

## **Primera entrega proyecto de Ingeniería de Datos (Corregida)**

### **Realizado por:**

-Juan Diego Sosa Barreto

-Juan José Gómez Orozco

**Base de datos seleccionada:** *Spotify and youtube* (Link de acceso: [Spotify and Youtube | Kaggle](#) ).

**Repositorio:** [GitHub - TDarkwing/Datos-canciones-youtube-spotify: Proyecto de base de datos hecho por Juan Jose Gomez y Juan Diego Sosa](#)

### **Problema de negocios:**

Dado a que en un video musical se gasta mucho dinero en inversión (incluso dos veces más que producir la canción), se necesita saber en qué plataforma sale más conveniente subir canciones para su distribución. Ya que si sale más conveniente en Spotify al tratarse de una plataforma exclusiva de audio podemos obviar la realización del video musical y por lo mismo ahorrarse el gasto de la producción del video, por el contrario, si sale más conveniente publicarla en YouTube se puede decir que el gasto en videos musicales estaría justificado y que este genera una buena ganancia.

### **Reglas del negocio:**

- El ID del artista será un valor único y no nulo.
- El nombre del artista será un valor único y no nulo.
- El ID del álbum será un valor único y no nulo.
- El nombre del álbum será un valor no nulo.
- El tipo de álbum será un valor no nulo y solo se podrán poner los tipos “álbum” y “single”.
- El ID de la canción será un valor único y no nulo.
- El nombre de la canción será un valor no nulo.
- El tempo de una canción será un valor no nulo.
- La duración de una canción será no nula.
- El número de palabras de una canción será no nulo (en caso de no tener palabras se pondrá el número 0).
- El balance de una canción será un valor no nulo.
- La energía de una canción será un valor no nulo.
- El ruido de una canción será un valor no nulo.
- La llave de una canción será un valor no nulo.
- La acústica de una canción será un valor no nulo.
- La instrumental de una canción será un valor no nulo.

- Un artista puede crear varios álbumes.
- Un álbum solo puede ser creado por un artista.
- Un artista puede crear varias canciones.
- Una canción puede ser creada por varios artistas (Como en el caso de las colaboraciones).
- A cada artista le corresponde uno o varios links de página de Spotify.
- A cada artista le puede corresponder uno, varios o ningún canal en YouTube.
- A cada canción le corresponde un grupo de datos en Spotify.
- A cada canción le corresponde un grupo de datos en YouTube.
- La url de la canción en Spotify es única y no nula.
- El stream de una canción en Spotify será un valor no nulo.
- La url de la canción en YouTube será un valor no nulo y único.
- El título de una canción en YouTube será un valor no nulo.
- La descripción de una canción en YouTube será un valor no nulo.
- El número de vistas de una canción en YouTube será un valor no nulo (en caso de no tener será 0).
- El número de likes de una canción en YouTube será un valor no nulo (en caso de no tener será 0).
- El número de comentarios de una canción en YouTube será un valor no nulo (en caso de no tener será 0).

### **Entidades y atributos:**

- Artista: Entidad que guardara ciertos datos de un artista en particular.
  - Id: Entero único que nos servirá para identificar a un artista en particular.
  - Nombre: Nombre del artista.
  - Canal\_Youtube: Canal de YouTube oficial del artista. (En caso de no tenerlo se deja null).
  - Url\_Spotify: Url de la página oficial del artista en Spotify.
- Álbum: Entidad que nos identificará un álbum de un artista en particular.
  - Nombre\_album: Nombre del álbum.
  - Tipo\_Album: El tipo de álbum en específico (Regular, conceptual o recopilación).
- Canción: Entidad que nos identificará los datos de una canción en particular.
  - Nombre: Nombre de la canción.
  - Duración: Tiempo de duración de la canción en segundos.
  - Tempo: Velocidad rítmica de la canción medida en beats por minuto.

- **Danzable:** Mide encuestas del público y da un valor numérico de que tan bailable es una canción.
- **Energía:** Mide encuestas del público y da un valor numérico de que tan energética es una canción.
- **Llave:** Tonalidad de la canción dando una equivalencia numérica a cada tonalidad posible.
- **Ruido:** Mide encuestas del público y da un valor numérico de que tan ruidosa es una canción.
- **N\_palabras:** Mide el número de palabras dichas durante una canción.
- **Acústica:** Medida de confianza de 0 a 1 que determina si una pista es acústica.
- **Instrumental:** Predice si una canción no tiene voz. (Las siglas “Ohh” o “Ahh” no son contadas en este apartado.)
- **Audiencia:** Mide que tan probable es que la canción haya sido grabada en vivo y con audiencia.
- **Balance:** Medida del 0 al 1 que detecta que tan animada puede ser una canción. (Entre más cerca de 1 este se dice que la canción es más feliz o eufórica, mientras que entre más cerca de 0 este se dice que la canción es más triste o melancólica.)

