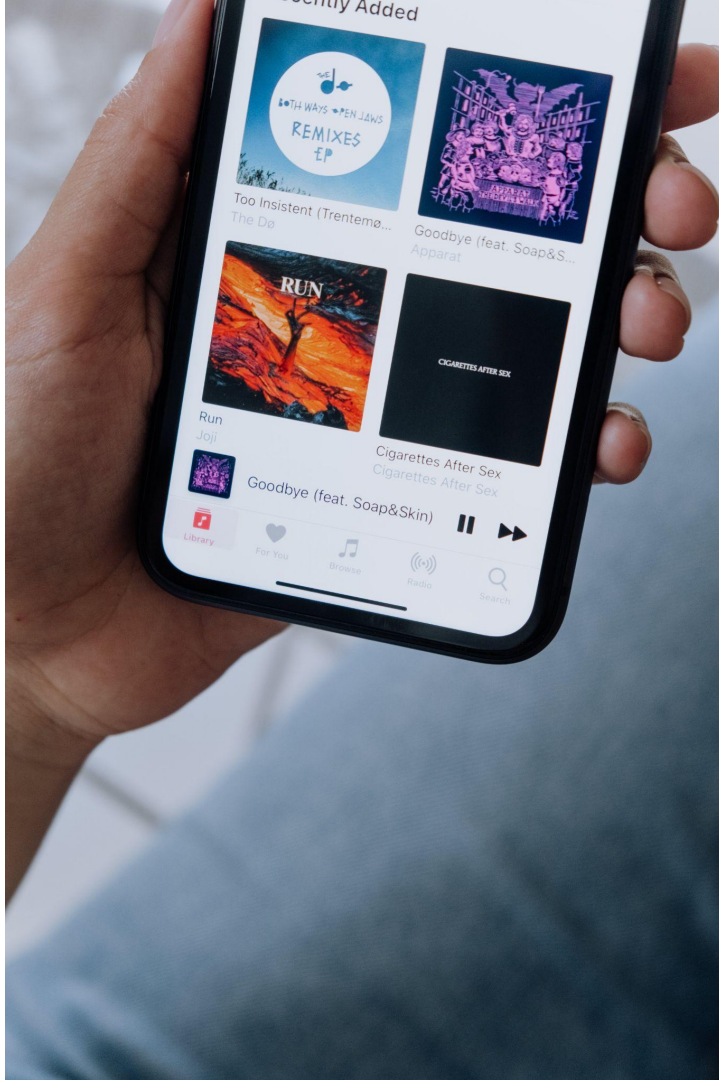
K-Means to cluster?

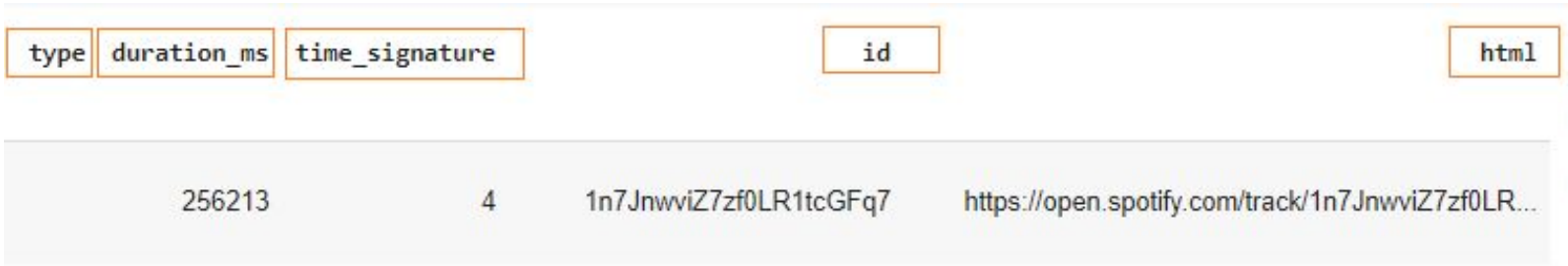
Goal → group songs based on similarities in their audio features.



