



# Analysis of Hip-Hop/Rap Music Lyrics over Time



Khemet Robinson  
SML 310 Final Project





# Motivation

“

“Hip-hop has done more damage to young African Americans than racism in recent years”

- Geraldo Rivera, Fox News

- Negative Perception of Hip-Hop/Rap Music !
- Hip-Hop/Rap still trending...





# Goals

## Goals

1. Assess the uniqueness of diction and profanity usage in hip-hop/rap music.
2. Uncover trends of sentiment in hip-hop/rap music.
3. Analyze content changes over time.



# Related Work

## Related Work

Genre  
Classification  
Using Music  
Lyrics




Sentiment  
Analysis of  
Music  
Lyrics

Topic  
Analysis  
of Music  
Lyrics





# Approach

- **3-Part Analysis:**
  1. Classification 
  2. Regression 
  3. Topic Modelling 
- **Primary Contribution: New Dataset !!**

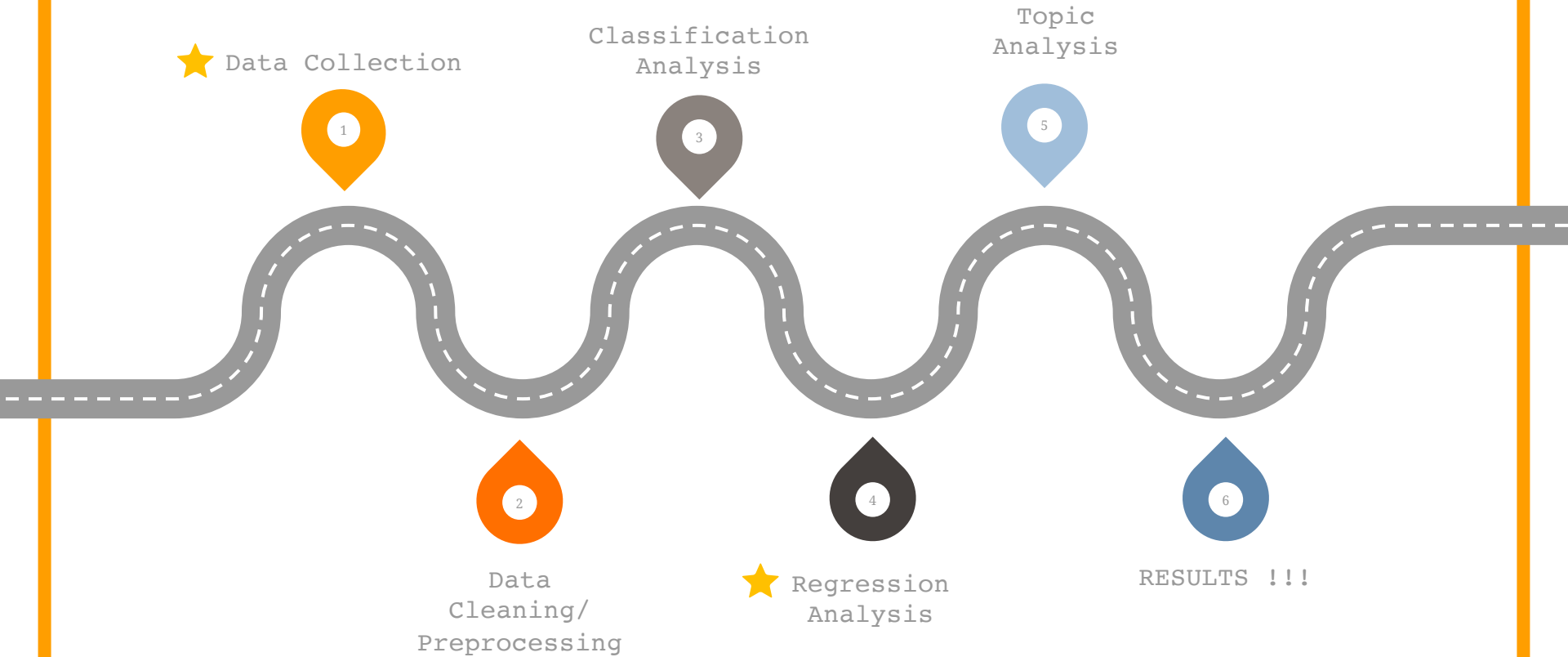
	RANK	TITLE	ARTIST	YEAR	LYRICS	sentiment	profanity total	Hip-Hop/Rap	Not Hip-Hop/Rap
3143	1	Blinking Lights	The Weeknd	2020	yeah ive tryna call ive long enough maybe show...	0.8065	0	0	1
3144	2	Circles	Post Malone	2020	oh oh oh oh oh oh oh oh oh oh oh couldnt turn ...	0.8539	4	0	1
3145	3	The Box	Roddy Ricch	2020	pullin coupe lot told em fuck 12 fuck swat bus...	-0.9908	22	1	0
3146	4	Don't Start Now	Dua Lipa	2020	dont wanna see full 180 crazy thinking bout wa...	-0.9656	0	0	1
3147	5	Rockstar	DaBaby	2020	woo woo pull like pull baby pull oh oh oh pull...	0.9979	21	1	0



# Implementation








# Project Roadmap



# Data Collection


2020 [See More Year-End Charts](#)

1		<b>Blinding Lights</b> The Weeknd	▼
2		<b>Circles</b> Post Malone	▼
3		<b>The Box</b> Roddy Ricch	▼
4		<b>Don't Start Now</b> Dua Lipa	▼
5		<b>Rockstar</b> DaBaby Featuring Roddy Ricch	▼

Billboard

GENIUS

COLLECTIVE SOUNDS | GENIUS LIVE | FEATURED | CHARTS | VIDEOS | COMMUNITY | SHOP | [f](#) [t](#) [@](#) [v](#)