-Datos\_S: Entidad que nos permite ver los datos en Spotify de una canción.

- **Url:** Url oficial en Spotify de la canción.
- **Stream:** Número de reproducciones mensuales de la canción.

-Datos\_Y: Entidad que nos permite ver los datos en YouTube de una canción.

- **Url:** Url oficial en YouTube de la canción.
- **Título:** Nombre de la canción dentro de la plataforma en YouTube.
- **Vistas:** Número de reproducciones totales de la canción en YouTube.
- **N\_likes:** Número de “me gusta” en una canción en YouTube.
- **N\_comentarios:** Número de comentarios de una canción en YouTube.
- **Descripción:** La descripción oficial de la canción en YouTube.
- **Licencia:** Valor booleano que nos dice si es verdadero que la canción tiene permiso de ser publicada en YouTube.
- **Video\_oficial:** Valor booleano que nos dice si es verdadero que estamos viendo el video oficial de la canción. Falso si la canción no tiene video oficial conocido.

## **Relaciones:**

- **Artista publica álbum:** Relación de uno a varios porque un artista puede publicar varios albums en su trayectoria.

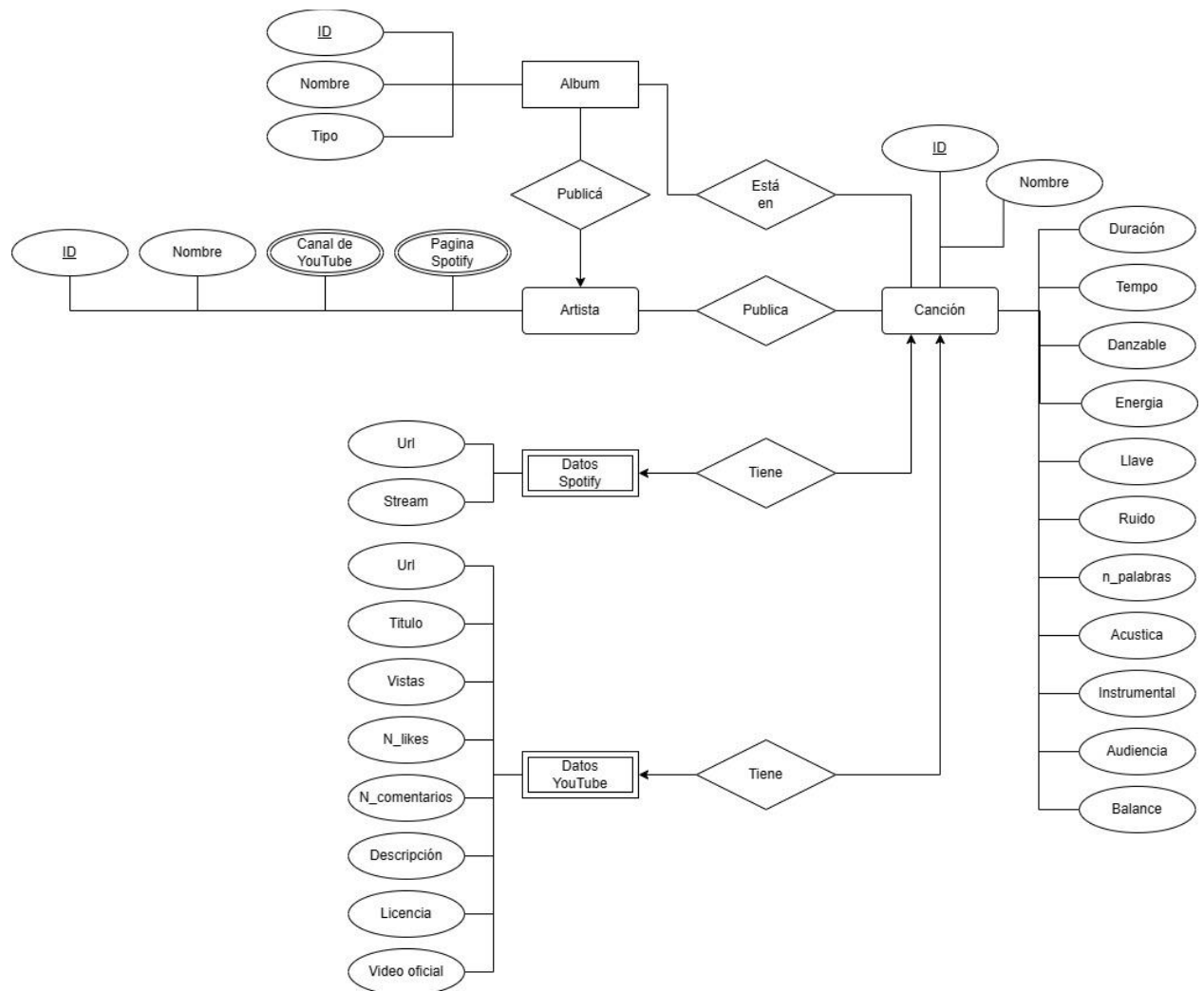
- **Artista crea canción:** Relación de varios a varios debido a que un artista puede crear varias canciones y también hay colaboraciones donde dos o más artistas crean una canción.

