Purple 3:
Akanksha, Eduardo, Jane,
Orijeet, Mollie, Yixie (Eddie)

Descriptive and Predictive Analysis on Spotify Data Set

Goal of the Project: Predict whether a song would be a hit or a flop by examining audio features of songs from 1960-2010. If a music studio were to start a new project, which features such as loudness, acousticness should they be mindful for when creating new songs?

Description of the Dataset: The Kaggle Spotify hit dataset contains 6 datasets partitioned by decade.

Name	**Type**	**Description**
Track	character	The name of the track.
Artist	character	The name of the Artist.
Uri	character	The Spotify url for the track.
Danceability	double	Describes how suitable a track is for dancing. 0 to 1 indicates Low to High
Energy	double	Measure of intensity and activity. 0 to 1 indicates Low to High
Key	integer	The key of the track. 0 to 11 indicates C note to B note
Loudness	double	Loudness of track in decibels
Mode	integer	Modality(Major or Minor) of a track
Speechiness	double	Presence of Spoken words in a track. 0 to 1 indicates Low to High
Acousticness	double	Measure of the acousticness of track. 0 to 1 indicates Low to High
Instrumentalness	double	Likelihood of a track containing no vocals. 0 to 1 indicates Low to High
Liveness	double	Presence of an audience in the recording.0 to 1 indicates Low to High
Valence	double	Musical positiveness conveyed by a track. 0 to 1 indicates Low to High
Tempo	double	Tempo of a track in BPM(Beats per Minute)
Duration_ms	integer	Duration of the track in milliseconds
Time_signature	integer	Overall time signature of a track (1,2,3,4 or 5)
Chorus_hit	double	Timing of the start of chorus
Sections	integer	Number of sections of a particular track

Target	integer	Whether the track is a hit (1) or not (0)

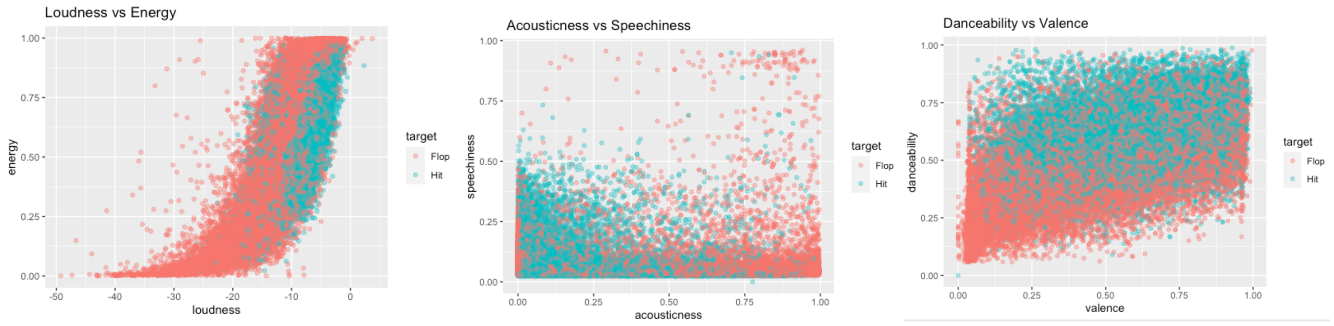
The steps that we took to pre-process the data are as follows :

1. Added decade column to each dataset to compare differences between the decades for exploratory analysis.
2. Combined data sets: We then combined all 6 decade datasets together and saved the file. At the end we had over 40,000 unique values.
3. While building the predictive models, we removed artist, track, decade, and uri as they were not relevant predictors for our analysis.

Data Visualization and Exploratory Analysis

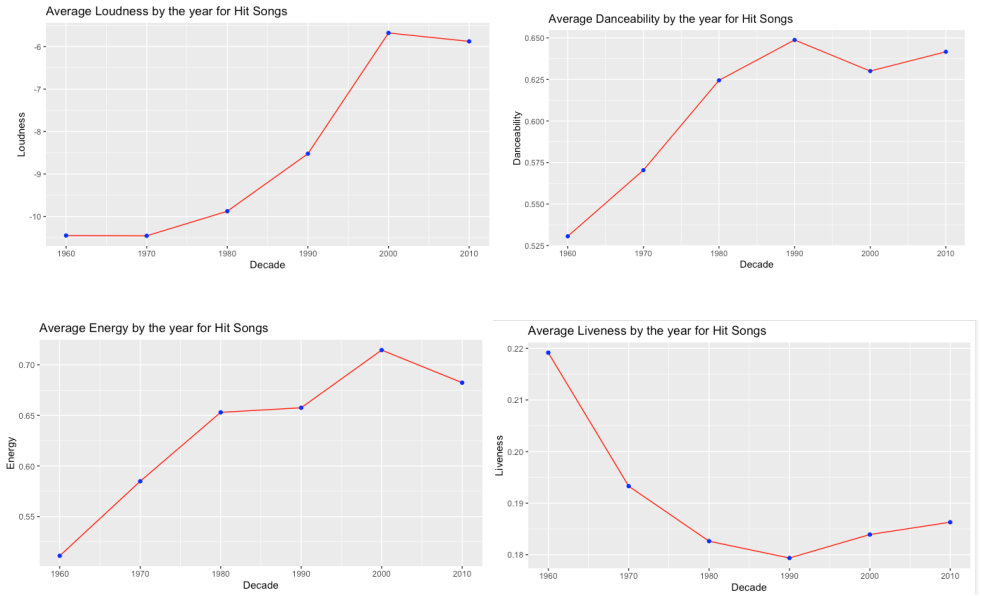
We started with exploratory analysis in order to find correlating factors with the Target value. We then graphed scatter plots using combinations of predictors which further showed that most hit songs in the dataset had higher loudness, energy and danceability and much lower acousticness whereas speechiness and valence had no effect on the target value *(Figure 1.1)*.

