

Gestión y Almacenamiento de Datos

Spotify

Sergio Masso Giraldo.
Daniel Otero Erazo
Alejandro Parra Palacios





Spotify Data

Problema: Limitada Diversidad en las Recomendaciones Musicales Personalizadas

• La falta de apoyo institucional y la concentración en ciertos géneros musicales pueden llevar a una falta de diversidad en las recomendaciones de plataformas de streaming de música como Spotify.



Impacto:



Los usuarios pueden perder la oportunidad de descubrir música nueva y diversa



Los artistas emergentes pueden enfrentar desafíos para llegar a su audiencia.



Inicio



Extracción Datos



Transformación Datos



Analisis Exploratorio



Analisis Exploratorio 2



Analisis Exploratorio 3



SQL



Solución Problematica

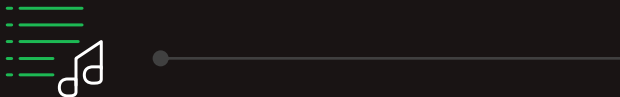
Spotify Data

API

- Creacion Cuenta
- Extraccion Datos
- Canciones, Artistas, Variables

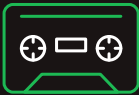
DF

- Creacion del DataFrame
- Union de DF



	artista_id	nombre artista	id_cancion	nombre_cancion	popularidad
0	2LRowlKmiHjgvgdNGBHNo	Feid	7bywjHOCdWsjGGbJ04xbVi	Luna	94
1	1TA5sGRlKUJBNA4ZyuDIX	Blessd	7chTH3GBijAyagf0SPCZDL	Si sabe fenexo	84
2	4q3ewBCK7sLwD24euV69X	Bad Bunny	7IQXYTyuG13aoeHxGG28Nh	Perro negro	91
3	3Me3SAWHCGqW4sZ7bWWJt1	Xavi	0R6NfOILzLj405VbYSJAjf	La diabla	94
4	1r4hU1h58CWwUQe3MxPuuu	Maluma	4bw8mcDUSRWfQo63ZTYRnU	Según quién	87
...
1696	07YUOmWijBTXwIseAUd9TW	Sebastian Yatra	2rCJl9naJYhaqlSf88uM	Pareja del año	70
1697	3OCvS8P2SGYMBvLdzY6g3e	Silvestre Dangond	6f4UPdBOONKJBRqwZGjaI	Cásate conmigo	62
1698	1zo2ucFHzu58HkcmIpSQs	Soley	1hf43vYDCkaqiclywCkGp	Todos mienten	48
1699	6ZvYVrrfpb1Z7kICDyxWQE	DEKKO	1k94CNVliqDsd0qYhkG8iV	12x3	71
1700	3ygJTpJIK7eEeC2EFRI9D	Alcoliryo	00jmsvshmuTF8m3tc7xit	Medellificación	60

1513 rows x 5 columns



artista_id	nombre artista	id_cancion	nombre_cancion	popularidad	imagen_album_url	popularidad artista	genero artista	seguir
0	2LRowlKmiHjgvgdNGBHNo	Feid	7bywjHOCdWsjGGbJ04xbVi	Luna	94	https://i.scdn.co/image/d67616d0000273faadL...	[colombian pop, reggaeton, reggaeton, reggaeton]	91
1	1TA5sGRlKUJBNA4ZyuDIX	Blessd	7chTH3GBijAyagf0SPCZDL	Si sabe fenexo	84	https://i.scdn.co/image/d67616d000027396346...	[reggaeton, urbano latino]	77
2	4q3ewBCK7sLwD24euV69X	Bad Bunny	7IQXYTyuG13aoeHxGG28Nh	Perro negro	91	https://i.scdn.co/image/d67616d0000273761f1c5...	[reggaeton, trap latino, urbano latino]	95
3	3Me3SAWHCGqW4sZ7bWWJt1	Xavi	0R6NfOILzLj405VbYSJAjf	La diabla	94	https://i.scdn.co/image/d67616d0000273750f1c7...	[bad sierrano]	85
4	1r4hU1h58CWwUQe3MxPuuu	Maluma	4bw8mcDUSRWfQo63ZTYRnU	Según quién	87	https://i.scdn.co/image/d67616d0000273d88885...	[latin pop, reggaeton, reggaeton colombiano, L...]	85



