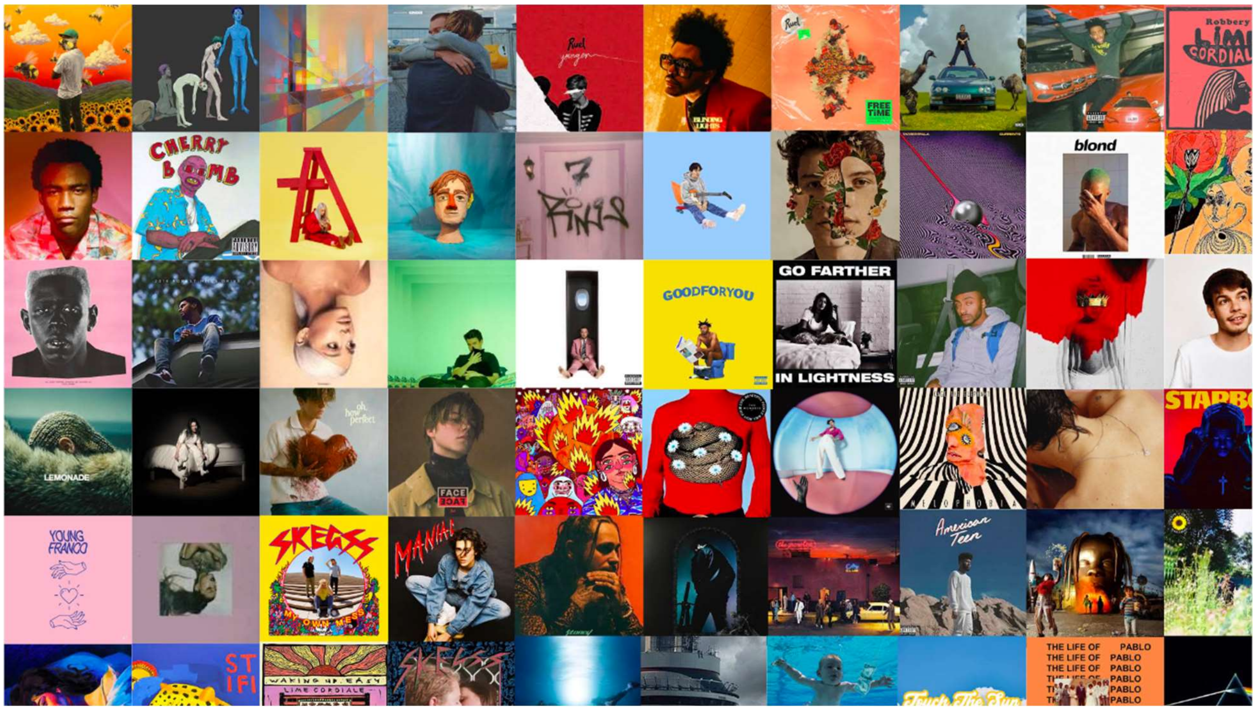
What Makes Modern Music Popular?

An Analysis of Song Characteristics and their Effect on Popularity



STAT 3220-001
Group 13
Joseph Park, Maya Qureshi, John Vithoulkas
Due: October 18, 2020

Introduction

As Spotify and other streaming services have become more popular it has become easier to analyze the popularity of music based solely on the number of streams, rather than analyzing sales data from vinyl, CD, and iTunes sales. Even listeners are easily able to view the popularity of their favorite songs, as Spotify provides rapidly updated information on the number of monthly listeners artists have, a scale of popularity shown with each song, and an artist's top hits with the number of streams each song has. As musicians make less money from each stream than they would have from sales, it is increasingly important for artists to compete wisely in the music business and analyze what makes modern music popular. This leads us to our general question: **what makes a song popular?**

Spotify tracks many elements of the songs on their platform including dancabilty. One of the questions we will focus on is - **do popular songs generally have a higher danceability?** Graphics made and data compiled by Columbia Business School[1] indicates that while top hits tend to be different from each other, there are some characteristics that recent hits seem to share - one of which is danceability. Additionally, we will look at how explicit language in a song affects its popularity - **do explicit lyrics make songs more attractive to modern audiences or does the stigma around foul language detract listenership from such songs?** Explicit songs without editing can be streamed more on Spotify and modern streaming services than they could be on the radio due to censorship of bad language. Through spotify, we can more easily see how the popularity of a song is affected by bad language, as it is not edited out. **Finally, does a high energy level make a song more popular?** When we think of current popular music, rap and pop genres are generally the most prevalent. These genres generally include high energy tracks, which could be a factor in their popularity. Overall, these factors seem especially important in measuring the popularity of a modern popular song and are of increasing importance due to the movement from radio and CDs to modern streaming services.

