

1. Proceso de extracción de datos

El proceso de extracción de datos comenzó con la obtención de un archivo en formato .xml desde Rekordbox, un software especializado en la organización de colecciones musicales y tracks utilizados por DJs. Este archivo contiene la base de datos interna del sistema, incluyendo información detallada tanto de la colección musical como del historial de DJ sets.

Extracción de datos desde Rekordbox

En una primera etapa, se extrajo la colección completa de tracks a partir de la estructura jerárquica (árbol) presente en el archivo XML. Para cada track se recuperaron todos los nodos y atributos disponibles, tales como nombre del artista, nombre de la canción, fecha de incorporación, BPM, entre otros metadatos relevantes proporcionados por Rekordbox.

Posteriormente, se extrajo el historial de DJ sets, el cual se encuentra representado en Rekordbox como un tipo especial de playlist. Este historial incluye la fecha de cada DJ set y los identificadores internos de los tracks utilizados, permitiendo vincular cada presentación con los elementos de la colección musical.

Limpieza inicial de datos

Una vez finalizada la extracción, se llevó a cabo un proceso de limpieza de datos, orientado a eliminar valores atípicos, registros inconsistentes y elementos irrelevantes o corruptos que pudieran afectar la calidad del análisis posterior.

Enriquecimiento de datos mediante la API de Spotify

Con la información limpia proveniente de Rekordbox, se realizó una integración con la API de Spotify para enriquecer los datos musicales. Para ello, se utilizaron dos campos clave extraídos de Rekordbox: nombre del artista y nombre del track. Estos campos se emplearon para construir consultas (queries) enviadas a la API de Spotify.

Los resultados devueltos por la API fueron evaluados mediante un índice de similitud textual, calculado a partir de funciones de comparación de texto implementadas en Python. Este enfoque permitió determinar el mejor match entre los registros de Rekordbox y Spotify, asegurando que la coincidencia seleccionada correspondiera efectivamente al track correcto. Paralelamente, los resultados fueron categorizados para monitorear y evaluar la calidad del proceso de matching.

Limpieza y validación de datos de Spotify

Una vez obtenida la información desde Spotify, se realizó una segunda fase de limpieza de datos. En esta etapa se eliminaron registros inconsistentes, coincidencias con baja similaridad y resultados que, a simple vista, no representaban correctamente el track original. El objetivo principal fue asegurar la confiabilidad de los datos enriquecidos, poniendo especial énfasis en la obtención correcta de los Spotify Track IDs.

Obtención del número de reproducciones mediante web scraping

Dado que el contador de reproducciones no está disponible a través de la API oficial de Spotify, se implementó un proceso de web scraping utilizando Selenium y BeautifulSoup. A partir de cada Spotify Track ID válido, se accedió a la página correspondiente del track y se extrajo el número de reproducciones de manera automatizada.

Consolidación final de la información

Finalmente, se realizó un proceso de consolidación de datos, generando dos conjuntos principales:

Master Collection: archivo que integra la colección completa de Rekordbox junto con toda la información enriquecida obtenida desde Spotify.

Master History: archivo que cruza el historial de DJ sets con los datos de Spotify, permitiendo analizar el uso de tracks en presentaciones junto con sus métricas externas.

Este proceso permitió construir una base de datos consolidada, enriquecida y validada, adecuada para análisis posteriores y visualización de información.

2. Proceso de extracción de datos

El dashboard desarrollado como parte de la práctica se encuentra estructurado en tres secciones principales, cada una orientada a analizar distintos aspectos del desempeño y características de la colección musical y del historial de DJ sets.

1. Sección Colección

Está enfocada en el análisis global de la colección musical del DJ, permitiendo comprender su tamaño, composición y características técnicas.

Métricas generales de la colección, se presentan cuatro métricas clave:

- Cantidad total de tracks, que permite dimensionar el tamaño de la colección.
- Cantidad de géneros, utilizada como indicador de diversidad musical.
- Duración total de la colección (en horas), relevante para evaluar la amplitud del material disponible.
- Promedio de BPM, que entrega una primera aproximación al perfil musical del DJ en términos de energía y estilo.

Estas métricas funcionan como una vista ejecutiva inicial que resume el estado general de la colección.

Evolución temporal de la colección, Se incluye un gráfico de series de tiempo denominado Crecimiento total de la colección y Top 4 géneros, el cual muestra la evolución histórica del número de tracks, tanto a nivel total como para los cuatro géneros más representativos.

Esta visualización permite analizar:

- El ritmo de crecimiento de la colección a lo largo del tiempo.
- Cambios en las preferencias musicales del DJ.
- La consolidación o declive de determinados géneros en distintas etapas.

Duración de tracks por género

Mediante boxplots se representa la distribución de la duración de los tracks para los Top 4 géneros de la colección. Esta visualización permite identificar:

- Rangos típicos de duración por género.
- Presencia de outliers.
- Grado de variabilidad entre estilos musicales.

