

PR1 – Visualización de Datos

Spotify Tracks Dataset:

Conjunto de datos de canciones de Spotify con más de 125 géneros distintos y diferentes características asociadas a cada canción.

Autor: *Maharshi Pandya (2022)*

Fuente: *Kaggle* - <https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset>

Descripción datos (20 columnas):

- ☐ *track_id (String)*: ID de Spotify de la canción.
- ☐ *artists (String)*: Nombres de los artistas que interpretaron la canción. Si hay más de un artista, se separan por un ;.
- ☐ *album_name (String)*: Nombre del álbum en el que aparece la canción.
- ☐ *track_name (String)*: Nombre de la canción.
- ☐ *popularity (Integer)*: Valor entre 0 y 100, siendo 100 la más popular. Calculo mediante un algoritmo basado en el número total de reproducciones que ha tenido la canción y lo recientes que son esas reproducciones.
- ☐ *duration_ms (Integer)*: Duración de la pista en milisegundos (ms).
- ☐ *explicit (Boolean)*: Si la canción tiene o no letra explícita (verdadero = sí la tiene; falso = no la tiene o desconocido).
- ☐ *Danceability (Numeric)*: Adecuación de una pista para bailar. Se basa en una combinación de elementos musicales que incluyen el tempo, la estabilidad del ritmo, la fuerza del compás y la regularidad general. Valor de 0,0 es el menos bailable y 1,0 el más bailable.
- ☐ *Energy (Numeric)*: Percepción de intensidad y actividad. Medida de 0,0 a 1,0.
- ☐ *Key (Integer)*: Tonalidad de la canción. Los números enteros se asignan a los tonos utilizando la notación estándar Pitch Class. Por ejemplo, 0 = Do, 1 = Do#/Db, 2 = Re, y así sucesivamente. Si no se detecta ninguna tonalidad, el valor es -1.

- *loudness (Numeric)*: Sonoridad general de una pista en decibelios (dB).
- *mode (Boolean)*: Modalidad (mayor o menor) de una pista, el tipo de escala del que se deriva su contenido melódico. Mayor se representa con 1 y menor con 0
- *speechiness (Numeric)*: La locuacidad detecta la presencia de palabras habladas en una pista. Cuanto más exclusivamente hablada sea la grabación, más se acercará a 1,0 el valor del atributo.
 - Los valores superiores a 0,66 describen pistas que probablemente estén compuestas en su totalidad por palabras habladas.
 - Valores entre 0,33 y 0,66 describen pistas que pueden contener tanto música como voz, ya sea en secciones o en capas.
 - Valores inferiores a 0,33 representan probablemente música y otras pistas no habladas.
- *acousticness (Numeric)*: Medida de confianza de 0,0 a 1,0 para determinar si la pista es acústica. 1,0 representa una confianza alta en que la pista es acústica.
- *instrumentalness (Numeric)*: Predice si una pista no contiene voces. Cuanto más se acerque el valor de instrumental a 1,0, mayor será la probabilidad de que la pista no contenga voces.
- *liveness (Numeric)*: Detecta la presencia de público en la grabación. Los valores más altos representan una mayor probabilidad de que la pista haya sido interpretada en directo.
 - Valor superior a 0,8 indica una gran probabilidad de que la pista se haya grabado en directo.
- *valence (Numeric)*: Medida de 0,0 a 1,0 que describe la positividad musical que transmite una pista. Las pistas con una valencia alta suenan más positivas mientras que las pistas con una valencia baja suenan más.
- *tempo (Numeric)*: Tempo general estimado de una pista en pulsaciones por minuto (BPM). Velocidad o el ritmo de una pieza determinada y se deriva directamente de la duración media de los tiempos.
- *time_signature (Integer)*: Un compás estimado que oscila entre 3 y 7, lo que indica compases de 3/4 a 7/4.

