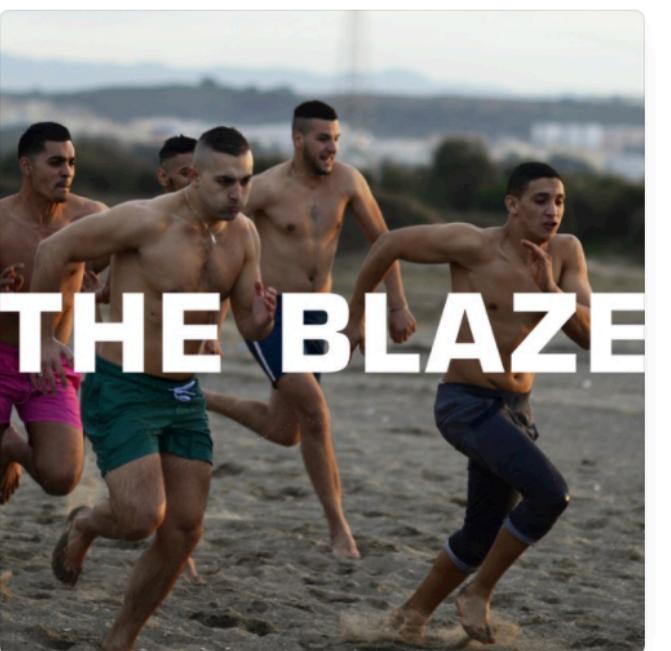
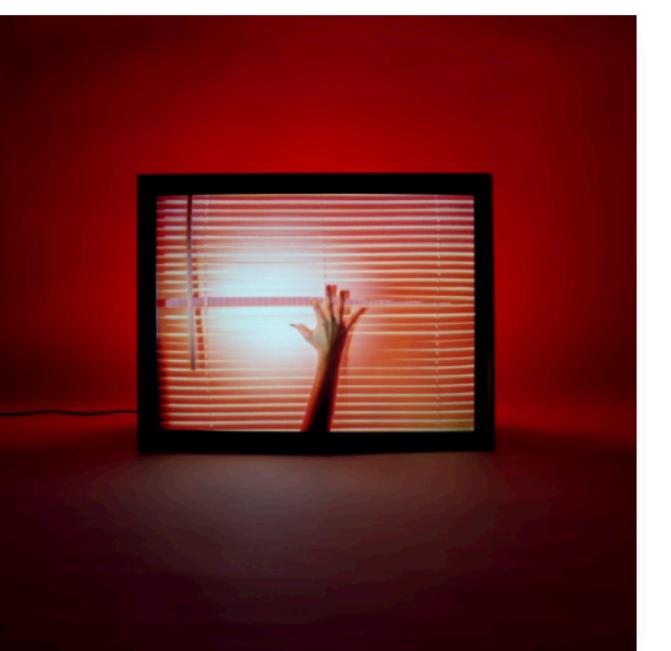




I Forget Where We Were  
Ben Howard



Territory - EP  
The Blaze



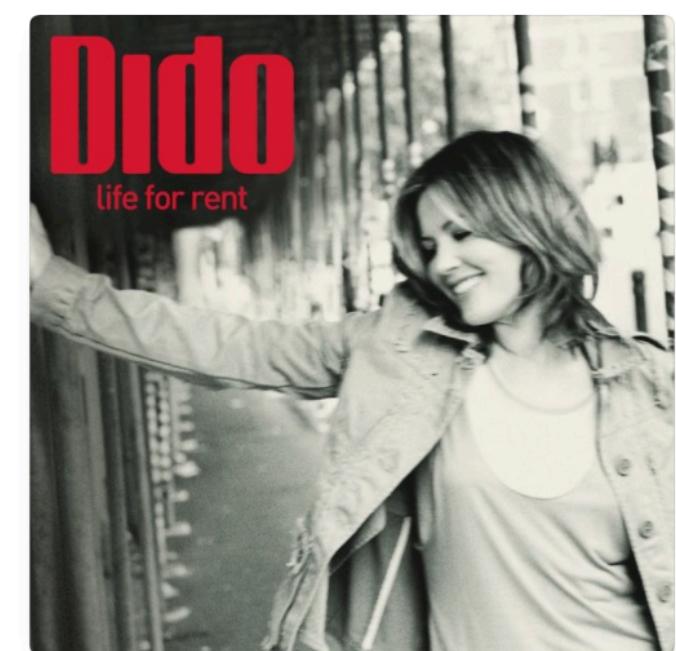
Screen Violence  
CHVRCHES



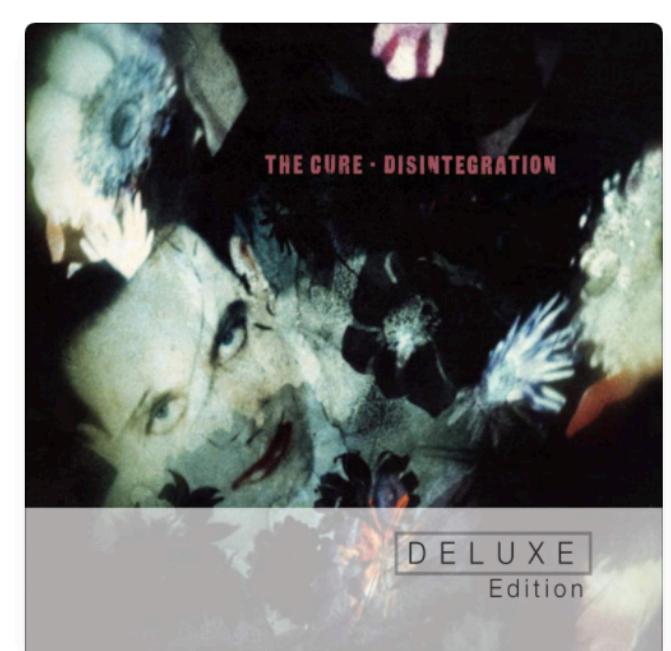
E  
Bootstraps  
Bootstraps



Christine and the Queens / Chaleur Humaine  
Christine and the Queens



Dido  
life for rent  
Life for Rent



THE CURE · DISINTEGRATION  
DELUXE Edition  
Disintegration (Deluxe Edition)  
The Cure



