

Investigating The Foremost Factors Determining Artists' Success

Om Fale
ofale1@binghamton.edu
Binghamton University
Binghamton, NY, USA

Disha Shetty
dshetty3@binghamton.edu
Binghamton University
Binghamton, NY, USA

Yashaswi Hasarali
yhasara1@binghamton.edu
Binghamton University
Binghamton, NY, USA

Harshitha Guru Raj
hgurura1@binghamton.edu
Binghamton University
Binghamton, NY, USA

Shubham Patil
spatil36@binghamton.edu
Binghamton University
Binghamton, NY, USA

ABSTRACT

In today's dynamic and ever-evolving music industry, understanding the factors that contribute to an artist's success is of paramount importance. This interdisciplinary project combines the power of social media and data science through the utilization of two prominent APIs, Reddit, and Spotify, to explore the key determinants behind artists' achievements. This project aims to analyze vast datasets from these platforms to uncover patterns, trends, and insights that shed light on the factors influencing an artist's trajectory in the music industry. Within this virtual sea of text data, including posts, comments, and discussions in subreddits like 'r/spotify,' 'r/MusicRecommendations,' 'r/musician,' 'r/singer,' and more, there is a valuable subset of content that revolves around Spotify music, tracks, albums, and artists. We will leverage Reddit data to gain a deep understanding of fan engagement, sentiment analysis, and emerging trends within the online music community. Additionally, we will tap into Spotify's extensive music streaming data to assess metrics such as play counts, follower growth, and playlist placements.

KEYWORDS

Reddit API, Spotify API, MongoDB, Data Collection, Python, REST, ModerateHatespeech, OAuth2

ACM Reference Format:

Om Fale, Disha Shetty, Yashaswi Hasarali, Harshitha Guru Raj, and Shubham Patil. 2023. Investigating The Foremost Factors Determining Artists' Success. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

1 INTRODUCTION

The contemporary music industry stands at a crossroads of unprecedented change, driven by technological advancements, digital platforms, and evolving consumer preferences. In this era of rapid transformation, understanding the intricate factors that determine

an artist's success has never been more crucial. Artists, both emerging and established, navigate a complex web of variables that shape their careers. From captivating crafting melodies to cultivating a devoted fan base and securing lucrative partnerships, the road to success is as diverse as it is challenging. To gain insights into this intricate web, we employ a fusion of social media analysis and data science methodologies, harnessing the capabilities of two prominent APIs: Reddit and Spotify. Reddit, as one of the largest and most diverse online communities, provides a treasure trove of fan discussions, reviews, and opinions. We delve into this wealth of data to discern patterns in fan engagement, sentiment analysis, and emerging trends within the music community. In parallel, we leverage Spotify's vast music streaming data, which encompasses play counts, follower growth, and playlist placements, offering a window into an artist's performance in the digital realm. In this project, we'll be gathering real-time music-related data to determine what musical style or performer is popular and trending when - we'll combine all this information from these sources within a certain time frame.

2 WHAT IS SPOTIFY AND REDDIT API?

The Spotify API is a set of tools and protocols provided by Spotify, a popular music streaming service. It allows developers to access and integrate Spotify's music catalog, user data, and various features into their own applications or services. Through the API, developers can retrieve information about tracks, albums, artists, playlists, and user profiles. They can also perform actions such as playback control, playlist management, and user authentication, enabling the creation of diverse music-related applications, ranging from custom players to recommendation systems. The Spotify API operates on a RESTful architecture, making it accessible over standard HTTP protocols, and it requires authentication through OAuth2 for secure interactions with user data. Developers use the API to enhance their applications with Spotify's vast music library and to create more interactive and personalized music experiences for users.

The Reddit API is a set of tools and protocols provided by Reddit, a popular social media platform. This API allows developers to interact with Reddit's platform programmatically, enabling them to retrieve and submit data, such as posts, comments, and user information. Through the API, developers can create applications or services that integrate with Reddit, offering functionalities like accessing subreddit content, submitting posts, fetching user details, and more. The API operates on a RESTful architecture, making it accessible over standard HTTP protocols. Authentication is typically

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA

© 2023 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

required, and Reddit often employs OAuth for secure interactions. Developers leverage the Reddit API to build diverse applications, including custom Reddit clients, data analysis tools, and bots that automate certain interactions on the platform. Overall, the Reddit API facilitates the integration of Reddit's features and content into external applications, enhancing the platform's accessibility and functionality for users.

