



Spotify Data Análisis

Alumnos :

Sebastian Trejo
Sofia Larriva



1- Abstract:

Spotify es un servicio de música, podcasts y vídeos digitales que te da acceso a millones de canciones y a otro contenido de creadores de todo el mundo. Las funciones básicas, como escuchar música, son totalmente gratis y también tiene un sistema premium de pago.

2- Descripción de la temática del trabajo:

Hemos tomado una base de datos pública (desde Kaggle.com) ,de la cual , extrajimos datos de la plataforma Spotify . Encontramos interesante poder analizar , partiendo del tecnicismo musical , cuáles han sido los niveles de energía / cadencias / duración de los temas / acústica de los temas rankeados como más escuchados. También el impacto global que puede llegar a tener una canción y definir si algunas de estas características pueden influir al momento de componer un Hit.

3. Hipótesis :

En el siguiente trabajo analizaremos cuáles han sido los géneros más escuchados hasta el 2020 en la plataforma Spotify a nivel mundial interpretado por su popularidad, entre otras variables.

A su vez podremos develar a través de sus variables cuáles han sido los artistas/ canciones con mayor demanda por el público comparando su tipo de clave, niveles de energía, intensidad de sonido y así encontrar alguna relación con los géneros.

4- Data Set :

<https://docs.google.com/spreadsheets/d/1SI5eFNSqo2OqhMoizsObXsCaGUI920n89UxCNNrbSY/edit?usp=sharing>



Tablas:

Artistas:

Tipo de Clave	Campo	Tipo de Campo
fk	id_cancion	int
	genero	varchar
	nombre_artista	varchar
	nombre_cancion	varchar
	popularidad	int
	fecha_cancion	date
pk	id_artista	int

Música:

Tipo de Clave	Campo	Tipo de Campo
pk	id_cancion	int
fk	genero	varchar
	acustica	int
	duracion_ms	int
	energia	int
	llave	varchar



	vivacidad	int
	volumen	int
	escala	varchar

Géneros:

Tipo de Clave	Campo	Tipo de Campo
	id_cancion	int
pk	genero	varchar
	nombre_cancion	varchar
	id_artista	int

ID Genero:

Tipo de Clave	Campo	Tipo de Campo
pk	id_genero	int
fk	genero	varchar

Artistas (2):

Tipo de Clave	Campo	Tipo de Campo
pk	nombre_artista - copy	varchar

calendario fecha canciones:

Tipo de Clave	Campo	Tipo de Campo
---------------	-------	---------------



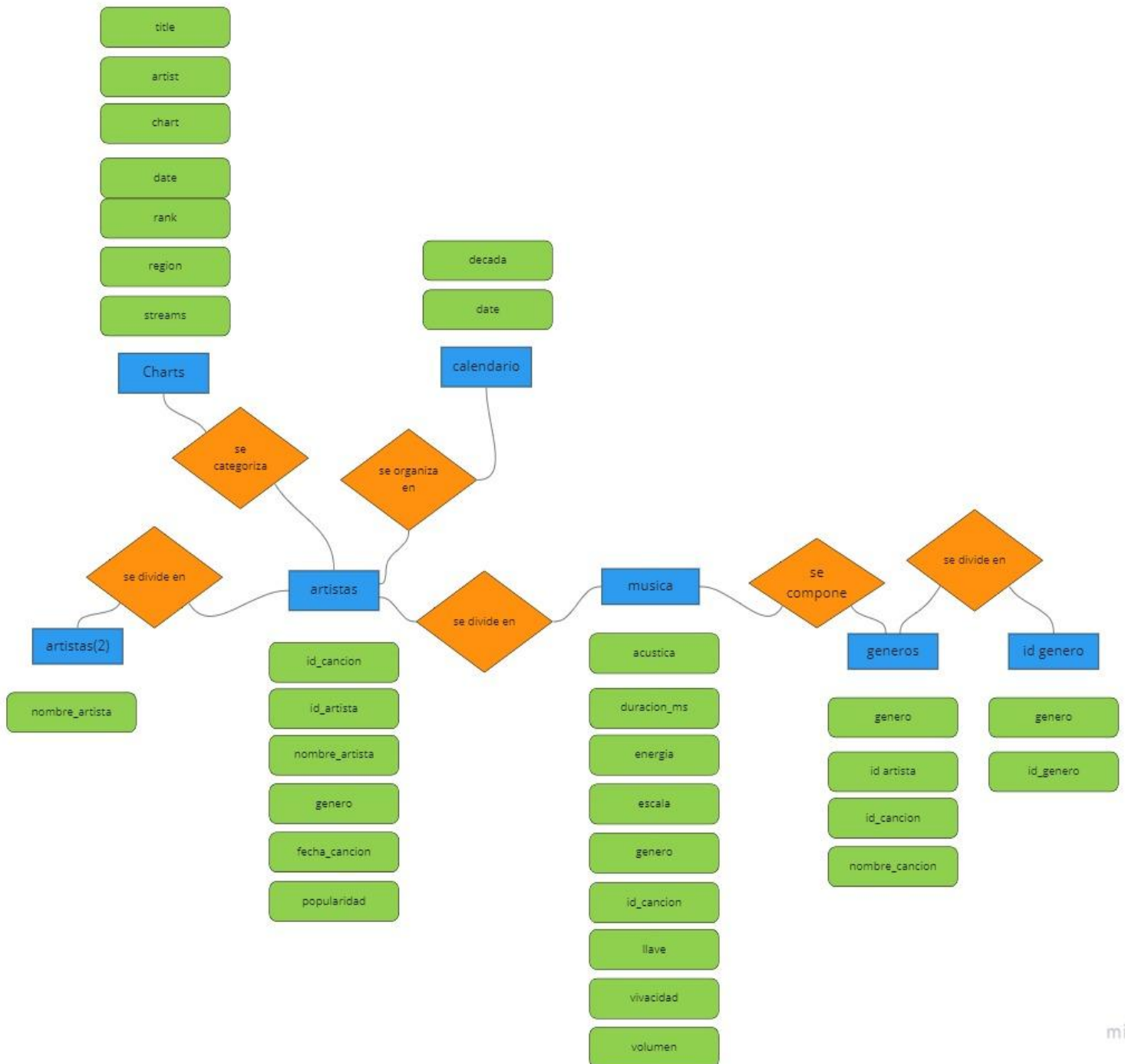
pk	date	int
	decada	int

Charts:

Tipo de Clave	Campo	Tipo de Campo
pk	title	varchar
fk	artist	varchar
	date	date
	rank	varchar
	region	varchar
	streams	int
	chart	varchar



