

Moosic



Unsupervised Machine Learning Project



Main Questions

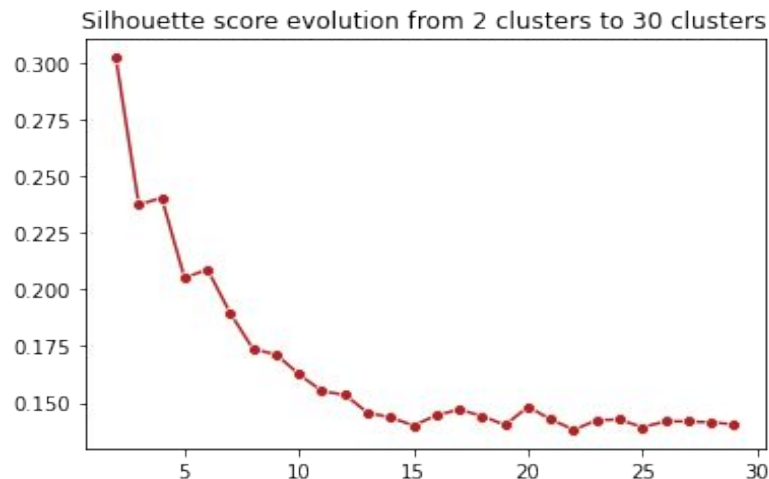
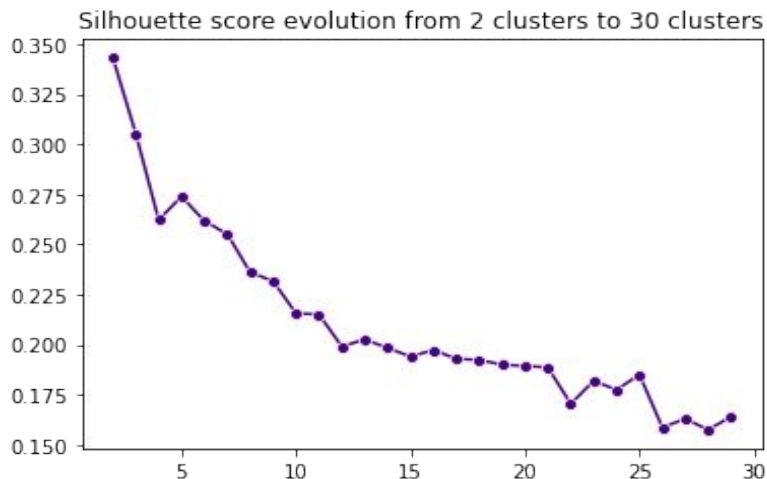
- Are Spotify's audio criteria sufficient for playlist generating?
- Is KMeans an appropriate method for unsupervised ML?

The Data

- Dataset with approx. 5200 songs to cluster into various playlists.
- After reviewing the data, we decided to drop the following columns:
 - 'id'
 - 'html'
 - 'artist'
 - 'type'
 - 'key'
 - 'mode'
 - 'time_signature'
 - 'duration_ms'

