How Well Do Track Properties Measure Popularity

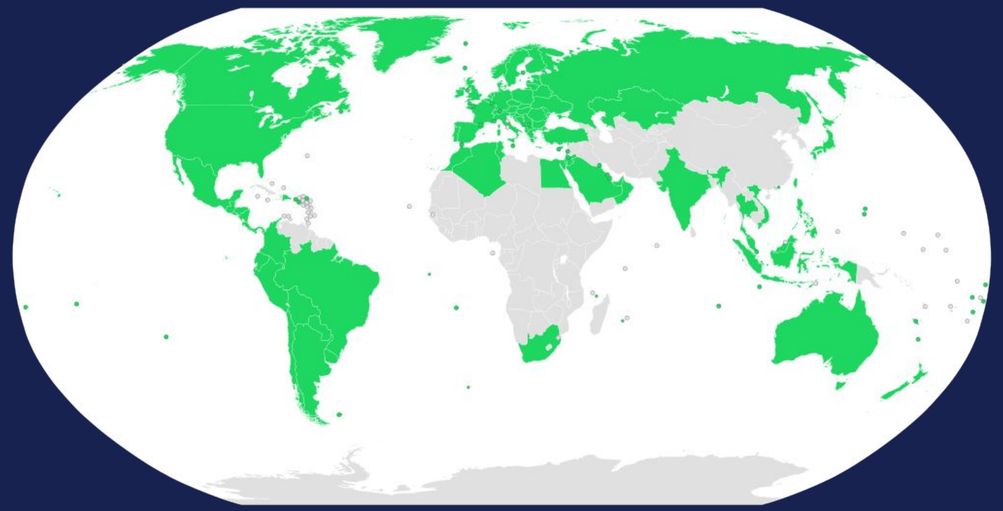
Nerd’s guide to making a popular song!

Brett Nesfeder, Qianhui (Sophia) Lian,
Siddhant Chauhan, David Cruz

Spotify®

Project Goals

- Determine the key factors in for estimating a song's popularity
- Dataset includes Billboard's top songs from 2010-2019