Figure 1.1: Scatter Plot.



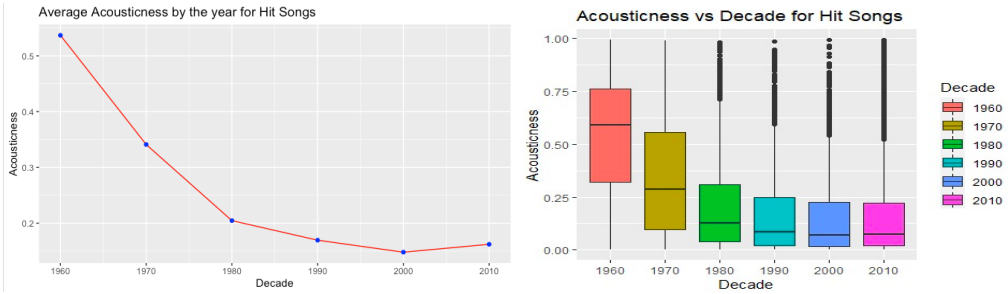
To examine general trends of characteristics across decades, we took all numerical predictors as they were and computed an average score for all across each decade (appendix 1.1). In addition, we examined averages only for those that were labeled as hit songs. It was observed that over the years average danceability, loudness and energy of hit songs has increased tremendously however liveness, acousticness and valence have decreased drastically *(Figure 1.2 & Figure 1.3)*.

Figure 1.2:



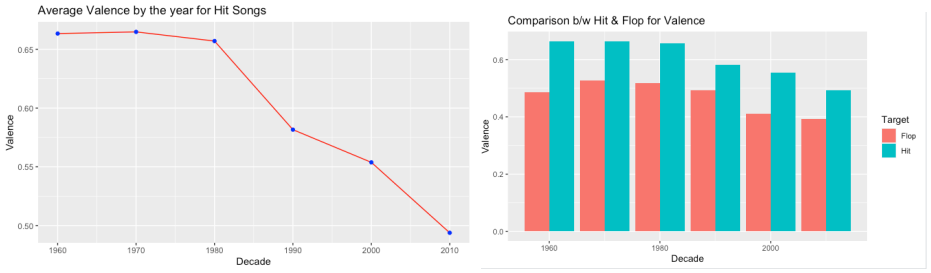
We then graphed the overall boxplot of these components to gain a better understanding of the dataset and observed that while median of acousticness decreased over the years but in the last 3 decades there were exceptions of hit songs with higher acousticness *(Figure 1.3)*.

Figure 1.3:



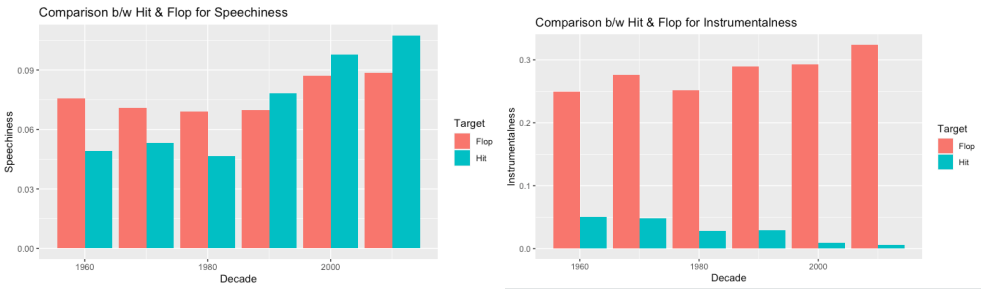
Further comparisons of the Hit and Flop songs were done using continuous bar graphs where we observed that even though average valence has decreased over the years, average valence of hit songs have always been higher than that of flop songs i.e. positive songs were always favored by the audience*(Figure 1.4)*.

