Moosic

Is K-Means a good method to create playlists?

Group 4
24.07.2023

Agenda

- Overview
- Data
- Inertia drop & Silhouette score evolution
- Silhouette Diagrams (Knife Chart)
- Heat Map
- Spider Chart
- Conclusion

Overview

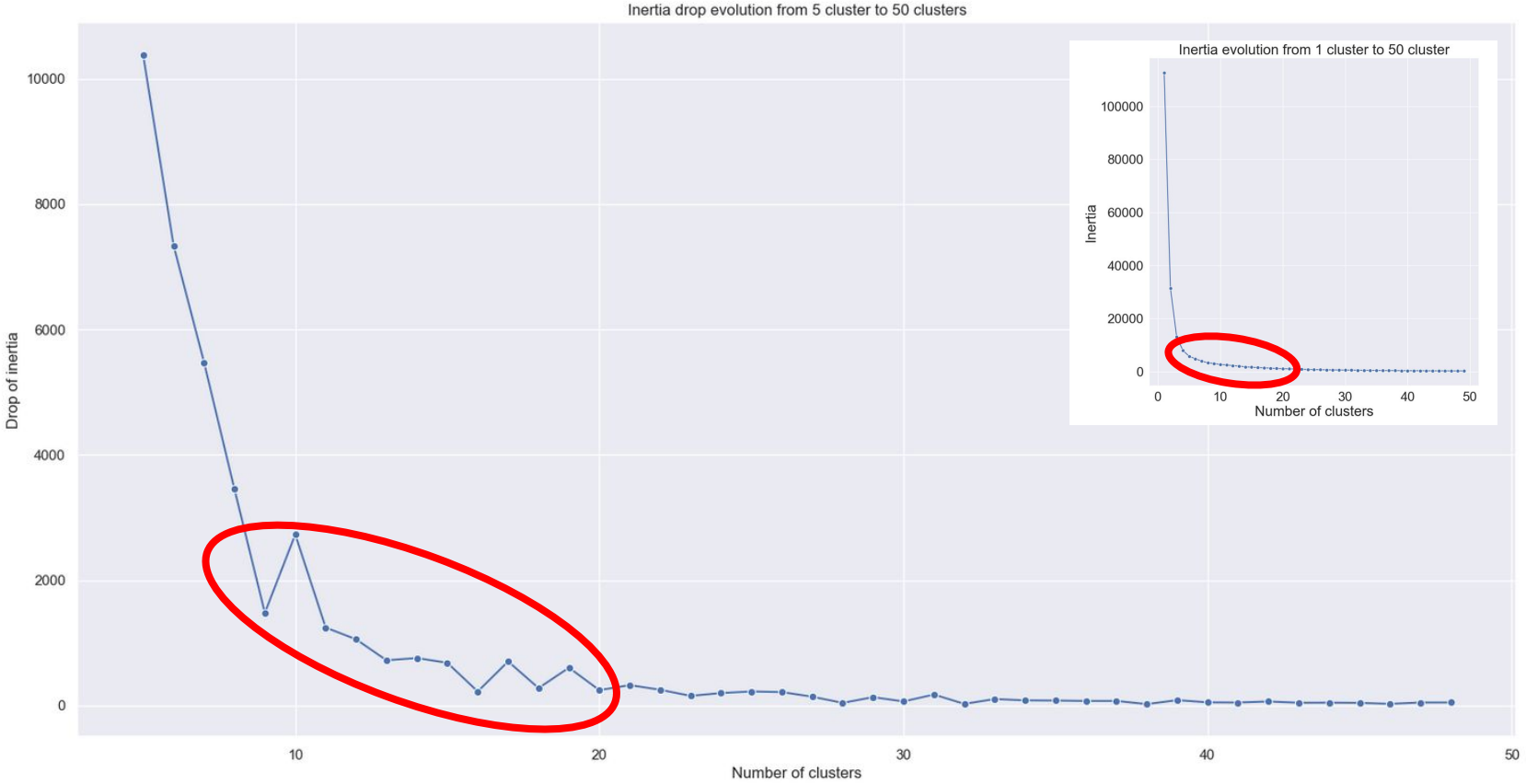
- Are Spotify's audio features able to identify "similar songs", as defined by humanly detectable criteria?
- Is K-Means a good method to create playlists? Would you stick with this algorithm moving forward, or explore other methods to create playlists?

