# HW7 - K-Means (100 points)

## 1 Introduction (Due 3/31/2025)

Take the code that was written in class for K-means and make it work for a 3Ddatasets. Using the Spotify_YouTube.csv dataset, read in the following three columns: Liveness, Energy, Loudness. Using an elbow graph, find the optimal number of K and use that to visualize the data and groups based on that K. Graphs should be appropriately labeled with an x, y, and z axis along with a title and legend. Then write what your results might mean to you.

Then, run hierarchical clustering on the same three variables. Are there any distinct groups? If so, how would you define each group? Graphs should be appropriately labeled. Then write what your results might mean to you.

## 2 Grading (Out of 100 points)

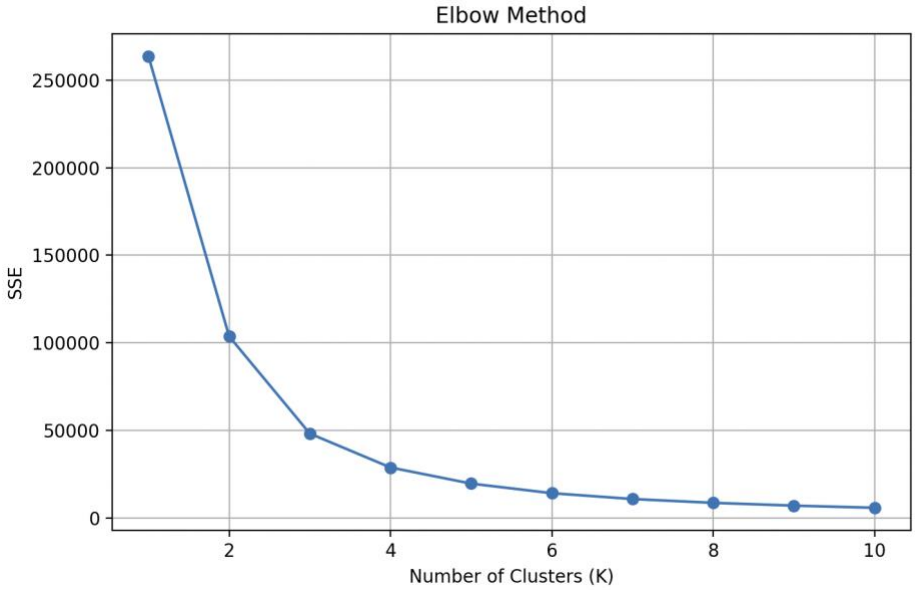
• 60 points: Cover what you did to update the code to work for 3D data and visualization along with what number of K you found to be optimal and the graphs showing the results of running K-means and what your results mean.

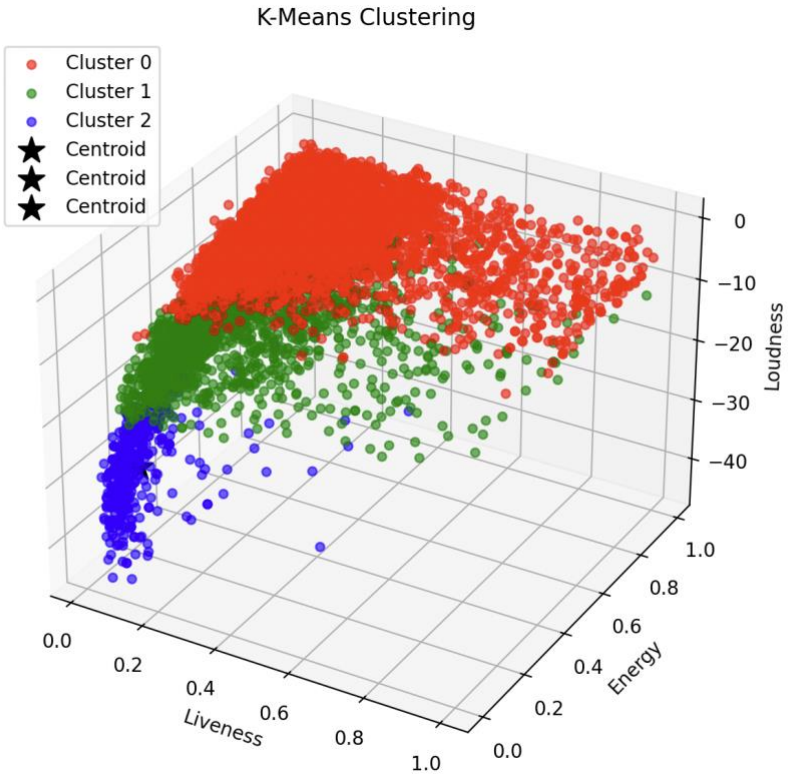
• 40 points: Report your findings for running Hierarchical clustering.

