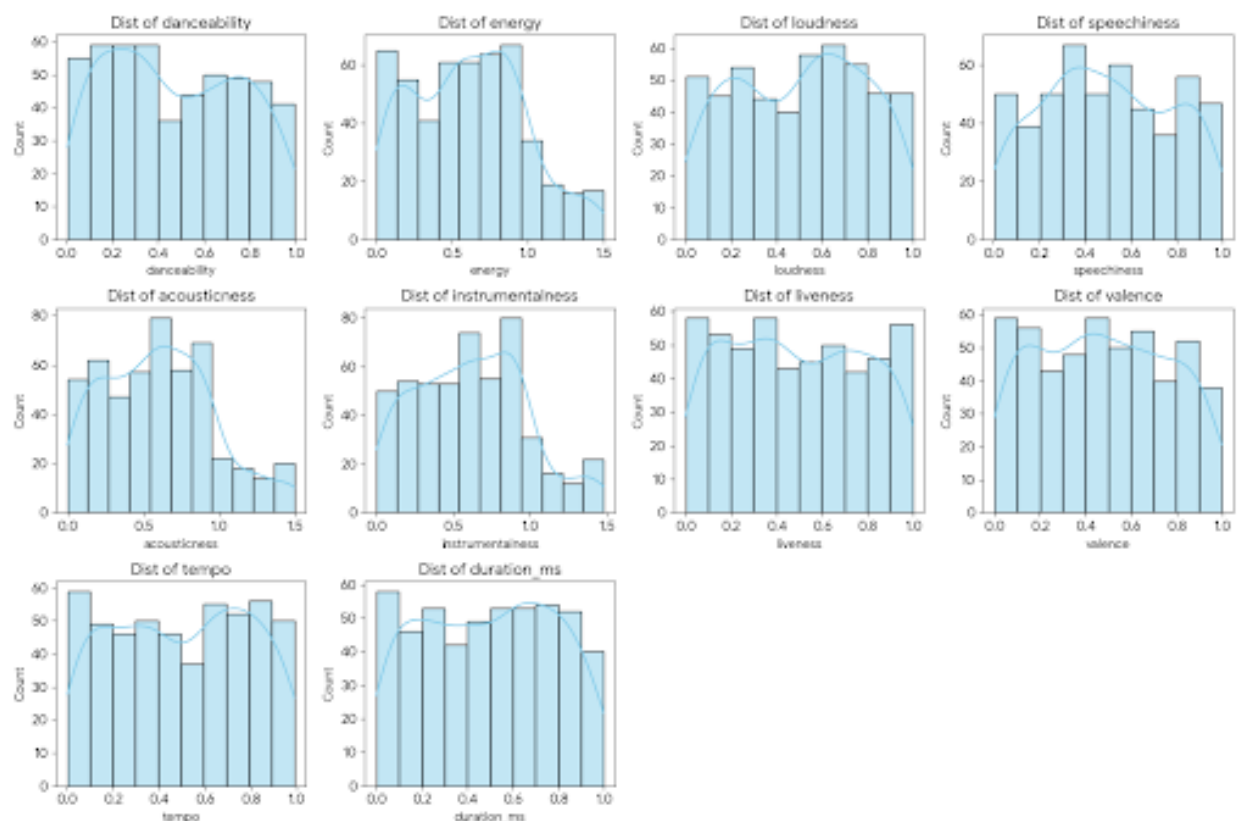


Project Report: Amazon Music Audio Feature Clustering

Objective: To implement unsupervised machine learning to group unlabeled music tracks into sonically similar clusters for recommendation and playlist generation.

1. Data Exploration & Feature Distribution

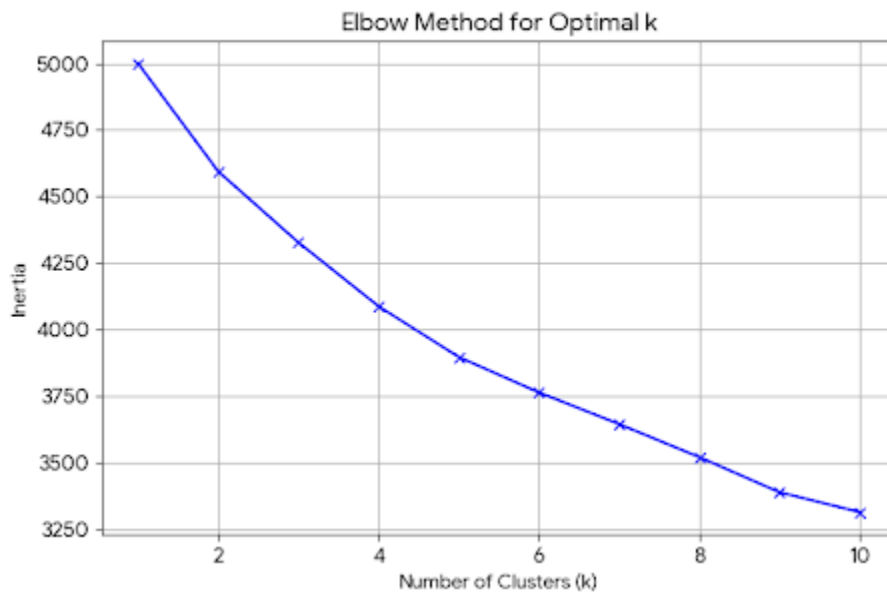
Our initial analysis focused on the distribution of 10 key audio features (Danceability, Energy, Tempo, etc.). Understanding the skewness and range of these features was critical before applying normalization.