Esta información es relevante para la planificación de DJ sets, ya que contar con tracks de distintas duraciones facilita la adaptación a diferentes contextos y momentos de una presentación.

Distribución armónica (Rueda de Camelot)

Se incorpora la rueda de Camelot, representando las claves musicales (1 a 12, A y B) y la frecuencia de tracks por cada una. Esta visualización permite evaluar el balance armónico de la colección.

- Un análisis de esta distribución resulta clave para:
- Detectar sobre-representación o escasez de determinadas claves.
- Guiar decisiones de adquisición de nuevos tracks.
- Facilitar mezclas armónicas más variadas y coherentes durante los sets.

Tracks por formato de archivo

Se presenta una visualización que muestra la cantidad de tracks por formato de archivo (MP3, AIFF, WAV y FLAC). Esta información es fundamental desde un punto de vista técnico, ya que los formatos sin compresión o de mayor calidad (AIFF, WAV, FLAC) son preferibles para presentaciones en sistemas de sonido profesionales.

Esta visualización permite identificar posibles riesgos en términos de calidad de audio y orientar acciones para mejorar la colección.

Distribución de BPM

Se incluye un histograma de BPM, el cual permite analizar el perfil energético del DJ. BPMs más altos suelen asociarse a géneros más intensos (por ejemplo, techno o hard techno), mientras que BPMs más bajos se relacionan con estilos más relajados (como Afro House u Organic House).

Tanto esta visualización como la de formatos pueden filtrarse por género, lo que habilita un análisis más granular de la colección.

2. Sección Colección vs Spotify

La segunda sección se enfoca en evaluar la calidad del enriquecimiento de datos obtenido desde Spotify y en analizar la popularidad de los tracks.

Se presentan métricas clave como:

- Cantidad de tracks encontrados en Spotify.
- Porcentaje de la colección con match exitoso.
- Similaridad promedio.
- Cantidad de tracks con duración coincidente entre Rekordbox y Spotify.

Estas métricas permiten evaluar la confiabilidad del proceso de integración entre ambas fuentes de datos.

Distribución del índice de similaridad

Se incluye un histograma del índice de similaridad, el cual permite identificar qué tan confiables son los matches realizados. Una distribución concentrada hacia valores cercanos a 1 indica alta calidad en la extracción de información.

Popularidad en Spotify

Se presenta un histograma del índice de popularidad de Spotify. Para DJs de música electrónica, mantener un nivel de popularidad promedio relativamente bajo puede ser deseable, ya que está asociado a una selección musical más underground y diferenciada.

Top 5 de tracks con más reproducciones

Se muestra un ranking de los Top 5 tracks con más reproducciones en Spotify, con un sistema de colores (verde, amarillo y rojo) que indica el nivel de confianza del match. Esta visualización es principalmente informativa y permite identificar los tracks más populares dentro de la colección.

Adicionalmente, esta sección cuenta con filtros por género para facilitar el análisis segmentado.

3. Sección Historial de DJ Sets

La tercera sección está orientada al análisis del uso real de la colección durante las presentaciones.

Métricas generales del historial, Se incluyen indicadores como:

- Cantidad total de DJ sets realizados.
- Porcentaje de la colección utilizada.
- Duración promedio de los DJ sets.
- Cantidad promedio de tracks por set.

Estas métricas permiten evaluar el grado de aprovechamiento de la colección.

Evolución temporal de DJ sets

Un gráfico de series de tiempo muestra la cantidad de DJ sets realizados por año, permitiendo identificar períodos de mayor o menor actividad.

Análisis anual de BPM y duración, Mediante filtros por año, se presentan:

- El BPM promedio de los DJ sets.
- Un histograma de la duración de los sets.

Estas visualizaciones permiten analizar la evolución del estilo del DJ a lo largo del tiempo, evidenciando cambios de género, intensidad y estructura de las presentaciones.

Top 5 de tracks más utilizados

Finalmente, se presenta el Top 5 de tracks más utilizados en los DJ sets, identificando aquellos tracks que funcionan como “cartas seguras” o elementos recurrentes en la construcción de las presentaciones.

3. Explicación de los insights/conclusiones alcanzados y reflejados en las visualizaciones realizadas.

A continuación, se presentan los principales insights y conclusiones derivados del análisis visual, organizados por sección del dashboard.

1. Insights de la sección Colección

En primer lugar, el tamaño total de la colección no permite, por sí solo, extraer conclusiones cualitativas relevantes. Sin embargo, la cantidad de géneros sí aporta información

significativa. La colección cuenta con 16 géneros distintos, lo cual representa un nivel de diversidad elevado. Considerando que la mayoría de los DJs profesionales suelen especializarse en un rango reducido de entre 4 y 5 géneros, este resultado indica una colección amplia y versátil.

Respecto al promedio de BPM, con un valor cercano a 125 BPM, se observa un perfil musical equilibrado y representativo de estilos como house, tech house y techno suave. Este valor sugiere que el DJ no se posiciona en extremos de energía, sino que mantiene un estilo consistente y ampliamente aceptado dentro de la escena electrónica.

