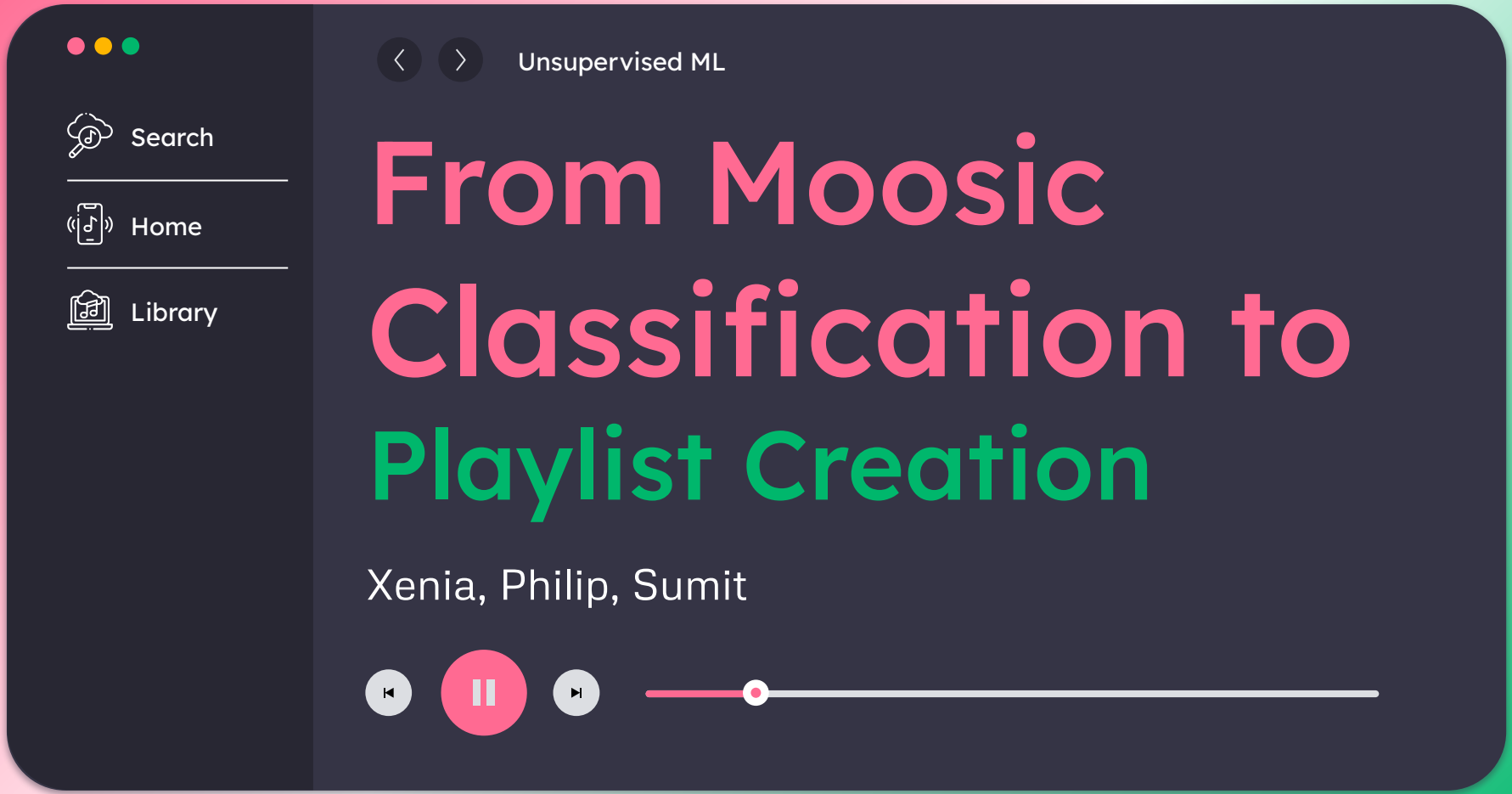

Search
Home
Library
Unsupervised ML
From Moosic Classification to Playlist Creation
Xenia, Philip, Sumit