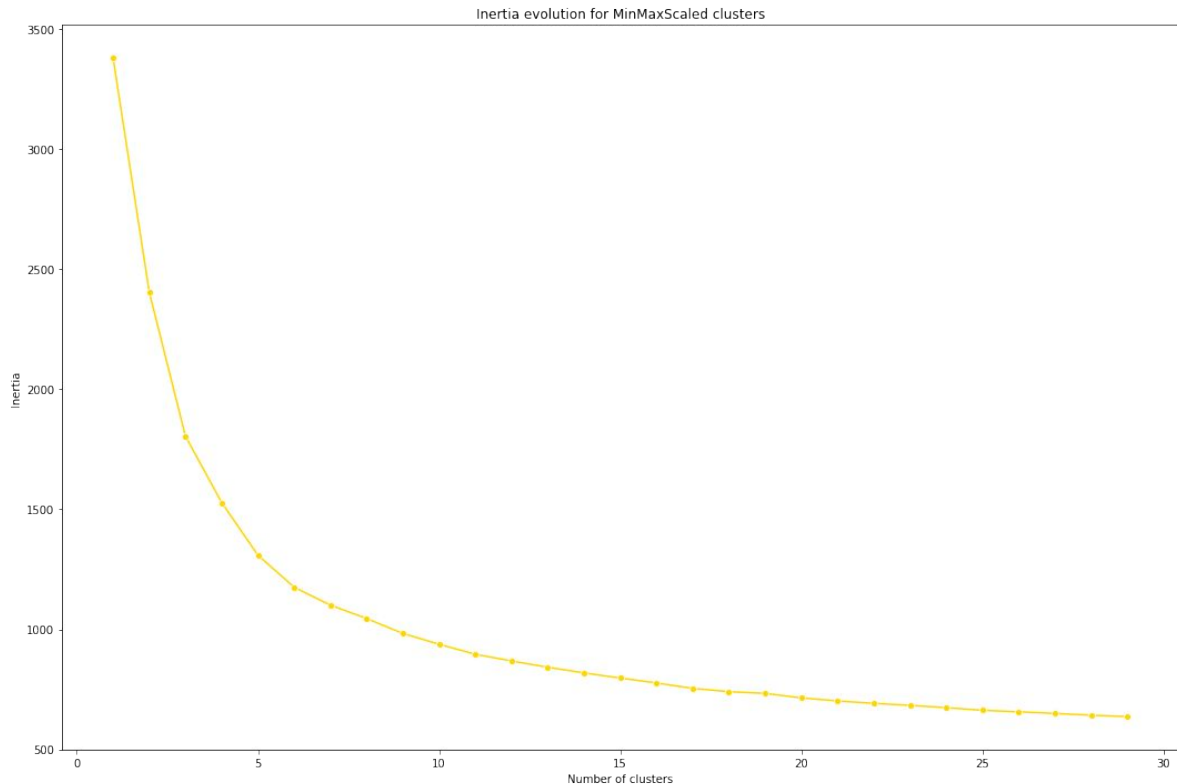
Scaling

- We decided to use the MinMaxScaler
 - Comparison between MinMax & Standard:



- Afterwards, we also used the Elbow chart with Inertia to decide the number of clusters

Number of Clusters



- Combining the findings of the inertia and the silhouette score we decided to go with 12 clusters / playlists.
- More clusters would not gain a lot, less clusters would mean extremely long playlists.

Comparing the Centers

	danceability	energy	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo
0	0.381047	0.161738	0.701988	0.047427	0.911344	0.048799	0.184679	0.216479	0.496382
1	0.306937	0.925627	0.869017	0.133307	0.010036	0.110967	0.193768	0.208344	0.566413
2	0.299504	0.107040	0.604758	0.048655	0.948021	0.887867	0.127047	0.125143	0.481671
3	0.800713	0.673231	0.864340	0.109451	0.138252	0.009680	0.134733	0.772939	0.504864
4	0.254557	0.908428	0.840993	0.116204	0.007304	0.822982	0.199477	0.205580	0.564583
5	0.588966	0.389938	0.801053	0.059365	0.677136	0.026352	0.165717	0.476887	0.541351
6	0.629036	0.844435	0.891159	0.093779	0.069793	0.022119	0.180330	0.723385	0.636284
7	0.635183	0.702570	0.878298	0.079028	0.100904	0.015474	0.165387	0.351813	0.569132
8	0.665244	0.643264	0.845068	0.085437	0.570846	0.017837	0.184953	0.810983	0.579521
9	0.571089	0.308529	0.718894	0.056325	0.837387	0.792827	0.178269	0.475637	0.543518
10	0.565132	0.816700	0.860306	0.072388	0.061516	0.750094	0.219936	0.539999	0.604515
11	0.617482	0.801350	0.882591	0.124881	0.151597	0.013617	0.658713	0.533389	0.575641

