Elucidation of the Relationship Between a Song's Spotify Descriptive Metrics and its Popularity on Various Platforms

Laura Colley[1,†], Andrew Dybka[1,†], Adam Gauthier[2,*], Jacob Laboissonniere[1,*],
Alexandre Mougeot[2,*], Nayeeb Mowla[1,*], Kevin Dick[1], Hoda Khalil[1], Gabriel Wainer[1]
[1]Department of Systems & Computer Engineering
[2]Department of Electrical Engineering
Carleton University, Ottawa, Ontario
[†,*] equal contribution
{lauracolley, andrewdybka}@cmail.carleton.ca,
{kevin.dick, hoda.khalil}@carleton.ca, gwainer@sce.carleton.ca

Keywords - Spotify, music, popularity, metrics

***Abstract*—This project involved searching for correlation between descriptive song metrics provided by Spotify and song popularity on various internet platforms. Each platform has a different definition of what makes a song popular, so relationships between other platform metrics and Spotify metrics were analyzed. The platforms worked with were Google, Twitter, YouTube, TikTok, WhoSampled and Billboard. Additional data was pulled from each platform and combined with a Spotify data set. After analysis of the data, there was no significant correlation between any metrics and a song's popularity on a platform. https://github.com/Jacoblab1/SYSC4906 [1]**

I. Introduction

The purpose of this project is to determine if there are certain features of music that cause them to be popular on various platforms. Success on various platforms can greatly assist a song's popularity, so the researchers seek to learn whether specific qualities in music can impact this success. Spotify, the music streaming service, provides descriptive metrics for songs on its platform. Determining what makes songs popular on different platforms can assist artists in targeting platforms for success. The different platforms looked at are Google, Twitter, TikTok, YouTube, WhoSampled and Billboard. Each platform is given its own section about data collection and processing, analysis, and results.

II. Problem Definition

Songs are used for different reasons on various platforms. A platform is an Internet service such as a search engine or social network. Spotify is a music streaming service that provides various metrics describing qualities of the song [2]. Understanding what makes a song popular on different platforms where each platform has a different definition of popularity can provide insight to what kind of song becomes popular. Understanding what makes a song popular on different platforms can provide artists a framework to create new songs for a better chance of them becoming popular. A data set containing songs from Spotify along with its metrics was used as a base [3] and data for additional platforms was collected separately.

III. Platforms

A. Spotify

1) Design & Methodology: The Spotify data was analyzed to determine the distribution of features and their covariance. To begin, the data was checked for null values and descriptive statistics were performed to understand the average and spread of different features. Histograms were plotted to visualize the features and a correlation matrix was used to determine the amount of dependence between features. The data was copied and preprocessed by dropping the non-numeric columns and discrete columns in order to do regression analysis.The values were standardized and an 80:20 training and testing set was created. The features were used to predict the Spotify song popularity using linear regression and a number of more complicated models. The criteria for success was to create a model that predicts the popularity with low error.

2) Results & Discussion: Figure 1 shows the individual distributions for most features in the Spotify dataset. As we can see in the figure, each feature follows a different distribution. From the histograms we noticed that the majority of songs were not explicit, had low instrumentation, had low speechiness and were mostly recent songs The correlation matrix showed that the features were linearly uncorrelated, although some features showed some correlation such as loudness and energy which is expected. Linear regression was performed to start as the most basic and interpretable model and for each regression model the mean absolute error was computed for the testing set. The linear model showed that the year feature had the largest coefficient and was most responsible for the popularity. This is likely due to users preferring songs that are newer and were produced closer to the present, since old music is often out of style. The mean absolute error for the linear model was 0.6158 and the Pearson correlation coefficient

