

MOOSIC STARTUP PROJECT



Spotify Playlist

MOOSIC PROJECT

WE TRY TO ANSWER MAIN QUESTIONS:

Are Spotify's audio features able to identify "similar songs", as defined by humanly detectable criteria and are these similarities detectable using the audio features from Spotify? —

Is K-Means a good method to create playlists or there is a need to explore other methods to create playlists?

Step we followed & techniques we used to create playlist:

Interia and elbow method

The silhouette score

Exploring data with clusters

Creating radar chart

PCA technique



Data cleaning and preparation

Dropped features:

duration_ms
time_signature
key
type



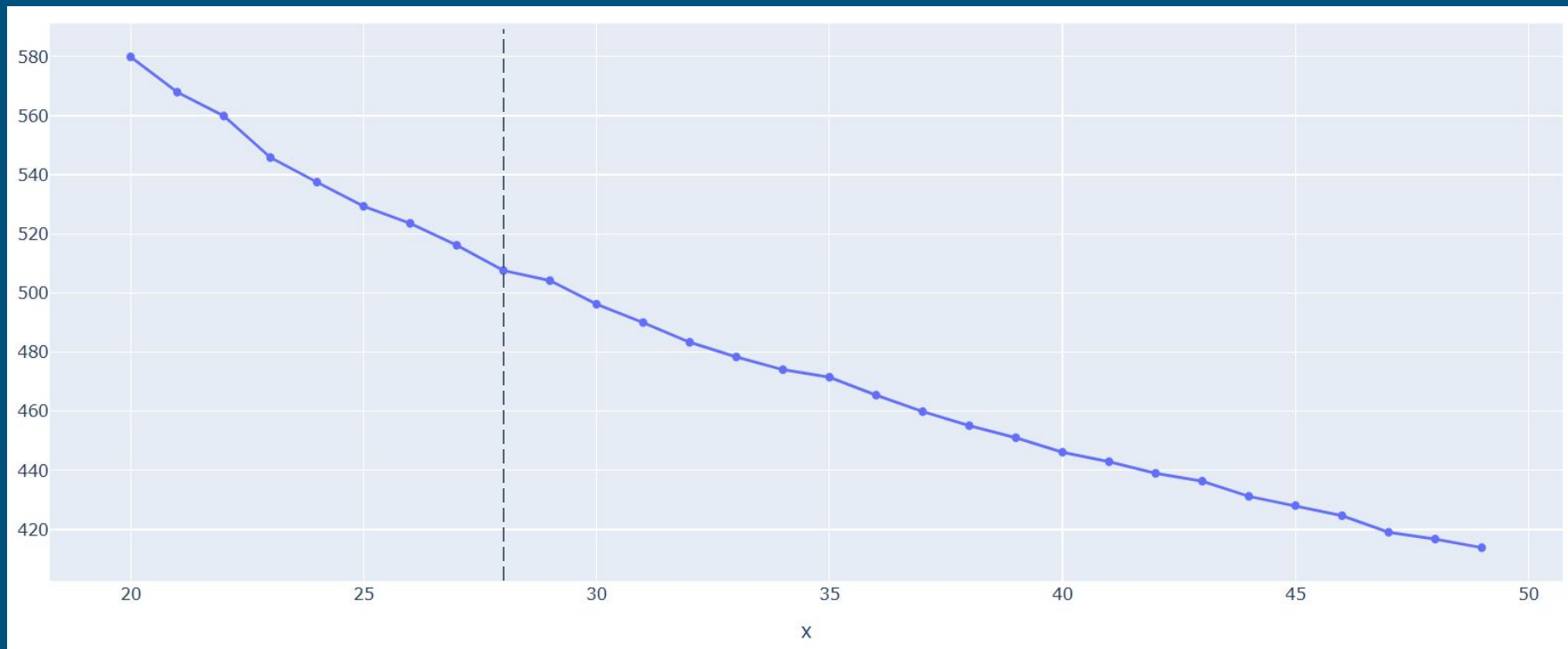
Used features from spotify:

danceability
energy
loudness
mode
speechiness
acousticness
instrumentalness
liveness
valence
tempo