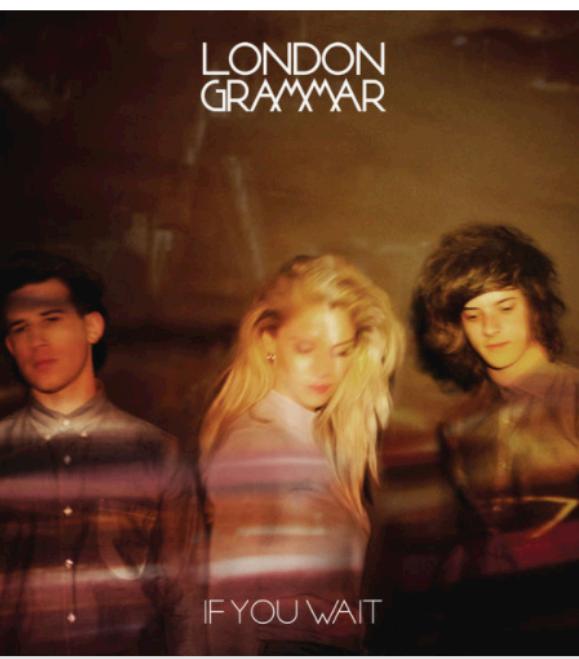
Recovery  
Eminem



DAMN.  
Kendrick Lamar



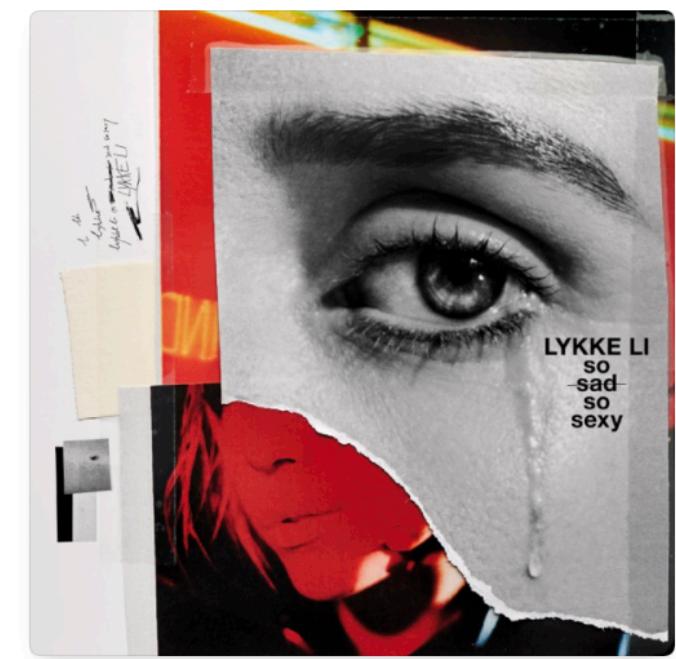
E  
Minutes to Midnight  
LINKIN PARK



London Grammar  
If You Wait



Cinema  
Ludovico Einaudi



LYkke Li  
so sad so sexy  
so sad so sexy



KANYE WEST  
808s & Heartbreak  
Kanye West



18  
Moby



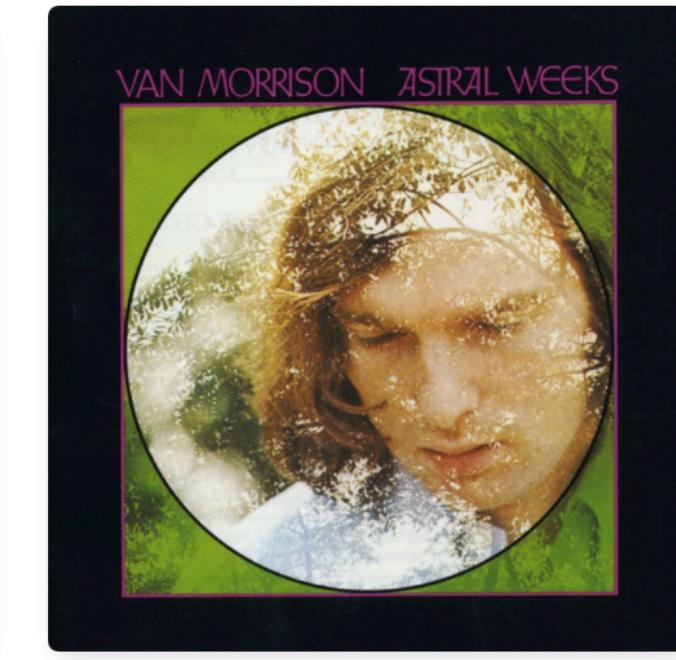
7  
Paul Kalkbrenner



Alla mia età  
Tiziano Ferro



U2  
Achtung Baby



VAN MORRISON  
Astral Weeks  
Van Morrison



The 1975  
The 1975



ANTONELLO VENDITTI  
sotto la pioggia  
Sotto la Pioggia  
Antonello Venditti

# Recommended Frequencies

A Music Recommendation Engine  
by Lucas Chizzali



Listen to music

---



Analyse music

# Recommended Frequencies

## What?

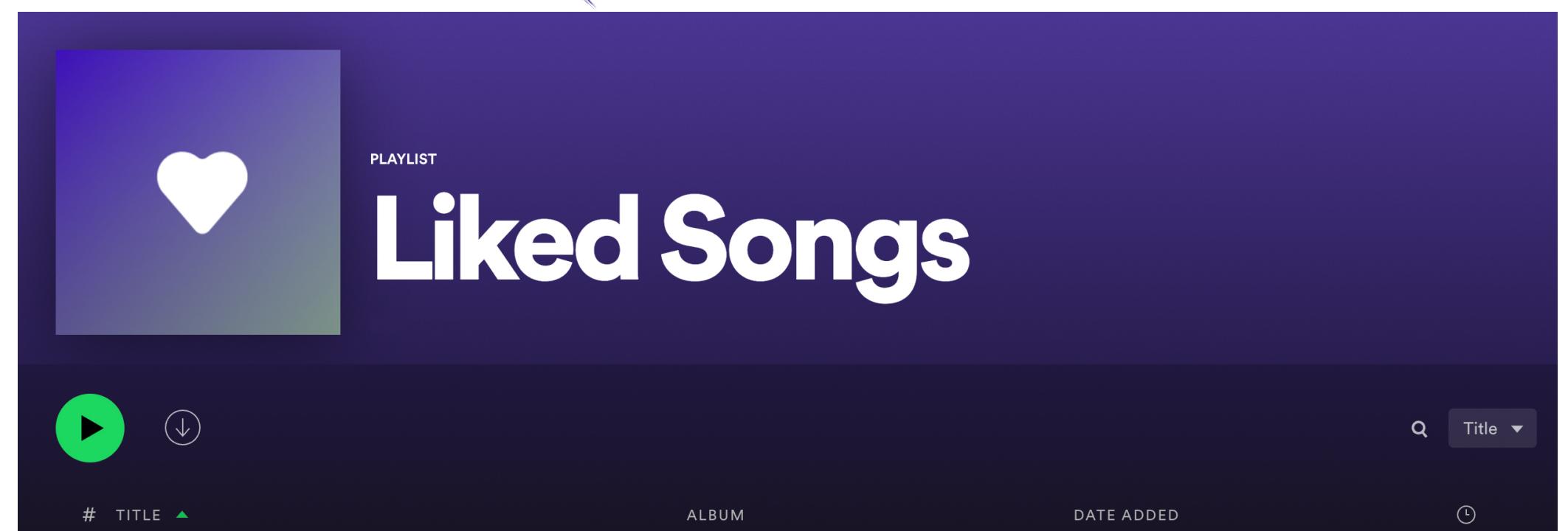
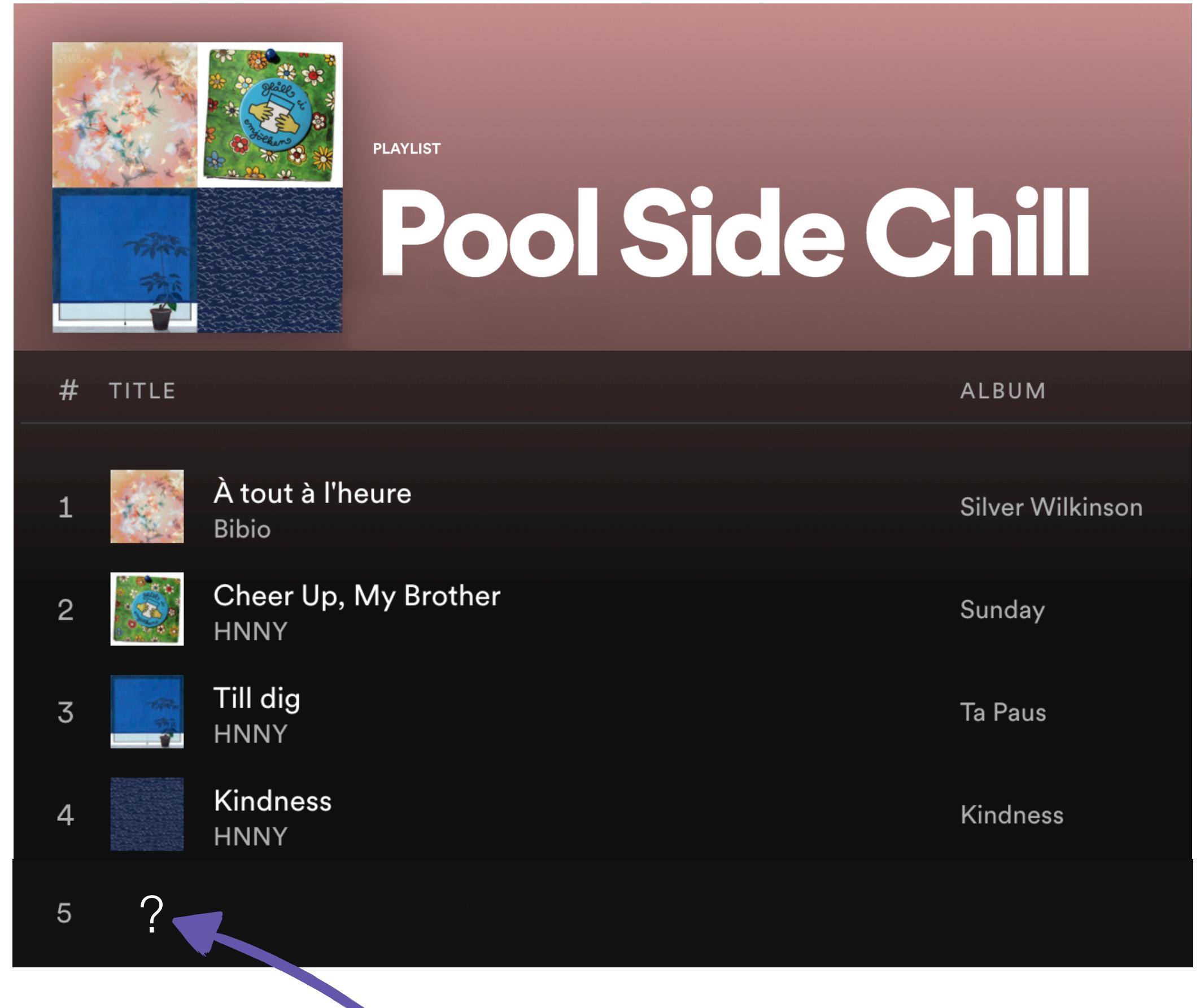
Help manage playlists

→ Identify songs in a person's music library that may fit well into a selected playlist

Multifaceted and difficult task since music is subjective

Specifically, a playlist may represent

- Mood (Summertime 🍉)
- Memory (High School 🎓)
- Genre (Synthwave 🌄)
- Era (80s 🎵)
- etc.



# Recommended Frequencies

## Why?

Playlists are an integral means to enjoying music

- Anecdotal evidence
- 4 bn playlists vs 350 m users on Spotify\*

The screenshot shows a dark-themed interface for a music service. At the top, there's a play button and a title 'Pool Side Chill'. Below it, a section titled 'Recommended' is labeled 'Based on what's in this playlist'. It lists three songs: 'Dream Machine - Kaskade Mix' by Mark Farina, Kaskade; 'Don't Call It Love - 12" Version' by Zero 7; and 'Hello' by Axel Boman. To the right of each song are three categories: 'To Do', 'On My Own', and 'Family Vacation', each with an 'ADD' button.

Music recommendation is nothing new

Nevertheless

- DIY is fun
- Re-discovering your own music



The screenshot shows a dark-themed mobile application. At the top, there's a purple header with the text '#DUBOLT' in white. Below it, a main area has the text 'Discover new music with the help of old favorites.' in purple. There's a search bar labeled 'Search by Track' and a small note 'Artists | Tracks | Reset' at the bottom right.

