



Data Analytics

Alumno: Mena, Jeremías

Profesor: Parente Aun, Ruben Alfredo

Tutor: Álvarez, Sara Belen

Comisión: #57435



Índice

Índice.....	2
Introducción.....	3
Alcance.....	4
Metodología.....	5
Usuario final y nivel de aplicación del análisis.....	6
Descripción del data set.....	7
Hipótesis.....	8
Diagrama Entidad-Relación.....	10
Descripción de las tablas.....	11
Dashboard.....	14
Conclusión.....	34

Introducción

En nuestra rutina diaria, es común disfrutar de momentos de ocio como leer un libro, ver una película, pasar tiempo con amigos o simplemente escuchar música. Este último, en particular, tiene un impacto significativo en nuestra vida cotidiana.

Estudios científicos indican que al escuchar música, nuestro cerebro produce dopamina, un neurotransmisor asociado al placer y la recompensa. La cantidad exacta de dopamina generada no puede determinarse con certeza, ya que varía según diferentes factores, como las preferencias musicales individuales, el contexto de la escucha y la intensidad emocional de la experiencia musical. Sin embargo, se reconoce que su influencia es mayormente positiva y, en muchos casos, ha sido beneficiosa para tratar trastornos como la depresión.

En el ámbito empresarial de la reproducción de archivos multimedia, Spotify destaca como una de las principales compañías. Con alrededor de 8,360 empleados y presencia en 187 países, Spotify ofrece anualmente a sus usuarios un resumen que detalla estadísticas como los artistas más escuchados, las canciones más reproducidas y el total de minutos dedicados a la música.

Este informe se enfoca en el análisis de datos relacionados con las canciones más populares en Spotify. Utilizando Power BI como herramienta principal, exploraremos cómo esta plataforma nos permite desglosar, comprender y aprovechar estos datos para identificar patrones en las tendencias musicales. También examinaremos las etapas clave para diseñar y desarrollar un panel de análisis efectivo en Power BI.

Alcance

El objetivo principal del proyecto es analizar y extraer información relevante de un dataset que incluye detalles sobre canciones, artistas, géneros musicales, streams y otros datos relacionados. Este análisis busca identificar tendencias, patrones e insights que puedan ser utilizados para mejorar la toma de decisiones en la industria.

Partiendo del objetivo principal, definiremos los siguientes **objetivos específicos**:

Análisis de las Canciones

- Identificar las canciones más populares utilizando diversas métricas.
- Identificar las canciones más populares de cada estación del año.
- Identificar la canción que más veces entró al top 10 de canciones globales.

Análisis de Artistas

- Identificar a los artistas más populares utilizando diversas métricas.
- Identificar las canciones más populares de cada estación del año.
- Examinar la diversidad de los artistas teniendo en cuenta el género musical en el que crean música y su país de origen.

Análisis de Géneros Musicales

- Investigar los géneros musicales más populares y cómo su popularidad ha cambiado con el tiempo.
- Estudiar la correlación entre ciertos géneros musicales y su éxito en términos de reproducciones.

Metodología

Para realizar un análisis exhaustivo y preciso de esta base de datos se ha establecido una metodología estructurada. La misma está diseñada para garantizar la calidad de los datos, descubrir insights significativos y presentar resultados claros y útiles.

Recolección de datos

El primer paso consiste en la exportación y limpieza de los datos desde el dataset en cuestión. Este proceso incluye la normalización de los datos para garantizar su coherencia e integridad, eliminando posibles errores y redundancias.

Análisis exploratorio de datos

En esta etapa, se emplean distintas técnicas para comprender las principales características de los datos.

Modelado y análisis

Aquí, se aplican técnicas de análisis de datos para identificar patrones y tendencias. Se utilizan herramientas como SQL y Power BI para la visualización y análisis, facilitando así una comprensión más profunda de los mismos.

Visualización de resultados

Finalmente, se crean dashboards interactivos y reportes visuales para comunicar los insights de la mejor manera.

Usuario final y nivel de aplicación del análisis

El destinatario principal de este análisis es la industria musical, que abarca diversas entidades del sector. Este análisis se aplica a un nivel estratégico, ofreciendo información y conocimientos que respaldan la toma de decisiones a largo plazo y con un impacto significativo en la industria musical.

Descripción del data set

El conjunto de datos propuesto se centra en las canciones que han alcanzado el top de Spotify, recopilando información variada sobre las más populares en años recientes.

Este dataset abarca una amplia gama de variables relevantes que ofrecen detalles sobre música, como el número de reproducciones de una canción específica, la frecuencia con la que ha alcanzado el top y su posición dentro de él. Además, podría incluir datos contextuales adicionales, como el número de artistas involucrados en una canción, su país de origen o sexo.

Al analizar este conjunto de datos, es posible identificar tendencias y patrones en los gustos musicales de los usuarios de Spotify a lo largo del tiempo. Se pueden comprender qué géneros son los más populares, qué canciones son las más reproducidas y cuáles han logrado mantenerse en el top 10 durante períodos prolongados.

El objetivo principal de este dataset es ofrecer a los usuarios una visión más amplia de las tendencias musicales recientes. Aplicando técnicas de análisis de datos y herramientas, es posible transformar esta información en visualizaciones atractivas y comprensibles.

Hipótesis

En el mundo digital actual, la música ha alcanzado un papel central en la vida de las personas, con plataformas de streaming como Spotify que ofrecen un acceso sin precedentes a una amplia gama de canciones.

La hipótesis que guía este análisis sugiere que, más allá de la calidad musical intrínseca, ciertos elementos como el género, la colaboración artística y la promoción pueden desempeñar roles significativos en la ascensión de una canción al estatus de "éxito". Explorar esta hipótesis nos permitirá comprender mejor las dinámicas detrás de las canciones más populares de los últimos años y ofrecerá perspectivas valiosas para los usuarios, la industria musical y el análisis de tendencias. A continuación, se presentan algunos elementos que podrían ser relevantes en este contexto:

Temporada

La estacionalidad podría influir en la popularidad de una canción. Por ejemplo, se podría comparar el número de reproducciones de canciones lanzadas en invierno con aquellas lanzadas en verano para evaluar si las canciones de verano tienen más streams.

Géneros populares

A sabiendas de que cada canción tiene un género en particular, se podría investigar a cuáles de estos pertenecen la mayoría de canciones.

Artistas populares

Si analizamos qué artistas contribuyen a cada canción, podríamos elaborar una lista de los más populares.

