



Group 2
Kayla Malik, Orjada Mucollari,
Tobias Aurer, Daniel Danielides

Agenda

- Overview
- Data
- K-Means
- Choice of Features
- Number of Clusters
- Evaluation

Overview



Can Algorithms categorize Music?

Is K-Means a good method to create playlists?



Data

- Number of Songs:
 - 5235
- Scaling Method:
 - Quantile Transforming to Gaussian Distribution



Data Preparation

- Importing
- Cleaning
- Feature-Choice



Modelling

- Scaling
- Finding best k
- Creating final clusters



Cluster Exploration

- Exploring descriptive stats of clusters
- Manually label clusters

K-Means Clustering

- K-means clustering is one of the simplest and popular unsupervised machine learning algorithms
- Given k , the k-means algorithm is implemented in four steps:
 - Partition objects into k nonempty subsets
 - Compute seed points as the centroids of the clusters of the current partition (the centroid is the center, i.e., mean point, of the cluster)
 - Assign each object to the cluster with the nearest seed point
 - Go back to step 2, stop when no more new assignments.

"Without music, life would be a mistake."
- Friedrich Nietzsche



Features' overview



Tempo:

The overall estimated tempo of a track in beats per minute (BPM).



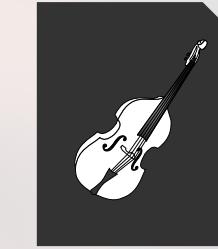
Valence:

Tracks with high valence sound more cheerful and euphoric, while tracks with low valence sound more sad and depressed.



Speechiness:

The more exclusively speech-like the recording the closer to 1.0 the attribute value.



Acousticness:

A measure from 0.0 to 1.0 of whether the track is acoustic.



Energy:

A perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy.



Loudness:

Averaged values across the entire track. Values typical range between -60 and 0 db.



Danceability:

Based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.

Number of Clusters

Level One:

Inertia and the Elbow Method

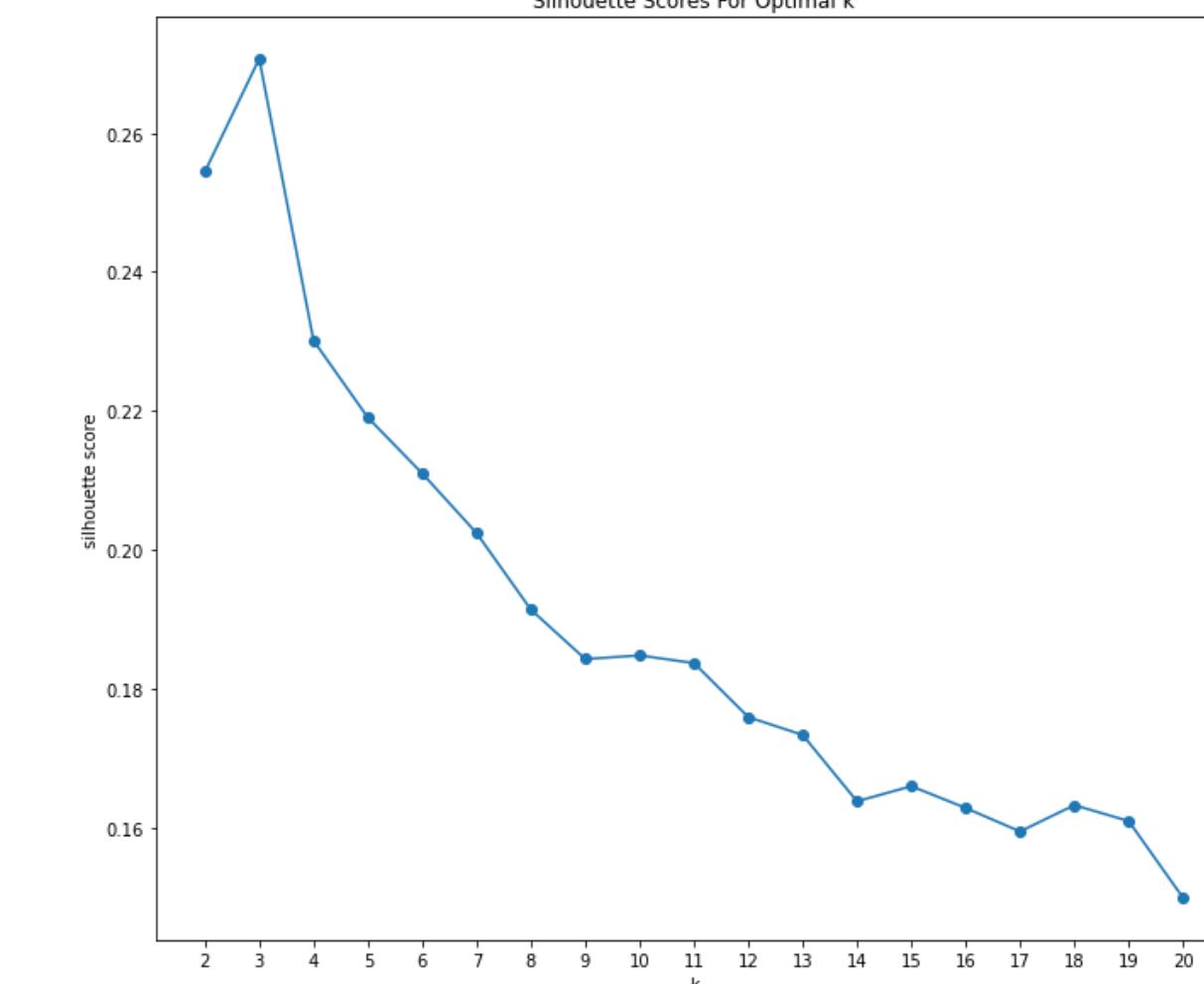
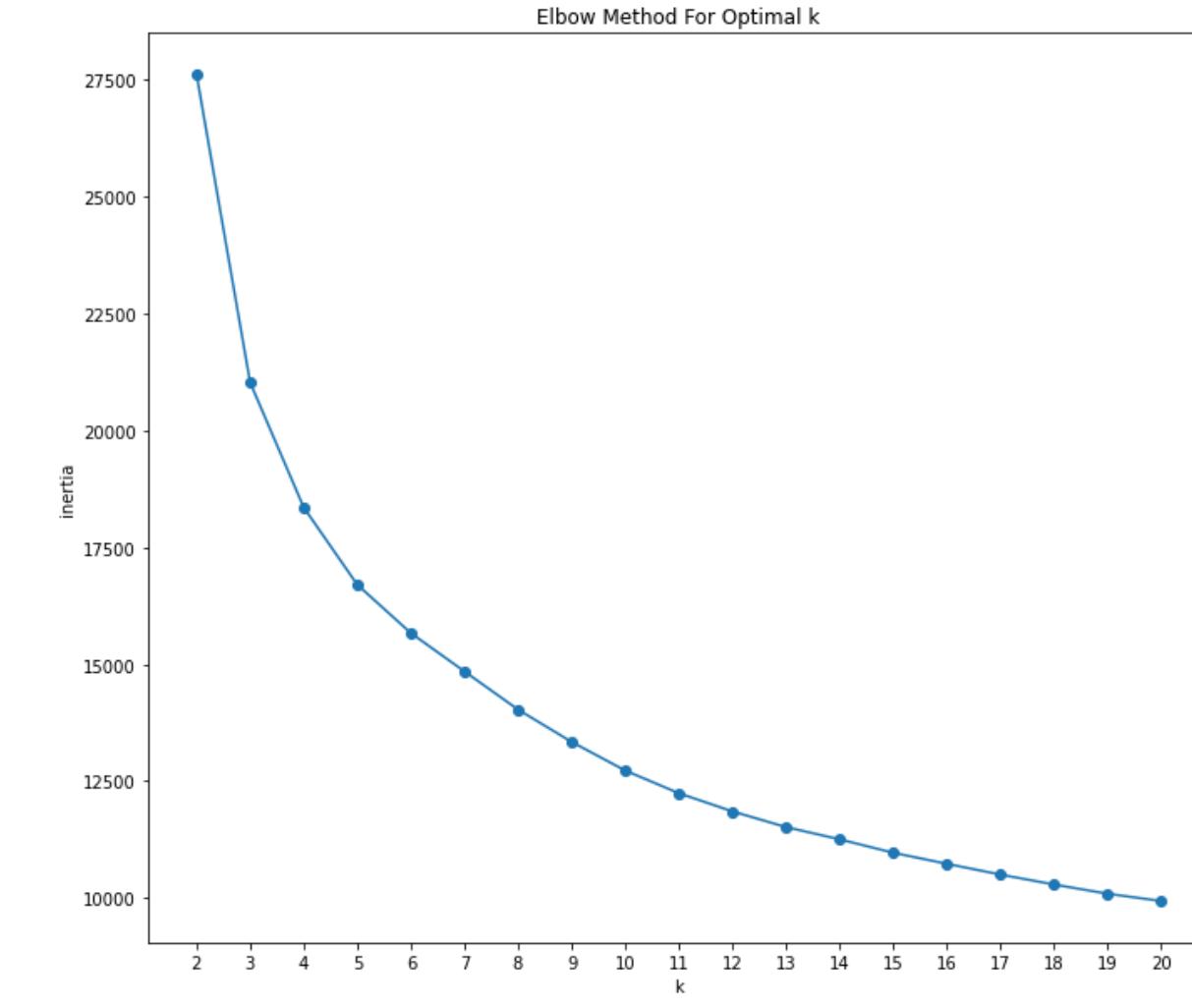
- $k = 3$ is chosen for the first layer

