Smart
Playlist creation

.. Scale it Up!

Is K-Means a good method for creating playlists?

YES, to an extent

Is K-Means a good method for creating playlists?

✅K-Means can create cohesive playlists by grouping songs with similar numerical audio characteristics

❌But it has limitations when it comes to musical nuance and user taste.

Is K-Means a good method for creating playlists?

Pros

✅ **Easy to Implement.**

✅ **Scalable**

✅ **Fast Convergence .**

Cons

❌ **Needs Predefined K**

❌ **Sensitive to Initialization**

❌ **Doesn’t Handle Non-Spherical Shapes Well**

❌ **Sensitive to Outliers**

What is the right amount of playlists?

Inputs

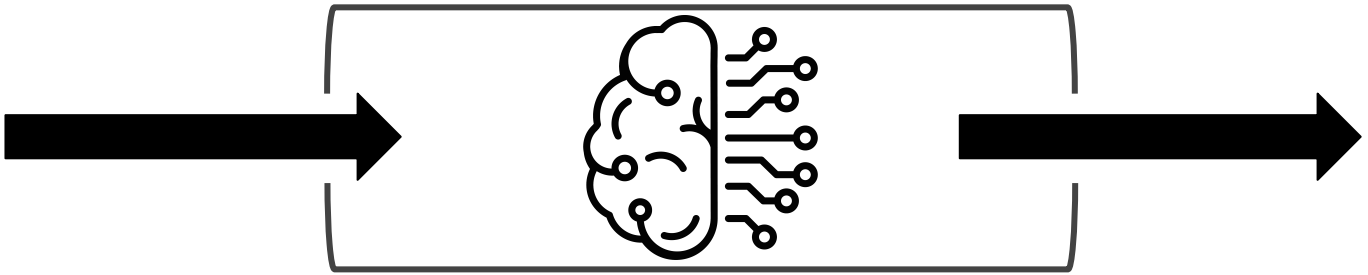
- List of 5235 songs
- 50 -250 songs/playlist
- Numerical features

Preprocessing

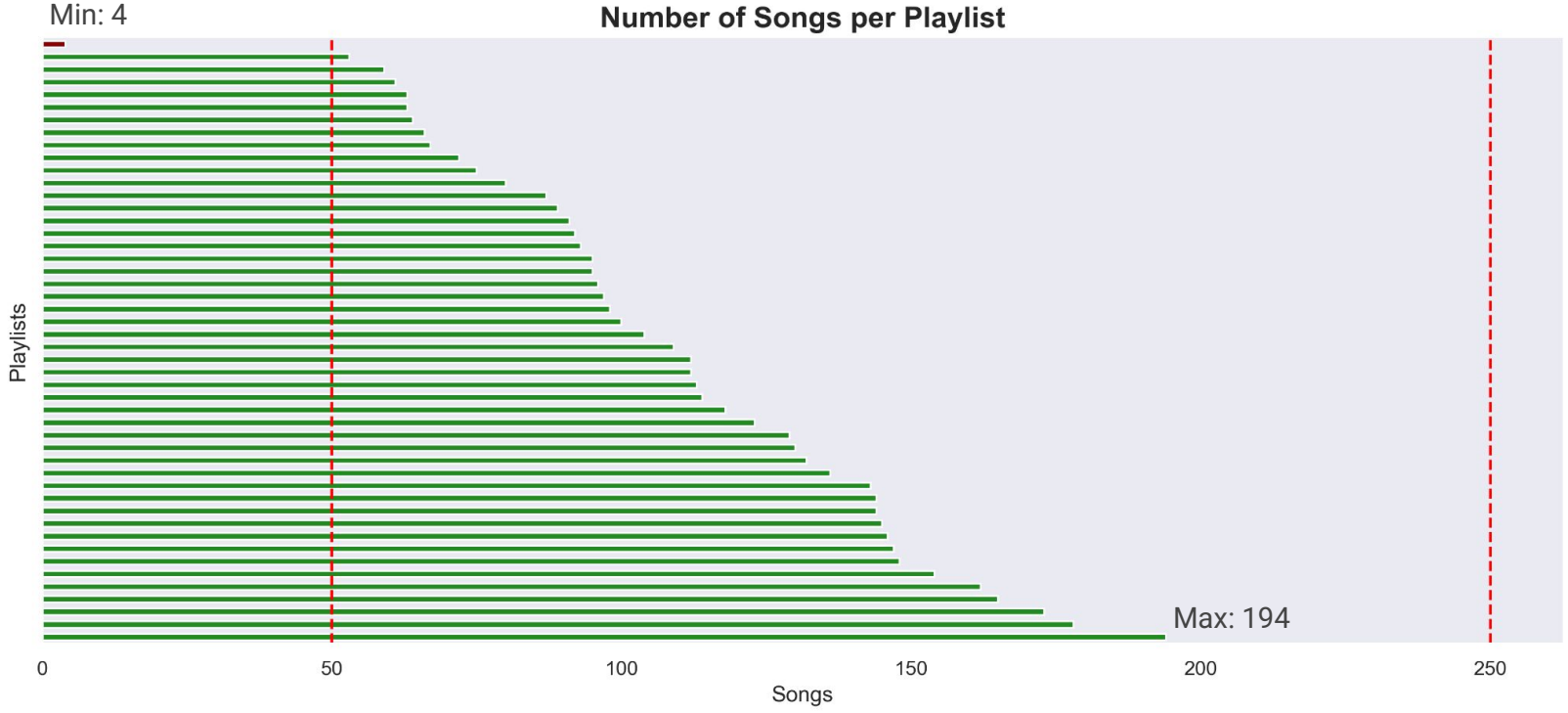
- Data scaling
- Principal Component Analysis (PCA)
- Inertia & silhouette scores
- Knee Locator