- *track_genre* (String): El género al que pertenece la pista.

track_id	artists	album_name	track_name	popularity	duration_ms	explicit	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	time_signature	track_genre
5S0i0kRyPMVoIQDjUgSV	Gen Hoshino	Comedy	Comedy	73	230666	FALSE	676	461	1	-6746	0	143	0.0322	1.01e-06	358	715	87917	4	acoustic
4cPND8W13p13qLC10K3A	Ben Woodward	Ghost (Acoustic)	Ghost - Acoustic	55	149610	FALSE	0.42	166	1	-17235	1	0.0763	924	5.55e-06	101	267	77489	4	acoustic
1uBS7n7YYzMBEGzK9b	Ingrid MichaelsonZAYN	To Begin Again	To Begin Again	57	210826	FALSE	438	359	0	-9734	1	0.0657	0.21	0.0	117	0.12	76332	4	acoustic
6tlq3CG4kTtEgTopyCyx	Kina Gramis	Crazy Rich Asians (Original Motion Picture Soundtrack)	Can't Help Falling In Love	71	201933	FALSE	266	0.0696	0	-18515	1	0.0363	905	7.07e-05	132	143	181.74	3	acoustic
5vLSffmIP28QGSW6N2K	Chord Overstreet	Hold On	Hold On	62	198853	FALSE	618	443	2	-9681	1	0.0626	469	0.0	0.0029	167	119949	4	acoustic
01MV09KVTNFBU9Jdc	Tyrene Wells	Days I Will Remember	Days I Will Remember	58	214240	FALSE	688	481	6	-8807	1	105	289	0.0	189	666	98017	4	acoustic
9vdcwAMxK9AM7WJcE37N	A Great Big WorldChristina Aguilera	Is There Anybody Out There?	Say Something	74	229400	FALSE	407	147	2	-8822	1	0.0355	857	2.88e-06	0.0913	0.0765	141284	3	acoustic
1EzrEDxmMHG43AXT1y1pA	Jason Mraz	We Sing, We Dance, We Scream Things	I'm Yours	80	242946	FALSE	703	444	11	-9331	1	0.0417	559	0.0	0.0973	712	150.96	4	acoustic
0k4bUovGnvd3AWczQ8	Jason MrazColbie Calliat	We Sing, We Dance, We Scream Things	Lucky	74	189613	FALSE	625	414	0	-8.7	1	0.0369	294	0.0	151	669	130088	4	acoustic
7h6Qu7YpA2q9yEdwEw2	Rosa Copperman	Hunger	Hunger	56	205594	FALSE	442	632	1	-6.77	1	0.0295	426	0.00419	0.0735	196	78899	4	acoustic
4mzP5mHRvGxhdGdAH7EJ	Zack Tabudlo	Episode	Give Me Your Forever	74	244800	FALSE	627	363	8	-8127	1	0.0291	279	0.0	0.0628	301	99005	4	acoustic

Figura 1. Ejemplo de los valores de las primeras filas del conjunto de datos *Spotify Tracks*.

Justificación de la selección del conjunto de datos

La selección del conjunto de datos basado en las canciones de la popular plataforma de streaming musical Spotify, presenta una gran relevancia y riqueza para el análisis de la música contemporánea a nivel global, ya que, contiene características de audio asociadas a cada canción.

La música es una de mis grandes pasiones y gracias a la plataforma Spotify, está presente en gran parte de mi día a día. Es por eso por lo que la visualización de estos datos ofrece una oportunidad invaluable para explorar tendencias en el consumo musical, las preferencias de los usuarios y potencialmente contribuir al desarrollo de sistemas de recomendación musical más efectivos.

Relevancia del conjunto de datos en su contexto

Los datos del conjunto no son del todo actuales, ya que fueron recogidos hace 1 año mediante la API Web y código Python, y la plataforma musical de Spotify actualiza continuamente la biblioteca musical y las métricas asociadas. Aun así, los datos radican en la actualidad y la exhaustividad de las canciones disponibles por lo que, resulta relevante para colectivos amantes de la música, artistas, productores e incluso personas que desarrollan aplicaciones relacionadas con la música, pero sobre todo, el conjunto de datos es relevante por la investigación en sistemas de recomendación de música, ya que se enfatiza la necesidad de introducir información compleja, así como características de audio y géneros integrados en el conjunto.

Es importante la disponibilidad de estos datos para comprender las tendencias y preferencias de los usuarios según la popularidad.

