Can machine learning help Moosic scale up operations?

Data

Requirements

- 50 – 250 songs playlists

- 5.235 songs ≈ 20 – 100 playlist clusters

Analysis
Cleaning

danceability
energy
key ✕
loudness
mode ✕
speechiness
acousticness
instrumentalness
liveness
valence
tempo
type ✕
duration_ms ✕
time_signature ✕
id ✕

Clean Dataset:
5.159 songs, 9 columns

Design

- Data scaling: Quantile Transformer
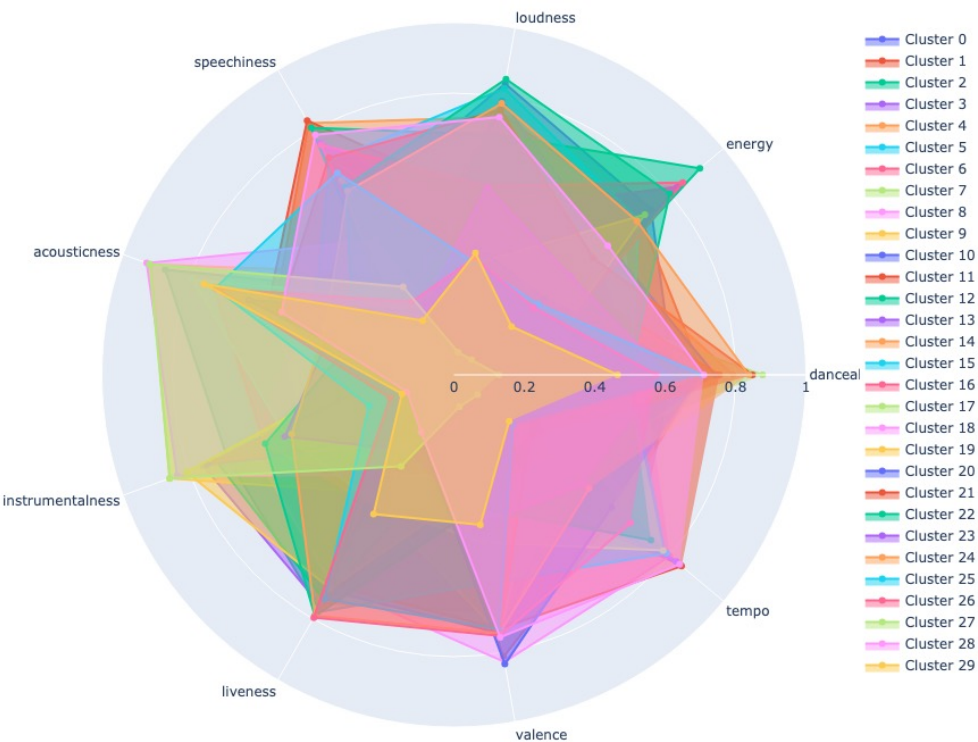
- Songs clustering: KMeans