The analysis will start with a hypothesized model relating all of the quantitative variables together and performing tests to pick out significant predictors. Then, in stage 2, the qualitative predictors will be analyzed and added to the model we got from stage 1. In stage 3, we will add the interactions between the quantitative and qualitative variables and will keep the ones found significant through an individual t-test. These 3 stages of model building will be used for the analysis. Furthermore, an analysis on the final model will be analyzed through tests to find cases of violation of regression assumptions, multicollinearity, outliers and influential points, etc. The alpha value that is used throughout this project will be 0.05.

Data Summary

Data was collected using the Spotify Web API any Spotipy, a python module for Spotify web servers. As a self taught data scientist, Yamac Eren Ay[2] collected and organized the data into numeric (non-integer values), categorical (text), and dummy variables (0 or 1) categories. Over 160,000 songs were analyzed, with release dates ranging from the 1920s to 2020.

While each individual data point is consistent with the set, a few text based variables are unrecognizable. Despite it being difficult to read the title and/or artist information, these data points will be kept, as no actual analysis is being done using text based quantifiers. In addition,

[1] Morris, Colin. n.d. "What Makes a Hit." Columbia Business School. https://www8.gsb.columbia.edu/articles/projects/what-makes-a-hit/.

[2] Yamac, Eren Ay. n.d. "Spotify Data 1921-2020." Kaggle. https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks.

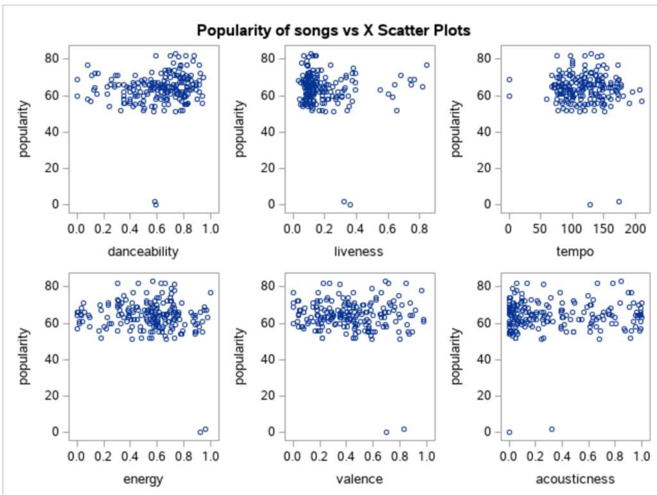
all data for songs released before 2015 were omitted, as our research question focuses on what makes a song popular in the present day, not in the mid 1900s. A simple random survey of 200 points was taken using the proc surveyselect function in SAS.

Data Dictionary

Field Name	Data Type	Units?	Description	Example
Danceability	Quantitative (0 to 1)	No	Suitability of a track for dancing based on a combination of musical elements	.25
Liveness	Quantitative (0 to 1)	No	Probability that the track was performed live	.5
Tempo	Quantitative	Yes	Estimated tempo of a track in beats per minute	110 bpm
Energy	Quantitative (0 to 1)	No	Represents a perceptual measure of intensity and activity	.35
Valence	Quantitative (0 to 1)	No	The musical positiveness conveyed by a track	.75
Mode	Qualitative (1 or 0)	No	The modality (major or minor) of a track	1
Acousticness	Quantitative(0 to 1)	No	A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.	.35
Explicit	Qualitative (1 or 0)	No	Whether or not track is ‘clean’ or explicit - explicit songs are a 1	1
Duration	Qualitative (1 or 0)	No	The length of the song broken up into 3-5 minutes or any other duration (1 represents songs that are below three minutes and above 5, while 0 represents songs 3-5 minutes long)	1

