

# ENTREGA FINAL DEL PROYECTO ~Data Analytics~

/\*DOCUMENTACIÓN

--Giordano.Santiago

Fecha.Presentación[25 de abril, 2022]

\*/

# ÍNDICE:

1. Introducción.....	Pág. 3
2. Descripción de la temática de los datos.....	Pág. 3
3. Hipótesis.....	Pág. 5
4. Objetivo.....	Pág. 5
5. Herramientas tecnológicas implementadas.....	Pág. 5
6. Sobre la base de datos:	
6.1 Dataset.....	Pág. 6
6.2 Diagrama Entidad-Relación.....	Pág. 6
6.3 Listado de tablas.....	Pág. 6
6.4 Listado de columnas por tabla.....	Pág. 7
7. Sobre la visualización:	
7.1 Alcance.....	Pág. 8
7.2 Transformación de datos.....	Pág. 8
7.3 Medidas calculadas.....	Pág. 11
7.4 Detalle de las fórmulas.....	Pág. 12
7.5 Segmentaciones elegidas.....	Pág. 13
7.6 Análisis funcional del tablero.....	Pág. 14
8. Conclusión.....	Pág. 20
9. Futuras líneas.....	Pág. 20

Versión	Fecha
Versión 1.0	31/3/2022
Versión 2.0	31/3/2022
Versión 3.0	3/4/2022
Versión 3.1	6/4/2022
Versión 4.0	13/4/2022
Versión 5.0	16/4/2022
Versión 6.0	20/4/2022
Versión 7.0	25/4/2022

[Tabla de versiones]

## 1. Introducción:

En el presente trabajo se detallará la documentación del proyecto final del curso “Data analytics”, el cual es parte de las *Carreras de Data analyst y Data Scientist* ofrecidas por Coderhouse. Primero se describirá la temática, la cual fue elegida luego de un proceso de propuesta y descarte de diversas opciones. Luego se irá documentando parte por parte todo el proceso, desde la lectura y limpieza del Dataset elegido en Excel, hasta la confección final de un tablero de control en Power BI.

## 2. Descripción de la temática de los datos:

La temática a analizar en este caso será el catálogo del servicio de streaming de música Spotify. Nuestro Dataset cuenta con el título de la canción, el género, el artista, la duración, la popularidad, la fecha de lanzamiento, el tempo, y una serie de categorías que refieren a atributos como “bailabilidad” (danceability), “instrumentalidad” (instrumentalness), entre otros (speechiness, acousticness, liveness, loudness, energy).

Spotify Premium cuenta con la posibilidad de ajustar una búsqueda de canciones utilizando estos atributos denominados “audio features”. Éstos son explicados en el [Blog para desarrolladores](#) de la plataforma de la siguiente forma:

*-Acousticness:* Una medida del 0.0 al 1.0 sobre si la canción es acústica. 1.0 representa que con alta confianza la canción es acústica.<sup>1</sup>

*-Danceability:* Describe qué tan adecuada es una canción para bailar basado en una combinación de elementos musicales incluido el tempo, la estabilidad del ritmo, la fuerza del “beat”, y una regularidad general. Un valor de 0.0 es menos bailable y 1.0 es muy bailable.<sup>2</sup>

*-Energy:* Es una medida de 0.0 a 1.0 y representa una percepción de la intensidad y la actividad. Típicamente, las canciones energéticas se sienten rápidas, fuertes y ruidosas.<sup>3</sup>

*-Instrumentalness:* Predice si una canción no contiene palabras cantadas. “Ooh” y “aah” son tratadas como instrumentos en este contexto. El rap y la palabra hablada son consideradas

---

<sup>1</sup> A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic. -Las traducciones son propias-

<sup>2</sup> Describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.

<sup>3</sup> Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy.

como palabras cantadas. Cuanto más cerca esté el valor a 1.0, más probabilidad hay de que la canción no contenga palabras cantadas. Los valores arriba de 0.5 pretenden representar canciones instrumentales, pero la confianza es mayor a medida de que el valor se acerca a 1.0.<sup>4</sup>

-*Liveness*: Detecta la presencia de una audiencia en la grabación. Valores más altos representan una probabilidad incrementada de que la canción fue tocada en vivo. Un valor arriba de 0.8 provee alta probabilidad de que la canción sea “en vivo”.<sup>5</sup>

-*Speechiness*: Detecta la presencia de palabras habladas en una canción. Cuanto más exclusivamente “hablada” es la grabación (libro audible, poesía, etc.), más cercano a 1.0 será el valor del atributo. Valores arriba de 0.66 describen grabaciones que probablemente estén totalmente hechos de palabras habladas. Valores entre 0.33 y 0.66 describen grabaciones que pueden contener tanto música como palabra hablada, sea en secciones o en capas, incluido el caso del rap. Valores debajo de 0.33 más probablemente representan música u otras grabaciones no habladas.<sup>6</sup>

