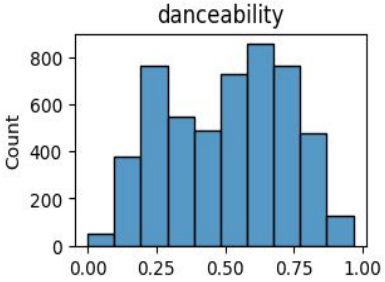
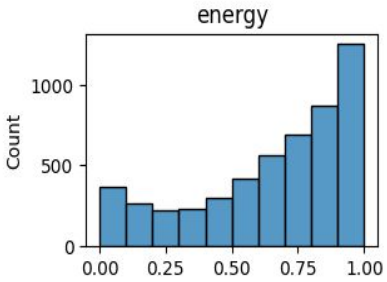
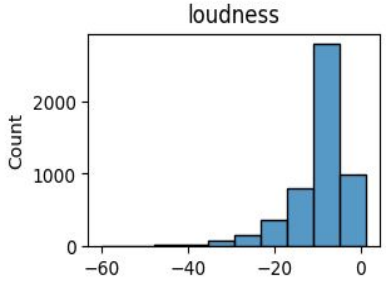
Clustering Songs into Spotify Playlists

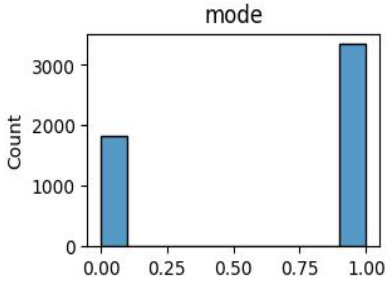
Sofía Tello, Ivan Curmi, Asif Siddiqi, Cathrin Rahn

