"Music happens to be an art form that transcends languages, age, gender, culture, stereotypes and prejudice."

ANALYSIS AND RECOMMENDER SYSTEM BASED ON YOUR MUSIC LISTENING HABITS.

Introduction

Overview

Music is inherently something that binds people together. Its common practice for someone to ask new friends' what kind of music they listen to as one of the ice breakers questions. Just talking about your favorite songs, the lyrics, the meaning behind the themes and how you personally relate to what the artist is singing about can lead to you forming a deeper relationship with someone.

Personally, I think music is a big part of who I am as a person. The artists you decide to follow and the music you listen to can really change you as a person. I sometimes wonder what kind of a person I would be if I never found my favorite artist. So, for this project I wanted to focus on building an interface that would find and recommend you the most similar songs to your picked song as well as have a way to compare and contrast songs based on artists.

Design Outline

Data Sets

We found two datasets on Kaggle which we utilized for our analysis. I did the initial analysis using the first dataset which is the "Spotify Top Chart Songs 2022". This dataset provides a list of all songs that have appeared on the weekly top chart where the data has been collected from Spotify's top global charts which are publicly available on their website. I used this one because I was already familiar with the data and I wanted to test the algorithm using a smaller dataset. The second dataset I used was a much larger dataset. It had 54000 songs from all around the world which I will be using for the final analysis. I did some initial cleaning on the data in Excel. I took out any duplicating cells as well as removed rows with any missing and/or incorrect values. There were quite a lot of rows where the song data was not parsed properly when pulling from the API so I decided to remove these songs as well. For this project I decided to only use the songs which has a popularity index of greater than 70 (out of 100). In the future, I would add a toggle function where the user can decide if they only want to look at and analyze popular songs or all songs.

These datasets include attribute information that is needed for further analysis, from the basic song attributes like track name and artist name, as well as specific information about the songs danceability, tempo, instrumentalness, etc.

