LUISS

Department of Business and Management
Chair of Business and Marketing Analytics

Prediction of Song Popularity with a Textual and Audio Based Approach

Prof. Francisco Villarroel Ordenes
Supervisor

Alessio Barboni
Candidate

Academic Year 2020/2021

For my family

1. Introduction

In the digital age, thanks to the availability and ease of access granted by digital mediums, books, movies and songs are increasingly being consumed. The year 2020 has been one of the best for companies operating in those industries (at least for the ones operating online, e.g. through streaming platforms). This is mainly due to the Covid-19 pandemic that forced many people to stay at home for months and limited the possibility of participating in social events in most countries around the world. As a result, sales of e-books and audiobooks had double-digit growth in the 12 months to the end of September 2020 (The Economist 2020); Netflix has added 10.1M new subscribers globally in the second quarter of 2020, a number even larger than its forecast of 8.3M (The Economist 2021); Spotify enjoyed record subscriber growth, adding 3M subscribers just in the first three months of the year 2021 (Financial Times 2021).

This research will focus on the music industry, whose total revenue (made of live, recorded and publishing music) was worth $62 billion in 2017, and it is forecasted to reach $131 billion by 2030 (Goldman Sachs 2017). The growing popularity of streaming platforms like Spotify and Apple Music is one of the main reasonable explanations (Goldman Sachs 2017). The pandemic is impacting live performances the most, lowering short-term forecasted profits for the live-music industry (the concert industry lost more than $30B according to Forbes 2020, and musicians themselves may have lost about two-thirds of their income in 2020 according to the BBC), but keeping the long-term growth outlook of the whole music industry still intact (Goldman Sachs 2020).

In this large market, some songs become popular while others do not. Popularity can be measured in various ways (e.g., the total number of streams, sales or live radio plays per week), and it is often summarized in music charts, in which tracks are listed by rank. Being able to predict popularity would benefit the majority of stakeholders (i.e., artists, streaming platforms, labelling companies, users, etc). E.g., a song predicted to become popular could be recommended to users liking the genre (good recommendations should be correlated with user's overall satisfaction of the platform), the artists producing it should incur fewer risks of not succeeding, the labelling companies could focus advertising efforts on it (discarding the non-promising ones based on expected successfulness), the streaming platform companies overall could provide a better service thanks to this *data-driven-decision* systems (e.g., it could be employed in the playlist creation and updates process).

These platforms have been reshaping the music business for a long time. The fundamental difference is in artists' compensation, which before was tied to sales, and now depends on the number of plays. Hence, pressure on artists is higher than ever, given that revenues per stream have always been low for songwriters and thus millions of plays are needed to make decent money. Additionally, a third of all streams tends to depend on the inclusion in the company's playlist, which is an automated process handled by algorithms as mentioned above. Hence, composers are adapting to what they think is being looked for, or better they have to, if they want to pursue this career. As a result, hit songs became shorter, intros truncated, choruses are starting sooner, etc (The Economist 2019). It is fundamental to understand what makes a product or artist valuable, with the ultimate goal of finding talent, which is the most important thing for a record company (s.c. A&R Process). This knowledge could also be used, for example, to decide the best moment to release a new album, or to promote marketing strategies, concentrating efforts on pieces not yet expected to become popular or switching the focus from unpromising cases into more promising ones (Araujo et al., 2019). Moreover, it could be useful even when writing new songs from scratch, to match the current successful trends.

This paper proposes a blended approach combining some textual based approaches presented in Berger and Packard 2018 and 2020, with an audio based one suggested in Lee and Lee

2018. The textual ones argue that similarity between cultural items (e.g., books, movies and songs) may help to shape success, and that cultural items activating personal connections with other people tend to be more successful respectively; While the audio based one claims that song popularity can be predicted with a combination of audio features (e.g., Complexity features, Arousal, MFCCs). An overview of past approaches in the area of data mining is presented in Section 2, reviewing studies in the area of text and audio mining respectively. Research Questions follow in Section 3. The data collection, feature engineering and modelling are covered in Section 4. Section 5 follows with models' results which are further discussed in the next and final section, i.e. Section 6.

2. Literature Review

Table 1: Literature Review

Year	Authors	Title	Scope	Results
2007	N. Camelin, F. Béchet, G. Damnati, R. De Mori	Speech Mining in Noisy Audio Message Corpus	Audio mining on phone calls for removing unreliable utterances from the audio corpora	Statistical significance
2009	V. Dhar, E.A. Chang	Does chatter matter? The impact of user-generated content on music sales	Text mining on blogs UGC for predicting song popularity	Future sales are predictable using blog post volume and percentage changes in Myspace friends
2016	B. Shulman, A. Sharma, D. Cosley	Predictability of popularity: Gaps between prediction and understanding	Text mining on Last.fm UGC for predicting song popularity	81% Accuracy on test set
2018	J. Berger, G. Packard	Are Atypical Things More Popular?	Text mining on song lyrics for predicting song popularity	More differentiated lyrics from the genre tend to be more popular
2018	J. Lee, J.S. Lee	Music Popularity: Metrics, Characteristics, and Audio-Based Prediction	Audio Mining on songs for predicting song popularity	Features like Chroma, Arousal and MFCCs showed statistical significance
2019	J. Berger, A. Humphreys, S. Ludwig, W.W. Moe, O. Netzer, D.A. Schweidel	Uniting the Tribes: Using Text for Marketing Insight	An overview of text mining, with a focus on marketing insights generation	.
2019	C.V.S. Araujo, M.A.P. Cristo, R. Giusti	Predicting music popularity using music charts	Audio Mining on songs for predicting song popularity	AUC of 80% on test set
2020	J. Berger, G. Packard	Thinking of You: How Second-Person Pronouns Shape Cultural Success	Text mining on song lyrics for predicting song popularity	Second person pronouns enhanced song
2021	X.S. Wang, S. Lu, X.I. Li, M. Khamitov, N. Bendle	Audio Mining: The Role of Vocal Tone in Persuasion	Audio mining on Kickstarter's video pitches for measuring the impact of vocal tones on funding success	Audience decision can be influence by a focused, low-stress and emotionally stable vocal tone

2.1 Overview of Text Mining Applications

Either when written down or pronounced while speaking, words are the most straightforward way to convey any sort of message. In this era, thanks to the digitization of information, words have undoubtedly become more valuable for businesses, as they are part of almost every marketplace interaction (Berger and Packard 2018). Online activities generate vast amounts of data that can be further processed to extract information, and part of this data consists of textual content ("80%–95% of all business data is unstructured, and most of that unstructured data is text"; Gandomi and Haider 2015). These activities could range from users writing an online review or just surfing the web, to firms communicating to the customers through advertising. The data generated could benefit all stakeholders, as in the case of a platform ecosystem (e.g., TripAdvisor), or a subgroup of them (e.g., the company producing the browser and the website owner) as in the case of web browsing. Regardless of this, words can provide insights in many ways, as they can generate different reactions in the reader. For example, they could induce the user to perform an action or trigger certain thoughts or emotions, that even if not explicitly stated would be reflected by the user's online activities (e.g., the cultural items she watches, reads, or listens to). Hence user activity tends to have an impact on items' popularity, making the prediction of their success worth attempting. To pursue this goal of extracting meaningful insights *Text Mining* comes to the aid. It is defined exactly as "the process of transforming unstructured text into a structured format to identify meaningful patterns and new insights, by applying advanced analytical techniques" (IBM 2020).

