

Título: Canciones más escuchadas en Spotify a nivel global

Autor: Martín Somoza

Temática: La base de datos (BD) contiene la información de las 50 canciones más escuchadas día a día en Spotify para 73 países del mundo durante el período abarcado entre el 18 de Octubre de 2023 hasta el 1 de Julio de 2024. Se incluye información de cómo varía la posición de cada tema en el top 50 de manera diaria y semanal. Se muestra además el artista, el álbum que contiene la canción (junto con su fecha de publicación) como así también características de la canción en sí como popularidad o presencia de contenido explícito en las letras. La BD muestra también características más técnicas incluyendo la tonalidad, el modo, la duración, el tempo, el compás y diversos índices que dan una idea de lo bailable, acústica, “hablada”, instrumental, “energética”, “positiva”, intensa (en términos de volumen) que pueda ser la canción.

Hipótesis: Las canciones más escuchadas varían según el país debido a las preferencias musicales de las poblaciones arraigadas en su base cultural e historia musical. Estas preferencias se verán reflejadas en los índices agregados de las canciones que compongan el top de cada país. La variación en el ranking de las canciones del top también será distinta según el país refiriendo a la movilidad de las canciones que más se escuchan.

Fuente: <https://www.kaggle.com/datasets/asaniczka/top-spotify-songs-in-73-countries-daily-updated>

Objetivo: Identificar las preferencias musicales de cada país en términos de los diferentes indicadores técnicos en base a las canciones que componen el top 50 en el tiempo.

Usuario final y aplicación del análisis: Actores de la industria musical pueden obtener información valiosa acerca de las preferencias musicales en cada país a la hora de orientar el producto en base a las mismas, hablando desde un punto de vista comercial. Además, en el caso de un producto que ya existe, podrá tomarse la decisión de publicitar/iniciar campañas de marketing de canciones con parámetros similares a las preferencias en cada uno de los países.

Estructura de las tablas:

Referencias:

PK: Primary Key; FK: Foreign Key

Tabla: Snapshots:

Columna	Tipo de dato	Clave
snapshot_id	Int	PK
spotify_id	Varchar(22)	FK
daily_rank	Int	-
daily_movement	Int	-
weekly_movement	Int	-
country_id	Varchar(2)	FK
snapshot_date	Date(DD/MM/AAAA)	-
popularity	Int	-

snapshot_id: Identifica el registro único tomado de un Snapshot, constituye la **PK** (no puede adquirir valores nulos).

spotify_id: Es el identificador único de la canción, **FK** de la tabla **Canciones** (donde es la PK).

daily_rank: Posición que tiene el registro en el ranking diario. Toma valores del 1 al 50.

daily_movement: Movimiento del registro en el ránking con respecto al día anterior. Toma valores de -49 a 49.

weekly_movement: Movimiento del registro en el ránking con respecto a la semana pasada (7 días antes). Toma valores de -49 a 49.

country_id: ID del país donde del registro, según el código ISO de dos letras. Es la **FK** de la tabla **Países**.

snapshot_date: Fecha en la cual se tomó el registro.

popularity: Medida de la popularidad de la canción a nivel global en ese día. Toma valores de 0 a 100.

Tabla: Canciones:

Columna	Tipo de dato	Clave
spotify_id	Varchar(22)	PK
nombre_cancion	Varchar (350)	-
artist_id	Int	FK
explicit	Boolean	-
duration_ms	Int	-
album_id	Int	FK
danceability	Float	-
energy	Float	-
key_id	Int	FK
loudness	Int	-
mode_id	Int	FK
speechiness	Float	-
acousticness	Float	-
instrumentalness	Float	-
liveness	Float	-
valence	Float	-
tempo	Float	-
time_signature	Int	-

spotify_id: Es el identificador único de la canción, **PK** de la tabla.

nombre_cancion: Nombre de la canción.

artist_id: ID del artista asociado al registro, **FK** de la tabla **Artistas**.

explicit: Indica si la canción contiene lenguaje explícito. Al ser un dato de tipo booleano toma valores de True (Verdadero) o False (Falso).

duration_ms: Duración en milisegundos (1000 milisegundos= 1 segundo) de la canción.

album_id: ID del álbum asociado al registro, **FK** de la tabla **Albunes**.

danceability: Indica cuán “bailable” es la canción. Puede tomar valores entre 0 (no se puede bailar) y 1 (no se puede parar de bailar).

energy: Indica la “energía” (combinación entre intensidad y actividad) de la canción. Puede tomar valores entre 0 y 1.

key_id: Indica la tonalidad en la cual se encuentra la canción, **FK** de la tabla **Tonalidad**.

loudness: Indica cuán “ruidosa” es la canción (a partir de los decibeles).

mode_id: Indica el modo de la canción (mayor o menor), **FK** de la tabla **Modo**.

speechiness: Indica cuán “hablada” es la canción (presencia de palabras). Puede tomar valores entre 0 y 1.

acousticness: Indica la calidad acústica de la canción. Puede tomar valores entre 0 y 1.

instrumentalness: Indica la probabilidad de que la canción no contenga voces (resultando por ende instrumental). Puede tomar valores entre 0 y 1.

liveness: Medida de la presencia de audiencia en vivo de la grabación. Puede tomar valores entre 0 y 1.

valence: Indica la positividad emotiva de la canción. Puede tomar valores entre 0 (emociones negativas) y 1 (emociones positivas).

tempo: Indica el tempo de la canción (en bpm: pulsos por minuto)

time_signature: Indica la métrica del compás de la canción (cuántas negras por compás).

