Textual Analysis of Spotify's Musical Map

Brian Yeh
Rutgers University
New Brunswick, NJ, USA
by146@scarletmail.rutgers.edu

Gerard de Melo
Rutgers University
New Brunswick, NJ, USA
gdm@demelo.org

Abstract

Spotify is one of the most popular commercial music streaming services worldwide, which gives it access to valuable data about the musical tastes of its users. It publishes data-driven articles through its project called Spotify Insights, one of which is a Musical Map consisting of thousands of cities [1]. Each city is represented by a playlist of songs that are most distinctly listened to there, which is updated by Spotify on a weekly basis. In this paper, we use natural language processing to study the lyrics across all songs contained in each city's playlist as they vary geographically or over time. Music is oftentimes selected based on a person's mood or mindset. By considering the playlist for a city as the overall sentiment of its populace, we propose that this is an effective survey of its inhabitants' thoughts and feelings. Our analysis shows that there are distinct differences in sentiment scores and themes between cities and countries, as well as a change in these metrics over time for a single city.

1. Introduction

The use of natural language processing on lyrics has been done before in multiple contexts. Some of these include aiding the classification of music and gaining further insight into lyrical trends for popular songs. The playlists for each city in Spotify's Musical Map allow us to now consider lyrics in a geographical and societal context. Each playlist is comprised of a list of songs that Spotify's data team considers to be distinct to the city it represents. When it comes to determining what songs are distinctively listened to in a city, the team at Spotify Insights defines "distinctive" as music that people in that city listen to a lot but is not listened to very much in other cities. More specifically, according to Insights team member Eliot Van Buskirk, "Distinctive popularity is a combination of absolute local popularity and that local popularity relative to global popularity, plus a bit of a bonus for newness" [1]. It is important to note that due to that formula, it is possible that a song exists in multiple city playlists. Each of these playlists are updated on a weekly basis, with some songs being replaced and others retaining their position.

The music a person listens to is often representative of their ideals or mood. The overall positivity or negativity of a playlist can also be calculated and provide a picture of the overall mood for each city. We can postulate that a city or country embroiled in societal unrest will probably have a more negative sentiment score. When considering a geographic collective of people, we find that the lyrics for songs popular in that region are representative of its demographics. For example, cities in the Bible Belt of the United States will likely have playlists that contain more songs of the Christian genre, and thus more mentions of "Jesus". However, this can also be influenced by the most popular genres listened to in each city, as some lean more lyrically positive or negative. Finally, we can make use of the concept of word embeddings by first defining a theme or keyword, and then comparing its vector representation against the words in lyrics to calculate a similarity score.

2. Data Sources

The main data source contains the lyrics and additional information for nearly 45,000 songs. 18,000 of these songs are in English, with the rest being a mix of other languages such as Spanish, French, or Chinese. Each row in the main lyrics dataset represents a song and its attributes, including its unique Spotify ID, the artist, title, lyrics, language, and audio features.

Another data source contains all the playlists associated with the Musical Map and their unique Spotify IDs. These IDs then allow us to query the Spotify Web API for all the songs within each playlist. Every two weeks a new data source is generated containing information about all the songs in the Musical Map and the playlists that they are associated with that particular week. Each playlist contains up to 100 songs, and the total amount of songs in one of these data sources is typically over 200,000. Keep in mind that the same song may appear across different playlists, so the number of unique songs is closer to 50,000. The same song may also appear across different songs data sources if they remain popular enough over time. In this data source, each song is associated with a location, the playlist it belongs to, and other information about the song such as its Spotify ID, artist, and title.

Additional datasets are generated through our analysis of songs, such as one containing the genres associated with each song's artist and one containing the calculated VADER sentiment scores for each song that has lyrics.

3. Obtaining the Data and Data Cleaning

The Spotify Web API allows us to obtain a user's public playlists as well as the songs within them. We pull the data for each city every two weeks using the database we have of all the Musical Map playlists. Then, lyrics are obtained by web-crawling LyricWiki and Genius by the use of third-party APIs [2, 3]. For Asian cities and songs, an additional script web crawls the KKBOX website for lyrics in Asian languages. The langid Python package is used to identify the language of the lyrics [4]. Not every song has attainable lyrics in this fashion, but out of more than 40,000 songs pulled every other week, we can obtain around 50% of the lyrics.