One important usage of this technology is to gain insights on the creator of the text, either individuals or corporations, given that it "reflects and indicates something about the text producer" and the context she is in (Berger and Packard 2018). Many attributes of the personality can be inferred by a simple post on social media, in addition to their current emotional state and thoughts. In some sense, it could be seen as a "fingerprint or signature of the text producer" (Pennebaker 2011). Moreover, the relationships within people or the attitude towards certain items can be forecasted, and even the future possible actions that the individual might undertake (e.g., the probability of defaulting on a loan can be inferred by analyzing the language used in the application; Herzenstein et al., 2019). This can be applied in the same manner for corporations or political organizations to understand the brand personality in the first case and the leadership style in the second. Furthermore, institutions or larger social groups can be analyzed by aggregating the textual contents of their components. In this way, different cultures can be studied, since texts reflect information about the contexts in which they were produced, along with their evolution through time (e.g., the way minorities are perceived in two different time periods).

In addition, whenever an audience consumes a text, there will be an impact on that audience, as people react to the above-cited inferred characteristics of the text creator. It could influence the brand reputation in both directions, affect future purchases or even the fact that people will talk about the brand. For example, it has been shown that the language of newspapers tends to change customers' attitudes (Humphreys and LaTour 2013), the way in which criticisms are handled by firms can avoid social media firestorms (Grewal et al. 2019), and even the lyrics of a song can shape its market success (Berger and Packard 2018; Berger and Packard 2020). Thus, text mining potentially provides "insights that may not be cost-effectively obtainable through other methods" (Berger and Packard 2018).

2.1.1 Text based Approaches for Song Popularity Prediction

Keeping aside the part of research arguing that popularity is random, some approaches belonging to the field of psychological science, and based on the psychological foundations of culture, argue that success can be forecasted to some extent, even with today's technology.

The first, by Berger and Packard 2018, argues that success can be shaped by analyzing the similarity between cultural items, for example songs. Since the novelty of a song is subjective and depends on people's experiences, one hypothesis is that the more differentiated and atypical a song is, the more it would be liked and likely become popular (i.e., liked and purchased more). In the case of songs, thus the differentiation is being bounded by genres, hence the analysis aims to verify whether, among songs belonging to the same genre, the most differentiated ones in terms of lyrics are also the most successful. Those songs are "different from the prototype but not so different as to be outside the genre" (Berger and Packard 2018).

The second approach, also by Berger and Packard 2020, claims that if songs are meant to promote "feelings of social connection", the ones that succeed in doing that would also be the most popular. Specifically, that aim is studied analyzing the usage of second person pronouns (i.e., you, your, yours, yourself, yourselves), that function as a “signal attention focus” for the listeners. Unlike first and third person pronouns, they signal that the speaker is “directly addressing cognitively or physically present people or their things”, thus stimulating the involvement of those people, conveying norms and imperatives. Or alternatively, they underline people’s relationships, which may also happen between the author and other humans different from the audience. The authors argue that in those cases, rather than encouraging the audience to see the singer’s own perspective, the audience is encouraged to outline someone present in their own lives and address them impersonating the songwriter's perspective (as shown by Green & Brock, 2000; Hartung, Burke, Hagoort, & Willems, 2016). For example, using an example from Berger and Packard 2020: “Rather than thinking Queen is going to rock them, listeners imagine another person or persons they want to ‘rock’ (e.g., an opposing sports team”). This provides people with a new way of seeing their lives, through the author’s lenses. Both approaches have returned statistically significant results on their datasets, underlying the possibility of predicting popularity with the above-cited variables.

Dhar and Chang 2009 instead employed a strategy based on analyzing user generated content (UGC), specifically online comments taken from social media and blogs. Days since release, the type of label (whether major or independent), the average number of reviews and average rating, blog post volume and the weekly change in blog post chatter, and finally Myspace data about friends were also taken into account. Findings suggest that based on their data future sales can be predicted by blog post volume and percentage changes in Myspace friends. Shulman et al. 2016 also took a similar approach on the Last.fm platform. The goal of the model here was to predict if a song would get higher than average interaction. The logistic regression achieved about 81% accuracy.

2.2 Overview of Audio Mining Applications

Audio data is becoming increasingly available along with text and visual data. Apart from online music streaming platforms, that have existed for many years now, new digital products solely based on audio content are starting to draw a lot of attention (e.g., podcasts, or social media like ClubHouse). To analyse these large chunks of data, Audio Mining comes to the aid. It has many applications, for example in the field of music information retrieval (MIR), automatic speech recognition (ASR), keyword spotting (KWS).

In the ASR framework, tasks can become very challenging when the quality of audio tracks is poor, and unfortunately it is very often the case when the data collected comes from a call centre or corporate surveys conducted by telephone. The tracks often contain surrounding noises, but the problem could rely on the speaker herself, as she could speak unclearly (e.g., mumbling in a low voice tone, or using peculiar accents), or even just showing many hesitations and corrections in the end. Bechet et al. 2007 proposed an audio mining approach for dealing with telephone surveys, specifically in the analysis of answers to a recorded message asking if they were satisfied with the service they had received. Their scope was to remove from the corpora of audio the "unreliable utterances" that may be considered just noise. They proposed a method based on the Kullback-Leibler divergence.

Bendle et al. 2021 relied on this technology for analyzing Kickstarter's video pitches, specifically the impact of vocal tones on the persuasion of the audience (measured in terms of funding outcomes). The authors argue that the audience's funding decision can be affected by the *persuader*'s vocal tone, as the latter can influence the persuader's perceived competence and thus the probability that the proposed project will be delivered. A successful vocal tone should be denoted by focus or task engagement, low stress or confidence, and stable emotions (extreme emotional levels could denote a lack of realism). Video characteristics and other control variables were added to the model, and they proved the statistical significance of the hypothesis (with a dataset restricted to musical projects).

2.2.1 Audio based Approaches for Song Popularity Prediction

Early approaches in this field have used song's features like pitch, timbre and loudness, with little success, while others returned better results relying on musical complexity, or Mel-frequency cepstral coefficients (*MFCCs*). Araujo et al. 2019 followed a blended approach combining metadata features (e.g., artist and song names, rank, duration, date) with acoustic ones (i.e., *MFCCs*, spectral centroid, spectral flatness, zero crossings, and tempo). The classification was performed with an *SVM* classifier and reached an *AUC* of more than 80%. Pleus and Rossi tried an approach based on Deep Convolutional Neural Networks (*CNNs*) performing a multiclass classification over three classes (low, mid or high popularity), feeding it with the extracted spectrograms images of each song. They managed to obtain about 61% test accuracy.

