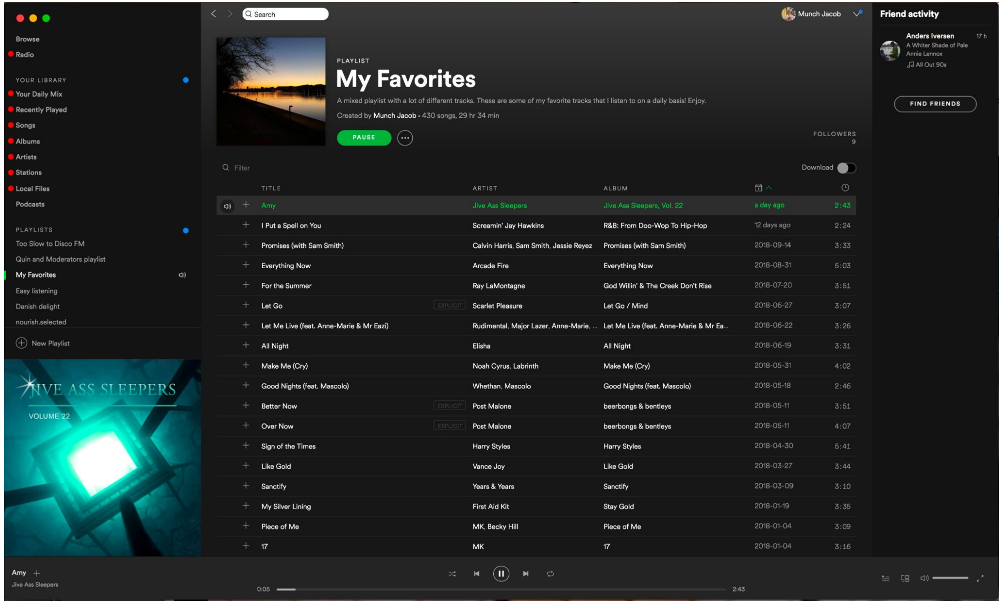
Spotify Algorithm

By Eun Lee, Joa Nguyen, Gyujin Hong, Suebeen Noh

Table of contents

1. Introduction
2. Motivation/Related works
3. Data
4. Methods
5. Result
6. Discussion



PLAYLIST

My Favorites

A mixed playlist with a lot of different tracks. These are some of my favorite tracks that I listen to on a daily basis! Enjoy.

Created by Munch Jacob • 430 songs, 29 hr 34 min

TITLE		ARTIST	ALBUM		
Amy		Jive Ass Sleepers	Jive Ass Sleepers, Vol. 22	a day ago	2:43
I Put a Spell on You		Screamin' Jay Hawkins	R&B: From Doo-Wop To Hip-Hop	12 days ago	2:24
Promises (with Sam Smith)		Calvin Harris, Sam Smith, Jessie Reyez	Promises (with Sam Smith)	2018-09-14	3:33
Everything Now		Arcade Fire	Everything Now	2018-08-31	5:03
For the Summer		Ray LaMontagne	God Willin' & The Creek Don't Rise	2018-07-20	3:51
Let Go	EXPLICIT	Scarlet Pleasure	Let Go / Mind	2018-06-27	3:07
Let Me Live (feat. Anne-Marie & Mr Eazi)		Rudimental, Major Lazer, Anne-Marie, ...	Let Me Live (feat. Anne-Marie & Mr Ea...	2018-06-22	3:26
All Night		Elisha	All Night	2018-06-19	3:31
Make Me (Cry)		Noah Cyrus, Labrinth	Make Me (Cry)	2018-05-31	4:02
Good Nights (feat. Mascolo)		Whethan, Mascolo	Good Nights (feat. Mascolo)	2018-05-18	2:46
Better Now	EXPLICIT	Post Malone	beerbongs & bentleys	2018-05-11	3:51
Over Now	EXPLICIT	Post Malone	beerbongs & bentleys	2018-05-11	4:07
Sign of the Times		Harry Styles	Harry Styles	2018-04-30	5:41
Like Gold		Vance Joy	Like Gold	2018-03-27	3:44
Sanctify		Years & Years	Sanctify	2018-03-09	3:10
My Silver Lining		First Aid Kit	Stay Gold	2018-01-19	3:35
Piece of Me		MK, Becky Hill	Piece of Me	2018-01-04	3:09
17		MK	17	2018-01-04	3:16