En cuanto al crecimiento de la colección, se evidencia una evolución altamente positiva. La cantidad total de tracks aumentó de 362 en octubre de 2021 a 710 en septiembre de 2025, lo que implica que la colección casi se duplicó en un período aproximado de cuatro años. Este crecimiento sostenido refleja una inversión constante en la expansión del material musical.

Analizando los géneros más representativos, se observa que Tech House es el género predominante, seguido por Techno, Melodic Techno, Minimal y Deep Tech. En relación con la duración de los tracks, el promedio se sitúa alrededor de los 400 segundos (aproximadamente 6 minutos), un valor estándar dentro de la música electrónica. Se identifica que los tracks de Techno y Melodic Techno tienden a tener mayor duración, mientras que los de Tech House y Minimal son relativamente más cortos, lo cual es coherente dada la cercanía estilística entre estos últimos géneros.

La distribución armónica, representada mediante la rueda de Camelot, muestra una clara concentración en la escala A, con una menor presencia de la escala B. Este comportamiento es consistente con prácticas comunes en la producción musical, ya que las claves menores suelen ser más utilizadas. No obstante, se detecta una ligera subrepresentación de la clave 12A, lo que podría ser considerado como una oportunidad de mejora para lograr un mayor balance armónico.

En relación con los formatos de archivo, se identifica una alta proporción de tracks en formato MP3, lo cual representa un punto crítico a mejorar. Desde una perspectiva técnica, sería recomendable incrementar la proporción de formatos de mayor calidad como AIFF, WAV o FLAC, especialmente considerando escenarios de reproducción en sistemas de sonido profesionales.

El histograma de BPM confirma que el DJ opera principalmente en torno a los 125 BPM, con una zona de confort aproximada entre 122 y 126 BPM. Fuera de este rango, la presencia de tracks es significativamente menor, lo que refuerza la idea de un estilo definido y consistente.

2. Insights de la sección Colección vs Spotify

El proceso de integración con Spotify presenta resultados positivos. Se logró un 95% de match entre la colección y Spotify, con un índice de similaridad promedio de 0.88, lo cual indica que la información enriquecida es confiable en términos generales.

El histograma del índice de similaridad se encuentra claramente cargado hacia valores cercanos a 1, reforzando la calidad del proceso de matching y validando el uso de la información proveniente de Spotify para análisis posteriores.

Respecto al índice de popularidad, la colección presenta un promedio de 26 puntos sobre 100, lo cual es un indicador positivo dentro del contexto del DJ de música electrónica. Este resultado sugiere una selección musical orientada al underground, con tracks poco masivos y diferenciados, alineados con el rol del DJ como curador musical.

El análisis de tracks por año de lanzamiento revela una concentración significativa de la colección entre los años 2016 y 2020, junto con una disminución notable a partir de 2023. Este patrón sugiere la necesidad de actualizar la colección incorporando lanzamientos más recientes para mantenerse alineado con la evolución del género.

3. Insights de la sección Historial de DJ Sets

Desde el punto de vista del uso de la colección, el DJ ha realizado un total de 43 DJ sets, utilizando aproximadamente el 47% de su colección. Este valor es relativamente bajo y representa una oportunidad clara de mejora, ya que idealmente se busca maximizar la rotación de tracks y evitar una dependencia excesiva de un subconjunto reducido.

En términos de duración promedio, los DJ sets tienen una media cercana a los 99 minutos, lo que sitúa al DJ por encima del estándar de los sets comerciales (alrededor de una hora) y más alineado con el formato de sets underground, que suelen extenderse entre una hora y media y dos horas.

El análisis de DJ sets realizados por año muestra una disminución significativa en 2025, con solo 6 sets realizados, lo cual constituye un punto de atención y una posible área de mejora en términos de actividad y exposición.

El BPM promedio de los DJ sets se mantiene consistente con la colección, nuevamente alrededor de 125 BPM, lo que indica coherencia entre el material disponible y su uso en presentaciones.

Finalmente, el análisis de la duración de los DJ sets confirma que la mayoría se concentra en torno a los 99–100 minutos, con algunos valores atípicos de menor duración que no representan el comportamiento general.

Conclusión

El análisis realizado a partir de las visualizaciones permitió comprender en profundidad la composición, evolución y uso de la colección musical y del historial de DJ sets. Se identificó una colección diversa, con un crecimiento sostenido en el tiempo y un perfil musical claramente definido en torno a los 125 BPM, alineado con géneros como house y techno. La integración con Spotify presentó altos niveles de confiabilidad y evidenció una orientación musical más cercana al underground que a lo comercial. No obstante, el análisis también reveló oportunidades de mejora, principalmente en la actualización de tracks recientes, la calidad de los formatos de audio y el aprovechamiento efectivo de la colección en los DJ sets. En conjunto, el dashboard se consolida como una herramienta útil para la toma de decisiones y la evolución artística del DJ.