

Moosic for Motivation!!

Unsupervised Machine Learning: Clustering Songs

The Challenge

Moosic - Startup

Creates curated playlists through music experts & provide playlists to their subscribers via an App

Have a list of songs from a wide variety of styles

Songs have audio features with ratings

Requirements

Using unsupervised ML we need to create playlists(from 5000 songs) that can be shared with users

The playlists should not be too large and should be themed through the audio features

Tasks

Data cleaning of song data set

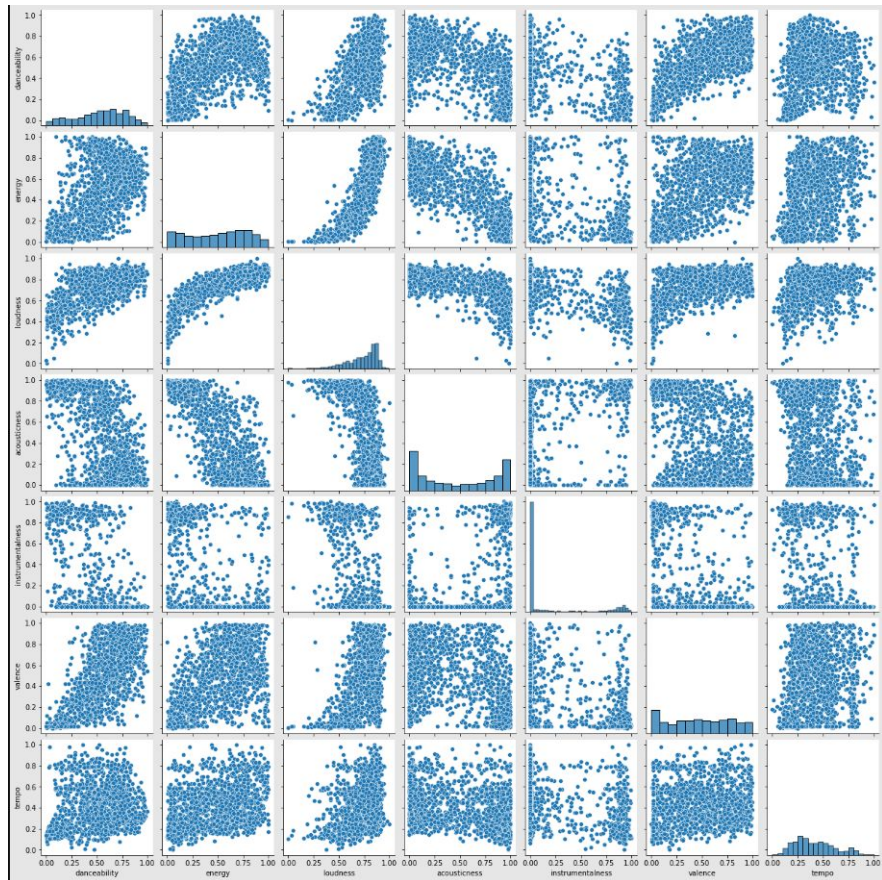
Selection of required columns (e.g. integer or float type data only)

Clustering of songs based on the ratings stated in the audio features to provide new Playlists

Step 1

Familiarisation - we explored:

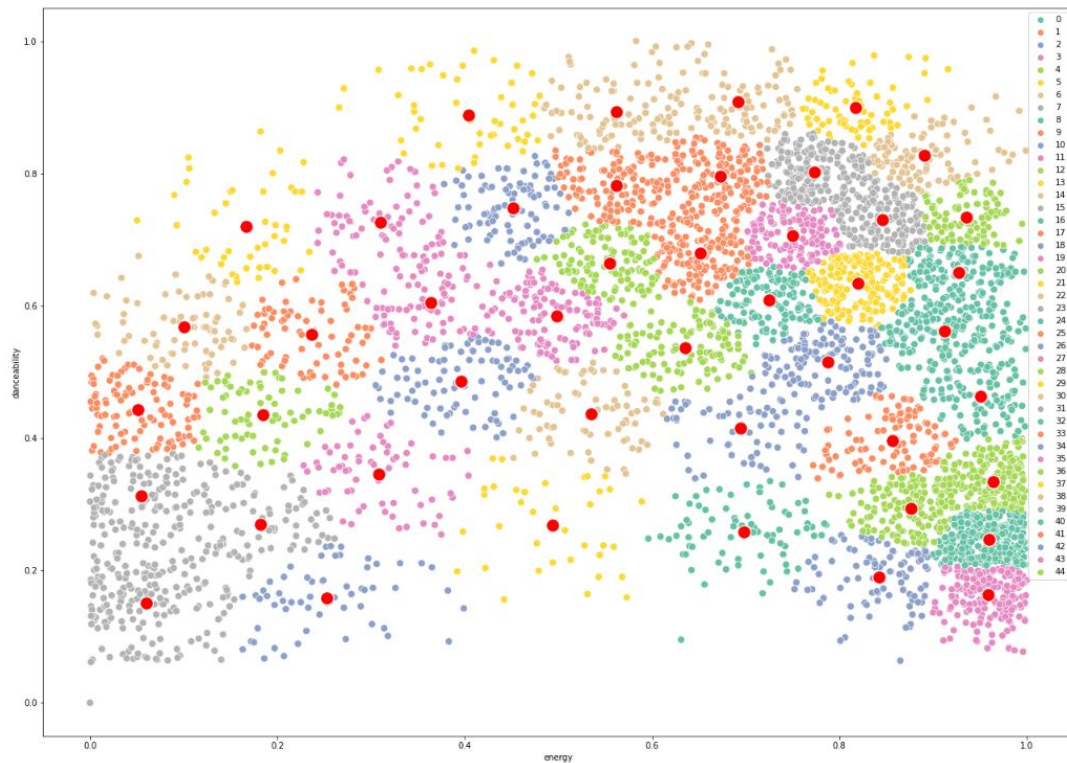
- the audio features of the songs
- calculated Euclidean distance
- used Min Max Scaler as data scaling algorithm
- Pairplotting to see relationships between features
- Noticed that some features such as loudness & energy increased together



Step 2

Which methods to use

- used K-means algorithm to set the number of clusters
- looked at the pairing /grouping of different features and the formation of clusters among them

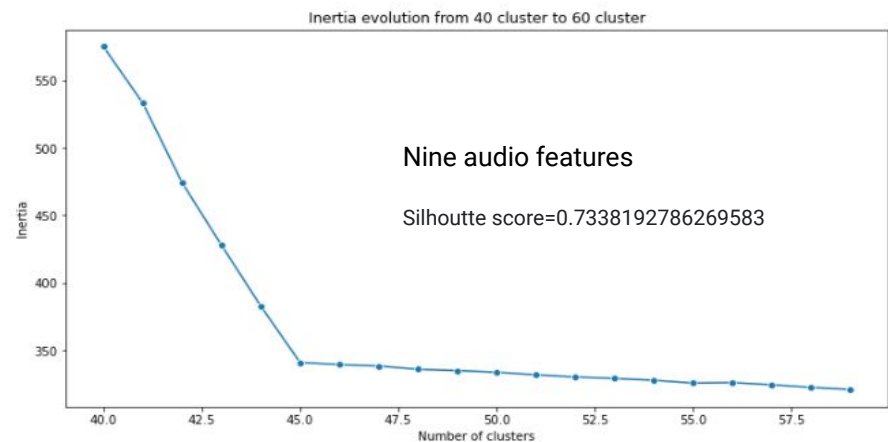
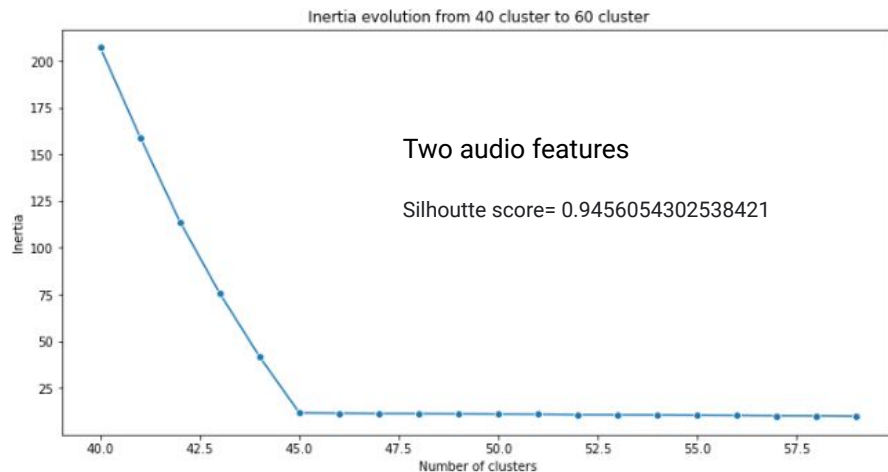