Introduction

- Create cluster models using song data from spotify to find songs similar to one another
- Music listeners all have different tastes and preferences when it comes to their music, and being able to find similar songs will help aid users in finding their next favorite song
- Our group plans to create cluster models (K Means, Mixture of Gaussian) to group similar songs and see which songs may be recommended to a Spotify user after seeing their playlists/favorite songs

Motivation

❏ Song varies depending on the day’s mood and using the unsupervised learning to recommend similar taste of music is more convenient than having to find each individual songs

Related works

❏ K-means (unsupervised)

❏ PCA (unsupervised)

❏ Mixture of Gaussians (unsupervised)

Related Works - Literature

PCA with Recommending systems

- Dimension reduction

Yin, C.-X., & Peng, Q.-K. (2012). A careful assessment of recommendation algorithms related to dimension reduction techniques. *Knowledge-Based Systems*, *27*, 407–423. https://doi.org/10.1016/j.knosys.2011.11.022

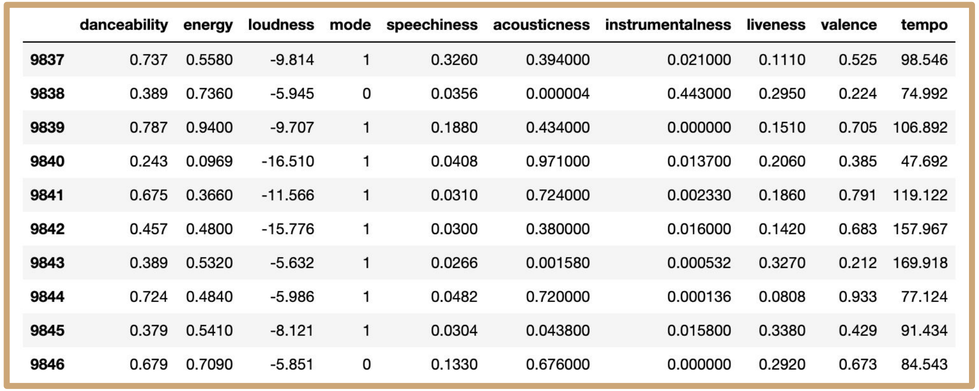
Mixture of Gaussians

- Clustering within multi dimensional model
- Convergence to local minimum
- Elliptical boundaries with covariance matrices

Agarwal, N., Sikka, G., & Awasthi, L. K. (2022). Evaluation of web service clustering using Dirichlet Multinomial Mixture model based approach for Dimensionality Reduction in service representation. *Information Processing & Management, 57,* 102238. https://www.sciencedirect.com/science/article/pii/S0306457320300492

Data

- Imported songs using Spotify API with the help if a python library called spotipy
- Imported data from a 10,000 song playlist called “Top 10,000 songs”
    - Has enough variety in song genre/styles to be a usable dataset
- Dropped columns we determined were not needed for song similarity