Data Scaling

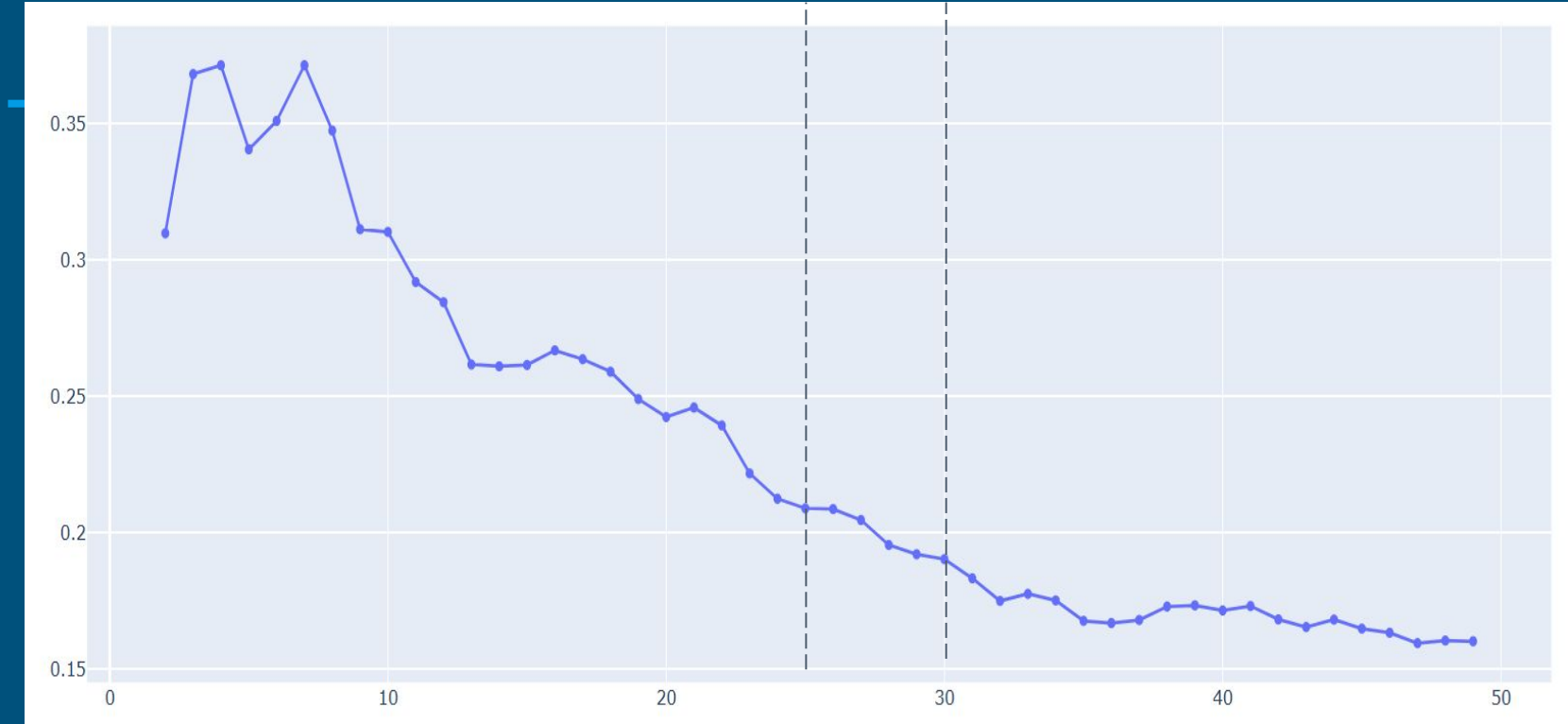
We are scaling **'loudness'**, **'tempo'**
using ***'MinMaxScaler'*** which scales the
data to a range between 0 and 1.

