



THE ONE STOP FOR EVERYTHING MUSIC

# SPARK-Y ROCK ANTHEM

Presented by Zaid, Khalid, Mahasmrti, Prajna, Niegil

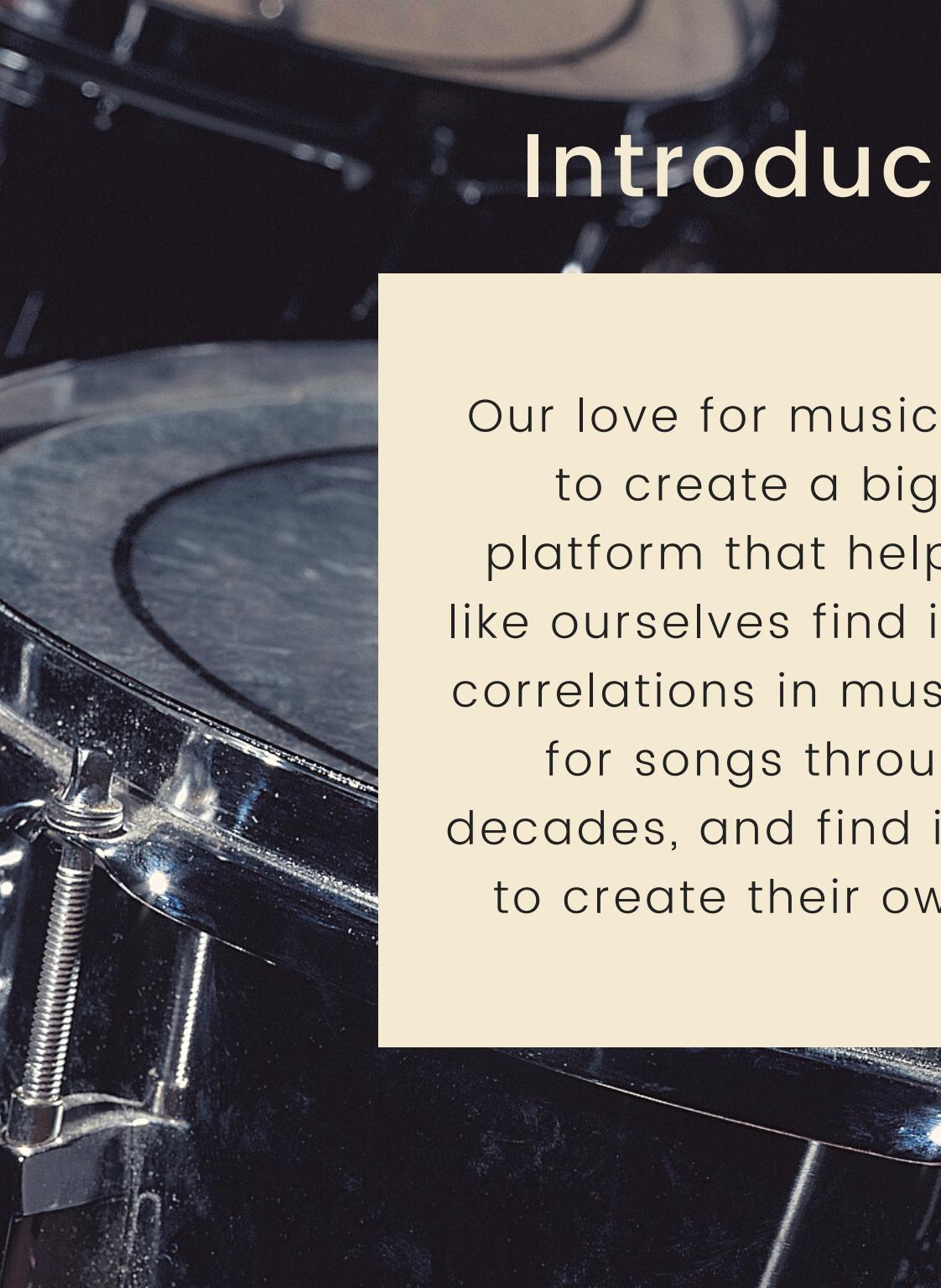




# Topics

- 
- Introduction
  - Project Flow
  - Data Scraping
  - Pre-processing
  - EDA
  - Song Search
  - Lyric Generator
  - Conclusion
  - Future Improvements

# Introduction



Our love for music drove us to create a big-data platform that helps others like ourselves find interesting correlations in music, search for songs through the decades, and find inspiration to create their own music



# PROJECT FLOW

## DATA SCRAPER

Top Albums &  
Artists  
Spotify track  
data  
Lyric data

## PRE PROCESS

General  
processing and  
cleaning of lyrics

## EDA

Relationships  
between  
ratings,  
popularity,  
followers, and  
song features