★ The index is set to **'name'** & **'artist'**

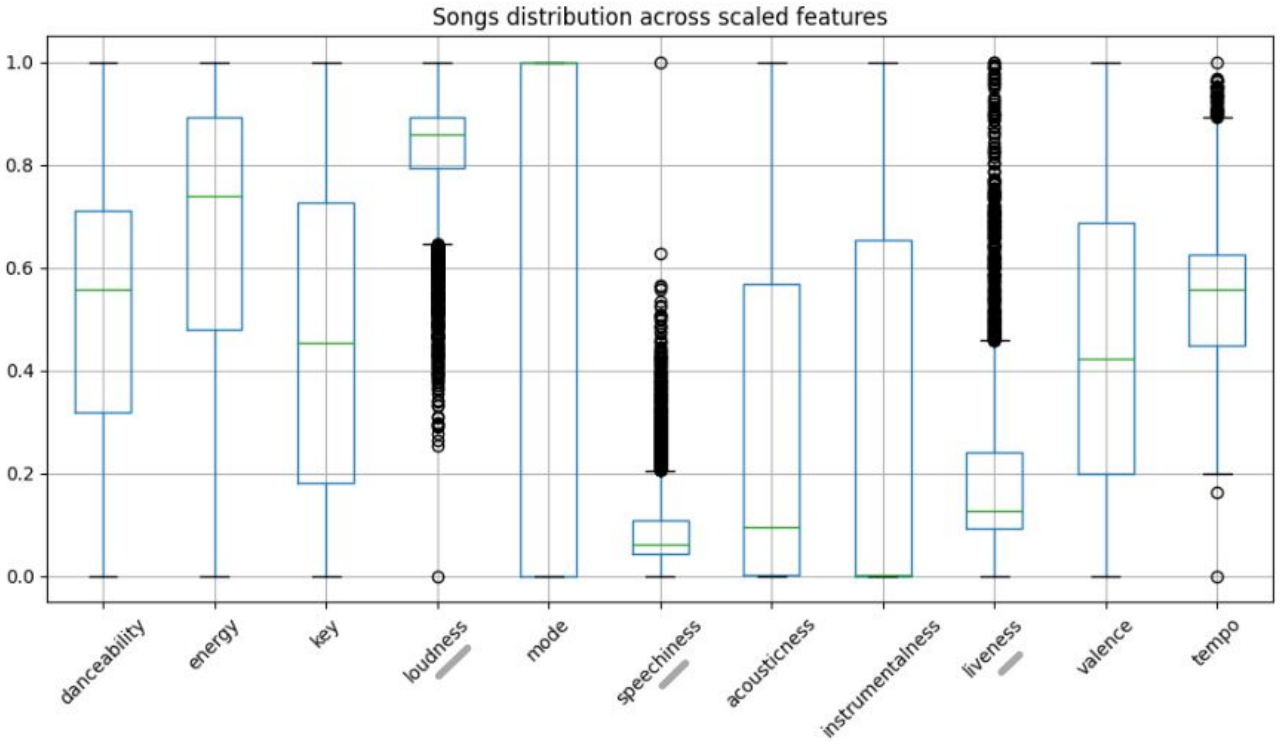
★ The following features have been removed (they do not impact our model).

name	artist	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	type	duration_ms	time_signature	id	html
Se Eu Quiser Falar Com Deus	Gilberto Gil	0.6580	0.25900	11	-13.141	0	0.0705	0.694	0.000059	0.9750	0.3060	110.376	256213	4	1n7JnwviZ7zf0LR1tcGFq7	https://open.spotify.com/track/1n7JnwviZ7zf0LR...	