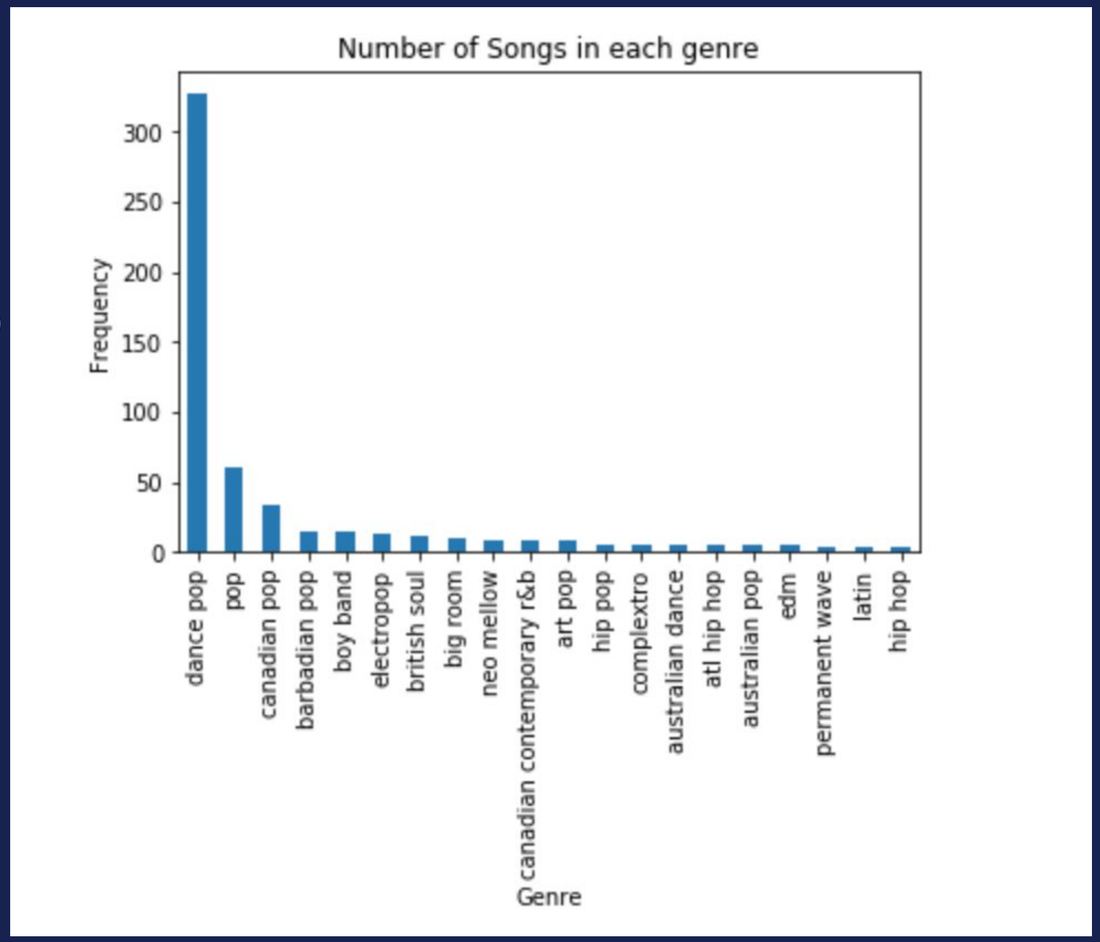


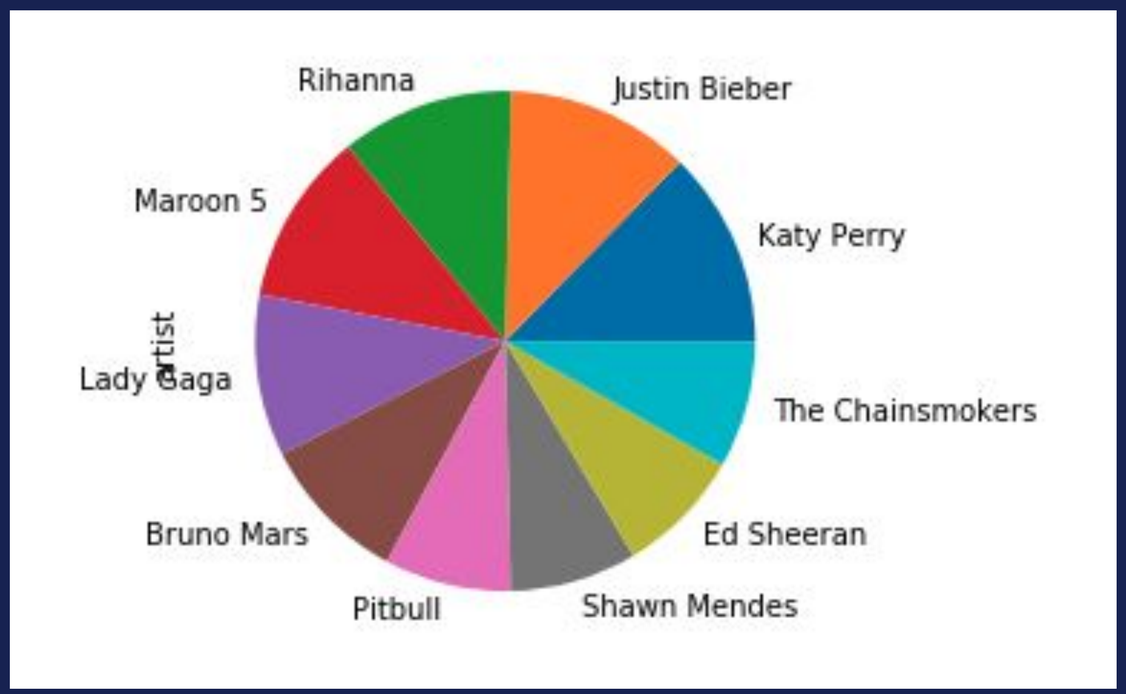
Why is this important?