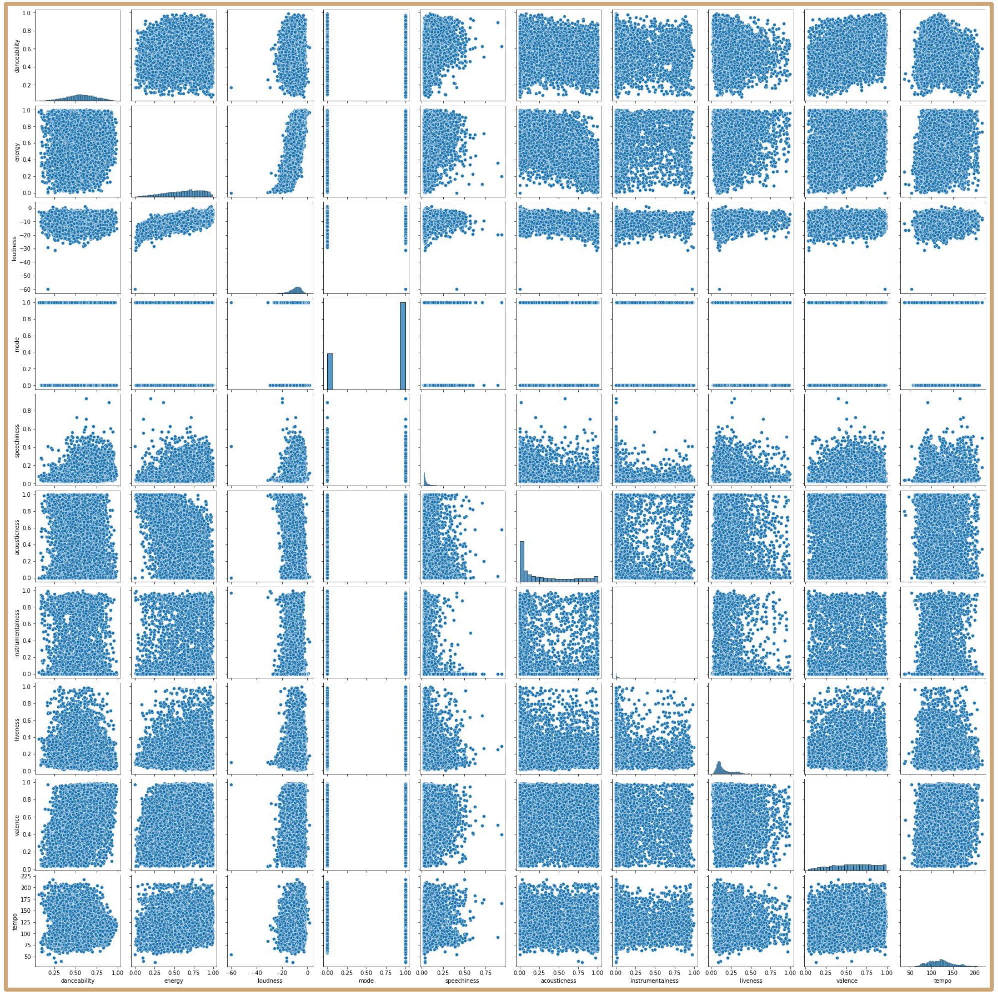


	danceability	energy	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo
9837	0.737	0.5580	-9.814	1	0.3260	0.394000	0.021000	0.1110	0.525	98.546
9838	0.389	0.7360	-5.945	0	0.0356	0.000004	0.443000	0.2950	0.224	74.992
9839	0.787	0.9400	-9.707	1	0.1880	0.434000	0.000000	0.1510	0.705	106.892
9840	0.243	0.0969	-16.510	1	0.0408	0.971000	0.013700	0.2060	0.385	47.692
9841	0.675	0.3660	-11.566	1	0.0310	0.724000	0.002330	0.1860	0.791	119.122
9842	0.457	0.4800	-15.776	1	0.0300	0.380000	0.016000	0.1420	0.683	157.967
9843	0.389	0.5320	-5.632	1	0.0266	0.001580	0.000532	0.3270	0.212	169.918
9844	0.724	0.4840	-5.986	1	0.0482	0.720000	0.000136	0.0808	0.933	77.124
9845	0.379	0.5410	-8.121	1	0.0304	0.043800	0.015800	0.3380	0.429	91.434
9846	0.679	0.7090	-5.851	0	0.1330	0.676000	0.000000	0.2920	0.673	84.543

Data

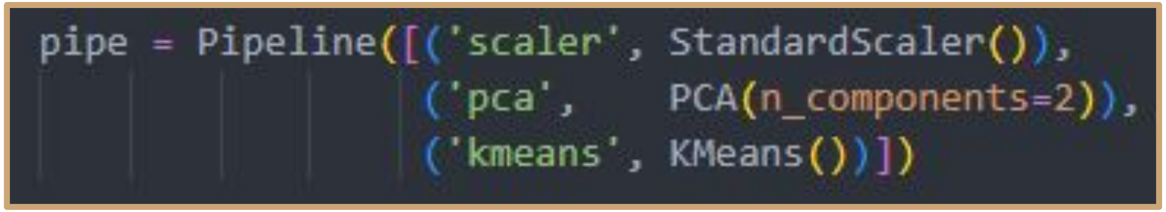
❑ PCA to decrease the dimensionality to 2 principle components (for data visualization) and scaled the data using StandardScaler()

