Term Project – Team 3

Topic – Spotify Top Tracks

Katrina Sutton, Rilov Kulankara, Sean Lin



Contents

[1. Introduction 3](#_Toc38031398)

[1.1 Objective 3](#_Toc38031399)

[1.2 Data Overview 3](#_Toc38031400)

[2. Data Preparation 5](#_Toc38031401)

[3. Data Analysis 5](#_Toc38031402)

[3.1 Attribute Correlation 5](#_Toc38031403)

[3.1.1 Duration 6](#_Toc38031404)

[3.1.2 Loudness 8](#_Toc38031405)

[3.1.3 Tempo 9](#_Toc38031406)

[3.1.4 Time Signature 9](#_Toc38031407)

[3.1.5 Tonality 10](#_Toc38031408)

[3.1.6 Danceability 12](#_Toc38031409)

[3.1.7 Energy 14](#_Toc38031410)

[3.1.8 Acousticness 15](#_Toc38031412)

[3.1.9 Instrumentalness 16](#_Toc38031413)

[3.1.10 Valence 17](#_Toc38031414)

[3.2 Common Features for Popular Songs 18](#_Toc38031415)

[3.3 Predicting Song Popularity 18](#_Toc38031416)

[4. Conclusions 19](#_Toc38031417)

Spotify 2018 Top Songs

# 1. Introduction

The data we have chosen to use is a list of Spotify’s most downloaded songs in 2018. It includes the name of the artist, song title as well as multiple song attributes created by Spotify to understand different variables about the songs. These song attributes are referred to as audio features and were extracted from the Spotify Web API.

## 1.1 Objective

The first objective we had for this assignment was to see if we could determine what in particular made a song popular. We wanted to know if there were any audio features that drove song popularity, then extended it further to see what those similarities were. Once we observed the similarities between attributes, we felt this would give us an understanding of why people liked these songs, and what people wanted to listen too. Lastly, we wanted to use our findings to see if we could use predictive modelling techniques to see if we could calculate whether or not a song would be popular based on these attributes.

## 1.2 Data Overview

The 2018 Spotify dataset we are using was derived from Kaggle. As part of the dataset there was categorical values such as the artist name and track title, but there were also numerous audio features that represented a numerical ranked value. These audio features were defined from Spotify; listed and defined below:

# 2. Data Preparation

In order to ensure accuracy and validity in the data there were a few steps we took to clean and scrub the dataset. The first step was removing categorical values such as artists and song names. We felt that in order to determine popularity, we wanted to use only numerical values so we could quantify the popularity metric. Since this was a strong structured dataset, we luckily did not have to transform missing values or dealing with null types. All values were populated, therefore making the cleaning process quite straightforward. Because of the lack of missing values and the data source, we felt this was a very strong dataset to work with. In addition, since the data derived from Kaggle in csv format, we were able to important data into Jupyter.

The first challenge we faced with the dataset was the number of attributes that were available to perform analysis on. Because there were so many different attributes to choose from, we felt that there was so much information and so many ways to slice the data to answer our objective. In order to understand where to start, we created a heat map which best illustrated where we should look first, allowing us to find a strong starting point in our analysis.

Another challenge we faced was organizing the data in a way that would allow us to easily show the correlation between all the different song attributes. By looking at each data attribute individually, we were able to determine which attributes were linked to popularity. Then from there we leverage machine learning to build a predictive model.

# 3. Data Analysis

## 3.1 Attribute Correlation

When starting off our understanding to see what attributes were correlating together, we decided to use a heat map to plot all the different song attributes.

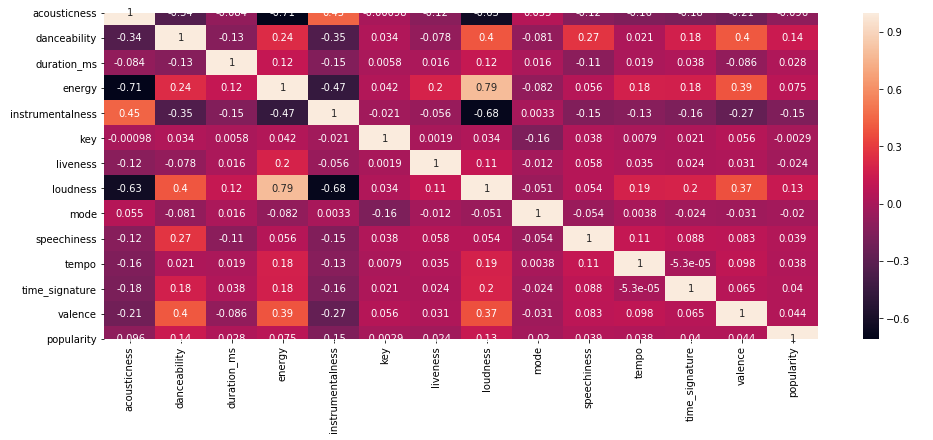


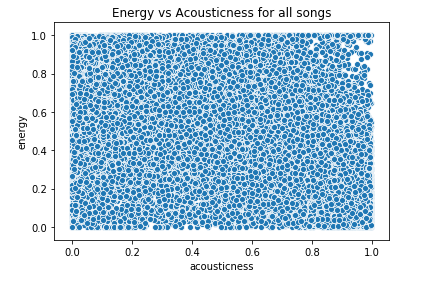
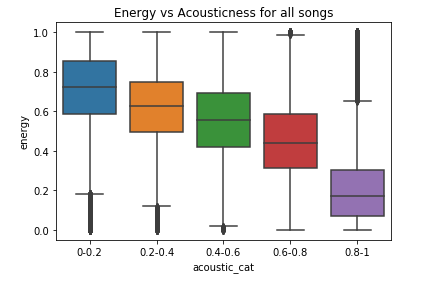
Figure – Heatmap showing correlation coefficients between all attributes

We were able to derive quite a few insights from this heat map. The first thing we wanted to investigate was whether or not there was a linear relationship between particular attributes. We know that when the value of correlation is close to zero, generally between the values of (-0.1 and +0.1), the variables are said to have a weak or no linear relationship. In addition, we know whether the attributes were correlated, which essentially means there was a relationship between two variables. We also wanted to understand whether these variables had a positive correlation, meaning that if popularity increased so would that variable set, or if it had a negative correlation, meaning there was an inverse relationship between them.

Our conclusions from this heatmap told us why people liked the most popular songs. Our findings told us that danceability, energy, and loudness all had a positive correlation to popularity. This meant as the score for each of these attributes increased, the song popularity increased as well, and were essentially the reason why people listened to these songs.

