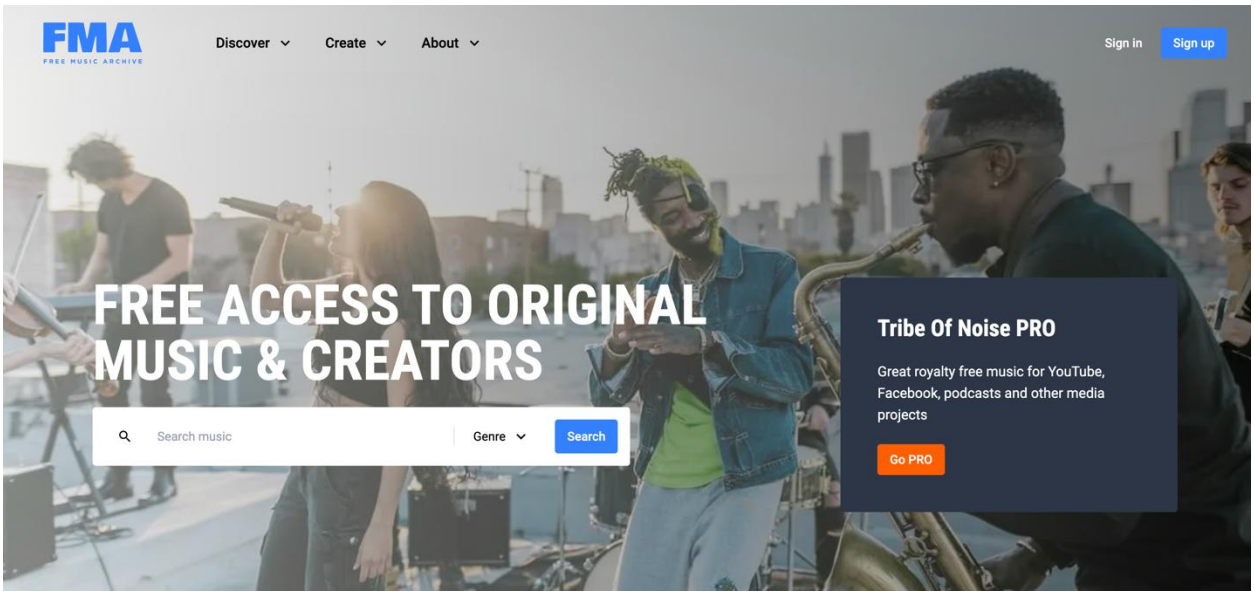
HW Assignment 4 (A4) Part 3: Clustering of FMA
CS6140: Machine Learning Fall 2024
Due Date: Thursday, _____
(10 points)



Scenario

Integrating clustering techniques with dimension reduction in unsupervised learning presents a fascinating study area. Dimension reduction, which is the process that streamlines complex, high-dimensional datasets into a more manageable form, is essential for efficient data analysis and visualization. Applying clustering methods such as k-means, c-means, hierarchical clustering, DBSCAN, and HDBSCAN to dimensionally reduced datasets offer an understanding of how these algorithms can identify patterns and groupings effectively. This approach facilitates a practical application of these algorithms and deepens the knowledge of their collective impact in enhancing data analysis, particularly within unsupervised learning.

Tasks

The last part of the Free Music Archive (FMA) dataset assignment focuses on the clustering, visualization of the results, and then conducting post-hoc analysis.

1. Machine Learning Pipeline Implementation: Implement standard machine learning pipeline techniques. This should include normalization of your data to ensure uniformity and outlier analysis to identify and address any anomalies in your dataset. You may have already done this step in Part 2 of the assignment.

2. Clustering: With your data preprocessed and dimensionally reduced, apply clustering techniques. Choose appropriate clustering algorithms to identify patterns and groupings in your dataset.

3**. Post-hoc Analysis of Clusters:** Finally, conduct a thorough post-hoc analysis of the clusters you have identified. This analysis should aim to interpret and understand the clusters, drawing insights from the patterns that emerge in the FMA data.