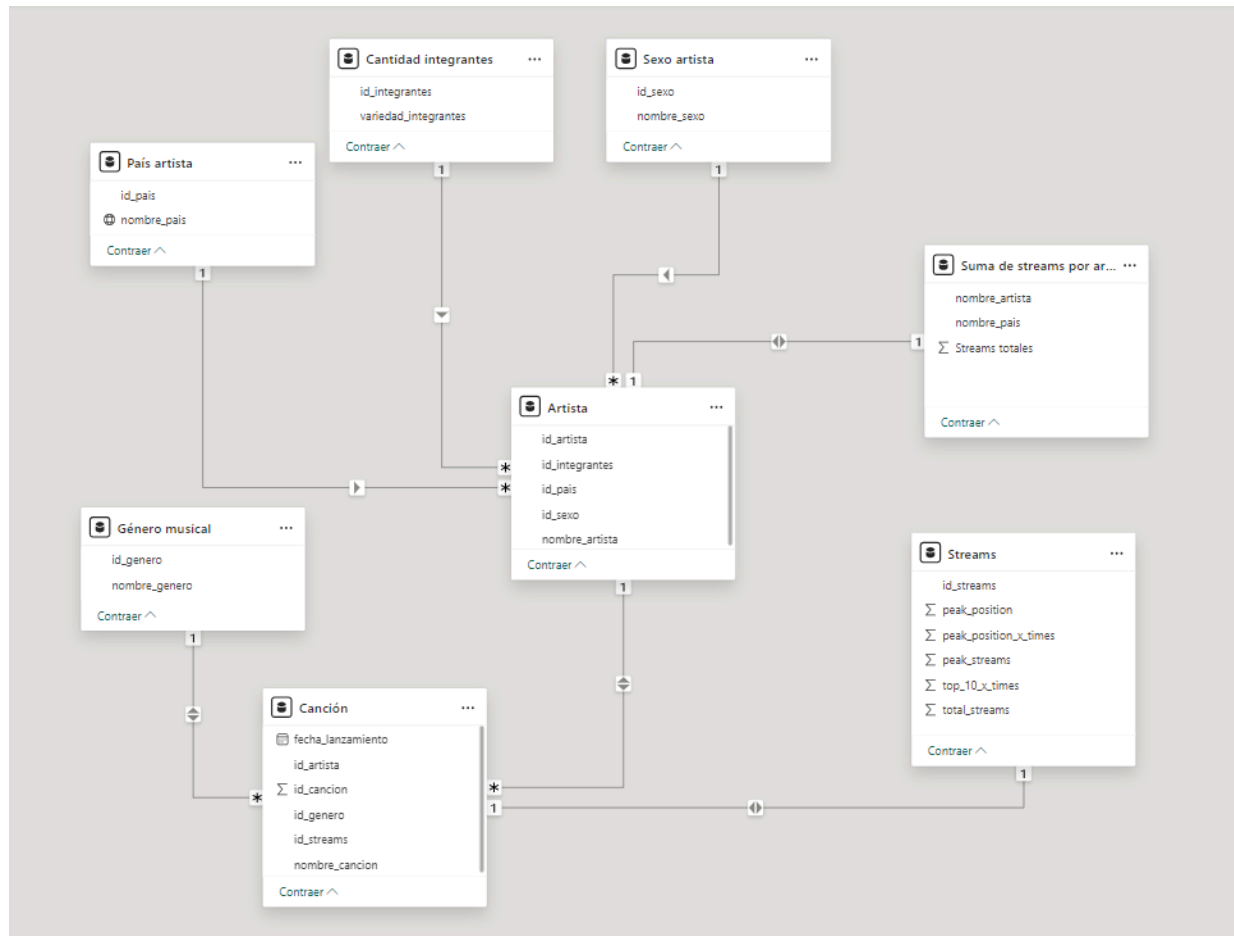
Países con más artistas

Conociendo a qué país pertenece cada uno de los artistas del dataset, se podría conocer cuáles son los países que tienen más artistas y determinar si USA es el país que más tiene.

Tendencias

Se podría hipotetizar si los géneros de las canciones han sufrido un aumento o disminución en su popularidad a lo largo de un año, analizando las reproducciones que obtuvieron en cada una de las estaciones del año.

Diagrama Entidad-Relación



Descripción de las tablas

A continuación se realizará un listado del contenido de cada tabla

Artista	
Variable	Descripción
id_artista	Es la clave primaria de la tabla. Identifica unívocamente a cada artista.
nombre_artista	Hace alusión al nombre de cada artista.
id_pais	Clave foránea. Relaciona esta tabla con la tabla pais_artista .
id_genero	Clave foránea. Relaciona esta tabla con la tabla genero_musical .
id_sexo	Clave foránea. Relaciona esta tabla con la tabla sexo_artista .
id_integrantes	Clave foránea. Relaciona esta tabla con la tabla cantidad_integrantes .

Canción	
Variable	Descripción
id_cancion	Es la clave primaria de la tabla. Identifica unívocamente a cada canción.
nombre_cancion	Hace alusión al nombre de cada canción.
fecha_lanzamiento	Hace alusión a la fecha de lanzamiento de la canción.
id_artista	Clave foránea. Relaciona esta tabla con la tabla artista .
id_fecha	Clave foránea. Relaciona esta tabla con la tabla calendario .
id_streams	Clave foránea. Relaciona esta tabla con la tabla streams .

Cantidad integrantes	
Variable	Descripción
id_integrantes	Es la clave primaria de la tabla. Identifica unívocamente a cada tipo de integrantes.
variedad_integrantes	Tipo de integrantes.

Género musical	
Variable	Descripción
id_genero	Es la clave primaria de la tabla. Identifica unívocamente a cada género.
nombre_genero	Hace alusión al nombre de cada género musical.

País artista	
Variable	Descripción
id_pais	Es la clave primaria de la tabla. Identifica unívocamente a cada país.
nombre_pais	Hace alusión al nombre de cada país.

Sexo artista	
Variable	Descripción
id_sexo	Es la clave primaria de la tabla. Identifica unívocamente a cada sexo.
nombre_sexo	Hace alusión a cada tipo de sexo.

Streams	
Variable	Descripción
id_streams	Es la clave primaria de la tabla. Identifica unívocamente a cada stream.
top_10_x_times	Cantidad de veces que la canción estuvo en el top 10.
peak_position	Mejor posición en la que estuvo la canción en el ranking de popularidad.
peak_position_x_times	Cantidad de veces que estuvo en esa posición antes descrita.
peak_streams	Cantidad máxima de reproducciones que la canción alcanzó.
total_streams	Cantidad total de reproducciones de la canción.

Dashboard

A continuación se describirán las distintas medidas, columnas y tablas dinámicas generadas para la funcionalidad del tablero.

Nombre	Descripción	Código
Artistas más populares		
Nombre del artista más popular	Esta medida obtiene el nombre del artista más popular basándose en la suma de las reproducciones que obtuvo en las canciones que produjo.	<pre> Artista más popular = VAR Nombre = LOOKUPVALUE (Artista[nombre_artista], 'Suma de streams por artista'[Streams totales], MAX('Suma de streams por artista'[Streams totales])) RETURN IF(ISBLANK(Nombre), "No se encontró resultado", Nombre) </pre>
País del artista más popular	Esta medida determina el país del artista más popular basándose en la suma de las reproducciones que obtuvo en las canciones que produjo	<pre> País del artista del artista más popular = VAR N_Artista = LOOKUPVALUE (Artista[nombre_artista], 'Suma de streams por artista'[Streams totales], MAX('Suma de streams por artista'[Streams totales])) VAR Id_Pais = LOOKUPVALUE (Artista[id_pais], Artista[nombre_artista], N_Artista) VAR Pais = LOOKUPVALUE ('País artista'[nombre_pais], 'País artista'[id_pais], Id_Pais) RETURN IF(ISBLANK(Pais), "No se encontró resultado", Pais) </pre>
Sexo del artista más popular	Esta medida determina el sexo del artista más popular basándose en la suma de las reproducciones que	<pre> Sexo del artista del artista más popular = VAR N_Artista = LOOKUPVALUE (Artista[nombre_artista], 'Suma de streams por artista'[Streams totales], MAX('Suma de streams por </pre>