Before we could further analyze which attributes influenced popularity, it was first necessary to bin many of the continuous attributes into discrete ranges. As the dataset we were working with contained over 100,000 data points and most attributes were only weakly or moderately correlated with each other, visualization was difficult.

Figure – Acousticness vs energy. According to the heatmap, these attributes have a correlation of R = -0.71, but this relationship is not visible without first discretizing one of the attributes

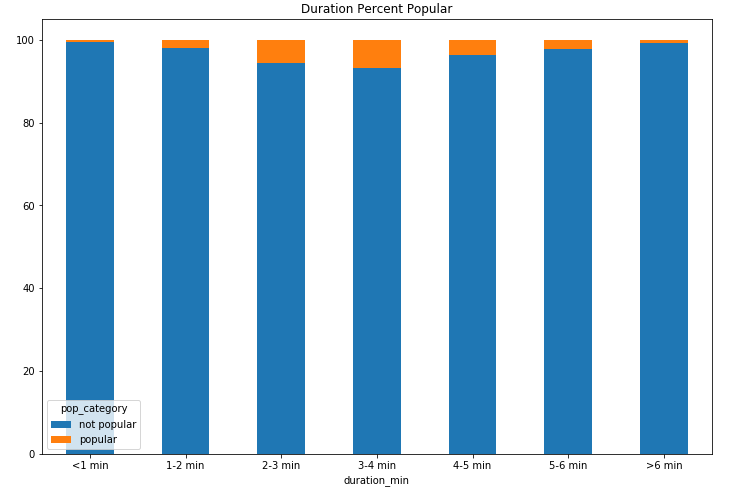
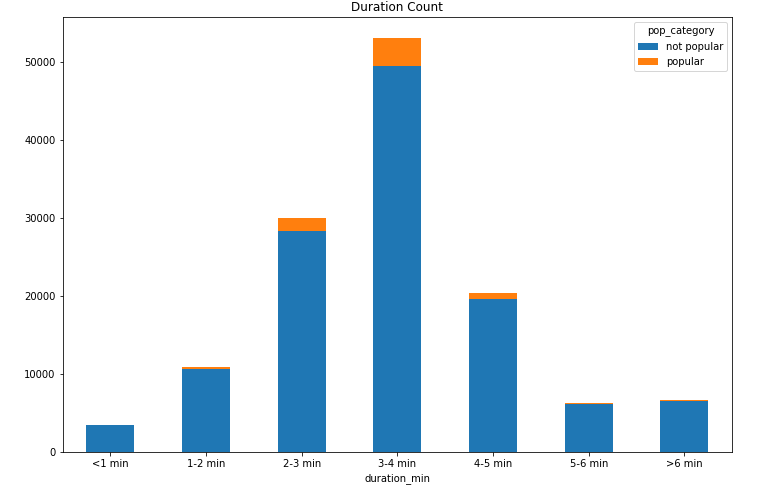


Continuous attributes that were given on a scale of 0-1 were binned into discrete ranges of 0.2, allowing for easier analysis and visualization using boxplots and bar graphs. Additionally, as popularity was not strongly correlated with any attribute, we defined a popular song to be one with a popularity value greater than 60. This corresponded to approximately the top 5%, or 6500, of all songs.

Next, we examined the impact of each attribute on song popularity.

### 3.1.1 Duration

Song lengths were given as an integer value in milliseconds. To improve legibility, we converted these numbers to time ranges in minutes.



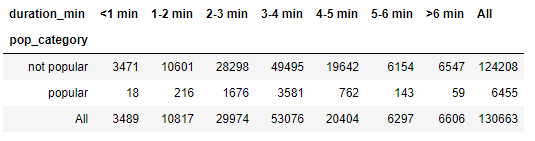


Figure – Frequency and rate of popularity for songs compared to song length

The data shows that medium-duration songs of 3-4 minutes are the most common category in absolute numbers, with a steep drop in frequency both on the shorter and longer sides. We then normalized each category by their song count to obtain the proportion of popular vs. non-popular songs for each respective category. This revealed that the 3-4 minute range also contained the highest proportion of popular songs, at 6.75%, while both extremely short and long songs experienced less than 1% “popular” rate.

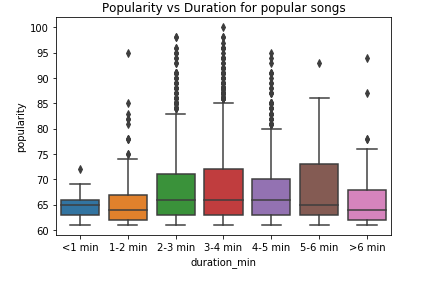
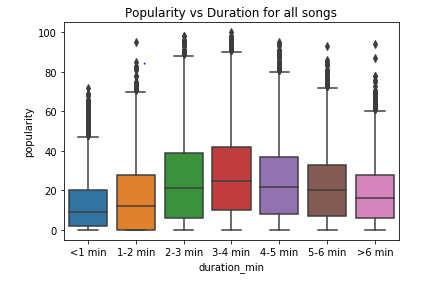
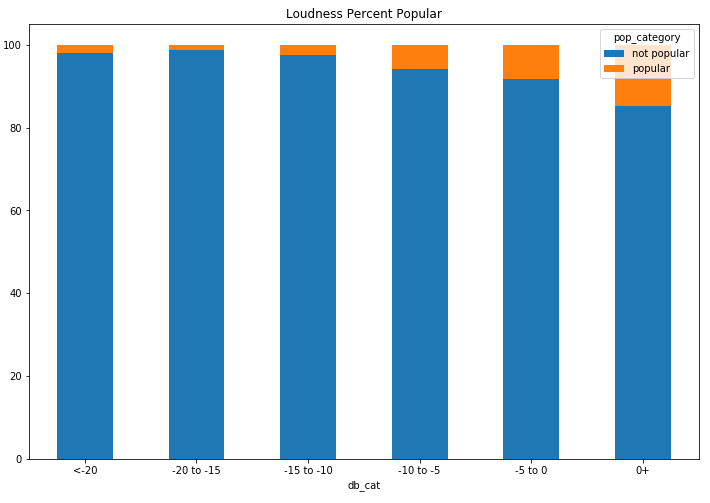
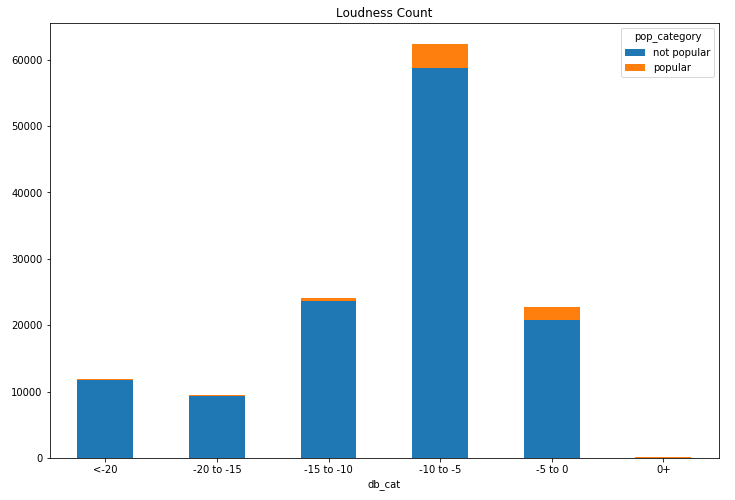


