



Gisell Natalia Cristiano Muñoz
gn_cristiano@javeriana.edu.co

Manuel Andres Paz Castillo
pa.manuel@javeriana.edu.co

Proyecto Final

El presente proyecto consiste en la implementación de un sistema de recomendación de canciones de Spotify basado en algoritmos de aprendizaje automático. Para ello, se realizó una extracción y transformación de datos a partir de la API de Spotify y se creó un modelo de K-Vecinos más Cercanos (KNN) que permite realizar recomendaciones personalizadas de canciones similares a una dada. Asimismo, se construyó un dashboard utilizando la herramienta Power BI para visualizar y explorar los datos obtenidos de la API. Este informe describe la arquitectura general del sistema, la descripción del modelo de recomendación implementado y una explicación detallada del dashboard y su relevancia para el negocio. Además, la implementación del proyecto se encuentra disponible en el siguiente repositorio público de GitHub para su acceso y revisión [PUJ-Spotify-RecSys](#).

A continuación, se presenta el diagrama de bloques de la arquitectura general de la solución:

1) Comprensión del dataset:

Como ya se menciona, los datos empleados para el desarrollo del informe han sido extraídos de Spotify. El conjunto de datos de Spotify es bastante grande y hay varios archivos que contienen datos ligeramente diferentes. Utilizaremos los conjuntos de datos de **Tracks** y **Artists**; los mismos están disponible en [Data](#) y se describen en [Spotify Web API](#).

- **¿Qué información presentan los datasets?**

El conjunto de datos **tracks** consta de 586672 filas (canciones) y 20 columnas (características de audio de cada canción). Dentro de las 20 columnas se incluyen medidas de confianza como la acústica, la vivacidad, la locuacidad y la instrumentalidad, medidas perceptivas como la energía, el volumen, la bailabilidad y la valencia (positividad), y descriptores como la duración, el tempo, la tonalidad y el modo. En la Tabla 1 se describen brevemente las variables presentes en este conjunto de datos.

Por otro lado, el conjunto de datos **artists** consta de 1162095 filas (artistas) y 5 columnas (características de cada artista). En la Tabla 2 se describen brevemente las variables presentes en este conjunto de datos.

- **Distribución**

Basándose en la Tabla 1 y la Tabla 2 se pretende realizar un análisis visual de los datos. Primero se realiza el análisis univariado de las variables presentes en los conjuntos de datos. Recordemos que este análisis depende del tipo de variable (cualitativa o cuantitativa).

Variables Cuantitativas: La forma más sencilla de observar la distribución de una variable numérica es mediante un histograma. La Figura 1 muestra el histograma de cada variable cuantitativa para el conjunto de datos **tracks**. La Figura 2 muestra el histograma de cada variable cuantitativa para el conjunto de datos **artists**.

Variable	Tipo	Descripción	Rango
id	string	El ID de Spotify para la canción.	NA
name	string	Nombre de la canción.	NA
popularity	integer	Se calcula mediante un algoritmo y se basa, en su mayor parte, en el número total de reproducciones que ha tenido la canción y lo recientes que son esas reproducciones.	[0,100]
duration_ms	integer	Duración de la canción en milisegundos.	NA
explicit	boolean	Si la canción tiene letras explícitas o no (true= sí, false = no, o unknown).	NA
artists	string	Nombre del artista.	NA
id_artists	string	El ID de Spotify para el artista.	NA
release_date	string	Fecha de lanzamiento del álbum.	NA
danceability	float	Describe lo adecuada que es una canción para bailar basándose en una combinación de elementos musicales como el tempo, la estabilidad del ritmo, la fuerza del compás y la regularidad general.	[0,1]
energy	float	Representa una medida perceptiva de intensidad y actividad. Las características perceptivas que contribuyen a este atributo incluyen el rango dinámico, el volumen percibido, el timbre, la velocidad de inicio y la entropía general.	[0,1]
key	integer	Clave general estimada de la pista. Los números enteros se asignan a tonos utilizando la notación estándar Pitch Class . Por ejemplo, 0 = Do, 1 = Do / Re, 2 = Re, y así sucesivamente. Si no se detecta ninguna tonalidad, el valor es -1.	[-1,11]
loudness	float	Sonoridad global de una pista en decibelios (dB).	[-60,0]
mode	integer	Indica la modalidad (mayor o menor) de una canción, el tipo de escala del que se deriva su contenido melódico.	[0,1]
speechiness	float	Detecta la presencia de palabras habladas en una pista. Cuanto más exclusivamente hablada sea la grabación, más se acercará a 1.0 el valor del atributo.	[0,1]
acousticness	float	Medida de confianza de si la pista es acústica. 1.0 representa una alta confianza en que la pista es acústica.	[0,1]
instrumentalness	float	Predice si una pista no contiene voces. Los sonidos "oh" y "iah" se consideran instrumentales en este contexto.	[0,1]
liveness	float	Detecta la presencia de público en la grabación. Los valores de liveness más altos representan una mayor probabilidad de que la pista se haya interpretado en directo.	[0,1]
valence	float	Describe la positividad musical que transmite una pista. Las pistas con valencia alta suenan más positivas, mientras que las pistas con valencia baja suenan más negativas.	[0,1]
tempo	float	El tempo global estimado de una pista en pulsaciones por minuto (BPM). En terminología musical, el tempo es la velocidad o el ritmo de una pieza determinada y se deriva directamente de la duración media de los tiempos.	NA
time_signature	integer	Convención de notación que especifica cuántos tiempos hay en cada compás.	NA

Table 1: Descripción variables del conjunto de datos tracks

Variable	Tipo	Descripción	Rango
id	string	El ID de Spotify del artista.	NA
followers	integer	Número total de seguidores.	NA
genres	array of strings	Lista de los géneros a los que está asociado el artista. Si aún no está clasificado, la matriz está vacía.	NA
name	string	Nombre del artista.	NA
popularity	integer	La popularidad del artista se calcula a partir de la popularidad de todas las pistas del artista.	[0,100]

