Student Name: Lifei Wang

HW7: K-Means

Introduction

I analyzed 8,999 songs from Spotify/YouTube using 3 features:

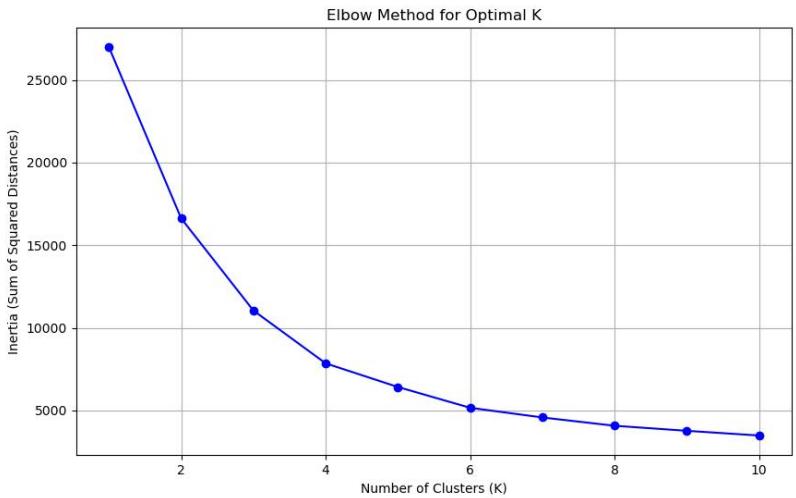
- Liveness (how much it sounds like a live recording)
- Energy (how intense/exciting the song is)
- Loudness (how loud the song is)

I used two methods to group the songs: K-means and Hierarchical clustering.

Part 1: K-means Clustering (60 points)

1. Finding the Best Number of Groups

- Used the Elbow Method to test 1–10 groups
- Best number was K=3 because adding more groups didn't help much after this point
- Graph saved in : Elbow Plot.png



2. The 3 Song Groups Found:

1) Group 0 (6,333 songs):

Medium liveness (0.16)

High energy (0.71)

Loud (–6.6 dB)

Example: Pop and rock songs

2) Group 1 (767 songs):

High liveness (0.66)

Medium energy (0.71)

Medium loudness (–7.3 dB)

Example: Live concert recordings

3) Group 2 (1,899 songs):

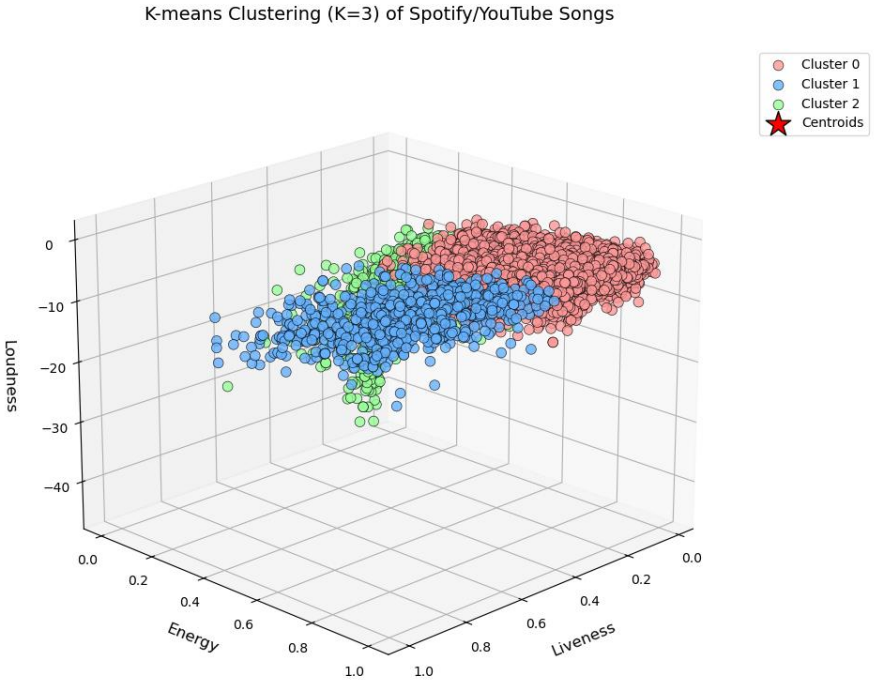
Medium liveness (0.14)

Low energy (0.27)

Very quiet (–16.4 dB)