Tabla: Artistas

Columna	Tipo de dato	Clave
artista_id	Int	PK
artista	Varchar(300)	-

artista_id: Es el identificador único del artista, **PK** de la tabla.

artista: Nombre completo del artista.

Tabla: Albunes

Columna	Tipo de dato	Clave
album_id	Int	PK
album_name	Varchar(300)	-
album_release_date	Date(DD/MM/AAAA)	-

album_id: Es el identificador único del álbum, **PK** de la tabla.

album_name: Nombre completo del álbum.

album_release_date: Fecha de lanzamiento del álbum.

Tabla: Países

Columna	Tipo de dato	Clave
country_id	Int	PK
country_name	Varchar(2)	-

country_id: Es el identificador único del país según el código ISO de dos letras, **PK** de la tabla.

country_name: Nombre completo del país.

Tabla: Modo

Columna	Tipo de dato	Clave
mode_id	Int	PK
mode	Varchar(5)	-

mode_id: Es el identificador único del modo, **PK** de la tabla.

mode: Indica el modo de la canción (menor o mayor).

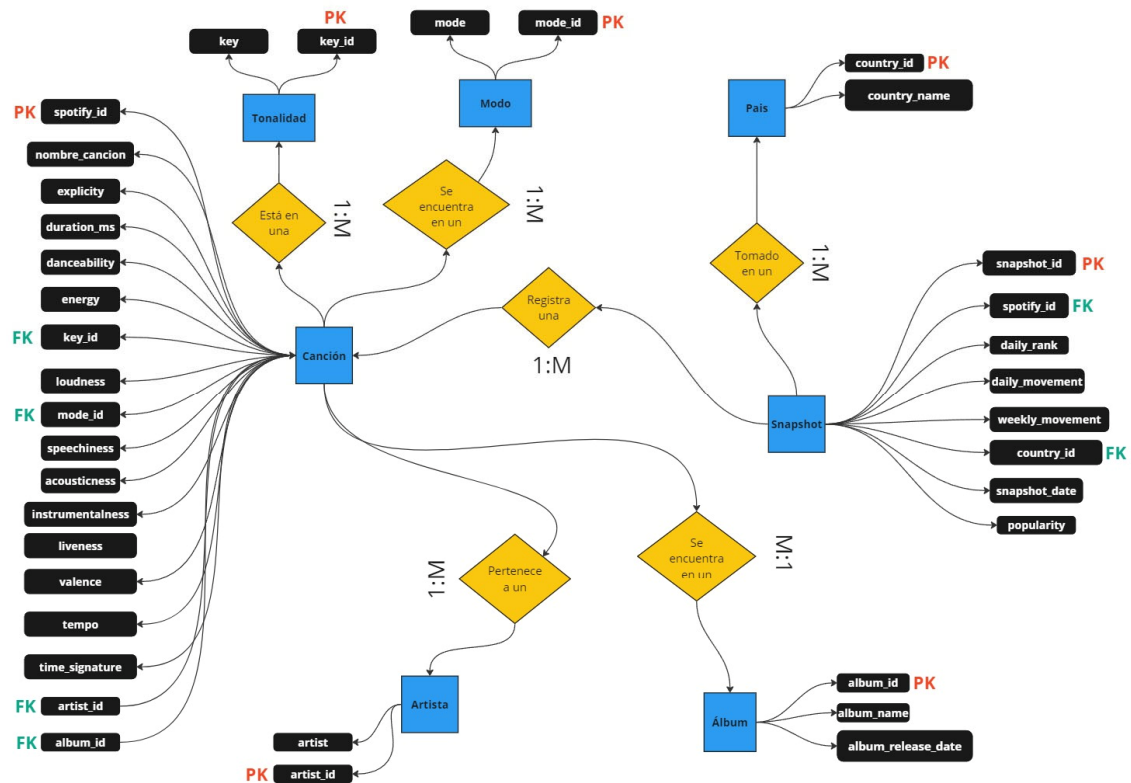
Tabla: Tonalidad

Columna	Tipo de dato	Clave
key_id	Int	PK
key	Varchar(2)	-

key_id: Es el identificador único de la tonalidad, **PK** de la tabla.

key: Indica la tonalidad de la canción en cifrado americano.