Inertia and the elbow method



based on that -> elbow lied on $k = 28$

The silhouette score



based on that -> k-cluster between 25-30

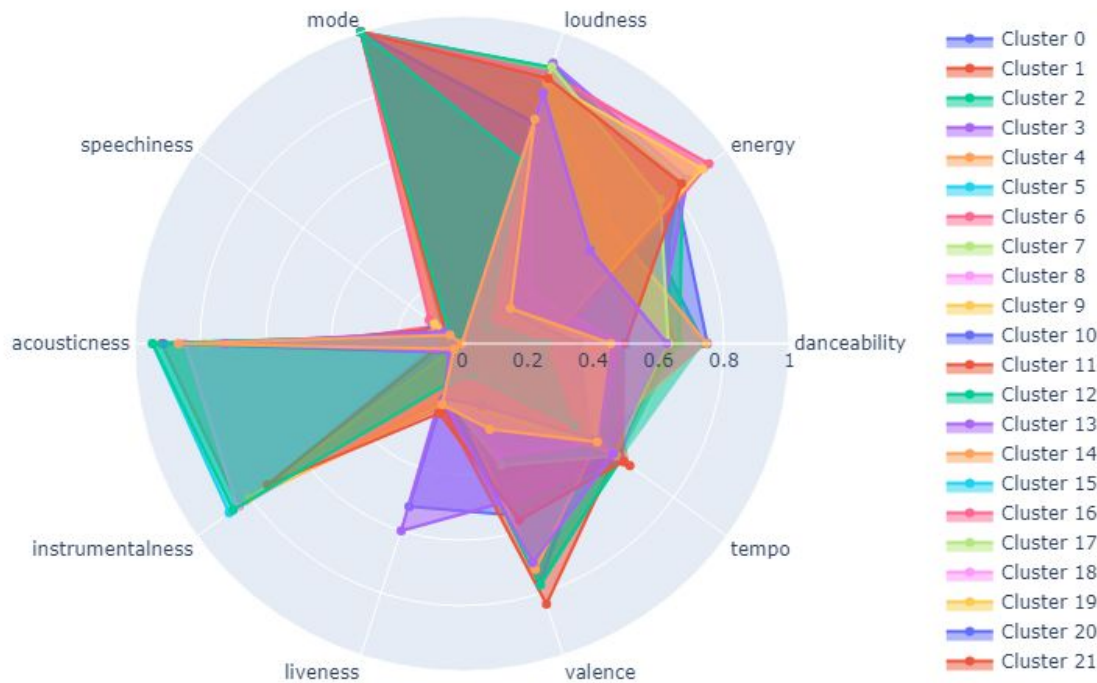


Based on

the silhouette score,
Inertia and the elbow method,

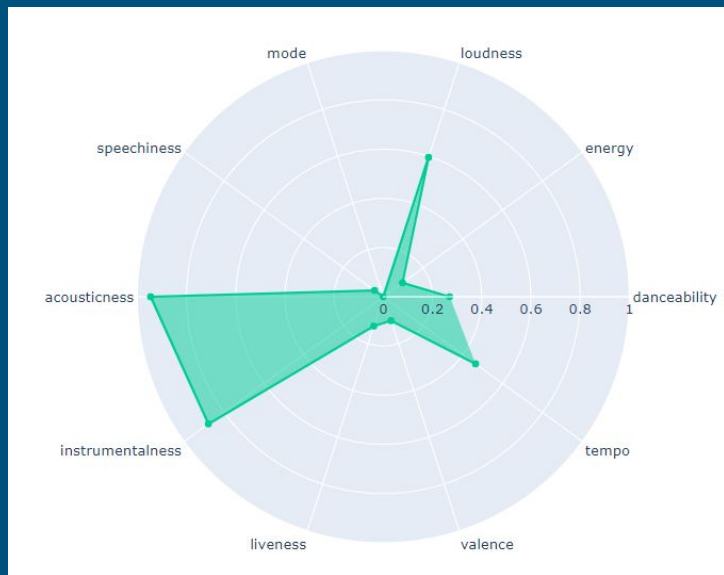
we decided suitable number of
clusters is
25 in our case

