

EXPLORING POPULARITY AND EXPLICITY OF 2010s MUSIC THROUGH MUSICALITY.

Eric Rios, Emma Wang, Pragya Raghuvanshi, Lorna Aine

2022-11-27

Abstract

The rise of music streaming, algorithms and the ubiquity of ear buds and headphones in the 2010s decade forever changed the way in which we consume music. Beyond the feelings it invokes in us, music is a combination of different qualities that often cannot be easily discerned by the untrained human ear. In contrast, producers and artists, the ones with trained ears, brought forth musical innovations during this advent of technology that changed the way how different aspects of music are recorded, successfully adapting to the new streaming market. Spotify, for example, is the main streaming website that doubles as a gauge on what musical aspects, translated through a lens of streaming metrics, most influence popularity, as well as other indicators of a song's performance (Spotify, 2022). For example, a non-vocal track will likely have a strong presence of instruments, which will be reflected in Spotify's instrumentalness metric.

For this study, we subset a Spotify dataset, with an assortment of descriptors focusing on the 2010s decade to investigate which qualities have been key to making songs popular and if they can be used to classify explicit and non explicit music, a phenomenon that has been rising during the past decade (Tayag, 2017). We found that a song's loudness, for every unit increment, affects a song's average popularity the most by increasing 1.6 units, while acousticness produced a smaller increase of 1.06 units. For the other predictors, for every 0.1 unit increase danceability and instrumentalness, there were increases of 1.13 and 1.14 respectively, which are relatively small but meaningful. For every 0.001 unit increase in a song's energy and instrumentalness, a song's average popularity will decrease by 0.03 respectively. Additionally the model's predictors could only almost 31%, the R^2 value, of the change in a song's popularity.

For the second question, as the tempo becomes slower, a song is 0.996 times less likely to be explicit. For every percentage unit increase in danceability, energy, and speechiness, the odds of a song being explicit compared to being non-explicit is 1.022, 1.006, and 1.09 times more likely. As opposed to preconceived notions, the impact is small, especially for speechiness, despite the increase. For example, if speechiness increases by 10%, the likelihood of a song being explicit increases by 9%. This is plausible because music is comprised of many elements, and enhancing these specific aspects may not specifically produce a large effect. We might need better predictors than what Spotify can provide to be able to find more impactful odds ratios. This conclusion was further validated when using the model from our train set, songs from 2010-2017, on our test set, songs from 2018-19. Overall, the model at the 0.5 threshold had an accuracy score of 81%, with a 14% success rate of identifying explicit songs. The AUC value was a good 79% when grouping positives and negatives, and this score is relatively high. In contrast, when adjusting the cutoff to the best threshold of 0.087, the overall accuracy dropped to 72%, but the ability to predict explicit songs rose to 73.2%. For non-explicit songs, the identification accuracy dropped from 97.9% to a 71.8%. This behavior was expected due to having such an extremely high number of non-explicit songs in our dataset. Also, we need better and more robust predictors to classify explicitness in songs.

Introduction

With over 433 million, monthly active users as of July 2022 (and increasing), Spotify is one of the largest music streaming service providers (Spotify, 2022). They ushered the music streaming market by using algorithms

to optimize for user preferences and monetizing the music streams and ads (ProductBlogs). Labels and producers always look to the Billboard and streaming charts to produce songs that will sonically mimic successful songs or alter trends, and Spotify's place in the market adds even more charting information to this strategy (Hogan, 2018). Although Spotify's algorithm remains one of the unrivaled enigmas of music recommendation, the company builds large databases of tracks on the platform with granular details on the tracks' sonic characteristics, which are synthesized by the algorithm. With these informational touchstones, we were able answer our research questions.

The subset of data used in this research contained 104,767 of observations/ tracks and 23 of variables for songs between 2010 and 2019, obtained from a larger spotify dataset that contained 586,672 tracks. These variables describe the sonic characteristics of songs and were defined below from Spotify's API documentation (Spotify, 2022).

-**Popularity** is calculated by an algorithm that is based on how many times a track has been played and how recent those plays were. This is the response variable of interest for research question 1 (Spotify, 2022).

-**Explicitness** is whether a song contains inappropriate words such as curse words and sexually explicit content that are unacceptable to play in some public settings. 1 is the value identifying a song as explicit, while 0 implies that a song is non-explicit. This is the dependent variable for the second research question (Spotify, 2022).

-**Acousticness** is a confidence measure from 0.0 to 1.0 of how much of the track is composed with acoustic instruments. 1.0 represents high confidence that the track is acoustic (Spotify, 2022).

-**Danceability** is a rating of a track's suitability for dancing. This metric is 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 (Spotify, 2022).

-**Energy** is a perceptual measure of intensity and activity. Energetic tracks typically feel fast, loud, and noisy (Spotify, 2022).

-**Instrumentalness** pertains to whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumentals in this context, while Rap or spoken word tracks are considered "vocal". If the instrumentalness values is greater than or equal to 0.5, the track is very likely to have no vocal content (Spotify, 2022).

-**Tempo** refers to 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, which derives directly from the average beat duration (Spotify, 2022).

-**Loudness** measures 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 association of physical strength (amplitude). The values typically range between -60 and 0 db (Spotify, 2022).

-**Speechiness** detects the presence of spoken words in a track. The more speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attributed 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 (Spotify, 2022).

Research Questions and Justification of Variable Selection

For this research, we sought to investigate two research questions. The first was which musical attributes gauged the popularity of songs in the 2010s, and the following predictors or variables were chosen : acousticness, danceability, energy, instrumentation, tempo, loudness, and speechiness. The second was to what extent could the danceability, energy, speechiness, and tempo of a song predict whether it was explicit or non-explicit. The different variables combinations for the relationships of interest were chosen based on prior, domain knowledge of music. What will follow is the justification for this "a priori" selection for both research questions.

The reasoning for choosing the specific predictors to predict song popularity is due to the weight of their

importance for the first research question. Popular songs, whether they are an emotional ballad or a dance track, all have certain features to keep the listeners engaged and interested in replaying these tracks (Leviatan, 2017). The tempo, energy and loudness indicate the pacing and sonic impact and pleasantness of the track. The speechiness, danceability and instrumentation (which also includes acoustic choices or “acousticness”) dictate melody choices, chord progressions, instrument choices, wordings, vocal lines and more types of sonic layers or arrangements. However, the latter is very nuanced because it pertains to the genre choice of the producers. There are very popular songs with high instrumentation, no words and low danceability, such as songs from classical music. On the other hand, Pop and Rock songs vary their levels of instrumentation and acousticness and speechiness. Lastly, if the song is aimed towards a festive audience, such as a club song, then prioritizing danceability governs the levels of instrumentation and speechiness and lack of acousticness, and this prioritization varies by genre (Androids, 2017). In conclusion, the interplay of these factors influences the popularity of songs by making them memorable and enticing.

