

# A Content-Based Song Recommender for Spotify

*J. Garrecht Metzger, PhD*



# J. Garrecht Metzger

PhD Geochemist



 Washington University in St. Louis

# J. Garrecht Metzger

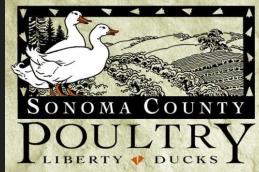
PhD Geochemist

Farmer

Quack



Washington University in St. Louis



# J. Garrecht Metzger

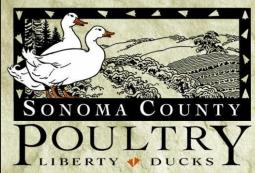
PhD Geochemist

Farmer

Musician



Washington University in St. Louis



METIS

# J. Garrecht Metzger

PhD Geochemist

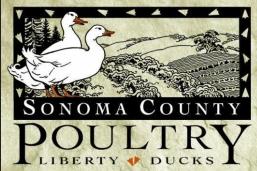
Farmer

Musician

Data Scientist



Washington University in St. Louis



# What is Spotify?

# What is Spotify?



# What is Spotify?

## FAST FACTS

Number of subscribers (as at December 31, 2020)

**155** Million

Revenue paid to rightsholders (since launch)

**€21+** Billion

Number of tracks

**70** Million +

Number of playlists

**4** Billion +



METIS

# What is Spotify?

That's a lot of material to recommend from!



# What is Spotify?

That's a lot of material to recommend from!



VASO  
VAGAL

## FAST FACTS

Number of subscribers (as at December 31, 2020)

**155** Million

Revenue paid to rightsholders (since launch)

**€21+** Billion

Number of tracks

**70** Million +

Number of playlists

**4** Billion +



METIS

Top Result

Vaso Vagal

ARTIST

- Go to Artist Radio
- Unfollow
- Report
- Share >

SEE ALL Artists

Songs

Image	Name	Artist
	Necksweat	Vaso Vagal
	Rudderless	Vaso Vagal
	Ausländer	Vaso Vagal
	Soar	Vaso Vagal

PLAYLISTS

- Necksweat Content...
- The Sound of Ever...
- Vaso Vagal Immolate
- Vaso Vagal ⓘ
- PURDIE SHUFFLE
- Karate
- The Same Stars - ...
- Local Drive
- 20190529 205205...
- 2021Workout

SEE ALL

SEE ALL

Albums

Image	Name	Artist
	Immolate	Vaso Vagal
	VasoVagal	Vaso Vagal
	Vaso Vagal	Vaso Vagal
	Vaso Vagal Immolate	Vaso Vagal

SEE ALL

Playlists

0:15 5:00

New Playlist

Necksweat

Vaso Vagal

METIS

Top Result

Vaso Vagal ARTIST

Go to Artist Radio

Unfollow Report Share >

Songs SEE ALL Artists SEE ALL

	Necksweat	Rudderless	Vaso Vagal
	Vaso Vagal		
	Soar	Ausländer	
	Vaso Vagal		

SEE ALL Playlists SEE ALL

	Immolate	VasoVagal	Vaso Vagal
	Vaso Vagal		
	Immolate		

0:15 5:00

Necksweat Vaso Vagal

Immolate Vaso Vagal

VasoVagal Vaso Vagal

Vaso Vagal Vaso Vagal

Vaso Vagal Immolate Vaso Vagal

Play Pause Stop

Next Previous

Shuffle

Loop

Volume

Redacted



METIS

redacted

Home

Browse

Radio

YOUR LIBRARY

Made For You

Recently Played

Liked Songs

Albums

Artists

Local Files

Podcasts

PLAYLISTS

Necksweat Content...

The Sound of Ever...

Vaso Vagal Immolate

Vaso Vagal ⓘ

PURDIE SHUFFLE

Karate

The Same Stars – ...

Local Drive

20190529 205205...

2021Workout

---

New Playlist

Necksweat ❤️

Vaso Vagal

0:57 5:00

METIS

redacted

Home

Browse

Radio

YOUR LIBRARY

Made For You

Recently Played

Liked Songs

Albums

Artists

Local Files

Podcasts

PLAYLISTS

Necksweat Content...

The Sound of Ever...

Vaso Vagal Immolate

Vaso Vagal ⓘ

PURDIE SHUFFLE

Karate

The Same Stars – ...

Local Drive

20190529 205205...