The screenshot shows the homepage of Spotibot.com. At the top, there's a green header with the text 'Spotibot.com'. Below it, a main area has the text 'Discover Something New.' in green. It explains that 'Spotibot uses the listening habits of millions of people to help you find your new favourites.' and encourages users to 'Simply enter the name of a band you already like:'. There's a search input field labeled 'Band/Artist:' with a green border, and two buttons below it: 'Generate Playlist!' and 'More Options...'. To the right, there's a graphic of a stack of dark cards with a green musical note icon on the top card.

The screenshot shows the 'Similar Songs Finder' page. At the top, there's a green header with the text 'Similar Songs Finder' and a subtext 'Playlist generator based on what you love'. Below it, a main area has a search input field labeled 'Type a song name' and a dropdown menu labeled 'Song'. A red 'Generate' button is to the right. At the bottom, there's a link 'Options +'.

# Let's research the



# State of Sound the Art

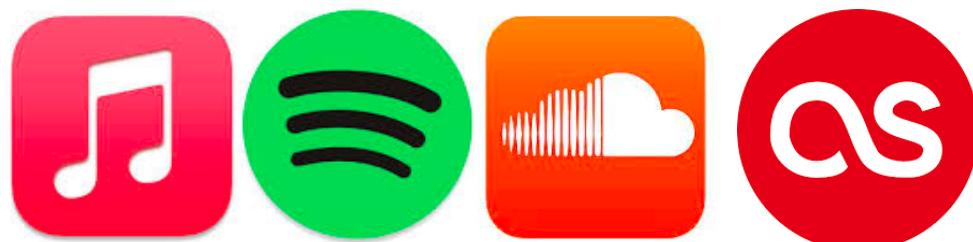
222,704 monthly listeners



# Music Recommendation

## How?

- Popular research domain
- Competitive edge for streaming services
- Usage **vs** content data

	Usage data	Content data
Approaches	Recommender Systems	Music Information Retrieval
Example of technique	Collaborative filtering [1]	Deep Learning [2]
Example of datasource		Million Song Dataset

[1] Song, Yading, Simon Dixon, and Marcus Pearce. "A survey of music recommendation systems and future perspectives." In 9th international symposium on computer music modeling and retrieval, 2012.

[2] Van den Oord, Aaron, Sander Dieleman, and Benjamin Schrauwen. "Deep content-based music recommendation." Advances in neural information processing systems 26 (2013)

# Music Recommendation Implications

	Usage data	Content data
Approaches	Recommender Systems	Music Information Retrieval
Example of technique	Collaborative filtering [1]	Deep Learning [2]
Example of datasource	 A row of four circular icons representing different music platforms: Apple Music (red square with white note), Spotify (green circle with three horizontal lines), SoundCloud (orange square with white cloud and waveform), and Last.fm (red circle with white lowercase letters 'as').	 A dark blue horizontal brushstroke underline.

- Not easy to come by
- Proprietary / limited access to the public
- Large amounts of data

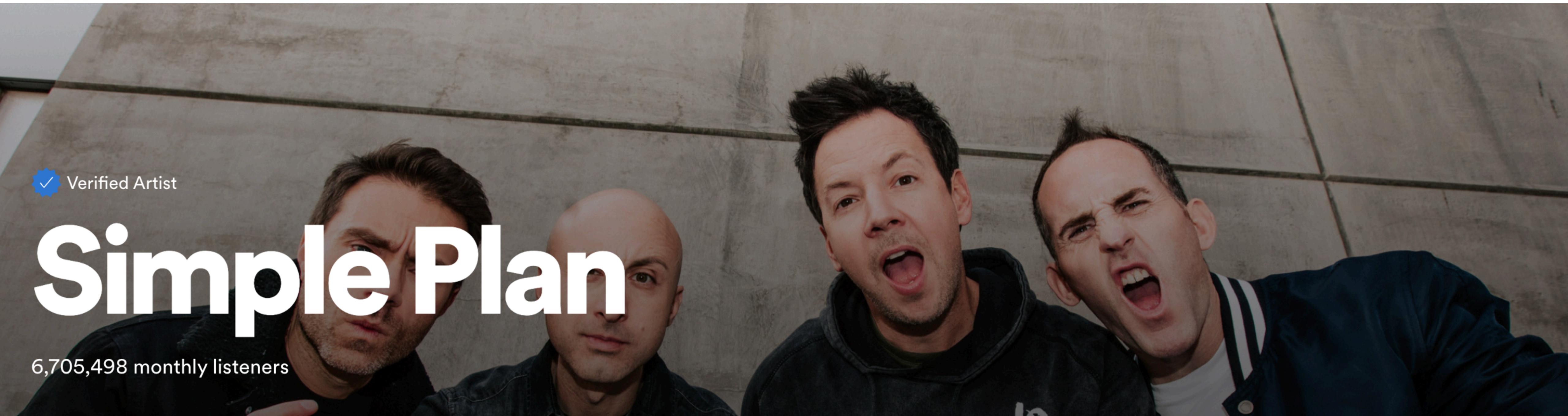
- Limited number of songs available
- Pre-trained models for audio → features
- Audio not straightforward to come by (e.g. copyright)

For now, let's devise a



# Simple Plan

6,705,498 monthly listeners





# Simple Plan

6,705,498 monthly listeners

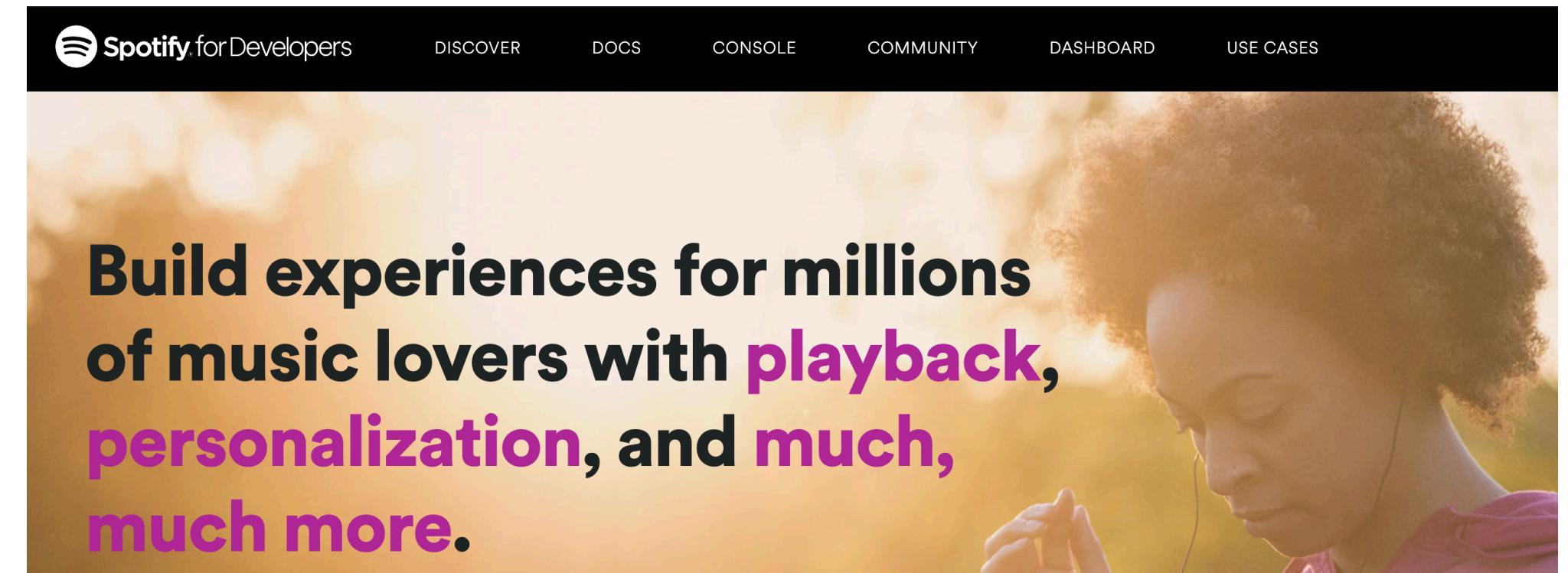


- Only use information relating to a person's music library
- Work with readily available audio features
- Quickly build an easy-to-use tool to consume recommendations

# Recommended Frequencies

## How?

### Use readily available audio features



### Spotify API

- Free
- Popular
- Audio features + metadata
- Python library

A screenshot of the Spotify API documentation for the Spotipy library. The left sidebar has a dark theme with white text and a blue header. It lists sections like "Welcome to Spotipy!", "Features", "Installation", "Getting Started", "Authorization Code Flow", "Client Credentials Flow", "IDs URIs and URLs", "Customized token caching", "Examples", "API Reference", and "client Module". The main content area has a light background with a large image of a hand holding a smartphone displaying the Spotify login screen. Above the image, it says "Docs » Welcome to Spotipy!" and "Edit on GitHub". Below the image, it says "Welcome to Spotipy!" and "Spotipy is a lightweight Python library for the Spotify Web API. With Spotipy you get full access to all of the music data provided by the Spotify platform.".

### Build a dashboard for interacting with results

### Streamlit

- “The fastest way to build and share data apps”

A screenshot of Streamlit. On the left, a terminal window titled "MyApp.py" shows Python code for a simple application:

```
import streamlit as st
import pandas as pd

st.write("""
# My first app
Hello *world!*"""
)

df = pd.read_csv("my_data.csv")
st.line_chart(df)
```

On the right, a browser window titled "My App • Streamlit" shows the resulting "My first app" page with the text "Hello world!" and a line chart.

Data, or should I say...