As for the second research question, the explicitness of tracks is strongly swayed by other factors. A very logical approach to predicting explicitness was first looking at the high levels of speechiness in songs. For example, rap songs rank high in this metric because the verses are composed of a spoken word format over a series of 8 or 16 bars, and each bar is a rap line (Edwards), while singing doesn't have to adhere to the “1 bar = 1 line” rule; thus, speechiness became the metric of most importance. In addition, songs in this genre tend to include explicit content, often sexual, in the lyrics (Tayag, 2017). Second to this measure, the other predictors of danceability, energy and tempo were considered as helpful in predicting explicitness. The energy and danceability of the song collude with speechiness to infer if a track could have explicit language. For example, a song with low energy and low danceability may or may not be less likely to have explicit language than a song with high energy and danceability, holding the speechiness level constant, and this is a relationship we wish to investigate as well. As for tempo, music genres that are known to include explicit language follow specific tempos. For instance, Trap songs usually have a tempo of 140 bpm (Burchell, 2019).

Methods

Data

The original data set was cleaned so that only tracks and variables relevant to the relationships of interest remained. To isolate the songs from the 2010 decade, we extracted all of the release years from the release date column. To distinguish songs from podcasts and other non-musical audio, we used the speechiness and tempo metrics. Speech tracks with speechiness values greater than 0.66 are podcasts and poetry, and tracks with tempo values of 0 are real songs with missing tempo values or tracks of rain sounds and white noise. Following the completion of these cleaning processes, our final subset contained a total of 104,767 records. We then used charts, graphs, and a table to provide a detailed examination of each of the variables in relation to the research questions.

Models

For the relationship between popularity and acousticness, danceability, energy, instrumentation, tempo, loudness, and speechiness, we used multiple linear regression. Popularity, the response variable, was regressed on the predictors. The predictors were mean centered to improve the interpretability of the model results, with an increase in any predictor causing a change to the average popularity of a song. Additionally, predictors energy, danceability, speechiness and instrumentalness were scaled because their values were recorded with higher precision. The seven predictors were put through a backward selection process with AIC to determine the best indicators of popularity. We choose AIC, as it imposes a lower penalty for having multiple independent variables, and were seeking to maximize the possibility of unknown relationships in our data set. The model with the lowest AIC was selected and fitted.

For the relationship between explicitness and danceability, energy, speechiness, and tempo, we chose to use multiple logistic regression because explicitness was a categorical variable that followed a binomial distribution, meaning variable can only take two values. In this case, a song was either explicit or non-explicit, only two options. The explicitness variable was factored as 1 for explicitness and as 0 for non-explicitness. For the predictors, to improve interpretability as well, danceability, energy, and speechiness were transformed into

percentages. Lastly, the data was split between a train set using tracks released from 2010-2017 and tested on a test set comprised of tracks released from 2018 to 2019.

Model Assessment

The linear regression was checked for the assumptions of linearity, equal variance and independence of error terms, and normality of residuals. Using cook's distance we determined and removed influential points, and the model was refit to accommodate the changes. A Variance Inflation Factor (VIF) test was run to investigate if there was multicollinearity in the model. For the final step, the model coefficients, p values, t values and confidence intervals were assessed and the model results were interpreted.

For the logistic regression, the process breaks down into four steps: assessing the significance of the predictors in the model as well as their insights, checking potential multicollinearity via VIF scores, evaluating the model's accuracy and the model assumptions. For the first, the exponentiated coefficients of our statistically significant predictors indicate the odds ratios that our song may be explicit as compared to non-explicit. For the second, we run a VIF test to verify that our variables are not influencing each other to a high extent because they are assumed to be independent of each other. If there is at least one predictor with a score higher than 5, it suggests high multicollinearity, and the model shall be refit without that predictor. For the third, the model generated from the train set is used to predict the likelihood of a song being explicit or non-explicit in the test set, which is the final step in logistic regression. We then gauge the effectiveness of our model by observing the general accuracy values, generating a confusion matrix and the Area Under the Curve (AUC) value, another measure of how well the model is predicting the negatives and positives in a study. For the fourth and final step, we shall also examine the independence assumption and the linearity assumption via the binned residuals plot, to verify that the data falls within the band of a 95% confidence interval. If the majority of the points are not contained within those bands, an appropriate transformation will be made.

Results

Exploratory Data Analysis

Before beginning EDA, we created a correlation matrix (*see Appendix*) to ascertain the weight of association amongst the predictors, and the variables were not strongly correlated except for acousticness and energy, which had a fairly strong correlation. To better understand the data set in the context of the first research question, all songs with popularity below 80 were compared to those with popularity above 80, ideally those between 100 and 80 popularity index score (*see Appendix*). In this comparison we sought to understand what set apart the top 20% of the popular songs. In Figure 1, we observed that more popular songs have a higher average danceability than less popular songs. In Figure 2, we observed that, as songs became more popular, the energy remained evenly distributed, but the instrumentalness is reduced with the exception of a few outliers, although the relationship between the two variables became insignificant.