Prototype

- 30 playlists

- Average 172 songs
  - Min 112
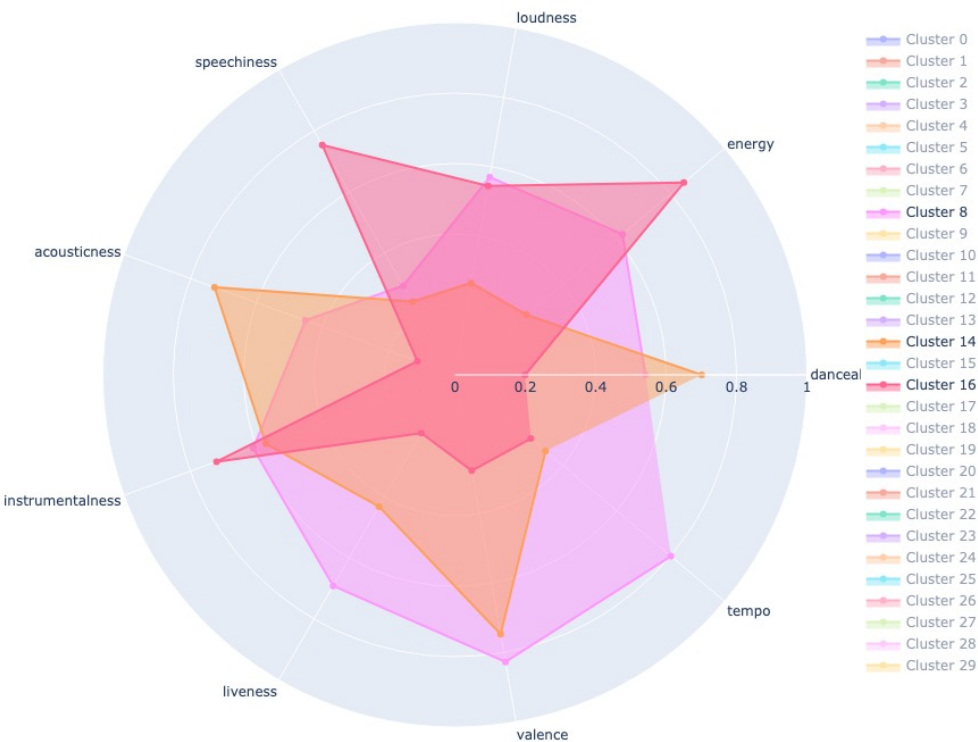 - 25% 137
  - 50% 162
  - 75% 209
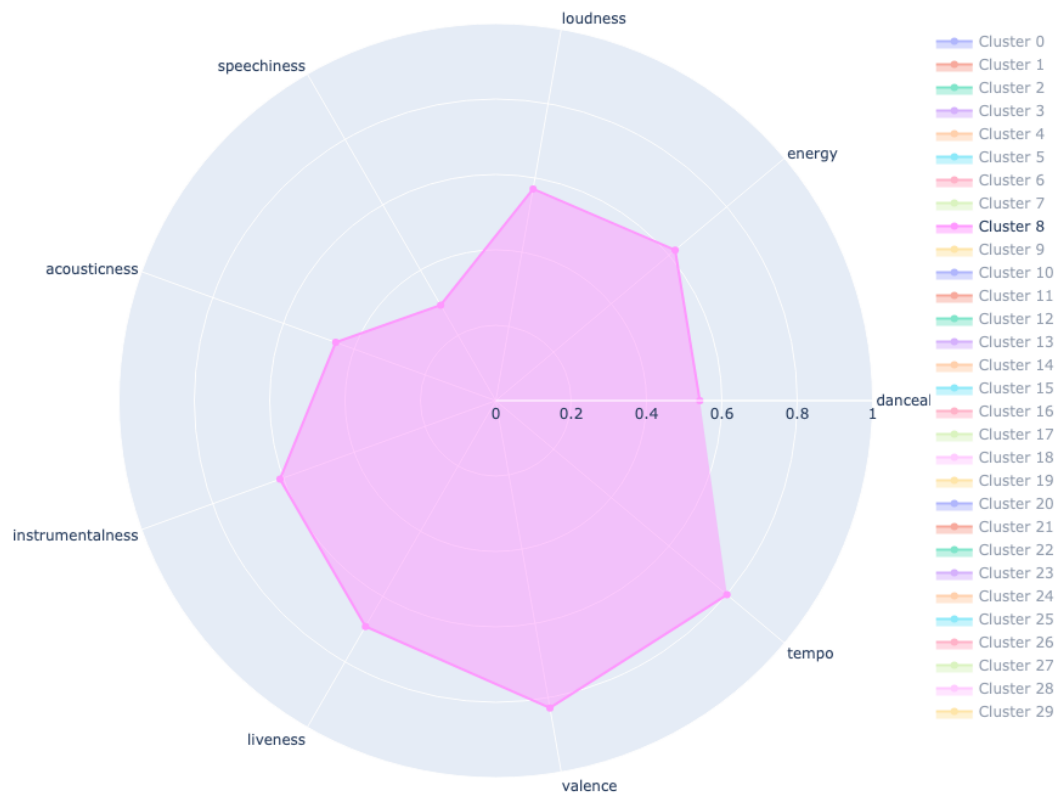 - Max 256

Clusters

Radar chart of songs by cluster

Radar chart of songs by cluster

Playlist (1/3)