	obtuvo en las canciones que produjo	<pre> artista'[Streams totales])) VAR Id_Sexo = LOOKUPVALUE(Artista[id_sexo], Artista[nombre_artista], N_Artista) VAR Sexo = LOOKUPVALUE('Sexo artista'[nombre_sexo], 'Sexo artista'[id_sexo], Id_Sexo) RETURN IF(ISBLANK(Sexo), "No se encontró resultado", Sexo) </pre>
Streams totales del artista más popular	Esta medida obtiene la cantidad de reproducciones totales del artista más popular.	Streams totales del artista más popular = <pre> LOOKUPVALUE('Suma de streams por artista'[Streams totales], 'Suma de streams por artista'[nombre_artista], [Nombre del artista más popular]) </pre>
Top 10 artistas más populares		
Nombre del artista Top N	Esta medida obtiene el nombre del artista según el top seleccionado en el filtro.	Nombre del artista Top N = <pre> VAR TopN_Value = SWITCH(SELECTEDVALUE('Top 10 artistas'[Top 10], "Top 10 artistas más populares"), "Primer artista", 1, "Segundo artista", 2, "Tercer artista", 3, "Cuarto artista", 4, "Quinto artista", 5, "Sexto artista", 6, "Séptimo artista", 7, "Octavo artista", 8, "Noveno artista", 9, "Décimo artista", 10, 1) VAR TopN_Artista = TOPN(1, TOPN(TopN_Value, </pre>

		<pre> 'Suma de streams por artista', 'Suma de streams por artista'[Streams totales], DESC), 'Suma de streams por artista'[Streams totales], ASC) VAR N_Artista = SELECTCOLUMNS(TopN_Artista, "Artista", 'Suma de streams por artista'[nombre_artista]) RETURN N_Artista </pre>
País del artista Top N	Esta medida obtiene el país del artista según el top seleccionado en el filtro.	<pre> País del artista Top N = VAR Nombre_Artista = [Nombre del artista Top N] VAR Pais = LOOKUPVALUE('Suma de streams por artista'[nombre_pais], 'Suma de streams por artista'[nombre_artista], Nombre_Artista) RETURN Pais </pre>
Foto del artista Top N	Esta medida obtiene la url de la foto del artista según el top seleccionado en el filtro.	<pre> Foto del artista Top N= VAR Selected = SELECTEDVALUE('Top 10 artistas'[Top 10]) VAR Foto = LOOKUPVALUE('Top 10 artistas'[URL Imagen], 'Top 10 artistas'[Top 10], Selected) RETURN Foto </pre>
Canciones más populares		
Nombre de la canción más popular	Esta medida obtiene el nombre de la canción más popular teniendo en cuenta las reproducciones que	<pre> Nombre de la canción más popular = VAR Nombre = LOOKUPVALUE('Canción'[nombre_cancion], Streams[total_streams], [Cantidad de streams]) RETURN </pre>

	obtuvo.	<code>IF(ISBLANK(Nombre), "No se encontró resultado", Nombre)</code>
Nombre del artista de la canción más popular	Esta medida obtiene el nombre del artista de la canción más popular.	<p>Nombre del artista de la canción más popular =</p> <pre> VAR Id_Artista = LOOKUPVALUE('Canción'[id_artista], Streams[total_streams], [Cantidad de streams]) VAR Nombre = LOOKUPVALUE(Artista[nombre_artista], Artista[id_artista], Id_Artista) RETURN IF(ISBLANK(Nombre), "No se encontró resultado", Nombre) </pre>
Fecha de lanzamiento de la canción más popular	Esta medida obtiene la fecha de lanzamiento de la canción más popular	<p>Fecha de lanzamiento de la canción más popular =</p> <pre> VAR Cancion = [Nombre de la canción más popular] VAR Fecha = FORMAT(LOOKUPVALUE('Canción'[fecha_lanzamiento], 'Canción'[nombre_cancion], Cancion), "DD/MM/YYYY") RETURN Fecha </pre>
Cantidad de streams de la canción más popular	Esta medida obtiene la cantidad de reproducciones que tuvo la canción más popular.	<p>Cantidad de streams de la canción más popular = <code>MAX(Streams[total_streams])</code></p>
Género de la canción más popular	Esta medida obtiene el género musical de la canción más popular.	<p>Género de la canción más popular =</p> <pre> VAR Id_Artista = LOOKUPVALUE('Canción'[id_artista], Streams[total_streams], [Cantidad de streams de la canción más popular]) VAR Id_Genero = LOOKUPVALUE('Canción'[id_genero], Artista[id_artista], Id_Artista) VAR Genero = LOOKUPVALUE('Género musical'[nombre_genero], 'Género musical'[id_genero], Id_Genero) </pre>

		<pre> RETURN IF(ISBLANK(Genero), "No se encontró resultado", Genero) </pre>
Nombre de la canción que más veces entró al top 10	Esta medida obtiene el nombre de la canción que más veces entró en el top 10 de canciones globales.	<pre> Nombre de la canción que más veces entró al top 10 = VAR Id_Streams = LOOKUPVALUE(Streams[id_streams], Streams[top_10_x_times], MAX(Streams[top_10_x_times])) VAR Nombre_Cancion = LOOKUPVALUE('Canción'[nombre_cancion], 'Canción'[id_streams], Id_Streams) RETURN Nombre_Cancion </pre>
Géneros más populares		
Género musical más popular	Esta medida obtiene el nombre del género más popular, basándose en la suma de las reproducciones que obtuvieron cada una de las canciones pertenecientes al mismo.	<pre> Género musical más popular = VAR Genero = LOOKUPVALUE('Género musical'[nombre_genero], 'Suma de streams por género musical'[Streams totales], MAX('Suma de streams por género musical'[Streams totales])) RETURN IF(ISBLANK(Genero), "No se encontró resultado", Genero) </pre>
Streams totales del género musical	Esta medida obtiene la cantidad de reproducciones que obtuvo el género seleccionado en el filtro.	<pre> Streams totales del género musical = VAR SelectedGenre = SELECTEDVALUE('Género musical'[nombre_genero]) VAR StreamsGenre = LOOKUPVALUE('Suma de streams por género musical'[Streams totales], 'Género musical'[nombre_genero], SelectedGenre) RETURN StreamsGenre </pre>
Promedio de streams por género musical	Esta medida obtiene el promedio de las	Promedio de streams por género musical =