- **Canción en álbum:** Relación de varios a varios debido a que un álbum puede contener varias canciones y también una canción puede estar en varios álbumes.

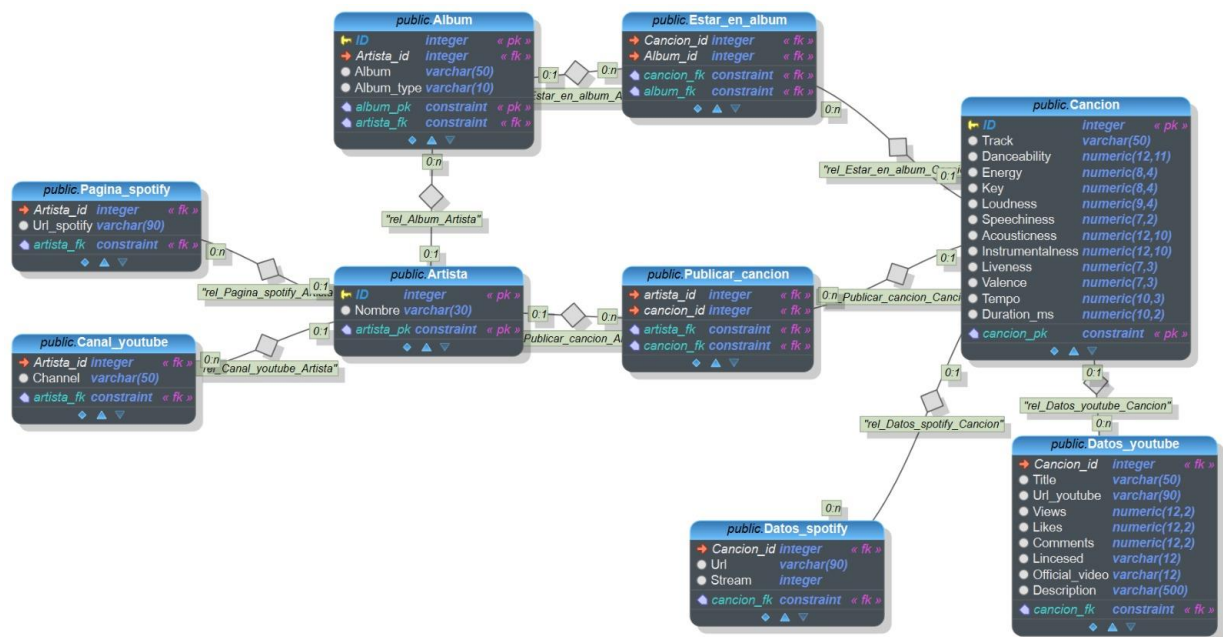
- **Canción subida a Datos\_Y:** Relación de uno a uno porque a cada canción le corresponden solo un conjunto de datos en YouTube.

-Canción subida a Datos\_S: Relación de uno a uno porque a cada canción le corresponde un solo conjunto de datos en Spotify.

### Modelo entidad relación:



### Modelo relacional:



## Análisis

Para ver que tan sustentable puede ser publicar una canción en Spotify o en YouTube se tendrán 4 análisis basados en los datos ingresados en nuestra base de datos.

Para las primeras 2, sacaremos una muestra de 15 artistas al azar y compararemos ciertos valores en Spotify y YouTube, los artistas seleccionados son los siguientes:

1. Gorillaz
2. Snow Patrol
3. 50 Cent
4. Metallica
5. Coldplay
6. Grupo Laberinto
7. Ennio Morricone
8. Seu Jorge
9. Leonard Cohen
10. Good Charlotte
11. Nirvana
12. Maroon 5
13. Romeo Santos
14. Yandel
15. Foo Fighters

**Primera consulta:** Para poder ver el número de reproducciones en total que tiene un artista en YouTube y en Spotify, se sumaran el total de sus reproducciones tanto en YouTube como en Spotify

de cada uno de los artistas seleccionados y se compararan para ver donde ha tenido más reproducciones totales.

**Segunda consulta:** Para poder estimar la aceptación de un artista dentro de YouTube, se seleccionará cada uno de los 15 artistas y se sumarán las vistas que tiene en YouTube y los likes totales. Se compararán los dos valores para poder darnos una idea de la aceptación de las personas al escuchar la canción.

Ahora, para las últimas dos consultas, se seleccionarán las primeras 15 canciones anexadas en la base de datos, y se mirara si estas canciones son lanzadas en álbum o en forma de sencillo. Con el fin de ver qué tipo de canciones consumen más los usuarios en cierta plataforma.

**Tercera consulta:** Para poder estimar una posible aceptación por parte del público a una canción, de cada una de las canciones se compararán las vistas y los likes en YouTube, además, se agrupará por canción lanzada en álbum y lanzada en single para ver la aceptación de cada single.

