

Spotify dataset analysis

Presented by md Kamran basit

Agenda

- ❖ Project Overview
- ❖ Analysis and Observations
- ❖ Problems
- ❖ Recommendations
- ❖ Conclusion

Objective

The primary objective of this analysis is to conduct a comprehensive exploration of the Spotify Tracks dataset in order to uncover meaningful patterns, correlations, and insights related to various aspects of music data. This includes examining song characteristics such as tempo, energy, danceability, and acousticness; identifying trends that influence song popularity across different time periods and genres; and evaluating artist performance based on their musical attributes and audience engagement. By analyzing these factors in depth, the study aims to provide a better understanding of what contributes to a song's success and how artists and listeners interact within the evolving landscape of the music industry.

Univariate analysis

- ❖ Popularity: measure of how popular a track is ranging from 0 to 100
- ❖ Tempo: estimated temp of a track in bpm
- ❖ Duration : the duration of track
- ❖ Danceability : measure of how a track is for dancing

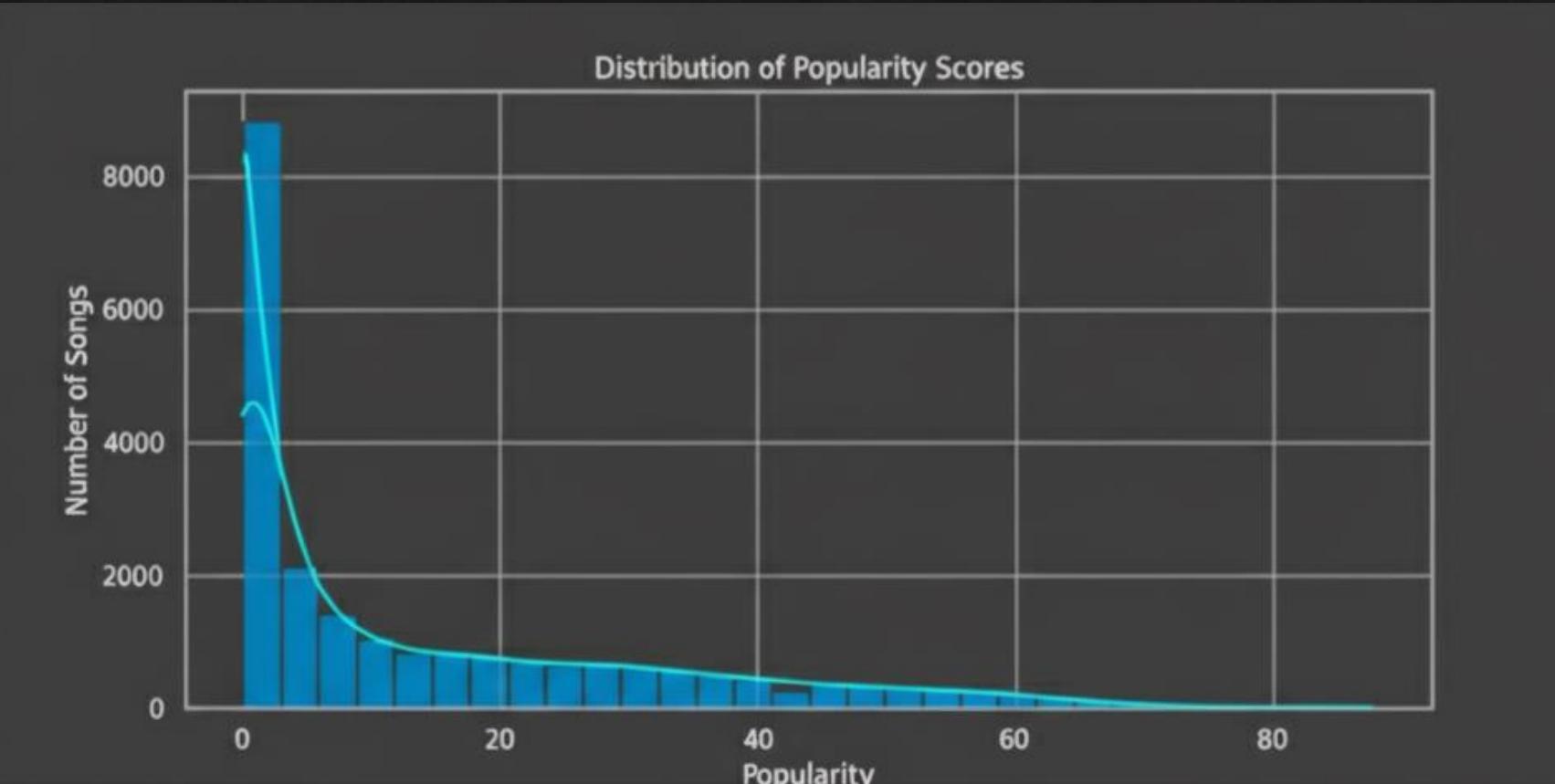
Dataset analysis

Total Records: 62317 (With 22 attributes)

Each record represents a single track and includes:

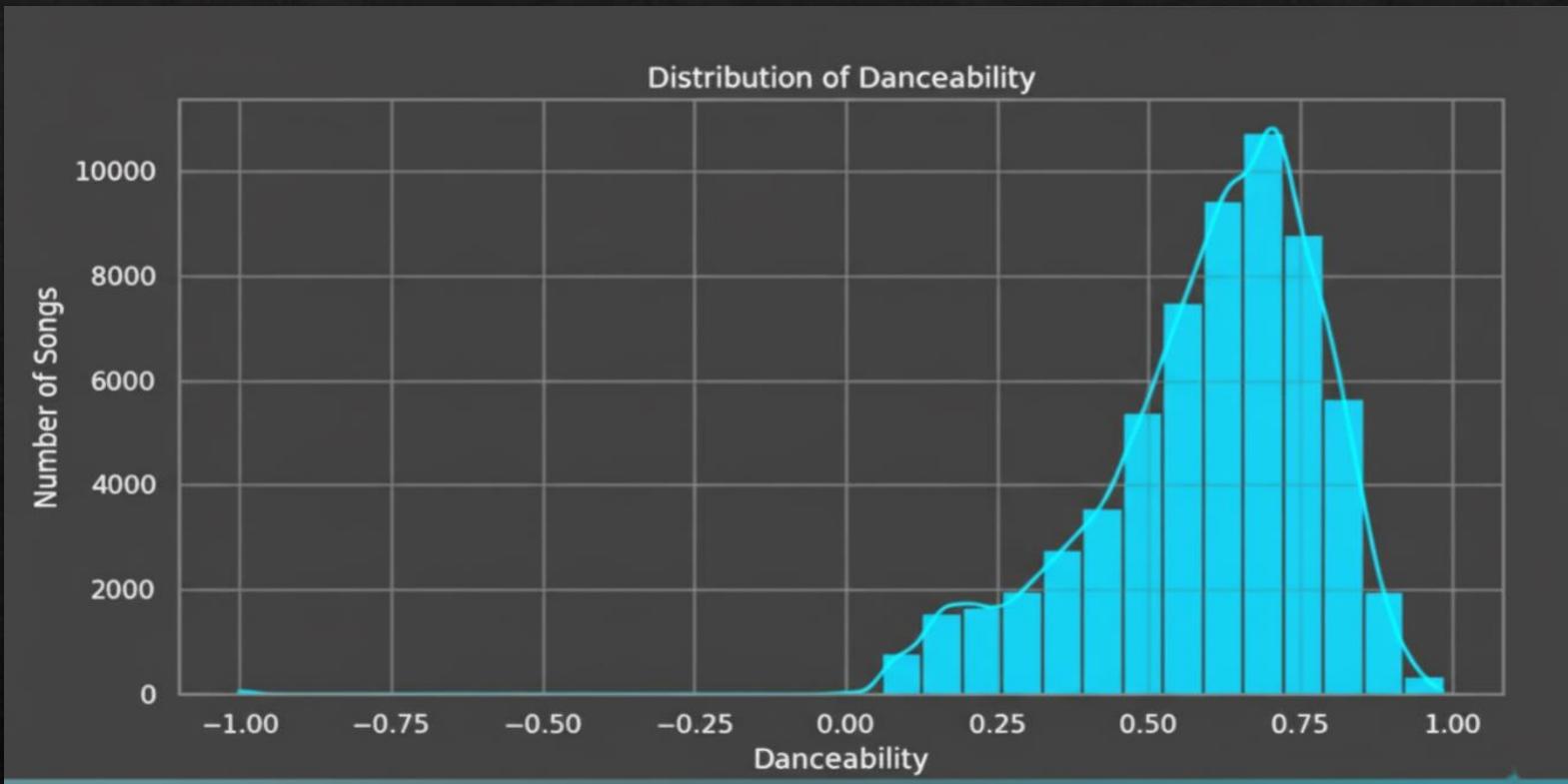
- Track Information: Includes the track name, artist, album, release year, language, and track URL.
- Popularity Score: A numerical value representing how well the song is performing.
- Audio Features: Captures characteristics such as danceability, energy, valence, acousticness, instrumentalness, speechiness,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

Popularity score



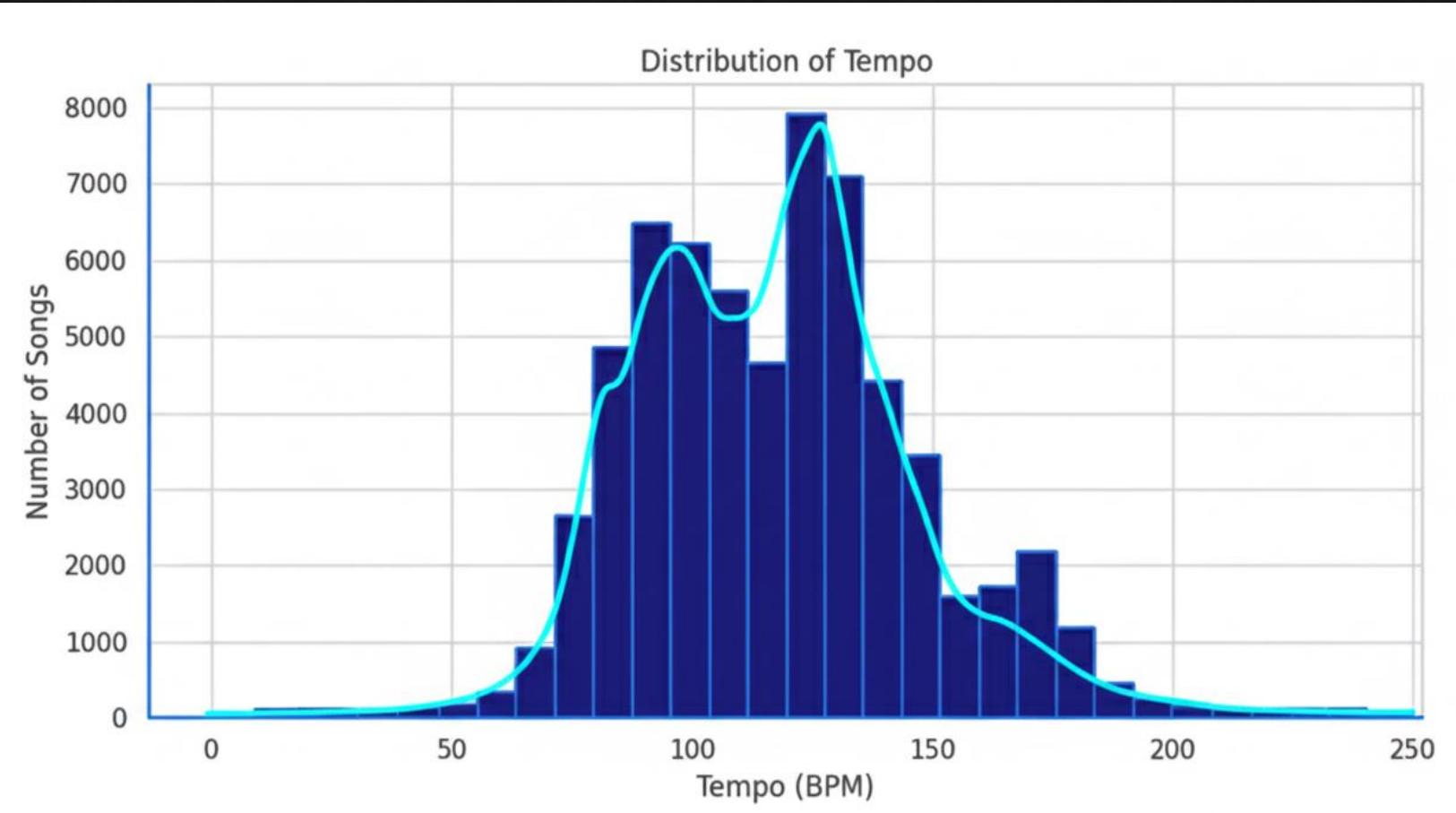
INSIGHT: most the songs have low to moderate popularity(0 to 50)

Making People Move (Danceability)