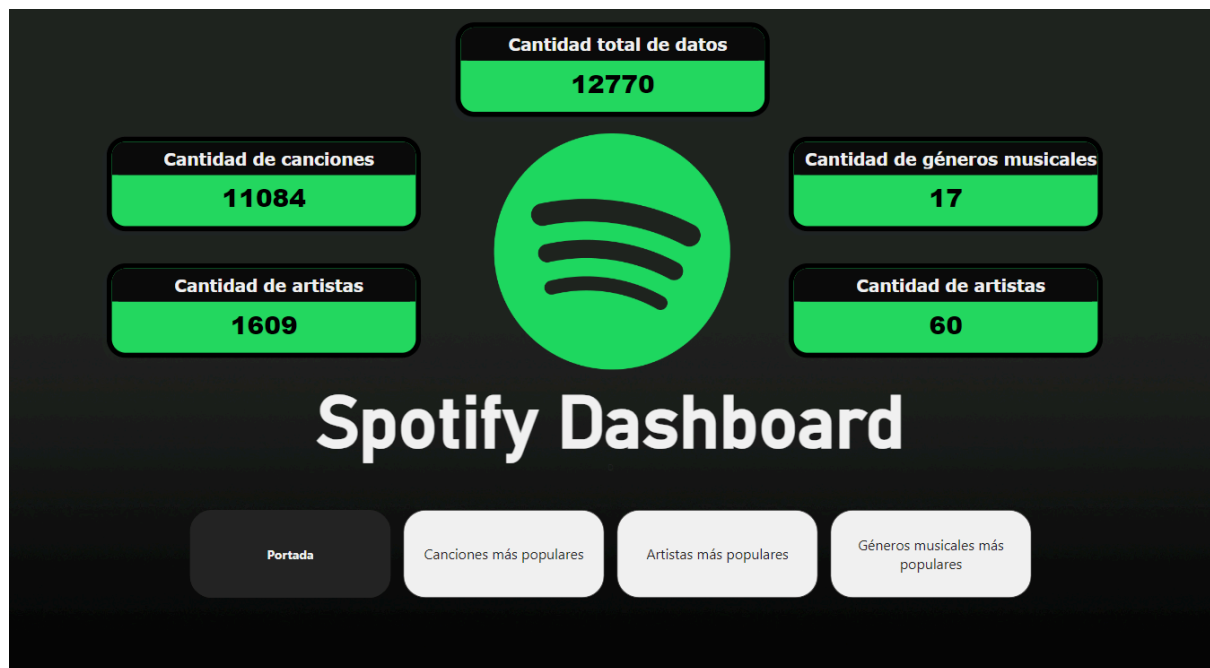
	reproducciones que obtuvo el género seleccionado en el filtro.	<pre> VAR SelectedGenre = SELECTEDVALUE('Género musical'[nombre_genero]) VAR Id_Genero = LOOKUPVALUE('Género musical'[id_genero], 'Género musical'[nombre_genero], SelectedGenre) VAR Promedio_Streams = CALCULATE(AVERAGE(Streams[total_streams]), 'Canción'[id_genero] = Id_Genero) RETURN Promedio_Streams </pre>
Datos analizados		
Cantidad de artistas del dataset	Esta medida calcula la cantidad total de artistas existentes en el dataset.	Cantidad de artistas del dataset = <code>COUNT(Artista[nombre_artista])</code>
Cantidad de canciones del dataset	Esta medida calcula la cantidad total de canciones existentes en el dataset.	Cantidad de canciones del dataset = <code>COUNT('Canción'[nombre_cancion])</code>
Cantidad de géneros musicales del dataset	Esta medida calcula la cantidad total de géneros musicales existentes en el dataset.	Cantidad de géneros musicales del dataset = <code>COUNT('Género musical'[nombre_genero])</code>
Cantidad de países del dataset	Esta medida calcula la cantidad total de países existentes en el dataset.	Cantidad de países del dataset = <code>COUNT('País artista'[nombre_pais])</code>
Cantidad total de datos	Esta medida calcula la cantidad total de datos que tiene el dataset.	Cantidad total de datos = <code>[Cantidad de artistas del dataset] + [Cantidad de canciones del dataset] + [Cantidad de géneros musicales del dataset] + [Cantidad de países]</code>
Tablas calculadas		
Suma de streams por artista	Esta realiza la suma de las reproducciones que obtuvo cada artista.	Suma de streams por artista = <code>SUMMARIZE(Artista, Artista[nombre_artista], 'País</code>

		<code>artista'[nombre_pais],"Streams totales", SUM(Streams[total_streams]))</code>
Suma de streams por género musical	Esta realiza la suma de las reproducciones que obtuvo cada género musical.	Suma de streams por género musical = <code>SUMMARIZE('Género musical', 'Género musical'[nombre_genero],"Streams totales", SUM(Streams[total_streams]))</code>

Portada del dashboard

Al inicio de nuestro tablero nos encontraremos con la portada que se muestra a continuación. La misma cuenta con un display de botones que le permiten al usuario navegar a través del dashboard. Además, se puede observar información adicional sobre los datos analizados.

Cabe destacar que cada página contiene 2 botones, los cuales le permiten al usuario ir a la próxima página o bien volver a la anterior.



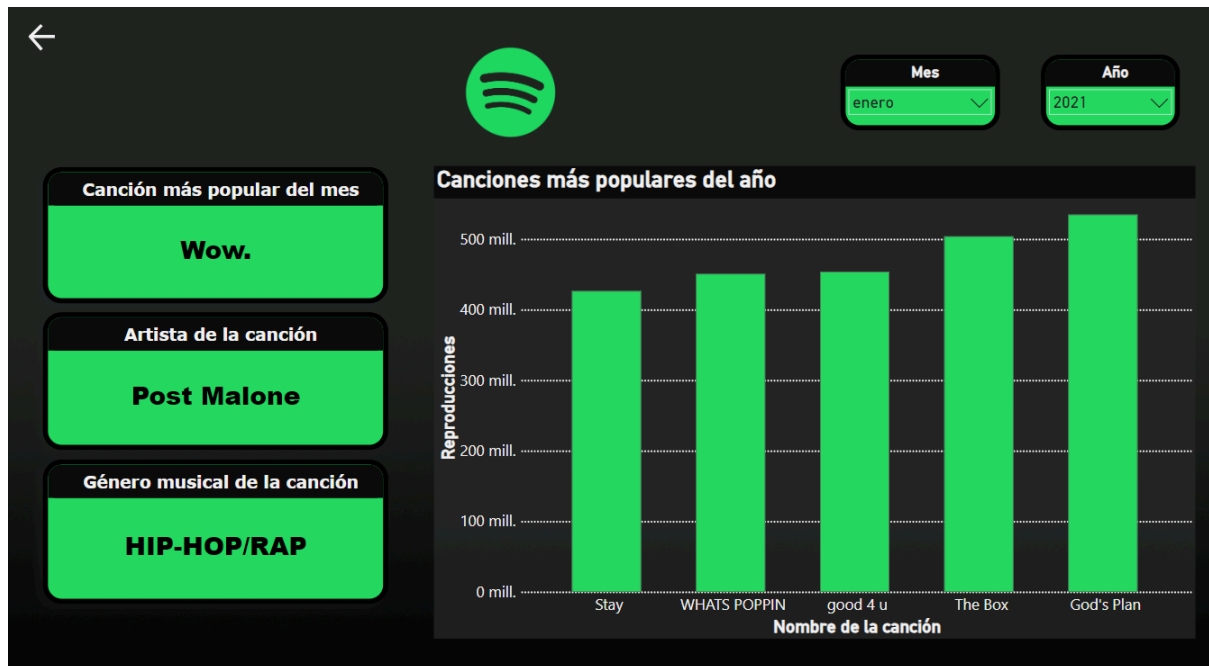
Página 1 - “Canciones más populares”

En esta sección del tablero, contamos con dos métricas de suma importancia; la canción más popular del dataset y la canción que más veces entró al top 10 de canciones globales. Además, podemos interactuar con tres botones que nos permitirán ver más información acerca de las canciones más populares del dataset.



