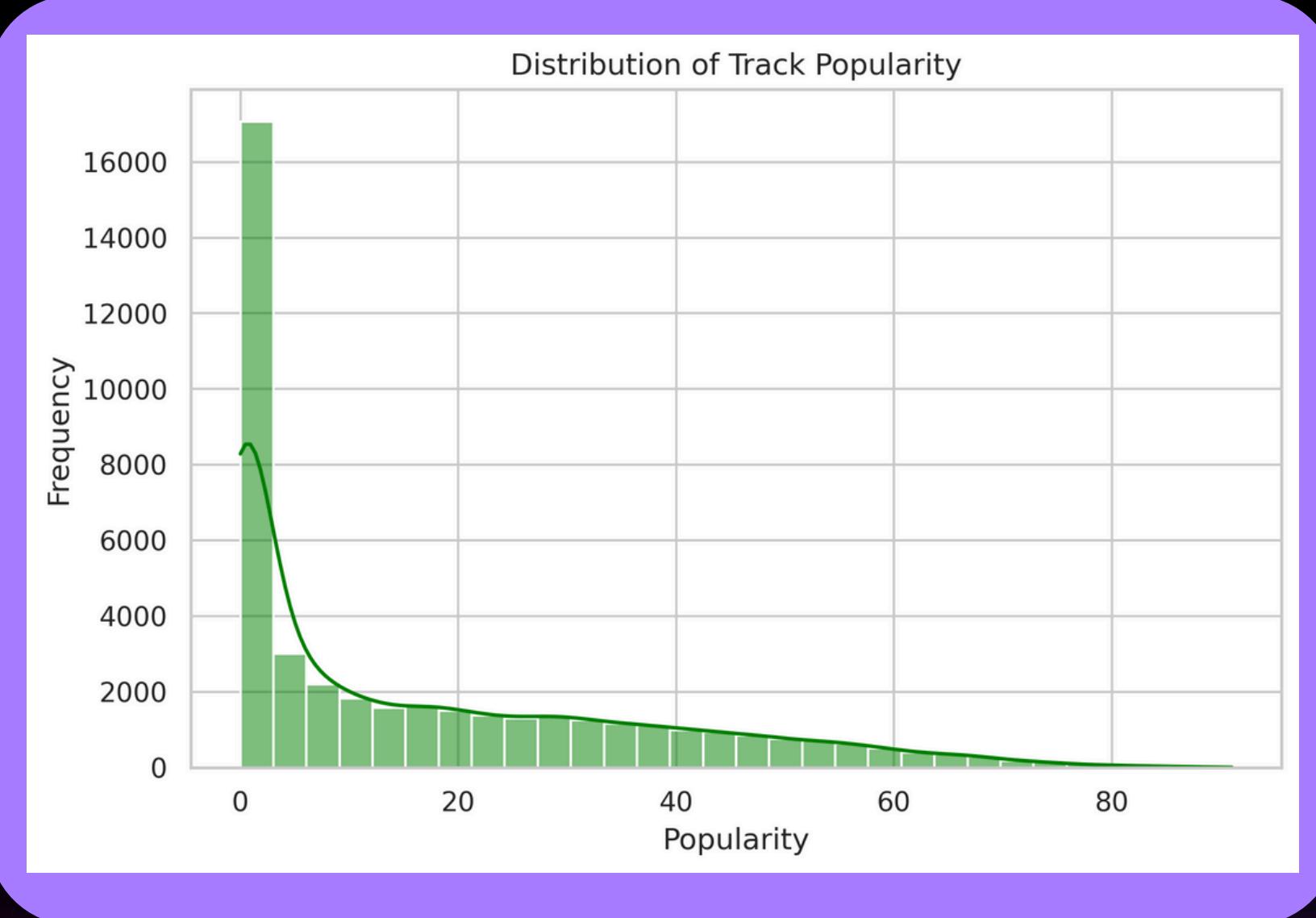
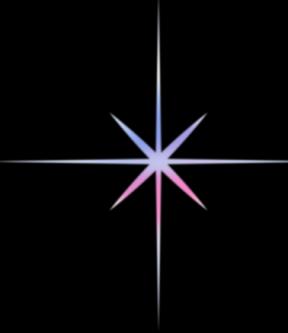


