

Primer Concepto -Análisis del estado de las bandas musicales en la escena colombiana

Rodrigo Castillo, Alejandra Campo, Gran Maestro Chamán.

October 22, 2021

Contents

1	Qué base de datos escogieron.	2
1.1	Metodología de la recolección de datos sobre bandas en la escena colombiana	2
1.1.1	SIMUS	2
1.1.2	BOMM (Bogotá Music Market)	3
1.1.3	Bandas o cantantes famosos	4
1.1.4	Sin embargo, lo que se pretende con estos datos es extraer solamente los nombres de las bandas activas en la escena colombiana, pues, posteriormente, a partir de estas bandas se obtendrá la información de spotify, por lo que unificaré los datos.	5
1.2	Merge con Spotify:	6
1.2.1	Concepto	6
1.2.2	Proceso de enriquecer los datos:	7
2	Razón por la cuál elegimos esta base de datos:	7
2.1	1	7
2.2	2	7
2.3	3	8
3	Qué métodos pueden potencialmente aplicarse con dicha base de datos.	8
3.1	Respecto a la comparación:	8
3.2	Respecto a la clasificación:	8
3.3	Respecto al clustering:	8
3.4	Respecto a la reducción de dimensionalidad:	8

1 Qué base de datos escogieron.

La idea de este proyecto es hacer un análisis del estado de las bandas musicales en la escena musical colombiana, sin embargo, aunque existen muchas bases de datos y formas diversas de recolectar datos sobre música, no existe un dataset sólido sobre la música en Colombia que cuente con suficientes features para aplicar análisis estadístico de datos en ellos. La mayoría de las plataformas (como spotify) que cuentan con una API no dan información sobre la región en donde se encuentra un artista.

Sin embargo, no tiene gracia hacer análisis de datos sobre grandes datasets que no tienen influencia sobre los miembros del grupo y sobre los cuales ya existen análisis muy complejos hechos por gente mas experimentada, por lo que el grupo se dispuso a hacer su propio dataframe de la música en Colombia.

1.1 Metodología de la recolección de datos sobre bandas en la escena colombiana

1.1.1 SIMUS

Existen diversos datos sobre bandas colombianas , por ejemplo existe una entidad que se llama el SIMUS en el cuál las bandas musicales que tienen contrataciones tienen que registrarse legalmente, esta base de datos cuenta con 882 bandas que se han registrado legalmente.

```
import pandas as pd
datos_simus = pd.read_excel("bandas_simus.xls")
print(datos_simus.head(10))
print(datos_simus.shape)
```

	Nombre	Tipo Agrupación
0	Ensamble musical casa de la cultura San Alberto	Conjuntos
1	Charlie Rueda & The Jazzmorgans	Agrupaciones de Música Popular Urbana
2	Gente pescaito	Conjuntos
3	los Andes	Conjuntos
4	CORO MUNICIPAL DE MALAMBO	Coros
5	CORO ATRES	Coros
6	A TRES TRIO	Trios
7	CUERPO Y ALMA LATIN ROCK BAND	Agrupaciones de Música Popular Urbana
8	banda escuela 16 de julio	Bandas
9	gaita casa de la cultura de san estanislao	Agrupaciones de Música Tradicional

(882, 6)

En esta base de datos cuenta con las columnas:

- nombre de la banda
- tipo de agrupación
- país(todas son colombianas)
- Departamento del país
- Municipio
- Estado de la banda

1.1.2 BOMM (Bogotá Music Market)

Existe una revista llamada BOMM es una plataforma de promoción y circulación organizada por la Cámara de Comercio de Bogotá, como parte de su programa de apoyo a las Industrias Creativas y Culturales, en esta página existe un catálogo de las bandas que han participado en eventos en el año 2021 (<https://www.bogotamusicmarket.com/Artistas?year=2021>) (no existen mas años y este input es vulnerable a RCE injection por si alguien quiere decirle a los de BOMM). En BOMM no existe propiamente un Dataset, sin embargo, están todas los artistas inscritos, por lo que se puede hacer un simple programa que los scrapee:

```
from selenium import webdriver
import pandas as pd
from time import sleep

#crea una session de un driver de google chrome
driver = webdriver.Chrome()

bomm_artist_url = "https://www.bogotamusicmarket.com/Artistas?year=2021"
driver.get(bomm_artist_url)

#espera 1 segundo para que el driver seguro cargue la página
sleep(1)

not_done = True
bomm_list = []
while(not_done):
    #agarro todos los nombres de una pagina específica
    for i in range(1,17):
        #si no se puede agarrar el elemento es porque no hay mas bandas
        try:
            first_artist = driver.find_element_by_xpath(f"/html/body/div[1]/div/div/di
```


bién scrapeado de <https://www.ranker.com/list/bands-from-colombia/reference>

nota:el script que scrapeo estos datos está en el repositorio pero no se puede llamar desde el notebook pues es un poco mañoso

```
import json
import pandas as pd
with open("famous_data.json") as file:
    data = json.load(file)
famous_dataframe = pd.DataFrame(data)
print(famous_dataframe)
```

```
      name
0    Shakira
1    Juanes
2  Mateo Camargo
3  Alfredo Gutierrez
4  Alex González
..      ...
100  Gabriel Torregrosa
101  Rafael Rodríguez
102  Fredys Arrieta
103  Alejandro Palacio
104  David Escobar Gallego
```

```
[105 rows x 1 columns]
```