Figure 1.4:



In figure 1.4, we observe that in the first 3 decades songs with lower speechiness became hits, however from 1990s onwards songs with higher speechiness were more popular. Further, from figure 1.5 we can conclude that not only songs with lower instrumentalness are favored by the audience but also over the years average instrumentalness has reduced significantly.

Figure 1.5:



Data Modeling and the Methodology

Our initial approach for building the model was to select all the relevant variables and assess their predictive abilities across all decades. We then decided to test the predictive power of each of these models when we separate the data set into individual decades. The seed was set to 100 and kept a 30-70 split for the validation-training data set across all models. In our modeling approach we used Logistic Regression, Decision Trees, and Random Forest, due to their ability to handle categorical analysis. The remaining models we've discussed in class, multiple linear regression (MLR) and k-means clustering, were excluded since MLR can only be used for continuous target values. K means was excluded since our data contains a target variable and we do not have a reason to run an unsupervised model.

Model I: Logistic regression

Method: Since mode, time signature, key, and decade are categorical variables, we factored these variables. We re-coded mode 0 as minor and 1 as major and left the base level as minor. The default base level for time signature was set as 4, since 4/4 time is one of the most common time signatures when it comes to music. For key, we converted the numerical values from the source data (0 to 11), to their corresponding keys where 0 = C. The default base level of 0 = C was kept.

We executed 3 regression runs. The first run contained all the data over the 6 decades. The second run comprised data over the last 3 decades i.e. 2010s, 2000s and 1990s. Our final run contained data over the earlier 3 decades i.e. 1960s, 1970s and 1980s. After running the model on the validation set, we used the validation set to plot the ROC curve, showing the tradeoff between sensitivity and specificity. Additionally, we used the coords() function to determine the value for the best threshold represented on the ROC curve as the point furthest away from the line of no discrimination (appendix 2.1).

For Dataset containing songs of all Decades (1960-2019):

Reviewing the results of the logistic regression (appendix 2.2) and p values, danceability, energy, loudness, mode, speechiness, acousticness, and instrumentalness were most significant predictors because their p-values were the smallest. Other significant variables include valence, tempo, and time signature. Hit songs are more likely to have higher danceability, valence, loudness, and are faster. Additionally, hit songs are more likely to be songs with 4 beats/measure, major key, and in keys C#, D#, F, F#, G#, A#, and B compared to songs in the key C. Correspondingly, songs with higher energy, speechiness, acousticness, and instrumentalness tend to be flops.

For Dataset containing songs of the past 3 decades (1990-2019) :

Reviewing the results of the logistic regression (appendix 2.3), danceability, energy, loudness, mode, speechiness, acousticness, liveness, valence, time signature, and instrumentalness were most significant

predictors. Hit songs are more likely to have higher danceability, loudness, and instrumentalness. Additionally, hit songs are more likely to be songs with 4 beats/measure and in major key. Correspondingly, songs with higher energy, acousticness, liveness, valence, and time signature (3) tend to be flops.

For Dataset containing songs of the earliest 3 decades (1960-1989):

Reviewing the results of the logistic regression (appendix 2.5) and p values, danceability, loudness, mode, tempo, speechiness, acousticness, time signature, instrumentalness were most significant predictors because their p-values were the smallest. Hit songs are more likely to have higher danceability, loudness, and are faster. Additionally, hit songs are more likely to be songs with 4 beats/measure and major key. Correspondingly, songs with higher speechiness, acousticness, instrumentalness, and songs with 5 beats per bar tend to be flops.

After running confusion matrices (appendices 2.1, 2.4, 2.6) for all groups described above, we see an accuracy of 73.31% for all six decades, 79.65% for 1990-2019, 71.62% for 1960-1989.

Model II: Classification Decision Tree Model

Method: To predict whether a song is Hit or Flop, we looked at the following scenarios (all with the same target variable).

- Predictors for scenario 1: decision tree for individual decades
- Predictors for scenario 2: decision trees for 30-year-groups
- Predictors for scenario 3: decision tree for all 6 decades combined (appendix 3.1)

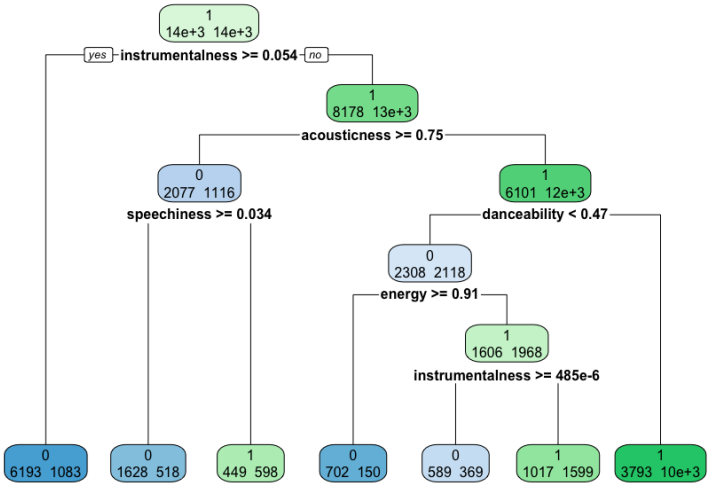
Results:

Scenario 1: Instrumentalness appears to be the most important predictor for whether a song is a hit across all decades. The accuracy of the model for each decade from 1960s to 2010s are 72.78%, 72.49%, 75.59%, 82.14%, 81.61%, and 79.48% respectively. In the 1960s and 1970s, the acousticness was the

second most important predictor. In the 1980s and 1990s, the duration of the song became an important indicator. In the 2000s and 2010s, loudness, and danceability have become more important predictors

Scenario 2: When we compare the datasets in 30-year-groups, the accuracy for 1960s-1980s is 71.55%, and the most important variables are instrumentalness and acousticness. The accuracy for 1990s-2010s is 79.26%, and the most important predictors are instrumentalness and danceability.

Figure 2.1



For Scenario 3: Figure 2.1 shows the classification tree for songs from 1960s-2010s. For songs in the past 6 decades, instrumentalness seems to be the most relevant predictor, with acousticness closely following. Thus, hit songs tend to have a instrumentalness less than .054 and acousticness less than .75. Instrumentalness measures whether a track contains no vocals. The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks. We can interpret that songs that have more vocals are more likely to be a hit song. Qualities of what makes a song a hit or a flop might have changed throughout the decades. For example, danceability is valued more for songs made in more recent decades.

Model III: Random Forest

Since a decision tree can encounter over-fitting issues and it can also ignore explanatory predictors when there's a small sample size and a large p-value, we decided to create and test a random forest model.

Method: For this model, we didn't have to partition the dataset into training and validation since the random forest picks a sample with replacement from the dataset as its training set for every individual tree. The records that are not used to build a tree are referred to as Out-Of-Bag (OOB), and are going to be used as our validation dataset. At first, we ran the model with the default values (number of trees 500, variables to consider at each splitting node 4) for the whole database and got an accuracy of 79.25%. We hypothesized that different tastes for music over time could be undermining the effects of the predictors on the target variable (the characteristics that made a hit in the 60s might be different from those in the 2010s). Thus, we used the audio features for the most recent three decades (1990-2010) as they have similar qualities and high combined accuracies. To determine the number of trees, we plotted the Out-of-Bag errors by the number of trees considered in the model (appendix 4.1) and found that 500 trees seem to be a proper number of trees since we can observe convergence, thus the errors are not decreasing anymore in a meaningful way by adding more trees. To determine the number of variables to consider at each splitting node, we computed the OOB errors for 1-10 variables (appendix 4.2). Considering 3 variables was optimal since it brings the lowest OOB error.

Results: A confusion matrix is provided in (appendix 4.3), we can observe that there was a 79.61% of accuracy in predicting flops and 88.97% of accuracy in predicting hits. The overall OOB error for this model is 15.71% (84.29% accuracy). As seen in (appendix 4.4), instrumentalness is the most important factor, followed by danceability and acousticness. Time_signature, mode, and chorus_hit seem to be the least important factors.

Modeling Conclusions: A song that has more vocals (less instrumentals), is more suitable for dancing (temp, rhythm, beat, regularity), and is less acoustic tends to produce a hit. From all three models we can conclude that instrumentalness has the highest hit song predictive power followed by danceability and acousticness. There is some variability throughout the decades. For example, acousticness was more

important between 1960s to 1980s. And danceability and loudness have become more important in the past 30 years. We found that the accuracy was higher when we group the songs in 30-year groups. Random forest has the highest prediction accuracy of 84.29%, with classification trees at 79.26%, and classification with logistic regression at 79.65%.

Results and Insights

Who can benefit from the data?

- Musicians: using predictive analysis, they can factor hit audio attributes (such as more vocals or more deancible) into their songs to help increase their popularity. They could also try to avoid adding elements that are more likely to predict a flop
- Record labels: filter through artists and try to only sign artists who use specific predicted hit song audio features more frequently.
- Historical musicologist: Musicology is the scholarly analysis and research-based study of music.They can make assessments of how "hit" song qualities have changed over the decades. It could give insights into values during that time period, political movements, or just overall trends of popular music genres/audio features of a song over time.