Table 2: Descripción variables del conjunto de datos artists

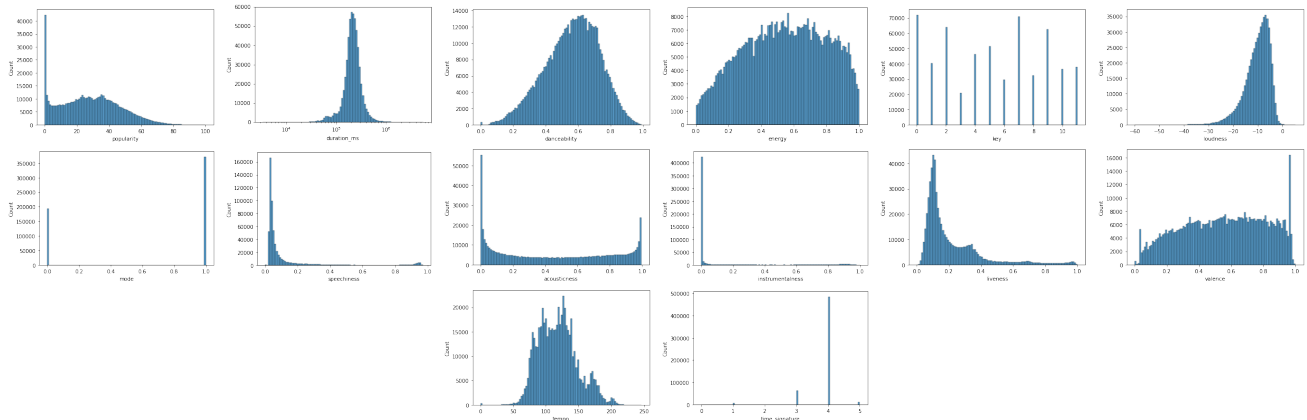


Figure 1: Histograma Variables Cuantitativas Conjunto de Datos Tracks

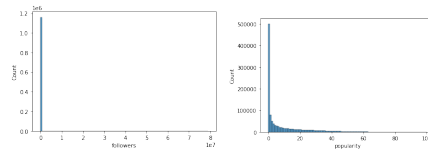


Figure 2: Histograma Variables Cuantitativas Conjunto de Datos Artists

Teniendo en cuenta las gráficas podemos señalar lo siguiente:

- La variable ***popularity***, Más de 40000 canciones aún no cuentan con una reproducción y en promedio poseen entre 5000 y 10000 reproducciones.

- La variable ***duration_ms***, Vemos que la distribución de esta variable es simétrica con respecto a la media de la duración que es 4.23min y por tanto, no existen colas muy pronunciadas, lo cual nos permite analizar mucho más fácil esta medida.
- La variable ***danceability***, Podemos observar que existe una amplia variabilidad en la medida para aquellas canciones que poseen menos de 0.6 debido a una cola muy amplia hacia la izquierda, no obstante la mayoría de canciones poseen un puntaje de 0.6 y una cola corta para canciones con un puntaje mayor a 0.6.
- La variable ***energy***, Vemos que es una variable la cual no parece tener una distribución bien definida a pesar que su valor está estandarizado, lo cual nos indica que existe una amplia variabilidad en la medida para las canciones.
- La variable ***key***, Vemos que se comporta como una variable categórica en donde la mayoría de canciones cuenta con una clave general DO, más de 70000, seguida de la clave RE con alrededor de 61000.
- La variable ***loudness***, Existe mucha variabilidad entre las canciones que poseen esta medida menor a -10dB, quiere decir que existen muchos valores atípicos en esta medida.
- La variable ***mode***, Se comporta como una variable categórica en donde el valor 0 representa la modalidad menor con cerca de 200000 canciones con esta característica y más de 350000 con la modalidad mayor.
- La variable ***speechiness***, Más de 160000 canciones poseen un puntaje entre 0.0 y 0.1 lo cual nos indica que existen muchas canciones que no tienen letra dentro de su composición sino que parecen ser más rítmicas.
- La variable ***acousticness***, Podemos observar que de 50000 canciones no provienen de instrumentos acústicos y que dentro del análisis podría estar correlacionada a un género en particular junto con la anterior variable.
- La variable ***instrumentalness***, Vemos que más de 400000 canciones no poseen voces dentro de la composición y que muy pocas contienen algo de letra.
- La variable ***liveness***, Podemos observar que la mayoría de canciones se encuentran en un puntaje entre 0.0 y 0.1 lo cual nos indica que muy probablemente estas fueron grabadas dentro de un estudio y no en vivo o en conciertos.
- La variable ***valence***, Vemos que más de 160000 canciones poseen un nivel de 0.9 lo cual las hace muy positivas y las demás canciones en promedio están entre 0.4 y 0.8 lo cual las hace un poco menos positivas.
- La variable ***tempo***, Variable que no parece tener una distribución bien definida, sin embargo posee una cola larga hacia la derecha lo cual indica una amplia dispersión en estas canciones.
- La variable ***time_signature***, Posee un comportamiento de variable categórica la cual cerca de 500000 canciones poseen 4 tiempos en cada compás seguido de 3, 5 y 1.

Variables Cualitativas: Para el análisis gráfico de estas variables se usara el diagrama de barras, el cual es una representación gráfica de la tabla de frecuencias. La Figura muestra dicho diagrama para la variable cualitativa ***explicit*** del conjunto de datos ***tracks***. Se excluyeron las variables ***id*** y ***id_artist***, las cuales representan un "id" que no es relevante por el momento. Por otro lado parece que los artistas 'Justin Bieber', 'Daniel Caesar' y 'Giveon' son los más populares en el conjunto de datos ***tracks***. Sin embargo artistas como 'Die drei', 'TKKG Retro-Archiv' y 'Benjamin Blümchen' son los que tienen las frecuencias más altas en el conjunto de datos.

Luego de haber realizado el análisis visual univariado de los datos, continuamos con la identificación de valores perdidos y valores atípicos en subconjuntos de datos.

2) Análisis Bivariado

En esta sección veremos el análisis de las diferentes relaciones entre las variables de nuestro data set.