Final number of playlists

48



What is the right amount of playlists?



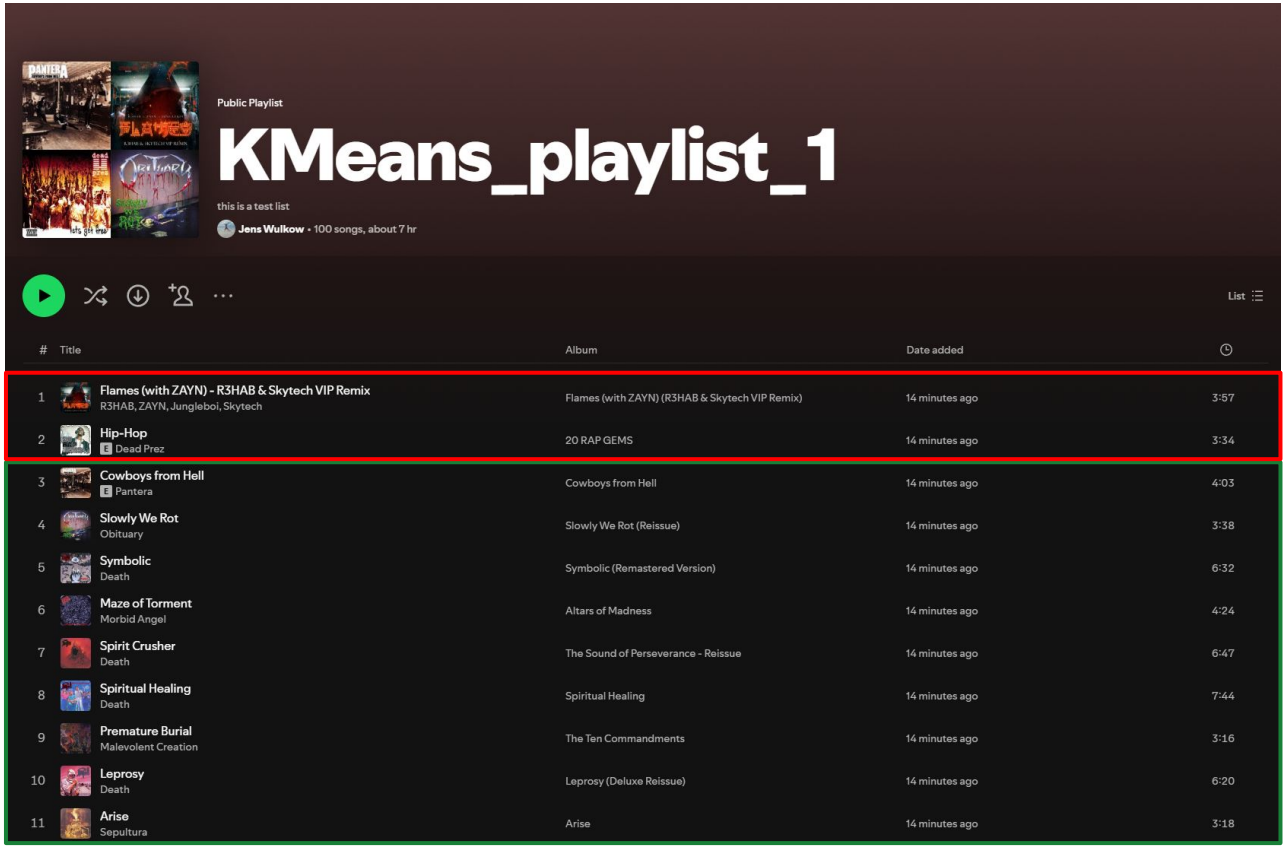
challenges

🧩Difficulty Finding the "Perfect" Number of Clusters (k)

🧩Some Playlist Sizes Not Within Desired Range

🧩No Overlapping Songs (Songs Only in One Cluster)

Are the Playlists cohesive?



Are the Playlists cohesive?

KMeans_playlist_2

#	Title	Album	Date added	
35	No Diggity Blackstreet, Dr. Dre, Queen Pen	No Diggity: The Very Best Of Blackstreet	18 minutes ago	5:05
36	Started From the Bottom E Drake	Nothing Was The Same (Deluxe)	18 minutes ago	2:54
37	Always On Time E Ja Rule, Ashanti	Pain Is Love	18 minutes ago	4:05
38	Bottoms Up (feat. Nicki Minaj) E Trey Songz, Nicki Minaj	Passion, Pain & Pleasure	18 minutes ago	4:02
39	Big Poppa - 2005 Remaster E The Notorious B.I.G.	Ready to Die (The Remaster)	18 minutes ago	4:12
40	Money Maker E Ludacris, Pharrell Williams	Release Therapy	18 minutes ago	3:50
41	Shake Ya Tailfeather (feat. Nelly & Murphy Lee) Diddy, Murphy Lee, Nelly	Bad Boys II	18 minutes ago	4:53
42	Just A Lil Bit E 50 Cent	Ultimate R&B Love 2008 (International Version)	18 minutes ago	3:58
43	There It Go (The Whistle Song) E Juelz Santana	What The Game's Been Missing!	18 minutes ago	3:00
44	Shake That E Eminem, Nate Dogg	Curtain Call (Deluxe)	18 minutes ago	4:34
45	All Eyez On Me (ft. Big Syke) E 2Pac, Big Syke	All Eyez On Me	18 minutes ago	5:07
46	Nuthin' But A "G" Thang E Dr. Dre, Snoop Dogg	The Chronic	18 minutes ago	3:58