❑ Pairplot to see any relationship between the variables



Methods

- Pipeline includes…
  - Standardization method:
    - StandardScaler()
  - PCA
  - Cluster Model Methods
    - K-Means: KMeans()
    - Mixture of Gaussians: GaussianMixture()

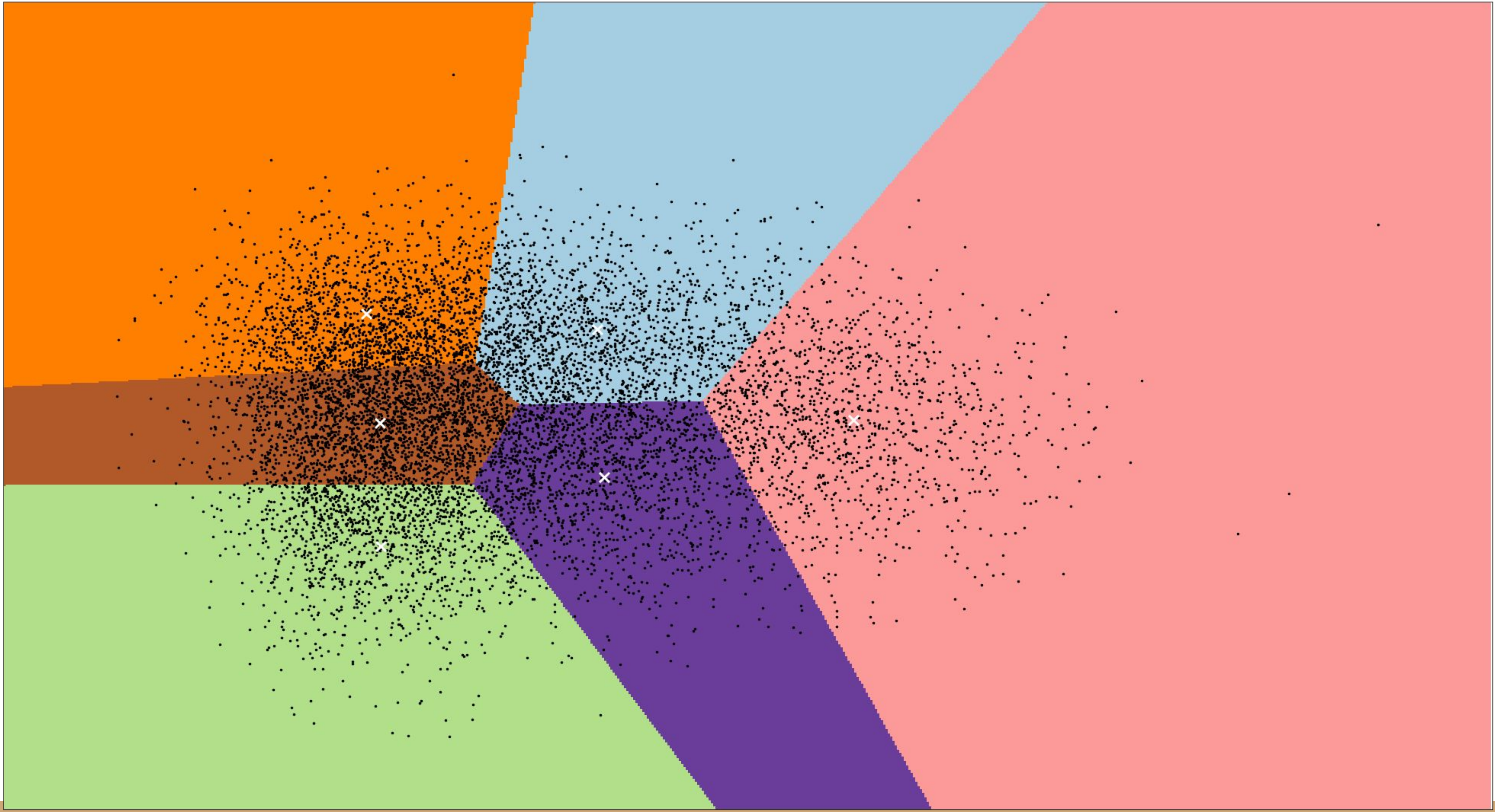


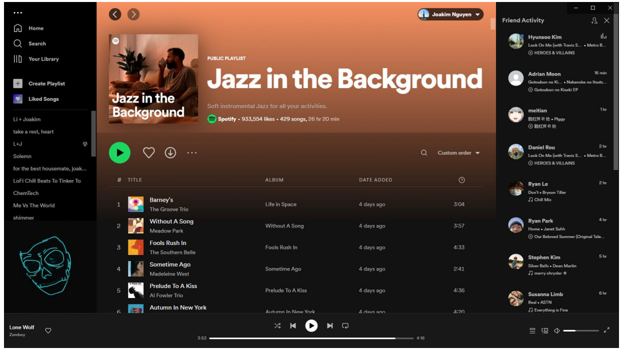
```
pipe = Pipeline([('scaler', StandardScaler()),
                 ('pca',    PCA(n_components=2)),
                 ('kmeans', KMeans())])
```

Methods - K Means

- GridSearchCV() method to determine the number of means to use for the model
  - Found that data was so clumped it was difficult to find optimal k-means using gridsearch
  - After adding additional playlists (sad, happy, jazz, etc.) we found 6 fit the model the best
- Matplotlib to create a meshgrid to visualize model
- Added the 6 new playlists onto the model fitted by the original data

K-means clustering on the digits dataset (PCA-reduced data)
Centroids are marked with white cross





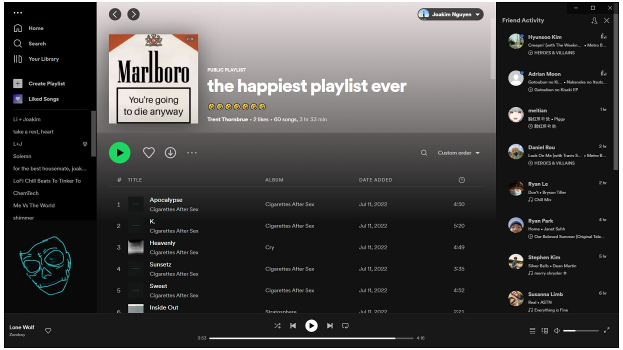
PUBLIC PLAYLIST