Data Processing

#1 Data Cleaning
Fix strings, drop duplicates & outliers

#2 Separate Data
Descriptive vs. numerical features

#3 Scaling
Relevant features using MinMax scaler

#4 Song Classification
Using a layered strategy

#5 Push Playlists
To Spotify

Layered Classification: Steps

1. **Main clusters**
    a. Identify relevant features for the main clustering
        i. PCA
        ii. Exploratory analysis
    b. Select the number of main clusters (Elbow + Silhouette plots)
    c. Create main clusters (*num_clusters = 5*)
2. **Subclusters**
    a. Select the number of subclusters
    b. Create subclusters (*num_clusters = 5*)

Step 1: Identify Distinguishing Features

	PC_1	PC_2	PC_3
danceability	-0.141786	-0.018445	-0.145324
energy	-0.213625	-0.053883	0.185191
key	-0.036622	0.091124	0.013083
loudness	-0.115993	-0.027323	0.032489
mode	0.133547	-0.477993	0.010980
speechiness	-0.018823	-0.001102	0.016863
acousticness	0.276062	0.062073	-0.229004
instrumentalness	0.273114	0.084209	0.221462
liveness	-0.019377	-0.003109	0.016083
valence	-0.171626	-0.053570	-0.152796
tempo	-0.041197	-0.011335	0.027233

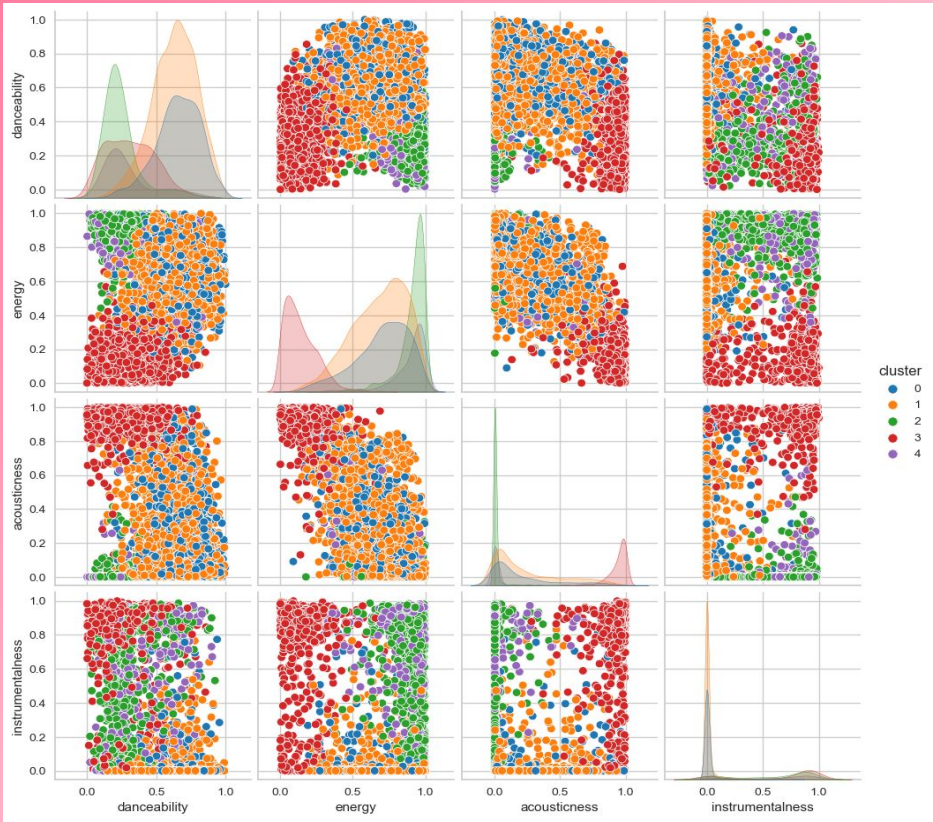
PCA

For PC_1

- Positively loading features in blue
- Negatively loading features in pink

Note: we drop *valence* from the first step as it needn't drive broad categorizations