**Cuarta consulta:** Para cada una de las 15 canciones, se compararán las vistas de la canción en YouTube y en Spotify, a su vez se agruparán por canciones lanzadas en álbum y canciones lanzadas como sencillo para ver cual tipo tiene más vistas acumuladas en las dos plataformas.

### Hipótesis

**Para la primera consulta:** Teniendo en cuenta que en Spotify se suele consumir música de forma más esporádica, podemos suponer que los oyentes recurrentes de Spotify no van a ser muy fieles a un artista en particular. Por lo tanto, suponemos que las vistas acumuladas van a ser mayores en YouTube que en Spotify.

**Para la segunda consulta:** Podemos suponer que la aceptación de una artista varia con respecto al género musical que se dedica a crear. Por ejemplo, suponemos que artistas que realizan canciones del género pop (Maroon 5, Coldplay) tendrían más likes en proporción con las vistas que canciones de genero metal (Metallica, Foo fighters). Por lo tanto, creemos que la proporción entre likes y vistas será mayor o menor dependiendo del género al que se dedique el artista.

**Para la tercera consulta:** Dado a que en YouTube a parte de la canción se vende también un video musical y que las canciones publicadas como sencillo suelen tener más probabilidad de ser publicadas con videoclip, se esperaría que las canciones de tipo sencillo tengan mayor aceptación (proporción vistas-likes) que las canciones publicadas en álbum.

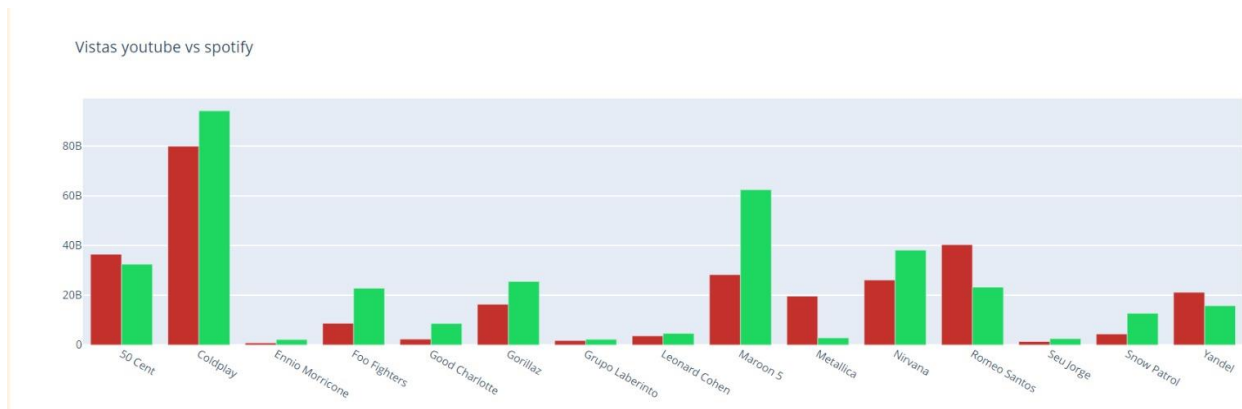
**Para la cuarta consulta:** Como se dijo anteriormente, en Spotify se suele consumir música de forma más esporádica, esto implica que el oyente promedio de Spotify podría consumir más canciones lanzadas en forma de sencillo porque estas no necesitan de contexto o apoyarse de otras canciones para garantizar la transmisión de su mensaje. Como tal esta hipótesis defiende que en total las vistas de canciones lanzadas en álbum son mayores en YouTube, mientras que las vistas de las lanzadas como sencillo son mayores en Spotify.

### Resultados en la base de datos

#### Consulta 1:

artista	stream_spotify	vistas_youtube
50 Cent	32476531312	36481925816.0
Coldplay	94227825888	79978223072.0
Ennio Morricone	2128969376	820454504.0
Foo Fighters	22824827960	8693929840.0