3 DATA SET DESCRIPTION

3.1 Data Source

We will be using Reddit data from Reddit Stream API wherein we'll be collecting information about top 10 streaming songs and the artists and for this we will be using Spotify API.

3.2 SPOTIFY API

We will be using the Spotify API, to check for fetch the data from a Spotify URL. We will extract a unique track ID and send that as a request parameter to query the Spotify URL and in response we will get the comprehensive data about the artist's name, album, popularity, ranking, followers, and such. Later, with the help of this data we can check the posts related to the artists that are currently trending on Reddit. We can use this information to determine which artist are trending on reddit post, the no. of discussions happening to the related artists posts and the trending songs.

GET v1/playlists/37i9dQZF1DXcBWIGoYBM5M: fetches the top 50 songs on the playlist provided by Spotify (this depends on several factors such as monthly listeners, popularity index, number of minutes the song played etc.)

GET v1/artists: fetches us the artist details whose songs are currently trending on Spotify. We will only be using the genre the artist follows and the popularity.

3.3 Reddit API

The Reddit REST API provides access to user-generated content on reddit.com, including discussions and stories that have been rated by the community. We plan to use this API to collect user reviews and opinions about songs trending on Reddit, gaining insights into public sentiment and engagement. The API also offers advanced features such as access to user account information and sub-reddit moderation, enabling us to understand user demographics and target specific music-related communities for analysis. This in general, gives data and insights about the trends and discussions on the platform. We have collected Reddit data using https://www.reddit.com/api/v1/access_token endpoint and are using subreddits r/Music, r/Spotify and r/popheads to collect pertinent data that can be stored in the database. From Spotify API, we fetch the artists name and then pass it to the reddit API https://oauth.reddit.com/r/artist_name/ to get the number comments from the posts we get as response from the API.

3.4 API Methods

We will access data resources through standard HTTPS requests to API endpoints. These APIs primarily use HTTP methods like GET and POST. Reddit APIs allow us to find, retrieve, and engage with various resources, primarily utilizing the GET method for data

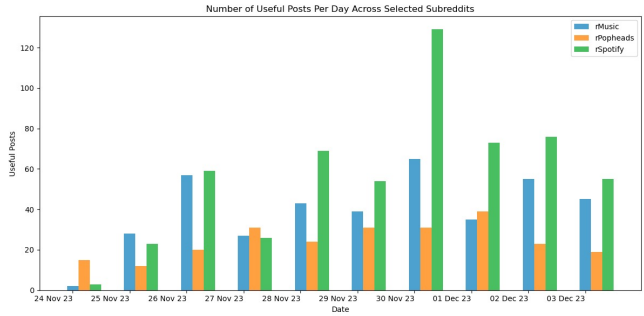


Figure 1: Useful Posts coming from Data Stream after filtering

retrieval. Spotify follows REST principles, where data is retrieved using GET requests, and the response data is typically in JSON format.

3.5 DATA EXPLORATION

For this part of the project, we collected data from all three main subreddits and from the artist's official subreddit page starting from 25th Nov 2023. During this duration, we also noted down the top_hits that changed from the Spotify API. We used Factory for scheduling the number of posts that come in from all the subreddits and have successfully stored it in the database. Additionally, we applied scheduling for the number of comments that people made, saving them in the database. We filtered and cleaned useful posts related to our case study, obtaining sufficient data for our research and saving it in our database. This process allowed us to discover relevant facts and insights concerning the three major purposes of this project with the aid of combined data from two separate sources and certain data science tools and approaches. We used MongoDB as a database tool for storing posts from Reddit and matplotlib for plotting graphs. For data cleaning and analysis, we utilized NumPy and Pandas. The requests library was used to fetch data from Reddit and Spotify APIs. Along with these, we also used date, pymongo, tabulate, plotly, spotipy, etc., libraries to complete the project.

