# "HITNESS": WHAT DEFINES A HIT SONG?


By: Jae, Wendy, Sudhish, Hagen

# ABSTRACT

Why do some songs do well and other's don't?

# DATA COLLECTION

| | hitness | danceability | energy | key | loudness | mod |
|---|---|---|---|---|---|---|
| 1 | TRUE | 0.683 | 0.3750 | 0 | -13.056 | 1 |
| 2 | TRUE | 0.451 | 0.2400 | 1 | -14.014 | 1 |
| 3 | TRUE | 0.724 | 0.2760 | 7 | -12.250 | 1 |
| 4 | FALSE | 0.339 | 0.2140 | 4 | -11.714 | 1 |
| 5 | TRUE | 0.760 | 0.4790 | 2 | -5.574 | 1 |
| 6 | TRUE | 0.752 | 0.4880 | 6 | -7.050 | 1 |
| 7 | TRUE | 0.442 | 0.5850 | 0 | -10.332 | 0 |
| 8 | TRUE | 0.889 | 0.4960 | 4 | -6.365 | 0 |
| 9 | TRUE | 0.657 | 0.7110 | 0 | -5.355 | 1 |
| 10 | TRUE | 0.572 | 0.3850 | 7 | -6.362 | 1 |
| 11 | TRUE | 0.511 | 0.3630 | 4 | -7.650 | 0 |
| 12 | TRUE | 0.575 | 0.7580 | 1 | -5.029 | 0 |
| 13 | TRUE | 0.494 | 0.6320 | 5 | -6.890 | 1 |
| 14 | TRUE | 0.680 | 0.5780 | 10 | -5.804 | 1 |
| 15 | FALSE | 0.383 | 0.6370 | 3 | -6.993 | 1 |
| 16 | TRUE | 0.842 | 0.8010 | 8 | -4.167 | 0 |
| 17 | TRUE | 0.792 | 0.7430 | 7 | -2.806 | 1 |
| 18 | TRUE | 0.575 | 0.5710 | 1 | -7.906 | 1 |
| 19 | TRUE | 0.703 | 0.7230 | 9 | -5.450 | 0 |
| 20 | TRUE | 0.731 | 0.6570 | 0 | -8.560 | 1 |
| 21 | TRUE | 0.629 | 0.4640 | 4 | -8.720 | 1 |
| 22 | FALSE | 0.495 | 0.3900 | 0 | -7.375 | 1 |
| 23 | TRUE | 0.656 | 0.5420 | 7 | -7.358 | 1 |
| 24 | FALSE | 0.901 | 0.4640 | 5 | -9.789 | 0 |
| 25 | TRUE | 0.876 | 0.7860 | 10 | -4.884 | 0 |
| 26 | TRUE | 0.499 | 0.7970 | 2 | -3.770 | 1 |
| 27 | FALSE | 0.360 | 0.2330 | 5 | -14.716 | 0 |
| 28 | FALSE | 0.761 | 0.6730 | 1 | -5.887 | 1 |
| 29 | TRUE | 0.666 | 0.7280 | 7 | -5.808 | 1 |
| 30 | FALSE | 0.797 | 0.2930 | 5 | -7.479 | 1 |
| 31 | FALSE | 0.568 | 0.6360 | 11 | -4.265 | 1 |

BILLBOARD HOT 100

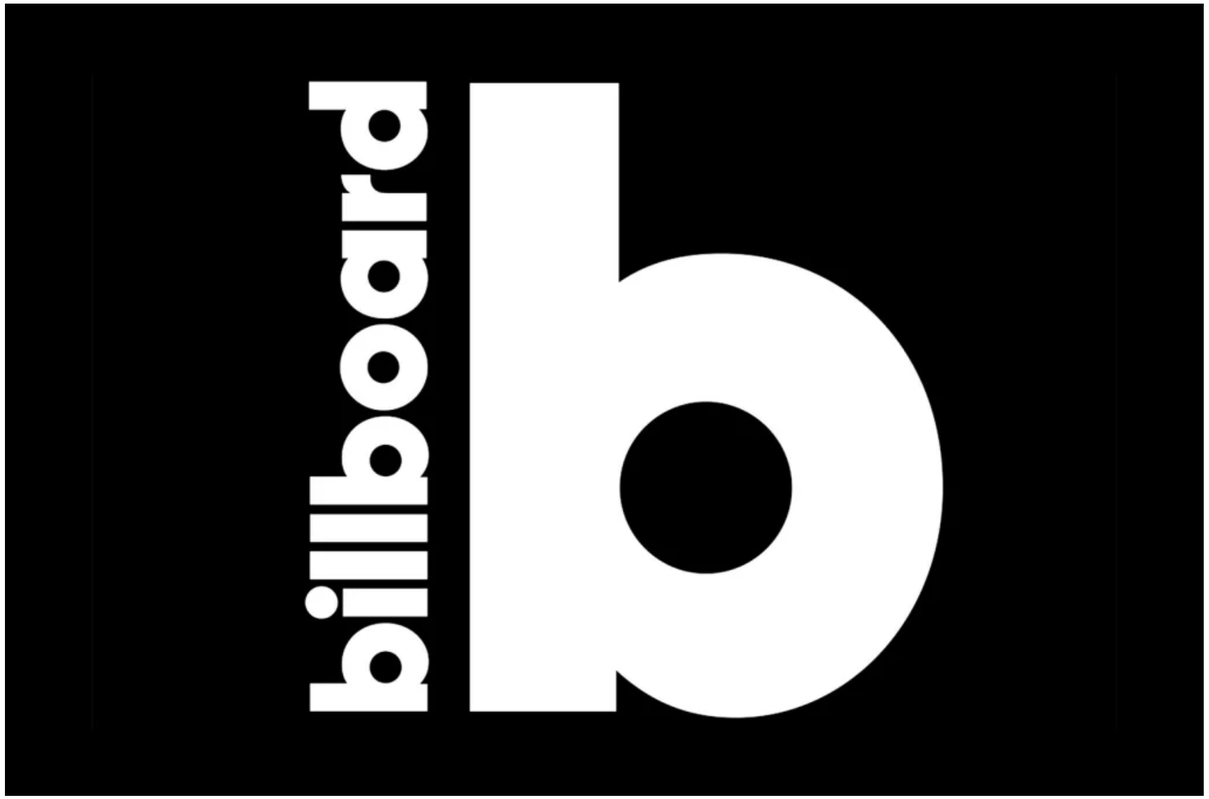
SPOTIFY

MUSICOSET

GENIUS

# BILLBOARD HOT 100

Sourced from Dhruvil Dave on Kaggle
From 1958 to 2021



# SPOTIFY AUDIO FEATURES

| | | | |
|---|---|---|---|
| acousticness | danceability | energy | key |
| liveness | loudness | mode | speechiness |
| tempo | time_signature | valence | |

