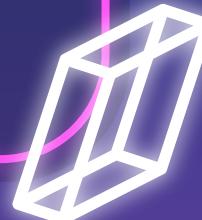


BOPPiNg

Through the Decades

Chad Richter
DSI907



TRACKLIST

01

PROBLEM
STATEMENT

02

DATA COLLECTION
& CLEANING

03

ANALYSIS

04

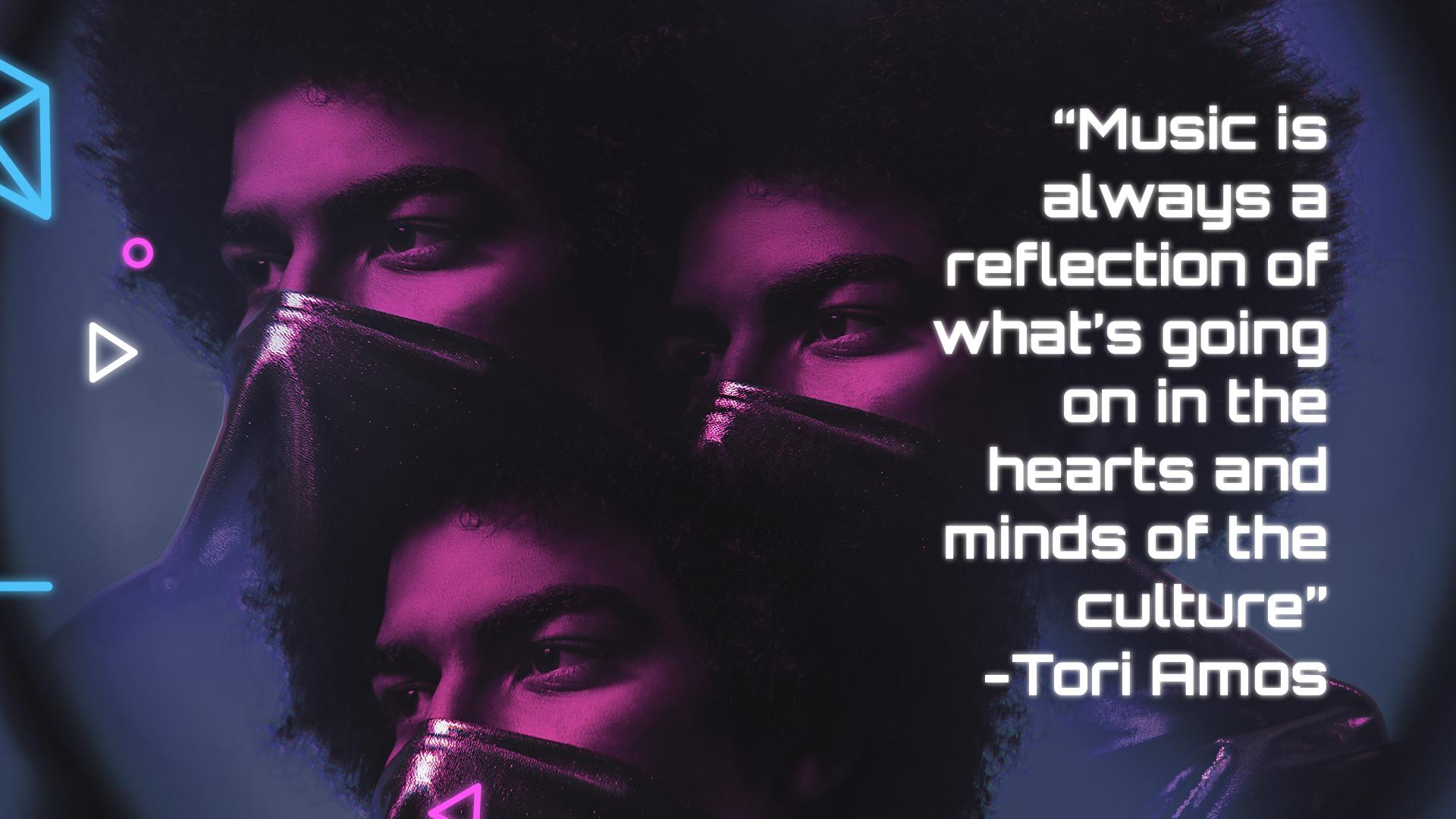
MODELING



01

PROBLEM STATEMENT

“I got ninety-nine problems”
JAY-Z



**"Music is
always a
reflection of
what's going
on in the
hearts and
minds of the
culture"**

-Tori Amos

“

“The times they are a-changin”

-BOB DYLAN

PROBLEM STATEMENT

Can the **lyrics** of our most popular songs be used determine which **decade** they were released?