Index	ID	Title	Artist	Release_Dat	Year	Duration_ms	Danceability	Energy	Tempo	Loudness	Liveness	Speechiness	Acousticness	Instrumental	Valence	Key	Popularity
1	41SwdQIX8H	10%	['KAYTRANA	12/13/2019	2019	186680	0.794	0.757	107.99	-6.644	0.0621	0.123	0.0267	0.000306	0.615	6	73
2	5HsSsEeP3h	3:00 AM	['Finding Ho	10/27/2017	2017	201951	0.55	0.147	81.991	-18.675	0.286	0.0404	0.563	0.00102	0.0346	0	72
3	2gNMXUDKR	3:59 AM	['DIVINE']	12/3/2020	2020	272197	0.669	0.588	97.026	-6.902	0.179	0.339	0.177	0	0.237	2	70
4	1nX9KhK3Fff	4:00 AM	['2 Chainz', "	6/16/2017	2017	255560	0.796	0.5	75.012	-7.21	0.155	0.425	0.118	0	0.227	1	70
5	7KbF6AdprO	11:00 PM	['Maluma']	5/17/2019	2019	175733	0.777	0.712	95.692	-4.84	0.091	0.277	0.217	0	0.68	10	79
6	3ipliatTDndC	7	['Ufo361', 'B	1/1/2021	2021	148915	0.828	0.697	167.997	-7.962	0.106	0.252	0.0147	0.0264	0.629	2	71
7	6X29iaaazw	8	['Billie Eilish	3/29/2019	2019	173202	0.735	0.235	62.446	-13.239	0.107	0.315	0.698	0.00604	0.462	3	74
8	034KyQEk7ih	17	['Pink Sweat	2/28/2020	2020	160737	0.558	0.364	129.17	-8.663	0.109	0.0502	0.169	9.94E-05	0.252	5	71
9	21hbZ0ylIYO	18	['Anarbor']	6/4/2013	2013	157360	0.724	0.854	98.056	-2.779	0.0805	0.0997	0.0603	0	0.88	11	73
10	3JjnGLK8lxkN	18	['One Directi	11/17/2014	2014	248360	0.681	0.582	124.038	-7.76	0.119	0.0313	0.22	0	0.221	6	81
11	3JKgcAa7r07	21	['Polo G']	5/15/2020	2020	163798	0.783	0.479	176.027	-10.646	0.101	0.541	0.149	0	0.533	7	84
12	6ce41Va6KF	22	['Taylor Swif	10/22/2012	2012	232120	0.661	0.729	103.987	-6.561	0.0477	0.0376	0.00215	0.0013	0.668	7	73
13	20FZrLI9z5b	23	['Mike WiLL	1/1/2013	2013	252520	0.859	0.599	140.005	-5.254	0.13	0.0614	0.000362	8.20E-06	0.319	9	71
14	0NipDzamAt	44	['Bad Gyal',	3/18/2021	2021	185467	0.801	0.76	102.061	-5.948	0.207	0.107	0.252	9.17E-06	0.508	5	74
15	2sqgNScWya	100	['Anuel AA',	1/22/2021	2021	215293	0.769	0.455	95.097	-5.698	0.251	0.155	0.554	2.07E-06	0.293	10	73
16	3KJ3opyV292	120	['Bad Bunny	11/27/2020	2020	151506	0.636	0.656	171.918	-5.542	0.287	0.0676	0.451	7.97E-06	0.439	2	76
17	16EMONI2v	212	['Azealia Bar	11/7/2014	2014	204957	0.846	0.772	125.998	-5.735	0.0769	0.259	0.0123	0.000106	0.622	11	71
18	2XWjPtKd65s	365	['Zedd', 'Katy	2/14/2019	2019	181900	0.745	0.693	98.039	-5.479	0.123	0.034	0.339	0	0.248	2	70
19	0Fl4PBflwC5	501	['Bilal Wahit	2/25/2021	2021	127592	0.613	0.796	186.102	-3.672	0.292	0.203	0.162	0	0.749	8	77
20	0BxE4FqsDD	505	['Arctic Monl	4/22/2007	2007	253587	0.526	0.866	140.266	-5.822	0.0945	0.0568	0.00287	7.78E-05	0.248	0	81
21	58ge6dfP91c	505	['Arctic Monl	4/24/2007	2007	253587	0.52	0.852	140.267	-5.866	0.0733	0.0543	0.00237	5.79E-05	0.234	0	78
22	5jzEwSyyym	512	['Mora', 'Jha	2/5/2021	2021	193636	0.818	0.673	87.961	-7.542	0.172	0.148	0.348	0	0.695	4	78
23	0sKEilrw1Grl	865	['Morgan Wa	1/8/2021	2021	190680	0.532	0.584	88.003	-5.254	0.101	0.0248	0.723	0	0.367	8	76
24	2chIkORxBLp	911	['Damso']	9/17/2020	2020	172293	0.685	0.706	155.989	-7.02	0.0986	0.0545	0.0894	0.0287	0.64	11	72
25	2TksvaRivgA	911	['Sech']	2/5/2021	2021	215891	0.795	0.655	93.029	-3.815	0.296	0.0422	0.076	1.76E-06	0.744	7	91
26	3bHZeMMos	911	['Ellise']	10/12/2018	2018	190000	0.698	0.611	130.098	-6.754	0.0587	0.0686	0.571	5.30E-06	0.507	10	71
27	6qI0MU175C	911	['Lady Gaga'	5/29/2020	2020	172133	0.789	0.813	116.007	-4.563	0.519	0.0374	0.0294	2.07E-06	0.42	5	74
28	3R9KvtovDyP	1942	['Ant Wan']	2/25/2021	2021	185263	0.72	0.47	85.496	-6.922	0.138	0.224	0.626	3.70E-05	0.208	10	70
29	0CZ8IquoTX2	1950	['King Prince	2/23/2018	2018	225133	0.6	0.535	71.912	-7.043	0.162	0.107	0.648	4.70E-06	0.269	4	78
30	5oQcOu1om	1985	['Bowling Fo	6/26/1905	1905	193000	0.606	0.887	119.974	-4.404	0.337	0.0598	0.000242	0	0.903	4	74
31	2BgEsaKNfH	2002	['Anne-Marie	4/27/2018	2018	186987	0.697	0.683	96.133	-2.881	0.137	0.117	0.0372	0	0.603	1	84