ON AIR



Inicio



Extracción Datos



Transformación Datos



Analisis Exploratorio



Analisis Exploratorio 2



Analisis Exploratorio 3



SQL



Solución Problematica

```
<class 'pandas.core.frame.DataFrame'>
Index: 1513 entries, 0 to 1700
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   artista_id            1513 non-null   object
1   nombre_artista        1513 non-null   object
2   id_cancion            1513 non-null   object
3   nombre_cancion        1513 non-null   object
4   popularidad           1513 non-null   int64
5   imagen_album_url      1513 non-null   object
6   popularidad_artista   1513 non-null   int64
7   generos_artista       1513 non-null   object
8   seguidores_artista    1513 non-null   int64
dtypes: int64(3), object(6)
memory usage: 118.2+ KB
None
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1513 entries, 0 to 1700
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   artista_id            1513 non-null   string
1   nombre_artista        1513 non-null   string
2   id_cancion            1513 non-null   string
3   nombre_cancion        1513 non-null   string
4   popularidad           1513 non-null   int64
5   imagen_album_url      1513 non-null   object
6   popularidad_artista   1513 non-null   int64
7   generos_artista       1513 non-null   object
8   seguidores_artista    1513 non-null   int64
dtypes: int64(3), object(2), string(4)
memory usage: 118.2+ KB
None
```



Frente a la data extraída, realizaremos una transformación de esta, con el fin de poder tener el tipo de variable correcta para poder hacer el análisis.