# MUSICOSET

Sourced from Mariana O. Silva, Laís M. Rocha, Mirella M. Moro

From 1964 to 2018

| | artist_id | year_end_score | is_pop | year |
|---|---|---|---|---|
| 1 | 3TVXtAsR1Inumwj472S9r4 | 115282 | True | 2018 |
| 2 | 0LyfQWJT6nXafLPZqxe9Of | 78376 | True | 2018 |
| 3 | 6eUKZXaKkcviH0Ku9w2n3V | 60969 | True | 2018 |
| 4 | 7dGJo4pcD2V6oG8kP0tJRR | 57511 | True | 2018 |
| 5 | 53XhwfbYqKCa1cC15pYq2q | 56375 | True | 2018 |
| 6 | 3WrFJ7ztbogyGnTHbHJFl2 | 52849 | True | 2018 |
| 7 | 0du5cEVh5yTK9QJze8zA0C | 47598 | True | 2018 |
| 8 | 1Xyo4u8uXC1ZmMpatF05PJ | 45579 | True | 2018 |
| 9 | 2YZyLoL8N0Wb9xBt1NhZWg | 40403 | True | 2018 |
| 10 | 4YLtscXsxbVgi031ovDDdh | 39973 | True | 2018 |
| 11 | 2ye2Wgw4gimLv2eAKyk1NB | 39556 | True | 2018 |
| 12 | 4dpARuHxo51G3z768sgnrY | 38742 | True | 2018 |
| 13 | 2QsynagSdAqZj3U9HgDzjD | 38319 | True | 2018 |
| 14 | 15UsOTVnJzReFVN1VCnxy4 | 37904 | True | 2018 |

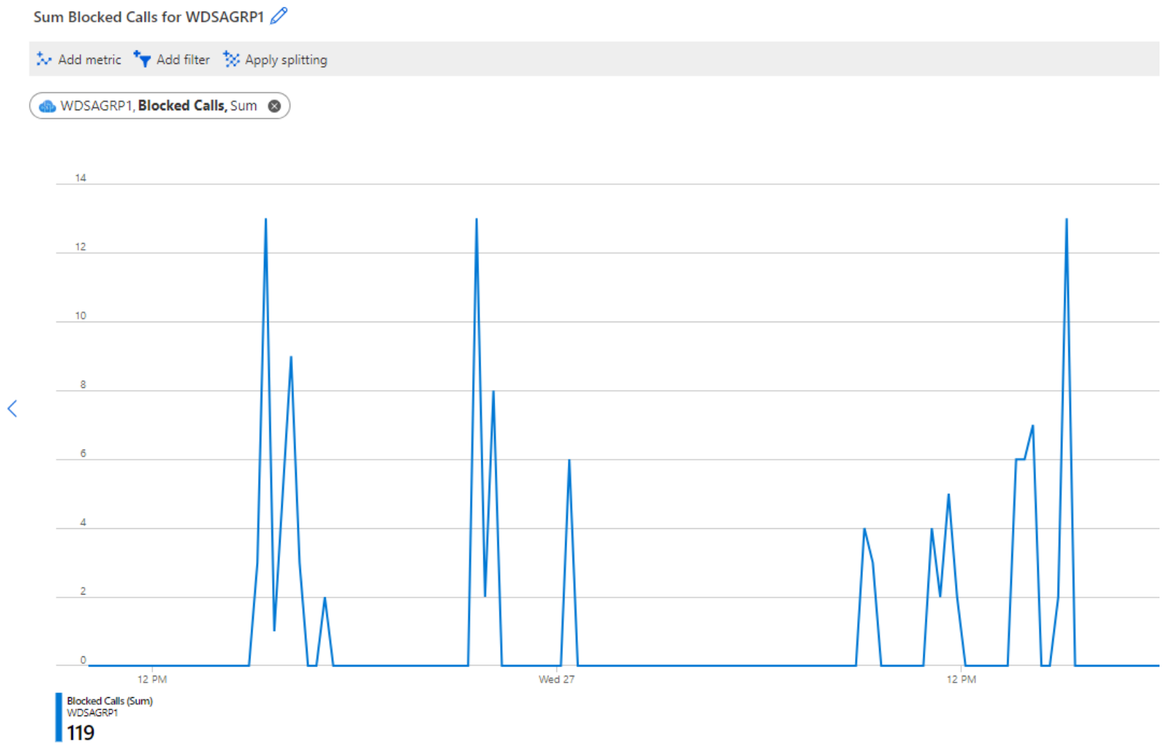
# ARTIST POPULARITY (COLLABORATIONS AND FEATURES)



# GENIUS

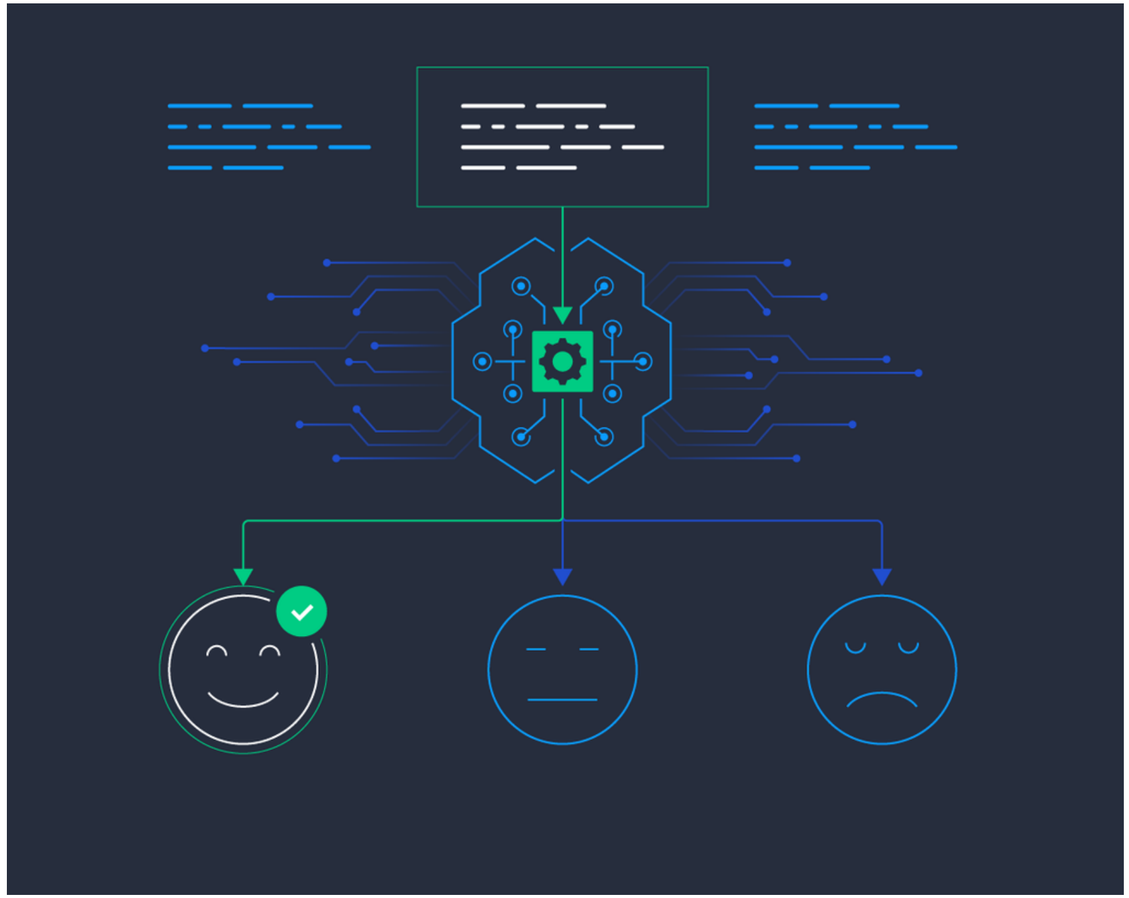
## LYRIC GATHERING

The Genius API in combination with the Python library lyricsgenius which directly utilizes HTML parser BeautifulSoup