The approach present in Lee and Lee 2018 instead is based on a mixture of complexity features, *MFCC* features and *MPEG* features. Complexity features are calculated by measuring the temporal changes of the components (i.e., Harmony or Chroma, Timbre and Rhythm), through a structural change algorithm presented in Mauch and Levi 2011 The *MFCC* features are used for calculating the spectral characteristics of audio signals, while the *MPEG* is an audio analysis tool providing valuable insights on various spectral and temporal characteristics. Results have shown statistical significance on their dataset for some predictors even when used in isolation, and definitely better when used in aggregation.

3. Research Questions

The task of predicting music popularity is far from being new, as one could start finding academic papers on the topic already for the year 2005 (e.g., “Automatic Prediction of Hit Songs” from Dhanaraj and Logan), and potentially even before. Undoubtedly the state of the art on this task can be improved by the upcoming research and will be, but innovating in the field has started to require strong foundations in either modelling or theorizing, and this has been reflected by the rising number of related articles in recent years. Better regression or classification models would certainly help, and the rising availability of data coming from users even more. Theories as well benefit from the mere existence of such data, as most of their propositions were not easily verifiable in the past, since it would have required empirical testing (Berger and Packard 2018), that basically implies to ask people directly. Hence the infeasibility to scale such an approach underlines the fact that it has long remained “unclear whether these aspects truly drive behavior” (Berger and Packard 2020). As of today, many theories have proven to be statistically significant, once the technology and the data availability made the testing of huge amounts of data possible.

The goal of this paper is to combine some methods chosen from the state of the art research in this field, aiming to reach optimal results in the final prediction task, and trying to potentially tackle eventual research gaps and limitations. The selected academic papers (i.e., Berger and Packard 2020; Berger and Packard 2018; Lee and Lee 2018) were published in the last 3 years and thus constitute a good approximation for the most recent research in the field. Other recent approaches, mostly social media based, were not chosen for the cumbersomeness of data collection, which tends to limit both the repeatability and generalizability of the studies. The required data for this research instead are obtainable with little web scraping skills.

The chosen research questions are the following:

1) Are the propositions present in Berger and Packard 2020 and Berger and Packard 2018 verified for statistical significance, as well as the ones in Lee and Lee 2018?
2) How could best performances be achieved on the song popularity prediction task? Which variables should the best model include?

.

4. Method

4.1 Data Description

The data collected was the one used in Packard and Berger's studies (Berger and Packard 2018; Berger and Packard 2020). It was scraped from *Billboard's digital download rankings*, taking one week every three months for a three years period (i.e., 2014-2016), for seven major genres appearing on the website (i.e., christian, country, dance, rock, pop, rap, r&b). In total, given that each chart contains exactly 50 ranked songs, there are 4200 songs originally ranked from 1 to 50 and subsequently reverse coded for letting positive coefficients describe a positive relationship with "audience engagement" (Berger and Packard 2018). Digital downloads charts were chosen because they are "more likely to be driven by consumer preferences rather than by institutional actors, e.g., radio DJs, professional critics, or awards" (Berger and Packard 2020).

Hence, the columns obtained from scraping by now are the following: song name, artist name, genre, date and rank. For the textual analysis, the song lyrics are needed, and they were added using the *Genius* API.

For the audio analysis, all the songs were downloaded in *mp3* format from *Youtube*.

The regression models for answering the second research question will use the data already made available by previous feature engineering steps.

4.2 Measurement Development

Most of the variables needed by the models had to be feature engineered. This is true especially for the audio data, since the audio tracks themselves have little use when fed as a whole to a regression or classification model.

For the Berger and Packard 2020 and Berger and Packard 2018 studies, additional word features were extracted from the lyrics using the *Linguistic Inquiry and Word Count (LIWC)* software (namely: cognitive, affect, social, perceptual, motivation, temporal, swear and relativity word metrics); Latent Dirichlet Allocation (*LDA*; Blei 2012) using Gibbs sampling at 5,000 iterations was performed on lyrics to define 10 topics (the choice of K=10 is further explained in the supplemental online materials of Berger and Packard 2018), and word distribution per topic; Before applying this step, lyrics were first preprocessed with some basic techniques (*Case Normalization*, *Stop words* removal, *Lemmatization*). From the LDA analysis, the song topic composition and average topic composition per genre were computed for calculating the Linguistic Style Matching (*LSM*). The LSM equation was taken from Pennebaker and Ireland 2010, and it was readapted for the topic composition and for measuring differentiation rather than matching. The final version for a given song is the following:

$$(1) \qquad \frac{\sum_{x=1}^{K} \frac{|songTopicComp_x - avgTopicComp_x|}{songTopicComp_x + avgTopicComp_x + 0.0001}}{K}$$

(Where *K* is the number of topics; *songTopicComp*$_i$ is the proportion of the i-th topic in a song; *avgTopicComp*$_i$ is the average proportion of the i-th topic in the song’s genre).

The Second Person Pronouns were also extracted using LIWC (i.e., variable “You” in the models below).

For the Lee and Lee 2018 study, there were no features besides the track itself, as written above. Hence first the Complexity Features were extracted, specifically Chroma and Timbre. The former represents “the instantaneous harmony at a particular moment, which is one of the 12 chords (*C, C#*, etc)” (Lee and Lee 2020), while the latter is defined as “the quality given to a sound by its overtones”. The first component was calculated using the *librosa* library in Python, while the second was calculated averaging the *MFCC*s of the 256 frames resulting from splitting the windowed audio segments, obtained moving a 2.97 seconds *Hamming window* one second at a time over the tracks (as presented in Lee and Lee 2018).

Finally the *Structural Change* of the complexity features was calculated as described in Mauch and Levy 2011, using the efficient implementation presented in section 2.2. Then MFCC and Arousal features were added as described in Lee and Lee 2018., the former measures the spectral characteristics of songs, while the latter, based on psychoacoustic theory, is linked to the power of songs to generate emotions.

4.3 Modelling

First, the statistical significance of feature engineered variables was checked using an Ordinary Least Squares linear regression. For the Berger and Packard 2020 study, the first 7 models presented were built, starting from Model 1, which is a simple linear regression of Second Person Pronouns onto Song's Rank, and testing different combinations of Second Person Pronouns and other predictors in the following models. Model 2 adds the number of times that song appeared on charts (i.e., "Times Charted"), the number of genres the song belongs to (i.e., "Count_genres"), and whether it appeared on the radio airplay chart of Billboard in that period (i.e., "Radio Airplay"). Model 3 instead includes some random effects for the "artist" variable. Model 4 considers the different genres, Model 5 the time variable, Model 6 the 10 different topics extracted from LDA and Model 7 considers the LIWC metrics extracted from the lyrics ("cognitive words", "affect words", etc). Finally, Model 8 incorporates all the previous 7 models; For the Berger and Packard 2018, the first 3 models were rebuilt, in which Model 1 measures the link between atypicality ("Lyrical Differentiation", extracted with the LSM modified equation) and Song's Rank, Model 2 adds once again "Times Charted", "Radio Airplay" and "Count Genres" variables (with the extracted topics and the "Artist_Song" as control variables), and Model 3 adds LIWC metrics and a control for time to the previous model. Further details can be found in Tables 2, 3, and 4 in the Appendix section; For the Lee and Lee 2018 study, 5 models were built, of which the first four consider the different predictors in isolation (i.e., Chroma, Timbre, Arousal and MFCCs), and the last one encompasses all the predictors together.