47	Nein Mann - Radio Edit Laserkraft 3D	SLAM! FM Presents Grand Slam! 2009 Volume 3 & 4	18 minutes ago	3:48
48	Hey Mama (feat. Nicki Minaj, Bebe Rexha & Afrojack) David Guetta, AFROJACK, Bebe Rexha, Nicki Minaj	Listen	18 minutes ago	3:12
49	Uptown Top Ranking - Remastered 2001 Althea And Donna	Uptown Top Ranking	19 minutes ago	3:49
50	007 (Shanty Town) Desmond Dekker	This Is Crucial Reggae - Desmond Dekker	19 minutes ago	2:33
51	Engine 54 The Ethiopians	Let's Do Rocksteady: The Story Of Rocksteady 1966-68	19 minutes ago	2:37
52	This Is Reggae Music Zap Pow	Reggae Chartbusters, Vol. 4	18 minutes ago	3:01

Are we capturing the **human** idea of **“similarity”** with spotify’s features?

Generally, we get **fairly cohesive** playlists, but sometimes we also encounter **really bad** matches.

What kind of **data** might help us create **better playlists**?

To detect **genre** or other **human** ideas of similarity, we could obtain labeled data from existing **human**-made Spotify playlists

- By scoring song pairs based on how often they co-occur—especially as neighbors—we can define a **playlist cohesion score** for newly generated playlists.
- This can serve as a cost function for **supervised ML**.

Next Steps

- **Exploring Other Algorithms**: HDBSCAN
- Adding **selection** layers to throw out misfits:
  - Chat-GPT can judge playlists quite efficiently
  - Some **humans**, too
- **Supervised Learning for Genres**
  - Train a model using **labeled data**

Thank You for Your attention!