Spotify Music Playlist Recommender



Claire Medina, Ferdie Taruc, Psalm Masaya



Claire Medina

MINUTES LISTENED 32,934

TOP GENRE Lo-Fi Beats

TOP ARTIST **Jhené Aiko**



Ferdie Taruc

MINUTES LISTENED

TOP GENRE Alt R&B

TOP ARTIST **Dijon**



Psalm Masaya

MINUTES LISTENED **54,046**

TOP GENRE Show Tunes

TOP ARTIST Elton John



Context

• Streaming services revolutionized accessibility of music







Research Question:

Given user preferences, how can we create a playlist that can encompass the interests of different people?

Motivation

- Creation of shared playlists with others
- Catering to different interests
- Current Projects: Recommending Songs based on mood, etc.
- This Project: Appealing to Different Users

Definition of "Playlists"

- Unit composed of songs in some ordering
- Subjective
- Definition: Group of songs that fit a certain theme or genre



Popular playlists



RapCaviar New music from Lil Baby and Lil Wayne featuring Drake.



Feelin' Myself When you feel like you're 100% that b****



Get Turnt Mode: Turnt 5,947,610 FOLLOWERS

Limitations to this Definition

D

С

W

- Why this definition?
 - Difficulty inDimensionReduction
- Can no longer capture:
 - Context of the playlist
- Content -based filtering

SHUFFLE PLAY	SHUFFLE PLAY	
	Never Gonna Give You Up	
ownload	Rick Astley • Whenever You Need Someb	
	Never Gonna	
The Internet, KAYTRANADA • Ego Death	The Stoned • Never Gonna	
t's Get It Started		
e Black Eyed Peas • Let's Get It Started	 Let You Down	
ome Over Here	Seether + Let You Down	
shua Ticsay • Rhythm and Greens	 Never Coppa	
	The 808 • The 808	
e Down याटाग James Arthur • James Arthur (Deluxe)		
	Run-Around	
n My Bed	 Blues Traveler • Travelogue: Blues Traveler	
Revolver • My Revolver		
ke Off Your Dress	and	
fton Chenier • Bogalusa Boogie	∧ Desert You	1
ait	 Audrey Morgan	1
MBe • Wait		1
	Home Browse Search Radio You	1

RICK ASTLEY

Music Recommendation Systems

- Collaborative Filtering:
 - Exploits interactions between users and items
 - Identifies similarities between users
- Content Based Filtering:
 - Intrinsic Audio Features
 - Textual Metadata
- We will use the latter method of filtering for our model.

Introduction to Approaches

take .json files

mpd.slice.835000-835999.json
mpd.slice.836000-836999.json
mpd.slice.837000-837999.json
mpd.slice.838000-838999.json
mpd.slice.839000-839999.json
mpd.slice.840000-840999.json
mpd.slice.840000-840999.json



convert dataset into dataframe



find optimal k given data, cluster using k -means



given song input, create a playlist of similar songs

	Song Name	Artist Name	eu_distances
7258	Torn Apart	Judah & the Lion	30.535039
16000	I'm Amazed	My Morning Jacket	32.789714
8020	Rewind	Wingtip	24.220736
12847	Under You	Nick Jonas	34.717765
8623	Mary, Did You Know?	Pentatonix	22.541530
10824	What About Love	Austin Mahone	23.477687

Spotify Web API

duration_ms	int	The duration of the track in milliseconds.
key	int	The estimated overall key of the track. Integers map to pitches using standard Pitch Class notation . E.g. $0 = C$, $1 = C \sharp/D_b$, $2 = D$, and so on. If no key was detected, the value is -1.
mode	int	Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0.
time_signature	int	An estimated overall time signature of a track. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure).
		A confidence measure from 0.0 to 10 of whether the track is acoustic 1.0

A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic. The distribution of values for this feature look like this:

acousticness

float



- Spotify's Web API
- Spotipy
- Other interesting audio features it provides:
 - Danceability, energy, instrumentalness, liveness, loudness, speechiness, valence (positiveness), tempo (BPM)
 - Explicitness, Release Date of Song

More information about audio features: https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/