Plot: Two features danceability & energy provide clear clustering

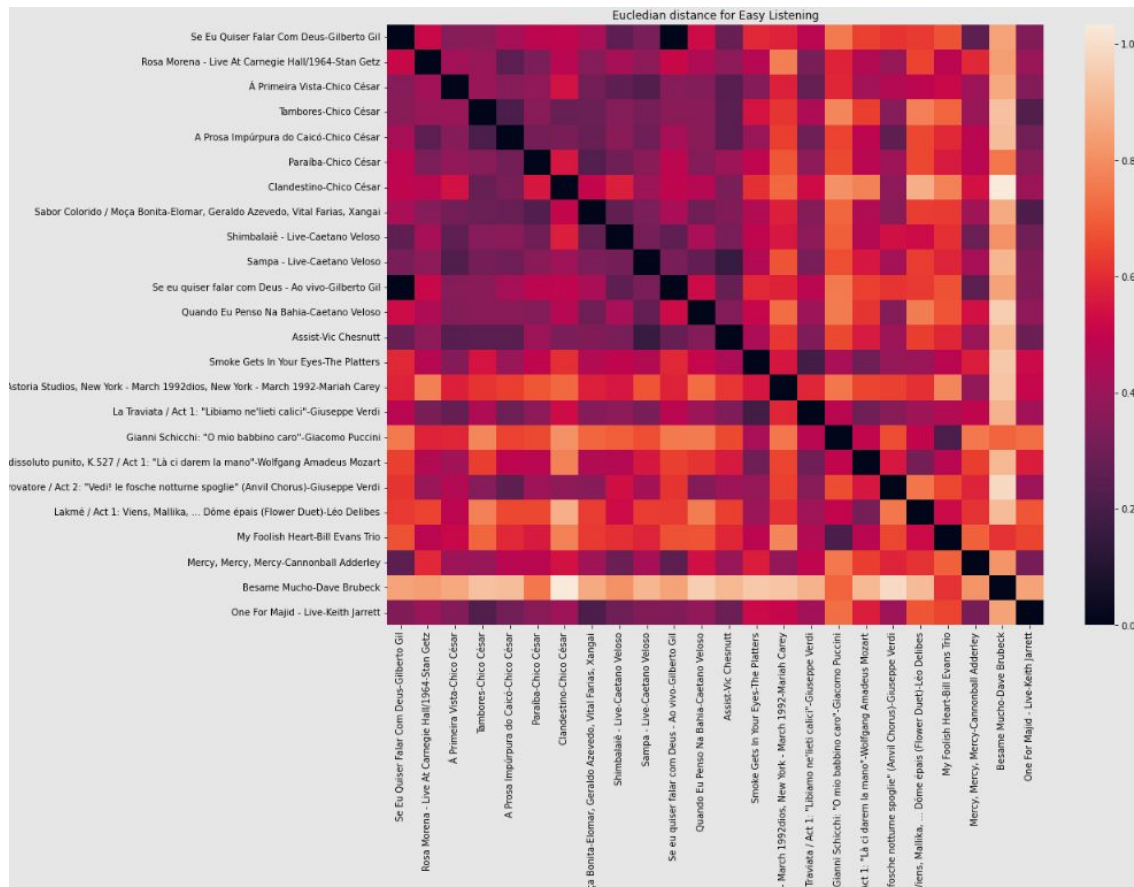
Step 3

Developed our Model

- decided on the main 9 features to use
- applied various K-values to find the best fit for number of clusters (K=45)
- silhouette score helped to confirm K-value was a good fit
- we then checked our clusters for fit against songs and anomalies
- while clusters produced were not distinct we also found similarities



How did we check our playlists?