We chose the above quantitative variables based on reasoning and prior knowledge, as well as research from popular music analyses like the Columbia study referenced in the introduction. Many of the variables within the data included genre, artist, song title, etc. that would not be easily analyzed in determining the popularity of a song, even though variables such as artist clearly affect popularity. Additionally, the qualitative variables explicitness and mode were already grouped as qualitative variables in the study, but we created duration based on research about how movement to streaming services changed the duration of songs. Previously, popular songs found on the radio were generally between three and five minutes due to time limitations and audience engagement. With the movement to streaming services artists now have more freedom with the duration of their songs so we split duration into a qualitative variable representing typical radio length or not.

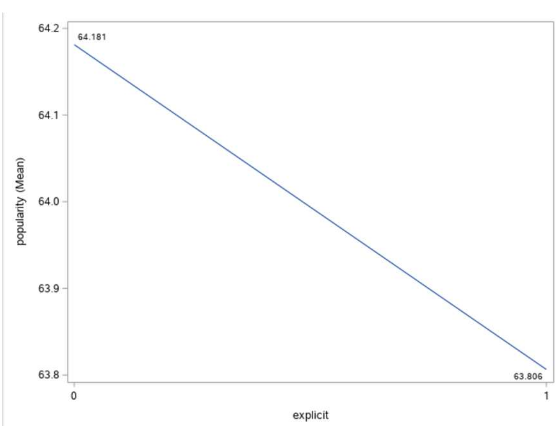
EDA



SAS Code:

```
proc sgscatter data=FILE NAME;
    title "Popularity of songs vs X Scatter Plots";
    plot Popularity*(Danceability Liveness Tempo Energy Valence Acousticness);
run;
```

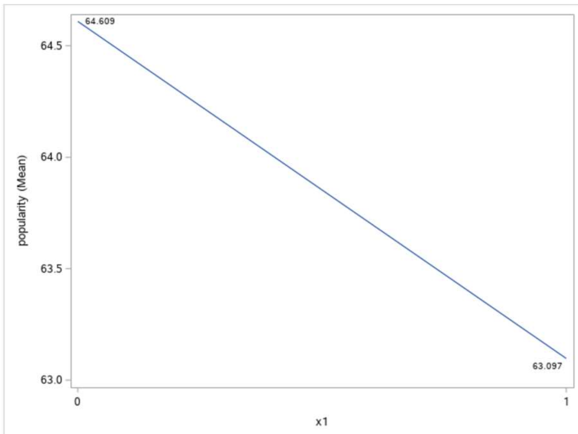
Description: The image above plots the popularity of songs against variables. As seen by the popularity and liveness plot, the majority of songs from 2015 to 2020 have low liveness levels, meaning that most tracks have a low probability of being performed live. In addition, the tempo of songs is generally somewhat high and most tracks have low acousticness, although there is a large range of acousticness levels in the sample. Finally, our sample covers tracks with a variety of different levels of energy, valence, and danceability.