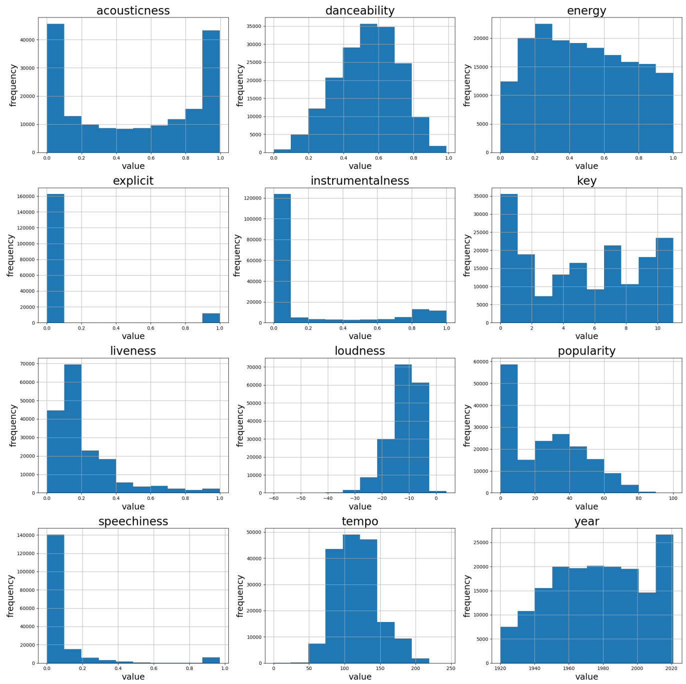

Fig. 1. Spotify Features

squared was computed to be 0.3586. Other regression models like the polynomial regression, support vector machines, nearest neighbors, decision tree, and random forests were tested and their absolute error computed for comparison. The random forest model had the best results when tested using the Spotify data and the linear model had the worst results. This is expected since the linear model is the most basic and has the least flexibility. Overall, none of the models performed that strongly as the R^2 score for the random forest was 0.65. This indicates that the Spotify metrics were not strong predictors of the song popularity on Spotify. This may be due to the method by which Spotify produces these metrics, as this analysis relies on their data generating methods. Overall, the models could be improved by attempting hyperparameter tuning. It would also be valuable to try a neural network instead to compare the performance against the other models.

B. Google

1) Design & Methodology: Google trends is a platform that records and analyzes popular search results through Google. Trending data can be analyzed over time, by region and by category. Trending data is available from 2004 until present. The API used to collect the data was PyTrends [4], an unofficial open source interface. The PyTrends API is used by sending payloads of up to 5 keywords at a time. Requests to the trending information is rate limited to around 2000 requests per day, however the actual amount is unspecified. After reaching the rate limit, requests must be sent with 60 seconds spacing. Since the Spotify data was large, a subset of songs with popularity greater than 75 was taken in order to focus on the popular songs that have a high probability of

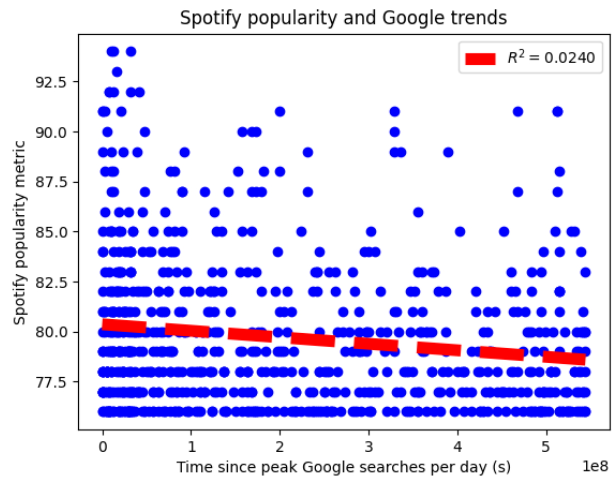

Fig. 2. Google Trends correlation with Spotify popularity

being searched. 1000 songs were randomly sampled from this set and the interest over time data was collected from Google trends. Histograms and correlation matrices were plotted to view how well the sampled dataset represented the total dataset. For some songs, no searches were performed resulting in no trending result. This is likely due to the song name having additions such as features or atypical characters. Songs that did not return a result were discarded leaving 770 songs. From the interest over time data, the date where the search term peaked was recorded and the time difference between the peak day and the present was computed. In theory, songs that peak in searches closer to the present should be more popular on Spotify. A linear regression was performed to find if the popularity could be predicted from the date it had the most Google searches.