Given the language limitations of the sentiment analysis tools we used, analyses were only done on songs with English lyrics. Depending on what analysis was being performed, some cities would be dropped if they did not have enough songs with English lyrics. Before each analysis, the lyrics are cleaned to remove tags such as "[Verse]" and "[Chorus]".

Additional features of songs were obtained through the Spotify Web API. For each song, Spotify provides audio features, including its musical valence which is a measure of how positive or negative a song sounds [5]. There are also analyses involving genres, which are obtained through the artist for each song because Spotify does not provide genres for an individual song. Each artist is associated with a list of zero or more genres.

4. Approach

4.1 Sentiment Scores using VADER

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool developed by Hutto and Gilbert [6]. Using this tool, each song is assigned a sentiment score that ranges from -1.0 to 1.0. Because VADER is more attuned to smaller texts such as social media posts, the entire lyrics for a song is not passed all at once into the sentiment analyzer. Instead, each line of the lyrics is analyzed individually and assigned the single, unidimensional composite score that is returned by VADER. The composite score is derived from the sum of the valence score of each word in the lexicon that is then normalized to be between -1 and +1. The average of all line scores is then calculated as the overall score for a song. Lines with a score of 0 are not factored into the mean as this typically indicates that there are no matches in VADER's lexicon for that line. The distribution of VADER scores across all songs that we have English lyrics for is shown in Figure 1.

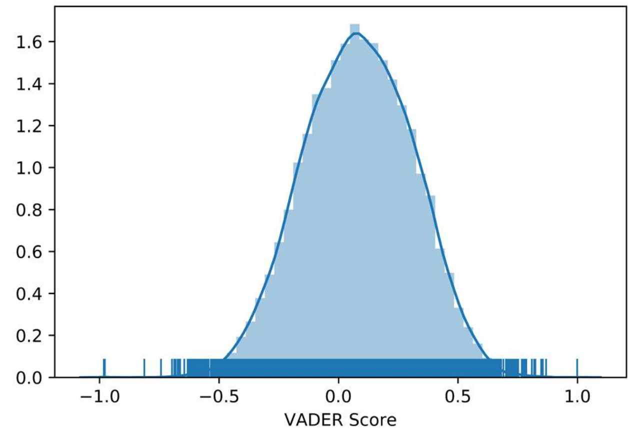

Figure 1. Distribution of VADER scores.

Because Spotify also provides us the musical valence for songs, we can see if there is any correlation between musical and lyrical positivity or negativity. Using the Pearson correlation coefficient, we find a Pearson r value of 0.055, with a p-value <.001. This indicates that there is no correlation between the musical and lyrical sentiment scores.

After obtaining the VADER scores for each song that we have lyrics for, we can now associate these scores with cities and countries in a weekly songs dataset. When calculating the entries for the following tables, locations that do not have at least twenty VADER scores are excluded. Given the week of May 1, 2019, we can find the average VADER score for a country by grouping playlists based on the country that the city is located in, and then taking the average of all songs across those playlists. Table 1 shows the top five most positive countries by VADER score, while Table 2 shows the most negative.

Table 1. Top 5 countries with the most positive VADER score for the week of May 1, 2019

Country	VADER Score
Kuwait	0.254273
Indonesia	0.223280
Saudi Arabia	0.217972
Qatar	0.202125
United Arab Emirates	0.195555

Table 2. Top 5 countries with the most negative VADER score for the week of May 1, 2019

Country	VADER Score
Italy	-0.094795
Taiwan	0.017980
Chile	0.042754
Jamaica	0.047072
France	0.061008

It is also possible to use the Python package called Folium to create a choropleth map that shows each country's VADER score [7]. Figure 2 provides an overview of how sentiment scores vary across all the countries we have data for.