4 BACKGROUND AND RELATED WORK

The proliferation of the internet has given rise to various music distribution platforms, particularly music streaming services, which have significantly increased the availability of music for regular consumption. These platforms not only offer a vast library of songs but also accumulate valuable user data, often treated as confidential by the companies. Despite the abundance of datasets related to music, such as the million-song dataset (MSD), they often become outdated. Recognizing the need for current and openly accessible data for developing and evaluating recommender systems, especially for emerging artists, we identified an opportunity. Many users share their music preferences on platforms like Spotify and engage in discussions on Reddit. Leveraging the free Spotify and Reddit APIs, we collected real-time data on the latest music trends and artists. This approach addresses the challenge of obtaining up-to-date information for the development and assessment of recommender

Artist	Song
Tyla	Water
Jack Harlow	Lovin On Me
Tate McRae	greedy
	exes
Dua Lipa	Houdini
	Dance The Night - From Barbie The Album
Beyoncé	MY HOUSE
Kenya Grace	Strangers
SZA	Snooze
	Kill Bill
Taylor Swift	Cruel Summer
	You're Losing Me (From The Vault)
	Is It Over Now? (Taylor's Version) (From The Vault)
Jung Kook	Standing Next to You
	Seven (feat. Latto) (Explicit Ver.)
Bad Bunny	MONACO

Figure 2: Top 10 Trending Artists and their Songs given by Spotify API for 30th November 2023

systems in the dynamic landscape of the music industry. By utilizing a dataset derived from these sources, we implemented and evaluated a music analysis system based on filtering methods. Our findings indicate that this approach holds promise for capturing and understanding the evolving landscape of music preferences and trends in real-time.

Earlier we just collected posts from all the subreddits, but we have made the modification of also collecting the comments from subreddits r/spotify, r/popheads and r/Music (Changes done in the implementation of reddit.py file). r/spotify provides a lot of data which helps us a lot for the analysis which will provide us with more robust data for trending artists in music. We used cloud generators like word cloud to find the words related to music and on those, we tried to find the related subreddit.

As a requirement, we have also added moderate hate speech API, to find out the toxicity in the comments and get sentimental analysis of the posts and comments; create a word cloud although we don't have a lot coming out, but it does help us to understand what kind of language is used by the people for positive, negative, and neutral sentiments for discussions – anything related to their opinion about any song or the genre they like).

5 DATA FLOW

We first retrieve the data from the Spotify and Reddit APIs using those respective APIs. We save the relevant information in mongo DB database. We will utilize the data such as artist's information, genre, trending songs and albums, music information for analysis and visualization. We will use python "requests" library to fetch the information from the API and we will use the same for authorization of the respective APIs required. Using this track ID, we can then send a GET request to the Spotify API to obtain all the track

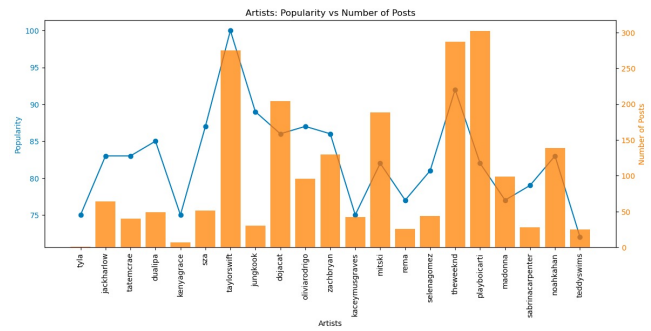


Figure 3: Artists - Popularity vs Number of Posts

information, including the artist, album, genre, and other details that we will need to evaluate our data. Using certain keywords, we will get the information from the Reddit.

6 OBJECTIVES

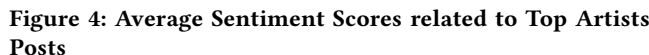
Our project aims to gain a deeper understanding of people's music preferences, identifying trending artists and albums, and exploring how these trends evolve over time. A primary focus is placed on analyzing the top trending artists on Reddit in conjunction with the data collected from Spotify. Our approach involves selecting the top artists and music on Spotify, examining their popularity on Reddit, and investigating the factors contributing to their success. Through this comparative analysis, we seek to uncover insights into the alignment of music trends between Spotify and Reddit, shedding light on the key factors influencing the success of trending artists.

