

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,*}, Nayeef Mowla^{1,*}, Kevin Dick¹, Hoda Khalil¹, Gabriel Wainer¹

¹Department of Systems & Computer Engineering

²Department of Electrical Engineering

Carleton University, Ottawa, Ontario

^{†,*} equal contribution

{kevin.dick, hoda.khalil}@carleton.ca, gwainer@sce.carleton.ca

Abstract—The music industry and personal music consumption have evolved dramatically with the advent of streaming platforms. In this evolving landscape, there is considerable interest in understanding what factors contribute to a song's popularity. Extrinsic (*i.e.* non-acoustic) features of a given song, such as the record label, and/or intrinsic (*i.e.* acoustic) features such as its energy may contribute to popularity on a given digital platform. In this work, we, for the first time, sought to systematically study how a song's Spotify acoustic descriptive features correlated with popularity metrics on various Internet platforms. Since each platform defines "popularity" according to platform-specific metrics, a large-scale correlation-based analysis was generated. The digital platforms considered in this article are Google Trends, WhoSampled, TikTok, Twitter, YouTube, and the Billboard Top-100. Platform-specific scrapers were created and all data was aggregated with the Spotify Echo Nest dataset of descriptive acoustic metrics. While the majority of correlations were unremarkable considering both Spearman and Pearson coefficients, a number of corroborating and contradictory findings resulted, with notable implications for acoustic features on various digital platforms. Notably, the YouTube view count was found to be positively correlated to the Spotify song popularity ($\rho = 0.822$), year ($\rho = 0.600$), and energy ($\rho = 0.455$) and moderately negatively correlated to acousticness ($\rho = -0.542$) and instrumentality ($\rho = -0.345$). All reproducing code and aggregated data from this work are open-source for use by the broader research community.

Keywords Song Popularity · Spotify · Web Scraping · Acoustics

I. INTRODUCTION

With the eruption of technology and social media promoting information sharing and distribution, the music industry and the way individuals consume music has evolved drastically in recent years. A key inflection point in 2018 saw streaming become the main form of music consumption for the first time, accounting for 47% of the music market, as reported in the International Federation of the Phonographic Industry annual report [1].

Music popularity has evolved from the tracking of album sales to now considering more nuanced factors such as the number of times a song is streamed, shared, searched for, and sampled. Amid these changes, music remains big business: the

global recorded music market produced US \$21.6 billion of revenue in 2020, with \$13.4 billion of this total coming from streaming platforms such as Spotify [2]. It follows that there exists substantial financial motivation to investigate the factors that contribute to a song's popularity. To that end, a number of features characterizing a given song can be considered including non-acoustic features (*e.g.* the record label), acoustic features (*e.g.* song *energy*), or both.

Investigation of non-acoustic features and their impact on song popularity is an ongoing area of research. A 2008 study compared factors such as record label, song genre, and a "star" variable to song survival time on Japanese charts. The study concluded that an artist being a so-called "star" had a positive effect on survival time, but that the record label did not have a significant impact [3]. The advent of the Internet age and the resultant proliferation of data has expanded opportunities for research into musicology and factors influencing popularity within this realm. For example, it has been reported that extrinsic factors influence future album sales, including a positive correlation with the volume of album-related blog posts [4] and that a song's presence on Spotify's Top-50 list can be predicted, with 88.49% accuracy, from factors such as the song's previous ranks, its "explicit" flag, and the artist name [5]. Related studies have also investigated the utility of song recommendation based on user Tweets [6]. While these studies have yielded promising results, the open question remains of whether intrinsic (*i.e.* acoustic) properties of a given song influence popularity on a given digital platform.

The impact of acoustic properties upon a song's popularity has been studied by members of the research community over recent decades. Notable work done in this area includes Berlyne's seminal work on Aesthetics and Psychobiology [7]. Therein, Berlyne hypothesizes that individuals prefer moderately arousing music, such as songs with a moderate degree of energy [7]. According to this theory, music with either too low or too high levels of "arousal" will be more broadly disliked, leading to a so-called inverted-U relationship between music popularity and arousal potential [7]. Although

this theory is contentious [8], it has nevertheless been used as a basis for subsequent research on music popularity. North *et al.* investigated this hypothesis while studying the impact of song energy on popularity in both the UK [9] and the US [10]. They found differing results between the countries: while energy and popularity seemed to have a U-shaped relationship in the UK (rather than an inverted-U [9]), results in the US were mixed [10]. While intriguing, these studies are ultimately limited by their narrow focus on song energy alone.

Investigating more nuanced acoustic factors beyond song energy may lead to more actionable insights. Work by Pachet and Roy in 2009 utilized statistically-generated song features from the MPEG format [11]. This work found that while some subjective labels such as mood could be reasonably predicted from these features, popularity could not [11]. These results were corroborated by Lee *et al.*, who also found that automatically-generated features from the MPEG format and from perceived frequencies did not perform well as a predictor of Billboard popularity [12]. Lee *et al.* also found, however, that their own calculated features of songs, particularly *harmony* and *arousal*, performed well as singular popularity predictors via Billboard charts [12]. Another study utilized musical features as calculated by AcousticBrainz and measured popularity via the UK Top-100 Singles chart. This study found that successful songs were *happier* and had a brighter *timbre* than non-successful songs [13].

Some of the most promising results in this space have resulted from utilizing musical features provided by Echo Nest, a musical intelligence company owned by Spotify since 2014. These features include, but are not limited to: *acousticness*, *danceability*, and *energy*. These features are leveraged within Spotify's recommendation algorithms to suggest new user-specific songs. Research has found that where these features indicate more upbeat songs, such as songs with higher tempo and danceability, the result is often a higher peak position on the Billboard Top-100 chart and a longer duration upon the chart [14]. Machine learning models trained on binarized Echo Nest features have been reported to obtain accuracy, precision, and Area under the Receiver Operating Characteristic Curve (AUC) metrics above 80% in all cases [15]. Promising results have also been obtained by adding non-acoustic features, such as the "star" variable and previous chart data to these acoustic ones [5], [13], [14], with accuracy increasing progressively.

