Spotify Audio Feature Analysis

Spotify Audio Feature Analysis: Exploring the Relationship Between Musical Attributes and Popularity.

Python Libraries Used:

Pandas, NumPy, Matplotlib, Seaborn

Domain:

Data Analytics / Business Analytics

Dataset:

Public Spotify Dataset (Kaggle)

Project Overview

Spotify provides multiple audio features for each song such as danceability, energy, loudness, tempo, valence, acousticness, and popularity. This project analyzes whether these musical attributes influence a song's popularity and explores the relationships between different audio features.

Objectives

- Identify the Top 5 Most Popular Artists
- Identify the Top 5 Most Popular Tracks
- Identify artists with the most danceable songs
- Analyze the relationship between audio features and popularity

Dataset Overview

Key Features Used:

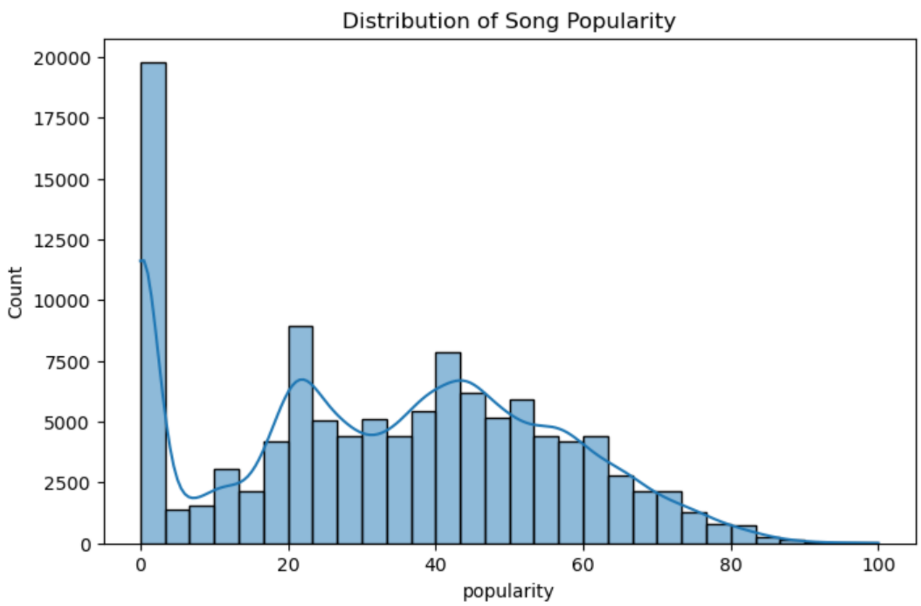
- **Danceability** – How suitable a track is for dancing (0–1).
- **Energy** – Intensity and activity level.
- **Valence** – Musical positivity (happy vs sad).
- **Loudness** – Overall loudness in decibels.
- **Tempo** – Beats per minute.
- **Acousticness** – Confidence measure of acoustic sound.
- **Popularity** – Spotify popularity score (0–100).

Data Cleaning Steps:

- Removed duplicates
- Checked for missing values
- Selected relevant variables for analysis

Exploratory Data Analysis (EDA)

1. Distribution Analysis: Plotted histogram to understand distribution of features. Observed whether features were skewed or normally distributed.



Interpretation

Popularity distribution is right-skewed. Most songs have low popularity, and only a small number reach very high popularity. This reflects a long-tail distribution where a few songs dominate streams.

The histogram of song popularity indicates that the distribution is **positively skewed** (right-skewed). A large proportion of songs have low popularity scores, while only a small number of songs achieve very high popularity. There is a noticeable concentration of songs in the lower popularity range (0–20), suggesting that most tracks in the dataset do not reach mainstream success. The frequency gradually decreases as popularity increases, forming a long right tail toward scores close to 100. This pattern suggests:

- Popularity is not normally distributed

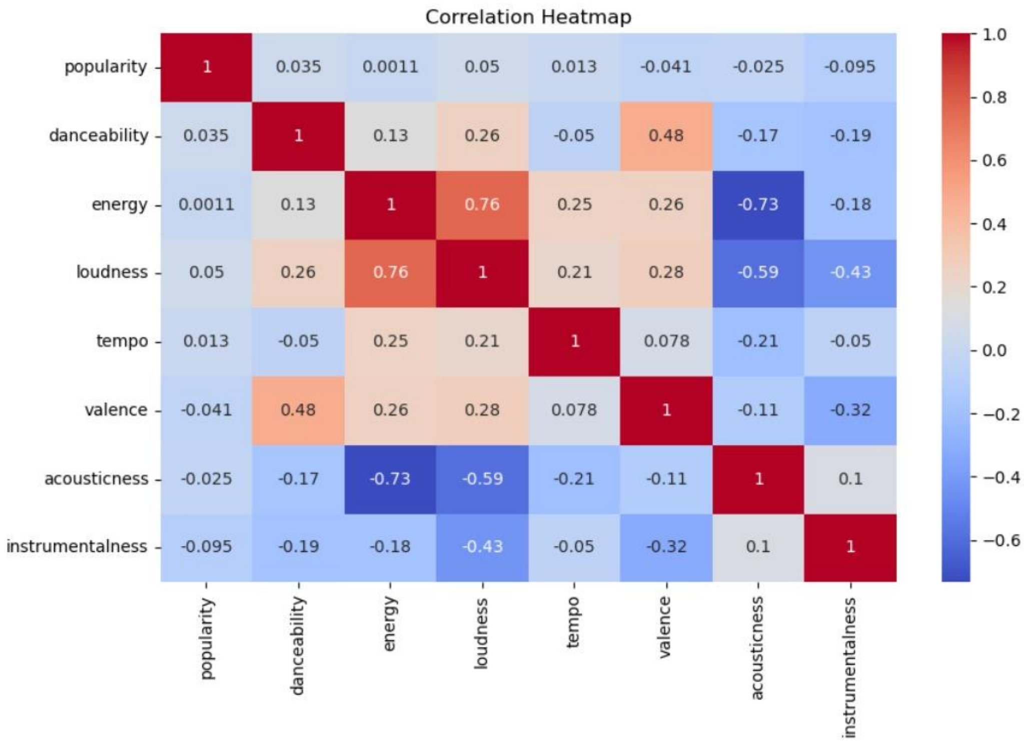
- The dataset contains many less-streamed or niche songs

• A small fraction of songs dominates listener attention

The Spotify dataset demonstrates a long-tail distribution, where a majority of songs have low popularity, and only a few songs achieve very high popularity.

Correlation with Popularity

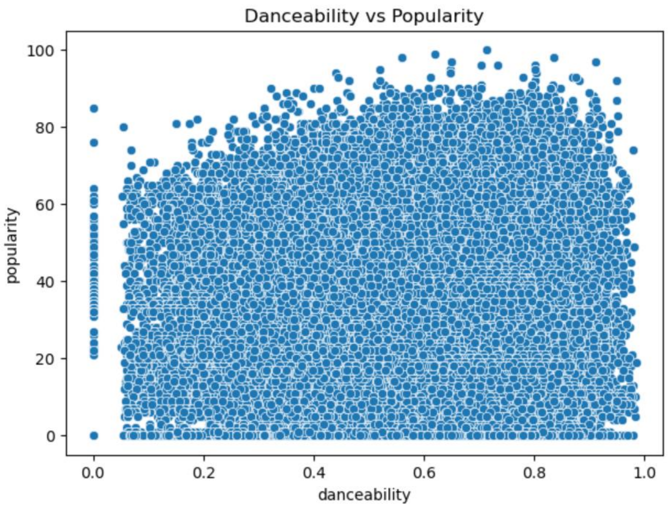
2. Correlation Analysis: Used correlation heatmap to identify relationships.
Focused primarily on: **Danceability vs Popularity, Energy vs Popularity** & **Loudness vs Popularity**.



Interpretation

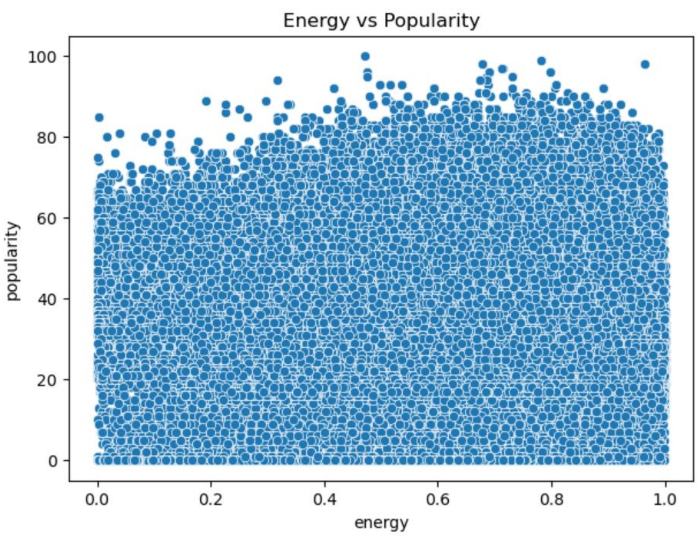
No strong correlation exists between popularity and any single audio feature. Being energetic, danceable, loud, or happy does not guarantee popularity. Popularity depends on multiple external factors.

3.Scatter Plot: Scatter Plots Visualized relationships between features and popularity. Checked for linear or non-linear patterns.



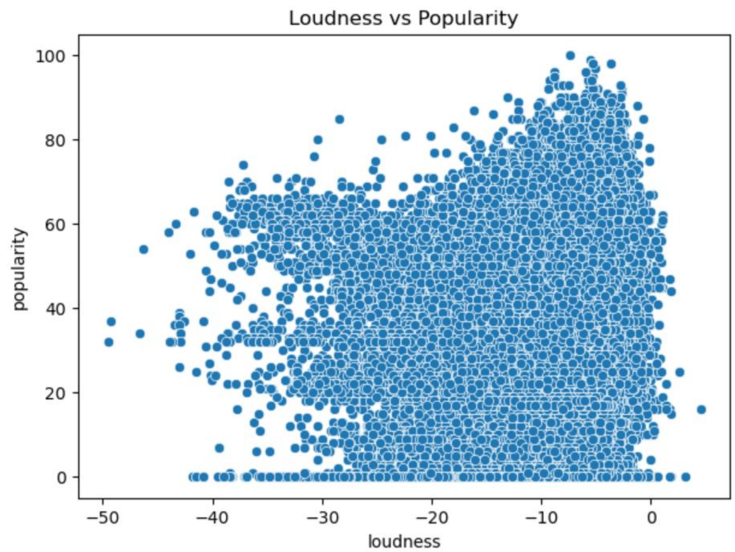
Popularity & Danceability (0.035)

There is a very weak positive correlation between danceability and popularity. This indicates that songs that are more dance-friendly are not necessarily more popular. **Conclusion:** Danceability alone is not a strong factor in determining a song's success.



Popularity & Energy (0.001)

The correlation between energy and popularity is almost zero. This suggests that high-energy songs do not automatically perform better in terms of popularity.
Conclusion: Energy level does not significantly influence a song's popularity.



Popularity & Loudness (0.05)

There is a very weak positive relationship between loudness and popularity.
While slightly louder songs may show a marginal tendency toward higher popularity, the effect is minimal.
Conclusion: Loudness has a negligible impact on overall popularity.

Overall Conclusion on Popularity

No strong relationship was found between popularity and any individual audio feature.
This indicates that popularity is multi-dimensional and influenced by several external factors such as:

- Artist brand value
- Marketing efforts
- Social trends
- Virality
- Listener behavior
- Playlist algorithms

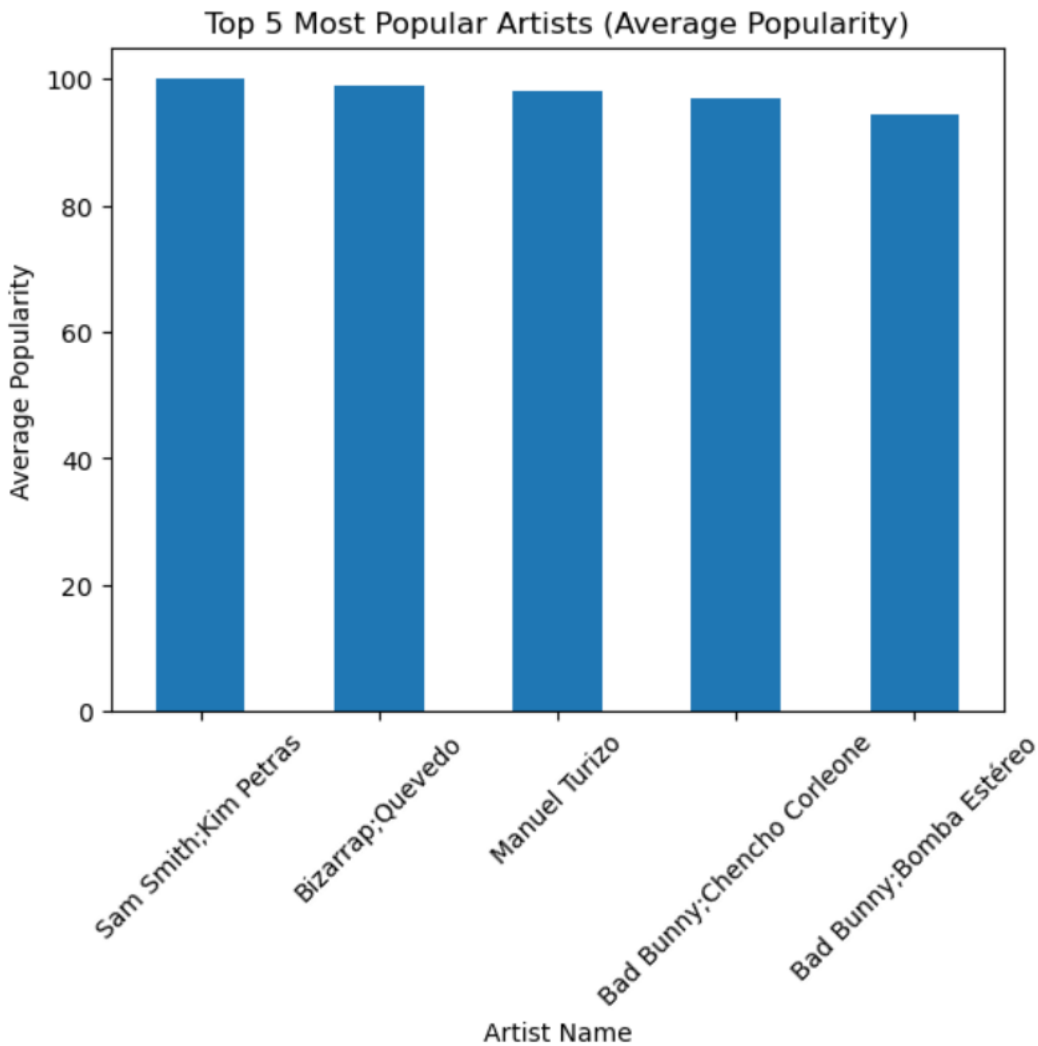
Key Insight
Audio features alone are not sufficient to predict whether a song will become a hit.

Relationships Between Audio Features

- Energy & Loudness (0.76): Strong positive relationship.
- Energy & Acousticness (-0.73): Strong negative relationship.
- Danceability & Valence (0.48): Moderate positive relationship.

Musical attributes are structurally connected, but these structural relationships do not directly translate into popularity.

Top 5 Most Popular Artists



The horizontal bar chart highlights the five most popular artists in the dataset:

- Sam Smith & Kim Petras
- Bizarrap & Quevedo
- Manuel Turizo
- Bad Bunny & Chencho Corleone
- Bad Bunny & Bomba Estéreo

Key Observations

High Popularity Concentration:
All artists have popularity scores between 95–100, reflecting extremely high listener engagement.

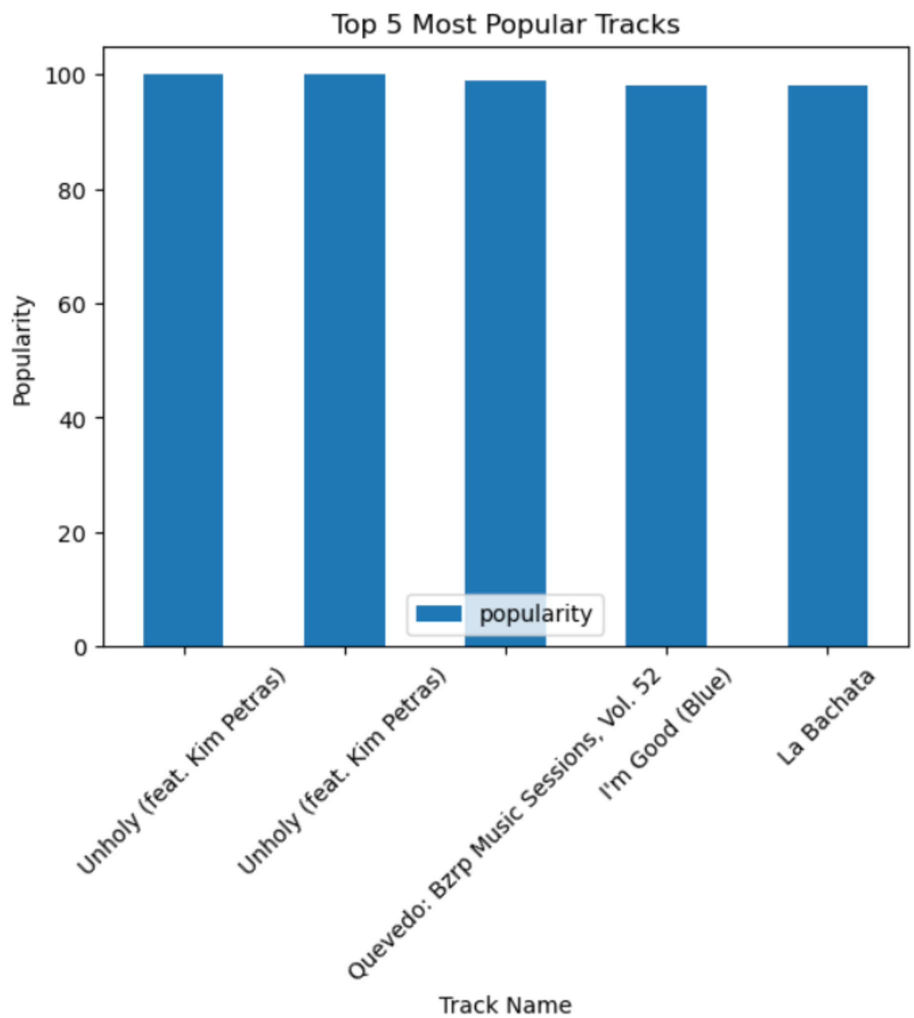
Minimal Variation:
The difference in scores among the top five is very small, showing intense competition at the top level.

Strong Presence of Collaborations:
Most top entries are collaborative tracks, suggesting that partnerships may help expand reach and boost streams.

Conclusion

Popularity in the dataset is concentrated among a small group of elite artists. The dominance of collaborations indicates that strategic partnerships can enhance visibility and streaming performance.

Top 5 Most Popular Tracks

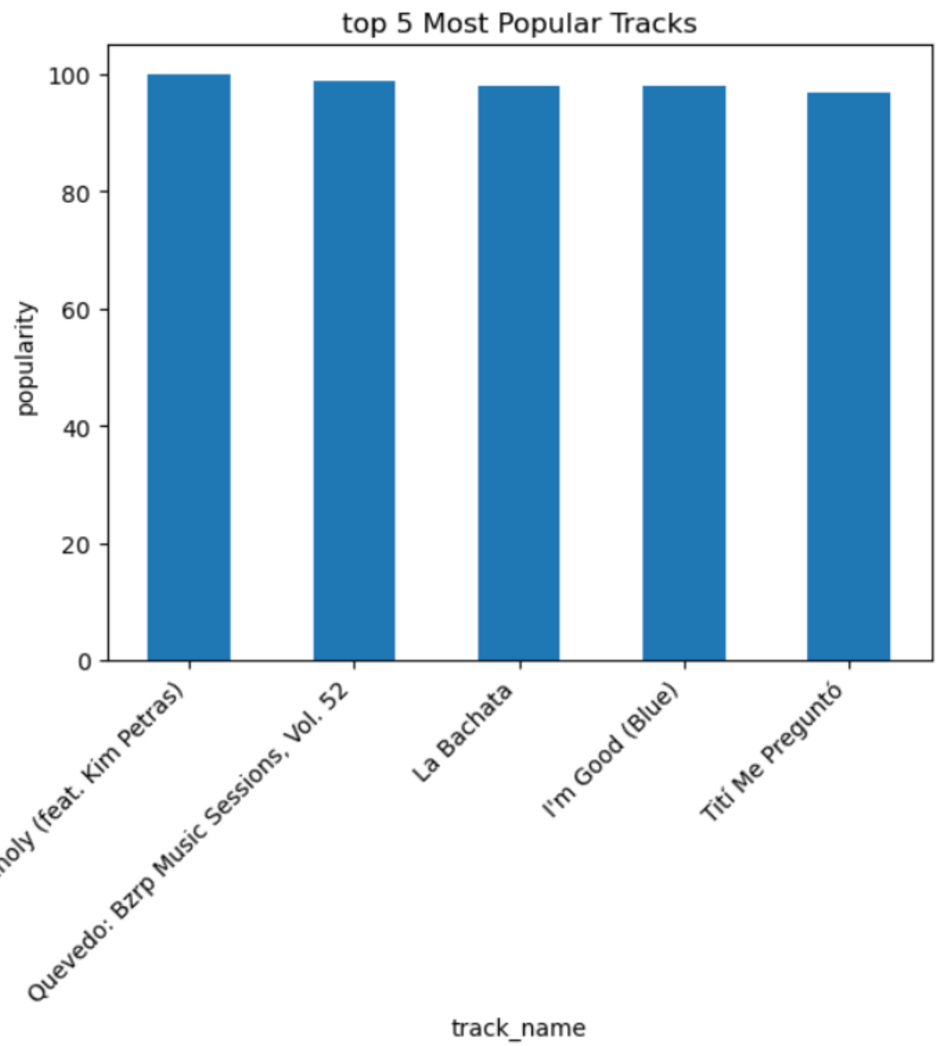


The bar chart displays the five tracks with the highest popularity scores in the dataset. All five songs have scores between 98–100, indicating extremely high listener engagement.

The difference in popularity among these tracks is very small, showing that top-performing songs cluster closely at the upper end of the scale.

The song *"Unholy"* appears twice, which Indicate Duplicates.

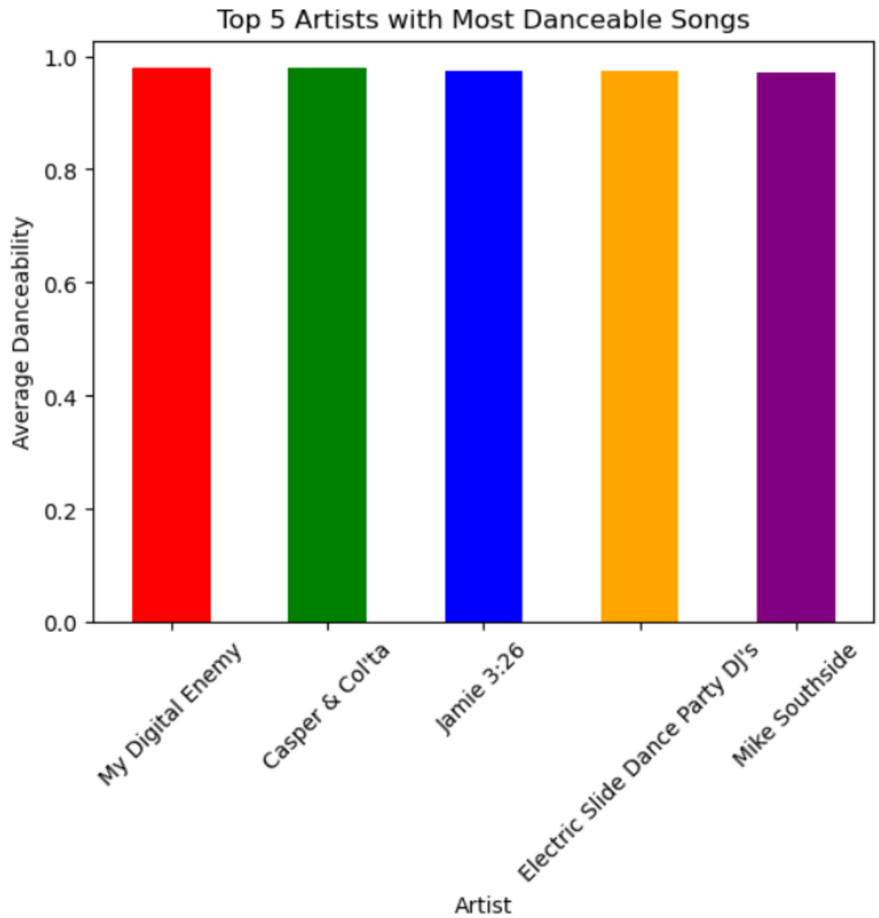
Duplicates Removed



Conclusion

The minimal variation among the top five tracks suggests that once a song reaches peak popularity, differences in scores become marginal.

Top 5 Artists with the most danceable songs



The chart highlights the top five artists with the highest average danceability scores. All selected artists have danceability values between **0.97 and 0.99**, showing consistently high rhythm-focused production.

The small variation among them indicates that they specialize in creating highly dance-friendly music, likely suited for clubs, parties, and energetic environments.

Conclusion

The narrow difference in average danceability confirms that these artists focus on producing strong, movement-oriented tracks.

However, as seen in earlier analysis, high danceability does not strongly correlate with popularity. This means producing highly danceable music does not necessarily guarantee greater commercial success.

Overall conclusion

Music features are structurally interconnected, but popularity is multi-dimensional. Hit songs are driven by branding, marketing, trends, and platform algorithms — not just sound characteristics.