Does Twitter chaos effect popularity on Billboard?

Sadhana Desai
State University of New York at Binghamton
Masters in Computer Science
sdesai13@binghamton.edu

Kasturi Vartak
State University of New York at Binghamton
Masters in Computer Science
kvartak2@binghamton.edu

ABSTRACT

People can freely express their opinions on any matter at one platform that is Social Media. Firstly, Twitter, Instagram, and Facebook are three main social networking sites where people produce a huge amount of data about users behaviors,news, new trends,etc. A large amount of user generated content is available, leading to a lot of changes in the world.Twitter undoubtedly has held its firm position among all social networking sites with an exponential number of users every year.Many studies in healthcare industry, politics, sports, and music industry are carried out investing the power of Twitter data.Music lovers are prone to interact with their favorite songs and artists through social media, which provides enormous troves of insight not on just individual song and artists but also on how music consumers perceive any song.This chaos can help to get the information on the next week's popular artists or songs.Findings of this study are beneficial to the music industry to discover song performance by real-live update trends on social media in order to propose an appropriate strategy for hit and non-hit songs.

KEYWORDS

Twitter, Sentiment Analysis, Artists, tweets, Mentions, Hashtags, Songs, Billboard charts, Flask, Plotly, Bokeh, Interactive dashboard, Graphs

1 INTRODUCTION

Nowadays, in our booming era, the freedom of expression on the internet has been simplified than ever. The widespread adoption of social media has empowered users to freely express themselves and engage in far-flung conversations with people from all over the world. Given the significance of social media, understanding user-generated content provides a rich source for management and strategists in various industries. It is no exception to the billion music industry, which has experienced an exponential growth thanks to the introduction of streaming music over the last ten years. The fast growth of technology has transformed traditional music platforms into online music services where everyone is brought closer together. Social media platforms are where music audiences naturally congregate, building their own communities, and sharing experiences of songs and artists that subsequently is a catalyst to encourage users' engagement to the music industry.

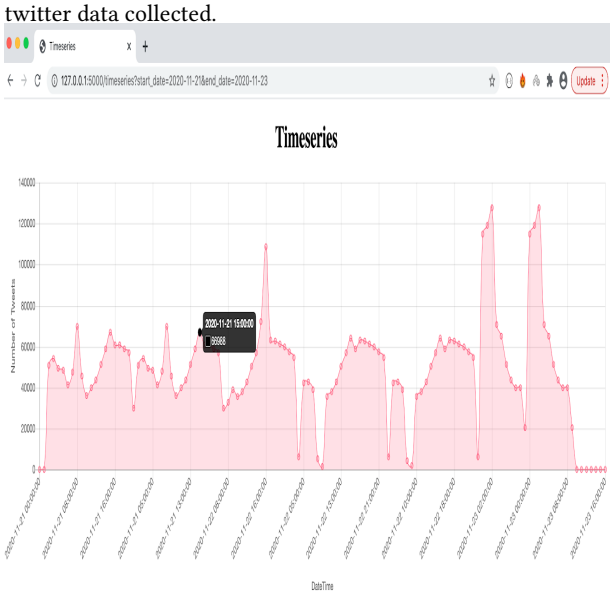
This paper marks the third stage of our project, that is building an interactive data visualization dashboards or charts as an application from measurements and analysis plan that we did in the previous stage. In the previous stage of our project, we cleaned the collected data and performed sentiment analysis using Vadar Sentiment Analyzer. We proposed multiple chart as our analysis such as Total number of tweets collected hourly which was a mandatory chart, total number of tweets by 'Hashtags'(#), 'Mentions'(@) and combined with respect to entries in Billboard. After sentiment analysis, we displayed charts such as total number of positive, negative and neutral tweets by 'Hashtags'(#), 'Mentions'(@) and combined with respect to entries in Billboard.