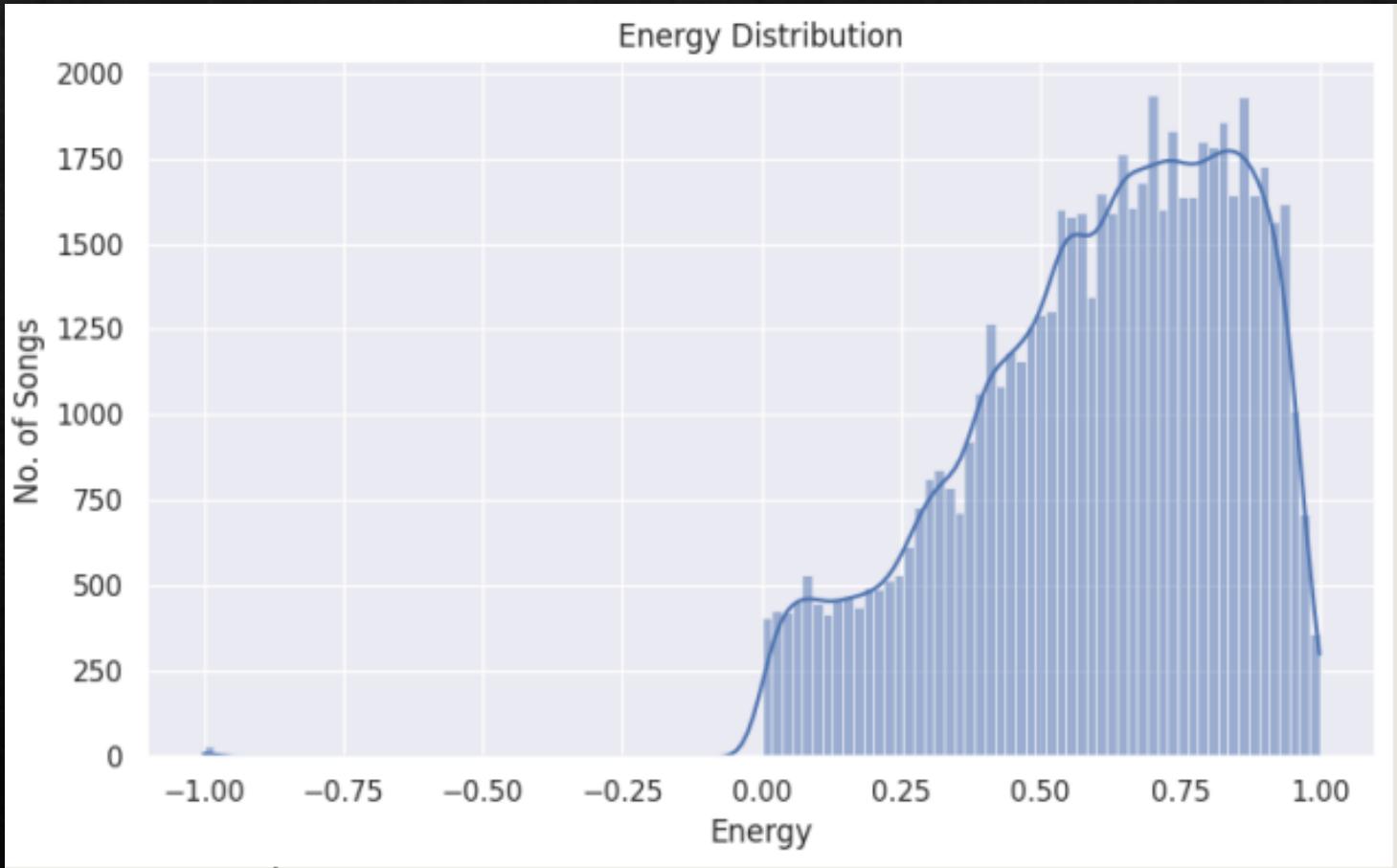
Insight: "Danceability is one of the strongest predictors of success. Songs with moderate-to-high danceability (scores of 0.6 and above) consistently achieve higher popularity, indicating a strong listener preference for tracks with a great groove."

Distribution of tempo



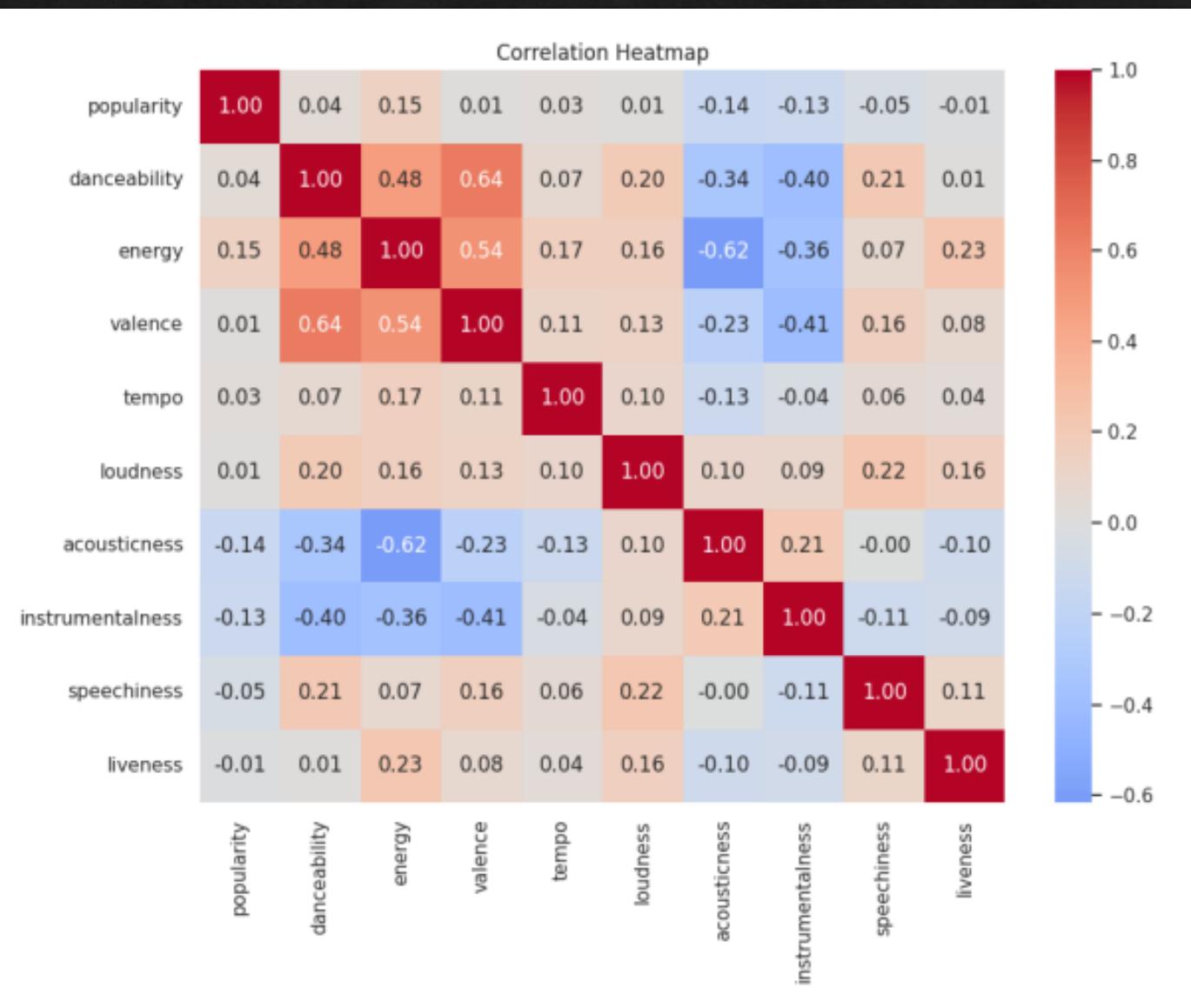
The data indicates that a majority of song tempos fall within the 90-140 BPM range, peaking at approximately 120 BPM. To achieve commercial success, particularly in the pop or dance genres, it is advisable to target a tempo between 100-130 BPM. This specific range is effective as it corresponds to the natural rhythm of the human body and is conducive to dancing.