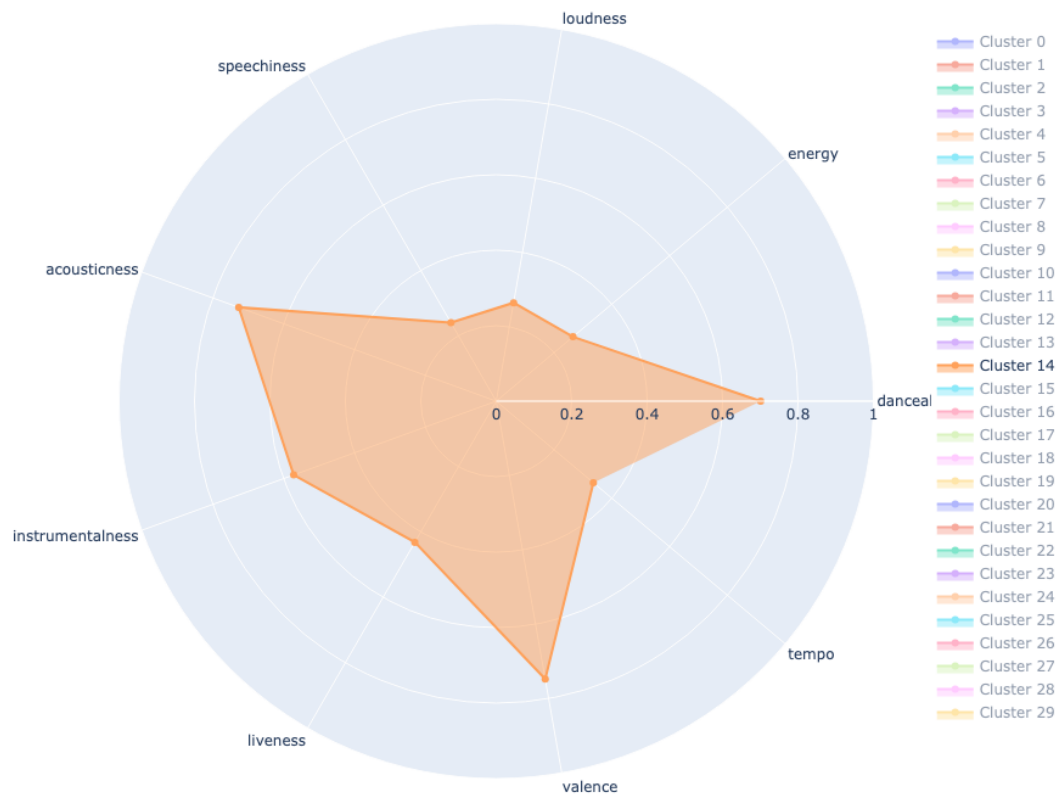"Happy listening" (123 songs)

Radar chart of songs by cluster

Legend:
- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5
- Cluster 6
- Cluster 7
- **Cluster 8**
- Cluster 9
- Cluster 10
- Cluster 11
- Cluster 12
- Cluster 13
- Cluster 14
- Cluster 15
- Cluster 16
- Cluster 17
- Cluster 18
- Cluster 19
- Cluster 20
- Cluster 21
- Cluster 22
- Cluster 23
- Cluster 24
- Cluster 25
- Cluster 26
- Cluster 27
- Cluster 28
- Cluster 29

Axes: loudness, energy, danceal, tempo, valence, liveness, instrumentalness, acousticness, speechiness

Name	Artist
Maroca	Mundo Livre
Já Sei Namorar - 2004 Digital Remaster	Tribalistas
Alegria	Batida
Dualize	L.A.
Baby Lee	Teenage Fanclub
Come Saturday	The Pains Of Being Pure At Heart
The Luckiest Guy On The Lower East Side	The Magnetic Fields
Your Fucking Sunny Day	Lambchop
Just Like Heaven	The Cure
Days	The Drums
Good for Nothin'	Scott Walker
Cali in a Cup	Woods
The Stars Keep On Calling My Name	Mac DeMarco
Red Eyes	The War On Drugs
Talking Backwards	Real Estate
Whispering or Singing	Boomgates
Natural Life	Cool Ghouls
Thine Eyes	Pure Bliss
Comin' Through	The Pastels
I Apologise	Teenage Filmstars
Ciencia Ficcion	Los Planetas
...	...