2021Workout

---

New Playlist

Necksweat ❤️

Vaso Vagal

0:57

5:00

Wheel of Doom



Spotify

Necksweat

Vaso Vagal

0:57

5:00

METIS

# Motivation

# Motivation

- Spotify uses an ‘ensemble’ recommender

# Motivation

- Spotify uses an ‘ensemble’ recommender
  - Includes audio + user content (‘likes’)

# Motivation

- Spotify uses an ‘ensemble’ recommender
  - Includes audio + user content (‘likes’)
  - User control is limited and indirect

# Motivation

- Spotify uses an ‘ensemble’ recommender
  - Includes audio + user content ('likes')
  - User control is limited and indirect
- Small artists are excluded from the recommendation system



# Motivation

- Spotify uses an ‘ensemble’ recommender
  - Includes audio + user content ('likes')
  - User control is limited and indirect
- Small artists are excluded from the recommendation system
- A more general recommender



# Motivation

- Spotify uses an ‘ensemble’ recommender
  - Includes audio + user content (‘likes’)
  - User control is limited and indirect
- Small artists are excluded from the recommendation system
- A more general recommender
  - Based on audio content



# Motivation

- Spotify uses an ‘ensemble’ recommender
  - Includes audio + user content (‘likes’)
  - User control is limited and indirect
- Small artists are excluded from the recommendation system
- A more general recommender
  - Based on audio content
  - Customizable feature selection



# Motivation

- Spotify uses an ‘ensemble’ recommender
  - Includes audio + user content (‘likes’)
  - User control is limited and indirect
- Small artists are excluded from the recommendation system
- A more general recommender
  - Based on audio content
  - Customizable feature selection



*A song recommender for everyone!*

# Problem: *Data Selection*

## Problem: *Data Selection*

- >70,000,000,000 tracks!

# Problem: *Data Selection*

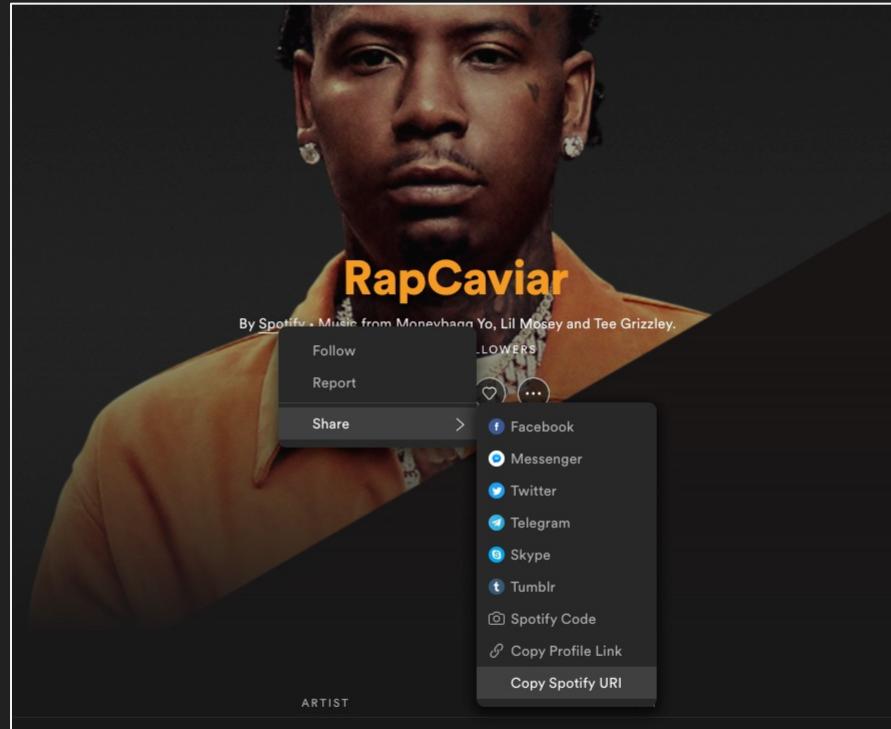
- >70,000,000,000 tracks!
  - Computationally prohibitive

# Problem: *Data Selection*

- >70,000,000,000 tracks!
  - Computationally prohibitive
  - No random sampling possible

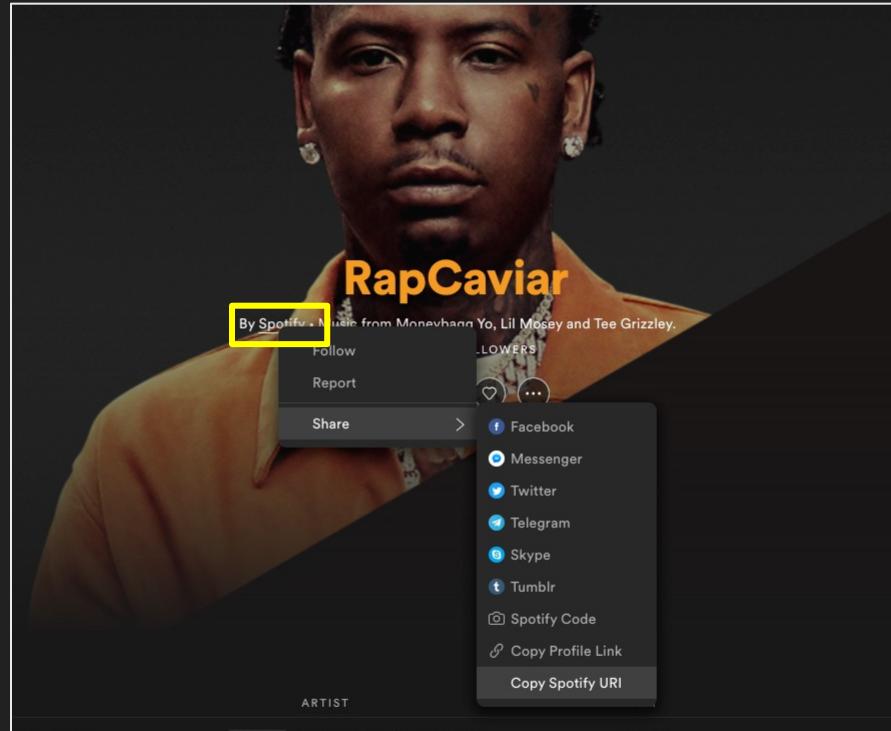
# Problem: *Data Selection*

- >70,000,000,000 tracks!
  - Computationally prohibitive
  - No random sampling possible
- Big 'get' was finding the official user account