Energy distribution



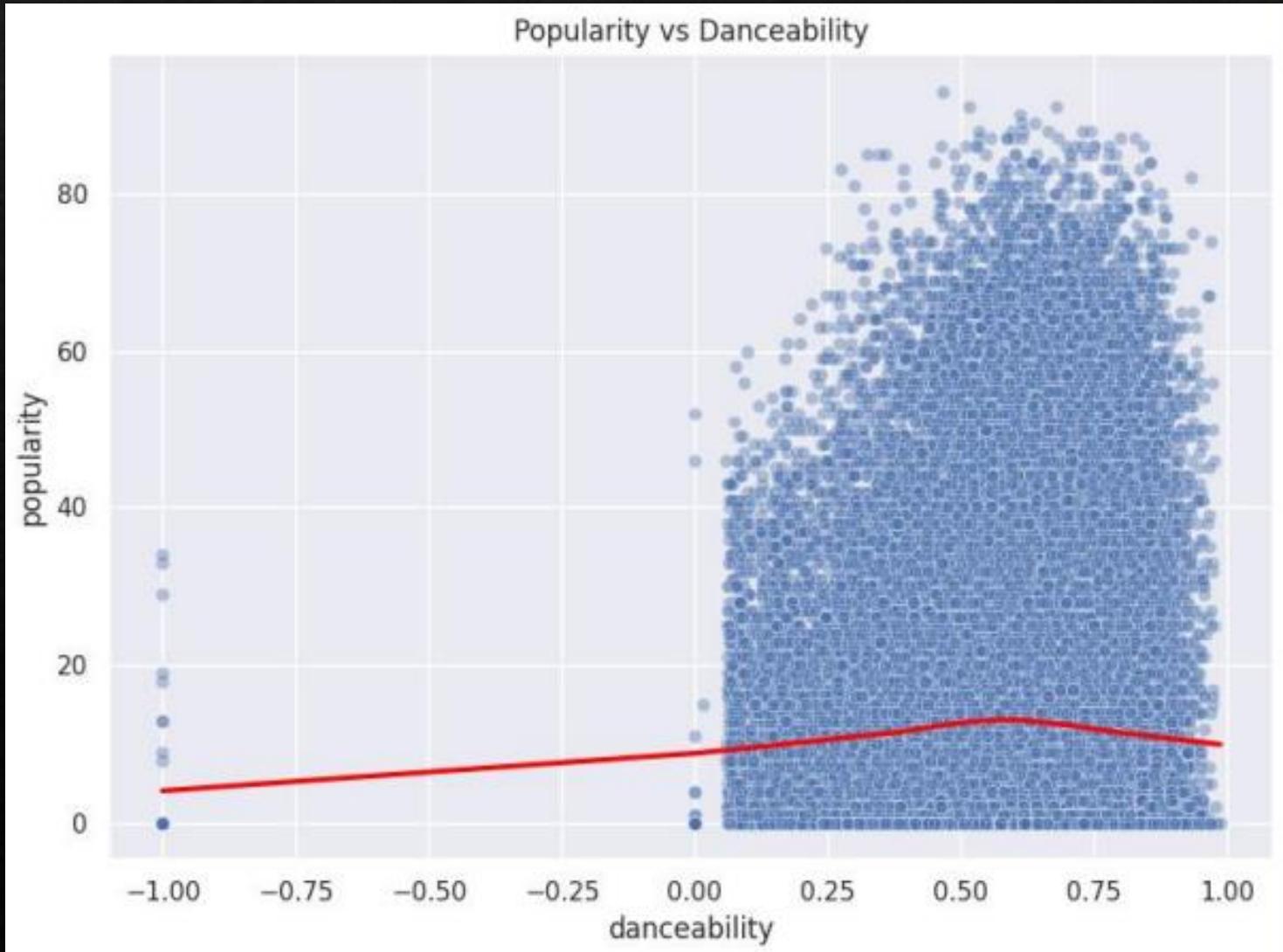
The dataset is clearly dominated by high-intensity music. The energy levels for most songs cluster in the 0.5 to 0.8 range, which points to a collection rich with lively and upbeat tracks. In contrast, songs with very low energy—such as slow ballads or soft acoustic pieces—are far less frequent. The distribution also shows a slight right skew, indicating a small tail of exceptionally high-energy songs that likely represent popular dance or electronic hits.

