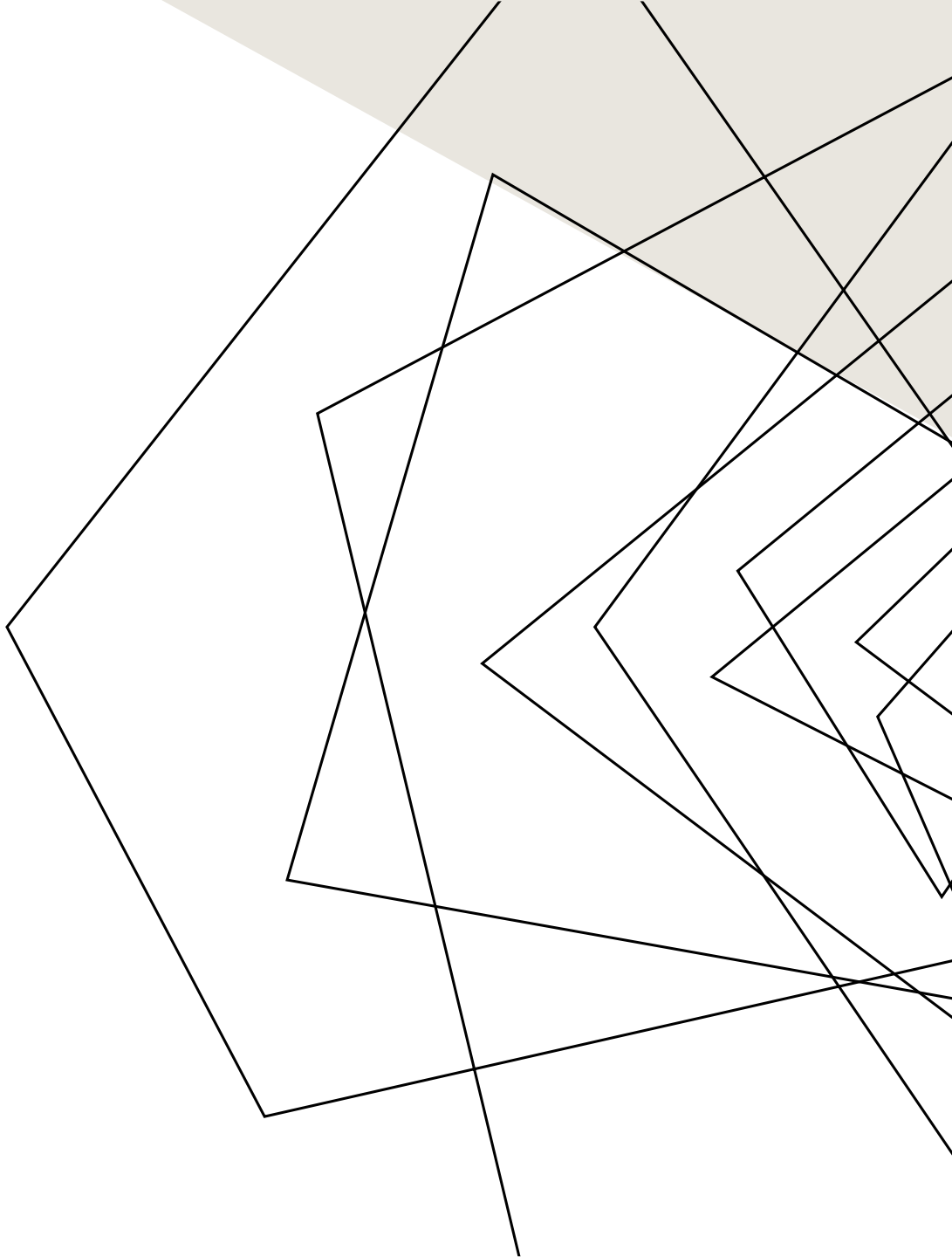
SPOTIFY DATA ANALYSIS

An Exploratory Data Analysis (EDA) Story to Uncover Reliable Insights

Presented by: Md Saquib

AGENDA

1. Project Overview
2. Dataset Analysis
3. Insights and Observations
4. Conclusion



OVERVIEW

Objective

The objective of this analysis is to explore the Spotify Tracks dataset to uncover meaningful patterns and insights related to song characteristics, popularity trends, and artist performance.

Since the objective is to understand the factors that contribute to a track's popularity and to examine how these features vary across languages, genres, and time periods.

WE NEED TO STUDY THE FOLLOWING FEATURES:

- **Danceability** – Measures how suitable a track is for dancing, considering rhythm stability, beat strength, and tempo.
- **Energy** – Reflects the intensity and activity level of a song, capturing how dynamic or lively it feels.
- **Valence** – Represents the musical positivity of a track; high valence sounds happy or cheerful, while low valence feels sad or moody.
- **PCA Scatter** – Visualizes the distribution of tracks based on their key audio features, helping to identify clusters, similarities, and patterns that influence popularity.
- **Acousticness** – Estimates how acoustic or non-electronic a track is; higher values indicate fewer synthesized or electronic sounds.

ANALYSIS AND OBSERVATIONS

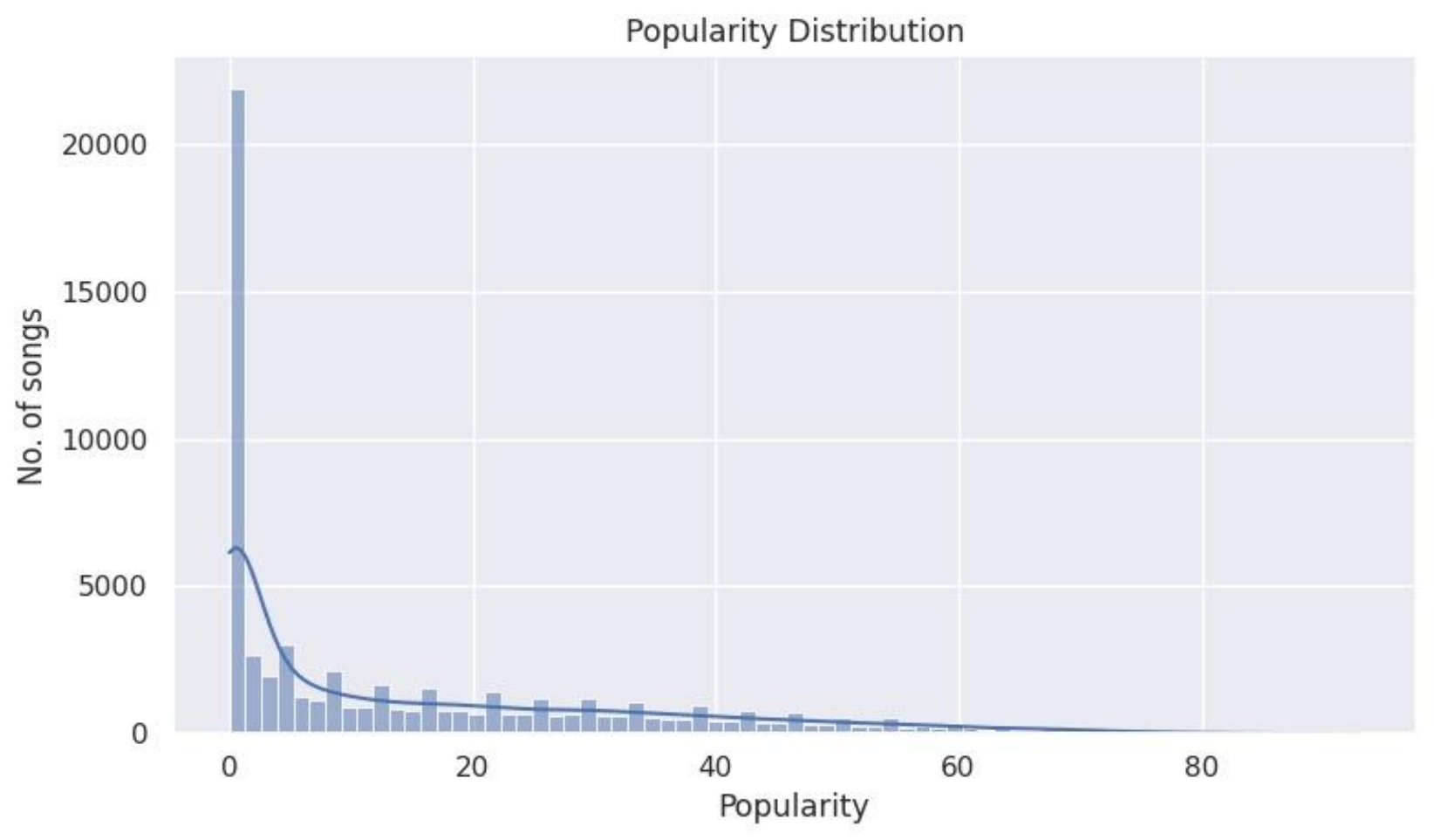
DATASET ANALYSIS

Total Records: 62317 (With 22 attributes)

Each record represents a single track and includes:

- **Track Information:** Includes the track name, artist, album, release year, language, and track URL.
- **Popularity Score:** A numerical value representing how well the song is performing.
- **Audio Features:** Captures characteristics such as danceability, energy, valence, acousticness, instrumentalness, speechiness, tempo, loudness, liveness, and more.
- **Musical Properties:** Details the key, mode, and time signature of the track.

POPULARITY DISTRIBUTION

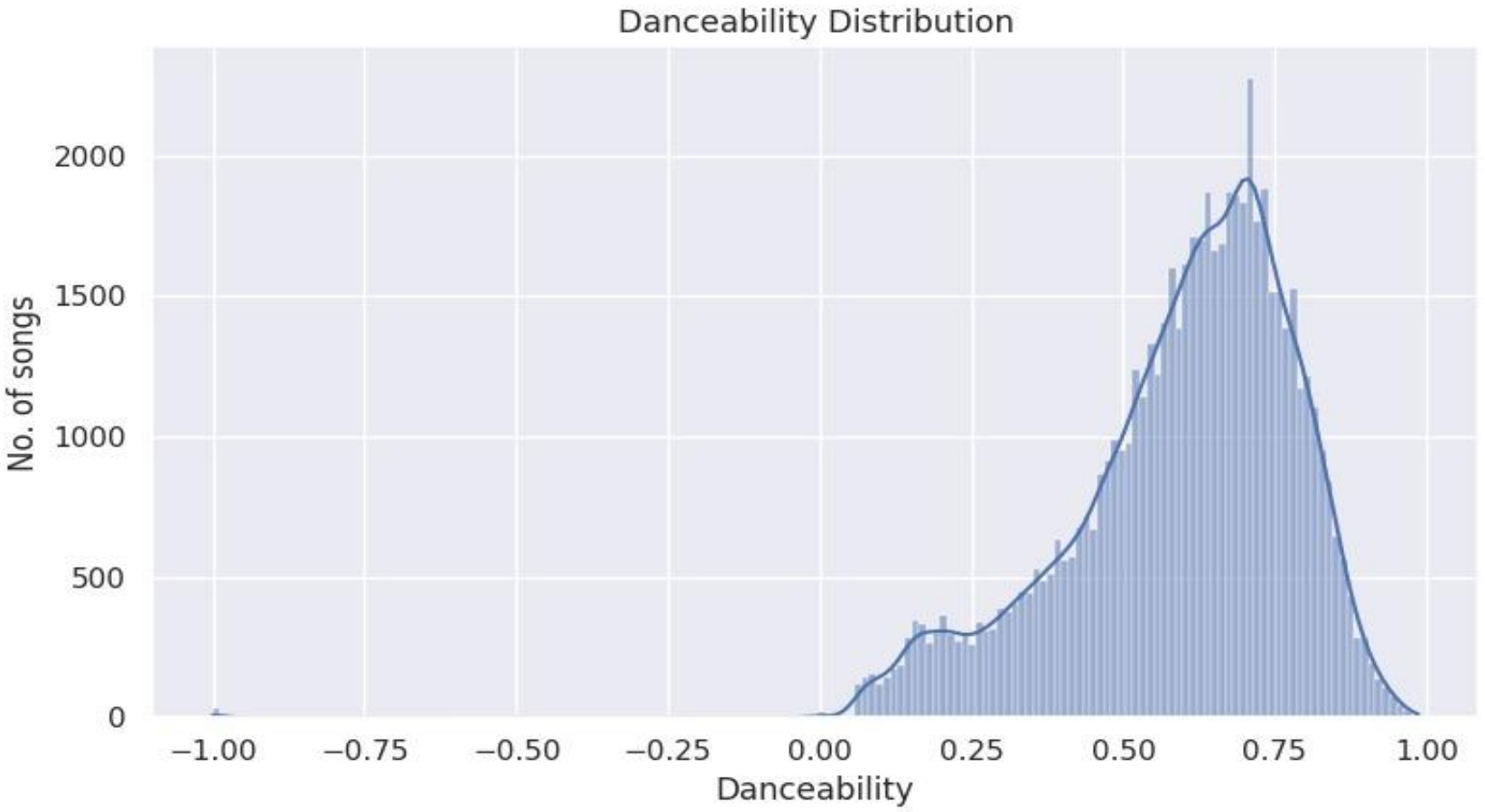


- Most songs cluster in the **moderate popularity range**, indicating that extremely popular hits are relatively rare.

- The distribution is **right-skewed**, with a small number of songs achieving very high popularity compared to the majority.

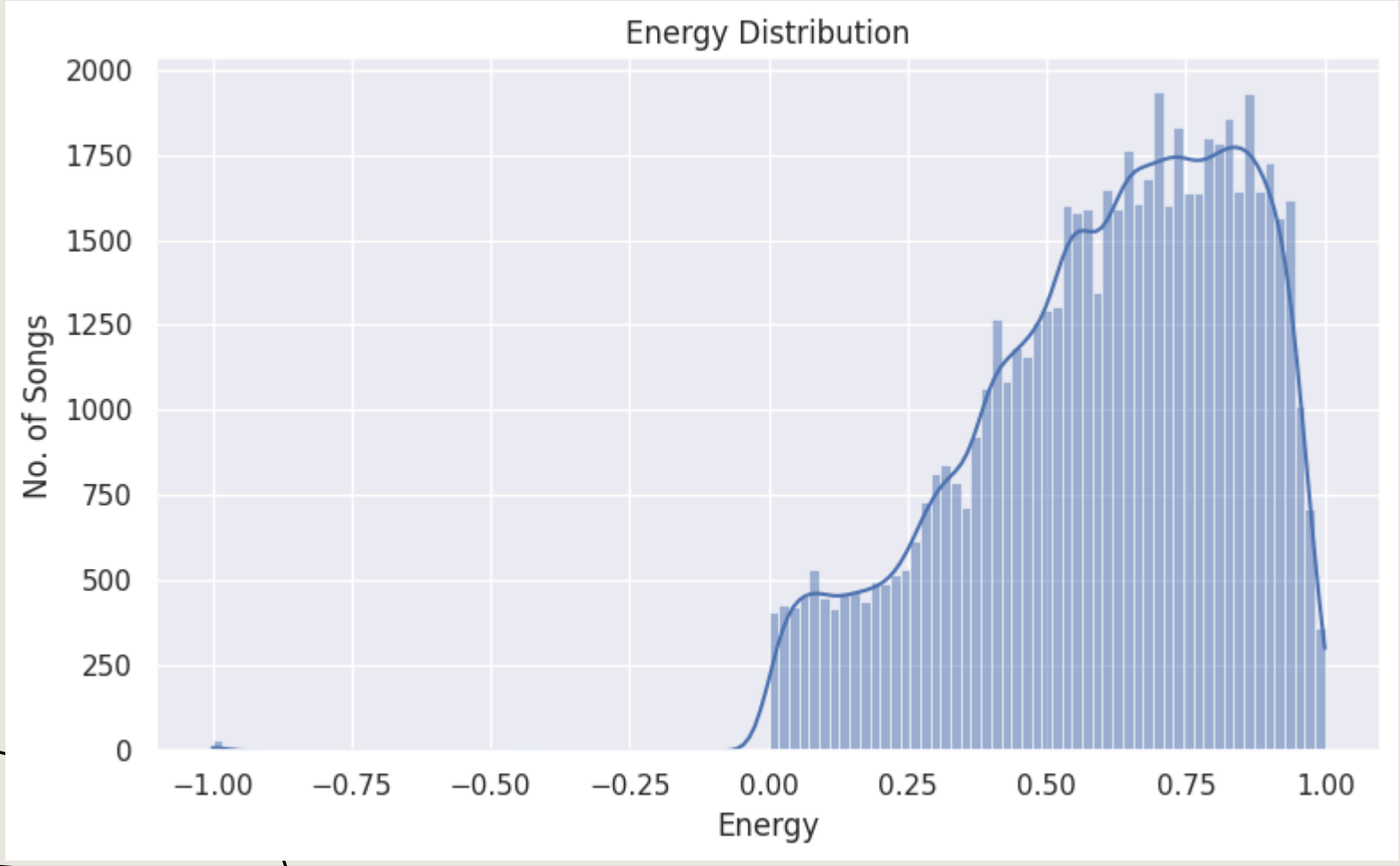
- The KDE curve highlights a **single main peak**, suggesting that most tracks receive a similar, average level of listener engagement rather than extreme popularity.

DANCEABILITY DISTRIBUTION



The danceability distribution shows that most tracks in the dataset have **moderate danceability**, clustering around 0.5–0.7. Extremely low or high danceable songs are rare, indicating the dataset is dominated by mainstream tracks with balanced rhythm and energy.

ENERGY DISTRIBUTION

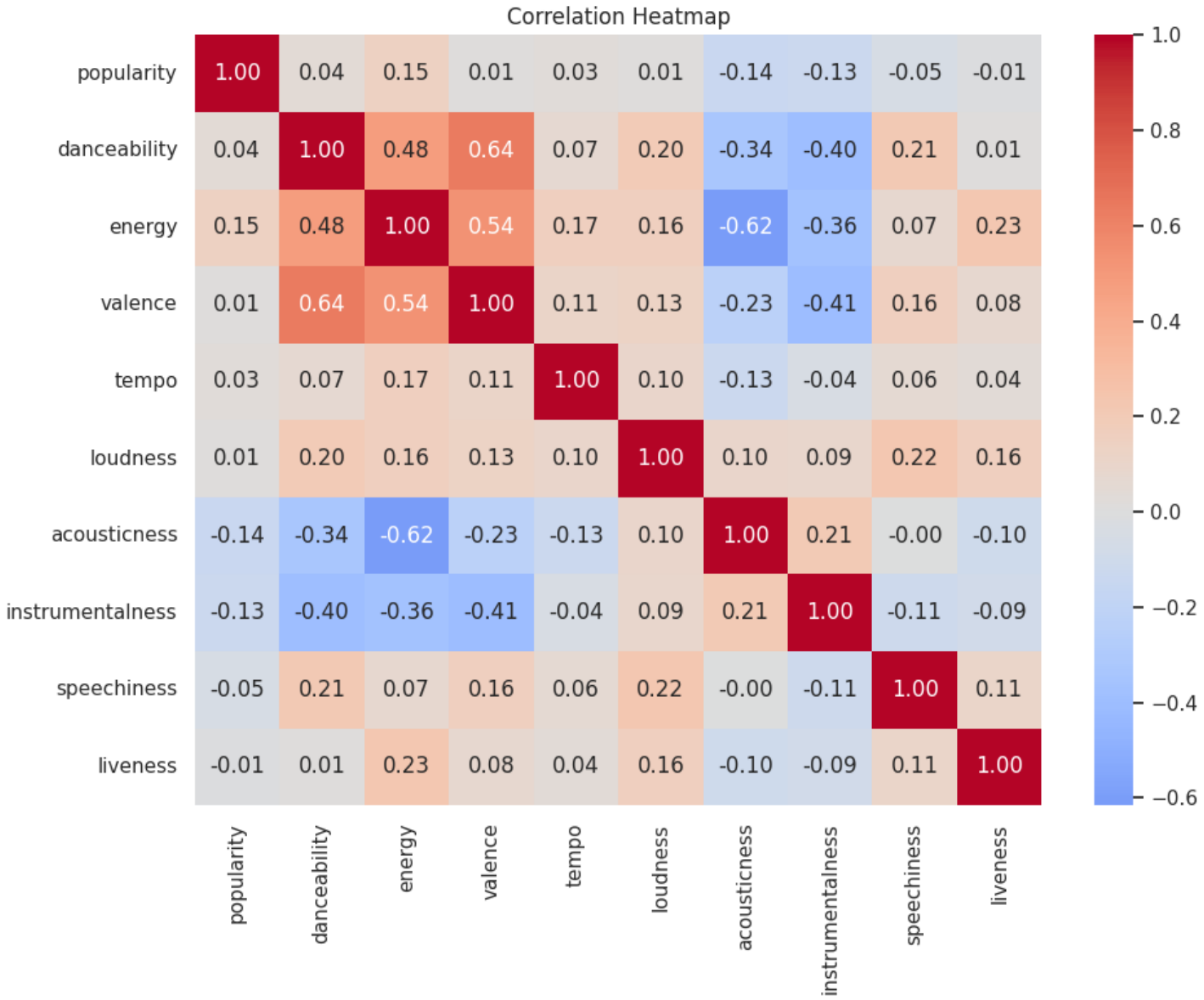


- Most songs have **moderate to high energy** (typically around 0.5–0.8), suggesting the dataset includes a majority of upbeat or lively tracks.

- Very low-energy songs (calm or soft tracks) are **less common**, indicating fewer slow ballads or acoustic tracks

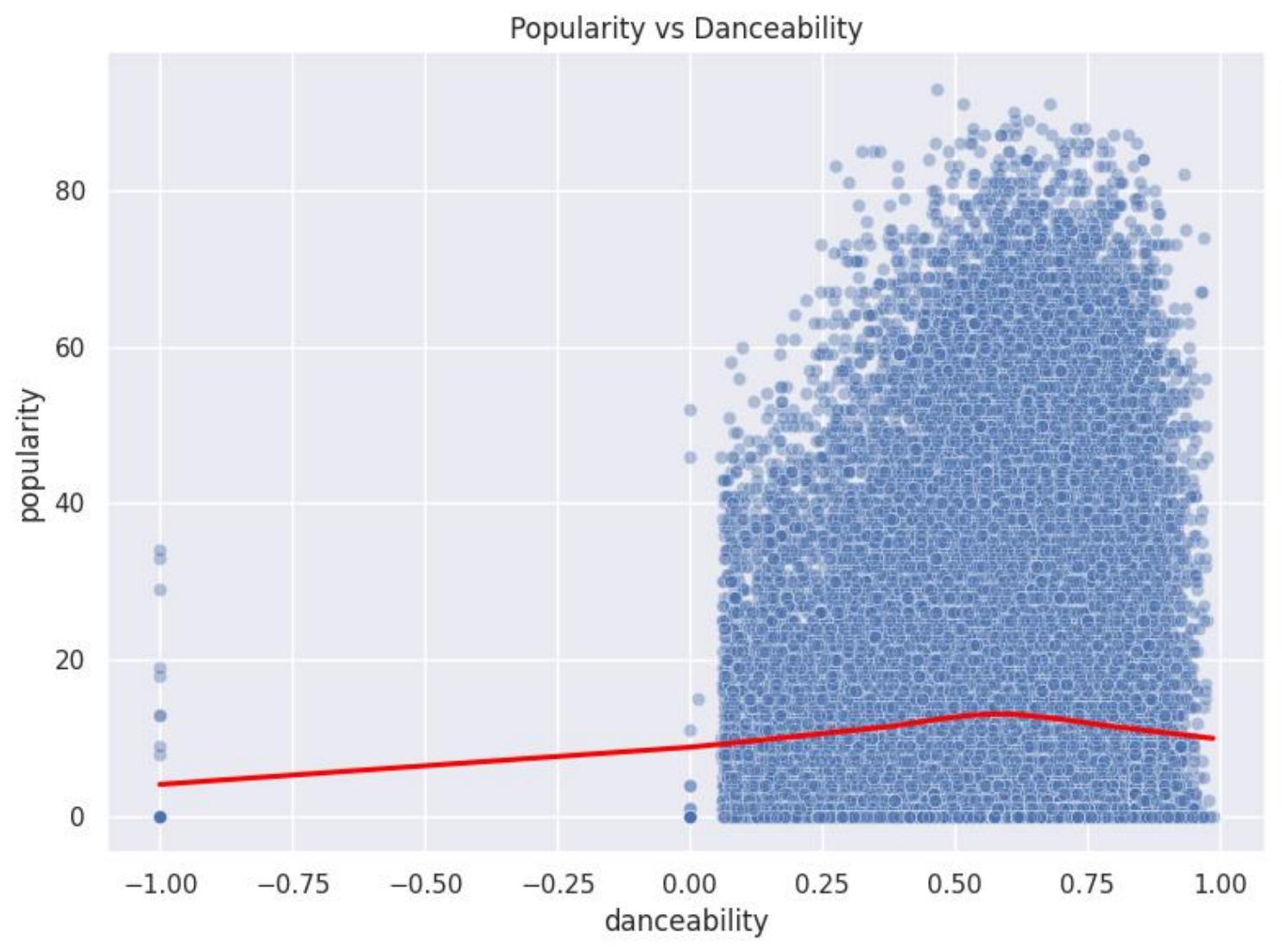
- The distribution may show a slight **right skew**, meaning a small portion of songs are extremely energetic, possibly representing dance, electronic, or pop hits.

CORRELATION HEATMAP



The correlation heatmap shows that **energy, loudness, and danceability are positively related**, while **acousticness and instrumentalness tend to be negatively correlated** with these features. Popularity shows only **moderate links** with audio attributes, indicating that hit songs are influenced by multiple factors rather than a single feature.

POPULARITY VS DANCEABILITY

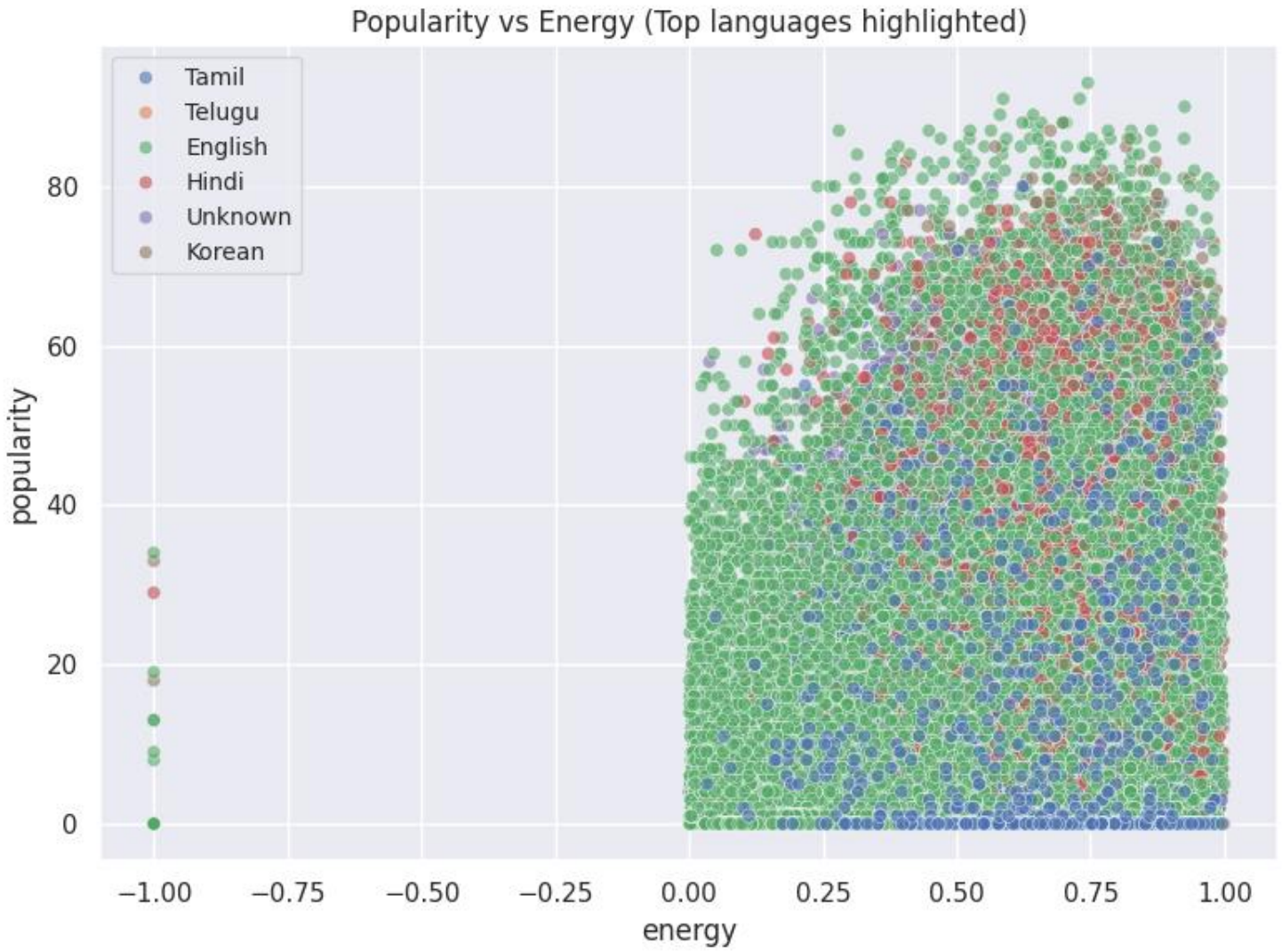


- **Slight positive trend**: The LOWESS curve suggests that songs with higher danceability tend to be somewhat more popular, but the relationship is not very strong.

- **Wide spread of popularity**: Many moderately danceable songs have both low and high popularity, indicating danceability alone does not guarantee a hit.

- **Outliers**: A few extremely popular songs exist across different danceability levels, showing that other factors (like artist, genre, or marketing) also influence popularity.

POPULARITY VS ENERGY (BY LANGUAGE)

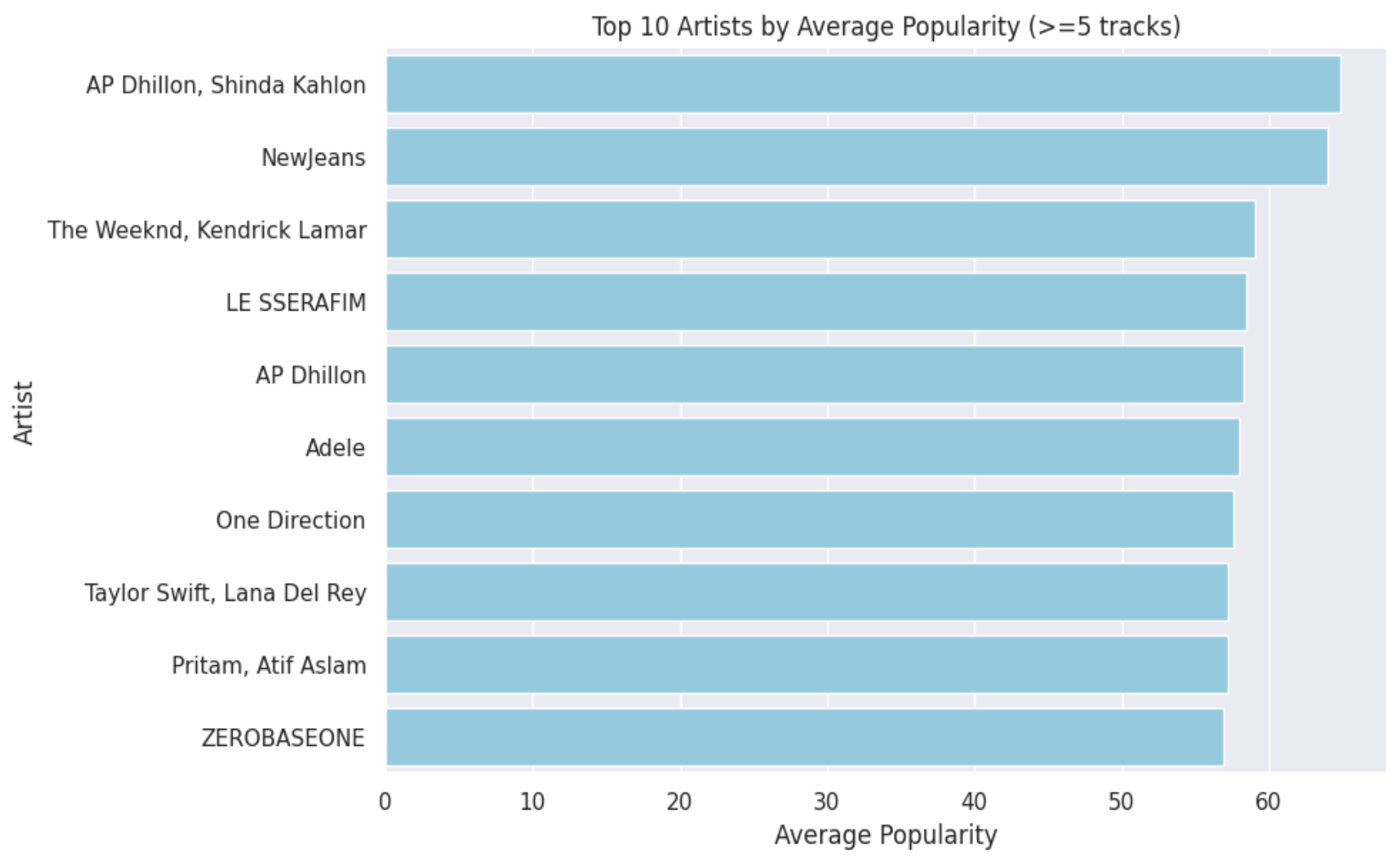


- **Moderate positive trend**: Songs with higher energy tend to have slightly higher popularity across most languages, though the trend is not very strong.

- **Language-specific patterns**: Some languages (e.g., English or Spanish) dominate the higher popularity and higher energy region, suggesting hits are more frequent in these languages.

- **Diverse spread**: Popularity varies widely at all energy levels, indicating that energy alone does not determine a song's success, and language may play a role in audience reach.

TOP ARTISTS BY AVG POPULARITY

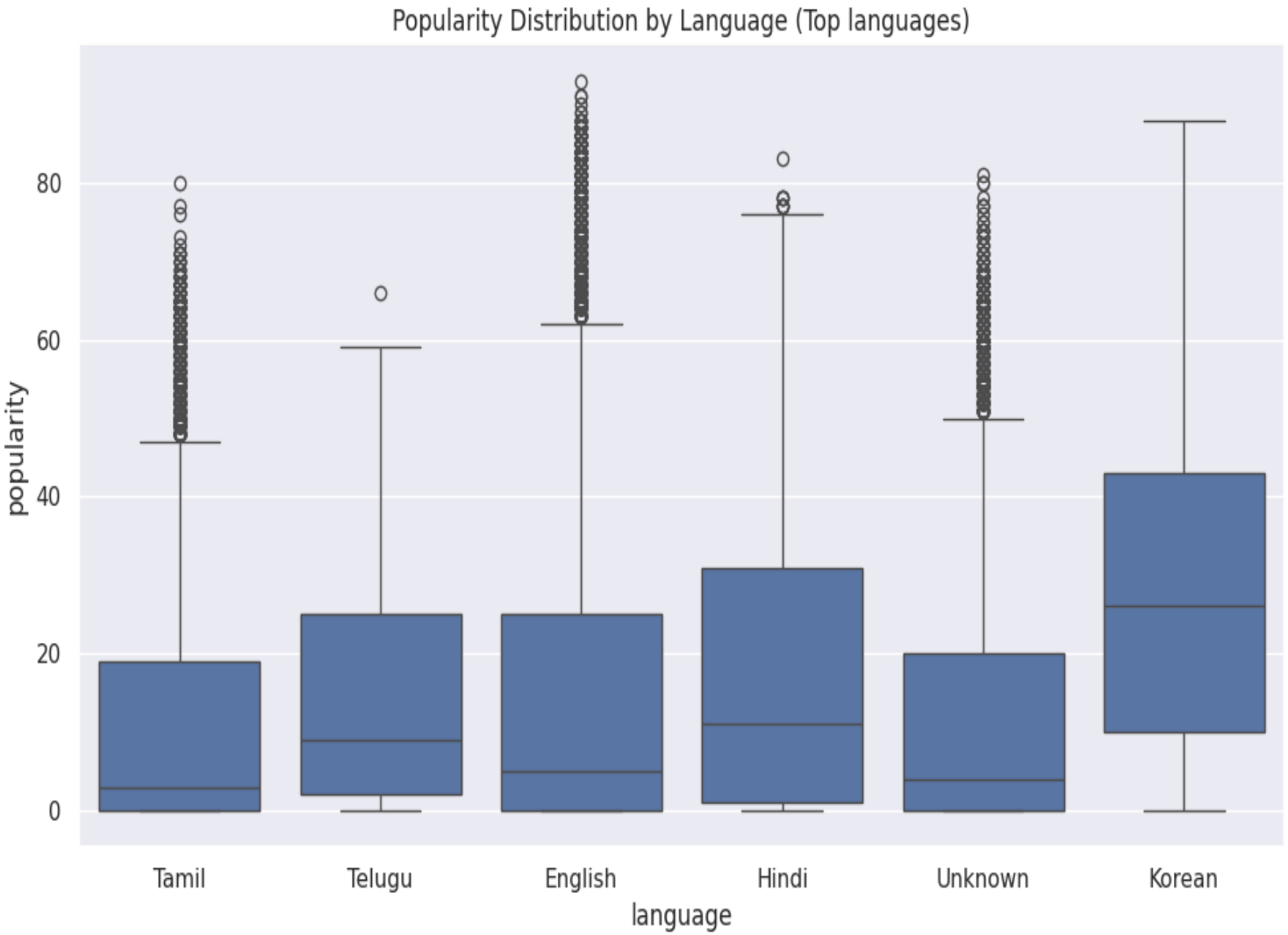


- **Top artists dominate**: A few artists consistently achieve higher average popularity across their tracks, highlighting their strong fanbase or mainstream appeal.

- **Track count filter**: Only artists with **5 or more tracks** are considered, ensuring the ranking reflects consistent performance rather than a single hit

- **Popularity gap**: There may be noticeable differences between the top-ranked artist and others, indicating that certain artists significantly outperform peers in audience reach.

LANGUAGE VS POPULARITY (BOXPLOT)

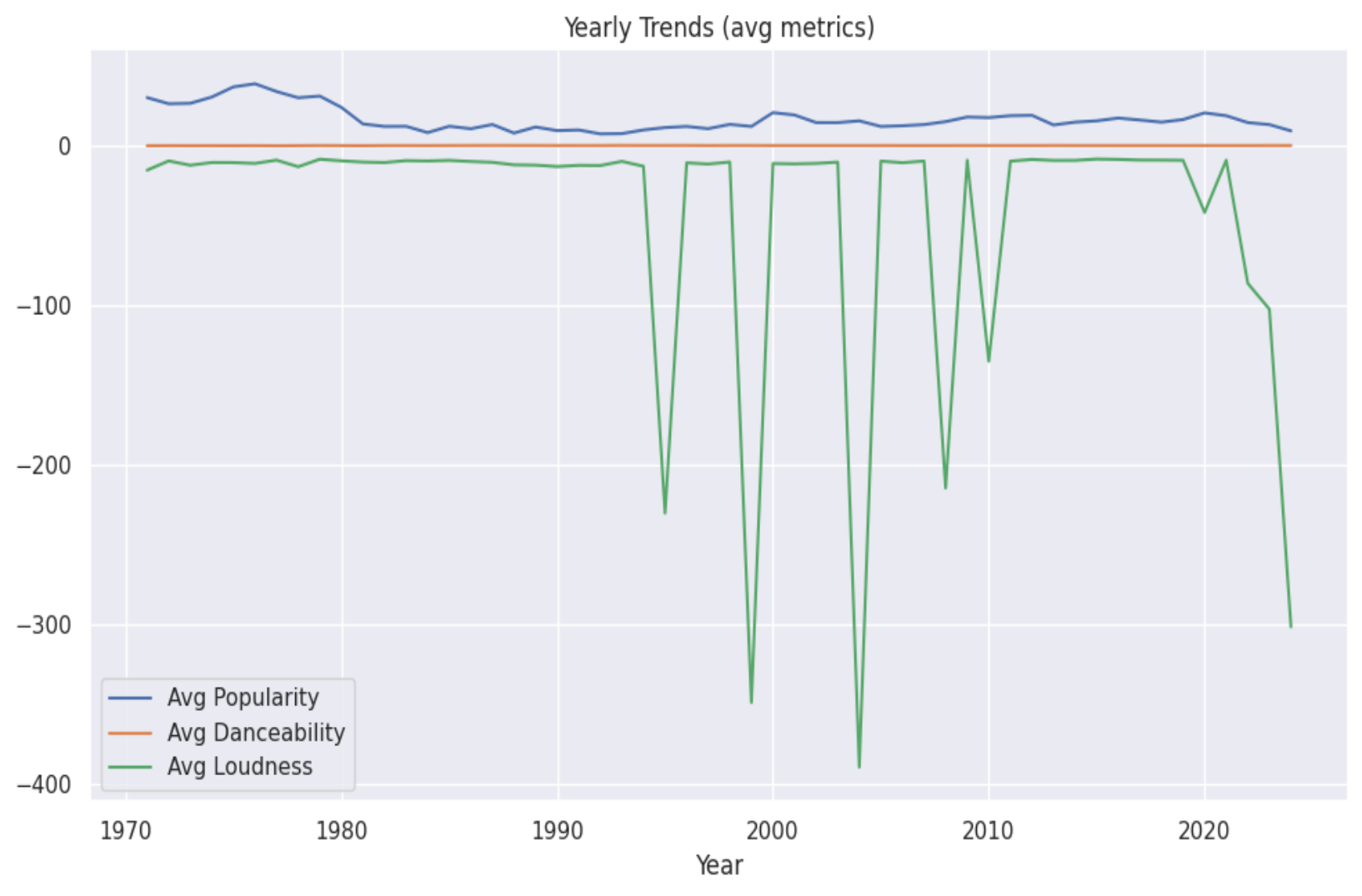


- **English dominance**: English tracks often show **higher median popularity** compared to other languages, reflecting their wider global audience.

- **Wide variability**: Popularity varies significantly within each language, indicating that hits can appear in any language but with different likelihoods.

- **Outliers**: Some songs achieve **exceptionally high popularity** regardless of language, highlighting that individual hits can transcend general trends.

YEARLY TRENDS

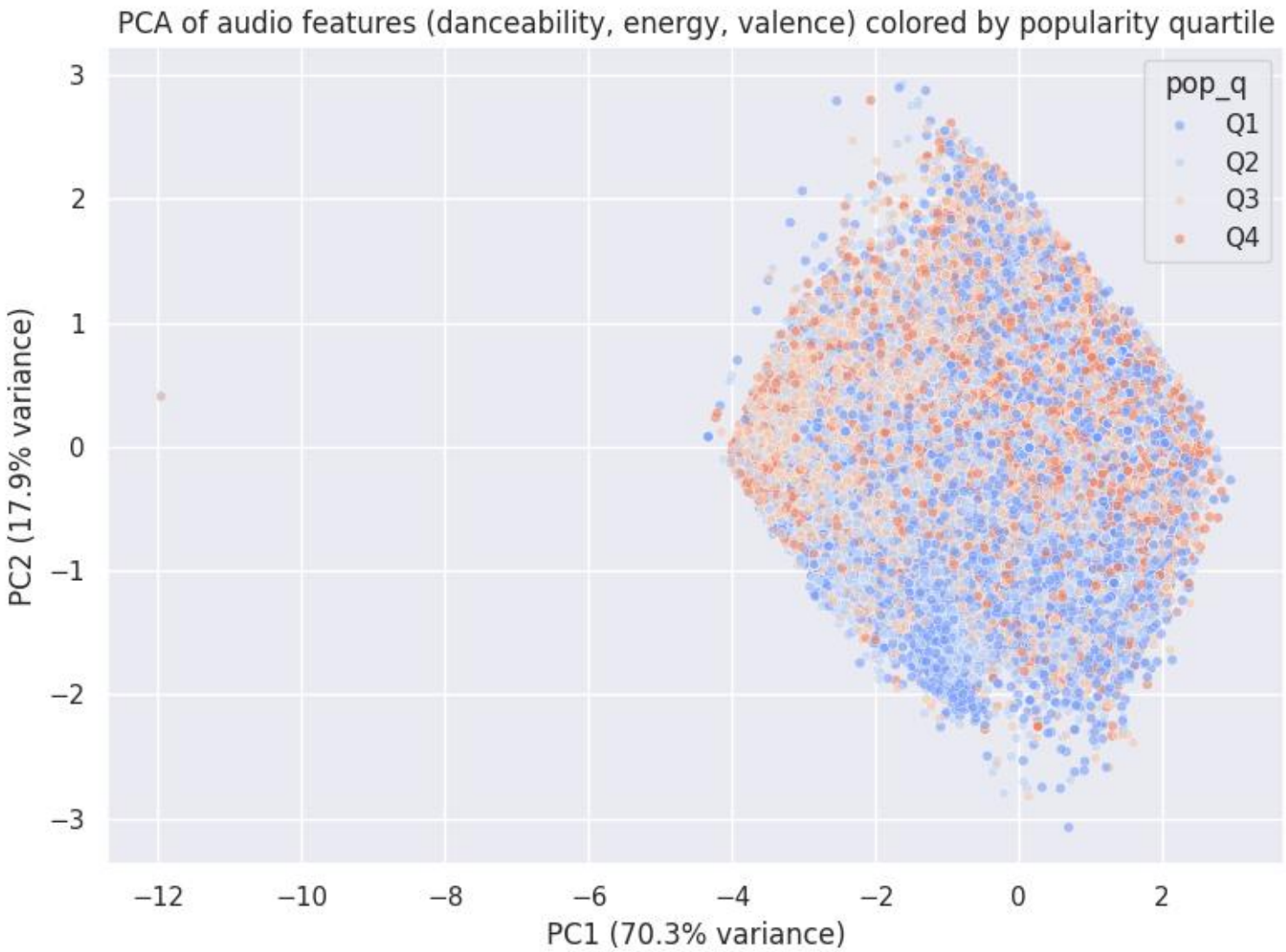


- **Popularity trends**: The average popularity may fluctuate over the years, reflecting changing listener preferences or trends in mainstream music.

- **Danceability evolution**: Average danceability can show gradual changes, indicating shifts in production style—songs may be getting slightly more or less danceable over time.

- **Loudness changes**: Average loudness may increase or decrease over the years, hinting at trends like the “loudness war” or evolving mastering practices in the music industry.

PCA SCATTER

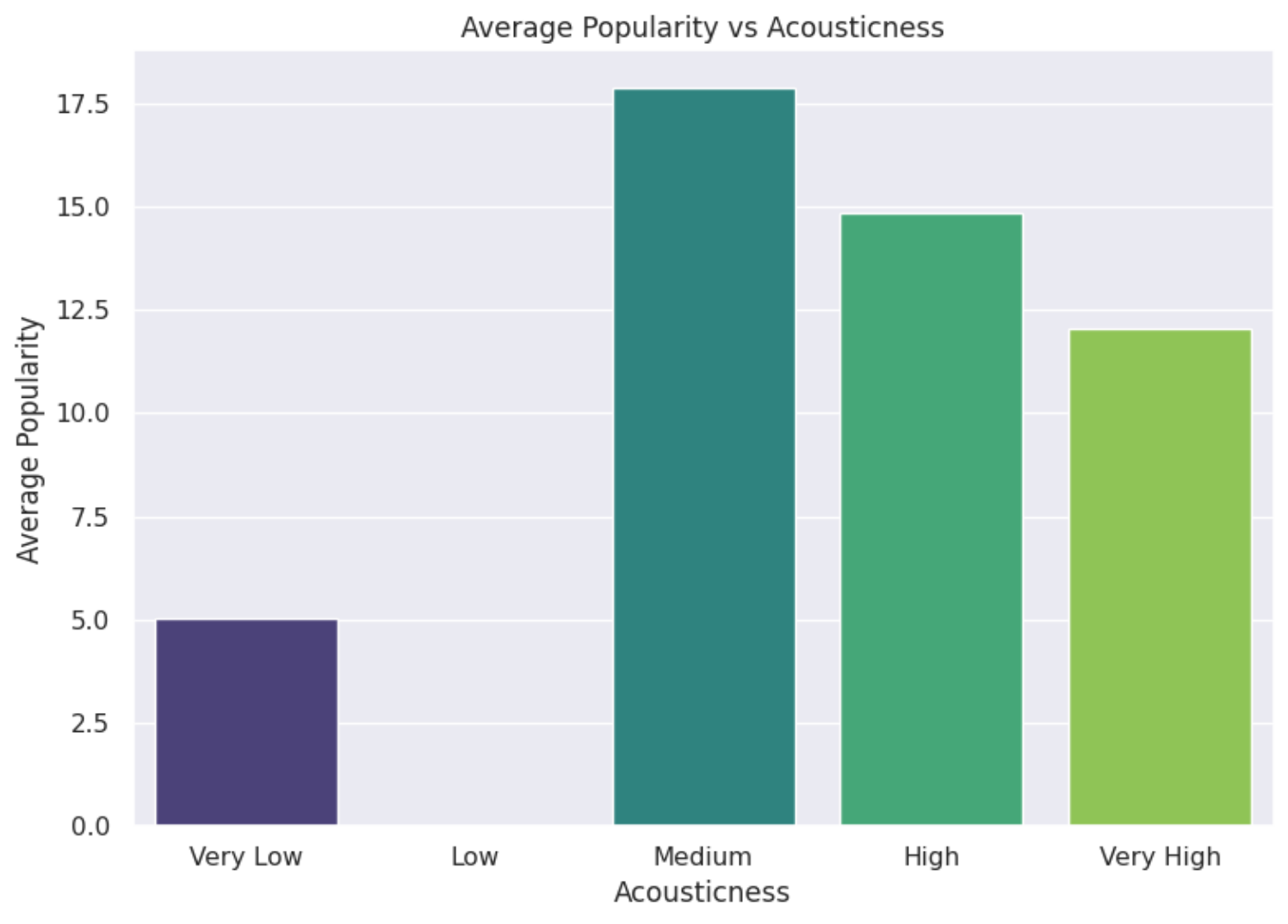


- **Feature clustering**: Songs cluster along PC1 and PC2 based on combinations of **danceability, energy, and valence**, showing natural groupings in audio characteristics.

- **Popularity spread**: Popularity quartiles (Q1–Q4) are somewhat mixed across the PCA space, suggesting that while audio features influence popularity, they are **not the sole determinant**.

- **Variance explanation**: PC1 explains the largest portion of variance, indicating that one combination of features (likely energy and danceability) captures most differences among songs, while PC2 adds nuance from valence or contrasting features.

AVERAGE POPULARITY VS ACOUSTICNESS



- **Low Acousticness → Moderate Popularity**
  Heavily produced or electronic tracks tend to achieve moderate popularity.

- **High Acousticness → Lower Popularity**
  Purely acoustic or unplugged songs generally attract fewer listeners.

- **Moderate Acousticness → Highest Popularity**
  Songs with a balanced mix of acoustic and produced elements tend to be the most popular.

CONCLUSION

1. Popularity Distribution: Most tracks have moderate popularity (scores between 20–50).
   Only a few songs reach global hit status (80+).

2. Danceability: Majority of songs are moderately to highly danceable (0.5–0.8).
   Indicates Spotify’s catalog leans toward upbeat and rhythm-driven music.

3. Energy Levels: High-energy songs dominate the dataset.
   Very few calm or low-energy tracks exist, showing preference for lively music.

4. Feature Correlations: Energy, loudness, and danceability are positively correlated.
   Acousticness and instrumentalness are negatively correlated with these.
   Popularity has only moderate correlation with audio features — multiple factors influence a hit.
5. Language Impact: English tracks generally have higher popularity.
   However, exceptional hits exist in all languages.

CONTD...

6. Artist Performance: Top artists consistently maintain higher average popularity.
   Sustained success depends on consistent audience engagement, not one-time hits.

7. Yearly Trends: Popularity, danceability, and loudness vary over the years.
   Reflects evolving listener preferences and production styles.

8. Acousticness Insight: Moderate acousticness (a mix of acoustic and electronic) yields the highest popularity.
   Purely acoustic or overly electronic songs are less favored.

9. Overall Observation: Popularity depends on a balance of musical features rather than extremes.
   Spotify listeners generally prefer energetic, danceable, and emotionally positive tracks.