Jazz in the Background

Soft instrumental Jazz for all your activities.

Spotify • 933,554 likes • 429 songs, 26 hr 20 min

#	TITLE	ALBUM	DATE ADDED	
1	Barney's The Groove Trio	Life in Space	4 days ago	3:04
2	Without A Song Meadow Park	Without A Song	4 days ago	3:57
3	Fools Rush In The Southern Belle	Fools Rush In	4 days ago	4:33
4	Sometime Ago Madeleine West	Sometime Ago	4 days ago	2:41
5	Prelude To A Kiss Al Fowler Trio	Prelude To A Kiss	4 days ago	4:36
6	Autumn In New York	Autumn In New York	4 days ago	4:20

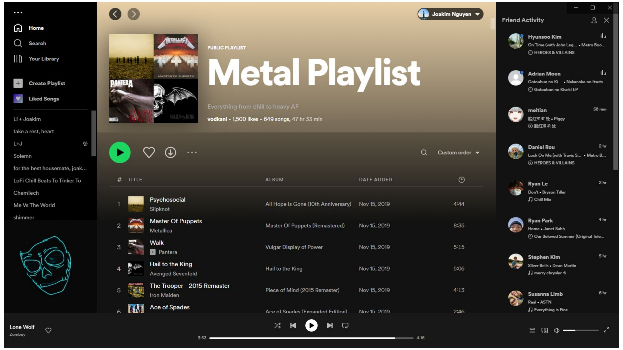


PUBLIC PLAYLIST

the happiest playlist ever

Trent Thornbrue • 2 likes • 60 songs, 3 hr 33 min

#	TITLE	ALBUM	DATE ADDED	
1	Apocalypse Cigarettes After Sex	Cigarettes After Sex	Jul 11, 2022	4:50
2	K. Cigarettes After Sex	Cigarettes After Sex	Jul 11, 2022	5:20
3	Heavenly Cigarettes After Sex	Cry	Jul 11, 2022	4:49
4	Sunsetz Cigarettes After Sex	Cigarettes After Sex	Jul 11, 2022	3:35
5	Sweet Cigarettes After Sex	Cigarettes After Sex	Jul 11, 2022	4:52
6	Inside Out	Stratosphere	Jul 11, 2022	3:21