-*Tempo*: El tempo general estimado de una canción en “beats” por minuto (BPM). En términos musicales, el tempo es la velocidad o el pulso de una pieza y deriva directamente del promedio de duración del “beat”.<sup>7</sup>

-*Valence*: Una medida entre 0.0 y 1.0 que describe la “positividad musical” que transmite la canción. Canciones con valores altos suenan más “positivas” (felices, eufóricas, etc.), mientras que canciones con valores bajos suenan más “negativas” (tristes, deprimentes, enojadas, etc.).<sup>8</sup>

-*Loudness*: Lo “fuerte” que suena en promedio una canción en decibels (dB). Los valores son promediados sobre toda la canción y son útiles para comparar entre canciones. Refiere a la

---

<sup>4</sup> Predicts whether a track contains no vocals. “Ooh” and “aah” sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly “vocal.” The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.

<sup>5</sup> Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.

<sup>6</sup> Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.

<sup>7</sup> The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.

<sup>8</sup> A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).

cualidad de un sonido como correlato psicológico primario de “fuerza física” (amplitud). Los valores típicamente están en el rango entre -60 y 0 dB.<sup>9</sup>

-*Key*: La clave musical en la que está la canción. Cada clave está representada por un entero: 0 = C, 1 = C#/Db, 2 = D, etc. Si no se detectó una clave, el valor es -1.<sup>10</sup>

-*Time Signature*: Una métrica de tiempo estimada. Es una convención de notación para especificar cuántos “beats” hay en cada compás. Tiene un rango entre 3 y 7, indicando métricas de “3/4” a “7/4”.<sup>11</sup>

-*Mode*: El modo indica la modalidad (mayor o menor) de una canción, el tipo de escala de la cual deriva su contenido melódico. Mayor es representado por un 1 y menor por un 0.<sup>12</sup>

Queremos analizar cuáles son las relaciones entre género, popularidad, y los diversos atributos detallados en el Dataset.

### 3. Hipótesis:

Como punto de partida consideramos que habrá alguna correlación significativa entre la “popularidad” y los diversos atributos de cada canción. En este sentido consideramos en un principio que valores como “danceability” (qué tanailable es la canción), “valence” (qué tan “positiva” / “alegre” suena la canción) e “instrumentalness” (en qué medida la canción es instrumental o contiene palabras habladas o cantadas), serán algunos de los atributos más destacados en su correlación con la popularidad de la canción.

### 4. Objetivo:

El objetivo de este trabajo es conocer los diversos pasos necesarios para el análisis de un Dataset, en este caso más particularmente a partir del análisis del Dataset de Spotify que hemos elegido. En este proceso veremos cuáles son las herramientas a utilizar y cómo aplicarlas a los datos, para eventualmente generar un tablero de control. El objetivo final será poder lograr derivar alguna conclusión en base al análisis llevado a cabo.

<sup>9</sup> The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typically range between -60 and 0 db.

<sup>10</sup> The key the track is in. Integers map to pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C#/Db, 2 = D, and so on. If no key was detected, the value is -1.

<sup>11</sup> An estimated time signature. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure). The time signature ranges from 3 to 7 indicating time signatures of “3/4”, to “7/4”.

<sup>12</sup> Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0.

## 5. Herramientas tecnológicas implementadas:

Para el presente trabajo se utilizaron las siguientes herramientas:

- Microsoft Excel: lectura y limpieza del Dataset.
- Presentaciones de Google: creación del Mockup, del Diagrama Entidad-Relación y de los fondos para el informe de Power BI.
- Microsoft Power BI Desktop: creación del tablero de control.

## 6. Sobre la base de datos:

### 6.1 Dataset:

El archivo del Dataset con las bases de datos organizadas y estructuradas utilizadas en el presente trabajo se adjunta a la presente documentación en el siguiente link: [DB Spotify](#).

### 6.2 Diagrama entidad-relación:

A continuación, se detalla el diagrama entidad-relación creado:



### 6.3 Listado de tablas:

En este apartado haremos mención a cada una de las tres tablas con las que trabajamos, junto a su descripción y a la definición de las diversas claves (primarias y foráneas).

-Tabla 1 -"Song" ("Canción"): contiene los títulos de las canciones con sus respectivos "id", además de otros datos como el id del artista/autor, la duración, si contiene o no contenido

explícito, su fecha de publicación, su popularidad, su id de género y la cantidad de tracks que componen el disco al que pertenece.

- PK: Id\_song
- FK: Id\_artist
- FK: Id\_genre

-Tabla 2 - “Genre” (“Género”): contiene los 7 géneros musicales en los que Spotify subdivide cada canción (R&B, latin, electronic dance music, pop, hip hop, rock, rap), con sus respectivos id.