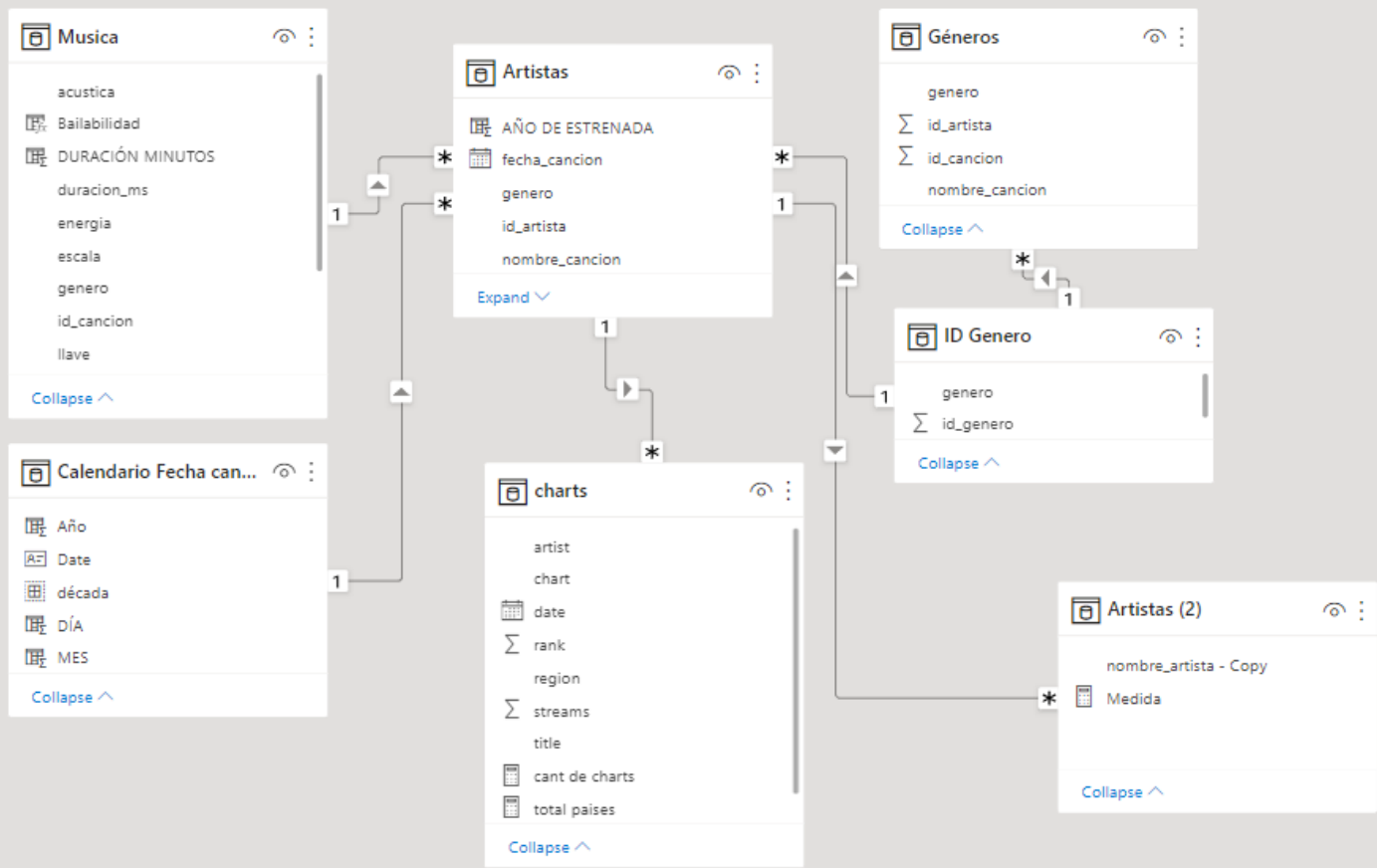
Diagrama Entidad Relación



Nuestro diagrama entidad - relación se conectan todas las tablas directamente con artistas, actualizada luego de incorporar nuevas tablas puentes (id genero, artistas(2)), tablas complementarias como “tabla calendario” y “charts” que ayuda a graficar la cantidad de streams que tienen las canciones a nivel global junto al ranking.



Transformación de datos



Cargamos el dataset a power bi y lo organizamos tal como se ve en la imagen anterior, agregamos una tabla para alimentar la base de spotify con streams(Charts) y sus puntos de alcance a nivel global. fue creada una nueva tabla para conectar la tabla de género con la tabla de artistas llamada "ID Genero" resultando ser la tabla puente al igual que la tabla "Artistas (2)".



Medidas Calculadas

1: Total artista = `DISTINCTCOUNT(Artistas[id_artista])` cantidad total de artistas

2: total paises = `DISTINCTCOUNT(charts[region])` cantidad total de países en el chart

3: cant de charts = `DISTINCTCOUNT(charts[chart])` cantidad total de charts

4: total streams = `SUM(charts[streams])` cantidad total de Streams

5:total genero = `COUNT('ID Genero'[genero])` cantidad total de generos

6:Promedio de Mediana de DURACIÓN MINUTOS por genero =

```
AVERAGEX(  
  
    KEEPFILTERS (VALUES ('Musica'[genero])),  
  
    CALCULATE (MEDIAN ('Musica'[DURACIÓN MINUTOS]))  
)
```

medida para calcular la mediana de duración de minutos por genero

7:Promedio de Recuento de DURACIÓN MINUTOS por id_cancion =

```
AVERAGEX(  
  
    KEEPFILTERS (VALUES ('Géneros'[id_cancion])),  
  
    CALCULATE (COUNTA ('Musica'[DURACIÓN MINUTOS]))  
)
```

promedio de recuento de duración de los minutos por canción

8:Promedio de Recuento de genero por DURACIÓN MINUTOS =

```
AVERAGEX(  

```




```
KEEPFILTERS (VALUES ('Musica' [DURACIÓN MINUTOS])),  
  
CALCULATE (COUNTA ('Musica' [genero]))  
  
)
```

Medida que calcula el promedio de duración en minutos de canciones.

Columnas agregadas

```
1: AÑO DE ESTRENADA = 2022 - 'Artistas'[fecha_cancion].[Año]
```

calculamos los años que lleva de estreno cada cancion.

```
2: Año = Year('Calendario Fecha canción'[Date])
```

Columna dividiendo los años en tabla calendario

```
3: Día = DAY([Date]) Columna dividiendo los días en tabla calendario
```

```
4: MES = MONTH([Date]) Columna dividiendo los meses en tabla calendario
```

```
5: Bailabilidad =  
if(and(Musica[vivacidad]>10000000,Musica[acustica]>=100),"bailable","no bailable")
```

columna agregada en tabla “Musica”, divide los resultados entre “bailable” y “no bailable” tomando como referencia la medida de “vivacidad” que responde a esta incógnita.

```
6: DURACIÓN MINUTOS = Musica[duracion_ms]*1.66666666666667E-05
```

se agrega una columna a la tabla musica donde transformamos la “duracion_ms” a tiempo (minutos) utilizando una formula de transformación.