02

DATA COLLECTION & CLEANING

“So fresh and so clean, clean”
-OUTKAST

COLLECTING THE SONGS

Scrape Billboard Top 100

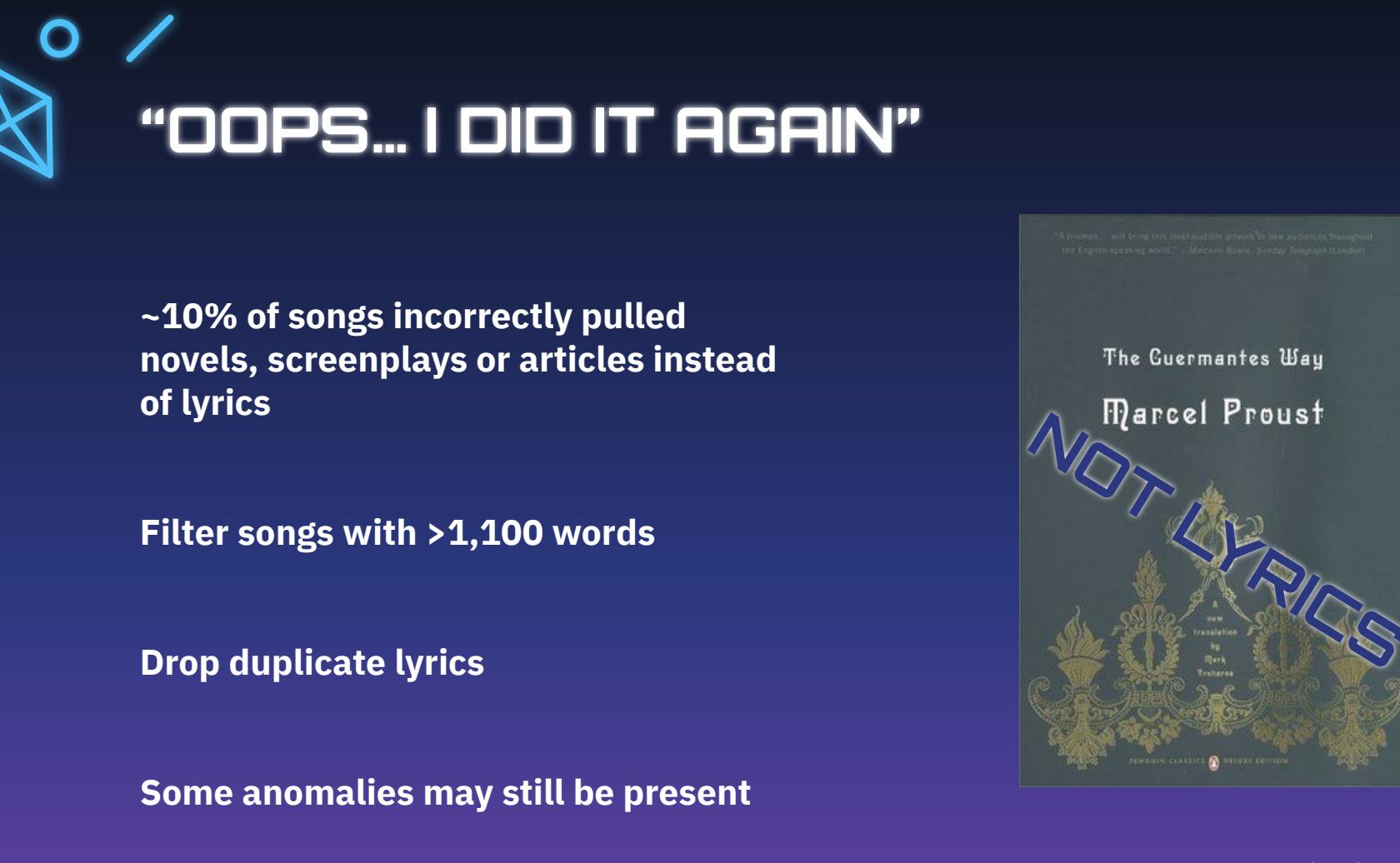
1960 - 2009

Genre Agnostic

Lyric Genius API



IMAGE: Wikimedia Commons & Rolling Stone



“OOPS... I DID IT AGAIN”

~10% of songs incorrectly pulled novels, screenplays or articles instead of lyrics