- PK: Id\_genre

-Tabla 3 - “Artist” (“Artista/Autor”): contiene los nombres de los artistas/autores y sus respectivos id.

- PK: Id\_artist

#### 6.4 Listado de columnas por tabla:

A continuación, se hace mención a cada tabla, a las columnas que posee, junto con su tipo de campo y su clave.

**Tabla 1: Song**

Tipo de clave	Campo	Tipo de campo
PK - index	Id_song	varchar(10)
-	Name_song	text
-	Name_artist	text
-	Duration_mm	decimal (3,2)
-	Explicit_content	tinyint
-	Release_Date	date
-	Release_date_precision	varchar(10)
-	Total_tracks	tinyint
-	Popularity	tinyint
FK	Id_genre	varchar(10)
FK	Id_artist	varchar(10)

**Tabla 2: Genre**

Tipo de clave	Campo	Tipo de campo
PK - index	Id_genre	varchar(10)
-	Name_genre	varchar(100)

**Tabla 3: Artist**

Tipo de clave	Campo	Tipo de campo
PK - index	Id_artist	varchar(10)
-	Name_artist	varchar(100)

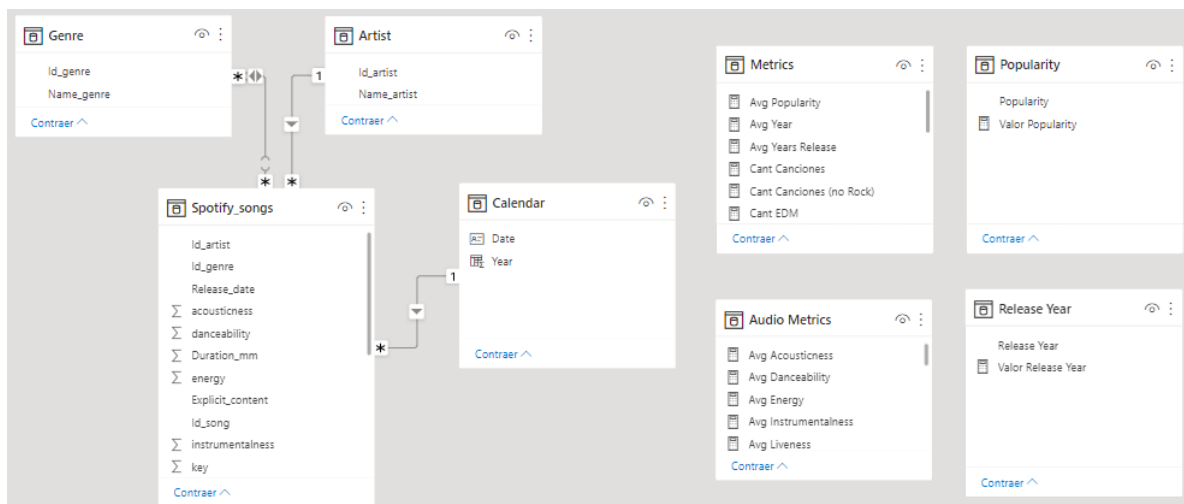
## 7. Sobre la visualización:

### 7.1 Alcance:

El *dashboard* se ha diseñado pensando en que cualquiera (usuario final) que lo observe pueda hacer un análisis introductorio de la temática (nivel de aplicación), buscando mostrar relaciones y correlaciones entre diversos datos, teniendo como variable principal los valores de “popularidad” asignados en la base de datos.

### 7.2 Transformación de datos:

A continuación, se muestra el modelo relacional actualizado y se documentan todos los cambios realizados sobre la base de datos y el modelo relacional desde Power BI y Power Query:



1. Base de datos: la conformación de la base de datos fue alterada, pasando de contener las tablas “Song”/”Genre”/”Audio features” a contener las tablas “Song”/”Genre”/”Artist”. Esto fue una corrección, dado que la tabla “Song” y la tabla “Audio features” contenían la misma clave primaria. Por esta razón se reincorporó (Union) la tabla “Audio features” dentro de la tabla “Song”. A continuación, se creó una nueva tabla “Artist”, la cual contiene los nombres de los artistas y una nueva clave primaria “Id\_artist”. Luego en Power Query se reemplazaron en la tabla “Song” los nombres de los artistas con sus respectivos “Id\_artist” (aquí como clave foránea), y luego se eliminó de esta tabla la columna “Artist\_name”.