Figure – Popularity distribution for different song lengths, for all songs (left) and popular songs (right)

Looking at the mean, median and distribution of popularity values for all song lengths, the trend appears to be that medium duration songs have slightly higher popularity. Songs of medium length had slightly wider distributions that were shifted up, while those with very short or long lengths have narrower distributions that were shifted down. Statistically speaking, there is not a significant variation between each subgroup as the boxplots all overlap.

### 3.1.2 Loudness

It was expected that louder songs would be more popular, given the occurrence of the loudness war in the music industry since the advent of digital media. Songs were categorized based on their average loudness measurement on the digital dB scale (LUFS), with 0 being the loudest and quieter songs being more negative.



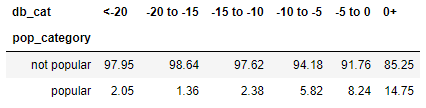
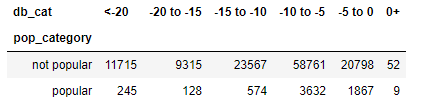


Figure – Frequency and rate of popularity for songs compared to loudness

Most songs fell between -10 and -5 dB in loudness, perhaps indicating that artists and producers prefer to write and master songs to have some dynamic range, rather than be loud constantly. 61 songs were also measured to have loudness values greater than 0, suggesting a significant amount of clipping and distortion as a result of poor mastering. After normalizing for each sub-population, it was revealed that as songs got louder, they progressively became more popular. Songs louder than -5 dB were most frequently popular, with a rate of 8.24% compared to the 1.36% exhibited by the -20 to -15 dB category. Boxplots show a similar trend, with louder songs having a wider and up-shifted distribution of popularity values. Given the tiny sample size of songs louder than 0 dB, no meaningful conclusions can be drawn for that subgroup.

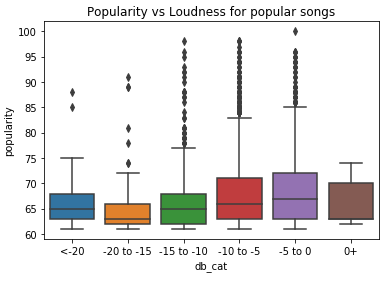
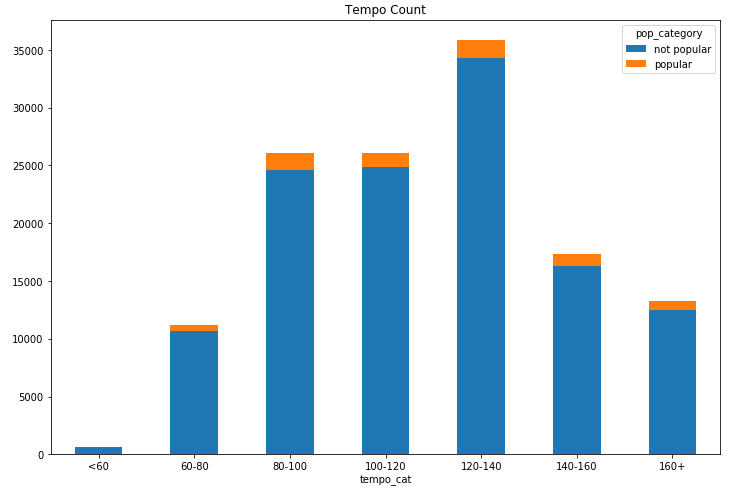
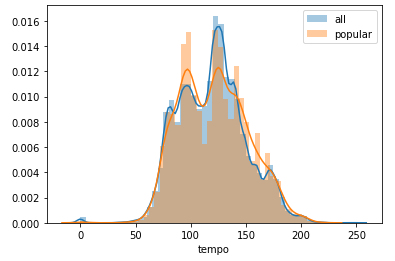
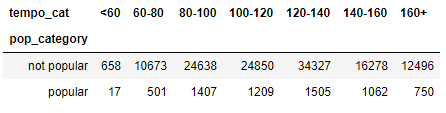
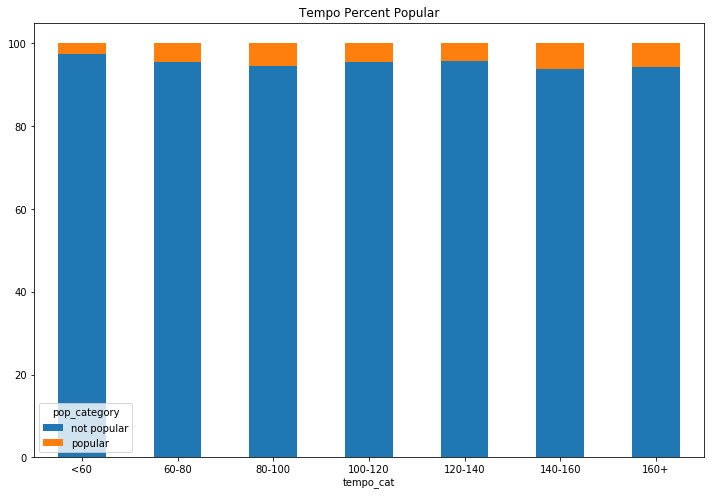


Figure – Popularity distribution for different loudness categories, for all songs (left) and popular songs (right)

### 3.1.3 Tempo

Songs were categorized based on their perceived tempo in beats per minute (bpm). A histogram plot for all songs and popular songs show two distinct peaks, located at just under 100 and 130 bpm. These likely correspond to the prevalence of medium-slow tempo ballads common to many genres, and up-tempo tracks that tend to be faster than 120. Songs slower than 60 bpm were rare.