# SENTIMENT ANALYSIS TRIALS

| TextBlob | Microsoft Azure |
| --- | --- |
| Unreliable Results | Blocked calls |



# VADER

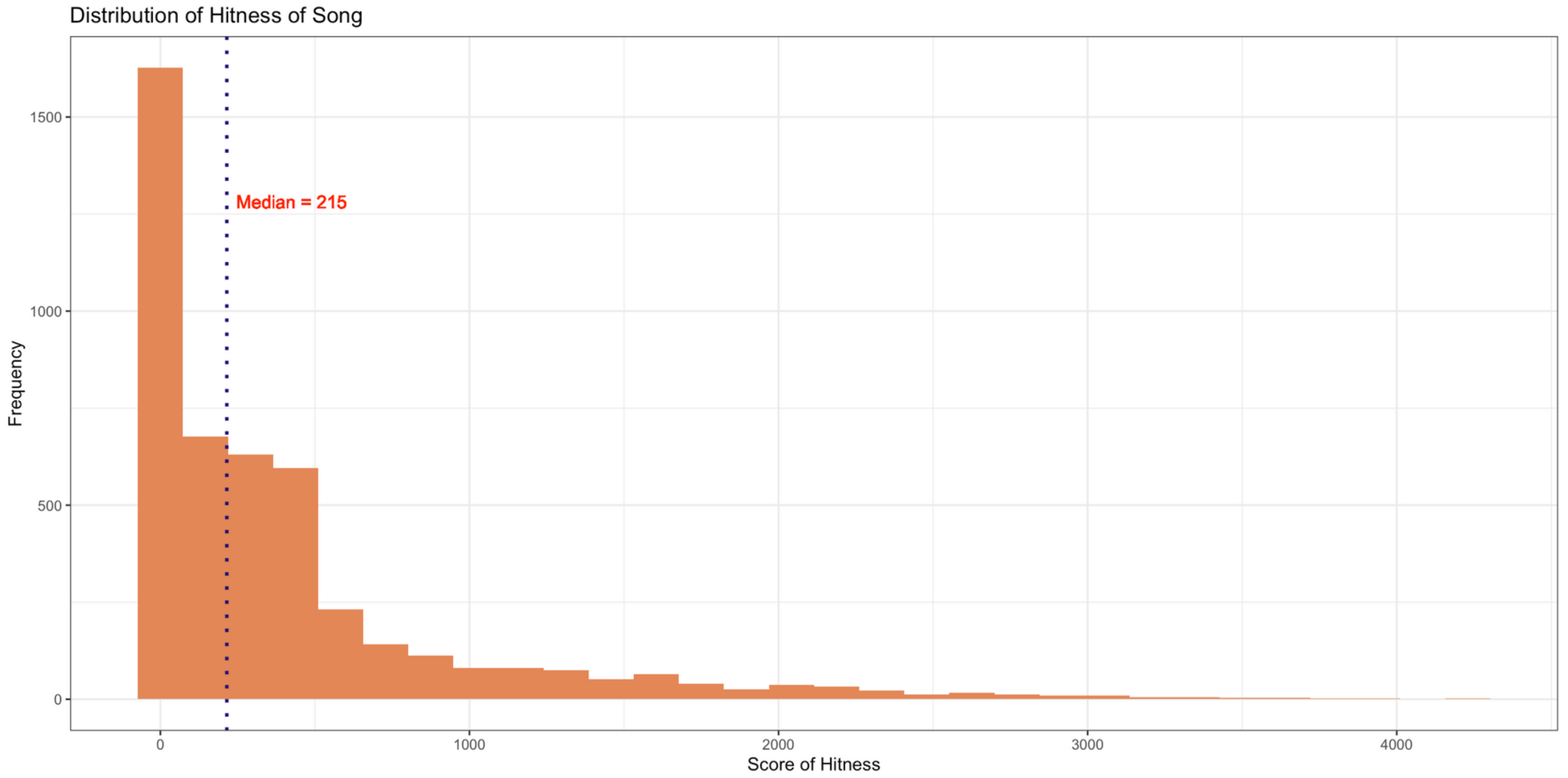
Lexicon and rule-based sentiment analysis tool