Diagrama de Entidad Relación:



Transformaciones realizadas:

Tabla Albums:

let

```
Source = Excel.Workbook(File.Contents("Canciones mas escuchadas en Spotify a nivel global - Martin Somoza.xlsx"), null, true),
```

```
Albumes_Sheet = Source[Item="Albumes",Kind="Sheet"][Data],
```

```
#"Promoted Headers" = Table.PromoteHeaders(Albumes_Sheet, [PromoteAllScalars=true]),
```

```
#"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"album_name", type text}, {"album_id", Int64.Type}, {"album_release_date", type date}}),
```

```
#"Removed Blank Rows" = Table.SelectRows(#"Changed Type", each not List.IsEmpty(List.RemoveMatchingItems(Record.FieldValues(_), {"", null})))
```

in

```
#"Removed Blank Rows"
```

Tabla Artistas:

let

```
Source = Excel.Workbook(File.Contents("Canciones mas escuchadas en Spotify a nivel global  
- Martin Somoza.xlsx"), null, true),
```

```
Artistas_Sheet = Source([Item="Artistas",Kind="Sheet"])[Data],
```

```
#"Promoted Headers" = Table.PromoteHeaders(Artistas_Sheet, [PromoteAllScalars=true]),
```

```
#"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"artist", type text},  
{"artist_id", Int64.Type}}),
```

```
#"Removed Blank Rows" = Table.SelectRows(#"Changed Type", each not  
List.IsEmpty(List.RemoveMatchingItems(Record.FieldValues(_), {"", null})))
```

in

```
#"Removed Blank Rows"
```

Tabla Canciones:

let

```
Source = Excel.Workbook(File.Contents("Canciones mas escuchadas en Spotify a nivel global  
- Martin Somoza.xlsx"), null, true),
```

```
Canciones_Sheet = Source([Item="Canciones",Kind="Sheet"])[Data],
```

```
#"Promoted Headers" = Table.PromoteHeaders(Canciones_Sheet, [PromoteAllScalars=true]),
```

```
#"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"spotify_id", type  
text}, {"nombre_cancion", type text}, {"artist_id", Int64.Type}, {"explicit", type logical},  
{"duration_ms", Int64.Type}, {"album_id", Int64.Type}, {"danceability", type text}, {"energy",  
type text}, {"key_id", Int64.Type}, {"loudness", type text}, {"mode_id", Int64.Type},  
{"speechiness", type text}, {"acousticness", type text}, {"instrumentalness", type text},  
{"liveness", type text}, {"valence", type text}, {"tempo", type text}, {"time_signature",  
Int64.Type}}),
```

```
#"Replaced Value" = Table.ReplaceValue(#"Changed  
Type", ".", "", Replacer.ReplaceText,{"danceability"}),
```

```
#"Changed Type2" = Table.TransformColumnTypes(#"Replaced Value",{{"danceability", type  
number}}),
```

```
#"Replaced Value1" = Table.ReplaceValue(#"Changed  
Type2", ":", "", Replacer.ReplaceText,{"energy", "loudness", "speechiness", "acousticness",  
"instrumentalness", "liveness", "valence", "tempo"}),
```

```
#"Changed Type3" = Table.TransformColumnTypes(#"Replaced Value1",{{"energy", type  
number}, {"loudness", type number}, {"speechiness", type number}, {"acousticness", type  
number}, {"instrumentalness", type number}, {"liveness", type number}, {"valence", type  
number}, {"tempo", type number}}),
```

```
#"Removed Blank Rows" = Table.SelectRows(#"Changed Type3", each not  
List.IsEmpty(List.RemoveMatchingItems(Record.FieldValues(_), {"", null}))),
```

```

#"Removed Duplicates" = Table.Distinct(#"Removed Blank Rows", {"spotify_id"}),

#"Changed Type1" = Table.TransformColumnTypes(#"Removed Duplicates",{{"danceability",
type number}})

in

#"Changed Type1"

```

Tabla Modalidad:

```

let

Source = Excel.Workbook(Canciones mas escuchadas en Spotify a nivel global - Martin
Somoza.xlsx), null, true),

Modalidad_Sheet = Source[{Item="Modalidad",Kind="Sheet"}][Data],

#"Promoted Headers" = Table.PromoteHeaders(Modalidad_Sheet,
[PromoteAllScalars=true]),

#"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"mode_id",
Int64.Type}, {"mode", type text}})

in

#"Changed Type"

```

Tabla Paises:

```

let

Source = Excel.Workbook(Canciones mas escuchadas en Spotify a nivel global - Martin
Somoza.xlsx), null, true),

Paises_Sheet = Source[{Item="Paises",Kind="Sheet"}][Data],

#"Changed Type" = Table.TransformColumnTypes(Paises_Sheet,{{"Column1", type text},
{"Column2", type text}}),

#"Renamed Columns" = Table.RenameColumns(#"Changed Type",{{"Column1",
"country_id"}, {"Column2", "country_name"}}),

#"Removed Top Rows" = Table.Skip(#"Renamed Columns",1),

#"Removed Blank Rows" = Table.SelectRows(#"Removed Top Rows", each not
List.IsEmpty(List.RemoveMatchingItems(Record.FieldValues(_), {"", null})))

in

#"Removed Blank Rows"

