MOOSIC

ANALYSIS AND SOLUTIONS

PLAYLIST CREATION

Jakub Czerniawski, Ramya Sri, Emiljan Ndokaj, Minwan Lee

PROTOTYPE CREATION

The deleted Features

TIME SIGNATURE BECAUSE OF NOT DEEMED IMPACT

DURATION BECAUSE OF IRRELEVANCE TO OUR ANALYSIS

MODE BECAUSE OF REDUNDANCY WITH THE "KEY" FEATURE

SPEECHINESS BECAUSE OF WRONG UNIT OF MEASURE

PROTOTYPE CREATION

Scaling data

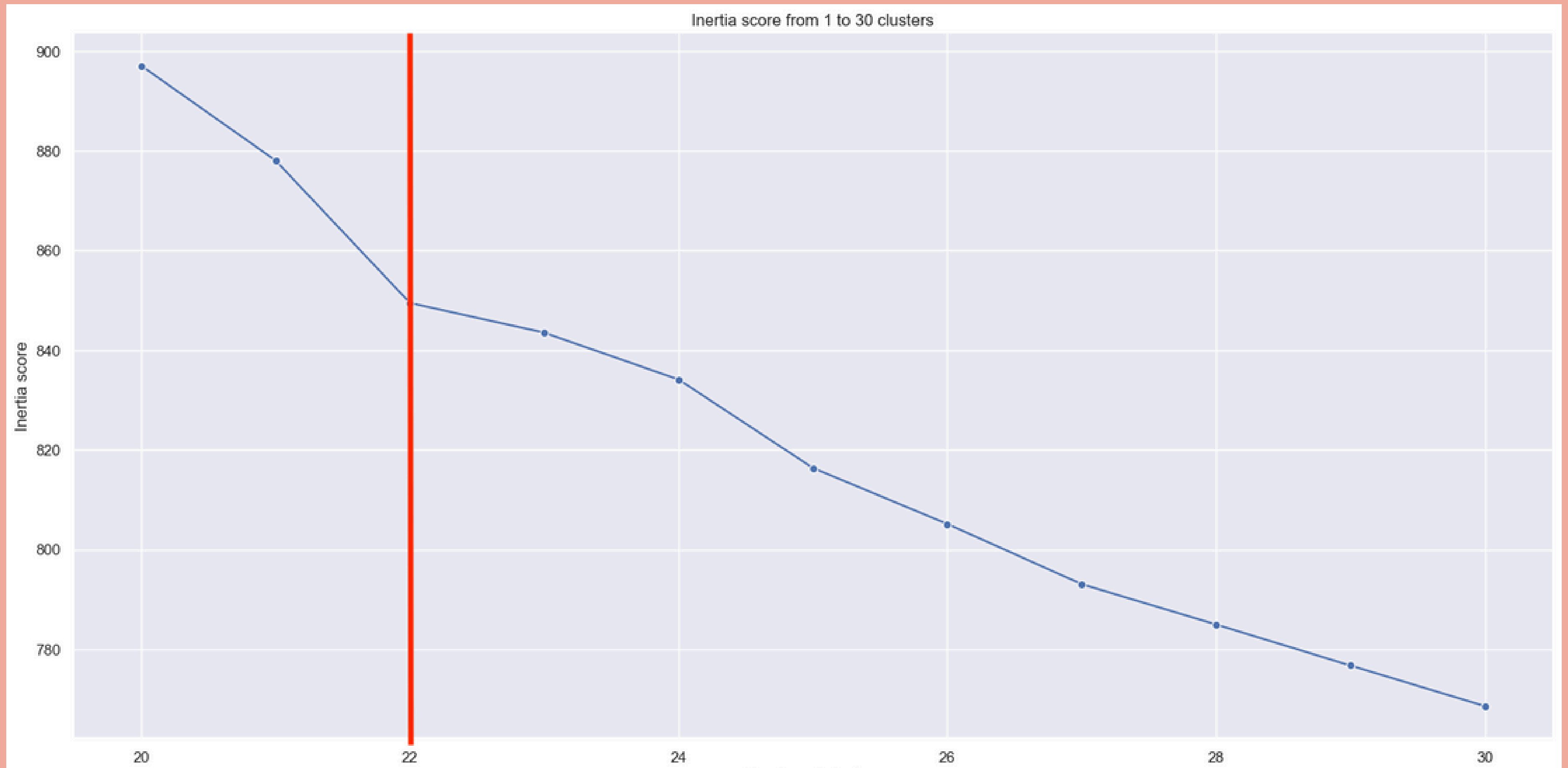
MINMAX BECAUSE OF IT SIMPLICITY TRANSFORMING WITH 0 TO 1

PROTOTYPE CREATION

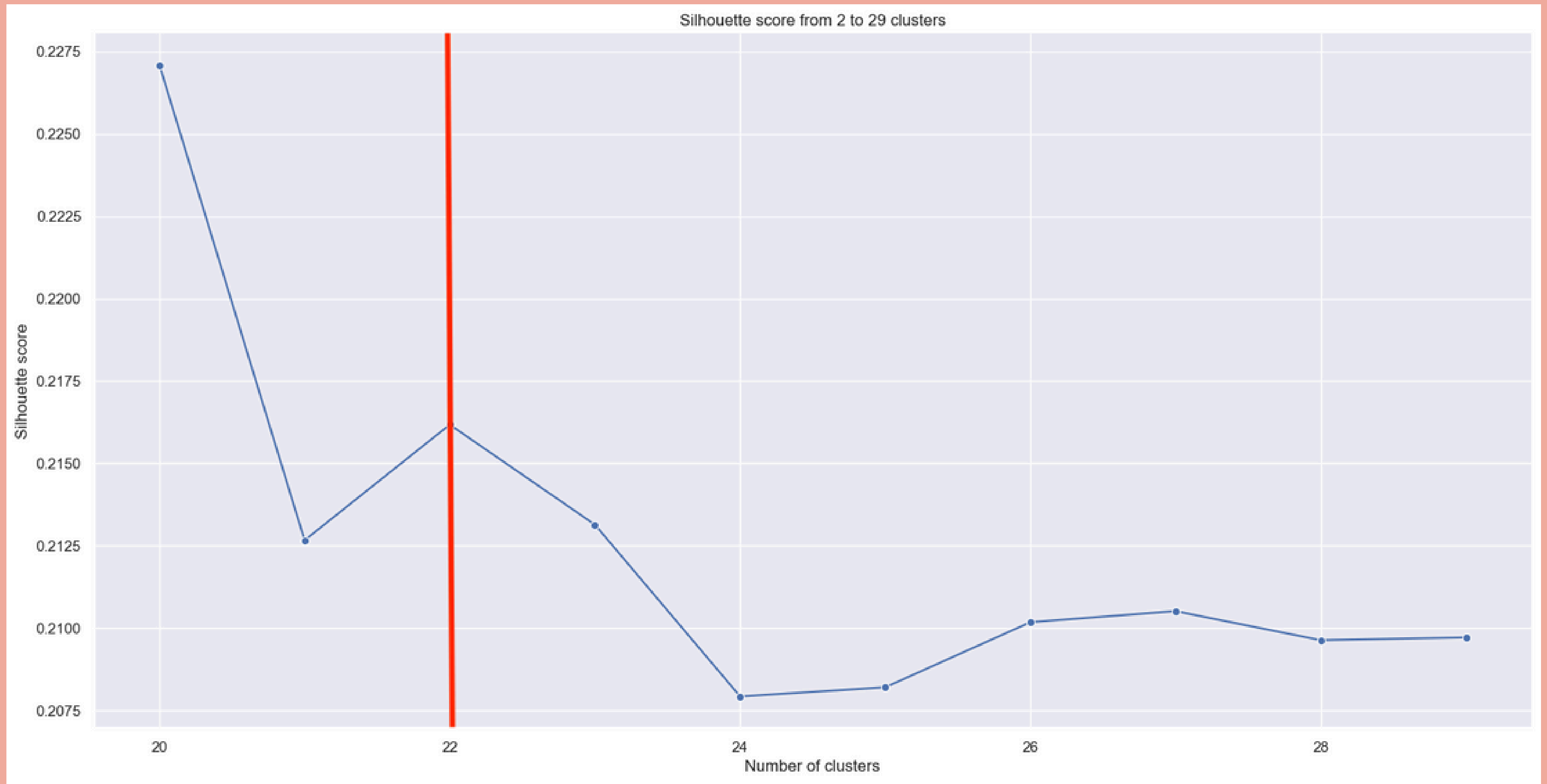
CLUSTERING USING K-MEAN

ESTIMATION OF 22 CLUSTERS BECAUSE OF:

Inertia Score



Sihouette Score



CREATING PLAYLISTS

HOW WE APPROACHED CREATING PLAYLISTS WITH 25 SONGS?