Filter songs with >1,100 words

Drop duplicate lyrics

Some anomalies may still be present

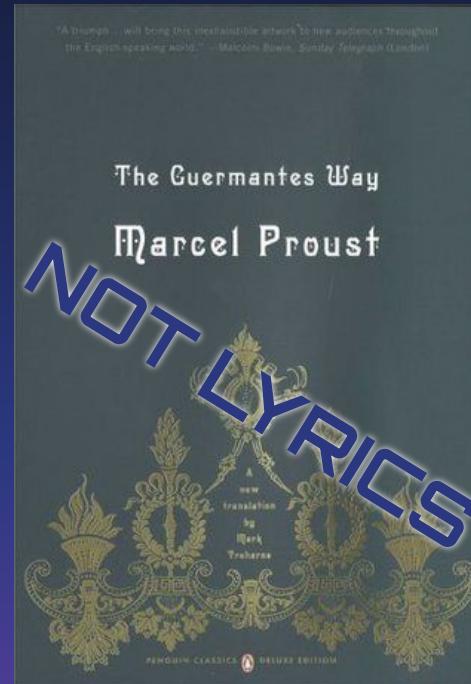
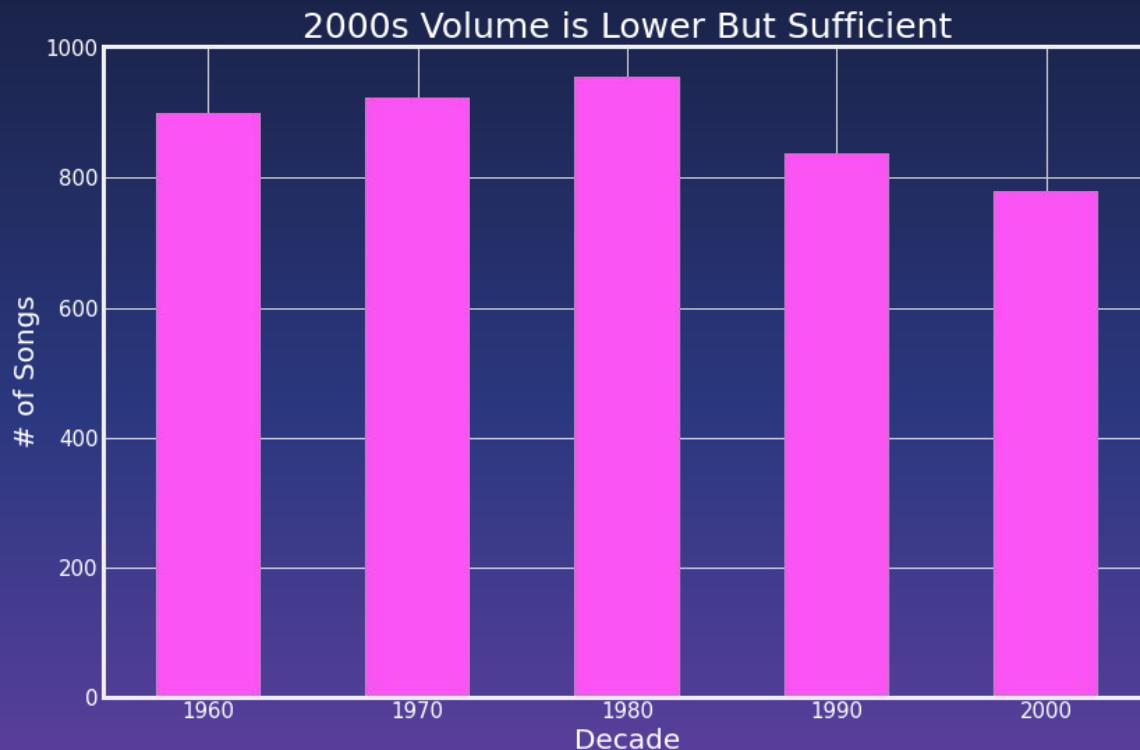


IMAGE: GoodReads

ALL OF THE HITS



FEATURE ENGINEERING

Total number of words in
the song

WORD COUNT

Whether the lyrics are
happy or sad

**SENTIMENT
ANALYSIS**

Average number of letters
per word in the song

**AVERAGE
WORD LENGTH**

Are the lyrics profane,
derogatory or suggestive

DIRTY?