Verified Artist

# JSON Derulo

34,123,266 monthly listeners



# Spotify API

## Audio Features

### ▼ Body

#### ▼ **audio\_features** array of objects

##### **acousticness** number<float>

A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.

>= 0 <= 1

##### **danceability** number<float>

Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.

##### **energy** number<float>

Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.

##### **instrumentalness** number<float>

Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.

application/json

required

##### **liveness** number<float>

Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.

##### **loudness** number<float>

The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typically range between -60 and 0 db.

##### **speechiness** number<float>

Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.

##### **tempo** number<float>

The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.

##### **valence** number<float>

A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).

>= 0 <= 1

# Spotify API Song Metadata

**release\_date** string

required

The date the album was first released.

**genres** array of strings

A list of the genres the artist is associated with. If not yet classified, the array is empty.

A lot of genres exist (>5k)!

global genres  
local genres  
**random genres**  
any genre

finnish metal  
pop house  
reggae en espanol  
christian alternative rock  
pop argentino  
banda carnavalera  
haryani pop  
turkish pop  
norwegian indie  
art pop  
christian indie  
german techno  
cubaton  
circuit  
chinese drama ost  
hindi hip hop  
canadian underground hip hop  
cologne hip hop  
new romantic  
dutch hip hop

afripop  
atmospheric  
background  
blues  
brazil  
children  
christian  
classical  
comedy  
country  
edm  
electronic  
folk  
france  
hip hop  
india  
indie  
japan  
jazz  
latin  
metal  
mexican  
oldies  
pop  
punk  
r&b  
reggae  
rock  
soul  
soundtrack  
slovene

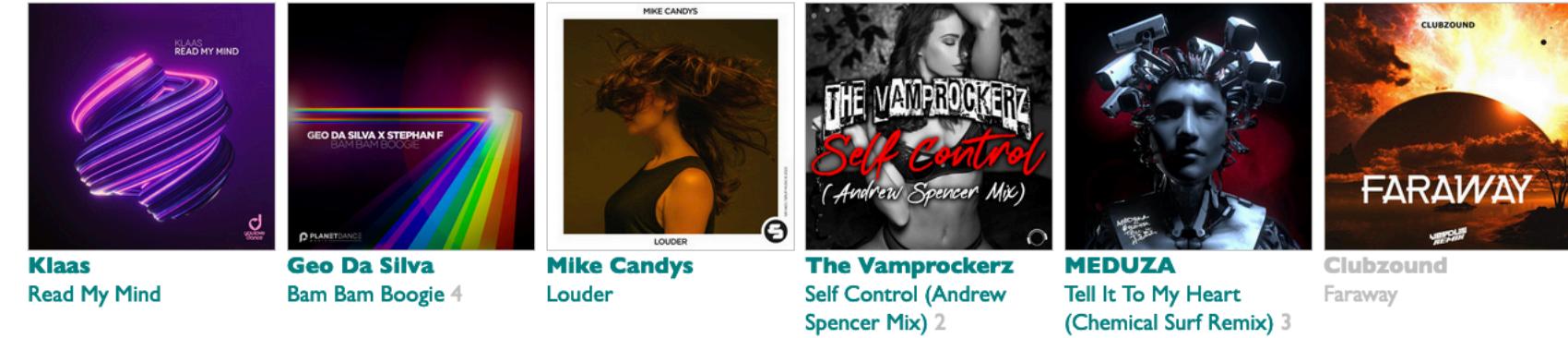
world folk  
world hip hop  
world pop  
world punk  
world rock

21st century classical  
432hz  
5th wave emo  
8-bit  
8d  
a cappella  
a3  
aberdeen indie  
abstract  
abstract beats  
abstract hip hop

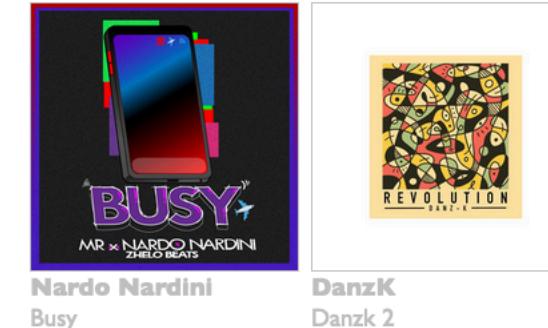
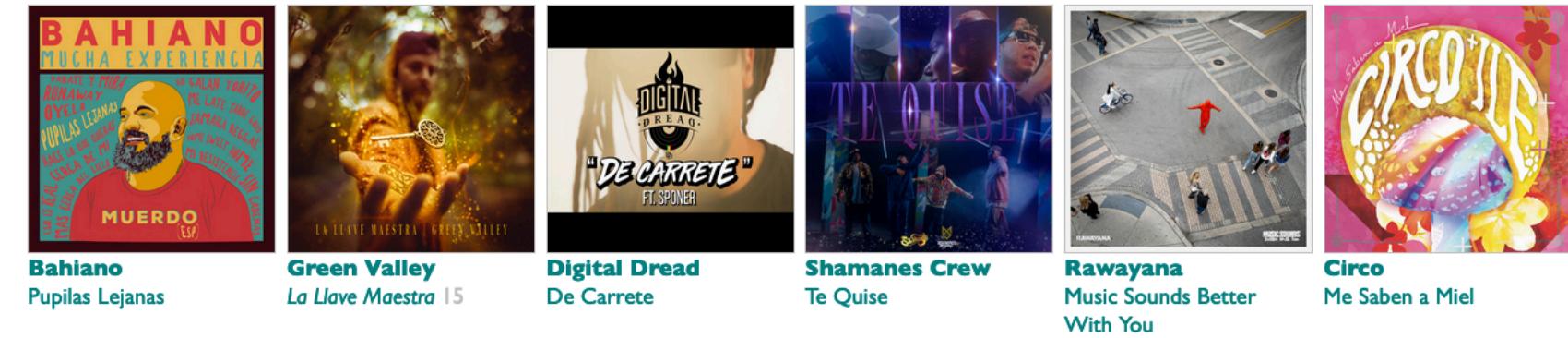
**finnish metal** 6



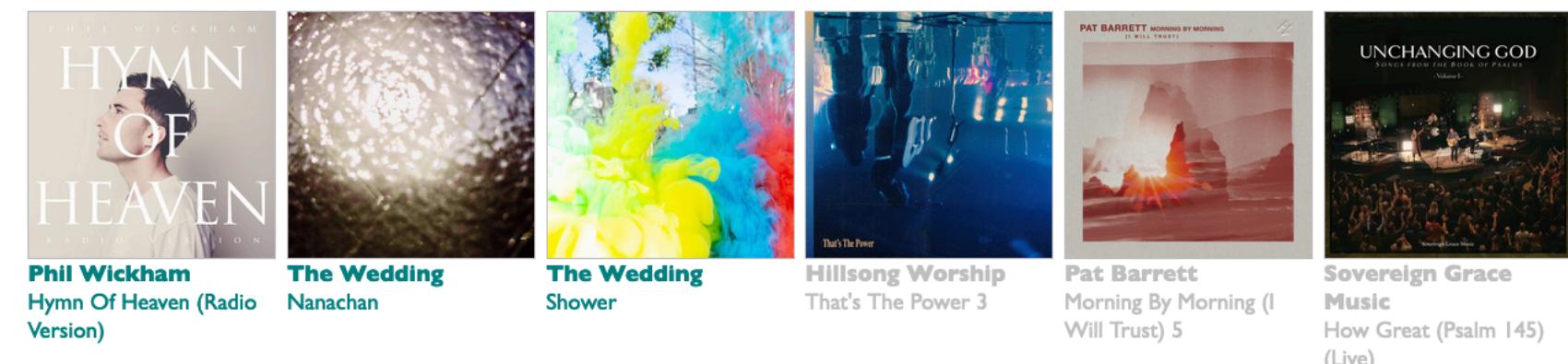
**pop house** 6



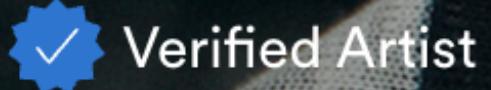
**reggae en espanol** 14



**christian alternative rock** 11



# Features



# Machine Gun Kelly → Learning

18,242,348 monthly listeners



Look at this graph...