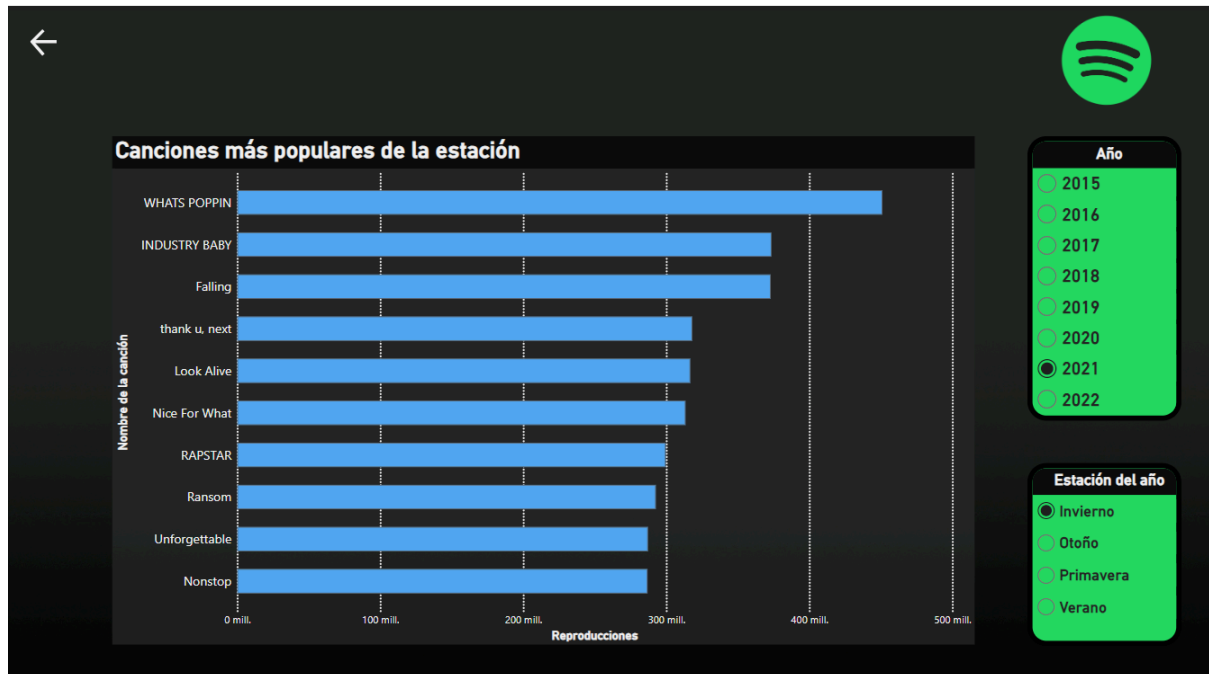
Página 1.1 - “Canciones más populares por mes y año”

Esta página del dashboard muestra la canción más popular del mes seleccionado junto con el artista y el género musical de dicha canción. En la parte superior derecha, se pueden ajustar los filtros “Mes” y “Año” para visualizar la información correspondiente al periodo deseado. Además, podemos observar un gráfico que nos muestra las cinco canciones más populares del año seleccionado. El mismo cuenta con una herramienta que permite ver más información acerca de las canciones.



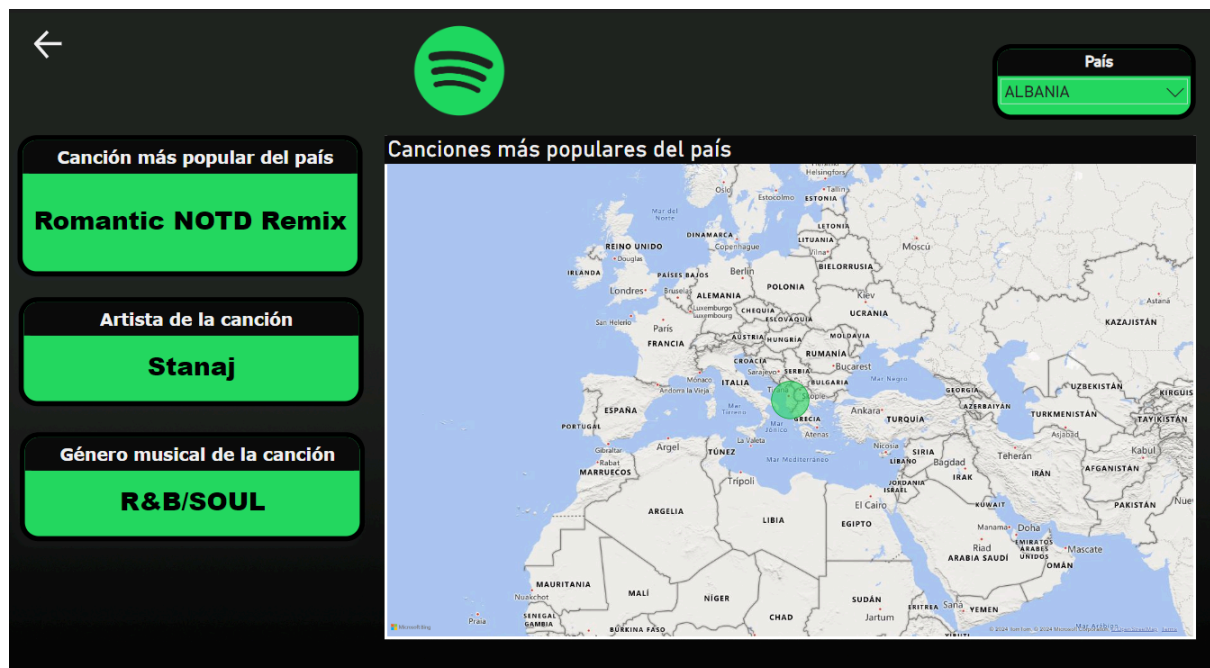
Página 1.2 - “Canciones más populares por estación del año”

Esta página del dashboard muestra un gráfico de barras que presenta las canciones más populares por estación del año, basándose en el número de reproducciones. Los filtros a la derecha permiten seleccionar el año y la estación del mismo para ajustar la visualización del gráfico.