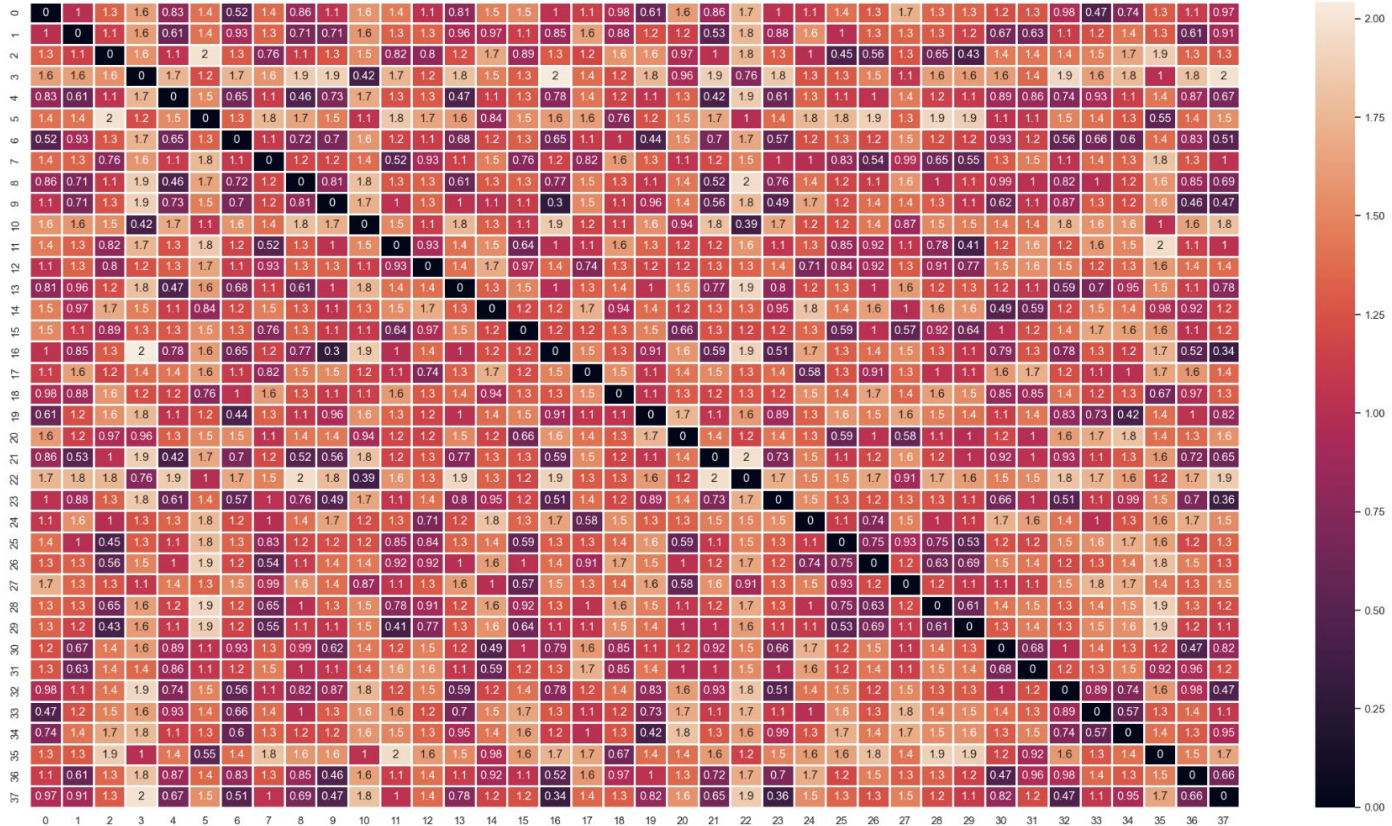


Preprocessing

- Standardising the data
- MinMaxScaler



Pairwise feature comparison(remove)



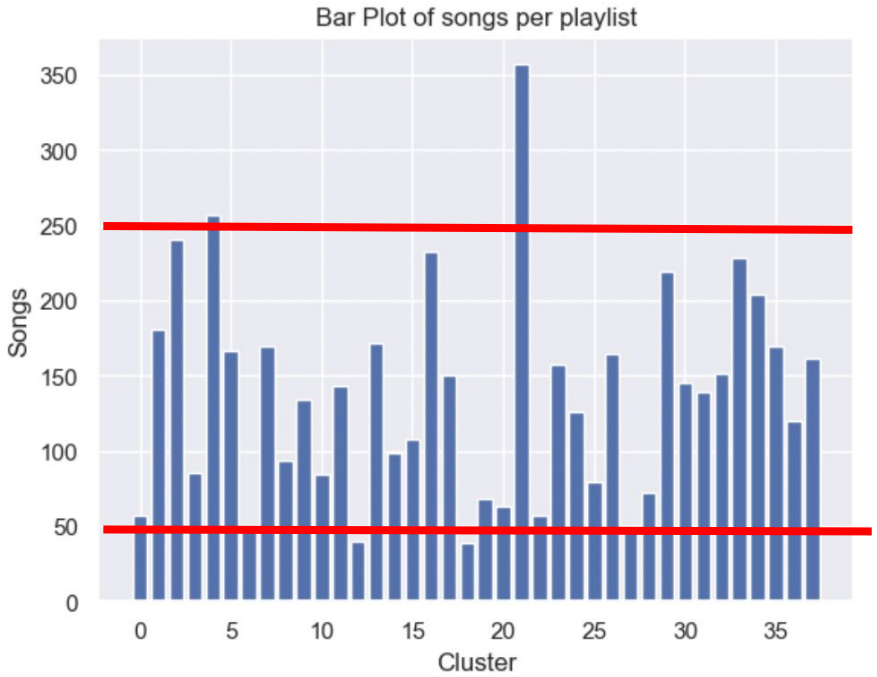
Are Spotify's audio features able to identify "similar songs", as defined by humanly detectable criteria?