... every time it makes me laugh

# Quantifying Song Attributes

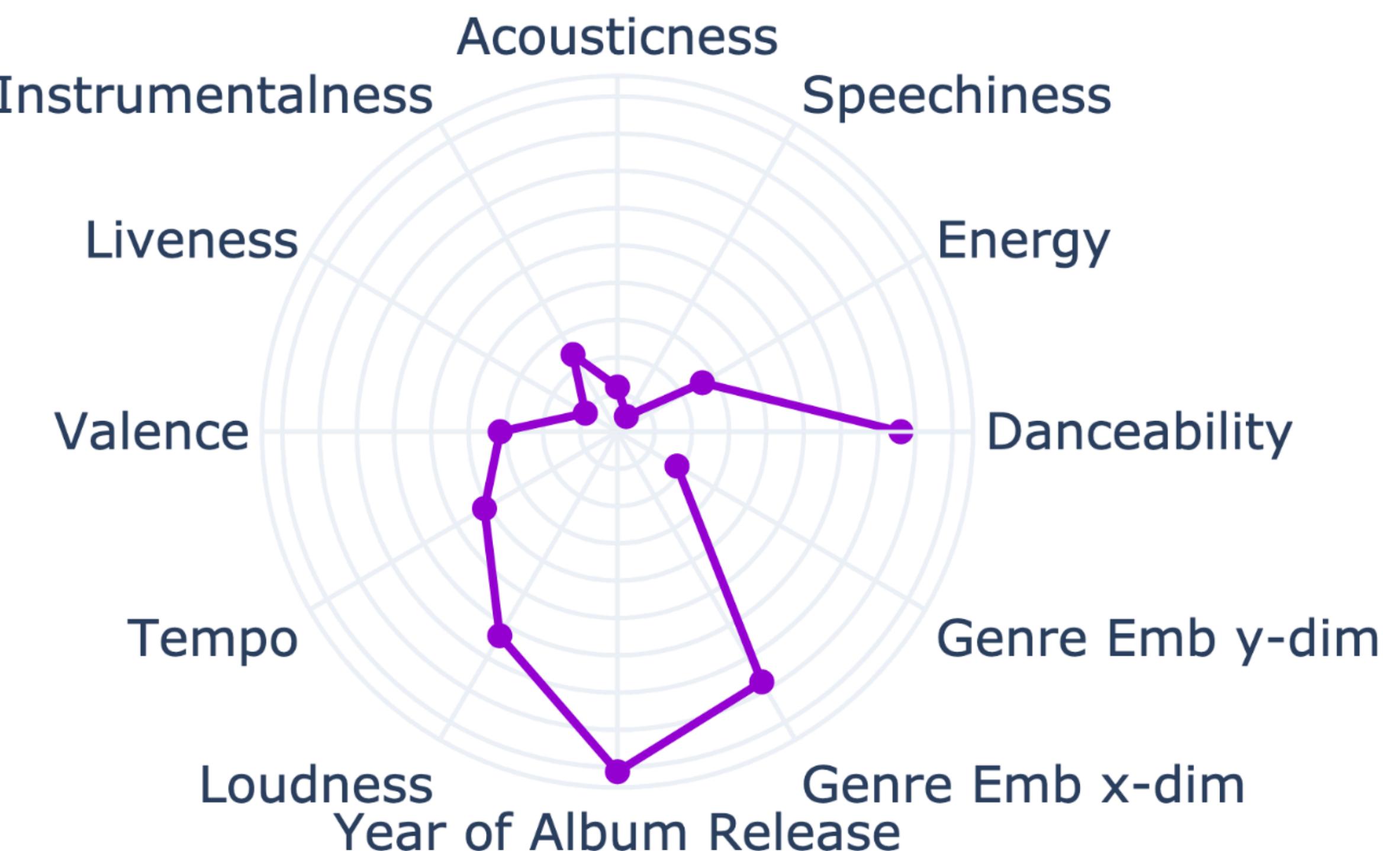
Feature Vector  $s \in \mathbb{R}^d$  with  $d = 12$

- 9 audio features
- 1 year of album release
- 2 coordinates of genre embedding

All features are scaled to fall into interval  $[0, 1]$

Genre embeddings are obtained from  
[everynoise.com](http://everynoise.com)

Feature vector for song "Cheer Up, My Brother" by HNNY



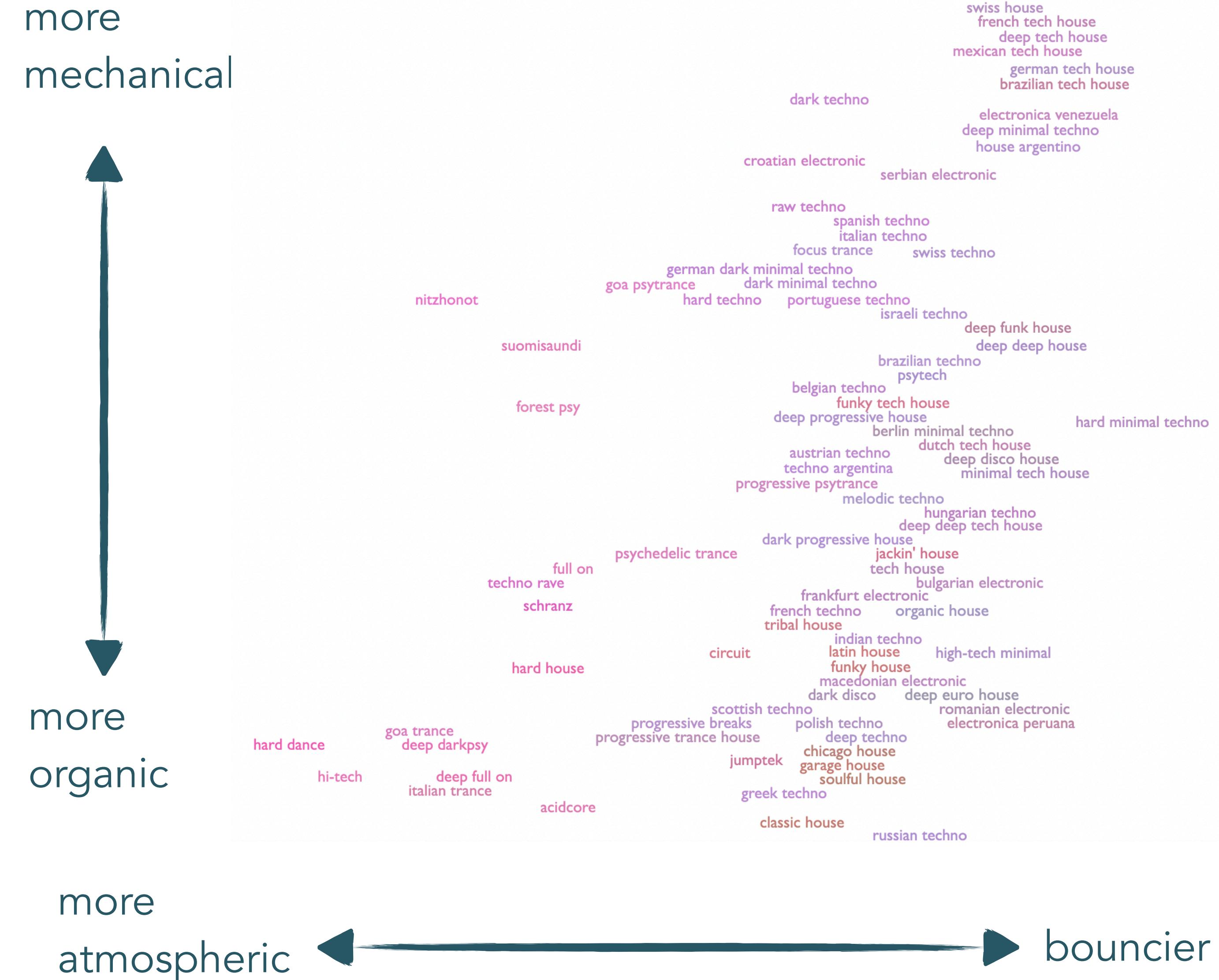
# Quantifying Genre

## Every Noise at Once

- [everynoise.com](http://everynoise.com) provides an “algorithmically-generated [...] scatter-plot of the musical genre-space” of Spotify
- Embeddings equal (x, y) coordinates of genre in plot, which can be obtained from the HTML source

```
><div id="item2627" preview_url="https://p.scdn.co/mp3-preview/8f54913108882b  
af589270d12840b087ae33adf0" class="genre scanme" scan="true" style="color: #d  
382c9; top: 0px; left: 1427px; font-size: 101%" onclick="playx("5K7Dmv4712ml7  
tbzvljdTA", "latin tech house", this);" title="e.g. Hector Couto "Salimo"">...
```

- Song's genre embedding equals  $(\bar{x}, \bar{y})$ , i.e. centroid of all associated genres
- Similar genres are supposed to be close in this embedding space