## SONG SEARCH

MinHashLSH to  
search for  
songs via lyric  
input

## LYRIC GEN

Generate  
lyrics for any  
song, artist or  
album via  
exact search

# DATA SCRAPING

---

---

17 GB of data

Years 2000–2023

5 charts - 1040 albums

812577 songs

3104 genres

646480 lyrics



## SOURCES

Rate-your-music -> album, artists

Spotify -> album, artist, track data

Genius Lyrics -> lyrics

Vagalume -> lyrics

Song-List -> artists/songs

A-Z lyrics -> artists/lyrics

# PRE-PROCESSING

## General

REMOVING DUPLICATE DATA

HANDLING OTHER NANS

## Lyrics

STANDARDIZING LYRIC FORMAT

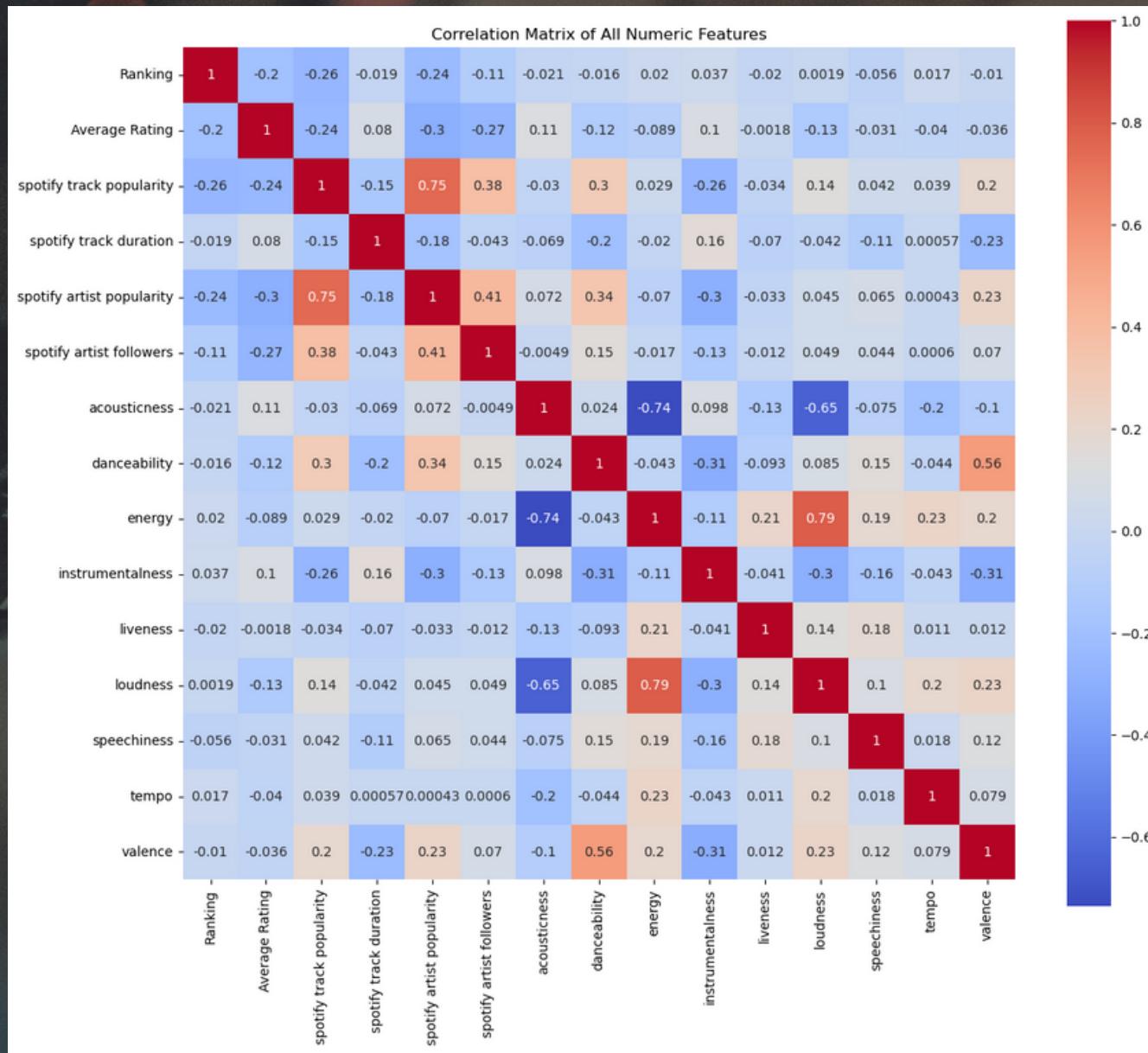
REMOVING STOPWORDS

TOKENIZING SENTENCES

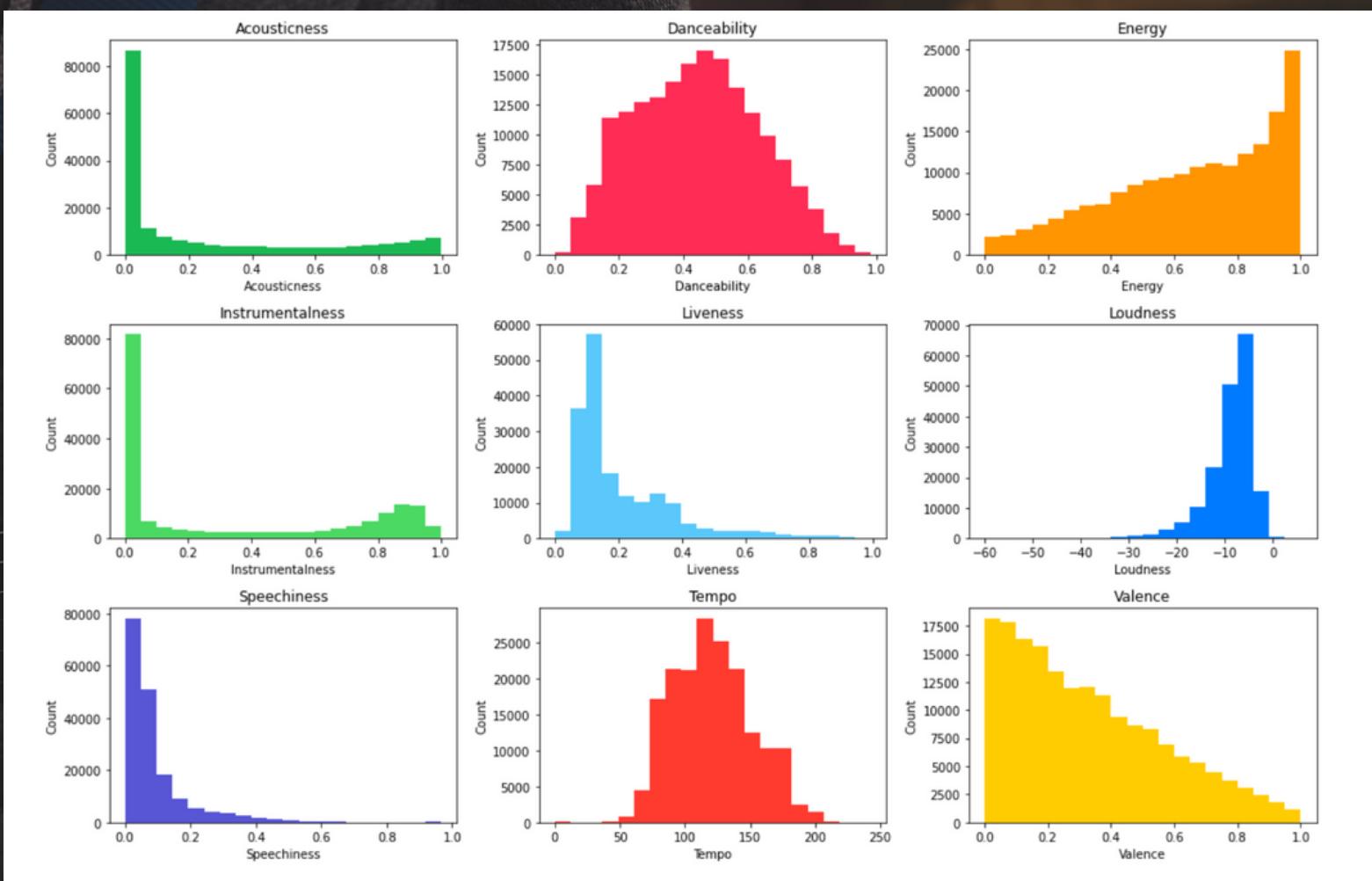
VECTORIZATION

# EDA

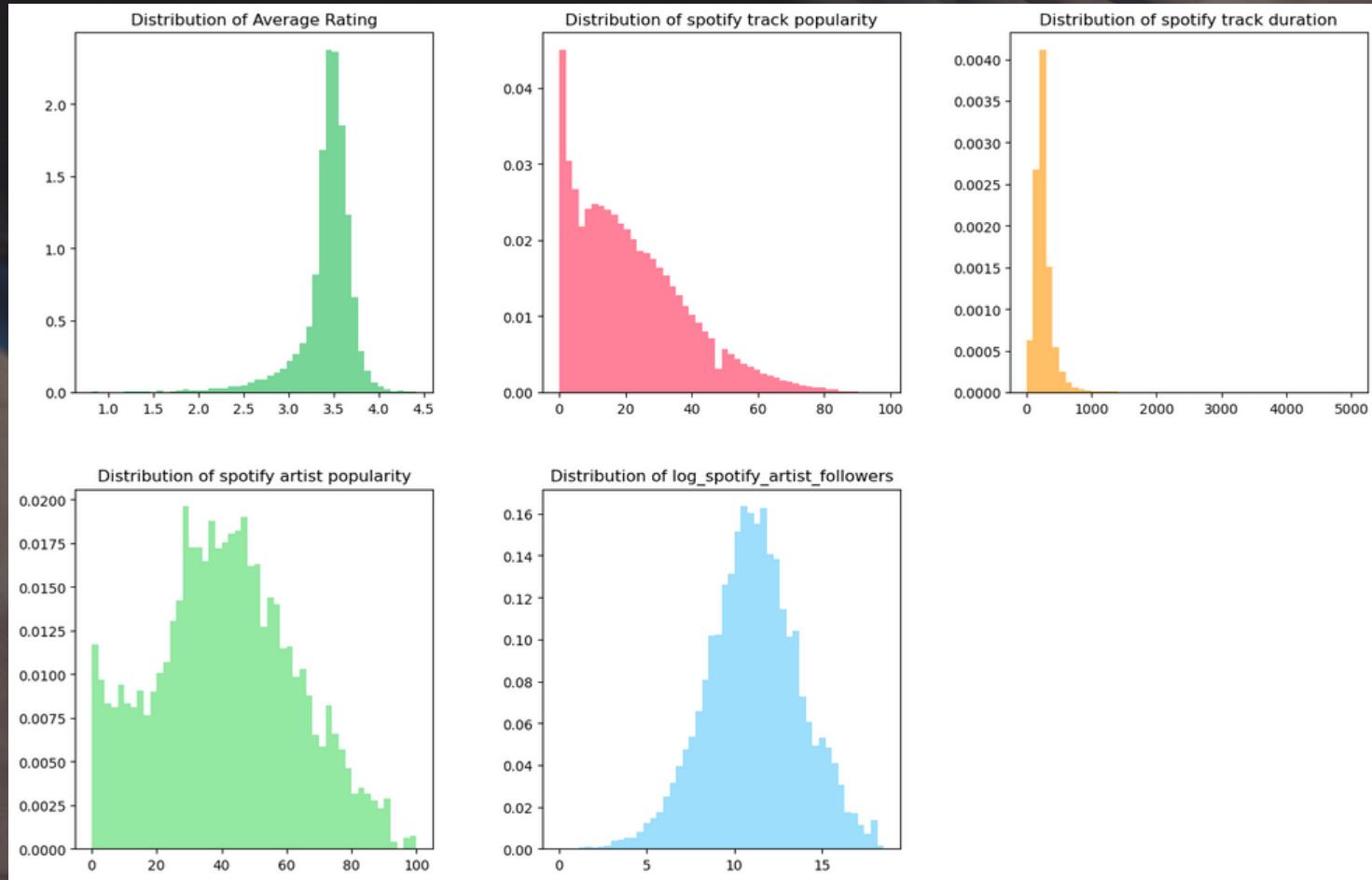
## RELATIONSHIPS YOU NEVER THOUGHT EXISTED



