

# K-Means Clustering for Music Recommendations

Name: Kenneth Gutierrez King  
Project Track: Machine Learning  
Major: Data Analytics

## Introduction

The growth of digital music makes discovering new songs increasingly difficult. Music streaming platforms use machine learning to group songs and generate personalized recommendations. This project leverages Spotify audio features, such as energy, valence, and danceability, to cluster songs with similar characteristics and provide content-based recommendations.

## Objective

The project aims to build a recommendation system that suggests songs based on their musical attributes. Using K-Means clustering on Spotify audio features, the system identifies groups of similar tracks. Users can input a song and artist to receive recommendations, optionally filtered by popularity.



Streamlit App

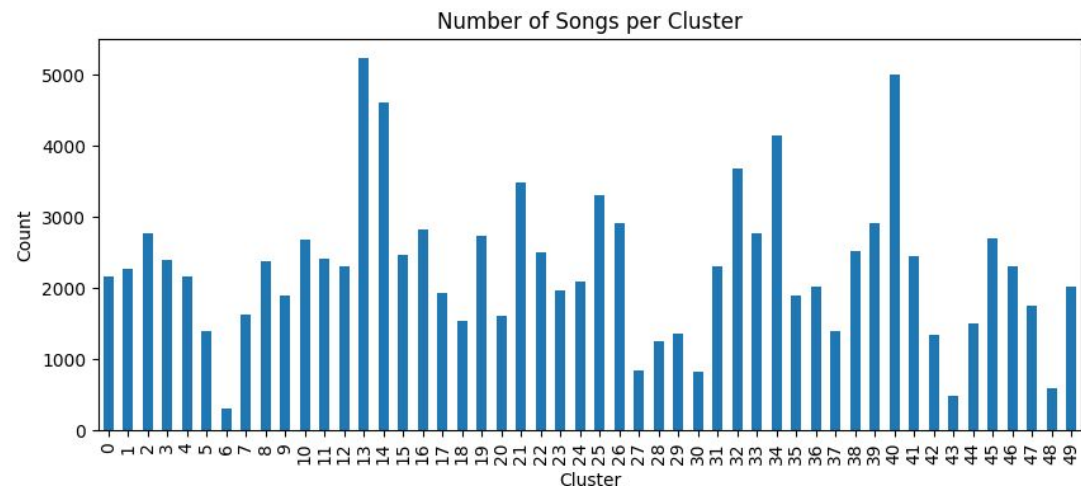
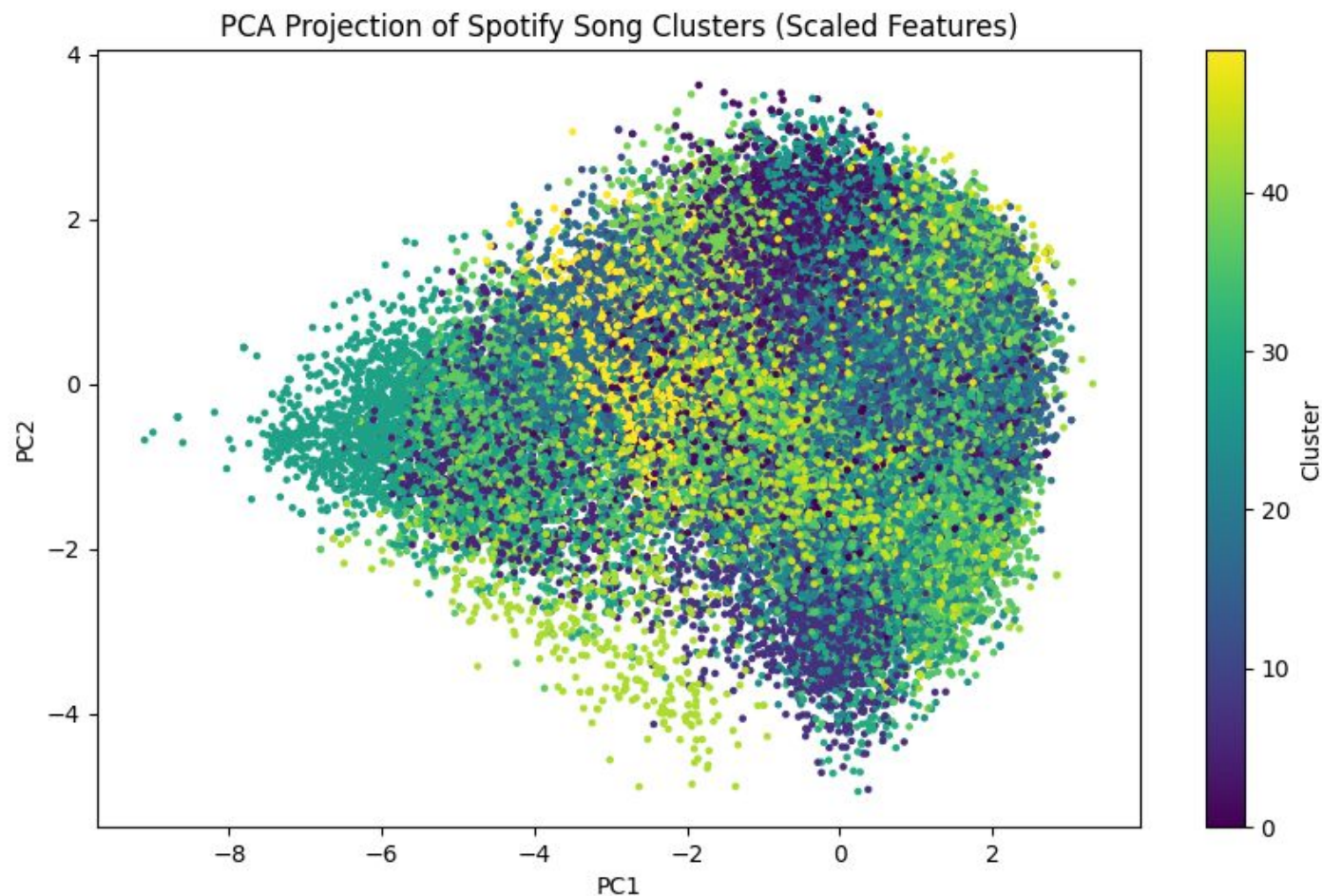
## Methodology:

### Dataset

Spotify API data from Kaggle containing song metadata (title, artist, album) and audio features (danceability, energy, valence, etc.) used for clustering and recommendations

### Modeling

Songs were grouped using K-Means clustering based on audio features such as energy, valence, and danceability. Due to the large dataset, 50 clusters were selected to balance detail and interpretability. A PCA scatter plot was used to visualize songs in 2D space, while a heatmap of cluster centroids summarized the average characteristics of each cluster.



### Tools used

- Python: programming language
- Pandas/NumPy: data handling
- Scikit-learn: clustering & PCA
- Joblib: save/load models
- Matplotlib/Seaborn: visualizations
- Streamlit: interactive dashboard

## Results:

A content-based music recommendation system was built using K-Means clustering on Spotify audio features. The model groups songs with similar musical characteristics, allowing users to receive personalized recommendations by selecting a song and artist. The system demonstrates how unsupervised learning can organize music and suggest tracks that match user preferences.

## Sources and Links:

Kaggle Dataset:  
<https://www.kaggle.com/datasets/solomonameh/spotify-music-dataset>

My LinkedIn:  
<https://www.linkedin.com/in/kennethgutierrezking>