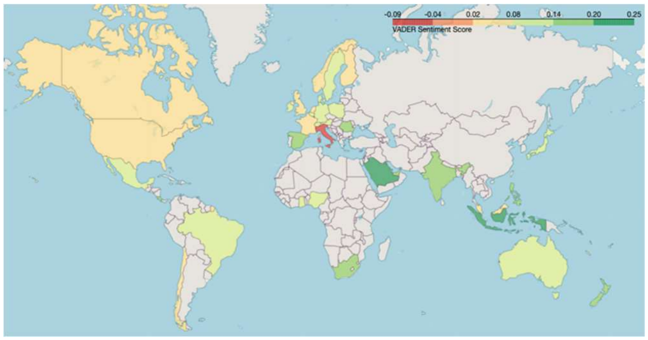

Figure 2. Choropleth Map of VADER scores for the week of May 1, 2019

We can also compare this to a choropleth map of the data from the week of March 15, 2019, shown in Figure 3.

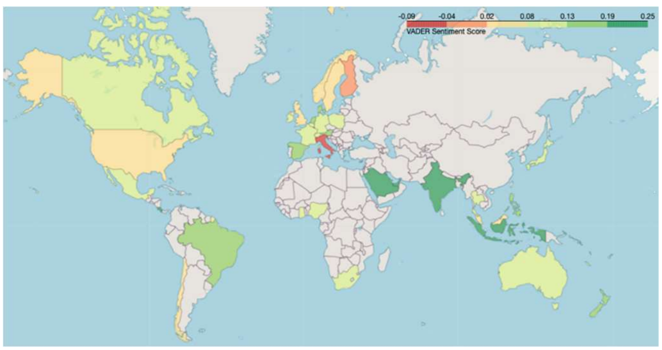

Figure 3. Choropleth Map of VADER scores for the week of March 15, 2019

By comparing the two maps we can see how VADER scores change over time across all countries, perhaps as a result of changes in its inhabitants' mood and musical tastes.

Given sufficient data for a multitude of cities in the United States, we can create a choropleth map for states within the continental United States of America, shown in Figure 4.

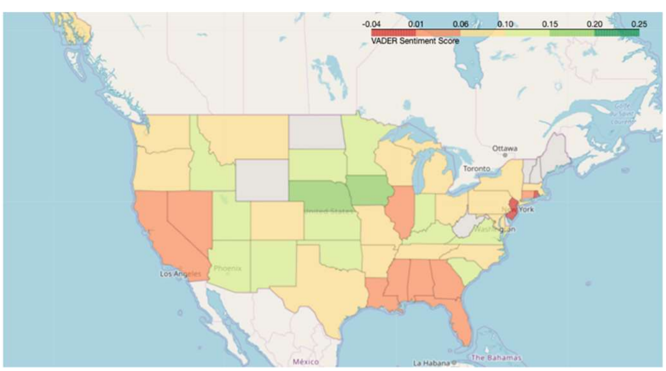

Figure 4. USA Map of VADER scores for the week of May 1, 2019

4.2 Word and Entity Tokenization

For the purposes of finding the most commonly used words and entities in lyrics, we use the Python package called spaCy [8]. It provides a word tokenizer and named entity recognition which we use for our analyzations. Because spaCy is well-suited for even very large corpuses, the entire lyrics for a song can be passed into its model all at once. For word tokenization and entity recognition, we use one of Spacy's pre-trained statistical models called en_core_web_md. This is an English multi-task convolutional neural network trained on OntoNotes [9]. In the tokenization process, stop words are removed according to the list of default stop words in spaCy, and the words are returned in their lemmatized form.

In general, the variety of most common words used when comparing different locations is a lot less than the variation in named entities. We can take a major city such as New York City, and compare it to Murfreesboro, Tennessee, which is a much less populated city in a different region of the United States. Table 3 and Table 4 compares the top five most commonly used words between these two cities for the week of March 1, 2019.