32	2u8AXJyk8pl	2002	['Sido', 'Apac	9/26/2019	2019	237157	0.657	0.691	160.106	-7.907	0.099	0.236	0.213	0	0.493	8	70
33	4Pbg79cTBu	2002	['Anne-Marie	4/19/2018	2018	186987	0.697	0.683	96.133	-2.881	0.137	0.117	0.0372	0	0.603	1	71
34	6dFn6my1sH	2009	['Mac Miller'	8/3/2018	2018	347987	0.533	0.496	155.158	-9.363	0.0969	0.517	0.77	0.000164	0.173	6	71
35	0YETkmonJ7	2019	['Martin Jens	2/26/2021	2021	131381	0.876	0.765	115.072	-6.554	0.0826	0.269	0.119	5.20E-06	0.729	9	71
36	2ZItjIqztEpZt	3005	['Childish Ga	12/10/2013	2013	234215	0.472	0.448	166.077	-7.272	0.0914	0.326	0.113	0	0.636	6	78
37	02NM6vucOvn	11-Jul	['BeyoncV©'	11/24/2014	2014	213507	0.747	0.705	136.024	-5.137	0.126	0.126	0.0128	0	0.56	9	72
38	30QR0ndUdi	...And to Those I Love, Thanks for !	['Suicideboy	2/14/2020	2020	168490	0.792	0.511	113.983	-6.876	0.14	0.0409	0.124	9.04E-05	0.111	2	80
39	3MjUtNVVq	...Baby One More Time	['Britney Spe	1/12/1999	1999	211067	0.759	0.699	92.96	-5.745	0.443	0.0307	0.202	0.000131	0.907	0	80
40	2yLa0QULdQ	...Ready For It?	['Taylor Swif	11/10/2017	2017	208187	0.613	0.764	160.015	-6.509	0.197	0.136	0.0527	0	0.417	2	73
41	4xkOaSrkexN	'Till I Collapse	['Eminem', 'I	5/26/2002	2002	297787	0.548	0.847	171.447	-3.237	0.0816	0.186	0.0622	0	0.1	1	85
42	5QTxFnGygV	(Don't Fear) The Reaper	['Blue Vñyste	5/29/1905	1905	308120	0.333	0.927	141.466	-8.55	0.297	0.0733	0.0029	0.000208	0.385	9	77
43	1Eb90nmqTr	(Everything I Do) I Do It For You	['Bryan Adan	9/24/1991	1991	393640	0.533	0.36	131.229	-12.633	0.0631	0.0282	0.0805	1.03E-05	0.299	1	77
44	2PzU4IB8DrE	(I Can't Get No) Satisfaction - Mo	['The Rolling	7/30/1965	1965	222813	0.723	0.863	136.302	-7.89	0.128	0.0338	0.0383	0.0317	0.931	2	78
45	7ojJ4XvqBhB	(I Can't Help) Falling In Love With	['UB40']	1/1/1993	1993	207440	0.642	0.722	172.406	-13.031	0.122	0.0375	0.0444	0.00388	0.836	2	70
46	4ByEFOBuLX	(I Just) Died In Your Arms	['Cutting Cre	1/18/1986	1986	280400	0.625	0.726	124.945	-11.402	0.0625	0.0444	0.0158	0.000169	0.507	11	74
47	4BFMQ15vXi	(I've Had) The Time of My Life - Fi	['Bill Medley	7/2/2004	2004	289067	0.7	0.781	108.695	-8.255	0.059	0.0507	0.0874	0.00342	0.385	6	70
48	3zBhihYUHB	(Sittin' On) the Dock of the Bay	['Otis Reddir	5/21/1905	1905	163756	0.769	0.367	103.621	-11.226	0.081	0.0312	0.684	1.62E-05	0.535	2	76
49	2g2GkH3vZH	(What A) Wonderful World	['Sam Cooke	9/26/2000	2000	125440	0.691	0.507	128.518	-10.23	0.457	0.0343	0.627	0	0.885	11	73
50	5K09WxKdfS	(You Make Me Feel Like) A Natura	['Aretha Fran	1/22/1968	1968	165333	0.603	0.271	110.89	-10.083	0.141	0.0284	0.679	0	0.405	5	70
51	4TBBPZks71	(Your Love Keeps Lifting Me) High	['Jackie Wils	6/28/1905	1905	181067	0.631	0.69	94.574	-6.676	0.121	0.0531	0.176	0	0.938	2	70
52	5qyq1H5OPk	(Your Love Keeps Lifting Me) High	['Jackie Wils	5/20/1905	1905	179120	0.634	0.677	95.792	-7.19	0.0902	0.0448	0.318	0	0.939	2	70
53	7JgZbix8xDa	[oops]	['potsu']	5/17/2017	2017	142183	0.843	0.256	68.997	-13.565	0.133	0.0752	0.538	0.114	0.655	7	70
54	01TuObJVd7	#thatPOWER	['will.i.am',	1/1/2013	2013	279507	0.797	0.608	127.999	-6.096	0.0748	0.0584	0.00012	7.66E-05	0.402	6	70
55	5Cbo7oz78gc	+	['Aitana', 'Ca	12/17/2019	2019	219367	0.561	0.666	160.019	-4.243	0.123	0.0322	0.106	0	0.675	8	70
56	4MZSqgwIE7	<3	['Bad Bunny'	2/28/2020	2020	157967	0.631	0.665	80.353	-7.589	0.152	0.229	0.602	1.70E-06	0.634	10	72
57	4jtyUzZm9W	0 To 100 / The Catch Up	['Drake']	7/14/2013	2013	275227	0.559	0.716	176.618	-3.964	0.252	0.39	0.737	0	0.504	11	73
58	4AlihYDqxXsl	00:00 (Zero O,Âô Clock)	['BTS']	2/21/2020	2020	250296	0.566	0.548	157.999	-6.456	0.11	0.0579	0.0802	0	0.196	1	74
59	5tz669p7tJuG	1-800-273-8255	['Logic', 'Ale	5/5/2017	2017	250173	0.62	0.574	100.023	-7.788	0.19	0.0479	0.569	0	0.357	5	79