2. Base de datos: en la tabla "Song" se reemplazaron los valores "True"/"False" (BOOLEAN) de la columna "Explicit\_content" por los valores "Contenido explícito"/"Sin contenido explícito" (TEXT), ya que en los informes de Power BI resultará más intuitivo leer los valores de esta forma.
3. Base de datos: se reorganizó la tabla "Song" para que las tres claves (1 PK y 2 FK) estén ubicadas en las primeras tres columnas.
4. Base de datos: los valores de las claves primarias de las tablas "Song" y "Artist" fueron modificados, pasando de ser valores enteros (INT) a ser caracteres variables (VARCHAR). Se realizó esta modificación ya que al crear los nuevos "Id\_artist" resultó una "buena práctica" que sus valores se diferencien de los "Id\_song". En consiguiente, los "Id\_song" pasaron de ser "1"- "5785" a ser "S0001"- "S5785" ("S" de "Song"), y los "Id\_artist" pasaron de ser "1"- "2867" a ser "A0001"- "A2867" ("A" de "Artist"). A su vez, otro resultado positivo de esta modificación es que, al ordenar de forma ascendente o descendente, no se genera ningún error en el orden resultante.
5. Tabla 1: se cambió el tipo de dato de la columna "Release\_date" (TEXT->DATE).
6. Tabla 1: el paso anterior volvió innecesaria la columna "Release\_date\_precision" y por tal razón ésta fue eliminada.
7. Tabla 1: la columna "Total\_tracks" hasta este punto ha resultado innecesaria, por tal razón fue eliminada.
8. Tabla 1: se cambió el tipo de datos de la columna "Duration\_mm", de Decimal a Decimal Fijo (luego este cambio también fue realizado en la base de datos).
9. Tabla 2: se eliminó la última fila, la cual estaba vacía (null).
10. Parámetro: se generó un parámetro con la ruta del archivo.
11. Modelo relacional: fue necesario crear la conexión entre la tabla "Song" y la tabla "Genre" dado que Power BI no la generó automáticamente.
12. Nuevas tablas: se crearon dos tablas, "Top 20" y "Bottom 20", donde se conservaron las primeras veinte filas con las 20 canciones con valores de popularidad más alto y más bajo, respectivamente.
13. Tabla de referencia: se creó una tabla de referencia para la tabla "Genre", para evitar errores en el modelo relacional.
14. Tabla 1: se cambió la columna "Release\_date" por la columna "Release\_year", donde figura sólo el año de lanzamiento, sin día y mes, ya que tanta precisión en el dato no resulta necesaria (este paso debería haberse llevado a cabo junto al paso 2).
15. Tablas "Top 20" y "Bottom 20": se llevó a cabo la misma transformación que en el paso anterior.
16. Tablas 1 a 3: se renombraron las tablas 1 a 3, ahora "Spotify\_songs", "Genre" y "Artist", prescindiendo de la numeración.
17. Tabla Calendario: se creó la tabla "Calendar", agregando además las conexiones respectivas entre esta tabla y las tablas "Spotify\_songs", "Top 20" y "Bottom 20" en el modelo relacional.
18. Nuevas medidas calculadas: se crearon cuatro nuevas medidas calculadas, "Avg\_popularity" y "Avg\_loudness" (AVERAGE) en la Tabla "Spotify\_songs", "Avg\_top\_loudness" en la Tabla "Top 20", y "Avg\_bottom\_loudness" en la Tabla "Spotify\_songs". Se eligió trabajar sobre las medidas de "loudness" porque en análisis

preliminares fueron uno de los valores más interesantes a tener en cuenta. Luego se evaluará cuáles otras medidas calculadas serán necesarias.

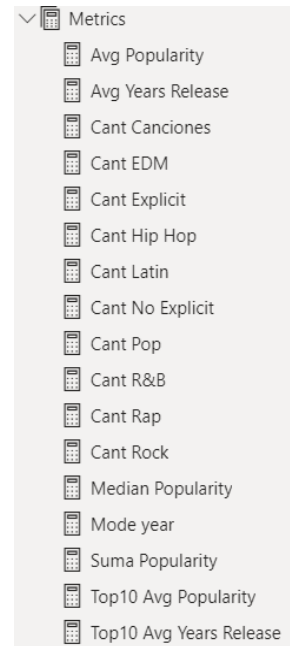
19. Nuevas columnas calculadas: se crearon dos columnas calculadas, "Years\_since\_release" (DATEDIFF), una en la Tabla "Top 20" y otra en la Tabla "Bottom 20". Además, en la Tabla "Spotify\_songs" se creó la columna calculada "Top\_or\_bottom", donde se señala (IF) cuáles entradas son más populares que el promedio y cuáles son menos populares que el promedio.
20. Nuevas medidas calculadas: a partir de las dos nuevas columnas calculadas se crearon dos nuevas medidas calculadas, "Avg\_top\_since\_release" en la Tabla "Top 20", y "Avg\_bottom\_since\_release" en la Tabla "Bottom 20".
21. Eliminación de tablas: se eliminaron del modelo las Tablas "Top 20", "Bottom 20" y la Tabla de referencia "Genre (Top/Bottom)". Para considerar que una canción está en el top 20, se requiere de un contexto. Es información, no dato. En este sentido no resulta correcto tener esta información como Tabla.
22. Parámetros: se crearon dos parámetros, uno para "Popularity" y otro para "Release Year".
23. Tablas de Métricas: se crearon dos tablas de métricas, una con métricas específicas sobre "audio features", y otra con el resto.
24. Métricas: en ambas tablas de métricas se crearon una serie de métricas básicas usando funciones de suma, cuenta, promedio, mediana y modo.
25. Métricas con Variables: se crearon dos métricas de una variable, "Top10 Avg Popularity" y "Top10 Avg Years Release".
26. Tabla Spotify\_songs: se agregaron dos columnas a la tabla, una calcula los años que pasaron desde que cada canción fue publicada y el día de hoy (Years\_release), la otra vuelve positivos los valores de la columna Loudness. Notamos que la correlación negativa que encontrábamos al comparar estos valores con otros radicaba en que éstos estaban expresados como números negativos, por esta razón creamos esta columna donde cada valor se multiplica por (-1) para convertirlos en números positivos.
27. Tabla Spotify\_songs: se eliminó la columna original "loudness", y seguidamente se renombró la nueva columna "loudness (positive)" como "loudness", reemplazando de esta forma a la primera.
28. Medidas calculadas: se eliminaron los promedios de cada "audio feature" dado que resultaron innecesarios.

