Title: *Amazon Music Clustering — Genre & Mood Segmentation*

by Kanish Midhun K

Objective:

To group songs from Amazon Music into meaningful categories (genres or moods) using **unsupervised machine learning**.

Tools & Libraries:
Python, Pandas, Scikit-learn, Seaborn, Matplotlib, Streamlit

Key Techniques:
EDA · Feature Scaling · K-Means Clustering · PCA · Cluster Profiling · Visualization

Problem Statement

With millions of songs on streaming platforms, manually tagging genres or moods is not scalable.

Goal:
Use audio features (danceability, energy, loudness, tempo, etc.) to automatically cluster songs that *sound alike*.

Applications:
- Personalized playlist generation 🎧
- Mood-based recommendation systems 🎶
- Artist trend analysis 📈

Dataset Overview

Dataset: single_genre_artists.csv

Features Used:

- danceability, energy, loudness, speechiness, acousticness
- instrumentalness, liveness, valence, tempo.

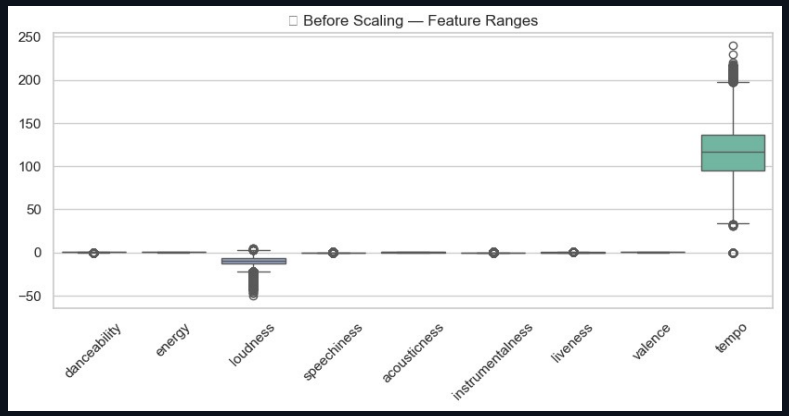
Removed Columns:

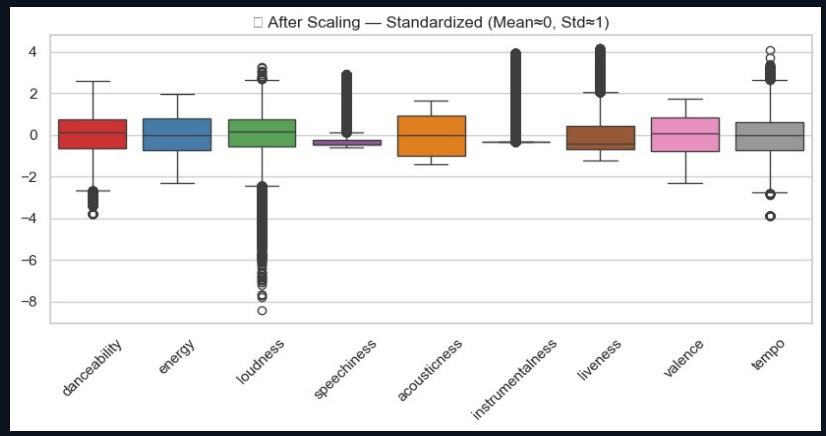
- track_id, track_name, artist_name (used later for mapping results)

Data Preprocessing

Steps Performed:

1. **Handled missing values** (if any)
2. **Selected relevant features** for clustering
3. **Applied StandardScaler** — standardized all features to mean=0, std=1
4. **Visualized scaling** using before/after boxplots





K-Means Clustering

Why K-Means?

- Simple and effective for numerical features
- Finds centroids based on Euclidean distance

Optimization:

Tested k = 2 → 10

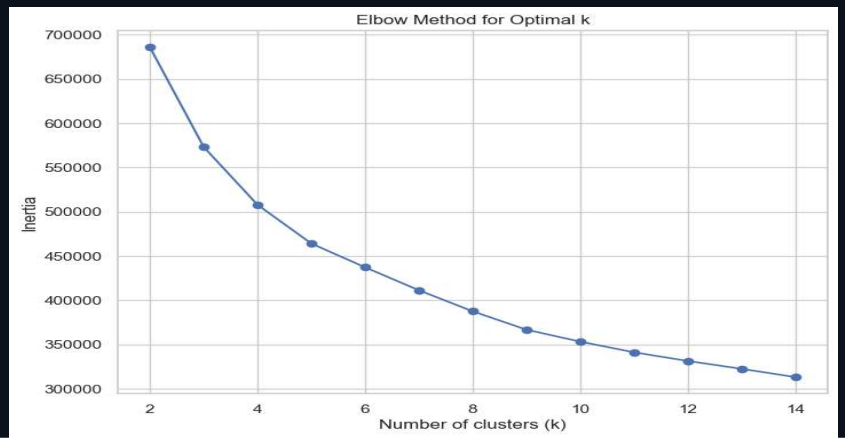
Evaluated using:

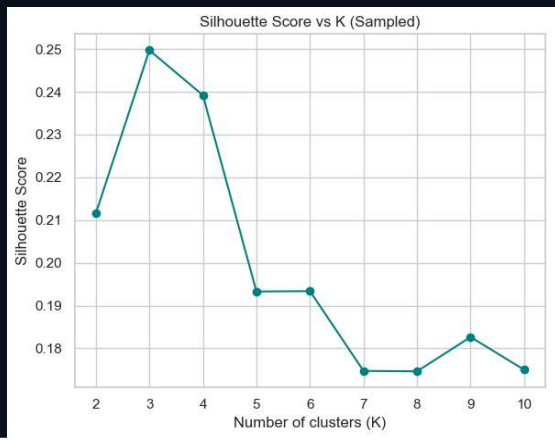
- Silhouette Score (↑ better)
- Davies–Bouldin Index (↓ better)
- Inertia (Elbow Method)

