Demystify Trends of Artists with Twitter, Reddit and Spotify API

Ketaki Khade
kkhade1@binghamton.edu
Binghamton University
Binghamton, NY, USA

Nikita Shriram Navkar
nnavkar1@binghamton.edu
Binghamton University
Binghamton, NY, USA

Swarali Patil
spatil32@binghamton.edu
Binghamton University
Binghamton, NY, USA

Tanmayee Kulkarni
tkulkar1@binghamton.edu
Binghamton University
Binghamton, NY, USA

Yash Jawale
yjawale1@binghamton.edu
Binghamton University
Binghamton, NY, USA

ABSTRACT

This project's main objective is to analyze data on currently popular musicians in the context of numerous social trends. Due to web's global presence and accessibility, social media platforms are crucial in boosting the popularity of any kind of music and music-based events that have an emotional impact on people. Reddit and Twitter are social media web platforms that millions of people use to share their ideas and opinions on a variety of topics. As a result, a stream of text data is produced that is made up of a huge number of tweets, posts and comments. In this project, we collected millions of posts tagged with "r/spotify", "r/music", "r/musician", "r/singer," etc. using Reddit API and also related tweets that refer to certain Spotify tunes, music, albums and artists from Twitter API. We analysed those tweets, 'r/spotify', 'r/music', 'r/musician', 'r/singer', posts, and comments and gather track-related data using the Spotify API. Finally, we identified the artists who are now trending by using the information we have acquired in combination with a particular time frame. We gathered a large number of tweets and posts from Twitter and Reddit, then we extracted the specific information, such as the URLs of popular songs or tracks, and we used those URLs to obtain artists from Spotify. We used the Spotify API to analyse those tweets and posts and extract information about the tracks. We plotted diffferent type of graphs in this analysis. Taken together, our results provide a datadriven, analysis of tranding artist in perticular time frame in music industry.

KEYWORDS

Twitter API, Reddit API, Spotify API, MongoDB, Data Collection, Snowball Sampling

1 INTRODUCTION

People of all ages value music greatly. It is a means of communicating and expressing human feeling. Popular music evolves in response to current social events. For instance, many candidates use music during election season to appeal to voters for a number of reasons, and campaign songs play a significant role. The music industry is currently undergoing ongoing change, with several transitional music genres experiencing popularity swings. There are a number of various platforms and sources for online music streaming, including Spotify, Apple Music, YouTube Music, Reddit, and others. Everybody has different musical tastes, both in terms of genre and artist. Even if it's not always better, if we hear something new, we usually share it with our friends, peers, and relatives. This eventually contributes to the creation of the music that is now popular or trendy. Using various hashtags, we frequently have a tendency to tweet, post, or share about it on social media, making it widely apparent. This unlimited, never-ending flow of knowledge about music from many sources could be used in creative ways in academia and the music industry. We've acquired real-time music-related data for this project from social networking sites like Twitter and Reddit as well as other music streaming services like Spotify, which offers a lot of music-related data. By collecting the data from these three sources over a specific period of time, we conducted analysis to discover which artists are popular and trending.

2 DATASET DESCRIPTION

2.1 Data Source

Here, in this project we used Twitter and Reddit dada from their respective API's to collect tweets and posts. Using that data, we collected information about top trending songs. For this we used Spotify web API to fetch data from a Spotify URL. We extracted track ID from the URL and sending it as a parameter to the Spotify API to retrieve details of any available track. For Twitter we used streaming API. And fro Reddit we collected data using $https://www.reddit.com/api/v1/access_token$ endpoint.

2.2 API METHODS

We will access data resources through standard HTTPS requests to an API endpoint. Web API uses appropriate HTTP verbs like; GET, POST, PUT and DELETE methods and returns all response data as a JSON object. As Twitter and Reddit API's allows us to find, retrieve and engage or create different resources, for this it

uses two HTTP methods: POST and GET. In this case using GET method we can retrieve data from Twitter and Reddit API's. As Spotify Offers us API methods based on REST principles. We begin receiving a stream of data and then the data will be in JSON format in response.