# HITNESS

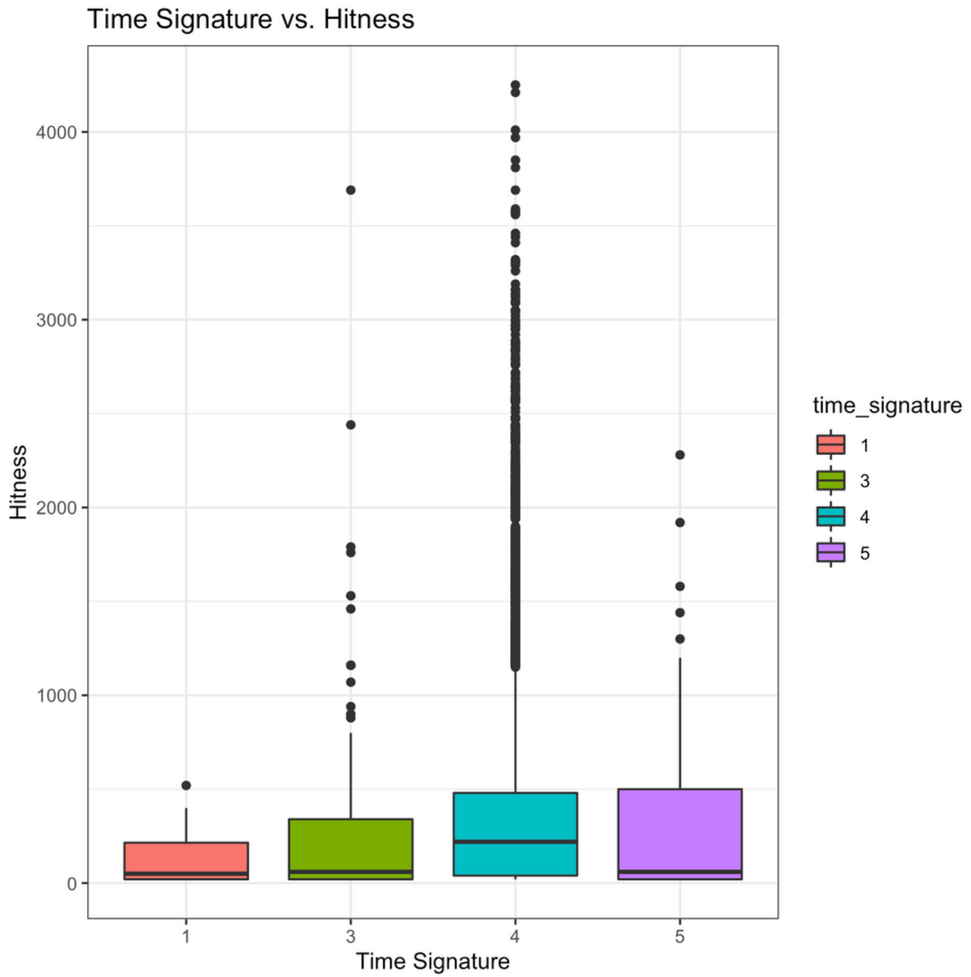

'Hitness' of a song is scores by sum of charting position for every week on Billboard Hot 100

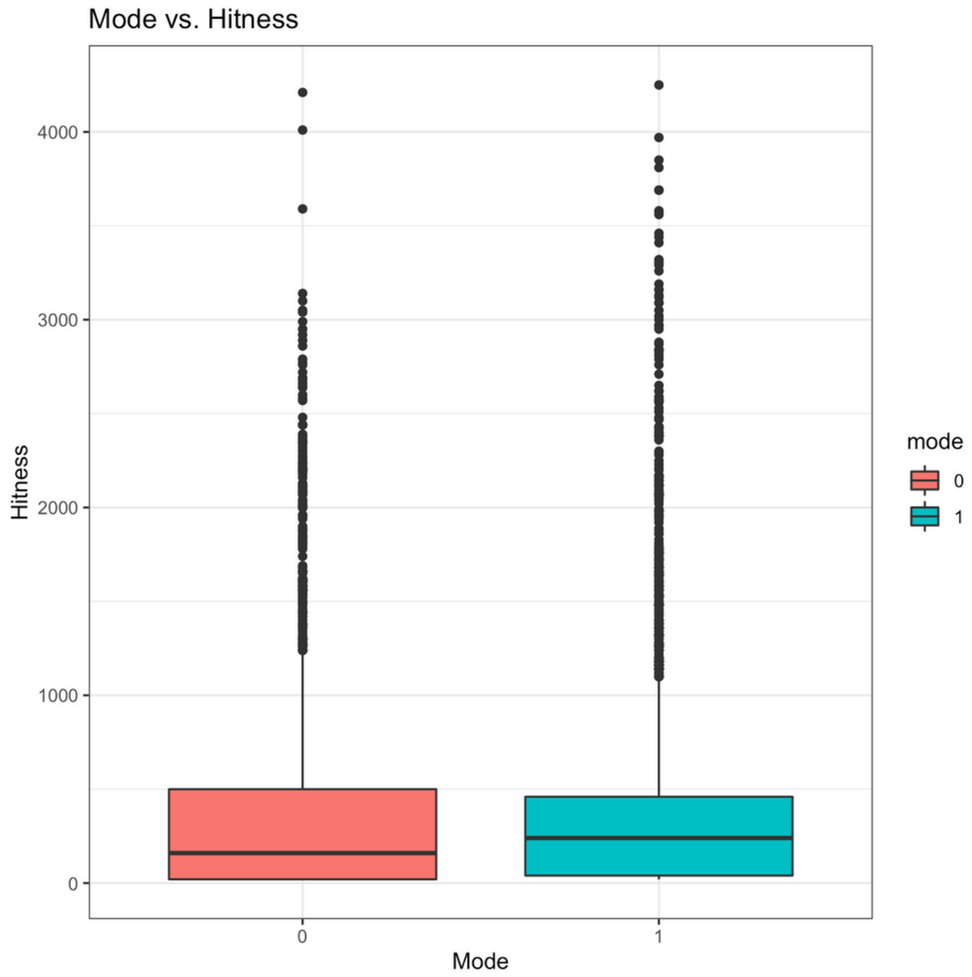
| Top 1 | Top 5 | Top 10 | Top 20 | Top 40 | Top 100 |
|---|---|---|---|---|---|
| 100 | 90 | 80 | 60 | 40 | 20 |