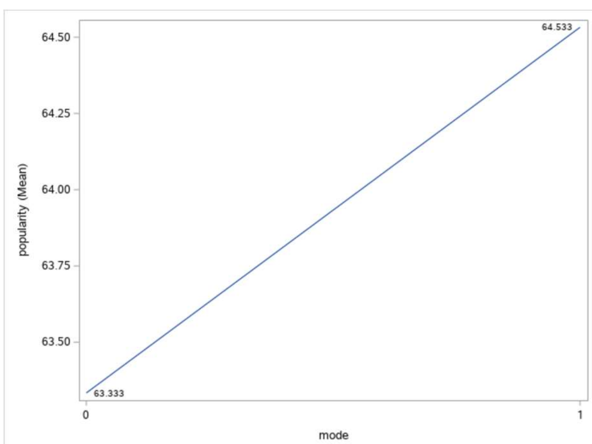


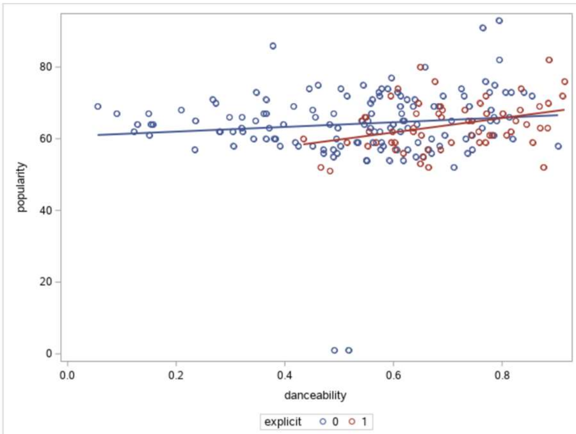
SAS Code:

```
proc sgplot data=samplesrs2;
        vline explicit / response=popularity datalabel stat=mean;
Run;
```

Description: The V-line between explicitness and popularity indicates a small variation in popularity between explicit and non-explicit songs, with explicit songs being approximately 0.4 percent less popular on average.

	**SAS Code:**
	proc sgplot data=samplesrs2;
	vline x1 / response=popularity datalabel stat=mean;
	Run;
	
	Description: The V-line between the dummy variable for duration and popularity demonstrates a small variation in popularity based on the duration of the song. The average popularity of a song with a duration between three and five minutes is approximately 1.5 percent more popular than those with other durations.

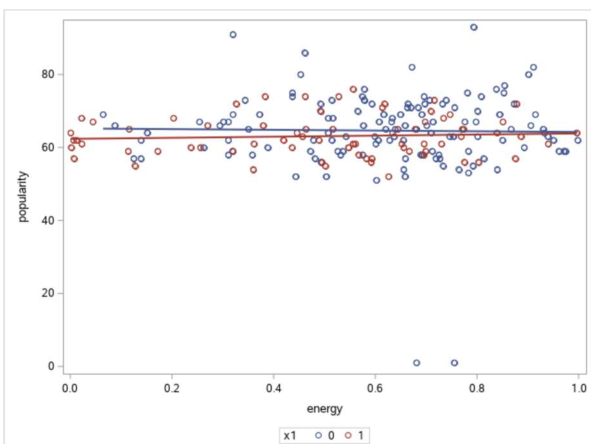
	**SAS Code:**
	proc sgplot data=samplesrs2;
	vline mode / response=popularity datalabel stat=mean;
	Run;
	
	Description: The V-line between mode and popularity demonstrates a small variation in popularity based on the modality of a song. The average popularity of a song starting with a major chord progression is approximately 1.2 percent more than other songs.



SAS Code:

```
proc sgplot data=samplesrs2;
scatter x=danceability y=popularity/
group=explicit;
reg x=danceability y=popularity /
group=explicit;
Run;
```

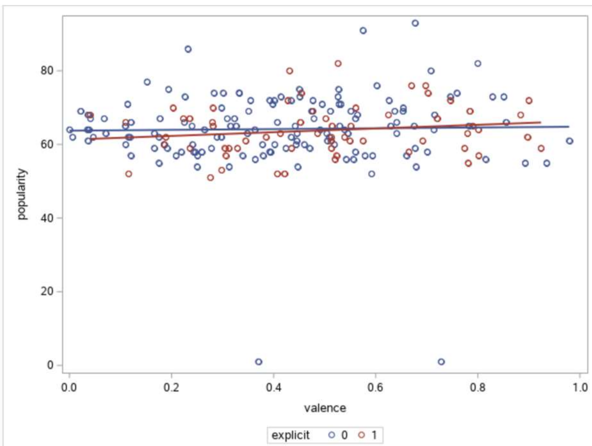
Description: The plot indicates an interaction between danceabilty and explicitness due to the varying slopes of the lines. Songs with explicit lyrics tend to become more popular at high levels of danceability.



SAS code:

```
proc sgplot data=samplesrs2;
scatter x=energy y=popularity/ group=x1;
reg x=energy y=popularity / group=x1;
run;
```

Description: The plot appears to indicate that there is not an interaction between energy and duration.