Table 3. Top 5 most used words for Murfreesboro, Tennessee for the week of March 1, 2019

Word	Frequency
like	186
know	160
love	141
go	139
get	136

Table 4. Top 5 most used words for New York, New York for the week of March 1, 2019

Word	Frequency
oh	133
know	90
get	72
love	66
yeah	62

We can see that there is not much of a difference between the two cities.

However, if we compare named entities that spaCy has identified in the lyrics, we can see much more of a difference. Table 5 and Table 6 list the most common named entities for the two cities. Note that spaCy's entity recognition relies heavily on capitalized words, so some words may have been misidentified as named entities. Attempting to transform all words into lowercase prior to passing them into the model results in an even more inaccurate recognition. In the end, we retained the capitalization of words as they were passed into spaCy. As a result, there are some false positives in the top five named entities that we present.

Table 5. Top 5 Named Entities for Murfreesboro, Tennessee for the week of March 1, 2019

Word	Frequency
Jesus	29
Morocco	13
Bubba	12
Tennessee	11
Hank	10

Table 6. Top 5 Named Entities for New York, New York for the week of March 1, 2019

Word	Frequency
Cupid	13
Time	5
Broadway	4
Comin	3
Eating	3

By examining the entities for Murfreesboro, we can see the influence of the demographics of that area. Murfreesboro is located in the Bible Belt of the United States, so it follows that "Jesus" is the most frequently occurring entity in songs that are popular there.

In addition to differences between cities, we can also see that entities change significantly over time as shown in Table 7 for the city of New York.

Table 7. Most Frequently mentioned Named Entities for New York, New York over time

Dates	Entities
2019-3-1	Cupid, Time, Broadway, Comin, Eating
2019-3-15	UN, Woah, Breaking, kick, Woah-oh
2019-4-1	Baby, Chase, Broadway, Mexico, the Cross Island Parkway
2019-4-15	Rosie, Moe $halizi, Moe, Rosalita, Halizi
2019-5-1	American, Brooklyn, Love, Rosie, Pedal

4.3 The Effects of Genres on Sentiment

Genre is important when it comes to classification of songs, so we ventured to see what insights we could gather with that in mind. Unfortunately, Spotify does not provide genres for an individual song, and instead assigns genres to artists. While it may not be perfectly accurate to assign the artist's genres directly to all of their songs, we can generally assume that there is not much deviation in genres for a specific artist's catalog of songs. Thus, each song in our lyrics data source is assigned the genres that its artist is associated with by querying the Spotify Web API. The song's VADER score is then assigned to each of those genres for our calculations. In Figure 5, we can see the average VADER score for each genre with at least twenty songs that have valid VADER scores.