K-mean







We created
25 K-means cluster

Playlist













<https://open.spotify.com/playlist/1Ja9OBWBj4iyb9ZejiPbhW>



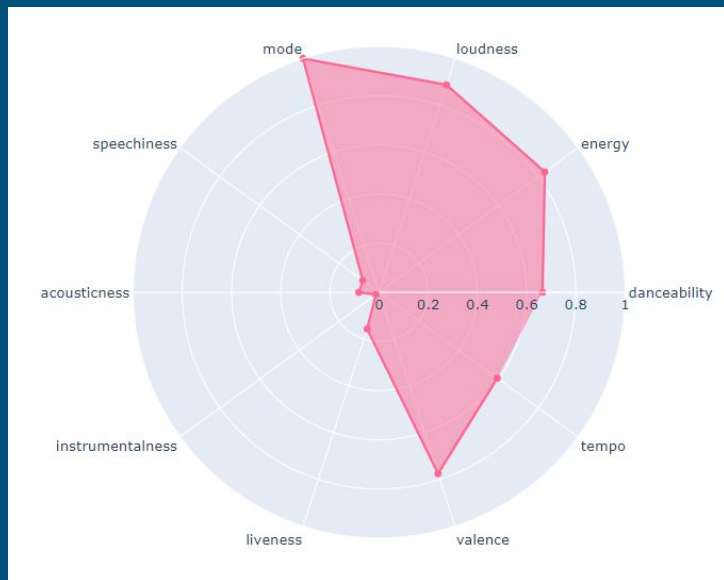
Playlist

Melancholic music

Hyoni • 10 songs, 2 hr 9 min

#	Title	Album
1	 Albatross - 2018 Remas... Fleetwood Mac	The Pious Bird of Good ...
2	 Two Skies Dirk Maassen	Ocean
3	 I. Adagio sostenuto Ludwig van Beethoven, Igo...	Piano Sonata No. 14 in C...
4	 Brotin Eydis Evensen	Bylur
5	 Concerto for Orchestra ... Witold Lutosławski, Polish ...	Lutoslawski: Symphoni...
6	 6 Chorale Preludes, BV ... Johannes Brahms, Ferrucci...	Encounter
7	 Symphony No. 1 in B flat... William Walton, Sir Colin Da...	Walton: Symphony No. 1
8	 Symphony No. 6 in A M... Anton Bruckner, London Ph...	Bruckner: Symphony No...
9	 Symphony No. 6 Allan Pettersson, Norrköpin...	Pettersson, A.: Sympho...
10	 A Flock Descends Into T... Toru Takemitsu, Ensemble T...	Takemitsu: Quatrain; A F...

Playlist



<https://open.spotify.com/playlist/5JN5tYEJkyIjp2WUKCpJfI>

Playlist

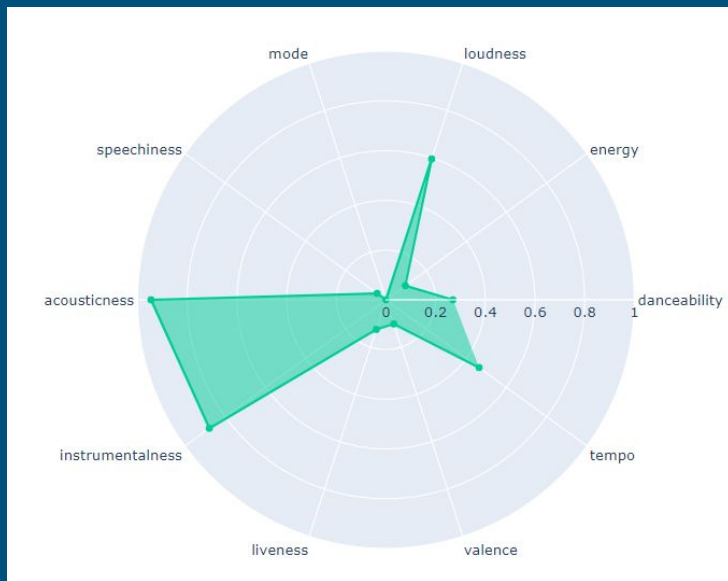
good vibe dance music

Hyoni • 11 songs, 38 min 28 sec

#	Title	Album
1	Alegria Batida	Batida (Soundway Reco...
2	Gecko (Overdrive) - Ra... Oliver Heldens, Becky Hill	Gecko (Overdrive) [Radi...
3	Unwritten Natasha Bedingfield	Unwritten
4	Chulin Culin Chunfly (fe... E Voltio, Calle 13	Voltio
5	Nie allein TEY	Nie allein
6	Aufstehn! (feat. CeeLo ... Seed, CeeLo Green	Aufstehn!
7	Magic Chucho	Tejido De Felicidad
8	Mary E Bausa	Mary
9	All The Lies (with Felix J... Alok, Felix Jaehn, The Vamps	All The Lies (with Felix J...
10	Vampiresa E Johnny Prez	The Money Machine Pre...
11	Ein Stern (der deinen N... DJ Ötzi, Nik P.	Ein Stern (der deinen Na...