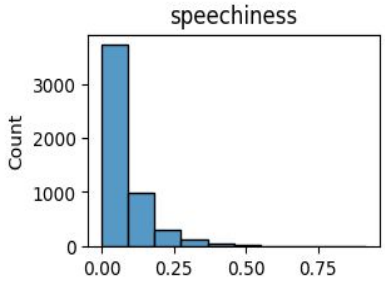
Feature Selection

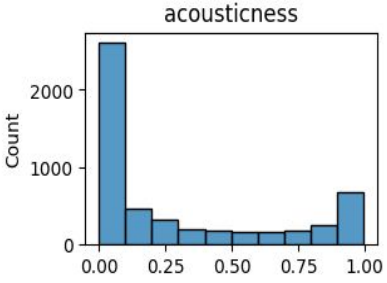


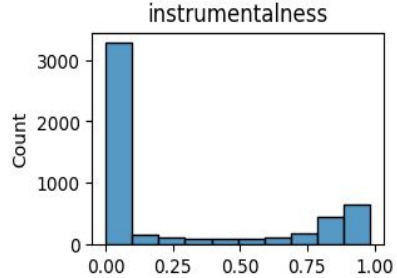


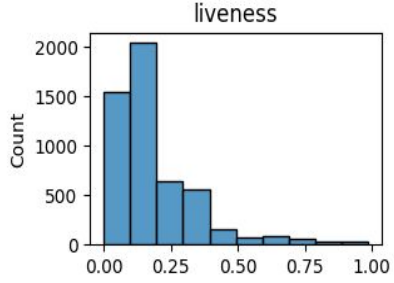


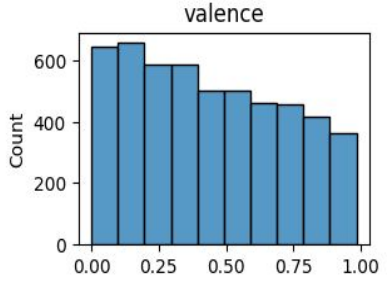


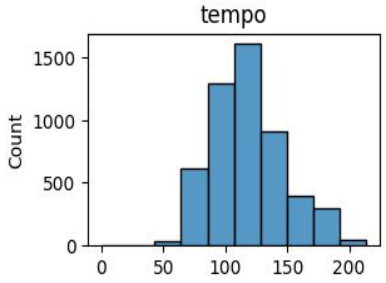