```

Tabla Snapshots:

let

```
Source = Excel.Workbook(File.Contents("Canciones mas escuchadas en Spotify a nivel global  
- Martin Somoza.xlsx"), null, true),
```

```
Snapshots_Sheet = Source{[Item="Snapshots",Kind="Sheet"]}[Data],
```

```
#"Promoted Headers" = Table.PromoteHeaders(Snapshots_Sheet, [PromoteAllScalars=true]),
```

```
#"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"snapshot_id",  
Int64.Type}, {"spotify_id", type text}, {"daily_rank", Int64.Type}, {"daily_movement",  
Int64.Type}, {"weekly_movement", Int64.Type}, {"country_id", type text}, {"snapshot_date",  
type date}, {"popularity", Int64.Type}}),
```

```
#"Removed Blank Rows" = Table.SelectRows(#"Changed Type", each not  
List.IsEmpty(List.RemoveMatchingItems(Record.FieldValues(_), {"", null})))
```

in

```
#"Removed Blank Rows"
```

Tabla Tonalidad:

let

```
Source = Excel.Workbook(File.Contents("Canciones mas escuchadas en Spotify a nivel global  
- Martin Somoza.xlsx"), null, true),
```

```
Tonalidad_Sheet = Source{[Item="Tonalidad",Kind="Sheet"]}[Data],
```

```
#"Promoted Headers" = Table.PromoteHeaders(Tonalidad_Sheet, [PromoteAllScalars=true]),
```

```
#"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"key_id",  
Int64.Type}, {"key", type text}})
```

in

```
#"Changed Type"
```

Tabla Rangos:

```
Rangos = ADDCOLUMNS(Paises,
```

```
"AvgDanc", CALCULATE(AVERAGE(Canciones[danceability])),
```

```
"AvgDur", CALCULATE(AVERAGE(Canciones[duration_ms])),
```

```
"AvgVol", CALCULATE(AVERAGE(Canciones[loudness])),
```

```
"AvgInst", CALCULATE(AVERAGE(Canciones[instrumentalness])),
```

```
"AvgPos", CALCULATE(AVERAGE(Canciones[valence])),
```

```
"AvgEnergy", CALCULATE(AVERAGE(Canciones[energy])),
```

```
"AvgAcus", CALCULATE(AVERAGE(Canciones[acousticness])),
```



```

"AvgTempo", CALCULATE(AVERAGE(Canciones[tempo])),
"AvgVoz", CALCULATE(AVERAGE(Canciones[speechiness])),
"AvgLive", CALCULATE(AVERAGE(Canciones[liveness])),
"AvgMovil", CALCULATE(AVERAGE(Snapshots[daily_move_abs]))
}

```

Tabla Calendario:

```

TablaCalendario = ADDCOLUMNS (
    CALENDAR (DATE(1900,1,1), DATE(2030,01,01)),
    "Id Fecha", FORMAT ( [Date], "YYYYMMDD" ),
    "Año", YEAR ( [Date] ),
    "Mes", FORMAT ( [Date], "MM" ),
    "Periodo", FORMAT ( [Date], "YYYY/MM" ),
    "Periodo2", FORMAT ( [Date], "YYYY/mmm" ),
    "Mes/Año", FORMAT([Date], "mm/yy"),
    "Año-Mes", FORMAT([Date], "yy-mm"),
    "Nombre Corto Mes", FORMAT ( [Date], "mmm" ),
    "Nombre Mes", FORMAT ( [Date], "mmm" ),
    "Día Semana número", WEEKDAY ( [Date] ),
    "Día Semana ", FORMAT ( [Date], "dddd" ),
    "Día Semana Corto", FORMAT ( [Date], "ddd" ),
    "Trimestre", "Q" & FORMAT ( [Date], "Q" ),
    "Año/trimestre", FORMAT ( [Date], "YYYY" ) & "/" & "Q" & FORMAT ( [Date], "Q" ),
    "Es Día Hábil", IF(WEEKDAY([Date], 2) < 6, "Si", "No"),
    "Month Name English", SWITCH(MONTH([Date]),
        1, "January",
        2, "February",
        3, "March",
        4, "April",
        5, "May",
        6, "June",

```

```

7, "July",
8, "August",
9, "September",
10, "October",
11, "November",
12, "December"
),
"Day of Week English", SWITCH(WEEKDAY([Date], 2),
1, "Monday",
2, "Tuesday",
3, "Wednesday",
4, "Thursday",
5, "Friday",
6, "Saturday",
7, "Sunday"
)
)
)

```

Medidas calculadas generadas:

AlbumesTotales = COUNT(Albumes[album_id])

ArtistasTotales = COUNT(Artistas[artist_id])

