MOOSIC

Data Detectives: Cracking the Playlist Case with Audio Clues and K-Means

THE SOUND OF SIMILARITY:
CAN ALGORITHMS MATCH HUMAN INTUITION?

✅ **Yes! K-Means successfully grouped songs in a way that aligns with human perception of similarity.**

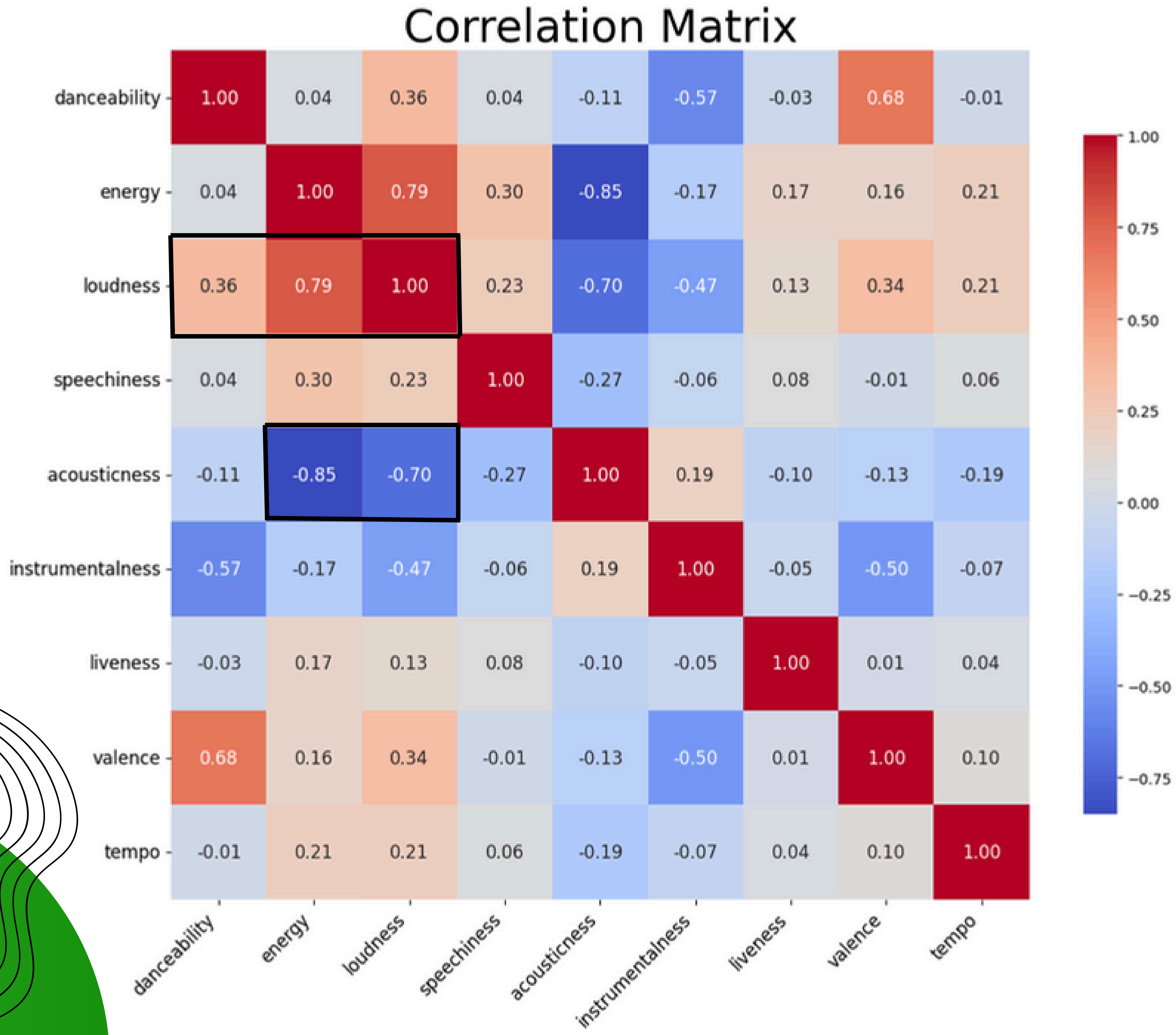
- Audio features like tempo, energy, and valence helped cluster songs with shared characteristics.
- Clustered songs conveyed similar moods.
- Human review confirmed a very high degree of perceptual similarity within clusters.