# Methodology, a.k.a

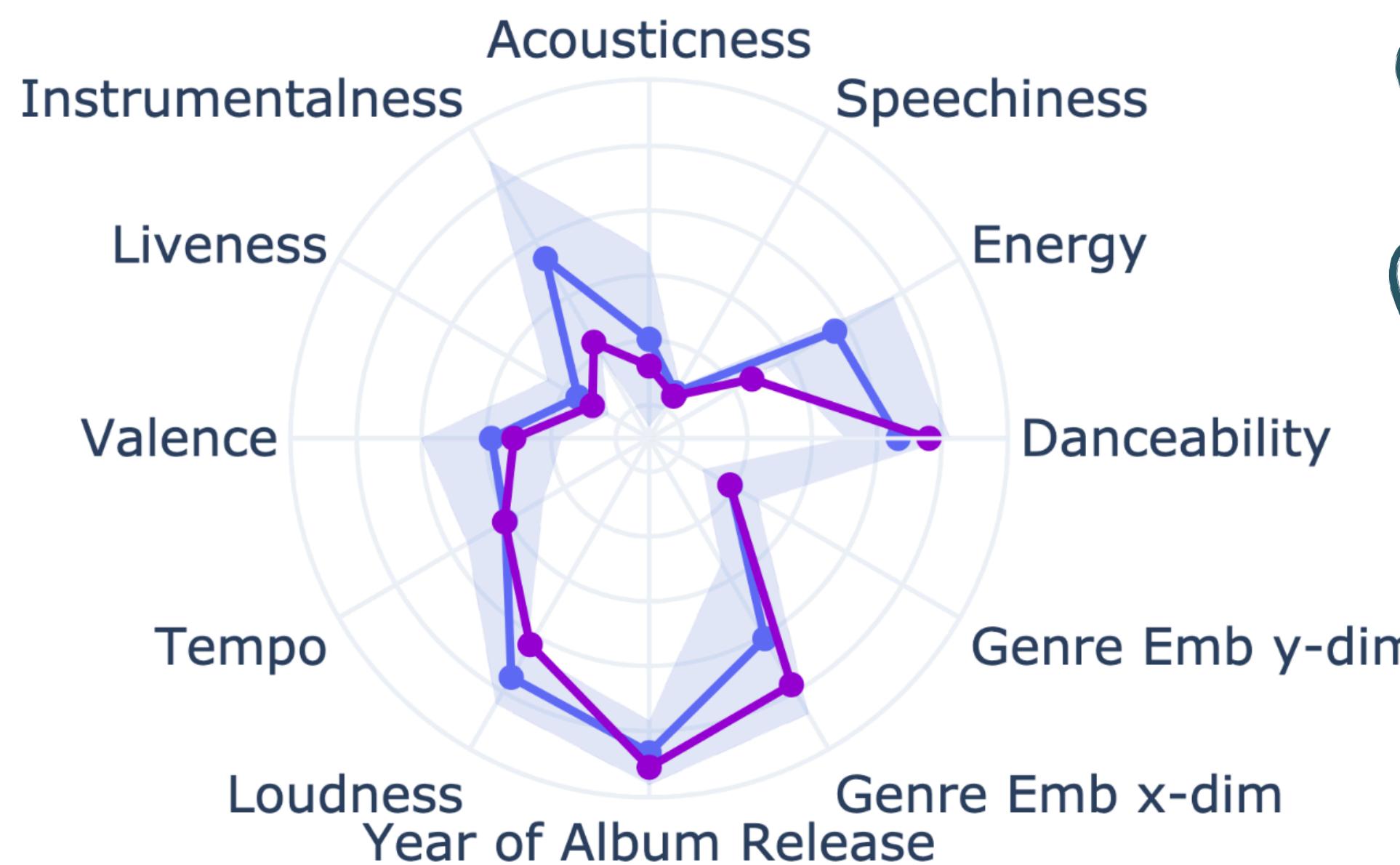
 Verified Artist

# Machine Gun Kelly → Learning

18,242,348 monthly listeners

# Song Similarity

Baseline: Euclidean Distance



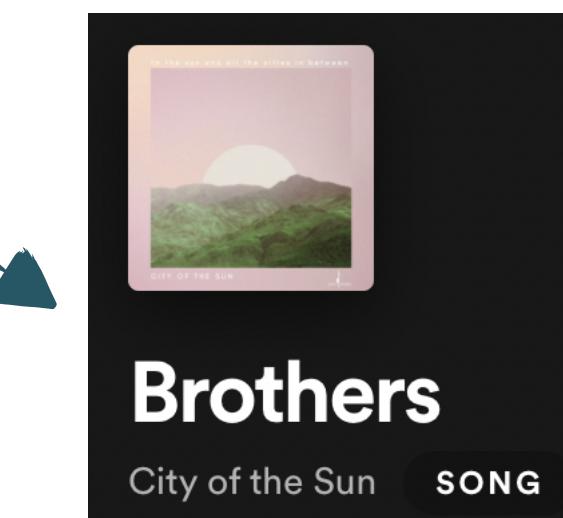
- Average of Playlist
- Song
- +/- Standard Deviation

PLAYLIST

## Classico - Adagio

# TITLE ALBUM DATE ADDED

#	TITLE	ALBUM	DATE ADDED
1	Berlin Song Ludovico Einaudi	Nightbook (Exclusive)	Jul 27, 2021
2	Two Trees Ludovico Einaudi	In A Time Lapse (Deluxe Edition)	Jul 27, 2021
3	Experience Ludovico Einaudi, Daniel Hope, I Virtuosi Italiani	In A Time Lapse	Jul 27, 2021
4	Fly - Reimagined by Mercan Dede and Dexter Crowe Ludovico Einaudi, Mercan Dede, Dexter Crowe	Reimagined. Volume 1, Chapter 1	Jul 27, 2021



Top 1 match

**Observation:** Euclidean distance between song features does not capture similarity very well!

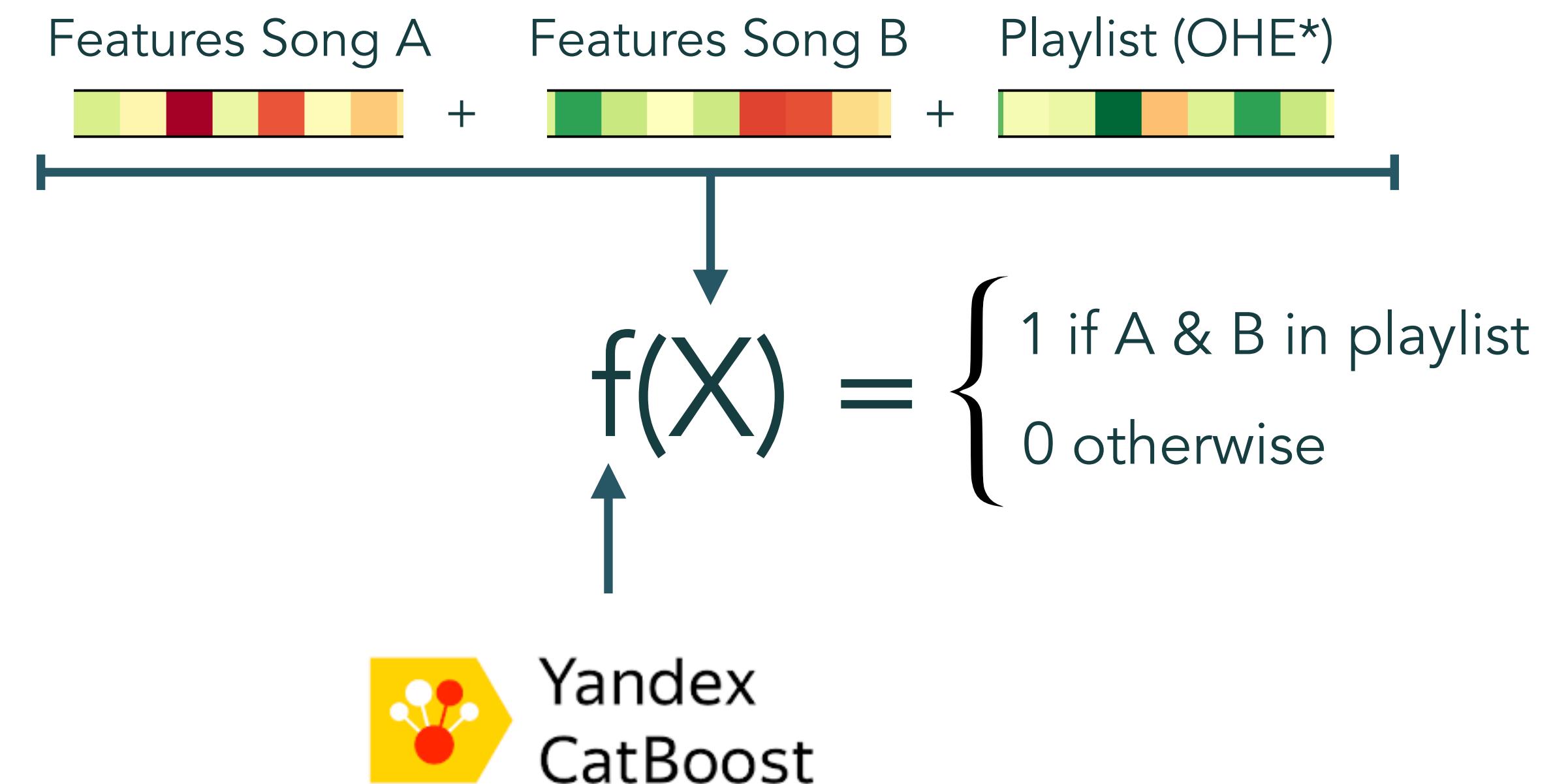
# Song Similarity

## Catboost

- Simple distance measure for feature vectors fails to capture subtleties in music similarity
- Similarity of songs is also conditional on playlist (to some degree)

### Solution

- Machine Learning Model → **Catboost**
- Trained by creating positive and negative song pairs
  - Positive: songs in the same playlist
  - Negative: songs in **very** different playlists



\*One-hot-encoded

YOU BROKE  
ME FIRST



Verified Artist

Dash Berlin board

1,356,807 monthly listeners



# Recommended Frequencies: A Recommendation System For Playlists

This app allows users to identify songs in their library that may fit well into a selected playlist