Página 1.3 - “Canciones más populares por país”

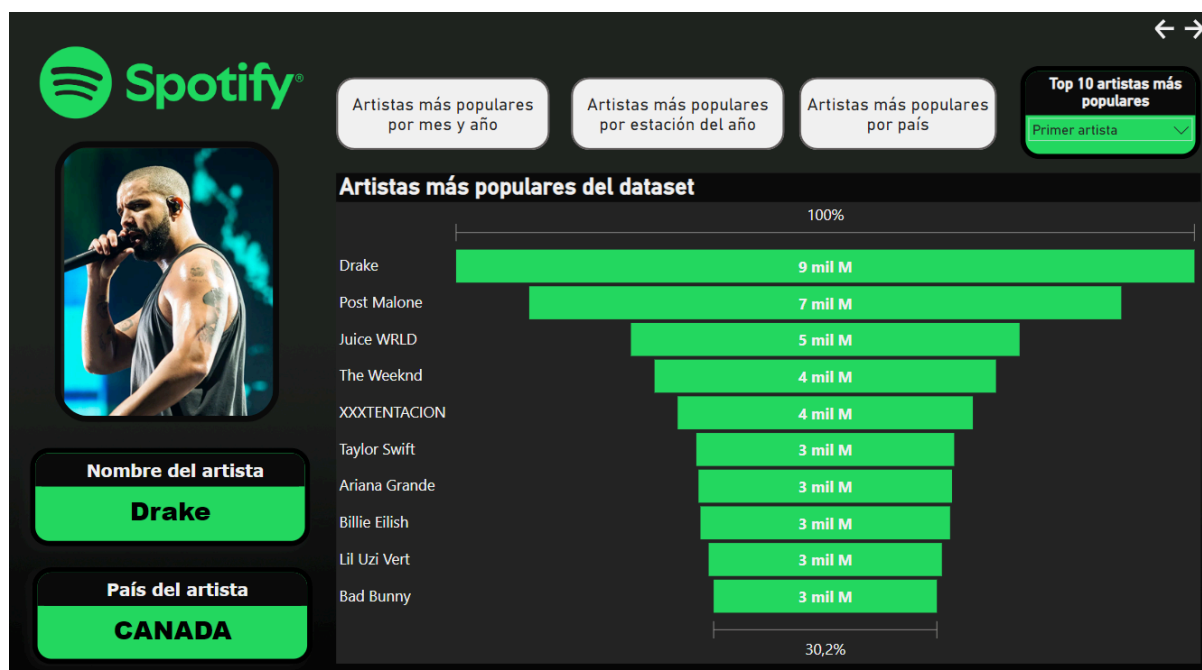
Esta página del dashboard muestra las canciones más populares del país. En el lado izquierdo, se presentan los detalles de la canción más popular del país seleccionado: el nombre de la canción, el artista y el género musical de la misma. A la derecha, se visualiza un mapa que destaca las cinco canciones más populares del país seleccionado a través de una herramienta. El filtro en la esquina superior derecha permite seleccionar el país,



Página 2 - “Artistas más populares”

Esta sección del dashboard nos permite ver un top diez de los artistas más populares del dataset mediante un gráfico. Este mismo cuenta con una herramienta que muestra la evolución de las reproducciones de cada artista a lo largo de los años.

En la esquina superior derecha podemos interactuar con un filtro, el cuál irá actualizando los valores de las métricas “Nombre del artista” y “País del artista”, además de la foto. Luego, a la izquierda de este filtro, podemos interactuar con tres botones diferentes que nos llevarán a otras páginas para conocer más información acerca de los artistas más populares.



Página 2.1 - “Artistas más populares por mes y año”

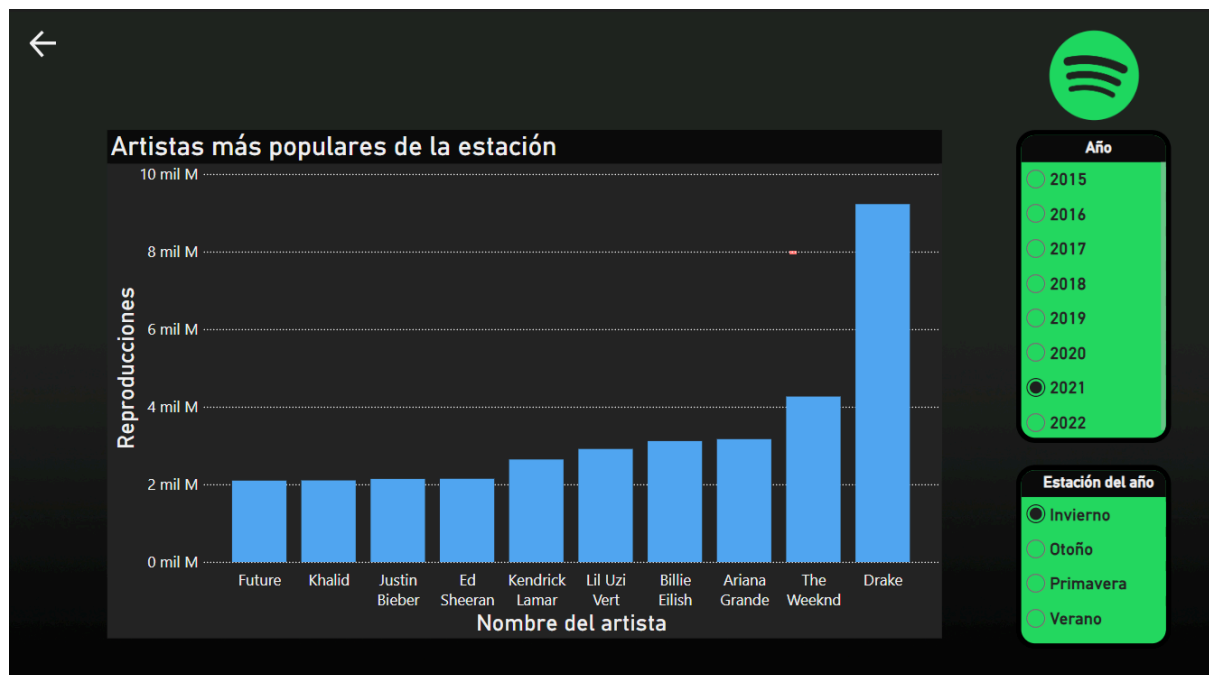
En esta sección del dashboard, a la derecha, podemos observar un gráfico que nos muestra los 5 artistas más populares del año seleccionado en el filtro “Año”. Mientras que a la izquierda podemos ver cuál fue el artista más popular del mes seleccionado en el filtro “Mes”, junto con su país y sexo.



Página 2.2 - “Artistas más populares por estación año”

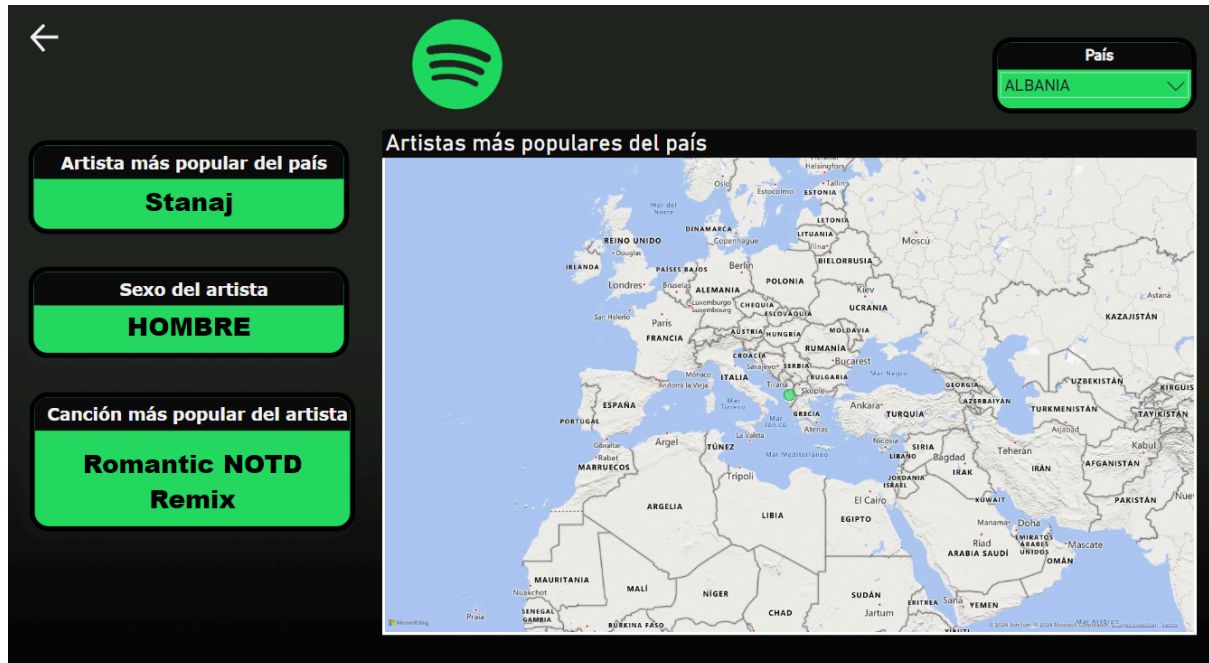
En esta página del dashboard encontraremos más información al respecto de los artistas más populares, teniendo en cuenta la estación de un año en particular.