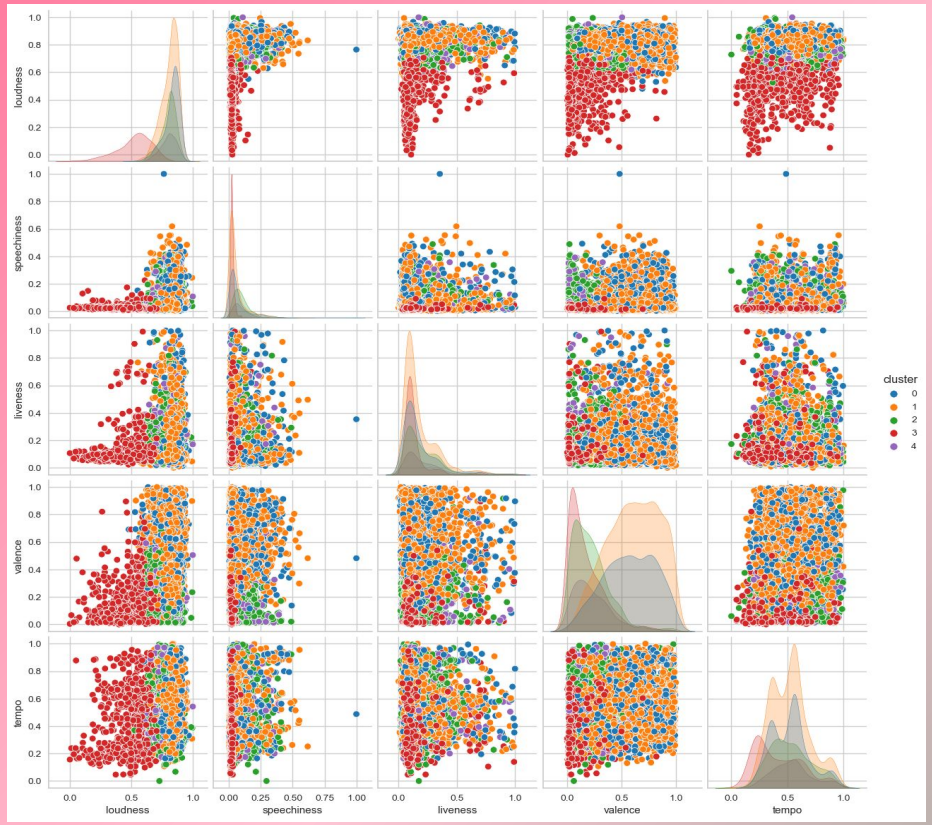
Step 1: Identify Distinguishing Features



Exploratory analysis

These four are also able to distinguish clusters!