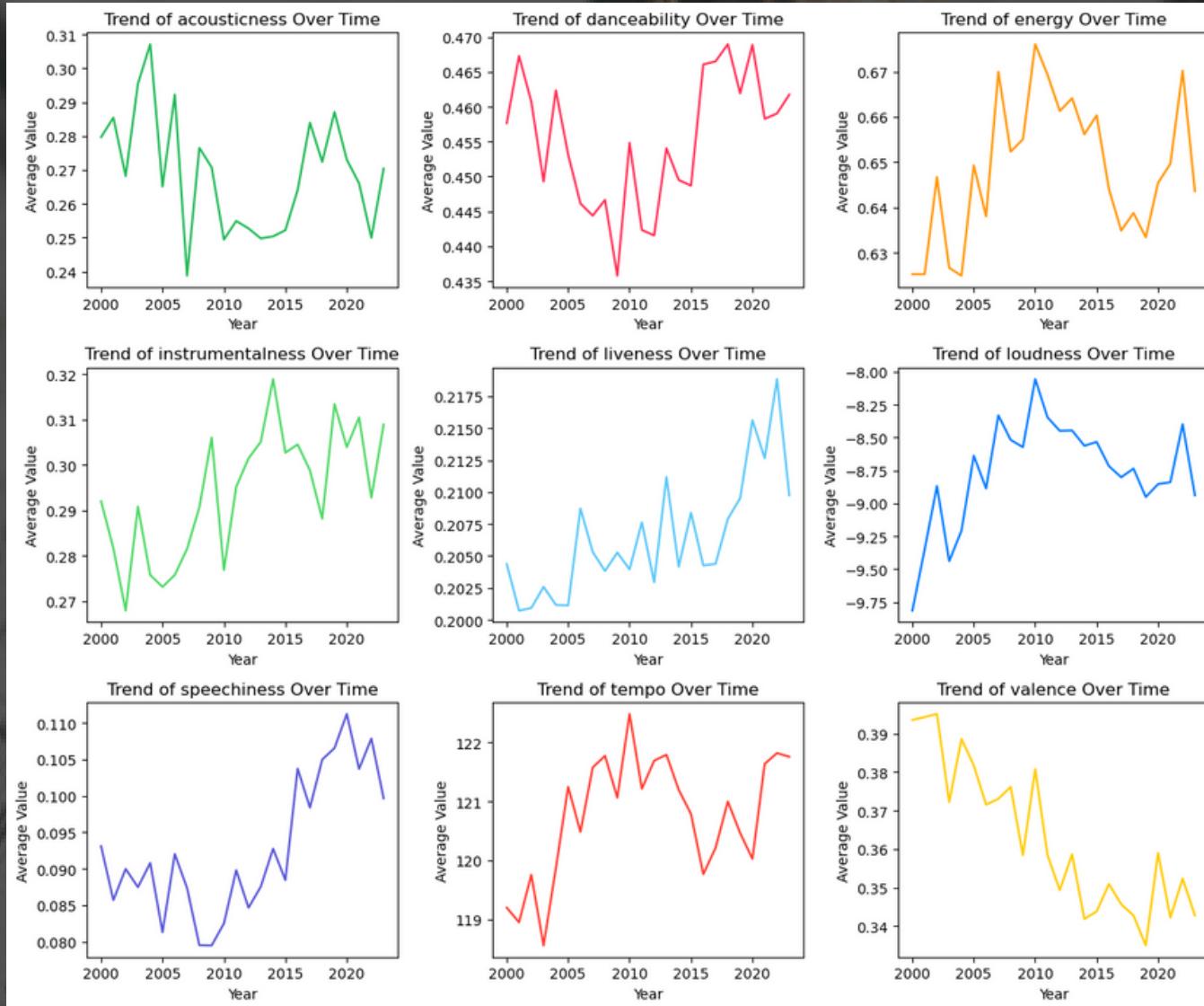
# DISTRIBUTION OF AUDIO FEATURES



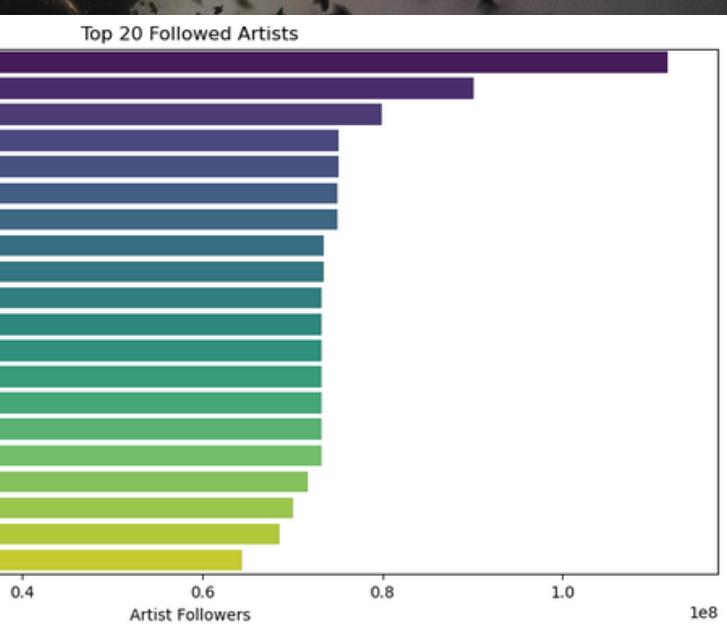
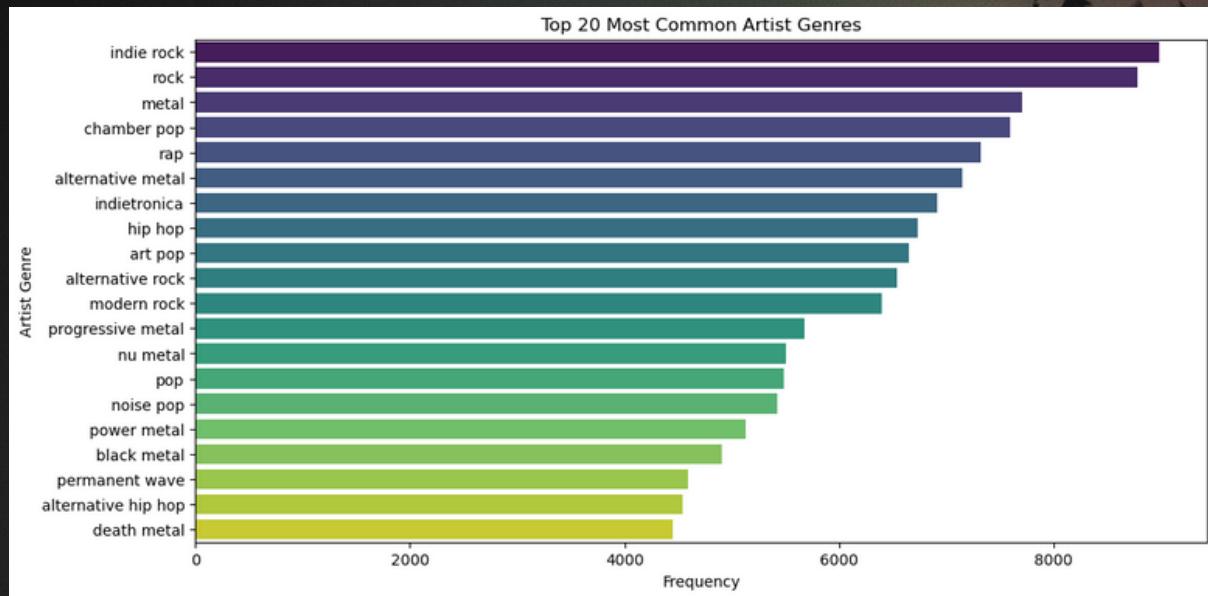
# DISTRIBUTION OF RATINGS, POPULARITY, ARTIST FOLLOWERS, TRACK DURATION



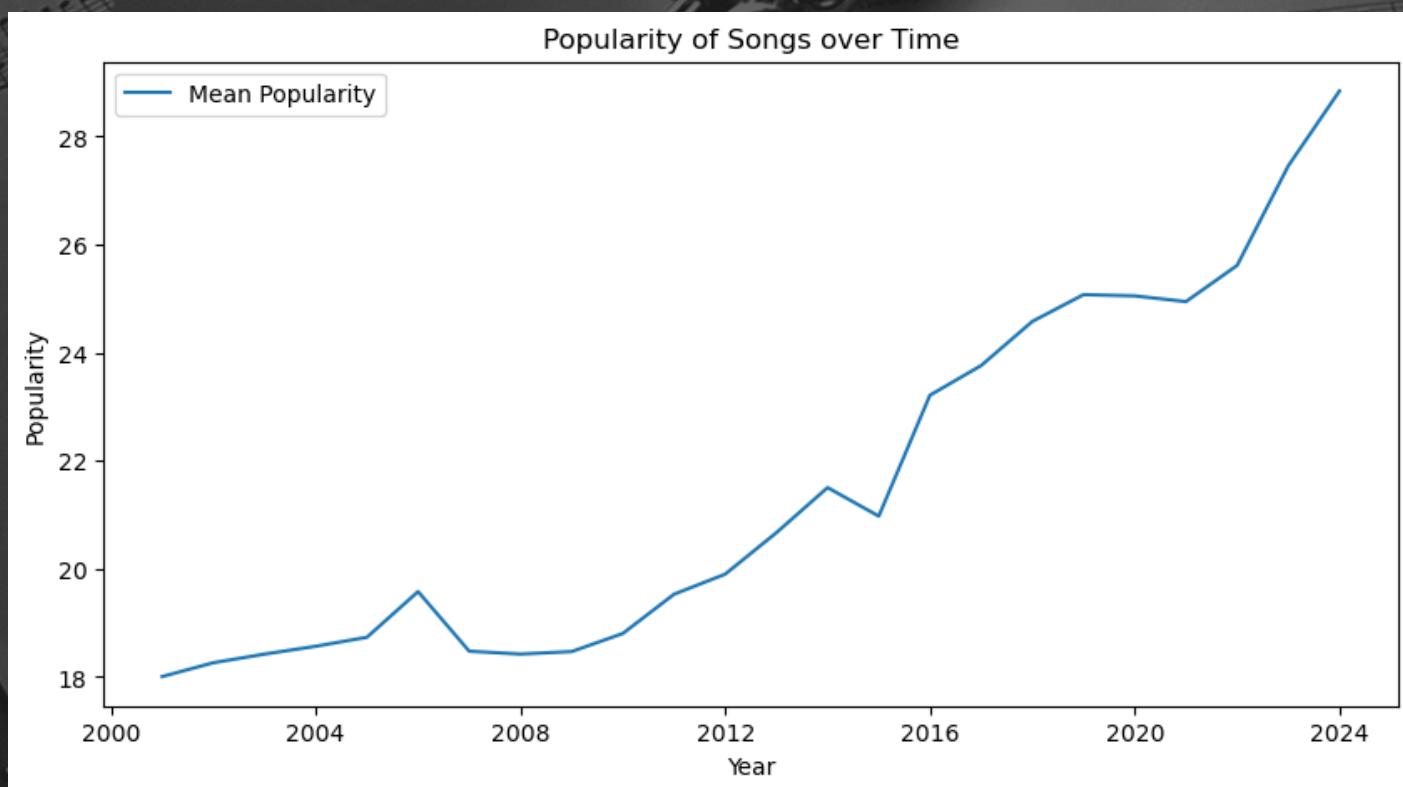
# HOW DID THE MUSICAL FEATURES VARY OVER TIME?