 **Blinding Lights**  
The Weeknd  
Produced by Oscar Holter, Max Martin & The Weeknd  
Album After Hours  
VERIFIED LYRICS BY The Weeknd

BLINDING LIGHTS LYRICS

[Intro]  
Yeah

[Verse 1]  
I've been tryna call  
I've been on my own for long enough  
Maybe you can show me how to love, maybe  
I'm going through withdrawals  
You don't even have to do too much  
You can turn me on with just a touch, baby

[Pre-Chorus]  
I look around and  
Sin City's cold and empty (Oh)  
No one's around to judge me (Oh)  
I can't see clearly when you're gone

About "Blinding Lights" 7 contributors

"Blinding Lights," the second single to The Weeknd's fourth studio album *After Hours*, following the record's lead single "Heartless," which released two days prior on November 27, 2019. The track finds Abel in a constant state of distraction that he only gets relief from when in the presence of a significant other. The song is an up-tempo electropop track that features... [read more »](#)

+409

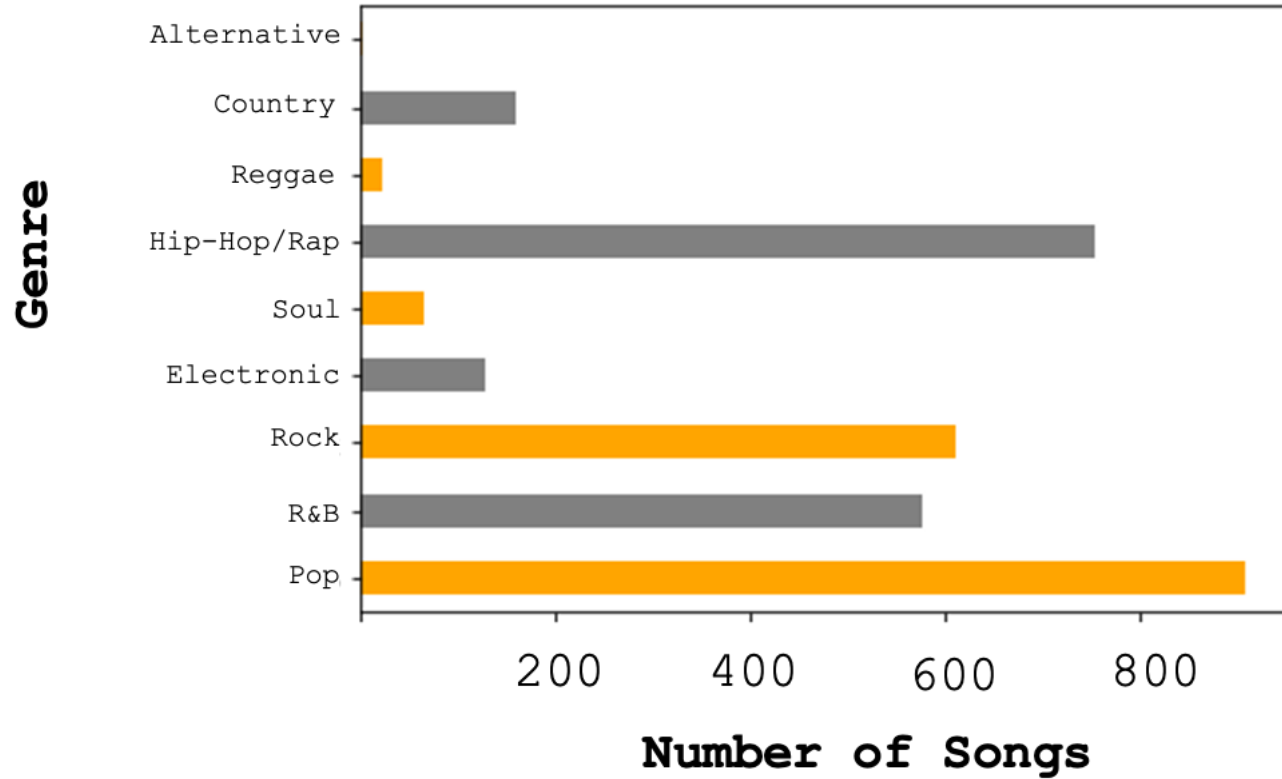
Translations

- Albanian Translation
- Danish Translation
- Dutch Translation
- Farsi Translation
- French Translation
- German Translation
- Hebrew Translation
- Italian Translation
- Japanese Translation
- Russian Translation