A la izquierda, podemos ver un gráfico de barras que cuenta con una herramienta que permite visualizar la cantidad de reproducciones que obtuvo el artista en la estación seleccionada en el filtro. A la derecha, se encuentran los filtros que actualizarán el gráfico.



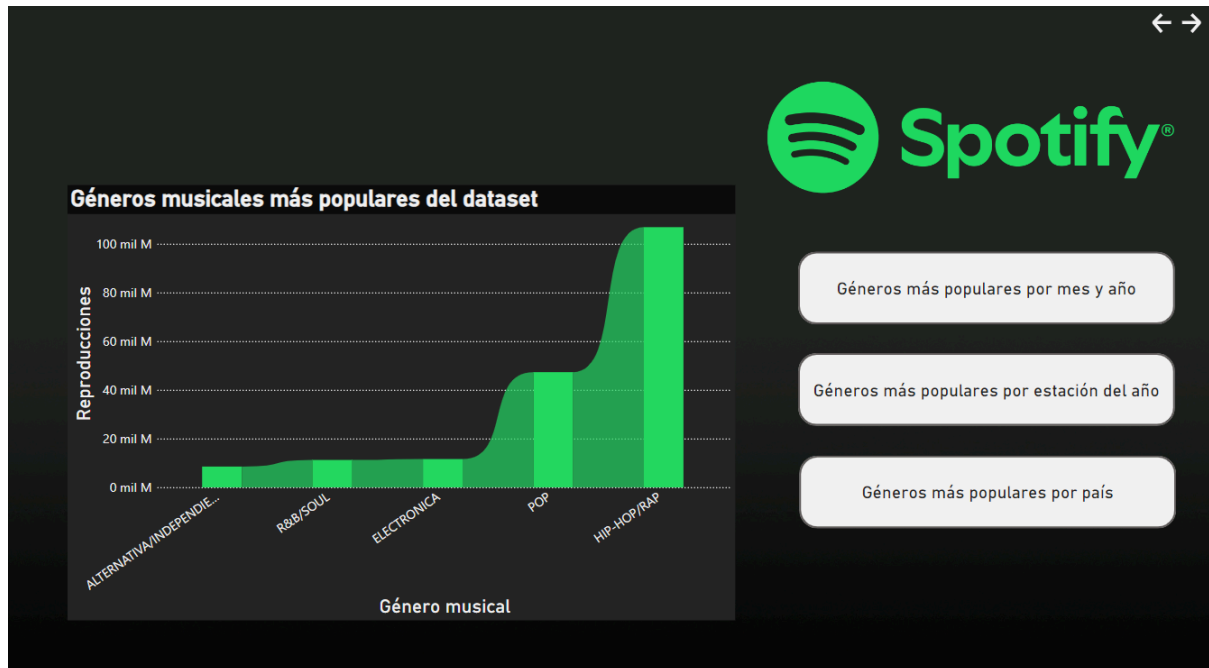
Página 2.3 - “Artistas más populares por país”

En la parte izquierda de la página de este dashboard, podemos encontrar información relativa al artista más popular del país seleccionado en el filtro “País”. Mientras que en la parte derecha, veremos un mapa el cuál cuenta con una herramienta que permite conocer los cinco artistas más populares del país.



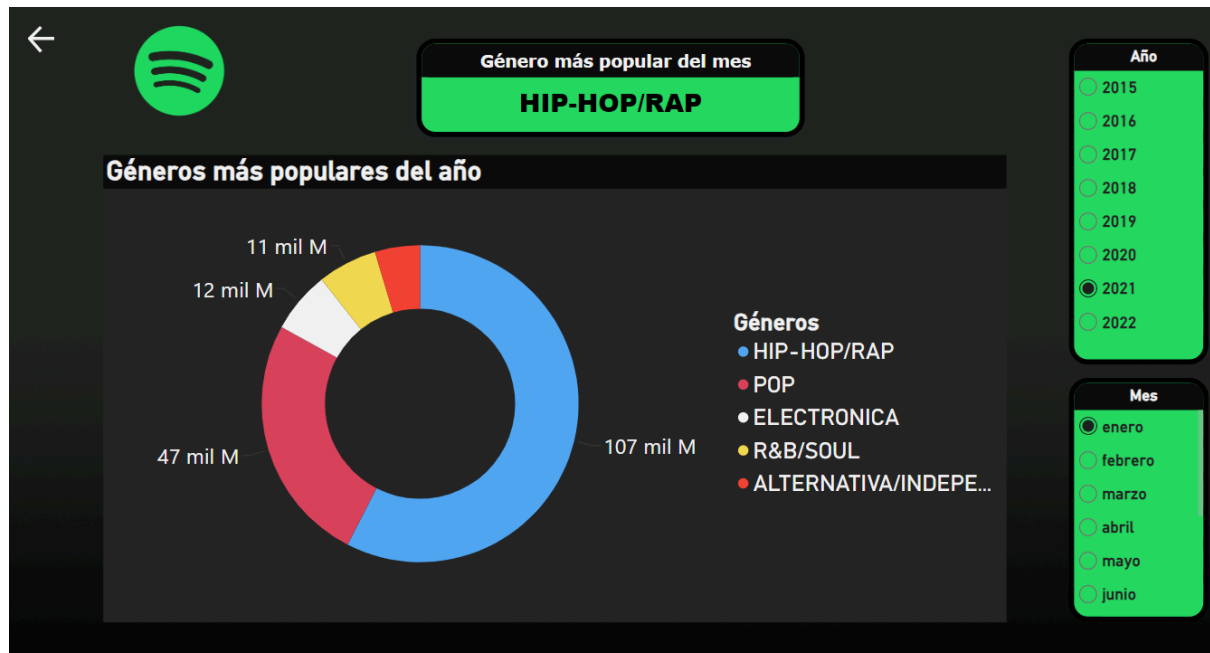
Página 3 - “Géneros musicales más populares”

En esta página del dashboard, nos encontraremos con cuatro elementos. Un gráfico, que permite conocer cuáles son los cinco géneros más populares del dataset y tres botones, los cuáles nos permiten saber más sobre los géneros populares.



