

Examining the Relationship between Spotify Metrics and Musicians

Kaylin Rutkowski and Prapti Patel

2024-05-14

1.0 Introduction

1.1 Project Description For our project we have investigated different artists and their songs. With this description our domain of study is music and the several aspects that impact how songs sound. Spotify uses certain algorithms in order to calculate different features of music and applies them to a quantitative scale. The specific question that we will answer in our project is: How do metrics that Spotify measures help to predict the artist? In other words: What is the relationship between Spotify metrics to the musician producing the song?

1.2 Background In this study, we are examining the relationship between Spotify metrics and the artists producing the songs. It is important to understand what we mean when we refer to “Spotify Metrics”. To begin, Spotify is a large and almost infinite platform that houses endless amounts of songs, podcasts, books, and other audio media. Users of Spotify are able to listen to any song at the touch of their fingertips and discover new musicians by simply listening to one of their many generated collections featuring different artists and genres. Further, Spotify can help listeners curate playlists and find new songs based on their existing preferences. In order to do this, it has to measure various aspects of songs. The metrics we have observed and refer to include: danceability, energy, loudness, speechiness, acousticness, instrumentality, liveness, valence, and tempo. All of these variables will be explained further in the following section.

2.0 Data Description

Our data was downloaded by utilizing the R library “SpotifyR”. This library gives access to functions which can import Spotify artist and song data into the R software. To start we decided to focus on many different artists in order to get unique metrics for every song and artist. These artists include: Ariana Grande, The Beatles, Drake, Fleetwood Mac, Taylor Swift, Coldplay, Lana Del Rey, Luke Bryan, and SZA. When analyzing these artists as a whole, there are far too many observations, 3290 to be exact, and some artists have drastically different discography lengths than others which is shown in the table below.

Table 1: Artist Frequency Distribution

Artist	Songs in Discography
SZA	98
Luke Bryan	168
Lana Del Rey	183
Ariana Grande	185
Coldplay	193
Drake	308
Taylor Swift	577
Fleetwood Mac	760
The Beatles	818

In order to compensate for these differences, we split the data set into two subsets, one containing artists with less than 200 songs and the other with more than 200 songs. This divided the artists into Ariana Grande, Coldplay, Lana Del Rey, Luke Bryan, SZA (827 observations) in one group and The Beatles, Drake, Fleetwood Mac, and Taylor Swift (2463 observations) in the other. In a different R file, we cleaned the data and re-exported it so that we wouldn’t have to reload it using SpotifyR every time it is opened (due to the time it takes to run). When we cleaned the data we got rid of several columns that were not necessary to the analysis like album name, song name, url, key, etc. All we mainly care about in our analysis are artists and quantitative metrics, nothing else. The artist names were also changed to factors so they could be a categorical measure. The metrics that we are focusing on in this analysis are the following (most of them are measures created by Spotify while others are normal music measures):

- **duration_ms** data type: numeric
 - measured in milliseconds for the length of a song.
- **tempo** data type: numeric
 - measured in beats per minute for evaluating the tempo of a song.
- **valence** data type: numeric
 - measured in units of valence, a unique metric created by spotify to evaluate the happiness level of a song.
- **liveness** data type: numeric
 - measured in probability, another spotify metric which evaluates the probability that a song was performed/recorded live.
- **instrumentalness** data type: numeric
 - measured from 0-1.0, evaluates the amount of instrumental presence within the song, higher value is more instrumental.
- **acousticness** data type: numeric
 - measured from 0-1.0, evaluates how acoustic a song is, higher value is more acoustic.

2.0 Data Description (Cont.)

- **speechiness** data type: numeric
 - measured from 0-1.0, evaluates the amount of spoken word within a song, higher value is more spoken word.
- **loudness** data type: numeric
 - measured in decibels (dB), evaluates the overall loudness averaged across the song.
- **energy** data type: numeric
 - measured from 0-1.0, evaluates a perceptual measure of intensity and activity, higher value is more energetic.
- **danceability** data type: numeric
 - measured from 0-1.0, evaluates how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity, higher value is more danceable.

2.0 Data Description (Cont.)

While all of these quantitative measures are key in understanding the features and identity of a song, this is an example of the range of values and frequencies by using a histogram. To help a quantitative visualization of the valence variable, there is also a table including the descriptive statistics of valence.

