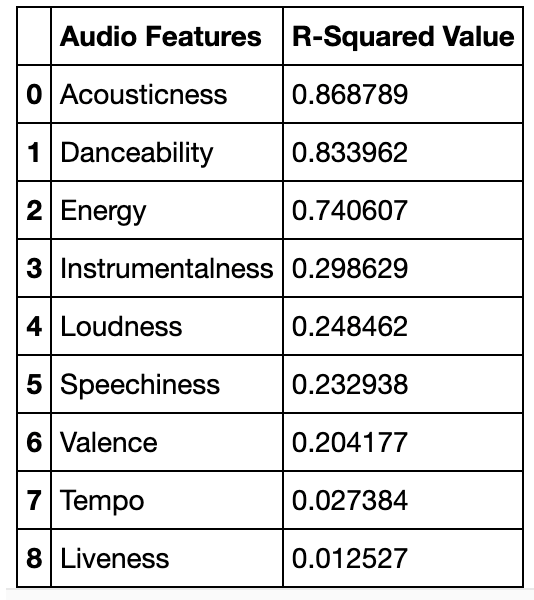
# SPOTIFY DATA ANALYSIS

Breckin Shoemaker, Julie Baker, Lydia Deterding

## WHAT CATEGORIES BEST PREDICT POPULARITY ACCORDING TO R2 VALUES?



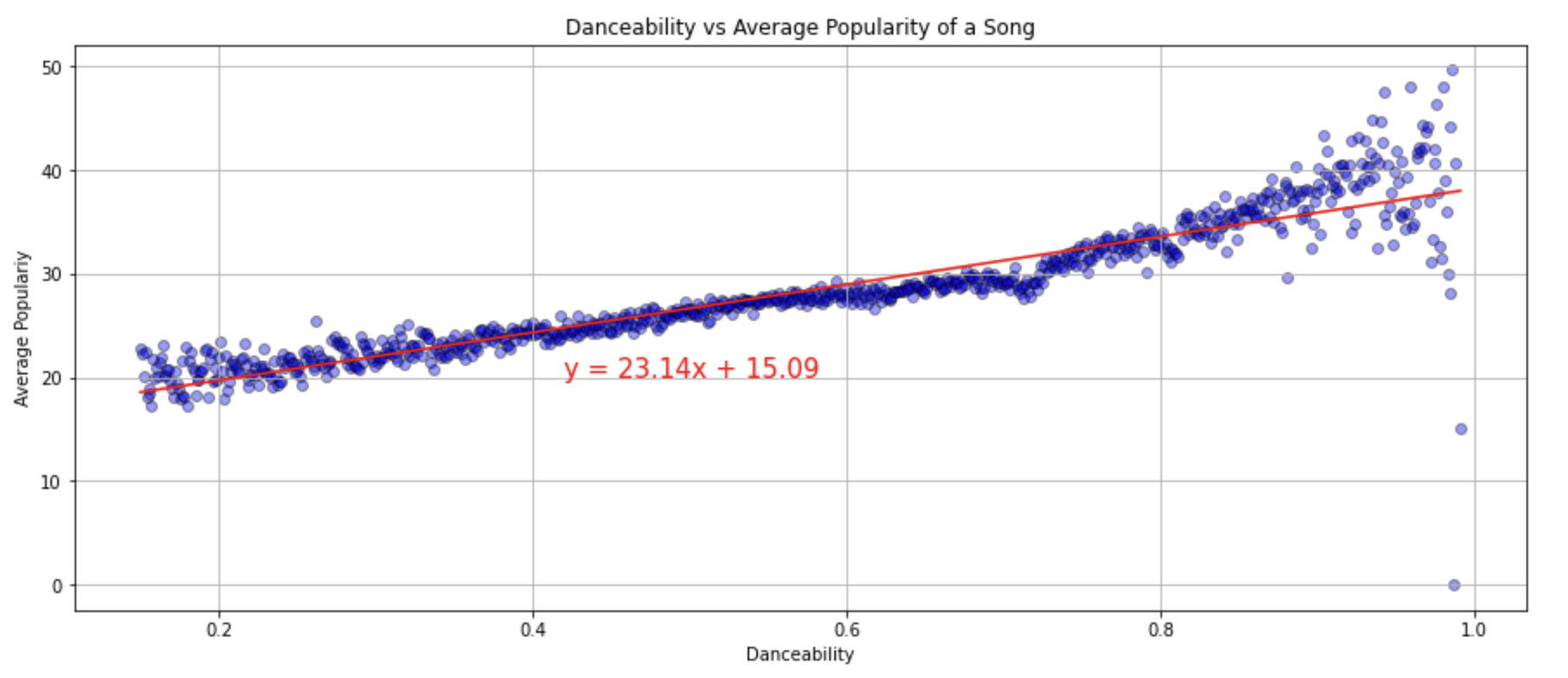
In order to answer this question, we calculated R2 values and compared them among the different audio features that we analyzed. Acousticness, danceability, and energy all have high correlations to popularity. Instrumentalness, loudness, and speechiness are weakly correlated to popularity. Valence, tempo, and liveness are all not correlated to popularity so we are not going to discuss these audio features in this writeup.