2) Results & Discussion: Figure 2 shows the correlation between the Spotify popularity and the time since the song had the most searches in one day. From the linear fit, we can see the Spotify popularity showed no dependence on the Google trends data as the fit had a slope of approximately zero. The Pearson correlation coefficient squared was nearly zero, indicating that there was no correlation between these variables. This result may be due to the method the data was sampled from the larger data set. If songs with lower popularity were included a relationship might emerge, however, songs with low population are not frequently searched and are unlikely to provide Google trends data. Another issue was that many songs have generic names which are used for many things. Therefore, the peak search date may not represent song searches. The Google trend peak search time was compared against the other features in the Spotify dataset and showed no correlation.

C. WhoSampled

1) Design & Methodology: WhoSampled.com provides information about song's samples, covers, and remixes [5]. The goal for this subsection is to determine if there is a correlation between a Spotify metric and WhoSampled categories. These categories include the number of samples a song contains, how many times a song is sampled, if the song is a cover of another, if the song has been covered, if the song is a remix of another or if the song has been remixed. This would identify trends for what type of songs are most common in each of the 6 categories mentioned. To solve this problem, data was searched for and scrapped from WhoSampled from the list of songs in the Spotify data set then combined. Requests were sent to WhoSampled endpoints containing song and artist information. The HTML response was searched for the information needed, the data found was appended to the row of data used to make the request.

When building and testing the scraping code, the home Internet IP in use was blocked after too many requests were made too quickly. WhoSampled only allows access to their metadata for commercial use or academic use of PhD level or higher. From here, requests were sent using the Carleton VPN and a 4 second delay was placed between each request, resulting in 20,000 songs from the 100,000+ song Spotify dataset. Some were also duplicates, so only about 10,000 songs were collected for analysis.

Correlation trends were searched for between all metrics and all WhoSampled categories and comparisons of metrics were made between songs that fell under a WhoSampled category and those that did not. Spearman coefficients were used to better handle outliers. Scatter plots of raw data and line graphs of averages were plotted for relations with the highest spearman value. Average Spotify metrics were also compared in tables between songs which were in WhoSampled categories and songs which were not.

2) Results & Discussion: There was no clear correlation between a song in a WhoSampled category and a Spotify metric. This may be due to lack of data, since most songs did not fall under WhoSampled categories and there was limited data collected from WhoSampled due to the request rate. There was also no significant difference between average Spotify metrics for songs under WhoSampled categories and songs that were not.

Metric	Sampled	+/-	Not Sampled	+/-
accousticness	0.22	0.014	0.22	0.006
danceability	0.65	0.009	0.60	0.003
energy	0.63	0.011	0.68	0.004
instrumentialness	0.04	0.010	0.17	0.006
liveness	0.19	0.010	0.23	0.004
loudness	-7.09	0.193	-7.38	0.070
popularity	54.41	1.520	37.91	0.644
speechiness	0.11	0.006	0.11	0.002
tempo	120.32	1.552	124.02	0.555
valence	0.47	0.0135	0.45	0.005

TABLE I
AVERAGE SPOTIFY METRICS FOR SONGS THAT HAVE BEEN SAMPLED AND NOT BEEN SAMPLED

D. TikTok

1) Design & Methodology: The popularity of a song on TikTok can be measured via multiple factors: share count, comment count, play count, and like count ("diggCount"). In this analysis, all four of these TikTok metrics were potential measures of popularity. The linear correlation between each TikTok metric and each Spotify metric was measured to determine any potential links. Success was defined as determining any existing correlations.