CancionesTotales = COUNT(Canciones[spotify_id])

PaísesTotales = COUNT(Países[country_id])

AlbumesPorFecha =
 CALCULATE(COUNT(Albumes[album_release_date]),USERRELATIONSHIP(Albumes[album_release_date],TablaCalendario[Date]))

PresenciasArtistaFiltradas =

IF(HASONEFILTER(Artistas[artist]),
 [CantidadSnapshots], "")

CantidadSnapshots = COUNT(Snapshots[snapshot_id])

AñoLanzamiento = YEAR(Albumes[album_release_date])

IndiceAntigüedad =
 CALCULATE(AVERAGE(Albumes[AñoLanzamiento]),USERRELATIONSHIP(Albumes[album_release_date],TablaCalendario[Date]))

daily_move_abs = ABS(Snapshots[daily_movement])

Movilidad = AVERAGE(Snapshots[daily_move_abs])

3/4 = (CALCULATE(COUNTROWS(Canciones),
 Canciones[time_signature]=3))/([CancionesTotales])

4/4 = (CALCULATE(COUNT(Canciones[spotify_id],
 Canciones[time_signature]=4))/([CancionesTotales])

Otra Metrica = 1-[3/4]-[4/4]

A = (CALCULATE(COUNT(Canciones[spotify_id]), Canciones[key_id]=9))/([CancionesTotales])

A# = (CALCULATE(COUNT(Canciones[spotify_id]), Canciones[key_id]=10))/([CancionesTotales])

B = (CALCULATE(COUNT(Canciones[spotify_id]), Canciones[key_id]=11))/([CancionesTotales])

C = (CALCULATE(COUNT(Canciones[spotify_id]), Canciones[key_id]=0))/([CancionesTotales])

C# = (CALCULATE(COUNT(Canciones[spotify_id]), Canciones[key_id]=1))/([CancionesTotales])

CantidadDePaises = DISTINCTCOUNT(Snapshots[country_id])

D = (CALCULATE(COUNT(Canciones[spotify_id]), Canciones[key_id]=2))/([CancionesTotales])

D# = (CALCULATE(COUNT(Canciones[spotify_id]), Canciones[key_id]=3))/([CancionesTotales])

E = (CALCULATE(COUNT(Canciones[spotify_id]), Canciones[key_id]=4))/([CancionesTotales])

F = (CALCULATE(COUNT(Canciones[spotify_id]), Canciones[key_id]=5))/([CancionesTotales])

F# = (CALCULATE(COUNT(Canciones[spotify_id]), Canciones[key_id]=6))/([CancionesTotales])

G = (CALCULATE(COUNT(Canciones[spotify_id]), Canciones[key_id]=7))/([CancionesTotales])

G# = (CALCULATE(COUNT(Canciones[spotify_id]), Canciones[key_id]=8))/([CancionesTotales])

Explicito = ((CALCULATE((COUNTROWS(Canciones),
 Canciones[explicit]=TRUE())))/[CancionesTotales])*100

No Explicito = ((CALCULATE((COUNTROWS(Canciones),
 Canciones[explicit]=FALSE())))/[CancionesTotales])*100

Mayor = CALCULATE(COUNT(Canciones[mode_id]), Canciones[mode_id]=1)

Menor = CALCULATE(COUNT(Canciones[mode_id]), Canciones[mode_id]=0)

PromedioAcusticidadGlobal = AVERAGE(Canciones[acousticness])

PromedioBailabilidadGlobal = AVERAGE(Canciones[danceability])

PromedioDuracionGlobal = AVERAGE(Canciones[duration_ms])

PromedioEnergiaGlobal = AVERAGE(Canciones[energy])

PromedioInstrumentalidadGlobal = AVERAGE(Canciones[instrumentalness])

PromedioPositividadGlobal = AVERAGE(Canciones[valence])