spotify-2000s-data | radar | Sheet1 | +

Ready — Accessibility: Investigate — 100%

We decided to focus on the following attributes:
1. **Track Name**
2. **Artist Name**: The artist or list of artists who made the song.
3. **Release Date**: The date the song was released.
4. **Duration (ms)**: The length of the song.
5. **Danceability**: 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.
6. **Energy**: 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. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.
7. **Tempo**: 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.
8. **Loudness**: 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.
9. **Liveness**: 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.

10. **Speechiness**: Speechiness detects the presence of spoken words in a track. 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.
11. **Acousticness**: A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
12. **Instrumentalness**: Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Values above 0.5 are intended to represent instrumental tracks.
13. **Valence:** The higher the value, the more positive mood for the song.
14. **Key**: The key the track is in. Integers map to pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C♯/D♭, 2 = D, etc. If no key was detected, the value is -1.

Developmental Tools

For the algorithm to find similarity between the songs I used Python to write and test the code. Even some of the static images are built using the matplotlib in Python. After finding the list of sings that is the most similar, I used Excel to make the radar charts.

I wrote a program that computes the similarity scores between one selected song and all the other songs to find the songs that are the most similar using its attributes. Also creating a cluster of similar songs with a fixed value of K.

Input: Spotify 2000's MegaSet Data - spotify-data.csv
Output:
1. List of songs similar to a selected song.
2. Cluster of similar songs using K Means Clustering.

I made sure to write plenty of comments so that if you want to play around with the data it should be easy to understand.

Link to Github: https://github.com/ishamahadalkar/spotifyvis

Here are some of visuals I mapped out:

Click on each data point (each song) to show the song features using the radar charts.
↳ it opens them under the chart.

Song 1   Song 2
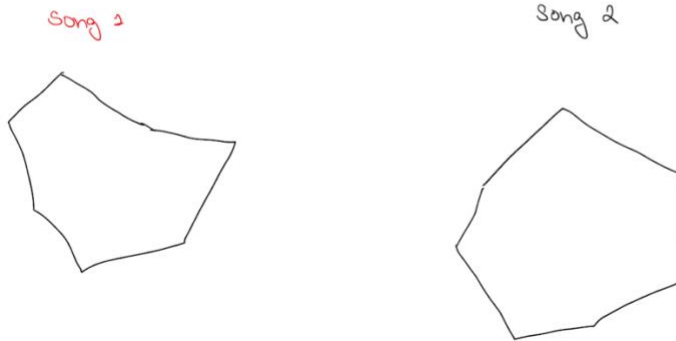
⇒ Find songs that similar to each other

① link them on the chart.

Artist 1
Artist 2

#Streams
↓
popularity

highlight

⇒ comparing the song features to find the closest match.

Year
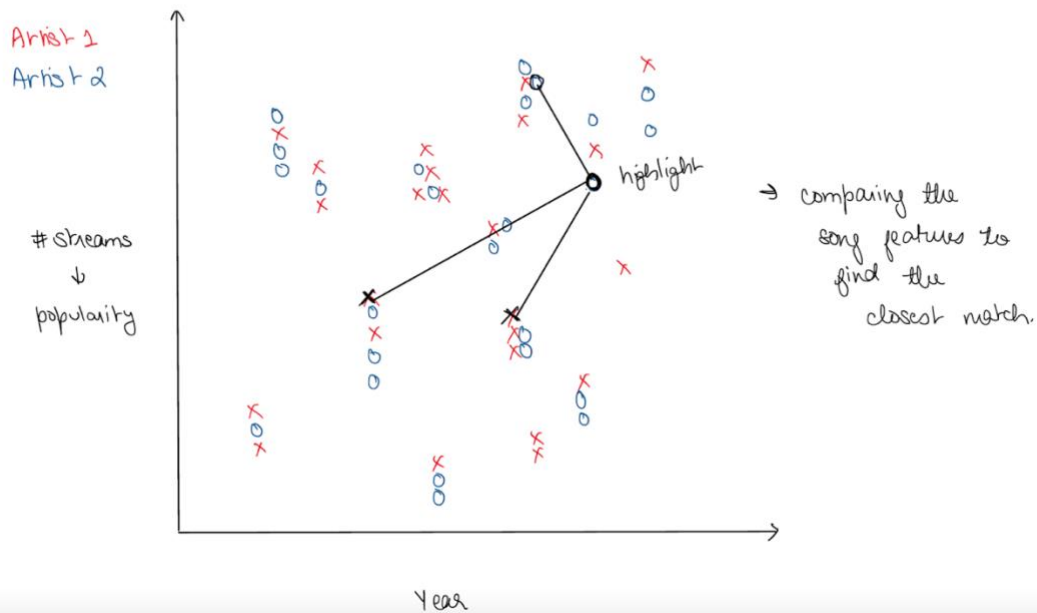
Future Work

Currently since it is a work in progress, I only have static images but ideally, I would make it into an interactive application using Processing and D3 where the user can click on each song data point and see the radar chart for that song comparing it with radar charts of different songs.