Step 1: Identify Distinguishing Features

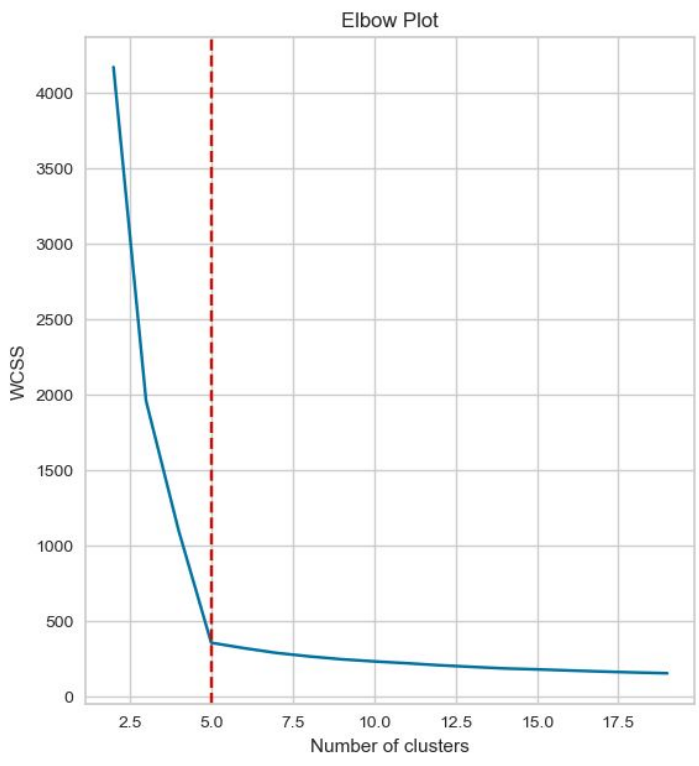


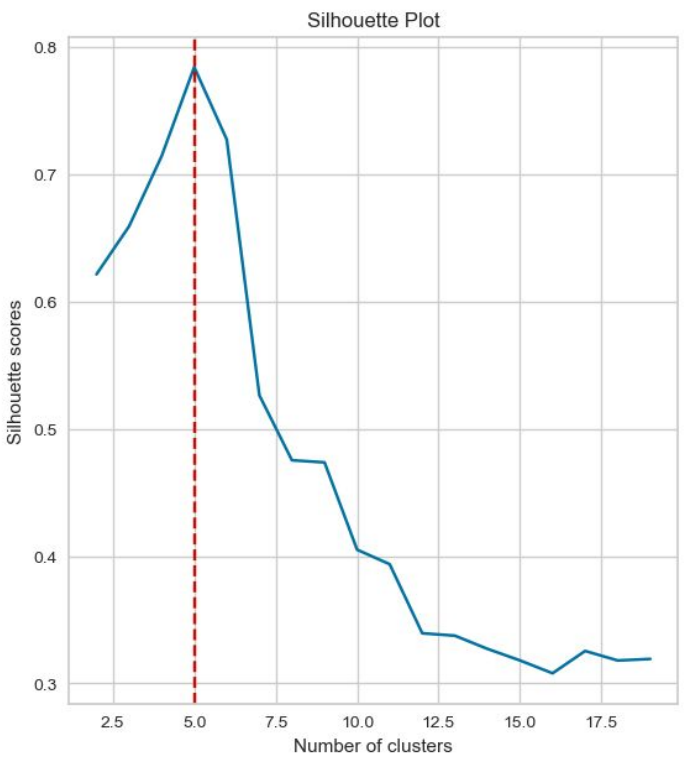
Exploratory analysis

The remaining features do not do an equally good job.

→ Hence dropped from the first clustering layer, considered later!