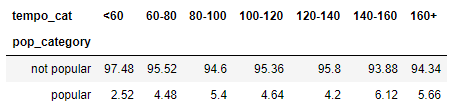
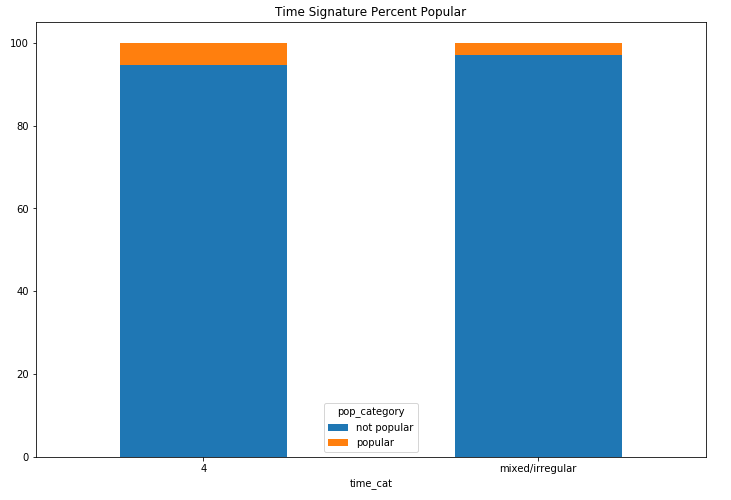
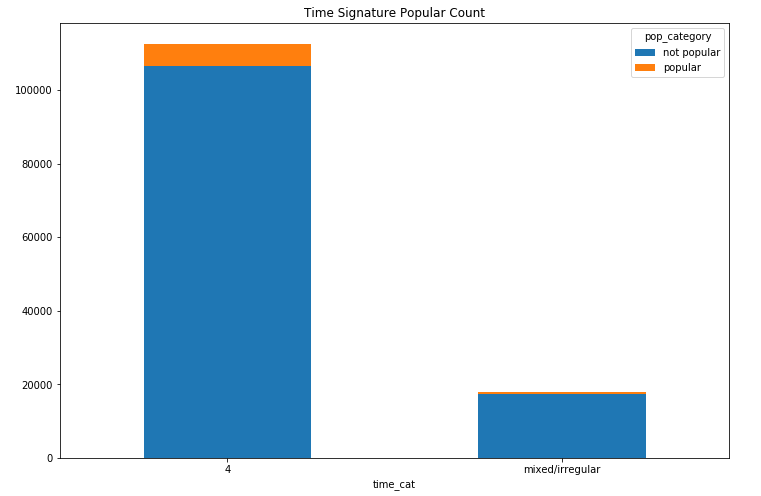


Figure – Histogram for all songs and popular songs vs. tempo (top left), frequency of songs vs tempo (top right), rate of popularity vs tempo (bottom left), tables for frequency and rate of popularity (bottom right)

Looking at absolute frequency, songs in the 120-140 bpm range were most common overall, though popular songs showed similar numbers in the 80-100 and 120-140 ranges. Normalizing for population, numbers were similar across the board with the 140-160 category having the highest popular rate of 6.12%. Surprisingly, the 120-140 range was less popular than all others except for extremely slow songs.

### 3.1.4 Time Signature

We then analyzed the influence of time signature on overall song popularity. In the dataset, songs were given a number based on their meter (4/4, 3/4 etc.). It also appeared that Spotify could not identify songs with changing or varied time signatures, as the attribute contained only one integer, and no odd time signatures except for 5/4. Nonetheless, as expected, 4/4 was by far the most common time signature.



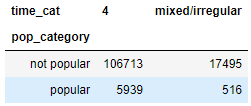
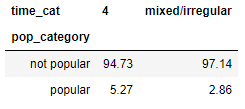
 

Figure – Frequency and rate of popularity for songs vs. time signature

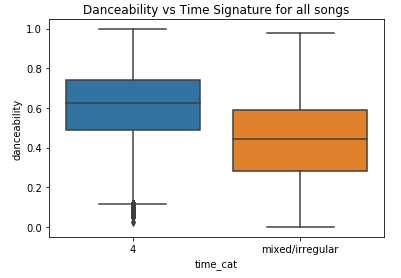
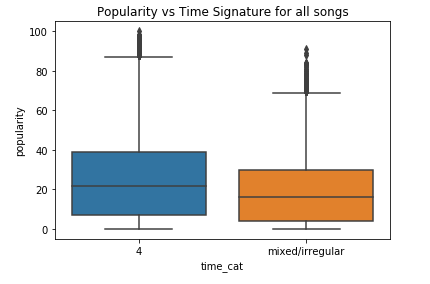


Figure – Popularity distribution vs. time signature (left), danceability vs. time signature (right)

Given the prevalence of 4/4 time with over 100,000 instances, songs were categorized as either 4/4 or non-4/4. When normalized by the size of each subgroup, 4/4 time showed a popular rate of 5.27%, almost twice as high as the non-4/4 group at 2.86%. Boxplots showed that the 4/4 group has a slightly higher median popularity with higher distribution as well. This is likely due to 4/4 being a much easier time signature to dance to, as evidenced by the median danceability of the 4/4 group being higher than the 75th percentile of the non-4/4 group.

### 3.1.5 Tonality

Next, songs were analyzed based on their tonality, a combination of their mode (major or minor) and the key they were based in (one of 12, with C = 0, C# = 1 etc.). This leads to 24 possible combinations. One flaw in this data was that each song only had one given key and mode, and thus if a song contained a key change, this information was not captured.

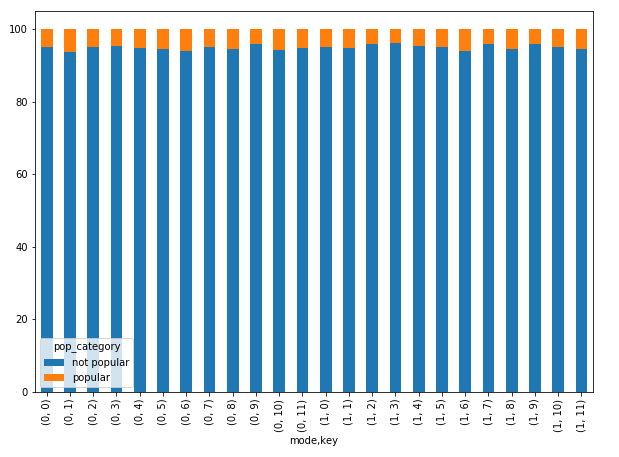
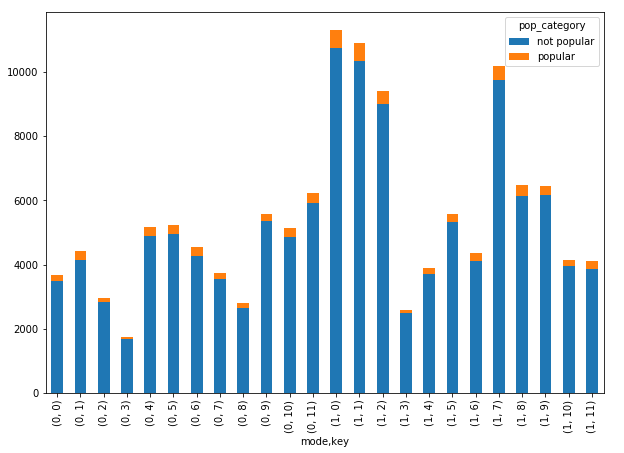
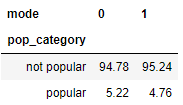
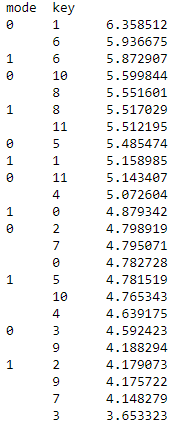
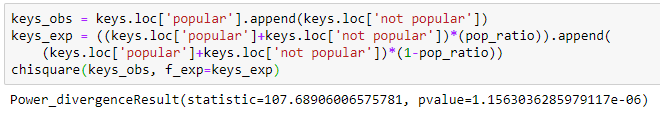
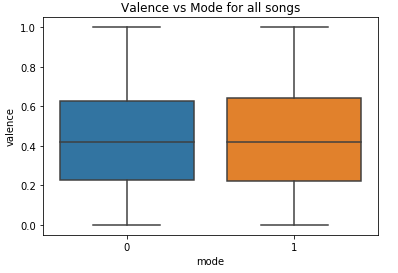


