

SI330W18 Project Proposal
Libby Murray

I'm interested in how a music artist is affected by what they do and say online. For this project, I want to look at how a music artist's influence on social media impacts their success. More specifically what effect does the activity of an artist, via the content they post, have on the popularity of their music. I will look at the artist on Spotify and use Twitter feed to look at social media influence. I propose looking at Twitter feed to gauge artist social influence and Spotify's top music data to measure an artist's success. Many users of music streaming platforms post what they are listening to on Twitter, therefore this is a great platform to display the social influence of an artist.

The main question I hope to answer is:

- Do more involved active artist have more success?

Additional questions I want to answer include:

- Can exposure trump biases; such as biases for musicians' music and fashion style?
- How does an artist social media platform dictate the dialogue surrounding the artist?
- Is the language used in an artist's music reflected in the language used on social media?

Data source 1: Spotify artists on Twitter

- Name: Spotify artists (81323 artists on Spotify)
- Size: 4.15 MB
- Location: <https://www.kaggle.com/ehcall/spotify-artists/data>
- Format: CSV
- Access method: download CSV file from Kaggle

Data source 2: Artist success on Spotify

- 1
 - a. Name: Top Spotify tracks of 2017 (Spotify's playlist of the songs streamed most often over the course of 2017)
 - b. Size: 12.84 KB
 - c. Location: <https://www.kaggle.com/nadintamer/top-tracks-of-2017/data>
 - d. Format: CSV
 - e. Access method: download CSV file from Kaggle
- 2
 - a. Name: Spotify's worldwide daily song ranking (200 daily most streamed songs in 53 countries) -- How a songs popularity varies over time
 - b. Size : 43.07MB
 - c. Location: <https://www.kaggle.com/edumucelli/spotify-worldwide-daily-song-ranking/data>
 - d. Format: CSV
 - e. Access method: download CSV file from Kaggle

Data manipulation:

Using the Spotify artist file, I will make a requested to Twitter Streaming API and pull content from and about a specific artist. Then group that information with each artist to develop the sense of the type and level of influence each artist has online.

I will combine data sets: "Spotify's worldwide daily song ranking" and "Top Spotify tracks of 2017," to compile more details on the popularity of a specific artist's song; details such as, how many streams that song had and when the song was at peak popularity.

The new information I can gain from combining the datasets include:

- Of artist that have the top 10% tweets posted, how many of their song reach the top charts
- Of artist that have the top 10% tweets posted, how many genres are represented
- Of artist in a specific genre, have the top 10% tweets posted
- What are the most common words used for arts in a specific genre
- the number of tweets/ day for an artist compared to the number of streams their top ranked songs had that day

Data visualization

- Number of tweets about the artist & number of streams on spotify
- Number of tweet from artists & number of streams on spotify