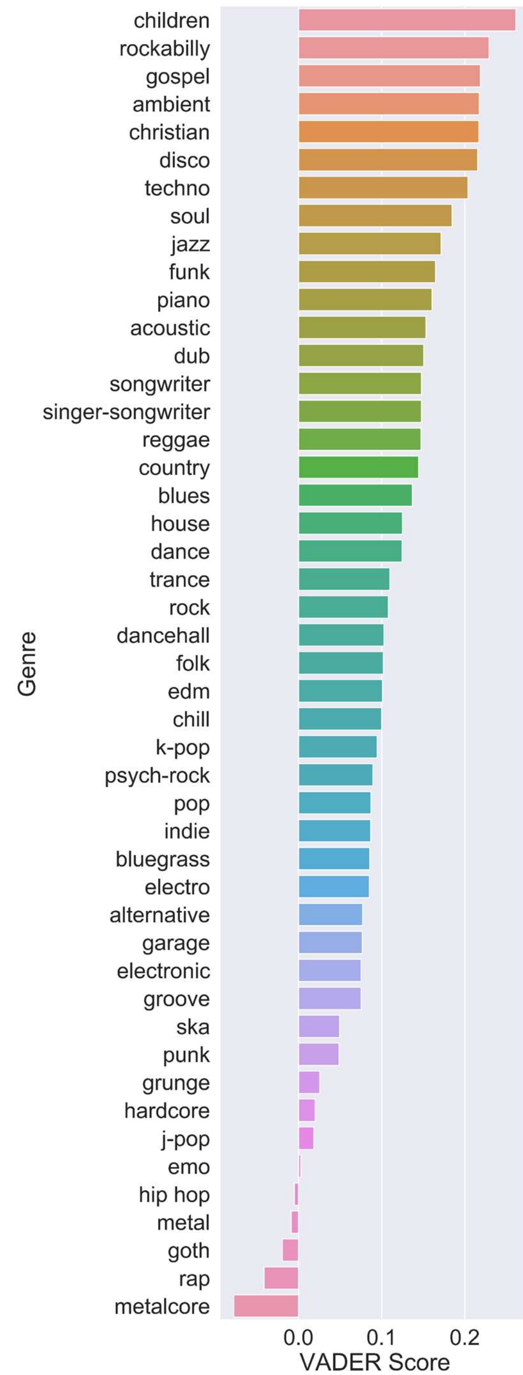

Figure 5. Average VADER scores for each genre

The average VADER score for a genre is rather intuitive. Genres such as Children, Gospel, and Christian are overwhelmingly positive, while Emo, Rap, and Metalcore are much more negative.

Now that we have determined that VADER scores can vary a lot by genre, we can take a look at how this might influence a city or country's overall score. Figure 6 and Figure 7 show the top five most popular genres for Indonesia and France in the week of April 15, 2019.

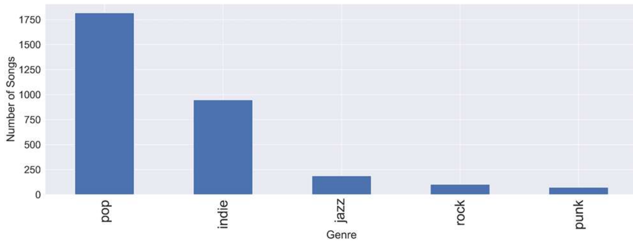

Figure 6. Top five most popular genres in Indonesia for the week of April 15, 2019

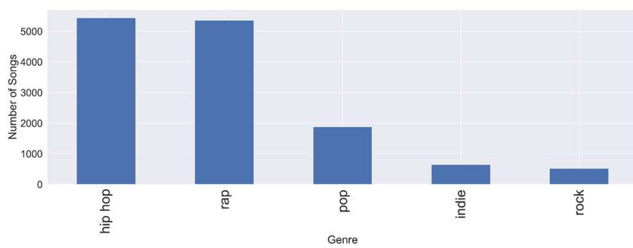

Figure 7. Top five most popular genres in France for the week of April 15, 2019

As we have determined before, Indonesia in in the top five countries with the most positive VADER scores, while France is in the top five most negative. When considering the average VADER score for a particular genre, we can see that the most popular genres in Indonesia tend to be more positive, while the most popular genres in France tend to be more negative. Though this is just a conjecture, it gives us a good insight into how a populace's taste in genres may significantly influence its overall sentiment score.

4.4 Identifying Themes in Lyrics

Word2vec is a commonly used word embedding method that assigns a vector representation for a word. It is very useful for detecting similarities between vectors, often by calculating the cosine similarity. Using this concept, we can specify a keyword or theme and check for any similarities among the lyrics in our data source.

For example, we can take the vector representation of the word "crime" and compare it against every word in the lyrics of a song. In our approach, we assign the maximum cosine similarity encountered to be the score for a song. For instance, if the lyrics were to contain the word "crime", the cosine similarity would be a value of 1, and thus the overall score for that song would be 1. Subsequently, we can take the average of all the scores for the songs in a playlist and have that be the thematic score for a specific location. Our approach is implemented with spaCy and its en_core_web_lg model. This model contains 685,000 unique GloVe vectors that are trained on Common Crawl [9, 10]. By using its similarity function, we can calculate the cosine similarity between two word vector representations. To test how effective this is, we present the following lyrics with their respective cosine similarity scores to certain themes.