3   IMPLEMENTATION FOR PROJECT 3

We are planning to use either python Flask or Django for the web app development, and call our function from our python script to dynamically give us the data, which we will then use to generate the plots. We have used matplotlib for plotting, and pymongo to dynamically fetch data from mongodb when the function is called. We then use html to show the image files generated by the python code. Our webapp then displays these images on localhost.

4   RESEARCH QUESTION

In this project implementation we are going to present 'Top trending artist in a given time frame on Twitter and reddit with respect to spotify'.

5   ANALYSIS

- 1. Top trending artist on twitter
- 2. Top trending artist on reddit
- 3. Compare data between twitter and reddit

6   DATA FLOW

As we can see in Figure 8, we first fetch data from twitter API using the Twitter Stream API. We will use a JS script to achieve this, namely using the Needle library which provides us an easy way of recursively sending API requests to the twitter Stream API. We then can the text field of the received data for relevant key words like 'spotify' and save these tweets in our MongoDB database. We can then further filter the Tweets in the MongoDB database to extract the URL and check of the URL is in a valid format, as the URL will provide us with the Spotify track ID, which we will then use to fetch the track information using Spotify API to get details like the artist name, genre, etc. We can then save these details in a separate database in MongoDB and use these in our analysis and visualization.

7   OBJECTIVES

The main objective of our project is to get more clarity and information on what type of music people listen to, which artist and albums are more trending and how these trends change with time. Moreover, we are focusing on to analyze top trending artists on Twitter and Reddit with respect to our data collected on Spotify. We will take top artists/music on Spotify, and compare whether they are trending on Twitter and on Reddit. (RESAERCH QUESTIONS)

8   CONCLUSION

As per our observation, we observed that, trending artist on Twitter are differenet from trending Artist on Reddit. Different people have different prefereneces to express their music interests. And as per our anaysis, Twitter is most used platform as comapred to Reddit. We also observed that people ask about song suggestions on Twitter than on Reddit.
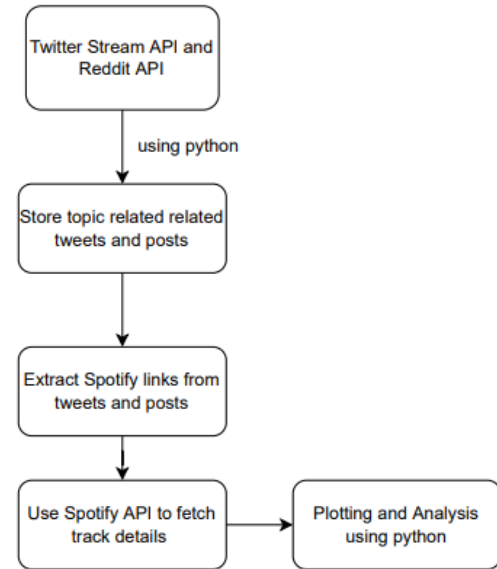

Figure 1: Data Flow Diagram

9   REFERENCES

- Gerlitz, C. Rieder, B. (2013). Mining one percent of Twitter: collections, baselines, sampling. M/C Journal, vol. 16, no. 2.
- Alina Campan, Tobel Atnafu Joseph Nolan (2018). Is Data Collection through Twitter Streaming API Useful for Academic Research? 2010 IEEE Tenth International Conference on Peer-to-Peer Computing (P2P). Spotify – Large Scale, Low Latency, P2P
- Abel, C. Hauff, G.-J. Houben, R. Stronkman, and K. Tao. Semantics+Filtering+Search=twitcident. Exploring Information in Social Web Streams. In Proc. of the 23rd ACM Conference on Hypertext and Social Media, pages 285–294. ACM, 2012.
- Dooms, T. De Pessemier, and L. Martens. MovieTweetings: a Movie Rating Dataset Collected From Twitter. In Workshop on Crowdsourcing and Human Computation for Recommender Systems, 2013.
- T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, and P. Lamere. The million song dataset. In A. Klapuri and C. Leider, editors, Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011), pages 591–596. University of Miami, 2011.
- M. Schedl, P. Knees, and G. Widmer. Investigating web-based approaches to revealing prototypical music artists in genre taxonomies. In Proceedings of the 1st International Conference on Digital Information Managemen