Blinding Lights  
The Weeknd  
0:00 - 3:21

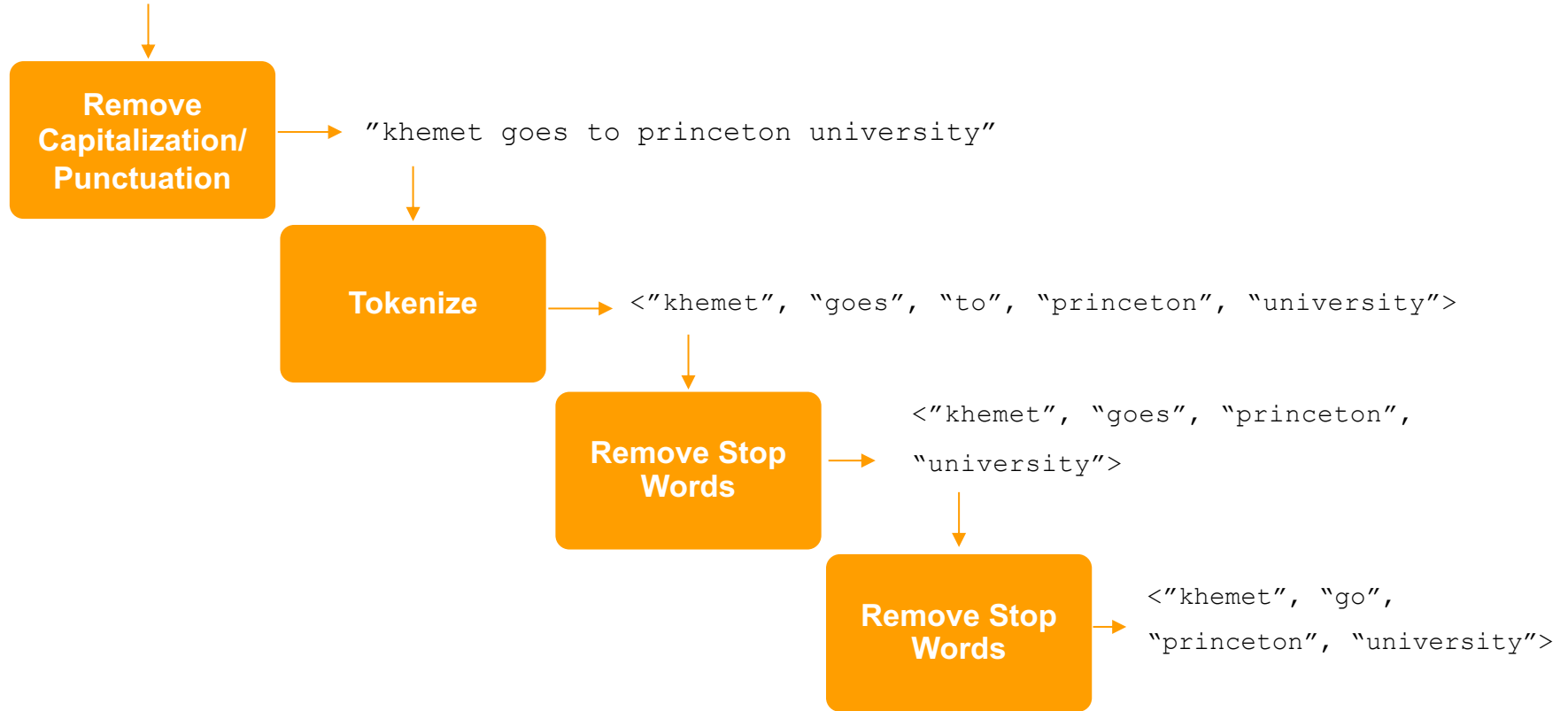
Genius

## Data Collection

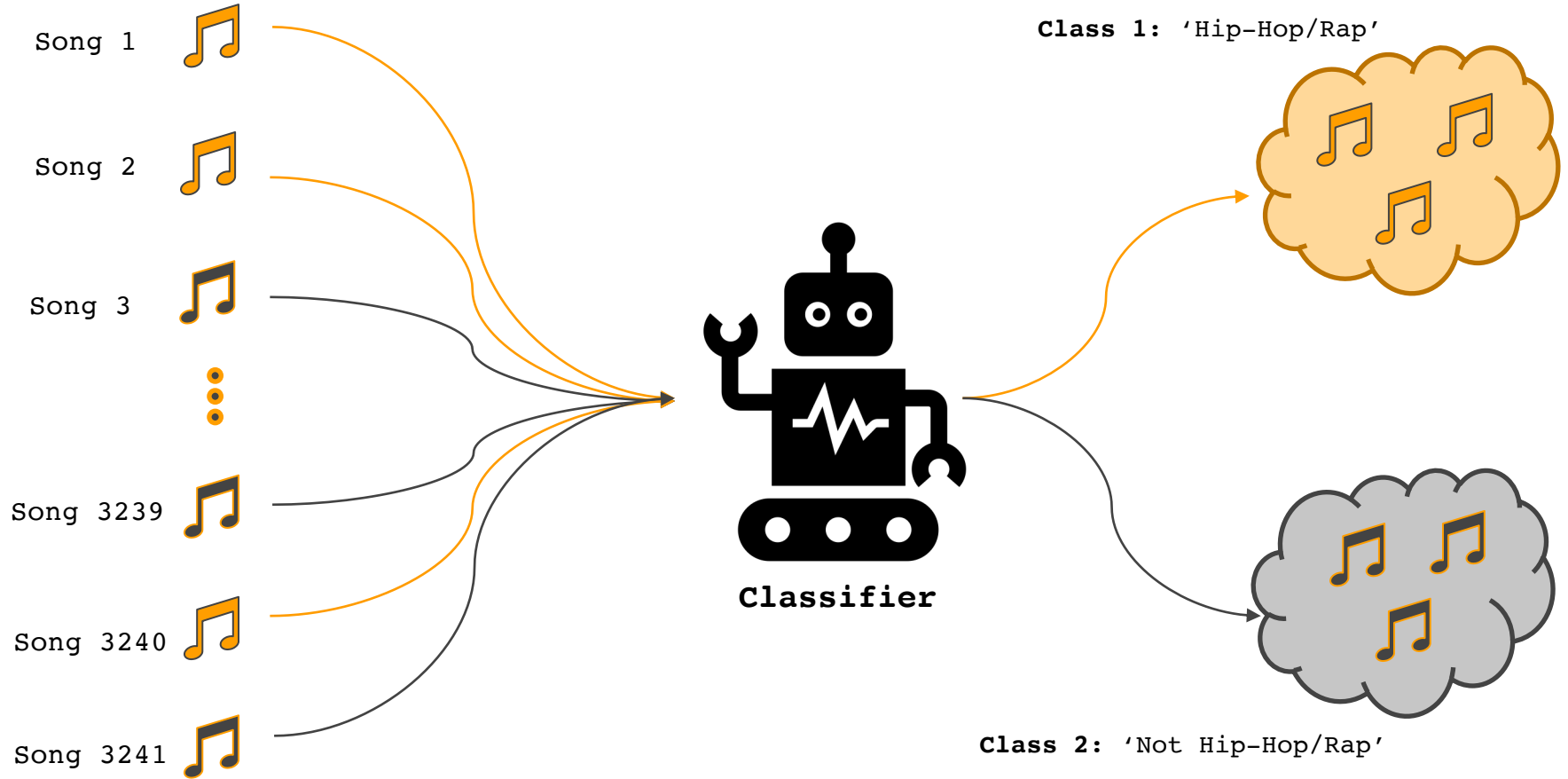


# Data Preprocessing

"Khemet goes to Princeton University."



