culturales emergentes o los hábitos sociales asociados a determinadas épocas del año. Todo ello provoca que ciertas canciones, géneros o artistas alcancen picos de popularidad en momentos concretos, mientras que en otros períodos experimentan descensos pronunciados.

El objetivo principal de la visualización desarrollada en Power BI es mostrar estos patrones de manera clara, interactiva y comparativa, permitiendo explorar la información desde distintas perspectivas: filtrando por año, por mes, por género musical o por artista concreto. Gracias a la interactividad de Power BI, donde se puede identificar variaciones significativas en el comportamiento del consumo musical y comprender qué factores pueden influir en los aumentos o disminuciones del número de streams. En conjunto, el dashboard ofrece una visión completa del fenómeno de la estacionalidad musical, proporcionando un análisis detallado del consumo en Spotify a lo largo del tiempo.

2.2. Preparación de la base de datos

Para poder analizar la estacionalidad musical en Power BI fue necesario realizar previamente un proceso de limpieza, transformación y estandarización del conjunto de datos utilizando Python. El dataset original presentaba problemas: formatos erróneos, géneros almacenados como listas dentro de cadenas de texto, fechas sin transformar, valores numéricos incorrectos y entradas erróneas tanto en artistas como en géneros. Por ello, el primer paso fue cargar y explorar los datos para identificar estos problemas y planificar su corrección.

En primer lugar, se transformó la columna Genre, que contenía listas codificadas como texto, convirtiéndolas en listas reales y extrayendo únicamente el primer género relevante. Posteriormente, se convirtió la columna Date en formato datetime y se generaron las columnas Year y Month, importantes para el análisis. Tras ello, se eliminaron columnas innecesarias como Date y Position, y se comprobó que los valores de Streams estuvieran correctamente convertidos a formato numérico.

Se aplicó también una limpieza adicional de texto para corregir valores problemáticos: artistas representados por números, nombres que contenían años, géneros numéricos o entradas que incluían mensajes de error del servidor. Estos valores fueron eliminados o sustituidos al no aportar información válida para el análisis.

Uno de los pasos más importantes fue la normalización semántica de los géneros musicales. El dataset contenía muchos tipos de géneros diferentes, por lo que se agruparon en categorías principales como pop, rap, trap, latin, edm, jazz o urban. Esta normalización permitió reducir la complejidad del conjunto de géneros.

Finalmente, se realizó una agregación mensual, sumando los streams por canción, artista, género, año y mes, generando así un archivo compacto y estructurado llamado `spotify_mensual_limpio.csv`. Este dataset final fue el utilizado para construir el dashboard interactivo centrado en el análisis de la estacionalidad musical.