2 WORKING

We have the data for both the sources Twitter and Billboard.As mentioned before, our steps for analysis were, we first cleaned the data and started filtering out the tweets using hashtags and mentions,performed sentiment analysis for the collected twitter data. Below are the graphs which we got using the data we collected:

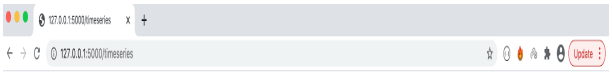
2.1 Timeseries plot of Twitter

Collection of stream of tweets is done using Python code.We had used the Twitter Streaming API to collect real time tweets.The resource URL used is https://api.twitter.com/2/tweets/sample/stream. We have data of tweets collected over a period of time. The graph shows the variation of twitter data streaming count.In this graph, we are displaying number of tweets collected binned hourly before filtering by Hashtags and Mentions.The url needed is http://127.0.0.1:5000/timeseries?start_date=2020-11-21&end_date=2020-11-23. Parameters needed are start date and end date to be mentioned in the url.Here, users can modify date as and when required to see the twitter data collected.



Below exceptions which we can get if the url is wrong or any other errors in parameters.For all other combinations of parameters in URL, gives error: Error : Enter correct URL. Find examples below:

- http://127.0.0.1:5000/timeseries
- http://127.0.0.1:5000/timeseries?start_date=2020-11-21
- http://127.0.0.1:5000/timeseries?&end_date=2020-11-23



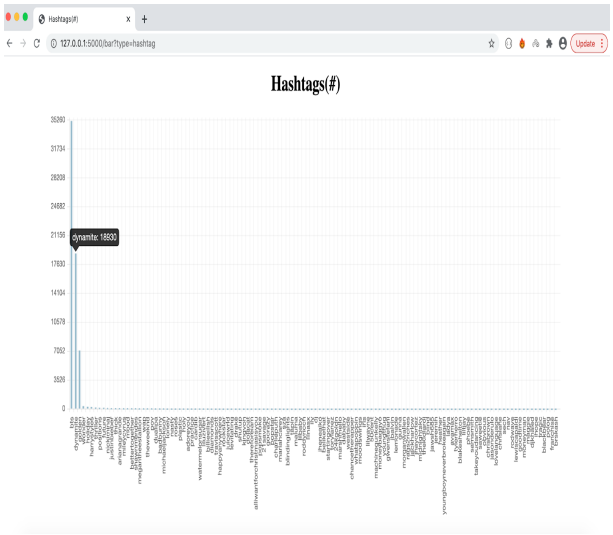
Error : Enter correct URL

2.2 Hashtags and Mentions

Hashtags are used a medium in social media to tag and wrap up the whole content of the tweet or posts. It is indicated by word preceding with a # symbol. It is used to index keywords or topics on twitter. Here in our dataset, we have used the Billboard top lists or the songs and artists names to clean and filter out the data for our analysis.Mentions in a tweet contains another username anywhere in the body of the tweet by using symbol @ to notify users that they are mentioned in the tweet. Here, there are usernames or accounts on the names of the artists, their groups or if the song reaches some popularity level then even songs have a username.

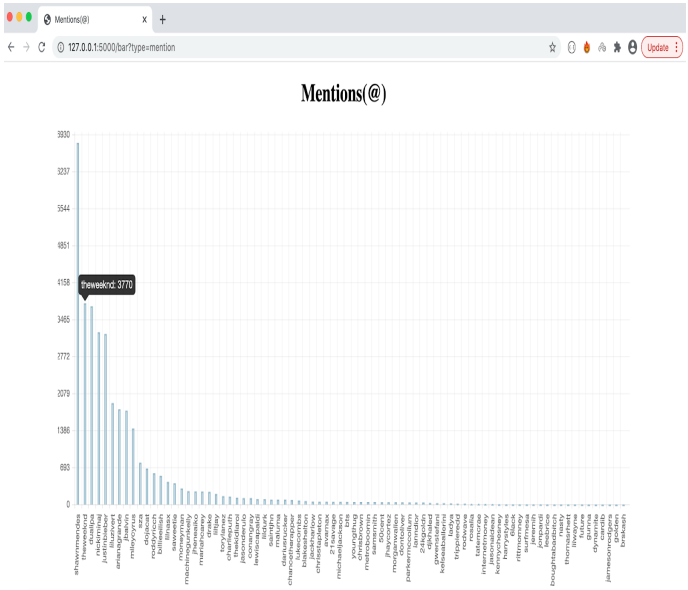
For the data filtered using hashtags, below graph shows the analysis as per billboard data which songs were mentioned in hashtags more and shows the chart with rankings of the songs and groups.