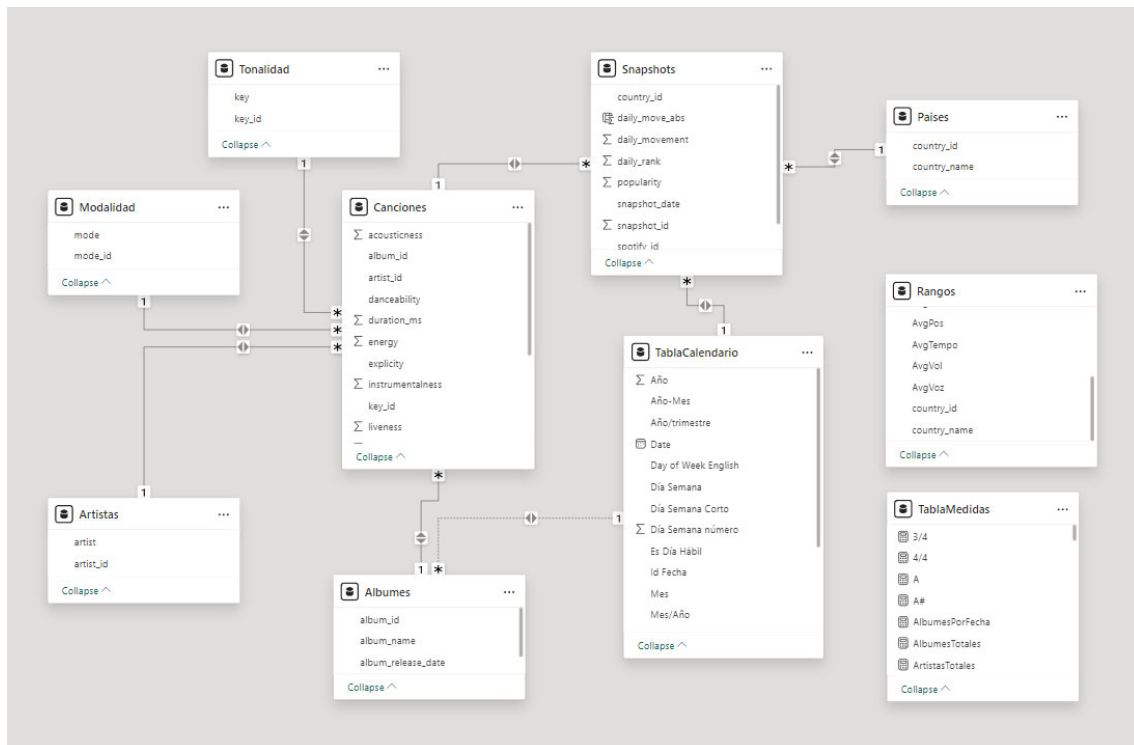
PromedioTempoGlobal = AVERAGE(Canciones[tempo])

PromedioVivoGlobal = AVERAGE(Canciones[liveness])

PromedioVolumenGlobal = AVERAGE(Canciones[loudness])

PromedioVozGlobal = AVERAGE(Canciones[speechiness])

Modelo de datos relacional de Power Bi:



Solapas del Dashboard:

Portada:

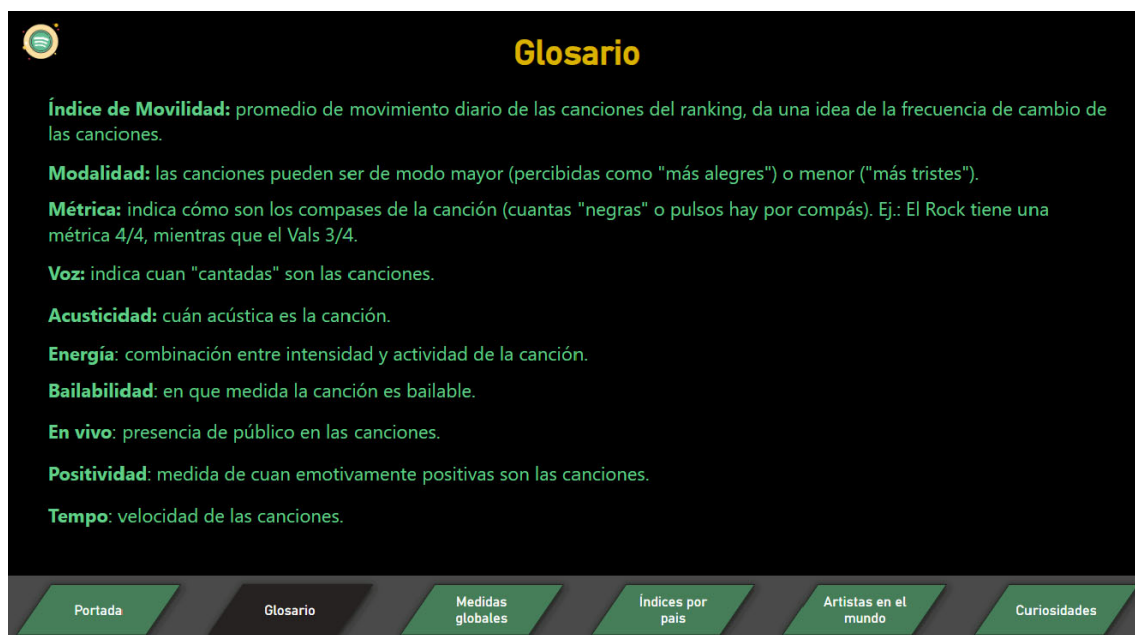
En la portada visualizamos el título del Dashboard y el período en el que la información fue recabada, junto con los datos de la comisión y autoría.

En el margen superior observamos una barra con botones de navegación a través de todas las solapas del Dashboard. Esta barra se encuentra presente en todas las solapas para facilitar la accesibilidad.