# Classification Analysis





## Regression Analysis

$$\text{sentiment} = \beta_0 + \beta_1 \cdot \text{YEAR} + \beta_2 \cdot \text{profanity\_total} + \beta_3 \cdot \text{HipHop/Rap}$$

- sentiment - the sentiment score given by VADER
- YEAR - the year that the song appears in the top 100
- profanity\_total - the # of occurrences of profanity words in song
- Hip-Hop/Rap - whether the song is hip-hop/rap (1) or some other genre (0)

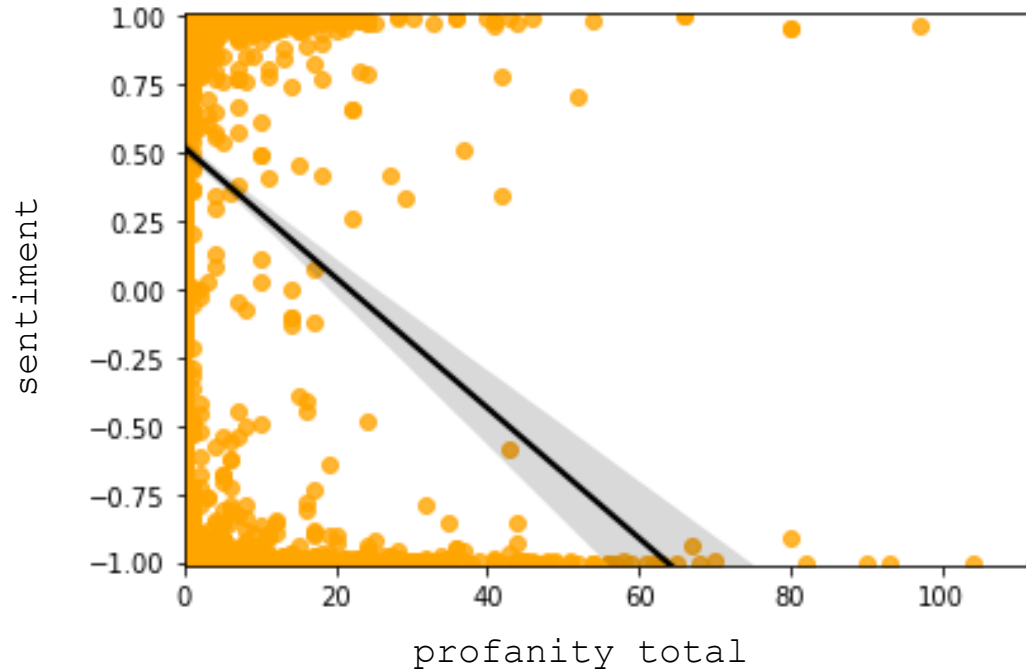


# Exploratory Data Analysis

## Exploratory Data Analysis

$$\text{sentiment} = \beta_0 + \beta_1 \cdot \text{YEAR} + \beta_2 \cdot \text{profanity\_total} + \beta_3 \cdot \text{HipHop/Rap}$$

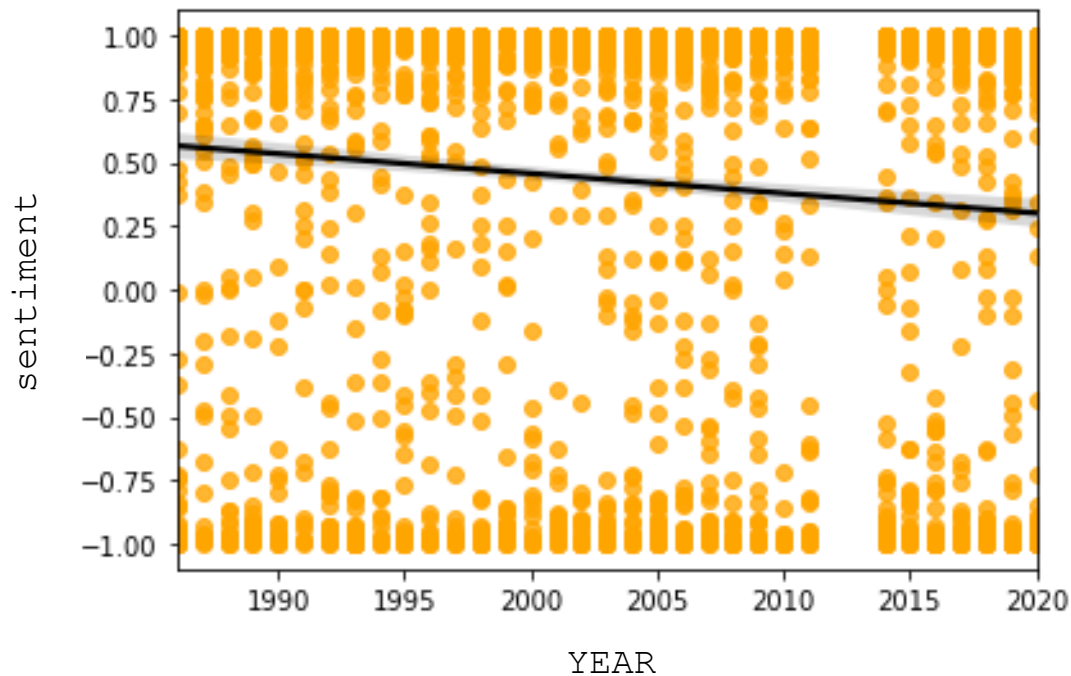
### Initial Visualizations



## Exploratory Data Analysis

$$\text{sentiment} = \beta_0 + \beta_1 \cdot \text{YEAR} + \beta_2 \cdot \text{profanity\_total} + \beta_3 \cdot \text{HipHop/Rap}$$