# EXPLORATORY DATA ANALYSIS

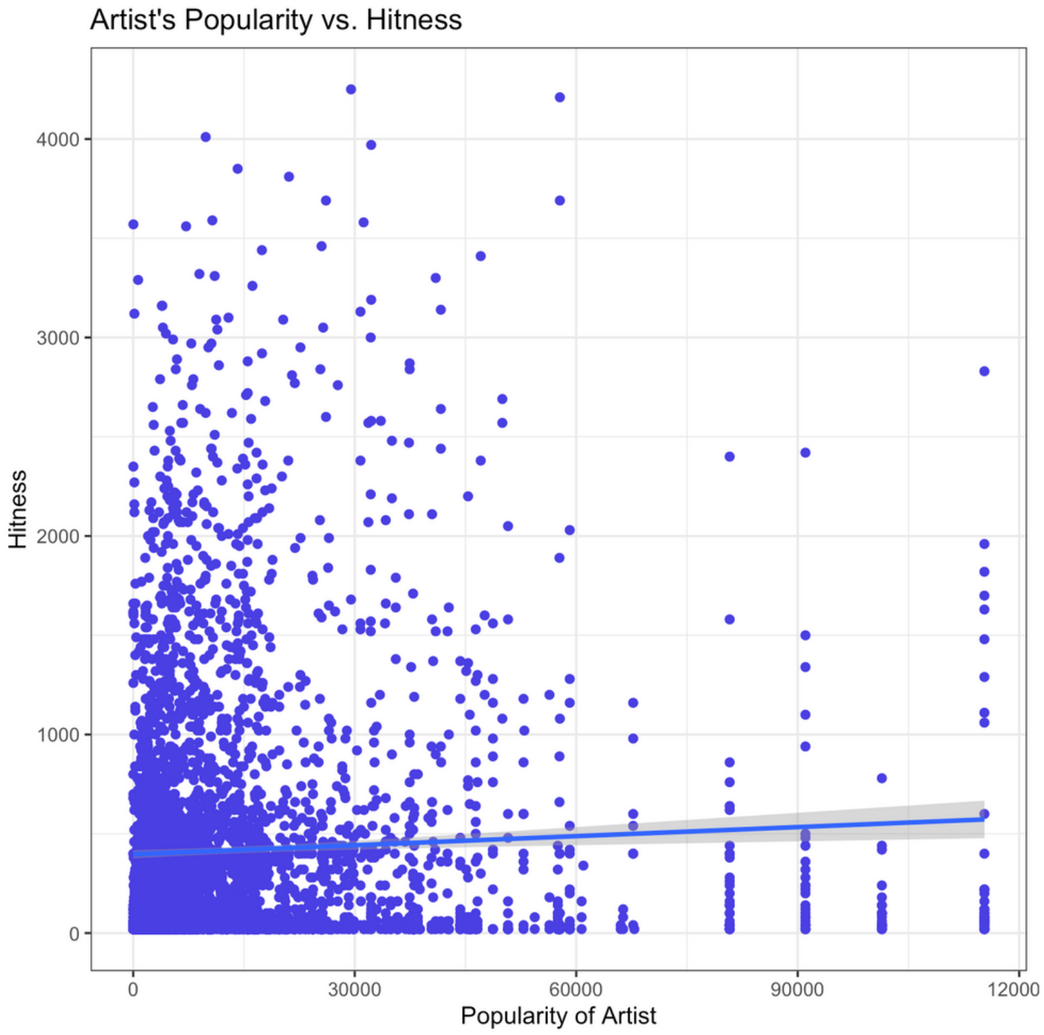


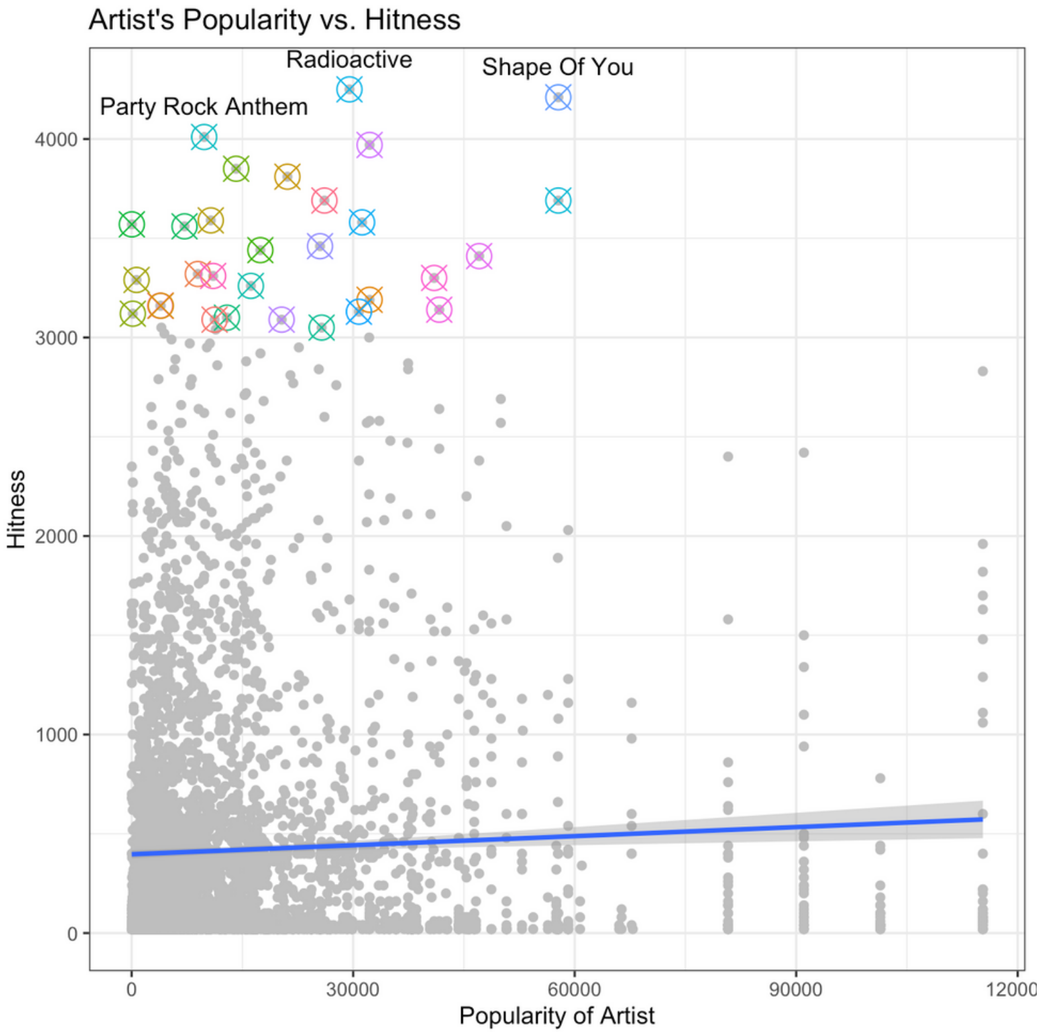
# EXPLORATORY DATA ANALYSIS



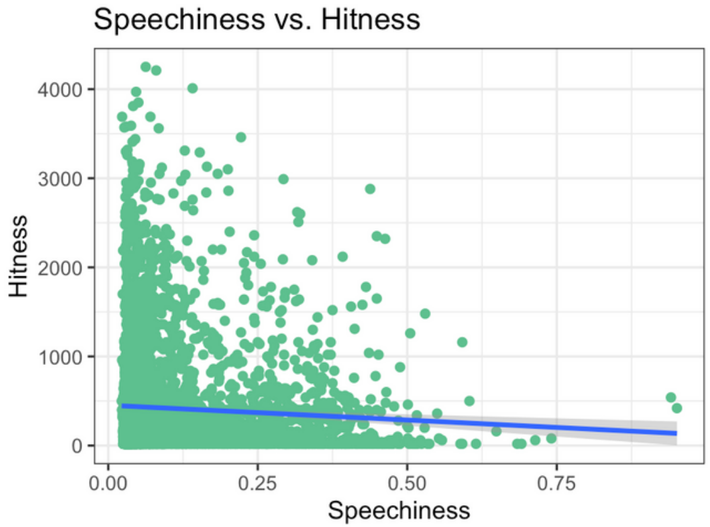



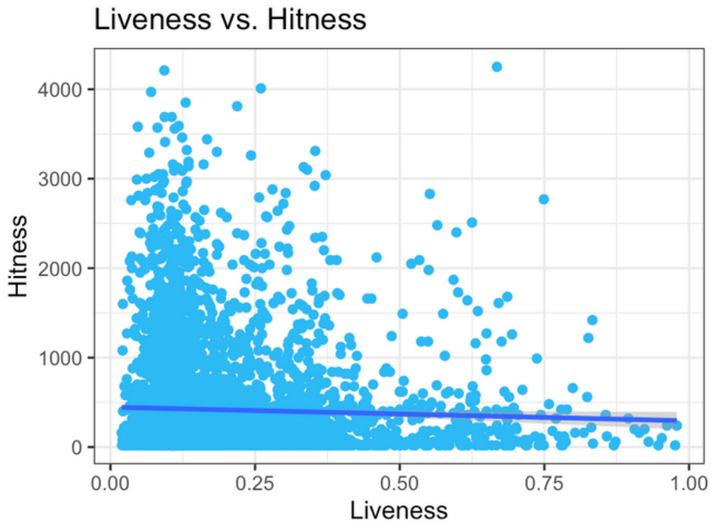
# EXPLORATORY DATA ANALYSIS

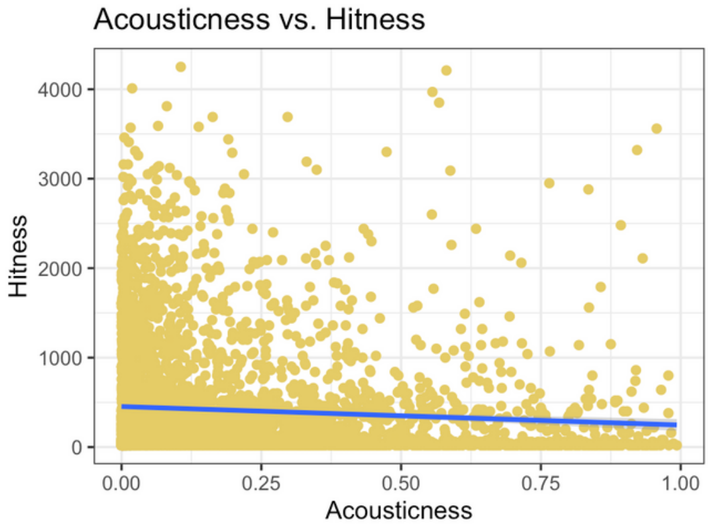