Figure – Frequency and rate of popularity plots for different keys (left), rate of popularity for all keys (top right), rate of popularity for minor vs. major (bottom right)

Bar graphs showed that major keyed songs are generally more common, with C, Db, D, and G major being especially common. C, G and D being popular can likely be attributed to them being relatively easy keys to play on piano and guitar, both of which see widespread usage in many genres of music. It is unclear why Db major would also be very common. Discounting these 4 keys, major and minor keyed songs show a similar level of prevalence. After normalizing, C# minor exhibited the highest rate of popular songs at 6.36%, while Eb major had the lowest at 3.65%. As one would expect all keys to have a similar rate of popularity, this is a puzzling result. To check the likelihood of this occurring given random distribution of popular songs, we conducted a chi-squared test:



The null hypothesis assumes that each of the 24 keys would have approximately the same proportion of popular and non-popular songs as the global rate, given by *pop\_ratio.* The expected frequency of both popular and non-popular songs occuring in each key is stored in *keys\_exp*, and these values were compared against the observed frequency (*keys\_obs*). The resulting chi-squared number was 107.69, with a p-value of 1.16 x 10-6 , indicating that the null hypothesis can be rejected and that the distribution is not random.

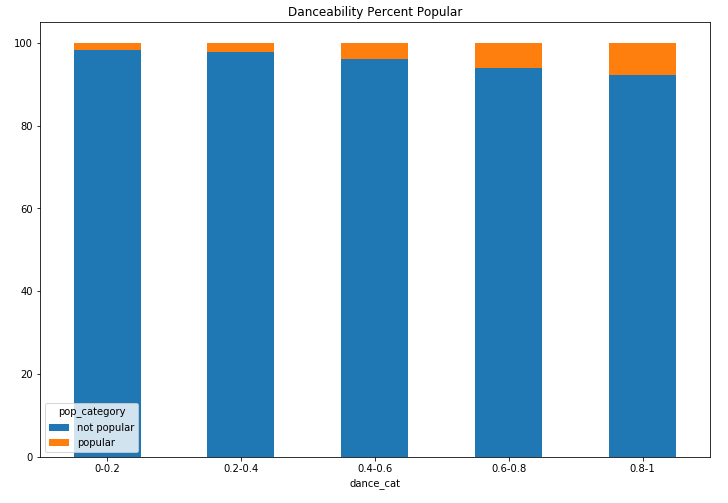
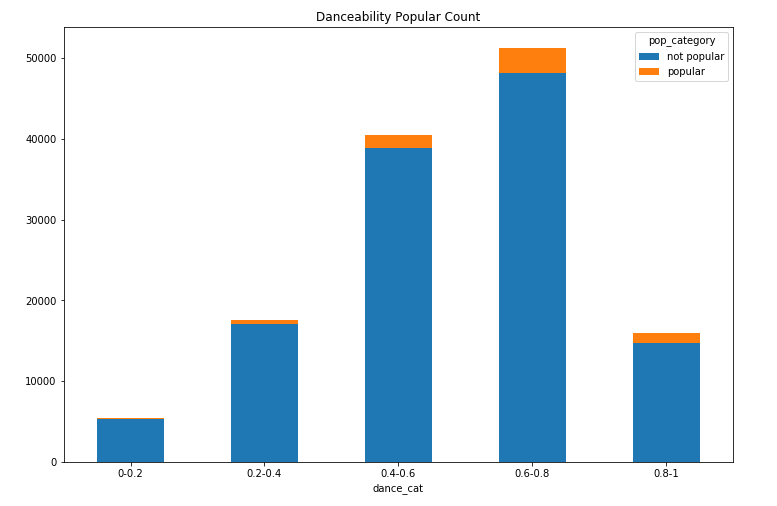


Finally, we examined the relationship between valence and mode. This boxplot shows that the valence distribution for both major and minor songs is virtually identical. As such, the cliché that major songs are happy and minor songs are sad can be disproven, as valence is influenced much more by factors such as loudness and tempo.

Figure – Valence distribution vs. mode

### 3.1.6 Danceability

Danceability describes how easy it is to dance to a song, based on elements such as tempo, rhythmic stability, beat strength, and overall regularity.



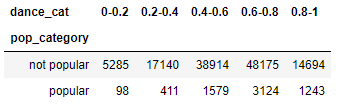
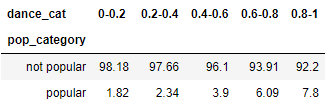
 

Figure – Frequency and rate of popularity vs. danceability

Danceability follows a slightly right-shifted normal distribution, with medium-high danceability songs being most common. Looking at popularity ratio, there is an overall increasing trend with more danceable songs having a higher rate of popular songs as well. Songs between 0.6 and 1 danceability have a much higher rate of popularity compared to those between 0 and 0.6. This is consistent with the fact that danceability is one of the attributes with the highest positive correlation with popularity. Thus, it can be said that people enjoy dancing and songs that are easy to dance to.