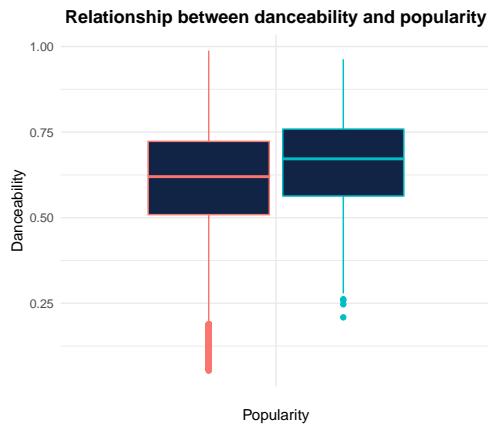


Fig 1

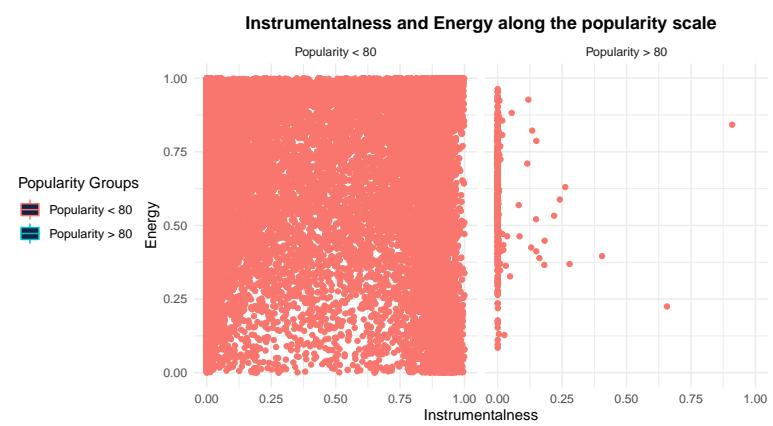
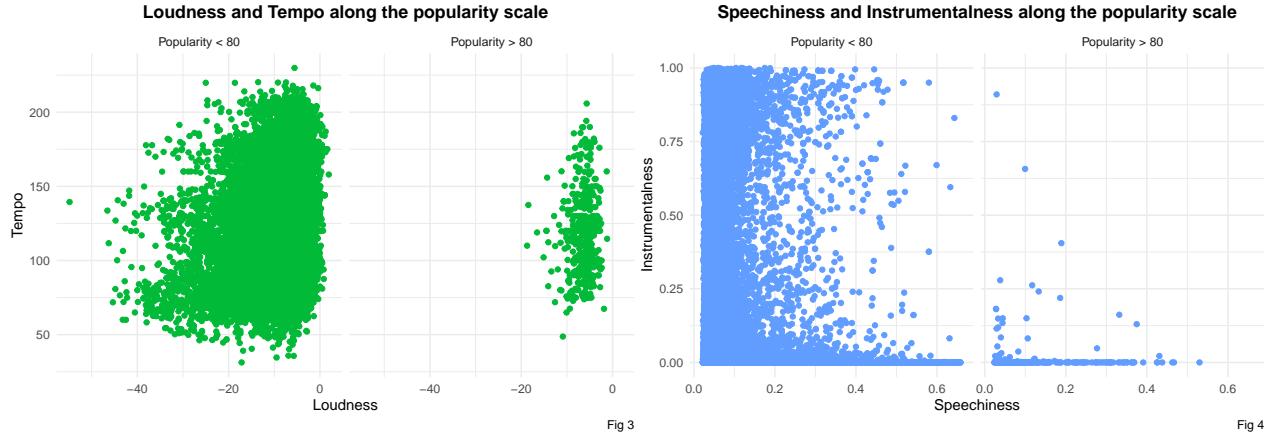
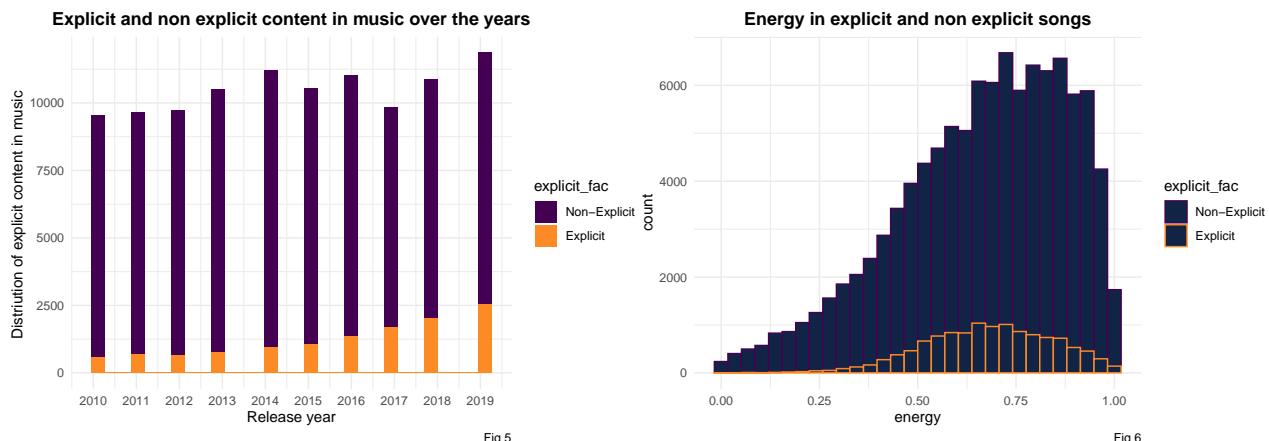


Fig 2



For Figure 3, the relationship between tempo and loudness revealed a very specific recipe for the most popular songs. The popularity greater than 80 group had tempo between 50-200 BPM (beats per minute) and -20 to 0 db, while other groups' tempo and loudness profiles remained spread out across the axes of tempo and loudness. Similarly, for Figure 4, although the speechiness values remained evenly distributed, the most popular songs exhibited a massive reduction in values in instrumentation. Overall, the insights showed that there seems to be an assortment of musical attributes that define an extremely popular song, yet they seem to be less pronounced for the less popular groups of songs. In other words, the most popular song profiles demonstrated an absence of instrumentalness and a specific range of tempo and loudness values.

For the second research question, we explored the basic statistics of the variables of interest. In the table (*see Appendix*), we categorized the mean, median, standard deviation, minimum and maximum values of variables by explicit and non explicit. While mean values of acousticness, instrumentalness, tempo are higher for non explicit songs, values for energy, danceability, loudness and speechiness were higher for explicit songs. In addition, we could infer that popularity of explicit songs is quite higher than non explicit songs. High values for standard deviation for tempo and popularity indicate that the data points were spread out in relation to the mean value, whereas low values for danceability, energy, speechiness indicate that the data points were clustered around the mean. Lastly, nearly equal values of median and mean for danceability, energy and tempo indicate that the data points were more or less evenly distributed. For Figure 5, it is clear that the explicit content in music rose over the past decade. As for Figure 6, the energy in explicit songs centered around a mean of 0.680, while the non explicit songs was skewed to a higher energy metric. In terms of Figure 7, the danceability of songs was centered around the mean of 0.687 for explicit songs and 0.599 for non-explicit songs. As for Figure 8, on average, explicit songs were found to be less speechy than non-explicit songs. In conclusion, the chosen variables would give us a good metrics to classify explicit and non explicit songs moving forward.



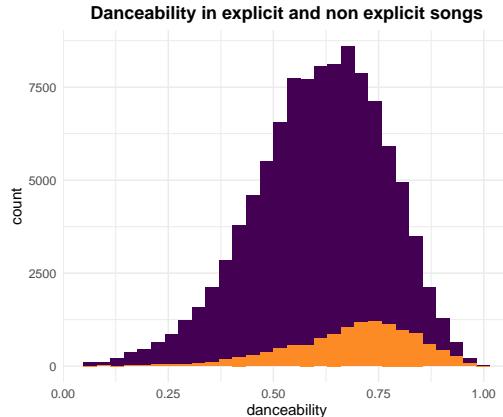


Fig 7

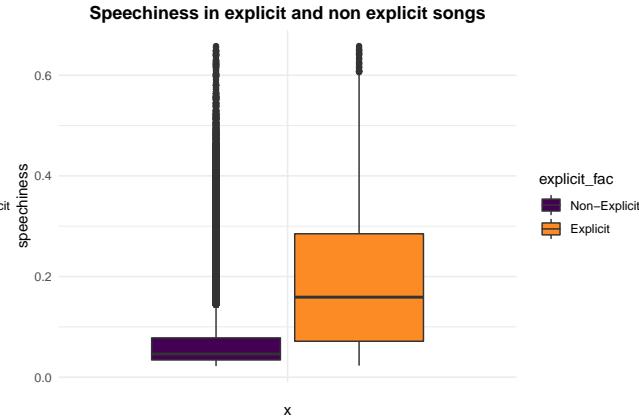


Fig 8

Model Results

RQ1 : Linear Regression Model

Observations	98380
Dependent variable	popularity
Type	OLS linear regression

F(7,98372)	6195.8388108
R ²	0.3059827
Adj. R ²	0.3059334

	Est.	S.E.	t val.	p
(Intercept)	39.3291358	0.0491853	799.6114795	0.0000000
acousticness_centered	1.0624778	0.2365748	4.4910860	0.0000071
danceability0.1_centered	1.1297289	0.0343530	32.8858805	0.0000000
energy0.001_centered	-0.0276904	0.0004289	-64.5647871	0.0000000
instrumentalness0.001_centered	-0.0327584	0.0002508	-130.6255204	0.0000000
tempo_centered	0.0008505	0.0018030	0.4717060	0.6371377
loudness_centered	1.6027811	0.0251106	63.8288884	0.0000000
speechiness0.1_centered	1.1367334	0.0574669	19.7806511	0.0000000

