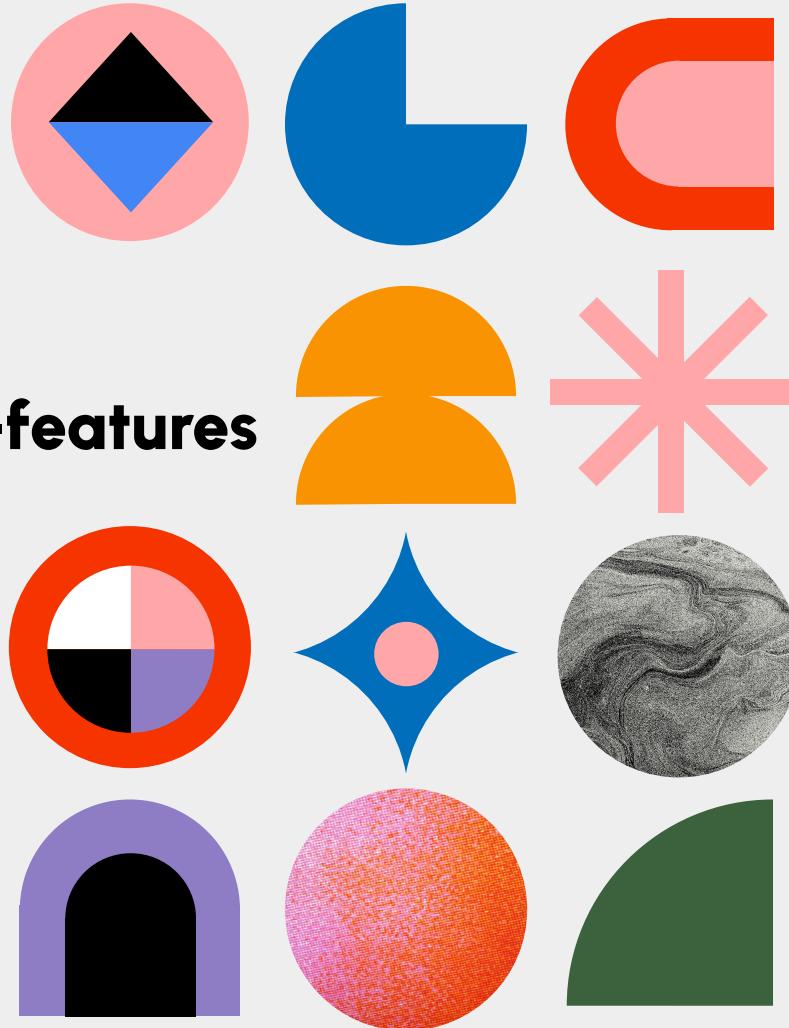


C!

Creating playlists based on audio-features

Anna
Dana
Haikel
Paula





Intro to the project



This project focuses on analyzing a dataset of **5,235 songs** sourced from **Spotify**, using **unsupervised machine learning** techniques to uncover hidden **patterns** and group similar songs into **clusters**. The primary objective is to better understand the **musical characteristics** of each cluster and identify **potential playlists** or groups based on these **similarities**.

Key Characteristics of the Dataset:

- The dataset contains a **mix of highly diverse musical attributes**: from **energy levels** to **speechiness** and **instrumentalness**.
- Some preprocessing steps, such as **cleaning column names**, **removing duplicates** (76 were identified and **dropped**), and **scaling numerical data**, were essential to prepare the data for clustering.

Dataset Overview:

- Total Songs: **5,235**
- Features: The dataset contains **19 features**, categorized into:
 - ◆ **Numerical Features (9)**: These include attributes such as danceability, energy, key, tempo, valence, and acousticness, which quantitatively describe the music.
 - ◆ **Categorical/Descriptive Features**: Song names, artists, and IDs are included for identification purposes.
- Target of Analysis: The project focuses on **numerical features** to create meaningful **clusters** of songs.