We have been asked to create playlists and keep the size of all the playlists between 50 and 250 songs (between 20 and 100 playlists)

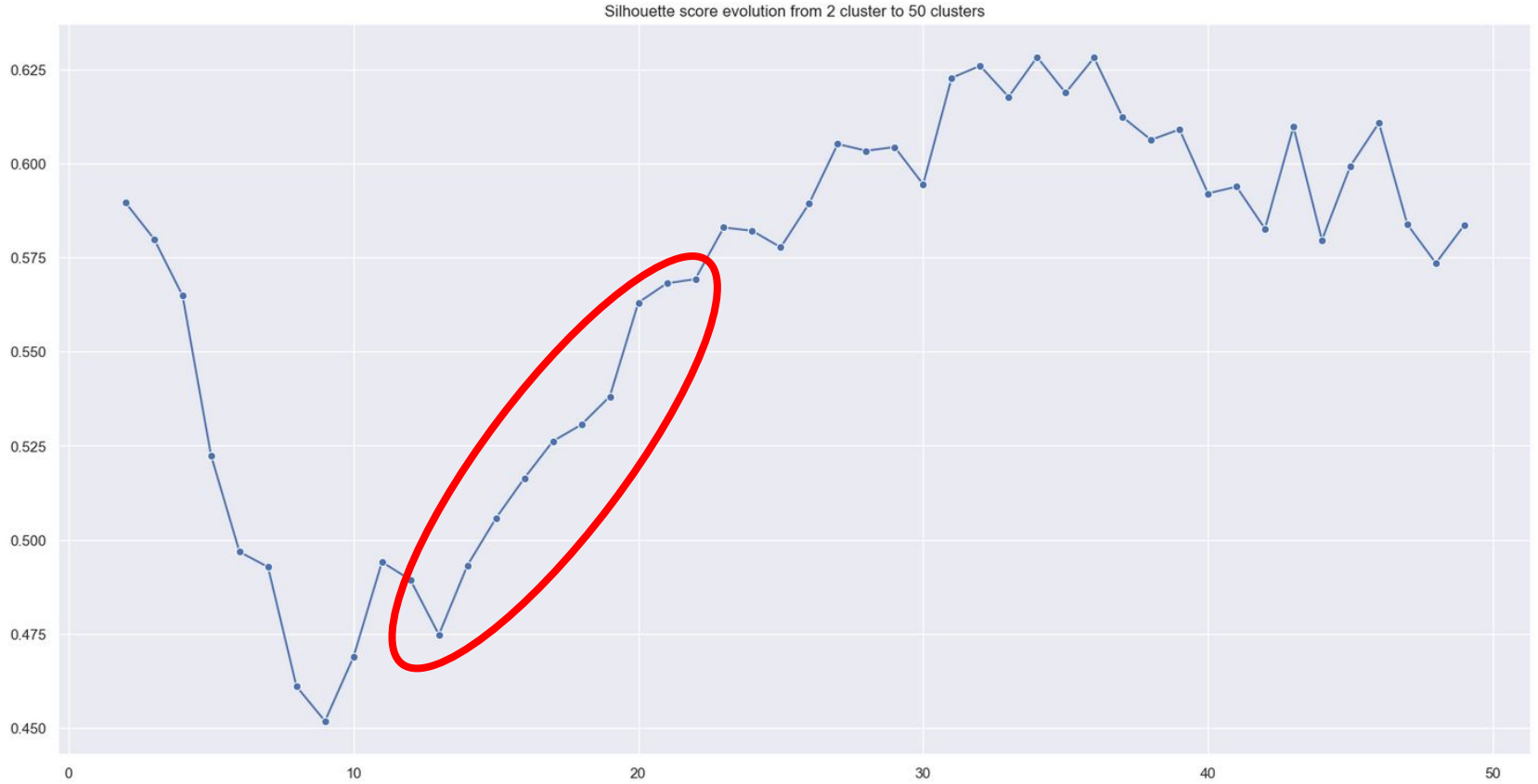
Data

- Number of songs: 5235
- Used audio features:
  - *danceability*
  - *energy*
  - *acousticness*
  - *instrumentalness*
  - *valence*
  - *tempo*
- Scaling Method: Min-max scaler
- Clustering method: K-Means
- Clustering by: style/mood

