

DATA ANALYTIC

Coderhouse – Comisión 32505

Profesora: Rebeca Figueroa

Tutor: Juan Pablo Acuña González

Tema: Top 200 semanal de las mejores canciones de Spotify entre
el 04/02/2021 y el 14/07/2022



Integrantes:

Carlos Mujica
Bruno Mangione

INDICE

| | |
|--|----|
| Introducción..... | 3 |
| Objetivo..... | 3 |
| Alcance..... | 3 |
| Data Set..... | 3 |
| Tablas..... | 3 |
| Gráfico Entidad Relación..... | 4 |
| Modelo Relacional..... | 6 |
| Paleta de Colores..... | 6 |
| Documentación de transformaciones..... | 7 |
| Tabla de versionados..... | 10 |
| Mock Up..... | 11 |
| Creación de Dashboard..... | 11 |
| Portada..... | 12 |
| Análisis Región País..... | 13 |
| Análisis Por Canción..... | 15 |
| Conclusión..... | 17 |

INTRODUCCION

El presente trabajo se realizó para cumplimentar el trabajo final del Curso Data Analytic de CoderHouse dictado por Rebeca Figueroa Soriano.

Los integrantes Carlos Mujica y Bruno Mangione llevaron a cabo la organización y análisis de datos bajo el tema “Top 200 semanal de las mejores canciones de Spotify entre el 04/02/2021 y el 14/07/2022”.

Se llevará a cabo un tablero de indicadores modelando los valores de las cantidades de streaming de canciones por países, regiones y a nivel mundial. Se verán reflejados cuales son los artistas y canciones más escuchados, la cantidad de reproducciones por canción, El género musical asignado por el algoritmo, como así también Idioma predominante en el país donde es escuchado el artista

Además se verán datos de diferentes características de las canciones, como que tan bailable es el tema, energía, Humor, si hubo collab o no, entre otras.

OBJETIVO

El objetivo es analizar el comportamiento de las tendencias musicales en el periodo especificado, ver como es la tendencia en Argentina y compararla a Nivel Regional y Global.

ALCANCE

El Dashboard es a Nivel Estratégico, ya que de acuerdo a las tendencias musicales de los diferentes países, se puede estudiar los gustos de las personas y de esa forma establecer metas y objetivos a largo plazo por parte de las productoras y artistas.

DATASET

Dejamos a continuación en enlace del Excel con las base de datos organizadas y estructuradas utilizadas en el presente trabajo:

<https://docs.google.com/spreadsheets/d/1eBGG8tly8fdPSUCKBFM1gumXPMm9Ex3E/edit?usp=sharing&ouid=103239331595329505531&rtpof=true&sd=true>

TABLAS

A continuación se muestran las tablas armadas al inicio del proyecto, de acuerdo al análisis inicial que se hizo de acuerdo a la base de datos.

Tablas

| Tracks | | |
|--------|----------------|--------------|
| PK | track_id | int |
| | track_name | varchar (50) |
| | artist_names | int |
| | genre | int |
| | weeks_on_chart | int |
| | streams | int |
| | release_date | date |
| | source | int |
| | country | int |
| | region | int |
| | language | int |

| Artists | | |
|---------|--------------|--------------|
| PK | track_name | int |
| | artist_id | int |
| | artist_names | varchar (50) |
| | genre | int |
| | country | int |
| | region | int |
| | language | int |

| Streams | | |
|---------|------------|-----|
| PK | streams_id | int |
| | tracks | int |
| | artist | int |
| | genre | int |

| weeks_on_chart | | |
|----------------|----------------|-----|
| PK | weeks_on_chart | int |
| | streams | int |
| | track_id | int |
| | artist_id | int |

| Region | | |
|--------|-----------|-----------|
| PK | region_id | int |
| | region | text (50) |
| | country | int |
| | language | int |
| | track | int |
| | artist | int |

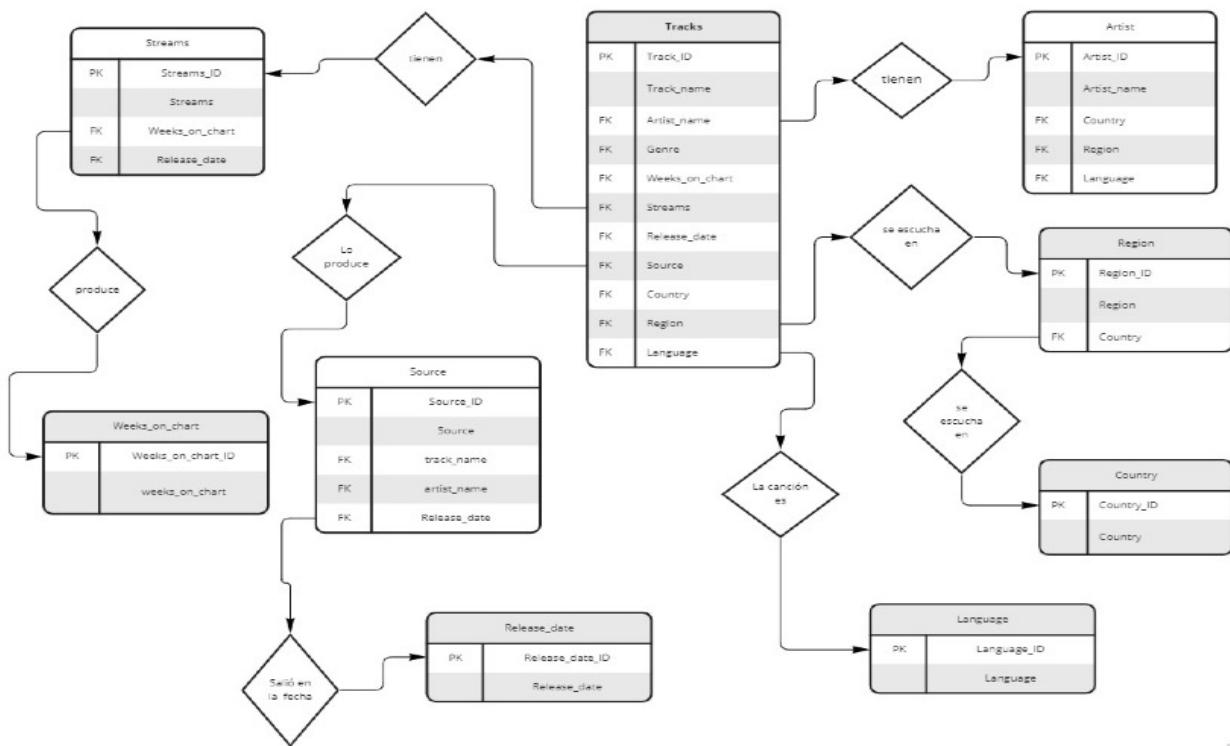
| source | | |
|--------|-----------|--------------|
| PK | source_id | int |
| | source | varchar (50) |
| | track | int |
| | artist | int |