Other Relevant Data Needed & Future Improvements

While the insights we obtained are insightful, understanding what makes a song a hit based on only the audio features does not give us a big enough lens to accurately predict its popularity. Even if there were highly accurate guidelines for audio features that made a song popular, there are still additional relevant factors needed to actually produce a hit song. If an artist produces a song with predicted hit audio features, it is not guaranteed that song will become a hit. In reality, factors such as genre, monthly listeners, lyrics, number of albums produced, and number of previous hit songs, all play a deciding factor in the success of a song. For example, analyzing lyrics, overall sentiment, amount of vocals, and

synthetic sounds used could help the artists gain a better idea of the song's composition. Another example would be if Adele releases a new song, it will most likely become a hit song solely because she has an already established fan base, has other top hit songs, and writes lyrics that resonate with her audience. Thus even if she produces a song with less popular audio features, say a song that is more instrumental and has less vocals, the song might still become a hit. Therefore, we need this type of data to be able to conclusively say what makes a song a hit.

However, even if we were able to gather all these features above, it would be quite hard to analyze. For example it would be hard to confidently say that only specific artists will produce a hit, there are more factors to consider.

During our research our team did try to obtain this additional information from other datasets, but when we combined it with our original data set, we lost a significant portion of our data. This was largely due to different data sets having different lists of hit songs or having multiple songs with the same song title. Thus, obtaining this specific data and ensuring all the different data sets have the same list of hit songs could be another limitation to this type of project.

Appendix

References

Data Source: https://www.kaggle.com/theoverman/the-spotify-hit-predictor-dataset

Data source with genre: https://www.kaggle.com/vicsuperman/prediction-of-music-genre

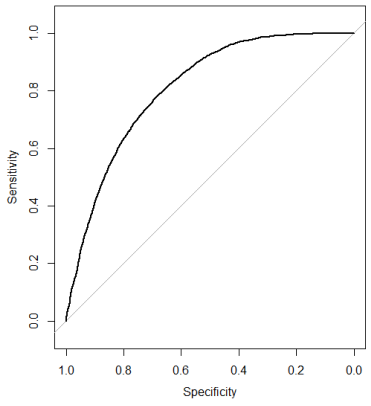
Visualizations

Appendix 1.1: Averages of Characteristics Across Decades

Danceability	Energy	Key	Loudness	Mode	Speechiness	Acousticness	Instrumentalness	Liveness	Valence	Tempo	Duration	Time Signature	Chorus Hit	Sections	Decade
0.5305823	0.5111524	5.140014	-10.449623	0.8389262	0.04931604	0.5367900	0.050243434	0.2191625	0.6634241	118.7427	165681.8	3.841240	38.68984	8.288822	1960
0.5703700	0.5848804	5.134690	-10.456478	0.7638424	0.05314100	0.3410477	0.048342900	0.1932934	0.6648517	119.7044	233918.1	3.940252	39.37795	10.473088	1970
0.6244435	0.6530307	5.289519	-9.876752	0.7084540	0.04663466	0.2045788	0.028438554	0.1826191	0.6570605	121.8678	257119.8	3.971917	39.28731	11.336422	1980
0.6487688	0.6575423	5.450725	-8.522604	0.6608696	0.07815243	0.1695109	0.029813553	0.1793557	0.5816171	117.8067	263748.3	3.962681	40.01892	11.328623	1990
0.6300490	0.7146076	5.260559	-5.677296	0.6910763	0.09778471	0.1480076	0.008858022	0.1839072	0.5537908	120.2836	238862.9	3.970027	38.84372	10.476839	2000
0.6416333	0.6823673	5.303532	-5.875753	0.6608315	0.10752441	0.1621728	0.006057000	0.1862975	0.4938130	123.5246	220502.7	3.979056	39.35369	9.888403	2010

Data Modeling

Appendix 2.1 : ROC Curve and Confusion Matrix, all decades




```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 4210 1320
         1 1972 4830

               Accuracy : 0.7331
                 95% CI : (0.7251, 0.7408)
    No Information Rate : 0.5013
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.4662

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.7854
            Specificity : 0.6810
         Pos Pred Value : 0.7101
         Neg Pred Value : 0.7613
             Prevalence : 0.4987
         Detection Rate : 0.3917
   Detection Prevalence : 0.5516
      Balanced Accuracy : 0.7332

       'Positive' Class : 1
```