To gather TikTok data, the tiktok-scraper by drawrowfly on GitHub was used [6] which utilizes the TikTok API. Originally, the planned approach was to cross-reference each song in the given Spotify dataset with the most recent TikToks that utilized the song and determine correlations between the metrics. However, this approach was not found to be feasible due to the limitations of the TikTok scraper. The scraper's "music" method, which gathers videos that use any given song, takes a unique numeric song ID as an identifier; these IDs are not searchable within the TikTok desktop app. Instead, the TikTok scraper's "trend" method was instead used to obtain the top 20,000 currently trending videos; these videos were then filtered to remove those which used original sound. The videos that remained were cross-referenced with the Spotify dataset to gain corresponding song metrics. Outliers that contained diggCount, playCount, commentCount, or shareCount metrics more than three standard deviations from the mean were removed.

In order to visualize correlation, scatter plots for each combination of metrics were created. In addition to this, a correlation matrix containing all variables was obtained.

This methodology undoubtably introduced bias towards more popular videos, as it only looks at TikTok's trending videos. This also excludes many of the songs in the Spotify dataset. The method by which the TikTok algorithm determines what is "trending" is also unknown, and could also be a source of bias.

There were no risks or ethical concerns present within this methodology.

2) Results & Discussion: No significant correlation was discovered between any of the Spotify song metrics and any of the TikTok popularity metrics. This can be particularly visualized using the playCount. This was identified as the most interesting TikTok popularity metric because it was highly correlated with diggCount (Pearson coefficient = 0.86), and requires the lowest effort from TikTok users. The scatter plots of playCount vs. Spotify metrics are shown in Figure 3.

E. Twitter

1) Design & Methodology: The "engagement" of a song on Twitter can be defined by the amount of Tweets mentioning that song as a topic. The engagement can also include the total number of likes, quotes, replies, and retweets that the Tweets mentioning a song get.

Using our Spotify dataset, we wanted to answer the following question:

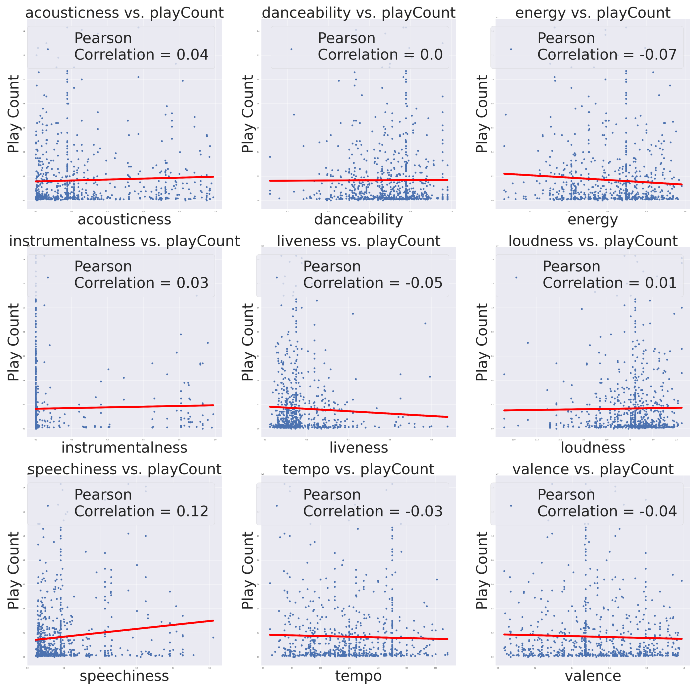

Fig. 3. TikTok playCount vs. Spotify song metrics

Do correlations exist between a song's Twitter engagement and its Spotify metrics?