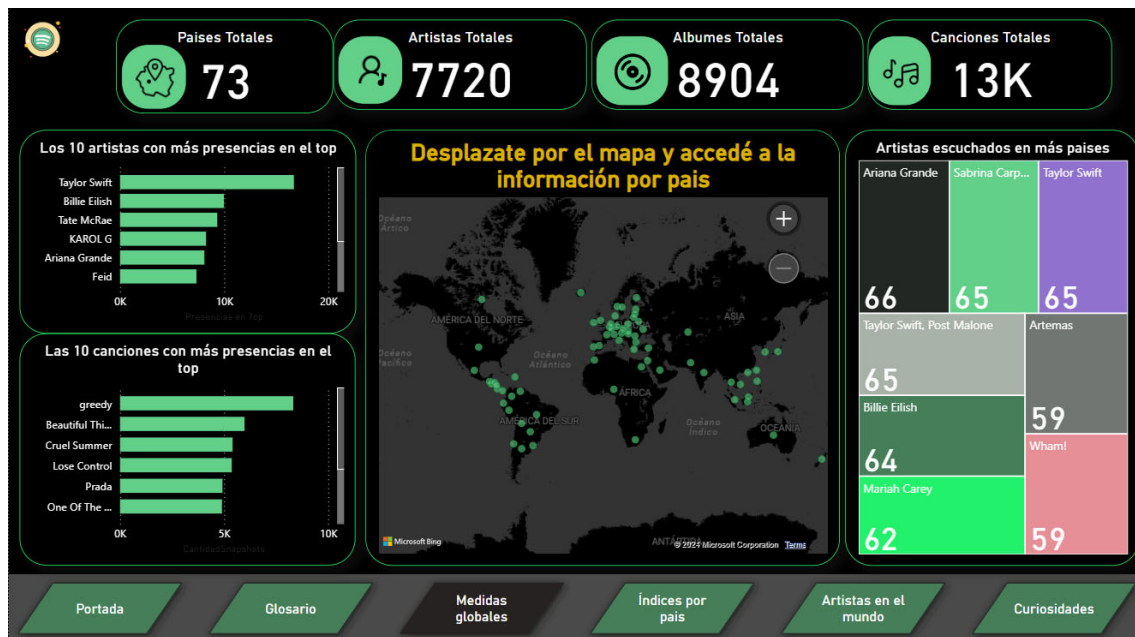
Glosario:

Aquí se encuentran definiciones de terminología utilizada en el Dashboard en vistas de facilitar la comprensión por parte del usuario.



Medidas globales:

En esta solapa encontraremos información global como las medidas de países totales de donde se recabó la información, cantidad total de artistas, cantidad total de álbumes y de canciones presentes en los rankings. A su vez, disponemos de un top 10 de los artistas y las canciones con más presencias en el ranking. Por otra parte, podemos visualizar cuáles son los artistas más globales, siendo escuchados en una cantidad mayor de países del mundo. El mapa es interactivo y permite ir visualizando la información país al clicar sobre él.

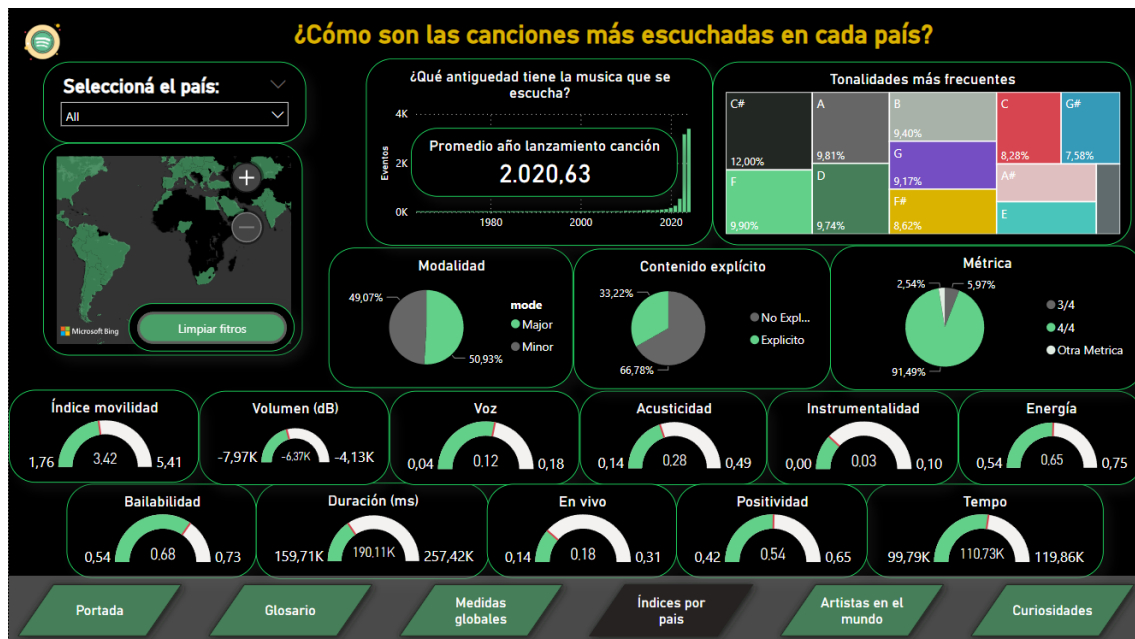


Índices por país:

En esta sección podemos visualizar cómo son las canciones más escuchadas en cada país: qué artistas son los que más se escuchan, qué tonalidades se utilizan más frecuentemente, la antigüedad de las canciones que pertenecen al ranking, el índice de movilidad (promedio de cuánto varían las canciones en el ranking en forma diaria), la modalidad, la presencia de contenido explícito en las letras, la métrica y todos los índices que caracterizan a las canciones (bailabilidad, volumen, duración, instrumentalidad, energía, acusticidad, tempo, voz, en vivo, positividad).

En el margen izquierdo superior se puede seleccionar el país en el menú desplegable.

Los valores en rojo en los indicadores corresponden al valor promedio global de esas variables. De esta manera, al seleccionar un país podemos ver fácilmente si se encuentra por debajo encima del mismo.



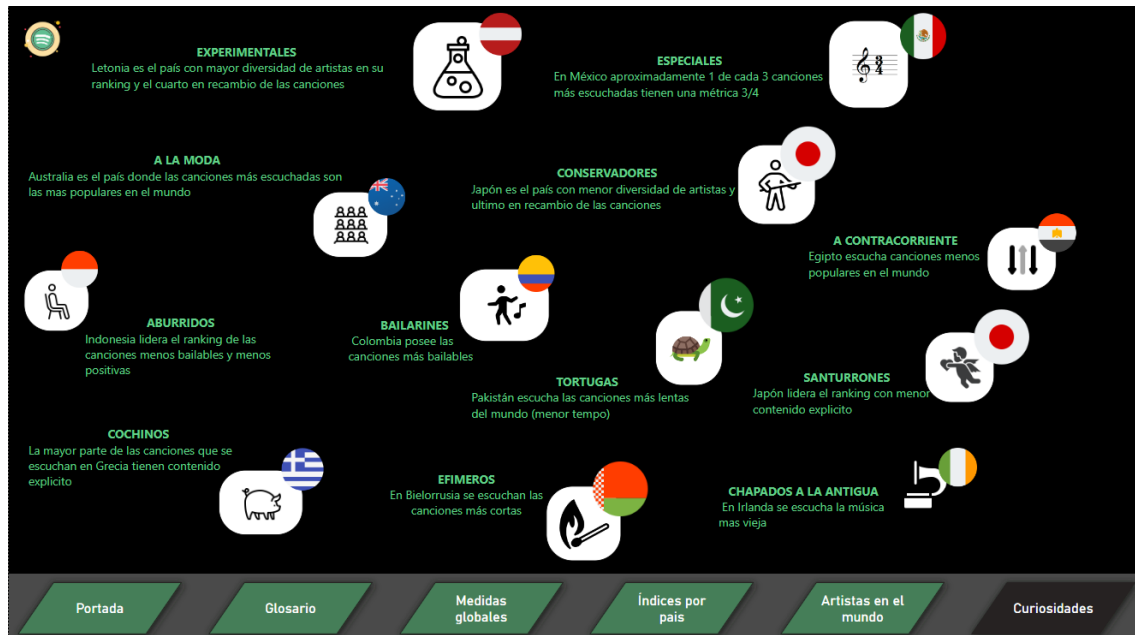
Artistas en el mundo:

En esta sección podemos seleccionar un artista y visualizar la distribución en el mundo de oyentes. También podemos filtrar con un segmentador el período a visualizar para comprender la dinámica en el tiempo de la distribución. El tablero nos muestra además cuales son los países con mayor cantidad de presencias del artista en el ranking, así como también las canciones más presentes. En el margen superior derecho observamos una tarjeta con el recuento de los países donde el artista se encuentra presente en el top.



Curiosidades:

En esta pestaña se mostrará información sobre características salientes de la música de diversos países como aquellos que poseen índices registrados como mínimos o máximos o una combinación de características (ej: bailarines, conservadores, a la moda, etc).



Conclusión:

A partir de lo analizado y visualizado en el Dashboard podemos concluir que efectivamente los países poseen preferencias musicales (reflejadas en los indicadores) y dinámicas temporales diversas.

Futuras líneas:

El análisis podría extenderse a incluir información respectiva a otras plataformas de música (Youtube Music, SoundCloud, Apple Music, Deezer, etc.) para poder abarcar y comprender el mercado en su totalidad. Se debe tener en cuenta que las poblaciones que utilizan las plataformas pueden variar entre sí y escuchar música distinta.

Podría también agregarse información del país de origen de cada uno de los artistas y analizar si la gente escucha más o menos a los artistas locales en contraposición con los extranjeros.

Otra variable a incorporar también podría ser el lenguaje local del país o el continente en el cual se encuentran, permitiendo encontrar así otras tendencias o patrones de valor.