| Country | | |
|---------|------------|-----------|
| PK | country_id | int |
| | country | text (50) |
| | region | int |
| | track | int |
| | artist | int |

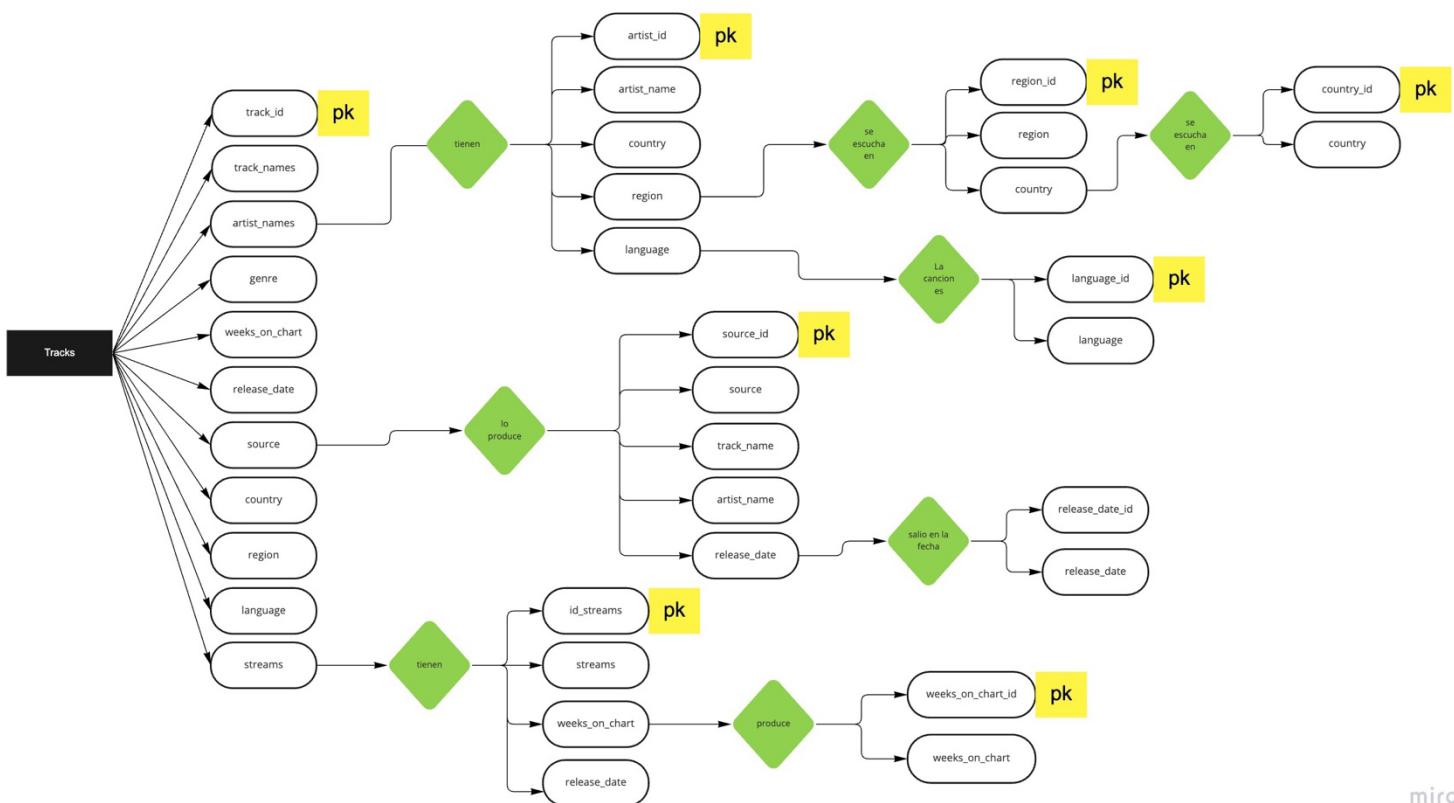
| release_date | | |
|--------------|--------------|------|
| PK | release_date | date |
| | streams | int |
| | track | int |
| | artist | int |

GRAFICO DE ENTIDAD RELACION

En los siguientes gráficos se muestran como en principio relacionamos las variables de la base de datos y como creímos en un principio la mejor forma de analizar la información e identificar las variables de las diferentes tablas.