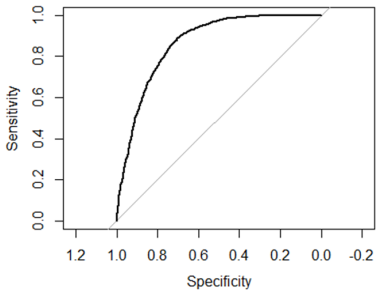
Appendix 2.2 : Logistic Regression, all predictors, all decades

```
Call:
glm(formula = target ~ ., family = "binomial", data = train.df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4788  -0.9454   0.3644   0.8730   3.1622

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       1.025e+00  1.618e-01   6.333 2.40e-10 ***
danceability      3.198e+00  1.114e-01  28.705  < 2e-16 ***
energy           -1.891e+00  1.202e-01 -15.733  < 2e-16 ***
keyC#             1.995e-01  6.460e-02   3.088 0.002012 **
keyD             -1.459e-01  5.730e-02  -2.546 0.010888 *
keyD#             2.318e-01  8.672e-02   2.673 0.007507 **
keyE             -2.949e-02  6.294e-02  -0.468 0.639459
keyF              1.364e-01  6.152e-02   2.218 0.026562 *
keyF#             1.755e-01  7.323e-02   2.397 0.016538 *
keyG             -5.475e-02  5.612e-02  -0.976 0.329248
keyG#             2.266e-01  7.124e-02   3.181 0.001468 **
keyA             -1.778e-02  5.795e-02  -0.307 0.758923
keyA#             3.762e-01  6.834e-02   5.505 3.70e-08 ***
keyB              7.235e-02  6.740e-02   1.073 0.283060
loudness          1.019e-01  5.289e-03  19.272  < 2e-16 ***
modeMajor         4.245e-01  3.203e-02  13.254  < 2e-16 ***
speechiness      -3.186e+00  1.879e-01 -16.957  < 2e-16 ***
acousticness     -1.404e+00  6.358e-02 -22.087  < 2e-16 ***
instrumentalness -3.394e+00  8.041e-02 -42.208  < 2e-16 ***
liveness         -2.035e-01  8.279e-02  -2.458 0.013957 *
valence           3.681e-01  7.210e-02   5.106 3.30e-07 ***
tempo             1.939e-03  5.098e-04   3.804 0.000142 ***
duration_ms      -8.316e-07  3.258e-07  -2.552 0.010698 *
time_signature0   1.452e+00  1.276e+00   1.138 0.255084
time_signature1  -7.782e-01  1.854e-01  -4.197 2.71e-05 ***
time_signature3  -2.031e-01  5.369e-02  -3.782 0.000156 ***
time_signature5  -3.405e-01  1.399e-01  -2.433 0.014963 *
chorus_hit       -2.431e-03  8.391e-04  -2.897 0.003771 **
sections         -2.151e-04  7.724e-03  -0.028 0.977779
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 39889  on 28773  degrees of freedom
Residual deviance: 30335  on 28745  degrees of freedom
AIC: 30393

Number of Fisher Scoring iterations: 5
```

Appendix 2.3: Spotify Merged Dataset (2010s, 2000s, 1990s) :

```
Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       2.491e+00  2.920e-01   8.531  < 2e-16 ***
danceability      4.982e+00  1.923e-01  25.900  < 2e-16 ***
energy           -3.974e+00  2.132e-01 -18.635  < 2e-16 ***
keyC#             1.330e-01  1.013e-01   1.313 0.189144
keyD             -2.286e-01  1.014e-01  -2.254 0.024212 *
keyD#             2.790e-01  1.488e-01   1.874 0.060881 .
keyE              1.327e-01  1.081e-01   1.227 0.219804
keyF              1.595e-01  1.088e-01   1.466 0.142777
keyF#             2.351e-01  1.116e-01   2.106 0.035201 *
keyG              1.532e-01  9.801e-02   1.563 0.118135
keyG#             3.273e-01  1.183e-01   2.768 0.005638 **
keyA              9.216e-03  1.024e-01   0.090 0.928282
keyA#             3.769e-01  1.186e-01   3.178 0.001482 **
keyB              3.700e-02  1.081e-01   0.342 0.732192
loudness          2.350e-01  1.095e-02  21.457  < 2e-16 ***
modeMajor         3.246e-01  5.238e-02   6.197 5.74e-10 ***
speechiness      -7.643e-01  2.616e-01  -2.921 0.003490 **
acousticness     -2.191e+00  1.262e-01 -17.363  < 2e-16 ***
instrumentalness -5.079e+00  1.966e-01 -25.840  < 2e-16 ***
liveness         -6.811e-01  1.443e-01  -4.721 2.35e-06 ***
valence          -4.659e-01  1.226e-01  -3.801 0.000144 ***
tempo             1.894e-03  8.291e-04   2.285 0.022335 *
duration_ms       9.527e-07  5.432e-07   1.754 0.079475 .
time_signature0  -6.626e+00  1.329e+02  -0.050 0.960252
time_signature1  -5.891e-01  3.251e-01  -1.812 0.070006 .
time_signature3  -4.926e-01  1.100e-01  -4.477 7.57e-06 ***
time_signature5  -9.942e-03  2.171e-01  -0.046 0.963476
chorus_hit       -2.791e-03  1.354e-03  -2.062 0.039253 *
sections         -1.464e-02  1.242e-02  -1.178 0.238684
---
```

Appendix 2.4: ROC Curve, Best Threshold Value, and Confusion Matrix (2010s, 2000s, 1990s)