IS ACOUSTICNESS RELATED TO POPULARITY?

## 

## As you can see from the plot, as the acousticness increases from 0.1 to 1.0, the popularity steadily decreases and then there is a large decline when the song reaches 1.0 (fully acoustic). This trend is most closely correlated to popularity.

IS DANCEABILITY RELATED TO POPULARITY?



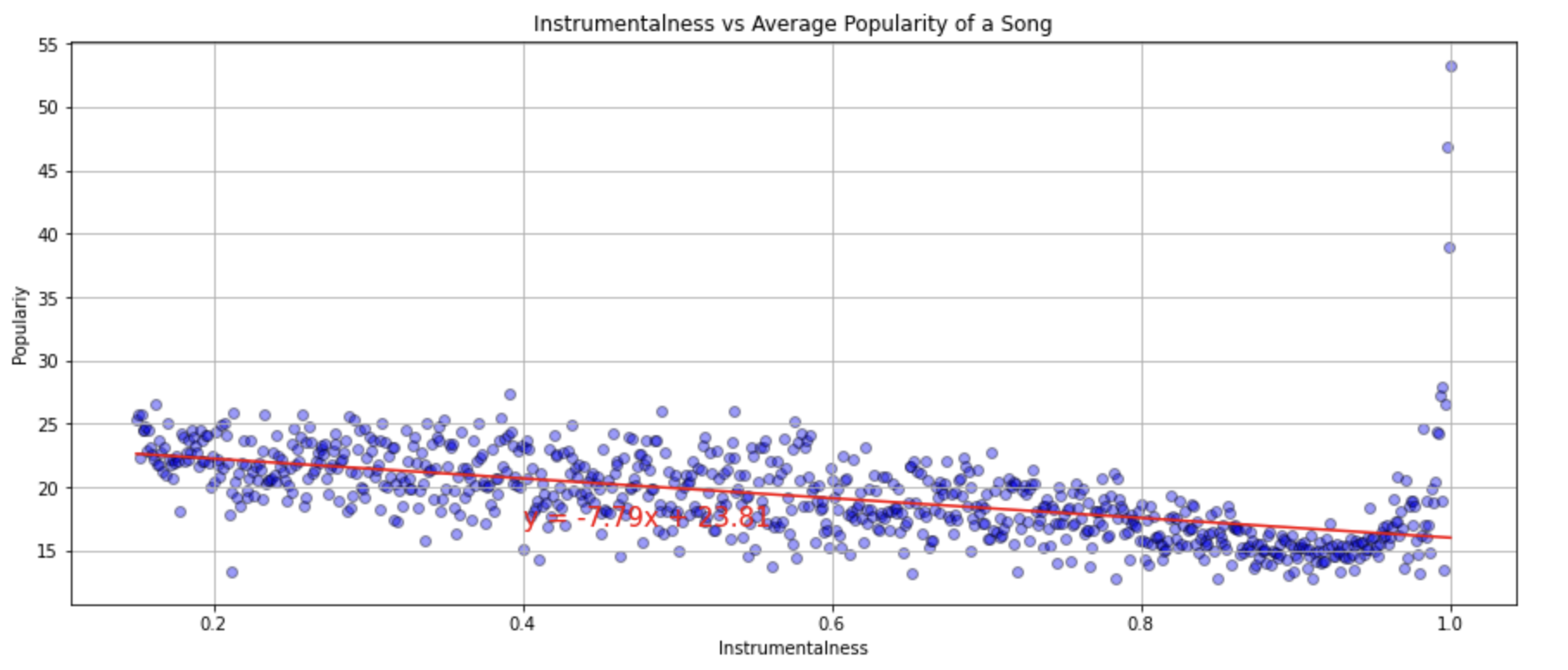
Danceability has a substantial correlation with popularity. Danceability uses different factors of the song to determine how suitable it is for dancing. More danceable songs have values close to 1. As the plot shows, popularity increases as danceability increases.