# Problem: *Data Selection*

- >70,000,000,000 tracks!
  - Computationally prohibitive
  - No random sampling possible
- Big 'get' was finding the official user account
  - 'hiding in plain sight'.



# API Results

- Official 'spotify' user account

# API Results

- Official ‘spotify’ user account
  - 1,398 Playlists

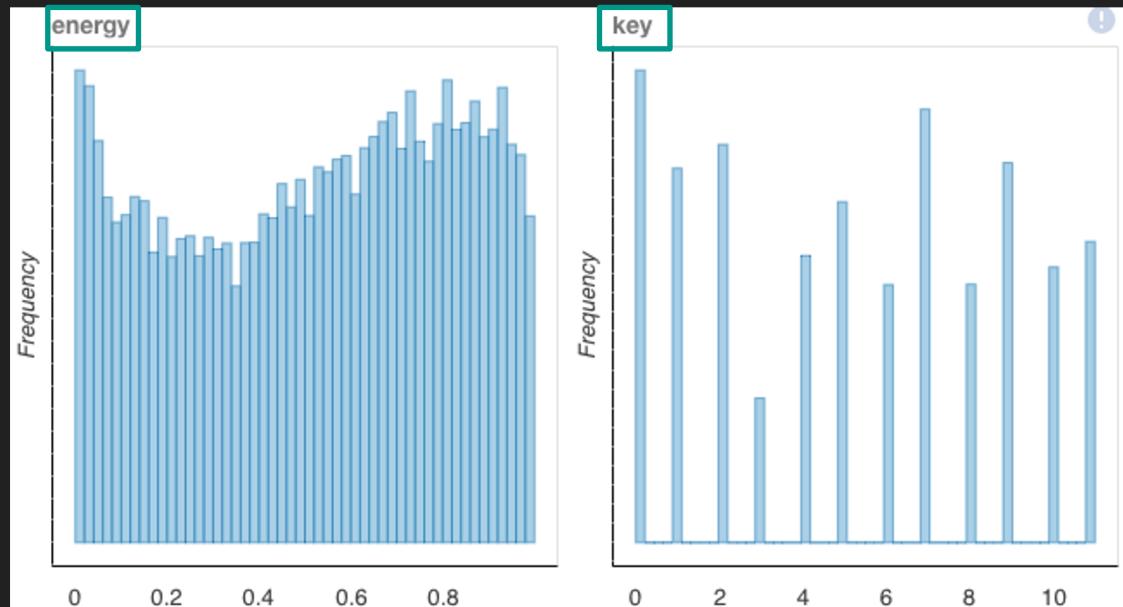
# API Results

- Official ‘spotify’ user account
  - 1,398 Playlists
  - 81,648 Tracks

# API Results

- Official 'spotify' user account
  - 1,398 Playlists
  - 81,648 Tracks

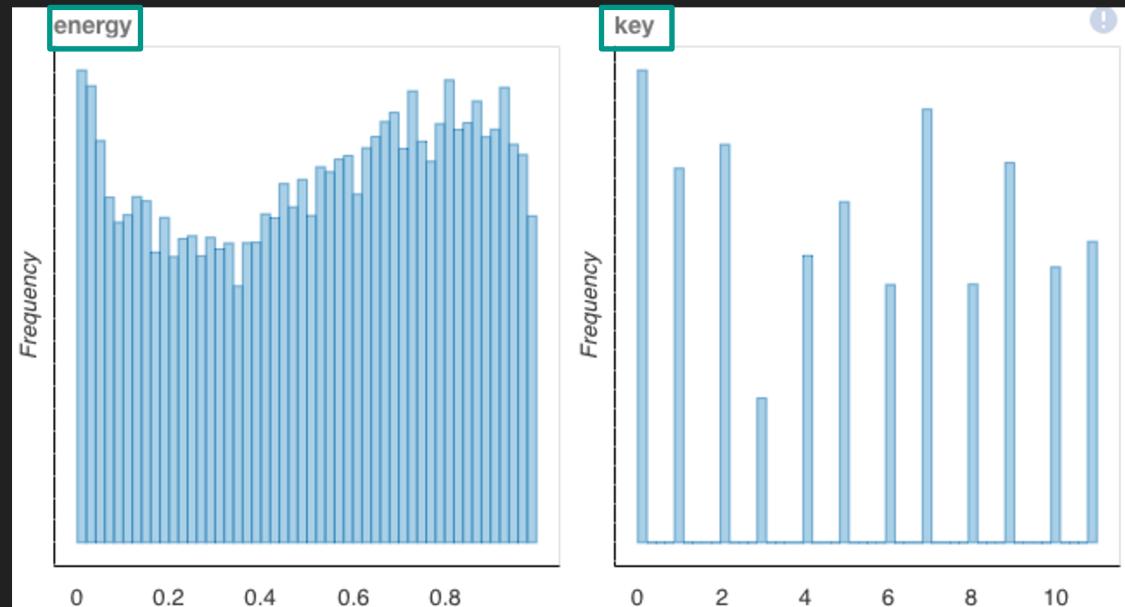
*Audio content 'features'*



# API Results

- Official ‘spotify’ user account
  - 1,398 Playlists
  - 81,648 Tracks
- Add *any* songs available on Spotify