- Number of tweets which has hashtags by Title song on Billboard entry in descending order are displayed in the below graph. The url to be used is http://127.0.0.1:5000/bar?type=hashtag
- Below url can be used to display or see the data for top songs only for the billboard entries hashtags. The url is http://127.0.0.1:5000/bar?type=hashtag&top_songs=12



Mentions also show the count of how many times the users exclusively included song or the user in their tweets.We also showed the analysis of mentions in the graphs.Similarly as hashtags we can analyze and see the mentions graphs.

- Number of tweets which have mentions by Title song on Billboard entry in descending order are displayed in the below graph. The url to be used is http://127.0.0.1:5000/bar?type=mention
- Below url can be used to display or see the data for top songs only for the billboard entries mentions. The url is http://127.0.0.1:5000/bar?type=mention&top_songs=12

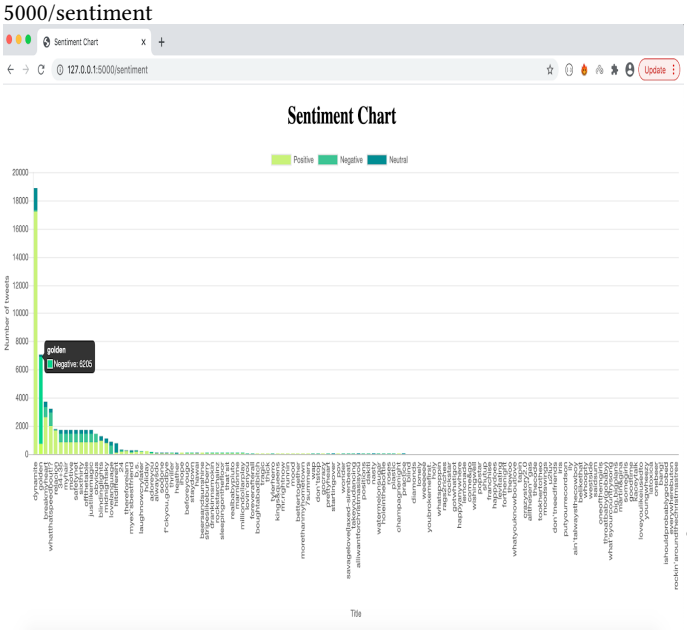


2.3 Sentiment Analysis Graph

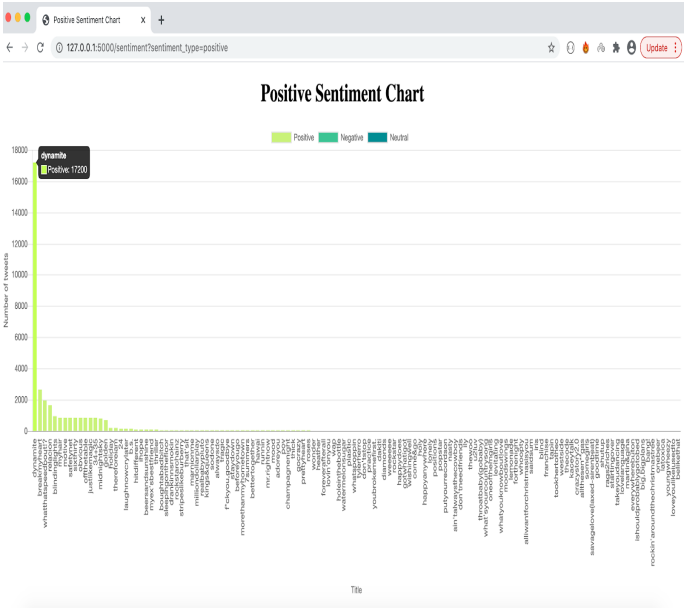
The major part of our research was to find the sentiment analysis of the tweets for the music industry.Sentiment analysis is performed for Twitter using a tool called Vader. These sentiments must be

present in the form of comments, tweets, retweets, or post descriptions, and it works well on texts from other domains also. VADER sentiment model is so designed which extracts features from twitter data, formulate the sentiment score, and classifies them in positive, negative, neutral classes. Sentiment-related variables are calculated on tweets related to songs only.