In this work, we build upon the results reported above by investigating potential correlations between the Echo Nest musical factors utilized by Spotify and various metrics for song popularity as defined upon different platforms as part of a systematic analysis across all acoustic factors. The platforms considered are Google Trends, WhoSampled, TikTok, Twitter, YouTube, and the Billboard Top-100. While previously cited works have made substantial use of the Billboard Top-100 as an indicator of popularity, our systematic analysis of song popularity across other contemporary Internet platforms is considered a novel application.

II. PLATFORMS, DATA, & METHODOLOGY

The data acquisition, aggregation, transformation, and analysis pipeline used in this work is visualized in Fig. 1. A data set containing songs from Spotify along with its metrics was used as a base [16] and data for additional platforms was collected separately. Important to this work was the consideration of employing ethical scraping practices [17]. All scrapers and data are released in the GitHub repository to ensure the replicability of this work [18].

A. Spotify

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 computed to observe the mean and variance of different features. Histograms were plotted to visualize the features and a correlation matrix was used to determine the interdependence between features. The data was copied and pre-processed by dropping the non-numeric columns and discrete columns in order to perform regression analysis. The values were standardized and an 80:20 training and testing set was created. The features were used to generate models to predict the Spotify song popularity. These models entailed linear regression, polynomial regression, Support Vector Machines, Nearest Neighbors, decision tree, and random forests. The criteria for success was producing a model that predicts the Spotify popularity with sufficiently high accuracy.

B. Google Trends

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 the present. The API used to collect the data was PyTrends [19], an unofficial open source interface. The PyTrends API is used by sending payloads of up to five keywords at a time. Requests to the trending information is rate-limited to around 2,000 requests per day, however the actual amount is unspecified. After reaching the rate limit, requests must be sent with a spacing of 60 seconds. Since the Spotify dataset was large, a subset of songs with popularity greater than 75 was taken in order to focus on popular songs with a high probability of being searched.

We randomly sampled a representative 1,000 songs 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 definitive trending result. This is likely due to the song name having additions such as features (*e.g.* "feat.") 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

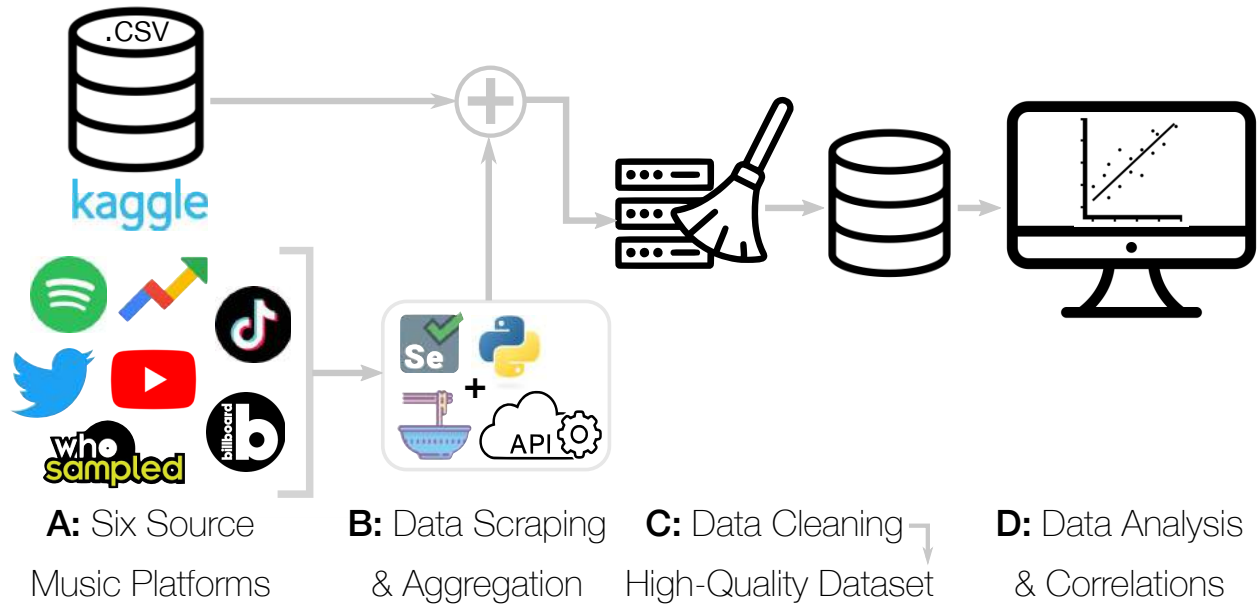


Fig. 1. **Conceptual Overview of the Data Acquisition, Aggregation, Transformation, and Analysis Pipeline.** In (A) & (B), platform-specific scrapers leveraged Cloud APIs, the Python programming language, and the Selenium and BeautifulSoup libraries to acquire song-specific popularity metrics. These were concatenated with the Spotify acoustic descriptive metrics and in (C), were post-processed to a high-quality dataset. These were then systematically analyzed in (D) to identify any notable correlations indicative that particular metrics might be predictive of song popularity on a given platform.

was computed to find if the popularity could be predicted from the date a particular song had the most Google searches.

C. WhoSampled

WhoSampled.com provides information about songs' samples, covers, and remixes [20]. The goal of this part of the study was to determine if there was a correlation between any Spotify metric and any WhoSampled category. 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 analysis would identify acoustic trends for what type of songs are most common in each of the six categories mentioned. To address this problem, data was queried and scraped from WhoSampled from the list of songs in the Spotify data set and then aggregated. Requests were sent to WhoSampled endpoints containing song and artist information. The HTML response was queried and the resultant data was aggregated.