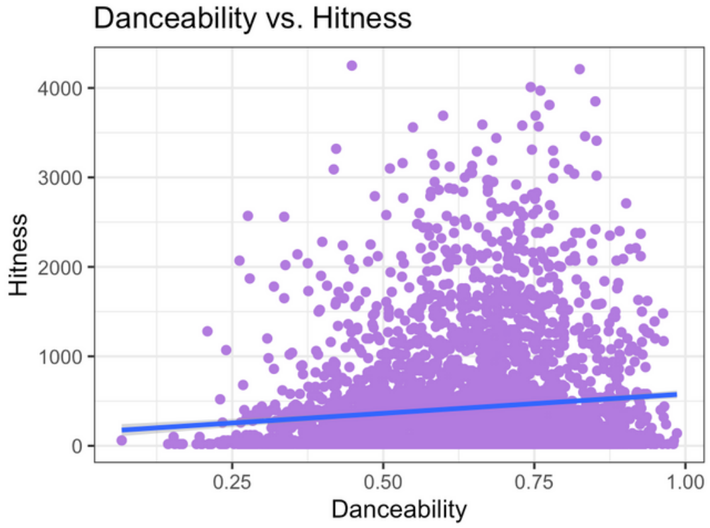



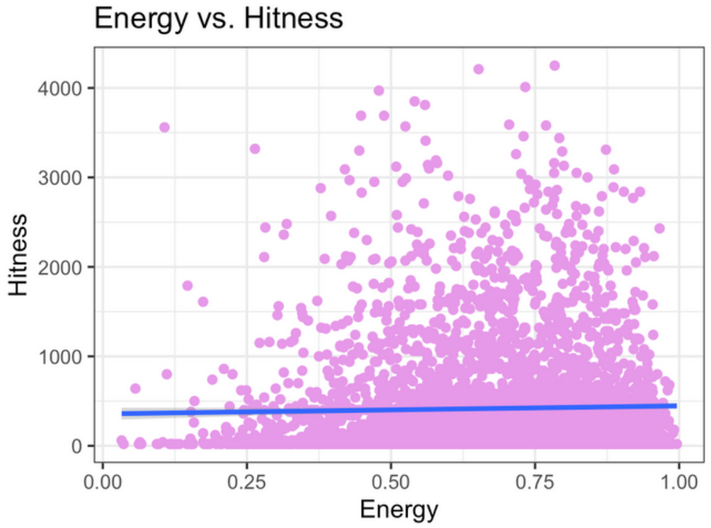
# EXPLORATORY DATA ANALYSIS

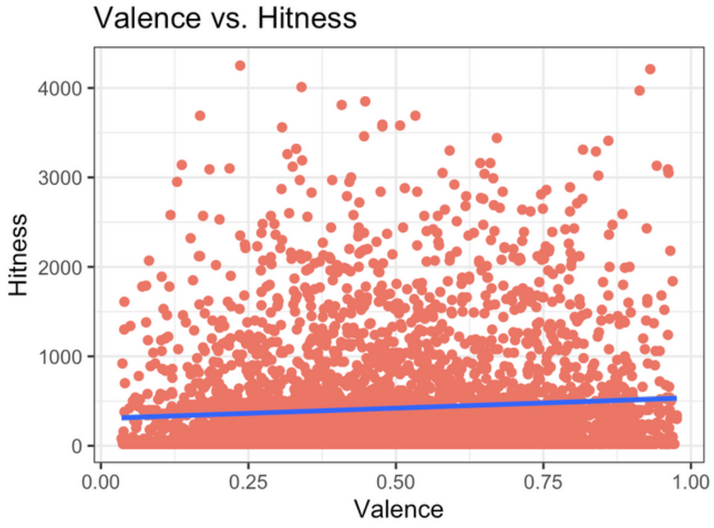


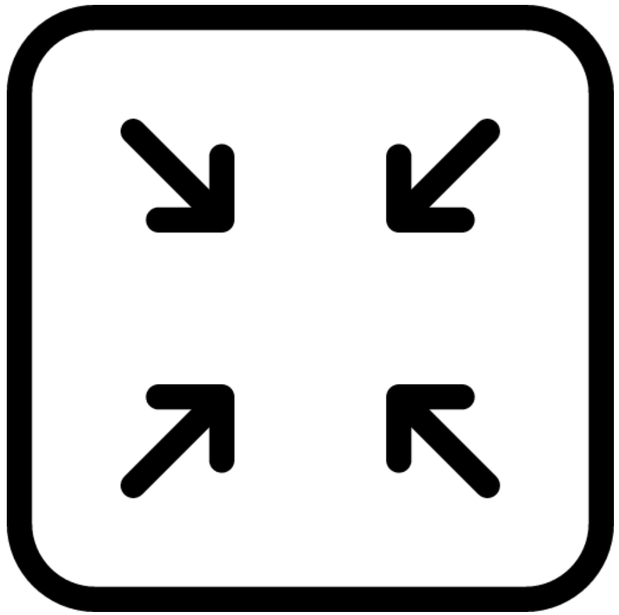





# DATA PREPARATION

# LINEAR MODEL (OLS)

Log Scaled

Anova Table (Type II tests)