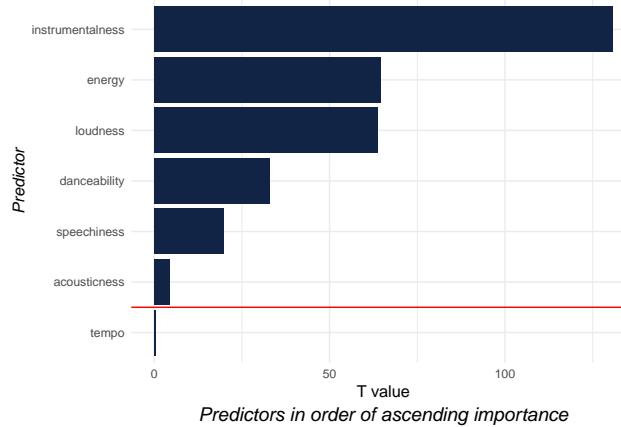
Standard errors: OLS

From the above model results, acousticness, danceability, energy, instrumentalness, loudness and speechiness are all significant predictors of a song's popularity score at the $\alpha = 0.05$ significance level ($p < 0.005$). At a mean popularity of 39, a song has an average value of 0.29 for acousticness, 0.61 for danceability, 0.66 for energy, 0.09 for instrumentalness, 122.28 for tempo, -7.25 for loudness, and 0.09 for speechiness. For every 1 unit increase in acousticness and loudness, a song's average popularity will increase by 1.06 and 1.6 respectively. For every 0.1 unit increase in danceability and speechiness, the song's average popularity will increase by 1.13 and 1.14 respectively, and, for every 0.001 unit increase in a song's energy and instrumentalness, a song's average popularity will decrease by 0.03 respectively in all cases holding all other factors constant. Additionally the model fit has R^2 value of 0.3059, meaning that almost 31% of the change in a song's popularity is explained by these predictors.

We are 95% confident that the intervals in the table below contain the true value for the significant predictors. The predictors in this model were found to be important in the order visualized below with significance cut off at acousticness.

Table 1: Confidence interval ranges for popularity model

	2.5 %	97.5 %
(Intercept)	39.2327332	39.4255385
acousticness_centered	0.5987940	1.5261617
danceability0.1_centered	1.0623974	1.1970604
energy0.001_centered	-0.0285310	-0.0268498
instrumentalness0.001_centered	-0.0332500	-0.0322669
tempo_centered	-0.0026834	0.0043844
loudness_centered	1.5535646	1.6519975
speechiness0.1_centered	1.0240989	1.2493679



RQ2 : Logistic Regression Model

Table 2: Odds Ratio and Confidence intervals

	Odd_ratios	X2.5..	X97.5..
danceability	1.0223159	1.0204765	1.0241663
energy	1.0063125	1.0049602	1.0076739
speechiness	1.0949458	1.0926451	1.0972606
tempo	0.9956723	0.9947489	0.9965945

From the model results above, the model predictors are all statistically significant in predicting a song's explicitness at $p < 0.005$. In terms of interpreting the coefficients, holding all other variables constant, the odds of a song being explicit compared to being non-explicit is 1.022 times more likely for every percentage unit increase in the danceability; 1.006 times more likely for every percentage unit increase in the energy predictor; 1.09 times more likely for every percentage unit increase in the speechiness; 0.996 times less likely for every unit increase in the tempo. We are 95% confident that the true exponentiated coefficients for our predictors lie between the values in our confidence interval.

