



Explicación Dataset

El dataset cuenta con los campos principales de las canciones, géneros y valores medibles para orientar al usuario a escuchar mas música que le agrade

Se busca poder lograr crear uno o muchos modelos que permitan entregar recomendaciones al usuario en base a sus preferencias, para esto se tomaran en cuenta columnas relacionadas con la energía, tonalidad, duración, sonoridad, etc.

Para explorar los datos de manera mas efectivo, responderemos a estas preguntas en el segundo punto llamado 'Data Understanding'.

1) ¿Cuántas canciones hay por genero de música?

2) ¿Cuál es la duración promedio de cada canción y que genero es el mas común?

3) ¿Cuál es el artista con mas canciones en la plataforma?

4) ¿Cuál es la canción mas popular según el genero musical?

- **track_id:** El ID de Spotify para la pista.
- **artists:** Los nombres de los artistas que interpretaron la pista. Si hay más de un artista, están separados por un ;.
- **album_name:** El nombre del álbum en el que aparece la pista.
- **track_name:** Nombre de la pista.
- **popularity:** La popularidad de una pista es un valor entre 0 y 100, donde 100 es lo más popular. La popularidad se calcula mediante un algoritmo y se basa, en su mayor parte, en el número total de reproducciones que ha tenido la pista y en cuán recientes son esas reproducciones. En términos generales, las canciones que se están reproduciendo mucho ahora tendrán una popularidad más alta que las canciones que se reprodujeron mucho en el pasado. Las pistas duplicadas (por ejemplo, la misma pista de un sencillo y de un álbum) se califican de manera independiente. La popularidad del artista y del álbum se deriva matemáticamente de la popularidad de las pistas.
- **duration_ms:** La duración de la pista en milisegundos.
- **explicit:** Indica si la pista tiene letras explícitas (true = sí las tiene; false = no las tiene o desconocido).
- **danceability:** La capacidad de baile describe cuán adecuada es una pista para bailar, basándose en una combinación de elementos musicales como el tempo, la estabilidad del ritmo, la fuerza del pulso y la regularidad general. Un valor de 0.0 es menos bailable y 1.0 es más bailable.
- **energy:** La energía es una medida de 0.0 a 1.0 que representa una medida perceptual de intensidad y actividad. Típicamente, las pistas enérgicas se sienten rápidas, ruidosas y fuertes. Por ejemplo, el death metal tiene alta energía, mientras que un preludio de Bach puntúa bajo en la escala.
- **key:** La tonalidad de la pista. Los números enteros se asignan a tonos utilizando la notación estándar Pitch Class. Por ejemplo, 0 = C, 1 = C#/D♭, 2 = D, y así sucesivamente. Si no se detecta ninguna clave, el valor es -1
- **loudness:** La sonoridad general de una pista en decibelios (dB).
- **mode:** El modo indica la 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:** La locuacidad detecta la presencia de palabras habladas en una pista. Cuanto más exclusivamente hablada sea la grabación (por ejemplo, programa de entrevistas, audiolibro, poesía), 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. Los valores entre 0,33 y 0,66 describen pistas que pueden contener tanto música como voz, ya sea en secciones o en capas, incluyendo casos como la música rap. Los valores inferiores a 0,33 representan probablemente música y otras pistas no habladas.

- **acousticness:** Una medida de confianza de 0,0 a 1,0 sobre si la pista es acústica. 1,0 representa una confianza alta en que la pista es acústica.
- **instrumentalness:** Predice si una pista no contiene voces. Los sonidos «ooh» y «aah» se consideran instrumentales en este contexto. Las pistas de rap o spoken word son claramente «vocales». Cuanto más se acerque el valor de instrumental a 1,0, mayor será la probabilidad de que la pista no contenga voces.
- **liveness:** Detecta la presencia de público en la grabación. Los valores de liveness más altos representan una mayor probabilidad de que la pista se haya interpretado en directo. Un valor superior a 0,8 indica una gran probabilidad de que la pista se haya grabado en directo.
- **valence:** 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 (por ejemplo, felices, alegres, eufóricas), mientras que las pistas con una valencia baja suenan más negativas (por ejemplo, tristes, deprimidas, enfadadas).
- **tempo:** El tempo global estimado de una pista en pulsaciones por minuto (BPM).
En terminología musical, el tempo es la velocidad o el ritmo de una pieza determinada y se deriva directamente de la duración media del compás
- **time_signature:** Un compás estimado. El compás es una convención de notación para especificar cuántos tiempos hay en cada compás. El compás oscila entre 3 y 7, indicando compases de 3/4 a 7/4.
- **track_genre:** Género al que pertenece la pista