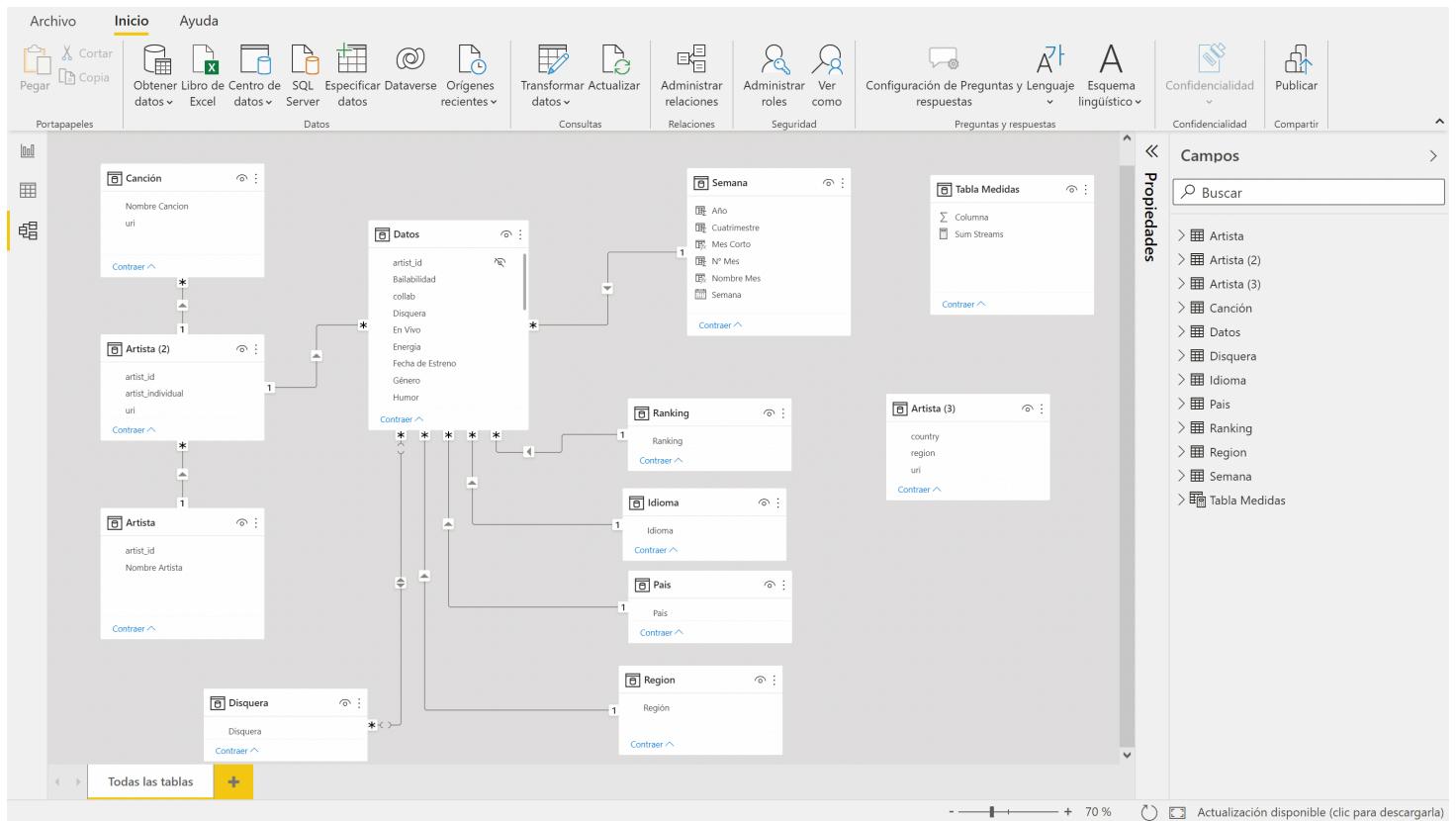
miro



miro

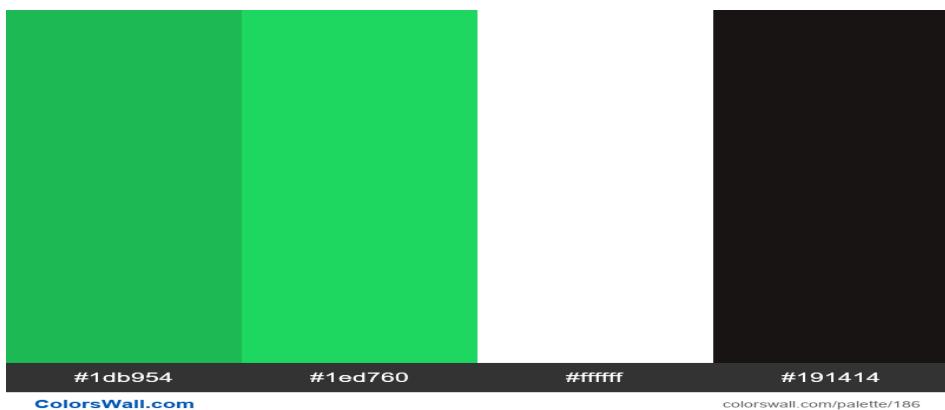
MODELO RELACIONAL

Luego de ir trabajando y analizando la base de datos y ver como podíamos relacionar los datos, el modelo relacional nos quedó de la siguiente manera.