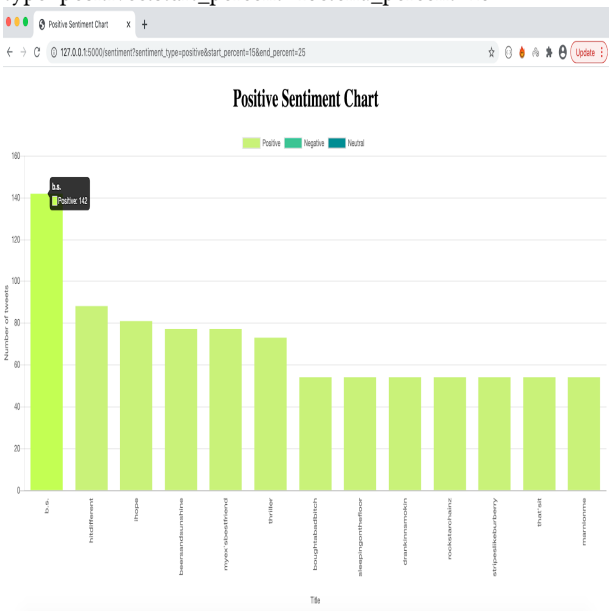
- Sentiment analysis result with positive, negative and neutral tweets in a stacked bar graph. The url used is http://127.0.0.1:5000/sentiment



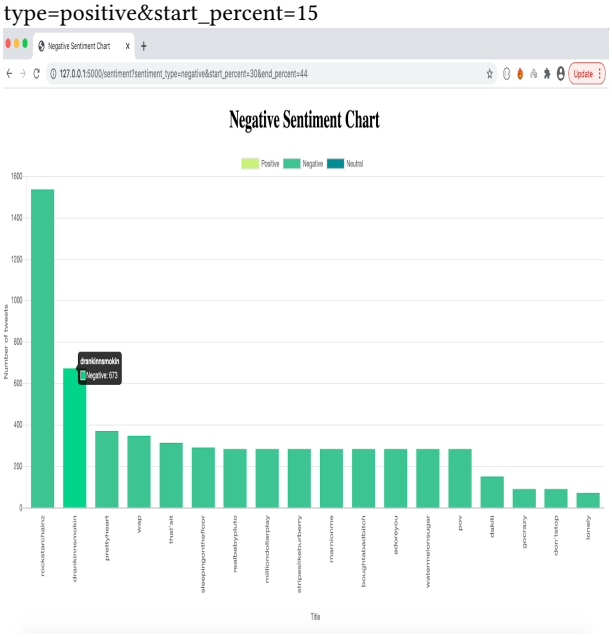
- Sentiment analysis result with positive tweets in descending order. The url used is http://127.0.0.1:5000/sentiment?sentiment_type=positive



- Sentiment analysis result with positive tweets between 15 to 25 percentage of total positive tweets, Same for Negative and neutral tweets just change the type to whichever is required. The url used is http://127.0.0.1:5000/sentiment?sentiment_type=positive&start_percent=15&end_percent=25

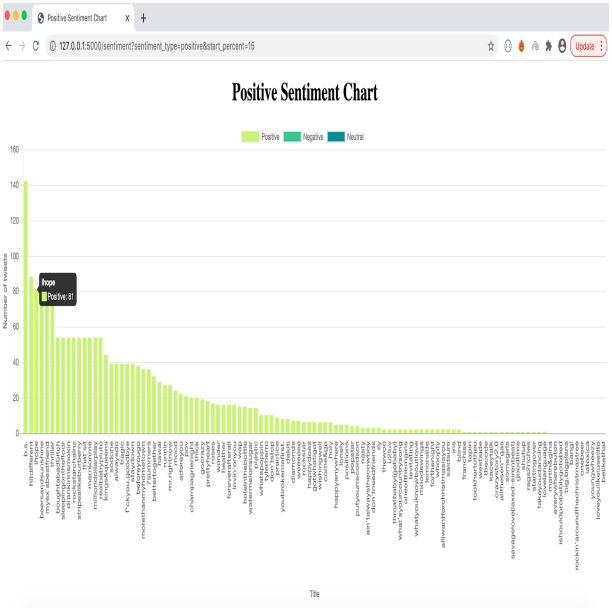


- Sentiment analysis result with positive tweets greater than 15 percent of total positive tweets, Same for Negative and neutral tweets just change the type to whichever is required. The url used is http://127.0.0.1:5000/sentiment?sentiment_type=positive&start_percent=15



- Sentiment analysis result with positive tweets less than 25 percent of total positive tweets, same for Negative and neutral tweets just change the type to whichever is required.