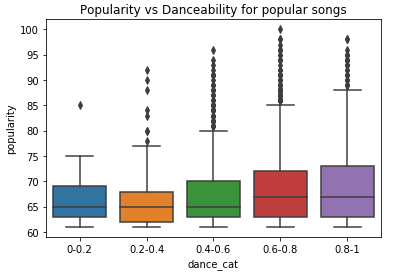
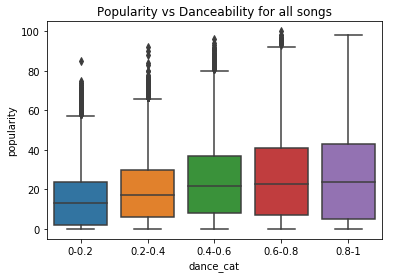
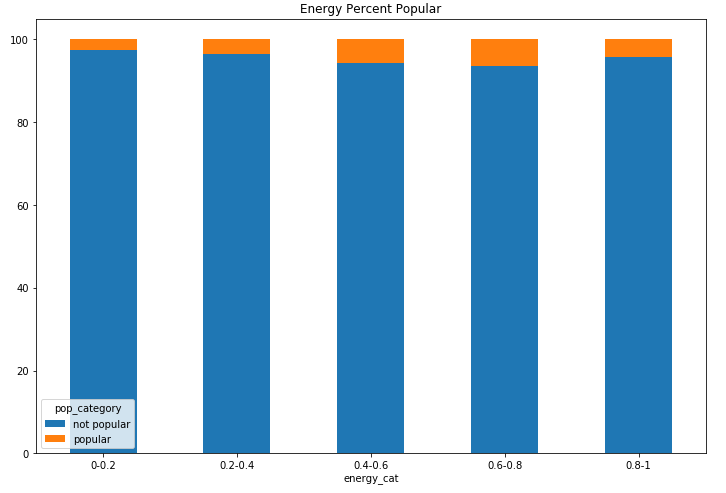
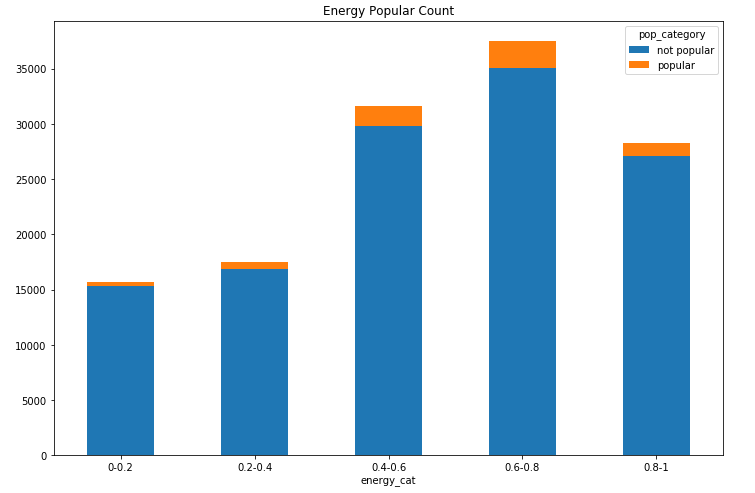


Figure – Popularity distributions for different danceability groups, for all songs (left) and popular songs (right)

Sliced another way, highly danceable songs tended to belong in more widely spread distributions, while songs with low danceability were clustered more closely in the medium-low popularity range.

### 3.1.7 Energy

Energy is a perceptual measure of intensity and activity in the song, where high energy usually indicates a faster, louder and noisy song.



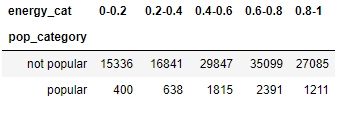
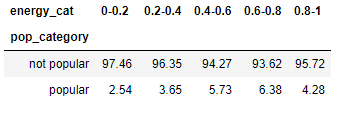
 

Figure – Frequency and rate of popularity vs. energy

These bar graphs show that higher energy songs are more common, and tend to have a higher proportion of popular tracks. However, there is a drop in rate of popularity, going from energy levels of 0.6-0.8 to 0.8-1. This is also reflected in the boxplots, where the 0.8-1 boxplot has a shorter whisker.

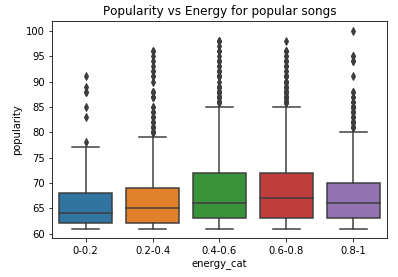
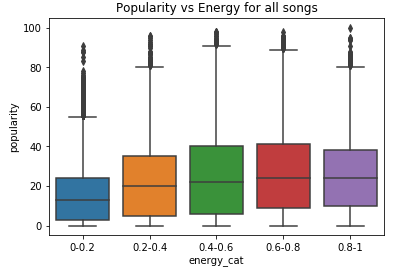
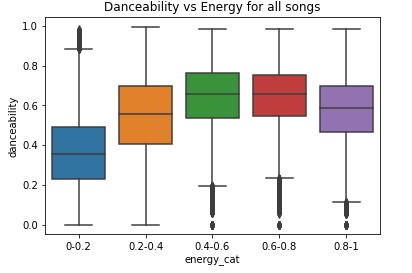


Figure – Popularity distributions for different energy groups, for all songs (left) and popular songs (right)

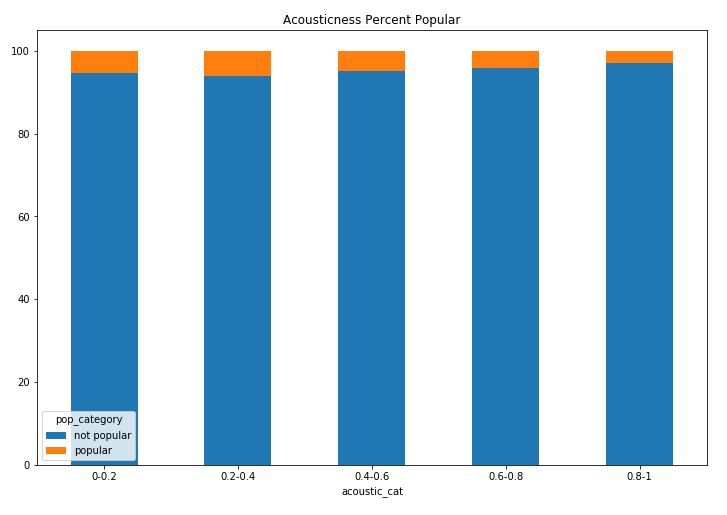
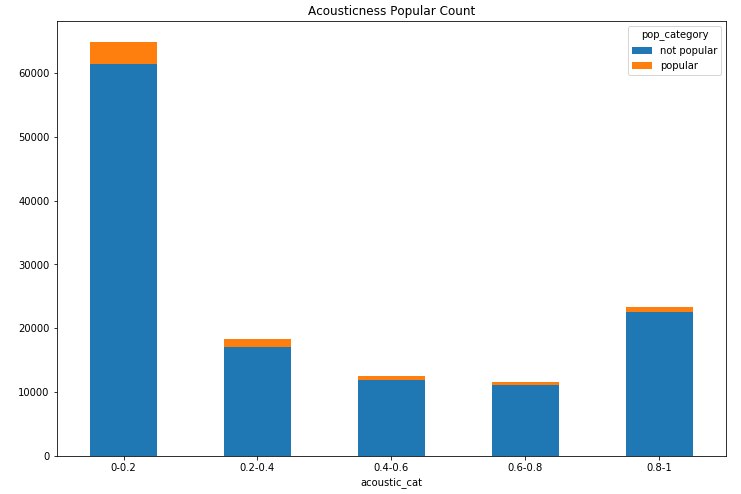
One possible explanation for this is that extremely high energy songs experience a drop in danceability, due to extremely fast tempos and having a less stable rhythm. The drop in popularity is also an indicator that people tend to enjoy songs with varying dynamics and mood (e.g. build-up and drop), rather than songs that are constantly hard-hitting.

