

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, Plotly, Jinja, Flask, Pandas, Matplotlib

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, 5 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

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

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/nnnnnnn.nnnnnnn>

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.

For this project, we have created an interactive dashboard that will be updated in real-time, showing results for the two research questions/areas based on our datasets and the type of data we have gathered. Our research objectives shall revolve around the factors influencing the popularity of Trending Artists on Spotify based on the data that we have collected in the from Nov 25th and comparing with the posts and discussions from the selected subreddits.

2 WHAT IS SPOTIFY AND REDDIT API?

The Spotify API [2] 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 [1] 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 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

Spotify and Reddit are the datasets that have been considered for collecting, analysis and visualization of the data for gathering 5000 useful posts from artists official subreddits and around 3000 useful posts regarding the trending artists on Spotify where the list has been changed every 3 to 4 days. In case of Tate McRae, who has been on the top hit's charts for 3 consecutive weeks.

3.1 Collection Methodology

We were successful in gathering pertinent data from the Spotify and Reddit platforms as part of Part 1 and 2 of the implementations. We had gathered information from all the mentioned platforms and filtered it as needed to create a viable data frame for analysis which is described below.

3.1.1 Spotify : We have used 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 artists 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 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.

3.1.2 Reddit: 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 collected Reddit data using the https://www.reddit.com/api/v1/access_token endpoint and focused on subreddits r/Music, r/Spotify, and r/popheads to gather relevant data that can be stored in the database.

From the Spotify API, we fetch the artist's name and then pass it to the Reddit API (https://oauth.reddit.com/r/artist_name/) to retrieve the number of comments from the posts we receive as a response from the API.

4 METHODOLOGY FOR DATA ANALYSIS

4.1 Spotify

- Firstly, a connection is established with MongoDB to retrieve and start the pre-processing of the available data.
- Since the list is the top 50, only the first 30 artists are processed.
- In the formed data frame, the independent and dependent variables are as follows:
 - Independent Variables: Artist Name, Posts Count, Images
 - Dependent Variables: Popularity, Followers
- After forming the data frame, pre-processing is performed.
- Following pre-processing, model training and testing are conducted for the data frame, which contains 58 columns.
- The data frame is divided into training and testing sets, and supervised learning modeling is employed, where a training set is used to instruct models to produce the desired results.
- Graphs are plotted to predict popularity based on other independent factors.
- An interactive feature is added, allowing the user to click the artist name retrieved from the MongoDB database. The artist name is auto-populated, displaying popularity and post count. Additionally, users can discover trending artists on that day.

4.2 Reddit

- For each of the subreddit data stored we have retrieved posts from 3 subreddits in addition to the official subreddit pages followed by formation of data frame.
- Further for descriptive analysis, number of comments were also collected from the 3 selected subreddits
- Lastly, data analysis was done for the selected research questions and a time series was plotted to determine the posts and comments on a particular date.

5 METHODOLOGY FOR CREATING DASHBOARD

For creating live data dashboard, there are several tools, library packages, web applications and software's available. From our data analysis project, we have answered and visualized the results of all our research questions that were: For this project, we have thought about and have answered three research questions/areas based on our datasets and the type of data we have gathered. Our research objectives shall revolve around the factors influencing the popularity of artists based on data collected from Reddit and Spotify. We

already have collected relevant information from project 1 using posts of subreddits r/Spotify, r/Music, r/Popheads and official subreddits of artists data. As far for Spotify we have used playlist id () to get the daily top hits and the artists details to get to know the genre they follow to and their popularity. Now it is time for us to play with these data.

QUESTION 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 the number of top hits and overall popularity?

Answer: We have extracted data from artists' posts and artists' details. From artists' posts, we have plotted the graph indicating the number of posts each artist receives per day. Additionally, we obtain corresponding artist details from the Spotify API and plot the graph. We check the correlation coefficient for *Artists: Popularity vs Number of Posts* to study the factors influencing an artist's success. A positive coefficient suggests a positive correlation (as the number of posts increases, popularity tends to increase). If the p-value is less than 0.05, we can reject the null hypothesis that there is no correlation.

QUESTION 2:

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)?

Answer: Extensive analysis, including data understanding, cleaning, and plotting, was conducted to address the research question. Utilizing sentiment analysis of comments and employing word clouds played a pivotal role in comprehending the patterns underlying the phenomenon. Through thorough examination, it has become evident that established artists such as Taylor Swift, Travis Scott, The Weeknd, Drake, and SZA are poised to emerge as the top trending artists on Spotify by the end of the year, considering all the factors taken into consideration.