Easy Listening

Plotted the cluster on heatmap

Searched for any anomalies - here we have Besame Mucho, Dave Brubeck - it has lowest tempo value therefore may be highlighting it

My Foolish Heart, Bill Evans Trio - is that we also have other instrumentals

Clandestino, Chico Cesar - speechiness has a low rating although the song has full lyrics, maybe because the voice blends with the music

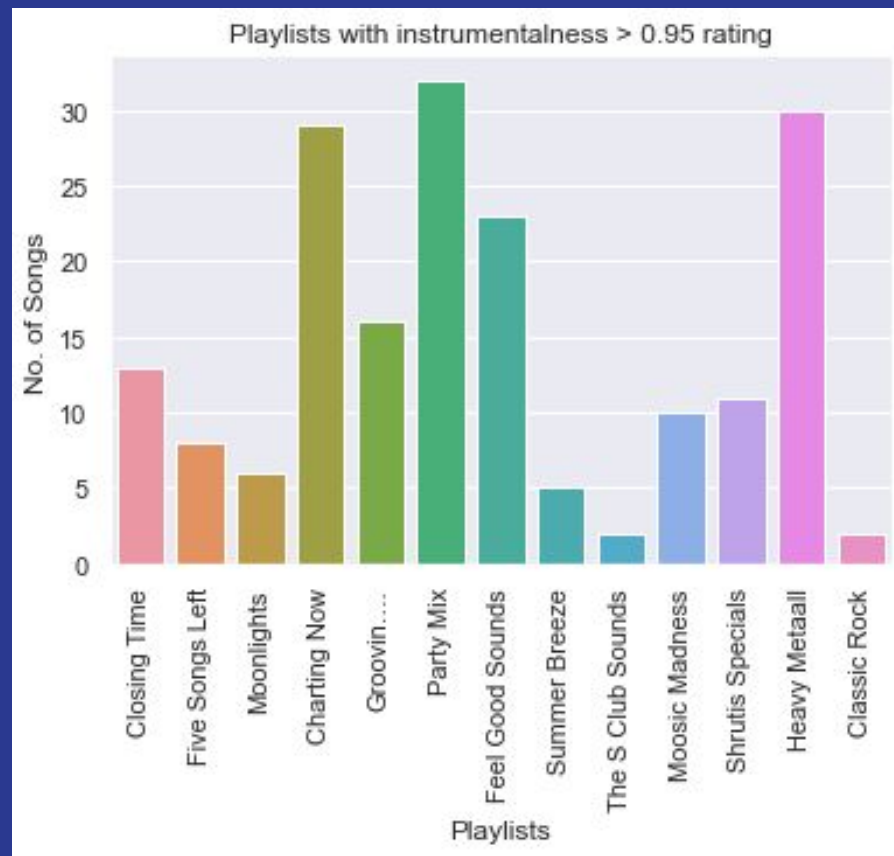
Our Playlists

- have a silhouette score of 0.733819
- we have 45 playlists
- mean of 113 songs per playlist
- manually named after listening to a selection of songs

Playlists	Songs No.	Playlists	Songs No.	Playlists	Songs No.
Acoustic Blues	162	Five Songs Left Folk, Blues & Beyond	91	Party Bash	128
Amazing Arias	48	Groovin....	236	Party Mix	80
<u>BeatBox</u>	105	Heartfelt	70	Play it by Ear	65
<u>Besta Bosa</u>	141	Heavy Metaall	150	<u>Poppin</u> Pop	117
Big Band Favs	78	Martins MashUp	75	Power Ballads	120
Charting Now	151	Melody Moods	52	<u>Runnin</u> Live	24
<u>Chillin</u>	48	Mighty Mix	158	Season Specials	159
Classic Rock	155	Moonlights	129	<u>Shrutis</u> Specials	54
Closing Time	74	Moosic Madness	86	Songs in the key of Z	100
Dance the Night Away	116	Movie Madness	124	Sounds of Summer	43
<u>Dancin</u> on the Ceiling	109	Night Clubbin	128	Sounds of the South	130
Dreamy Themes	44	Night Train	125	Summer Breeze	83
Drive Time Sounds	81	Out on the Weekend	210	Sunday Morning	153
Easy Listening	159	<u>Outta</u> this World	130	Techno Hops	165
Feel Good Sounds	176		137	The S Club Sounds	145

Learnings from the project

- We won't always get a complete fit based on the Audio features provided using unsupervised ML
- Exploring other methods that give improved categorisation



Conclusion

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

- Partially but ...
- Not guaranteed to cluster distinctly like humans

Is K-Means a good method to create playlists?

- No
- KMeans algorithm is not an optimum way of clustering
- the more features the less distinct the clusters become