For the second research question, three Random Forests and three Support Vector Machines (*SVM*) models were built for predicting song's rank, tuning the hyperparameters using Grid Search, and selecting the best one in terms of performances using cross-validation. Model 1 used textual variables only (taking all the covariates of the previously built models into account), Model 2 the extracted audio variables, and Model 3 a combination of the previous two respectively. Further details are in the appendix section.

5. Results

After running the models of Berger and Packard 2020, statistical significance was confirmed for the variable "You" representing the proportion of Second Person Pronouns, having *p-values* always below the 0.05 threshold in all the first seven models (cf. Table 3 and Table 4 in the Appendix section). Model 8, the one containing all the previous variable combinations together, returned a significant *p-value* of 0.024 (coef.=0.0014, SD=0.0015).

For the Berger and Packard 2018 study, *p-values* were confirmed to be all significant for "Lyrical Differentiation", this time below the 0.01 threshold, even for Model 3 that encompasses all the predictors and control variables together (cf. Table 2 in the Appendix section for further details); For the Lee and Lee 2018 study, the engineered features were first tested alone, showing statistical significance in just some of the variables (i.e., Chroma 2, Timbre 5, Timbre 6, Arousal SD, and MFCC 4, 8 and 9; Further details in Tables 5, 6, 7 and 8 in the Appendix section), then all together maintaining the same outcome in terms of significance.

Regarding the final prediction task, the Random Forest considering textual features only achieved an Mean Squared Error (*MSE*) of 0.817, and Mean Absolute Error (*MAE*) of 0.745; Considering Audio features only the *MSE* and *MAE* drop to 0.938 and 0.823 respectively; While the model encompassing all the variables achieved an *MSE* of 0.812 and an *MAE* of 0.739. These metrics are so small because the data was scaled to obtain a distribution with a mean value of 0 and a standard deviation of 1. This passage is not needed by the random forest models themselves, as ensemble methods in general are not sensitive to the variance in data, but it was performed to allow the comparison with Support Vector Machines (*SVM*), since they instead require scaled data to perform best. The *SVM* regressor achieved an *MSE* and *MAE* of 0.931 and 0.799 respectively with textual features only, again rising to 1.016 and

0.856 when taking only audio features, and finally obtaining an *MSE* of 0.931 and *MAE* of 0.798 with all the variables encompassed.

Some explainability-oriented plots, showing first of all the feature importance for the Random forest model (not including the artist/song control variable), were obtained using Shapley Additive Explanations (*SHAP*). More on this in the next section.

Figure 1.

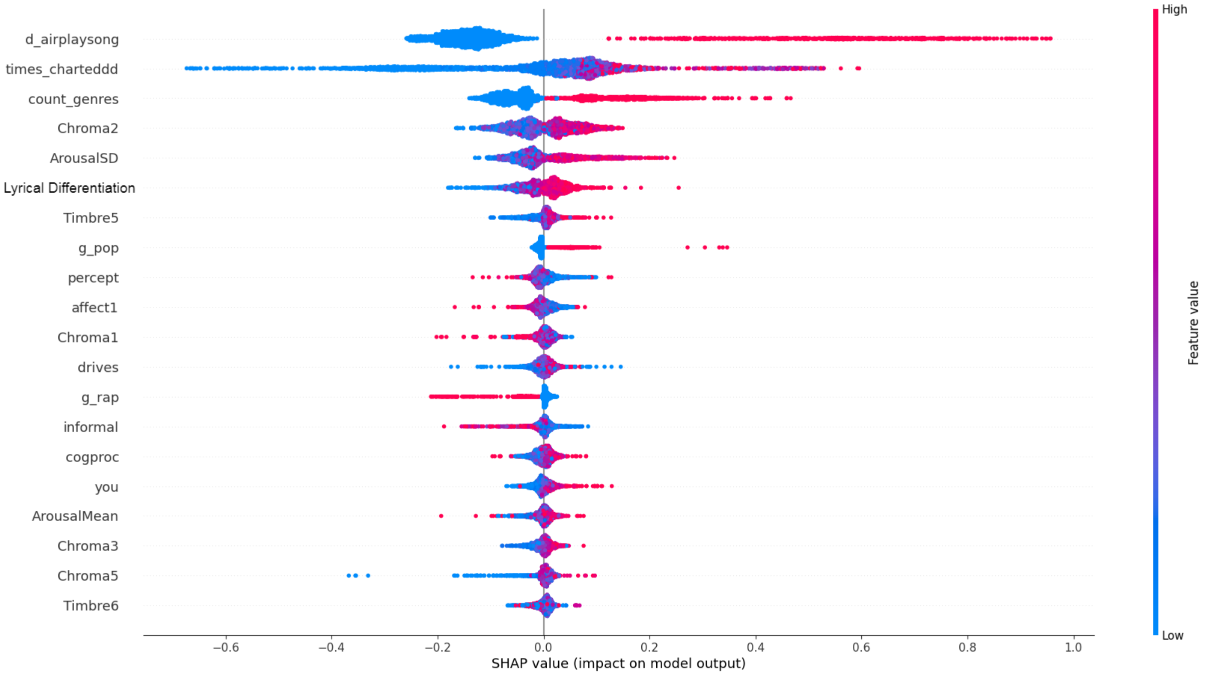


Figure 2.