The Silhouette Coefficient

- $k = 3$ would be optimal

Level Two:

Split into further clusters depending on the Elbow and Silhouette Scores



Different example playlists

PLAYLIST

Rock/Reggae

70s and 80s new wave, pop rock, and reggae rock

Kayla • 30 songs, 1hr 48 min

TITLE ALBUM

#	TITLE	ALBUM
1	Just Can't Get Enough Depeche Mode	Speak And Spell (Deluxe)
2	Cracklin' Rosie - Single Version Neil Diamond	Tap Root Manuscript
3	Valerie Steve Winwood	Talking Back To The Night
4	Burning Love Elvis Presley	Elvis (Fool)
5	I'm A Man Bo Diddley	Bo Diddley
6	Together In Electric Dreams The Human League	The Greatest Hits
7	Only the Lonely Roy Orbison	Sings Lonely and Blue
8	I Am a Rock Simon & Garfunkel	Sounds Of Silence

+30
songs/category

PUBLIC PLAYLIST

Weird Mix

Playlist with songs from Cluster 6

Tobias Aurer • 30 songs, 1 hr 55 min

II ⏪ ⏴ +2 ...

#	TITLE	ALBUM	DATE ADDED
1	You Can Call Me Al Paul Simon	Graceland (25th Anniversary Deluxe Edition)	3 hours ago
2	The Logical Song - Remastered 2010 Supertramp	Breakfast In America (Deluxe Edition)	3 hours ago
3	Ritual Tiësto, Jonas Blue, Rita Ora	Ritual	3 hours ago
4	Maneater Nelly Furtado	Loose	3 hours ago
5	Baby Jane - 2008 Remaster Rod Stewart	Body Wishes	3 hours ago
6	Bennie And The Jets - Remastered 2014 Elton John	Goodbye Yellow Brick Road (Remastered)	3 hours ago
7	Tie a Cherry CL	ALPHA	3 hours ago
8	LALISA LISA	LALISA	3 hours ago
9	Dancing On The Ceiling Lionel Richie	Dancing On The Ceiling (Expanded Edition)	3 hours ago
10	Du hast den schönsten Arsch der Welt Alex C., Yass	Euphorie	3 hours ago
11	Don't Bring Me Down Electric Light Orchestra	Discovery	3 hours ago

Evaluation

General categorization works

Overall, the algorithm achieves somewhat reasonable playlists.

Disadvantages

However, the algorithm adds completely unfitting songs to some playlists with regularity.

Further Experimentation

- Using different features.
- Applying a different clustering algorithm (e.g. K-medoids).
- Increase / Decrease layers.
- Use information like artist/genre to improve clustering

Thank you for your attention

"The time is gone, the song is over,
thought I'd something more to say"
- Pink Floyd