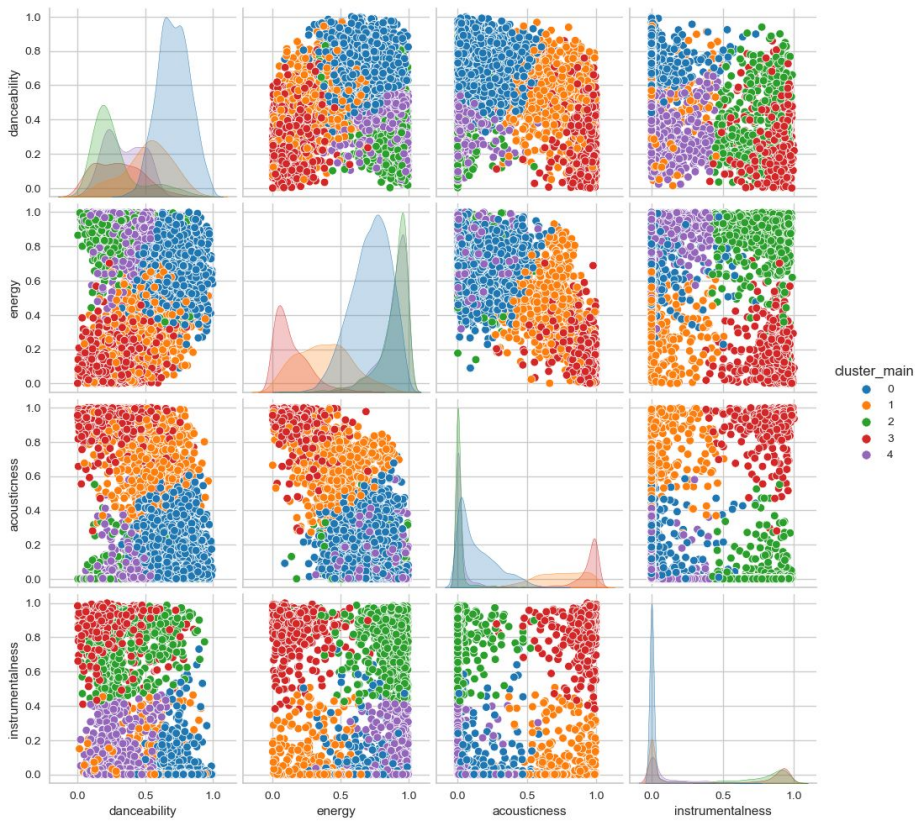
Step 2: Select number of main clusters





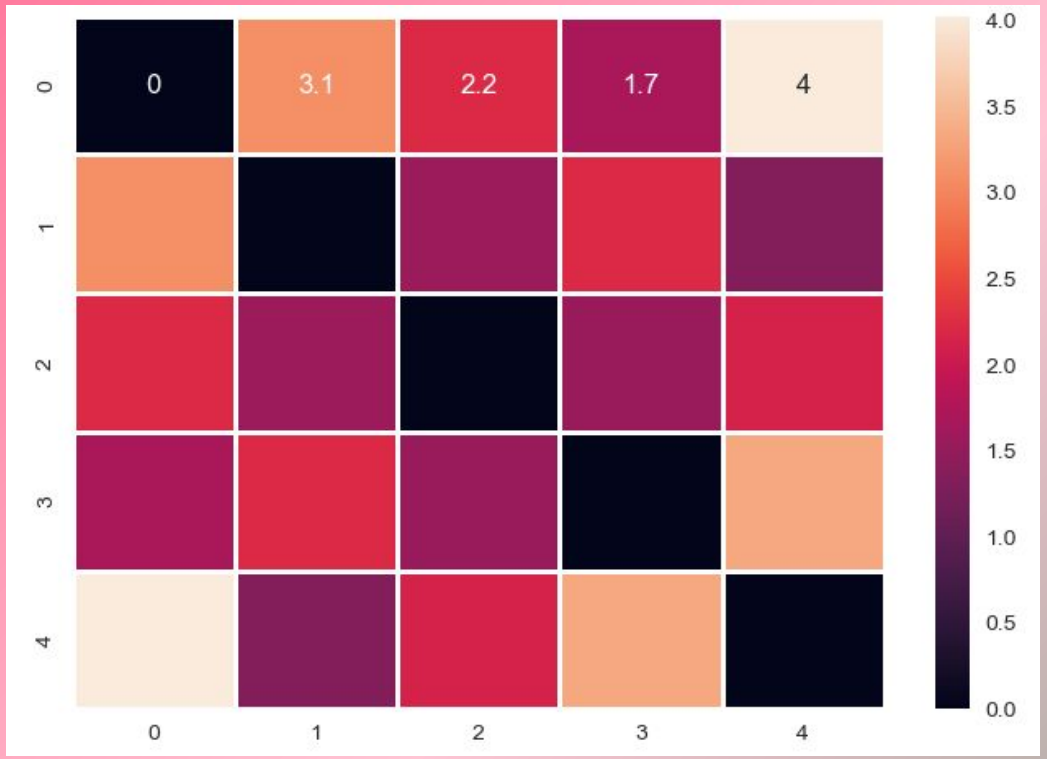
5 seems about right!

Step 3: Create 5 Main Clusters



Main clusters are distinguishable!