The url used is http://127.0.0.1:5000/sentiment?sentiment_type=positive&end_percent=25



3 HOW TO RUN

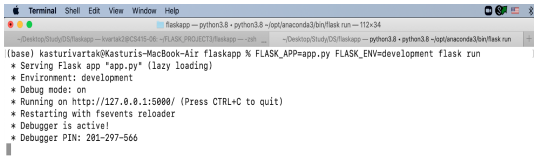
We have used Flask app and Plotly to built our interactive dashboard for displaying data content of our project part two.We planned on using Bokeh before but the documentation which we knew for it was deprecated hence we thought of using Plotly and it was comparatively better to use.The functionality of the graphs and their urls were mentioned above.Below you can find how to run the app and access the dashboard. To run from VM:

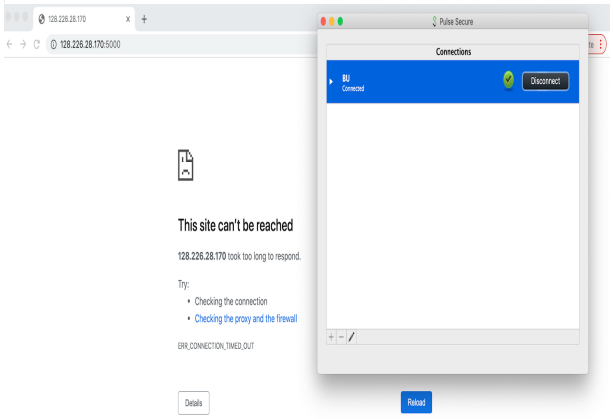
- Log in to VM
- Go to /FLASK_PROJECT3/flaskapp
- export FLASK_APP=app.py flask run –host 0.0.0.0
- Go to browser and run URLs mentioned(connect to VPN) http://128.226.28.170:5000

To run from local machine:

- Go to project directory
- FLASK_APP=app.py FLASK_ENV=development flask run
- Go to browser and run URLs mentioned in the description: For example, http://127.0.0.1:5000/sentiment

Flask App is running in our local machine.Flask App is also running on VM but when we get Timeout error connection when trying to access the via URL(connected to VPN).





4 CONCLUSION

We have built an interactive dashboard based on our project two data and analysis.We could come up with four graph ideas and our sentiment analysis graphs being complicated ones showing 5 types of variations including top analysis and percentage level analysis.We could find the sentiment analysis was effective to find out the positive,negative and neutral tweets from the tweets which we collected and graphs are proof to it.The interactive graphs with user given parameters can give idea about the data and represent the project analysis through graphs.

As a future scope, we found out working on a Geo-location of the tweets and analyzing that from where and how many users post a tweet and where the music content is more popular.Due to time constraint, we could not complete that part as such but would like to work on it as future scope.Also, we found out some prediction algorithms with which we could analyze our content and work on predicting new lists.also, we found out as per our analysis BTS was on top the next week after our analysis of twitter data of previous week in project two.This news proves we were successful in our analysis.That factor can also be added to our future scope list.

5 ACKNOWLEDGEMENT

We would like to Thank our Professor Jeremy H Blackburn for letting us work on this project and guiding us through and helping us with this idea.

REFERENCES

[1] https://api.twitter.com/2/tweets/sample/stream
[2] https://www.billboard.com/charts/hot-100
[3] https://flask.palletsprojects.com/en/1.1.x/
[4] https://developer.twitter.com/en/docs/twitter-api/tweets/sampled-stream/api-reference/get-tweets-sample-stream
[5] https://https://jsfiddle.net
[6] https://www.chartjs.org
[7] https://mode.com/blog/python-interactive-plot-libraries/