Response: hitness

| | Sum Sq | Df | F value | Pr (>F) | |
|---|---|---|---|---|---|
| danceability | 1.128 | 1 | 4.8006 | 0.028662 | * |
| energy | 0.493 | 1 | 2.0993 | 0.147658 | |
| key | 2.083 | 11 | 0.8062 | 0.634036 | |
| loudness | 0.638 | 1 | 2.7155 | 0.099665 | . |
| mode | 1.023 | 1 | 4.3561 | 0.037109 | * |
| speechiness | 3.950 | 1 | 16.8139 | 0.00004431 | *** |
| acousticness | 0.804 | 1 | 3.4209 | 0.064646 | . |
| instrumentalness | 0.001 | 1 | 0.0056 | 0.940470 | |
| liveness | 0.004 | 1 | 0.0173 | 0.895270 | |
| valence | 2.388 | 1 | 10.1667 | 0.001471 | ** |
| time_signature | 0.190 | 1 | 0.8092 | 0.368546 | |
| sentiment | 0.605 | 1 | 2.5760 | 0.108784 | |
| duration_min | 0.031 | 1 | 0.1331 | 0.715317 | |
| log_tempo | 0.000 | 1 | 0.0007 | 0.979001 | |
| log_popularity | 5.481 | 1 | 23.3293 | 0.000001561 | *** |
| Residuals | 254.897 | 1085 | | | |

---

Signif. Codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ‘ 1

Testing Accuracy (%)

Base: 59.8

Max Div: 60.2

Log: 62.1

# LOGISTIC REGRESSION (GLM)

Log Scaled

Analysis of Deviance Table (Type II tests)

Response: hitness

| | LR Chisq | Df | Pr(>Chisq) | |
|---|---|---|---|---|
| danceability | 5.0156 | 1 | 0.02512 | * |
| energy | 2.1564 | 1 | 0.14197 | |
| key | 9.1753 | 11 | 0.60571 | |
| loudness | 3.0262 | 1 | 0.08193 | . |
| mode | 4.5770 | 1 | 0.03240 | * |
| speechiness | 17.4543 | 1 | 0.00002943 | *** |
| acousticness | 3.4535 | 1 | 0.06312 | . |
| instrumentalness | 0.0029 | 1 | 0.95703 | |
| liveness | 0.0201 | 1 | 0.88733 | |
| valence | 10.1927 | 1 | 0.00141 | ** |
| time_signature | 0.9852 | 1 | 0.32091 | |
| sentiment | 2.6395 | 1 | 0.10424 | |
| duration_min | 0.1400 | 1 | 0.70827 | |
| log_tempo | 0.0025 | 1 | 0.95996 | |
| log_popularity | 24.4355 | 1 | 0.0000007684 | *** |

---

Signif. Codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ‘ 1

Testing Accuracy (%)

Base: 59.9

Max Div: 60.3

Log: 62.0

# LOGISTIC LASSO REGRESSION (GLM)

Log Scaled

Analysis of Deviance Table (Type II tests)

Response: hitness

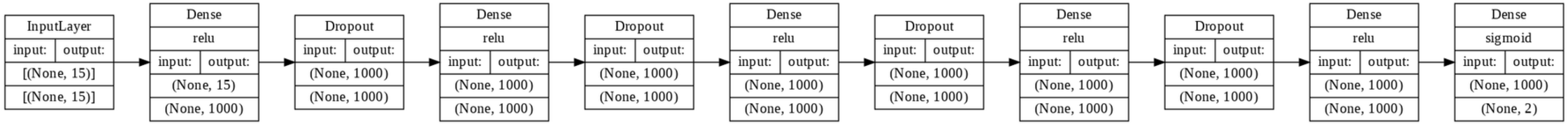
| | LR Chisq | Df | Pr (>Chisq) | |
|---|---|---|---|---|
| danceability | 6.7522 | 1 | 0.009363 | ** |
| loudness | 1.1385 | 1 | 0.285976 | |
| mode | 5.1728 | 1 | 0.022943 | * |
| speechiness | 19.7425 | 1 | 0.000008861 | *** |
| acousticness | 2.3131 | 1 | 0.128286 | |
| valence | 9.9641 | 1 | 0.001596 | ** |
| log_popularity | 26.2782 | 1 | 0.0000002956 | *** |

---

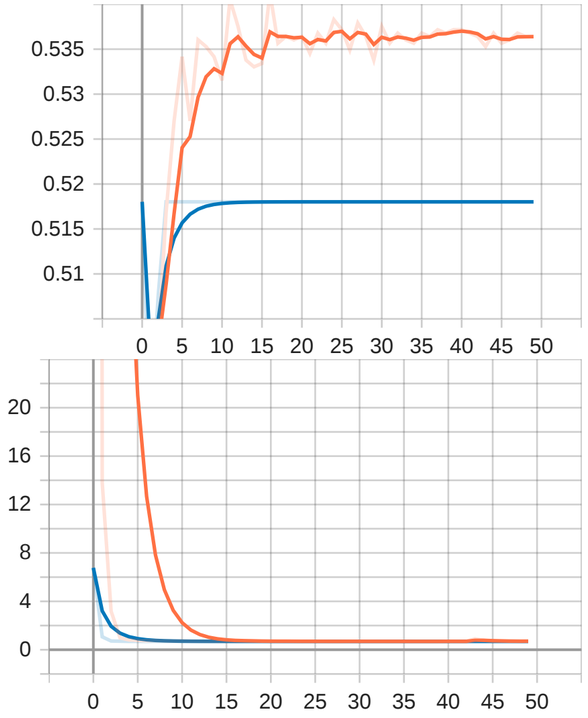
Signif. Codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ‘ 1