HOW DID WE APPROACH THIS MYSTERY?

- 📁 Step 1: Load the Evidence (5,000 Songs) in a CSV format
- ⚖️ Step 2: Scaled the continuous, numerical features (Danceability, Energy, etc.).
- 📊 Step 3: Using K-Means clustering, we grouped songs based on their shared features and perceived similarities.
- 🎶 Step 4: Assigned thematic playlist names based on cluster centroids.
- 🚀 Step 5: Uploaded playlists to Spotify via API

CLUE BREAKDOWN

- Kept Features: We kept only the most relevant numeric features that will lead us to the solution (clustering).
- Dropped Features: We eliminated distractions – non-numeric data that wasn't helpful for our investigation.
- Feature Scaling: Like leveling the playing field, we scaled the features to ensure no one clue (like loudness) could overshadow the rest in the analysis.



Positive Correlations:

loudness & energy

danceability & valence

Negative Correlations:

acousticness & energy

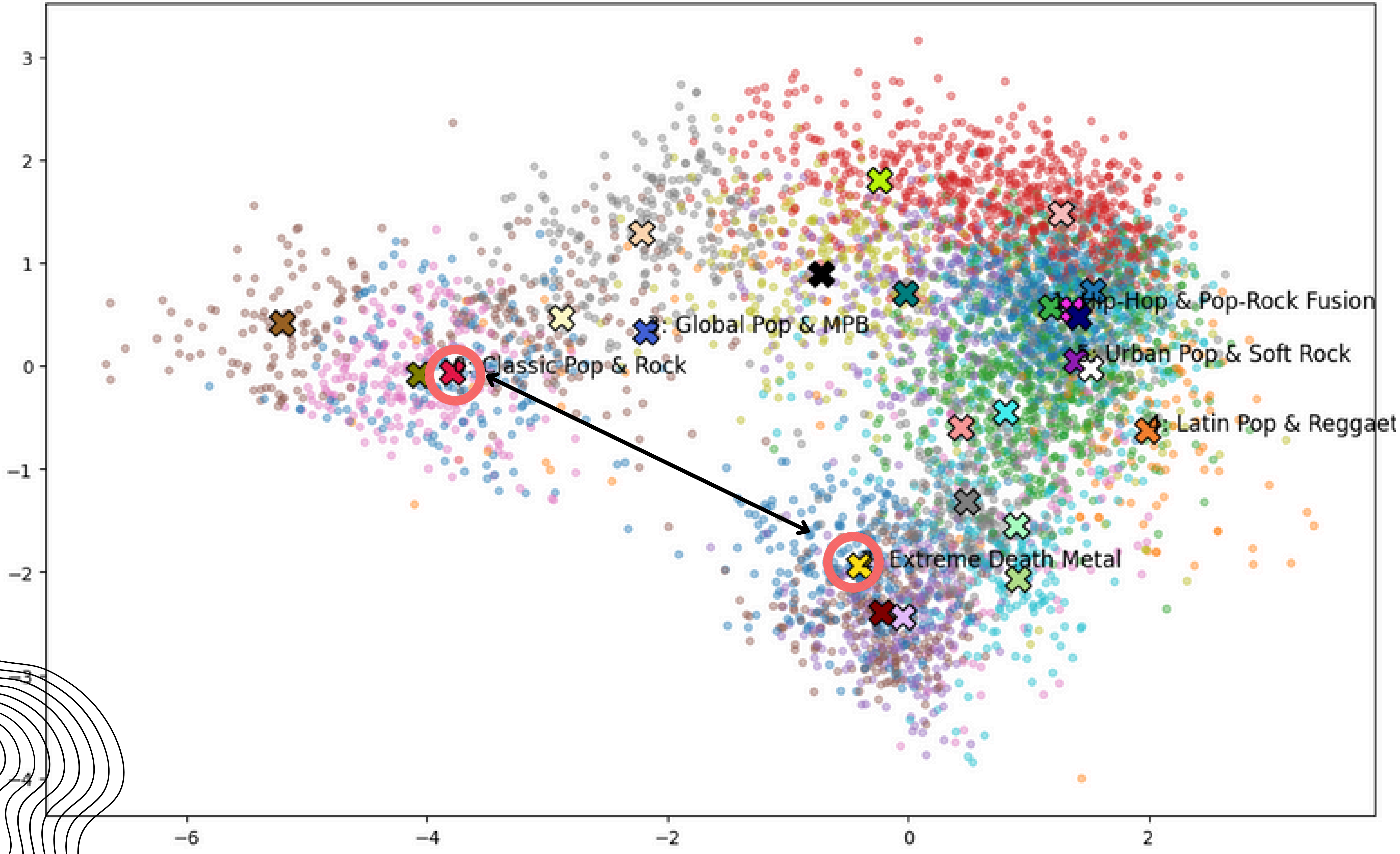
acousticness & loudness

instrumentalness & danceability

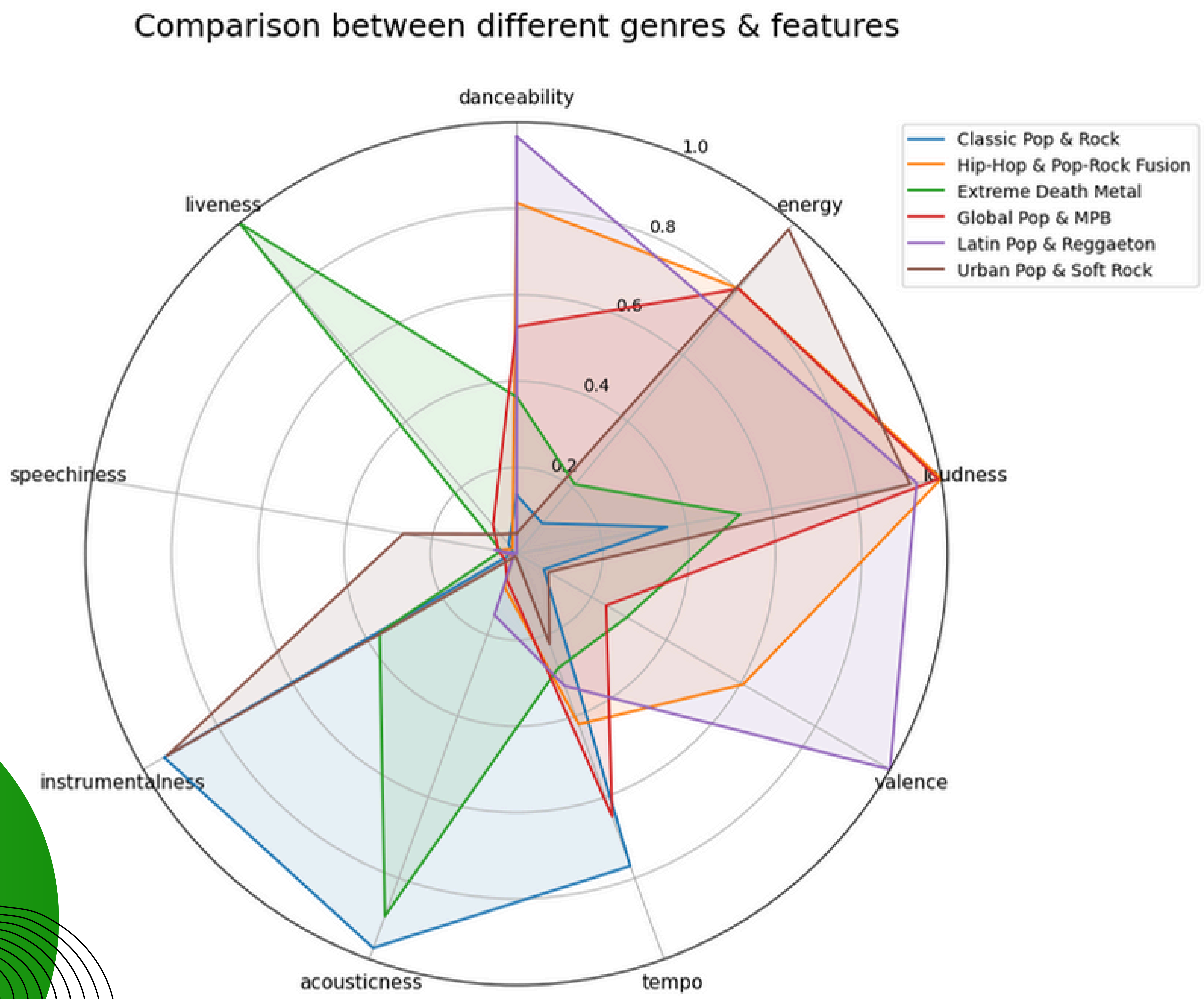
First Analysis:

- Energetic tracks tend to be loud
- acoustic and instrumental tracks are generally quieter and less energetic

Clustering and Mapping Music Genres with KMeans and PCA



- Each point is a Song colored by Cluster
- KMeans clustering (k=25) to group songs with similar sound profiles.
- PCA to reduce feature space to 2D for visualization.
- Large "X" = Centroids
- 6 clusters labeled for clarity
- Distant Clusters: distinct styles (Metal vs. Classic Pop & Rock)



A few observations:

- "**Hip-Hop & Pop-Rock Fusion**" has high speechiness and energy but low acousticness, indicating a vocal-driven, energetic, loud style with high danceability.
- "**Urban Pop & Soft Rock**" has a high level of energy, loudness and instrumentalness => this can speak for a great diversity in the genre
- "**Latin Pop & Reggaeton**" has a high level of danceability, loudness and valence

HOW DID WE APROACH THIS STUDY

METHODOLOGY

- Grouping songs with numerical features only
- Data scaling: using the Min Max approach, range [0,1]

- Clustering with KMeans (k = 25)
- random_state=42 ensures reproducibility

- We opted for K-means:

1. fits the Unsupervised ML - no need for
2. data labels
3. Easy to implement
4. Fast for large data sets
5. Easy to interpret

🎯 This approach does not use genre labels — it's purely based on sound features.

HOW DID WE APROACH THIS STUDY

METHODOLOGY (2)

- Thematic Playlists names assignment:

1. Artist profiling using ChatGPT
   - Pl 0: “Classic Pop & Rock”
   - Pl 1: “Hip-Hop & Pop-Rock Fusion” (Sample: Appendix, Table 1)
2. Manual checking of each genre

- Number of centroids (k) was determined using:
- The Silhouette score (Appendix, graph 1)
- backed with, the Elbow method (Appendix, graph2)

PROTOTYPE EFFECTIVENESS

METHODOLOGY (3)

Unnamed: 0	artist	cluster	html
1029	Dusty Springfield	0	https://open.spotify.com/track/1uLwVWTOpMqNkBs...
4880	Revolverheld	0	https://open.spotify.com/track/55VQvOY6dT63vHm...
2161	Eminem	1	https://open.spotify.com/track/4dK00wCxlqWEeN8...
4807	Tokio Hotel	1	https://open.spotify.com/track/2qQj72YfEUQ98t7...

- Sampled 2 songs per playlist:
- Pl0: “Classic Pop & Rock”
- Pl1: “Hip-Hop & Pop-Rock Fusion”
- Playlists show strong cohesion

CONCLUSION & NEXT STEPS

- **Spotify's audio features** can detect musical similarities recognizable by human listeners.
- **K-Means** is a good baseline for playlist generation, but exploring more advanced models can yield better personalization and musical flow.