Even for the scatterplot I would have it so that when a user clicks on one song it shows the information about the songs and then has links to the similar songs almost like a network diagram or a map while, ideally, playing the song in the background. I would also like to have a filter on Artists so the users can compare their favorite artists.

Results

Algorithm analysis:

I chose 3 songs to analyze from different genres for some popular artists:

1. "Dynamite" by BTS

Similar Songs:

Dynamite	['BTS']	11/20/2020
Best Of Me	['BTS']	8/24/2018
Cum on Feel the Noize	['Quiet Riot']	6/5/1905
Den fineste Chevy'n (feat. Maria Mena)	['Halva Priset', 'Maria Mena']	2/26/2021
Hurt You	['The Weeknd', 'Gesaffelstein']	3/30/2018
Intro : Persona	['BTS']	4/12/2019
Madrid (feat. Myke Towers)	['Maluma', 'Myke Towers']	8/21/2020
Slide (feat. Frank Ocean & Migos)	['Calvin Harris', 'Frank Ocean', 'Migos', 'Funk Wav']	6/30/2017
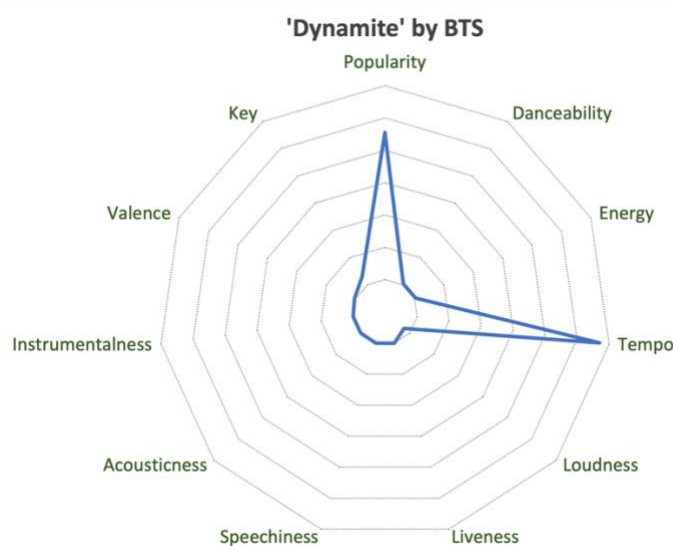
7 Most Similar Songs to 'Dynamite' by BTS

Legend:
- Best Of Me
- Cum on Feel the Noize
- Den fineste Chevy'n (feat. Maria Mena)
- Hurt You
- Intro : Persona
- Madrid (feat. Myke Towers)
- Slide (feat. Frank Ocean & Migos)

Axes: Popularity, Danceability, Energy, Tempo, Loudness, Liveness, Speechiness, Acousticness, Instrumentalness, Valence, Key
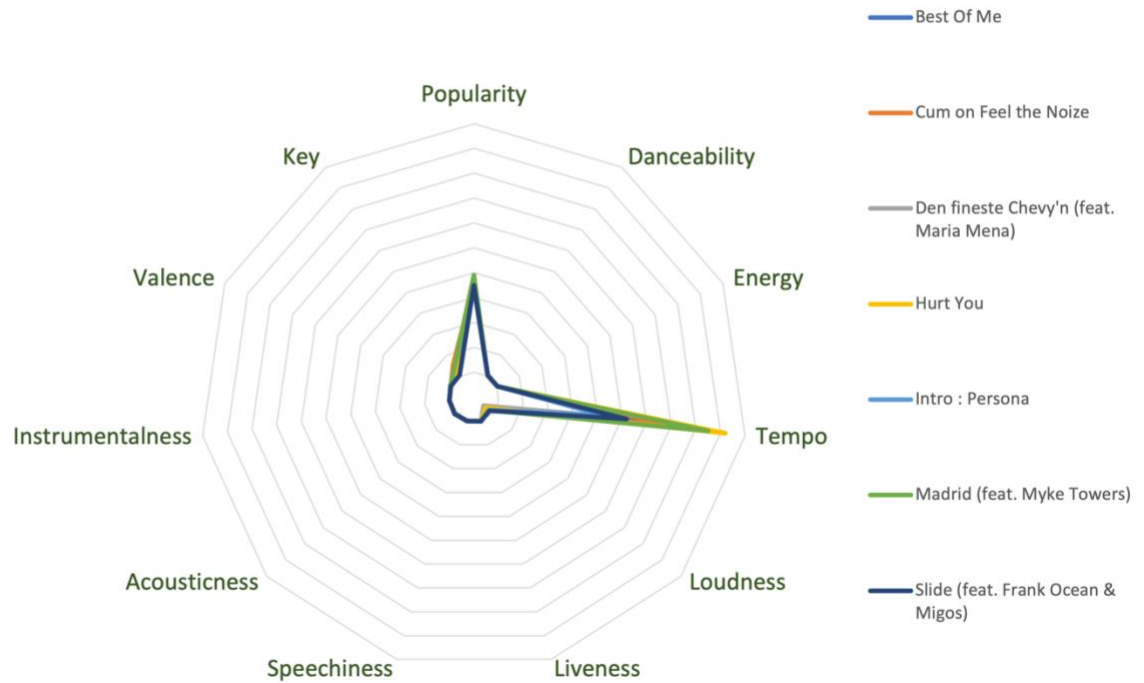
2. 'Money Trees' by Kendrick Lamar ft. Jay Rock