Spotify Million Playlist Dataset

- Enable research for music recommendation systems in 2018.
- Captures 1 million playlists of over 2 million unique tracks.
- Re-released September 2020 by AlCrowd

Dataset source: https://www.aicrowd.com/challenges/spotify-million-playlist-dataset-challenge

Dataset Statistics

- All playlists were generated by users from the United States:
 - Gender
 - 45% Male | 54% Female | 0.5% Unspecified | 0.5% Nonbinary
 - Age
 - **13-17: 10%**
 - 18-24: 43%
 - **25-34: 31%**
 - **35-44**: 9%
 - **45-54**: 4%
 - **55+:** 3%

Dataset: Playlist Criterion

- Playlists that meet the criteria are selected at random:
 - Contains at least <u>5</u> tracks and no more than <u>250</u> tracks
 - Contains at least <u>3 unique artists and at least</u> <u>2 unique albums</u>
 - Was created after January 1, 2010 and before December 1, 2017
 - Does not have an offensive title
 - Etc.
- Other modifications made to some playlists

Dataset Formatting

- Approx. 33 GB
- Stored as .json files
 - Approx. 1000 slices
 - 1000 playlists within each slice
- Huge Database

			20
😸 mpd.slice.835000-835999.json	32,992,511	5,671,529	JSON File
😸 mpd.slice.836000-836999.json	33,595,688	5,764,018	JSON File
🕷 mpd.slice.837000-837999.json	33,049,035	5,707,054	JSON File
😸 mpd.slice.838000-838999.json	34,928,190	6,006,748	JSON File
酸 mpd.slice.839000-839999.json	34,083,708	5,861,244	JSON File
酸 mpd.slice.840000-840999.json	34,143,923	5,922,166	JSON File
酸 mpd.slice.841000-841999.json	34,049,719	5,822,537	JSON File
😸 mpd.slice.842000-842999.json	34,189,274	5,914,169	JSON File
酸 mpd.slice.843000-843999.json	34,134,485	5,830,282	JSON File
酸 mpd.slice.844000-844999.json	33,041,376	5,672,722	JSON File
酸 mpd.slice.845000-845999.json	32,102,521	5,520,550	JSON File
酸 mpd.slice.846000-846999.json	33,588,870	5,824,659	JSON File
酸 mpd.slice.847000-847999.json	33,719,367	5,868,449	JSON File
😸 mpd.slice.848000-848999.json	32,718,774	5,649,143	JSON File
酸 mpd.slice.849000-849999.json	34,037,174	5,910,930	JSON File
😸 mpd.slice.850000-850999.json	35,894,609	6,208,639	JSON File
🕷 mpd.slice.851000-851999.json	32,932,954	5,657,991	JSON File
🕷 mpd.slice.852000-852999.json	31,667,557	5,486,026	JSON File
🕷 mpd.slice.853000-853999.json	34,413,134	5,926,061	JSON File
🕷 mpd.slice.854000-854999.json	32,984,590	5,709,862	JSON File
🕷 mpd.slice.855000-855999.json	34,045,739	5,864,183	JSON File
🚳 mpd.slice.856000-856999.json	33,537,070	5,815,300	JSON File
🕷 mpd.slice.857000-857999.json	33,157,538	5,634,929	JSON File
🚳 mpd.slice.858000-858999.json	32,432,861	5,695,758	JSON File
🕷 mpd.slice.859000-859999.json	34,041,795	5,862,611	JSON File
🚳 mpd.slice.860000-860999.json	33,993,906	5,856,889	JSON File
😸 mpd.slice.861000-861999.json	34,002,773	5,885,961	JSON File
😸 mpd.slice.862000-862999.json	34,150,688	5,896,727	JSON File
😸 mpd.slice.863000-863999.json	32,695,777	5,644,674	JSON File
🕷 mpd.slice.864000-864999.json	33,484,967	5,829,164	JSON File

Example of .json

- Playlist information provided:
 - Name of playlist
 - PID (unique identifier)
 - Collaborative
 - Last Modified
 - Etc.
- "URI" (song and artist identifier) used to query specific information from Spotify's API.