### 7.3 Medidas calculadas:

A continuación, se detallan las medidas calculadas creadas:

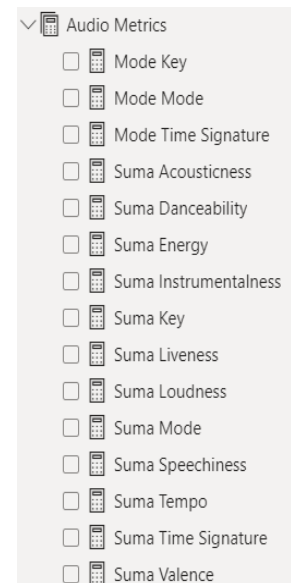
- *Tabla "Metrics":*

1. Avg Popularity: calcula el promedio de los valores de la columna Popularity de la tabla Spotify\_songs.
2. Avg Years Release: calcula el promedio de los valores de la columna Years\_release de la tabla Spotify\_songs.
3. Cant Canciones: cuenta la cantidad de canciones en la tabla Spotify\_songs.
4. Cant <genre>: calcula la cantidad de canciones en la tabla Spotify\_songs que son de cada género.
5. Cant Explicit/No Explicit: calcula la cantidad de canciones en la tabla Spotify\_songs que tiene contenido explícito o no explícito, respectivamente.
6. Median Popularity: calcula la mediana de popularidad desde el parámetro Popularity.
7. Mode Year: calcula el modo dentro de los valores de la columna Release\_date de la tabla Spotify\_songs.
8. Suma Popularity: suma los valores de la columna Popularity de la tabla Spotify\_songs.
9. Top10 Avg Popularity: utiliza una variable para calcular cuál es el promedio de popularidad del top 10 de canciones (las 10 canciones con mayor popularidad).
10. Top10 Avg Years Release: utiliza una variable para calcular cuál es el promedio de años que pasaron desde que la canción fue publicada y el día de hoy, para aquellas 10 canciones del top 10.



- *Tabla "Audio Metrics":*

1. Mode <Key/Mode/Time Signature>: calcula el modo de la columna < Key/Mode/Time Signature> de la tabla Spotify\_songs para aquellos valores que tiene sentido calcular como modo.
2. Suma <Audio feature>: suma los valores de la columna <Audio feature> de la tabla Spotify\_songs.



## 7.4 Detalle de las fórmulas:

### 1. Tabla "Metrics":

2. Avg Popularity = AVERAGE(Spotify\_songs[Popularity])
3. Avg Years Release = AVERAGE(Spotify\_songs[Years\_release])
4. Cant Canciones = COUNT(Spotify\_songs[Id\_song])
5. Cant <genre> = CALCULATE(  
     Metrics[Cant Canciones],  
     Spotify\_songs[Id\_genre] = "<genre>"  
 )
6. Cant Explicit = CALCULATE(  
     Metrics[Cant Canciones],  
     Spotify\_songs[Explicit\_content] = "Contenido explícito"  
 )  
 /  
 Cant No Explicit = CALCULATE(  
     Metrics[Cant Canciones],  
     Spotify\_songs[Explicit\_content] = "Sin contenido explícito"  
 )
7. Median Popularity = MEDIAN(Popularity[Popularity])
8. Mode year = MINX (  
     TOPN (  
         1,  
         ADDCOLUMNS (  
             VALUES ( Spotify\_songs[Release\_date] ),  
             "Frequency", CALCULATE( COUNT ( Spotify\_songs[Release\_date] ) )  
         ),  
         [Frequency],  
         0  
     ),  
     Year(Spotify\_songs[Release\_date])  
 )
9. Suma Popularity = SUM(Spotify\_songs[Popularity])
10. Top10 Avg Popularity =  
     VAR A = CALCULATETABLE(Spotify\_songs,  
         TOPN(10, Spotify\_songs, Spotify\_songs[Popularity], DESC))  
  