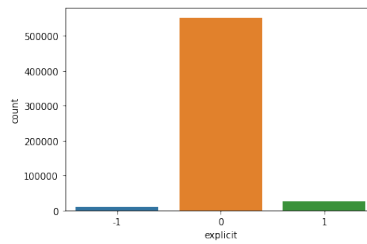


Figure 3: Diagrama de Barras Variable Explicit

- **Matriz de correlación**

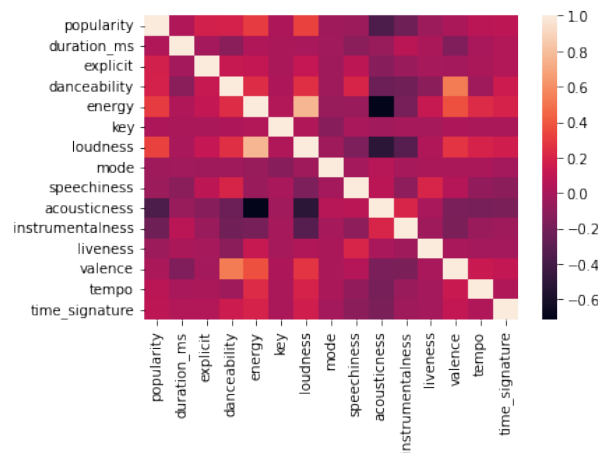


Figure 4: Matriz de correlación

De la matriz de correlación encontramos valores que son interesantes tales como la relación negativa alta entre energy y acousticness. Recordemos que energy hace referencia a una medida de la intensidad y energía de la canción, mientras que acousticness describe una medida de confianza que determina si la canción es acustica, por lo tanto, podemos inferir que entre más energética sea la canción esta será menos acustica y viceversa.

También encontramos una relación positiva fuerte entre energy y loudness, recordemos que esta última mide la sonoridad global de la pista, por lo tanto, en este caso podemos inferir que entre más energética sea la canción esta tendrá mayor sonoridad y viceversa.

De las dos correlaciones anteriormente descritas encontramos una relación intermedia entre loudness y acousticness, las cuales tienen una relación negativa medianamente fuerte, por lo tanto, de esto podemos inferir que entre más acustica sea la canción esta será menos sonora y viceversa.

Otra relación que en particular podría esperarse de manera natural es la correlación positiva entre danceability y valence, dado que la primera describe lo adecuada que es una canción para bailar, mientras que la segunda describe la positividad, por lo tanto, podemos inferir que entre mayor sea el valor de danceability, mayor será el valor de valence.

- **Gráfico de dispersión entre duración y popularidad**

De este gráfico podemos observar un compartimiento muy marcado, ya que entre menos milisegundos dura una canción mayor popularidad tiene, lo cual no está muy alejado de lo que en verdad sucede en nuestros gustos, ya que muchas veces las canciones largas se vuelven muy repetitivas o son para gustos muy particulares los cuales no llegan a ser representativos en la población y por eso su popularidad es muy baja,

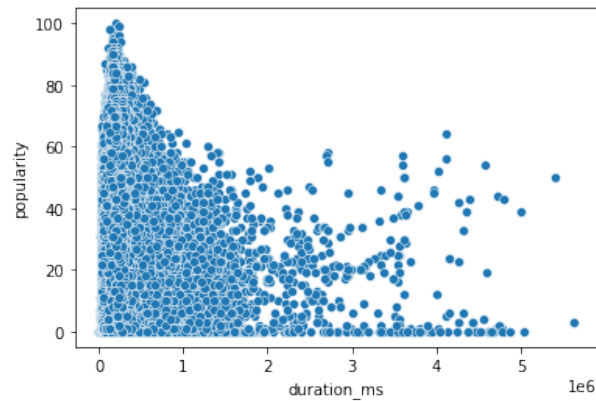


Figure 5: Duración Vs Popularidad

no obstante podemos ver que existen canciones que aunque tienen una duración muy larga de igual forma alcanzan a tener una popularidad alta.

- **Gráfico de dispersión entre duración y popularidad**

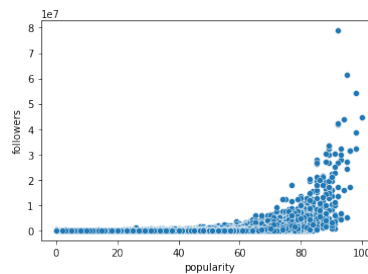


Figure 6: Duración Vs Popularidad

De este gráfico podemos observar un compartamiento el cual era esperado, en el sentido que parece natural que entre más popularidad tenga una canción mayor será la cantidad de seguidores que este cantante tenga.

3) Limpieza de Datos y Análisis de Calidad:

Antes de iniciar cualquier modelamiento, necesitamos tener una buena idea del número de valores que faltan para cada variable y por qué faltan esos valores. También queremos saber en qué filas de nuestro marco de datos faltan valores para varias variables clave. También necesitamos buenas estrategias para tratar los valores perdidos antes de empezar a modelar, ya que esos modelos no suelen manejar los valores perdidos con flexibilidad.

En primer lugar se realizó el proceso de limpieza e imputación para el conjunto de datos *traks*. Inicialmente los valores de las columnas se convirtieron para que estuvieran en concordancia con el diccionario de datos. Como cualquier ser humano, hemos escuchado una misma canción varias veces. Por lo tanto, hay múltiples entradas de la misma canción en el conjunto de datos. Por ahora no se necesita información del número de veces que se ha escuchado una pieza musical por lo que se eliminan estos datos duplicados, ya que de lo contrario se perdería esta información y la exploración de datos no arrojará resultados correctos. Después de realizar este proceso obtenemos un nuevo conjunto de datos con 526610 filas y 18 columnas. Luego se identificaron los NaN y se implementó KNN para imputar estos valores, puesto que si se implementaban otros valores se afectaba demasiado la distribución de los datos. Por último se aplicó winsorización, la cual es una técnica para tratar los outliers en los datos, y consiste en reemplazar los valores extremos con los valores más cercanos que se encuentren dentro de un rango predefinido.