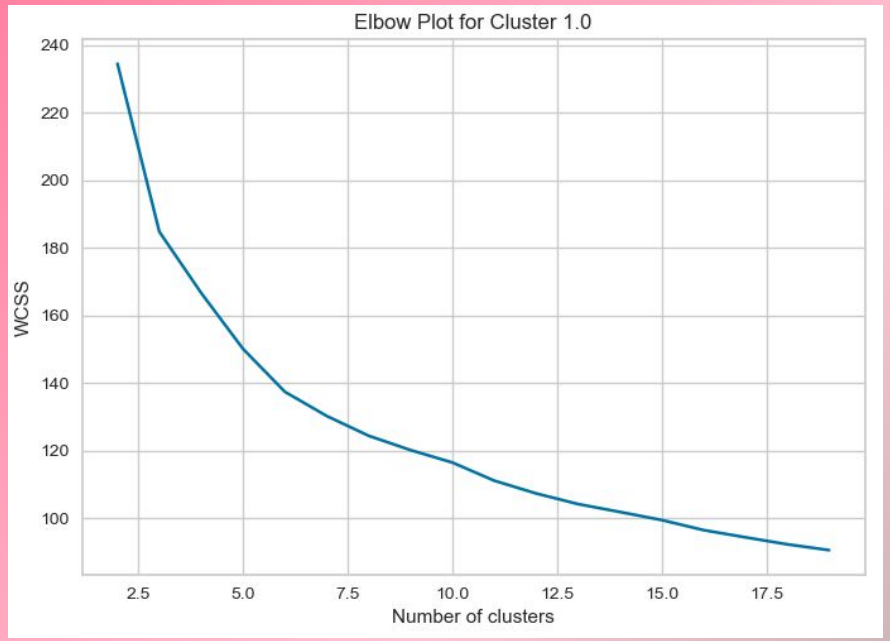
Step 3: Create 5 Main Clusters

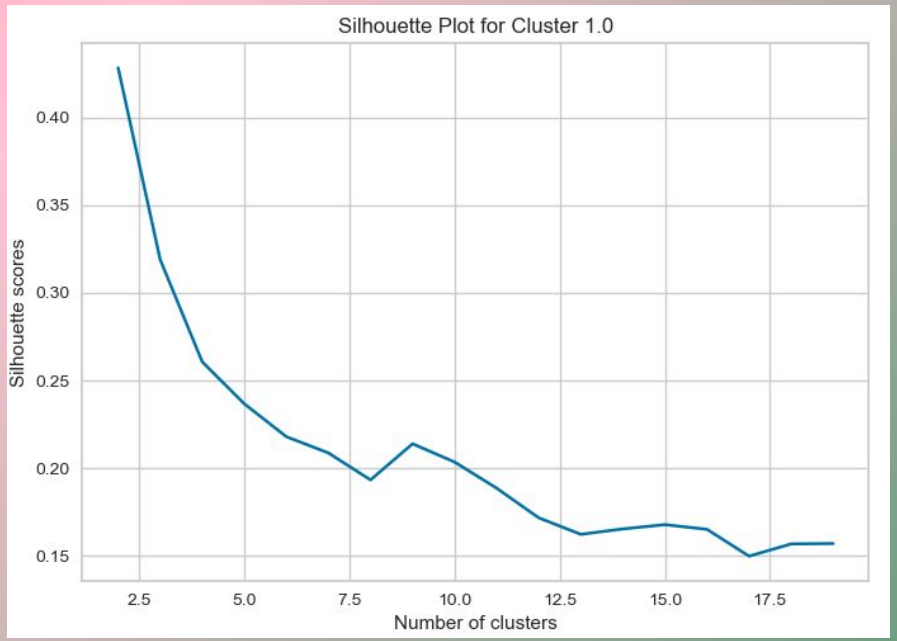


Centroids are distinct!

Steps 4+5: Create Subclusters

4-5 subclusters seem to work!

