Tracking Popularity of Music from 2018 to 2019

Derek Kartalian, Khalid Nadi, Elaine Sagalchik, Linda Zhang

Cornell University

May 14, 2019

**Abstract**

[margins on side]

**1 Introduction**

The popularity of music is a complicated and multifaceted measure to quantify. With the rise and domination of electronic music streaming, the accessibility of music has risen greatly; however, many other mediums are still used to listen to music, such as purchased electronic music, the radio, and even CD players. In addition to the multitude of listening platforms, music itself inherently has many different qualities and traits that serve to differentiate types of music, and give each piece its own unique sound, feel, and timbre.

**1.1 Objectives**

We are seeking to find what features of music make it ‘popular’. Spotify, one of the premier music streaming services, thrives off of its ability to utilize its proprietary recommendation systems and predictive technologies. Of particular interest, Spotify analyzes each song in its over 30 million song library with qualitative and quantitative ‘audio features’. These elements of songs range from intrinsic musical qualities such as tempo (beats per minute) and key, to more calculated and complex measures such as ‘danceability’.

[ more about our intentions of what we want to analyze / achieve based upon which of our analyses work lmao ]

**2 Parallels in Other Contexts**

**2.1 Movies, TV, and Other Forms of Entertainment**

Similar to Spotify, Netflix is an entertainment content provider that prospered against competition due to their innovativeness and superior use of technology. Beginning as a DVD rental service, Netflix’s competitive advantage arose from its subscription model and recommendation system that accounted for the tastes and preferences of content of its users. Netflix’s competition sought to leverage the most recent box office movies and other large titles, while Netflix remained focused on developing countless sub-genre categorizations to fit customer tastes.

In a similar fashion, Spotify has risen to be one of the most popular music streaming platforms in the world. What separates Spotify from music services of massive technology companies such as Apple, Google, and Amazon is its ability to provide tailored content to users through its customized playlists. Whether a user is seeking to discover new music and artists they have never heard, or dive deeper into a specific subgenre that they like, Spotify Music provides the best interface and services for these needs. This core capability is made possible by Spotify’s recommendation systems - powered by all the analytic tools on its API.

**3 Questions and Workflow**

**3.1 Motivation**

For this project, we utilized two datasets: one found on Kaggle, and one created by us.Both datasets were created to survey and analyze the top music tracks of each year - 2018 and 2019 respectively. The Kaggle dataset was created by a user surveying Spotify’s most popular tracks for 2018. This user was seeking to find whether the audio features of a track can predict other features, which features correlate the most, and for any patterns amongst the values. The second dataset was created by us - attempting to gather the same data as the Kaggle set, but for the most popular music of 2019.

**3.2 Composition**

**3.3 Collection Process**

For the creation of our 2019 dataset, we surveyed the Billboard Top 100 tracks for each week of 2019 so far since the year is not concluded, and Spotify itself has not yet released the most popular tracks of the year. Although the Billboard Top 100 is not Spotify’s released list of top tracks, Billboard’s metrics are more holistic, and take into account measures such as radio plays, streams, downloads, and record and CD sales. After compiling this list of music into a Spotify playlist, a python tool called “Exportify” allows users to export their own Spotify playlists into csv files. These csv files contain basic information such as the track, artist, and album name, along with when the song was added to the playlist. The essential value from this exportify program is the Spotify URI (unique resource identifier) for each track, as this URI can be utilized in Spotify’s Developer API.

Separating all of the track URIs from the playlist, we were able to put them into Spotify’s open source tool called “Get Audio Features for Several Tracks”. This tool from the API allows users to input song URIs and receive all of Spotify’s analytics on ‘audio features’. All of the output data was gathered in JSON format, which we converted into another csv file. Appending the track, artist, and album name values from the first csv, we created our own dataset containing all of Spotify’s audio features for the most popular tracks of 2019.

**3.4 Preprocessing/Cleaning/Labeling**

The preprocessing before consolidating all the data into one csv was quite simple. Some variable names were changed to replace white space with underscores for clarity.

**4 Development Process**

**4.1 Clustering Based on Audio Features**

We are interested in investigating clustering tendency in Spotify’s top music of 2018 and 2019. First, we would like to determine whether it is possible to group songs together based on audio features such as danceability, energy, key, acousticness, valence and tempo. After clustering these data points, we will observe whether we can group similar songs together based on just two key elements at a time. Then, we can determine if the content in each cluster is related: For instance, are there specific patterns? Do the songs in a cluster belong to the same artist or genre? Once we cluster our data, we will be able to observe whether content is related.