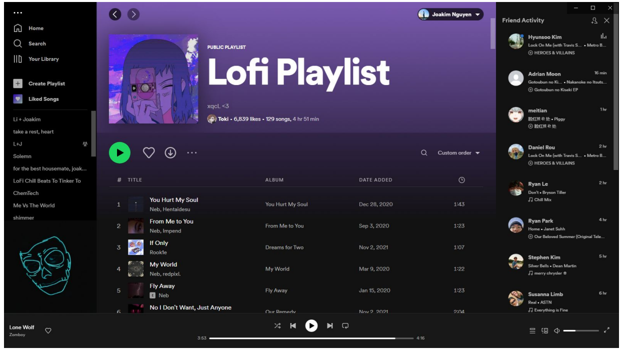
PUBLIC PLAYLIST

Metal Playlist

Everything from chill to heavy AF

vodkanl • 1,500 likes • 649 songs, 47 hr 33 min

#	TITLE	ALBUM	DATE ADDED	
1	Psychosocial Slipknot	All Hope Is Gone (10th Anniversary)	Nov 15, 2019	4:44
2	Master Of Puppets Metallica	Master Of Puppets (Remastered)	Nov 15, 2019	8:35
3	Walk Pantera	Vulgar Display of Power	Nov 15, 2019	5:15
4	Hail to the King Avenged Sevenfold	Hail to the King	Nov 15, 2019	5:06
5	The Trooper - 2015 Remaster Iron Maiden	Piece of Mind (2015 Remaster)	Nov 15, 2019	4:13
6	Ace of Spades	Ace of Spades (Expanded Edition)	Nov 15, 2019	2:46



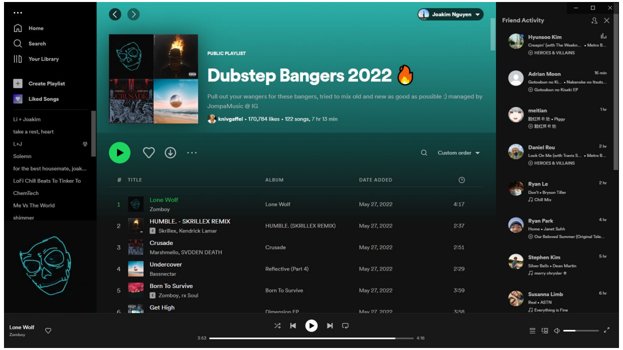
PUBLIC PLAYLIST

Lofi Playlist

xqcL <3

Toki • 6,839 likes • 129 songs, 4 hr 51 min

#	TITLE	ALBUM	DATE ADDED	
1	You Hurt My Soul Neb, Hentaideesu	You Hurt My Soul	Dec 28, 2020	1:43
2	From Me to You Neb, Impend	From Me to You	Sep 3, 2020	1:23
3	If Only Rookie	Dreams for Two	Nov 2, 2021	1:07
4	My World Neb, redpixl.	My World	Mar 9, 2020	1:22
5	Fly Away Neb	Fly Away	Jan 15, 2020	1:23
6	No I Don't Want, Just Anyone	Our Remedy	Nov 2, 2021	2:04



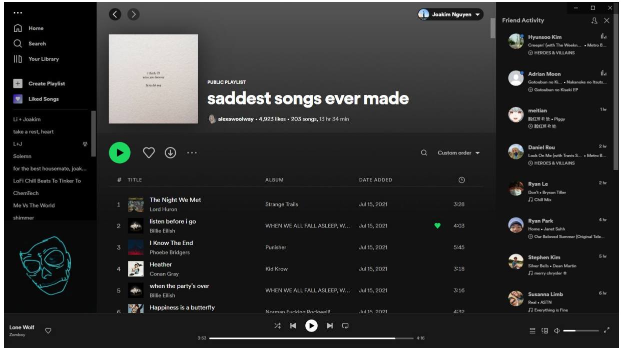
PUBLIC PLAYLIST

Dubstep Bangers 2022 🔥

Pull out your wangers for these bangers, tried to mix old and new as good as possible :) managed by JompaMusic @ IG

knivgaffel • 170,784 likes • 122 songs, 7 hr 13 min

#	TITLE	ALBUM	DATE ADDED	
1	Lone Wolf Zomboy	Lone Wolf	May 27, 2022	4:17
2	HUMBLE. - SKRILLEX REMIX Skrillex, Kendrick Lamar	HUMBLE. (SKRILLEX REMIX)	May 27, 2022	2:37
3	Crusade Marshmello, SVDDEN DEATH	Crusade	May 27, 2022	2:51
4	Undercover Bassnectar	Reflective (Part 4)	May 27, 2022	2:29
5	Born To Survive Zomboy, rx Soul	Born To Survive	May 27, 2022	3:59
6	Get High	Dimension EP	May 27, 2022	3:58