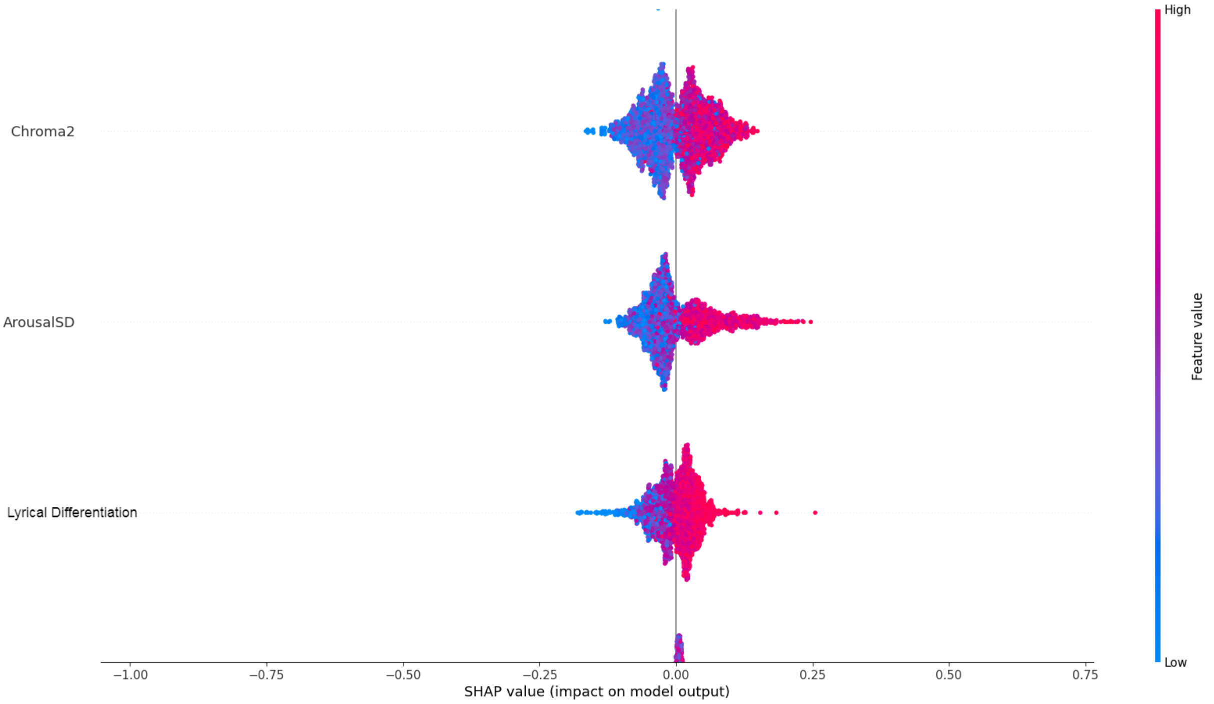
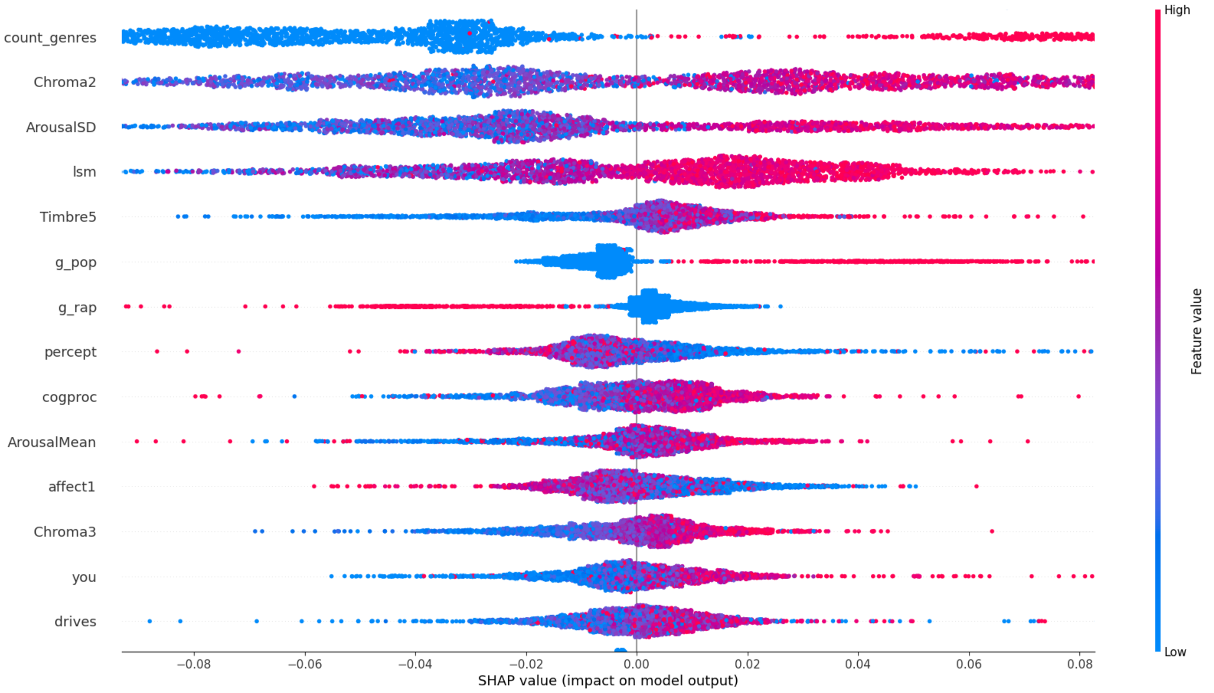


Figure 3.



6. Discussion

The results shown above confirmed the hypothesis of the authors.

First, the usage of more second person pronouns (2PPs) is linked with song popularity, as illustrated by Model 1, and this solely depends on consumer preference (denoted by variable "You", given that it was extracted from a dataset reflecting consumer choice), since it remained true in all the other settings tested. Model 2, for example, has proven that the result does not change when taking radio airplay into account, since the *p-value* of "You" remained significant. This discarded the possibility that the linkage between *2PP*s and popularity is instead caused by the fact that radio stations played, by chance, many songs containing high proportions of *2PP*s in that period, and that the radio is the real driver of popularity. If it was the case, it would have negatively affected the significance of the *p-value* of "You". The significance of the other *p-values* is not important in this sense (i.e., with the aim of proving that the real popularity driver is consumer preference), as they just measure the relationship between their respective variables and success. The same holds true for the other variables of the model, namely whether the song has appeared on charts before and if it belongs to more than one genre, as the above-cited linkage still persists afterwards. Model 3 confirmed that the linkage between *2PP*s and success is not caused by the fact that famous artists happened to put many *2PP*s by chance in their songs, as the *p-value* of "You" remains significant even when accounting for them. Models 4 and 5 showed that neither the genre nor the date seem to reject the original proposition. Model 6 proves that the topics do not question the idea that consumer preference is driving popularity either, as again the *p-value* of "You" remains significant, discarding the possibility that topics are instead the driver of popularity. Model 7 proves the same for a list of *LIWC* features. Finally, Model 8 confirms that, even when taking every variable into account, the original proposition on the popularity driver does not change, as the *p-value* on "You" remains once and for all significant.

Second, the differentiation of a song from its genre is linked with its popularity, as Model 1 has shown, since the significance of its *p-value* held through all the three models. This remained through also in Models 2 and 3, when adding other variables as covariates (More details in Table 3 of the Appendix section).

The audio analysis made clear that some complexity features are effective for predicting success, even when taken in isolation. The significance of *Chroma-2*'s *p-value*, for example, denotes that listeners seem to care about the structural change of the song's flow (intended as the song's melody, mood and chord). The same argument applies to the structural change of the other complexity features (i.e., Timbre 5 and 6). Data on the spectral characteristics of the tracks, represented by the extracted *MFCCs* features, has shown some significance, highlighting the fact that somehow such characteristics could help shape success. The same holds true for part of the arousal features, showing a linkage between success and the song's capability to evoke emotions and allowing to conclude that emotions impact music preference.