Playlist prototype creation I

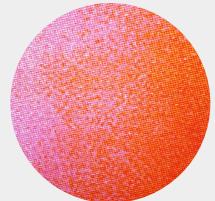
Used Features:

- Danceability
- Energy
- Vocals
- Acoustics
- Instrumental
- Live
- Mood- based
- Tempo
- Key

Dropped Features:

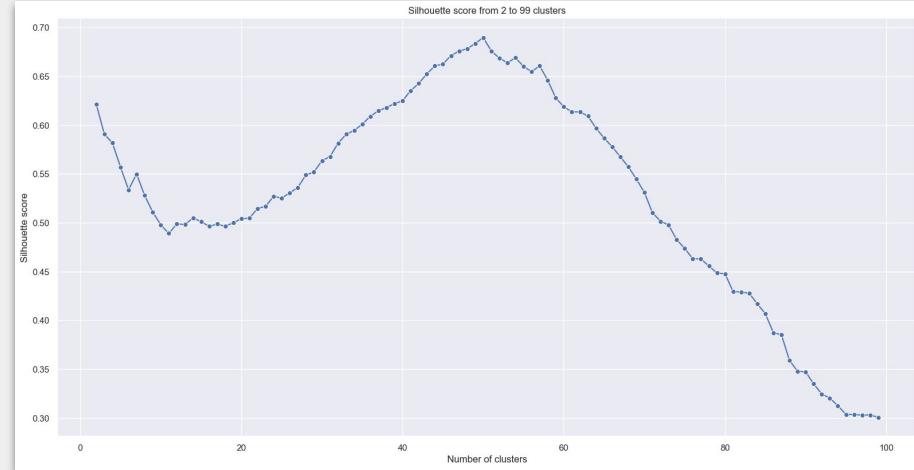
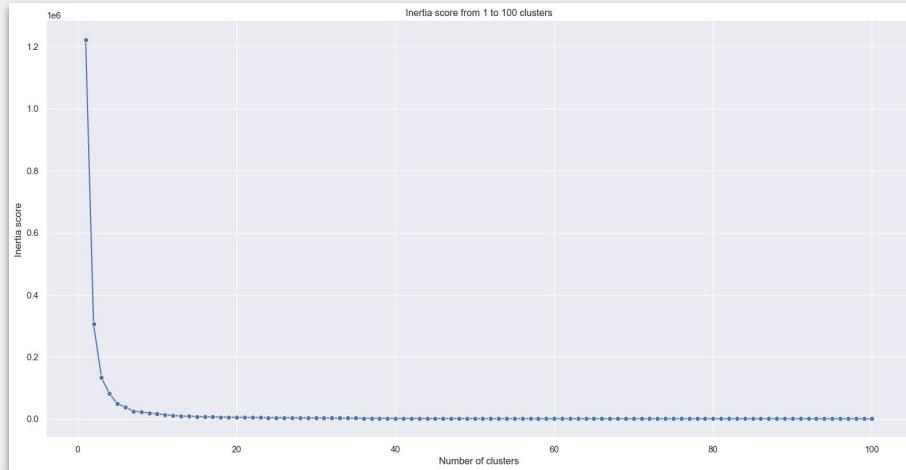
- Unnamed 0
- Song name
- Artist
- Mode
- Loudness
- Duration milliseconds
- Time signature
- Html
- ID

Combined index: song_name and artist





Playlist prototype creation II



Inertia Score:

- As low as possible
- Good fit starting at **15-20** clusters

Silhouette Score:

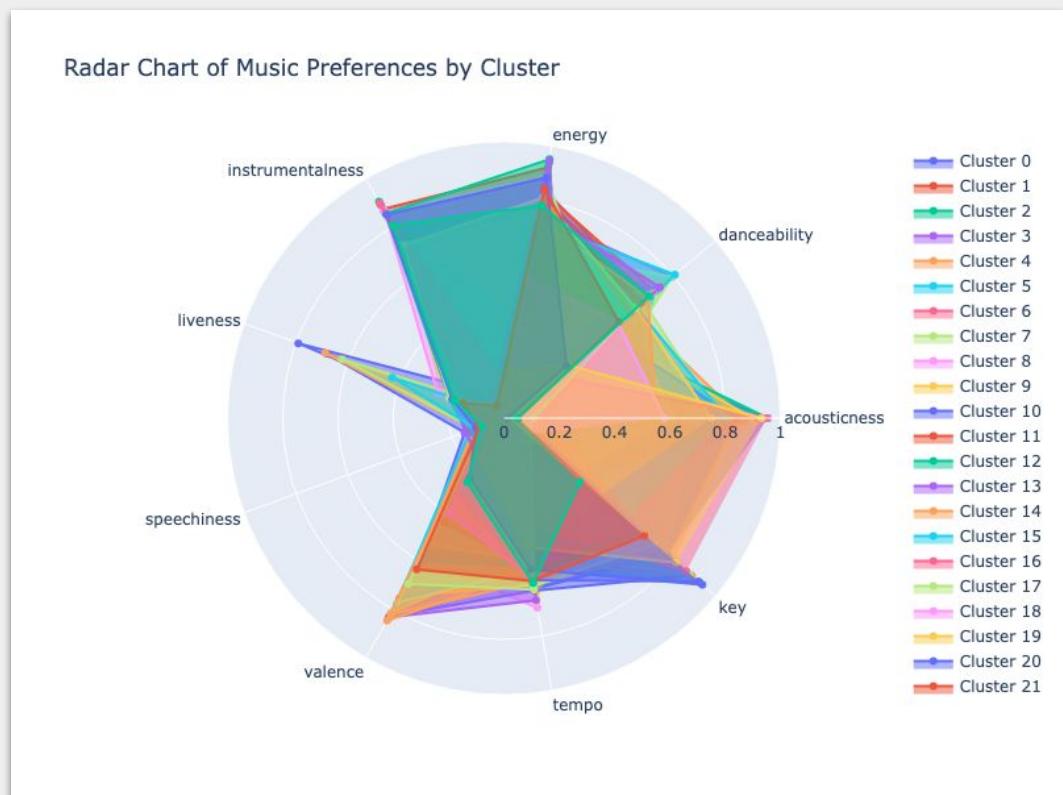
- The higher the better
- Peaking between **50-60** clusters