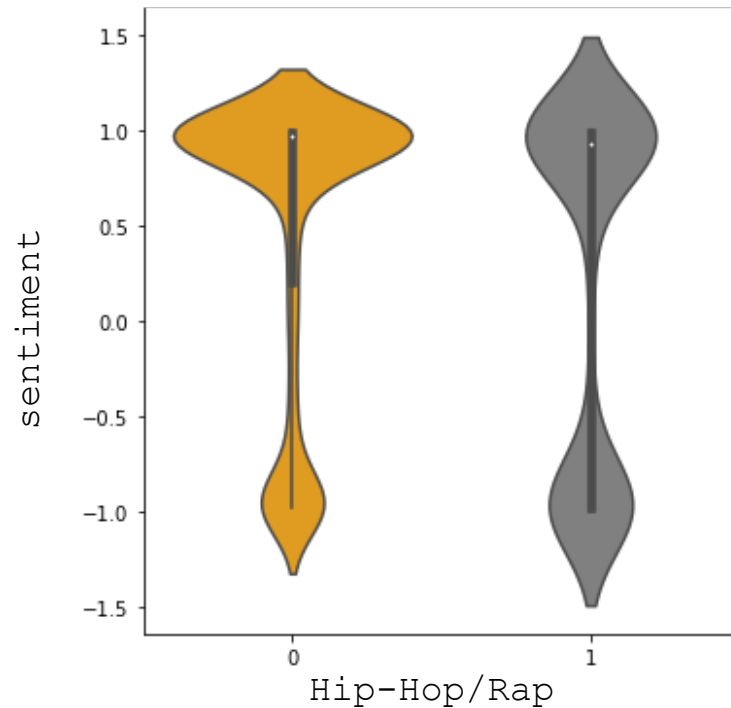
### Initial Visualizations



## Exploratory Data Analysis

$$\text{sentiment} = \beta_0 + \beta_1 \cdot \text{YEAR} + \beta_2 \cdot \text{profanity\_total} + \beta_3 \cdot \text{HipHop/Rap}$$

Initial Visualizations



### Modified Regression!

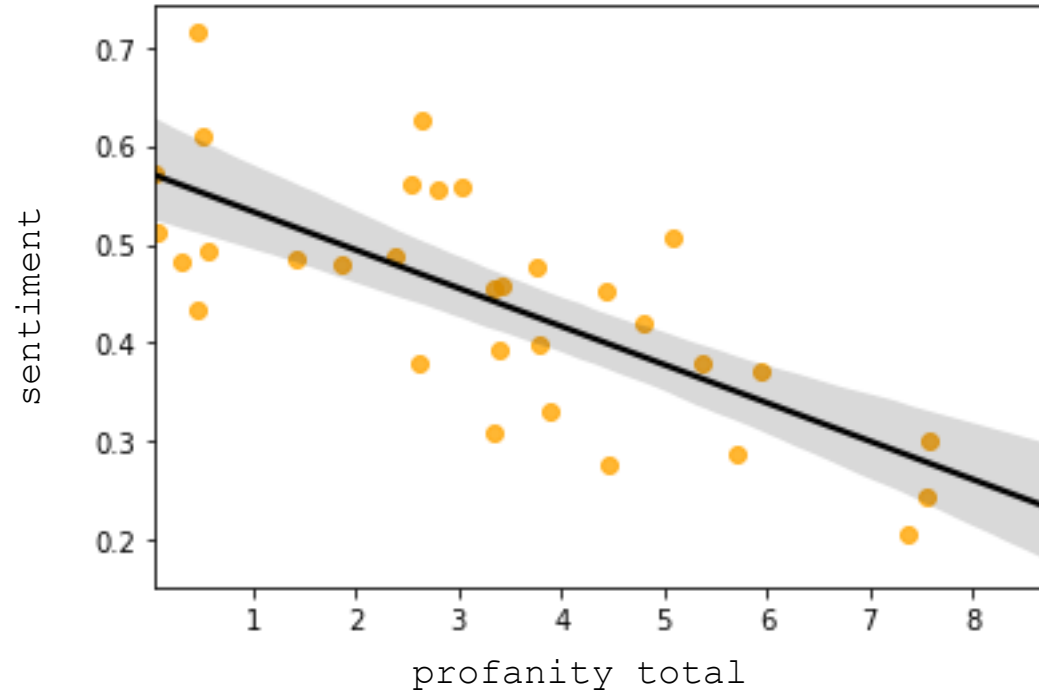
$$\text{sentiment} = \beta_0 + \beta_1 \cdot \text{YEAR} + \beta_2 \cdot \text{profanity\_total} + \beta_3 \cdot \text{HipHop/Rap}$$

- sentiment - average of sentiment scores of all songs for the year.
- YEAR - the year.
- profanity\_total - average # of profanity words per song for the year.
- Hip-Hop/Rap - # of hip-hop/rap songs in top 100 for the given year.