Testing Accuracy (%)
Base: 54.9
Max Div: 56.3
Log: 61.4

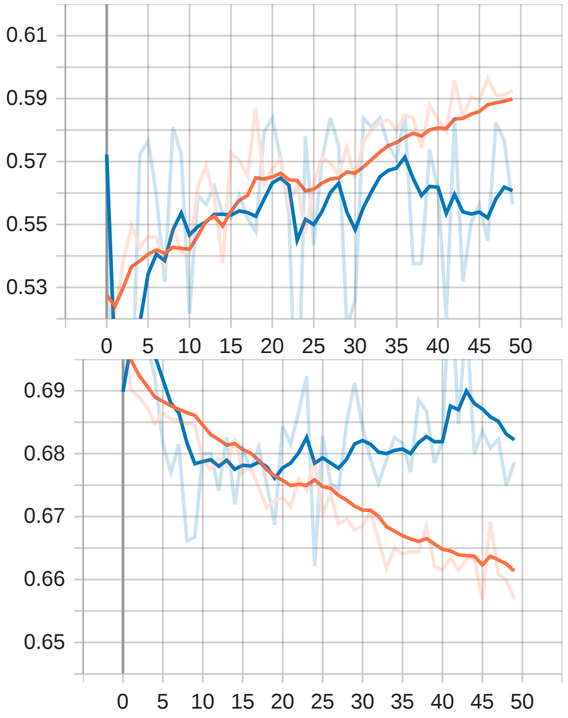
# DEEP NEURAL NETWORK



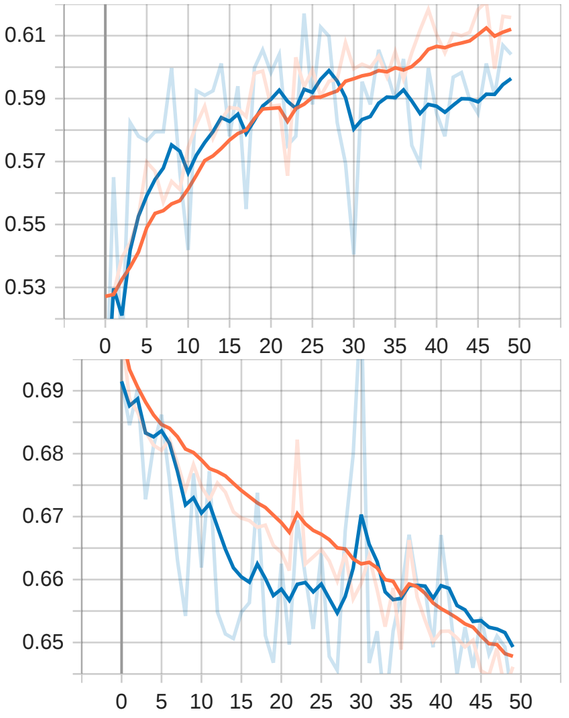

## Base (No Scaling)
## Testing Accuracy: ~54%




## Max Divide Scaling
## Testing Accuracy: ~57%

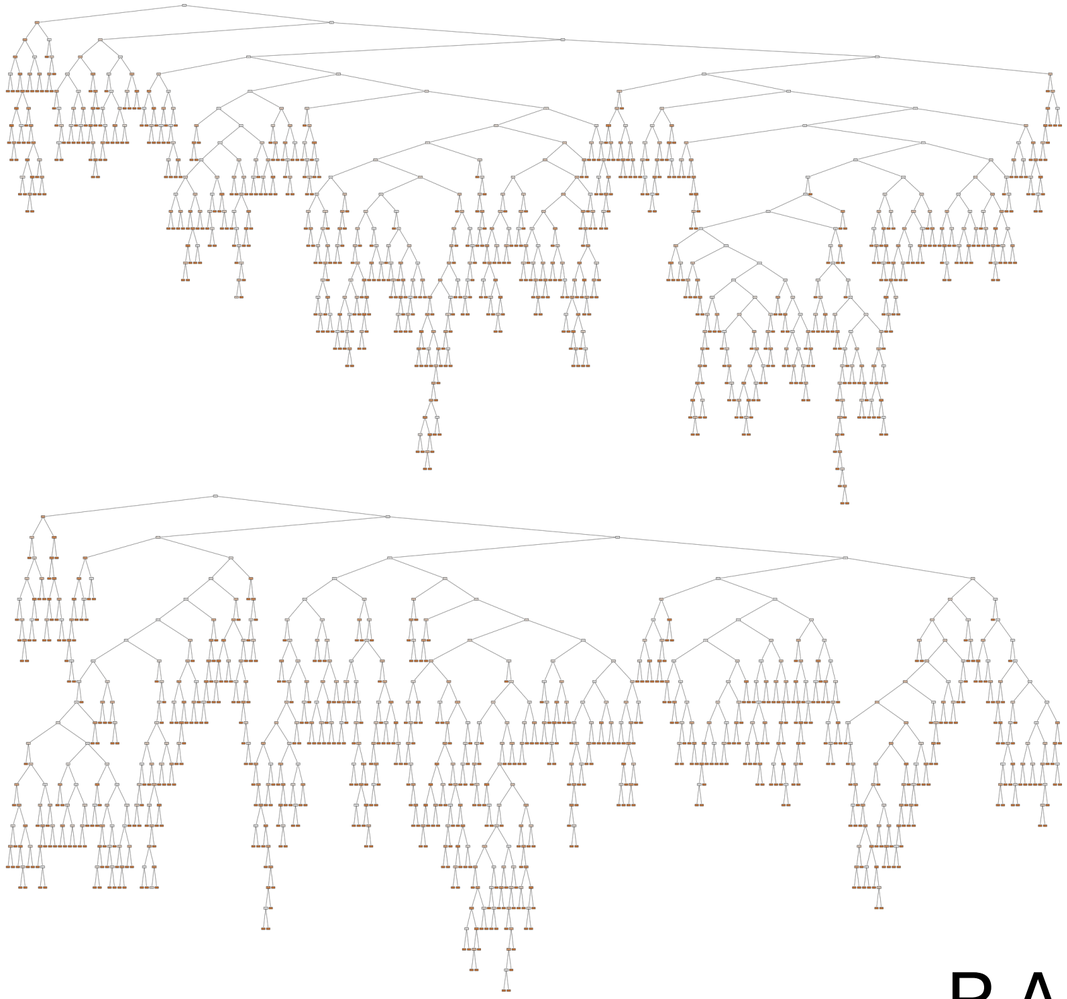



## Log Scaling
## Testing Accuracy: ~61%



# RANDOM FOREST

| Variable | Importance |
|---|---|
| log_popularity | 0.1277 |
| speechiness | 0.0974 |
| duration_min | 0.0957 |
| danceability | 0.0921 |
| valence | 0.091 |
| acousticness | 0.0822 |
| log_tempo | 0.0803 |
| loudness | 0.0798 |
| liveness | 0.079 |
| energy | 0.078 |
| key | 0.0428 |
| instrumentalness | 0.0402 |
| mode | 0.0092 |
| time_signature | 0.0048 |



Testing Accuracy : ~60% with all

# CONCLUSION

Testing Accuracy

| | No Scale (%) | Maximum Division Scale (%) | Log Scale (%) |
|---|---|---|---|
| Linear Model | 59.8 | 60.2 | 62.1 |
| Logistic Regression | 59.9 | 60.3 | 62.0 |
| Logistic LASSO Regression | 54.9 | 56.3 | 61.4 |
| Deep Neural Network | ~54 | ~57 | ~61 |
| Random Forest | ~60 | ~60 | ~60 |

# CITATIONS


Cjhutto. “CJHUTTO/Vadersentiment: Vader Sentiment Analysis. Vader (Valence Aware Dictionary and Sentiment Reasoner) Is a Lexicon and Rule-Based Sentiment Analysis Tool That Is Specifically Attuned to Sentiments Expressed in Social Media, and Works Well on Texts from Other Domains.” GitHub, https://github.com/cjhutto/vaderSentiment.

Dave, Dhruvil. “Billboard ‘The Hot 100’ Songs.” Kaggle, 9 Nov. 2021, https://www.kaggle.com/datasets/dhruvildave/billboard-the-hot-100-songs.

Silva, Mariana O. S. “MusicOSet An Enhanced Music Dataset for Music Data Mining.” MusicOSet - an Enhanced Music Dataset for Music Data Mining, https://marianaossilva.github.io/DSW2019/#downloads.

“Web Player: Music for Everyone.” Spotify, https://open.spotify.com/.