Both taken into consideration: 53 playlists will be created

Comparing playlists

Iterates through each playlist, retrieves the mean values for the categories, and creates a Scatterpolar object for each playlist.

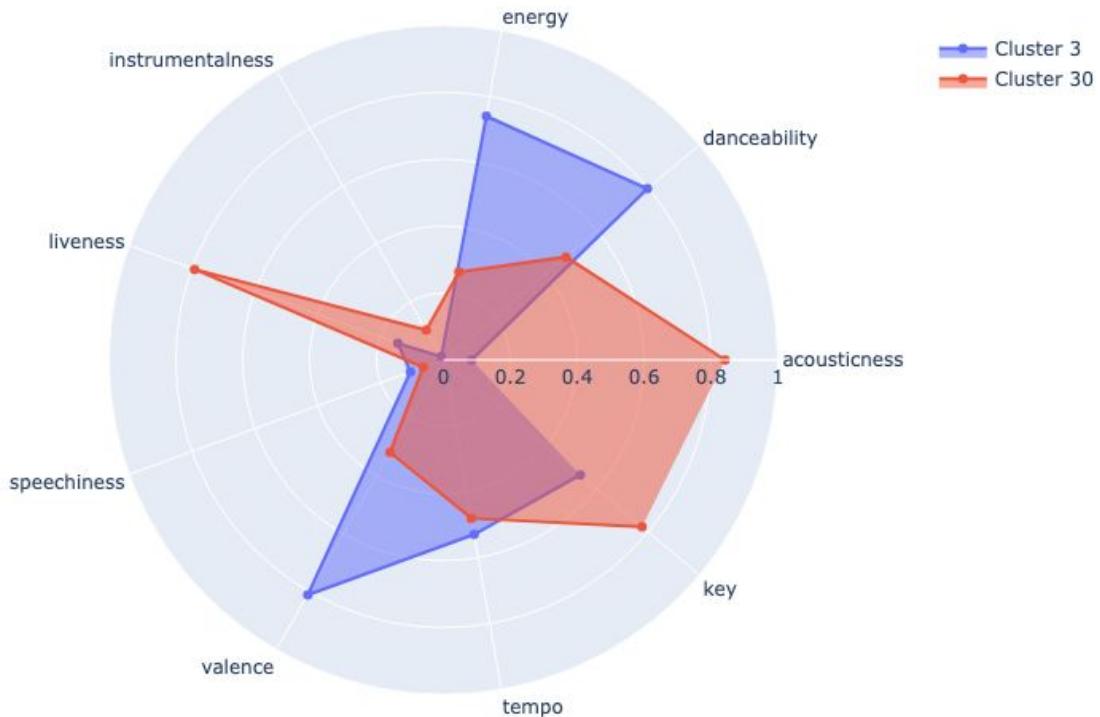
These objects define the individual radar plots for each of the 53 playlists