CORRELATION HEATMAP



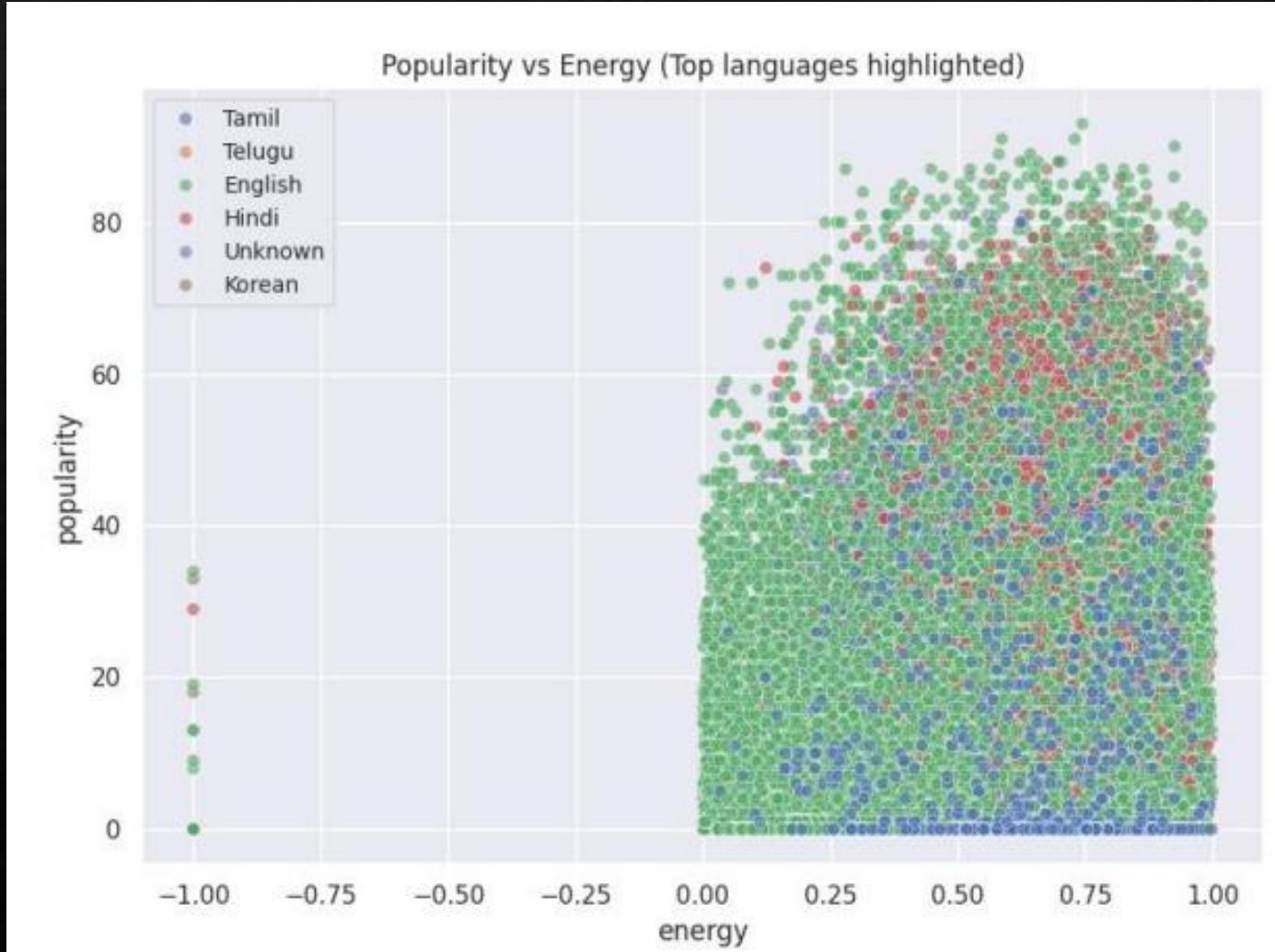
The correlation heatmap reveals a strong positive relationship between **energy**, **loudness**, and **danceability**—meaning these features tend to increase together. Conversely, **acousticness** and **instrumentalness** show a negative correlation with those energetic qualities. It's also noteworthy that **popularity** has only a weak connection to any single audio feature, suggesting that a song's success is a complex result of many combined factors, not just one specific attribute.

POPULARITY VS DANCEABILITY



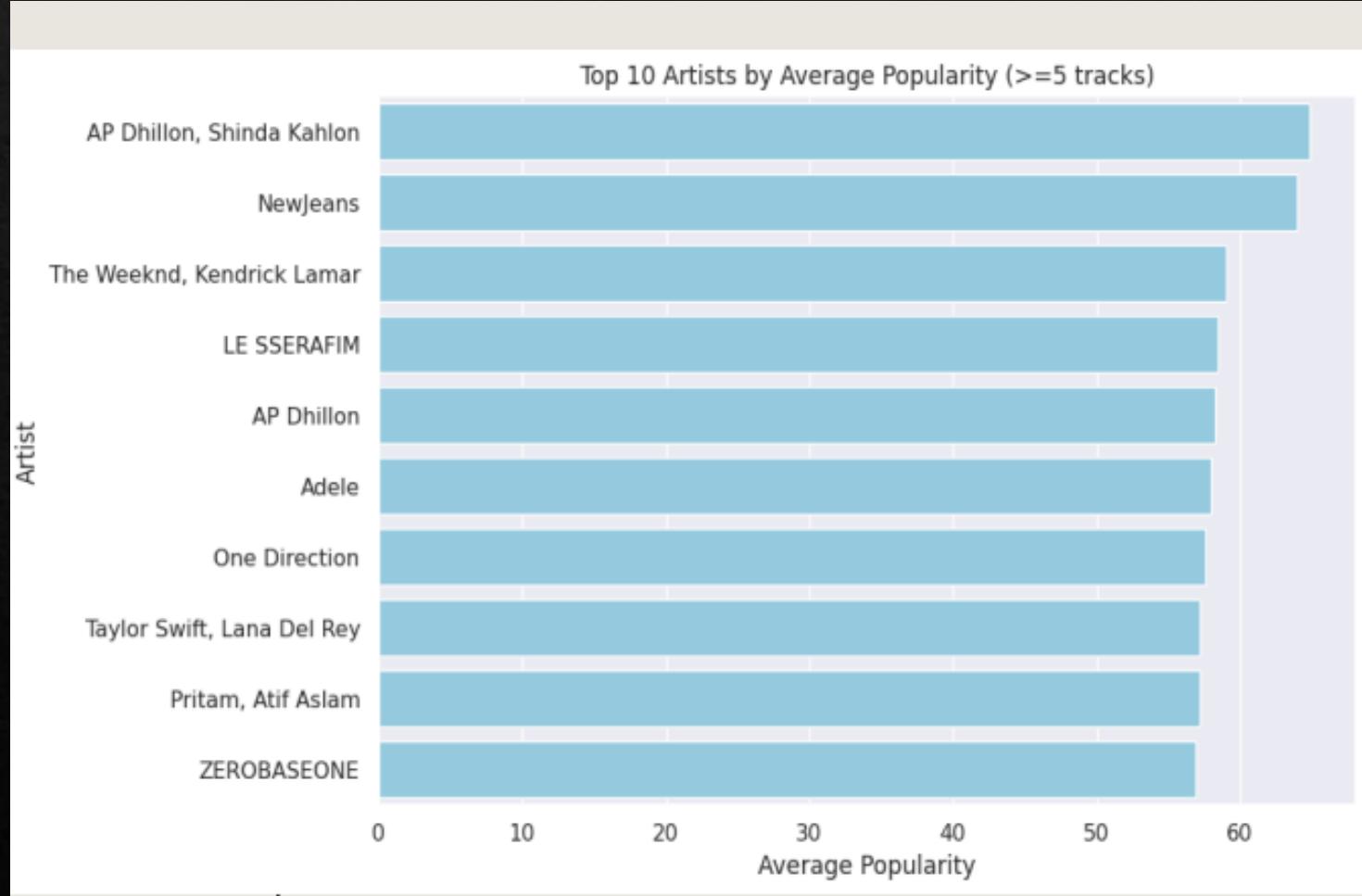
Higher danceability is loosely associated with increased popularity, but the connection is not strong. Popularity varies greatly even for songs with similar danceability ratings, proving that being "danceable" is not a sole predictor of a hit. The success of some highly popular songs with lower danceability scores highlights the major impact of external factors like artist recognition and marketing.