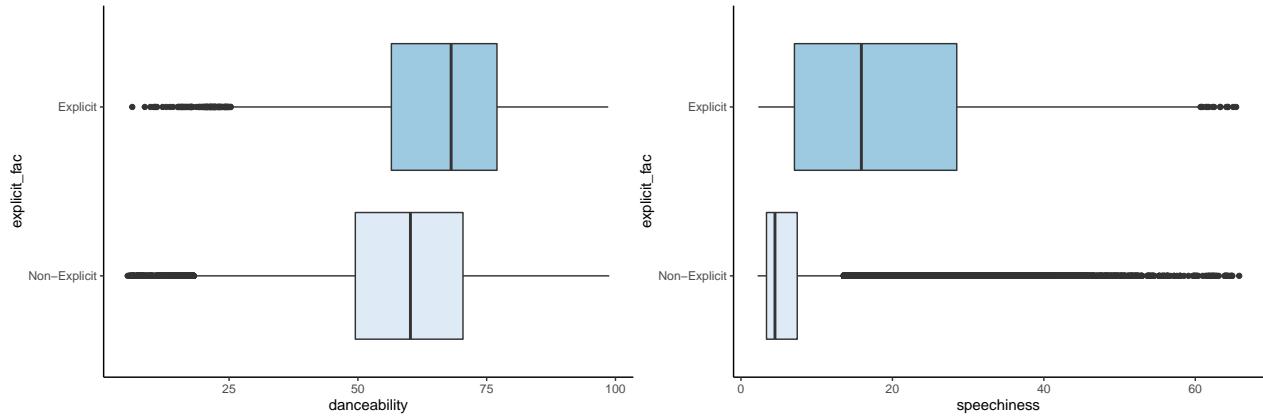


Figure 1. Danceability and Explicitness

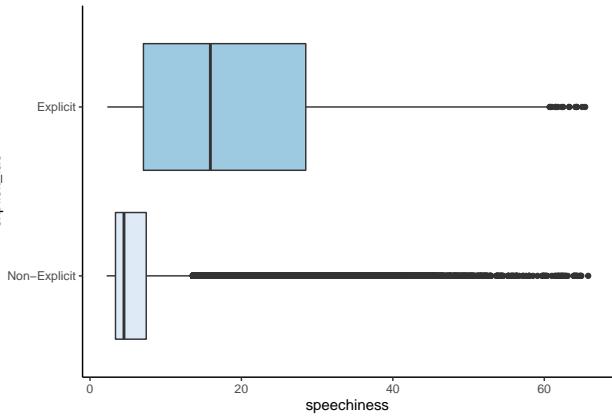


Figure 2. Speechiness and Explicitness

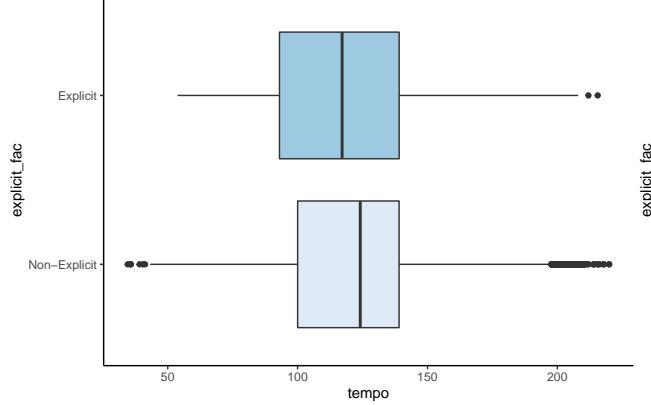


Figure 3. Tempo and Explicitness

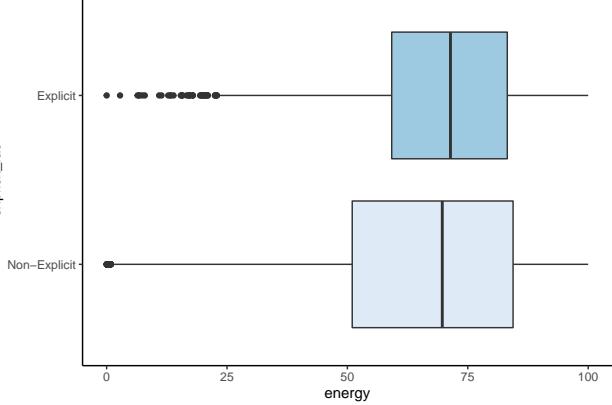


Figure 3. Energy and Explicitness

From the bar graphs above, we can see that only tempo has a slightly negative effect on explicitness. Unlike expectations, the tempo's odds ratio indicates that, as a song becomes slower, it will be less likely to be explicit. On the other hand, the speechiness, danceability, and tempo are positively related to explicitness. More specifically, we can see speechiness and danceability have stronger relationship with explicitness as compared to tempo and energy. Even though every predictor in the analysis has been proved to be statistically significant, showing the evident association between the predictors and the explicitness, the odds ratios do not seem to be impactful when examining their sizes. This makes sense given that music consists of many elements and only enhancing one aspect may not necessarily have a huge effect on explicitness. For instance, if danceability increases by 10%, it will only make the odds of being explicit increase by 6% ($1.006^{**}10 = 1.06$). In other words, very danceable records are technically more likely to have explicit lyrics, but the gamut of the impact isn't quite strong. Energy also implied a similar relationship. As for speechiness, its odds are best understood by thinking "the more words a song has, the more likely it is to have some explicit content", but it follows this trend of small impact as well.

Model assesment

Checking Assumptions for the linear Model

- Linearity:** The residuals vs fitted plot shows that there is no clear pattern between the two variables, and, the data points do not move in a non-linear manner.
- Independence of error terms :** The points on the residual vs fitted plot are randomly spread, so there was no clear violation of this assumption. In addition, it is also assumed that Spotify recorded this data meticulously during their data collection process, which is validated by this plot.
- Equal Variance of error terms :** The points on the residual vs fitted plot are equally spread around the ab line(0,0); therefore, we can conclude that there is homoscedasticity.
- Normality of residuals:** The points on the quantile-quantile plot are mostly clustered along the 45 degree line, even though a minimal amount of points are trailing off at the beginning and the end of

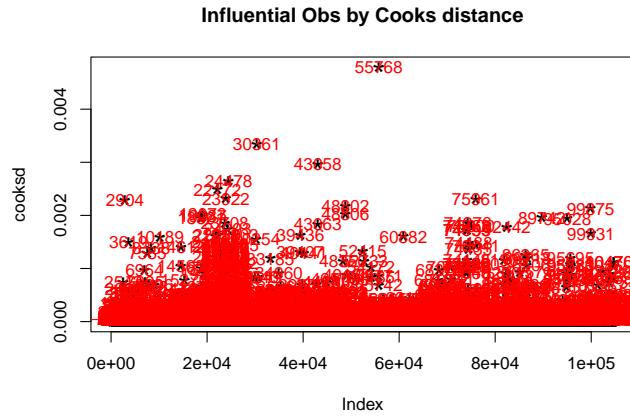
the line. However, there are no clear or strong violations of this assumption. In conclusion, the assumptions of the model do not appear to be strongly violated.

Multicollinearity in the linear model

The model predictors were tested for multicollinearity using the Variance Inflation Factor. In the table below, we observed that there was no multicollinearity as all the VIF scores were below 5.

Outliers and Influential points

We checked the data for outlier and influential points and compared the effect of removing them on the model. We used Cook's Distance to determine how much influence an observation has on the model with a higher distance meaning more influence as it uses the leverage and the residuals to calculate the score. All points having distance above $4/N$, where N is the sample size, are influential points, and those were excluded from the data points, which reduced the data set to 98,380 observations. After removing the influential points the model's R^2 increased from about 17% to 31%, and the tempo predictor became statistically insignificant.



Multicollinearity in the logistic model

The model's predictors were tested for multicollinearity using the Variance Inflation Factor. In the table below, no multicollinearity was found to be present, as all the VIF scores were hardly above 1, which means the predictors were independent.

Table 3: VIF Results for logistic regression

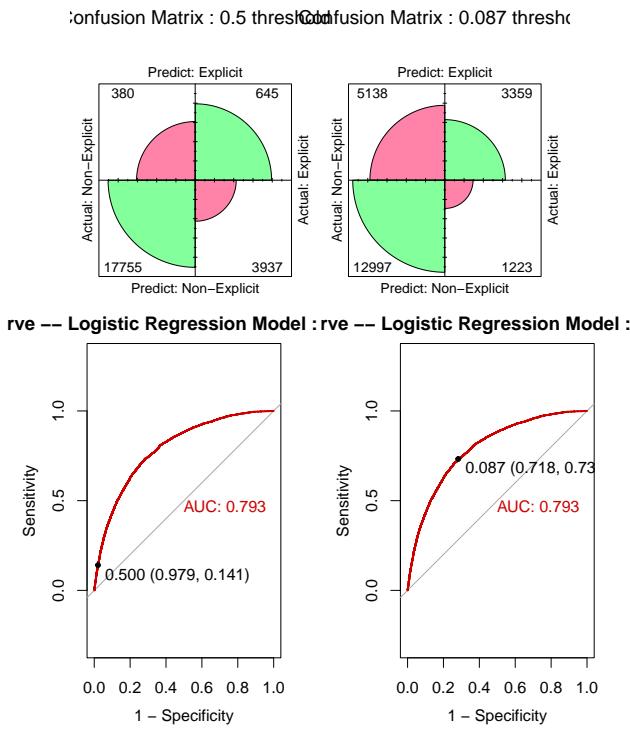
	x
danceability	1.019711
energy	1.031062
speechiness	1.006210
tempo	1.042395

Evaluating Model Testing the model

After training our model on songs from 2010 to 2017, the following step was to predict the a song's likelihood of being explicit and non-explicit for our test set, composed of songs from 2018-19. Shown below are the confusion matrix (see appendix) and the ROC curve (see appendix), as well as the overall accuracy values. One of our challenges was addressing the small number of explicit songs in our data, where it might hurt model accuracy. That concern was proven true when our model failed at correctly classifying 3,937 explicit songs. It had an overall, high accuracy of 81%, a sensitivity of 97.9% and a specificity of 14%. The sensitivity value signifies that our model is efficient at identifying non-explicit songs, while our specificity value demonstrates that our model is lacking when attempting to identify an explicit song. Also, the AUC value demonstrated a 79% success rate of distinguishing negatives and positives.