Include real-time-user feedback:

- geodata & time
- rates how often songs are skipped
- how much time spend the user listening to the playlist
- how much time spent on playlist

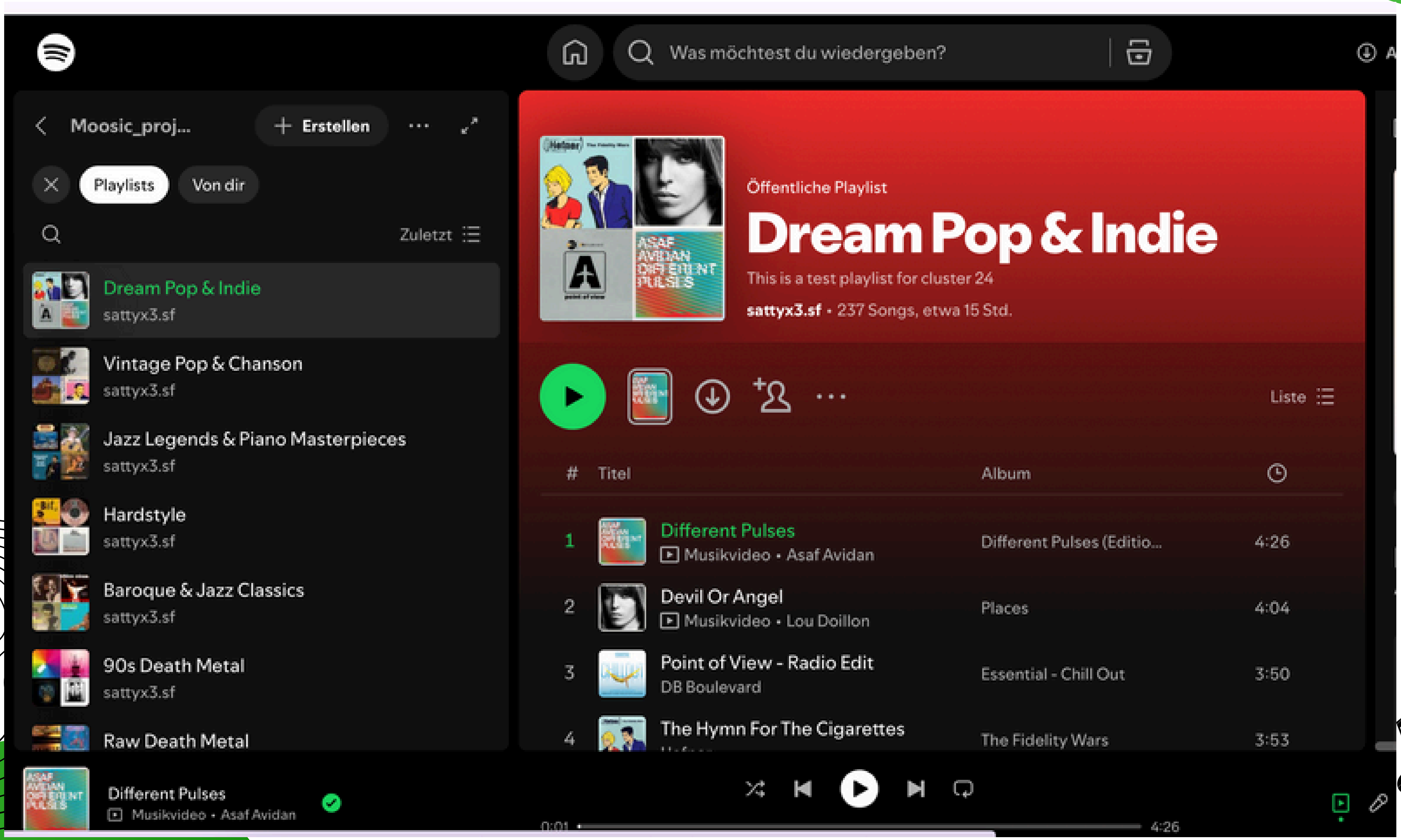
Personalized Playlists for each User Profile