IS ENERGY RELATED TO POPULARITY?

## 

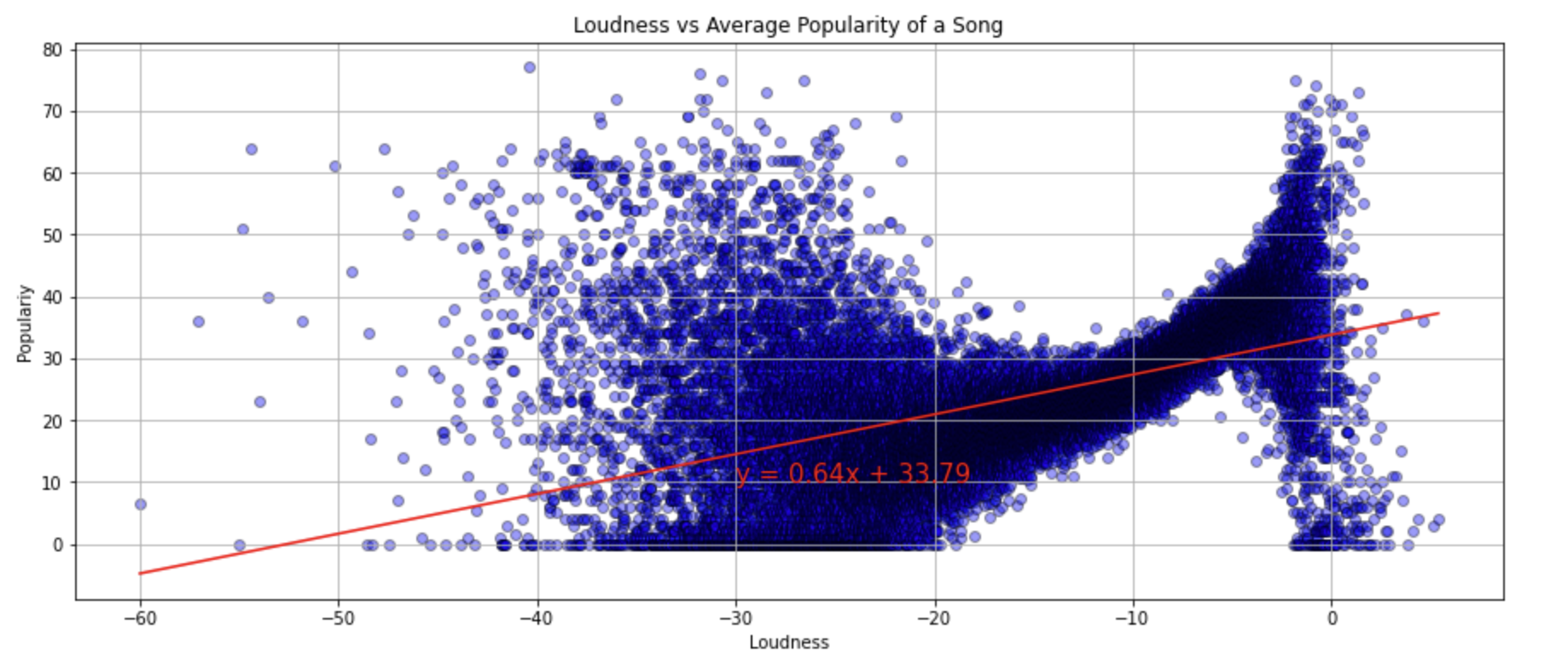
Energetic songs are generally more popular than the less-energetic songs. Popularity peaks at an energy rating of about 0.8, and at energies higher than that there is a small decline in popularity.

## IS INSTRUMENTALNESS RELATED TO POPULARITY?



Songs that have more vocals in them are more popular than songs that are instrumental. Instrumentalness does have a weak correlation with popularity. As the plot shows, the popularity steadily decreases by a small amount as the instrumentalness increases.