1.1.4 Sin embargo, lo que se pretende con estos datos es extraer solamente los nombres de las bandas activas en la escena colombiana, pues, posteriormente, a partir de estas bandas se obtendrá la información de spotify, por lo que unificaré los datos.

```
import pandas as pd
```

```
colombia_scene_list = []
#añado los datos del dataframe que hice de famosos previamente
for i in famous_dataframe["name"]:
    data = {"name":i,
            "lugar_extraido":"pagina_artistas_famosos"}
```

```

colombia_scene_list.append(data)

#añado los datos del dataframe que hice de BOMM previamente
for i in bomm_data["band_name"]:
    data = {"name":i,
            "lugar_extraido":"bomm"}
    colombia_scene_list.append(data)

#añado los datos que descargue del SIMUS
for i in datos_simus["Nombre"]:
    data = {"name":i,
            "lugar_extraido":"simus"}
    colombia_scene_list.append(data)

colombia_state_names = pd.DataFrame(colombia_scene_list)
#
#guardo los datos para vincularlos posteriormente con spotify y con youtube
colombia_state_names.to_csv("nombres_bandas_generados.csv")
print(colombia_state_names.tail(5))

```

	name	lugar_extraido
1239	Semillero escuela Municipal de música de Piendamó	simus
1240	Chirimia del Pacifico Colombiano	simus
1241	ESCUELA DE FORMACION MUSICAL PAEZ VIVE	simus
1242	Pal' Lereo Pabla	simus
1243	Mar a fuera	simus

1.2 Merge con Spotify:

1.2.1 Concepto

Spotify cuenta con una API bastante completa que suministra mucha información sobre las bandas que se encuentran en esta plataforma. Ya que se tiene un dataset con muchos nombres de bandas que suenan en la escena de la música colombiana, ahora lo que busco es enriquecer los datos que tengo buscando los nombres de las bandas en spotify y extrayendo sus features. En un proyecto anterior, junto con Juan Pablo, habíamos hecho una clase que funcionaba dentro de la API de spotify para relacionar nombres de artistas con sus respectivos ids en spotify

1.2.2 Proceso de enriquecer los datos:

previamente se almacenaron los datos en “nombresbandasgenerados.csv”, lo que busco es asociar a cada artista con su respectivo id de spotify con el fin de luego poder extraer la información correspondiente a este

fue muy pesado el script que hacía el merge por lo que me tocó correrlo como un script aparte que está en el repositorio pero no se puede correr desde el documento, de todas formas, los datos sobre los que se va a trabajar el proyecto son estos (aunque luego se podrían adjuntar matrices respectivas a las músicas de cada artista de ser pertinente)

```
import pandas as pd
final_data = pd.read_csv("final_data.csv")
print(final_data.head(4))
print(final_data.columns)
print(final_data.shape)
```

	Unnamed: 0	name_scrapped	id
0	0	Shakira	<built-in function id> https://open.spotify.com/art
1	1	Juanes	<built-in function id> https://open.spotify.com/art
2	2	Mateo Camargo	<built-in function id>
3	3	Alfredo Gutierrez	<built-in function id> https://open.spotify.com/art

```
[4 rows x 12 columns]
Index(['Unnamed: 0', 'name_scrapped', 'id', 'link_spotify', 'followers',
      'generos', 'imagenes', 'popularidad', 'nombre_en_spotify',
      'type_of_artist', 'lugar_extraido', 'index_on_last_dataset'],
      dtype='object')
(1244, 12)
```

2 Razón por la cuál elegimos esta base de datos:

2.1 1

porque junto con Juan Pablo nos divertimos haciendo proyectos relacionados con música ya que siempre descubrimos bandas excelentes. Además, porque tiene sentido estudiar la música que consumimos.

2.2 2

porque el hecho de que no existan datasets sobre el tema implica que es un tema que nadie ha estudiado con rigurosidad, y este análisis le puede servir a

promotores de la industria musical o a bandas de amigos que quieran tomar estrategias para aumentar su popularidad.

2.3 3

porque este proyecto puede revelar un panorama muy interesante de la música en Colombia

3 Qué métodos pueden potencialmente aplicarse con dicha base de datos.

3.1 Respecto a la comparación:

Se pueden usar métodos de comparación para ver que factores influyen en la popularidad de un artista en Colombia

3.2 Respecto a la clasificación:

Se puede intentar aplicar métodos de regresiones logísticas para saber si un artista, según sus features, va a ser popular o no

3.3 Respecto al clustering:

Se puede aplicar modelos de clustering como Kmeans para entender que tipos de bandas suenan en nuestro país

3.4 Respecto a la reducción de dimensionalidad:

Se puede reducir la dimensión de las matrices de los mel-espectrogramas de canciones de los artistas para poder visualizar en un plano un panorama completo de la música colombiana (que en realidad es mi meta personal con este proyecto.) , con esto, tener una visión mas clara de como se mueve la onda musical en Bogotá.