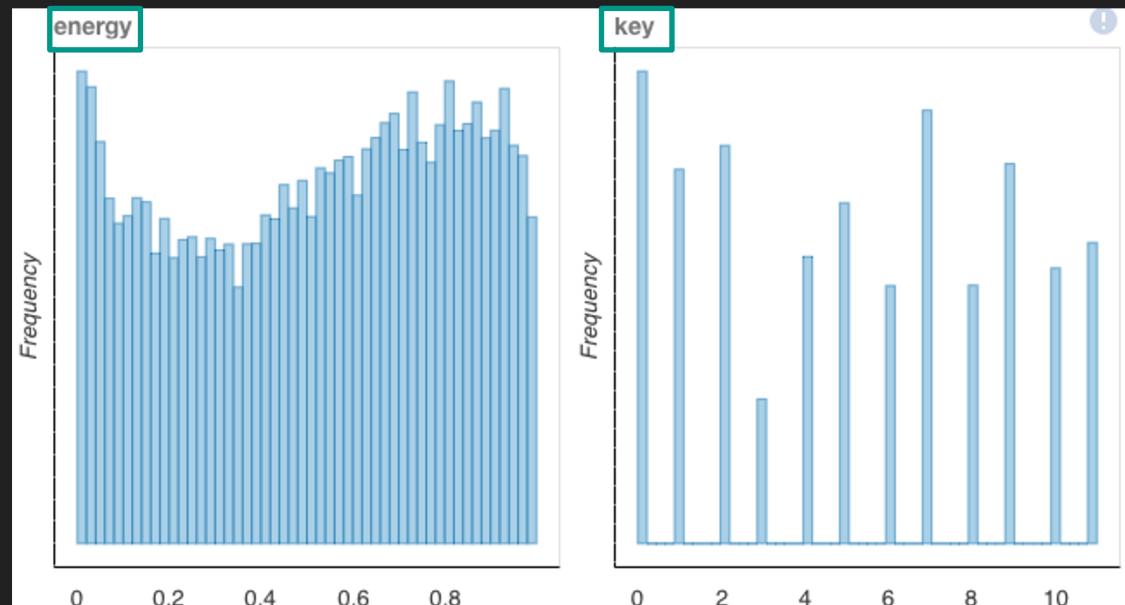
*Audio content ‘features’*



# API Results

- Official ‘spotify’ user account
  - 1,398 Playlists
  - 81,648 Tracks
- Add *any* songs available on Spotify
  - Allows *direct connection* between small artists and bigger ones

*Audio content ‘features’*



# Track Data

## Metadata

- Track Name
- Album Name
- Artist
- Release Date
- Artwork

# Track Data

## Metadata

- Track Name
- Album Name
- Artist
- Release Date
- Artwork



- Artist
- Album Name

# Audio Content Features

Subjective (0 – 1)

# Audio Content Features

## Subjective (0 – 1)

Acousticness

Danceability

Energy

Instrumentalness

Liveness

Speechiness

Valence

# Audio Content Features

## Subjective (0 – 1)

Acousticness

Danceability

Energy

Instrumentalness

Liveness

Speechiness

Valence

## Objective

# Audio Content Features

## Subjective (0 – 1)

Acousticness

Danceability

Energy

Instrumentalness

Liveness

Speechiness

Valence

## Objective

Duration (range, milliseconds)

Key (Western, 12)

Loudness (~ -60 to 0, dB)

Mode (0=minor, 1=major)

Popularity (0-100)

Tempo (0 - ~300)

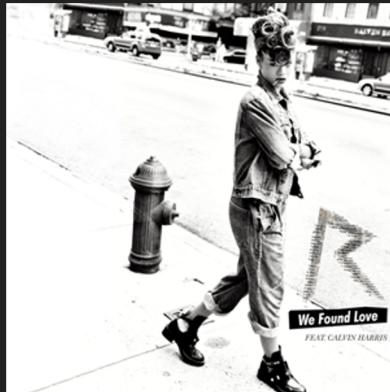
Time Signature

# Examples

# Examples

Rihanna: *We Found Love*

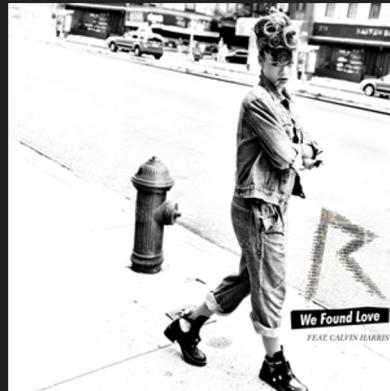
- Danceability: 0.74
- Energy: 0.77
- Tempo: 128