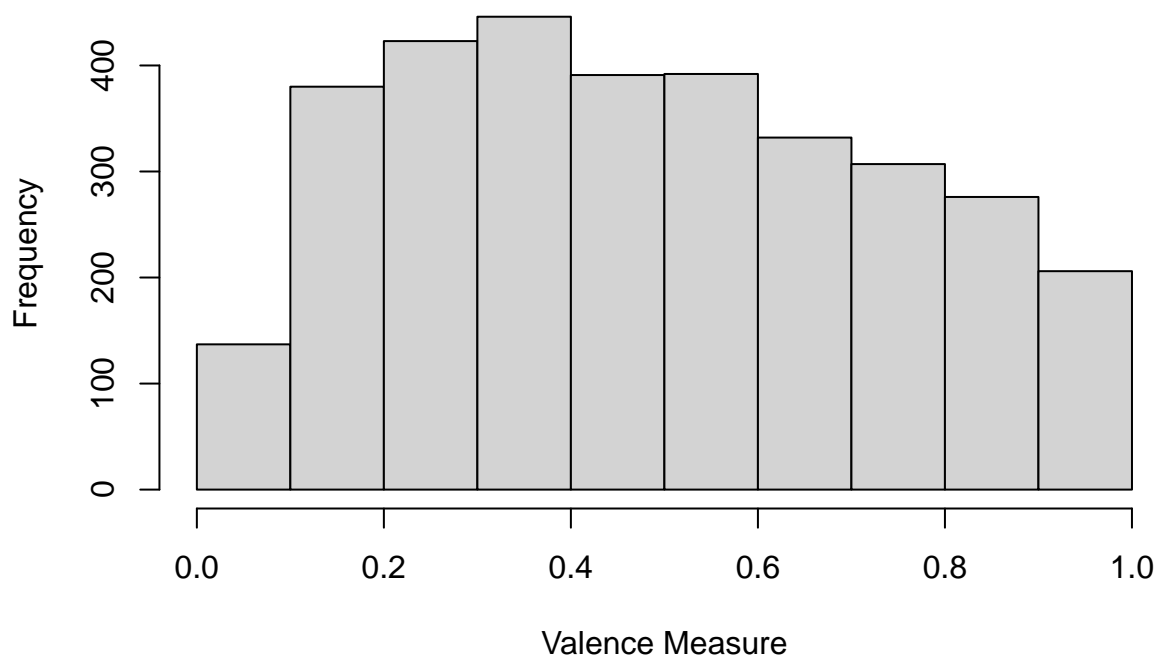


Figure 1: Distribution of Valence Values (all artists)

Table 2: Descriptive Statistics of Valence

Statistics	Values
Min.	0.0000
1st Qu.	0.2760
Median	0.4650
Mean	0.4853
3rd Qu.	0.6850
Max.	0.9800

3.0 Analysis

To recall our exploratory question, we wanted to analyze the relationship between Spotify metrics and the musicians of different songs. Essentially we want to see if there are similarities in the songs within a discography of an artist. For example, do most of Taylor Swift’s songs have the same tempo? Or does Drake produce mostly “speechy” songs? Could these characteristics categorize musicians? So to determine whether these relationships exist we used categorization techniques including decision trees and cluster analysis across the musicians and variables in our data set.

3.1 The first method we used were decision trees. We used the two different subsets of less songs and more songs as we split up earlier within the exploratory analysis of our data. Instead of using a decision tree with all of the variables, we used a random forest in order to determine which variables had the biggest impact. The random forest was made with 500 trees where we found danceability, acousticness, and loudness had the biggest impact for artists with less songs. While duration_ms, instrumentalness, and speechiness had the biggest impact for artists with more songs. So in single tree models, we used only these metrics respectively as independent variables to train data to predict the musicians (dependent variable). The data in each training and test set was split 75:25 (training:testing). So the less song data set had 622 training observations and 205 testing observations, while the more song data set had 1846 training observations and 615 testing observations.