CALCULATE DISTANCES:

- Distance Measure: For each song, we measure how close it is to the center of its cluster (called the centroid). This is done using Euclidean distance, which tells us how far the song is from the cluster's average characteristics.

SELECT TOP SONGS:

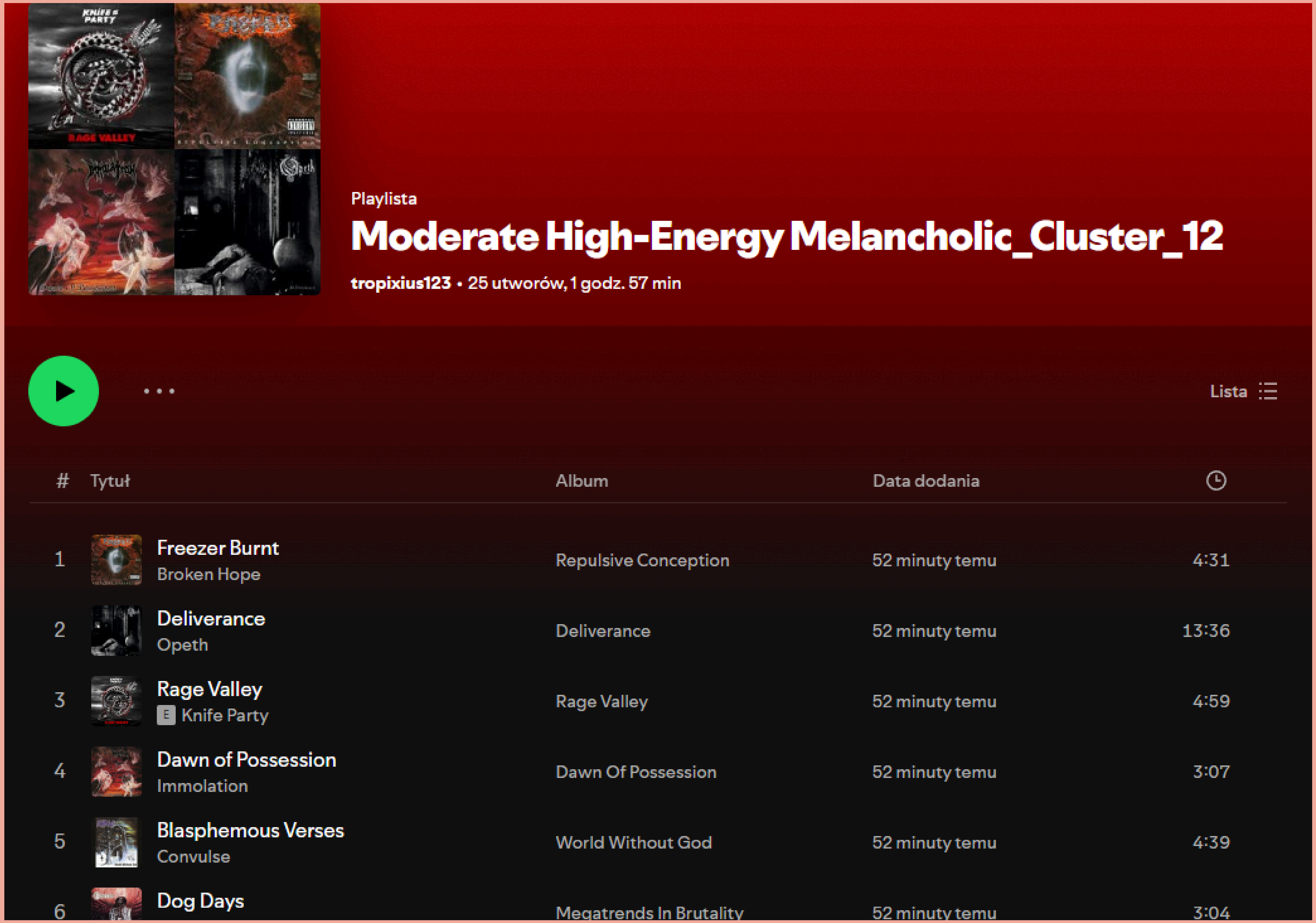
- We sort the songs by their distance to the centroid and select the top 25 closest songs for each cluster. These are the songs that best represent the cluster.