## IS LOUDNESS RELATED TO POPULARITY?



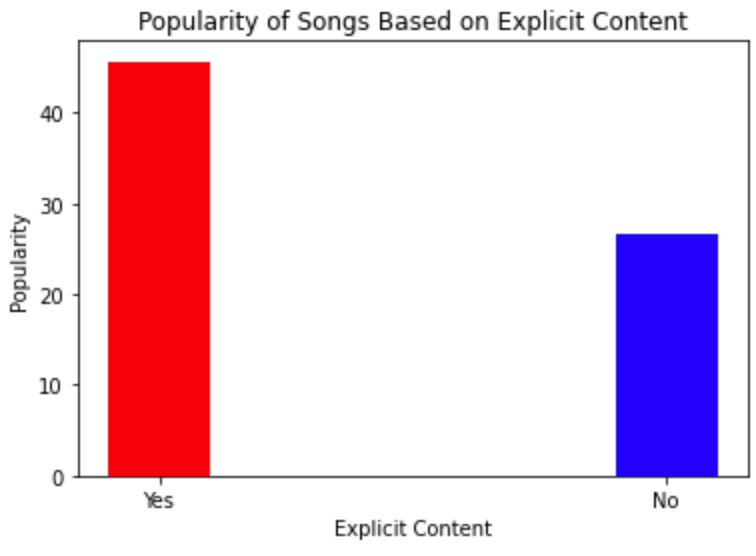
Songs that have more vocals in them are more popular than songs that are instrumental. Instrumentalness does have a weak correlation with popularity. As the plot shows, the popularity steadily decreases by a small amount as the instrumentalness increases.

IS SPEECHINESS RELATED TO POPULARITY?

## 

Speechiness has a low correlation with popularity. The higher the value of speechiness, the closer the song is to something like an audiobook. The majority of songs are low on the speechiness scale, and the songs that are almost all speech are low in popularity.

## ARE EXPLICIT SONGS MORE OR LESS POPULAR THAN NON-EXPLICIT SONGS?



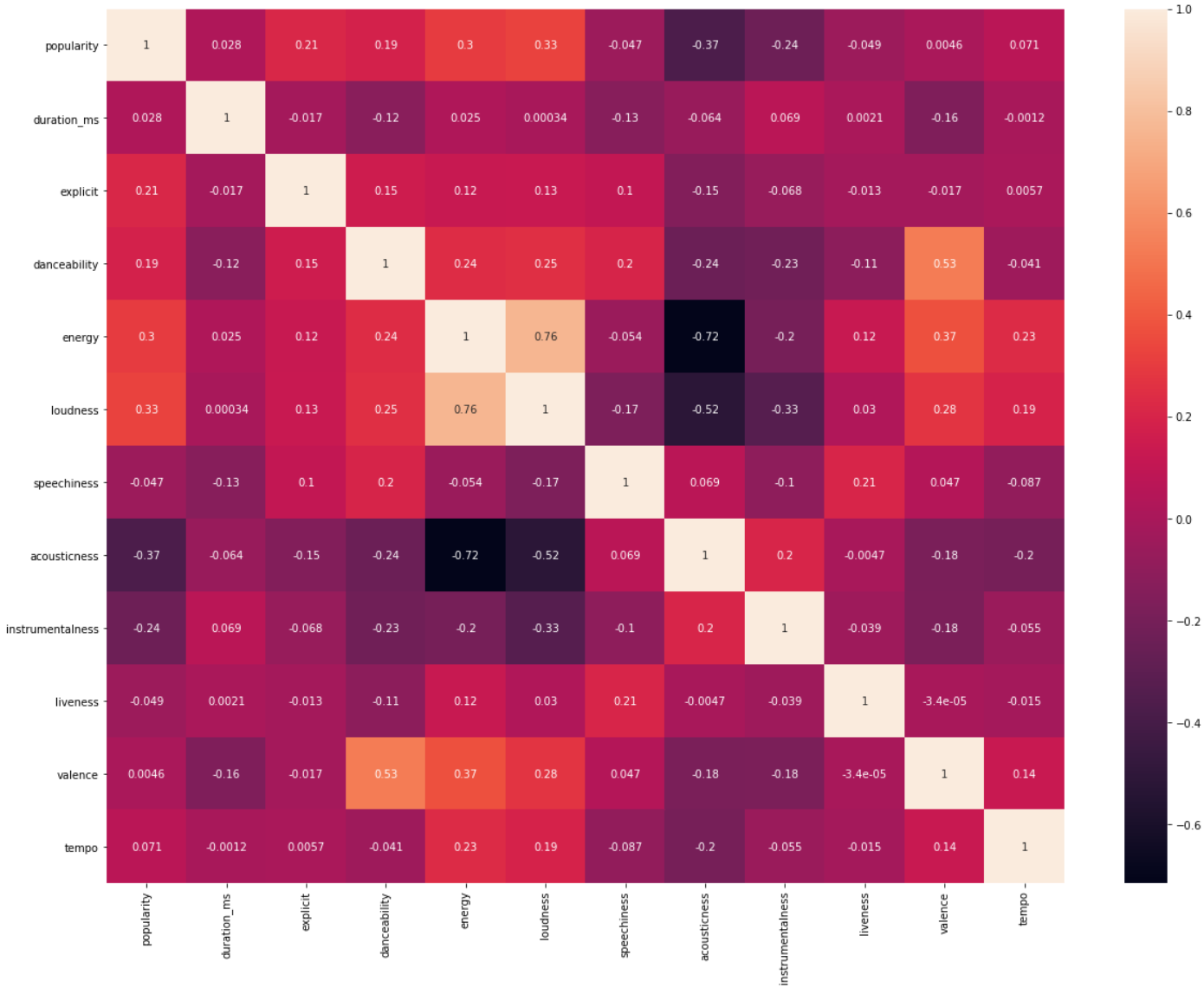
The bar chart shows that explicit songs are almost twice as popular, on average, than non-explicit songs. This is a general trend over the dataset of total songs, of which most were non-explicit. The explicit version of a song was not compared to its non-explicit counterpart.