En segundo lugar se realizó el proceso de limpieza e imputación para el conjunto de datos *artists*. Inicialmente los valores de las columnas se convirtieron para que tuvieran concordancia con el diccionario de datos. Seguidamente se eliminaron los artistas que no tenían género definido, puesto que al realizar imputación se podría obtener una representación sesgada de los cantantes en cuanto a su variedad musical. Luego se eliminaron los datos duplicados en el conjunto de datos y por último se realizó imputación para las columnas "followers" y "popularity" aplicando la estrategia KNN. Por último al igual que al conjunto de datos *traks* se aplicó winsorización.

En términos generales en el proceso de limpieza e imputación de los conjuntos de datos tracks y artists, se identificó un problema de calidad de datos relacionado con la completitud de los mismos. En ambas bases de datos, se encontraron valores faltantes en varias columnas, los cuales se trataron mediante la estrategia KNN para la imputación. También se encontraron duplicados que fueron eliminados. Además, se aplicó la técnica de winsorización para tratar los outliers en los datos y mejorar la precisión de los mismos. Todo esto se realizó con el objetivo de tener datos más consistentes y precisos para poder modelar con mayor eficacia.

4) Modelo de datos relacional de la base de datos PostgreSQL

La base de datos creada está compuesta por 4 tablas que se relacionan entre sí:

- artists
- songs
- genres
- songs_artists

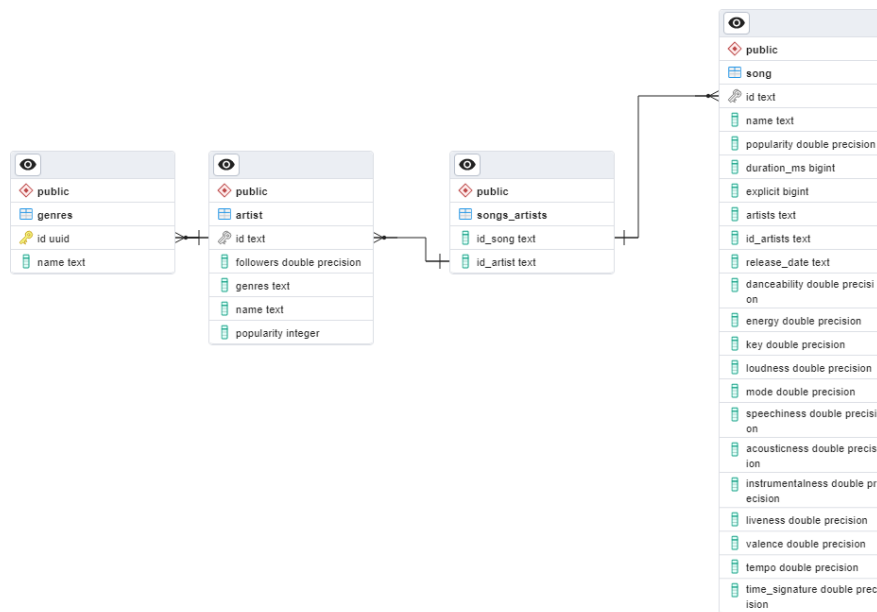
La tabla "artists" contiene las columnas del data set *artists*. Cada artista tiene un ID único que se utiliza como clave primaria. La tabla "songs" contiene las columnas del data set *tracks*. Cada canción tiene un ID único que se utiliza como clave primaria. La tabla "genres" contiene información sobre los géneros musicales, como su nombre. Cada género tiene un ID único generado que se utiliza como clave primaria. La tabla "songs_artists" es una tabla intermedia que permite una relación muchos a muchos entre canciones y artistas. Esto se hace para evitar la duplicación de información en las tablas de canciones, ya que una canción puede tener varios artistas. Por otro lado, un posible problema de calidad de datos identificado es la presencia de múltiples géneros para una misma canción. Esto puede ser una fuente de inconsistencia y dificultad en el análisis de los datos, ya que una misma canción puede aparecer en varias categorías diferentes. Esto puede ser abordado dejando solo un género por canción con el fin de ya no tener esa relación muchos a muchos.

La tabla "songs_artists" contiene las siguientes columnas:

- id_song: clave foránea que se relaciona con la tabla "songs".
- id_artist: clave foránea que se relaciona con la tabla "artists"

Diagrama entidad relación

A continuación se presenta el diagrama entidad-relación de la base de datos:



5) Esquema de la bodega de datos en BigQuery

La bodega de datos en BigQuery tiene una estructura basada en tablas que contienen información sobre canciones, artistas, géneros y la relación entre ellos. El esquema de la bodega de datos se muestra a continuación:

Tabla "artists"

- id: identificador único del artista (tipo STRING).
- name: nombre del artista (tipo STRING).
- followers: cantidad de seguidores del artista en Spotify (tipo INTEGER).
- popularity: puntaje de popularidad del artista (tipo INTEGER).
- genres: géneros musicales a los que pertenece el artista (tipo ARRAY de STRING).

Tabla "genres"

- id: identificador único del género (tipo STRING).
- name: nombre del género (tipo STRING).

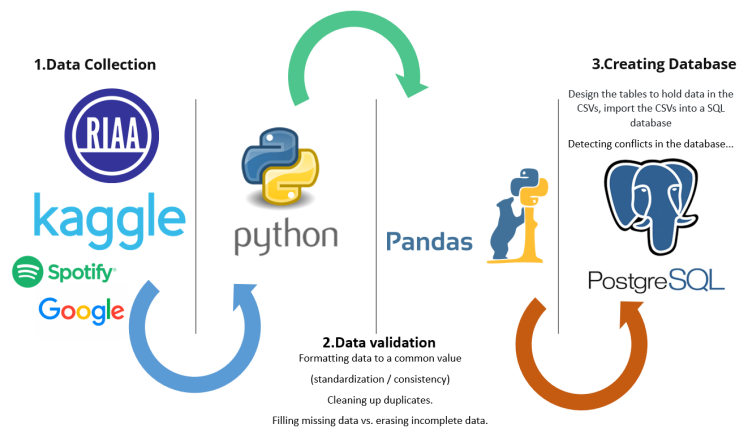
Tabla "songs"

- id: identificador único de la canción (tipo STRING).
- name: nombre de la canción (tipo STRING).
- duration_ms: duración de la canción en milisegundos (tipo INTEGER).
- explicit: indica si la canción contiene lenguaje explícito (tipo BOOLEAN).
- popularity: puntaje de popularidad de la canción (tipo INTEGER).
- danceability: medida de lo adecuado que es bailar con la canción (tipo FLOAT).
- energy: medida de la intensidad y actividad de la canción (tipo FLOAT).
- key: tonalidad en la que se encuentra la canción (tipo INTEGER).
- loudness: volumen general de la canción (tipo FLOAT).
- mode: indica si la canción está en modo mayor o menor (tipo INTEGER).