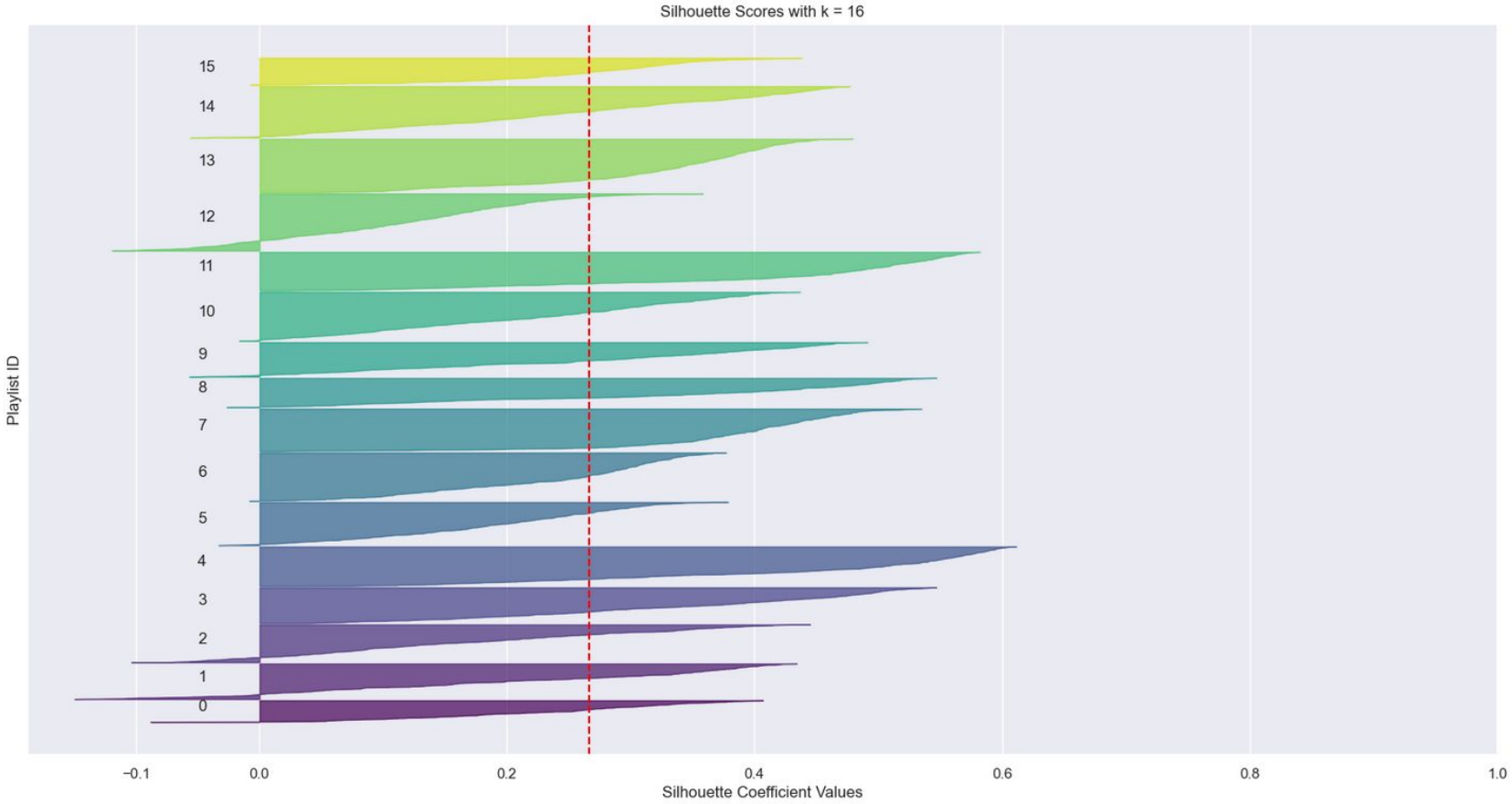
Inertia drop evolution



Silhouette score evolution