Inicio



Extracción Datos



Transformación Datos



Analisis Exploratorio



Analisis Exploratorio 2



Analisis Exploratorio 3



SQL

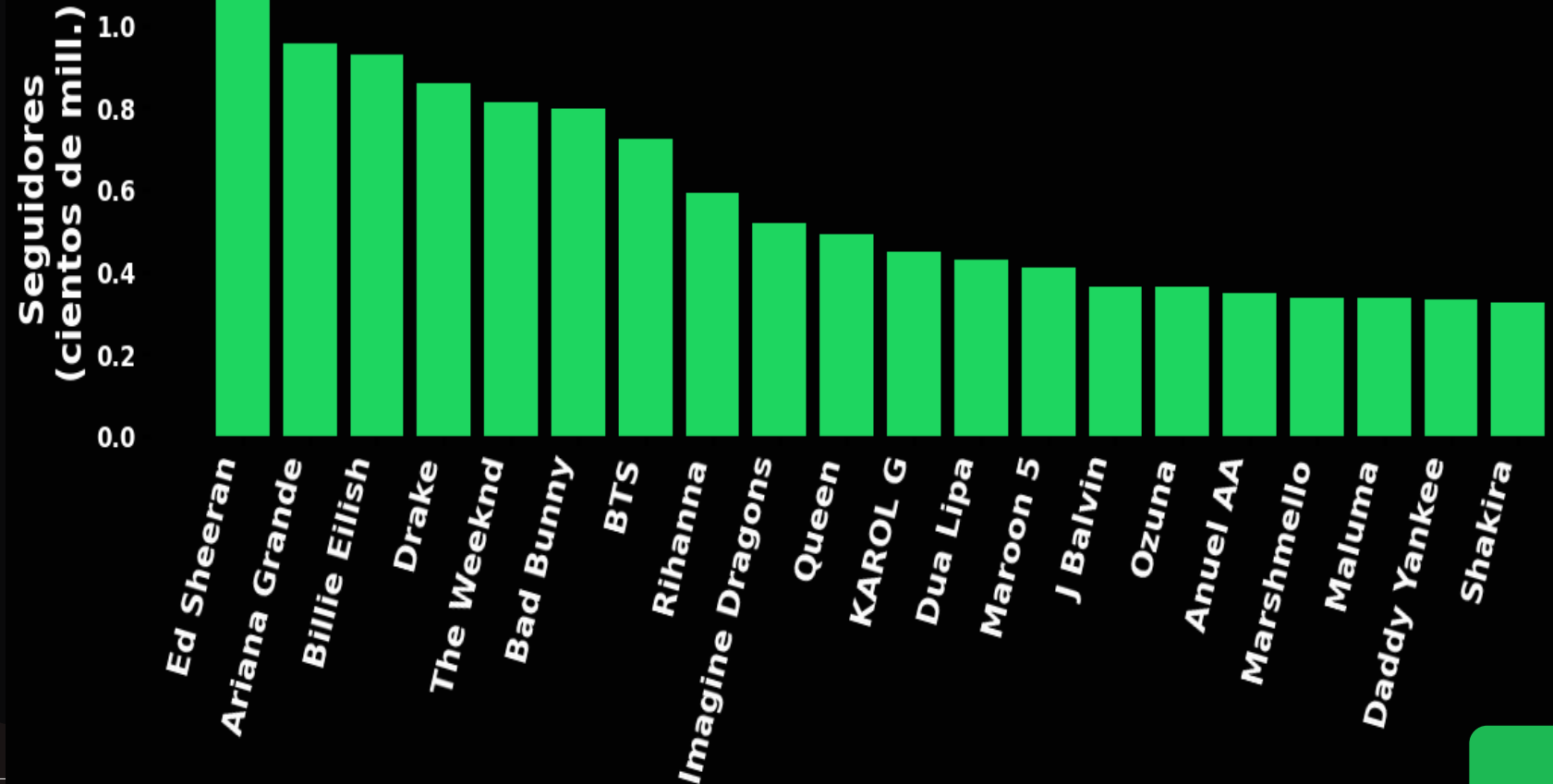


Solución Problematica

Spotify Data



Top artistas (según #seguidores)