In Figure 8, we can see that when there is a direct match to our theme, a maximum score of 1.0 is returned. In Figure 9, a very high score of 0.8340 is returned because of the word "economic", which is very closely related to our theme of "Economy" since it is in fact another form of the same word. This holds true when we change our theme to "Government", shown in Figure 10.

Title: Down

Artist: Pietro Lombardi

Cosine Similarity Score for "Economy": **1.0**

…

Don't you ever leave the side of me

And definitely, now probably,

And honestly, I'm down like the economy."

…

Figure 8. Relevant lyrics of "Down" by Pietro Lombardi for the theme of "Economy" [11]

Title: What Would Happen?

Artist: Pegz

Cosine Similarity Score for "Economy": **0.8340**

…

Would you make it worth your while, go tonight, take a bus, steal a posty's motorbike?

Would you globalize and let everybody know, we're all victims of economic growth?

…

Figure 9. Relevant lyrics of "What Would Happen?" by Pegz for the theme of "Economy" [12]

Title: Son of God

Artist: Chris Renzema

Cosine Similarity Score for "Government": **0.8198**

…

We have seen the glory of the Lord rise with broken lives restored

But not by governments nor fire

…

Figure 10. Relevant lyrics of "Son of God" by Chris Renzema for the theme of "Government" [13]

Title: Money Sprung (feat. Don Q)

Artist: A Boogie Wit da Hoodie

Cosine Similarity Score for "Government": **0.7393**

…

I'm ahead of the race

Flooring the gas, I'm passing the cops that I lead on a chase

If they come find out this weapon I placed

They might just give me a federal case

All of my folks has a residue taste

Rollie presidential when that presidential

Counting blue strips with the president face

…

Figure 11. Relevant lyrics of "Money Sprung" by A Boogie Wit da Hoodie for the theme of "Government" [14]

We can find more interesting results if we find the top similarity score that does not contain any of the word forms of "Government". In the lyrics for Figure 11, it is likely that our function picked up on words such as "federal" or "presidential" as words with a high cosine similarity to "Government".

Now that we have made sense of how our function is scoring these songs, we can move on to finding the overall thematic score for different locations. In order for us to have as much data as possible, we will calculate a single thematic score for a city or country using all the datasets that we have across all dates. Cities and countries that had the highest similarity score for the word "crime", are shown in Table 8 and Table 9 respectively. Note that places that do not have scores for at least thirty songs are excluded in the survey.

Table 8. Crime: Top 5 city similarity scores

City	Cosine Similarity Score
Sacramento California US	0.551192
Hergiswil CH	0.550793
San Francisco California US	0.541653
Secaucus New Jersey US	0.539204
Kingston JM	0.538189

Table 9. Crime: Top 5 country similarity scores

Country	Cosine Similarity Score
Jamaica	0.538189
Nigeria	0.495108
Switzerland	0.488291
Ghana	0.479834
United Kingdom	0.474853

We can also compare lyrics against another theme such as "economy", shown in Table 10 and Table 11.

Table 10. Economy: Top 5 city similarity scores

City	Cosine Similarity Score
Cricklewood GB	0.494464
Sumrall Mississippi US	0.490715
Edgware GB	0.489482
Kingston JM	0.488154
Surbiton GB	0.487170

Table 11. Economy: Top 5 country similarity scores

Country	Cosine Similarity Score
Jamaica	0.488154
Nigeria	0.480197
Ghana	0.472350
Malaysia	0.468016
United Kingdom	0.467734

What can be interpreted from these results is that the people of these cities and countries listen to songs with lyrics that express the themes of crime or economy the most out of any other locations. We can hypothesize that these themes are an important issue in these locations. To further illustrate this, all the American cities listed in Table 8 have an overall crime rate that is well above the national average or its state's average according to the FBI's Uniform Crime Reporting Program for 2017 [15].

One thing to note is that these average scores are not very high. One reason for this could be that in general, these themes are simply not very popular or often referred to when compared to other themes in music. To illustrate this, we can look at the top 5 countries and their scores for the much more commonly occurring theme of "Love".