# Examples

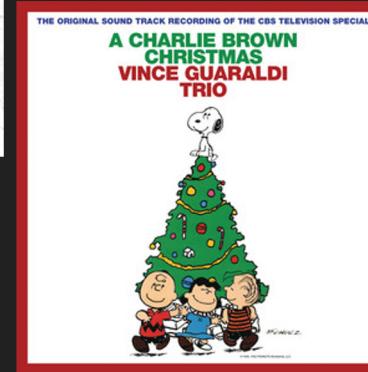
Rihanna: *We Found Love*

- Danceability: 0.74
- Energy: 0.77
- Tempo: 128



Vince Guaraldi Trio: *Für Elise*

- Danceability: 0.43
- Energy: 0.08
- Tempo: 130



# Examples

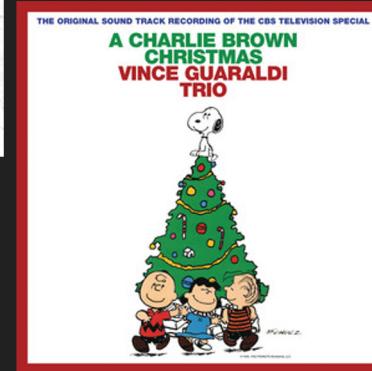
Rihanna: *We Found Love*

- Danceability: 0.74
- Energy: 0.77
- Tempo: 128



Vince Guaraldi Trio: *Für Elise*

- Danceability: 0.43
- Energy: 0.08
- Tempo: 130



REM: *Shiny Happy People*

- Danceability: 0.45
- Energy: 0.91
- Tempo: 125



# Examples

Rihanna: *We Found Love*

- Danceability: 0.74
- Energy: 0.77
- Tempo: 128



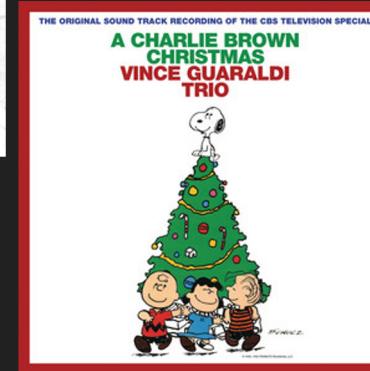
Vince Guaraldi Trio: *Für Elise*

- Danceability: 0.43
- Energy: 0.08
- Tempo: 130

REM: *Shiny Happy People*

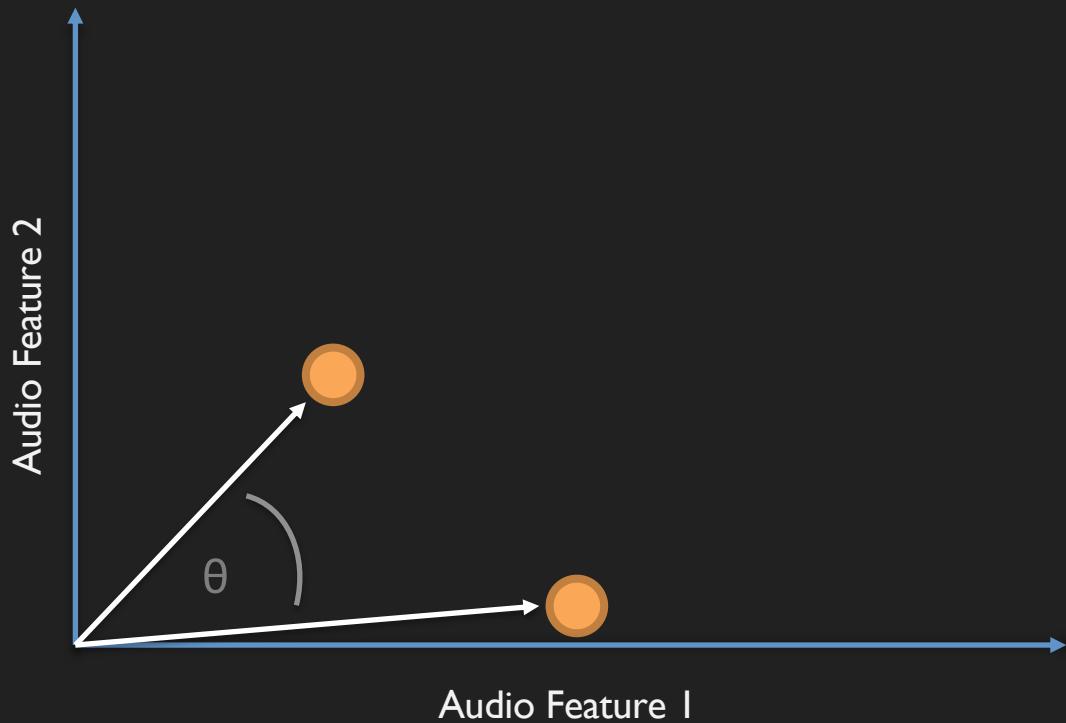
- Danceability: 0.45
- Energy: 0.91
- Tempo: 125