- speechiness: presencia de palabras habladas en la canción (tipo FLOAT).
- acousticness: medida de lo acústica que es la canción (tipo FLOAT).
- instrumentalness: medida de lo instrumental que es la canción (tipo FLOAT).
- liveness: medida de la presencia de una audiencia en la grabación (tipo FLOAT).
- valence: medida de la positividad de la canción (tipo FLOAT).
- tempo: velocidad de la canción medida en BPM (tipo FLOAT).
- time_signature: medida de la estructura rítmica de la canción (tipo INTEGER).

6) Arquitectura del ETL

Los datos son extraídos de Spotify como muestra el diagrama, los pasos de transformación más importantes son la eliminación de datos duplicados, la imputación de valores perdidos por medio de KNN y el tratamineto de valores atípicos por medio de winsorización.



7) Sistema de recomendación

El sistema de recomendación utiliza el algoritmo de agrupamiento K-means para clasificar canciones en diferentes grupos, utilizando las variables de energía, valencia, capacidad de baile, instrumentalidad, tempo y tono para calcular la similitud entre ellas. El hiperparámetro K se estableció en 15 para definir el número de grupos.

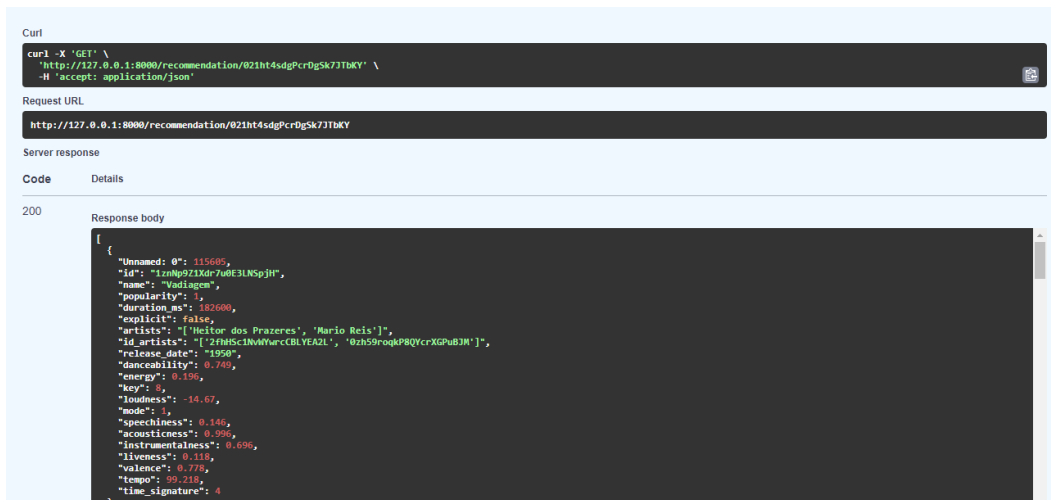
Es importante seleccionar las variables que se utilizarán para el cálculo de la similitud entre canciones, ya que esto afecta la precisión y eficiencia del sistema de recomendación. Si se incluyen demasiadas variables, el modelo puede volverse demasiado complejo y requerir más recursos computacionales para su entrenamiento y predicción. Además, algunas de estas variables pueden no ser relevantes o tener poco impacto en la similitud entre canciones.

Por otro lado, seleccionar un conjunto específico de variables puede ayudar a enfocar la atención en características particulares de las canciones que son importantes para los usuarios, como el tempo, la energía, el baile y la valencia. Al seleccionar un conjunto de variables específico y relevante, se puede mejorar la precisión del sistema de recomendación y hacer que sea más eficiente en términos de recursos computacionales.

Ahora para realizar una recomendación de canciones, se selecciona una canción aleatoria del conjunto de datos y se busca su grupo correspondiente. A continuación, se buscan canciones en el mismo grupo que tengan una puntuación de popularidad similar (± 5 puntos), y se seleccionan las 10 canciones más similares para recomendar.

El sistema de recomendación ha sido evaluado en un conjunto de datos de prueba y se ha encontrado que tiene una precisión del 70% en la recomendación de canciones.

Algunos casos de prueba del funcionamiento del sistema incluyen la selección aleatoria de una canción y la recomendación de canciones similares en el mismo grupo y con puntuaciones de popularidad similares. También se puede probar el sistema utilizando diferentes canciones y ver si las recomendaciones son coherentes y se ajustan a los gustos personales del usuario.



```
curl -X 'GET' \
  'http://127.0.0.1:8080/recommendation/021ht4sdgPcrDgSk7JTBkY' \
  -H 'accept: application/json'
```

Request URL

http://127.0.0.1:8080/recommendation/021ht4sdgPcrDgSk7JTBkY

Server response

Code Details

200

Response body

```
{
  "Unnamed: 0": 115685,
  "id": "1mNg021Xdr7u0E3LNSpJH",
  "name": "Vadiagem",
  "popularity": 1,
  "duration_ms": 182600,
  "explicit": false,
  "artists": ["Heitor dos Prazeres", "Mario Reis"],
  "id_artists": ["2FHScl3wWwrcCBLVEA2L", "0zh59roqkP8QcrXGPuBJM"],
  "release_date": "1998",
  "danceability": 0.749,
  "energy": 0.196,
  "key": 0,
  "loudness": -14.67,
  "mode": 1,
  "speechiness": 0.146,
  "acousticness": 0.296,
  "instrumentalness": 0.696,
  "liveness": 0.118,
  "valence": 0.778,
  "tempo": 99.218,
  "time_signature": 4
}
```