Inicio



Extracción Datos



Transformacion Datos



Analisis Exploratorio



Analisis Exploratorio 2



Analisis Exploratorio 3

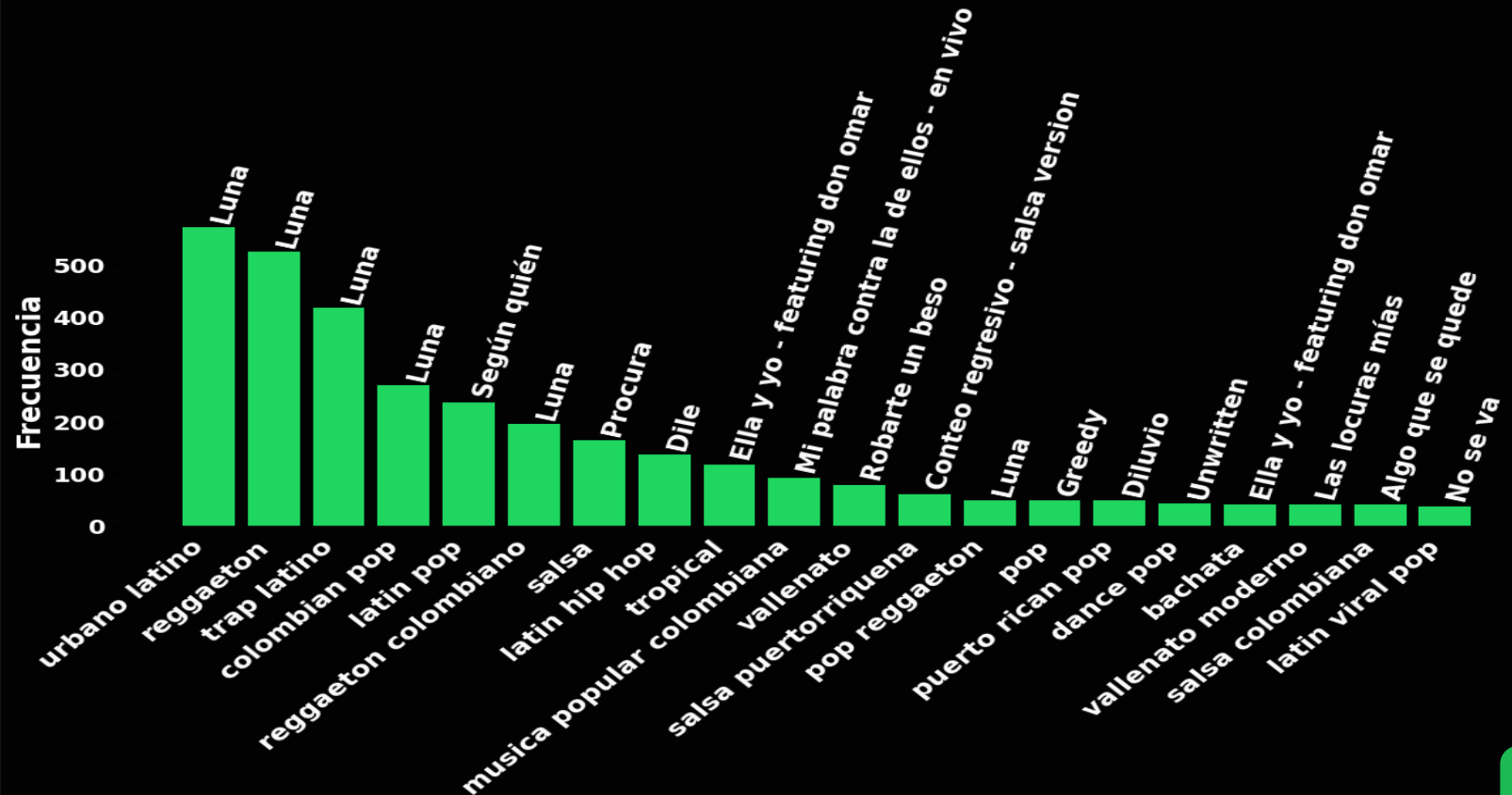


SQL



Solución Problematica

Top artistas por canciones mas populares en Colombia



Gestión y Almacenamiento de Datos



Inicio



Extracción Datos



Transformacion Datos



Analisis Exploratorio



Analisis Exploratorio 2



Analisis Exploratorio 3



SQL



Solución Problematica

Spotify Data



Analisis Exploratorio



Inicio



Extracción Datos



Transformacion Datos



Analisis Exploratorio



Analisis Exploratorio 2



Analisis Exploratorio 3



SQL



Solución Problematica

Spotify Data



Normalización de tablas Ajustes Dataframe Conexión MySQL

```
# conexion a la base de datos  
conexion = mysql.connector.connect(user = 'root', password = '****', host = 'localhost')
```

```
# Creamos el database 'spotify'  
cursor = conexion.cursor()  
query = 'CREATE DATABASE spotify'  
cursor.execute(query)
```

```
# Cambiamos el tipo de dato para poder hacer la relacion (tabla artist)  
conexion = mysql.connector.connect(user = 'root', password = 'chester', host = 'localhost', database = 'spotify')  
cursor = conexion.cursor()  
query = 'ALTER TABLE artist CHANGE artista_id artista_id VARCHAR(255);'  
cursor.execute(query)  
conexion.commit()
```

```
# Creamos las tablas con los dataframes creados  
user = 'root'  
password = 'chester' # Cambiar por la contraseña de su servidor  
host = 'localhost'  
database = 'spotify'  
  
url_de_conexion = f"mysql+mysqlconnector://{user}:{password}@{host}/{database}"  
motor = create_engine(url_de_conexion)  
  
artist.to_sql('artist', con = motor, if_exists = 'replace', index = False)  
tracks.to_sql('tracks', con = motor, if_exists = 'replace', index = False)  
tracks_variable.to_sql('tracks_variable', con = motor, if_exists = 'replace', index = False)
```


Spotify Data

```
# Cambiamos el tipo de dato para poder hacer la relacion (tabla artist)
conexion = mysql.connector.connect(user = 'root', password = 'chester', host = 'localhost', database = 'spotify')
cursor = conexion.cursor()
query = 'ALTER TABLE tracks CHANGE artista_id artista_id VARCHAR(255);'
cursor.execute(query)
```

```
# Creamos las relaciones de las tablas
conexion = mysql.connector.connect(user = 'root', password = 'chester', host = 'localhost', database = 'spotify')
cursor = conexion.cursor()
query = 'ALTER TABLE tracks ADD CONSTRAINT fk_artista_id FOREIGN KEY (artista_id) REFERENCES artist (artista_id) ON DELETE CASCADE ON UPDATE CASCADE;'
cursor.execute(query)
conexion.commit()
```

```
# Creamos las relaciones de las tablas
conexion = mysql.connector.connect(user = 'root', password = 'chester', host = 'localhost', database = 'spotify')
cursor = conexion.cursor()
query = 'ALTER TABLE tracks_variable ADD CONSTRAINT fk_tracks_id FOREIGN KEY (id) REFERENCES tracks (id_cancion) ON DELETE CASCADE ON UPDATE CASCADE;'
cursor.execute(query)
conexion.commit()
```

```
query_2 = 'ALTER TABLE tracks ADD PRIMARY KEY (id_cancion);'
query_3 = 'ALTER TABLE tracks_variable ADD PRIMARY KEY (id);'
cursor.execute(query_1)
cursor.execute(query_2)
cursor.execute(query_3)
conexion.commit()
```



Extracción Datos



Transformación Datos



Analisis Exploratorio



Analisis Exploratorio 2



Analisis Exploratorio 3



SQL



Solución Problemática

Spotify Data

Solución Problemática

- Mejorar los algoritmos de recomendación para incluir una mayor diversidad de géneros musicales.
- Incorporar datos adicionales sobre la diversidad de la música escuchada globalmente.
- Promover activamente artistas emergentes y géneros menos comerciales en los algoritmos de recomendación.

Esta problemática puede abordarse mediante mejoras en los algoritmos de recomendación de las plataformas de streaming de música.





Thank You