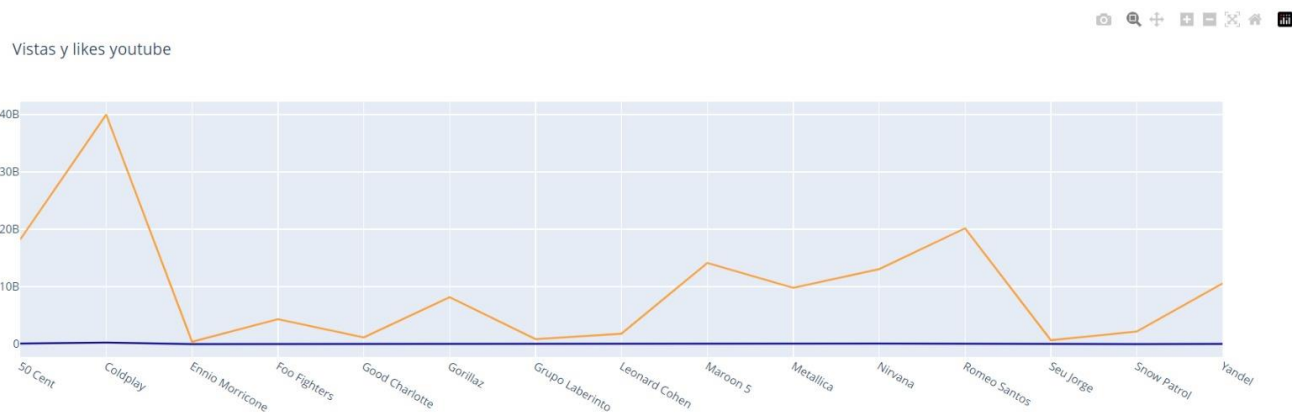
Good Charlotte	8623909648	2327214840.0
Gorillaz	25545395000	16351542768.0
Grupo Laberinto	2216071344	1737663496.0
Leonard Cohen	4618507096	3625199168.0
Maroon 5	62458504024	28260827296.0
Metallica	2813188792	19627692088.0
Nirvana	38153010304	26113999960.0
Romeo Santos	23226245688	40326150792.0
Seu Jorge	2506681984	1361613280.0
Snow Patrol	12688171760	4380332632.0
Yandel	15752641904	21161963728.0



## Consulta 2:

artista	vistas	likes
50 Cent	18240962908.00	110057520.00
Coldplay	39989111536.00	252631580.00
Ennio Morricone	410227252.00	2141712.00
Foo Fighters	4346964920.00	23191956.00
Good Charlotte	1163607420.00	7002784.00
Gorillaz	8175771384.00	78274108.00
Grupo Laberinto	868831748.00	2306456.00
Leonard Cohen	1812599584.00	9660508.00
Maroon 5	14130413648.00	94571020.00
Metallica	9813846044.00	47377592.00
Nirvana	13056999980.00	90806172.00
Romeo Santos	20163075396.00	66517352.00
Seu Jorge	680806640.00	4013308.00