The approach to solve the problem was to scrape Twitter to create a Twitter engagement dataset for the songs. This was done using Twitter API v2 which has a rate limit of 450 requests allowed every 15 minutes for Standard users. Each search query request also returns a maximum of 100 Tweets from the past 7 days [7]. Due to these limitations, the original Spotify dataset was first cleaned. The Spotify dataset was reduced to a subset of 2250 songs in the past 5 years and songs that have a popularity greater than 50. The rationale for this was that songs that are more recent and popular might have more noticeable engagement on Twitter in the past 7 days. From there, each song from the dataset was queried via the Twitter API, returning metrics of engagement. A Spearman correlation matrix and scatter plots were used to visualize if the Twitter metrics of songs correlate to the year, popularity, danceability, etc. described by Spotify.

This method uses the official Twitter API and does not involve any risks and/or ethical concerns.

2) Results & Discussion: Figure 4 shows no significant correlation was found between the engagement of songs on Twitter and any of Spotify's metrics from the data which was collected. This is seen as the correlation is never greater than +/-0.03 between any of the metrics. Some significant biases in this analysis that have to be considered is the fact that Twitter API only limits the search to Tweets from the past 7 days. 7 days is a very small window when considering discussions about songs in the past 5 years, which may have affected these results. A future extension of this analysis may involve using Twitter Premium / Enterprise API. This version allows for queries further back than only 7 days and allows for more requests.

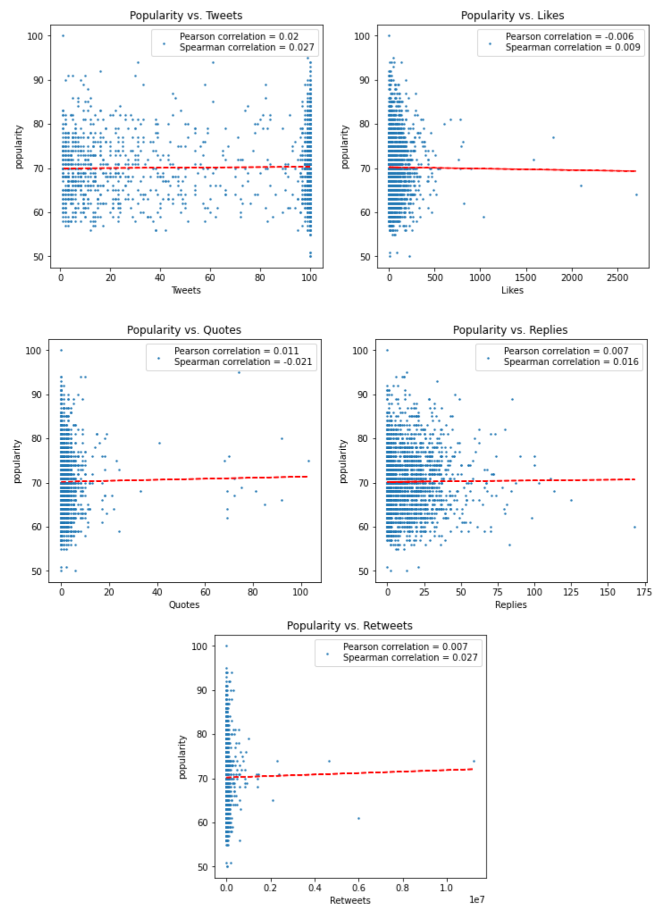

Fig. 4. Spotify's popularity metric vs. Twitter metrics for every song

F. YouTube

1) Design & Methodology: For this analysis, the popularity of a song on YouTube is the number of views for the top search result for the name of the song on Spotify. To compare the popularity of a song on YouTube with Spotify attributes, it is first necessary to gather the view count for each video.

Initially, the plan was to use the YouTube API to gather view counts for each song in the Spotify dataset. After further research, it was not used due to rate limitations that cap usage at 50 video searches per day [8].

With the official YouTube API out of the question, a web-scraper could be used to gather video view counts. YouTube is a JavaScript-heavy website, so scraping it requires a web-scraper that is capable of rendering JavaScript. Selenium Webdriver is a browser testing framework that can render JavaScript-based web pages. Selenium Webdriver was able to scrape the contents of a YouTube search results page without difficulty.