En cuanto a la perspectiva de género, no se presenta una información explícita, sino que tal y como se ha comentado anteriormente, se presentan características musicales asociadas a cada canción. De todas maneras, al considerar cada uno de los artistas y las preferencias musicales podría haber la oportunidad de explotar los datos teniendo en cuenta como ciertos géneros o artistas son recibidos en función del género de la persona u oyente, así como proporcionar información valiosa para diversos grupos demográficos en el contexto musical y tecnológico actual.

Esta exploración podría revelar indirectamente preferencias y tendencias culturales. Además, es pertinente en el contexto de la industria musical, ya que ayuda a comprender la evolución de los gustos y las demandas.

La complejidad (medida, variables disponibles, tipos de datos, etc.)

El conjunto de datos de las pistas de Spotify cumple con los requisitos de complejidad establecidos, ya que, contiene un total de 20 columnas/variables y 114000 filas en la cual se presentan alrededor de 125 géneros musicales diferentes entre otras variables tanto categóricas como cuantitativas, hecho que demuestra una cantidad significativa de datos y lo hace adecuada para análisis detallados y complejos.

Las variables disponibles se presentan al inicio de la práctica junto con un ejemplo de las filas iniciales del conjunto donde se observan los tipos de datos que contiene, combinando datos cuantitativos como la métrica de popularidad (cuantitativo), y datos categóricos como los nombres de artistas y álbumes (categóricos), etc. Se puede observar cómo es un conjunto de datos diverso y complejo, ya que, la combinación de estos datos permitirá identificar patrones de géneros, explorar relaciones entre múltiples variables, así como correlacionar características y realizar análisis de tendencias. Por lo que, se puede demostrar que es un conjunto de datos diverso y complejo.

La originalidad

Es posible que existan visualizaciones basadas en conjuntos de datos de Spotify, dada la popularidad y accesibilidad de los datos de la plataforma. Sin embargo, la originalidad no necesariamente se limita a la existencia o ausencia de visualizaciones previas, sino a cómo se aborda y presenta la información. Este conjunto de datos específico puede ofrecer perspectivas originales debido a su carácter exhaustivo y la amalgama de características que contiene, pudiendo permitir hallazgos e interpretaciones novedosas.

La incorporación de datos externos en este conjunto no se ha valorado, ya que el hecho de contener tantos valores para cada variable complicaría mucho la fusión con otras fuentes, y al ya tener características diversas y completas, se ha considerado una mejor solución el trabajar con los datos obtenidos des de la fuente de origen.

Las cuestiones que se responderán con la visualización de datos

Las preguntas planteadas consideran los puntos anteriores al abordar diversas dimensiones del conjunto de datos de Spotify, aprovechando su riqueza en características de audio, popularidad y géneros musicales. Se pueden llegar a explorar correlaciones entre variables, variaciones temporales y diferencias entre géneros, lo cual está en línea con la sugerencia de aprovechar la diversidad de variables disponibles y presentar una narrativa visual diferente.

Es posible que algunas de estas preguntas hayan sido abordadas en otras visualizaciones o proyectos debido a la popularidad y accesibilidad de los datos de Spotify. Sin embargo, tal y como se ha comentado anteriormente la originalidad no reside solo en las preguntas en sí, sino en la manera única en que se abordan, se presentan y se contextualizan a través de la visualización. Cada enfoque particular podría generar interpretaciones únicas y revelar matices específicos dentro del conjunto de datos, lo cual es clave para la originalidad y relevancia de la visualización propuesta.

En cuanto a su idoneidad para el conjunto de datos elegido, estas preguntas son adecuadas y relevantes. Aprovechan las diversas dimensiones del conjunto de datos de Spotify, incluyendo las características de audio, la popularidad y los géneros musicales, lo que permite un análisis amplio y detallado de múltiples aspectos de la música. Esto respalda la relevancia y adecuación de las preguntas planteadas para el conjunto de datos seleccionado.

Algunas preguntas planteadas para la visualización de datos son:

- ¿Existe una correlación entre la energía de una pista y su popularidad?
- ¿Cuáles son los géneros más comunes asociados con artistas populares?
- ¿Hay diferencias significativas en la duración de las pistas entre diferentes géneros?
- ¿Cómo se comparan las características de audio entre géneros contrastantes (por ejemplo, música clásica versus hip-hop)?