CREATE PLAYLISTS:

- Using these top songs, we create playlists that capture the essence of each cluster.

JK

EXAMPLE PLAYLISTS



https://open.spotify.com/playlist/7sbPTwmtq1VVnoLkN3zkWu?si=10c9011d165a41c2

EXAMPLE PLAYLISTS

https://open.spotify.com/playlist/2zzzCz19wth4RQA4PYiDjI?si=046bb4b0752d4ac9

ASSESSING AUDIO FEATURES

Spotify's audio features, such as Danceability, Energy, and Valence, are effective for grouping songs based on certain aspects of their mood and style. These features can capture and categorize songs within similar emotional and stylistic dimensions. However, they may not fully align with human perceptions of similarity, which are influenced by a broader range of factors.

To enhance playlist quality and better reflect human notions of similarity, incorporating additional features could be beneficial. These include:

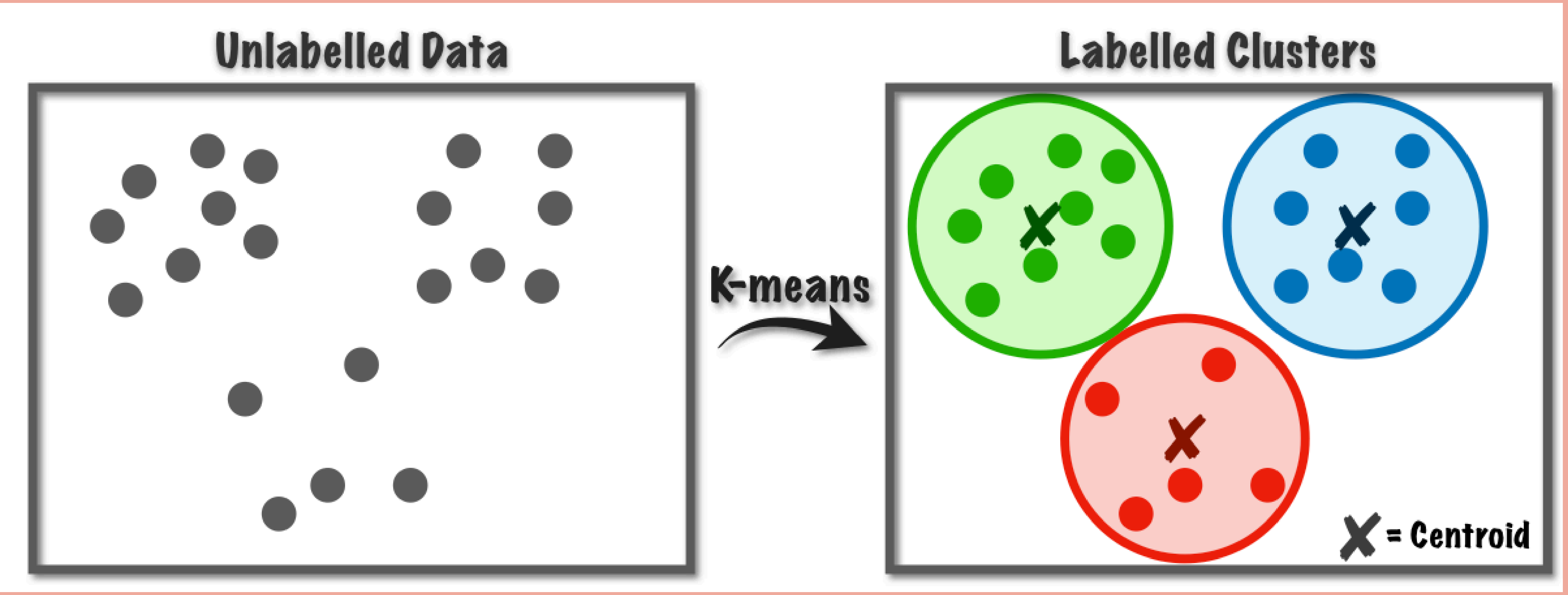
- Genre: Helps in grouping songs with similar musical styles.
- Lyrical Content: Captures thematic and emotional elements of the lyrics.
- Cultural Context: Considers the cultural and historical background of the music.
- Language: Affects the interpretation and emotional resonance of the lyrics.