Table 12. Love: Top 5 country similarity scores

Country	Cosine Similarity Score
Philippines	0.909973
Kuwait	0.909055
Northern Mariana Islands	0.904651
United Arab Emirates	0.897295
Guam	0.885811

We can see in Table 12 that there is a massive difference in maximum scores between "Love" and the other themes we presented.

Table 13. Similarity score distributions for different themes

Theme	Mean	Standard Deviation	Min	Max
Crime	0.4449	0.0234	0.4151	0.5382
Economy	0.4521	0.0113	0.4292	0.4882
Love	0.904651	0.0503	0.6825	0.9100

We can also look at the distribution statistics for these themes as well, shown in Table 13. The range, minimum, and maximum of values for “Crime” and “Economy” are much smaller than “Love”. The smaller standard deviation tells us that the significance of those themes is portrayed in much smaller numerical values.

5. Conclusions and Future Work

We have presented an analysis of song lyrics with an added geographical and societal context. Using natural language processing tools, we performed analyses that help us determine how the lyrics of songs that are uniquely popular to a specific population can reflect their sentiment or ideals. Musicality may also be a major factor especially when determining positive or negative sentiment as we have seen in the case of music genre. The results of this study may give more credence to the idea that music can be a universal language that reflects the voice of the people.

Future work would include lyrics in other languages, as this would be very important for the proper representation of locations where English is not the dominant language. As time goes on, we may also be able to improve upon the data available for English songs as well. The method in which Spotify chooses the songs for these playlists is weighted towards newer and less popular songs, which means that lyrics are often not available at the time when these playlists are populated. Thus, we have the opportunity to obtain more lyrics for existing datasets as they are eventually crowd-sourced on LyricWiki and Genius in the future. To further delve into the study of how these lyrics reflect society, we could also make comparisons between the sentiment expressed in these songs and other sources of sentiment such as Twitter or the news.

6. Acknowledgements

This report was conducted as part of an Independent Study in Computer Science at Rutgers University–New Brunswick under the guidance of Professor Gerard de Melo.

7. References

[1] Eliot Van Buskirk. 2015. Musical Map: Cities of the World. (July 2015). Retrieved May 10, 2019 from https://insights.spotify.com/us/2015/07/13/musical-map-of-the-world/

[2] Enrico Bacsis. 2018. LyricWikia. Retrieved from https://pypi.org/project/lyricwikia/

[3] John W. Miller. 2019. LyricsGenius. Retrieved from https://pypi.org/project/lyricsgenius/

[4] Marco Lui. 2016. Langid. Retrieved from https://pypi.org/project/langid/

[5] Eliot Van Buskirk. 2013. Plotting Music’s Emotional Valence, 1950-2013 (Nov. 2013). Retrieved May 10, 2019 from https://insights.spotify.com/no/2013/11/05/musics-emotion-over-time/

[6] Hutto, C.J. & Gilbert, E.E. 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

[7] Folium. Retrieved from https://pypi.org/project/folium/0.1.5/

[8] Honnibal, Matthew and Montani, Ines. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

[9] spaCy. Retrieved May 12, 2019 from https://spacy.io/models/en

[10] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. Retrieved May 14, 2019 from https://nlp.stanford.edu/projects/glove/

[11] Pietro Lombardi. “Down”. 2008. Kamaljit Jhooti, Dwayne Carter, Jared Cotter, J-Remy, Bobby Bass and Jonathan Perkins. Genius. Retrieved from https://genius.com/Pietro-lombardi-down-lyrics

[12] Pegz. “What Would Happen?”. 2005. Genius. Retrieved from https://genius.com/Pegz-what-would-happen-lyrics

[13] Chris Renzema. “Son of God”. 2018. Genius. Retrieved from https://genius.com/Chris-renzema-son-of-god-lyrics

[14] A Boogie wit da Hoodie and Don Q. “Money Sprung”. 2017. Retrieved from https://genius.com/A-boogie-wit-da-hoodie-money-sprung-lyrics

[15] United States Department of Justice, Federal Bureau of Investigation (2017). *Crime in the United States, 2017.* Retrieved May 14, 2019 from https://ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017