When building and testing the scraper, it was determined that the IP in use was blocked using a rate-limiter following an excessive number of requests. WhoSampled specifies access to their metadata for commercial use or academic use at or above the doctoral level. In accordance with the ethical scraping practices outlined in [17], requests were sent using an academic VPN and a four-second delay between each request, resulting in 20,000 songs from the 100,000+ song Spotify dataset. These were then deduplicated producing ~10,000 songs for analysis.

Correlation trends were computed 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. Both Pearson and Spearman coefficients were computed, with the latter serving to better account for 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.

D. TikTok

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, we used an open source scraper [21] that leveraged 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 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. Therefore, we used the TikTok scraper's "trend" method to obtain the top 20,000 currently trending videos.

The videos were filtered to eliminate videos that used original sound. The videos that remained then were cross-referenced with the Spotify dataset to gain corresponding song metrics. In the data cleaning process, we removed outliers that contained diggCount, playCount, commentCount, or shareCount metrics more than three standard deviations from the mean. This resulted in a final dataset of 358 unique songs matched to 783 total TikTok videos.

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 biases our analysis towards more popular videos as it only considered currently trending TikTok videos. The process also excluded many of the songs in the Spotify dataset. The TikTok algorithm that determined what was “trending” is also unknown and could be an additional source of bias.

E. Twitter

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 received.

Using our Spotify dataset, we wanted to answer the following question: *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 accomplished 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 seven days [22]. Due to these limitations, the original Spotify dataset was first sampled. The Spotify dataset was reduced to a subset of 2,250 songs in the past five years with popularity greater than 50. We rationalize this choice given that songs that are more recent and popular are likely to have more noticeable engagement on Twitter in the past seven days, enabling the elucidation of the putative trend. Thereafter, 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 correlated to the *year, popularity, danceability, etc.* described by Spotify.

F. YouTube

For this analysis, the popularity of a song on YouTube was defined as the number of views on the video that was 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 was first necessary to gather the view count for each video.

Initially, we sought to use the YouTube API to gather view counts for each song in the Spotify dataset; however, the API enforces request limitations that cap usage at 50 video searches per day [23]. Consequently, we developed a web-scraper to collect video view counts. YouTube is a JavaScript-heavy website requiring a web-scraper that was capable of rendering

JavaScript. To this end, we leveraged the Selenium Webdriver, a browser testing framework that can render JavaScript-based web pages, to scrape the contents of YouTube search results.

The Selenium Webdriver iterated through the Spotify dataset songs and retrieved their view count. Limitations due to scraping speed restricted view count collection to a random subsample of 50,000 of the 174,000 songs in the Spotify dataset. Finally, 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.

G. Billboard

The Billboard Top-100 is a weekly list that compiles the top 100 songs for each week in the United States. The rankings are based on both physical and digital sales, online streams, and radio plays in the United States. We sought correlations between what makes a song popular on the Billboard Top-100 and Spotify metrics. To that end, 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.

We compiled songs that had appeared on the Billboard Top-100 chart with data coverage from the early 1940’s to 2020 [24]. The dataset included five different metrics for each song but it was decided that only one would be used for analysis: the total amount of weeks a song has been on the chart.

Extensive data cleaning followed to reduce the redundancy of songs reoccurring in subsequent weeks upon the Top-100 chart. For the purpose of this analysis, the *last date a song made it onto the chart was the only instance that was considered*. Any songs that never appeared among the Top-100 were dropped. The Spotify data was then cleaned to ensure that songs with a popularity value of less than five were dropped due to their disproportionate representation in the dataset (resulting in a long-tailed skew of the data). Thereafter, all songs with matching titles and performers were merged into a singular set resulting in a total of 9,829 songs for use in this analysis.

III. RESULTS

A holistic understanding of the relationships between acoustic song metrics and platform popularity is needed to understand how song properties influence platform specific engagement. To summarize our findings across all metrics and platforms, we tabulated all Spearman and Pearson correlation coefficients in Table I. As expected, the majority of correlations between the metrics considered are non-existent (near zero), however notable trends emerge and are each discussed in the following sections.

A. Spotify

Fig. 2 illustrates the individual distributions for nine of the features in the Spotify dataset. Each feature follows a different distribution and generally, we note that the majority of songs were not explicit, have low instrumentation, have low speechiness, and were typically recent songs. Within the