Business requirement

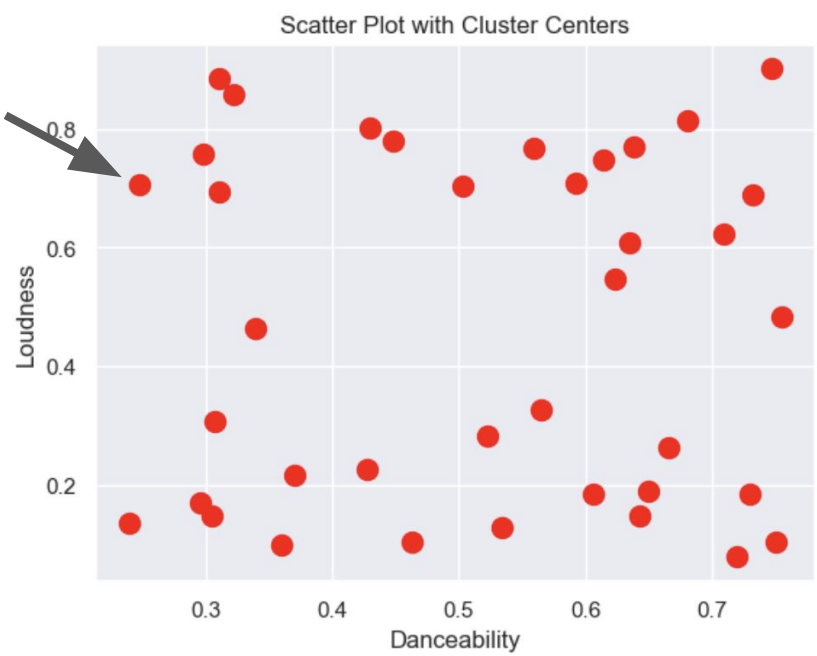
No. of playlists=38

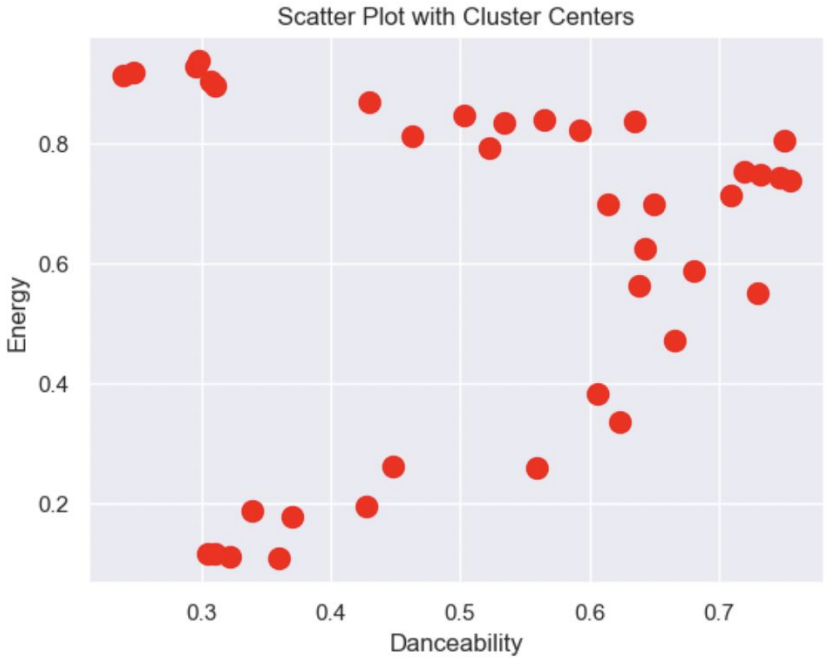
Most playlists contains

(50-250 songs)