     RETURN  
     CALCULATE([Avg Popularity] , A)
11. Top10 Avg Years Release =  
     VAR A = CALCULATETABLE(Spotify\_songs,  
         TOPN(10, Spotify\_songs, Spotify\_songs[Popularity], DESC))  
  
     RETURN  
     CALCULATE(Metrics[Avg Years Release], A)

## 2. Tabla "Audio Metrics":

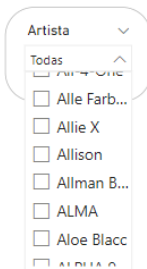
1. Mode <Key/Mode/Time Signature> = MINX (
 

```

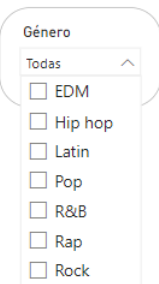
      TOPN (
        1,
        ADDCOLUMNS (
          VALUES ( Spotify_songs[<Key/Mode/Time Signature>] ),
          "Frequency", CALCULATE ( COUNT ( Spotify_songs[<Key/Mode/Time Signature>] ) )
        ),
        [Frequency],
        0
      ),
      Spotify_songs[<Key/Mode/Time Signature>]
    )
  
```
2. Suma <Audio Feature> = SUM(Spotify\_songs[<Audio Feature>])

## 7.5 Segmentaciones elegidas:

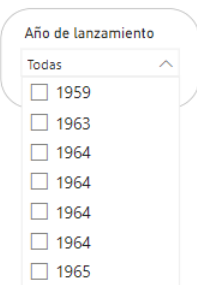
A continuación, se detallan las segmentaciones elegidas:



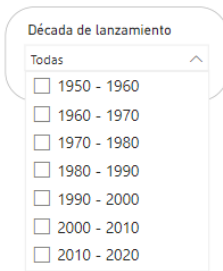
1. Artista: filtro desplegable con la lista de todos los artistas.



2. Género: filtro desplegable con la lista de todos los géneros.



3. Año de lanzamiento: filtro desplegable con la lista de todos los años de lanzamiento.



4. Década de lanzamiento: filtro desplegable con la lista de todas las décadas de lanzamiento.

5. A su vez, en las solapas "03" a "05" se utilizaron botones interactivos:

Instrumentalness	Speechiness	Acousticness	Valence
Energy	Danceability	Liveness	Loudness
Tempo	Key	Time Signature	Mode

### 7.6 *Análisis funcional del tablero:*

A continuación, se hace una breve explicación de la información incluida en cada solapa y qué se puede analizar con cada gráfico:



**Solapa "Portada":** es la portada del trabajo, donde se incluye el título y la temática del Dataset.



### Solapa "Presentación Dataset":

1. Filtros: por Artista, por Género, por Año de Lanzamiento, por Década de lanzamiento.
2. KPI/Tarjeta: Año con mayor cantidad de lanzamientos (Modo), Promedio de años desde el lanzamiento (Top 10).
3. Treemap: Proporción de canciones por género.
4. Gráfico de columnas apiladas: Cantidad de años desde que fue lanzada cada canción (Top 10).
5. Gráfico de barras apiladas: Nivel de popularidad de cada género.



### Solapa "Introducción Atributos":

1. Filtros: por Artista, por Género, por Año de Lanzamiento, por Década de lanzamiento.
2. Gráfico de anillos: Proporción de canciones que tienen y no tienen contenido explícito.
3. Treemap: Proporción de canciones que tienen contenido explícito por género.
4. Gráfico de líneas: Correlación entre popularidad y un atributo (audio feature "danceability") por canción (Top 10), Correlación entre popularidad y un atributo (audio feature "danceability") por género.