7 METHODOLOGY

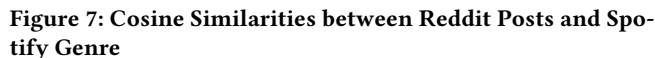
We are going to perform following analysis on the following dataset: Using data set we have collected in the database; we will be performing descriptive analysis with help us give a patterns and connections. We will retrieve details from the playlists API which we have selected - the daily top hits; and get hit playlists, artists details and music genres. On that we will also collect artists details and their associated genres and will do a descriptive analysis of what genre is currently followed and listed to in the given time frame. For comparison we will also gather details from the subreddits such as r/Music, r/Spotify and r/popheads to get to know what genre related discussions are going on at the same time frame. We will also be looking at the current posts on the artists official subreddit pages and investigate which tracks, albums and playlists are trending and the number of comments on those particular posts. We will analyze multiple factors like genre, popularity of artist and tracks and by collecting all the data related to the popular artists which we will plot a graph, which will give us a clear clarity onto change in trade with time.

8 DATA VISUALIZATION

Now that we have collected ample of data from various subreddits for a specific time, now it's time for us to start playing with the data and answer some questions. **Figure 1** shows the same that we have started collecting the data from November 25th around



Additionally, these posts are cleaned for our research purposes to analyze and answer the questions. Figure 3 shows the dual axis graph of artists popularity and the number of posts they get from their official subreddit pages which gives us the correlation between the number of posts in their official subreddit page and the popularity of each trending artists. We have used Pearson Coefficient to get the sentiment analysis on the posts related to the artists - **Figure 4** gives us the sentiment analysis of posts for each artist. Our second question also answers about the genre that is most discussed and songs recommended on Reddit. We have gathered all the posts and have performed scikit-learn metrics and extraction to extract the songs and the genre which is most discussed on Reddit Platform. **Figure 5** explains the artists genre of songs which are trending in spotify from the selected playlist. As we can see 11 artists have released pop genre which are currently on the charts and have led several discussions on Reddit. **Figure 6** gives us the occurrences of genre in the subreddit posts and the cosine similarities between the top 10 posts and the spotify genre (**Figure 7**) and also generate a word cloud as well (**Figure 8**) Moving ahead, we also want to answer the bigger question: who will be included in the Spotify



We have also done a correlation regarding the top trending artists on Reddit and check whether it does get align with the Spotify List

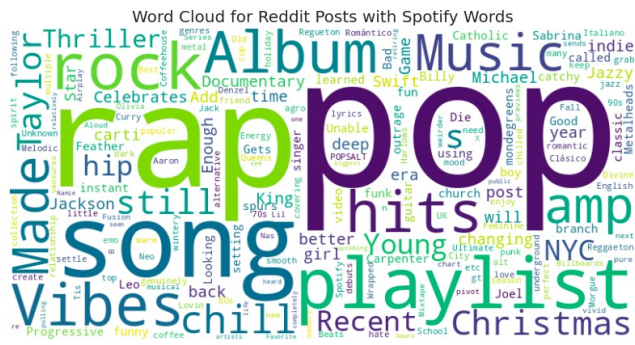


Figure 8: World Cloud for Reddit Posts with Spotify Genre

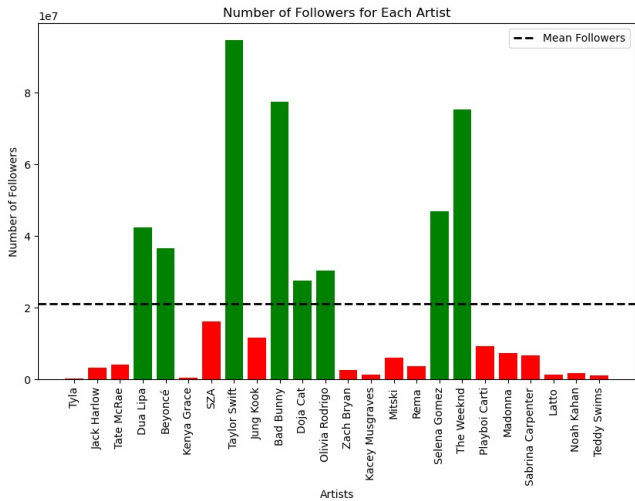


Figure 9: Number of Followers for each artist.

and for some artists it does. The correlation coefficient does match up and concludes that the analysis is positive.