PALETA DE COLORES

Buscamos la paleta de colores de Spotify para aplicar al trabajo. Los colores son los siguientes:
#1db954 #1ed760 #ffffff #191414



DOCUMENTACIÓN DE TRANSFORMACIONES

Se procede a cargar los datos en el power query para su normalización

Procedemos a normalizar los datos a eliminar las filas en blanco y las columnas que no nos sirven para realizar el análisis

URI (Uniform Resource Indicator) = seria una primary key

```
#"Columnas quitadas1" = Table.RemoveColumns(#"Valor reemplazado1", {"pivot", "duration", "tempo", "instrumentalness", "artist_img", "album_cover"}),
```

Se elimino la columna "album_cover" por no ser relevante para el análisis

Se elimino la columna "artist_img" por no ser relevante para el análisis

Se borró la columna tempo por que el tipo de datos era incompatible con power bi

Se eliminó la Columna de duración por que la transformación daba error al manejarla

Se eliminó la columna pivot por no ser relevante al proyecto de análisis de datos

Se eliminó la columna “instrumentalness” por que el tipo de datos era incompatible

Se normalizo la tabla previous_rank que tenía unos datos irregulares

- = Table.ReplaceValue(#"Columnas quitadas1", -1, 1, Replacer.ReplaceValue, {"previous_rank"})

Se reemplazó un valor que daba como resultado siempre un 0 a la izquierda que no correspondía en ranking previo

- = Table.ReplaceValue(#"Valor reemplazado", ".0", "", Replacer.ReplaceText, {"album_num_tracks"})

Se reemplazo el .0 que no correspondía por un espacio en blanco para poder interpretar la información correctamente en la columna “numero de canciones por album” también se hizo lo mismo en la columna numero de artistas

```
#"Valor reemplazado" = Table.ReplaceValue(#"Columnas quitadas", ".0", "", Replacer.ReplaceText, {"artists_num"}),
```

Se elimino la columna mode

Se elimino la columna Key

Se elimino la columna loudness

Se elimino la columna **acousticness**

Luego se procedió a cambiarle el nombre en inglés a las columnas por su equivalente al español para facilitar la manipulación y el análisis de las mismas

```

#"Columnas con nombre cambiado" = Table.RenameColumns(#"Columnas quitadas2", {"danceability", "Bailabilidad"}, {"energy", "Energia"}, {"speechiness", "Presencia de Palabras"}),

#"Columnas con nombre cambiado1" = Table.RenameColumns(#"Columnas quitadas3", {"liveness", "En Vivo"}, {"valence", "Humor"}, {"country", "Pais"}, {"region", "Región"}, {"language", "Idioma"}, {"week", "Semana"}, {"weeks_on_chart", "Semanas en el Top"}, {"previous_rank", "Ranking Previo"}, {"peak_rank", "Ranking Tope"}, {"source", "Disquera"}, {"album_num_tracks", "Numero de Canciones por Album"}, {"release_date", "Fecha de Estreno"}, {"track_name", "Nombre Cancion"}, {"artist_genre", "Género"}, {"artist_individual", "Nombre Artista"}, {"artists_num", "Num.de Artista"}, {"artist_names", "Nombre del Artista"}, {"rank", "Ranking"}}

let

    Origen = Excel.Workbook(File.Contents("C:\Users\Usuario\Downloads\dataset.xlsx"), null, true),
    Sheet1_Sheet = Origen{[Item="Sheet1", Kind="Sheet"]}[Data],
    #"Encabezados promovidos" = Table.PromoteHeaders(Sheet1_Sheet, [PromoteAllScalars=true]),
    #"Columnas quitadas" = Table.RemoveColumns(#"Encabezados promovidos", {"Column1"}),

    #"Valor reemplazado" = Table.ReplaceValue(#"Columnas quitadas", ".0", "", Replacer.ReplaceText, {"artists_num"}),

    #"Valor reemplazado1" = Table.ReplaceValue(#"Valor reemplazado", ".0", "", Replacer.ReplaceText, {"album_num_tracks"}),

    #"Columnas quitadas1" = Table.RemoveColumns(#"Valor reemplazado1", {"pivot", "duration", "tempo", "instrumentalness", "artist_img", "album_cover"}),

    #"Valor reemplazado2" = Table.ReplaceValue(#"Columnas quitadas1", -1, 1, Replacer.ReplaceValue, {"previous_rank"}),

    #"Columnas quitadas2" = Table.RemoveColumns(#"Valor reemplazado2", {"mode", "key", "loudness"}),

    #"Columnas con nombre cambiado" = Table.RenameColumns(#"Columnas quitadas2", {"danceability", "Bailabilidad"}, {"energy", "Energia"}, {"speechiness", "Presencia de Palabras"}),

    #"Columnas quitadas3" = Table.RemoveColumns(#"Columnas con nombre cambiado", {"acousticness"}),

    #"Columnas con nombre cambiado1" = Table.RenameColumns(#"Columnas quitadas3", {"liveness", "En Vivo"}, {"valence", "Humor"}, {"country", "Pais"}, {"region", "Región"}, {"language", "Idioma"}, {"week", "Semana"}, {"weeks_on_chart", "Semanas en el Top"}, {"previous_rank", "Ranking Previo"}, {"peak_rank", "Ranking Tope"}, {"source", "Disquera"}, {"album_num_tracks", "Numero de Canciones por Album"}, {"release_date", "Fecha de Estreno"}, {"track_name", "Nombre Cancion"}, {"artist_genre", "Género"}, {"artist_individual", "Nombre Artista"}, {"artists_num", "Num.de Artista"}, {"artist_names", "Nombre del Artista"}, {"rank", "Ranking"}}

in

    #"Columnas con nombre cambiado1"

