# Billboard Hot 100 Sentiment & Recs

Aadithya Anandaraj,
LaShay Fontenot,
Haritha Maheshkumar,
Catherine Miao,
 Sijo VM

Will the song make it to the Top 25 Billboard?

**BUSINESS**

**CONSUMER**

Which song should I listen to next?

# 1 SONG & LYRIC SCRAPING

Compile full list of songs from Billboard Top 100 and feed titles and artists into Genius to pull complete lyrics

# 2B SENTIMENT ANALYSIS

Assess sentiment of each song's lyrics using Vader (overall and across sentiments)

# 2A TOPIC MODELING

Determine list of topics across songs and topic percentage

# 3 RANK & RECS

Use outcomes of sentiment and topic modeling to attempt to predict song rank in the Top 100 and create song recommendations

# 11 years

Of Billboard Top 100 lists

# 1100 songs

Of varying genres with lyrics scraped from Genius

# 686 artists

From Taylor Swift to Luke Bryan

# VADER SENTIMENT ANALYSIS

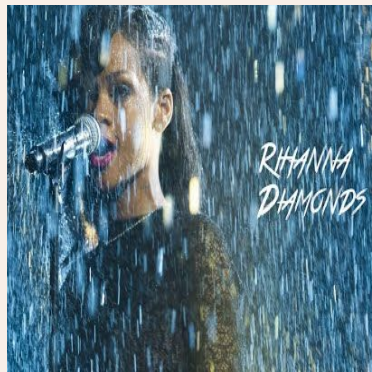Lexical method of obtaining sensitivity scores by building a lexicon or a 'dictionary of sentiment'

# POLARITY SCORES

The sentiment score of a song - calculated by summing up the sentiment scores of each VADER-dictionary-listed word in the lyrics

Scaled to -1 to 1 : most negative to most positive

# DIAMONDS

# iSPY

# ALL TIME LOW







**Pos: 0.564**

Neu: 0.436

Neg: 0

Pos: 0

**Neu: 0.993**

Neg: 0.007

Pos: 0.026

Neu: 0.247

**Neg: 0.728**

# EXTRACTION OF TOPICS IN LYRICS

Topic Modeling produced 4 topics across all song lyrics

### Topic 1

Party

Night, life, sexy, drink

### Topic 2

Love

Heart, love, baby, leave

### Topic 3

Dance

Booty, hump, lean, rock

### Topic 4

Money

Money, grill, best

Rank <= 25 ; High

| Logistic Regressions | Class Weight | Accuracy | ROC | F1 |
|---|---|---|---|---|
| 1. Lyrics | Balanced | 0.83 | 0.89 | 0.70 |
| 2. Topic & Sentiment | Balanced | 0.74 | 0.53 | 0.50 |
| 3. Lyrics, Topic, Sentiment | Balanced | 0.82 | 0.90 | 0.70 |

Baseline Accuracy: 75%

# SONG CLUSTERING

Using K-Means clustering we determined similar songs

### Cluster 0

- Drake
- Britney Spears
- XXXtentacion
- Imagine Dragons
- Adele

### Cluster 1

- Ed Sheeran
- Beyonce
- Miley Cyrus
- Rihanna
- Adele

### Cluster 2

- Taylor Swift
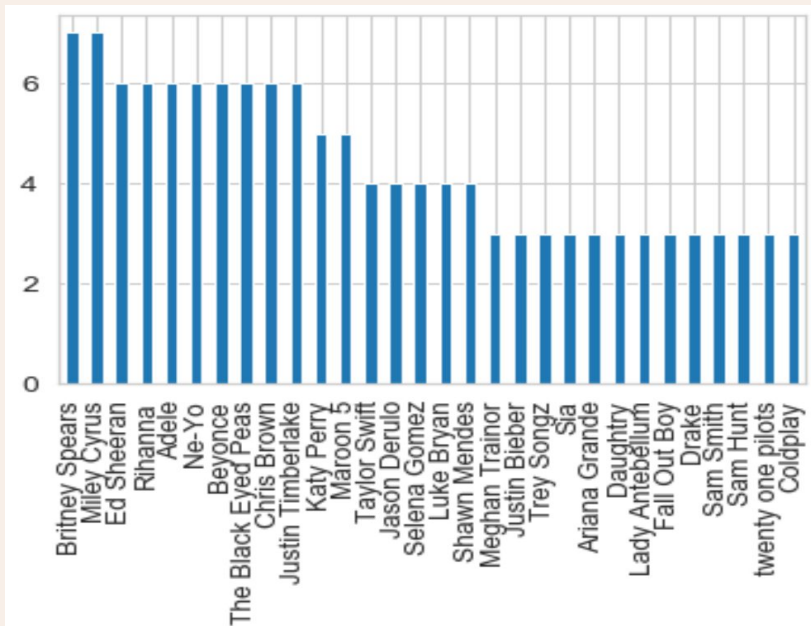- Bruno Mars
- Katy Perry
- Lady Gaga
- The Weeknd