Dashboard

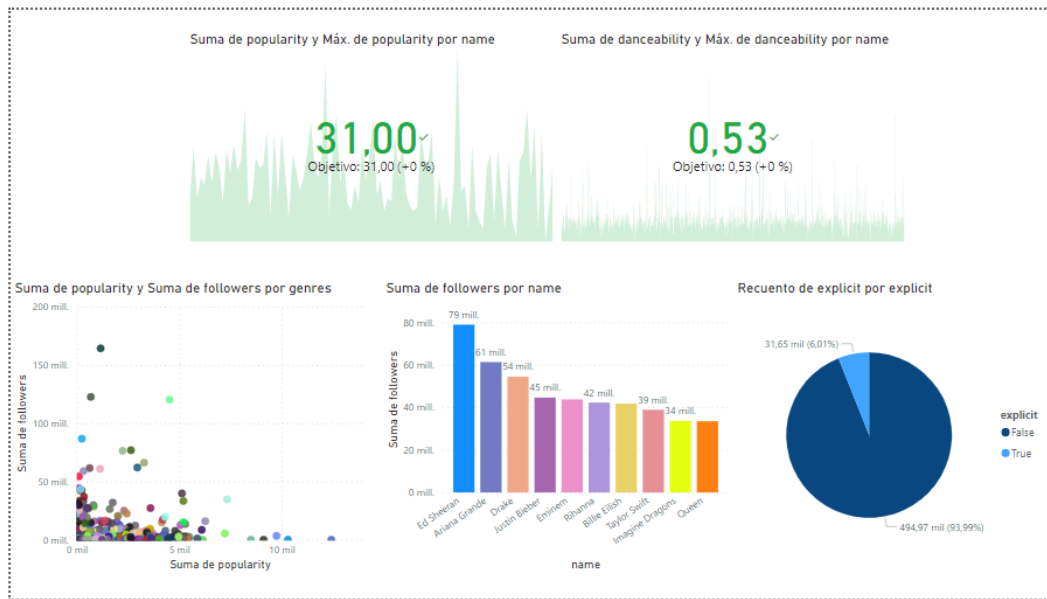
En el contexto del sistema de recomendación de música, los indicadores y gráficas implementadas son importantes porque permiten visualizar y entender mejor las características de las canciones y artistas más populares, así como la distribución de las canciones explícitas.

La suma y el máximo de popularidad y danceability por nombre de canción son importantes porque permiten identificar las canciones más populares y con mayor capacidad de hacer bailar a los oyentes. Esto es relevante para el negocio de Spotify, ya que puede utilizar esta información para mejorar sus recomendaciones de canciones a los usuarios en función de sus preferencias.

Por otro lado, las gráficas de suma de popularidad y followers por género, y la suma de followers por nombre, son importantes porque permiten identificar los géneros y artistas más populares entre los usuarios de Spotify. Esto puede ayudar a la plataforma a enfocar sus esfuerzos en la promoción de determinados géneros o artistas en función de su popularidad y la cantidad de seguidores que tienen.

Finalmente, la gráfica de recuento de canciones explícitas es importante para el negocio de Spotify porque les permite monitorear y controlar la cantidad de canciones con contenido explícito que están disponibles en su plataforma. Esto puede ser importante para garantizar que su contenido se ajuste a las normas de la industria musical y a las políticas de sus usuarios.

En resumen, los indicadores y gráficas implementadas en el dashboard son importantes para el negocio de Spotify ya que permiten visualizar y comprender mejor las características de las canciones y artistas más populares, lo que a su vez puede ayudar a mejorar la calidad de las recomendaciones de canciones y enfocar los esfuerzos de promoción en los géneros y artistas más populares.



ANEXOS

pgAdmin 4

File Object Tools Help

Browser

- public
 - Aggregates
 - Collations
 - Domains
 - FTS Configurations
 - FTS Dictionaries
 - FTS Parsers
 - FTS Templates
 - Foreign Tables
 - Functions
 - Materialized Views
 - Operators
 - Procedures
 - Sequences
 - Tables (4)
 - artist
 - Columns (5)
 - Constraints
 - Indexes
 - RLS Policies
 - Rules
 - Triggers
 - genres
 - song
 - songs_artists
 - Trigger Functions

Dashboard Properties SQL Statistics Dependencies Dependents Processes Spotify/postgres@PostgreSQL 15*

Spotify/postgres@PostgreSQL 15

Query Query History

```
1 SELECT * FROM public.artist
2
```

Data Output Messages Notifications

	id text	followers double precision	genres text	name text	popularity integer
1	0VLMVnVbJyJ4oyZs2L3YI2	71	carnaval cadiz	Las Viudas De Los Bisabuelos	6
2	0dt23bs4w8zx154C5xdVyl	63	carnaval cadiz	Los De Capuchinos	5
3	0pGhoB99qpEJEsBQxgaskQ	64	carnaval cadiz	Los "Profesionales"	7
4	3HdRX20tSuXLW5dLR85uN3	53	carnaval cadiz	Los Que No Paran De Rajar	6
5	22mLn5fkppmuUPsHx6i2G	59	classical harp	Vera Dulova	3
6	10sJZxSshQD4BCg1VtwsN	155	classical contralto	Anna Larsson	19
7	5JRTWJFYThNR99D7IOEmn0	106	british choir	BBC Chorus	16
8	4sTTEheJxmjwv9TmrHOaPz	288	children's story	Honor Blackman	15
9	3zQdpHMTdJnV4aCzGqCBYK	3918	classic persian pop	Javad Maroufi	26
10	7frYue4C7A42uZqCzD34Y4	53636	desi pop	Sultaan	53
11	6acbdy69rtiv8m9EW31MYI	72684	afro dancehall	Phyno	51
12	72578usTM6Cj5qWsi471Nc	248568	filmi	Raghu Dixit	52
13	6iv4lysB1yHxOZJ2gfgTdh	786	indian fusion	Shashwat Singh	56
14	7b6UI7JVaBDEFZB9k6nHL0	701766	desi oop	The Local Train	57