Snow Patrol	2190166316.00	12974036.00
Yandel	10580981864.00	50488756.00

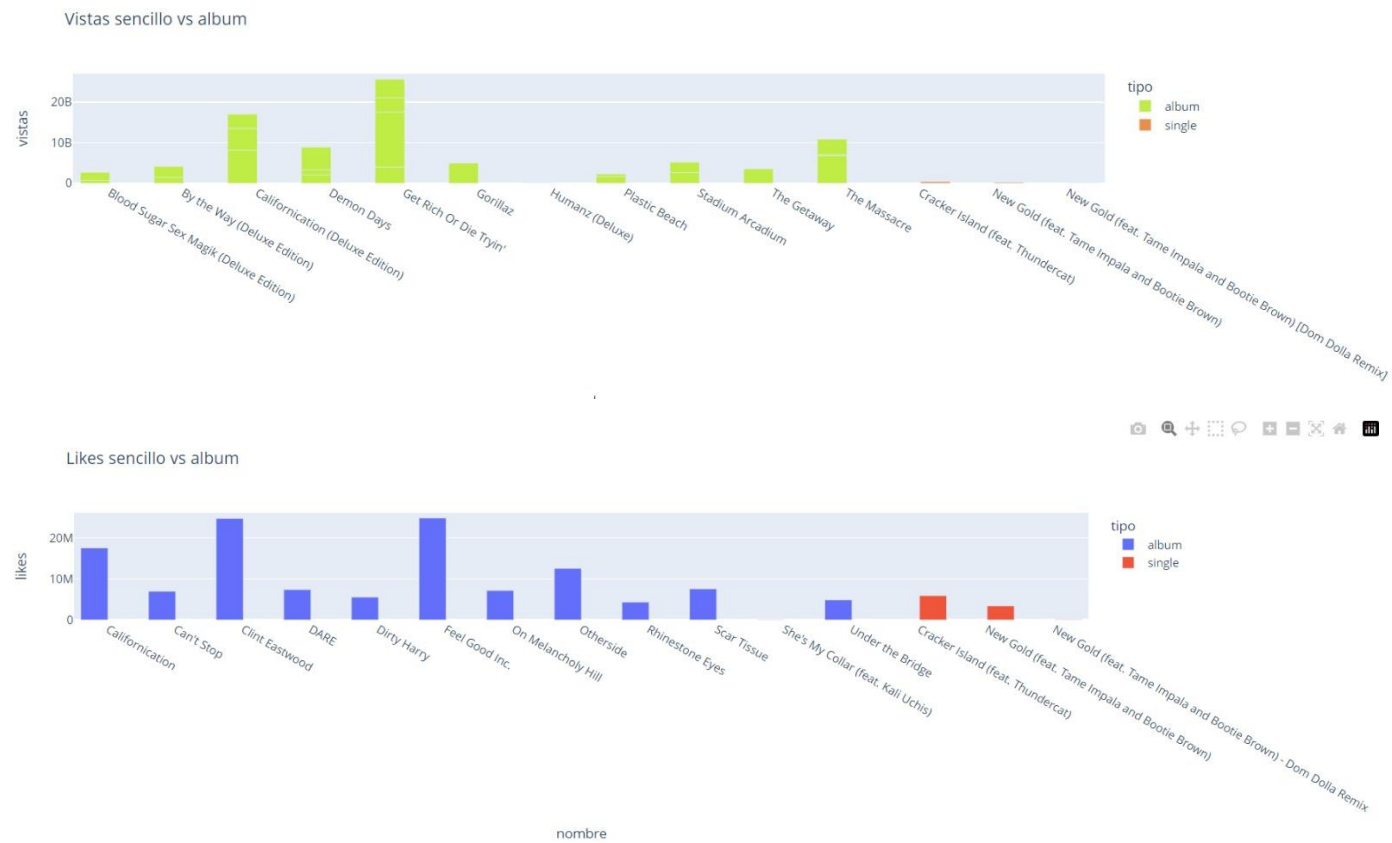


### Consulta 3:

track	album	album_type	vistas	likes
Californication	Californication (Deluxe Edition)	album	4075245036.00	17577884.00
Can't Stop	By the Way (Deluxe Edition)	album	1346543036.00	6960896.00
Clint Eastwood	Gorillaz	album	2473923832.00	24789272.00
Cracker Island (feat. Thundercat)	Cracker Island (feat. Thundercat)	single	195678560.00	5916216.00
DARE	Demon Days	album	1036084644.00	7378632.00
Dirty Harry	Demon Days	album	619044224.00	5547680.00
Feel Good Inc.	Demon Days	album	2774220884.00	24883584.00
New Gold (feat. Tame Impala and Bootie Brown)	New Gold (feat. Tame Impala and Bootie Brown)	single	101220660.00	3385704.00
New Gold (feat. Tame Impala and Bootie Brown) - Dom Dolla Remix	New Gold (feat. Tame Impala and Bootie Brown) [Dom Dolla Remix]	single	3615968.00	93488.00
On Melancholy Hill	Plastic Beach	album	847019808.00	7154308.00
Otherside	Californication (Deluxe Edition)	album	2694114624.00	12561424.00
Rhinestone Eyes	Plastic Beach	album	288046580.00	4316512.00
Scar Tissue	Californication (Deluxe Edition)	album	1740486120.00	7563600.00
She's My Collar (feat. Kali Uchis)	Humanz (Deluxe)	album	4043928.00	70700.00



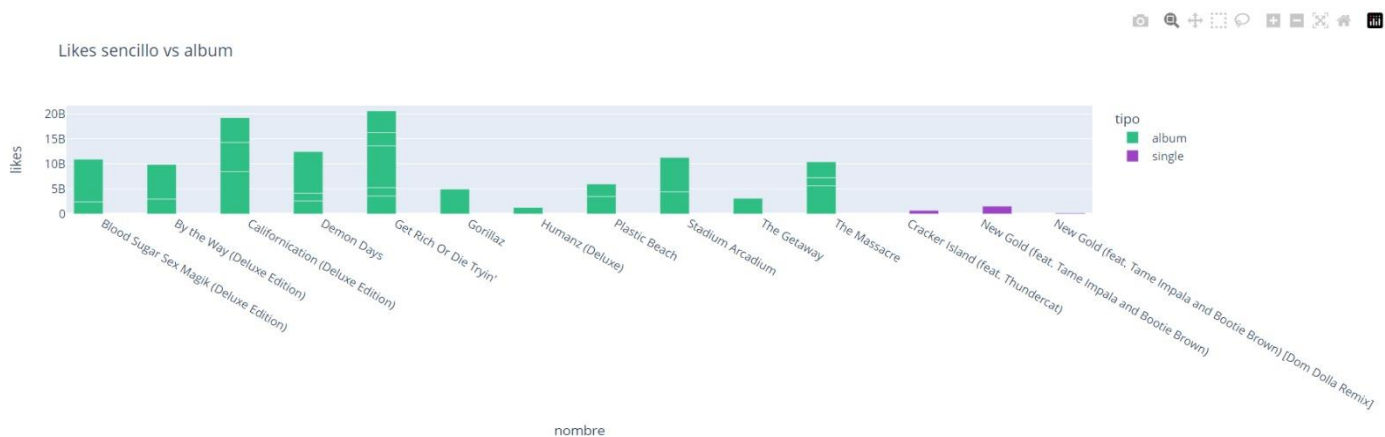
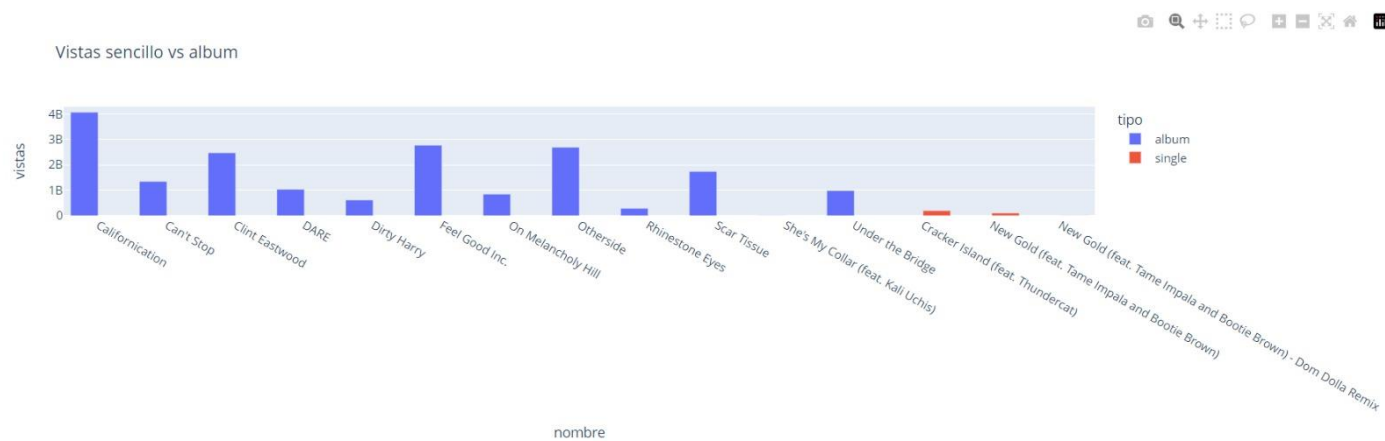
Under the Bridge	Blood Sugar Sex Magik (Deluxe Edition)	album	986750856.00	4854288.00
------------------	----------------------------------------	-------	--------------	------------