Silhouette diagrams



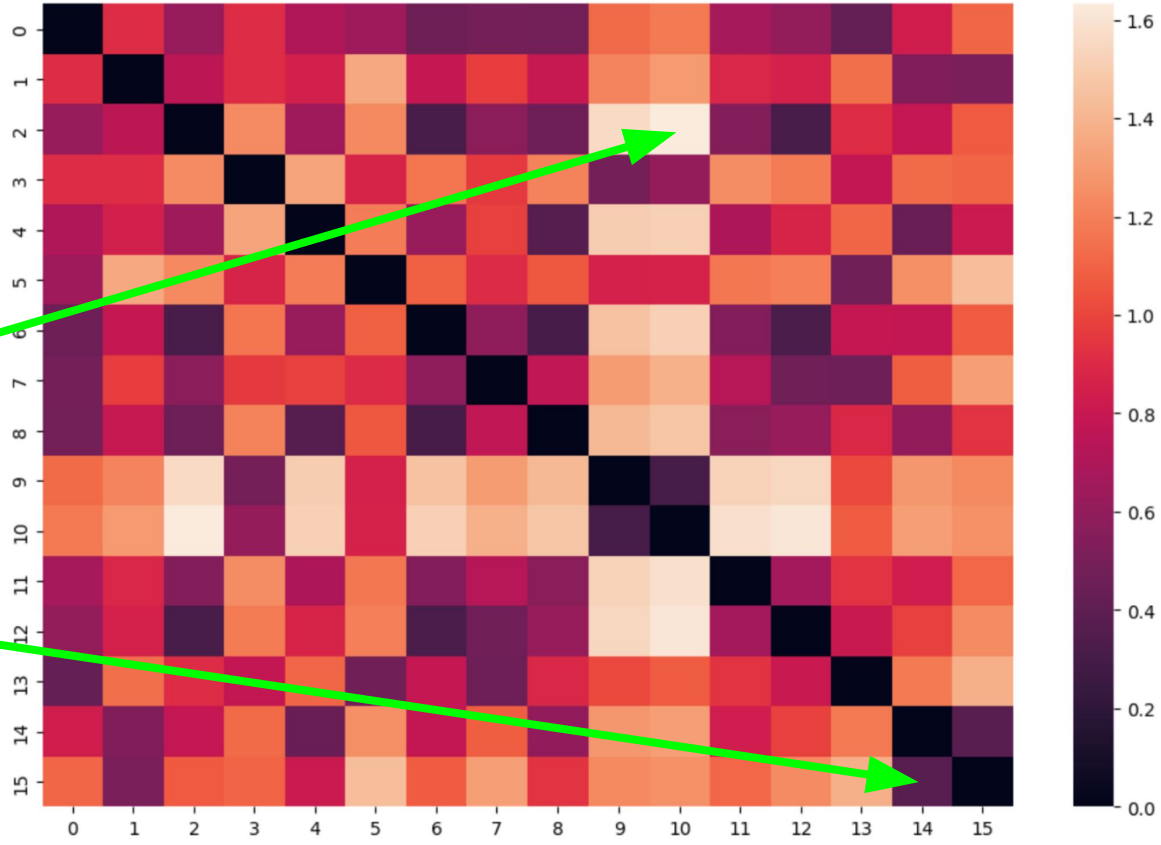
Heat Map for 16 playlists

What are we looking at?