Playlist (2/3)

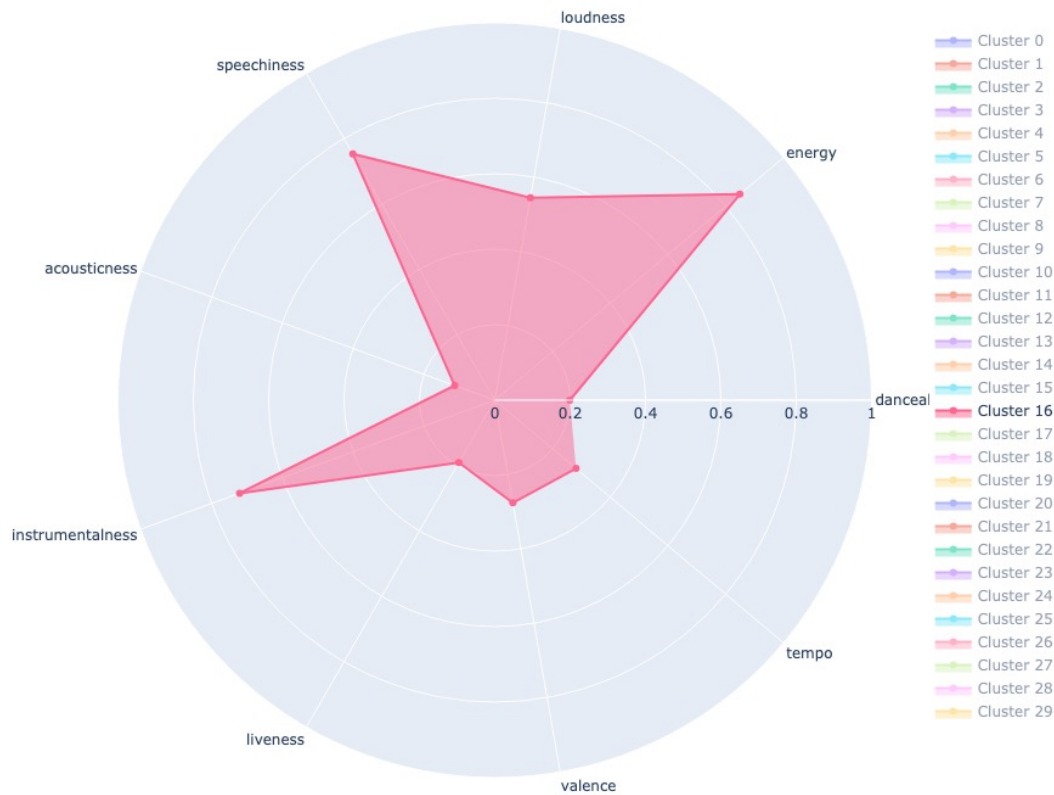"Jazzy" (202 songs)

Radar chart of songs by cluster

Legend: Cluster 0, Cluster 1, Cluster 2, Cluster 3, Cluster 4, Cluster 5, Cluster 6, Cluster 7, Cluster 8, Cluster 9, Cluster 10, Cluster 11, Cluster 12, Cluster 13, **Cluster 14**, Cluster 15, Cluster 16, Cluster 17, Cluster 18, Cluster 19, Cluster 20, Cluster 21, Cluster 22, Cluster 23, Cluster 24, Cluster 25, Cluster 26, Cluster 27, Cluster 28, Cluster 29

Name	Artist
Madalena	Maria Gasolina
Previsao	Bossacucanova
Blue Bossa - Remastered	Joe Henderson
Mas Que Nada	Rio Combo
Nova Bossa Nova	Marcos Valle
Rebel Rebel	Seu Jorge
Batida Diferente	Cannonball Adderley
Team Zissou	Seu Jorge
O Amor Em Paz	Cannonball Adderley
Once Again	Stan Getz
Corcovado - Remastered	Cannonball Adderley
August Day Song	Bebel Gilberto
Laia Ladia	Brazilian Jazz
Triste	Brazilian Jazz
Só Danço Samba	Stan Getz
O Que É, O Que É	Cesar Camargo Mariano & Romero Lubambo
Chega de Saudade	Jenny Chi
Coisa Feita	Toots Thielemans
Felicia and Bianca	Toots Thielemans
Menina Moca	Stan Getz
Santa Maria de Feira	Devendra Banhart
Blue Bossa	Tete Montoliu
...	...