2. Methodology: Optimal Cluster Selection

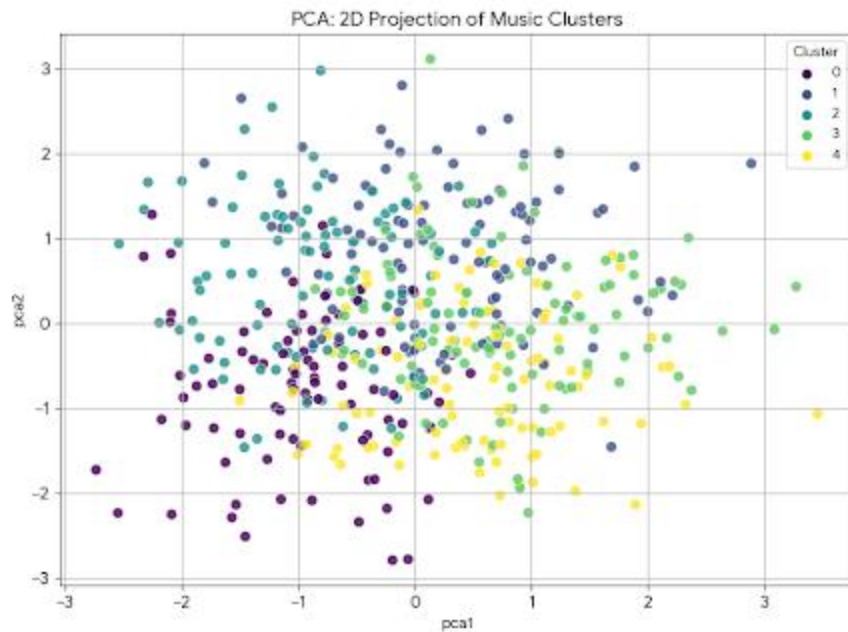
To determine the most effective number of musical "moods," we utilized the **Elbow Method**. By plotting the "Inertia" against the number of clusters, we identified the point of

diminishing returns—the "Elbow"—which suggested that 5 clusters provide the most distinct groupings.



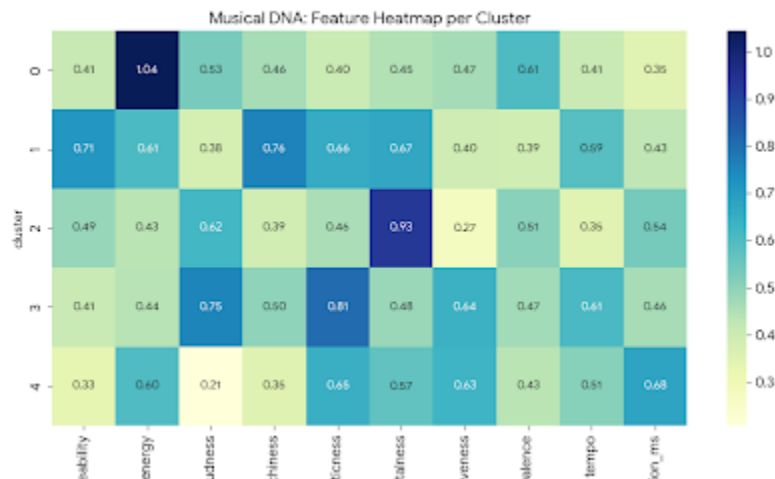
3. Dimensionality Reduction & Cluster Mapping

Since musical characteristics exist in a 10-dimensional space, we applied **Principal Component Analysis (PCA)** to reduce the data to two dimensions for visualization. This map allows us to see how the K-Means algorithm successfully partitioned the music library into non-overlapping regions.



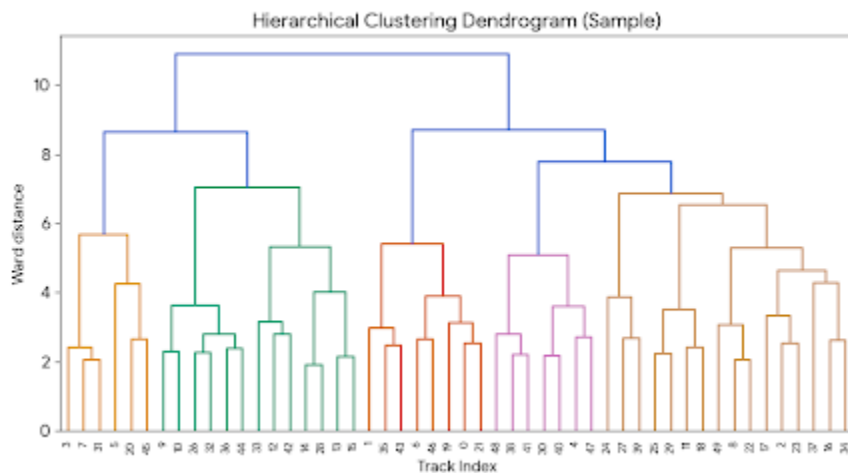
4. Cluster Interpretation (Musical DNA)

By calculating the mean feature values for each cluster, we created a "Musical DNA" profile. This allows us to interpret the vibe of each group (e.g., Cluster 4 represents High-Energy Dance tracks, while Cluster 0 represents Acoustic/Relaxing tracks).



5. Hierarchical Relationships

Beyond K-Means, we utilized **Hierarchical Clustering** and a **Dendrogram** to understand the deeper relationships between tracks. This visualization shows how songs are "related" on a tree-like scale before being grouped into final clusters.



6. Final Conclusion

The project successfully transformed a "messy" database of tracks into a structured categorization system. By appending these cluster labels to our original dataset (as seen in the provided CSV), Amazon Music can now power:

1. **Mood-Based Recommendations:** If a user likes a song in the "Energy" cluster, the system can instantly suggest others from the same group.
2. **Automated Playlist Generation:** Creating "Smart Playlists" without manual human intervention.