```

Se Crea una Tabla Calendario

Calendario = CALENDAR(FIRSTDATE(Datos[Semana]), LASTDATE(Datos[Semana]))

Se crea una tabla que contenga las fechas únicas del top

#"Duplicados quitados" = Table.Distinct(#"Columnas quitadas4"),

```
#"Filas filtradas" = Table.SelectRows(#"Duplicados quitados", each true)
```

Se crea una base de datos de Regiones y países para optimizar el procesamiento de los datos en power bi eliminando los duplicados de ambas

REGION

```
#"Otras columnas quitadas" = Table.SelectColumns(#"Columnas con nombre cambiado1", {"Región"}),
```

```
#"Duplicados quitados" = Table.Distinct(#"Otras columnas quitadas")
```

PAIS

```
#"Otras columnas quitadas" = Table.SelectColumns(#"Columnas con nombre cambiado1", {"Pais"}),
```

```
#"Duplicados quitados" = Table.Distinct(#"Otras columnas quitadas")
```

Se crea la Tabla Ranking

```
#"Otras columnas quitadas" = Table.SelectColumns(#"Tipo cambiado", {"Ranking"}).
```

```
#"Duplicados quitados" = Table.Distinct(#"Otras columnas quitadas").
```

```
#"Filas ordenadas" = Table.Sort(#"Duplicados quitados", {"Ranking", Order.Ascending}).
```

```
#"Filas inferiores quitadas" = Table.RemoveLastN(#"Filas ordenadas",1)
```

Se crea la Tabla Artista

```
##"Otras columnas quitadas" = Table.SelectColumns(#"Tipo cambiado", {"Nombre del Artista"})
```

```
#"Duplicados quitados" = Table.Distinct(#"Otras columnas quitadas")
```

Se crea la tabla Canción

```
#"Otras columnas quitadas" = Table.SelectColumns(#"Tipo cambiado", {"Nombre Cancion"}),  
    #"Tipo cambiado1" = Table.TransformColumnTypes(#"Otras columnas quitadas", {"Nombre Cancion", type text}),  
    #"Errores quitados" = Table.RemoveRowsWithErrors(#"Tipo cambiado1", {"Nombre Cancion"}),  
    #"Duplicados quitados" = Table.Distinct(#"Errores quitados")
```

Se crea la Tabla Idioma

```
#"Otras columnas quitadas" = Table.SelectColumns(#"Tipo cambiado", {"Idioma"}),  
    #"Tipo cambiado1" = Table.TransformColumnTypes(#"Otras columnas quitadas", {"Idioma", type text}),  
    #"Duplicados quitados" = Table.Distinct(#"Tipo cambiado1")
```

Se Crea la tabla disquera

```
#"Otras columnas quitadas" = Table.SelectColumns(#"Tipo cambiado", {"Disquera"}),  
    #"Duplicados quitados" = Table.Distinct(#"Otras columnas quitadas")
```

in

```
    #"Duplicados quitados"
```

Se crea una tabla calendario basandonos en las fechas obtenidas por las semanas construyendo 4 columnas, (año, mes, n° de mes, cuatrimestre)

Primero creamos tabla calendario para segmentar información y luego eliminamos la tabla, ya que nos dimos cuenta de cómo trabajar la información con los datos de fechas (semanas) que teníamos originalmente.