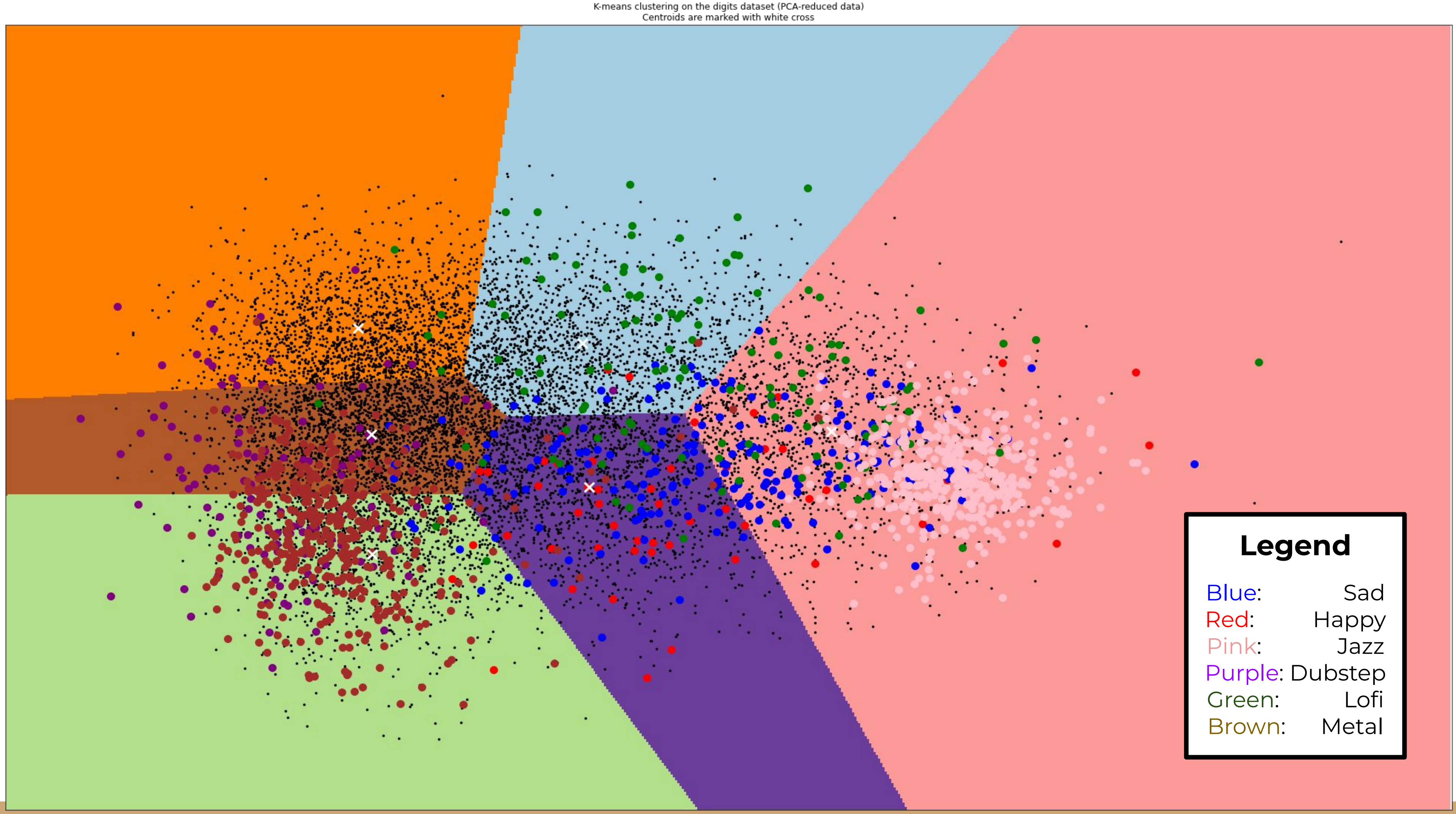
PUBLIC PLAYLIST

saddest songs ever made

alexawoolway • 4,923 likes • 203 songs, 13 hr 34 min

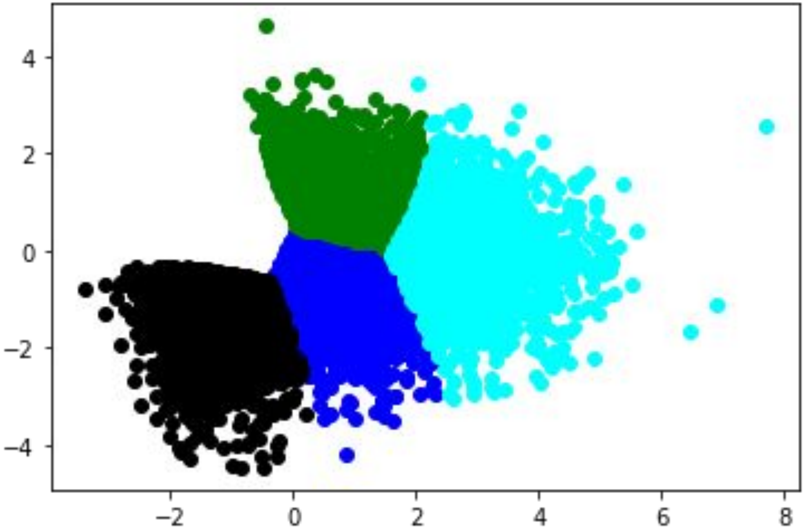
#	TITLE	ALBUM	DATE ADDED	
1	The Night We Met Lord Huron	Strange Trails	Jul 15, 2021	3:28
2	listen before i go Billie Eilish	WHEN WE ALL FALL ASLEEP, W...	Jul 15, 2021	4:03
3	I Know The End Phoebe Bridgers	Punisher	Jul 15, 2021	5:45
4	Heather Conan Gray	Kid Krow	Jul 15, 2021	3:18
5	when the party's over Billie Eilish	WHEN WE ALL FALL ASLEEP, W...	Jul 15, 2021	3:16
6	Happiness is a butterfly	Norman Fucking Rockwell!	Jul 15, 2021	4:32

6 unique playlists with their own genres


K-means clustering on the digits dataset (PCA-reduced data)
Centroids are marked with white cross
Legend
Blue: Sad
Red: Happy
Pink: Jazz
Purple: Dubstep
Green: Lofi
Brown: Metal

Methods - Mixture of Gaussians

- Added ('MoG', GaussianMixture()) into the pipeline
- Visualized Data using sklearn.cluster
- Ultimately found that since the song data is not randomly distributed, clusters were not gaussians



Result - K Means

- After visualizing the data, we compared the resulting graphs ranging from five to twelve cluster points (k) and concluded that six clusters (k=6) fitted our data best.

- After the best k value was determined, we found six different genres of playlists with distinctively strong features of song data.

- After overlaying the six playlists’ clusters identified by different colors on top of our k-means model on the 10,000 random songs, we tried to identify which k-cluster certain playlists overlapped with the best

Discussion

- Deeper understanding of:
    - Implementing spotipy (Spotify API wrapper), pipeline to find a recommender algorithm
    - How K-means and PCA's dimensionality reduction worked well to reduce noise
- Could have increase the principal components for a more accurate model
- Mixture of Gaussian isn't the best model since the data is not random
- Our playlist was only selected with sample of 10,000 songs from a single playlist (which could potentially be biased)

References

- https://www.sciencedirect.com/science/article/pii/S0167865509002323?casa_token=HXAcvkia2XMAAAAA:DbZtC3vPao5Js7T53qhcVI8svx_OiHZ9NEzfBwH_o2qlfEOSyJqsy48H6JjP_IAmeNvOI4iPrN0
- https://www.analyticsvidhya.com/blog/2019/10/gaussian-mixture-models-clustering/
- https://developer.spotify.com/documentation/web-api/
- https://spotipy.readthedocs.io/en/2.21.0/