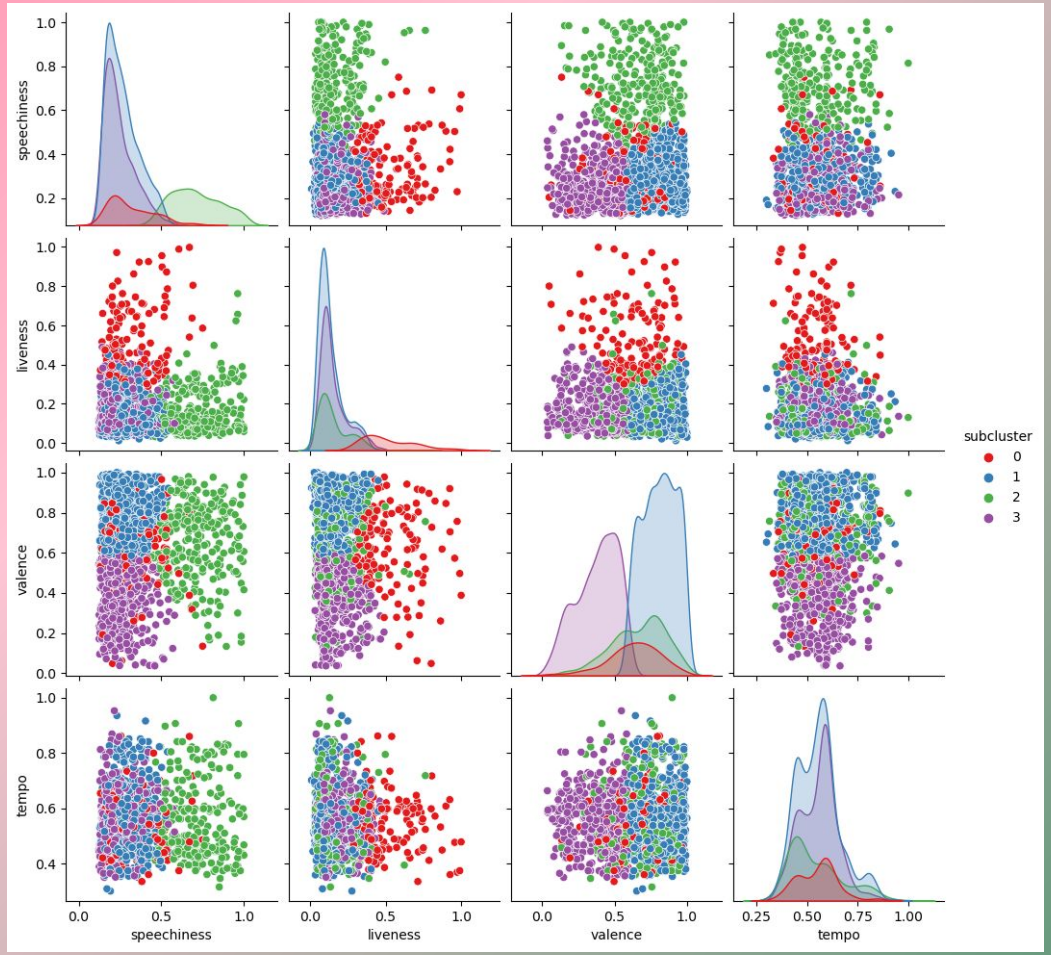




Steps 4+5: Create Subclusters

Subclusters well-separated for good choice of features for clustering!

Requires manual tests!



Example Playlists

main sub

playlist_0_1:

Pop & Rock Happy Vocals

works!

works!

not really...

sort of...

works!

sort of...

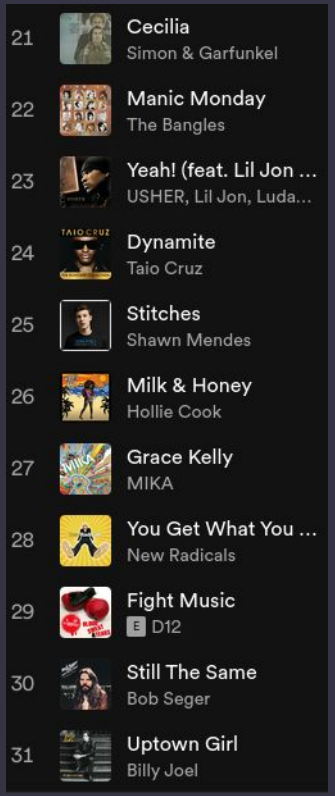
works!

works!

sort of...

works!

works!