Playlist (3/3)

"Metal" (219 songs)

Radar chart of songs by cluster

Axes: loudness, energy, danceal[bility], tempo, valence, liveness, instrumentalness, acousticness, speechiness

Scale: 0, 0.2, 0.4, 0.6, 0.8, 1

Legend: Cluster 0 – Cluster 29

Name	Artist
Cowboys from Hell	Pantera
Blood and Thunder	Mastodon
Pull the Plug	Death
Into The Grave	Grave
Chapel of Ghouls	Morbid Angel
Pierced from Within	Suffocation
Hammer Smashed Face	Cannibal Corpse
Spirit Crusher	Death
I Cum Blood	Cannibal Corpse
Spiritual Healing	Death
Leprosy	Death
Arise	Sepultura
For Victory	Bolt Thrower
Lake of Fire	Nocturnus
Mother Man	Atheist
Soulless - remastered 2003	Grave
Sinners Bleed	Entombed
After World Obliteration	Terrorizer
Stench of Paradise Burning	Disincarnate
Dead Embryonic Cells	Sepultura
Chronic Infection	Pestilence
...	...

KMeans Observations

+ Music clusters generally cohesive

+ Fast and simple

– Lacking cultural and lyrical nuance

Prototype refinements

• Test more techniques and compare results

• Enrich data (genre, lyrics analysis, user behavior insights)

• Analyze cluster data to enhance optimization

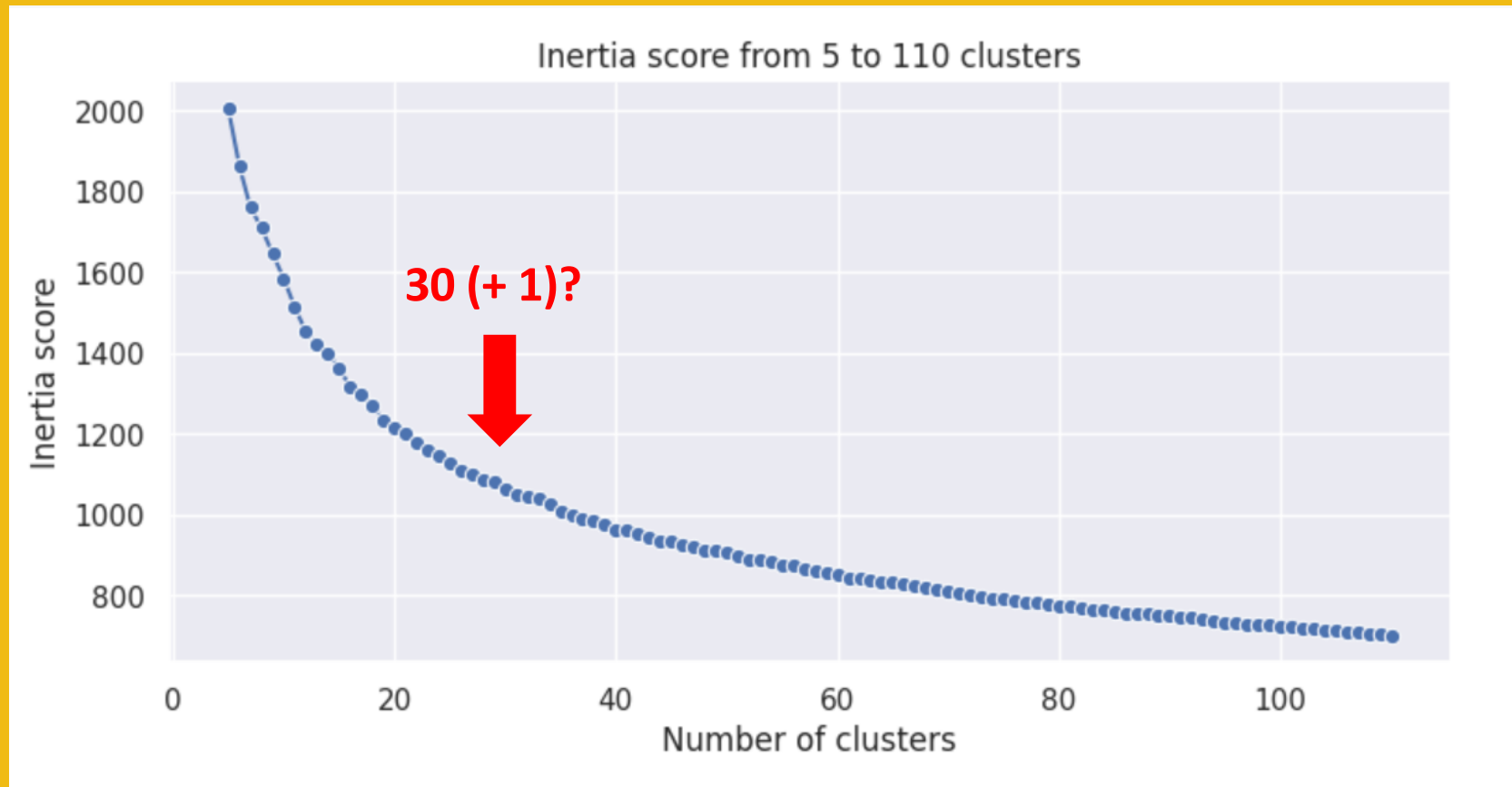• Explore deep learning for personalized recommendations