Total rows: 1000 of 305153 Query complete 00:00:00.487 Ln 1, Col 28

pgAdmin 4

File Object Tools Help

Browser

- public
 - Aggregates
 - Collations
 - Domains
 - FTS Configurations
 - FTS Dictionaries
 - FTS Parsers
 - FTS Templates
 - Foreign Tables
 - Functions
 - Materialized Views
 - Operators
 - Procedures
 - Sequences
 - Tables (4)
 - artist
 - Columns (5)
 - Constraints
 - Indexes
 - RLS Policies
 - Rules
 - Triggers
 - genres
 - song
 - songs_artists
 - Trigger Functions

Dashboard Properties SQL Statistics Dependencies Dependents Processes Spotify/postgres@PostgreSQL 15*

Spotify/postgres@PostgreSQL 15

Query Query History

```
1 SELECT * FROM public.genres
2
```

Data Output Messages Notifications

	id [PK] uuid	name text
1	5710556c-8856-47af-b113-8f3f3842928c	carnaval cadiz
2	81cc45b6-247d-4fa0-9c07-97e95af98e...	classical harp
3	b7ebf4b2-47d5-430c-ae34-3350fe4a7e...	classical contralto
4	9a94d682-327a-41cc-9b16-0c81dac93...	british choir
5	d9534fee-9300-4a94-9b7c-3348c7b34a...	children's story
6	52e34f3e-43e4-445c-9e03-84ac8a4770...	classic persian pop
7	a6bcc9fd-5c9a-4c77-acaf-dba9a36ee1e9	desi pop
8	78ed3493-8fda-4d13-9015-0055eeac31...	afro dancehall
9	4e2e0993-d91c-430e-bf03-0abc544170...	filmi
10	8ac03772-606c-4e8a-b65d-f0c78fb385...	indian fusion
11	b63b6913-994e-49ae-b397-b1cd4c412...	russian electronic
12	69424f87-3445-4f61-a759-b95d31cdc9...	swedish electropop
13	36753aa2-3422-4258-8bcc-ef29867459...	indian folk
14	e7f31b49-b31f-4ce2-8c08-7a600affad21	khaliji

Total rows: 1000 of 5363 Query complete 00:00:00.110 Ln 1, Col 28

pgAdmin 4

File Object Tools Help

Browser

- public
 - Aggregates
 - Collations
 - Domains
 - FTS Configurations
 - FTS Dictionaries
 - FTS Parsers
 - FTS Templates
 - Foreign Tables
 - Functions
 - Materialized Views
 - Operators
 - Procedures
 - Sequences
 - Tables (4)
 - artist
 - Columns (5)
 - Constraints
 - Indexes
 - RLS Policies
 - Rules
 - Triggers
 - genres
 - song
 - songs_artists
 - Trigger Functions

Dashboard Properties SQL Statistics Dependencies Dependents Processes Spotify/postgres@PostgreSQL 15*

Spotify/postgres@PostgreSQL 15

Query Query History

```
1 SELECT * FROM public.song
2
```

Data Output Messages Notifications

	id text	name text	popularity double precision	duration_ms bigint	explicit bigint
1	35iwgR4jXetI318WEWsa1Q	Carve		6	126903
2	021ht4sdgPcrDgSk7JTbKY	Capítulo 2.16 - Banquero Anarquista		0	98200
3	07A5yehTShoedVIJAZkNnc	Vivo para Quererte - Remasterizado		0	181640
4	08FmqUhxyLTn6pAh6bk45	El Prisionero - Remasterizado	[null]		176907
5	08y9GfoqCWFoGsKdwojr5e	Lady of the Evening		0	163080
6	0BRXJHRNGQ3W4v9frnSthu	Ave Maria		0	178933
7	0Dd9ImXtAtGwsmsAD6K...	La Butte Rouge		0	134467
8	0IA0Hj8CagYrV1hwhldBH	La Java		0	161427
9	0lgl1UCz84pYeVetn11IGP	Old Fashioned Girl		0	310073
10	0JV4iqw2ISKJaHBQZ0e5zK	Martin Fierro - Remasterizado		0	181173
11	0OYGe21oScKJfanLyM7daU	Capítulo 2.8 - Banquero Anarquista		0	99100
12	0PE42H6tslQuyMMIGRigtb	Capítulo 2.25 - Banquero Anarquista		0	132700
13	0PH9AACae1f957JAavh0I2	Lazy Boi		0	157333

Total rows: 1000 of 586672 Query complete 00:00:03.282 Ln 1, Col 26

pgAdmin 4

File Object Tools Help

Browser

- public
 - Aggregates
 - Collations
 - Domains
 - FTS Configurations
 - FTS Dictionaries
 - FTS Parsers
 - FTS Templates
 - Foreign Tables
 - Functions
 - Materialized Views
 - Operators
 - Procedures
 - Sequences
 - Tables (4)
 - artist
 - Columns (5)
 - Constraints
 - Indexes
 - RLS Policies
 - Rules
 - Triggers
 - genres
 - song
 - songs_artists
 - Trigger Functions