### 

Figure – Danceability distributions for different energy groups

### 3.1.8 Acousticness

Acousticness is a confidence measure of whether the track is entirely acoustic.



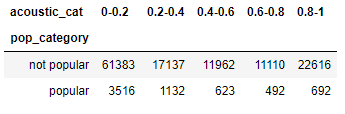
 

Figure – Frequency and rate of popularity vs. acousticness

As expected, most songs have low acousticness, as popular genres such as pop, rock, hip-hop and EDM all contain electronic elements. The slight increase in highly acoustic songs can be attributed to genres that traditionally don’t contain any electronic instruments, such as jazz and classical. Looking at the normalized graph, songs with 0.2-0.4 acousticness have the highest rate of popularity at 6.2%, while purely acoustic music has a rate of less than 3%. This can be explained through the correlation between acousticness, danceability, and loudness.

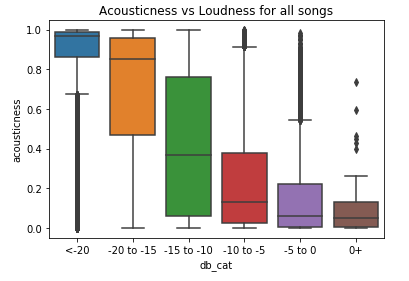
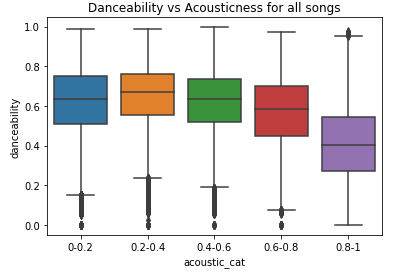
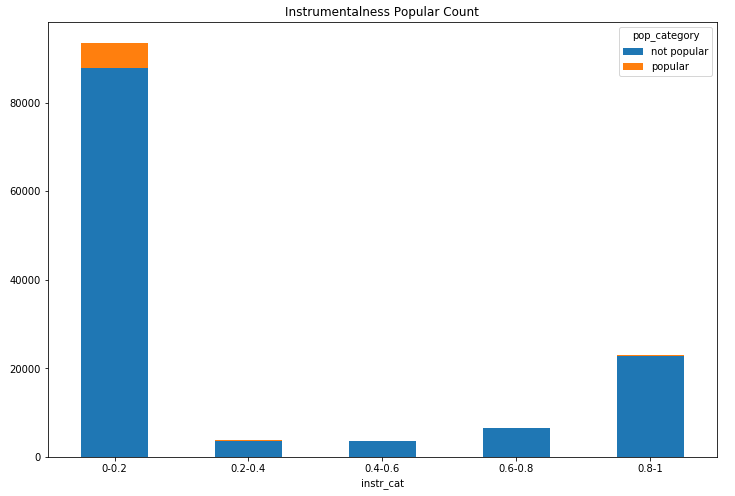
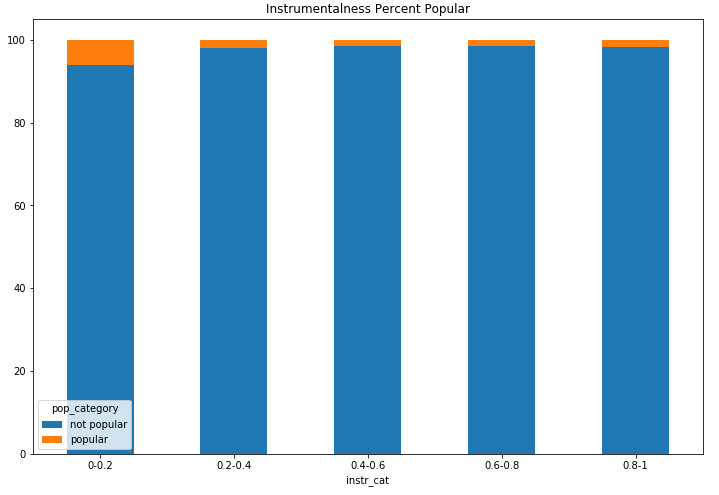


Figure – Danceability distribution vs. acousticness (left), acousticness distribution vs loudness (right)

These boxplots demonstrate that songs are typically more danceable when then have low to medium acousticness, and that louder music is generally low in acousticness. Since people enjoy loud and danceable music, they will naturally tend to prefer songs that are low in acousticness.

### 3.1.9 Instrumentalness

Instrumentalness is a predictive measure for the likelihood of a song containing no vocals (not including “oohs” and “aahs”), and not necessarily the amount of vocals a track has.

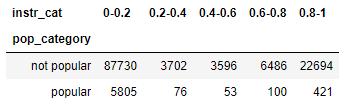
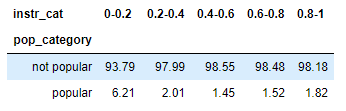
 

Figure – Frequency and rate of popularity vs. instrumentalness

The data shows that zero instrumentalness is very common, meaning that most songs do contain vocals. High instrumentalness is also relatively common compared to those with a medium amount, likely due to the prevalence of purely instrumental tracks in genres like classical and jazz. Once each subgroup has been normalized by population, it is apparent that tracks containing vocals are significantly more popular. This is reflected by the 0-0.2 subcategory being more than 3x more likely to contain a popular track than all other groups.

