



Natural Language Processing Case Study

What it takes to be in BillBoards
Top 100

James Brochhausen



Overview

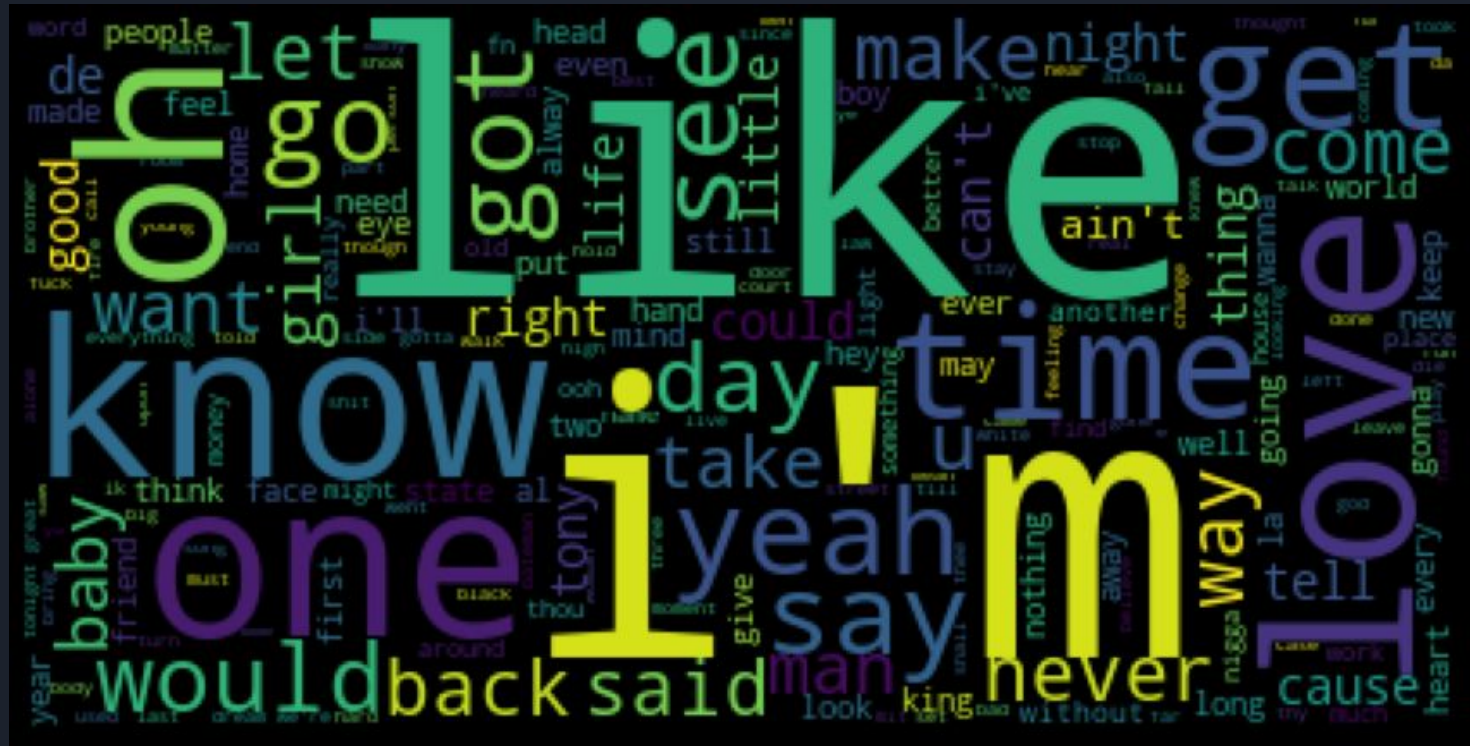
- Over the past two months, I've tasked myself in identifying and creating a model which helps predict and in interpret if the lyrics to a song are likely to get to you into the top 100 of the Billboard (BB) charts
- Along with this we'll be identifying trends within the music industry that correlate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