Tabla calendario

```
1: Calendario Fecha canción =  
calendar(FIRSTDATE('Artistas'[fecha_cancion]),LASTDATE('Artistas'[fecha_cancion]))
```

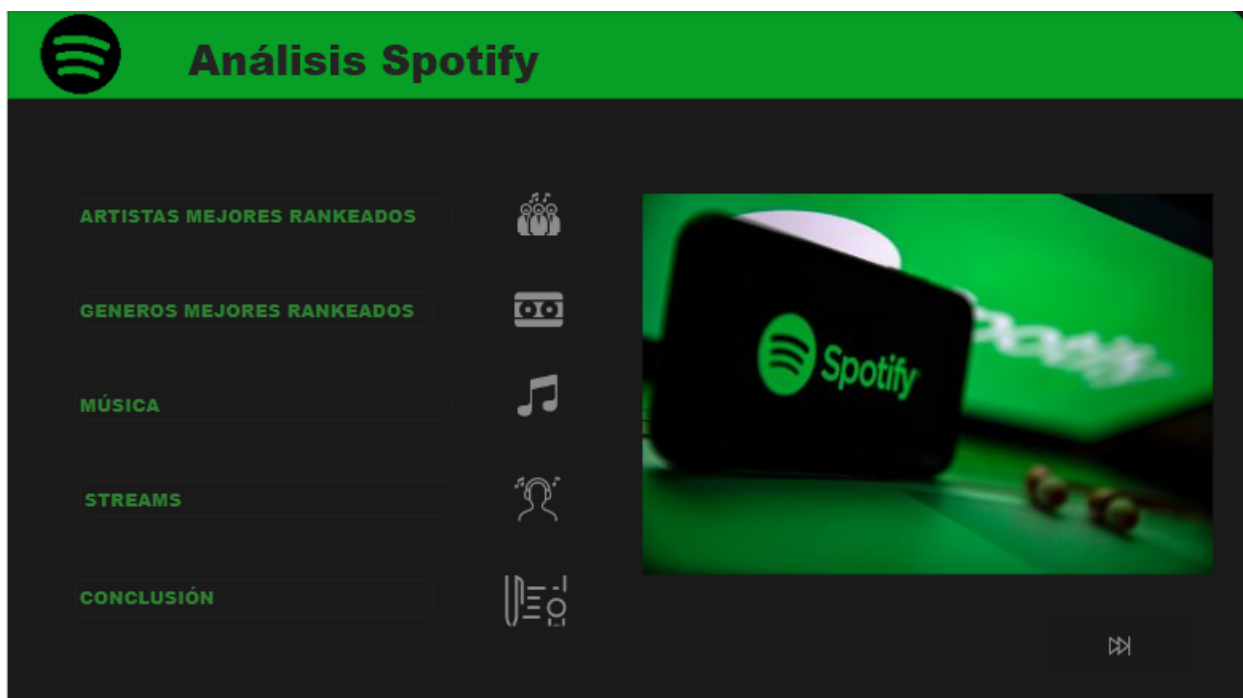
Tabla calendario creada tomando la primera fecha hasta la ultima de las canciones publicadas.

5. Visualización de los datos

Podrán encontrar un resumen detallado de cada solapa que conforma el tablero final.

5.1 Inicio : Partimos de una solapa que contiene botones creados para que el usuario pueda elegir desde un comienzo como quiere leer la historia que contamos. Además, incluimos un botón que simula flecha de "rewind" para que puedan avanzar o retroceder , simulando la experiencia de reproducir música.

Tablero de Inicio :





5.2 Artistas Destacados :

Se creó esta página con el fin de obtener un resumen de los artistas destacados y los títulos registrados por artistas mediante un gráfico de elementos claves en el cual se puede discriminar la información por popularidad y género . Además se incluye una tabla para observar un detalle de las canciones/género/ artista que contiene el data set, En el extremo superior derecho, se añadió una segmentación de datos. Dicho comando le permitirán al usuario visualizar la información en rangos de años , si así lo desea.



