

Spotify SNA

Cynthia Vidal Ovando
Data Engineering
Universidad Politécnica de Yucatán
Km. 4.5. Carretera Mérida — Tetiz
Tablaje Catastral 7193. CP 97357
Ucú, Yucatán. México
Email: 2109155@upy.edu.mx

Didier Omar Gamboa Angulo
Universidad Politécnica de Yucatán
Km. 4.5. Carretera Mérida — Tetiz
Tablaje Catastral 7193. CP 97357
Ucú, Yucatán. México

Abstract

This report presents a detailed analysis using the Spotify API to explore its capabilities in managing music and user data. The tools and methodologies used to obtain, process, and analyze the information are described. The results highlight my listening patterns and musical preferences, identifying popular artists and predominant genres among my music. Additionally, the limitations of the API, such as restrictions on access to sensitive data and request limits, are discussed. The analysis provides a better understanding of how to personalize user experience on music streaming platforms, proposing future research directions to optimize the use of this data.

Index Terms

Spotify API, Music streaming, User data analysis, Playlists management, Listening patterns, Popular artists, Music genres, OAuth 2.0, Data visualization, Machine learning, Personalization, Predictive modeling, Data processing, User experience, Network.



UNIVERSIDAD
POLITÉCNICA
DE YUCATÁN



Spotify SNA

I. INTRODUCTION

IN the following report it intended to cover the process from scratch of how you can perform an analysis for the network on Spotify. It presents a detailed analysis using the Spotify API. The goal is to explore the API's capabilities for managing and retrieving playlists, accessing user profile data, and obtaining information about the top tracks and artists that users listen to. The limitations of the API, such as restrictions on access to personal and sensitive data, and request rate limits, are also discussed.

Spotify is a leading music streaming platform that provides access to millions of songs across various genres and artists. Launched in 2008, Spotify allows users to create and share playlists, discover new music, and follow their favorite artists. The platform has grown significantly, now boasting over 400 million active users globally, with a vast amount of data generated daily by user interactions, including song plays, searches, and playlist creations.

The platform has grown significantly, now boasting over 400 million active users globally, with a vast amount of data generated daily by user interactions, including song plays, searches, and playlist creations.[1]

Spotify API offers functionalities such as retrieving playlists, accessing user profile information, and obtaining details about the most popular tracks and artists. By leveraging this API, developers can build applications that enhance the user experience by providing personalized music recommendations, analyzing listening habits, and more[2]

II. OBJECTIVES

The main goal of this analysis is to demonstrate how the Spotify API can be used to gather and analyze user data to extract valuable insights. These insights can be used to tailor music recommendations, identify trends in music consumption, and understand the preferences of different user segments, in this scenario we will be analyzing my own Spotify user data.

- Extracting data about user preferences, popular tracks, and artists.
- Analyzing listening patterns to identify trends and preferences.
- Exploring the limitations and constraints of the Spotify API in accessing and processing user data.

III. DEVELOPMENT

The methodology for this analysis involved these key steps, using a combination of Python, data processing and the Spotify API for the network analysis:

A. Data Collection

Data was collected using the Spotify Web API, which allows developers to send HTTP requests to access various types of data. The API supports functions such as:

- Retrieving details about playlists, including tracks and metadata.
- Accessing user profile information, such as recently played tracks and top artists.
- Searching for artists, tracks, and albums based on user queries.

To access the Spotify API, it was necessary to authenticate using OAuth 2.0, a secure authorization protocol that enables applications to interact with Spotify on behalf of users without exposing their credentials. Tokens were generated and used to authenticate each API request, ensuring secure access to the desired data.

```
Lana Del Rey: 0.0045
Taylor Swift: 0.0045
Selena Gomez: 0.0045
Bruno Mars: 0.0045
Nicki Minaj: 0.0532
Chayanne: 0.0000
Olivia Rodrigo: 0.0045
Marcos Menchaca: 0.0000
Aventura: 0.2832
Feid: 0.0312
Sam Smith: 0.0045
Adele: 0.0045
Camila Cabello: 0.0045
Cartel De Santa: 0.1754
Sabino: 0.0000
Lng Sht: 0.0000
Duki: 0.0000
Quevedo: 0.0000
Humbe: 0.0000
Chava Salazar: 0.0000
```

B. Get the data)

Here we used functions to get the data based on the Spotify API documentation:

- 1.1 Access Profile Data
- 1.2 Get Followed Artists

	name	degree	centrality
1	Taylor Swift	116068455	
4	Bruno Mars	56901415	
17	Adele	56588816	
3	Selena Gomez	50561026	
8	Olivia Rodrigo	37796740	
0	Lana Del Rey	37208366	
18	Camila Cabello	33561844	
5	Nicki Minaj	31737706	
16	Sam Smith	23976039	
2	Cartel De Santa	16528771	
13	Feid	12397519	
11	Duki	10807730	
12	Aventura	8197095	
7	Chayanne	8013637	
19	Quevedo	4784082	
10	Humbe	1584964	
6	Sabino	909898	
9	Marcos Menchaca	318832	
15	Lng Sht	193548	
14	Chava Salazar	18217	

1.3 Get Top Artists and Tracks

1.4 Get Related Artists

C. Network Analysis

I saved my data on csv files to avoid rate limit and with this data I analysed with this metrics:

- **Centrality Degree**
The degree of centrality measures the popularity of an artist based on how many other artists are connected to it
- **Closeness Centrality**
Evaluates how close an artist is to everyone else in the network. We need to construct a network to calculate this metric.

```
Lana Del Rey: 0.6338
Taylor Swift: 0.6338
Selena Gomez: 0.6338
Bruno Mars: 0.6338
Nicki Minaj: 0.6914
Chayanne: 0.5850
Olivia Rodrigo: 0.6338
Marcos Menchaca: 0.5070
Aventura: 0.7605
Feid: 0.6338
Sam Smith: 0.6338
Adele: 0.6338
Camila Cabello: 0.6338
Cartel De Santa: 0.4753
Sabino: 0.3236
Lng Sht: 0.3236
Duki: 0.4753
Quevedo: 0.4753
Humbe: 0.0526
Chava Salazar: 0.0526
```

- **Intermediation Centrality.**
Measures how many times an artist acts as a bridge in the network.
- **Diversity of neneres**
It analyzed how many different genres the artists follow.
- **Average popularity**
It calculates the average popularity of the artists following and those of the most common genres.

```
Lana Del Rey: 0.0045
Taylor Swift: 0.0045
Selena Gomez: 0.0045
Bruno Mars: 0.0045
Nicki Minaj: 0.0532
Chayanne: 0.0000
Olivia Rodrigo: 0.0045
Marcos Menchaca: 0.0000
Aventura: 0.2832
Feid: 0.0312
Sam Smith: 0.0045
Adele: 0.0045
Camila Cabello: 0.0045
Cartel De Santa: 0.1754
Sabino: 0.0000
Lng Sht: 0.0000
Duki: 0.0000
Quevedo: 0.0000
Humbe: 0.0000
Chava Salazar: 0.0000
```

```
Average Popularity on my artist: 76.25
verage Popularity on my artist in common genres: 79.39
```

- **Shortest path**
It calculates the shortest way from one artist to other based on genres or colaborqtions.

```
No path from Harry Styles to Natanael Cano
```

D. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was performed to understand the underlying patterns and trends within the data. This included:

- Visualizing the distribution of listening habits across different times of the day and week.
- Identifying the most popular artists and tracks among my tracks.
- Analyzing the diversity of genres listened to by different segments.

E. Modeling and Prediction

To predict which artists I might follow based on:

- Based on artist that I like.

```
related_artist      genres
0  Bad Bunny        reggaeton, latin pop, reggaeton
1  Jhay Cortez       reggaeton, latin pop, reggaeton
2  Jhay Cortez       reggaeton, latin pop, reggaeton
3  Jhay Cortez       reggaeton, latin pop, reggaeton
4  Jhay Cortez       reggaeton, latin pop, reggaeton
...
395 Edgardo Nuñez    corrido, corridos tumbados, sad sierrano, sier...
396 Feid              corrido, corridos tumbados, sad sierrano, trap...
397 Alvaro            mexican hip hop, trap, mexican
398 Dan Sanchez      corrido, corridos tumbados, sierrano
399 Feid              reggaeton, latin pop, reggaeton

popularity
0      70
1      71
2      71
3      71
4      71
...
395    72
396    72
397    72
398    72
399    72

[400 rows x 3 columns]
```

- Based on my top tracks.



Fig. 1. Network Artists based on genres

IV. RESULTS

Analysis:

The analysis of my musical network and patterns shows a preference towards pop and urban genres, in contrast to the initial expectation of a predominance of regional Mexican artists (since that is what I have listened to the most in the last few weeks, I guess the application takes into account a bit more history and not just the most recent). This result suggests that, although regional Mexican has a presence in my preferences, the more international and popular genres are the ones that really dominate my musical experience. The artist network also shows how you are connected to a variety of genres and artists, which shows that I have an eclectic and diverse musical taste.