- New upcoming artists that want to break into the industry and achieve success
- Music trends change rapidly and it is important for record labels to know what current trends are.
- Companies can use information to introduce new concepts and sounds to market

Initial Examinations

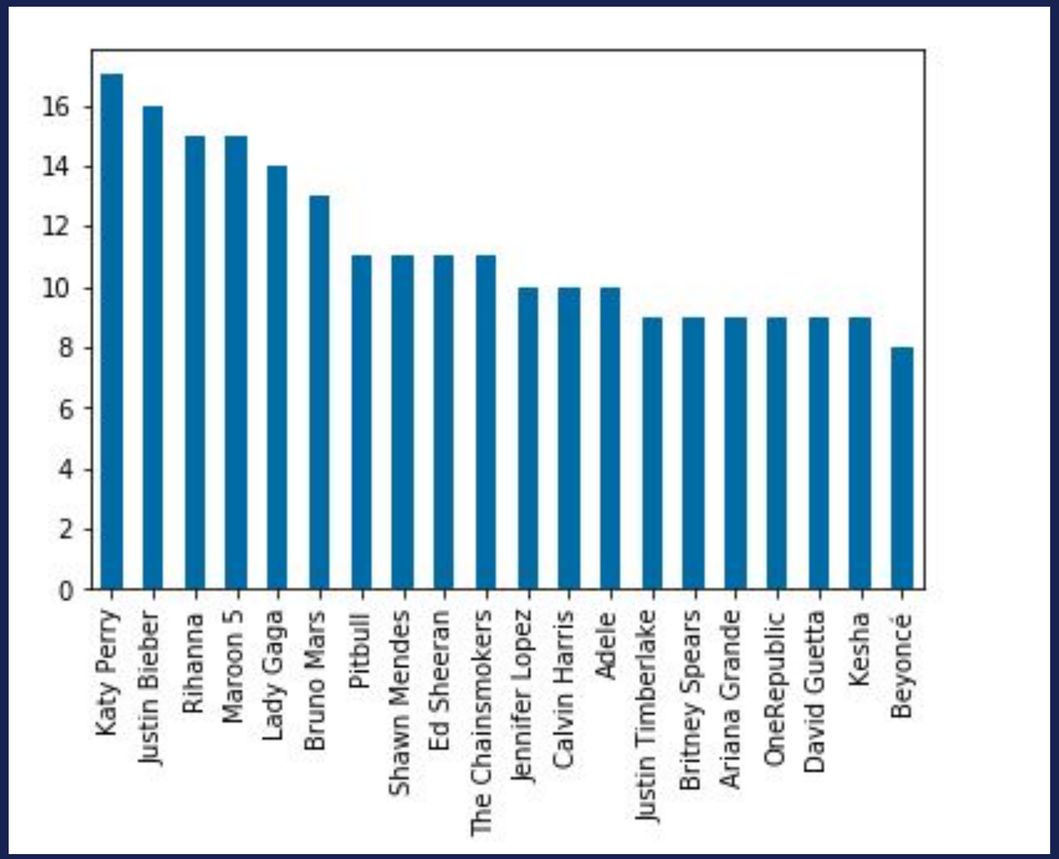
- Dance pop and other variations of pop were most popular about 80% of the data set.
- The top artists were mentioned evenly in the dataset