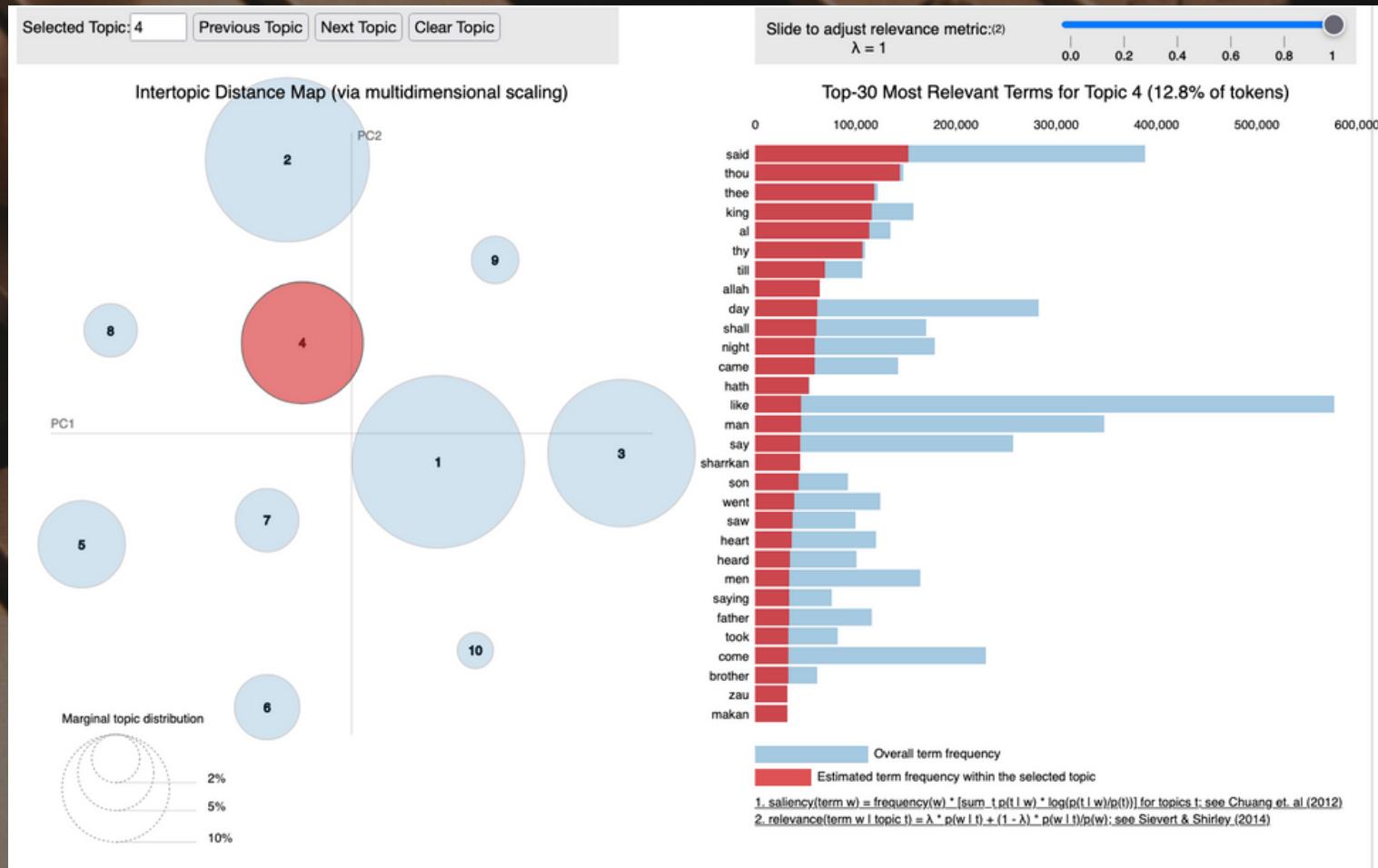
# HOW DID THE ARTISTS DO?



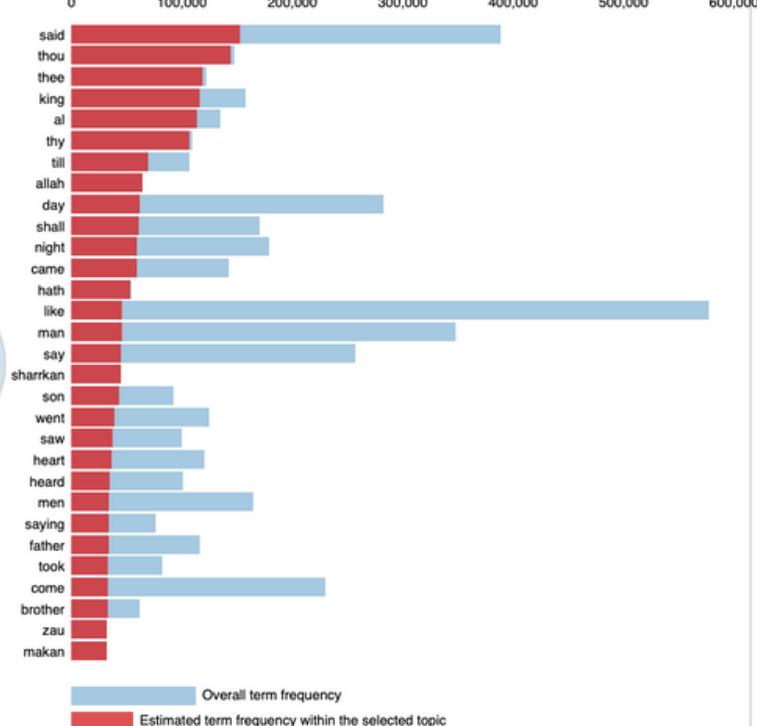
# WHY DID POPULARITY OF SONGS INCREASE OVER YEARS?