The actual song popularity prediction has shown some improvements with the combination of the two approaches, as both the *MSE* and *MAE* dropped using all the predictors. The *Random Forest Regressor* performed best in every model tried, when compared to the *SVM Regressor*, hence the final prediction, after retraining the model on the whole dataset and not just on the train-test split, will be computed using it. Figures 1 and 2, besides listing feature importance (here as a ranking in terms of mean absolute *shap values*), aim to provide a global view of the model with the aggregation of many *shap value* instances. The colour red/blue represents the value of the feature in a determined observation. E.g., the red points of the radio airplay feature (called "d_airplaysong" in the figures), indicate high values of radio airplay (namely one, since it is a categorical variable with two categories, that are one and zero), and the position over the x-axis denotes the *shap value* of the instance, that basically measures the impact on the model result of that particular observation (with high values suggesting that the instance pushed the model value up towards the response variable, i.e., song's rank). The points for radio airplay are perfectly separated, proving that the red points (corresponding to radio airplay equal to one) caused the model to predict a higher song rank, with a different but always positive impact, while the blue ones (corresponding to radio airplay equal zero) the opposite. This can be said, although the points are not perfectly split on the *shap value* equal zero, also for the other variables such as "lyrical differentiation", since as shown in Figure 2, the most differentiated songs having the most intense red colour lie mostly on the positive side of the x-axis, proving that lyrical differentiation is linked to more popularity for most of the instances of the dataset.

Further research could apply more advanced technologies, as for example Deep Neural Networks to get even more precise results. Better text and audio based approaches can be used as research proceeds and also totally different approaches could be tried, for example a social media based one, that will become more valuable every year as the mole of the accumulated data grows.

7. Citations and References

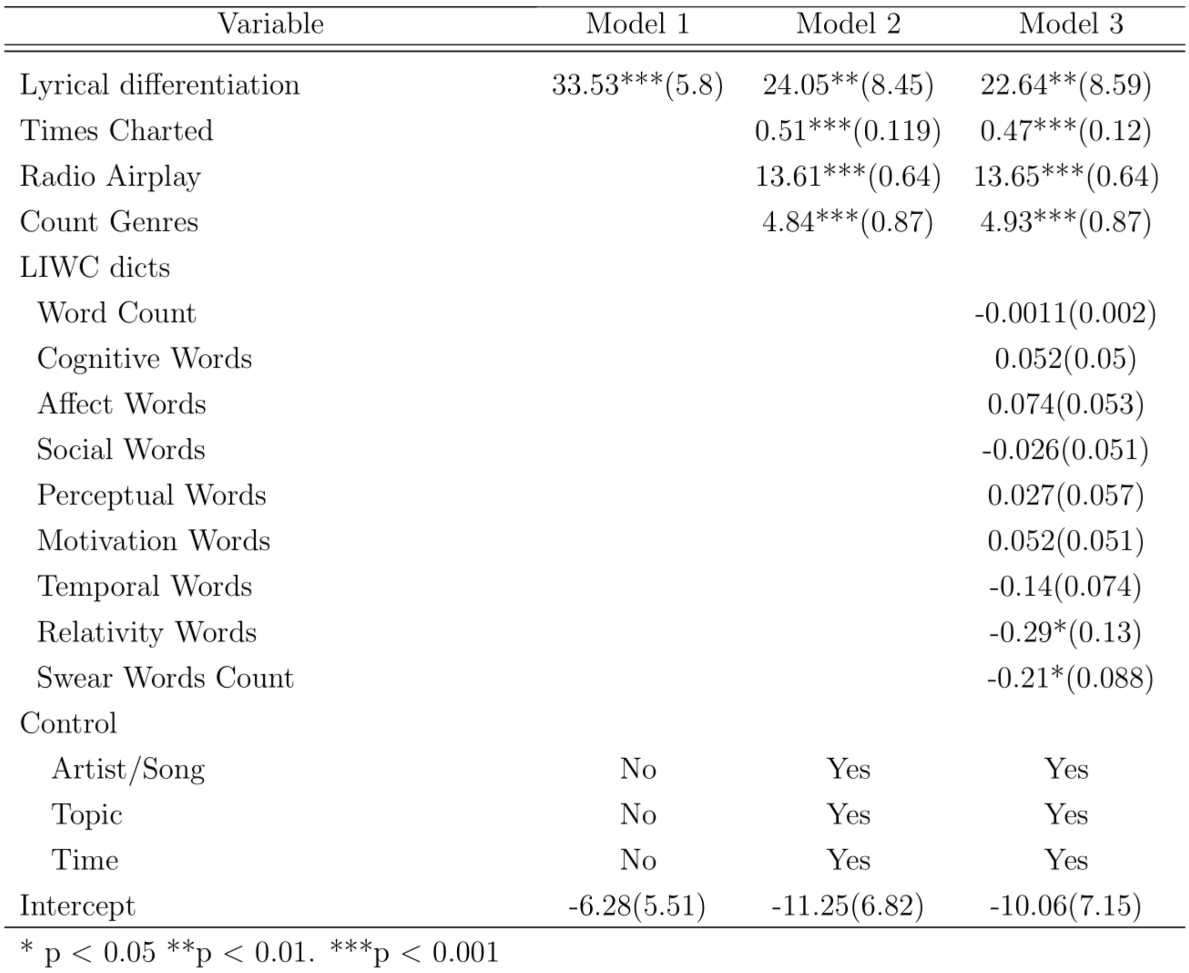
- Araujo, Carlos V. S., Marco A. P. de Cristo and Rafael Giusti (2019), "Predicting Music Popularity Using Music Charts", *Institute of Electrical and Electronics Engineers (IEEE)*.
- Berger, Jonah and Grant Packard (2018), "Are Atypical Things More Popular?", *Psychological Science*.
- Berger, Jonah and Grant Packard (2020), "Thinking of You: How Second-Person Pronouns Shape Cultural Success", *Psychological Science*.
- BBC (2020), "Musicians will lose two-thirds of their income in 2020", https://www.bbc.com/news/entertainment-arts-54966060.
- Bechet, Frederic, Nathalie Camelin, Geraldine Damnati and Renato De Mori (2007), "Speech Mining in Noisy Audio Message Corpus", *Interspeech*.
- Bendle, Neil, Xin Shane Wang, Shijie Lu, X. I. Li and Mansur Khamitov (2021), "Audio Mining: The Role of Vocal Tone in Persuasion", *Journal of Consumer Research*.
- Blei, David M. (2012), "Probabilistic topic models", *Communications of the ACM*.
- Burke, Michael, Peter Hagoort, Franziska Hartung and Roel M. Willems (2016), "Taking Perspective: Personal Pronouns Affect Experiential Aspects of Literary Reading", *Plos One*.
- Dhanaraj, Ruth and Beth Logan (2005), "Automatic Prediction of Hit Songs", *ISMIR*.
- Dhar, Vasant and Elaine A. Chang (2009), "Does Chatter Matter? The Impact of User-Generated Content on Music Sales", *Journal of Interactive Marketing*.
- Financial Times (2021), "Spotify added 3m subscribers in first three months of 2021", https://www.ft.com/content/d2050b0e-5ee5-436b-8cf1-9f70894a60a7.