**Solapa "Atributos/Popularidad #01":**

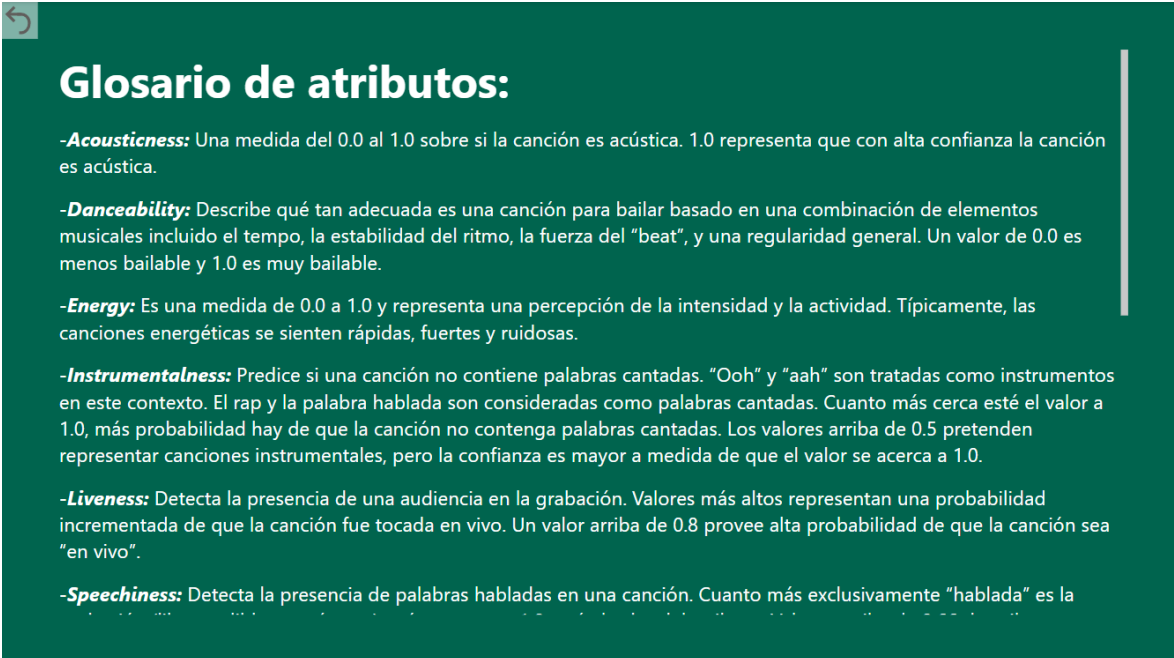
1. Filtros: por Década de lanzamiento y por "Audio Feature" (Botones interactivos)
2. Gráfico de líneas: Correlación entre popularidad y un atributo (audio features "instrumentalness", "speechiness", "acousticness", "valence") por género.

**Solapa "Atributos/Popularidad #02":**

1. Filtros: por Década de lanzamiento y por "Audio Feature" (Botones interactivos)
2. Gráfico de líneas: Correlación entre popularidad y un atributo (audio features "energy", "danceability", "liveness", "loudness") por género.

**Solapa "Atributos/Popularidad #03":**

1. Filtros: por Década de lanzamiento y por "Audio Feature" (Botones interactivos)
2. Gráfico de líneas: Correlación entre popularidad y un atributo (audio features "tempo", "key", "time signature", "mode") por género.



## Glosario de atributos:

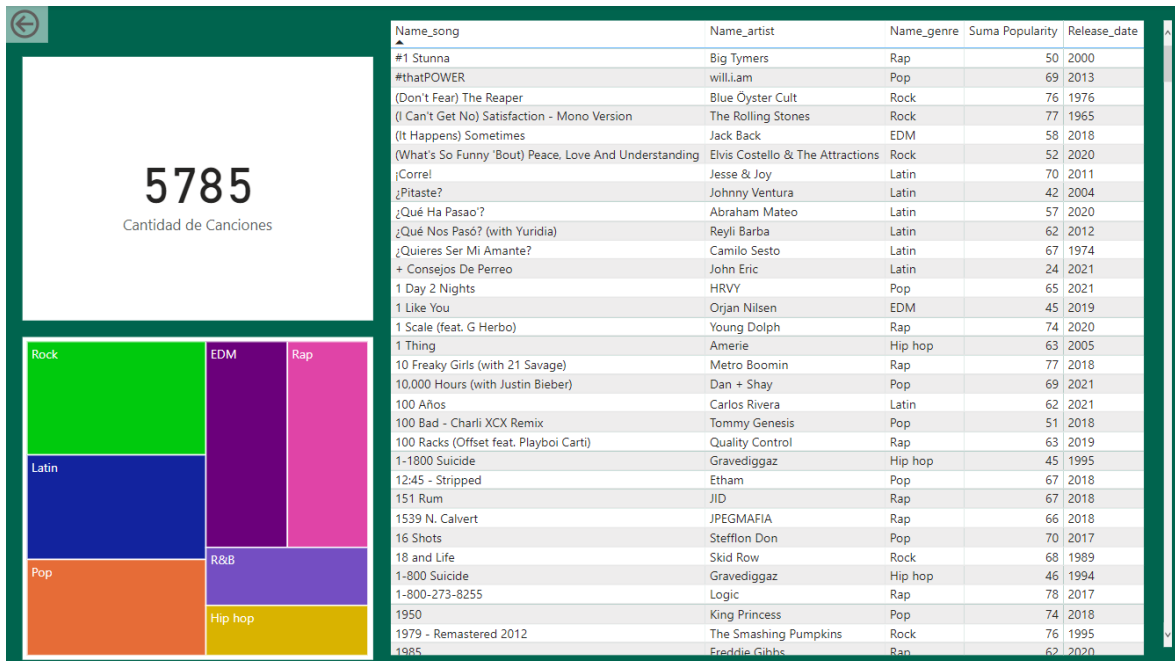
- Acousticness:** Una medida del 0.0 al 1.0 sobre si la canción es acústica. 1.0 representa que con alta confianza la canción es acústica.
- Danceability:** Describe qué tan adecuada es una canción para bailar basado en una combinación de elementos musicales incluido el tempo, la estabilidad del ritmo, la fuerza del "beat", y una regularidad general. Un valor de 0.0 es menos bailable y 1.0 es muy bailable.
- Energy:** Es una medida de 0.0 a 1.0 y representa una percepción de la intensidad y la actividad. Típicamente, las canciones energéticas se sienten rápidas, fuertes y ruidosas.
- Instrumentalness:** Predice si una canción no contiene palabras cantadas. "Ooh" y "aah" son tratadas como instrumentos en este contexto. El rap y la palabra hablada son consideradas como palabras cantadas. Cuanto más cerca esté el valor a 1.0, más probabilidad hay de que la canción no contenga palabras cantadas. Los valores arriba de 0.5 pretenden representar canciones instrumentales, pero la confianza es mayor a medida de que el valor se acerca a 1.0.
- Liveness:** Detecta la presencia de una audiencia en la grabación. Valores más altos representan una probabilidad incrementada de que la canción fue tocada en vivo. Un valor arriba de 0.8 provee alta probabilidad de que la canción sea "en vivo".
- Speechiness:** Detecta la presencia de palabras habladas en una canción. Cuanto más exclusivamente "hablada" es la