	**SAS code:**
	proc sgplot data=samplesrs2;
	scatter x=valence y=popularity/
	group=explicit;
	reg x=valence y=popularity / group=explicit;
	run;
	
	Description: The plot seems to indicate that there could be an interaction between valence and explicitness due to the differing slopes of the lines through their scatter plots. Songs with higher valences become more popular with explicit lyrics and less popular without explicitness.

<u>Analysis</u>

Stage 1

Since we had chosen many possible quantitative variables during our initial EDA, we chose to run a stepwise regression on the model to find the most useful quantitative variables to predict population in our model. The results of the tests with SLentry and Slstay values of .15 indicated that **danceability, energy, valence, acousticness, and instrumentalness** were the most important quantitative predictors in our model. Besides danceability, we chose energy and valence as our other quantitative variables as we had already looked at these during our initial EDA and they were also chosen during the stepwise regression. The hypothesized model is $y = \beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\varepsilon_i$ where x_1 is danceability, x_2 is energy, and x_3 is valence. The model was run using a simple random sample of size 200 from the data, indexed to songs released in 2015 and later.

This model resulted in an MSE of 157.804 and an adjusted R-squared value of -0.0008.

We performed a Global F Test testing the null hypothesis that all of the betas are equal to 0, and the alternative hypothesis that at least one of the beta is nonzero. Through our analysis of the data, we produced a test statistic of F=0.95 and a p-value of 0.4182. Therefore, we failed to reject the null hypothesis, getting to the conclusion that we do not have enough evidence to say that at least one of the coefficients is nonzero, therefore we cannot conclude that the model is useful in predicting popularity.

Although these quantitative variables were considered the most useful quantitative predictors in the piecewise regression analysis, alone in the model we did not have enough evidence to say they are useful in predicting popularity in the given model. The F-value associated with the test is extremely low and the p-value is nowhere near the .05 alpha cutoff we are using throughout the model building process. Additionally, the extremely low r-squared values are another indicator that the model is not useful in predicting the popularity of a song, since almost all of the variation in popularity is not explained by the model. Although we got conflicting results through the global utility test, we will move on with the end stage 1 model of : $y = \beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\varepsilon i$.

Stage 2

The qualitative predictors we have deemed most important to the model through EDA and research are duration and explicit. The duration of popular songs before Spotify tended to be between 3 and 5 minutes in order to conform to radio time limits, but artists now have more freedom in deciding on the length of their songs. Additionally, Spotify allows artists to share their work the way that they wrote and recorded it, rather than radio edits that would exclude explicit language. The hypothesized model is $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_5+\varepsilon_i$ where x_1 is danceability, x_2 is energy, and x_3 is valence, x_4 is the duration of the song with the dummy variable being 0 if the duration of the song is 3-5 minutes and 1 otherwise, and x_5 is explicit with 0 being non explicit and 1 meaning the song contains explicit lyrics.

This model resulted in an MSE of 91.409 and an adjusted R-squared value of 0.0039.

We performed a Global F Test testing the null hypothesis that all of the betas are equal to 0, and the alternative hypothesis that at least one of the beta is nonzero. Through our analysis of the data, we produced a test statistic of F=1.16 and a p-value of 0.3318. Therefore, we failed to reject the null hypothesis, getting to the conclusion that we do not have enough evidence to say that at least one of the coefficients is nonzero, therefore we cannot conclude that the model is useful in predicting popularity. As our end stage 2 model, we will remove all of the qualitative variables and end with the model: $y = \beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\varepsilon_i$.

Stage 3

We started by using a hypothesized model of $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_1x_4+\beta_6x_3x_4+\varepsilon_i$ where x_1 is danceability, x_2 is energy, and x_3 is valence, x_4 is explicit with 0 being non explicit and 1 meaning the song contains explicit lyrics. We added back in the x_4 (explicit) for a check to see if the interaction between the quantitative variable and the qualitative variable of explicit was significant enough to keep in our model.

This model resulted in an MSE of 86.026 and an adjusted R-squared value of 0.0628.

We performed a Global F Test testing the null hypothesis that all of the betas are equal to 0, and the alternative hypothesis that at least one of the beta is nonzero. Through our analysis of the data, we produced a test statistic of F=3.21 and a p-value of 0.0050. Therefore, we can reject the null hypothesis and can conclude that at least one of the beta parameters is different from 0, therefore the model is useful in predicting popularity.