## 3 How to turn in

Turn in the final report and code that you wrote to Githbub. Then, put a link to your github submission to Canvas.

• 60 points: Cover what you did to update the code to work for 3D data and visualization along with what number of K you found to be optimal and the graphs showing the results of running K-means and what your results mean.



I used the Elbow Method and found that K=3 was a good number of clusters. After running K-Means with K=3, the 3D plot showed three clear groups based on Liveness, Energy, and Loudness.

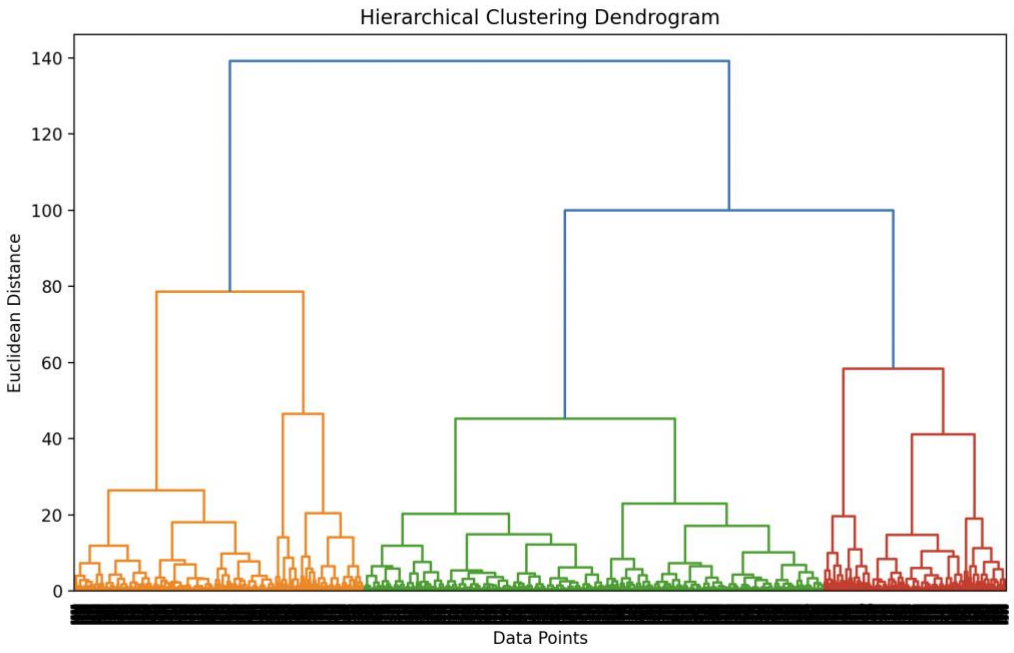
Cluster 0 has songs with higher Liveness and Energy, and the Loudness is not too low. These songs might be more lively or good for dancing.

Cluster 1 has songs with medium Energy and Liveness, and a bit louder sound. Maybe these are more energetic studio tracks.

Cluster 2 has songs that are quiet and low in Energy and Liveness. They might be calm or slow songs.

Overall, this shows that the songs in the dataset can be grouped by how energetic or loud they are. It could help to make playlists or understand song types better.

• 40 points: Report your findings for running Hierarchical clustering.




```
Cluster counts (Hierarchical):
Cluster 1: 2802 samples
Cluster 2: 4442 samples
Cluster 3: 1755 samples
```

I used hierarchical clustering with the same three features: Liveness, Energy, and Loudness. After scaling the data and plotting the dendrogram, I chose to cut the tree into 3 clusters to match what I saw in K-Means.

The dendrogram showed that there are some natural groupings in the data. After forming 3 clusters, I printed out the number of songs in each group. The clusters have different sizes, which shows some groups are more common in the dataset.

Just like in K-Means, one cluster includes songs that are louder and more energetic, and another one has quieter songs with low energy. The third group seems to be in the middle. This shows that both clustering methods found similar patterns.

Hierarchical clustering is helpful because it shows how close songs are to each other step by step, not just final group labels. It’s useful for visualizing structure in the dataset.