mpd	lice.835000-835999.json	
	"info": {	
	"generated_on": "2017-12-04 03:05:11.774401",	
	"slice": "835000-835999",	
	"version": "v1"	
	},	
	"playlists": [
	"name": "turn up music",	
	"collaborative": "false",	
	"pid": 835000,	
	"modified_at": 1508976000,	
	"num_tracks": 56,	
	"num_albums": 31,	
	"num_followers": 2,	
	"tracks": [
	"pos": 0,	
	"artist_name": "Lil Uzi Vert",	
	"track_uri": "spotify:track:5wyrQLBbmKdxY6yLNdienj",	
	"artist_uri": "spotify:artist:4015NlyKLIASxsJ0PrXPfz",	
	"track_name": "I Can Drive",	
	"album_uri": "spotify:album:6LX75kNicFqjjiAOeZgN67",	
	"duration_ms": 153808,	
	"album_name": "The Perfect LUV Tape"	
	},	
	"pos": 1,	
	"artist_name": "Young Thug",	
	"track_uri": "spotify:track:4cg1yakyRSI0jxKM2I7J1q",	
	"artist_uri": "spotify:artist:50co4Is1HCEo8bhOyUWKpn",	
	"track_name": "Digits",	
	"album_uri": "spotity:album:2z4c8M8aVz17CTobIp36KF",	
	"duration_ms": 176386,	
	"album_name": "Slime Season 3"	

Dataframe Example

	pos int64	artist_name object	track_uri object	artist_uri object	<pre>track_name object</pre>	album_uri obj
0	0	Bobby Darin	spotify:track:3E5ndy0f06	spotify:artist:0EodhzA6y	Mack The Knife	spotify:albur
1	1	Frank Sinatra	spotify:track:2y8Eez5cFF	spotify:artist:1Mxqyy3pS	Fly Me To The Moon	spotify:albur
2	2	Etta James	spotify:track:0CmIALzGn4	spotify:artist:0i0VhN3tn	At Last - Single Version	spotify:albur
3	3	Ella Fitzgerald	spotify:track:3Bbbz0IGOR	spotify:artist:5V0MlUE1B	Dream A Little Dream Of	spotify:albur
4	4	The Chordettes	spotify:track:433nBELv00	spotify:artist:62GnBjssW…	Mister sandman	spotify:albur
		danceability floa	t64 energy float64 k	ey int64 loudness float64	mode int64 speechiness	float64 acousticness
	e.,	0.5400000000000000000000000000000000000	0 500 0	10.001	0 0 100	0.76

5y	0.548999999999999999	0.529	3	-12.291	0	0.108	0.76
oS	0.657	0.26	7	-14.251	1	0.0521	0.474
tn	0.168	0.326	5	-9.981	1	0.0324	0.721
1B	0.631	0.127	0	-15.333	1	0.0949	0.939
sW	0.616	0.298	9	-13.191	1	0.052000000000000005	0.841

Dataframe Example (2)

	pos int64	artist_name object	track_uri object	artist_uri object	<pre>track_name object</pre>	album_uri obj
0	0	Bobby Darin	spotify:track:3E5ndy0f06	spotify:artist:0EodhzA6y	Mack The Knife	spotify:albur
1	1	Frank Sinatra	spotify:track:2y8Eez5cFF	spotify:artist:1Mxqyy3pS	Fly Me To The Moon	spotify:albur
2	2	Etta James	spotify:track:0CmIALzGn4	spotify:artist:0i0VhN3tn	At Last - Single Version	spotify:albur
3	3	Ella Fitzgerald	spotify:track:3Bbbz0IGOR	spotify:artist:5V0MlUE1B	Dream A Little Dream Of …	spotify:albur
4	4	The Chordettes	spotify:track:433nBELv00	spotify:artist:62GnBjssW	Mister sandman	spotify:albur



Artist Name	Song Name	genres	guadalajara indie	triangle indie
Bobby Darin	Mack The Knife	['adult standards', 'brill building pop', 'eas	0	0
Frank Sinatra	Fly Me To The Moon	['adult standards', 'easy listening', 'lounge']	0	0
Etta James	At Last - Single Version	['adult standards', 'jazz blues', 'soul', 'sou	0	0
Ella Fitzgerald	Dream A Little Dream Of Me	['adult standards', 'jazz blues', 'swing', 'vo	0	0
The Chordettes	Mister sandman	['adult standards', 'brill building pop', 'cla	0	0

Adding Artist Genres

- Approximately 5,071 distinct genres to describe artists in Spotify
- Echonest applies ML on audio features to classify genres
- Classify genre of song based on genre of artist (has limitations!)