# CAN I GET SOME TOPICS FROM SONG LYRICS?



Top-30 Most Relevant Terms for Topic 4 (12.8% of tokens)





# OUR METHOD

---



## CONSULTATION

Presentations are communication tools for reports and more.

## CURATION

Presentations are communication tools for reports and more.

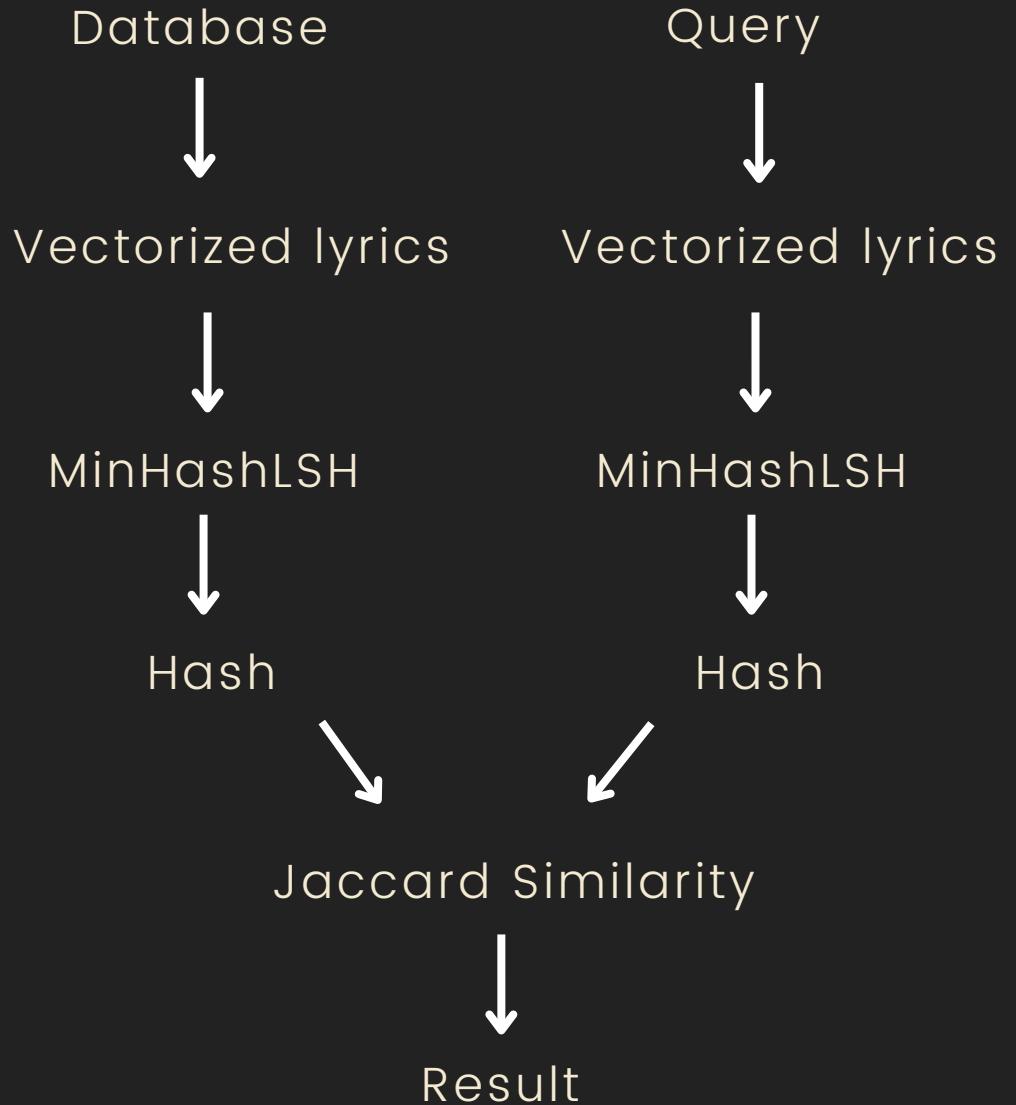
## MANAGEMENT

Presentations are communication tools for reports and more.





# SONG SEARCH



# SONG SEARCH



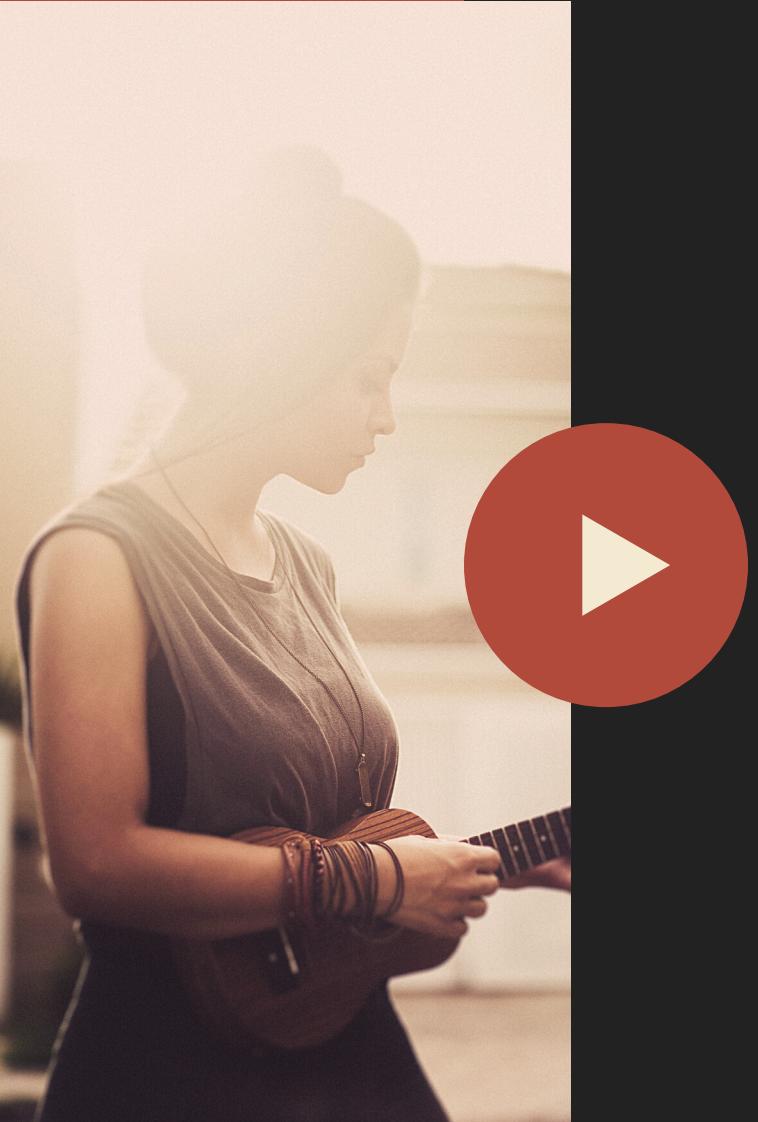
Enter a partial query: no magic words  
Seven years until the hex is broken  
Seven years to undo the curse  
And no magic words can be spoken  
Potions and prayers will make it worse  
The sadness comes with the setting sun  
The war is won when the fight is done  
Will you turn and run or can you stop right now