TABLE I
SUMMATIVE TABLE OF ALL SYSTEMATICALLY EVALUATED CORRELATIONS ACROSS CONSIDERED PLATFORMS

Platform	Platform Metric	Correlation Test	Acoustic ess	Dancea bility	Duration (ms)	Energy	Explicit	Instrumental ness	Key	Liveness	Loudness	Mode	Popularity	Speech ness	Tempo	Valence	Year
Spotify	Popularity	Spearman	-0.398	0.100	0.185	0.340	0.123	-0.286	0.001	-0.107	0.350	0.014	N/A	-0.177	0.094	0.065	0.507
		Pearson	-0.397	0.124	0.025	0.329	0.153	-0.301	0.002	-0.079	0.337	0.008	N/A	-0.195	0.095	0.063	0.513
Google Trends	Max Searches	Spearman	-0.077	-0.146	0.142	0.040	-0.147	0.033	0.008	0.030	-0.082	0.049	-0.160	-0.091	-0.038	-0.012	-0.410
		Pearson	-0.043	-0.157	0.121	0.042	-0.151	0.028	0.015	-0.003	-0.086	0.055	-0.155	-0.076	-0.027	0.012	-0.347
Who Sampled	Sampled Times	Spearman	-0.101	0.139	0.086	0.110	0.000	-0.057	-0.012	-0.017	0.111	0.034	-0.083	0.025	0.021	0.144	N/A
		Pearson	-0.091	0.091	0.092	0.108	-0.001	-0.067	-0.016	0.079	0.097	0.013	-0.052	0.043	0.024	0.110	N/A
	Covered Times	Spearman	-0.066	0.062	0.052	0.043	-0.126	-0.047	0.001	-0.015	0.049	-0.010	-0.069	-0.035	-0.018	0.104	N/A
		Pearson	-0.003	0.005	-0.006	-0.010	-0.093	-0.040	-0.039	0.043	-0.009	0.032	-0.094	-0.058	0.007	0.100	N/A
	Remixed Times	Spearman	-0.172	-0.116	0.064	0.184	-0.208	0.043	0.026	0.001	0.163	-0.065	-0.084	-0.172	-0.068	-0.038	N/A
		Pearson	-0.152	-0.054	0.021	0.159	-0.164	0.090	0.006	-0.003	0.100	-0.057	-0.089	-0.134	-0.072	-0.002	N/A
	Samples Contained	Spearman	-0.016	0.059	0.138	0.048	0.149	-0.030	-0.022	0.035	0.039	-0.023	-0.037	0.134	-0.041	0.007	N/A
		Pearson	-0.034	0.054	0.132	0.085	0.094	-0.015	-0.010	0.044	0.040	-0.015	-0.030	0.130	-0.060	0.049	N/A
TikTok	DiggCount ('Likes')	Spearman	0.099	-0.088	-0.032	-0.091	N/A	0.037	N/A	-0.019	-0.083	N/A	0.034	-0.009	-0.044	-0.073	0.022
		Pearson	0.086	-0.060	-0.064	-0.056	N/A	0.100	N/A	-0.009	-0.039	N/A	0.025	0.050	-0.058	-0.068	0.045
	ShareCount	Spearman	0.033	-0.025	0.043	-0.017	N/A	0.040	N/A	-0.054	-0.079	N/A	-0.109	-0.066	-0.026	0.039	-0.098
		Pearson	-0.024	0.015	0.028	0.021	N/A	0.087	N/A	0.006	-0.056	N/A	-0.085	-0.029	-0.020	0.039	-0.051
	PlayCount	Spearman	0.072	-0.049	-0.070	-0.068	N/A	0.005	N/A	-0.038	-0.060	N/A	0.065	0.039	-0.014	-0.052	0.072
		Pearson	0.043	0.005	-0.114	-0.072	N/A	0.026	N/A	-0.048	0.014	N/A	0.025	0.120	-0.031	-0.036	0.099
	Comment Count	Spearman	0.111	-0.124	0.031	-0.090	N/A	0.065	N/A	-0.011	-0.112	N/A	-0.054	-0.048	-0.074	-0.114	-0.054
		Pearson	0.106	-0.095	0.033	-0.073	N/A	0.086	N/A	-0.006	-0.062	N/A	-0.001	-0.047	-0.079	-0.088	-0.014
Twitter	Tweets	Spearman	0.001	-0.023	0.018	0.008	N/A	-0.013	N/A	-0.013	0.011	N/A	0.027	-0.008	-0.011	-0.027	0.023
		Pearson	-0.012	-0.007	0.013	0.007	N/A	-0.020	N/A	0.021	0.009	N/A	0.020	0.001	-0.003	-0.013	0.029
	Likes	Spearman	-0.021	-0.026	0.038	-0.006	N/A	-0.004	N/A	-0.027	-0.003	N/A	0.009	-0.008	-0.025	-0.016	-0.020
		Pearson	0.015	-0.038	-0.002	-0.034	N/A	0.030	N/A	-0.011	-0.040	N/A	-0.006	-0.005	0.001	-0.036	-0.017
	Quotes	Spearman	0.023	-0.012	0.012	-0.017	N/A	-0.021	N/A	-0.004	-0.018	N/A	-0.021	-0.001	-0.016	-0.014	-0.019
		Pearson	0.003	-0.001	0.004	-0.025	N/A	0.026	N/A	0.004	-0.025	N/A	0.011	0.016	0.040	0.019	-0.006
	Replies	Spearman	-0.029	-0.007	0.028	0.008	N/A	-0.012	N/A	-0.023	0.010	N/A	0.016	-0.012	-0.033	-0.002	0.003
		Pearson	-0.012	0.003	0.003	-0.003	N/A	-0.005	N/A	0.008	0.012	N/A	0.007	-0.008	-0.026	0.013	-0.007
	ReTweets	Spearman	-0.006	0.000	0.015	0.005	N/A	-0.039	N/A	0.004	0.025	N/A	0.027	0.009	0.003	-0.022	0.029
		Pearson	0.025	0.004	-0.023	-0.019	N/A	-0.012	N/A	-0.019	-0.005	N/A	0.007	0.010	0.030	0.000	0.030
YouTube	Views	Spearman	-0.542	0.173	0.271	0.455	N/A	-0.345	N/A	-0.115	0.440	N/A	0.822	-0.167	0.115	0.065	0.600
		Pearson	-0.134	0.094	0.013	0.117	N/A	-0.079	N/A	-0.030	0.150	N/A	0.236	-0.016	0.020	-0.005	0.197
Billboard	Weeks on Charts	Spearman	-0.175	0.141	0.179	0.112	-0.009	-0.080	0.007	-0.068	0.188	-0.039	0.329	-0.016	-0.005	-0.024	0.301
		Pearson	-0.180	0.126	0.096	0.110	-0.008	-0.072	0.007	-0.054	0.181	-0.036	0.315	-0.015	0.002	-0.028	0.292

correlation matrix (not shown; see Table I) we observed the features were not linearly correlated, with some expected exceptions including *loudness* and *energy*.