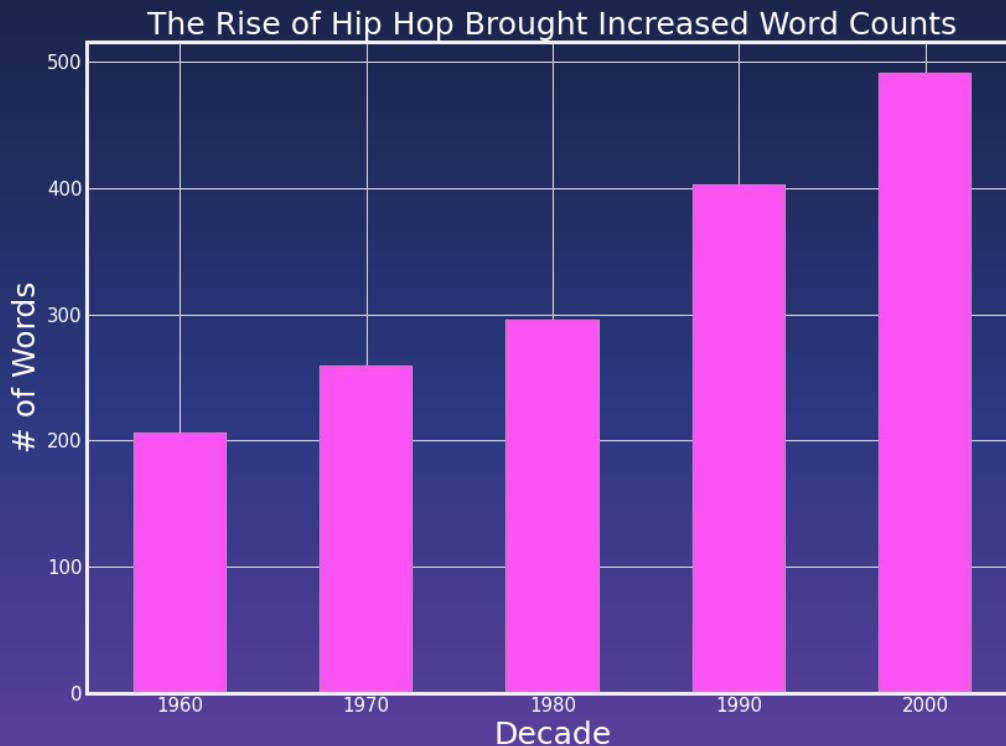
03

ANALYSIS

*“Gonna take some time to do the
things we never had”*
-TOTO / WEEZER



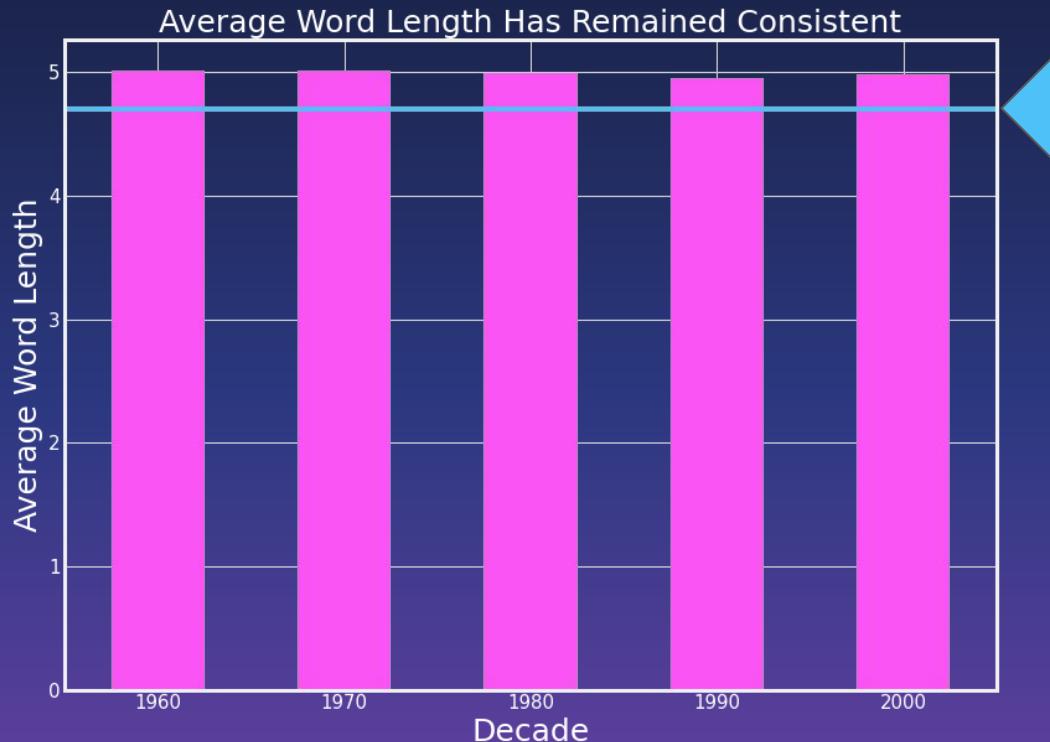
WORD COUNT OVER THE DECADES



*“Can I say my sh*t”*
-KANYE WEST



AVERAGE WORD LENGTH

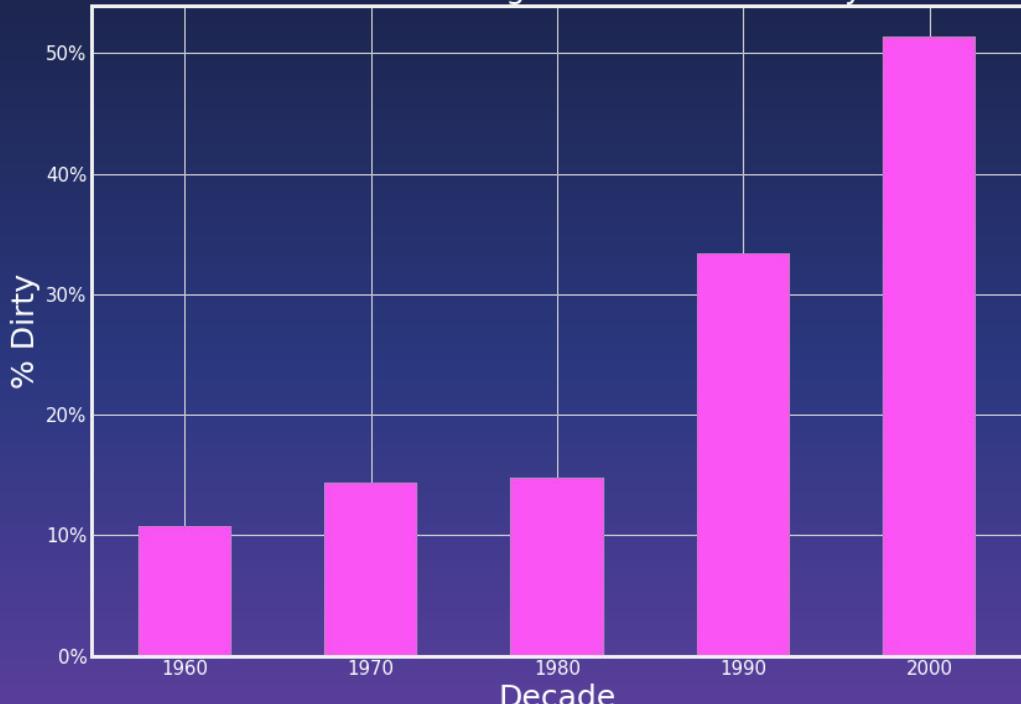


English Average

*“Same old song and
dance, my friend”*
-AEROSMITH

DIRTY LYRICS OVER THE DECADES

Over Half of Songs After 2000 Are Dirty



850 Words Considered:

Profane

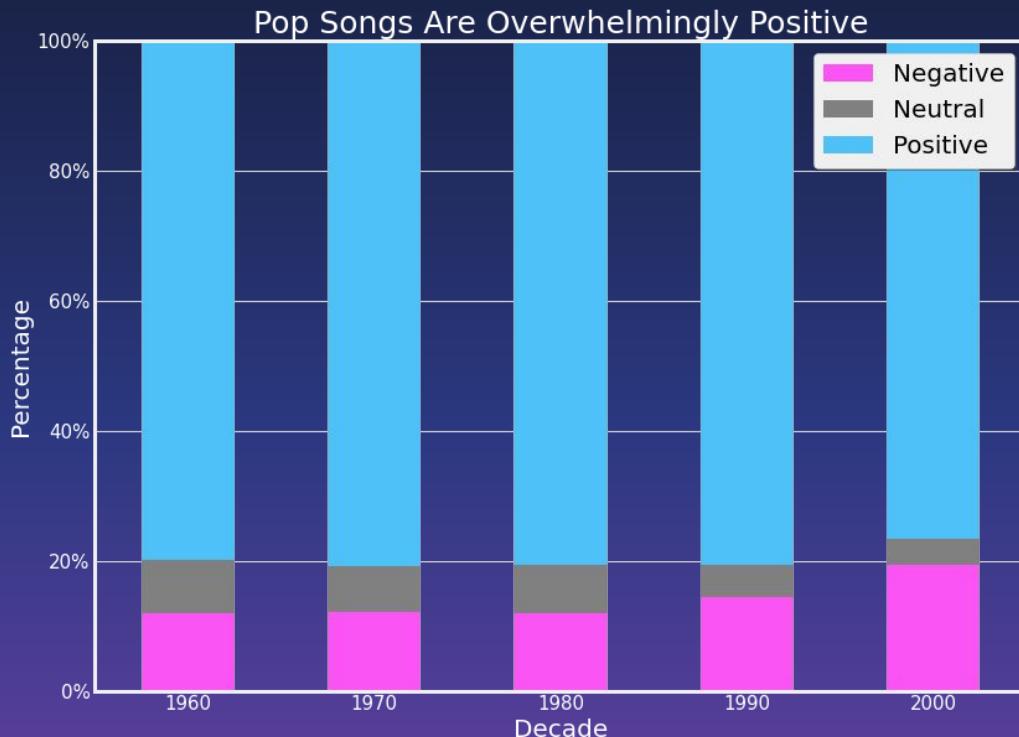
Derogatory

Suggestive

*"Because I'm bad, I'm
bad"*

-MICHAEL JACKSON

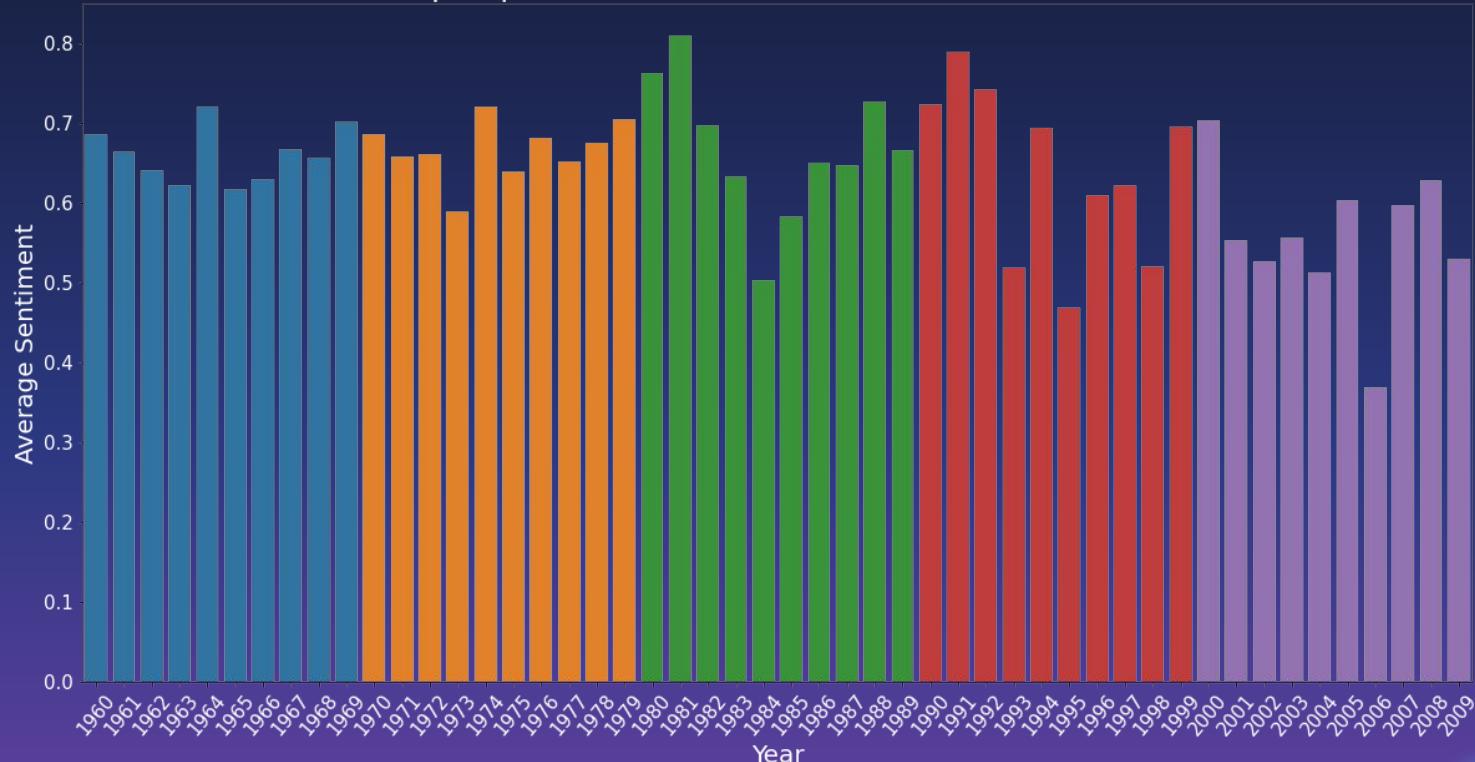
POSITIVE AND NEGATIVE SENTIMENT



"Why can't you just be happy?"
-TAKING BACK SUNDAY

SENTIMENT BY YEAR

Hope Sprints Eternal At the Turn of a New Decade



04

MODELING

*"I shake my little tush on the
catwalk"*

-RIGHT SIDE FRED

21.7%

BASELINE



MODEL PERFORMANCE



ACCURACY

WORDS

INTERPRET?