Ideally, we would like to use all key elements to cluster similar songs together - however, due to the multidimensionality of the data resulting from using multiple variables representing audio features, we would need to use algorithms such as Principal Component Analysis (PCA) to properly visualize our data in two-dimensions. Although we attempted to incorporate this step into our project, we ultimately decided that the complexity of the process rendered it outside the scope of this particular assignment. Consequently, we clustered based on two key elements at a time, observing whether we can group similar songs together, accounting for just energy and key, energy and acousticness, energy and valence, tempo and energy, and other pairings. We hypothesized that these particular elements had the most significant correlation and similarities between both artists and genres.

**4.2 Acclaim and Artist**

After clustering based on audio features, we began to question what it really means for a song to be popular. Spotify creates its “Top 100” lists by calculating the amount of times its user listen to a song and add it to the playlist accordingly (Fama). It analyzes the number of times each song is played, which means that Spotify users basically created this list themselves (Fama). Spotify just has the data analysis tools to crunch the numbers and compile all the songs onto one list.

However, if this is the case, then what does this mean for artists? Are artists who have one song make the Top 100 likely to have other songs place too?

In order to check this out, we made a list of all the artists whose songs made it onto the Top 100. This was to determine how many artists we need to keep track of - we figured out that for the 100 songs on the list, there are only 70 artists. After creating this list, we made a dictionary where the keys are artist names and the values are the number of their songs in the Top 100. We then made a bar graph of this data to see if there are any trends and/or surprises.

The bar graph was crowded but it showed that out of these 70 artists the vast majority of them only had one song on the list. This was shocking because it delineates that the name of the artist might not play that large of a role in securing a place in the Top 100. To better visualize this data, we created another, more directed dictionary which only included artists who had at least two songs in the Top 100. The resulting graph had a mere 18 artists, and yet the same trend persisted. The vast majority of these artists only had two songs. To further narrow this visual, we created another dictionary. This one only included artists who had at least three songs in the top 100. The results displayed five artists: Post Malone with six songs, XXXTentacion with six songs, Drake with four songs, Marshmello with three songs, and Ed Sheeran with three songs. This was an interesting finding because it suggests that something about the style of these songs - perhaps even artists - makes them so popular. This set us on a quest to dig deeper.

**4.3 Artist Hits and Other Variables**

We created a new variable called “Hits.” It is based on the number of songs an artist has in the Top 100 and is a whole number from 1-6. We then created scatter plots to determine if there are any clusters among this variable and the variables that we already have.

**5 Further Research**

**5.1 Social Media Influence**

Since Spotify determines a song’s popularity by the amount of times it is streames, it is important to consider what outside factors can lead someone to click play (Fama). Studies have shown that one’s upbringing and social circles play an important role in determining what kind of music one likes, and in this age of technology, it would be interesting to explore the impact that social media has on one’s music selection (Stevens). If someone’s social media feeds are filled with people raving over a new Drake song, then are they more likely to listen to it on Spotify? What function could the linked accounts between Facebook and Spotify take on, where people can see the music that their friends are listening to?

**5.2 Criminal Activity as an Element**

Like social media, current events are an impetus to talk about something, and the criminal actions of an artist are no exception. Discussing the unlawful behavior of an artist, even if it is a negative form of attention, still draws hype to an artist and his/her music. Another music streaming service, Soundcloud, has taken advantage of this phenomenon and pulls statistics from various sources to ensure that it propels the most talked about artists to the top, and these sources include news sites (Shah). Its algorithm picks up on what is trending but doesn’t recognize why, which is different from Spotify’s current algorithm, but if a song is trending on one streaming service, then how does it affect its popularity on another streaming service? Investigating artists who are making headlines due to criminal activity, and thus trending on various streaming services, could yield intriguing results as to why a song (also) gains traction on Spotify.

**5.3 Data Visualization in PCA**

We clustered data based on two variables at a time to see if we can group songs together and find correlations and/or common factors that point to reasons for their popularity. There were many factors, and it would be compelling to use algorithms such as Principal Component Analysis (PCA) to display its multidimensionality.

**References**

Fama, Amanda. “Your Top 100 Spotify Playlist For 2018 Includes Your Favorite Songs Of The

Year.” *Elite Daily*, Elite Daily, 7 May 2019,

www.elitedaily.com/p/how-did-spotify-pick-my-2018-top-100-playlist-its-based-off-you

-favorites-13256912.

Shah, Saqib. “SoundCloud Shows How Its Algorithms Influence Music Streams.” *Engadget*, 20

Nov. 2017,www.engadget.com/2017/11/20/soundcloud-algorithm-plays-stats/.

Stevens, Alison Pearce. “Blame Your 'Environment' for Your Taste in Music.” *Science News for*

*Students*, 28 Oct. 2016,

www.sciencenewsforstudents.org/article/blame-your-environment-your-taste-music.