5.2 Géneros Destacados :

En esta página mostramos la incidencia de los géneros en cuanto a la popularidad y su relación con los temas que componen el data set.

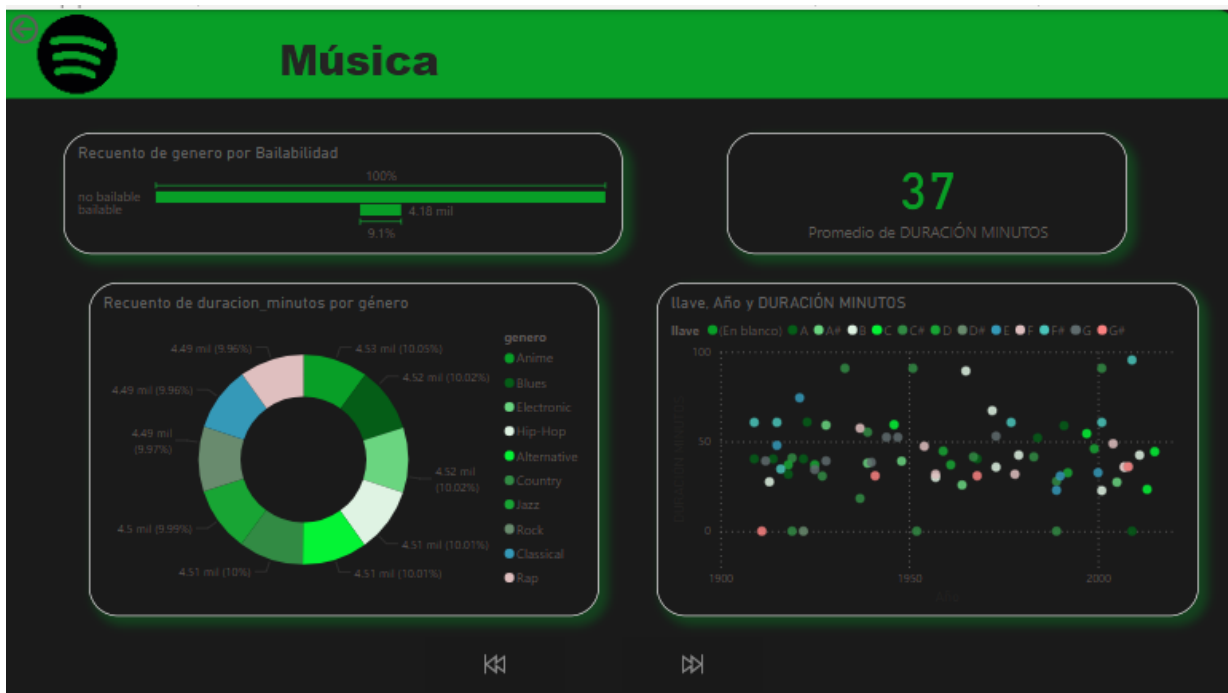


5.3 Música :

En esta página mostramos la incidencia de los géneros en cuanto a la popularidad y su relación con los temas que componen el data set.

Se muestra a su vez el promedio de duración de las canciones y también , usando una función condicional , hemos discriminado cuales son las canciones categorizadas como bailables o no bailables.

En el margen inferior izquierdo pudimos realizar un gráfico de dispersión en el cual mostramos con el correr de los años cómo ha variado la composición de las canciones en cuanto a duración respecta.