Example Playlists

works!

playlist_3_1: not really...

Relaxing to acoustical piano

works!

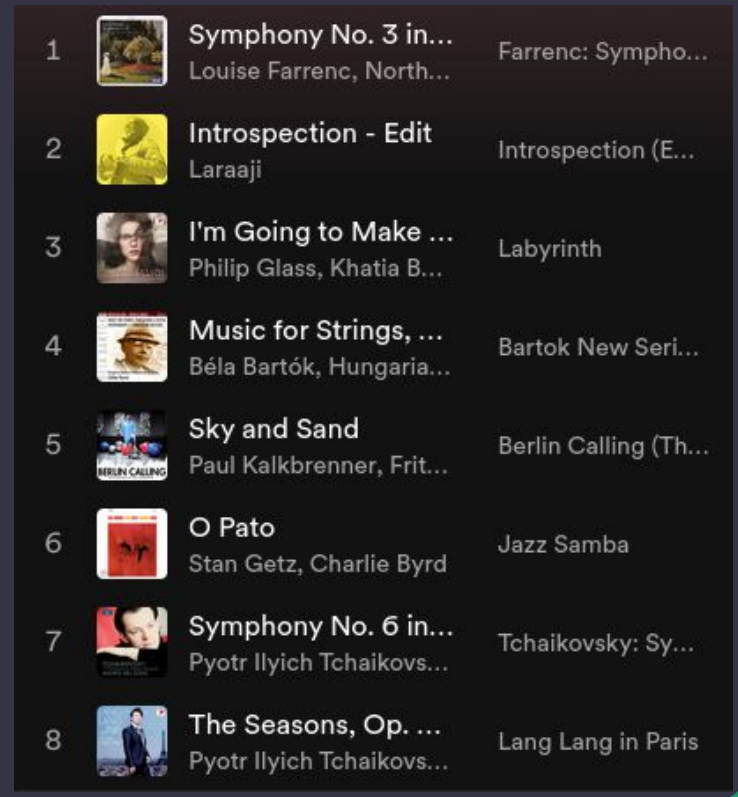
works!

what

not really...

works!

works!



Example Playlists

playlist_4_0:

works!

Obscure Death Metal

works!

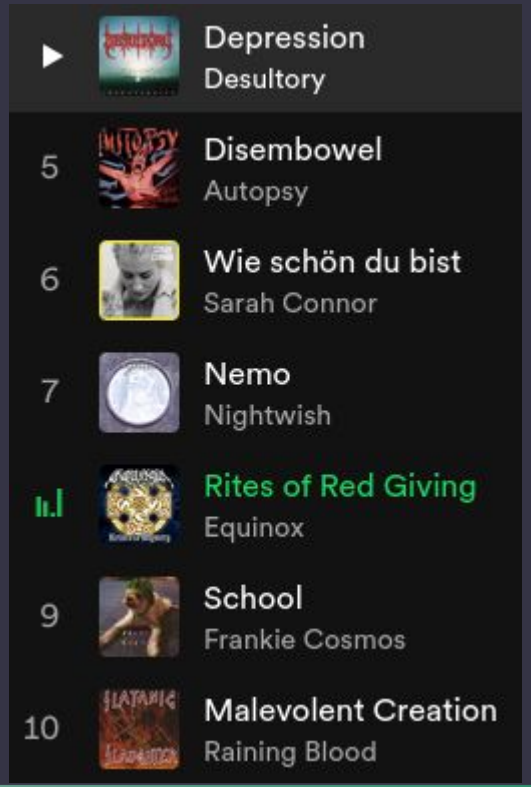
does not work!

sort of works...

works!

nope!

works!



Concluding remarks

01 **Are Spotify's audio features able to identify "similar songs", as defined by humanly detectable criteria?**

In principle yes - to a certain degree - but tricky!

02 **Is K-Means a good method to create playlists ?**

Clusters are sensitive to feature selection.

Requires a lot of manual fine-tuning and using relevant musical features/business logic!