Bayes

37.7%

2,000

Low

Regression

37.6%

1,000

High

Boosting

43.2%

5,000

Medium



MOST IMPACTFUL FEATURES

y=1960s top features		y=1970s top features		y=1980s top features		y=1990s top features		y=2000s top features	
Weight?	Feature	Weight?	Feature	Weight?	Feature	Weight?	Feature	Weight?	Feature
+4.982	<BIAS>	+2.306	<BIAS>	+0.646	hee	+1.247	word_count	+1.612	word_count
+0.663	broke	+0.621	bell	+0.608	crack	+0.524	jam	+0.697	shorty
+0.557	dear	+0.517	drop	+0.500	dancing	+0.521	cuz	+0.639	american
+0.548	trick	+0.499	mother	+0.499	miracle	+0.495	wishing	+0.627	club
... 462 more positive ...		+0.491	dancin	... 530 more positive ...		+0.492	shoo	+0.600	tryna
... 527 more negative 548 more positive 459 more negative ...		+0.450	blame	+0.556	damn
-0.532	like	... 441 more negative ...		-0.475	fallin	+0.439	smooth	+0.552	shawty
-0.537	quick	-0.478	watch	-0.479	rolling	... 519 more positive ...		+0.535	steady
-0.548	damn	-0.479	fine	-0.500	park	... 470 more negative 522 more positive ...	
-0.556	passion	-0.479	scream	-0.502	record	-0.424	heartache	... 467 more negative ...	
-0.584	wishing	-0.485	na	-0.502	dad	-0.431	butterfly	-0.509	hungry
-0.621	rainy	-0.514	kill	-0.520	lord	-0.453	seat	-0.532	running
-0.665	keeping	-0.527	dit	-0.552	bell	-0.539	chick	-0.538	love
-0.685	crack	-0.531	bum	-0.556	drop	-0.565	bar	-0.559	honey
-0.767	remember	-0.614	sick	-0.558	sittin	-0.620	ring	-0.740	dawn
-0.781	wanna	-0.643	green	-0.561	mother	-0.662	pick	-0.795	dancing
-1.769	word_count	-1.020	word_count	-0.787	broke	-2.495	<BIAS>	-4.638	<BIAS>