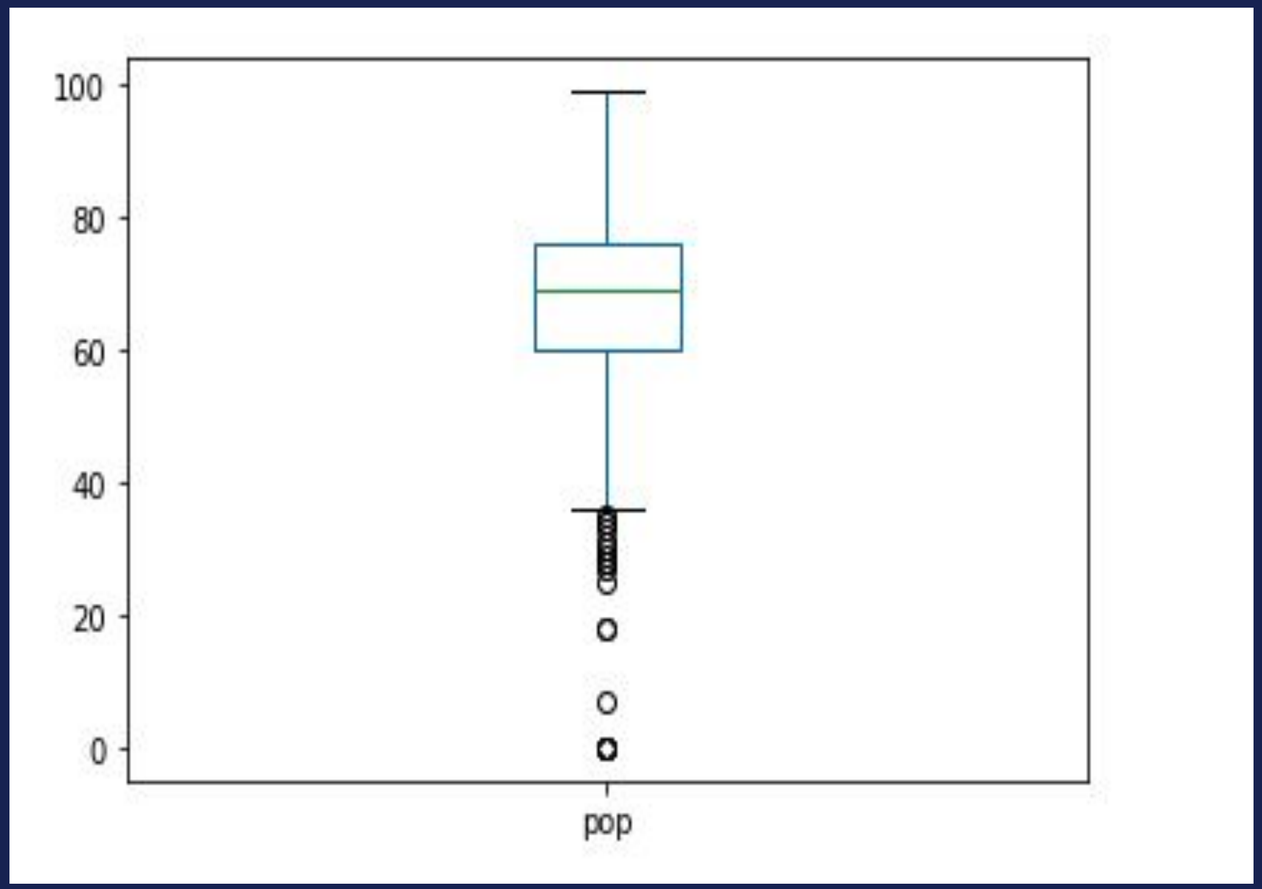




Other Observations

- Maroon 5 had the most popular song (Memories) within the dataset
- Most songs received a score around 60-80 in popularity





Models

- Linear Regression
- Logistic Regression
- Tree

Spotify®

Linear Regression

- Initial model with 10 predictors excluding artist name and genre.
- No multicollinearity
- R^2 = 11.7%
- Significant predictors: year, energy and loudness.

OLS Regression Results

Dep. Variable:	pop	R-squared:	0.117
Model:	OLS	Adj. R-squared:	0.102
Method:	Least Squares	F-statistic:	7.848
Date:	Sun, 09 Aug 2020	Prob (F-statistic):	6.84e-12
Time:	19:37:05	Log-Likelihood:	-2430.8
No. Observations:	603	AIC:	4884.
Df Residuals:	592	BIC:	4932.
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-2485.6127	473.080	-5.254	0.000	-3414.732	-1556.494
year	1.2752	0.234	5.449	0.000	0.816	1.735
bpm	0.0138	0.024	0.581	0.562	-0.033	0.060
nrgy	-0.1578	0.053	-2.958	0.003	-0.263	-0.053
dnce	0.0440	0.052	0.844	0.399	-0.058	0.146
dB	1.3033	0.252	5.177	0.000	0.809	1.798
live	-0.0333	0.044	-0.748	0.455	-0.121	0.054
val	0.0234	0.032	0.722	0.471	-0.040	0.087
dur	-0.0128	0.018	-0.717	0.473	-0.048	0.022
acous	-0.0189	0.034	-0.552	0.581	-0.086	0.048
spch	-0.0425	0.078	-0.548	0.584	-0.195	0.110