Page Settings

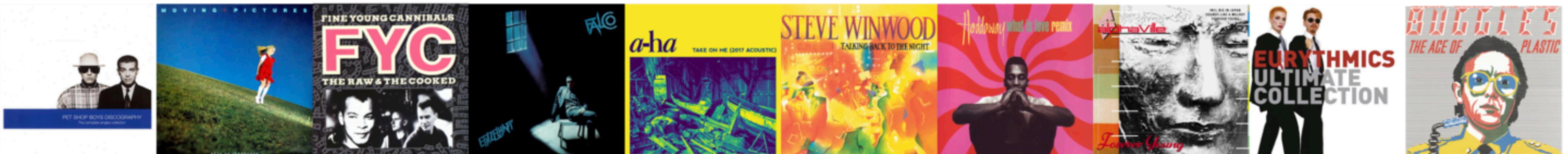
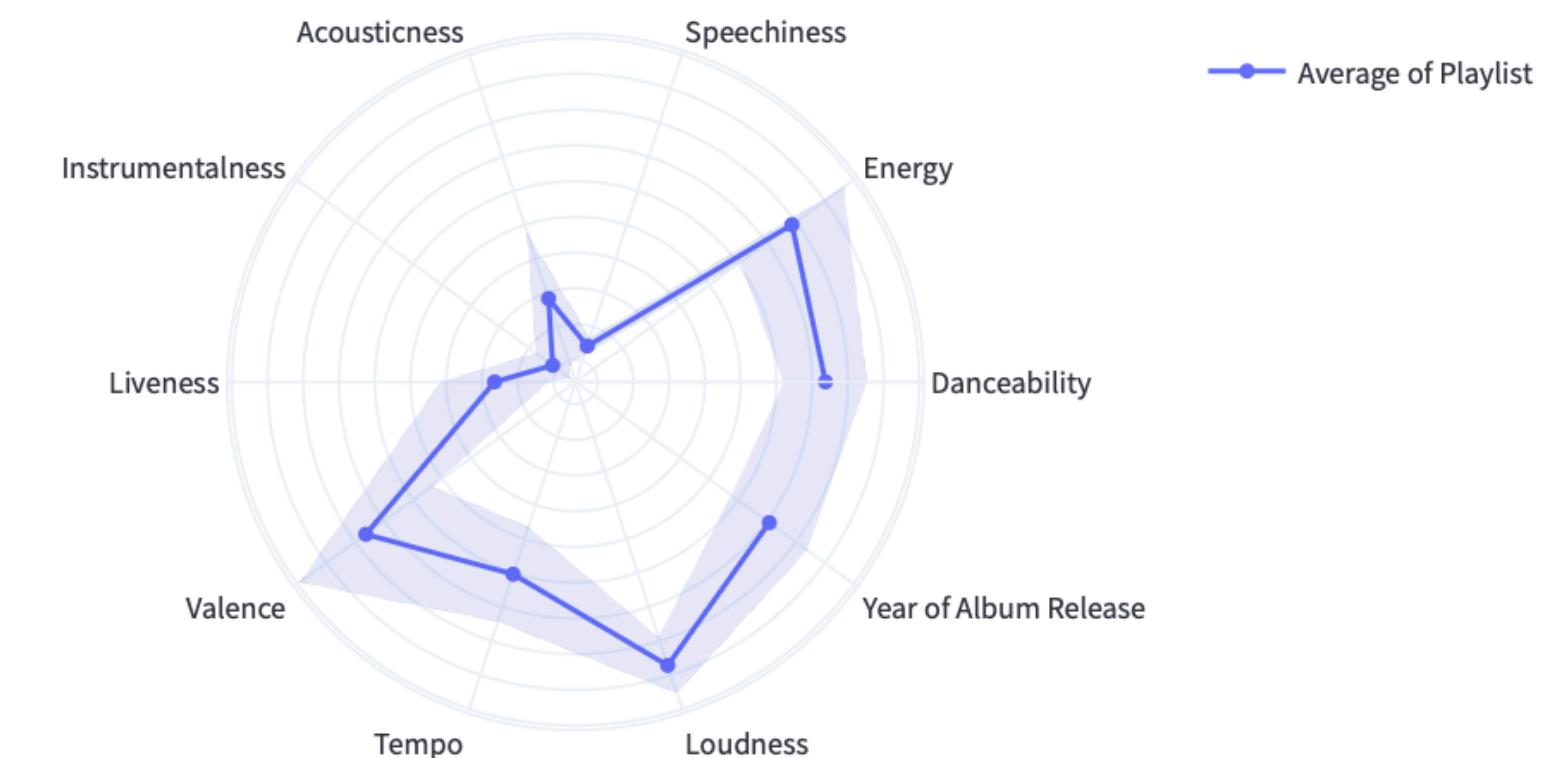
## 1. Choose a playlist

80's

### Some songs in playlist «80's»

	SongName	Artist	ID
0	Wouldn't It Be Good	Nik Kershaw	00FDHurakzVEiPutdUxXXq
1	Sledgehammer	Peter Gabriel	029NqmlySn1kOY305AAhxT
2	Always on My Mind	Pet Shop Boys	07ABETRdek3ACMpRPvQuaT
3	There Must Be An Angel	Luciano Pavarotti	0AeMfVr7exf7lf2RwcpPMc
4	200度	Sally Yeh	0CYelmjEps63DAuqLV9b6J
5	What About Me	Moving Pictures	0MHeOVBiRkg9mD29VMTkET
6	Stella Stai	Umberto Tozzi	0NUyAEi7WIinhF0SJGVavUG
7	What A Feeling	Irene Cara	0aAR5HogGoT68EWFbyRFqx
8	Tainted Love	Soft Cell	0cGG2EouYCEEC3xfa0tDFV
9	Ti Amo	Umberto Tozzi	0gAbf0NL9no1Urk1Wj8Uui

### Song attributes for playlist «80's»



## 2. Find similar songs

Similarity Settings

### Most similar songs to playlist «80's»

Result Page #

1

	SongName	Artist	ID	Similarity
0	Glory Days	Bruce Springsteen	2Y90nL1ohB4sgYELDs7uNx	98.3%
1	If You Love Somebody Set Them Free	Sting	5Xhqe9xu6bKRSqLj1mS1SB	97.8%
2	Walking On Sunshine	Katrina & The Waves	05wlrZSwuaVWhcv5FfqeH0	97.0%
3	Conga	Gloria Estefan	4aMT5LHe8A2ulc11H8Cx2m	96.4%
4	Hungry Heart	Bruce Springsteen	1Ksl8NEeAna8ZIdojI3FiT	96.3%
5	West End Girls - 2001 Remaster	Pet Shop Boys	2yzPBI5UXK2sqvnNM9QQ0	96.2%
6	Brandy (You're a Fine Girl) - Rerecorded	Looking Glass	4l4f4tl1JNGiXXynCoMe3j	95.8%
7	Englishman In New York	Sting	4KFM3A5QF2IMcc6nHsu3Wp	95.7%
8	Stars	Simply Red	75CgD6l7K4qMzZrn4CbZqz	95.6%
9	Holding Out for a Hero - From "Footloose" Soundtrack	Bonnie Tyler	5Hyr47BBGpvOfcykSCcaw9	95.5%