When setting an arbitrary threshold, 0.5 tends to be the most intuitive, like a coin toss, where heads and tails have equal chances of coming up. However, as previously stated, the explicit points were completely outnumbered by their non-explicit counterparts. One of the ways to counteract this deficit is to use the ROC curve's arguments to print the best threshold, which is a new 0.087. This best metric maximizes the sensitivity and the specificity and sensitivity as much as possible. Upon doing so, the sensitivity dropped to 71.8%, and the specificity rose to 73.2%, while overall model accuracy decreased to 72%. In addition, the model then identifies 3,359 explicit songs 12,997 non-explicit songs correctly (unlike the 17,755 non-explicit songs under the previous threshold). This is a conscious trade-off to increase model accuracy for the explicit songs.

For the binned residuals plot (see appendix), Overall, it shows that we have a good fit for our data since most points fall within the bands of the 95% confidence interval. However, we indeed need to be cautious about a few of outliers, where there are five having residuals larger than a 0.1 difference. Nonetheless, given that our sample size is large enough, we would not need further actions for only 5 outliers, so our linearity assumption appears to not have been violated. For our independence assumption, we surmise that it is completely satisfied because the data points were collected independently from each other, and there seems to be no pattern to the clustering of the data.



Conclusion

Key Insights

In the analysis of the first research question, which dealt with inferring the significant predictors impacting the popularity of the song, we have a finding that *instrumentalness* is the most important attribute influencing the popularity of the song in 2010s decade. A track with high levels of instrumentalness is likely to be least appealing to the audience. Also, other important predictors impacting the popularity are *acousticness*, *danceability*, *energy*, *loudness* and *speechiness*. For the popularity of the track to be high, instrumentalness and energy, the most significant musical attributes should have low values. A track that has high levels of acousticness, danceability, loudness and speechiness is likely to have a higher popularity score. Furthermore, loudness was the predictor with the highest coefficient, which is not surprising. During music mastering, adjusting loudness commands high prioritization due to how sensitive the human ear perceives changes in it (Sage Audio).

Based on the second research result, we can conclude that the four chosen predictors, danceability, energy, speechiness and tempo, are indeed related to explicitness and can predict explicitness to some extent. Overall, the model at the 0.5 threshold is a good fit with an acceptable, prediction accuracy score of 81%, with a 14% success rate of identifying explicit songs. The AUC value was 79%, a determinant of how well the model can classify positives and negatives, and this score is relatively high. On the contrary, when adjusting the cutoff to a best threshold of 0.087, the overall accuracy drops to 72%, but the ability to predict explicit songs rises to 73.2% while the capacity to classify non-explicit songs drops from 97.9% to a 71.8%. The gains and losses from this trade-off are evident, and we conclude that we need better and more robust predictors to have better accuracy. Yet, we still remain with a pressing question. Why do we prioritize correctly classifying explicit songs? If a song's capacity to play in a setting depended on this categorizing, then this model would have to be more accurate. An ill classified explicit song should not play in a setting for children. Conversely, a misidentified, non-explicit song should not be penalized and lose potential profits by not being allowed to play in a general, public setting. However, these arguments would mostly apply if our statistical model was a mission critical one, but it is not; Spotify most likely employs methods from the realm of natural language processing to flag the explicitness of songs, not statistical methods. This implies that they would ignore most of the musical aspects when employing their codification processes.

Limitations

For the first research question, even after removing the influential points, we observe the Adjusted-Rsquared value increases to only 0.3059. This could be a potential limitation as the predictors can only explain variability in the outcome, the popularity score, up to 30.59%. This indicates a possibility of other significant predictors, which have not been taken into account for the analysis. For example, besides the attributes of a track, popularity could also be quite dependent largely on who is the artist. Secondly, the different genres have different combinations of musical attributes that make them popular within their respective frameworks. Therefore, this analysis doesn't integrate genre classification into its context. Thirdly, the interest of our research deals with the 2010s decade and cannot be extrapolated to the popularity of a track from other timeframes. As, music is a very dynamic research area, the attributes that appeal to people in recent times may not have been significant in an earlier time period. Lastly, we initially theorized that energy and tempo seem to have the similar kinds of statistical impact, yet, according to the model analysis, while energy is an important and significant attribute, tempo is not. Also, even though it seems that popular songs should have higher energy levels to drive higher engagement with the audience, our model analysis suggests otherwise. These discrepancies might indicate a possibility of an interaction term, which we haven't considered in the analysis.

For the second research question, despite providing insights toward predicting explicitness, the analysis indeed has some limitations. First, our model is weak at identifying non-explicit songs, probably because there was a smaller distribution of explicit songs in our dataset, which is reflective of real word patterns in regards to music consumption. The majority of songs are simply non-explicit; however, such distribution might hinder the training process when desiring to correctly categorize explicit and non-explicit songs. Secondly, more or better predictors may have been needed for this particular research question, such as the language of the song, its genre, popularity, duration, amongst others. Thirdly, Spotify likely employs the use of speech recognition (natural language processing) models to flag the explicitness of songs, but we used statistical methods. Thus, the research where we used the four predictors to statistically predict the nature of a such a "latent" variable may have already been fraught with much difficulty since the beginning. To conclude, these results do prove that any variant of our model, even after adjusting the thresholds, is insufficient for predicting the explicitness of songs (there are 19% of errors in predicting for the former and 28% for the latter), but, at least, it is overall successful when classifying the explicitness and non-explicitness of songs. For this second question, attempting to predict explicitness while overlooking the fundamental human element. Songwriters are the people who are making songs explicit and non-explicit.

For both questions, there are two glaring limitations that were overlooked during the process of this study. First and foremost, Spotify translated a numerous set of musical attributes into a few variables for streaming algorithms. This means an immeasurable amount of musical information was lost, such as silences, crescendos, instrument classes, among others. Even singer vocal types could have served as a predictor. Key changes in a

song could have been another predictor. For the second question specifically, an oversight was attempting to predict explicitness while overlooking the fundamental human element. Songwriters are the people who are making songs explicit and non-explicit. Maybe predictors can be added that address the types of songwriters and the destinations for these tracks, which will be addressed in recommendations.

Recommendations/Future Studies

For the first research question, in future studies, we can achieve better explanatory power by including more predictors crucial to understanding the popularity of the tracks, which may have not been included in this investigation despite our most careful approach when selecting variables with prior domain knowledge. Duration is one example, since popular songs tend not to be very lengthy. We can also use genres to classify how different musical attributes affect the popularity track for different genres. This may give a better understanding of how the significance musical attributes vary when it comes to different genres. The time period of the research can be expanded to incorporate 1990s and 2000s decade for a better understanding of how the significance of musical attributes has changed over a period of time. We may also check the model for possible interaction terms that may influence the outcome variable, popularity, and subsequently improve the explanatory power of the model.