Playlist 3 Vs 30

Playlist 3 with a max of 244 songs and playlist 30 with 17

Radar Chart of Music Preferences for Clusters 3 and 30



Playlist 30

#	Titel	Album
1	 Se eu quiser falar com Deus Gilberto Gil	Luar
2	 Rosa Morena - Live At Carnegie Hall/1964 Stan Getz, João Gilberto	Getz/Gilberto #2
3	 À Primeira Vista Chico César	Aos Vivos
4	 Sabor Colorido / Moça Bonita Elomar, Geraldo Azevedo, Vital Farias, Xangai	Cantoria 2
5	 Shimbalaïé - Live Caetano Veloso	Multishow Ao Vivo Caetano e Maria Gadú
6	 Sampa - Live Caetano Veloso, Maria Gadú	Multishow Ao Vivo Caetano e Maria Gadú
7	 Zamba del Carnaval Pedro Aznar	Quebrado Vivo
8	 Shout, Pts.1&2 The Isley Brothers	Shout!
9	 I'll Be There (feat. Trey Lorenz) - Live at MTV Unplugged, ... Mariah Carey, Trey Lorenz	MTV Unplugged EP
10	 La traviata, Act I: Libiamo ne' lieti calici "Brindisi" Giuseppe Verdi, Illeana Cotrubas, Plácido Domingo, Chor der Ba...	Verdi: La Traviata
11	 Gianni Schicchi: O mio babbino caro Giacomo Puccini, Montserrat Caballé, José Collado	Romantic Caballé
12	 Don Giovanni, ossia Il dissoluto punito, K.527 / Act 1: "Là ... Wolfgang Amadeus Mozart, Ildebrando D'Arcangelo, Mojca Erd...	Mozart: Don Giovanni
13	 Il Trovatore / Act 2: "Vedi! le fosche notturne spoglie" (Aida) Giuseppe Verdi, The London Opera Chorus, National Philharmonic...	Verdi: Il Trovatore
14	 Lakmé / Act I: Viens, Mallika, ... Dôme épais (Flower Duet) Léo Delibes, Anna Netrebko, Elina Garanica, SWR Symphonieorc...	"Die Operngala der Stars - Live aus Baden-Baden"
15	 My Foolish Heart Bill Evans Trio	Waltz For Debby (Original Jazz Classics Remaste...



Playlist 3

#	Title	Album
1	Alegria Batida	Batida (Soundway Records)
2	Mas Que Nada Río Combo	Café Bossa Nova - Beaches And Bars, Samba And Sun
3	Agua De Beber Sophie Milman	Sophie Milman
4	Tirei o Chapéu Batida, Ikonoklasta	Batida (Soundway Records)
5	In Your Hands Charlie Winston	Hobo
6	Guerrera DELLAFUENTE, C. Tangana	Guerrera
7	Picky Joey Montana	Picky Back To The Roots
8	Hope Jonathan Taylor	I'm Here For You
9	Dance Party Kelly Price	Dance Party
10	Increase My Faith Brian Courtney Wilson	A Great Work
11	Tuya Soy Ivy Queen	Diva Platinum Edition





Usability of K-Means



Pro: relatively simple, yet can be tweaked with different pre-processing

Con: songs can't be re-used; not too precise

Possible solutions:

- Try different algorithms
- Change part of the pre-processing (one at the time)
- Further filtering or adding of columns that are of most interest