As you might see, of my big surprises was to notice that the group Aventura was a strong bridge in my top artists. As for the highest thought in the Degree of Centrality, I was not surprised that Taylor Swift was the one with the highest thought.

Artists such as Aventura and Nicki Minaj stand out in closeness centrality, meaning they are closer to other artists within the network, facilitating interaction and mutual influence. Artists such as Aventura and Cartel de Santa have high intermediation centrality, suggesting that they act as important bridges between different subgroups within the music network.

V. CONCLUSION

This project provides a deeper understanding of musical tastes and connections between artists, it is also an excellent opportunity for us to develop key skills. Throughout the project, various stages of the data science process were addressed, from data mining using the Spotify API, to network construction and network visualization. In my opinion, I really enjoyed creating the graphs and seeing the weight that the nodes had in terms of music as I am a music lover.

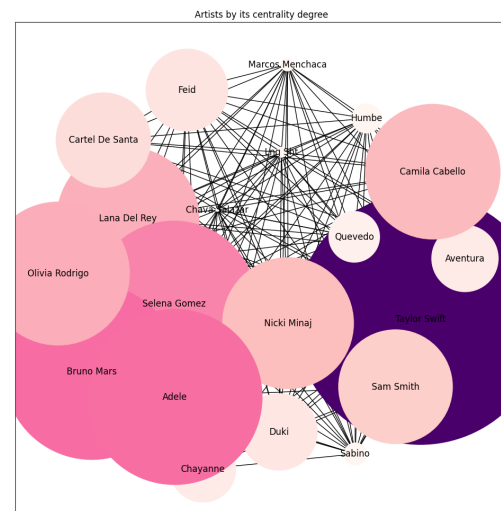


Fig. 2. Artists by its Centrality Degree

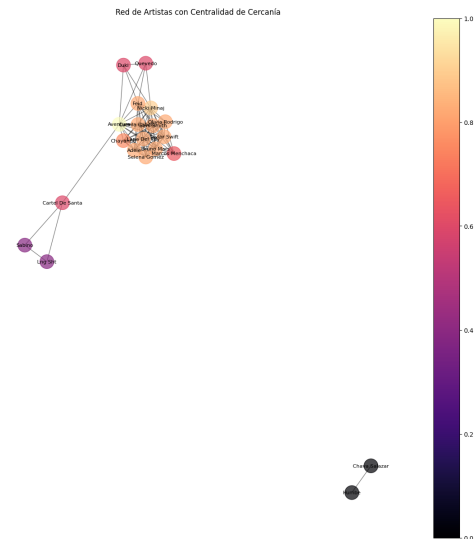


Fig. 3. Clonsnesse Centrality

REFERENCES

- [1] Unknown, *For all of the above, Roma is looking for a better organization in all areas of its business*. *BIEX UNAm*, Mexico
- [2] Emilio, N. (2024, 25 abril). Análisis y visualización de datos con Power BI — BiSmart: Análisis y visualización de datos con Power BI. biSmart. <https://blog.bismart.com/visualizacion-de-datos-con-power-bi>
- [3]

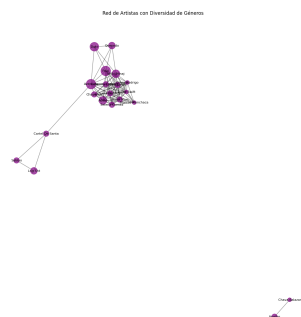


Fig. 4. Generes Diversity

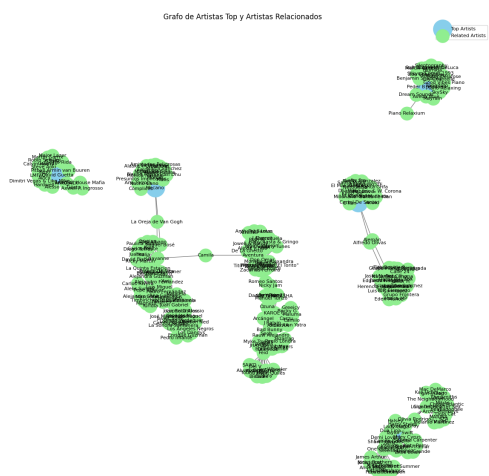


Fig. 5. Projection of artist that I might like