Similar Songs:

Money Trees	['Kendrick Lamar', 'Jay Rock']	10/22/2012
Blue Tacoma	['Russell Dickerson']	6/16/2017
El Borracho	['Los Dos Carnales']	11/20/2020
GOOD TIME	['Niko Moon']	2/14/2020
It's a Long Way to the Top (If You Wanna Rock 'N' Roll)	['AC/DC']	5/14/1976
Let Her Go	['Passenger']	7/4/1905
Money Talks (feat. Dave)	['Fredo', 'Dave']	1/28/2021
Passin' Me By	['The Pharcyde']	11/24/1992

'Money Trees' by Kendrick Lamar ft. Jay Rock
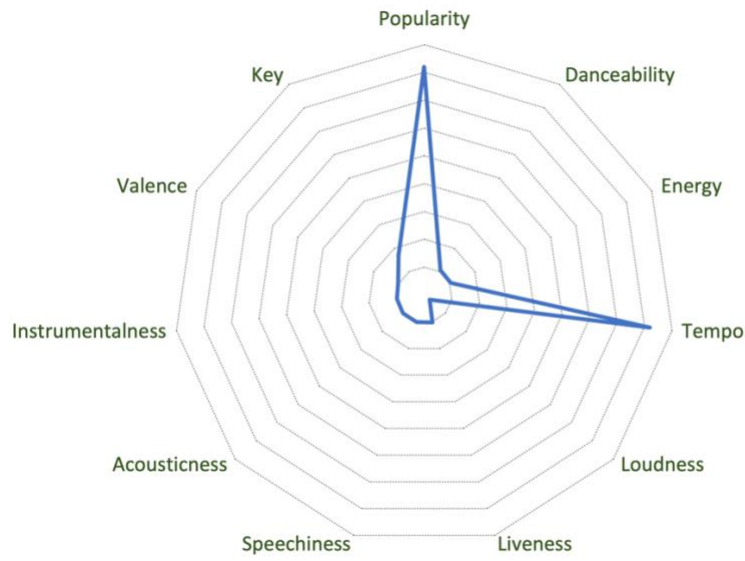
7 Most Similar Songs to 'Money Trees' by Kendrick Lamar ft. Jay Rock