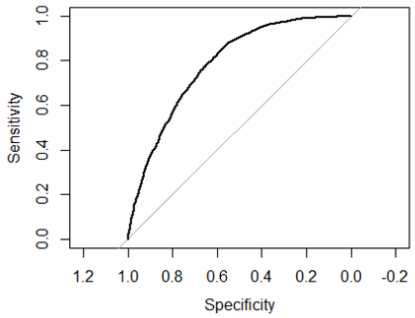



```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1858  304
         1  782 2393

               Accuracy : 0.7965
                 95% CI : (0.7855, 0.8072)
    No Information Rate : 0.5053
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.5922

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.8873
            Specificity : 0.7038
         Pos Pred Value : 0.7537
         Neg Pred Value : 0.8594
             Prevalence : 0.5053
         Detection Rate : 0.4484
   Detection Prevalence : 0.5949
      Balanced Accuracy : 0.7955

       'Positive' Class : 1

> |
```

Appendix 2.5: Spotify Merged Dataset (60s, 70s and 80s) :

```
Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)       3.378e-01  2.069e-01   1.633  0.10244
danceability      2.522e+00  1.536e-01  16.423  < 2e-16 ***
energy           -2.251e-01  1.584e-01  -1.422  0.15513
keyC#             2.560e-01  9.165e-02   2.793  0.00522 **
keyD             -2.286e-01  7.173e-02  -3.186  0.00144 **
keyD#             1.586e-01  1.091e-01   1.454  0.14595
keyE             -1.393e-01  8.078e-02  -1.724  0.08472 .
keyF             -1.710e-03  7.731e-02  -0.022  0.98236
keyF#             1.420e-01  1.073e-01   1.324  0.18539
keyG             -1.838e-01  7.079e-02  -2.596  0.00943 **
keyG#             7.303e-03  9.467e-02   0.077  0.93850
keyA             -1.083e-01  7.313e-02  -1.481  0.13864
keyA#             2.814e-01  8.817e-02   3.191  0.00142 **
keyB              1.945e-01  9.206e-02   2.113  0.03459 *
loudness          6.435e-02  6.810e-03   9.450  < 2e-16 ***
modeMajor         5.380e-01  4.336e-02  12.408  < 2e-16 ***
speechiness      -9.305e+00  4.682e-01 -19.874  < 2e-16 ***
acousticness     -1.147e+00  7.804e-02 -14.695  < 2e-16 ***
instrumentalness -2.808e+00  9.122e-02 -30.784  < 2e-16 ***
liveness          2.049e-03  1.072e-01   0.019  0.98475
valence           2.294e-01  1.033e-01   2.221  0.02638 *
tempo             2.977e-03  6.921e-04   4.302 1.69e-05 ***
duration_ms      -7.040e-07  4.372e-07  -1.610  0.10734
time_signature0   1.159e+01  1.195e+02   0.097  0.92274
time_signature1  -6.549e-01  2.131e-01  -3.073  0.00212 **
time_signature3  -1.099e-01  6.360e-02  -1.729  0.08388 .
time_signature5  -1.013e+00  2.216e-01  -4.574 4.78e-06 ***
chorus_hit        4.629e-04  1.146e-03   0.404  0.68625
sections         -1.645e-02  1.048e-02  -1.569  0.11664
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix 2.6 : ROC Curve, Best Threshold Value, and Confusion Matrix (60s, 70s and 80s)




```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 2243  749
         1 1236 2767

               Accuracy : 0.7162
                 95% CI : (0.7055, 0.7268)
    No Information Rate : 0.5026
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.432

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.7870
            Specificity : 0.6447
         Pos Pred Value : 0.6912
         Neg Pred Value : 0.7497
             Prevalence : 0.5026
         Detection Rate : 0.3956
   Detection Prevalence : 0.5723
      Balanced Accuracy : 0.7158

       'Positive' Class : 1
```

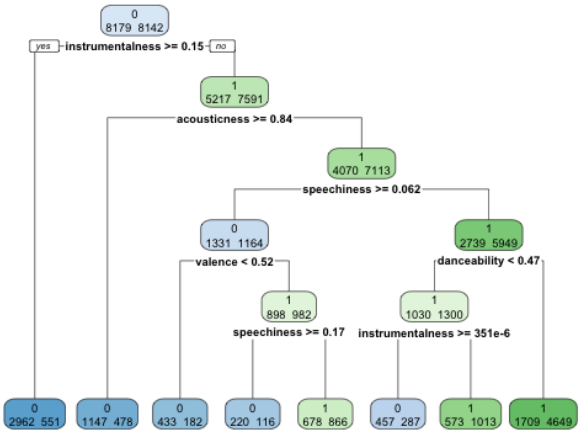
```
> coords(r, x = "best")
  threshold specificity sensitivity
1 0.5039852   0.6447255   0.7869738
> |
```

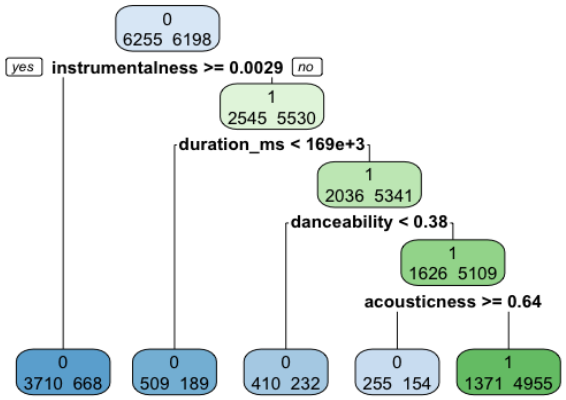
Appendix 3.1: Confusion Matrix and Variable Importance

