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 utilisation of two prominent APIs, Reddit and Spotify, to explore the key determinants behind artists' achievements. This project aims to analyse 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

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 crafting captivating 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 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.

2.1 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 on, 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 happing to the related artists posts and the trending songs. We have collected Spotify data using https://api.spotify.com/v1/playlists/id endpoint (note that id used here is 37i9dQZF1DXcBWIGoYBM5M (for Todays Top Hits - this is generated by Spotify with respect too maximum likes for the playlist). Later we take the artist ID from this API response and then send it as the request parameter of the following artist API end point https://api.spotify.com/v1/artists/id to get the popularity the artist and later we can go ahead with the sentimental analysis with respect to the reddit posts.

2.2 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 tar- get 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 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 utilising the GET method for data retrieval. Spotify follows REST principles, where data is retrieved using GET requests, and the response data is typically in JSON format.

4 IMPLEMENTATION

In this project, we will be repeatedly fetch data from Reddit API from subreddits r/Spotify, r/Music and r/popheads. The subreddits were chosen considering the overall factors such as the number of subscribers (i.e. the size of the audience who take part in the discussion), number of comments (I.e. people engaging in the conversation) as well as the discussion style for each subreddit. Let's take an example for r/Music; it was selected considering its large number of subscribers and being Reddit's 1 music community and with ranking within top 10 with respect to size. r/Spotify is used especially for our project purpose as the second source of data is from Spotify application. This subreddit allows the users to share their Spotify playlists, the unique style of discussion it offers and certain criteria it follows to be in the group. r/popheads is selected for analysis of artists with pop music genre as it has been ruling the top 50 chart, despite with small audience. Additionally, we will be keeping tack of the daily counter in the MongoDB database and no data gathering is required for Spotify API.For Spotify database we will be using the playlists API with gives us Top Hits of the day list by providing the playlist ID. Using this playlist ID, we then send a GET request to the Spotify API to obtain all the top hit with parameters such as artist ID, artist name, song name, and album name which will be useful to evaluate this data. We are using faktory for scheduling purposes to get data from Reddit API and store it in the database. Then, once stored we will be using this dataset, to use for our analysis purpose. In addition to that, we are also updating the MongoDB database with the number of posts which will be analyzed every second. Currently, we have not filtered out any data
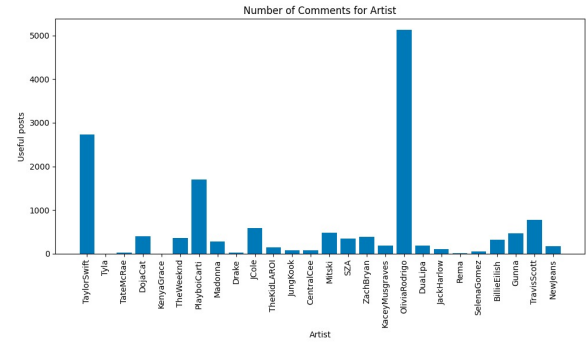

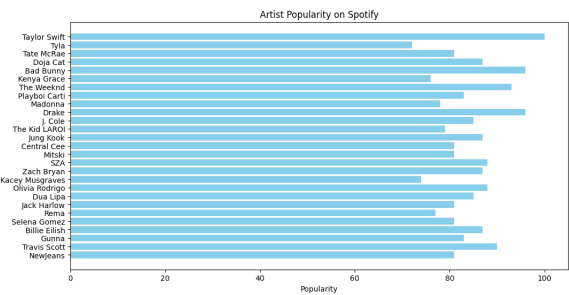**Figure 1: Graph for Number of Comments in the Artists official Subreddit page**

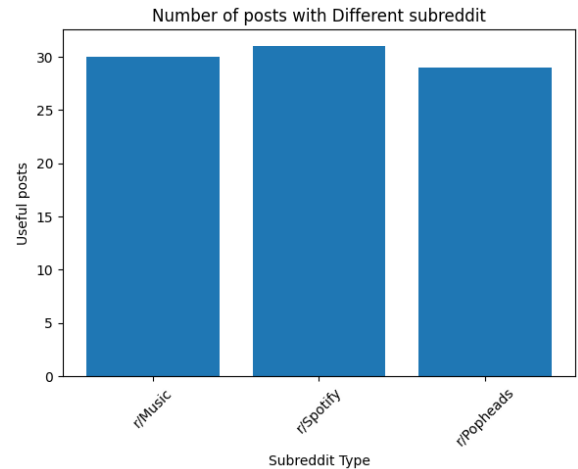**Figure 2: Trending Artists Popularity in Spotify**

Figure 3: Bar Graph for For useful posts against selected subreddit

from the reddit posts as majority of posts we get from reddit are useful to us.