(Stop right now)  
You can stop right now  
(Stop right now)  
Stop and start right now

Do it once just to answer the question  
Know how it feels to be like them  
But when the poison has left the system  
You'll have new reasons to try again

# SONG SEARCH

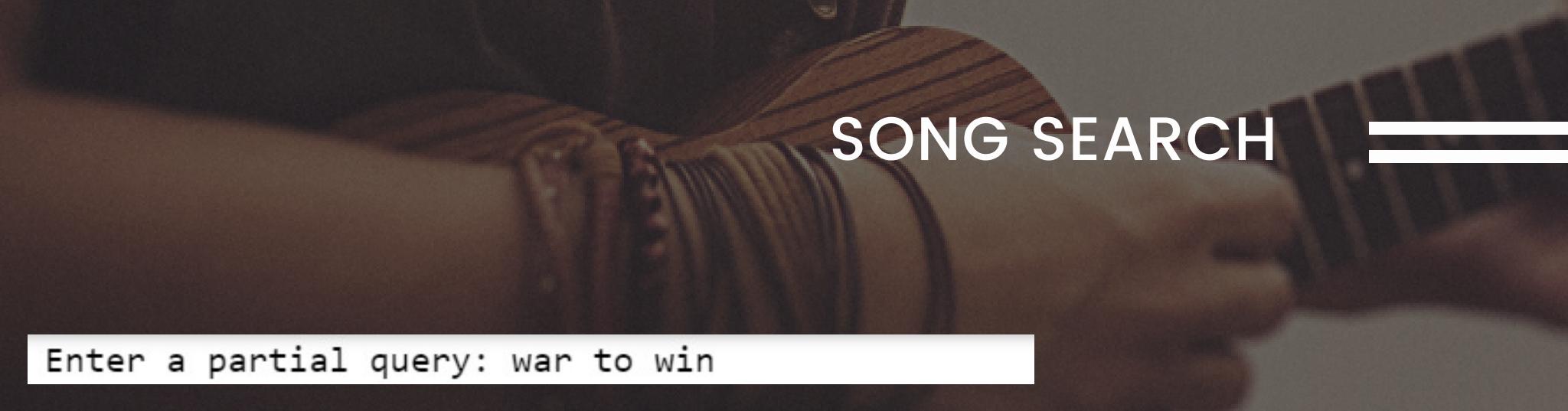
Enter a partial query: stop and start



Seven years until the hex is broken  
Seven years to undo the curse  
And no magic words can be spoken  
Potions and prayers will make it worse  
The sadness comes with the setting sun  
The war is won when the fight is done  
Will you turn and run or can you stop right now

(Stop right now)  
You can stop right now  
(Stop right now)  
Stop and start right now

Do it once just to answer the question  
Know how it feels to be like them  
But when the poison has left the system  
You'll have new reasons to try again  
The sickness comes like the rising sun  
Now your war is done, what have you become



# SONG SEARCH

---

Enter a partial query: war to win

third, and also in the twentieth book of his Histories, says 'The Celtæ sometimes have single combats at their entertainments. For being collected in arms, they go through the exercise, and make feints at, and sometimes they even go so far as to wound one another. And being irritated by this, if the bystanders do not stop them, they will proceed even to kill one another. But in olden times,' he continues, 'there was a custom that a hind quarter of pork was put on the table, and the bravest man took it; and if any one else laid claim to it, then the two rose up to fight, till one of them was slain. And other men in the theatre having received some silver or gold money, and some even for a number of earthen vessels full of wine, having taken pledges that the gifts promised shall really be given, and having distributed them among their nearest connexions, have laid themselves down on doors with their faces upwards, and then allowed some bystander to cut their throats with a sword.'

"And Euphorion the Chalcidian, in his Historical Memorials, writes as follows: 'But among the Romans it is common for five men to be offered to any one who chooses to take it, to allow his head to be cut off with an axe, so that his heirs might receive the reward: and very often many have returned their names as willingly, so that there has been a regular contest between them as to who had the best right to be beaten to death.'" These atrocious examples serve to show us something of the evil and the corruption of which Christianity cleansed the pagan world, although it is to be feared that the battle is not yet won, since it is a notorious and deplorable fact that at the hour of such a crisis as the Great War the respect due to human life became cheapened in men's eyes, with the consequence that murder and deeds of violence once more broke out in every direction showing that savage instincts were dominated indeed, but in many cases not wholly eradicated. It does not require a keen perception to see the direct agency of the devil here, and these atrocities which bred so callous and cruel a spirit are by no means altogether unconnected with the recrudescence of necromancy and black magic which foul arts once more grew green and were almost openly pursued on every side.

# Lyric Gen

## MODEL

GPT-2- Base by  
John Snow labs

## LYRIC O/P LENGTH

1000 characters

## INPUTS

Artist name or  
Album name or  
Song name

## QUERY

String match



# Lyric Gen

Searching by artist name

```
▶ inp = input("Enter the artist name: ")  
↳ Enter the artist name: katy perry  
  
[67] temp1 = temp.filter(lower(col("Artist Name")).contains(inp.lower())).sample(False, 0.1, seed = 0).limit(1)  
  
[68] print_data = temp1.toPandas()  
  
[69] try:  
    print(f"Getting AI generated lyrics for {print_data['spotify track name'][0]} by {print_data['Artist Name'][0]} from the album {print_data['Album'][0]}")  
except:  
    print("No song found!")  
  
Getting AI generated lyrics for Teenage Dream by Katy Perry from the album Teenage Dream
```

Input text:

You think I'm pretty without any makeup on  
You think I'm funny when I tell the punch line wrong  
I know you get me, so I let my walls come down, down  
Before you met me, I was alright  
But things were kinda heavy, you brought me to life  
Now every February, you'll be my Valentine, Valentine  
Let's go all the way tonight  
No regrets, just love