2.3. Descripción del dashboard

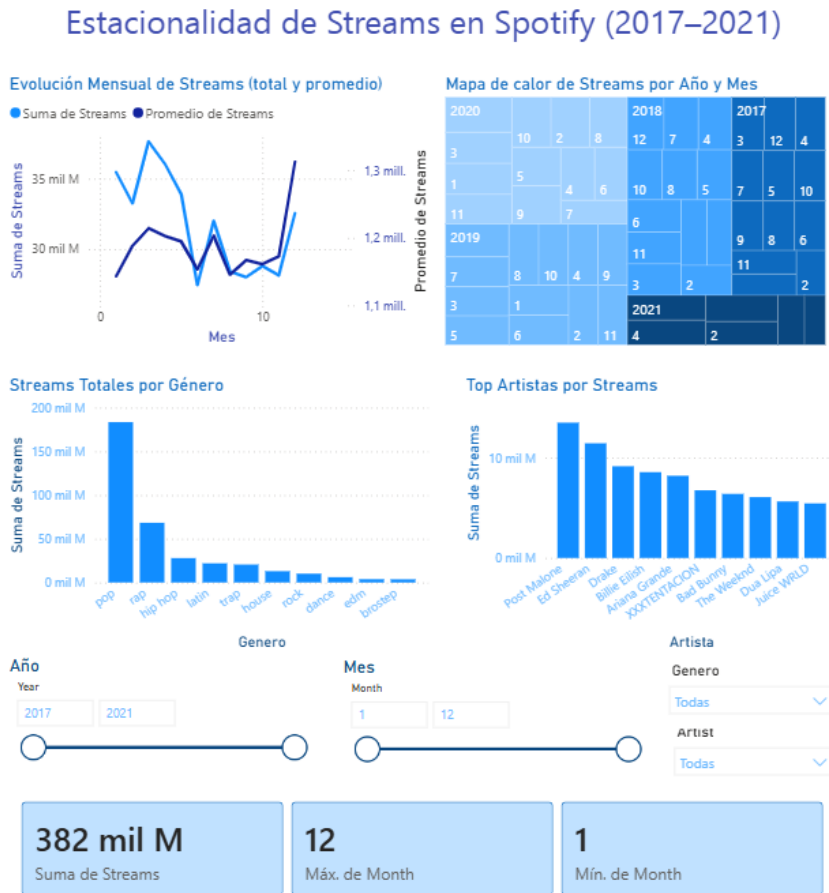


Figura 3: Dashboard de Power BI

Nuestro dashboard presenta diferentes tipos de visualizaciones, cada una orientada a representar aspectos clave de la estacionalidad y del comportamiento del consumo musical.

La primera visualización es una gráfica de línea que muestra la Evolución Mensual de Streams (total y promedio). Esta representación permite observar la evolución mensual del consumo musical durante todo el período analizado, identificar tendencias ascendentes o descendentes, detectar meses con comportamientos atípicos y reconocer patrones estacionales recurrentes, como picos de actividad en determinadas épocas del año. La combinación de ambas medidas, suma y promedio, ofrece una perspectiva más completa: la suma refleja el volumen total mensual, mientras que el promedio suaviza las variaciones y facilita la interpretación de la tendencia general.

La segunda visualización es un Mapa de calor de Streams por Año y Mes, que representa la suma de streams para cada combinación de mes y año. Su principal utilidad está en la capacidad para identificar de un vistazo cuáles son los meses con mayor nivel de actividad dentro de cada año y cómo evoluciona el consumo musical entre 2017 y 2021. Las diferentes intensidades de color permiten comparar fácilmente períodos de alto consumo con otros de menor actividad, revelando patrones cíclicos, tendencias repetidas y posibles anomalías temporales.

Junto a ello, se presenta un gráfico de barras que muestra los Streams Totales por Género, pensado para identificar qué artistas concentran el mayor volumen de escuchas en el periodo 2017–2021. Este gráfico resulta fundamental porque permite interpretar los picos de actividad observados en otras visualizaciones: muchos incrementos bruscos en el consumo mensual pueden deberse al lanzamiento de un nuevo álbum, a un fenómeno viral o a la popularidad de determinados artistas. Con

esta representación, se puede reconocer de inmediato cuáles son los artistas que dominan Spotify. También facilita comparaciones entre artistas, tanto en términos absolutos como relativos.

El dashboard incluye también un gráfico de barras con la Top Artistas por Streams, pensado para identificar qué artistas concentran el mayor volumen de escuchas en el periodo 2017–2021. Esta visualización es fundamental para interpretar los picos de actividad detectados en las gráficas anteriores, ya que muchos incrementos mensuales pueden darse por lanzamientos por la popularidad de determinados artistas. Además, permite comparar artistas y entender mejor el papel que desempeñan en la estacionalidad global del consumo musical.

Para proporcionar un alto nivel de interactividad, incorporamos segmentadores que permiten filtrar los datos de forma dinámica según cuatro variables clave: Año, Mes, Género y Artista. Estos filtros posibilitan la exploración de subconjuntos concretos de información, comparar la estacionalidad entre distintos años, observando cómo evoluciona un género específico mes a mes o analizar el comportamiento de un artista concreto en el contexto global.

Finalmente, el dashboard integra un conjunto de indicadores, diseñados para ofrecer una lectura rápida de las magnitudes más importantes del análisis. Entre ellos destacan el Total de Streams en el período seleccionado, así como el mes con mayor actividad y el mes con menor actividad. Estos indicadores sirven como puntos de referencia clave, permitiendo identificar de forma instantánea los momentos de mayor y menor consumo musical y facilitando la interpretación general del comportamiento estacional.