Omnibus:	180.471	Durbin-Watson:	0.480
Prob(Omnibus):	0.000	Jarque-Bera (JB):	516.067
Skew:	-1.464	Prob(JB):	8.66e-113
Kurtosis:	6.460	Cond. No.	1.72e+06

Linear Regression

- Second model with just year, energy, and loudness as predictors.
- R^2 = 10.9%

OLS Regression Results

Dep. Variable:	pop	R-squared:	0.109
Model:	OLS	Adj. R-squared:	0.104
Method:	Least Squares	F-statistic:	24.31
Date:	Mon, 10 Aug 2020	Prob (F-statistic):	7.45e-15
Time:	17:40:04	Log-Likelihood:	-2433.7
No. Observations:	603	AIC:	4875.
Df Residuals:	599	BIC:	4893.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-2630.9998	444.809	-5.915	0.000	-3504.574	-1757.426
year	1.3474	0.221	6.110	0.000	0.914	1.780
nrgy	-0.1300	0.041	-3.135	0.002	-0.211	-0.049
dB	1.3802	0.237	5.814	0.000	0.914	1.846

Omnibus:	176.883	Durbin-Watson:	0.453
Prob(Omnibus):	0.000	Jarque-Bera (JB):	488.605
Skew:	-1.449	Prob(JB):	7.96e-107
Kurtosis:	6.324	Cond. No.	1.60e+06

Linear Regression

	2010	
	coefficient	p-value
valence	0.2458	0.0743

	2011	
	coefficient	p-value
acoustic	-0.2932	0.0645
speech	-0.4579	0.0963

	2012	
	coefficient	p-value
energy	-0.4443	0.0819

	2013	
	coefficient	p-value
	no significant predictors	

	2014	
	coefficient	p-value
live	-0.472	0.0524

	2015	
	coefficient	p-value
acoustic	0.1852	0.0528

	2016	
	coefficient	p-value
dB	1.614	0.0028

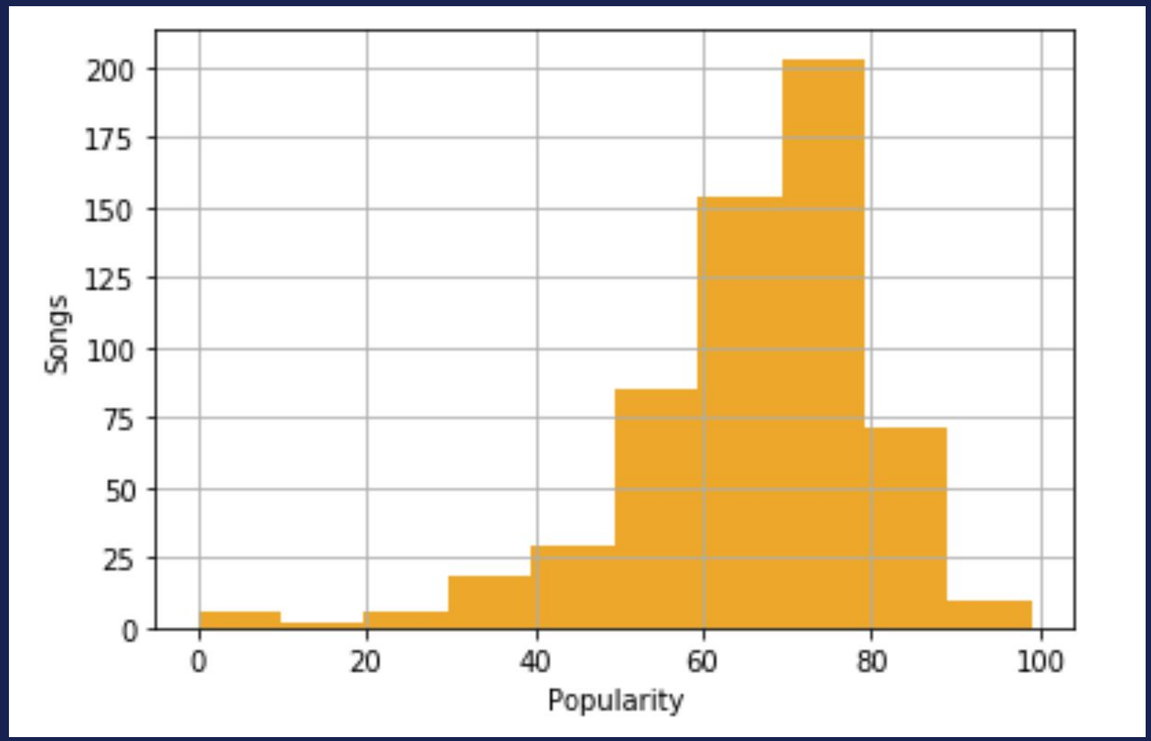
	2017	
	coefficient	p-value
dance	0.3185	0.012
dB	2.0587	0.0559
live	-0.3006	0.0484
speech	0.517	0.0107

