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

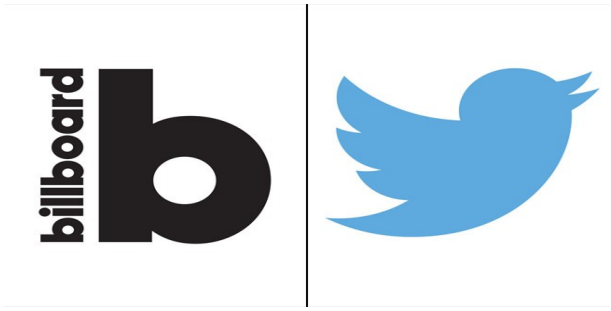

Figure 1: Lot of Twitter chaos,will it effect popular stars?

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, Datasets, Artists, Top lists, Songs, Billboard charts, Hashtags, Mentions, PostgreSQL.

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.
Twitter remains as the best indicator of the broader pulse of the world and overview picture of what is happening within.It has become one of the outlet for people talking about music.Given a considerable amount of music-related users' behaviors data, understanding Twitter is a potential source for the music industry in various perspectives. With all this we can say that the chaos going on in the twitter tweets, retweets, followers, mentions, hashtags can help to analyze the changes in the rankings of songs and artists of the Billboard charts.

2 DATA COLLECTION

This project has Twitter and Billboard as the platform from which data has been collected. Below two sections specify the details of data collection from the two sources:

2.1 Twitter Data

Collection of stream of tweets is done using Python code.We have 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 appended several parameters at the end of the URL which can be useful for the further analysis in our Project.

We plan to analyze the number of followers, mentions,retweet counts, hashtags and few factors to track the effect on Billboard charts.We have used the twitter fields available which are:

- Id - Unique identifier of the Tweet.
- Tweet - The content of the Tweet.
- CreatedAt - Creation time of the Tweet.

- Public Metrics - Engagement metrics for the Tweet at the time of the request such as count of retweets, replies and likes.
- Entities - Contains details about text that has a special meaning in a Tweet such as mentions, hashtags and other details.
- Annotations - Contains context annotations such as domain,domain name for the Tweet.
- Language - Language of the Tweet, if detected by Twitter.

The implementation of the collection program used Python's request library for sending the stream request. We have used Bearer token as a credential. This credential is generated once we have created a developer account on Twitter. The response request is sent using GET method and we collect the JSON response. We have collected every field required for each line in response.

2.2 Billboard Data

There is no official API for Billboard, so we have created a script as per discussion with the professor just to get HOT-100 chart entries using https://github.com/guoguo12/billboard-charts as a reference. URL: https://www.billboard.com/charts/hot-100

As we were getting the response in HTML format, we have used BeautifulSoup as an HTML parser. Billboard API updates every Tuesday afternoon (though the time is not specified anywhere).If there is a holiday on Monday then HOT-100 chart updates on Wednesday afternoon. Thus we are running our script every Tuesday and Wednesday afternoon at 12 PM. After getting the chart entries, we are inserting chart entry attributes into our database given below:

- Artist: The name of the artist, as formatted on Billboard.com.
- PeakPos: the track's peak position on the chart as of the chart date, as an int (or None if the chart does not include this information).
- LastPos: The track's position on the previous week's chart, as an int (or None if the chart does not include this information). This value is 0 if the track was not on the previous week's chart.
- Weeks: The number of weeks the track has been or was on the chart, including future dates (up until the present time).
- Rank : The track's current position on the chart.
- isNew: Whether the track is new to the chart.

2.3 Data Storage

After the data is collected, We need to store it in a database to preserve it so we can query it as and when data is required.To do this, we are using Postgres RDBMS.
Postgres is a free and open-source relational database management system (RDBMS) emphasizing extensibility and SQL compliance.It is designed to handle a range of workloads, from single machines to data warehouses or Web services with many concurrent users.

3 DATA EXPLORATION

3.1 Twitter

We have collected and appended the twitter data in a .csv file.We have around 37000 tweets collected in our analysis for around 15 mins timestamp. Below is one of the representation of the tweets collected:

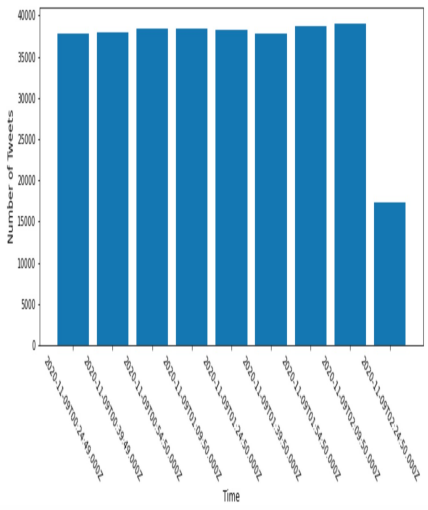

Figure 2: Tweets collected at a timestamp of 15 mins

3.2 Billboard

For the Billboard data, we have collected the charts and stored them directly in the database as it updates weekly. Below we have a sample of the data collected during our process:

[title, artist, peakPos, lastPos, weeks, rank, isNew]

['Positions', 'Ariana Grande', 1, 0, 1, 1, True]

['Forever After All', 'Luke Combs', 2, 0, 1, 2, True]

['Mood', '24kGoldn,iann dior', 1, 1, 12, 3, False]

['Laugh Now Cry Later', 'Drake,Lil Durk', 2, 3, 11, 4, False]

['Blinding Lights', 'The Weeknd', 1, 4, 48, 5, False]

['I Hope', 'Gabby Barrett,Charlie Puth', 6, 6, 44, 6, False]

['WAP', 'Cardi B,Megan Thee Stallion', 1, 2, 12, 7, False]

['Savage Love (Laxed - Siren Beat)', 'Jawsh 685,Jason Derulo', 1, 5, 20, 8, False]

['Lemonade', 'Internet Money,Gunna,Don Toliver,NAV', 9, 10, 11, 9, False]

['Holy', 'Justin Bieber,Chance The Rapper', 3, 9, 6, 10, False]

4 CONCLUSION

The objective of this project was to do the data collection which is the most important part as the collected data would be further used for analysis.Our future scope will be to analyze the twitter content using Billboard current data applying certain algorithms and measuring factors to observe the effect of the twitter chaos on Billboard charts next week.

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://github.com/guoguo12/billboard-charts
[4] https://developer.twitter.com/en/docs/twitter-api/tweets/sampled-stream/api-reference/get-tweets-sample-stream
[5] https://en.wikipedia.org/wiki/PostgreSQL