Discrepancy

Moosic Playlist 30

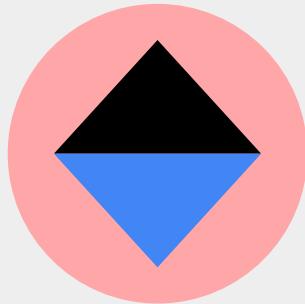
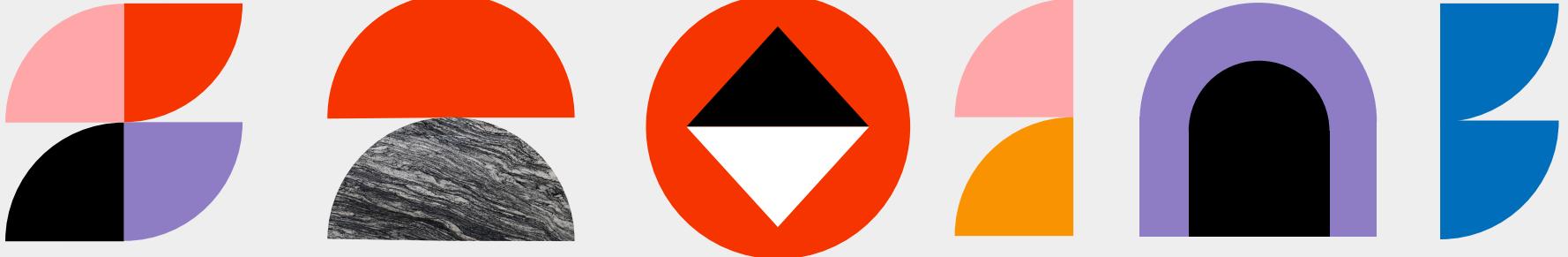
#	Titel	Album
1	Se eu quiser falar com Deus Gilberto Gil	Luar
2	Rosa Morena - Live At Carnegie Hall/1964 Stan Getz, João Gilberto	Getz/Gilberto #2
3	À Primeira Vista Chico César	Aos Vivos
4	Sabor Colorido / Moça Bonita Elomar, Geraldo Azevedo, Vital Farias, Xangai	Cantoria 2
5	Shimbalaïé - Live Caetano Veloso	Multishow Ao Vivo Caetano e Maria Gadú
6	Sampa - Live Caetano Veloso, Maria Gadú	Multishow Ao Vivo Caetano e Maria Gadú
7	Zamba del Carnaval Pedro Aznar	Quebrado Vivo
8	Shout, Pts.1&2 The Isley Brothers	Shout!
9	I'll Be There (feat. Trey Lorenz) - Live at MTV Unplugged, ... Mariah Carey, Trey Lorenz	MTV Unplugged EP
10	La traviata, Act I: Libiamo ne' lieti calici "Brindisi" Giuseppe Verdi, Ileana Cotrubas, Plácido Domingo, Chor der Ba...	Verdi: La Traviata
11	Gianni Schicchi: O mio babbino caro Giacomo Puccini, Montserrat Caballé, José Collado	Romantic Caballé
12	Don Giovanni, ossia Il dissoluto punito, K.527 / Act 1: "Là ... Wolfgang Amadeus Mozart, Ildebrando D'Arcangelo, Mojca Erd...	Mozart: Don Giovanni
13	Il Trovatore / Act 2: "Vedi! le fosche notturne spoglie" (A... Giuseppe Verdi, The London Opera Chorus, National Philharmonic...	Verdi: Il Trovatore
14	Lakmé / Act I: Viens, Mallika, ... Dôme épais (Flower Duet) Léo Delibes, Anna Netrebko, Elina Garanca, SWR Symphonieorc...	"Die Operngala der Stars - Live aus Baden-Baden"
15	My Foolish Heart Bill Evans Trio	Waltz For Debby (Original Jazz Classics Remaste...



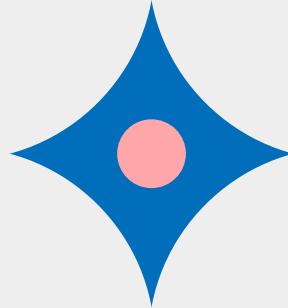
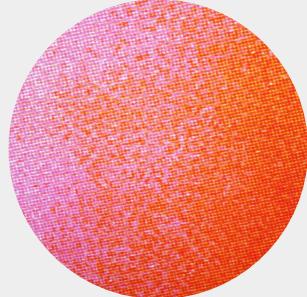
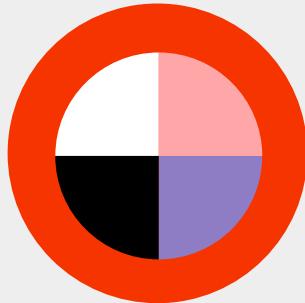
Conclusion



- Using UML was helpful creating overall the playlist, however our recommendation would be to emphasize the preprocess to make the playlists more homogeneous



THANK YOU!



STICKER SHEET

Playlist prototype creation I

Features used	Features dropped
danceability : Essential for determining the suitability of a track for dancing.	Unnamed: 0 , song_name and artist : Descriptive metadata, not useful for clustering algorithms.
energy : Helps categorize tracks based on intensity.	mode : While they describe musical attributes, less impactful for mood/style clustering.
speechiness : Useful for distinguishing between tracks with vocals versus instrumentals.	loudness : Correlated with energy; can be dropped to avoid redundancy
acousticness : Indicates how "acoustic" a track feels, helping cluster softer or more natural tracks.	duration_ms : Less important unless playlist length is a factor.
instrumentalness : Differentiates vocal-heavy tracks from instrumentals.	time_signature : Rarely varies significantly to impact clustering.
liveness : Detects live performance tracks, useful for certain playlist themes.	id and html : Metadata, not useful for clustering.
valence : Correlates with the emotional positivity of a track, ideal for mood-based playlists	type - object that can not be scaled
tempo : Captures the pace of the track, an important factor for playlist flow.	
key Represents the musical key in which the song is written	



Possible next steps



- Try different algorithms
- Change part of the pre-processing
(one at the time)
- Further filtering of columns that are
of most interest



Usability of K-Means



Pro: relatively simple, yet can be tweaked with different pre-processing

Con: songs can't be re-used; not too precise

Playlist

Moosic Playlist 30

annuschka13 • 1 Save • 17 Songs, 1 Std. 12 Min.