### Cluster 3

- Drake
- Rihanna
- P!nk
- Lorde
- Ke$ha

The results of this analysis could be used to place new music/artists near other popular songs when marketing

## SAME SAME BUT DIFFERENT

Clustering -> Longtail Curves (artists) -> Insights



### HYPOTHESES

- Differences in target market i.e. Britney Spears younger, party crowd vs Sam Hunt niche genre
- Differences in music quality and production ('danceability score' etc.)

**Music labels could focus on what could move songs further to the left, but be wary of making comparisons**

# RECOMMENDATION PROCESS

**1** Calculate Cosine Similarity among songs based on sentiment and topic modeling scores

**2** Select input song

**3** Produce recommendations using top similarity scores

# RECOMMENDATION EXAMPLE



**Song** : Boom Boom Pow by The Black Eyed Peas

**Sentiment** : Positive ( score .64)

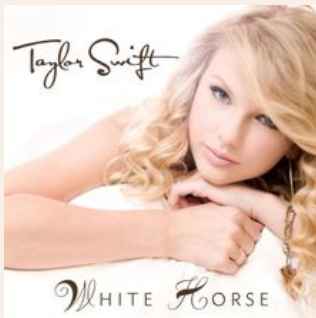**Topic** : Topic 1 (Party)

**1**

Vogue by Madonna

**2**

We Didn't Start the Fire by Billy Joel

**3**

Live Your Life by T.I. Ft. Rihanna

# RECOMMENDATIONS

**Song** : White Horse by Taylor Swift

**Sentiment** : Positive ( score .89)

**Topic** : Topic 4 (Money)

**1**
Everything By
Jody Watley

**2**
How Can I Fall?
By Breather

**3**
Just Like Jesse
James by Cher

## SO WHAT?

- Artists have freedom on choice of topics as they don't seem to drive position on billboard

- Lyrics are a big indicator of a song making it to Top 25

- Sentiment distribution does not vary across low and high ranking songs - hinting to the fact that feelings don't matter as long as the target market is right

***Success is less about the sentiment and topic of the song and more about the total package and the target audience***

# THANKS!

Does anyone have any questions?

# APPENDIX

# MUSIC INDUSTRY IN US

- Recorded music in the retail sector - **$9.8 Bn in 2018**

- **75%** of music industry's **total revenue** comes from **streaming** industry

- **Streaming services reached 50Mn subscriptions** in US - first time in 2018 (Source: RIAA report)

- **Seamless listening experience** would help capture a larger market - recommendation is key

Selected Punctuation Removal

Stop words removal

Lower case conversion

Lemmatization

# Example Lyric

"oh yes it is me again the crazy one"
{'neg': 0.198, 'neu': 0.579, 'pos': 0.223, 'compound': 0.0772}

## PUNCTUATION

"oh yes! it is me again!
the crazy one!"
(pos:0.276)

## CAPITALIZATION

"OH YES it is me again
the crazy one"
(pos:0.268)

## DEGREE MODIFIERS

"oh yes it is me again
the sort of crazy
one" (pos:0.188)

## CONNECTING WORDS

"oh yes it is me again
but the crazy one"
(neg: 0.239)

## POLARITY NEGATION

"oh yes it is me again
the not crazy one"
(pos:0.372,neg:0)

# RECOMMENDATIONS

**Song** : Viva La Vida by Coldplay

**Sentiment** : Negative ( score - .92)

**Topic** : Spread across Topics 2, 3, 4

**1**

Ayo Technology by 50 Cent Ft. Justin Timberlake

**2**

Right Round by Flo Rida

**3**

All or Nothing by Milli Vanilli