'old school rap francais', 'latintronica', 'new england emo', 'spanish punk', 'geek rock', 'canadian blues', 'barbershop', 'anglican liturgy', 'melbourne indie'. 'sega mauricien', 'sacramento hip hop', 'nintendocore'. 'jump up', 'benga', 'deep filthstep', 'alternative hip hop', 'bluegrass gospel', 'huapango', 'western americana'. 'jazz violin', 'baroque', 'glam rock', 'italian hip hop', 'latin tech house'. 'syrian pop', 'pittsburgh metal', 'family gospel', 'meditation', 'barnmusik', 'danish alternative rock'. 'swedish electronic'.

'shoegaze', 'virginia metal', 'folk-pop', 'american metalcore', 'norsk lovsang', 'progressive deathcore', 'african electronic'. 'nz metal'. 'pop electronico', 'ska punk', 'bubblegrunge', 'black metal', 'baltimore indie', 'rock drums', 'wonky', 'deep idm', 'la indie', 'italian disco', 'rock potiguar', 'psychedelic trance', 'korean ost'. 'garage rock', 'british power metal', 'show tunes'. 'indie punk', 'operatic pop', 'pinoy trap', 'jam band', 'post-hardcore', 'free folk', 'vintage western',

Adding Artist Genres (everynoise.com)

Mechanical instrumentation (techno music)



jangle rock dutch rock dansebund funeral doorn oi german of ack thrash christian hardenne. irish nock nost-nunk

atmospheric

bouncy/upbeat

Organic instrumentation (classical)

T-SNE Visualization



Instrumental/ More Classical

Radio Edited -Britney Spears -Drake

The Playlist Recommendation System

K-Means Clustering

Elbow Method



• • • def find_best_k(df): $max_diff = -9999999$ $best_k = dict()$ avg_sum_squares = [] for k in range(1, 10): kmeans = KMeans(n clusters = k, random state = 42) kmeans.fit(df) avg_sum_squares += [kmeans.inertia_] *if* k == 1: prev_inertia = kmeans.inertia_ else: slope = prev_inertia - kmeans.inertia_ prev_inertia = kmeans.inertia_ max_diff = max(max_diff, slope) if max_diff <= slope:</pre> best k[max diff] = k k = best_k[max_diff] return k

- Input = Song Name, Artist Name, Spotify DataFrame
- Randomized selection of 30000 rows from csv due to hardware limitations



- Train the KMeans model on the given DataFrame
- Append cluster labels to the DataFrame for each row using kmeans.labels_

Artist Name	 appalachian folk	blues	funky breaks	rock nacional	canzone napoletana	jazz blues	pagode baiano	indie folk argentino	cumbia ranchera	Cluster Labels
Frank Sinatra	 0	0	0	0	0	0	0	0	0	0
Louis Armstrong	 0	0	0	0	0	0	0	0	0	0
Nina Simone	 0	0	0	0	0	1	0	0	0	0
The Marvelettes	 0	0	0	0	0	0	0	0	0	0
James Brown & The Famous Flames	 0	0	0	0	0	0	0	0	0	0

Preprocessing for the input.

	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	 big band	appalachian folk
0	0.618	0.741	2	-3.95	1	0.031	0.17	0	0.0822	0.296	 0	0



- Filter out dataframe by cluster label
- Calculate euclidean distances and sort based on minimum distance
- Pick top n songs (n = 2 for example)

	Song Name	Artist Name	eu_distances
16835	My Love (feat. Drake) - Remix	Majid Jordan	22.389118
11010	Give My Life - Radio Edit	Army Of Lovers	27.922494
	•		

Building a Playlist: n -songs Input

Inputs:

- List of Lists → [[Song Name, Artist Name], [Song Name, Artist Name],...]
- Spotify DataFrame
- Number of Songs to Recommend / Song



Dancing Queen 🔿 ABBA





The Example

•••

song_list = [["Dancing Queen", "ABBA"], ["Just for a Moment", "Olivia Rodrigo"], ["Burning Up", "BTS"]]