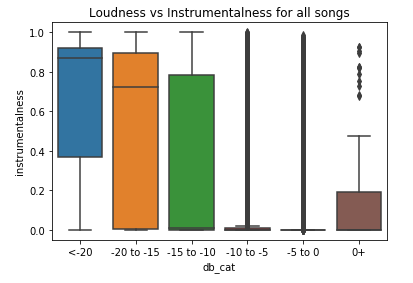
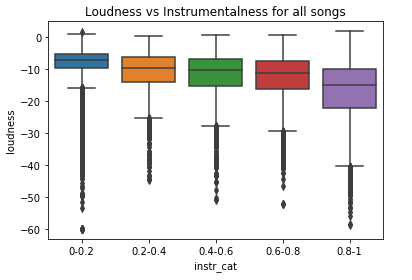
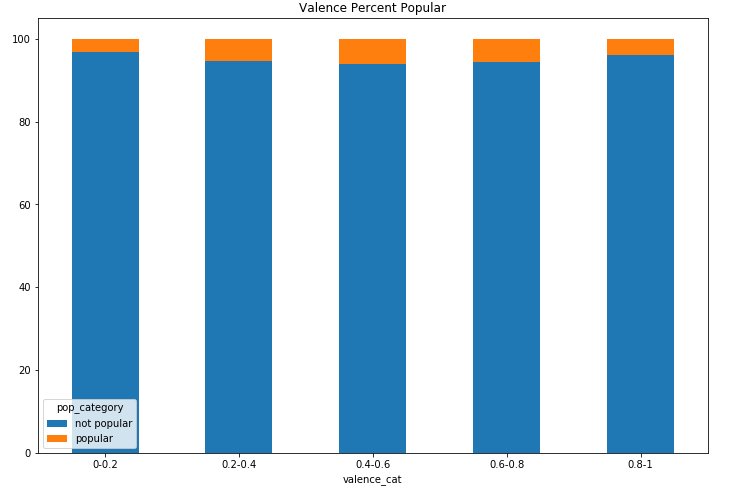
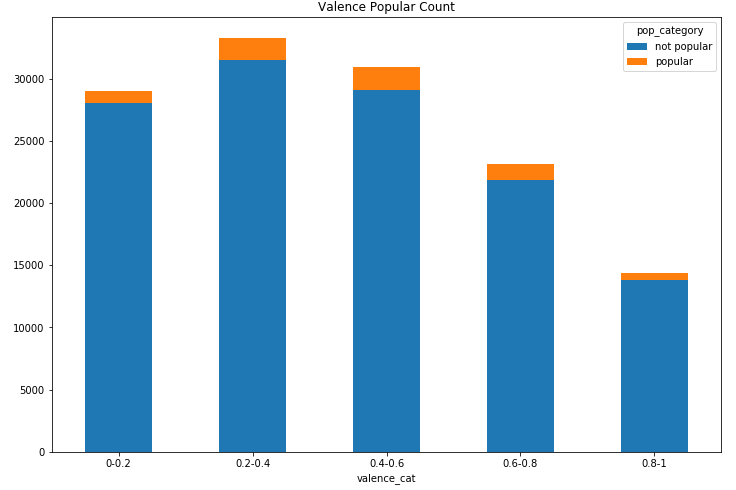


Figure – Loudness distribution vs. instrumentalness (left), instrumentalness distribution vs. loudness (right). High loudness songs demonstrate all 4 quartiles of data compressed to near-0 instrumentalness

This observation can be explained by examining these boxplots comparing instrumentalness and loudness, which show that tracks containing vocals tend to be produced much louder than instrumental tracks. Notably, songs between -10 and 0 dB in loudness almost always exhibit 0 instrumentalness as shown in their highly compressed boxplots. It is unclear which of these is the driving variable, but the result is the same—people enjoy loud music and thus also prefer music that contains vocals.

### 3.1.10 Valence

Finally, we examined valence, which is a quantitative description for the musical positiveness of a song. Tracks with high valence sound more positive (e.g. happy, cheerful, and euphoric), while low valence tracks sound more negative (e.g. sad, depressed, and angry).



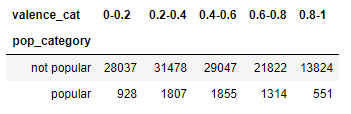
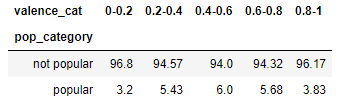
 

Figure – Frequency and rate of popularity vs. valence

The overall left-shifted distribution of songs shows that music tends to be more negative in the emotions they convey. After normalizing for population in each category, it was shown that the highest rate of popularity was associated with the 0.4-0.6 valence group, with 6%. In fact, the middle three bands were similar, with rates between 5.4% and 6%, while the edge groups both had rates under 4%. This indicates that people generally prefer ambivalent-feeling songs that are not overtly happy or sad.

## 3.2 Common Features for Popular Songs

While no single attribute was a strong contributor to song popularity, our analysis showed several trends that popular songs typically exhibited. Categories that contained the highest rates of popular songs were more frequently characterized by:

* Higher loudness and danceability, which usually meant:
  + they were faster than 60 bpm; most frequently between 80-100, or 140+
  + they were in 4/4 time
  + they had high but not excessive levels of energy
  + they contained some acoustic elements but were not entirely acoustic
* Presence of vocals
* Medium length (3-4 minutes)
* Not being overtly happy or sad

## 3.3 Predicting Song Popularity

We have leveraged machine learning to determine how to predict popularity based on song attributes. The first thing we did was note all the attributes that all popular songs had. In relation to our model we chose danceability, instrumentalness and loudness. From there we decided to label the data into two groups, 0 – 50 being “Non-Popular” and 51 – 100 being “Popular”. Next we had to define 80% of the dataframe for training and use the other 20% for training. After running the data we determined that we would not be able to accurately predict popularity based on the model we created. The AUC was only 50%, which tells us the model is making random guesses.

In order to make the model more accurate, we would need the attributes most common in popularity to have a stronger correlation. Alternatively, we could also use another model that wasn’t dependent on direct linear correlation, as the correlation coefficients observed between our attributes and song popularity were generally low.

# 4. Conclusions

In analyzing our data, we determined that while individual attributes were poor indicators of song popularity, popular songs do tend to exhibit certain characteristics more than others. Across the entire population, popular songs were usually louder and more danceable, which also meant that they were more often than not in 4/4 time, had tempos between 80-100 or 140+ bpm, had high but not excessive energy, and were sparse in acoustic elements. Popular songs also typically featured vocals, had a length between 3 and 4 minutes, and more frequently displayed an ambivalent mood as opposed to one that was overwhelmingly positive or negative. These findings were consistent with the initial comparison between all attributes to find their correlation with song popularity.

However, it proved difficult to use these findings to predict the popularity for any individual song with machine learning. Many attributes were not linearly correlated with popularity, and most of our boxplot analyses did not show statistically significant differences between sub-categories of song attributes.

Future work would also include further analysis into how these attributes influence each other, and not just popularity.