IS GENRE RELATED TO POPULARITY?

## 

The most popular genres are pop and rap. R&B comes in third, with metal and emo close behind. Hiphop is the least popular genre.

NOTABLE TRENDS AMONG AUDIO FEATURES



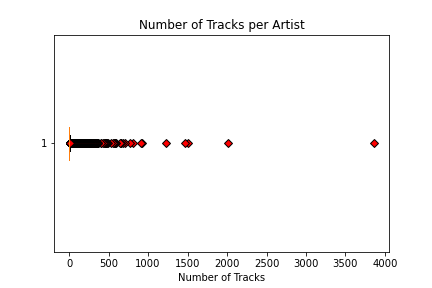
We created a correlation matrix and a heatmap to study the correlation between different audio features. This revealed that loudness and energy in fact are highly correlated. Acousticness is also highly correlated to energy. The less acoustic a song is, the more energy it has. Acousticness is also correlated to loudness, which makes sense. The more acoustic a song is, the quieter it will be. Valence and danceability are correlated, showing that happy songs are more danceable than less happy tunes. Valence also has a low correlation to energy, suggesting that happier songs have more energy.

Number of Tracks by an Artist and Popularity Scores

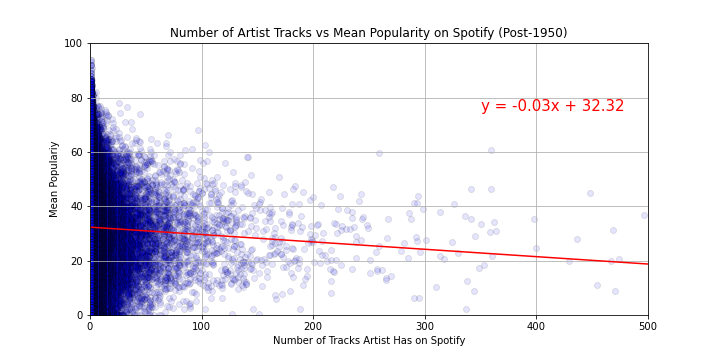
For this section, the data frame was reduced by removing duplicate tracks and only looking at tracks that were released starting in 1950.

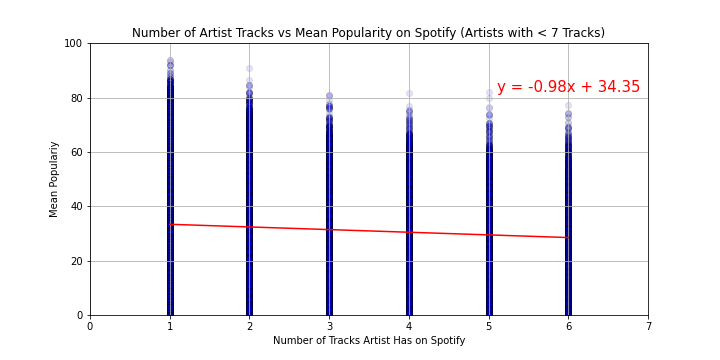
Tracks were marked as duplicates if the title and artists were listed as the same. Because duplicate tracks differed in popularity scores, the duplicate with the highest popularity of its set is what was returned to the dataset for analysis.