Selenium Webdriver iterated through the Spotify dataset songs and retrieved their view count. Limitations due to

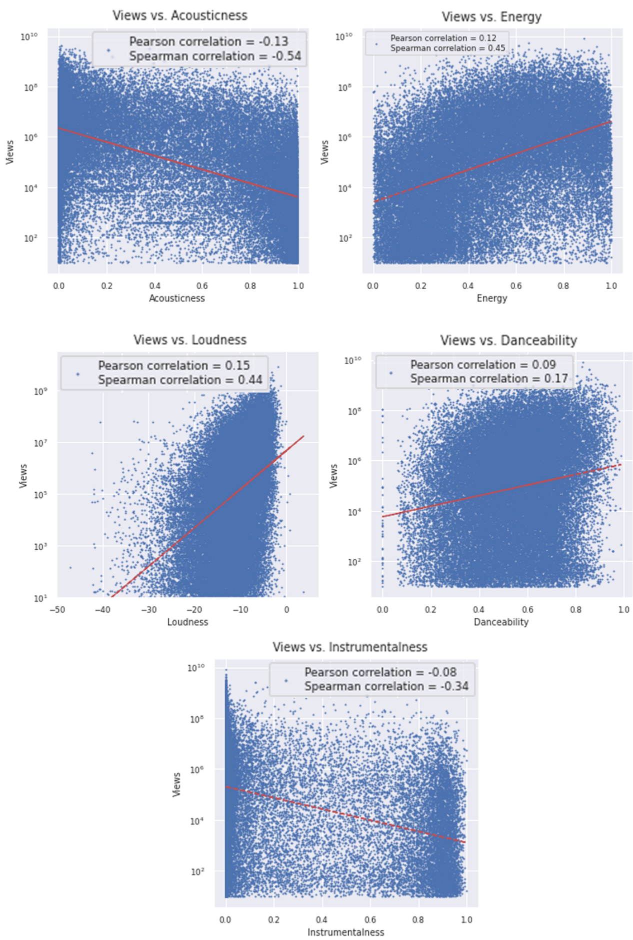

Fig. 5. YouTube Views vs. Spotify Attributes

scraping speed restricted view count collection to 50,000 of the 174,000 songs in the Spotify dataset.

YouTube view counts were then merged with the Spotify dataset for further analysis. Scatter plots and correlation tests were then used to further analyze the data.

2) Results & Discussion: A Pearson correlation matrix illustrated which Spotify attributes correlated with popularity on YouTube. Upon inspection of the correlation matrix, it appeared that acousticness, energy, loudness, danceability, and instrumental-ness had Pearson correlations of significance.

Scatter plots were then used to compare the YouTube view count with each of these attributes of significance. Labels on the scatter plots provide both the Pearson Correlation and the Spearman Correlation.

Figure 5 shows Pearson Correlations for each attribute that indicate a "very-low to no-level" of correlation. Figure 5 also displays Spearman Correlations which indicate that certain attributes have some correlation with YouTube popularity.

Energy and loudness both have a positive medium-level correlation with YouTube popularity. This suggests that loud, energetic songs are likely to garner more views on YouTube.

Acousticness and instrumentalness both have a negative medium-level correlation with YouTube popularity. This suggests that acoustic or instrumental songs are likely to receive fewer views on YouTube.

Danceability had a Spearman Correlation of 0.17, indicating a low level of correlation to YouTube popularity.

Spearman Correlation is less sensitive to extreme outliers than the Pearson coefficient. Certain videos have billions of views, and others have no views. It is likely that the Spearman Correlation better accounted for cases with extreme view counts.

G. Billboard

1) Design & Methodology: The billboard top 100 is a weekly list that compiles the top 100 songs of each week. The rankings are based on both physical and digital sales, online streams and radio plays in the United states. This subsection focuses on finding out if there is a correlation between what makes a song popular on the billboard top 100 and Spotify metrics. To do so, two different data sets were needed, a list of songs from Spotify and their respective metrics, and a list of songs that have made it onto the billboard top 100.