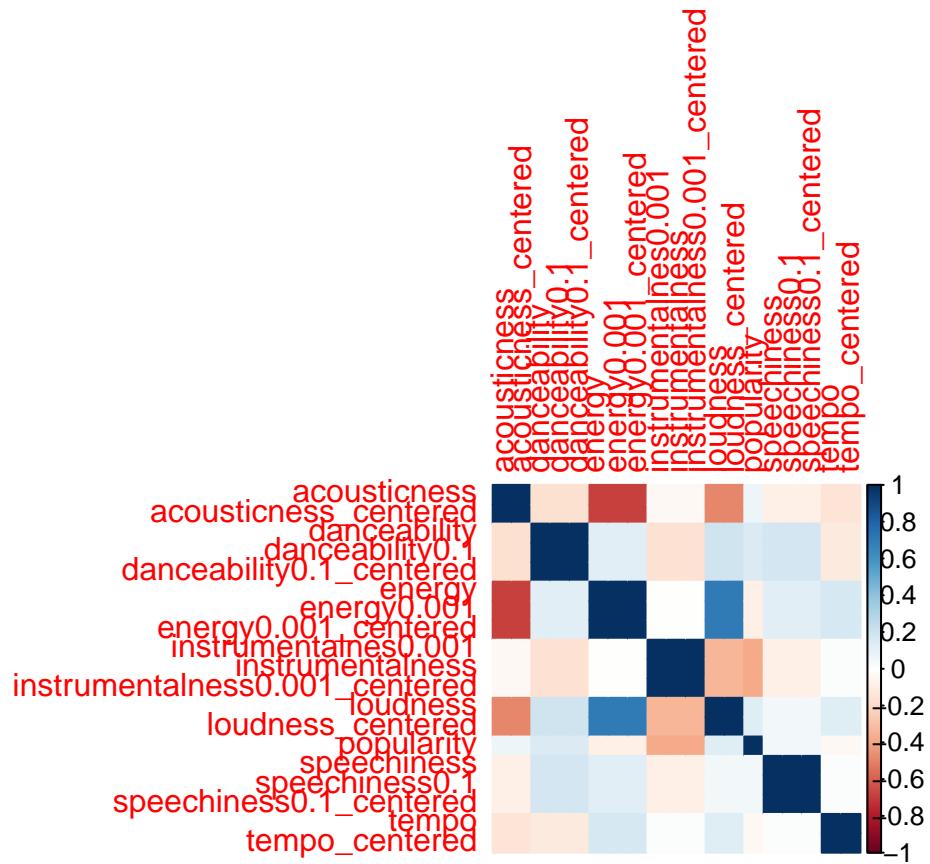
For our second research question, future work needs to pay more attention to the fundamental causes of wrongly flagging non-explicit songs as explicit songs. Incorporation of more songs in the dataset that are explicit will lead to a more balanced training process, which can lead to better predictive power. We can also incorporate predictors like language, duration, and genre to aid the predictive power of the model. Interaction terms may also be investigated for further improving the predictive power of the analysis. One that comes to mind is the interaction between speechiness and tempo, given that trap songs follow certain tempo and speechiness levels (Burchell, 2019). A final recommendation, specifically due to the hurdles faced with this research question, is potentially accessing another Spotify dataset that contains other variables for these songs. They may be able to shed light on better ways to handle prediction for this particular response variable. In this same thread of thinking, a way to address the songwriters who make these songs explicit and non-explicit is by adding variables that code their record labels, since the number of labels is finite when compared to the amount of songwriters. The destinations or purposes of these tracks can be recorded as well. Is it for a tv series, a commercial, a movie, a videogame or an album? Variable ideas like these can help make the predictions more accurate when we choose to refit the model.

As for other avenues of study, There are some variables that could have been relatively interesting in terms of prediction or inference. To illustrate, one could predict the scale of a song based on the popularity, in addition to other predictors. Duration may have been predicted from the genre, loudness, danceability, popularity, and instrumentality. Tempo itself could have also been a response variable, modeled by using duration, popularity, speechiness and energy. There are many possibilities. In other cases, some predictors are not good choices. Time signature, which was improperly coded, has an extremely limited set of plausible values (lacks signatures like 6/8). In this example, Spotify could do a better job at classifying the data, since it was able to identify scales relatively effectively. Other variables like Liveness, which parametrizes a song's performance as a live or studio quality recording, have not been taken into consideration, but can be probed for gaining more insights as the popularity of live songs seems to be increasing in recent times.

fix this # Appendix ## Table 1

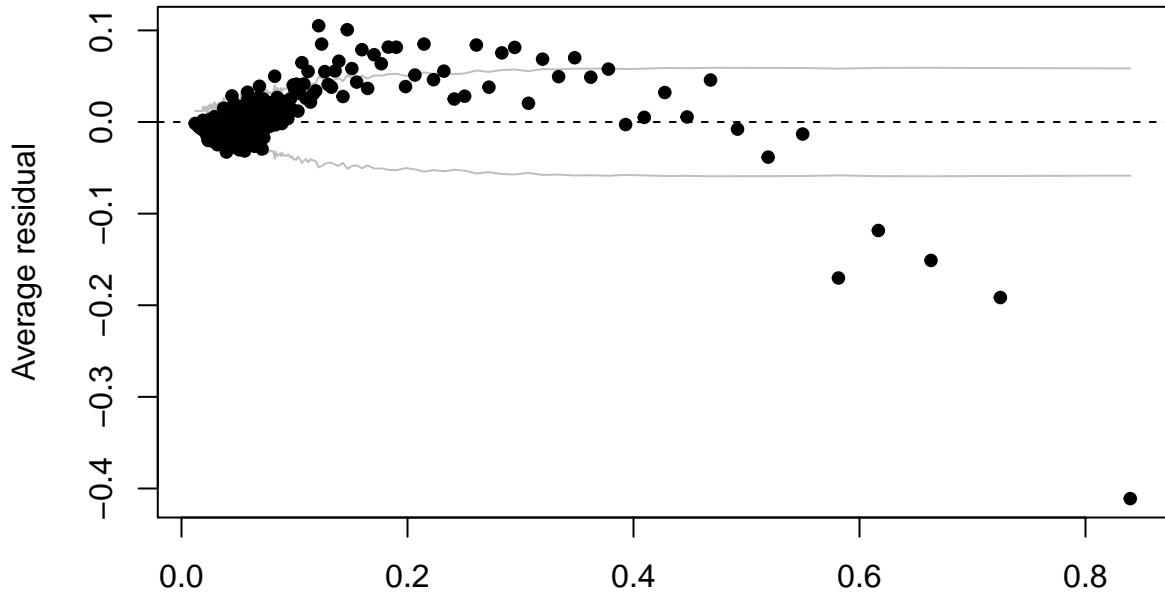
	Non-Explicit (N=92440)	Explicit (N=12327)	total (N=104767)
acousticness			
Mean (SD)	0.295 (0.300)	0.227 (0.224)	0.287 (0.293)
Median [Min, Max]	0.181 [0, 0.996]	0.151 [0.00000115, 0.993]	0.176 [0, 0.996]
danceability			
Mean (SD)	0.599 (0.154)	0.687 (0.147)	0.609 (0.156)
Median [Min, Max]	0.609 [0.0532, 0.988]	0.708 [0.0620, 0.986]	0.620 [0.0532, 0.988]
energy			
Mean (SD)	0.657 (0.224)	0.680 (0.162)	0.660 (0.217)
Median [Min, Max]	0.692 [0.0000203, 1.00]	0.685 [0.0000202, 1.00]	0.691 [0.0000202, 1.00]
instrumentalness			
Mean (SD)	0.0975 (0.254)	0.0212 (0.110)	0.0885 (0.243)
Median [Min, Max]	0.00000368 [0, 1.00]	0 [0, 0.989]	0.00000251 [0, 1.00]
tempo			
Mean (SD)	123 (28.2)	120 (29.4)	122 (28.3)
Median [Min, Max]	124 [31.3, 230]	120 [46.7, 220]	124 [31.3, 230]
loudness			
Mean (SD)	-7.32 (3.90)	-6.71 (2.56)	-7.25 (3.77)
Median [Min, Max]	-6.52 [-54.8, 1.93]	-6.40 [-29.0, 1.26]	-6.50 [-54.8, 1.93]
speechiness			
Mean (SD)	0.0752 (0.0754)	0.188 (0.130)	0.0885 (0.0912)
Median [Min, Max]	0.0461 [0.0220, 0.658]	0.159 [0.0229, 0.658]	0.0497 [0.0220, 0.658]
popularity			
Mean (SD)	38.1 (19.5)	47.7 (17.4)	39.2 (19.5)
Median [Min, Max]	42.0 [0, 92.0]	50.0 [0, 94.0]	42.0 [0, 94.0]