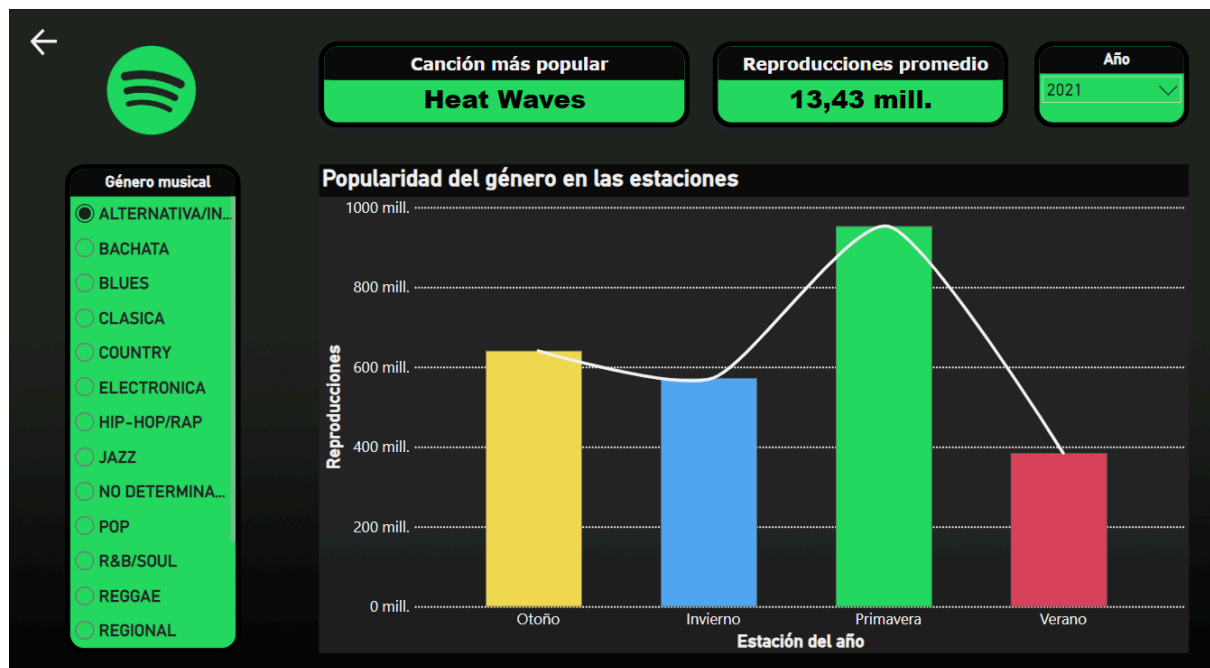
Página 3.1 - “Géneros musicales más populares por mes y año”

En esta página del dashboard, en la parte derecha, encontramos dos filtros: "Año" y "Mes". El filtro "Año" actualiza el gráfico central para mostrar información sobre los cinco géneros más populares de ese año. El filtro "Mes" actualiza la métrica "Género más popular del mes".



Página 3.2 - “Géneros musicales más populares por estación del año”

En la parte superior de esta sección del dashboard, hay tres elementos: dos métricas y un filtro. Las métricas muestran la canción más popular del género musical seleccionado y el promedio de reproducciones que obtuvo este género. A la izquierda se encuentra el filtro "Género musical". En la parte central, se presenta un gráfico que muestra la evolución de las reproducciones del género seleccionado a lo largo de las cuatro estaciones del año seleccionado.

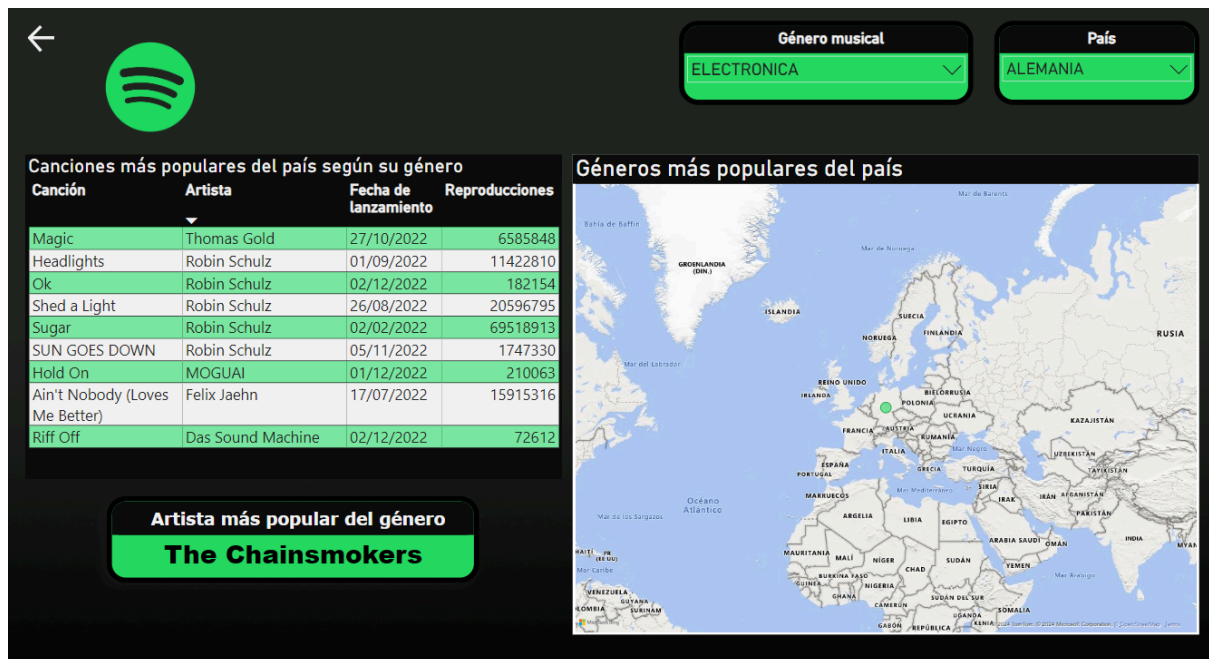


Página 3.3 - “Géneros musicales más populares por país”

En esta sección del dashboard encontraremos más información acerca de los géneros musicales teniendo en cuenta los países de cada artista.

En la parte izquierda podemos ver un top diez de las canciones más populares del país según el género seleccionado en el filtro “Género musical”. Además, contamos con una métrica que hace referencia al artista más popular del género seleccionado.

En la parte derecha, encontramos un mapa que cuenta con una herramienta que permite ver un top cinco de los géneros más populares del país seleccionado en el filtro “País”.



Conclusión

Tras realizar un análisis exhaustivo podemos realizar las siguientes afirmaciones sobre las canciones registradas en el dataset:

- La canción más popular del conjunto de datos y que más veces entró en el top diez de canciones globales es **Sunflower** del artista Post Malone.
- El artista más popular del dataset es **Drake** y cuenta con **9.228.928.893** de reproducciones en total.
- El género musical más popular del dataset es **HIP-HOP/RAP** y tiene más de cien mil millones de reproducciones.