We produced linear regression models and calculated the mean absolute error (MAE) for the testing set for each regression model. The linear model showed that the *year* feature had the largest correlation coefficient and was most responsible for the popularity ($\rho = 0.513$). This was likely due to users preferring songs that were written more recently since older music eventually recedes in popularity and with certain genres eventually falling out of style. The MAE for the linear model was 0.6158 and the R^2 was computed to be 0.359. Alternate regression models – polynomial regression, Support Vector Machines, Nearest Neighbors, decision tree, and random forests – were also 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 was expected since the linear model is considered the least *expressive* of all models considered. Overall, none of the models showed strong

correlation, with the strongest correlation coming from the random forest model with an R^2 score of 0.650. This indicated that the Spotify metrics were not strong predictors of the song popularity on Spotify. This could be due to the method by which Spotify produced these metrics, as this analysis relied on their data generation methods. Overall, the models could be improved further by expanding the hyperparameter tuning. It would also be valuable to consider neural-network models to in future consideration of this work.

B. Google Trends

Fig. 3 depicts both the Spotify popularity metric and the time since a given song had the most searches in a given day. From the linear fit, we can see the Spotify popularity showed no dependence on the Google trends data since the fit had a slope of approximately zero and an R^2 of nearly zero, indicative 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 popularity are not frequently searched and are

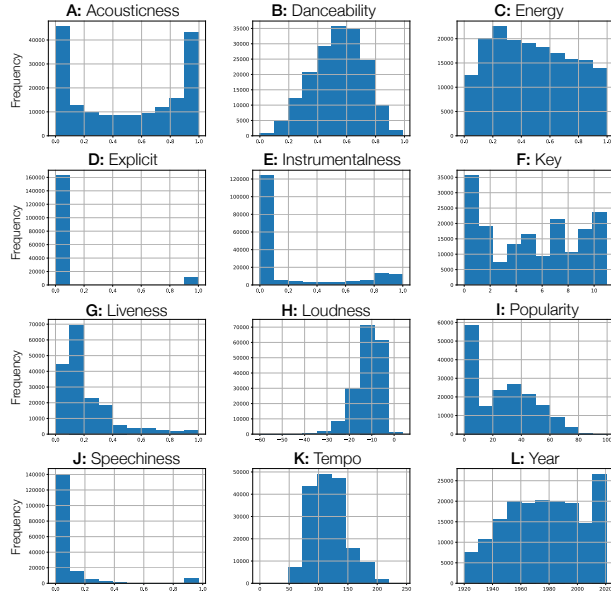


Fig. 2. **Distribution of Spotify Features as a Basis of Analysis.** Most songs are not explicit, have low instrumentation, have low speechiness, and are recent songs.

unlikely to provide Google Trends data. Another issue in this analysis was that many songs have generic names which are used for many things. Therefore, the peak search date may not represent that particular song, but may represent searches about other topics. The Google Trends peak search time was compared against the other features in the Spotify dataset and showed no correlation.

It is also to be noted that Google Trends itself is not, by design, a music hosting or streaming platform and therefore, represents a more generic (and search-centric) internet platform.

C. WhoSampled

In contrast to Google Trends, WhoSampled is a platform dedicated to the sharing, sampling, and the mixture of music. Interestingly, there was no clear correlation between a song in a WhoSampled category and a Spotify metric. We posit that this may be due to lack of sufficient data, since most songs did not fall under WhoSampled categories and there was limited data collected from WhoSampled due to the request rate. As a test for generalization, we noted there was no notable difference between average Spotify metrics for songs under WhoSampled categories and songs that were not (Table II).

D. TikTok

As a new video streaming platform, TikTok captures emergent trends in song popularity. Interestingly, no significant correlation was discovered between any of the Spotify song metrics and any of the TikTok popularity metrics. We visualized the results for one platform-specific metric in particular: the playCount. This was considered the most interesting

TABLE II
AVERAGE SPOTIFY METRICS FOR SONGS THAT HAVE BEEN SAMPLED COMPARED TO THE NON-SAMPLED

Metric	Sampled	+/-	Non-Sampled	+/-
acousticness	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
instrumentalness	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	0.014	0.45	0.005

TikTok popularity metric because it was highly correlated with diggCount, or the number of likes (Pearson coefficient = 0.86), and requires the lowest effort (*i.e.* platform engagement) from TikTok users. The scatter plots of playCount vs. Spotify metrics are depicted in Fig. 4.

E. Twitter

Similar to the Google Trends platform, Twitter is not a dedicated music hosting or streaming service; however, music popularity may nonetheless be represented based upon how users reference music on the platform. Our results are depicted in Fig. 5, again showing that 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. A limitation to this work that potentially introduces biases to this analysis is due to restrictions on the Twitter API. The API restricted the searches to tweets in the previous seven days. This restriction could be considered a threat to the validity of the analysis of the popularity

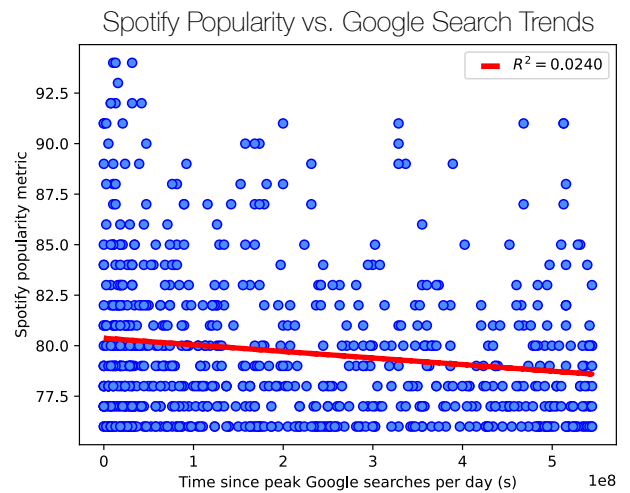


Fig. 3. **Google Trends Correlated with Spotify Popularity.** The linear regression's slope and the $R^2 = 0.024$ indicate that a correlation does not exist.

trends. A future extension of this analysis may involve using the Twitter Premium/Enterprise API. This version allows for queries further back than 7 days and allows for more requests. Additionally, expanding the window of recently released songs

beyond 5 years would also enable the quantitative measurement of how music popularity upon Twitter varies over time.