### Visualise similarity of proposed song

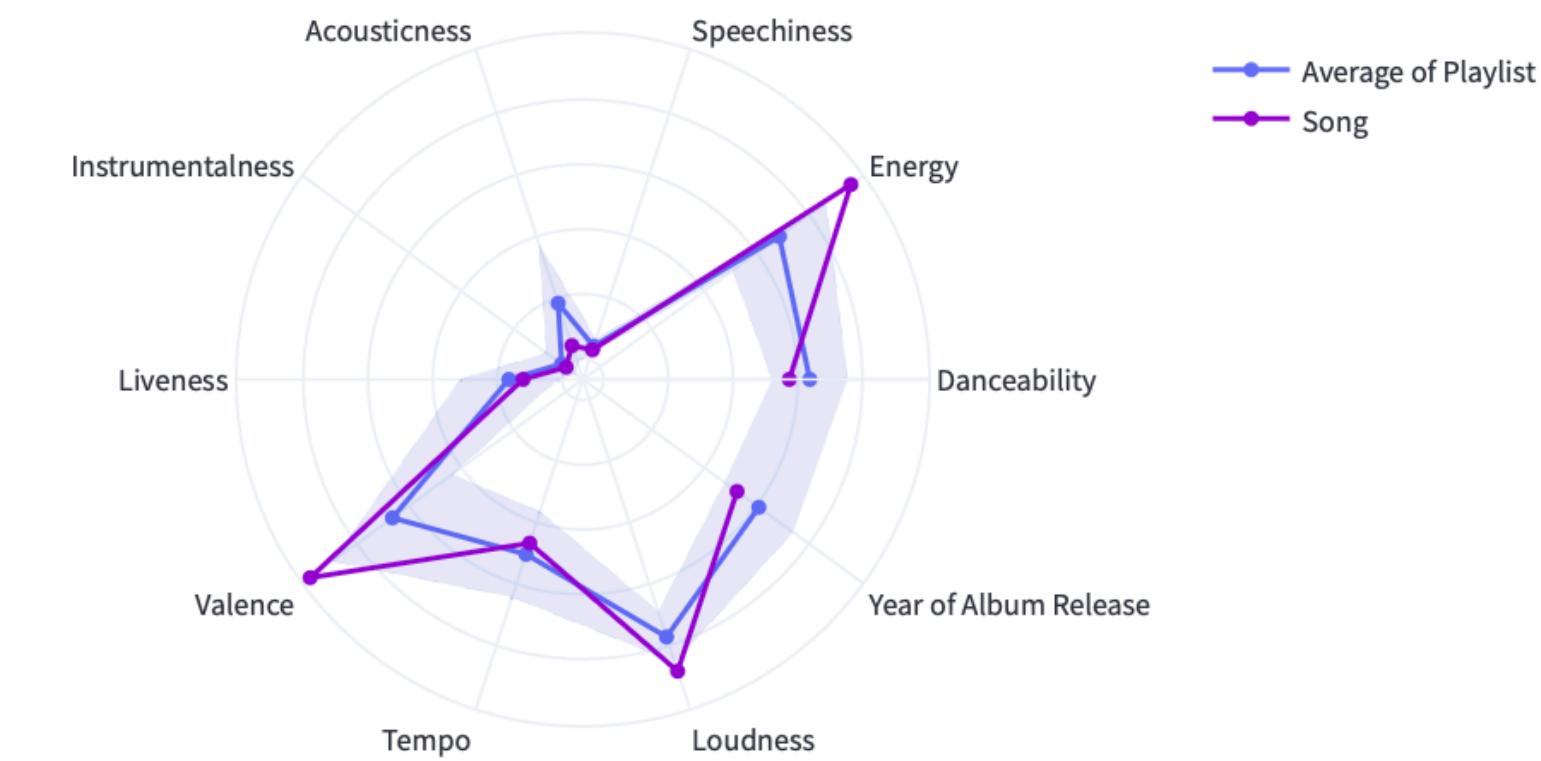
Search for song by

- ID
- Name

ID of song to visualise

2Y90nL1ohB4sgYELDs7uNx

Song: Glory Days by Bruce Springsteen  
Playlist: 80's



### Listen to proposed song



# Let's deep dive into some results



ALBUM

## Results May Vary



Limp Bizkit • 2003 • 18 songs, 1 hr 8 min

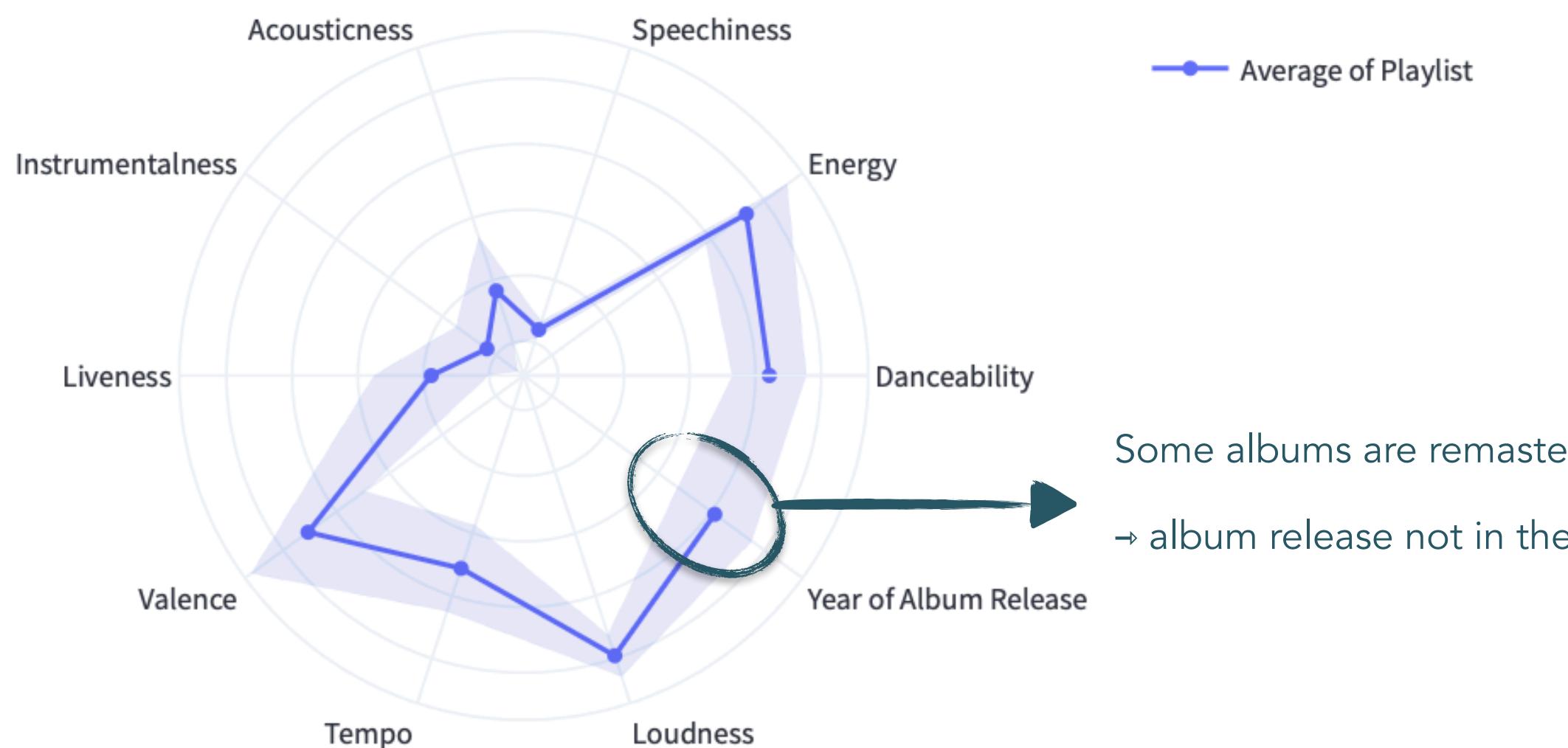
# «80's Retro»

## for that 80's feeling



Example of songs in the playlist

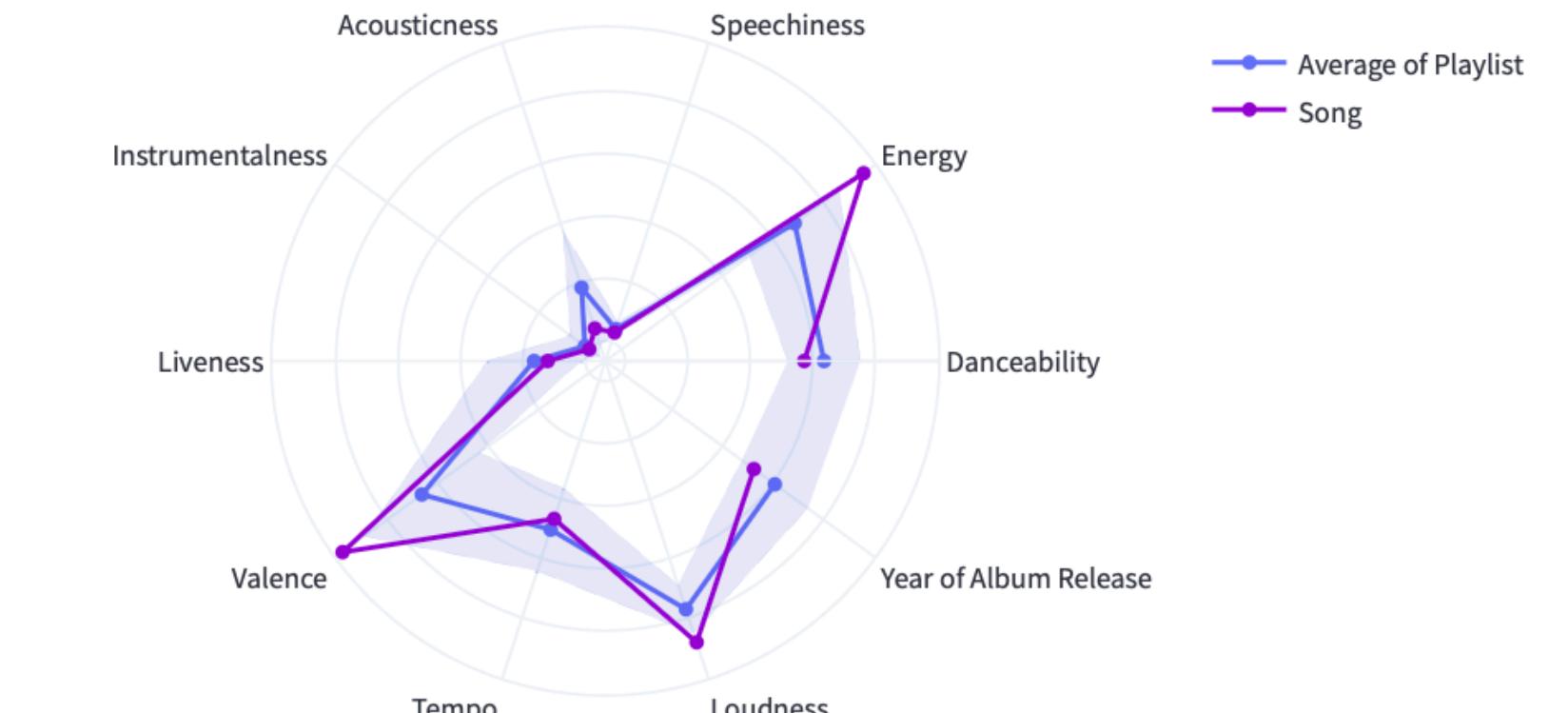
Sledgehammer	Peter Gabriel
200度	Sally Yeh
Stella Stai	Umberto Tozzi
Tainted Love	Soft Cell
The Rhythm of the Night	Corona
Just Can't Get Enough	Depeche Mode
I'm So Excited	The Pointer Sisters
Like a Prayer	Madonna
Two Tickets to Paradise	Eddie Money
Tarzan Boy	Baltimora



Top recommendations

🤔	Glory Days	Bruce Springsteen
🤔	If You Love Somebody, Set Them Free	Sting
👍	Walking On Sunshine	Katrina & The Waves
👌	Conga	Gloria Estefan
🤔	Hungry Heart	Bruce Springsteen
👌	West End Girls	Pet Shop Boys
🤔	Brandy (You're a Fine Girl)	Looking Glass
🤔	Englishman in New York	Sting
👍	Stars	Simply Red
👌	Holding Out for a Hero	Bonnie Tyler

Song: Glory Days by Bruce Springsteen  
Playlist: 80's





Song

# Takeaway(s)

The Chainsmokers

- Catboost model better picks up on subtleties of song similarity
- Highly depends on creating informative positive and negative song pairs
- Model works well for homogeneous playlists (e.g. same style, era, etc.) much less for say “best of library” playlists

# Some possible improvements for the



- Include language of songs to filter on
- Address issue with remastered albums by getting oldest release
- Allow specific filters (e.g. acceptable interval of valence)
  - Useful for playlists that capture certain mood

The End