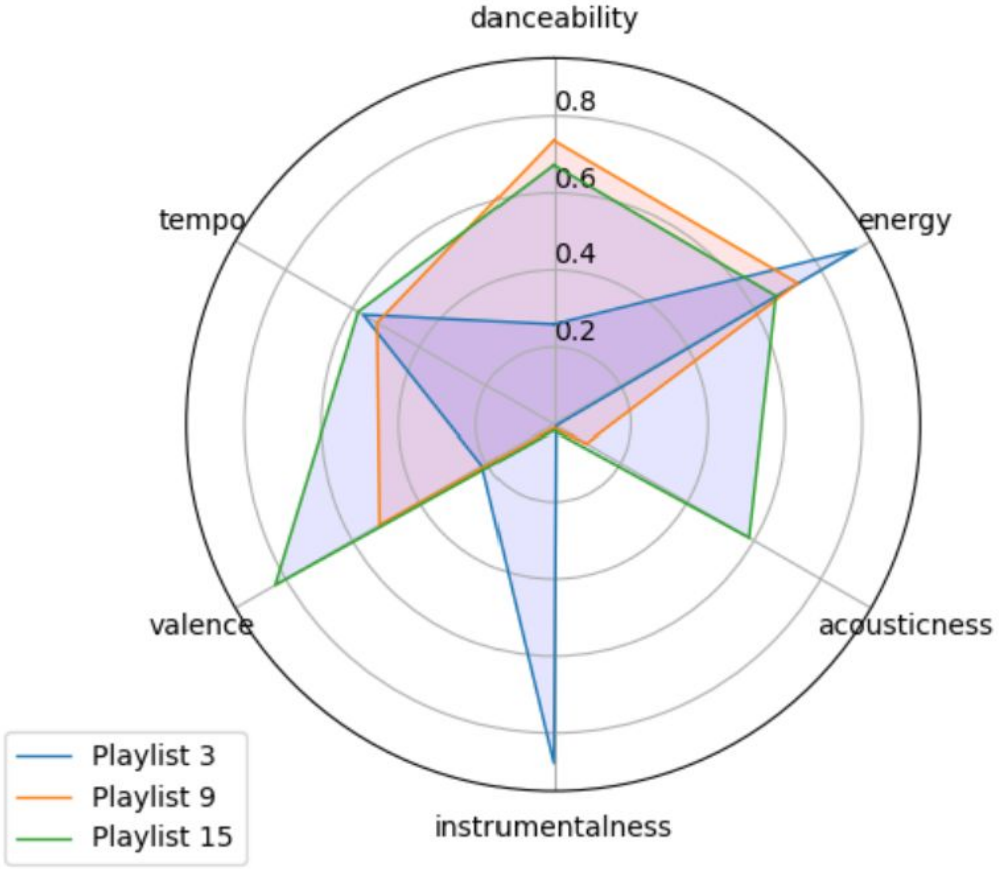
The centroids of each clusters, compared to each other, pairwise

lighter shades

darker shades



Spider Chart for 6 features



Playlist 3:

Sentenced - When the moment of Death arrives

Exhumed - Septicema (Festering Sphinctral Malignancy)

David Guetta - Sunshine

Playlist 9:

Ed Sheeran - Beautiful People

Eagles - Hotel California

Wu-Tang Clan - Gravel Pit

Playlist 15:

Paul Anka - Diana

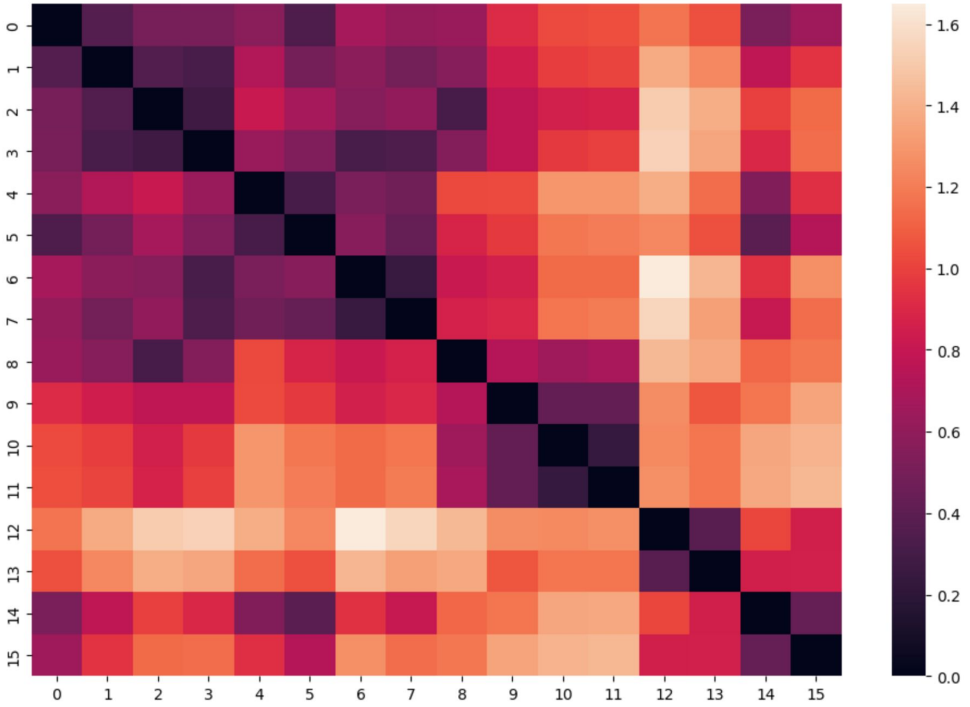
Alligatoah - Willst Du

Britney Spears - Sometimes

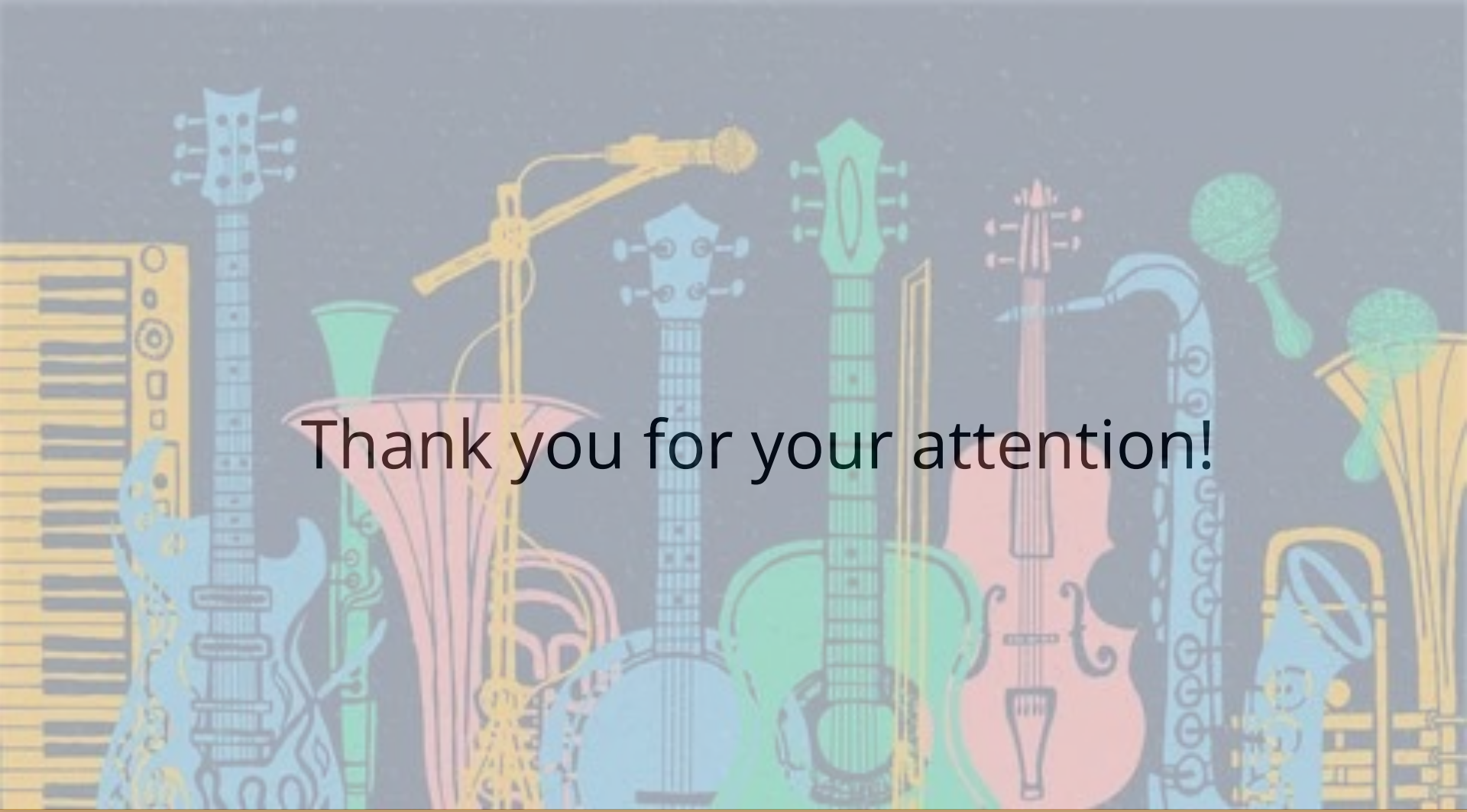
Conclusion

We can preselect playlists, but they have to be refined in a second step:

a) manually
b) subclustering / KMeans or BisectingKMeans
c) start with BisectingKMeans from the beginning (see heatmap)



Heatmap for Bisecting K-Means with "biggest_inertia" strategy


Thank you for your attention!