Agenda

- Common Lyrics
- Correlation & Genre
- Model

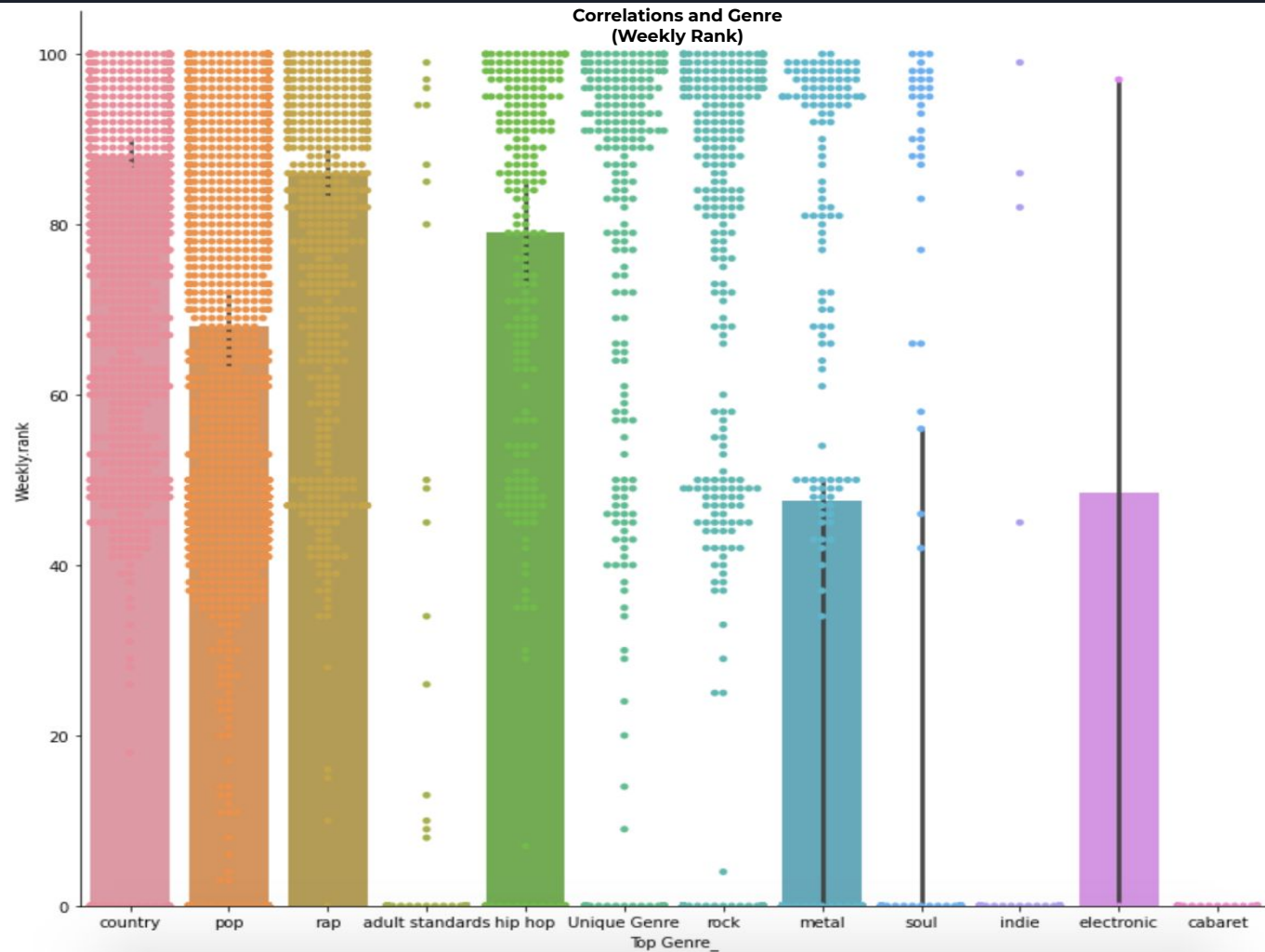
Common Words

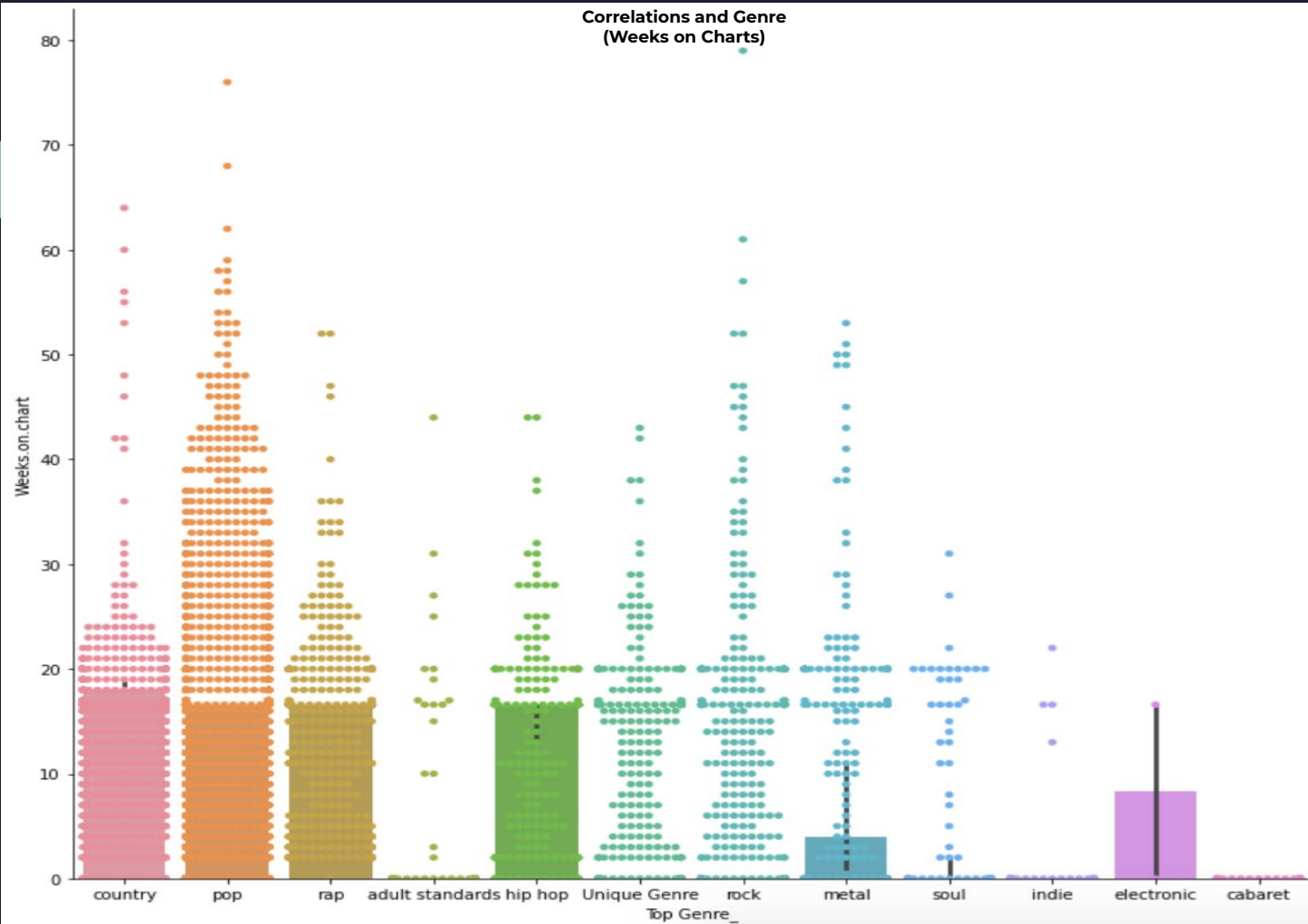


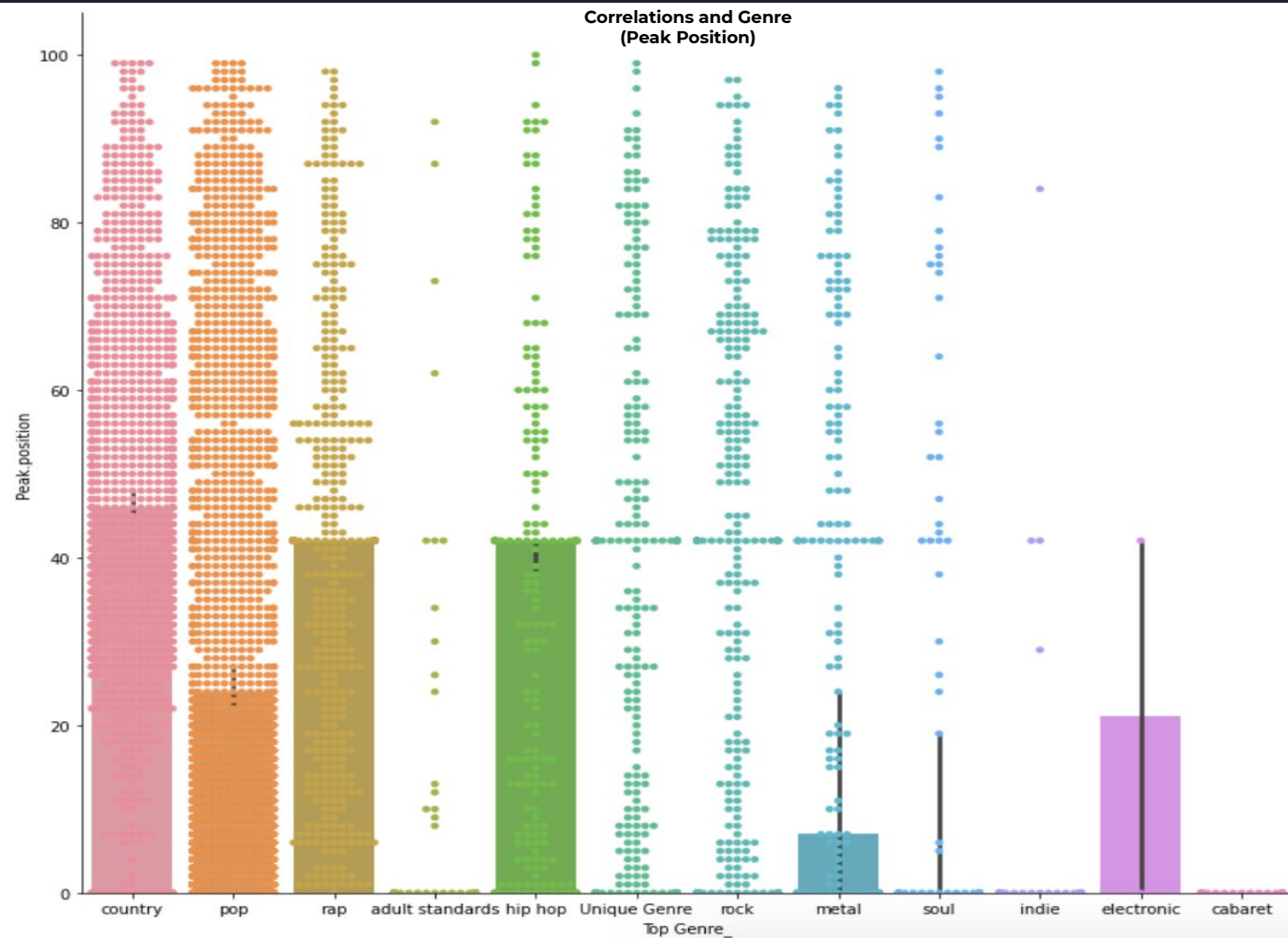


Common Words Continued

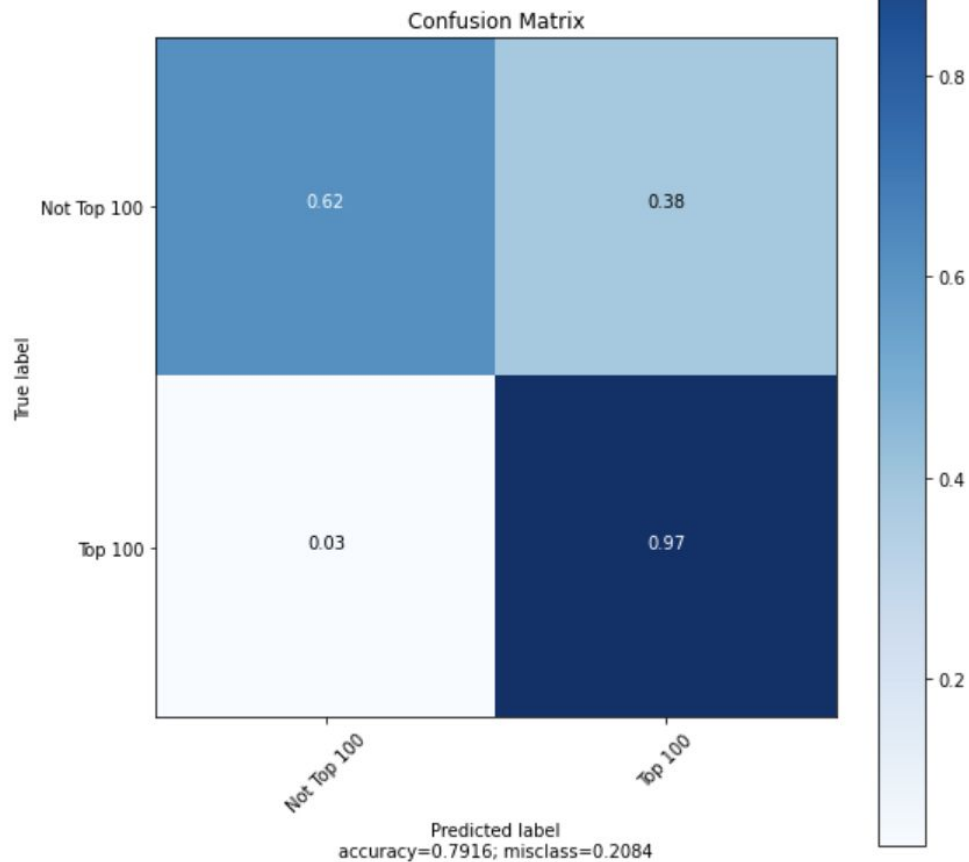
	Lyric	Freq
0	(oh, oh)	0.001056
1	(yeah, yeah)	0.000870
2	(la, la)	0.000575
3	(na, na)	0.000486
4	(zau, al-makan)	0.000378
5	(i'm, gonna)	0.000373
6	(mme, de)	0.000350
7	(let, go)	0.000340
8	(love, love)	0.000334
9	(de, guermantes)	0.000299
10	(feel, like)	0.000287
11	(hey, hey)	0.000278
12	(aaron, swartz)	0.000257
13	(know, i'm)	0.000256
14	(oh, yeah)	0.000251
15	(de, charlus)	0.000239
16	(cause, i'm)	0.000230
17	(united, states)	0.000228
18	(new, york)	0.000213
19	(know, know)	0.000209







Our Model



Recommendations

- **Have catchy hooks and finishes to your songs**
- **Talk about Love**
- **Stay within Country, Rap / Hip-Hop and Pop (High Barrier to Entry)**

Future Work

- View Data by the decades and how they trended
- Dive deeper into Genre rather than Lyrics.

Thank You

Contact:

James@flatironschool.com

123-456-7890