GPT Result:

You think I'm pretty without any makeup on  
You think I'm funny when I tell the punch line wrong  
I know you get me, so I let my walls come down, down  
Before you met me, I was alright  
But things were kinda heavy, you brought me to life  
Now every February, you'll be my Valentine, Valentine  
Let's go all the way tonight  
No regrets, just love

I'm so happy to be here  
I love you, I love you  
I'll be here for you, for you  
You're my best friend, I'll be there for you forever  
I can't wait to see you again  
I've been waiting for you for so long  
I want to be with you forever, I want to kiss you forever

Original Song:

You think I'm pretty without any makeup on  
You think I'm funny when I tell the punch line wrong  
I know you get me, so I let my walls come down, down  
Before you met me, I was alright  
But things were kinda heavy, you brought me to life  
Now every February, you'll be my Valentine, Valentine  
Let's go all the way tonight  
No regrets, just love  
We can dance until we die  
You and I, we'll be young forever  
You make me feel like I'm livin' a teenage dream  
The way you turn me on, I can't sleep  
Let's run away and don't ever look back  
Don't ever look back  
My heart stops when you look at me  
Just one touch, now, baby, I believe

# Lyric Gen

## Searching by album name

```
[85] inp = input("Enter the album name: ")  
Enter the album name: melodrama  
  
[86] temp1 = temp.filter(lower(col("Album")).contains(inp.lower())).sample(False, 0.1, seed = 0).limit(1)  
  
[87] print_data = temp1.toPandas()  
  
[88] try:  
    print(f"Getting AI generated lyrics for {print_data['spotify track name'][0]} by {print_data['Artist Name'][0]} from the album {print_data['Album'][0]}")  
except:  
    print("No song found!")  
  
Getting AI generated lyrics for Hard Feelings/Loveless by Lorde from the album Melodrama
```

Input text:

```
Go back and tell it
Please, could you be tender?
And I will sit close to you
Let's give it a minute before we admit that we're through
Guess this is the winter
Our bodies are young and blue
I'm at Jungle city, it's late and this song is for you
'Cause I remember the rush, when forever was us
GPT Result:
Go back and tell it
Please, could you be tender?
And I will sit close to you
Let's give it a minute before we admit that we're through
Guess this is the winter
Our bodies are young and blue
I'm at Jungle city, it's late and this song is for you
'Cause I remember the rush, when forever was us

I remember the cold, when we were in the middle of the night
I was in the jungle, I was in a jungle, we were all in the same place
I saw the sun, I saw the moon, I heard the wind
I heard the rain, I hear the wind, I see the sun
I see the moon
I hear the rain
I know the sun is coming, I know the moon is coming
I knew the sun was coming, we all knew the moon was coming
We all knew that we were going to die
We knew that the sun would come, we knew that it would come
We knew that we would die
I can't believe we're still alive
I've been here for so long
I don't know what to do
I just want to go back to my home
I want to be here
```

Original Song:

```
Go back and tell it
Please, could you be tender?
And I will sit close to you
Let's give it a minute before we admit that we're through
Guess this is the winter
Our bodies are young and blue
I'm at Jungle city, it's late and this song is for you
'Cause I remember the rush, when forever was us
Before all of the winds of regret and mistrust
Now we sit in your car and our love is a ghost
Well, I guess I should go, yeah, I guess I should go
Hard feelings
These are what they call hard feelings of love
When sweet words and fevers
All leave us right here in the cold, oh, oh
Alone with the hard feelings of love
God, I wish I believed ya
When you told me this was my home, oh, oh
You might also like
I light all the candles
Cut flowers for all my rooms
I care for myself the way I used to care about you
These days, we kiss and we keep busy
The waves come after midnight
I call from underwater
```

# Lyric Gen

Searching by song name

```
[105] inp = input("Enter the song name: ")  
Enter the song name: pink + white  
  
[106] templ = temp.filter(lower(col("spotify track name")).contains(inp.lower())).sample(False, 0.9, seed=0).limit(1)  
  
[107] print_data = templ.toPandas()  
  
[108] try:  
    print(f"Getting AI generated lyrics for {print_data['spotify track name'][0]} by {print_data['Artist Name'][0]} from the album {print_data['Album'][0]}")  
except:  
    print("No song found!")  
  
Getting AI generated lyrics for Pink + White by Frank Ocean from the album Blonde
```

Input text:

```
Yeah, yeah, um (Woo)  
Yeah, yeah, yeah  
That's the way every day goes  
Every time we've no control  
If the sky is pink and white  
If the ground is black and yellow  
It's the same way you showed me  
Nod my head, don't close my eyes  
GPT Result:  
Yeah, yeah, um (woo)  
Yeah, yeah, yeah  
That's the way every day goes  
Every time we've no control  
If the sky is pink and white  
If the ground is black and yellow  
It's the same way you showed me  
Nod my head, don't close my eyes  
  
I'm not going to let you go  
I'll be there for you, I'll be with you  
I will be there with you, and I'll make you happy  
I won't let you leave  
I don't want to leave  
You're my friend, I don't care if you're a girl or a boy  
I want you to be my friend  
I love you, you're my best friend  
You'll be my best girl  
I know you're going to be a girl, I know you'll be a boy, I'm going to make you a girl  
You know I'm not a girl anymore  
I've got to be the best girl I can be  
I can't be the girl anymore, I can't let go  
You can't leave  
It'll be okay, it'll be alright  
I just want you, it's okay, I just want to be with you  
I wanna be with ya, I wanna be in your arms
```

Original Song:

```
Yeah, yeah, um (woo)  
Yeah, yeah, yeah  
That's the way every day goes  
Every time we've no control  
If the sky is pink and white  
If the ground is black and yellow  
It's the same way you showed me  
Nod my head, don't close my eyes  
Halfway on a slow move  
It's the same way you showed me  
If you could fly, then you'd feel south  
Up north's getting cold soon  
The way it is, we're on land  
Still, I'm someone to hold true  
Keep you cool when it's still alive  
Won't let you down when it's all ruin  
Just the same way you showed me, showed me  
You showed me love  
Glory from above  
Regard, my dear  
It's all downhill from here  
You might also like  
In the wake of a hurricane  
Dark skin of a summer shade  
Nosedive into flood lines  
Tall tower of milk crates
```

# CONCLUSION

Concepts used:

- Filtering/Queries
  - pre-processing
  - EDA
- MinHashLSH
  - Song Search
- Spark for ML
  - Lyric Gen

Technologies Used:

- Spark
- Dask



# CONCLUSION

Why RDBMS?

- Data is structured
- Entries are related

Why is this Big Data?

- 17GB crashes the computer
- More data can be added
- All concepts are scalable



## FUTURE IMPROVEMENTS

- Scrape more data years 1890-2000
- Merge the Vagalume lyric data with the current dataset
- Train models on each genre