We are also performing a sentiment analysis of Reddit Comments and check the positive, negative, and neutral reactions of the posts by using Moderate Hate Speech and generate a Word Cloud of the Hate Speech or the comments which are flagged.

Moving ahead, we also want to answer the bigger question: who will be included in the Spotify 2023 Top Artists list. We will perform a predictive analysis of the trending artists as to who make it into the year-end Spotify Top Artists Charts. Several factors here come into picture (fan engagement, popularity index, number of posts on their official subreddit page, use of words or let's go with discussion of that particular artists) and with the data we have available we have managed to answer this question and this will also help us answer how the above metric impact the music industry by plotting a aggregate sentiment analysis on reddit comments over the time (Using VADER) and the language or sentiments which are used during discussion using word cloud.

We have also done a correlation regarding the top trending artists on Reddit and check whether it does get align with the Spotify List

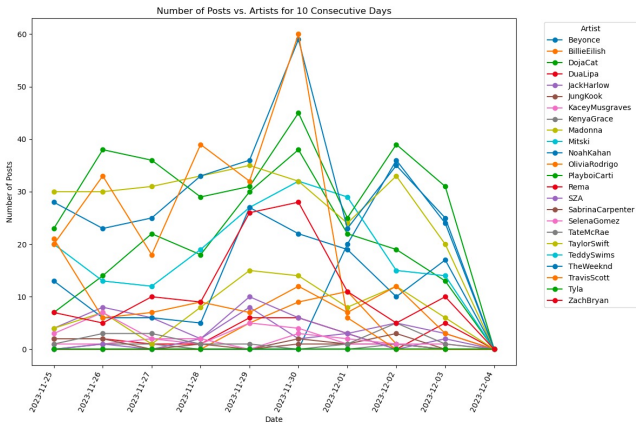


Figure 10: Number of Posts for the top trending artist for 10 consecutive days

Word	Frequency
song	202
spotify	199
songs	176
wrapped	158
music	146
taylor	135
new	118
top	110
tour	96
like	93

Figure 11: Common words found in Reddit Posts

and for some artists it does. The correlation coefficient does match up and concludes that the analysis is positive.

We are also performing a sentiment analysis of Reddit Comments and check the positive, negative, and neutral reactions of the posts by using Moderate Hate Speech and generate a Word Cloud of the Hate Speech or the comments which are flagged.

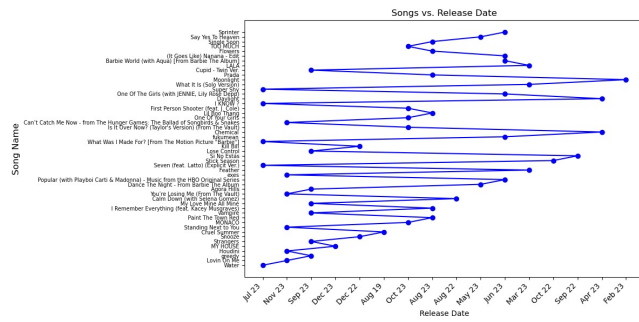


Figure 12: Trending Songs and their release dates

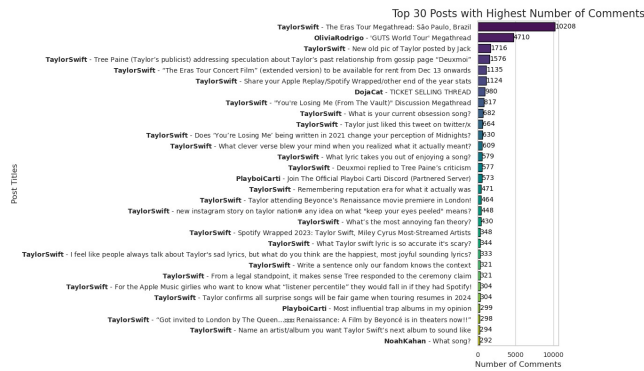


Figure 13: Top 30 Posts with highest number of comments.