TOP 3 MOST PREDICTIVE WORDS

1960s

Broke
Wanna
Remember

1970s

Bell
Green
Sick

1980s

Hee
Crack
Broke

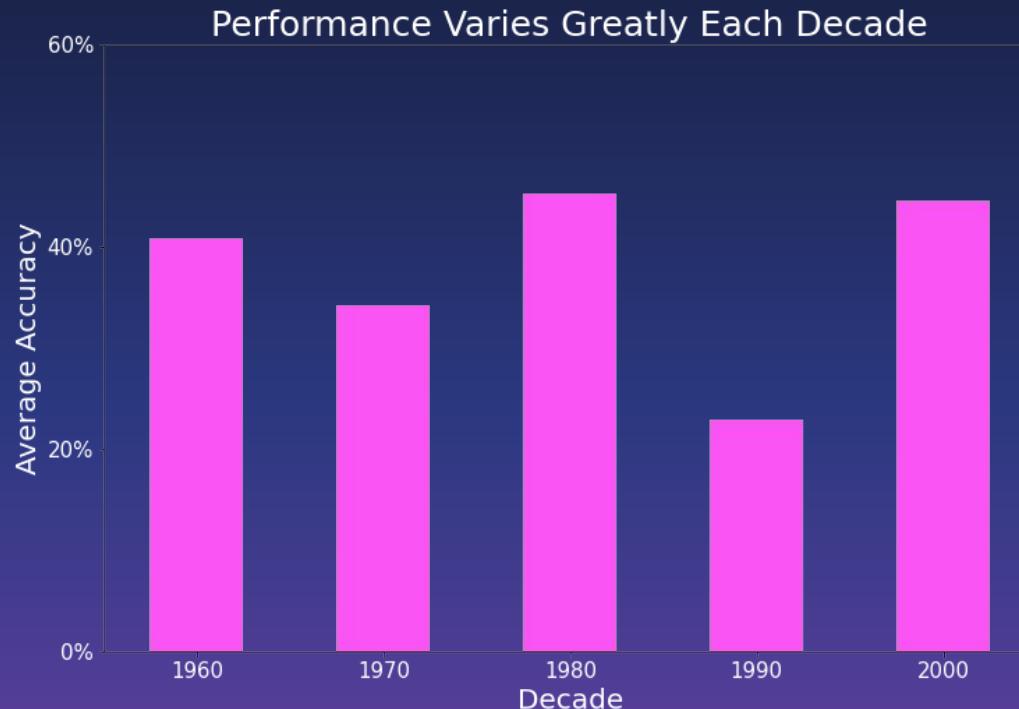
1990s

Jam
Pick
Ring

2000s

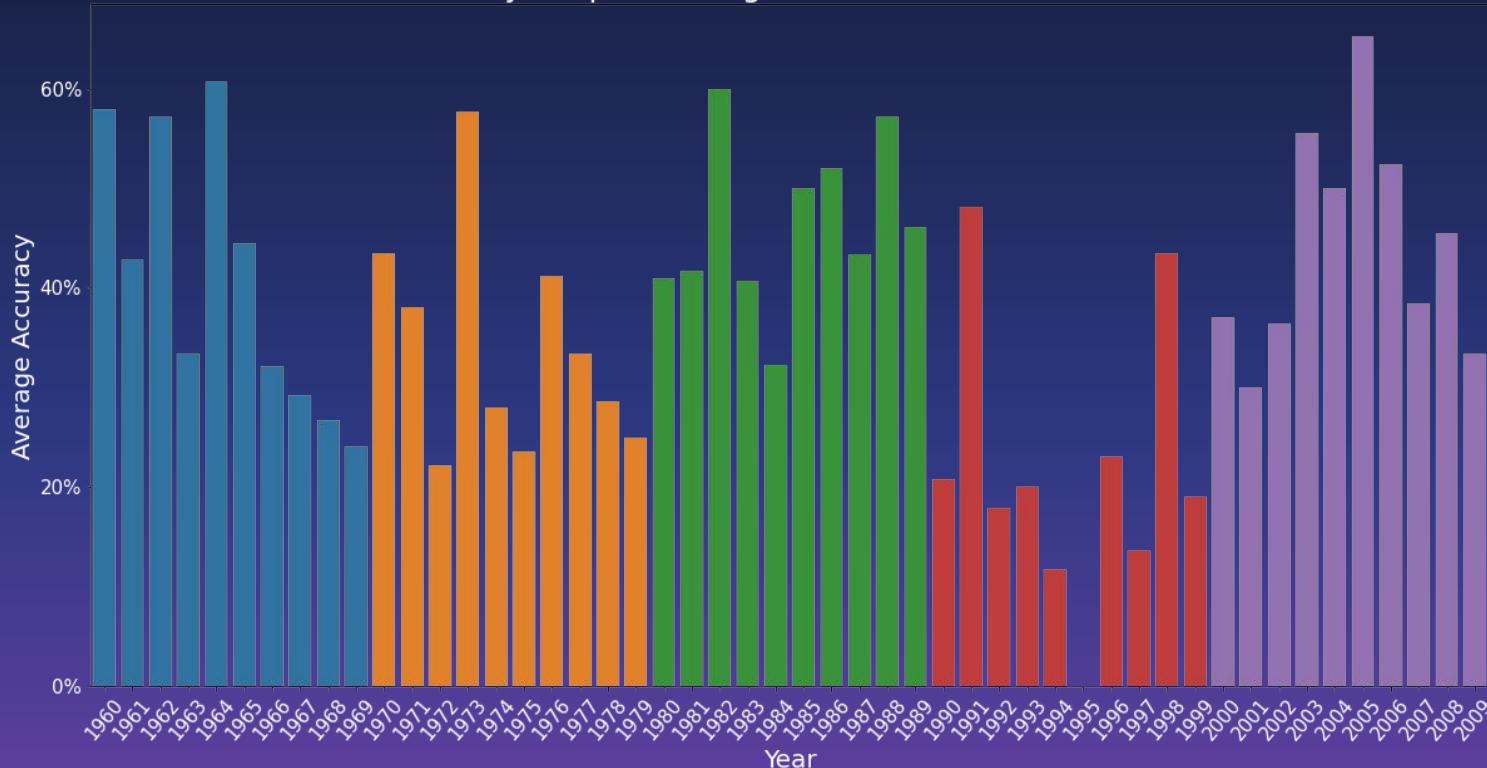
Shorty
Club
Dancing

BEST MODEL - ACCURACY BY DECADE

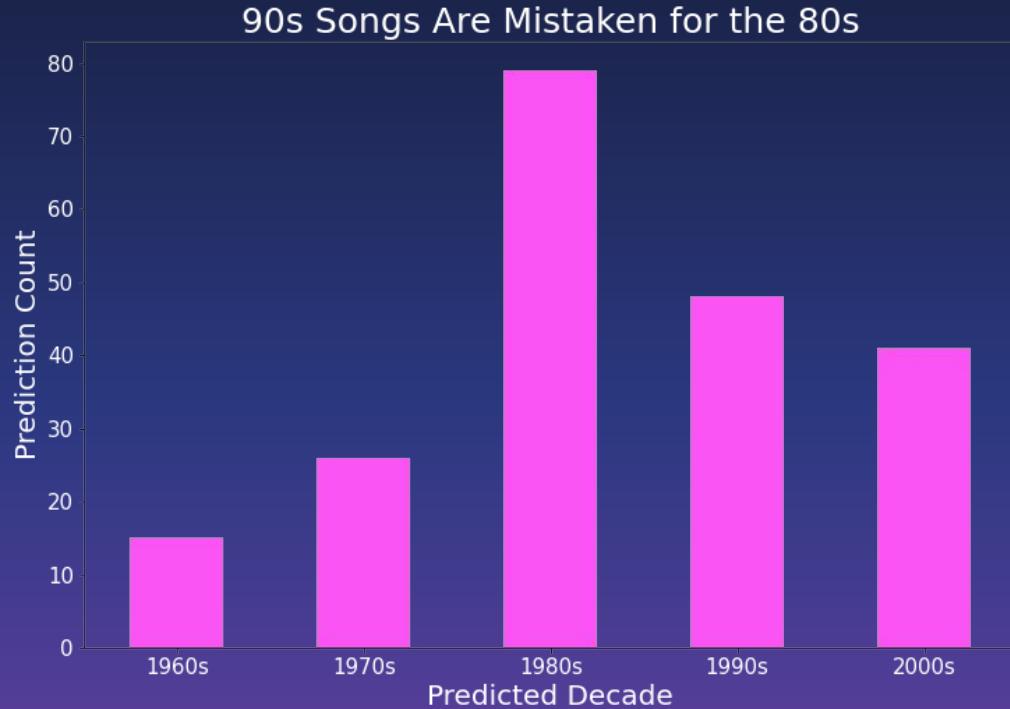


ACCURACY BY YEAR

Accuracy Drops As Songs Near the Turn of a Decade



WHAT HAPPENED TO THE 90s?