A8BA/ARRIVAL



Dancing Queen 🛇 ABBA





Output



Dancing Queen 🛇

	Song Name	Artist Name	eu_distances
28890	Love You Better	Dominique	3.100977
24554	Turn Your Love Around	George Benson	7.000762
16983	What's Luv? (feat. Ashanti) - Clean Version fe	Fat Joe	7.833364
14408	Is This Love	Chris Brown	13.203771
8828	4 Me	Maleek Berry	19.493569
6469	Live Life	Collie Buddz	20.350899
9733	Buried In Detroit - Lucas Löwe Remix	Mike Posner	8.407508
11925	Last First Kiss	One Direction	19.569077
21799	Do You Wanna Come Over?	Britney Spears	23.599895

Output



	Song Name	Artist Name	eu_distances
28890	Love You Better	Dominique	3.100977
24554	Turn Your Love Around	George Benson	7.000762
16983	What's Luv? (feat. Ashanti) - Clean Version fe	Fat Joe	7.833364
14408	Is This Love	Chris Brown	13.203771
8828	4 Me	Maleek Berry	19.493569
6469	Live Life	Collie Buddz	20.350899
9733	Buried In Detroit - Lucas Löwe Remix	Mike Posner	8.407508
11925	Last First Kiss	One Direction	19.569077
21799	Do You Wanna Come Over?	Britney Spears	23.599895

Output



	Song Name	Artist Name	eu_distances
28890	Love You Better	Dominique	3.100977
24554	Turn Your Love Around	George Benson	7.000762
16983	What's Luv? (feat. Ashanti) - Clean Version fe	Fat Joe	7.833364
14408	Is This Love	Chris Brown	13.203771
8828	4 Me	Maleek Berry	19.493569
6469	Live Life	Collie Buddz	20.350899
9733	Buried In Detroit - Lucas Löwe Remix	Mike Posner	8.407508
11925	Last First Kiss	One Direction	19.569077
21799	Do You Wanna Come Over?	Britney Spears	23.599895

Limitations of Dataset

- Only used 4 slices (500MB) of .json files
- Limited to only a small amount of songs
- Playlist criterion also creates a biased sample.
- Unable to represent all genres
- Smaller in comparison to larger recommendation systems like:



Recommended Songs A Based on the songs in this playlist					
ADD	BLEACH		BROCKHAMPTON	SATURATION III	4:33
ADD	GOLD		BROCKHAMPTON	SATURATION	4:26
ADD	RENTAL		BROCKHAMPTON	SATURATION III	3:34
ADD	See You Again (feat. Kali Uchis)		Tyler, The Creator, Kali Uchis	Flower Boy	3:00
ADD	FACE		BROCKHAMPTON	SATURATION	4:19
ADD	Nights		Frank Ocean	Blonde	5:07
ADD	Boredom (feat. Rex Orange County		Tyler, The Creator, Rex Orange	Flower Boy	
ADD	90210 (feat. Kacy Hill)	EXPLICIT	Travis Scott, Kacy Hill	Rodeo (Expanded Edition)	5:39

Conclusion

- Real-World Implications
 - Ability to create a playlist off of different user preferences



Next Steps: Ready for the Real-World?

- Ready for the Real -World?
 - \circ Scale Up
 - Playlist metadata
 - Ensemble of Collaborative Filtering with Content Based techniques

	Song Name	Artist Name	eu_distances
8770	RIP	Spice 1	13.393440
10024	Fool's Gold - B Side	Amy Winehouse	13.618472
12000	Love Me Better	James Blunt	14.597757
26138	Sweetest Thing	Allman Brown	18.666827
10887	You Are My Starship - Remastered	Norman Connors	15.588317
5067	The Ocean	Led Zeppelin	22.269343
12220	DOPE	T.I.	27.556257
2127	Miss Mysterious	Set It Off	28.335754
14338	Borders	St. Beauty	34.425192
19357	Jinsei No Merry Go Round	Yuki Niino	46.009973
18985	I Wanna Go Down With You	Russ	46.179388
18601	Demons	Sleigh Bells	48.748783

Thank You!