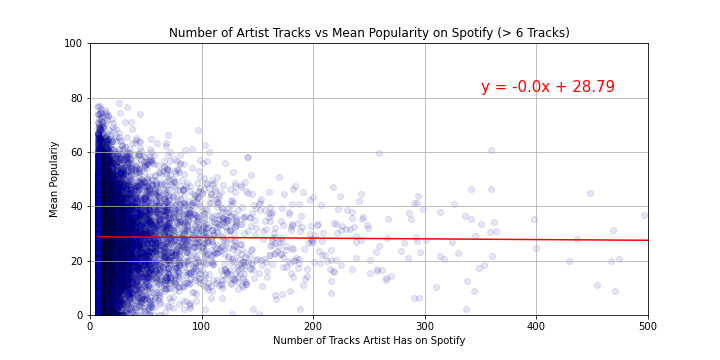
There was a wide range for the number of tracks an artist had on Spotify. Over half of the artists only had 1 track on spotify. The highest number of tracks for an artist is 3856, and any artist with more than 6 tracks is considered to be an outlier in the dataset.



Mean Popularity Based on the Number of Tracks an Artist Has

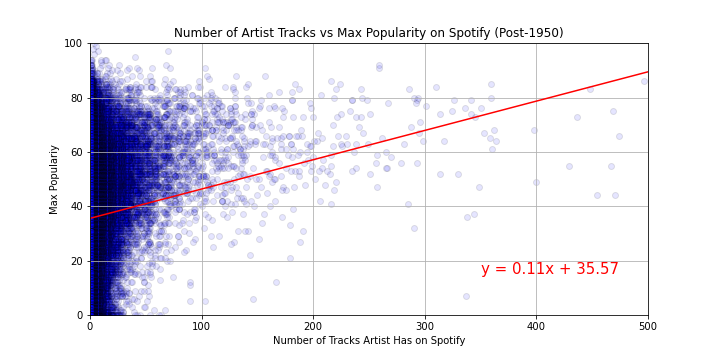


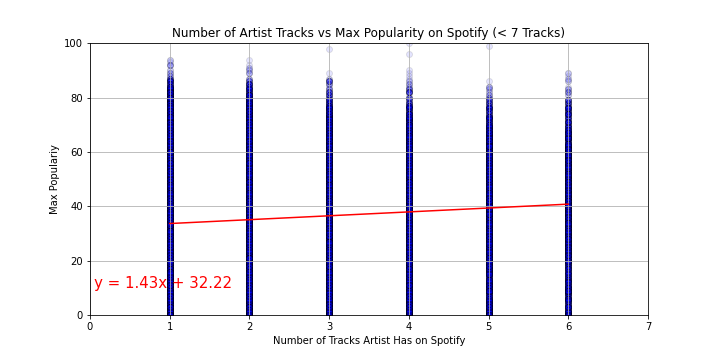


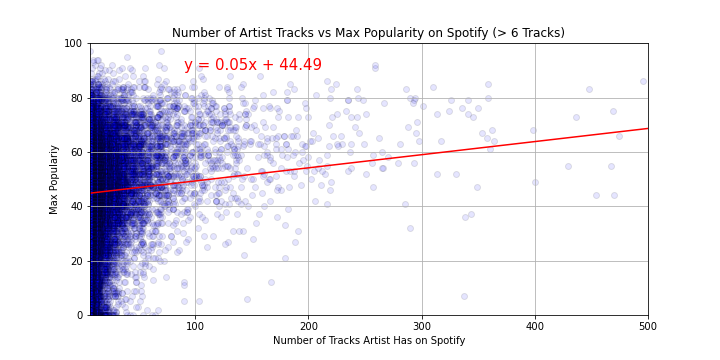


There is a slight negative trend in the non-outlier data (i.e., artists with one to six tracks) such that the more tracks an artist has, the lower the mean popularity is. When looking at artists with seven or more tracks, the trendline is flat, suggesting no relationship between the number of tracks and popularity in artists with many songs on Spotify.

Maximum Popularity of an Artist’s Tracks by Number of Release Years







With regard to maximum popularity, the data for all three groupings tended toward a positive relationship between the number of tracks an artist had and their maximum popularity score.