**Solapa "Glosario":** contiene el glosario de los atributos musicales contenidos en el Dataset.



### Solapa "Tooltip Canciones" (Oculta):

1. KPI/Tarjeta: Cantidad de canciones.
2. Tabla: Top 3 de Artistas.



### Solapa "Detalle Canciones" (Oculta):

1. KPI/Tarjeta: Cantidad de canciones.
2. Treemap: Géneros.
3. Tabla: tabla con nombre de la canción, nombre del artista, género, popularidad y año de lanzamiento.

## 8. Conclusión:

Como punto de partida, en nuestra hipótesis, consideramos que habría alguna correlación significativa entre la “popularidad” y los diversos atributos de cada canción. En este sentido habíamos planteado que valores como “danceability” (qué tan bailable es la canción), “valence” (qué tan “positiva” / “alegre” suena la canción) e “instrumentalness” (en qué medida la canción es instrumental o contiene palabras habladas o cantadas), serían algunos de los atributos más destacados en su correlación con la popularidad de la canción.

En nuestro análisis, hasta el punto que llegamos, hemos alcanzado dos primeras conclusiones tentativas:

1. Se puede observar con mayor claridad la correlación cuando segmentamos por géneros.
2. Hemos observado aparentes correlaciones fuertes entre diversos atributos y los valores totales de popularidad (de cada género), sin embargo, no tanto en aquellos que intuitivamente esperábamos.

Los atributos con mayor correlación serían: Tempo, Key, Time Signature y Mode. En los cuatro casos se trata de atributos ligados a conceptos de teoría musical, a diferencia del resto de atributos que son más relativos y/o ligados a conceptos construidos por Spotify.

Tentativamente uno podría pensar que a partir de los valores encontrados correlacionados con los valores altos de popularidad se podría derivar un modelo a seguir para quien quiera componer una canción potencialmente “popular”. Sin embargo, en caso de que eso sea posible, sería necesario un análisis más profundo, en tanto las correlaciones que hemos podido encontrar están ligadas a géneros y no a canciones en particular.

En resumen, este análisis representa un primer acercamiento a la muestra de este Dataset, donde pudimos comenzar a observar que habría correlaciones entre variables, y en tal sentido podría funcionar como punto de partida para un proyecto de análisis más profundo.

## 9. Futuras líneas:

En una eventual profundización del análisis llevado a cabo en este proyecto, podríamos plantear algunas potenciales líneas de investigación futuras:

1. ¿Por qué encontramos correlaciones más evidentes en la segmentación por género? ¿Es realmente confiable la correlación que se observa por género?
2. ¿Por qué la mayor correlación se observa en aquellos atributos que están más ligados a conceptos de teoría musical?
3. ¿Es posible derivar un modelo de creación de canciones populares? ¿Qué factores no considerados por Spotify tienen peso a la hora de que una canción se vuelva popular?
4. En resumen, ¿es posible predecir si una canción será o no popular?