- Forbes (2020), “Concert Industry Will Lose More Than $30B Because Of Pandemic, Pollstar Finds”, https://www.forbesmiddleeast.com/consumer/entertainment/concert-industry-will-lose-more-than-%2430-billion-because-of-pandemic-pollstar-finds.
- Gandomi, Amir and Murtaza Haider (2015), “Beyond the hype: Big data concepts, methods, and analytics”, *International journal of information management*.
- Green, Melanie C. and Timothy Brock C. (2000). “The role of transportation in the persuasiveness of public narratives”, *Journal of Personality and Social Psychology*, 79(5), 701-721.
- Grewal, Dhruv, Dennis Herhausen, Stephan Ludwig, Jochen Wulf and Marcus Schoegel (2019), “Detecting, Preventing, and Mitigating Online Firestorms in Brand Communities”, *Journal of Marketing*.
- Goldman Sachs (2020), “Music in the Air”, https://www.goldmansachs.com/insights/pages/infographics/music-in-the-air-2020/report.pdf.
- Herzenstein, Michal, Oded Netzer and Alain Lemaire (2019), “When words sweat: Identifying signals for loan default in the text of loan applications”, *Journal of Marketing*.
- Humphreys, Ashlee and Kathryn A. Latour (2013), “Framing the game: Assessing the impact of cultural representations on consumer perceptions of legitimacy”, *Journal of Consumer Research*.
- IBM (2020), “What is Text Mining?”, https://www.ibm.com/cloud/learn/text-mining.
- IFPI, Goldman Sachs Global Investment Research (2017), https://www.goldmansachs.com/insights/pages/infographics/music-streaming.

- Ireland, Molly E. and James W. Pennebaker, (2010). “Language style matching in writing: Synchrony in essays, correspondence, and poetry”. *Journal of Personality and Social Psychology*, 99, 549–571.
- Lee, Jong-Seok and Junghyuk Lee (2018), “Music Popularity: Metrics, Characteristics, and Audio-Based Prediction”, *IEEE Transactions on Multimedia.*
- Mauch, Matthias and Mark Levy (2011), “Structural change on multiple time scales as a correlate of musical complexity”, in *Proc. Int. Soc. Music Inf. Retrieval Conf.*, pp. 489–494.
- Pennebaker, James W. (2011), “The Secret Life of Pronouns,” *New Scientist*, 211 (2828), 42–45.
- Pleus, Mitch and Brian Rossi, “Music Popularity Prediction via Techniques in Deep Supervised Learning”, http://cs230.stanford.edu/projects_spring_2018/reports/8291085.pdf.
- Shulman, Benjamin, Amit Sharma and Dan Cosley (2016), “Predictability of popularity: Gaps between prediction and understanding”, *Tenth International AAAI Conference on Web and Social Media*, pp. 348–357.
- The Economist (2019), “The economics of streaming is changing pop songs”, https://www.economist.com/finance-and-economics/2019/10/05/the-economics-of-streaming-is-changing-pop-songs.
- The Economist (2020), “Will 2021 be another strong year for books?”, https://www.economist.com/the-world-ahead/2020/11/17/will-2021-be-another-strong-year-for-books.
- The Economist (2021), “The pandemic has shaken up the movie business”, https://www.economist.com/the-world-ahead/2020/11/17/the-pandemic-has-shaken-up-the-movie-business.

8. Appendix

Table 2: Results From the Models Testing the Link Between Atypicality and Song Ranking



Variable	Model 1	Model 2	Model 3
Lyrical differentiation	33.53***(5.8)	24.05**(8.45)	22.64**(8.59)
Times Charted		0.51***(0.119)	0.47***(0.12)
Radio Airplay		13.61***(0.64)	13.65***(0.64)
Count Genres		4.84***(0.87)	4.93***(0.87)
LIWC dicts			
Word Count			-0.0011(0.002)
Cognitive Words			0.052(0.05)
Affect Words			0.074(0.053)
Social Words			-0.026(0.051)
Perceptual Words			0.027(0.057)
Motivation Words			0.052(0.051)
Temporal Words			-0.14(0.074)
Relativity Words			-0.29*(0.13)
Swear Words Count			-0.21*(0.088)
Control			
Artist/Song	No	Yes	Yes
Topic	No	Yes	Yes
Time	No	Yes	Yes
Intercept	-6.28(5.51)	-11.25(6.82)	-10.06(7.15)

* $p < 0.05$ **$p < 0.01$. ***$p < 0.001$

Table 3: Results From the Models Testing the Link Between Second Person Pronouns and Song Ranking (*p< 0.05 **p < 0.01. ***p < 0.001)

Variable	Model 1	Model 2	Model 3 ()	Model 4	Model 5	Model 6
You	.0017**(.001)	.0023***(.01)	.001*(.001)	.0018**(.001)	.0017**(.001)	.0018**(.001)
Times Charted		0.4925***(0.069)				
Count_genres		3.8920***(0.453)				
Radio Airplay		9.8654***(0.524)				
Artist(song)			[Random effects incl.]			
Chr_genre				2.89***(.55)		
Cou_genre				3.12***(.55)		
Dan_genre				3.05***(.55)		
Rnb_genre				3.08***(.55)		
Rap_genre				3.26***(.55)		
Pop_genre				2.95***(.55)		
Roc_genre				3.20***(.55)		
q1_14					1.88*(.74)	
q2_14					1.89**(.74)	
q3_14					1.92**(.74)	
q4_14					1.91**(.74)	
q1_15					1.94**(.74)	
q2_15					1.93**(.74)	
q3_15					1.91**(.74)	
q4_15					1.85*(.74)	
q1_16					1.92**(.74)	
q2_16					1.87*(.74)	
q3_16					1.86*(.74)	
q4_16					1.87*(.74)	
ldatopic1						5.26***(1.38)
ldatopic2						3.60**(1.25)
ldatopic3						2.56*(1.14)
ldatopic4						1.20(0.65)
ldatopic5						1.44**(0.54)
ldatopic6						1.36*(0.62)
ldatopic7						2.21***(0.55)
ldatopic8						0.60(1.01)
ldatopic9						2.889**(0.96)
ldatopic10						1.68*(0.79)
Intercept	24.68***(0.35)	15.32***(0.71)	24.20***(0.41)	21.57***(0.31)	22.78***(.32)	22.81***(.36)

Table 4: Results From the Models Testing the Link Between Second Person Pronouns and Song Ranking (Contd.)