	2018	
	coefficient	p-value
dB	2.2527	0.0521

	2019	
	coefficient	p-value
bpm	-0.1095	0.0979
energy	-0.3797	0.0055
dB	2.577	0.028

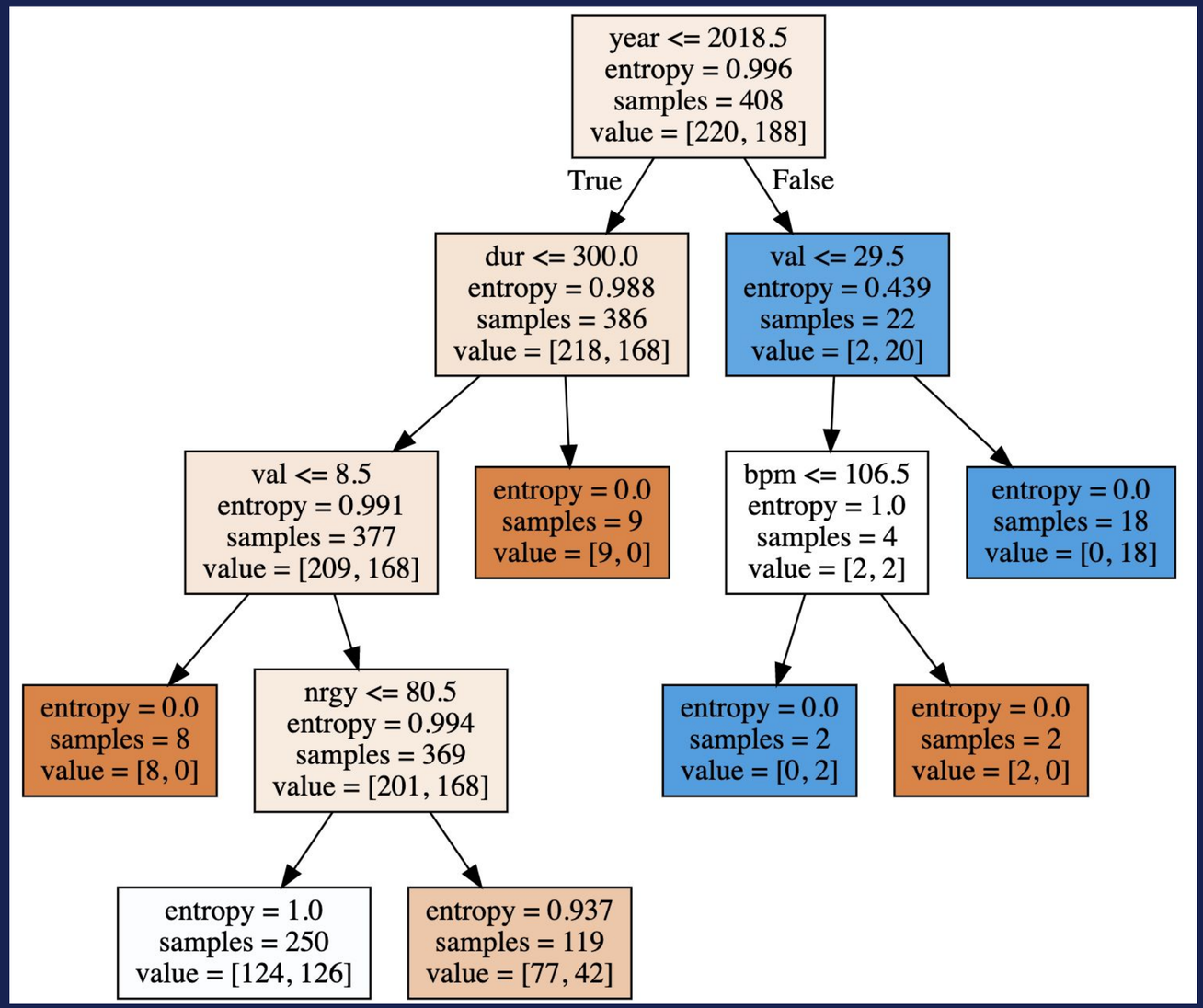
Logistic Regression

- Classifier = Popularity
  - Median ‘Pop’ = 69
- Significant Positive Values
  - dB and Spch
- Significant Negative Values
  - Nrgy