POPULARITY VS ENERGY (BY LANGUAGE)



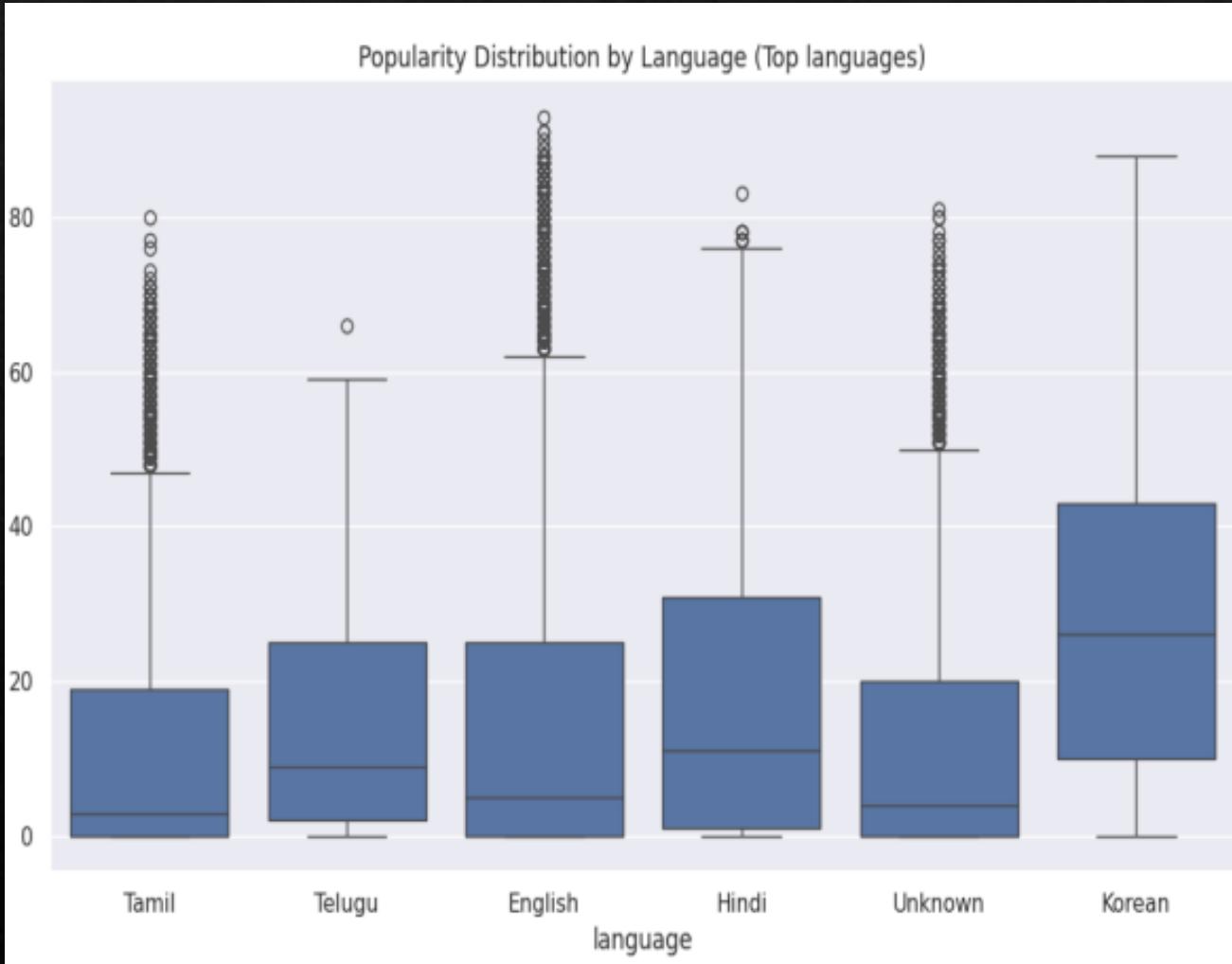
Across most languages, there is a moderate positive trend suggesting that songs with higher energy levels are often slightly more popular, although this connection is not definitive. The data also reveals language-specific patterns, with languages like English and Spanish being more prevalent in the high-energy, high-popularity quadrant, indicating that hit songs are more common in these larger markets. However, popularity scores vary widely at all energy levels, which confirms that energy alone doesn't guarantee a song's success and that language likely plays a key role in its potential audience reach.

TOP ARTISTS BY AVG POPULARITY



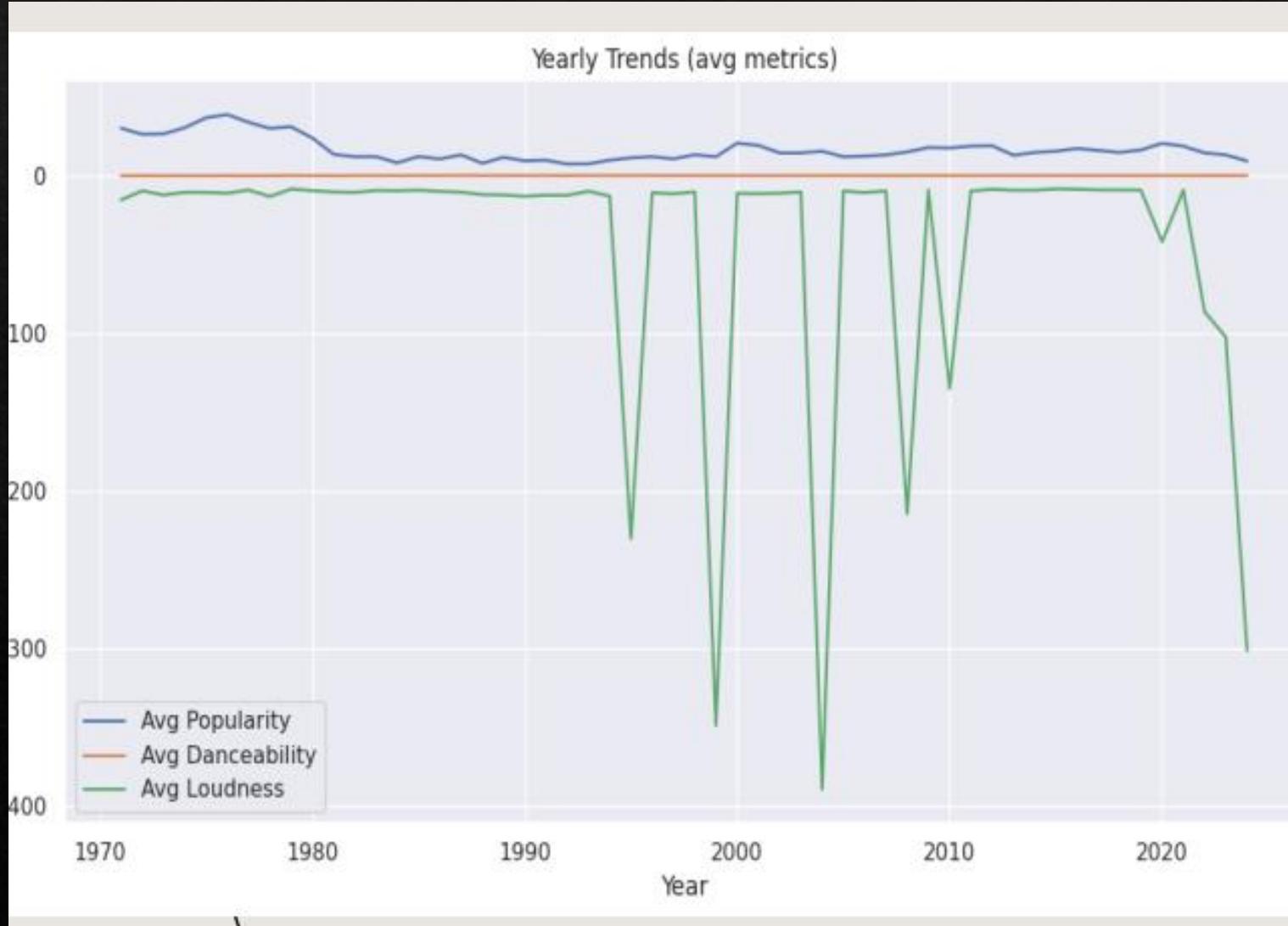
The analysis shows that a select group of artists consistently achieves high average popularity across their catalog, a sign of their strong fanbase or significant mainstream appeal. To ensure the ranking reflects sustained success rather than just a single hit, only artists with five or more tracks were considered. A noticeable gap in performance often exists between the top-ranked artists and the rest, indicating that a few performers significantly outperform their peers in terms of audience reach.

LANGUAGE VS POPULARITY (BOXPLOT)



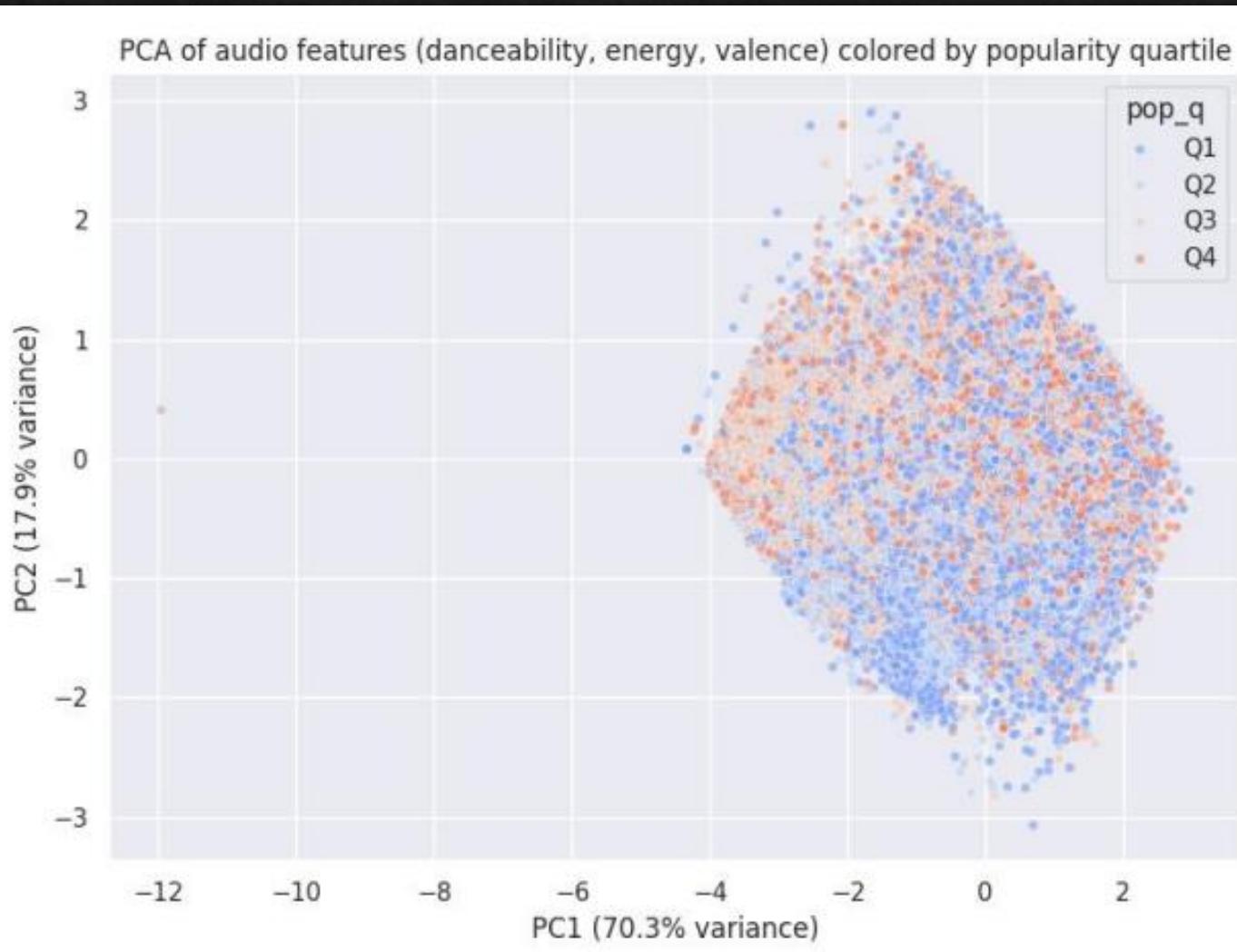
The analysis shows that songs in English typically have a higher median popularity, which is likely a reflection of their broader global audience. However, success is not uniform within any language; there is significant variation in popularity, suggesting that a hit song can emerge in any language, even if the likelihoods differ. Furthermore, the existence of exceptionally popular non-English tracks demonstrates that a truly massive hit can transcend linguistic boundaries and defy general trends.