Scatter Plot with Playlist Centers



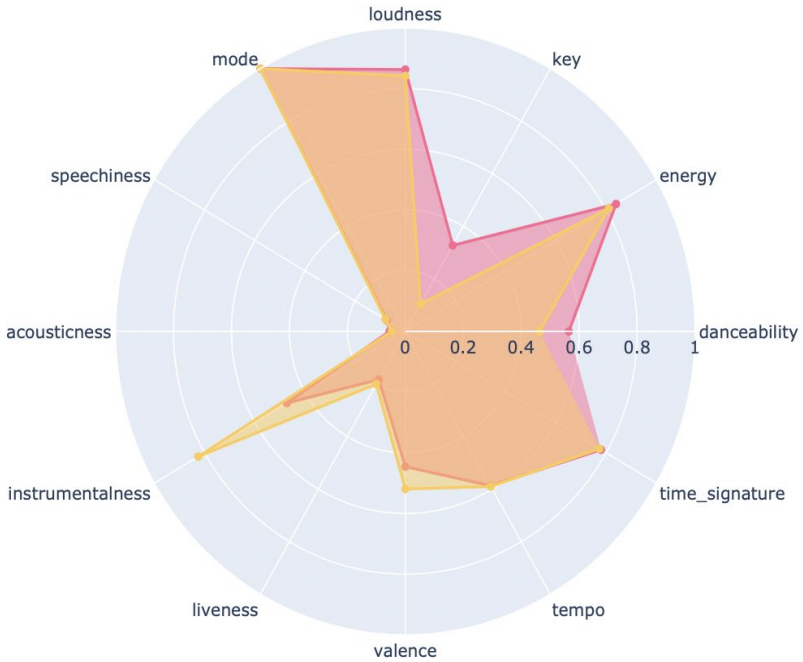


Similar Playlist Comparison

Feature Distribution

Playlists 6 & 19

Song	Playlist
Parabens by Marcos Valle	6
Lets Explode by Clem Snide	19

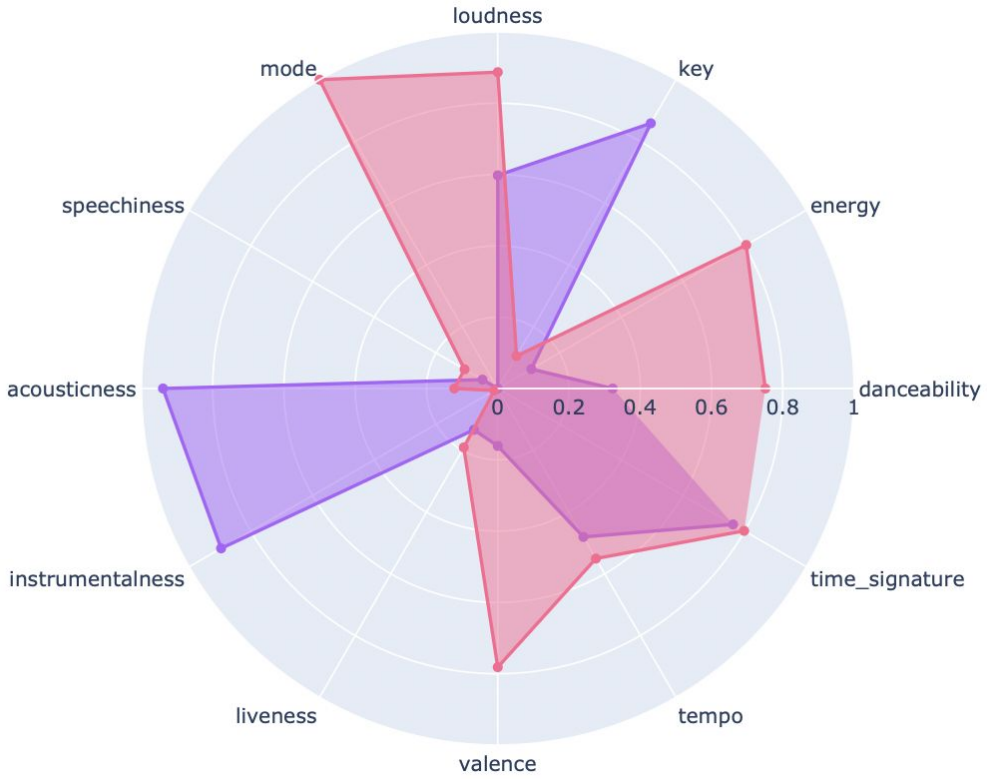


Distinct Playlist Comparison

Feature Distribution

Playlists 3 & 16

Song	Playlist
Brotin by Eydis Evenson	3
light my fire -mono by The Doors	16



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

1. YES, on the basis of data attributes
2. Could not work sometimes, as similarity in music is subjective
3. Important missing categorical information from the features

    a. Genre
    b. preference
    c. Lyrics

4. **Hard to identify songs that are not similar in a playlist**

Is K-Means a good method to create playlists?

Yes, but only to get initial playlists

Drawbacks:

- One song though it suits two playlists will be assigned to only one
- Deals with numerical data and ignores categorical data where human judgement comes into picture
- No way to test the playlists automatically
- Doesn’t deal with outliers

Conclusion

- ★ K-Means, a good way to make playlists initially based on numerical data!
- ★ Consider other techniques to improve efficiency!
- ★ More algorithms could be explored for better or combined with K-Means!