F. YouTube

Among digital music hosting and streaming platforms, YouTube is the oldest and most expansive among all platforms considered within this work. As expected, a Pearson correlation matrix (not shown; see Table I) identified which Spotify attributes correlated with popularity on YouTube, indicating that *acousticness*, *energy*, *loudness*, *danceability*, and *instrumentalness* had Pearson and Spearman correlations of significance. Scatter plots were then used to compare the YouTube view count with each of these attributes of significance.

While Fig. 6 shows Pearson correlations for each attribute that indicate a very low to nonexistent level of correlation, it also displays Spearman Correlations which indicate that certain attributes have some correlation with YouTube popularity.

More specifically, *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

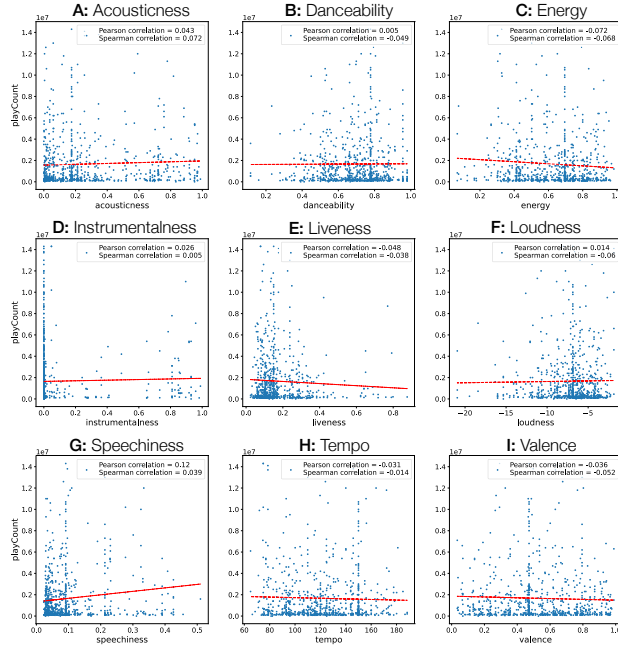


Fig. 4. **TikTok playCount vs. Spotify Song Metrics.** All sub-panels show varied distributions without a notable trend and correlation emergent among any of the considered variables.

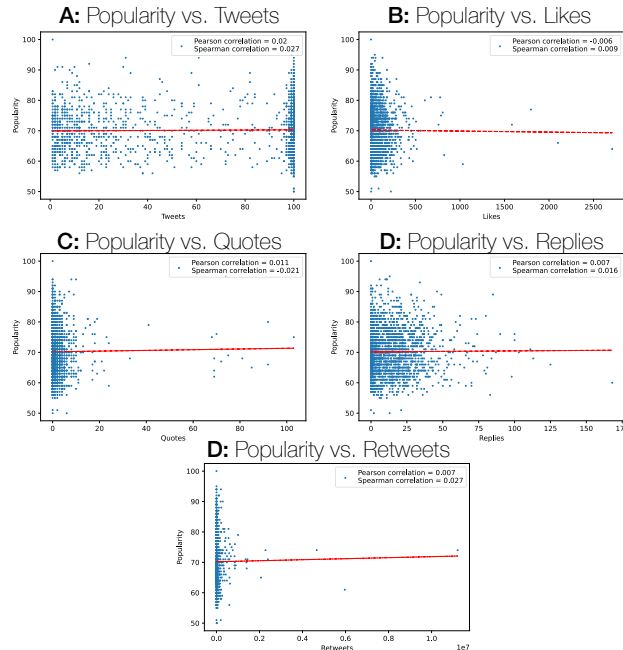


Fig. 5. **Spotify's Popularity Metric vs. Twitter Metrics.** Among the five metrics considered here, little to no relationship emerges.

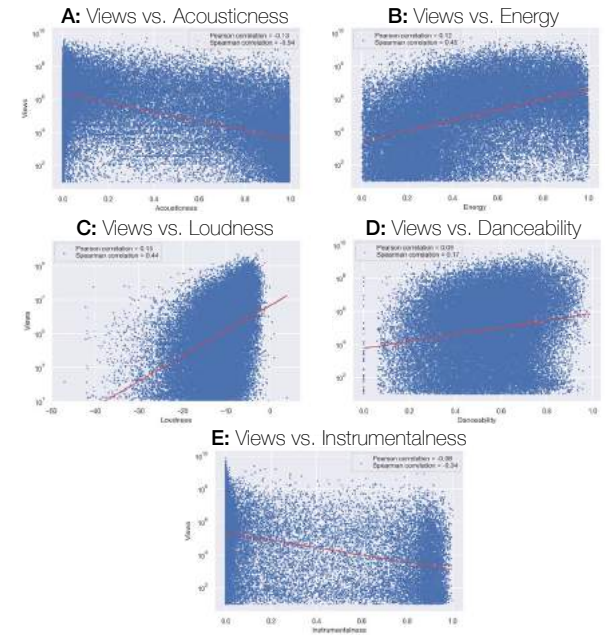


Fig. 6. **YouTube Views vs. Spotify Attributes.** Among the five factors considered, a notable linear trend emerges. The point distribution of Acousticness (A) and Energy (B) suggest that potentially a non-linear relationship exists with views.

views, and others have no views. Therefore, it is likely that the Spearman correlation coefficient better accounted for cases with extreme view counts.