## Exploratory Data Analysis

$$\text{sentiment} = \beta_0 + \beta_1 \cdot \text{YEAR} + \beta_2 \cdot \text{profanity\_total} + \beta_3 \cdot \text{HipHop/Rap}$$

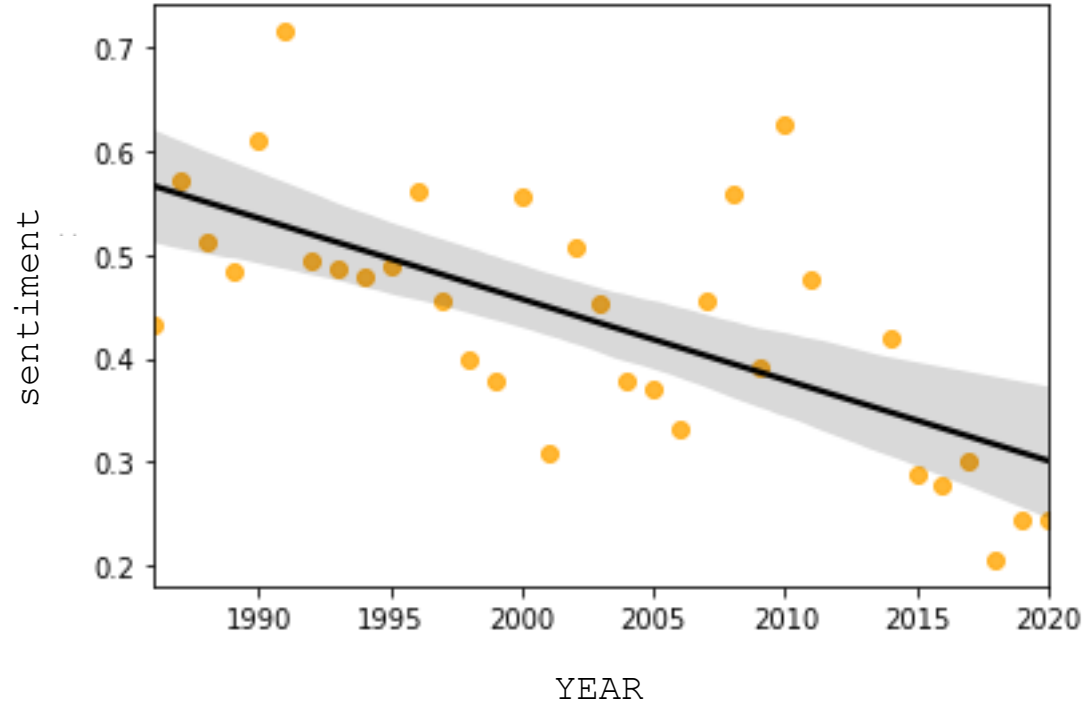
New Visualizations



## Exploratory Data Analysis

$$\text{sentiment} = \beta_0 + \beta_1 \cdot \text{YEAR} + \beta_2 \cdot \text{profanity\_total} + \beta_3 \cdot \text{HipHop/Rap}$$

### New Visualizations

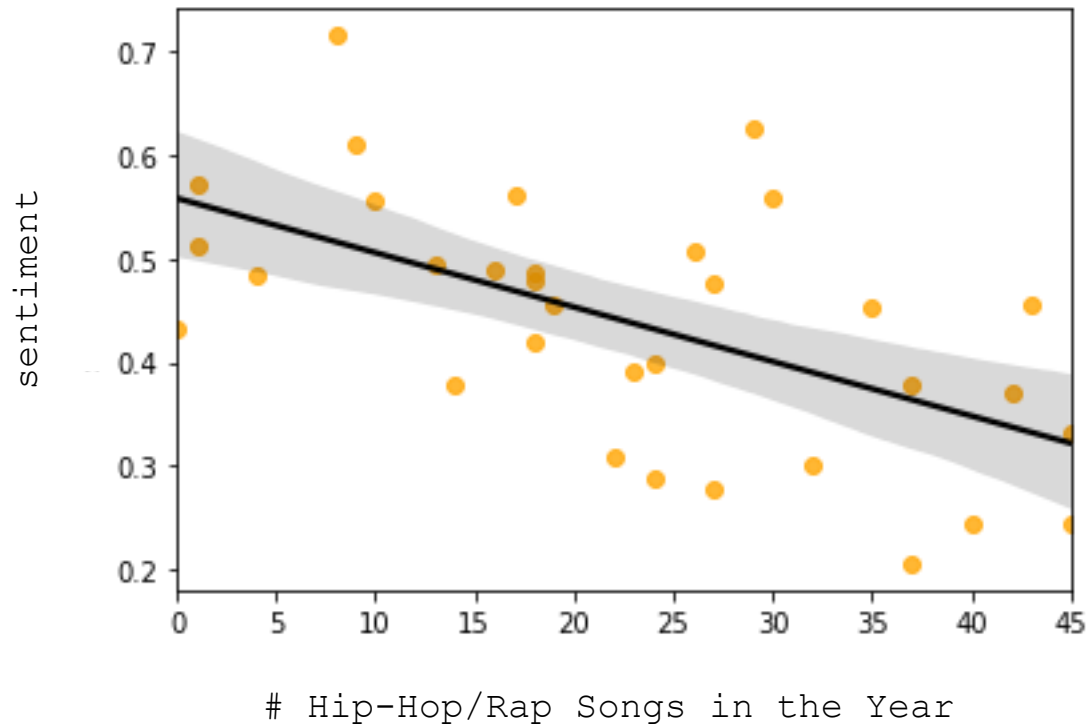




## Exploratory Data Analysis

$$\text{sentiment} = \beta_0 + \beta_1 \cdot \text{YEAR} + \beta_2 \cdot \text{profanity\_total} + \beta_3 \cdot \text{HipHop/Rap}$$