Variable	Model 7
You	.0016**(.001)
Cognitive Words	.0005(.0)
Affect Words	-.0012(.001)
Perceptual Words	-4.5e-5(.001)
Motivation Words	0.0001(.0)
Temporal Words	-.0015*(.001)
Swear Words Count	-0.0004(.0)
Intercept	24.87***(.88)

*p< 0.05 **p < 0.01 ***p < 0.001

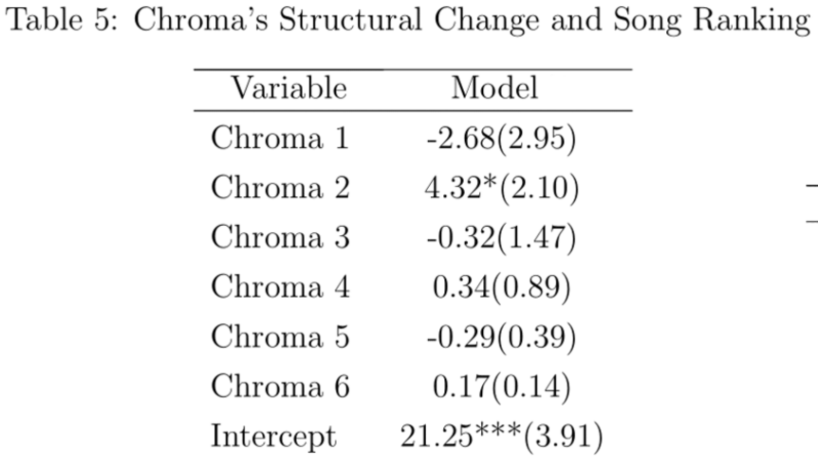


Table 5: Chroma's Structural Change and Song Ranking

Variable	Model
Chroma 1	-2.68(2.95)
Chroma 2	4.32*(2.10)
Chroma 3	-0.32(1.47)
Chroma 4	0.34(0.89)
Chroma 5	-0.29(0.39)
Chroma 6	0.17(0.14)
Intercept	21.25***(3.91)

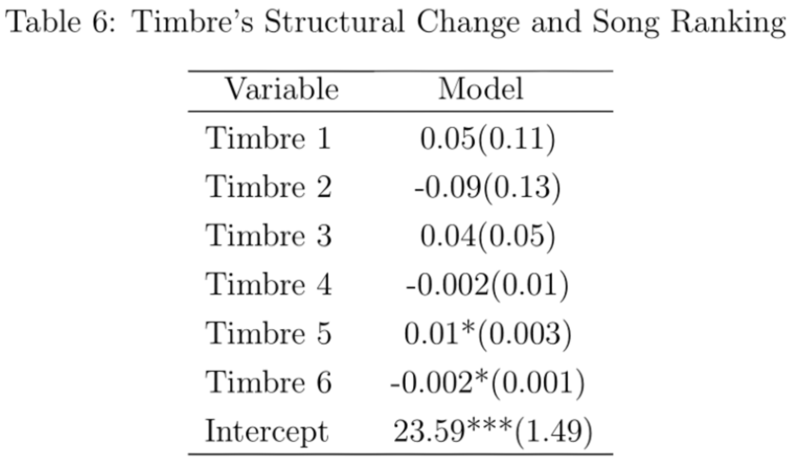


Table 6: Timbre's Structural Change and Song Ranking

Variable	Model
Timbre 1	0.05(0.11)
Timbre 2	-0.09(0.13)
Timbre 3	0.04(0.05)
Timbre 4	-0.002(0.01)
Timbre 5	0.01*(0.003)
Timbre 6	-0.002*(0.001)
Intercept	23.59***(1.49)

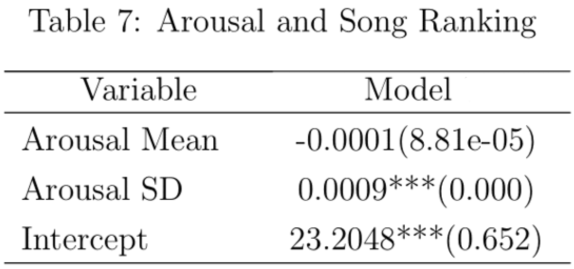


Table 7: Arousal and Song Ranking

Variable	Model
Arousal Mean	-0.0001(8.81e-05)
Arousal SD	0.0009***(0.000)
Intercept	23.2048***(0.652)

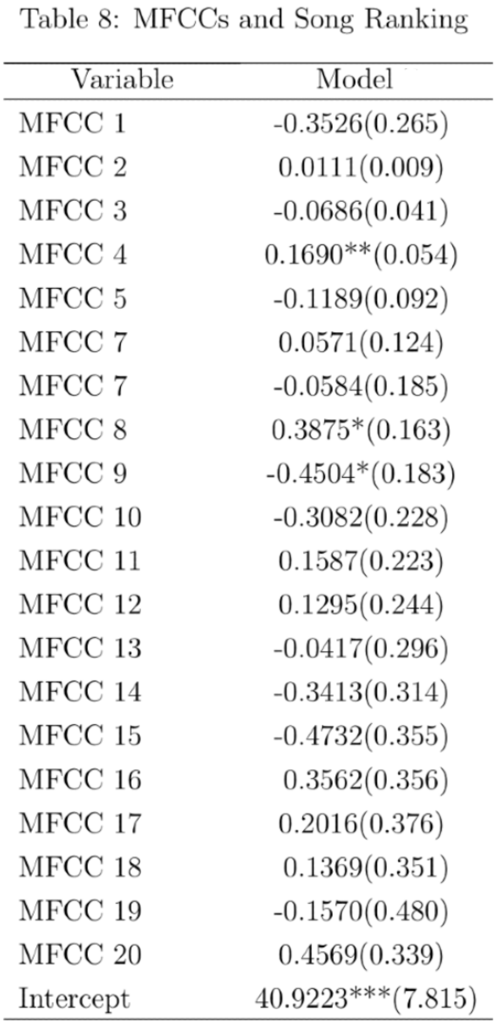


Table 8: MFCCs and Song Ranking

Variable	Model
MFCC 1	-0.3526(0.265)
MFCC 2	0.0111(0.009)
MFCC 3	-0.0686(0.041)
MFCC 4	0.1690**(0.054)
MFCC 5	-0.1189(0.092)
MFCC 7	0.0571(0.124)
MFCC 7	-0.0584(0.185)
MFCC 8	0.3875*(0.163)
MFCC 9	-0.4504*(0.183)
MFCC 10	-0.3082(0.228)
MFCC 11	0.1587(0.223)
MFCC 12	0.1295(0.244)
MFCC 13	-0.0417(0.296)
MFCC 14	-0.3413(0.314)
MFCC 15	-0.4732(0.355)
MFCC 16	0.3562(0.356)
MFCC 17	0.2016(0.376)
MFCC 18	0.1369(0.351)
MFCC 19	-0.1570(0.480)
MFCC 20	0.4569(0.339)
Intercept	40.9223***(7.815)

Link to the data sources :

https://drive.google.com/drive/folders/18FxAAgxQKcgvRJI5b6EHyC-3YPw4Phzr?usp=sharing

Page intentionally left blank