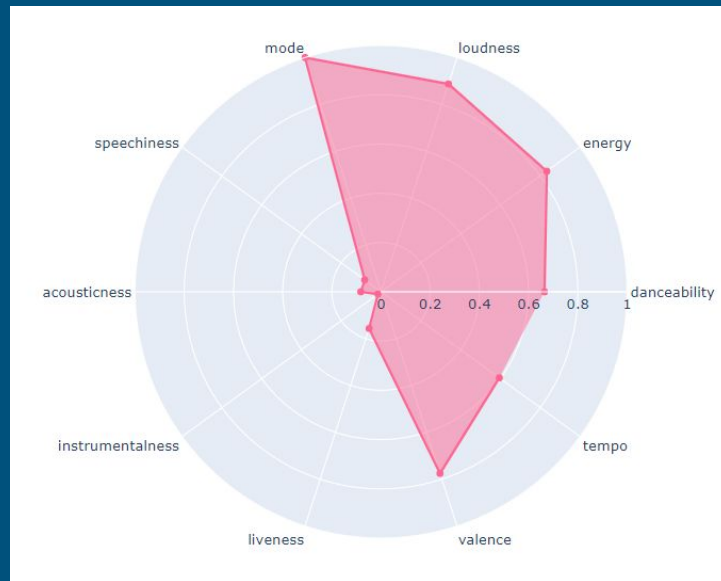
Playlist Comparison

Melancholic music



VS

Good vibe dance music

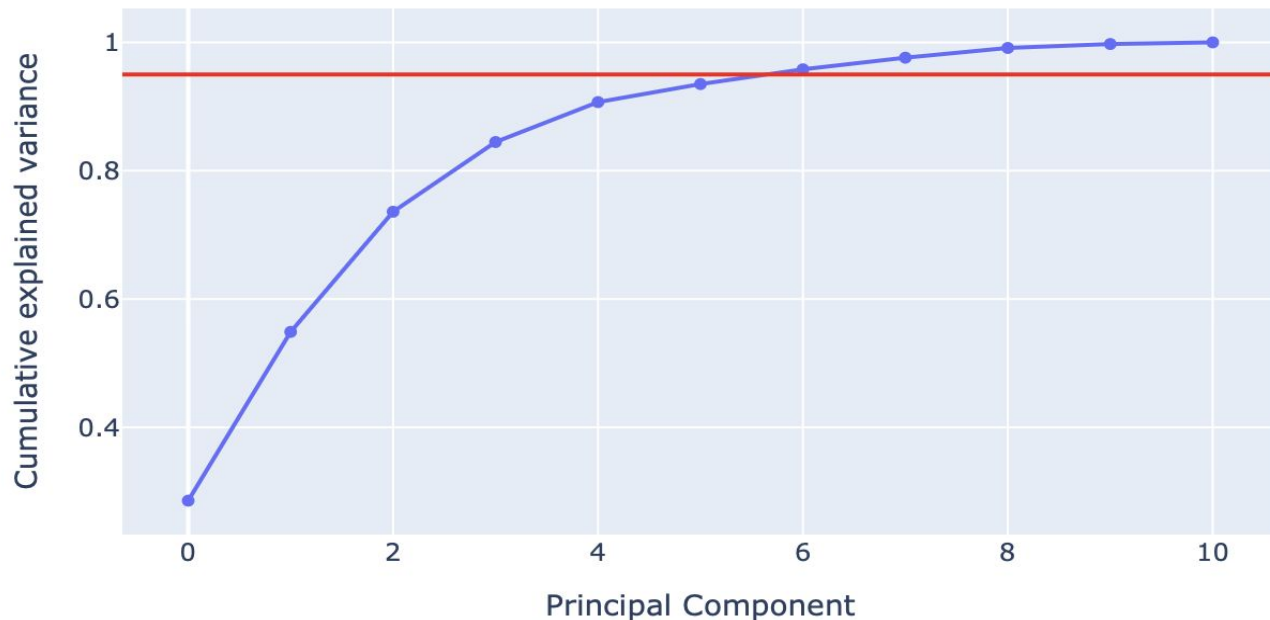


- PCA Analysis

- WE perform PCA on the scaled data to reduce dimensionality.
- The explained variance plot suggests that around 7 principal components explain 95% of the variance in the dataset.

Cumulative explained variance (CEV) is calculated by summing the explained variance of each Principal Component. The CEV plot allows up to compare cumulative explained variance to some benchmark value.

Cumulative explained variance of Principal Components



Comparison of Clustering Results:

- We compare the clustering results from the base model (without PCA) with the PCA-based model.

- It is evident that the cluster assignments differ between the base model and the PCA model. This suggests that dimensionality reduction via PCA has an impact on the clustering results

	Base model	PCA - 95%
1665	0	22
4667	0	18
1403	0	1
2860	0	23
573	0	11
...
2374	24	24
2373	24	5
2445	24	18
2381	24	5
2378	24	24

CONCLUSIONS

- K-means clustering helps reduce large song datasets but isn't fully accurate for creating playlist.
- Clustering method is useful but not enough on its own; other methods should be explored.
- Spotify's audio features don't completely capture human-perceived song similarity

Made by pretty girls:
Hyoen,
Kate,
Alla.