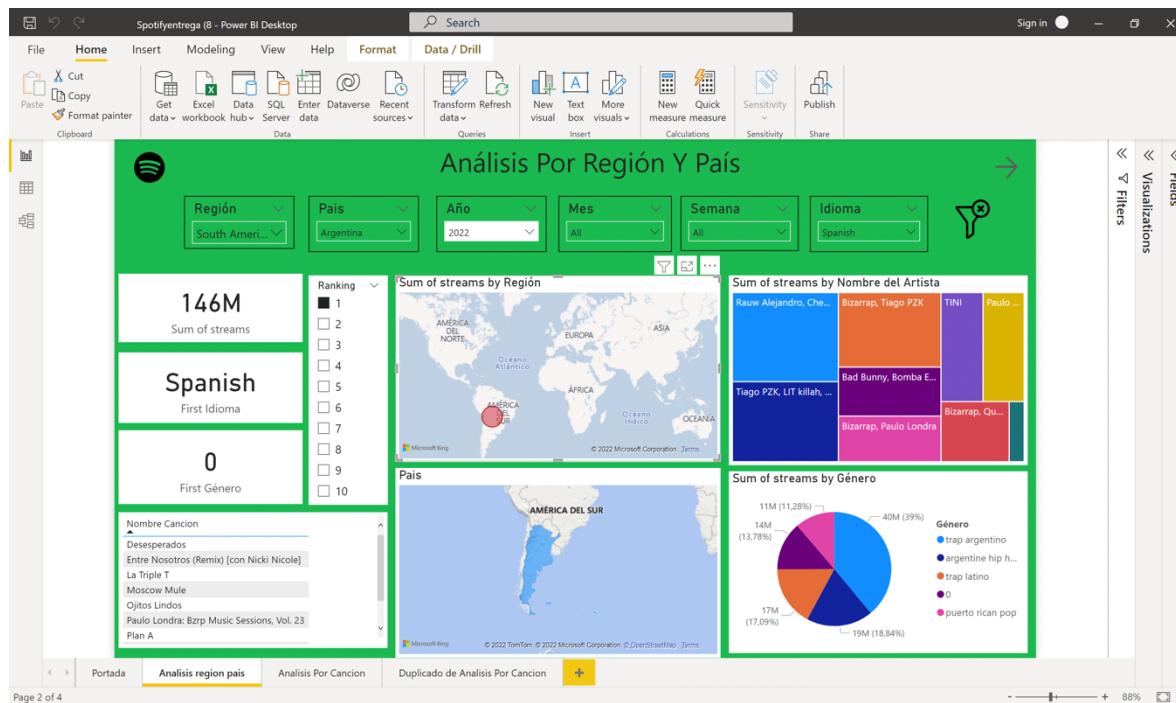
TABLAS DE VERSIONADOS

A continuación detallamos las versiones y fechas de las diferentes etapas del proyecto:

V 0.1 fecha 2/11/22
V 0.2 fecha 7/11/22
V 0.3 fecha 14/11/22
V 0.4 fecha 18/11/22
V 0.5 fecha 20/11/22

MOCK UP

A continuación mostramos la primer versión del tablero principal:



CREACIÓN DE DASHBOARD

Fecha 19/11/22

Filtros: Se podrá seleccionar y visualizar los siguientes elementos:

- Región
- País
- Idioma
- Año
- Mes
- Semana
- Ranking

Creación de gráficos y tarjetas: se crean las tarjetas y gráficos específicos para poder realizar los análisis de acuerdo a la funcionalidad que se le quiera dar a los mismos.

Botones: Se crean los botones a fines de darle navegabilidad al tablero.

Marcadores: Se crean los marcadores a fines de darle navegabilidad al tablero.

PORtADA

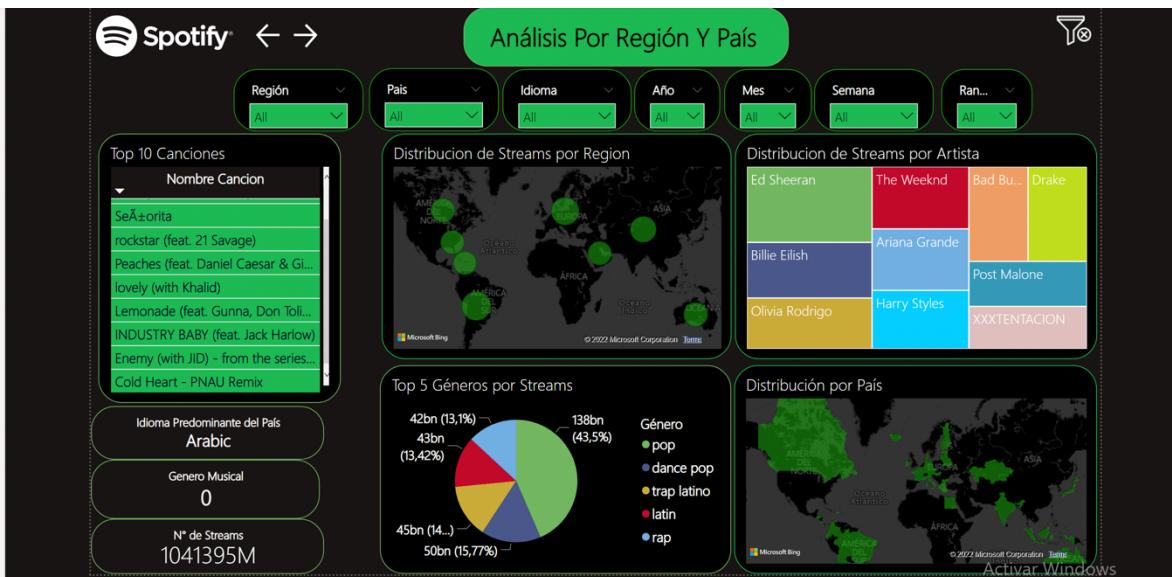
Al ingresar al tablero se vera la siguiente portada

The screenshot shows a dark-themed dashboard with a navigation bar at the top. The navigation bar has three tabs: 'Portada' (selected), 'Analisis region país', and 'Analisis Por Canción'. The main title 'Data Analytics' is displayed prominently in large white font. Below the title, the text 'Coderhouse – Comisión 32505' is shown. Underneath that, 'Profesora: Rebeca Figueroa' and 'Tutor: Juan Pablo Acuña González' are listed. A descriptive text block states: 'Tema: Top 200 semanal de las mejores canciones de Spotify entre el 04/02/2021 y el 14/07/2022'. Below this text, there is a section labeled 'Integrantes:' followed by the names 'Carlos Mujica' and 'Bruno Mangione'. At the bottom center of the dashboard is the Spotify logo.

Donde en las pestañas de arriba se puede acceder a las diferentes páginas del tablero presionando la tecla Ctrl + "click" en la pestaña.