5.1 LIVE DASHBOARD

This project involves the development of a web-based dashboard using Flask [8], a micro-web framework. The interactive dashboard developed for this project serves as a dynamic and user-friendly platform for real-time exploration and representation of correlations between artist subreddit engagement, Spotify metrics, and their resonance within Reddit communities. Leveraging the Flask framework and a versatile suite of Python libraries, including Plotly [10], Matplotlib [7], WordCloud [11], and Pandas [12], the dashboard offers a visually compelling experience. Its functionality empowers users to filter and customize displayed results based on their preferences, fostering a personalized and insightful exploration of the intricate relationships between artist-specific subreddit engagements and Spotify metrics. This user-centric approach enhances the overall usability, ensuring that users can easily navigate through diverse data representations. Beyond its role in visualizing research

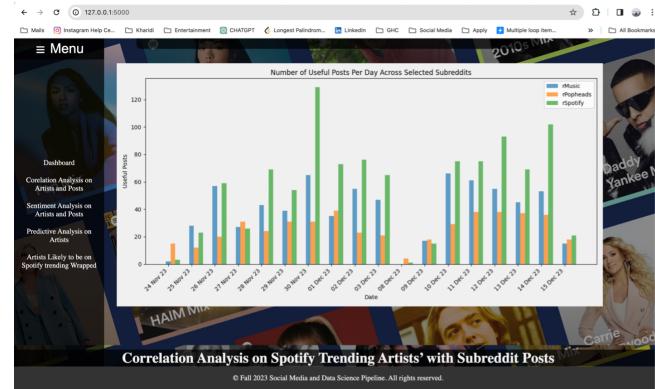


Figure 1: Useful Posts coming from Data Stream after filtering

findings, the interactive dashboard stands as an intuitive and engaging tool, contributing to a deeper understanding of music trends in real time.

5.2 Implementation

- The Dashboard in Flask has HTML Components module (dashboard.html) has a component for every HTML tag as well as keyword arguments for all the HTML arguments.
- Spotify and Reddit related graphs which were plotted in .ipynb file was converted in .py file using nbconvert and as an when the user starts to run the Flask application the images from the graph which were plotted get generated automatically in the static folder of the directory which is then rendered in the dashboard of the page and placed for our Correlation, Sentiment and Predictive Analysis.
- For Parameter variation, since the data for Spotify gets updated in like 3-4 days, we have basically created a drop down for artists and it shows the corresponding popularity of the artist which is fetched from Spotify and the number of useful posts from their official subreddit using Reddit API and displays it. This is only in consideration about the artists that it only shows the trending artists for that day.
- Below are the results of our implementation:
 - Figure 1:** shows a graph plot between useful number of posts by subreddit r/Spotify, r/popheads and r/Music (y-axis) vs date (x-axis) i.e. Number of posts by each sub-reddit is displayed. (we started fetching data from November 25th)
 - Figure 2:** shows a graph plot between total number of posts on Reddit official subreddit pages of the artists.
 - Figure 3:** shows the dual axis graph of the Popularity of Artists from Spotify and the Number of posts on Subreddit.
 - Figure 4:** shows the relationship about the number of posts for each artists vs the posts that have been incoming for that day.
 - Figure 5:** gives the graph for the top 30 posts which has highest number of posts and against for which artists: this gives us the relationship about the engaging activity and the fan following for that particular artist.
 - Figure 6:** gives the graph for the trending songs which are discussed the most on Reddit Platform

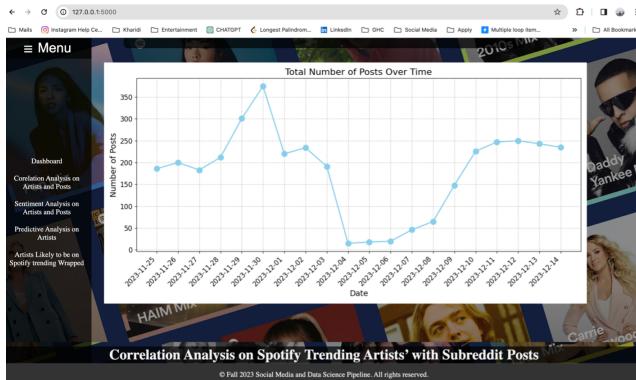


Figure 2: Useful Posts coming from Data Stream of the official subreddit pages of artists after filtering