Consulta 4:

album	track	album_type	vistas	stream
Blood Sugar Sex Magik (Deluxe Edition)	Give It Away	album	693103408.00	2415577272
Blood Sugar Sex Magik (Deluxe Edition)	Under the Bridge	album	1973501712.00	8494004176
By the Way (Deluxe Edition)	By the Way	album	1432042368.00	2939884064
By the Way (Deluxe Edition)	Can't Stop	album	2693086072.00	6931719608
Californication (Deluxe Edition)	Californication	album	8150490072.00	8445907184
Californication (Deluxe Edition)	Otherside	album	5388229248.00	5862196120
Californication (Deluxe Edition)	Scar Tissue	album	3480972240.00	4910709392

Cracker Island (feat. Thundercat)	Cracker Island (feat. Thundercat)	single	391357120.00	682750416
Demon Days	DARE	album	2072169288.00	2590802616
Demon Days	Dirty Harry	album	1238088448.00	1528597704
Demon Days	Feel Good Inc.	album	5548441768.00	8321878832
Get Rich Or Die Tryin'	21 Questions	album	4025988576.00	3568024352
Get Rich Or Die Tryin'	If I Can't	album	368768.00	1683788632
Get Rich Or Die Tryin'	In Da Club	album	13460931664.00	8333894464
Get Rich Or Die Tryin'	Many Men (Wish Death)	album	3529245024.00	2657403200
Get Rich Or Die Tryin'	P.I.M.P.	album	4581272520.00	4333828488
Gorillaz	Clint Eastwood	album	4947847664.00	4938077904
Humanz (Deluxe)	She's My Collar (feat. Kali Uchis)	album	8087856.00	1276847432
New Gold (feat. Tame Impala and Bootie Brown)	New Gold (feat. Tame Impala and Bootie Brown)	single	202441320.00	1513523208
New Gold (feat. Tame Impala and Bootie Brown) [Dom Dolla Remix]	New Gold (feat. Tame Impala and Bootie Brown) - Dom Dolla Remix	single	7231936.00	170658464
Plastic Beach	On Melancholy Hill	album	1694039616.00	3477308472
Plastic Beach	Rhinestone Eyes	album	576093160.00	2480669864
Stadium Arcadium	Dani California	album	2593829296.00	4400539128
Stadium Arcadium	Snow (Hey Oh)	album	2566969896.00	6885778528
The Getaway	Dark Necessities	album	3520303712.00	3085422984
The Massacre	Candy Shop	album	6711797200.00	5609991144
The Massacre	Disco Inferno	album	318607608.00	1619817496
The Massacre	Just A Lil Bit	album	3809067704.00	3161240880



### Descripción de gráficos

**Para consulta 1:** Tenemos una barra roja que representa el número de vistas acumulado de un artista en YouTube, y una barra verde que representa el número de vistas acumuladas en Spotify. Podemos darnos cuenta de dos cosas y es, primero, la predilección por el pop ya que los artistas Coldplay y Maroon 5 tienen los picos más altos de vistas. A su vez podemos notar como las barras en Spotify son más altas, por lo cual podemos concluir que se acumulan más vistas en la plataforma de Spotify.