APPENDIX

selecting number of clusters (k)
for KMeans

How many clusters?
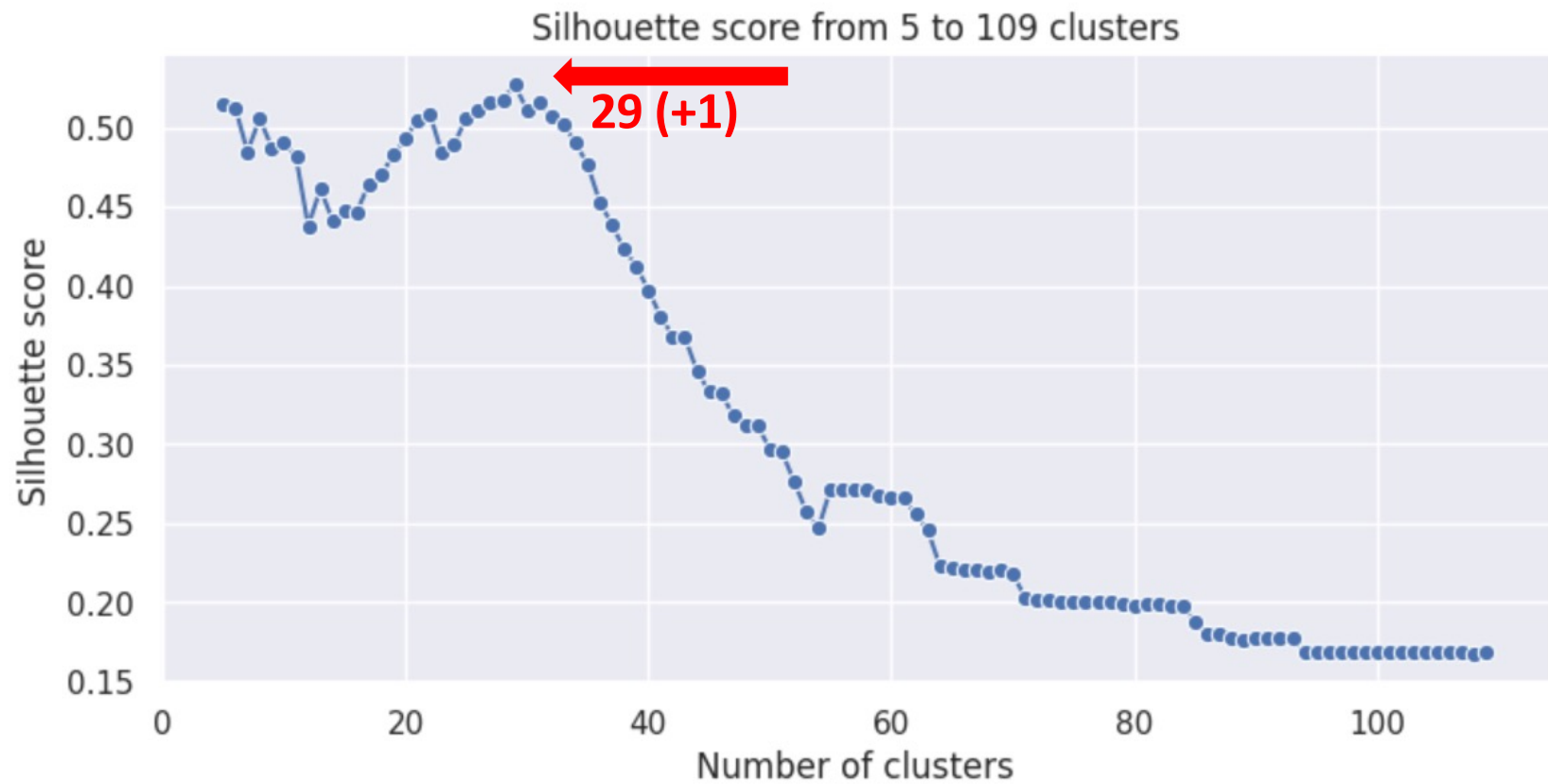
Quantile transformer, 9 features, no PCA
Inertia/elbow = 31 clusters?

