Natural Language Processing Case Study

What it takes to be in Billboard's Top 100

James Brochhausen

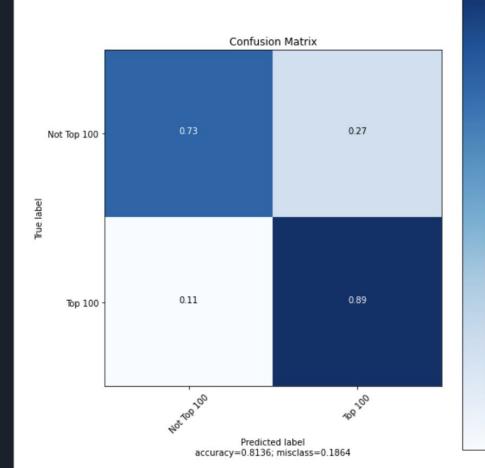
Overview

- Over the past two months, I've tasked myself in identifying and creating a model which helps predict and in interpret if the lyrics to a song are likely to get to you into the top 100 of the Billboard (BB) charts
- Along with this we'll be identifying trends within the music industry that correlate to being in the BB Top 100
- Started with 15k+ data entries from 1959-2019
 - Ended with 4k+ data entries and 18 classes

Agenda

- Our Model
- Common Lyrics
- Correlation & Genre

Our Model (SVM)

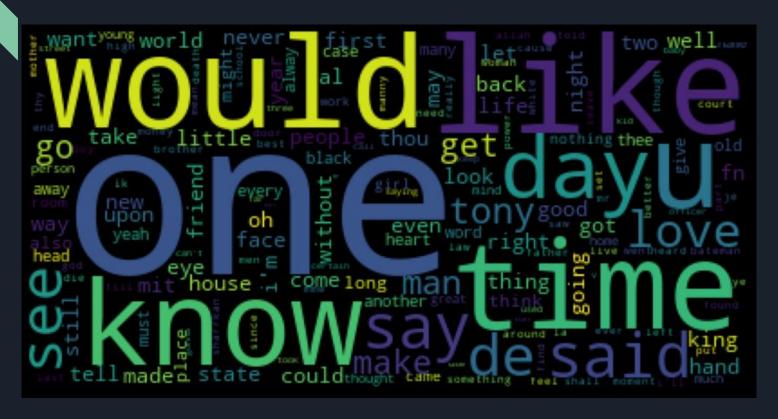


- 0.4

Words in Billboard Top 100



Words in Not Billboard Top 100



Our Model and Feature Importances

0 girl 2.881748 might 2.068768 front 1.818364 bout 1.802342 1.789969 wanna i'm 1.698252 keep 1.666101 cause 1.587499 1.536806 that's 1.536058 truth 1.512255 tonight 1.504143 even 1.492773 here's 1.414876 else 1.408737 good 1.394643 three 1.320400 made 1.292781 1.288751 let's 1.284280 1.279412 waiting beer 1.264831 making 1.257060 truck 1.246915 1.242812 i'ma 1.238310 like 1.225177 lucky 1.218651 1.211608 sexv 1.205099

Common Lyrics Found Together

Not in BB Top 100

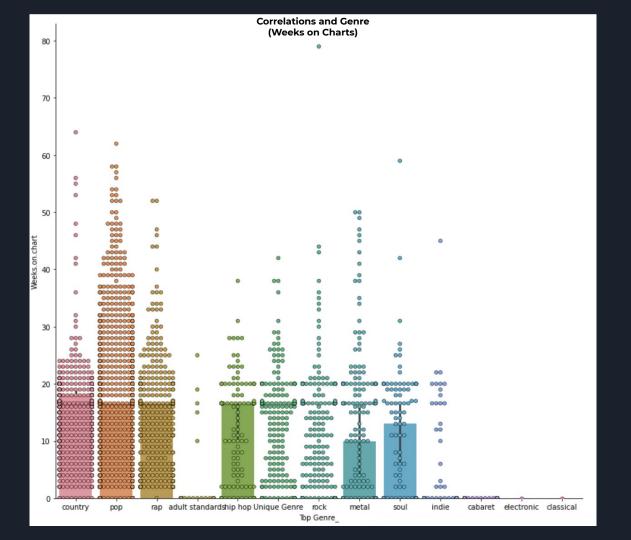
	Lyric	Freq
0	(zau, al-makan)	0.000510
1	(mme, de)	0.000473
2	(de, guermantes)	0.000403
3	(aaron, swartz)	0.000348
4	(de, charlus)	0.000322
5	(united, states)	0.000306
6	(oh, oh)	0.000219
7	(high, school)	0.000212
8	(yeah, yeah)	0.000205
9	(wazir, dandan)	0.000204
10	(mme, verdurin)	0.000202
11	(new, york)	0.000199
12	(I, E)	0.000196
13	(chris, brown)	0.000188
14	(na, na)	0.000185
15	(omar, bin)	0.000185
16	(la, la)	0.000179
17	(, tony)	0.000168
18	(auspicious, king)	0.000160
19	(one, day)	0.000160

BB Top 100

	Lyric	Freq
0	(oh, oh)	0.003457
1	(yeah, yeah)	0.002740
2	(la, la)	0.001824
3	(na, na)	0.001404
4	(i'm, gonna)	0.001125
5	(love, love)	0.001045
6	(let, go)	0.001010
7	(hey, hey)	0.000905
8	(know, i'm)	0.000801
9	(cause, i'm)	0.000786
10	(feel, like)	0.000782
11	(oh, yeah)	0.000743
12	(know, know)	0.000616
13	(da, da)	0.000566
14	(like, i'm)	0.000563
15	(every, time)	0.000554
16	(baby, baby)	0.000539
17	(ooh, ooh)	0.000512
18	(yeah, i'm)	0.000501
19	(go, go)	0.000470

Most Impactful Genres

- Country
- Pop
- Hip-Hop/Rap



Recommendations

- Leverage hooks, chorus and outro lyrics such as "oh, oh", "la, la", "hey, hey" etc.

Don't sing about specific locations

Stay within Country, Rap / Hip-Hop and Pop

Future Work

- View data by the decades and how they trended
- Look into other features such as BPM, Tempo, Danceability, etc.
- Dive deeper into Genre

Thank You

Contact:

<u>James@flatironschool.com</u>

123-456-7890