Finally, we posit that as the oldest platform most dedicated to the official hosting and sharing of music (and their related music videos), YouTube is the most established platform capable of monitoring and predicting trends of music popularity. An interesting consequence of this work suggests that the most popular songs on YouTube tend to be recent, low in acoustictness and instrumentalness, and high in energy loudness.

G. Billboard

Finally, the Billboard Top-100 chart, while not a dedicated music hosting, sharing, viewing, or communication platform, aggregates physical and digital interest in music. A Pearson correlation matrix was constructed (not shown; see Table I) to identify putative correlations between Spotify metrics and Billboard Top-100 chart presence (measured in the amount of weeks a song remained on the chart). Interestingly, no substantial correlations were found. The two metrics with the highest correlations, *popularity* and *year*, were plotted in Fig. 7.

The amount of weeks a song was on the Top-100 chart had a positive correlation of approximately 0.329 with the *popularity* metric of a song and a positive correlation of 0.292 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.

IV. DISCUSSION

Several of the Internet platforms investigated had weak Spearman and Pearson correlation coefficients for every acoustic and popularity metric. Namely, no strong correlations existed for Google Trends, WhoSampled, TikTok, or Twitter. The correlation of greatest magnitude among these was a -0.410 Spearman correlation coefficient between Google Trends' maximum searches metric and song year; this may suggest that older songs are Googled more, since they are less widely known in the present day. Broadly, however, these results suggest that popularity on these platforms is not dependent on acoustic features. This conclusion has intuitive merit. For example, videos on TikTok exist in a wide range of genres, and their popularity is likely influenced by other factors, such as the creator's prior popularity, that were not examined in this study. Song samples and remixes, as measured by WhoSampled, are likely created for a range of purposes and in a variety of contexts, which may have led to low correlations. Simply put, much of the data examined may be too broad to draw noticeable correlations.

The lack of correlations obtained for the Billboard Top-100 requires more examination. Some previous research ([13], [14], [15]) experienced success correlating the same Echo Nest acoustic metrics with charts such as the Spotify Top 50 and the Billboard Top-100. Our results contradict these findings.

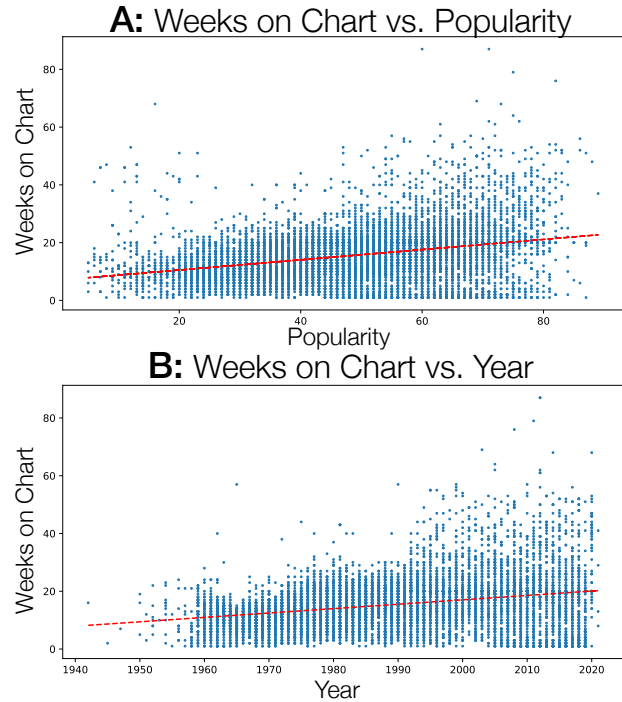


Fig. 7. **Weeks on Chart vs. Spotify Attributes.** A low-to-moderate positive relationship exists between two Spotify metrics, the popularity (A) and year of release (B).

The study in [15] differed from the current research in two ways. It sought to predict whether or not a song would appear on Spotify's Top 50 rankings, rather than a song's duration on the Billboard Top-100 chart; it also binarized most of the acoustic features used, which may explain the difference in results. Our study also differs from the research in [13] in two ways, which may account for the difference in results. Firstly, the study conducted in [13] measured success as a song simply appearing on the Top-100 Chart, and did not look at duration of time on the chart. Secondly, [13] also utilized a candidate set of unsuccessful songs in addition to successful ones, which our research did not do. As such, these two studies are fundamentally asking two different questions: [13] looks at acoustic features correlated with whether a song will appear on the Billboard Top-100, without duration of time on the chart considered. Our study, on the other hand, looks at songs that are already on the Billboard Top-100 and looks at their acoustic features as correlated with their duration on the chart. The differences between the results indicate that while acoustic features may be an effective predictor of a song appearing on the chart, they cannot effectively predict how long it will remain once it appears there.

It is interesting that Askin and Mauskopf [14] concluded some different results than the results achieved by our analysis although they used the Echo Nest metrics. However, several factors could have played a role in the contrasting results. For example, the data collected by the authors was for Billboard's

Hot 100 from 1958 to 2013, while our dataset spanned a longer time period (1940 to 2020). Also, the authors used 25,762 songs [14], while our cleaned data resulted in 9,829 songs. Whether the reason for the contrasting results was the different datasets or the used methodology would be an interesting research point in future work.

The other results of interest were those obtained for YouTube view count. Namely, this metric was found to be positively correlated to the Spotify song popularity ($\rho = 0.822$), year ($\rho = 0.600$), loudness ($\rho = 0.440$), and energy ($\rho = 0.455$) and moderately negatively correlated to acousticalness ($\rho = -0.542$) and instrumentalness ($\rho = -0.345$). Taken together, these results indicate that Youtube videos acquire more views when they use popular, current songs (as indicated by Spotify popularity and year) which are energetic (as indicated by the positive correlations with loudness and energy, and the negative correlations with acousticalness and instrumentalness). This may indicate a preference by users of this platform for contemporary, “happy” videos. Youtube’s deviation from the other platforms discussed herein, in having some notable correlations, may indicate that it is more affected by current musical trends.