- Blue Tacoma
- El Borracho
- GOOD TIME
- It's a Long Way to the Top (If You Wanna Rock 'N' Roll)
- Let Her Go
- Money Talks (feat. Dave)
- Passin' Me By
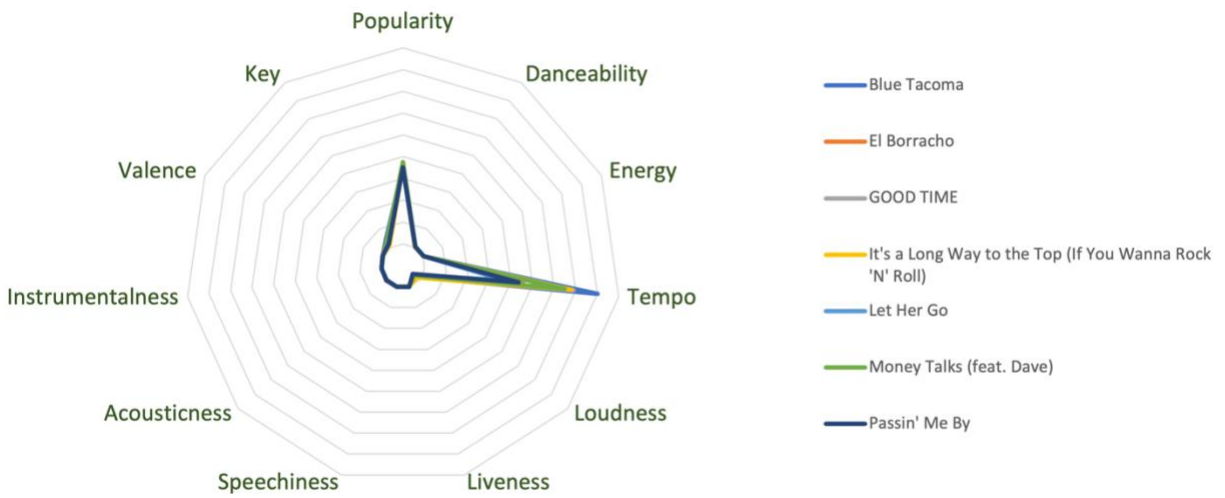
3. 'Dancing Queen' by ABBA
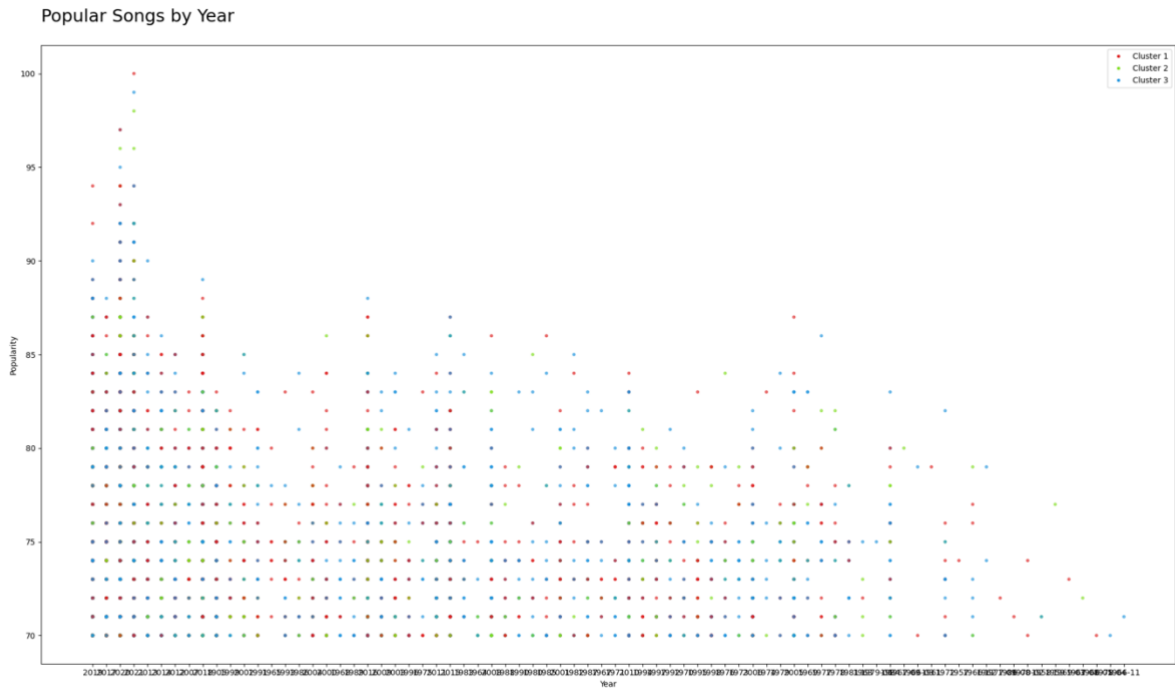
Similar Songs:

Dancing Queen	['ABBA']	5/29/1905
2 Hearts (feat. Gia Koka)	['Sam Feldt', 'Sigma', 'Gia Koka']	1/10/2020
Back to You (feat. Bebe Rexha & Digital Farm Animals)	['Louis Tomlinson', 'Bebe Rexha', 'Digital Farm Animals']	7/21/2017
Goosebumps - Remix	['Travis Scott', 'HVME']	1/15/2021
In the End	['Linkin Park']	10/24/2000
Kaikai Kitan	['Eve']	12/23/2020
On My Mind (Jorja Smith X Preditah)	['Jorja Smith', 'Preditah']	8/25/2017
Dancing Queen	['ABBA']	5/29/1905