*How do we recommend?*



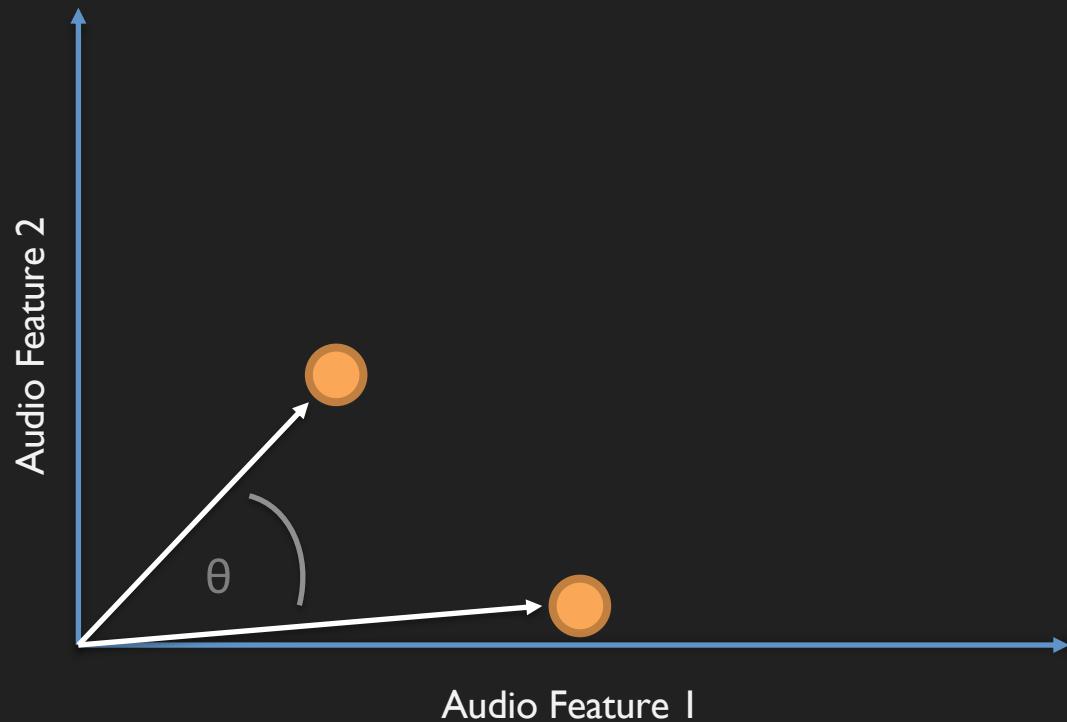
# Method: *Cosine Distance/Similarity*

# Method: *Cosine Distance/Similarity*



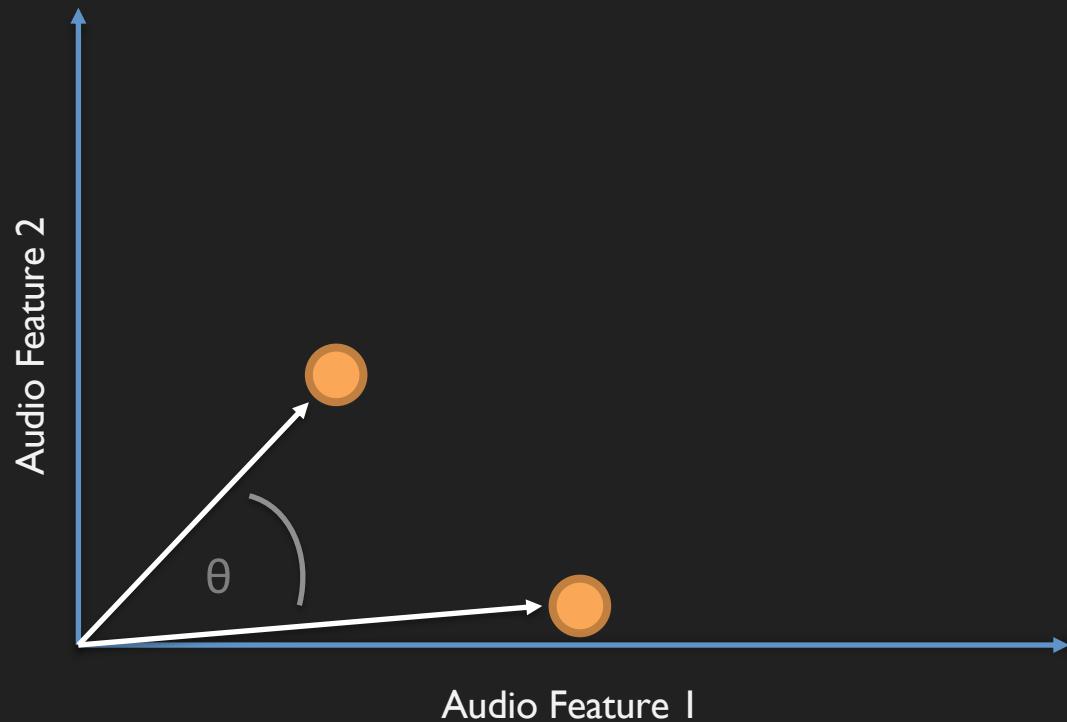
# Method: *Cosine Distance/Similarity*

- Distance Matrix



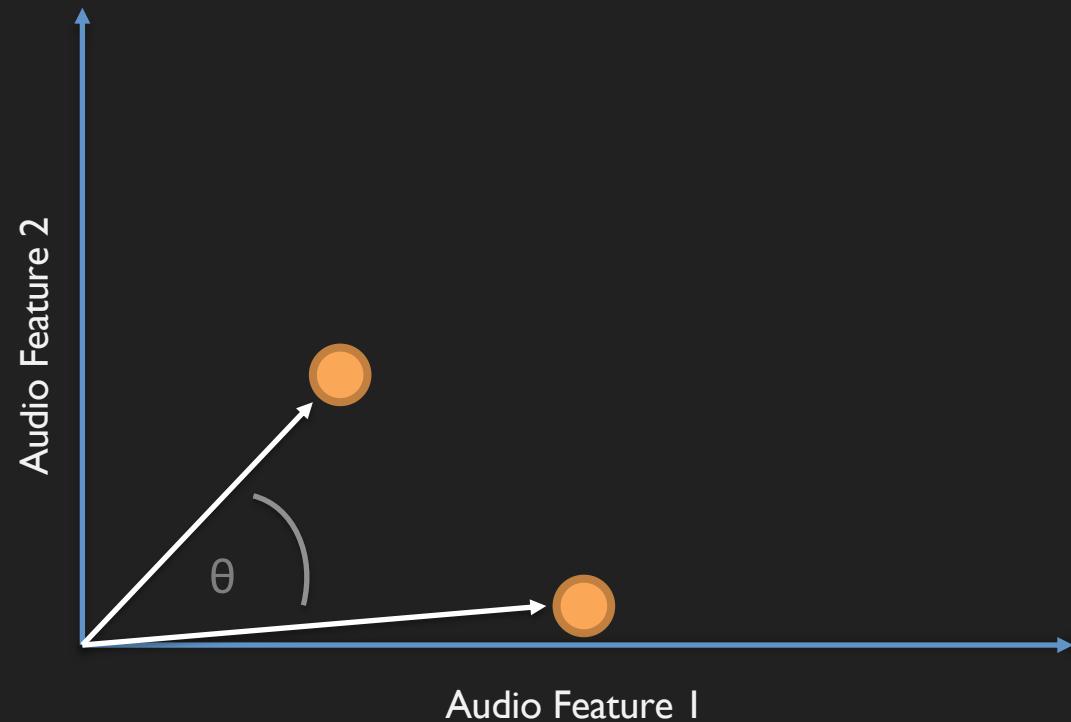
# Method: *Cosine Distance/Similarity*

- Distance Matrix
  - Filter



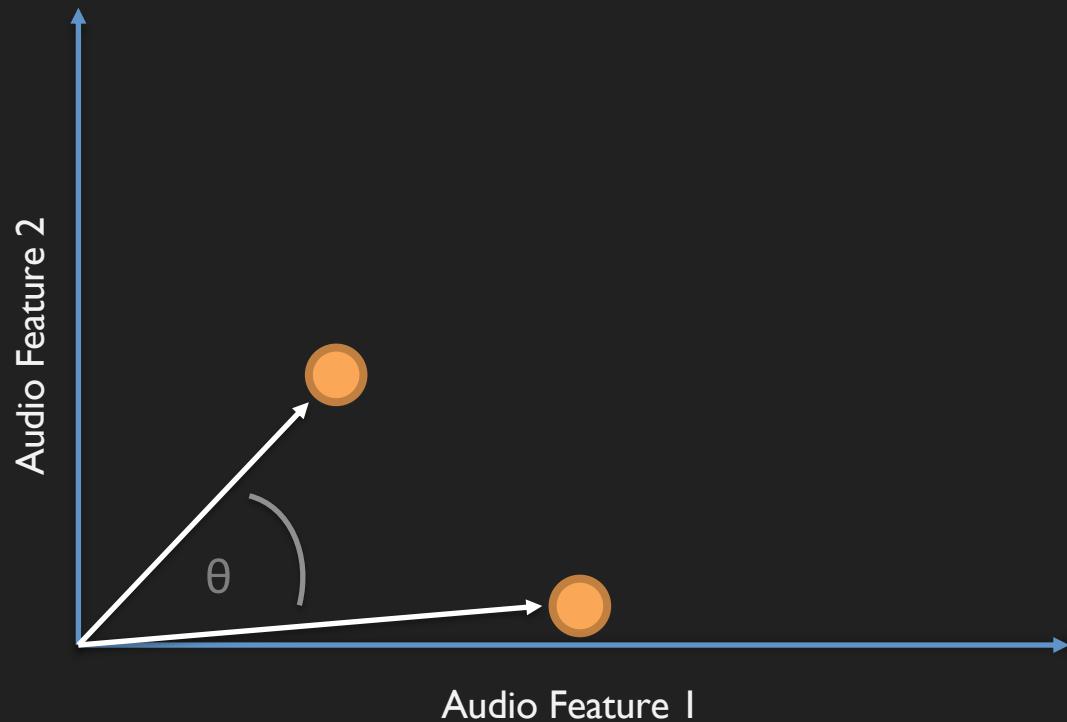
# Method: *Cosine Distance/Similarity*

- Distance Matrix
  - Filter
    - Remove largest values



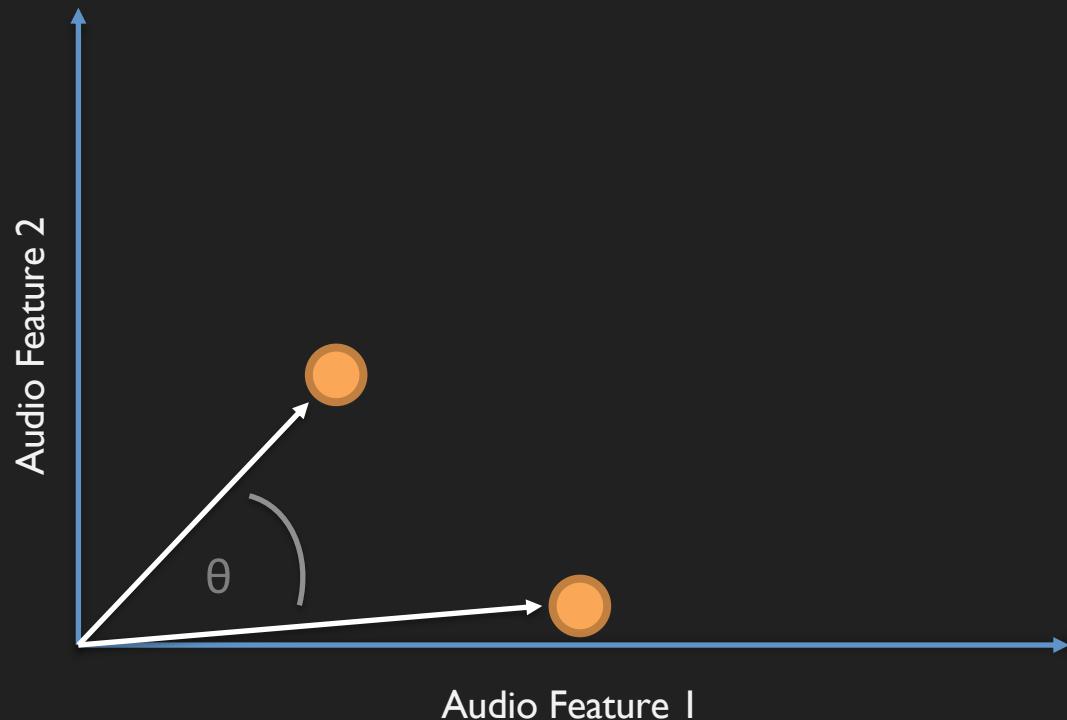
# Method: *Cosine Distance/Similarity*

- Distance Matrix
  - Filter
    - Remove largest values
    - Rapid computation



# Method: *Cosine Distance/Similarity*

- Distance Matrix
  - Filter
    - Remove largest values
    - Rapid computation



*Let's hear it!*

# Top Recommended Tracks

VASO  
VAGAL

IMMOLATE

Reference Track

# Content-Based Recommender: *Closing Thoughts*

# Content-Based Recommender: *Closing Thoughts*

- *Then:*