Word	Frequency
song	202
spotify	199
songs	176
wrapped	158
music	146
taylor	135
new	118
top	110
tour	96
like	93

Figure 14: Common words found in Reddit Posts

9 LIMITATIONS AND FUTURE WORK

This work has been limited in many angles. Predictive analysis will only be true for certain artists but not entirely. It is majorly due to the data set and the time duration (limited time) we have set and

Artist	Popularity
SZA	87
Eminem	88
The Weeknd	92
Lana Del Rey	88
Ariana Grande	87
Kendrick Lamar	86
Kanye West	89
Coldplay	86
Zach Bryan	86
Travis Scott	90
Bad Bunny	95
Taylor Swift	100
Justin Bieber	87
Jung Kook	89
Olivia Rodrigo	87
Drake	95
Billie Eilish	86
Doja Cat	86

Figure 15: Trending Artists on Reddit Platform using Spotipy

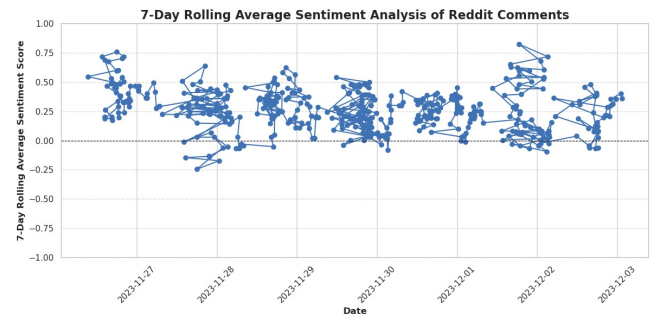


Figure 16: 7 day Rolling Average Sentiment Analysis of Reddit Comments

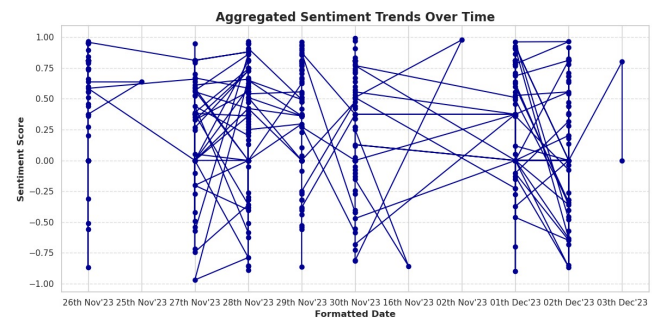


Figure 17: Aggregate sentiment trends from 26th November

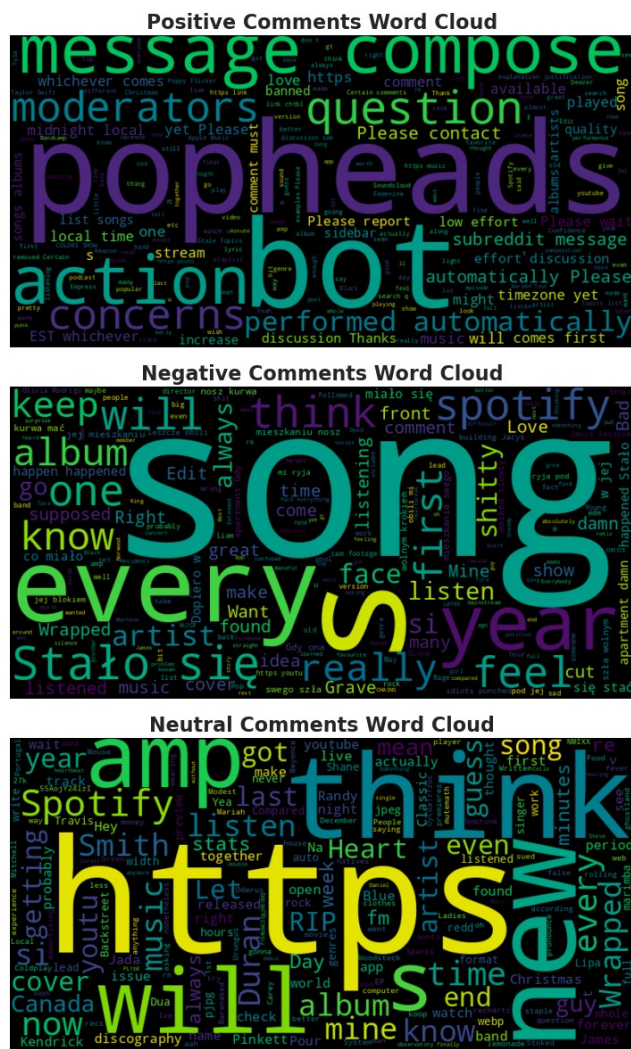


Figure 18: World Clouds with 3 different sentiment analysis on Posts

there are many external factors which can be taken into consideration such as Media Presence, Global Appeal, marketing campaigns, album releases, or external events and many more but due to the tools and the API we have used, there are many such outliers which can change the accuracy of the output. Furthermore, social media platforms such as Twitter, Instagram are more widely used for such discussions than reddit so the dataset might be little less than expected. Nevertheless, all the above factors can be taken into consideration by future studies.

10 CONCLUSION

In summary, we find that:

- 1) Is there a correlation between the level of activity on artist-specific subreddits (e.g., official subreddits) and the artist's success on Spotify, measured by metrics such as number of top hits and overall popularity?

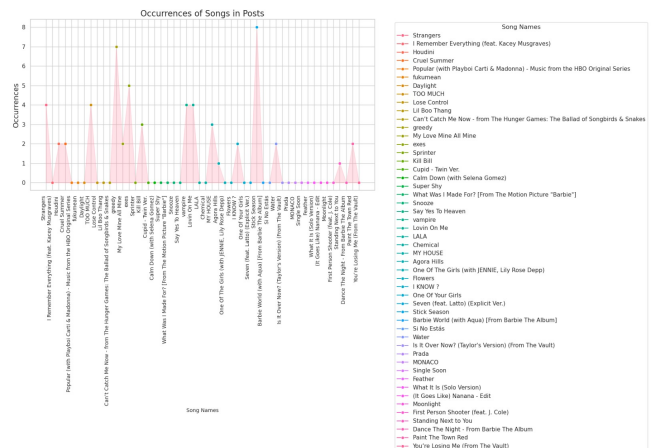


Figure 19: Trending songs on Reddit