Correlation Matrix

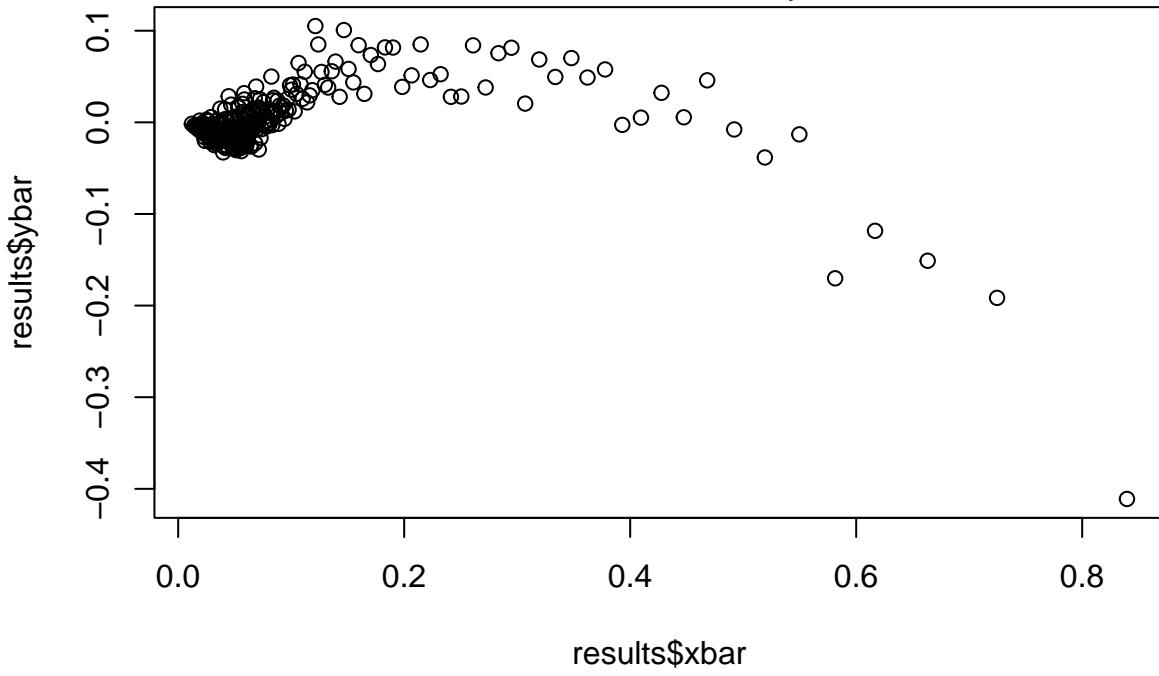


Binned Residual Plot

Binned residual plot



Predicted Probability



ROC Curve showing the 0.5 threshold

Bibliography

Androids (2017, October 13). An Idiot's Guide to EDM Genres. Retrieved October 20, 2022, from <https://www.complex.com/music/an-idiots-guide-to-edm-genres/>

Billboard. (2022, October 29). Billboard Hot 100. Billboard Media. Retrieved November 3, 2022. <https://www.billboard.com/charts/hot-100>

//www.billboard.com/charts/hot-100/2022-10-29/

Burchell, C. (2019, May 27). 10 Tips for Making Your First Trap Beat. Inverse. Retrieved October 20, 2022, from <https://flypaper.soundfly.com/produce/10-tips-for-making-your-first-trap-beat/#>

Edwards, E. (n.d.). Rap Song Structure Is TOO Important To Ignore. Retrieved October 20, 2022, from <https://rhymemakers.com/rap-song-structure/>

Hogan, M. (2018). Billboard Charts Change to Count Paid Streams More Than Free. Retrieved November 25, 2022, from <https://pitchfork.com/news/billboard-charts-change-to-count-paid-streams-more-than-free/>

Leviatan, Y. (2017, July 27). Making Music: The 6 Stages of Music Production. Waves. Retrieved October 20, 2022, from <https://www.waves.com/six-stages-of-music-production>

ProductBlogs. How Spotify Built a \$20 Billion Business by Changing How People Listen to Music. Retrieved November 25, 2022, from <https://producthabits.com/how-spotify-built-a-20-billion-business-by-changing-how-people-listen-to-music/>

Sage Audio. What is Loudness for Mastering? Retrieved November 25, 2022, from <https://www.sageaudio.com/blog/mastering/what-is-loudness-for-mastering.php>

Spotify. (2022). Spotify Web API Reference | Spotify for Developers. Retrieved October 20, 2022, from <https://www.waves.com/six-stages-of-music-production https://developer.spotify.com/documentation/web-api/reference/#/operations/get-audio-features>

Spotify. (2022, July 27). Q2 2022 Update. Retrieved November 25, 2022 from https://s29.q4cdn.com/175625835/files/doc_presentation/Q2-2022-Shareholder-Deck-FINAL.pdf

Tayag, Y. (2017, May 17). Expert on Male Psychology Explains How Pop Got Sexually Explicit. Retrieved October 20, 2022, from <https://www.inverse.com/article/31842-pop-music-sexually-explicit-lyrics-rap-hip-hop>

Yamac. (2016). Spotify Dataset 1921-2020, 600k+ Tracks. Spotify. Retrieved October 3, 2022, from <https://www.kaggle.com/datasets/yamaerenay/spotify-dataset-19212020-600k-tracks>