Spotify Songs' Genre Segmentation

Objective

To develop an automated system for clustering songs based on their audio features, enabling better music recommendations. The project aims to segment songs into groups sharing similar characteristics to personalize user experiences.

Dataset Information

- Source: Provided CSV file containing Spotify song data.
- Dataset Features:
    - Audio Features: Danceability, Energy, Tempo, Valence, Acousticness, Instrumentalness, Loudness, etc.
    - Metadata: Playlist Genre, Playlist Name.
- Size: Dataset size varies depending on the source (assumed multiple rows and attributes).
- Target: No explicit target variable (unsupervised learning).

Project Workflow

1. Data Preprocessing:
    - Missing Value Handling: Handled missing data using forward-fill or imputation techniques.
    - Feature Scaling: Standardized numerical features for uniform clustering.
    - Feature Encoding: Encoded categorical variables like playlist genres or names.
2. Exploratory Data Analysis (EDA):
    - Correlation Matrix:
        - Analyzed relationships among audio features.
        - Strong correlations observed, e.g., Danceability and Energy.
    - Visualizations:
        - Histograms and box plots for feature distributions.
        - Pair plots to explore relationships between features.
    - Key Insights:

        - Higher danceability and energy often align with upbeat genres.
        - Acousticness and instrumentalness are prominent in classical or ambient genres.

3. Clustering:
    - Clustering Algorithm: K-Means Clustering.
    - Optimal Clusters:
        - Determined using the Elbow Method (optimal clusters = 5).
    - Cluster Visualization:
        - Reduced dimensionality with PCA for 2D visualization.
        - Plotted clusters to show groupings of songs with similar features.
4. Model Results:
    - Songs were grouped into five distinct clusters based on audio features.
    - Cluster Characteristics:
        - Cluster 1: High energy, low valence (intense tracks).
        - Cluster 2: High acousticness and instrumentalness (ambient or classical tracks).
        - Cluster 3: Balanced features (general pop songs).
        - Cluster 4: High tempo and danceability (party tracks).
        - Cluster 5: Low energy and tempo (calm or chill tracks).
5. Final Output:
    - Labeled dataset with clusters assigned to each song.
    - Exported cluster information for integration into a recommendation engine.

Future Recommendations

1. Data Enrichment:
    - Include additional features like user preferences, streaming frequency, or popularity metrics.
2. Advanced Models:
    - Experiment with clustering algorithms like DBSCAN or hierarchical clustering for better segmentation.

3. Real-Time System:
    - Integrate with a streaming platform to provide real-time recommendations.
4. Dynamic Recommendations:
    - Use collaborative filtering with the clusters for better personalization.

Results

- Key Findings:
    - Songs grouped effectively into clusters based on audio features.
    - Each cluster represents a unique genre or mood.
    - Playlist genres align well with audio feature clusters.
- Visualization Highlights:
    - PCA Scatter Plot: Showed clear separation of clusters.
    - Elbow Plot: Helped identify the optimal number of clusters.

Conclusion

The project successfully implemented an unsupervised learning model to cluster Spotify songs based on audio features. The results can be used to build a music recommendation system that suggests songs within the same cluster or similar clusters, enhancing user satisfaction.