After this, we checked each interaction individually through the T test, starting off with the interaction between valence(x_3) and explicit(x_4). With our null hypothesis of $\beta_6=0$ and alternative hypothesis of $\beta_6\neq0$, we resulted in a test statistic of t=3.04 and a p-value of 0.0027. Since the p value is less than our alpha value, we reject the null hypothesis and conclude that the interaction term between valence and explicit is useful in predicting popularity in the given model. This interaction is then kept into the model, resulting in the same model of: $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_1x_4+\beta_6x_3x_4+\varepsilon_i$

This model resulted in an MSE of 85.628 and an adjusted R-squared value of 0.0671.

With the same hypothesized model of $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_1x_4+\beta_6x_3x_4+\varepsilon_i$ where x_1 is danceability, x_2 is energy, and x_3 is valence, x_4 is explicit with 0 being non explicit and 1 meaning the song contains explicit lyrics, we checked the interaction between danceability(x_1) and explicit(x_4) individually through the T test. With our null hypothesis of $\beta_5=0$ and alternative hypothesis of $\beta_5\neq0$, we resulted in a test statistic of t=-0.33 and a p-value of 0.7428. Since the p value is greater than our alpha value, we conclude that we do not have enough evidence to say that β_5 is significantly different from 0, therefore we cannot say it is a significant predictor of popularity in the model. Therefore, we conclude stage 3 by removing the interaction between danceability and explicit and end with an end stage 3 model of: $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_3x_4+\varepsilon_i$.

Summary of Model Building

End of Stage 1 Model: $y = \beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\varepsilon_i$

The end of stage 1 model included three predictors with an F value of 0.95, adjusted R square of -0.0008 and MSE of 157.804.

End of Stage 2 Model: $y = \beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\varepsilon_i$

The end of stage 2 model included three predictors with an F value of 0.95, adjusted R square of -0.0008 and MSE of 157.804.

End of Stage 3 Model: $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_3x_4+\varepsilon_i$

The end of stage 3 model included three predictors with an F value of 3.85, adjusted R square of 0.0671 and MSE of 85.628.

Evaluating the Model

Finally, we evaluated the model by checking each of the regression assumptions, multicollinearity, and looking for outliers/influential points.

To check for each regression assumption, we first looked at the assumptions of lack of fit and the unequal variance assumption. We created residual plots of the residuals vs each of the explanatory variables and the residual vs predicted plot and found that there were no clear trends in each of the explanatory residual plots. Furthermore, there was also no clear trend in the predicted residual plot as well as any fanning out or in pattern. Therefore, we could say that there was no violation in the mean zero assumption/lack of fit and the unequal variance assumption. To check for the normality assumption, We created both the histogram and QQ plot to look at the normality of the model. The histogram had a good bell curve, indicating a normal distribution, and the QQ plot also backed this up. It had no clear trends away from the line, no clusters of residuals in the middle and very spaced in the end, and the residuals seemed to align with the line pretty well. From this, we can say that there was no violation of the normality assumption. For the final regression assumption of residual correlation, we conducted the Durbin-Watson test with a null hypothesis of H_0: The errors are uncorrelated and an alternative hypothesis of H_a: The errors have positive or negative correlation. From our analysis of the test, we found that with Pr<DW being less than 0.0001 and Pr>DW = 1, there is evidence of positive correlation. Because we detected autocorrelation, we would add extra parameters to our model.

To find any multicollinearity from our model, we fit the model with the VIF statement in order to look at and analyze the variance inflation of each of the explanatory variables. No VIF value of any predictors were greater than 10 and the average VIF for all of the predictors seemed to be not much greater than 1. From this, we concluded that multicollinearity was not present in any of the variables.

Lastly, we looked for any outliers and/or influential points for the model. We performed influential diagnostics for the model and we found that observations 199 and 200 were outliers. We also found out that observations close to 200 had a popularity of less than 10, which could be a result in them being outlier observations. To correct this, we would not remove them, because we want to keep as many data points as possible, but rather present our conclusions and analysis keeping in mind that this concern may be a problem for future conclusions.