```
          Reference
Prediction    0    1
         0 3914  884
         1 2268 5266

               Accuracy : 0.7444
                 95% CI : (0.7366, 0.7521)
    No Information Rate : 0.5013
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.4891

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.6331
            Specificity : 0.8563
         Pos Pred Value : 0.8158
         Neg Pred Value : 0.6990
             Prevalence : 0.5013
         Detection Rate : 0.3174
   Detection Prevalence : 0.3891
      Balanced Accuracy : 0.7447

       'Positive' Class : 0
```

instrumentalness	acousticness	danceability	energy	loudness	valence
2481.6749961	823.7302141	663.2068066	510.7754648	349.6834610	324.8941983
speechiness	tempo	time_signature	duration_ms	sections	
169.8124770	26.0111575	14.0094306	1.3803325	0.7481088	

Appendix 3.2: Classification Tree for 1960s-1980s, 1990s-2010s.

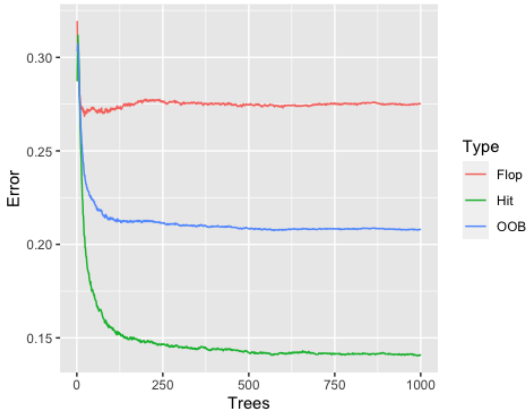





```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 2213  724
         1 1266 2792

               Accuracy : 0.7155
                 95% CI : (0.7048, 0.7261)
    No Information Rate : 0.5026
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.4305

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.6361
            Specificity : 0.7941
         Pos Pred Value : 0.7535
         Neg Pred Value : 0.6880
             Prevalence : 0.4974
         Detection Rate : 0.3164
   Detection Prevalence : 0.4199
      Balanced Accuracy : 0.7151

       'Positive' Class : 0
```

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 2078  545
         1  562 2152

               Accuracy : 0.7926
                 95% CI : (0.7814, 0.8034)
    No Information Rate : 0.5053
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.5851

 Mcnemar's Test P-Value : 0.6306

            Sensitivity : 0.7871
            Specificity : 0.7979
         Pos Pred Value : 0.7922
         Neg Pred Value : 0.7929
             Prevalence : 0.4947
         Detection Rate : 0.3894
   Detection Prevalence : 0.4915
      Balanced Accuracy : 0.7925

       'Positive' Class : 0
```

Appendix 4.1: OOB Error by Number of Trees



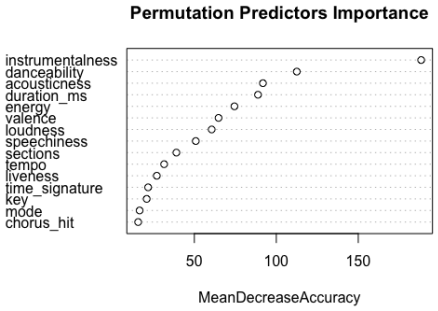
Appendix 4.2: OOB Errors When Considering 1 Variable to 10 Variables

Number of variables considered per split	OOB Error
1	0.1685779
2	0.1585722
3	0.1570545
4	0.1574480
5	0.1579593
6	0.1586847
7	0.1599775
8	0.1586847
9	0.1590781
10	0.160371

Appendix 4.3 : Confusion Matrix

```
 randomForest(x = predictorsf, y = target.900010, ntree = 500,      mtry = 3)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 3

        OOB estimate of  error rate: 15.71%
Confusion matrix:
     0    1 class.error
0 7082 1813   0.2038224
1  981 7914   0.1102867
```

Appendix 4.4 : Importance of Predictors



For interpretation purposes, the way these values are determined is by permuting all the OOB values from a predictor (so you break the real relationship with the target) and then try them with each tree. The bigger the difference between the number of correct predictions for the original values and the permuted ones, the more important the predictor is. This implies that the higher the Mean Decrease Accuracy (MDA), the more important.