Example: Acoustic songs

3D Visualization is shown as below



X–axis: Liveness

Y–axis: Energy

Z–axis: Loudness

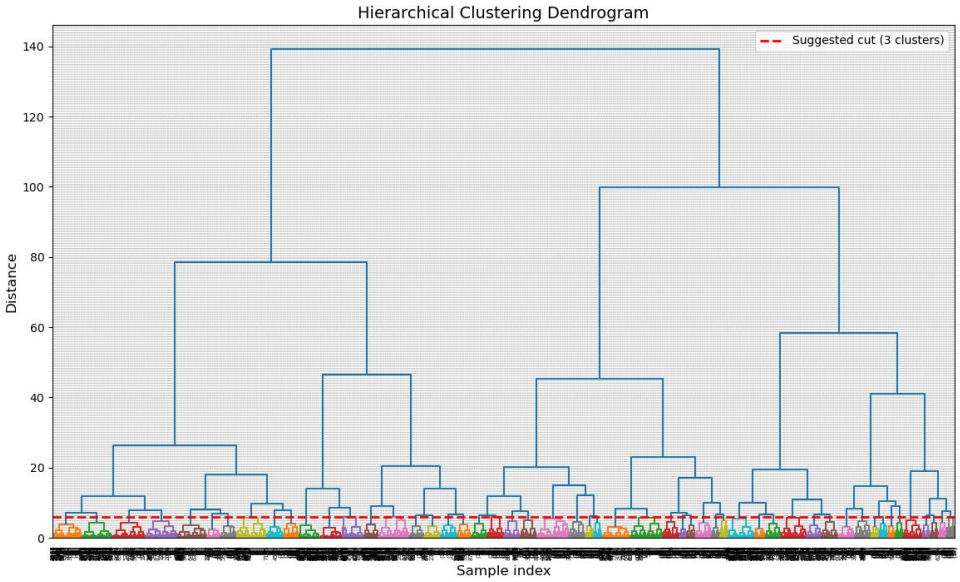
Colors show the 3 groups

Star symbols (*) show the center of each group

Part 2: Hierarchical Clustering (40 points)

1. Dendrogram Results

The tree diagram (dendrogram) also suggested 3 groups



2. The 3 Song Groups Found:

1) Group A (2,802 songs):

Low energy (0.35)

Very quiet (–14.2 dB)

2) Group B (1,755 songs):

Medium–high energy (0.73)

Medium loudness (–7.1 dB)

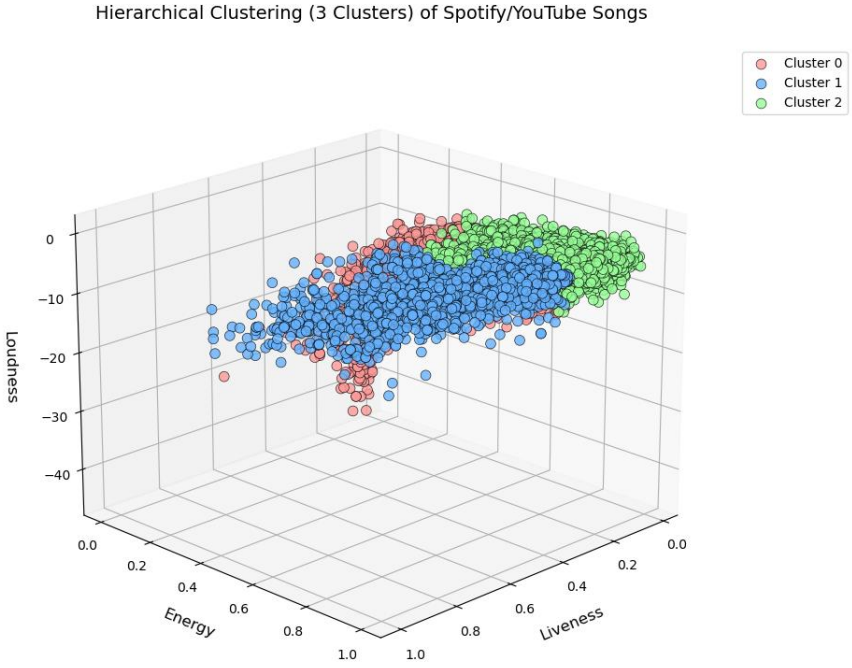
3) Group C (4,442 songs):

Medium energy (0.74)

Medium–high loudness (–6.0 dB)

3D Visualization is shown as below:

Same axes as K–means plot



Analysis & Conclusion

Both methods found 3 groups, but they disagree on how to group most songs:

1. For K-means:

Groups are based mainly on energy and loudness

Clear separation between high-energy and quiet songs

2. For Hierarchical:

Groups overlap more with each other

Harder to find simple patterns

3. Key Difference:

The low agreement (19.49%) shows songs can be grouped in multiple ways, and there is NO single "correct" way to categorize music