THANK YOU

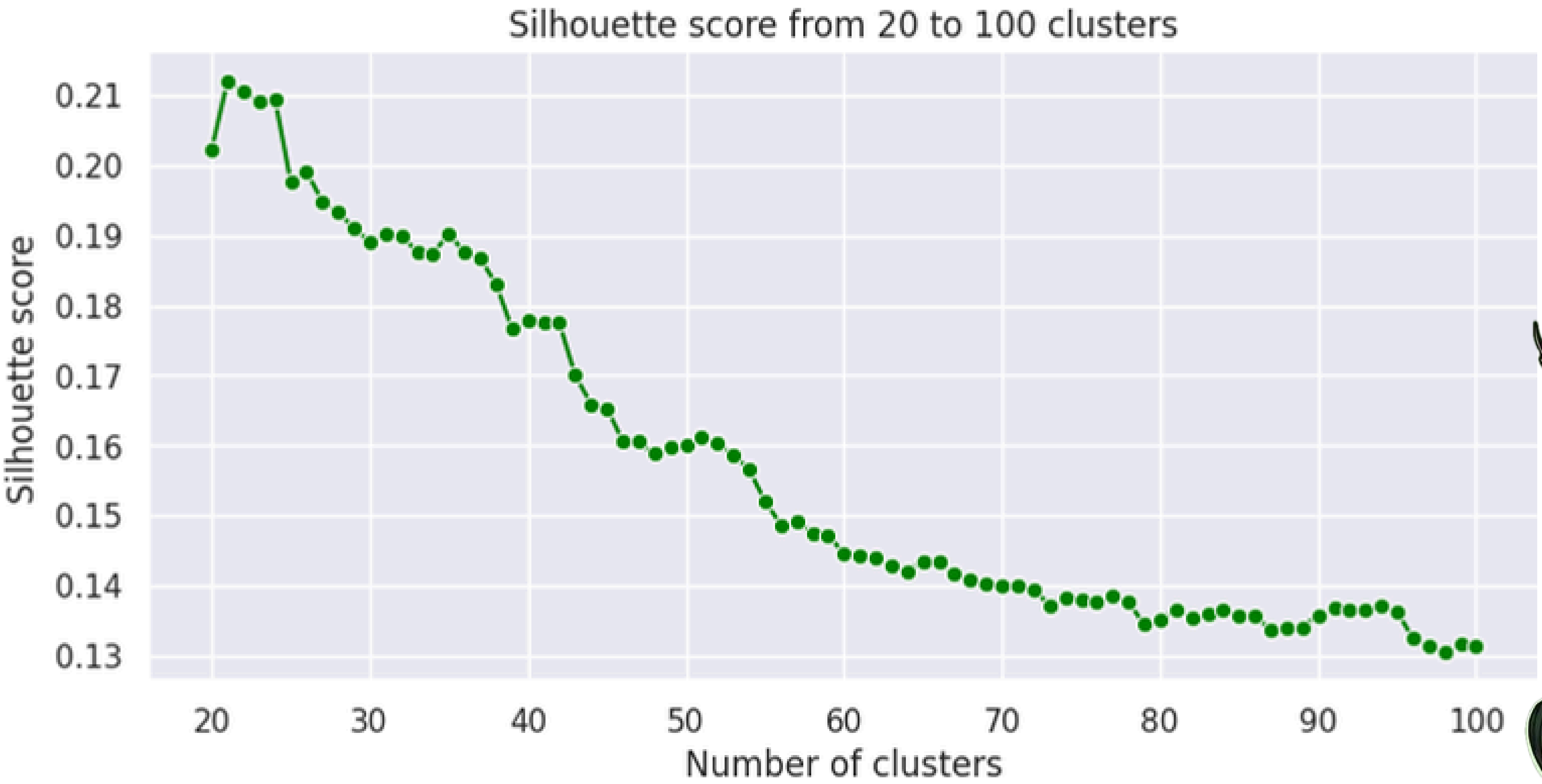
APRIL, 2025

NABIL
ANTHONY
JAN
SATYA

SPOTIFY UPLOAD



APPENDIX: GRAPH 1



APPENDIX: GRAPH 2

APPENDIX: TABLE 1

Playlist	Name
0	Classic Pop & Rock
1	Hip-Hop & Pop-Rock Fusion
2	Jazz & Neo-Classical Piano
3	Extreme Death Metal
4	Indie & World-Beat
5	Old-School Death Metal
6	Global Pop & MPB
7	Classical Masters
8	Country & Folk Rock Classics