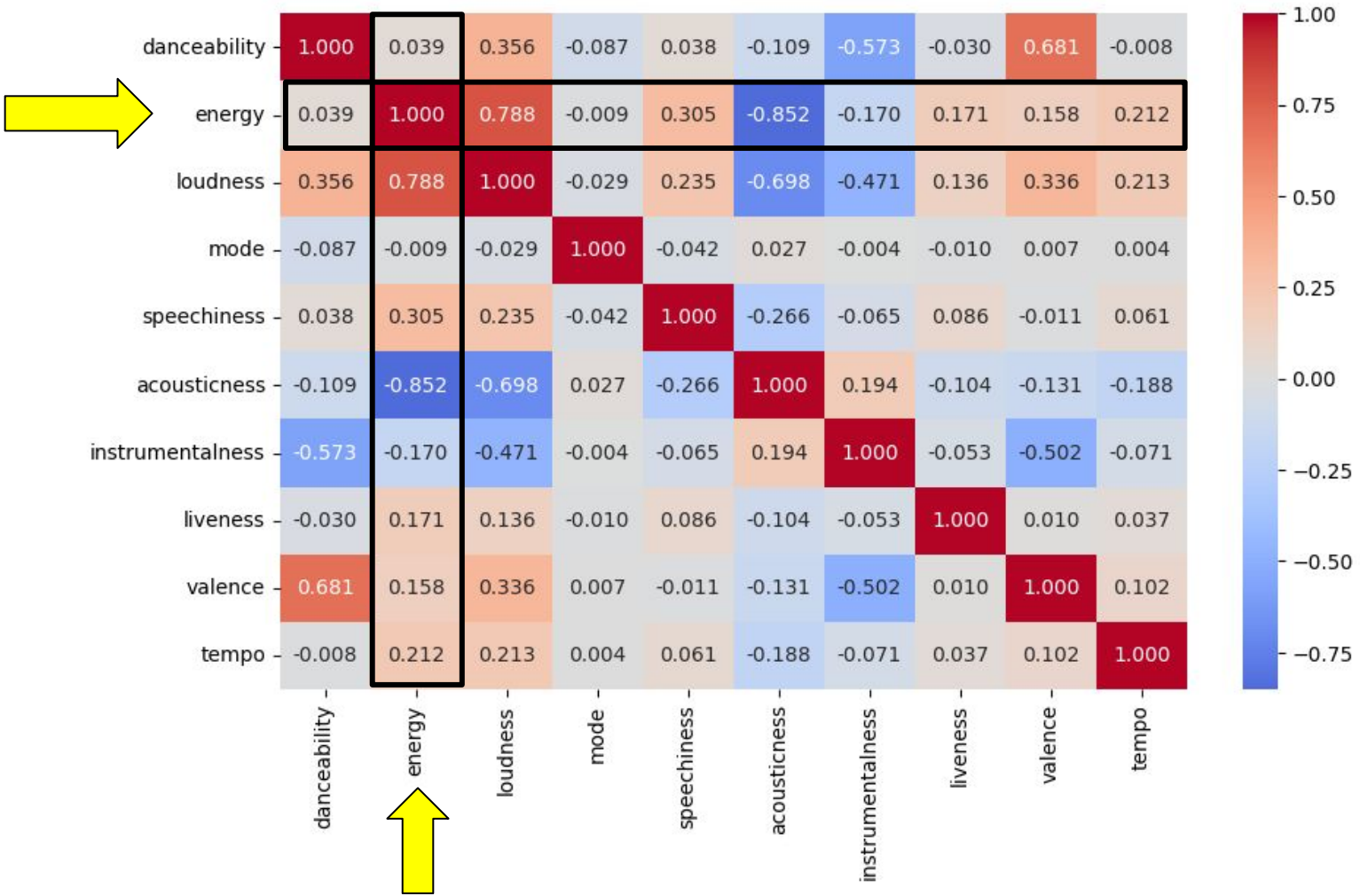


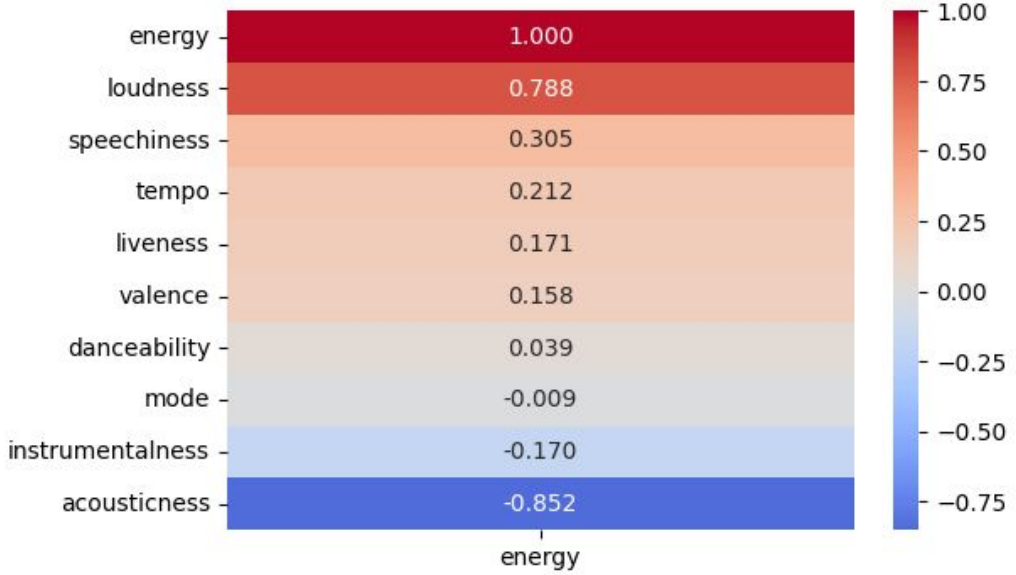


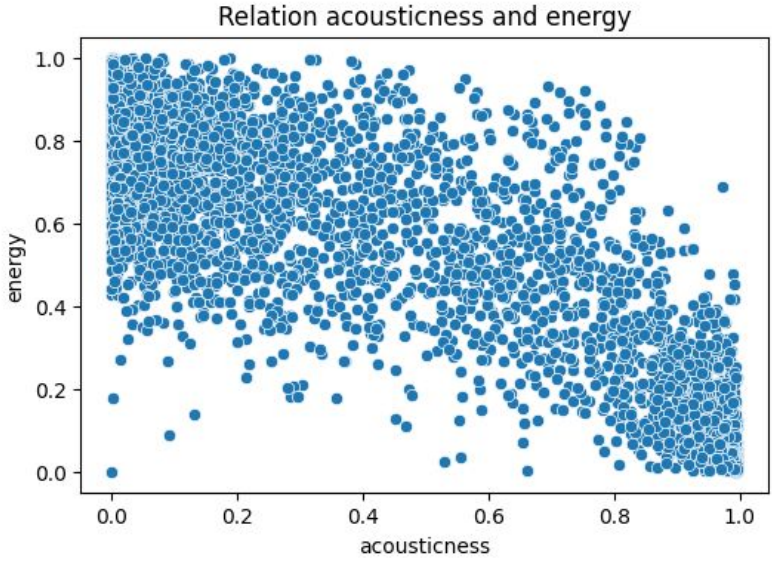


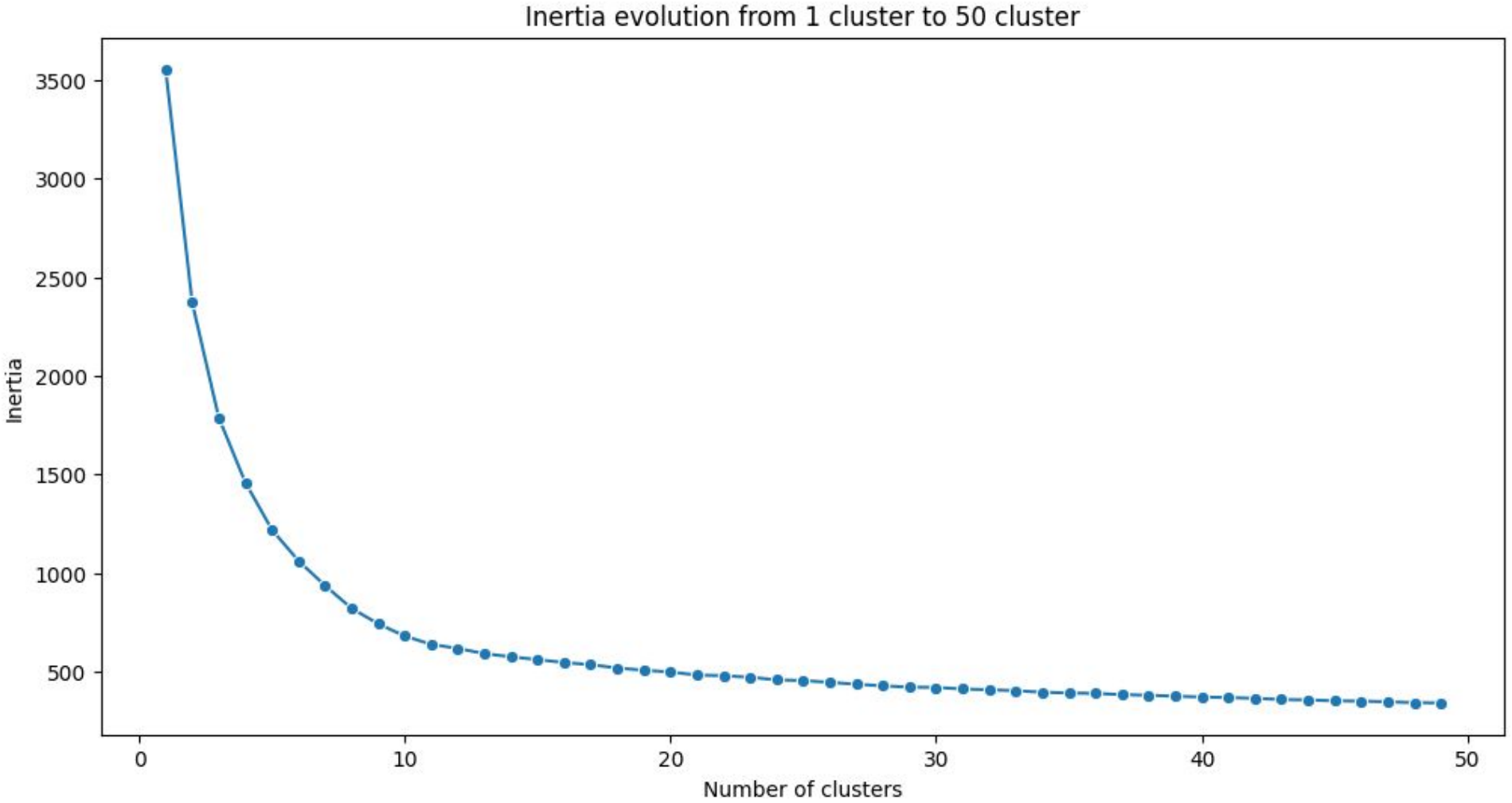


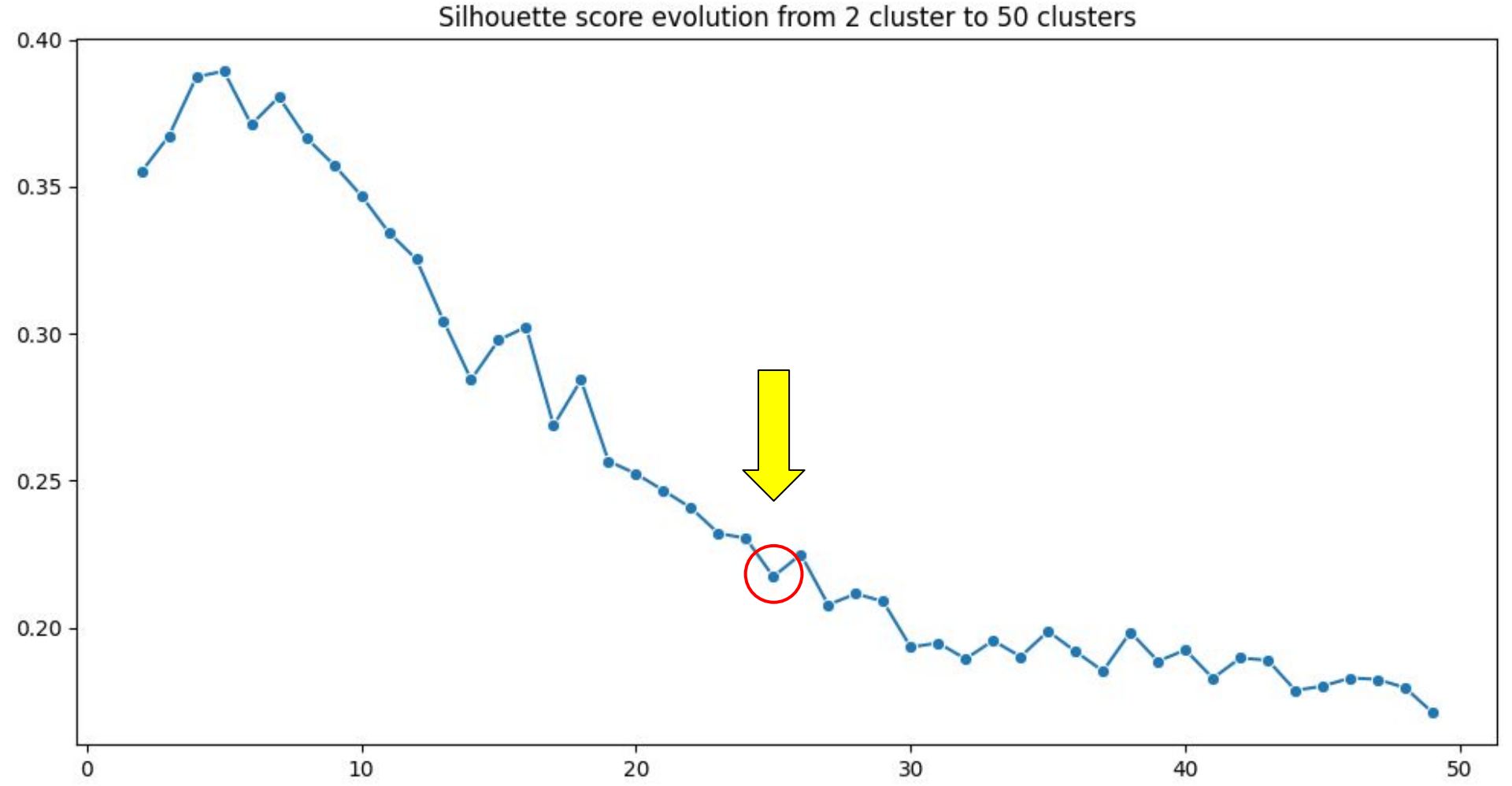


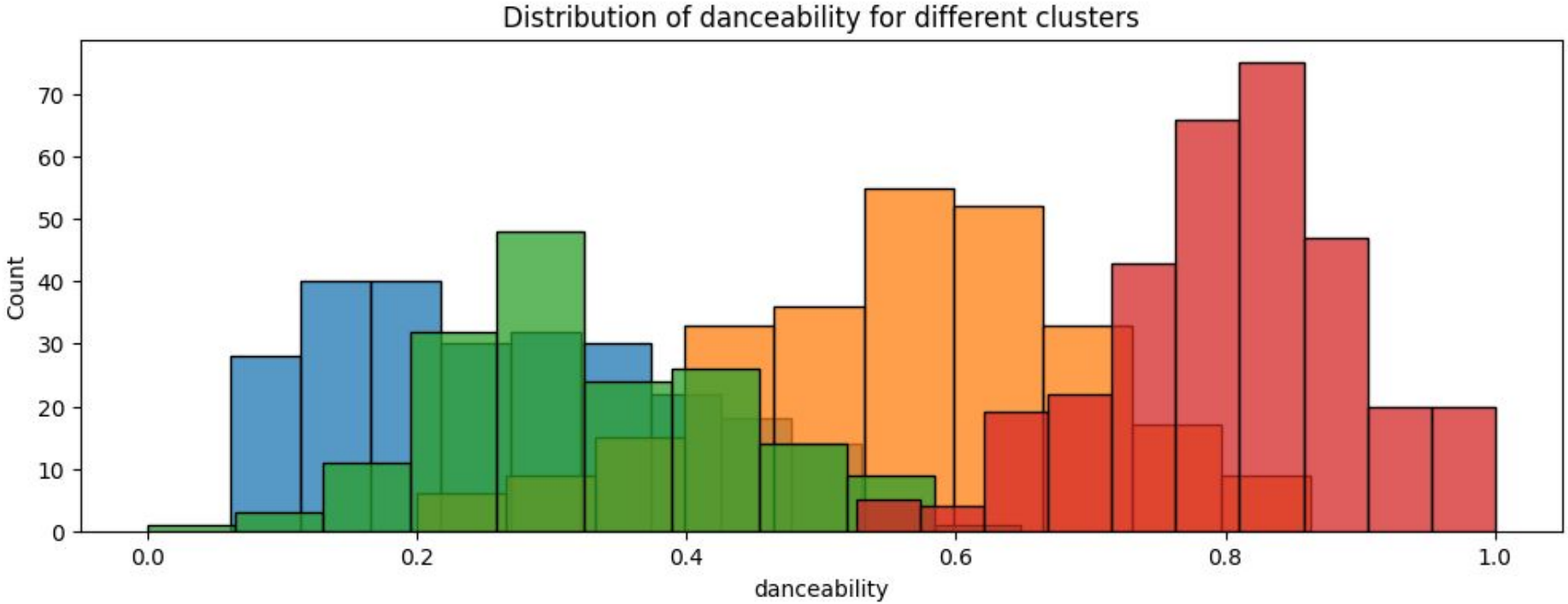

danceability 1.000 0.039 0.356 -0.087 0.038 -0.109 -0.573 -0.030 0.681 -0.008
energy 0.039 1.000 0.788 -0.009 0.305 -0.852 -0.170 0.171 0.158 0.212
loudness 0.356 0.788 1.000 -0.029 0.235 -0.698 -0.471 0.136 0.336 0.213
mode -0.087 -0.009 -0.029 1.000 -0.042 0.027 -0.004 -0.010 0.007 0.004
speechiness 0.038 0.305 0.235 -0.042 1.000 -0.266 -0.065 0.086 -0.011 0.061
acousticness -0.109 -0.852 -0.698 0.027 -0.266 1.000 0.194 -0.104 -0.131 -0.188
instrumentalness -0.573 -0.170 -0.471 -0.004 -0.065 0.194 1.000 -0.053 -0.502 -0.071
liveness -0.030 0.171 0.136 -0.010 0.086 -0.104 -0.053 1.000 0.010 0.037
valence 0.681 0.158 0.336 0.007 -0.011 -0.131 -0.502 0.010 1.000 0.102
tempo -0.008 0.212 0.213 0.004 0.061 -0.188 -0.071 0.037 0.102 1.000
danceability
energy
loudness
mode
speechiness
acousticness
instrumentalness
liveness
valence
tempo
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