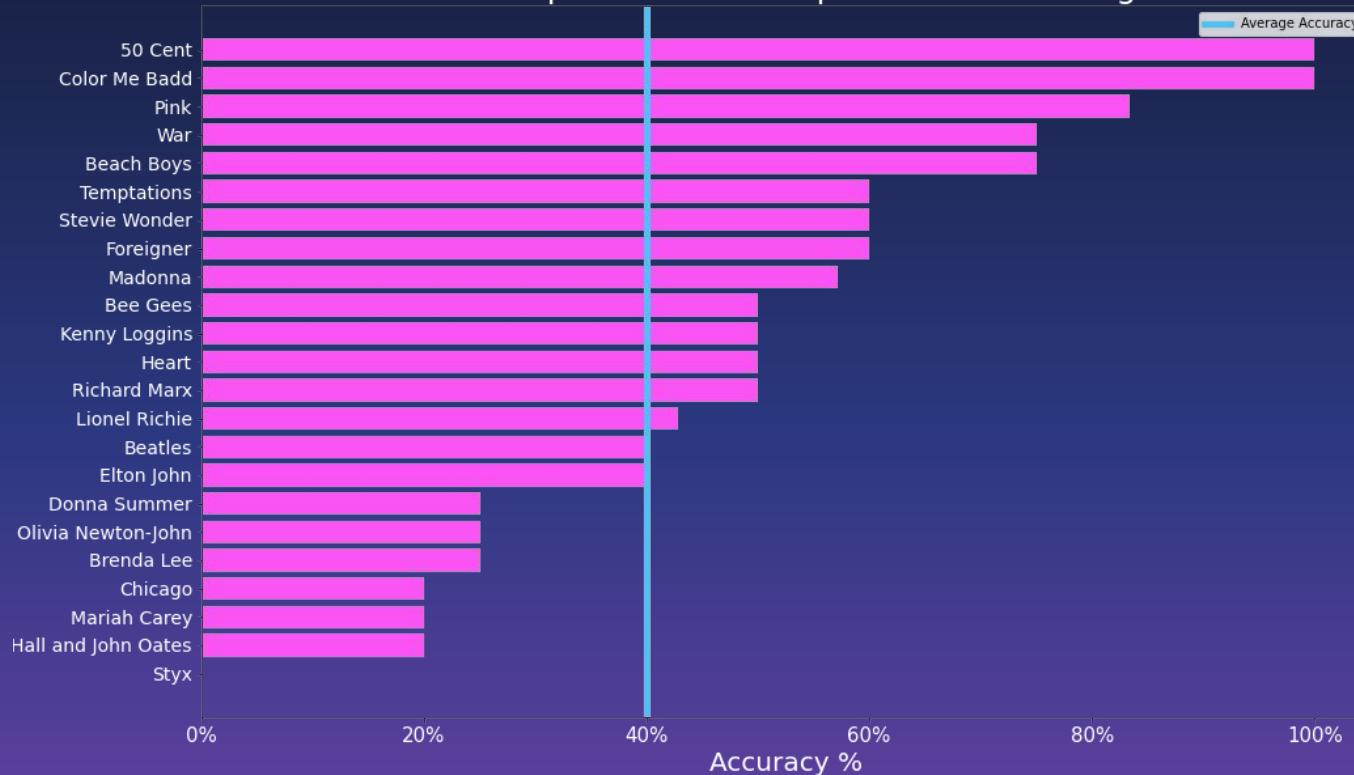


*"'Cause she's still
preoccupied with
1985"*
-BOWLING FOR SOUP



ARTISTS WITH > 3 TOP 100 Songs

More Popular Artists Outperform the Average



04

CONCLUSION

“This is the final countdown”
-EUROPE

RECALL OUR PROBLEM STATEMENT

Can we use the **lyrics** of a song
determine in which **decade** it was
released?



RECALL OUR PROBLEM STATEMENT

Can we use the **lyrics** of a song determine in which **decade** it was released?



KEY FINDINGS

WORD
COUNT IS
INCREASING

LYRICS ARE
BECOMING
DIRTIER

WORDS
INVENTED
OR FALL
OUT OF USE

POPULAR
ARTISTS
ARE
DISTINCT

POP SONGS
ARE
GENERALLY
HAPPY

LYRICS NEAR
THE TURN OF A
DECADE ARE
DIFFICULT TO
PREDICT

THANK YOU

Q&A

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.



Fonts & colors used

This presentation has been made using the following fonts:

Orbitron

(<https://fonts.google.com/specimen/Orbitron>)

IBM Plex Sans

(<https://fonts.google.com/specimen/IBM+Plex+Sans>)

#fda7f9

#fb54f4

#4fc1f9

#5f3f9e