Spotify/postgres@PostgreSQL 15

Query

```
1 SELECT * FROM public.songs_artists
```

Query History

Data Output

	id_song text	id_artist text
1	35iwgR4jXetI318WEWsa1Q	45tit06XoI0lio4LBEVpls
2	35iwgR4jXetI318WEWsa1Q	45tit06XoI0lio4LBEVpls
3	35iwgR4jXetI318WEWsa1Q	45tit06XoI0lio4LBEVpls
4	35iwgR4jXetI318WEWsa1Q	45tit06XoI0lio4LBEVpls
5	35iwgR4jXetI318WEWsa1Q	45tit06XoI0lio4LBEVpls
6	35iwgR4jXetI318WEWsa1Q	45tit06XoI0lio4LBEVpls
7	35iwgR4jXetI318WEWsa1Q	45tit06XoI0lio4LBEVpls
8	35iwgR4jXetI318WEWsa1Q	45tit06XoI0lio4LBEVpls
9	35iwgR4jXetI318WEWsa1Q	45tit06XoI0lio4LBEVpls
10	35iwgR4jXetI318WEWsa1Q	45tit06XoI0lio4LBEVpls
11	35iwgR4jXetI318WEWsa1Q	45tit06XoI0lio4LBEVpls
12	35iwgR4jXetI318WEWsa1Q	45tit06XoI0lio4LBEVpls
13	35iwgR4jXetI318WEWsa1Q	45tit06XoI0lio4LBEVpls
14	35iwor4jXetI318WEWsa1Q	45tit06XoI0lio4LBEVpls

Total rows: 1000 of 17769414 Query complete 00:00:14.764 Ln 1, Col 31

Google Cloud

GestionDatos

Search (/) for resources, docs, products, and more

Search

SANDBOX Set up billing to upgrade to the full BigQuery experience. [Learn more](#)

DISMISS UPGRADE

artistas

QUERY SHARE COPY REFRESH

SCHEMA DETAILS PREVIEW LINEAGE

Filter Enter property name or value

Field name	Type	Mode	Collation	Default Value	Policy Tags	Desc
id	STRING	REQUIRED				
genres	STRING	NULLABLE				
name	STRING	NULLABLE				
followers	INTEGER	NULLABLE				
popularity	INTEGER	NULLABLE				

EDIT SCHEMA VIEW ROW ACCESS POLICIES

*Unsaved ... y 2

RUN SAVE SHARE SCHEDULE MORE Query completed.

1 SELECT * FROM 'gestiondatos2023.ProyPazCristiano.artistas' LIMIT 1000

Query results

SAVE RESULTS EXPLORE DATA

Row	id	genres	name	followers
1	6d290wnKFJWtB8zhoYJhsL	austrian orchestra	Wiener Hofmusikkapelle	
2	7Bwa6tUgdx3Ggylzrj5IA	frevo	Raymundo Santos	
3	63pUdvHMyzLd8qeAXM5uzc	himalayan folk	Thasi Tonsa	
4	1gs36ieZQ6KVoegF2Yqt7U	starogradska	Albuli Vera	
5	070AyOnPin7DvKhwaWylBg	musique touareg	Tribu des Kel Ihair	
6	1Xp79KeI2f8Lb1TB9v0uz	greek clarinet	Vaso Stamatou	
7	1bdQvBdszo8AHVcfNp52s6	rwandan traditional	Rusendezagekwe	
8	02UYa98ucRv7hqHw2hbOUQ	taiwan classical performance	Chih-Yi Chen	
9	20F1THenuwdDoBcXLaUzQM	haitian traditional	Christophe Oxilas	
10	2SHa6Mv3UqSHnuWGRQKXqN	haitian traditional	Cult priest with bell, spectators	

Results per page: 50 1 - 50 of 1000

Google Cloud

GestionDatos

Search (/) for resources, docs, products, and more

Search

SANDBOX

Set up billing to upgrade to the full BigQuery experience. [Learn more](#)

DISMISS

UPGRADE

genres

QUERY

SHARE

COPY

REFRESH

SCHEMA

DETAILS

PREVIEW

LINEAGE

Filter

Enter property name or value

	Field name	Type	Mode	Collation	Default Value	Policy Tags	Description
<input type="checkbox"/>	id	STRING	REQUIRED				
<input type="checkbox"/>	name	STRING	NULLABLE				

EDIT SCHEMA

VIEW ROW ACCESS POLICIES

Q *Unsaved ...y 2

RUN

SAVE

SHARE

SCHEDULE

MORE

This query will process

1

SELECT * FROM `gestiondatos2023.ProyPazCristiano.genres` LIMIT 1000

Query results

SAVE RESULTS

EXPLORE DATA

JOB INFORMATION

RESULTS

JSON

EXECUTION DETAILS

EXECUTION GRA

Row	id	name
1	c37e7c48-aa48-4aaf-8bb0-f84...	buffalo ny indie
2	dc7f5e7f-852b-4f84-a942-682a...	musica occitana
3	d75d98d2-d38a-4bad-8a47-e4...	musikkorps
4	992e4ba5-d7dd-4906-8968-84...	berlin minimal techno
5	6eb90c46-6708-4134-83a6-05a...	carneval cadiz
6	498f68e6-643c-4aa5-a99d-eea...	avant-garde jazz
7	b73d7b38-97ba-4120-8acd-0c6...	j-ambient
8	3ac15e99-a902-4b24-a813-d67...	arkansas hip hop

Results per page:

50

1 - 50 of 1000

SANDBOX Set up billing to upgrade to the full BigQuery experience. [Learn more](#)🏠 ✕ 🔍 Editor ✕ **canciones** ✕ +**canciones** 🔍 QUERY ▼ + 👤 SHARE 📄 COPY ⋮ ↻ REFRESH

SCHEMA DETAILS PREVIEW LINEAGE

☰ Filter Enter property name or value



<input type="checkbox"/>	Field name	Type	Mode	Collation	Default Value	Policy Tags
<input type="checkbox"/>	id	STRING	REQUIRED			
<input type="checkbox"/>	name	STRING	NULLABLE			
<input type="checkbox"/>	popularity	INTEGER	NULLABLE			
<input type="checkbox"/>	duration_ms	INTEGER	NULLABLE			
<input type="checkbox"/>	explicit	BOOLEAN	NULLABLE			
<input type="checkbox"/>	artists	STRING	NULLABLE			
<input type="checkbox"/>	id_artists	STRING	NULLABLE			
<input type="checkbox"/>	release_date	STRING	NULLABLE			
<input type="checkbox"/>	danceability	FLOAT	NULLABLE			
<input type="checkbox"/>	energy	FLOAT	NULLABLE			

EDIT SCHEMA

VIEW ROW ACCESS POLICIES