ANALISIS REGION PAIS

Al ingresar en la primer pestaña vera lo siguiente:



A continuación explicamos los diferentes botones y visuales de la primer pestaña:



Desde el logo de Spotify se podrá volver a la portada inicial y con las flechas se puede navegar entre las distintas pestañas del tablero.



Con estos filtros se puede seleccionar las regiones, países, idiomas, años, meses, semanas y ranking que tuvieron las diferentes canciones. El objetivo de estos filtros es poder visualizar que canciones estuvieron en el top 200 semanal de Spotify en las diferentes regiones y países, como así también en un mes determinado del año. Esto también nos permite ver por regiones y países, que idiomas musicales predominan, cuáles son los géneros musicales y las cantidades de stream que tuvieron las canciones.

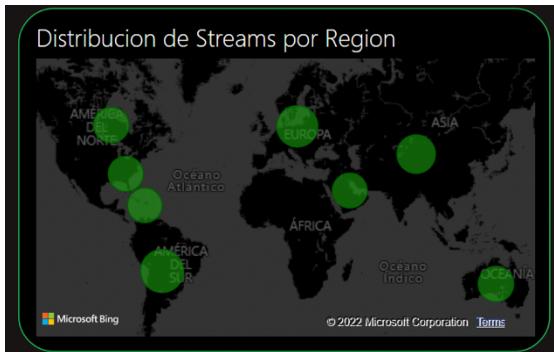


En estas visualizaciones, una vez seleccionados los filtros que queremos analizar, nos brinda cual es el top 10 de canciones que mayor cantidad de stream que tuvieron.

Si aquí seleccionamos algunas de las 10 canciones, en las visualizaciones debajo, nos dice que idioma predomina en el país, cual es el género musical de la canción seleccionada y cuál fue la cantidad de streams que tuvo la canción (de acuerdo al periodo de tiempo seleccionado).



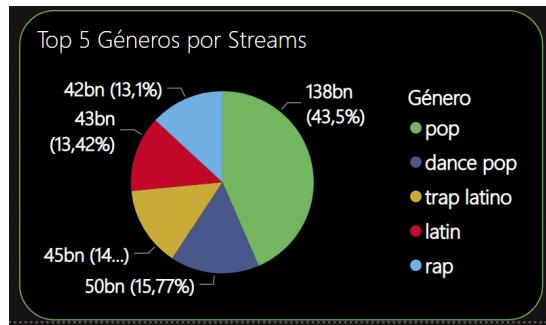
En esta visualización podemos ver la cantidad de stream que tuvo una canción determinada en una región determinada.



Aquí podemos ver los stream de canciones por artistas. Seleccionando un artista, podemos ver su top 10 de canciones, en qué región y país se escucha más, cual es el género musical que predomina y que tiene cada canción.



Esta visual, gráfico de cascada nos permite ver el top 10 de artistas por país y por cantidad de stream de las canciones de cada uno de ellos.



Este grafico en formato de circulo nos permite ver el top 5 de géneros musicales por cantidad de stream de canciones.



Este es un marcador para borrar todos filtros seleccionados.

ANALISIS POR CANCION

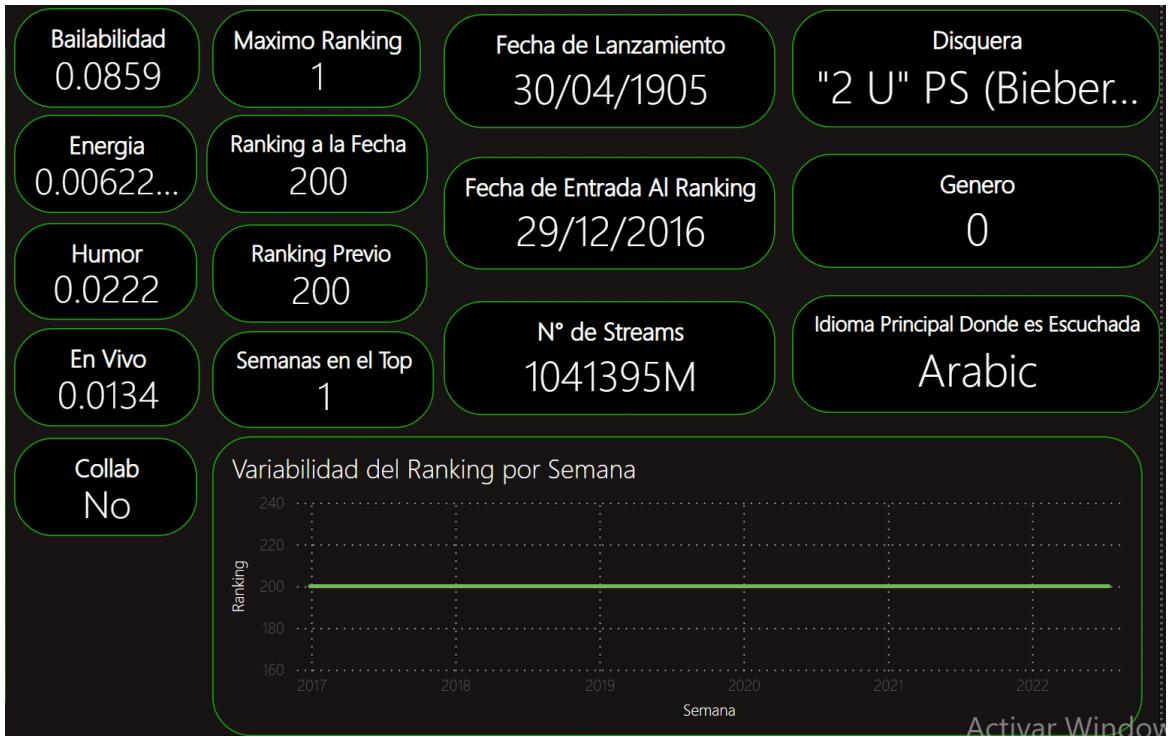
Al ingresar en la segunda pestaña vera lo siguiente:

The screenshot shows the 'Análisis Por Canción' (Analysis by Song) tab. At the top, there are six filter dropdowns: Región (All), País (All), Idioma (All), Año (All), Mes (All), and Semana (All). To the right of these are two buttons: 'Ran...' and a trash icon. Below the filters is a table with columns 'Nombre Cancion' and 'Nombre Artista'. The table lists several songs with their artists, such as 'China' by Anuel AA, 'Cold Heart - PNAU Remix' by Elton John, and 'happier' by Ed Sheeran. To the right of the table are several summary statistics in boxes: Bajabilidad (0.0859), Máximo Ranking (1), Fecha de Lanzamiento (30/04/1905), Disquera ("2 U" PS (Bieber...)), País (All), Idioma (All), Género (0), and N° de Streams (1041395M). Below these is a chart titled 'Variabilidad del Ranking por Semana' showing a horizontal line at a ranking of 200 from 2017 to 2022. At the bottom right is a 'Activar Windows' button.

Al igual que en la primer pestaña del tablero, en la parte superior, nos encontramos con el logo de Spotify, las flechas para navegar entre pestañas, los diferentes filtros que queremos realizar para hacer nuestro análisis como así también el marcador para eliminar los filtros seleccionados.

The screenshot shows the 'Análisis Por Canción' tab. At the top, there are six filter dropdowns: Región (All), País (All), Idioma (All), Año (All), Mes (All), and Semana (All). To the right of these are two buttons: 'Ran...' and a trash icon. Below the filters is a table with columns 'Nombre Cancion' and 'Nombre Artista'. The table lists several songs with their artists, such as 'China' by Anuel AA, 'Cold Heart - PNAU Remix' by Elton John, and 'happier' by Ed Sheeran. The interface is identical to the one shown in the previous screenshot, with the same filters and visualizations.

Como el objetivo de esta pestaña es analizar que produce una canción, al seleccionar los filtros nos va a traer un listado de canciones junto con el artista que la canta. Aquí podemos seleccionar una canción determinada y las visuales de la derecha nos traerá toda la información de la canción seleccionada.



En estas visualizaciones vamos a analizar cómo es cada canción:

Bailabilidad: nos va decir que tan bailable es el tema en una calificación de 0 a 1, siendo 0 una canción nada bailable y siendo 1 una canción totalmente bailable.

Energía: nos va a decir cuanta energía tiene o transmite la canción, en una calificación de 0 a 1, siendo 0 una canción sin energía y 1 una canción con la máxima energía.

Humor: nos va a decir que tan humorística es la canción, en una calificación de 0 a 1, siendo 0 una canción sin humor y 1 una canción totalmente humorística.

En Vivo: nos va a decir si la canción tiene partes en vivo, en una calificación de 0 a 1, siendo 0 una canción si parte en vivo y 1 una canción totalmente en vivo.

Collab: nos dice si en la canción hubo alguna colaboración con otro artista (feat), si hubo nos va a decir Si, de lo contrario será No.

Máximo Ranking: nos dice cuál fue el máximo ranking que tuvo la canción.

Ranking a la fecha: a la fecha determinada, nos dice en qué ranking se encuentra.

Ranking Previo: nos dice en qué ranking estuvo previamente del ranking a la fecha determinado.

Semanas en el top: nos dice cuantas semanas estuvo en el top 200 semanal.

Fecha de lanzamiento: nos dice en qué fecha fue lanzada la canción.

Fecha de entrada al ranking: nos dice en qué fecha entro al ranking del top 200 semanal.

N de Stream: nos dice la cantidad de reproducciones que tuvo esa canción.

Disquera: nos dice cuál es la productora que sacó el disco donde se encuentra la canción del artista.

Género: nos dice el género de la canción.

Idioma principal donde es escuchada: nos dice en qué idioma se escucha mayormente la canción.

Variabilidad del Ranking por semana: el gráfico nos muestra cómo varió la canción dentro del ranking del top 200 semanal.

CONCLUSION

Al iniciar el proyecto y durante el trabajo del mismo, nos encontramos con una base de datos que es mayormente cualitativa y no tan cuantitativa. Esto nos llevó a que no podíamos realizar demasiadas medidas y cálculos numéricos. A pesar de ello, pudimos analizar la información de la base de datos y ver las diferentes perspectivas musicales que existen en Argentina con el resto de la región y a nivel mundial.

Algunos puntos que analizamos fueron los siguientes:

En Argentina en el período analizado, se produjeron un total de 16.133 Millones de reproducciones de canciones, donde la tendencia musical es el trap latino y trap argentino, que abarcan más del 50% de los géneros musicales. Mientras que a nivel mundial, se produjeron 1.041.395 billones de reproducciones de canciones y el género más escuchado es el pop con un 43,5%.

En Argentina la bailabilidad de las canciones son del 0.278 en una calificación de 0 a 1, mientras que a nivel mundial, la bailabilidad es del 0.0859.