-From the Spotify API - after plotting the popularity and the number of posts we can see that for majority of the artists have a correlation and it goes hand in hand with the number of posts and the popularity. It maintains up to the mark on the other hand, artists such as Jungkook (Seven) and Tyla (Water) do not have much popularity and activity in their official subreddit due to various reasons.

- 2) How do the music genres discussed on Reddit correlate with the genres that users listen to on Spotify, and can we identify any trends in genre preferences from these discussions?

-Based on post frequency top genres were retrieved to calculate the occurrences of the songs and the genre in the post. This not for all subreddits go hand in hand as there are a lot of users who use reddit and have discussions on country music, rock and many such songs which even we haven't heard of. But according to our analysis after plotting graph, we have noticed that pop and rap songs have been in major discussion and have ruled with maximum number of comments.

- 3) To what extent can the engagement metrics from Reddit (e.g., post popularity, comments, sentiment) and Spotify (e.g., daily top hits, artist details, popularity index) predict the commercial success and overall popularity of artists, and are there specific patterns or trends that indicate a higher likelihood of success within the music industry (sentiment analysis on the comments)?

-This research question took an extensive analysis of understand the data, cleaning, plotting, and getting to know the pattern as to why that takes place. Sentiment analysis of the comments and Word cloud has helped us in this to answer the question and after thorough analysis we have seen that for all the factors that have been taken into consideration for answering this question established artists like Taylor Swift, Travis Scott, The Weeknd, Drake and SZA will make the year end top trending artists for Spotify.

11 REFERENCES

- (1) Reddit Documentation.
<https://www.reddit.com/dev/api/>
- (2) Spotify API Documentation.
<https://developer.spotify.com/documentation/web-api>