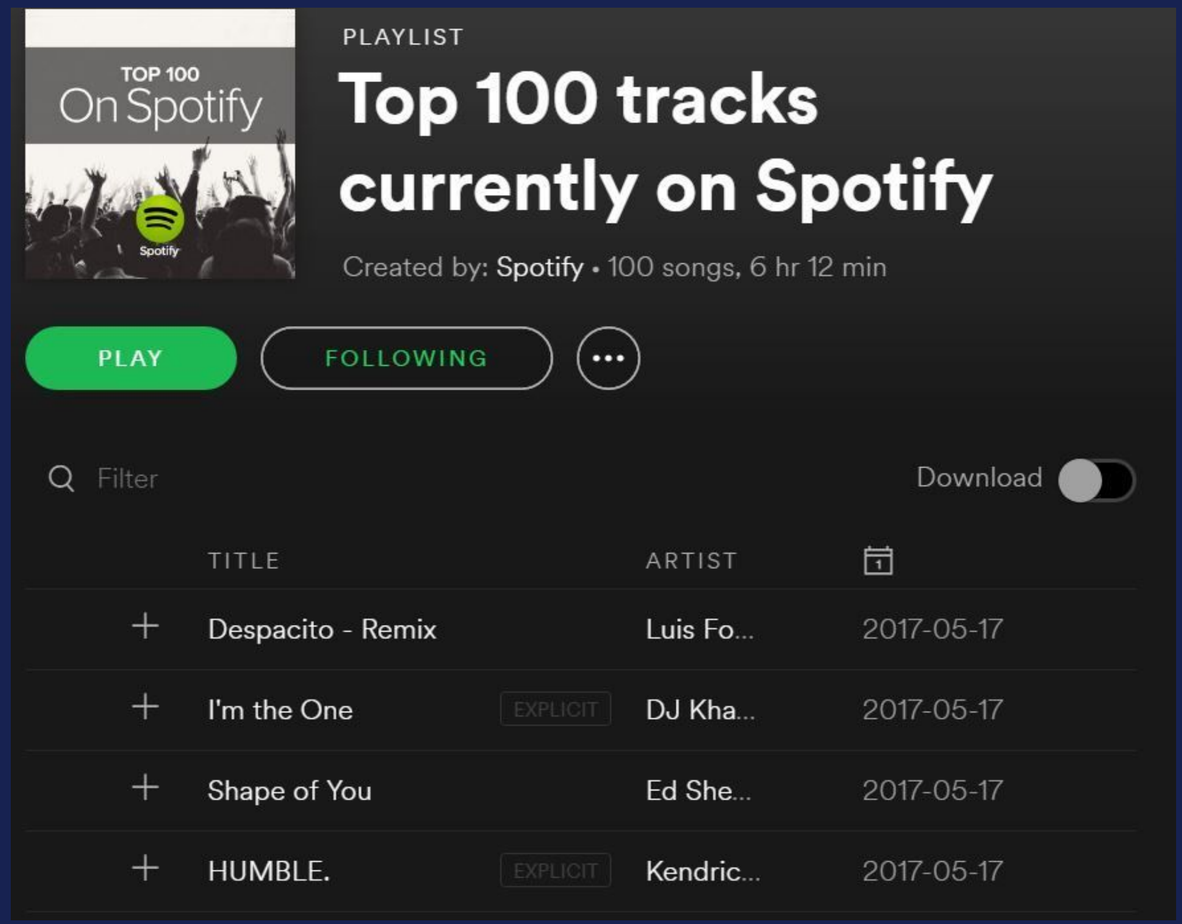
Tree

- Split by Year because of greatest reduction of entropy
- Max depth of 4 because of overfitting



Insights

- More songs from each year



Insights