Inertia score from 5 to 110 clusters

30 (+ 1)?

How many clusters?

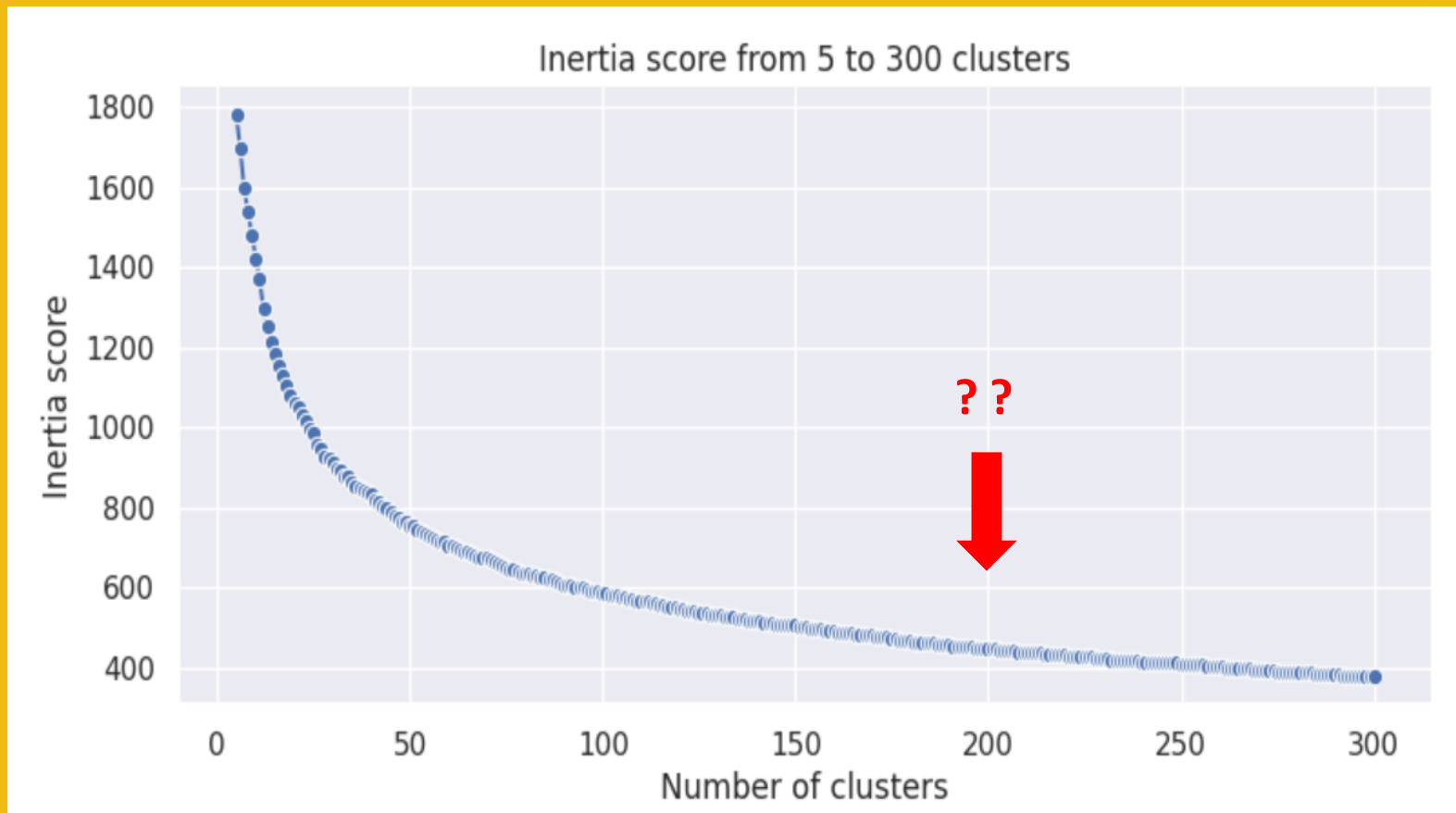Quantile transformer, 9 features, no PCA
Silhouette score= 30 clusters

How many clusters?

Quantile transformer, **7 PCAs**
Inertia/elbow = 200 clusters??

How many clusters?

Quantile transformer, **7 PCAs**
Silhouette score= 15 or 27