3.1.1 Decision Trees:

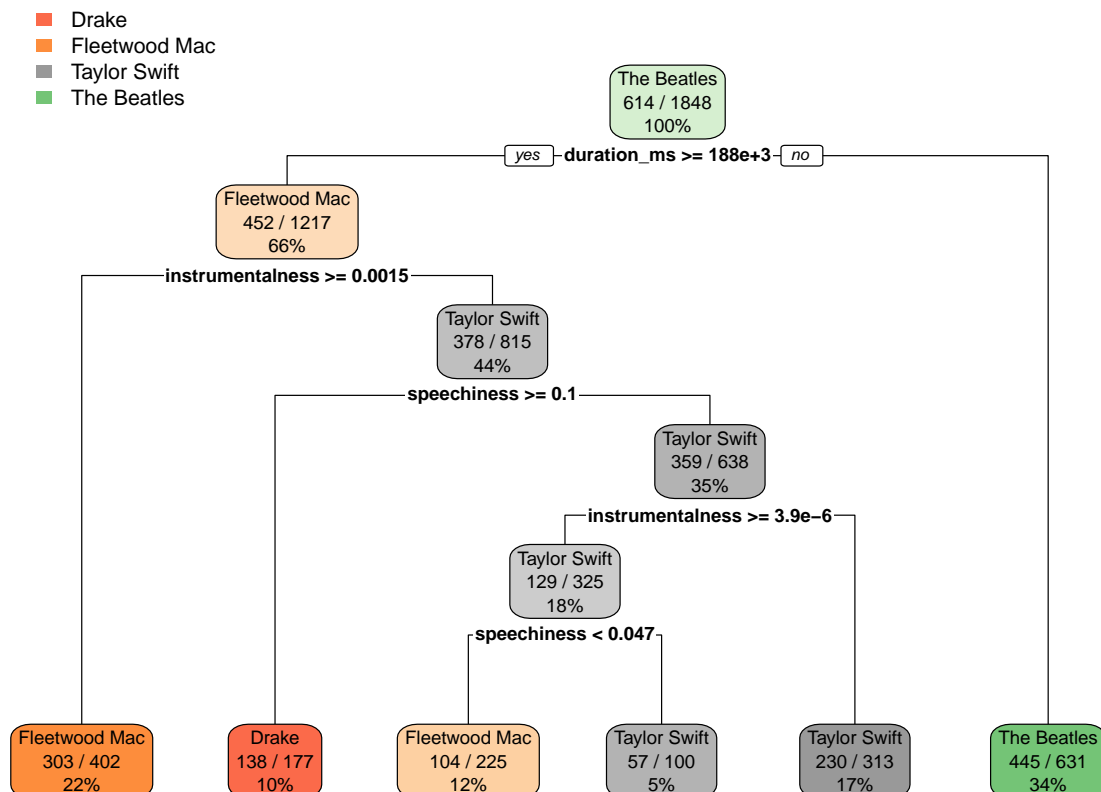


Figure 2: Prediction Decision Tree for Artists with More Songs

3.1.1 Decision Trees (Cont.):

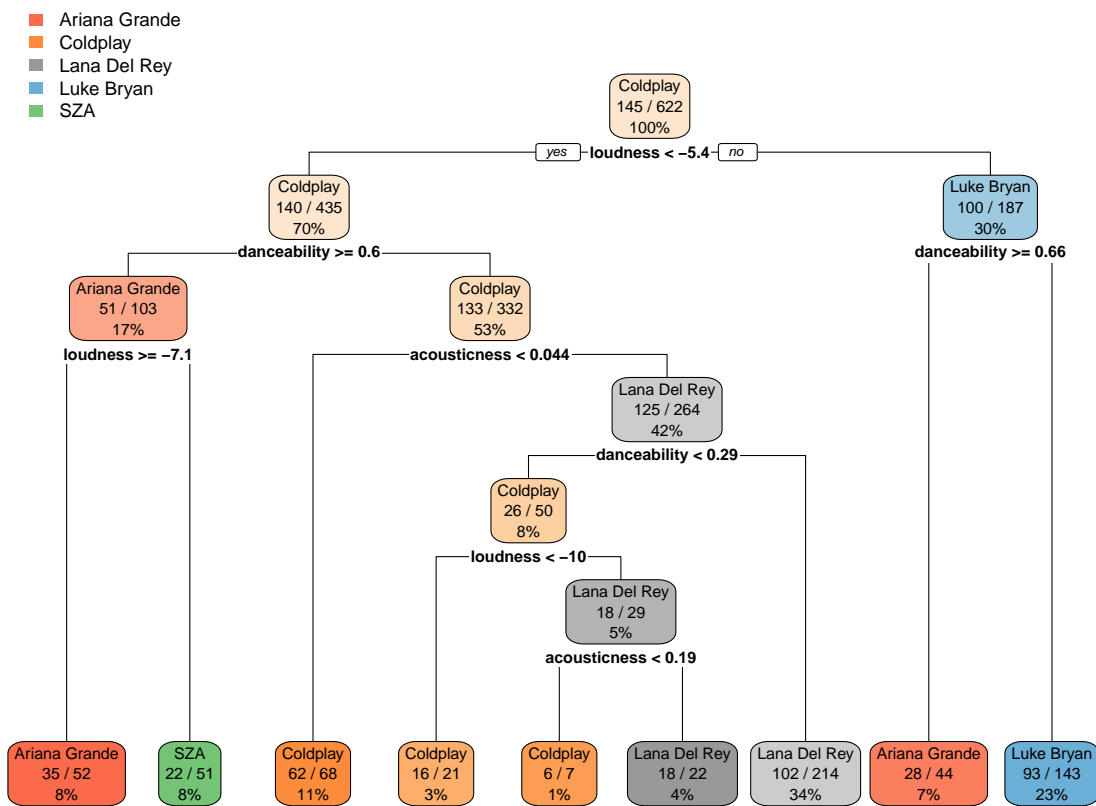


Figure 3: Prediction Decision Tree for Artists with Less Songs

Table 3: Accuracy and P-Value for Each Subset

Subset	Accuracy	P-Value
More Songs (>200)	0.6341	0.0002214
Less Songs (<200)	0.5610	0.0001060

The decision trees above are how we trained the data in order to complete the following analysis.

From using decision trees to train and test the categorization of our data we found that they were able to distinguish musicians in a statistically significant way for both the larger and smaller artists. While the accuracy rates were not too much better than a random model as they were both less than 20 percentage points greater than 50, the p-values in both training sets proved to be much less than .05 proving the analysis to be statistically significant in some sense.

3.2 The second method we used was clustering. Again the data was split up into the artists with less and more songs. We did this in both analyses in order to make the data easier to categorize and work with. Since there were over 2000 songs in total, it was easier to make the groups smaller so the models did not have to cluster into 9 different categories and instead to 5 and 4. To cluster the data we used a k-means model.

3.2.1

Table 4: Less Songs K-Means Clustering

Ariana Grande	Coldplay	Lana Del Rey	Luke Bryan	SZA
23	0	4	3	27
118	21	39	156	39
29	31	105	7	28
0	37	3	0	1
15	104	32	2	3

Table 5: More Songs K-Means Clustering

Drake	Fleetwood Mac	Taylor Swift	The Beatles
68	242	195	282
3	1	4	38
12	182	17	74
225	335	361	424

Table 6: K-Means Cluster Analysis

Subset	Total Withiness	Betweeness
More Songs (>200)	16771.410	7848.588
Less Songs (<200)	4427.692	3832.308

In our K-Means cluster analysis we were hoping to find that there would be clear and distinct clusters for each of the data points in our subsets. Unfortunately, none of our clusters had 100% purity or something that was even relatively close. Something that we were able to find was that the subset with less songs had better performance at creating contained clusters as they had a higher total-withiness, while the subset with more songs had better seperation between clusters as their betweeness was higher.

4.0 Conclusions

After a comprehensive analysis of all the data and the models we discovered the answer to our guiding question: What is the relationship between Spotify metrics to the artist producing the song? Contrary to our initial expectations, there are no major similarities in songs within a discography of an artist and these characteristics do not necessarily help to categorize musicians. While the models yielded statistically significant results the overall results were subtle and not useful. There are a few threats to validity in the analysis. One threat could be that the selection of artists including in the analysis may not be representative of all musicians on Spotify. This could lead to biased conclusions about the relationship between Spotify metrics and artists if certain genres or demographics are overrepresented or underrepresented. Another threat to validity is the choice of Spotify metrics used in the analysis. The features included may not capture all important aspects of music that influence the relationship with artists and their songs.

5.0 References

Spotify Developement Website (used to aquire the access key to the artist data sets): <https://developer.spotify.com/dashboard>