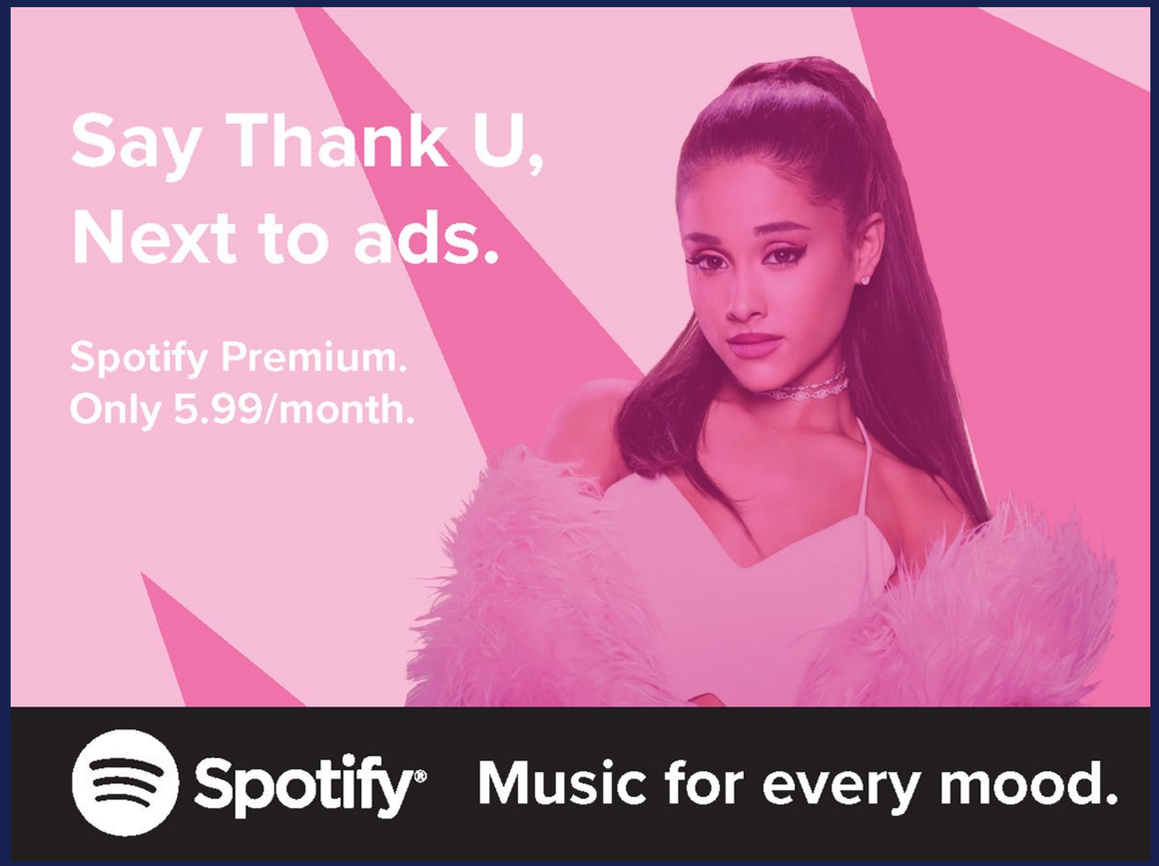
- Other measures to include to help build a more accurate model:

Popularity of Artist

Collaborations

Platforms of Promotion

Production/Advertising Budgets



Producer

Thank you

Spotify®