Figure 3: Artists - Popularity vs Number of Posts

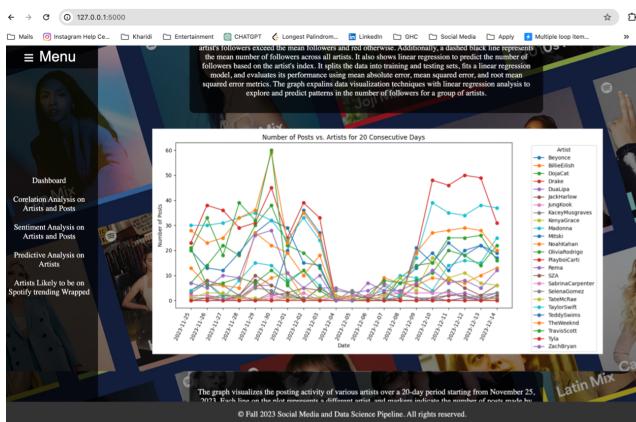


Figure 4: Number of Posts for the top trending artist beginning from 25Th November



Figure 5: Top 30 Posts with highest number of comments.

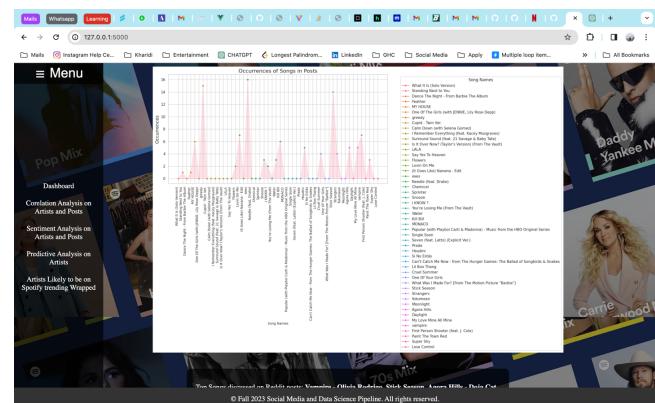


Figure 6: Trending songs on Reddit Platform

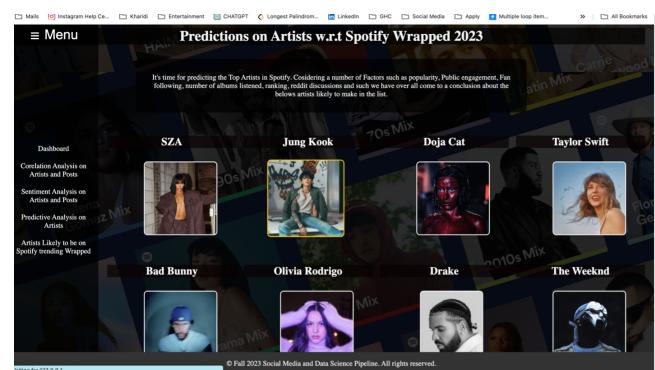


Figure 7: Predictive Analysis on the Top Artists for Spotify 2023 Wrapped

- Figure 7: Predictive Analysis on the artists who will make in the Top Artists list of Spotify Wrapped

6 CONCLUSION

After examining all these data sets for a range of artists and songs, we were able to predict the factors influencing the popularity of the trending artists on Spotify. And for this project we were able to

add some interactivity to basic data characterization by creating a live dashboard.

7 REFERENCES

- (1) Reddit Documentation. <https://www.reddit.com/dev/api/>
- (2) Spotify API Documentation. <https://developer.spotify.com/documentation/web-api/>
- (3) ModerateHatespeech Documentation. <https://moderatehatespeech.com/docs/>
- (4) Martin Pichl, Eva Zangerle, Gunther Specht. Combining Spotify and Twitter Data for Generating a Recent and Public Dataset for Music Recommendation, 2014.
- (5) Data Study on Popular Artists, Weijie Deng, Mar 27, 2018.
- (6) 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 Adessi, R Caterina, M Costa), pp. 960–966. Society for Music Perception Cognition/European Society for the Cognitive Sciences of Music.
- (7) 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.
- (8) Matplotlib Documentation - <https://matplotlib.org/stable/users/index.html>
- (9) Plotly Documentation - <https://plotly.com/python/>
- (10) Flask Documentation - The official Flask documentation provides comprehensive guidance on getting started, building applications, and using its various features - <https://flask.palletsprojects.com/en/2.1.x/>
- (11) Python Official Documentation - The official Python documentation offers comprehensive guides, tutorials, and references for Python programming - <https://docs.python.org/3/>
- (12) Plotly Library Documentation. <https://plotly.com/python/getting-started/#jupyterlab-support>
- (13) WordCloud Library Documentation. <https://pypi.org/project/wordcloud>
- (14) Pandas- <https://pandas.pydata.org/docs/>