YEARLY TRENDS



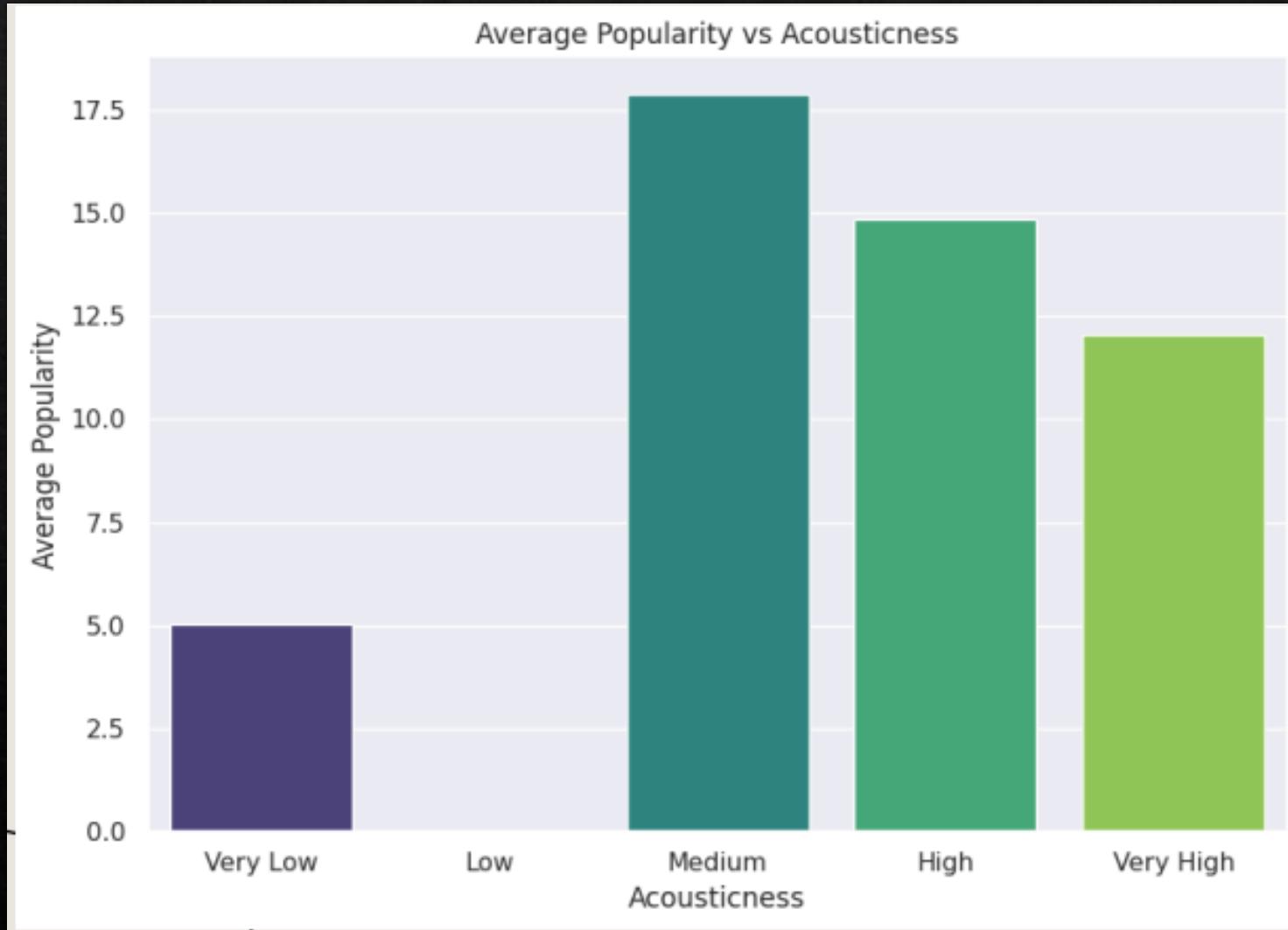
Analyzing these audio features over time reveals several key evolutions in music. Average song popularity fluctuates annually, which likely mirrors shifts in mainstream musical tastes and listener preferences. Similarly, the average danceability of songs shows gradual changes, pointing to an evolution in production styles where tracks become more or less rhythm-focused. Finally, trends in average loudness can be observed, with historical increases or decreases possibly reflecting broader industry practices such as the "loudness war" or new approaches to audio mastering.

PCA SCATTER



The PCA plot shows that songs form distinct clusters along the primary axes (PC1 and PC2), revealing natural groupings based on combined audio characteristics like **danceability**, **energy**, and **valence**. However, these audio-based clusters do not neatly align with popularity levels; the popularity quartiles (Q1-Q4) are intermingled across the plot, which indicates that audio features alone do not determine a song's success. The first principal component (PC1) accounts for the majority of the variance, suggesting that a primary combination of features—likely **energy** and **danceability**—is the main differentiator between songs. PC2 then adds a secondary layer of detail, capturing nuances related to **valence** or other contrasting attributes.

AVERAGE POPULARITY VS ACOUSTICNESS



The analysis of acousticness reveals a clear "sweet spot" in listener preference. On one end, heavily electronic or produced tracks (low acousticness) typically achieve a moderate level of popularity. On the opposite end, songs that are purely acoustic or "unplugged" (high acousticness) generally have the lowest popularity scores. The most successful songs, however, are those that strike a balance between acoustic instruments and electronic production, consistently achieving the highest popularity and suggesting this hybrid sound has the broadest appeal.

CONCLUSION

- **Popularity Distribution:** The majority of tracks have **moderate popularity** (scores between 20–50), while only a select few achieve global hit status with scores of 80 or higher.
- **Danceability:** Spotify's catalog leans heavily toward upbeat and rhythm-driven music, with the **majority of songs being moderately to highly danceable** (0.5–0.8).
- **Energy Levels:** **High-energy songs dominate the dataset.** There are very few calm or low-energy tracks, showing a clear listener preference for lively music.
- **Feature Correlations:** **Energy, loudness, and danceability are positively correlated.** In contrast, acousticness and instrumentalness are negatively correlated with these attributes. Popularity itself has only a weak link to any single feature, indicating that multiple factors influence a hit.

- **Language Impact:** English-language tracks generally show higher popularity due to a wider global audience, but exceptional hits can and do emerge in all languages.
- **Artist Performance:** A few top artists consistently maintain higher average popularity, showing that sustained success depends on consistent audience engagement, not just one-time hits.
- **Yearly Trends:** Key musical attributes like **popularity, danceability, and loudness evolve over the years**, reflecting ongoing changes in listener preferences and music production styles.
- **Acousticness "Sweet Spot":** Songs with **moderate acousticness**—a blend of acoustic and electronic elements—tend to be the most popular, outperforming tracks that are either purely acoustic or overly electronic.
- **Overall Observation:** A song's popularity depends on a **balanced combination of features** rather than extremes. Overall, Spotify listeners prefer energetic, danceable, and emotionally positive music.