Musical Characteristic Analysis of Top 200 Ranked Spotify Songs

Sam Roberts-Baca
University of Denver, COMP4449

PROBLEM STATEMENT AND RESEARCH QUESTIONS

A new, unsigned, independent music artist from the United States who has a variety of genres of music released on Spotify desires to know the best response in terms of number of streams. What kind of music in terms of musical characteristics (danceability, acousticness, etc.) should he make next to generate the most streams?

The artist would like to know what kinds of music in terms of its musical characteristics performs the best in terms of number of streams. They would also like to know what musical characteristics are shared between top charting songs, and also how their own music compares to that of top charting songs in terms of musical characteristics. To that end, there are several research questions we wish to answer through our analysis:

1. What commonalities are shared between top-charting songs in terms of musical characteristics? Between an individual artist and top-charting artists?
2. Can the musical characteristics of any given top-charting song predict the number of streams it has?

DATA DESCRIPTION

For this project, three datasets were used: The *Spotify's Worldwide Daily Song Ranking* dataset, the *Spotify Song Attributes* dataset, and a dataset of song attribute observations collected by myself using Spotify's API.

The *Spotify's Worldwide Daily Song Ranking* dataset was obtained from kaggle[1]. This dataset contains the daily ranking of the 200 most listened to songs in 53 countries from 2017 and 2018 by users of Spotify, an audio streaming and media services provider. This dataset comprises of 18,598 observations, where each observation represents a song that charted on a specified date between January 1st 2017 and January 9th 2018 in a specified country. Each observation includes the position on the charts at that date and county, as well as the track name, artist, and number of streams.

The *Spotify Song Attributes* dataset was also obtained from kaggle[2]. This dataset contains 2017 popular songs from 2017 that were selected from a large variety of genres, not necessarily within the top 200 most listened to songs from the *Spotify's Worldwide Daily Song Ranking* dataset. Each song observation in this dataset contains 13 musical attributes as retrieved by Spotify's API for each song. Each observation was also labelled by the data collector under a "target" column, based on the data collector's personal preference to the song, a "1" signifying that the data collector liked the song, and a "0" signifying that the data collector disliked the song. This dataset also included fields for the track name and artist. The musical attributes are described as follows, as defined by Spotify[3]:

---

[1] https://www.kaggle.com/datasets/edumucelli/spotifys-worldwide-daily-song-ranking
[2] https://www.kaggle.com/datasets/geomack/spotifyclassification
[3] https://developer.spotify.com/documentation/web-api/reference/#/operations/get-audio-features

- **Acousticness** - A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
- **Danceability** - Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.
- **Duration_ms** - The duration of the track in milliseconds.
- **Energy** - Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.
- **Instrumentalness** - Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.
- **Key** - The key the track is in. Integers map to pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on. If no key was detected, the value is -1.
- **Liveness** - Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.
- **Loudness** - The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typically range between -60 and 0 db.
- **Mode** - Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0.
- **Speechiness** - Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.
- **Tempo** - The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.
- **Time_signature** - An estimated time signature. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure). The time signature ranges from 3 to 7 indicating time signatures of "3/4", to "7/4".
- **Valence** - A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).

The third dataset was obtained using Spotify's API by collecting musical characteristic information from a personal playlist[4] which included all 165 songs that this researcher released on the platform, where each song also included the 13 musical attributes as determined by Spotify's API as listed above.

---

[4] https://open.spotify.com/playlist/4Abs2Ey8vgdxy81aSMReQt?si=efaa55397c9847b4

DATA PREPARATION

For this analysis, the *Spotify's Worldwide Daily Song Ranking* and *Spotify Song Attributes* datasets were merged together based on the "Track Name" column to produce two new datasets: one which aggregated songs by total streams globally, and one which aggregated songs by total streams in the United States (US). All observations from the *Spotify's Worldwide Daily Song Ranking* dataset which had Date values beyond December 31st 2017 were removed, as there were no 2018 songs included in the *Spotify Song Attributes* dataset. Observations in the *Spotify's Worldwide Daily Song Ranking* dataset where the region (country) value was set to "global" were also removed, as there were several songs present in this dataset where there was no "global" value, which would have significantly reduced the amount of songs available for analysis. There were no observations in either dataset with missing values. Tracks were then aggregated by total amount of streams across all dates in 2017 by region, as well as globally using the summation of streams in each region across all dates. For the purpose of this analysis, the songs with US region were also filtered out. After merging these two datasets, there were 471 globally charting songs and 168 US charting songs present with both streaming and musical characteristic data.

The third dataset which was collected by the researcher using Spotify's API was left intact without any modifications.

DATA ANALYSIS METHODOLOGY

First, an exploratory analysis was conducted where the distributions of musical characteristics from both the globally aggregated dataset and the US-only aggregated dataset were compared against eachother, as against the musical characteristics of the researcher's personal discography.

After the globally charting data was tested for multicollinearity, it was determined that the time signature and duration_ms variables exhibited strong covariance, and were removed from further analysis. Instrumentalness was also removed, as the mean of the distribution of instrumentalness was 0.03, suggesting that the overwhelming majority of top charting songs were not instrumental. The "target" variable was also removed, as we were not interested in the subjective preference of songs on behalf of the data collector of the *Spotify Song Attributes* for this study.

PCA was performed Songs in the globally aggregated dataset to determine the dimensionality of the data. A K-means, Affinity Propogation, Birch, and Gaussian Mixture model were used in attempt to categorize globally top charting songs by their musical characteristics after applying PCA. These four clustering models chosen provided a suitable range of clustering methods applicable for the dataset. Four components were used for initializing the K-means model. Affinity Propogation was conducted with a max iteration value of 250, and a damping value of 0.9. The Birch model used a threshold value of 0.01, and four clusters. The gaussian mixture model algorithm was conducted with four clustered. Each method was then visually compared and tested for accuracy using a logistic regression model to determine the most accurate model for categorizing song data by musical characteristics.

Finally, the global charting song data was split into a testing and training set with a test size of 0.33, and two multivariate log-linear regression models were fitted to the globally charting song data to determine if musical characteristics could predict number of streams for globally charting songs. The first model used all ten musical characteristics (acousticnes, danceability, energy, key, liveness, loudness, mode, speechiness, tempo, and valence) as predictor variables. The second model used a forward selection process to determine the best-fitting predictor variables. Both models used a log-transformed "Streams" response variable, as this variable followed a non-linear, exponential distribution.

RESULTS

Exploratory Analysis Results

From our initial exploratory analysis of both the global and US-charting songs on Spotify from 2017, we observed that the energy and loudness characteristics were strongly correlated in both global and US datasets. The energy and acousticness values were somewhat negatively correlated – in other words, it appears that songs with higher acoustic values tended to have lower energy or or overall intensity. Streams and position values were strongly negatively correlated, which is to be expected as the more streams a song has, the higher up it is on the charts (with a lower corresponding chart position).
The vast majority of chart topping songs from 2017 are in 4/4 time signature. Danceability of songs tended to follow a left skewed normal distribution for both global and US charting songs, with a mean of .66 for globally charting songs, and a mean of .67 for US charting songs. The duration of songs appeared to follow a somewhat right skewed normal distribution, with a mean of approximately 3.83 minutes for globally charting songs on Spotify, and approximately 3.77 minutes for US charting songs on Spotify.

Figure 1 shows the distributions of musical characteristics for both globally charting and US-only charting songs on Spotify for 2017. We observe that the US-only charting songs' musical characteristics are remarkably similar to those of globally charting songs.

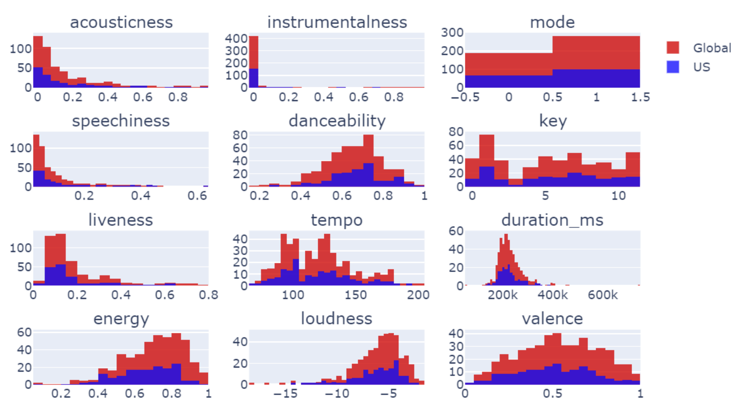

Figure 1: 2017 Global vs. US Chart Topping Song Musical Characteristic Distributions

Figure 2 shows the distributions of musical characteristics for songs in the researcher's discography compared to US charting songs for 2017. Instrumentalness, speechiness, and liveness appear to align between US charting songs and the researcher's discography. In general, the researcher's songs appear to be somewhat less danceable, more evenly distributed in terms of acousticness, and somewhat more negative sounding overall as determined by the valence value.

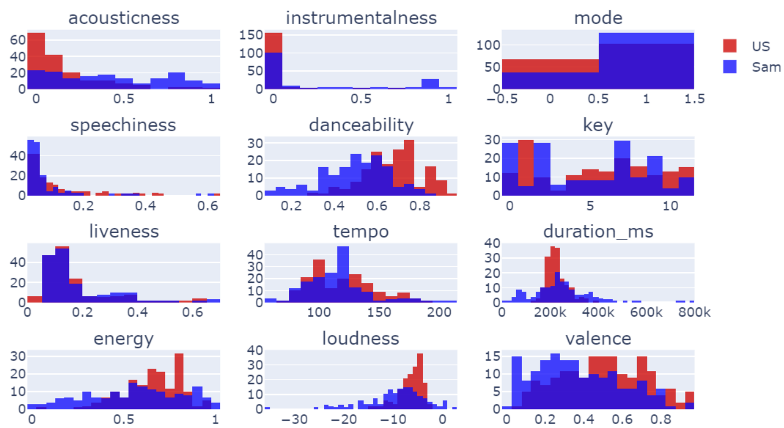

Figure 2: 2017 US Chart Topping Songs vs. Researcher's Discography Musical Characteristic Distributions

PCA and Clustering Algorithm Results

Figure 3 shows the elbow plot and hierarchical clustering dendrogram used to determine the appropriate number of dimensions to choose for our initial PCA. Four clusters were determined to be appropriate for our initial PCA analysis, to reshape the top globally charting songs in the dataset for use by four different clustering algorithms.

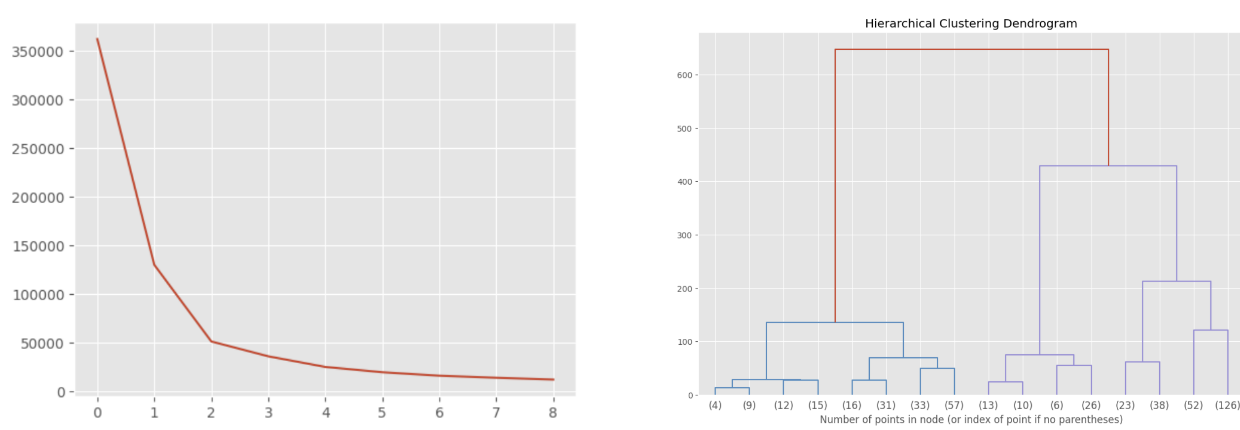

Figure 3: PCA Elbow Plot and Hierarchical Clustering Dendrogram

Figure 4 shows the characteristics determined using four clustering algorithms. K-Means clustered songs into four clusters, Affinity Propogation clustered songs into fourteen clusters, Birch clustered songs into two clusters, and GMM clustered songs into four clusters. Each model was tested with a logistic regression model using the ten selected musical characteristics to predict the cluster groups, and the accuracy scores determined for each clustering model were as follows: K-means: 1.0, Affinity Propagation: .8238, Birch: 0.9894, GMM: 0.9724. Clustering songs via K-Means clustering with four clusters provided the best results in terms of accuracy score.

After the multivariate log-linear model regression model was trained and used to predict the training and testing set outputs. All ten musical characteristics (acousticnes, danceability, energy, key, liveness, loudness, mode, speechiness, tempo, and valence) were used as predictor variables. The training set R^2

value was 0.05, and the testing set R^2 value was 0.09. The training mean-squared error (MSE) was 11.39 and the testing MSE was 10.41.

Figure 4: Musical Characteristics using K-Means, Affinity Propogation, Birch, and GMM Models

A second multivariate log-linear model was produced by applying forward selection. The training set R^2 value was 0.05, and the testing set R^2 value was 0.06. The training MSE was 10.07 and the testing MSE was 11.80.

DISCUSSION

These results show that the most influential variables that determine the number of streams any given song will have is danceability. This would appear to suggest that the more danceable a track is, the more likely it is to be well streamed, which is an interesting though perhaps somewhat obvious observation, supported by the fact that the top streamed songs of 2017. All of the clustering algorithms clustered songs into distinct groups distinguished most obviously by tempo. K-means performed the best in terms of accuracy in clustering songs, while affinity propagation was the worst performing model in terms of visually distinct clusters and overall accuracy as determined by the logistic regression model.

It was not clear that a relationship existed between musical characteristics and number of streams, as both multivariate log-linear models performed poorly in terms of providing a strong correlation between musical characteristic predictor variables and streams.