Play • + ... Liste

#	Titel	Album	Hinzugefügt am	
1	Se eu quiser falar com Deus Gilberto Gil	Luar	vor 2 Stunden	4:49
2	Rosa Morena - Live At Carnegie Hall/1964 Stan Getz, João Gilberto	Getz/Gilberto #2	vor 2 Stunden	4:04
3	À Primeira Vista Chico César	Aos Vivos	vor 2 Stunden	3:36
4	Sabor Colorido / Moça Bonita Elomar, Geraldo Azevedo, Vital Farias, Xangai	Cantoria 2	vor 2 Stunden	4:18
5	Shimbalaïê - Live Caetano Veloso	Multishow Ao Vivo Caetano e Maria Gadú	vor 2 Stunden	3:21
6	Sampa - Live Caetano Veloso, Maria Gadú	Multishow Ao Vivo Caetano e Maria Gadú	vor 2 Stunden	3:24
7	Zamba del Carnaval Pedro Aznar	Quebrado Vivo	vor 2 Stunden	5:28



Spotify audio features as criteria



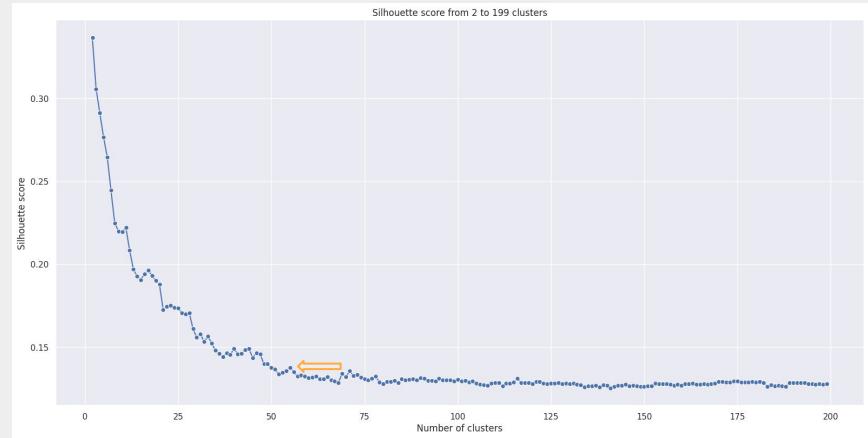
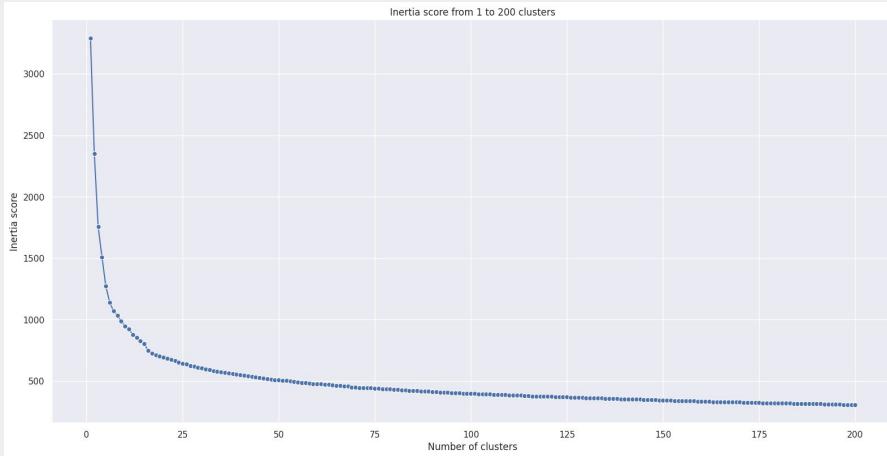
- The features given helped, but we would like to add some other features e.g. language, genre, etc.
- Try different algorithms
- Change part of the pre-processing (one at the time)
- Further filtering of columns that are of most interest



Playlist prototype creation II



Number of playlists: 53



Playlist #3 with 253 songs and Playlist #30 with 17 songs

Correlations between features