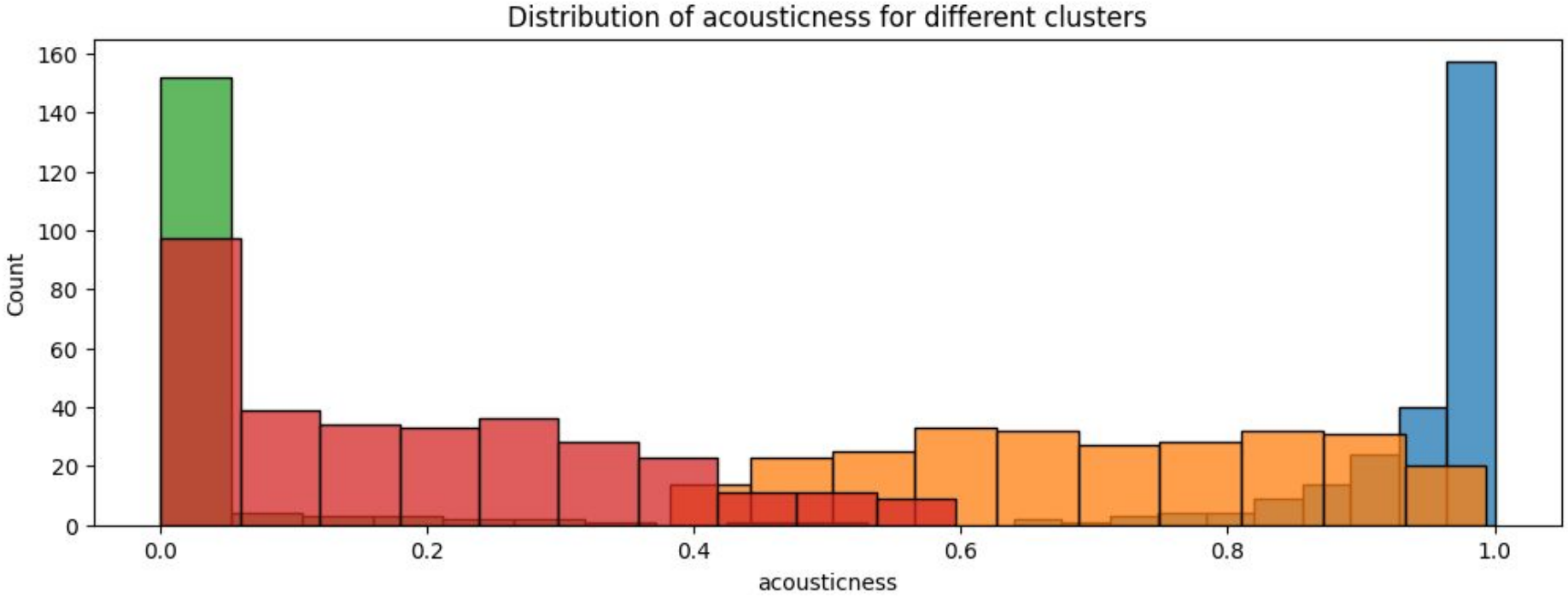

energy
loudness
speechiness
tempo
liveness
valence
danceability
mode
instrumentalness
acousticness
1.000
0.788
0.305
0.212
0.171
0.158
0.039
-0.009
-0.170
-0.852
energy
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75


Relation acousticness and energy
energy
acousticness
0.0
0.2
0.4
0.6
0.8
1.0


Inertia evolution from 1 cluster to 50 cluster
Inertia
3500
3000
2500
2000
1500
1000
500
0
10
20
30
40
50
Number of clusters


Silhouette score evolution from 2 cluster to 50 clusters
0.40
0.35
0.30
0.25
0.20
0
10
20
30
40
50


Distribution of danceability for different clusters
Count
0
10
20
30
40
50
60
70
0.0
0.2
0.4
0.6
0.8
1.0
danceability


Distribution of acousticness for different clusters
Count
160
140
120
100
80
60
40
20
0
0.0
0.2
0.4
0.6
0.8
1.0
acousticness

Exploring Genres

Cluster 0 Word Cloud

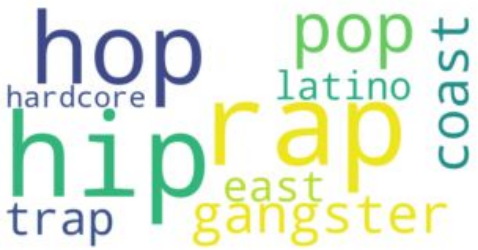
Cluster 3 Word Cloud

Cluster 1 Word Cloud

Cluster 4 Word Cloud

Cluster 2 Word Cloud

Cluster 5 Word Cloud



Conclusions

- Spotify audio features are able to identify similar songs, although there are some genres that are better classified than others

- K-Means is a good method to create playlists but it’s not enough - you still need external info to refine the job, like what is a good size for a playlist or to look at which elements definitely don’t belong to certain clusters

Thank you!