V. CONCLUSION

In this research, we conducted a novel systematic study of the correlation between the descriptive features of songs, as collected from the Spotify Echo Nest dataset, and the popularity metrics of the songs on various platforms. The studied platforms were Google Trends, WhoSampled, TikTok, Twitter, YouTube, and Billboard Top100.

Across all platforms, very few notable correlations were found between the Spotify metrics considered and the platform popularity metrics derived in this work. This reinforced the existing idea that music popularity can not be attributed solely to quantifiable acoustic elements that can be predicted accurately or optimized within a song. One notable exception was YouTube, for which some moderate correlations existed; this indicated that this platform is more influenced by current musical trends than others. Further investigation beyond this work could be done to define alternative metrics and identify other correlations. Furthermore, expanding the data scraped in this work promises increased resolution into platform-specific trends. For example, circumventing many of the digital scraping limiters set upon each platform would enable a more expansive query of songs and time-varying factors. To this end, we released all our code and data for use by the broader research community to expand this work.

REFERENCES

- [1] International Federation of the Phonographic Industry, “Global music report,” International Federation of the Phonographic Industry, www.ifpi.org, Tech. Rep., March 2019. [Online]. Available: <https://www.ifpi.org/resources/>
- [2] “IFPI Global Music Report 2021,” 2021, accessed: 2021-06-02. [Online]. Available: <https://www.ifpi.org/ifpi-issues-annual-global-music-report-2021/>
- [3] S. Asai, “Factors Affecting Hits in Japanese Popular Music,” *Journal of Media Economics*, vol. 21, pp. 97–113, 2008.
- [4] V. Dhar and E. A. Chang, “Does Chatter Matter? The Impact of User-Generated content on Music Sales,” *Journal of Interactive Marketing*, vol. 23, pp. 300–307, 2009.
- [5] C. V. S. Araujo, M. A. P. Cristo, and R. Giusti, “Predicting Music Popularity Using Music Charts,” *Proceedings of the 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 859–864, 2019.
- [6] M. Pichl, E. Zangerle, and G. Specht, “Combining Spotify and Twitter Data for Generating a Recent and Public Dataset for Music Recommendation,” *Proceedings of the 26th GIWorkshop on Foundations of Databases (Grundlagen von Datenbanken)*, 2014.
- [7] D. E. Berlyne, *Aesthetics and Psychobiology*. New York, New York: Appleton-Century-Crofts, 1971.
- [8] C. Martindale, K. Moore, and J. Borkum, “Aesthetic Preference: Anomalous Findings for Berlyne’s Psychobiological Theory,” *The American Journal of Psychology*, vol. 103, no. 1, pp. 53–80, 1990.
- [9] A. C. North, A. E. Krause, L. P. Sheridan, and D. Ritchie, “Energy, Typicality, and Music Sales: A Computerized Analysis of 143,353 Pieces,” *Empirical Studies of the Arts*, vol. 35, no. 2, pp. 214–229, 2017.
- [10] A. C. North, A. E. Krause, L. P. Sheridan, and D. Ritchie, “Popularity, Mood, Energy, and Typicality in Music: A Computerized Analysis of 204,506 Pieces,” *Psychology of Aesthetics, Creativity, and the Arts*, vol. 13, no. 1, pp. 89–109, 2019.
- [11] F. Pachet and P. Roy, “Hit Song Science is not yet a Science,” *Proceedings of the 9th International Conference on Music Information Retrieval*, pp. 355–360, 2008.
- [12] J. Lee and J.-S. Lee, “Music Popularity: Metrics, Characteristics, and Audio-based Prediction,” *IEEE Transactions on Multimedia*, 2018.
- [13] M. Interiano, K. Kazemi, L. Wang, J. Yang, Z. Yu, and N. L. Komarova, “Musical trends and predictability of success in contemporary songs in and out of the top charts,” *Royal Society Open Science*, vol. 5, no. 5, 2016.
- [14] N. Askin and M. Mauskopf, “Cultural Attributes and Their Influence on Consumption Patterns in Popular Music,” *Proceedings of the 6th International Conference, SocInfo 2014*, pp. 508–530, 2014.
- [15] C. V. S. Araujo, M. A. P. Cristo, and R. Giusti, “A Model for Predicting Music Popularity on Streaming Platforms,” *Journal of Theoretical and Applied Informatics*, vol. 27, no. 4, 2020.
- [16] 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>
- [17] A. Luscombe, K. Dick, and K. Walby, “Algorithmic thinking in the public interest: navigating technical, legal, and ethical hurdles to web scraping in the social sciences,” *Quality & Quantity*, pp. 1–22, 2021.
- [18] L. Colley, A. Dybka, A. Gauthier, J. Laboissonniere, A. Mougeot, N. Mowla, K. Dick, H. Khalil, and G. Wainer, “Spotify-Platform-Correlation,” 5 2022. [Online]. Available: <https://github.com/adybka/Spotify-Platform-Correlation>
- [19] “Pytrends,” accessed: 2021-04-03. [Online]. Available: <https://pypi.org/project/pytrends/>
- [20] “Whosampled,” accessed: 2021-04-04. [Online]. Available: <https://www.whosampled.com/>
- [21] “tiktok-scraper,” accessed: 2021-04-03. [Online]. Available: <https://github.com/drawrowfly/tiktok-scraper>
- [22] “Twitter API,” accessed: 2021-04-04. [Online]. Available: <https://developer.twitter.com/en/docs/twitter-api>
- [23] “Youtube API,” accessed: 2021-04-01. [Online]. Available: <https://developers.google.com/youtube/v3>
- [24] S. Miller, “Billboard hot weekly charts,” accessed: 2021-04-01. [Online]. Available: <https://data.world/kcmillersean/billboard-hot-100-1958-2017>