  Integrating these dimensions could provide a more comprehensive understanding of song similarity and improve the relevance of curated playlists.

K-MEANS EVALUATION

K-MEANS IS A CLUSTERING ALGORITHM WHICH DOES A PRETTY GOOD JOB TRYING TO COMBINE FEATURES WHICH ARE ALIKE.

It creates clusters which center around a centroid which is the middle point of all the instances



But is that good enough for our case in trying to create playlists?

LET'S FIND OUT BY TRYING TO BRING OUT THE PROS AND CONS ABOUT IT

PROS

SIMPLICITY:

Easy to implement and understand the process that's behind it.

SCALABILITY:

Handles large datasets efficiently.

COHESIVENESS:

Groups songs with similar audio features effectively.

CONS

FIXED CLUSTERS:

K-Means requires a pre-set number of clusters, which may not always match the natural grouping of songs. Although this can be solved by implementing PCA it would still cause a few issues.

Depending on the scaling that we use we get different results:

- Standard scaling does not reduce the cluster number but it maintains all the information necessary
- MinMax will reduce the number of columns but since the PCA Is based with variance then we will never know if the information lost is important or not.
- It becomes difficult to find which ones are the columns that should be removed since PCA naming is an issue

CONS

FEATURE SENSITIVITY:

The quality of clusters heavily depends on the selected features. We must be careful even in feature selection or even in data gathering of those features. It can cause that the whole clustering and selection of centroid to be wrong and skewed.

DOESN'T ACCOUNT VARIETY:

When creating a playlist rather than wanting similar songs there are people who would like various different categories of songs in one regarding to their listening habits . K-means is not a good clustering algorithm when it comes to trying not to group by similarity .

CONCLUSION

K-Means is a good method to create playlists but not the best if there's not too much information over the songs or the users listening behaviour. With enough data then K-Means is a great tool to create playlists.

FUTURE UPDATES

PERSONALIZING PLAYLISTS WITH USER DATA

Integrate user listening habits, favorite songs, and playlist compositions into the playlist generation process.

FEATURE-BASED GENRE CREATION

By combining key audio features like Danceability, Energy, and Valence, we can generate new, more descriptive genre labels.

FEATURE-BASED GENRE CREATION

Conduct A/B testing to compare the machine-generated playlists with human-curated ones. Measure user engagement metrics such as time spent listening, number of skips, and playlist saves to assess the effectiveness of the machine-generated playlists

The end

We are open for you questions

How we named the Playlists? (Appendix additional info)

Analyze Playlist Characteristics:

- Key Features: We looked at important features of the songs, such as danceability, energy, acousticness, valence, and tempo.
- Variability: For each playlist, we considered the variability in these features to understand its unique traits.

Determine Key Traits:

- Top Features: For each playlist, we identified the top three features with the most variability. This helps in understanding the main characteristics of the playlist.

Create Descriptive Titles:

- Descriptive Labels: Based on the values of these key features, we assigned descriptive labels. For instance:
    - Danceability: Low = "Chill", Medium = "Groovy", High = "Dancefloor"
    - Energy: Low = "Calm", Medium = "Energetic", High = "High-Energy"
    - Acousticness: Low = "Electric", Medium = "Blended", High = "Acoustic"
    - Valence: Low = "Melancholic", Medium = "Balanced", High = "Joyful"
- Combine Labels: We combined the labels of the top features to create a title that reflects the overall vibe of the playlist.

Calm Groovy Moderate_Cluster_2

- Features:
    - Energy: Calm (relaxed and soothing)
    - Danceability: Groovy (engaging rhythm)
    - Acousticness: Moderate (balanced sound)
- Explanation: This playlist features a relaxed mood with a rhythmic beat and balanced energy, offering a pleasant and engaging listening experience.