**Para consulta 2:** Por su parte, tenemos la tabla de relación entre vistas y likes acumulados de un artista en YouTube, de forma lineal no damos cuenta de que algunos artistas tienen más aceptación por parte de su público, por ejemplo, el artista Coldplay a pesar de tener más vistas acumuladas que el artista Romeo Santos se ve como Coldplay no tiene tanta aceptación en comparación a Romeo Santos, concluyendo que tener un gran número de vistas no es sinónimo de aceptación.

**Para consulta 3:** Podemos ver dos tablas, una que nos mide el número de vistas acumuladas en YouTube de una canción en particular y otra que nos mide el número acumulado de likes de la misma en YouTube. A su vez podemos hacer dos diferenciaciones entre canciones publicadas en álbum y canciones publicadas como sencillo. Ahora bien, tanto en vistas como en likes parece haber un mayor acumulado en canciones publicadas en álbum, esto podría indicarnos que este tipo de canciones son más convenientes para publicar. No obstante, posteriormente veremos la tasa de aceptación del público en las canciones para corroborar que esta idea sea verídica.

**Para consulta 4:** Podemos ver dos tablas, en una se muestra el número acumulado de vistas en YouTube de una canción en particular, mientras que en otro su número acumulado de vistas en Spotify se hace la diferencia también entre canciones del tipo sencillo y canciones publicadas en álbum, en la gráfica se puede ver un mayor acumulado en canciones de tipo álbum y cabe resaltar

que las vistas acumuladas en YouTube y Spotify suele variar bastante entre canciones, cosa que no pasaba cuando analizábamos a los artistas.

### Conclusión

**Para consulta 1:** En los resultados pudimos notar como de los 15 artistas, 11 tienen más vistas acumuladas en Spotify, dejándonos apenas un 4 con más vistas acumuladas en YouTube. Con esto podemos concluir que a pesar de ser una plataforma de consumo rápido en cuanto a música se refiere, Spotify al final sale más rentable para un artista que YouTube.

#### Para consulta 2:

artista	Aceptación
50 Cent	0,006033536747
Coldplay	0,006317509199
Ennio Morricone	0,005220794059
Foo Fighters	0,005335206616
Good Charlotte	0,006018167193
Gorillaz	0,009573911051
Grupo Laberinto	0,002654663582
Leonard Cohen	0,005329642622
Maroon 5	0,006692728349
Metallica	0,004827627394
Nirvana	0,006954596932
Romeo Santos	0,00329896857
Seu Jorge	0,005894930754
Snow Patrol	0,005923767481
Yandel	0,004771651313

En la tabla de aceptación podemos observar la proporción dada entre las vistas de una canción y su número de likes. A pesar de ser relativamente poco se pueden evidenciar cambios sustanciales entre artistas. El artista Gorillaz es el artista que más aceptación posee, con un 0.0095 en sus inicios Gorillaz solía hacer una fusión entre rock y rap, no obstante, últimamente se ha inclinado más por el género pop. Por su parte Romeo Santos es el artista quien menos aceptación tiene con un 0.0032, tratándose de un artista latino y haciendo un género que no es del agrado de la mayoría en este tiempo, esto evidenciándose también con Yandel quien tiene un 0.0047 de aceptación. A su vez Maroon 5 dedicándose a hacer pop almacena una aceptación de 0.0066.

Lo anterior parece apoyar a nuestra hipótesis de que los artistas con más tasa de aceptación suelen ser los artistas dedicados al género pop, a parte podemos añadir que asegura más aceptación ser un artista norteamericano.

#### Para consulta 3:

Tipo de album	Aceptación
album	0,006547807877
single	0,03126433663

Considerando la tabla anterior podemos ver que en YouTube la aceptación por canciones publicadas como sencillo es mucho mayor que las canciones publicadas en álbum. Por esto podríamos concluir que, tal y como decimos en la hipótesis, las canciones de tipo single son más aceptadas por el público de YouTube, por lo mismo salen más rentables a pesar de tener que invertir más dinero en videoclips.

#### **Para consulta 4:**

Podemos evidenciar, también, como se cumple la cuarta hipótesis, esto debido a que, a pesar de los videos musicales, las canciones publicadas como sencillo tienen más visualizaciones acumuladas en Spotify que en YouTube. Por lo mismo, para un artista sería más conveniente publicar una canción en Spotify que en YouTube.

#### **Conclusión general**

Podemos que concluir que, aunque el mercado entre YouTube y Spotify es bastante reñido, tratándose de un artista pequeño quien no tiene capital para invertir en un video musical podemos recomendarle que publique sus canciones en Spotify, ya que en este mercado se vende bien la canción sin necesidad de incluir video musical, por ende, ahorrándose el gasto de invertir en la realización del video con las mismas probabilidades de ganancia de dinero que si lo subiera en YouTube.