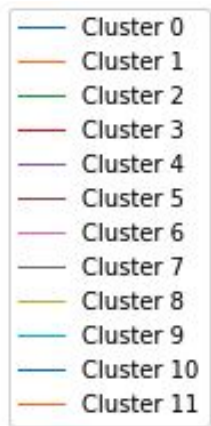
Loudness all similar values

Speechiness very small range

Liveness all similar values except cluster 11

Tempo all similar values

Comparing the Centers: Visually

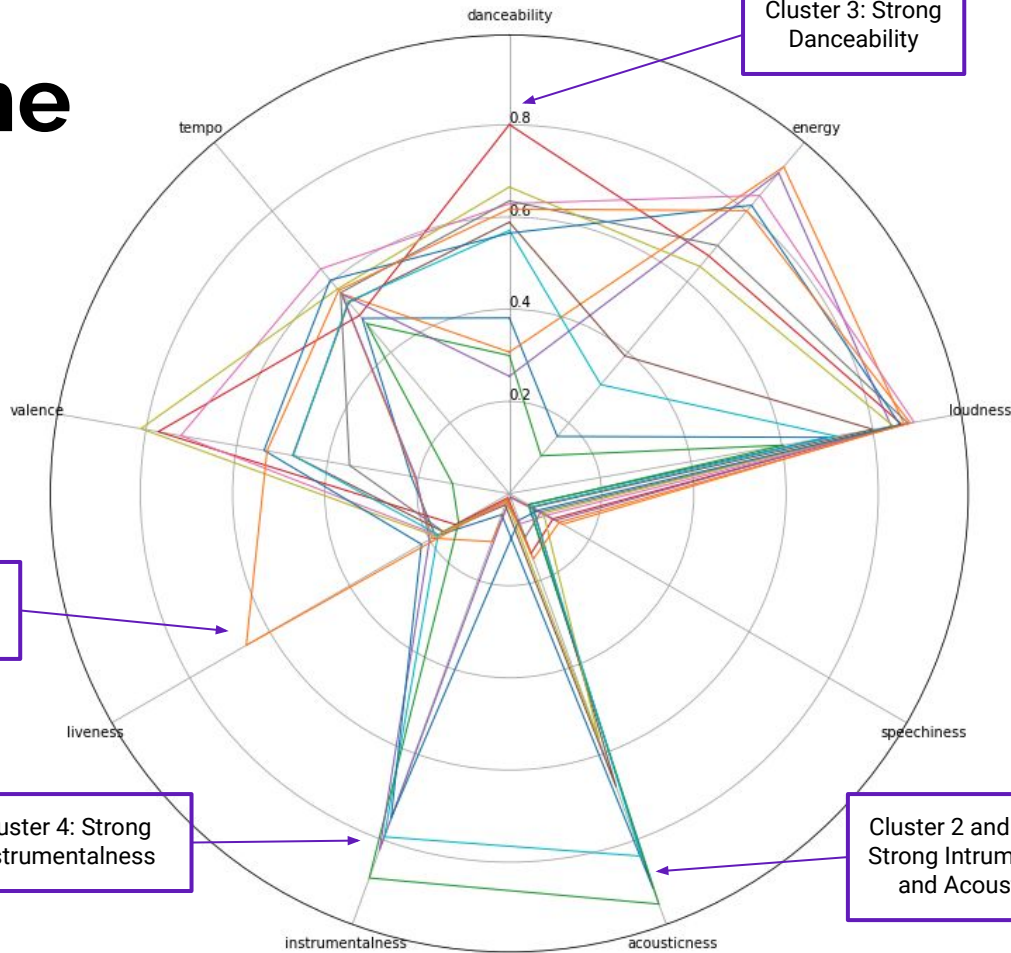


Cluster 11:
Strong Liveness

Cluster 4: Strong
Instrumentalness

Cluster 3: Strong
Danceability

Cluster 2 and Cluster 9:
Strong Instrumentalness
and Acousticness



In-depth Analysis of the Clusters

Example:

Death/Progressive Metal

- Very high tempo
- Very high energy
- Very low loudness

Outliers with
three indie rock tracks
and one EDM track
in sample

	danceability	tempo	energy	loudness	acousticness	instrumentalness	liveness
count	206.000000	206.000000	206.000000	206.000000	206.000000	206.000000	206.000000
mean	0.309606	0.892150	0.851651	0.010189	0.464738	0.199738	0.567743
std	0.110831	0.108305	0.049881	0.044427	0.103706	0.123852	0.121570
min	0.094209	0.492000	0.720827	0.000000	0.267005	0.044174	0.344787
25%	0.226474	0.853250	0.824594	0.000009	0.384772	0.100431	0.478569
50%	0.274560	0.926500	0.856143	0.000061	0.468528	0.160588	0.556183
75%	0.376680	0.970750	0.884602	0.000513	0.546954	0.278369	0.644261
max	0.635988	0.998000	1.000000	0.360442	0.658883	0.616008	0.901294

		danceability	tempo	energy	loudness	acousticness	instrumentalness	liveness
name	artist							
Let's Get Out Of This Country	Camera Obscura	0.279214	0.830	0.903264	0.009598	0.440609	0.143870	0.548222
The Waiting	Angel Olsen	0.328852	0.890	0.806234	0.000025	0.368528	0.077609	0.901294
Els dies s'allarguen	Renaldo & Clara	0.439504	0.714	0.790796	0.094679	0.475127	0.066768	0.489700
Dark Eternity	Therion	0.232678	0.811	0.832285	0.000002	0.647716	0.060588	0.525613
Phrenetic Tendencies	Fleshcrawl	0.241986	0.908	0.814401	0.000269	0.635533	0.064134	0.648731
I Abstain	Napalm Death	0.293692	0.994	0.895634	0.000014	0.493401	0.069605	0.434978
Never to Die Again	Fleshcrawl	0.202689	0.756	0.798556	0.001145	0.461929	0.096859	0.405869
Slaughterday	Autopsy	0.260600	0.994	0.887500	0.009177	0.513706	0.336373	0.495967
Condemned to Obscurity	Gorguts	0.264736	0.949	0.852206	0.000036	0.482234	0.122594	0.657227
Ghosts n Stuff - Sub Focus Remix	deadmau5	0.470527	0.824	0.906459	0.000247	0.496447	0.073658	0.813384
Why Do I Watch the Dawn	Cartilage	0.199586	0.973	0.933895	0.040060	0.529949	0.324215	0.660671
Regurgitated Guts	Death	0.199586	0.926	0.846842	0.000000	0.617259	0.099595	0.467237

Summary Playlists / KMeans

- KMeans was able to cluster with decent efficiency for some of the values
 - For other values, there was no consistency in clustering
 - There was no sense for some of the genre “borders”; most infamously one playlist was made up of mostly Death Metal & Electronic Music / Techno because of similarities in loudness, tempo and energy.
-
- My recommendation is to keep the human factor strong for now in establishing playlists and look into more refined ML solutions for the future.