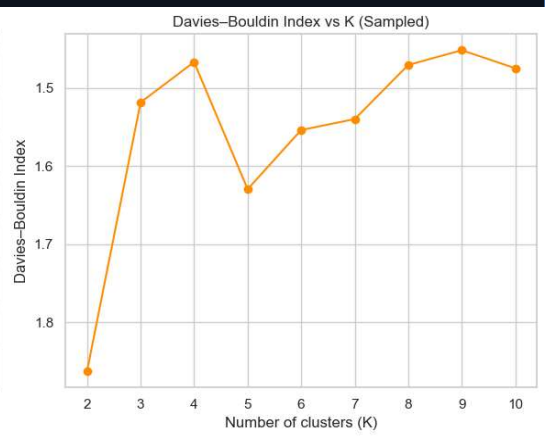
Final Model (k = 3)

Model: KMeans(n_clusters=3, random_state=42)

Interpretation:
Moderate separation — clear mood-based clusters form







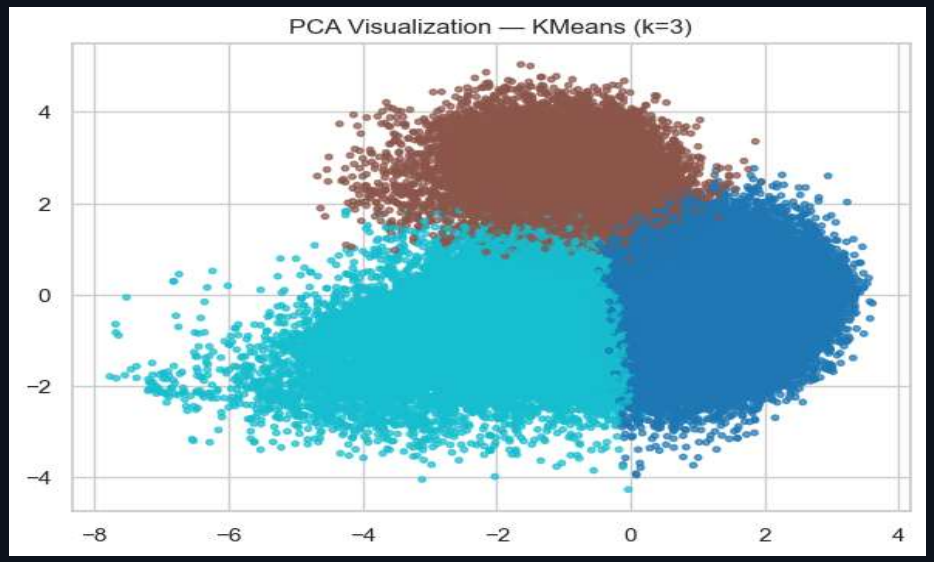
Cluster Profiles

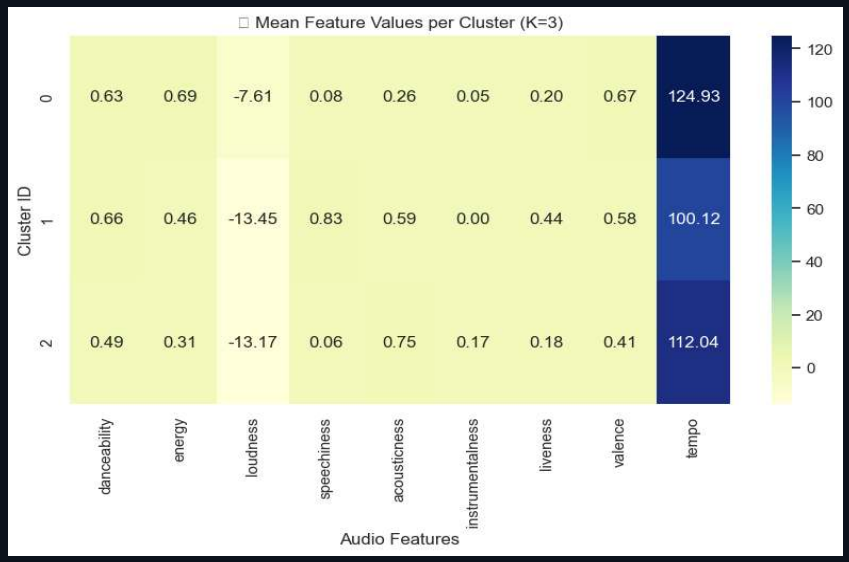
Cluster	Description	Key Traits
0	**Party / Upbeat** 🎉	High energy, high danceability, loud
1	**Chill Acoustic** 🌙	High acousticness, low energy
2	**Instrumental / Ambient** 🎧	High instrumentalness, low speechiness

Visualization

Techniques Used:

- PCA 2D Scatter Plot — visualize cluster separation
- Feature Heatmap — compare mean feature values per cluster





Key Takeaways

- ✅ Successful mood-based clustering of songs
- 🎧 3 natural groupings: *Upbeat*, *Chill*, *Instrumental*
- 🧠 Combined ML + Music Analytics + Visualization
- 💡 Ready for playlist recommendation integration