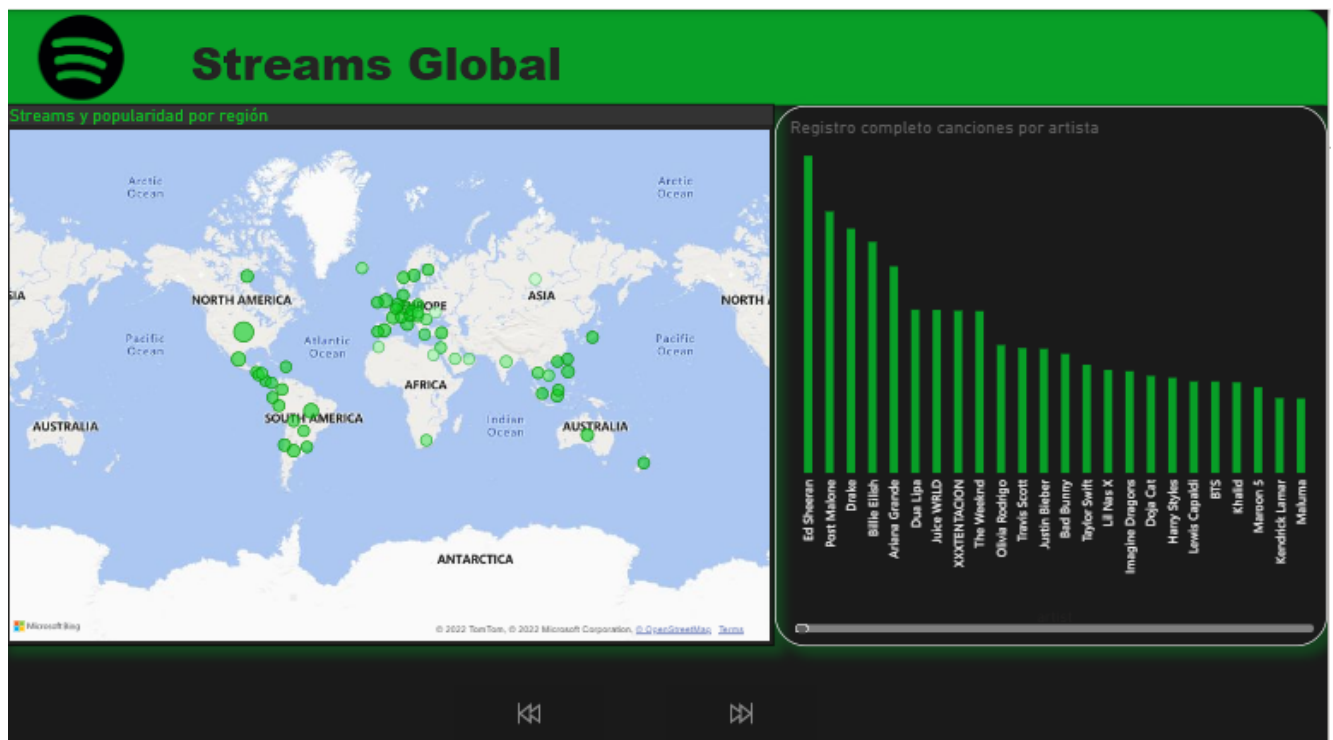


5.5 Stream Global :

La página no fue incluida en el mockup, pero se consideró relevante para mejorar la experiencia del usuario con el tablero. De esta forma, el mismo podrá visualizar datos actuales y sentir más familiaridad con la información que se expone.

Se agregaron 2 gráficos en el cual exponemos un gráfico de cintas donde se exponen la relevancia por artistas y títulos que hicieron que los mismos estén rankeados en el stream de spotify.

Además se incluyó un mapa para que se visualice en un mapa continental la relevancia del stream a nivel global.





5.6 Conclusión :

Se observó que la mayor cantidad de canciones son del genero rap lo cual hace sentido , ya que actualmente con la llegada de los streams , y la facilidad de reproducir las canciones con la plataforma spotify , hacen que el alcance sea exponencial y el género continúa siendo tendencia. Así mismo podemos asociar el género trap , 2022 ha sido un año donde la fusión rap/ free style ha llegado ser máxima tendencia a nivel mundial.

El rock es el género que le sigue. El mismo ha sabido ser desde los 70 a los 90 un género que trajo nuevos referentes musicales .Puede considerarse una de las corrientes musicales más influyentes, importantes y auténticas de la historia de la música.

En cuanto a la extensión de las canciones varía en función del medio de distribución que apoye al mercadeo. En la actualidad Spotify y otras plataformas similares son la forma favorita de escuchar música y como el modelo de pago es por cada vez que se escucha, es más rentable hacerlas cortas. Algunos raperos apuntan a canciones de 2 minutos/3 minutos. Dado a que nuestro dataset tiene datos desde 1908 en adelante podríamos concluir en que en el siglo XX el promedio de minutos de tema era considerablemente mayor (37 minutos) que el promedio actual (3 minutos) .

Podemos concluir en que dado a que actualmente dado a que nuestra capacidad de concentración es menor que en la antigüedad , la industria de la música se ha amoldado a estos consumos , Tras los discos de vinilo, aparecieron los CD's, que podían almacenar hasta 70 minutos de música. Los artistas podían desarrollar temas más largos, pero siguieron haciendo canciones de tres minutos para adaptarse a las exigencias de las emisoras de radio.



Conclusión

70

total países

1 bill.

total streams

streams por region