5 MEASUREMENTS AND ANALYSIS

We will be analysing data to identify trending artists or genres based on their popularity on Reddit social media platforms. We'll gather

data from both Reddit and Spotify APIs and continuously update our database in real-time as new information becomes available. To analyse the collected data, we'll periodically retrieve it from the database and create graphs or charts using Python's plotting and visualisation libraries. Additionally, we plan to conduct sentiment analysis by assessing positive and negative reactions in tweets related to trending artists or albums.Additionally, we plan to conduct sentiment analysis by assessing positive and negative reactions in posts and comments related to trending artists or albums.

6 PRELIMINARY EXPLORATION OF DATA

The information is all text and relevant posts are needed from subreddits such as r/popheads and r/spotify. We use the playlist ID of Todays Top Hits and send GET queries to the Spotify API and save it in the "artists_posts" collection. Once refreshed the number of comments increases depending on how active the discussion is going on with respect to that particular post. Similarly, we have also done for artists API and store it in database as "artists_details" collection An individual ID, the post content, the URL object and the number of comments on that post will all be contained in each MongoDB item. Additionally, we have a MongoDB storage location for the daily count as well.

7 DATA COLLECTION ESTIMATES

After the preliminary testing, not every post from the r/Music is essential for use (we might use it in the coming stages of this project). Only about 20 per posts are genuinely useful for gathering information that will be later used for the analysis.After successfully deploying on VM, now we are able to process a number of posts per day and are storing in the MongoDB database. Just 20 per of posts that include relevant terms are collected after 50 posts/hour, which can ultimately be used to extract significant data. For official subreddits of the artists, we get in approximately 20 posts in 10 minutes for the top 10 trending ones (depending on many factors such as new album release, tour dates released, number of streams etc.) So, with analysis of comments we check the trending posts or discussions against the singer and check who are trending on Reddit.

8 DATA FLOW AND IMPLEMENTATION

We first retrieve the data from the Spotify and Reddit APIs using those respective APIs. We save the relevant information in mongoDB 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 information, including the artist, album, genre, and other details that we will need to evaluate our data. Using certain keywords we will get the informations from the Reddit.

9 CHANGE COMPARED TO PROPOSAL

We have used relevant API endpoints of Reddit and Spotify. After looking through multiple subreddits we have finalized 3 subreddits which can actually give us the data we will be using it for
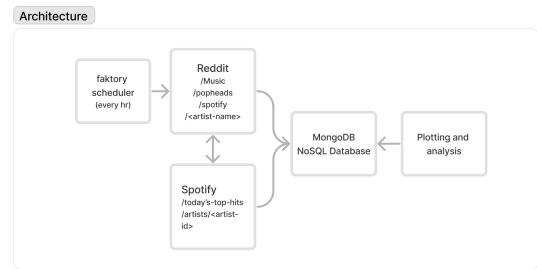

Figure 4: System Diagram of Data collection system

analysis purpose in this project. We have changed used r/Spotify, r/popheads and r/Music which will help us to get more relevant data.Additionally, we will be taking first 30 trending artists from Spotify API and look into their official subreddits page (if they have one; some of them don't have one) and display the number of useful posts and the comments under that post in the MongoDB database.

10 REFERENCES

Martin Pichl, Eva Zangerle, Günther Specht.Combining Spotify and Twitter Data for Generating a Recent and Public Dataset for Music Recommendation,2014.

Data Study on Popular Artists, Weijie Deng, Mar 27, 2018.

Greasley AE, Lamont AM. 2006. Music preference in adulthood: why do we like the music we do? In Proc. of the 9th Int. Conf. on Music Perception and Cognition (ICMPC9), Bologna, Italy, 22–26 August 2006 (eds M Baroni, AR Addessi, R Caterina, M Costa), pp. 960–966. Society for Music Perception Cognition/European Society for the Cognitive Sciences of Music.

Bhattacharjee S, Gopal RD, Lertwachara K, Marsden JR, Telang R. 2005. The effect of P2P, file sharing on music markets: a survival analysis of albums on ranking charts. NET Institute, Working Paper No. 05-26.