### New Visualizations



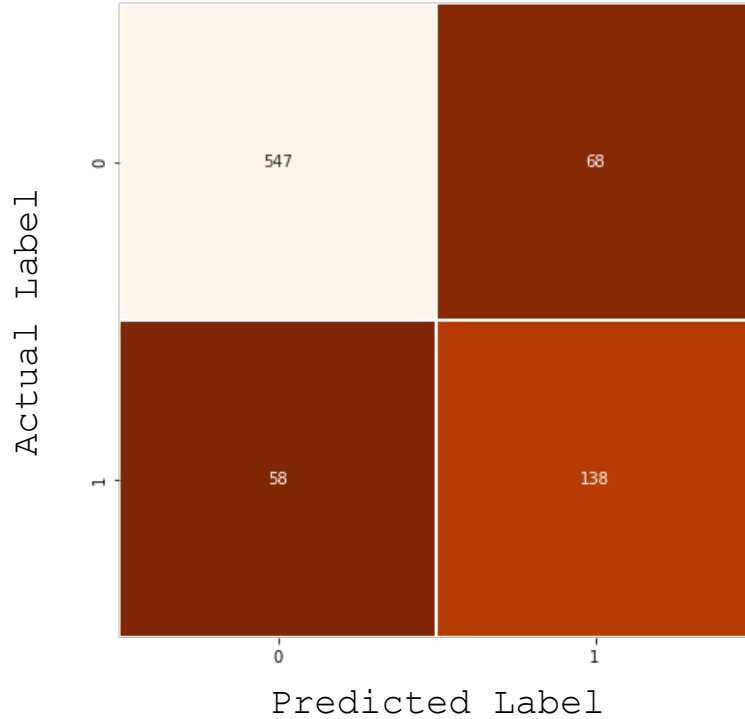
[illegible]



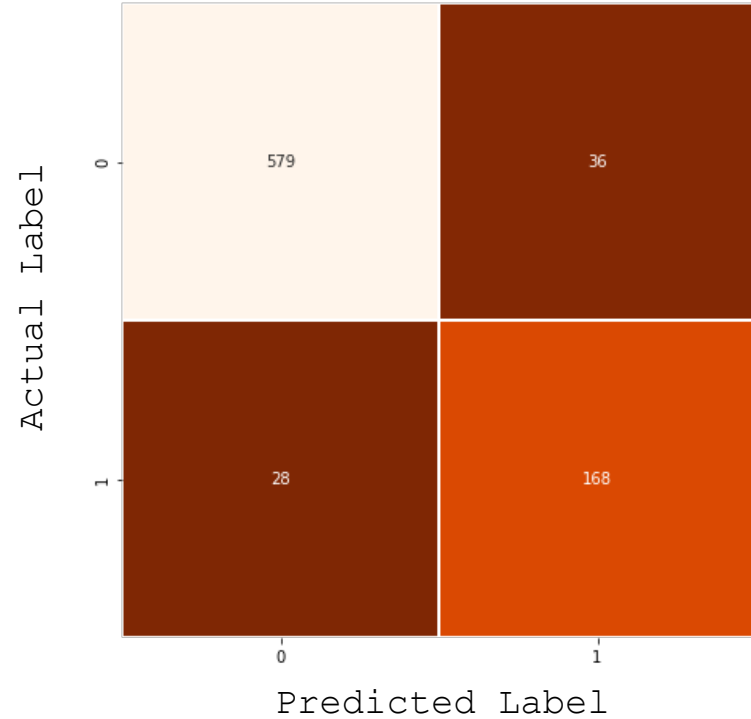
# Results

## Classification Results

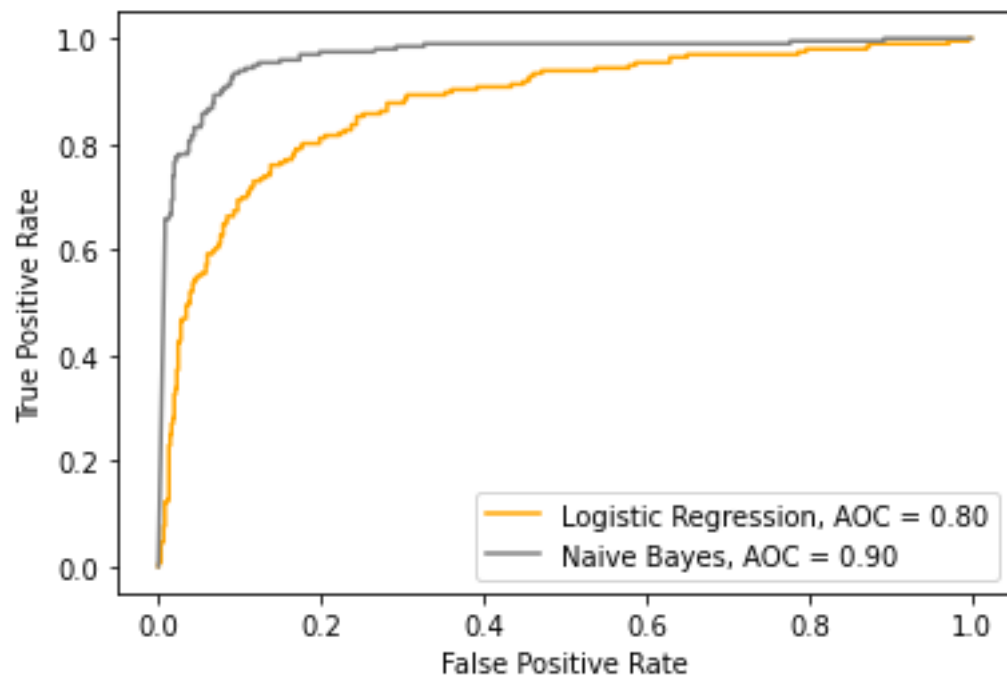
Logistic Regression  
Accuracy Score: **0.8446**