While doing research on different methods of collecting the billboard data, a data set was found. The data compiled songs that had appeared on the billboard top 100 chart from the early 1940's to 2020 [9]. It included five different metrics for each song but it was decided that only one would be used for analysis. The metric that would be used is the total amount of weeks a song has made it onto the chart.

The data would need to be cleaned to be used for this project since a song would be repeated each week that it made it onto the top 100 chart. For the purpose of this analysis, the last time a song made it onto the chart was the only instance that was needed. The data was cleaned only kept the last instance of a song making it onto the chart. All songs that didn't have a value in the weeks on the chart column were dropped as well. The Spotify data was then cleaned up, songs that had a popularity value of less than 5 were dropped as there was a disproportionate amount of them which skewed the data. From there, all songs with matching titles and performers were merged into a singular set. This left 9 829 songs that could be used for analysis.

2) Results & Discussion: After the data was cleaned, analysis began. It started off with a pearson correlation matrix to find if there were any interesting correlations between metrics to further look into. No substantial correlation between the top 100 chart metrics and the Spotify metrics were found. The two metrics with the highest correlations were made into scatter plots in figure 6.

The amount of weeks a song has made it onto the top 100 chart had a positive correlation of about 0.32 with the popularity metric of a song and a positive correlation of 0.29 with the year the song was released. From these plots, we can see that it seems that newer songs are lasting longer on the charts and that the popularity of a song affects how long a song will stay on the top 100 charts.

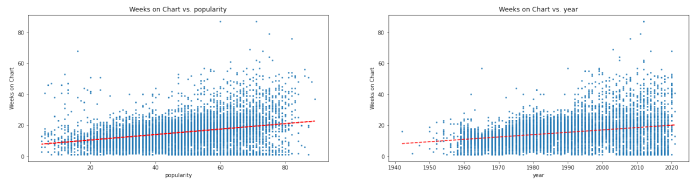

Fig. 6. Weeks on Chart vs. Spotify Attributes

IV. Conclusion

Across all platforms no significant correlation between any Spotify metrics and platform popularity metrics were found. This reinforces the existing idea that music popularity is not a science and cannot be predicted accurately. Further investigation could be done to investigate other correlations; or expanding the data to include more songs, or limiting songs to those released in a certain time frame, could potentially provide better results.

References

[1] "Sysc4906." [Online]. Available: https://github.com/Jacoblab1/SYSC4906
[2] Ashrith, "What makes a song likeable?" accessed: 2021-04-01. [Online]. Available: https://towardsdatascience.com/what-makes-a-song-likeable-dbfdb7abe404
[3] Y. E. Ay, "Spotify dataset 1921-2020, 160k+ tracks," accessed: 2021-04-01. [Online]. Available: https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks
[4] "Pytrends," accessed: 2021-04-03. [Online]. Available: https://pypi.org/project/pytrends/
[5] "Who sampled," accessed: 2021-04-04. [Online]. Available: https://www.whosampled.com/
[6] "tiktok-scraper," accessed: 2021-04-03. [Online]. Available: https://github.com/drawrowfly/tiktok-scraper
[7] "Twitter api," accessed: 2021-04-04. [Online]. Available: https://developer.twitter.com/en/docs/twitter-api
[8] "Youtube api," accessed: 2021-04-01. [Online]. Available: https://developers.google.com/youtube/v3
[9] S. Miller, "Billboard hot weekly charts," accessed: 2021-04-01. [Online]. Available: https://data.world/kcmillersean/billboard-hot-100-1958-2017

V. Appendix A

Unfortunately feedback was not as easy to incorporate. To formulate more specific questions for each platform a correlation between a metric and popularity was looked to be found to formulate a more specific question investigating the correlation. Since no platform had any significant correlation with a metric, the time was mainly spent searching for the correlation to base a more specific question.