'Dancing Queen' by ABBA

7 Most Similar Songs to 'Dancing Queen' by ABBA

Scatterplots:
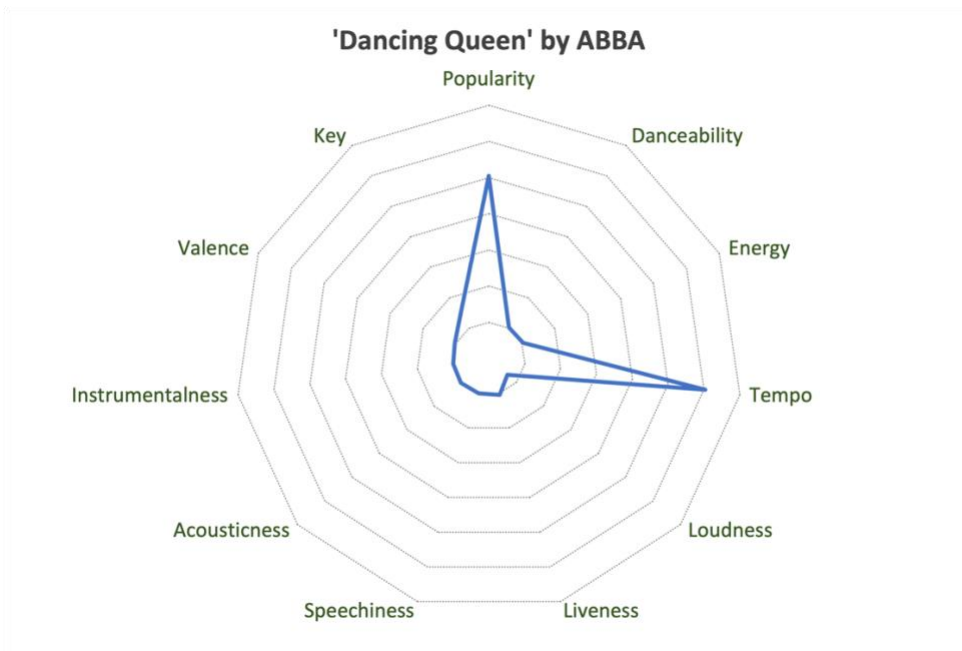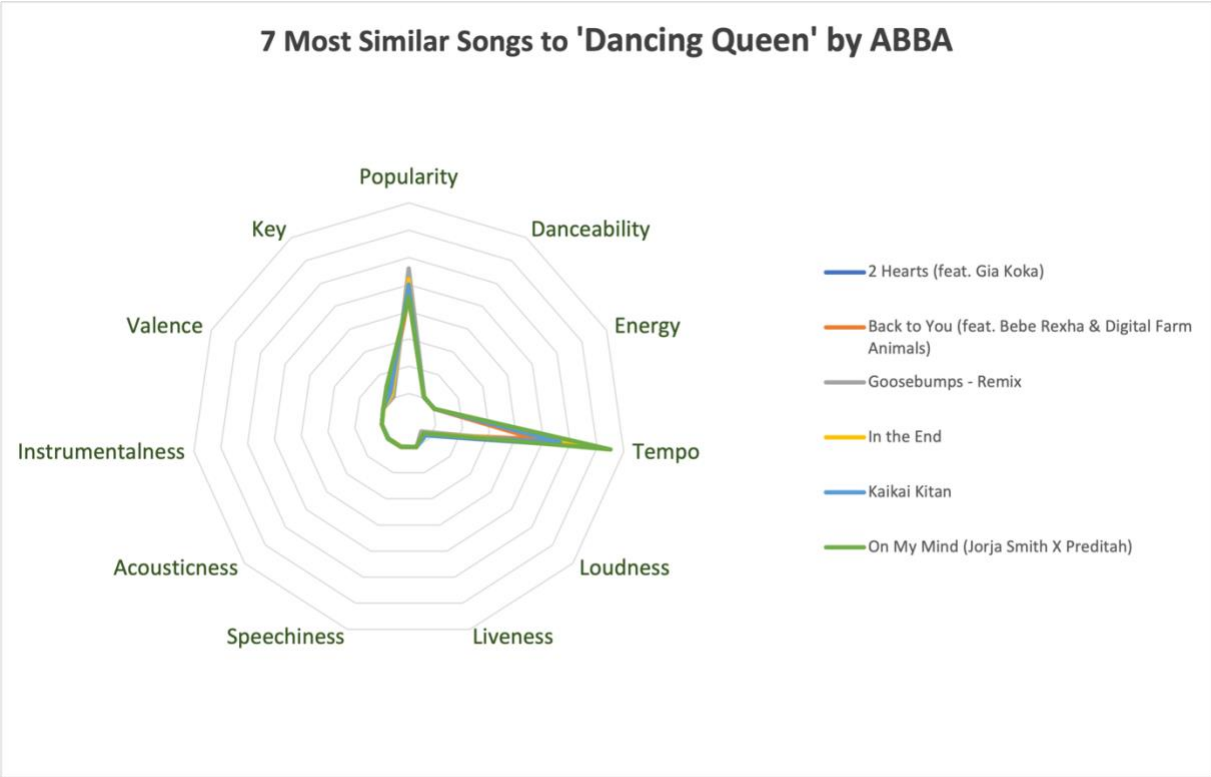
Popular Songs by Year

Note: The clusters will look "grouped" together on a 11-dimensional axis. In the interactive visual I would create 2D clusters for each individual attribute.

Popular Songs by Danceability