Naïve Bayes  
Accuracy Score: **0.9211**

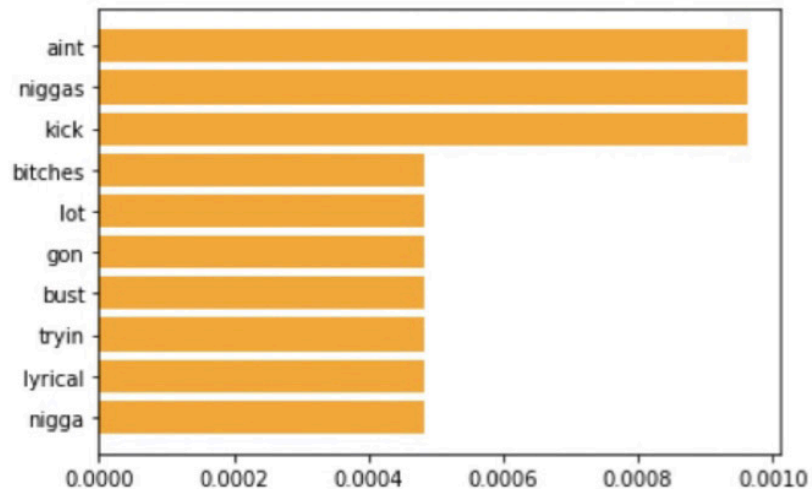


## CLASSIFICATION RESULTS



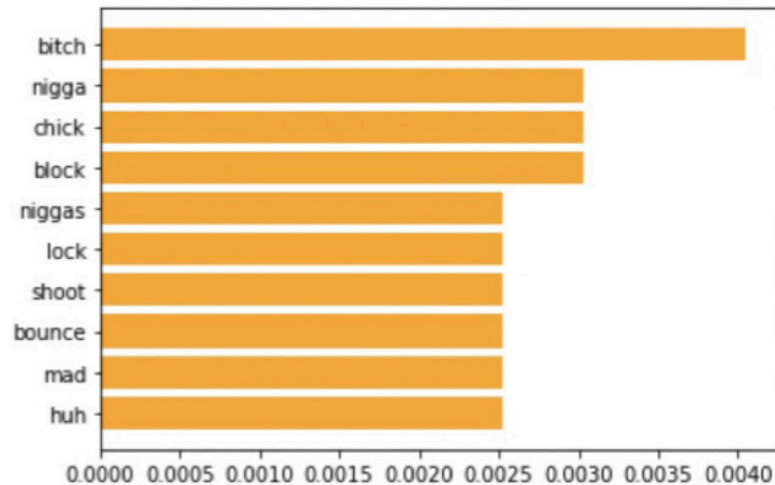
## CLASSIFICATION RESULTS

From Songs Before 2003



Feature Importances

From Songs Since 2003



Feature Importances

## Regression Results

**Original Regression:**  $\text{sentiment} = \beta_0 + \beta_1 \cdot \text{YEAR} + \beta_2 \cdot \text{profanity\_total} + \beta_3 \cdot \text{HipHop/Rap}$

**Modified Regression:**  $\text{sentiment} = \beta_0 + \beta_1 \cdot \text{YEAR}^* + \beta_2 \cdot \text{profanity\_total}^* + \beta_3 \cdot \text{HipHop/Rap}^*$   
, where \* indicates modified regression variables.

	Train R <sup>2</sup>	Test R <sup>2</sup>	Train MSE	Test MSE
Original	0.09328	0.04776	0.59890	0.64202
Modified	0.52258	0.60568	0.00634	0.00633

### Coefficient Estimates

	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
Original	5.4833	-0.0025	-0.0242	-0.0216
Modified	3.6188	-0.0015	-0.0303	-0.0002

## Topic Analysis Results

### All Hip-Hop/Rap Songs

	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
<b>Topic 1</b>	never	life	man	need	hey	tell	gotta	feel	put	could
<b>Topic 2</b>	bitch	shit	fuck	ayy	money	boy	walk	as	lil	gon
<b>Topic 3</b>	way	youre	hot	move	ride	jump	ill	body	shake	gonna

### All Non-Hip-Hop/Rap Songs

	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
<b>Topic 1</b>	baby	want	wanna	yeah	youre	got	ooh	girl	tell	cause
<b>Topic 2</b>	never	time	one	cant	say	youre	ill	away	let	way
<b>Topic 3</b>	yeah	get	got	let	come	back	aint	girl	hey	baby



## Topic Analysis Results

### Hip-Hop/Rap Songs Before 2007

	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
<b>Topic 1</b>	youre	cant	ill	need	boy	tell	gotta	thing	keep	life
<b>Topic 2</b>	nigga	shit	bitch	fuck	as	jump	hot	wit	money	give
<b>Topic 3</b>	move	hey	ride	bring	wild	woo	booty	play	hit	work

### Hip-Hop/Rap Songs Since 2007

	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
<b>Topic 1</b>	bitch	ayy	fuck	money	ooh	woo	lil	hey	put	big
<b>Topic 2</b>	baby	life	time	come	right	feel	youre	girl	watch	way
<b>Topic 3</b>	girl	baby	walk	talk	way	right	low	hot	club	take

# Conclusions

## Conclusions

- Average sentiment decreases by  $\sim 0.35\%$  each year since 1986.
- With each additional hip-hop/rap song, average sentiment decreases by  $\sim 0.05\%$ .
- With Naïve Bayes classifier, predict genre with  $> 92\%$  accuracy.
- Little changes in content/defining characteristics of hip-hop/rap music.

# Limitations

### Manually Genre Labelling:

- Which genre was the right one to choose?

### Sentiment Scores

- Can VADER capture slang, alternate meanings, or misspelled words?

## Future Work

### Follow-up Projects:

- Predicting Genre/Sentiment using different features
- Clustering

### Long-term:

- Effect of hip-hop/rap music on society (youth)

### Sources of Data/Code

- [https://en.wikipedia.org/wiki/Billboard\\_Year-End](https://en.wikipedia.org/wiki/Billboard_Year-End)
- <https://genius.com/>
- <https://github.com/AntiDeprime/billboard-year-end-wiki-scraper>
- <https://kvsingh.github.io/lyrics-sentiment-analysis.html>
- <https://github.com/johnwmillr/LyricsGenius>



**THANKS!**

*Any questions?*