In the end, we decided to keep our final model of $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_3x_4+\varepsilon_i$

Conclusion

Through our statistical analysis, we determined the prediction equation to be

$\hat{y}=63.86010 + (7.49346)x_1 + (-2.59856)x_2 + (-9.40693)x_3 + (-6.68650)x_4 + (21.49465)x_3x_4$

where x_1 is danceability, x_2 is energy, and x_3 is valence, x_4 is explicit with 0 being non explicit and 1 meaning the song contains explicit lyrics. When danceability, energy, valence and a song is not explicit, the estimated popularity of a song is 63.86010. As shown by the model, songs considered to be explicit decrease popularity by a minimum of 6.8 points compared, but are able to make up for this difference depending on the song's valence (represented by the 5th beta parameter). In addition, when danceability increases by one, popularity increases by 7.49. When energy increases by one, popularity decreases by 2.59. When a song's valence increases by one, popularity decreases by 9.40. If a song is explicit, popularity decreases by 6.69. However, when the interaction between explicit and valence increases by one, popularity increases by 21.49.

Despite being a statistically significant model, low R^2 and adjusted R^2 values decrease the effectiveness of the model. In practical terms, our R^2 value shows that roughly 9% of our data fit our regression line. In addition, a relatively high mean squared error also further highlights the ineffectiveness of our model. Based on the nature of Spotify's rating system (ratings from 0-1 with many points at all values), no outliers were detected in our initial analysis.

We randomly selected Tootsie Slide by Drake to test our model. The song has a danceability of .834, energy of .454, valence of .837 and is considered explicit. When applied to the prediction equation, we get a predicted popularity of 72.36. The actual popularity is 95, which leads us to another issue with our model.

If we were attempting to create the 'perfect' song, we could theoretically design a song to maximize popularity depending on the following factors: danceability, energy, valence, and explicitness. Taking the danceability of Vanilla Ice's 'Ice Ice Baby' (.98), energy of Don Quixote, Op. 35 (.0121), valence of the Barney Theme Song (.998), and enough curse words for the masterpiece to be considered explicit, the popularity of the song will be a 76.5. The fact that our model's popularity maximizes below 80 while a song's popularity ranges from 0 to 100 also contributes to its overall ineffectiveness.

Future Research

If future research were conducted on more song data, we would be interested in making a model specific to one genre as opposed to the industry as a whole. In doing so, we may be able to more accurately predict a song's popularity by analyzing trends among a specific genre. This would also create a smaller, more 'relevant' data set, lessening the need to manipulate data. However, this would require Spotify to add additional variables to their API and then added to the data set. In addition, it would be interesting to separate data based on the specific artists. This would allow us to see how songs that could be considered 'bad' still earn a high popularity score just because of the artist's track record.

Appendix

Durbin-Watson D	0.457
Pr < DW	<.0001
Pr > DW	1.0000
Number of Observations	200
1st Order Autocorrelation	0.661

```
proc reg data=work.samplesrs
plots(only) = (ResidualbyPredicted QQPlot ResidualPlot);
model popularity = danceability energy valence explicit interaction / dwprob;
run;
```


"Get Audio Features for a Track." n.d. Spotify for Developers. https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/.

Leight, Elias. 2017. "Producers, Songwriters on How Pop Songs Got So Slow." Rolling Stone. August 15, 2017. https://www.rollingstone.com/pro/features/how-did-pop-music-get-so-slow-197794/.

Morris, Colin. n.d. "What Makes a Hit?" Columbia Business School. https://www8.gsb.columbia.edu/articles/projects/what-makes-a-hit/.

Tauberg, Michael. 2018. "Music Is Getting Shorter." Medium. April 27, 2018. https://medium.com/@michaeltauberg/music-and-our-attention-spans-are-getting-shorter-8be37b5c2d67.

"What Makes a Song Likeable?" 2018. Towards Data Science. December 3, 2018. https://towardsdatascience.com/what-makes-a-song-likeable-dbfdb7abe404.

Yamac, Eren Ay. n.d. "Spotify Data 1921-2020." Kaggle. https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks.