Introduction:

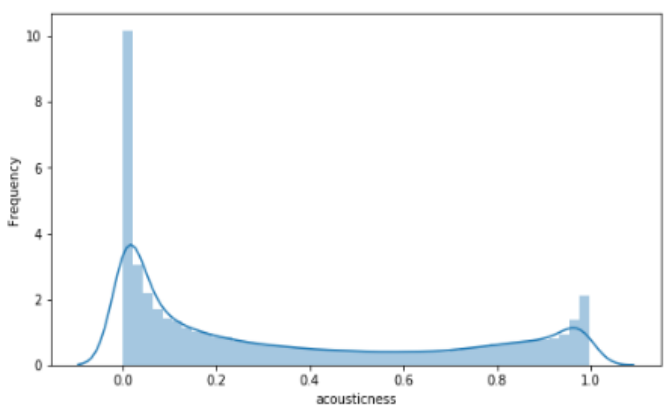
What exactly are the musical elements that create a hit song? Each and every song that exists has unique audible elements, but we aim to study whether or not there are driving similarities that exist between songs and artists. Through the use of an extensive data bank containing 18 different traits on 228,159 songs, we strive to determine the significance of these audible traits both within and between genres.

This data comes from Swedish music platform Spotify, that was launched in 2008. Spotify has amassed a user base of 217 million users, 100 million of which are paid subscribers, while the others must listen to ads in order to access their music. Spotify does have many competitors in the marketplace of free-to-access music, however using data from Spotify rather than a different service will not add any bias into our model. This stems from the fact that the traits in the data set (acousticness, danceability, energy, tempo, etc.) are specific to each song and therefore universal across all platforms.

This data was downloaded from Kaggle, an internet community centered around the world of data. As stated before, this data set consists of 228,159 songs from 26 different musical genres. The variables are all unique, and are of multiple types as well. For sake of our analysis, we have removed the variables that could be considered categorical in nature. The remaining variables in the data set are:

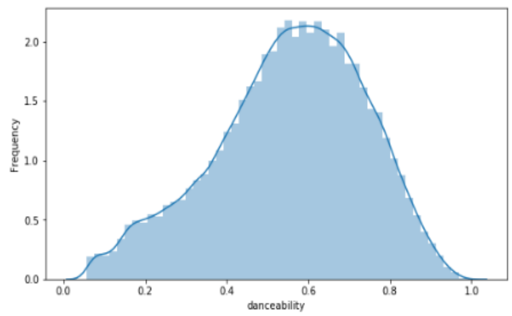
Acousticness:

A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.



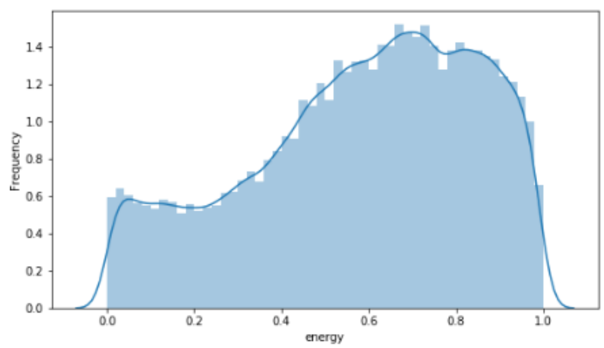
Danceability:

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.



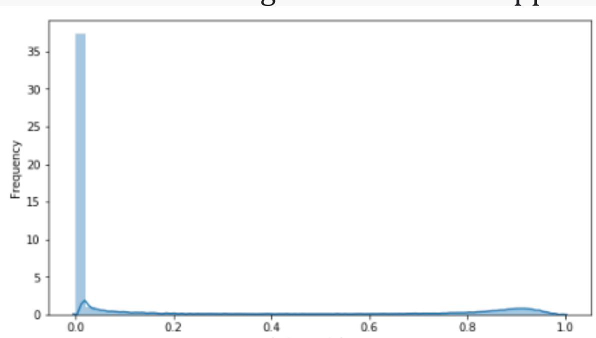
Energy:

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.



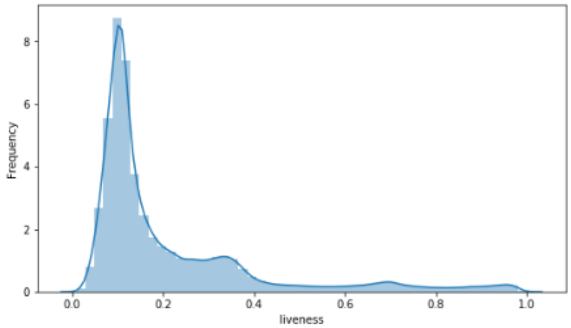
Instrumentalness:

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.



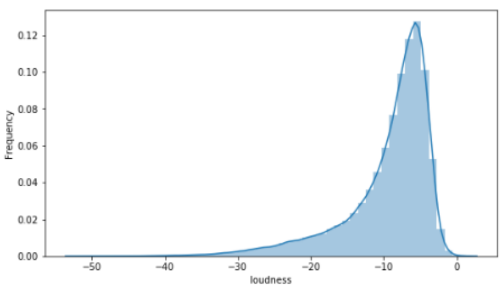
Liveness:

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.



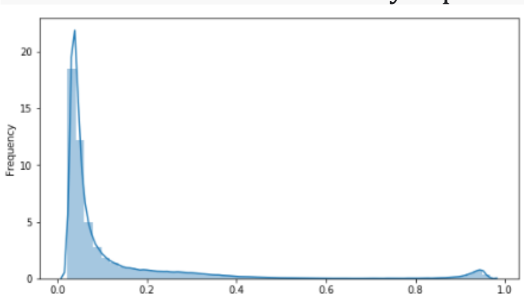
Loudness:

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 typical range between -60 and 0 db.



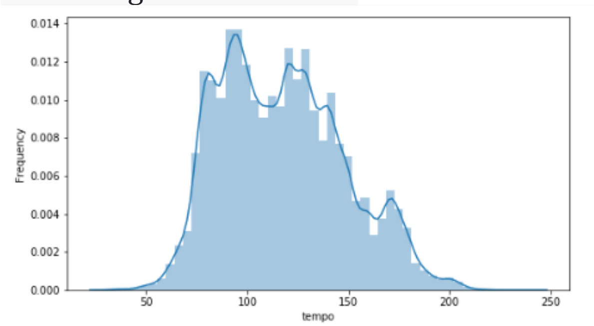
Speechiness:

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.



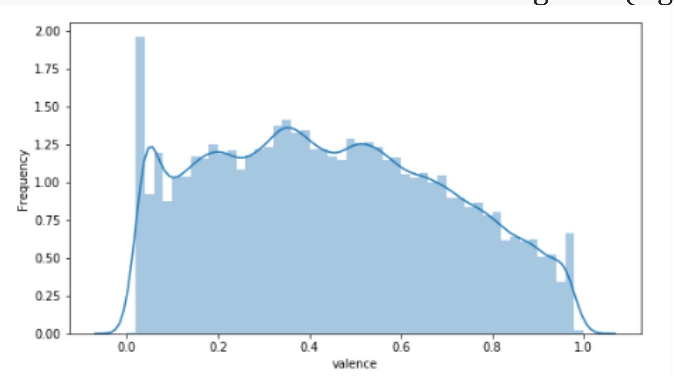
Tempo:

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.



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).



Source: https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/

- Text descriptions were sourced from this website, visualizations depicted via Python

Through the use of different methods of analysis, our group aims to determine what audio features determine the popularity of a song within its genre/subgenre.

Methods:

We started off by wrangling our data to limit our dataset and reduce noise. Making the data better suited for addressing our research question. The process entailed:

1. Getting rid of categorical variables
2. Reducing our scope to the top 50% of songs by popularity
3. Assigning each group member a genre to analyze

Given our large dataset with 26 genres, each of which containing multiple subgenres we suspected that we might find some discrepancies between the genre analysis and specific artists within it. We broke down our respective genres and found that there were indeed differences between the overall group trend and the trends for specific artists. This was basically just a realization of Simpsons paradox which explains how trends may become distorted or even completely disappear due to specific mis-grouping of data. While this was great insight, we needed to better understand the variations within our data.

We utilized a variety of visualization techniques, including; heatmaps, scatter plots, regression plots, scree plots, and histograms. Each allowed us to look at our data in a different way, which helped give us an idea of what the data was saying.

However, most of us can only visualize two or three dimensions, so the most logical next-step for our 10 dimensional data was reducing its dimensions. This makes the data easier to interpret and speeds up these computations. Dimensionality reduction also helps us find appropriate functions of the predictors that correspond to interesting features, get rid of unwanted predictors, and removes contamination from measurement noise. We proceeded using Principal Component Analysis (PCA) as this was a great way to quantify and pinpoint the variation within our dataset. PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. These new principal components act as new axes, which help reduce the dimensionality of our dataset without loss of information. To better interpret these components we used loading plots of our new axes and examined what variables were the most heavily weighted as a means of better understanding our dataset.

Since the goal was to find out what makes *any* song popular, we concluded our analysis by testing our findings on all 26 genres. We plotted each genre's median audio features against their median popularity in search of some trend or correlation between audio features and popularity.

Conclusion:

After running all the analysis, our fear of Simpson's paradox showed true on a much deeper scale than we thought. When first breaking down the data and looking for trends, it seemed likely that the overall dataset would not provide much light on what affects popularity. What was daunting, was also how little insight the data offered us when broken down by genre, which was our original idea of why Simpson's paradox would occur. This turned out to be unworthy and lead us into a deeper level of thought. While PCA score plots by genre did not yield any obvious sub-groupings within a particular genre, our intuition behind another Simpson's paradox within the genres did not waiver. To check this, we decided to look at separate bands and their trends between their song attributes and popularity. It finally confirmed what we had been looking for and gave us insight that we could elaborate on in order to identify important attributes.

Upon further scrutinizing, the attributes of tempo, danceability, and acousticness appeared to have the largest effect on a song and its popularity. This was relatively similar across all genres and artists within them, but what varies is how each change in attribute correlated with popularity. A change in the value of any of these 3 attributes could have adverse or advantageous results in popularity depending on the genre. Tempo's importance became much more clear in the breakdown of PCA, as it had the strongest loading on PC1, which contained anywhere from 96% to 99% of the data depending on the genre.

Overall, many of our suspicions about our data were confirmed. It is true that there is a relevant relation between popularity and some of a particular song's attributes even though the correlation may not be extremely strong. A weak correlation is not a problem though, and is actually what we expected to happen. If there were strong correlations between a particular music attribute and music popularity, artists would do whatever it took to adjust those attributes in order to maximize popularity and revenue. Since this does not happen in music today, we knew the correlation would be hard to find. In fact, we feel that it is a good thing that these correlations are weak as it creates little incentive to be similar in music and creates the diverse spectrum of sound that makes up our musical palette.