# Content-Based Recommender: *Closing Thoughts*

- *Then:*
  1. Native recommender excludes content

# Content-Based Recommender: *Closing Thoughts*

- *Then:*
  1. Native recommender excludes content
  2. User control is limited

# Content-Based Recommender: *Closing Thoughts*

- *Then:*
  1. Native recommender excludes content
  2. User control is limited
- *Now:*

# Content-Based Recommender: *Closing Thoughts*

- *Then:*
  1. Native recommender excludes content
  2. User control is limited
- *Now:*
  1. Found representative data

# Content-Based Recommender: *Closing Thoughts*

- *Then:*
  1. Native recommender excludes content
  2. User control is limited
- *Now:*
  1. Found representative data
  2. Developed rapid computational framework

# Content-Based Recommender: *Closing Thoughts*

- *Then:*
  1. Native recommender excludes content
  2. User control is limited
- *Now:*
  1. Found representative data
  2. Developed rapid computational framework
  3. Positioned for custom recommender web app

# Content-Based Recommender: *Closing Thoughts*

- *Then:*
  1. Native recommender excludes content
  2. User control is limited
- *Now:*
  1. Found representative data
  2. Developed rapid computational framework
  3. Positioned for custom recommender web app
  4. Direct way to relate content from smaller artists with bigger ones



# Tools Used



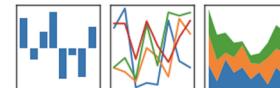
*SpotiPy*



**NumPy**

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



METIS

# References

Art: <https://www.pixilart.com/art/titanic-sinking-f947f8b9b5c953a>

Spotify: <https://github.com/plamere/spotipy/blob/2.17.1/docs/index.rst>

Official Spotify Documentation: <https://developer.spotify.com/documentation/web-api/reference/#endpoint-get-audio-features>

Spotify Facts: <https://newsroom.spotify.com/company-info/>

Cosine and kmeans figures: Metis lectures