EXPLORATORY DATA ANALYSIS OF SPORTIFY DATA TRACKS

Prepared By - **DEBMALYA DAS (CSE)**
GCECTB - R24 - 3010

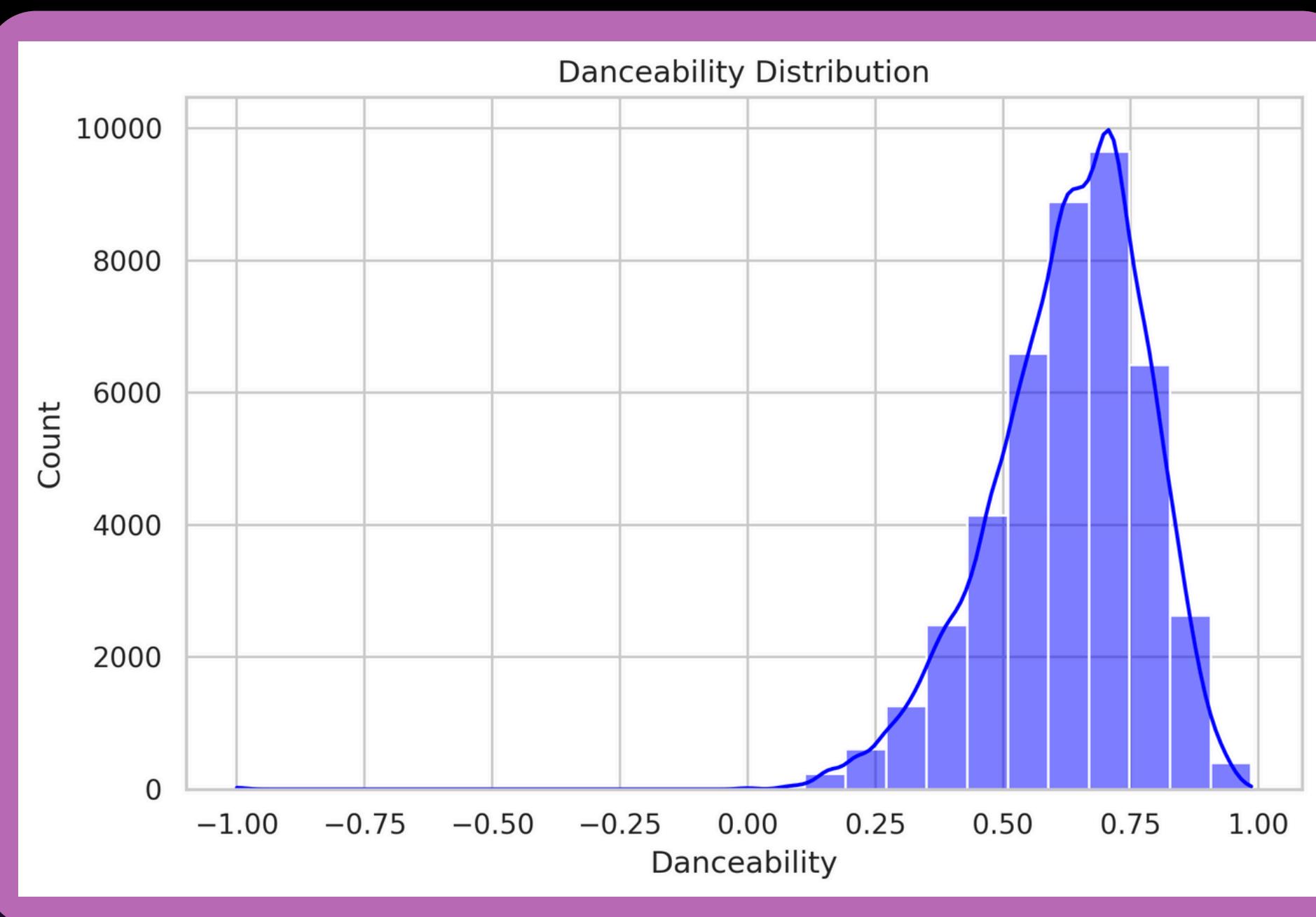
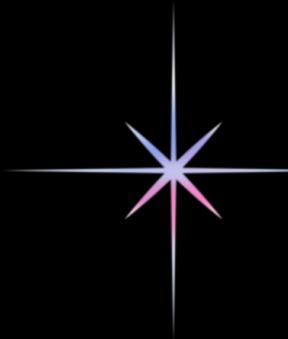


Popularity Distribution



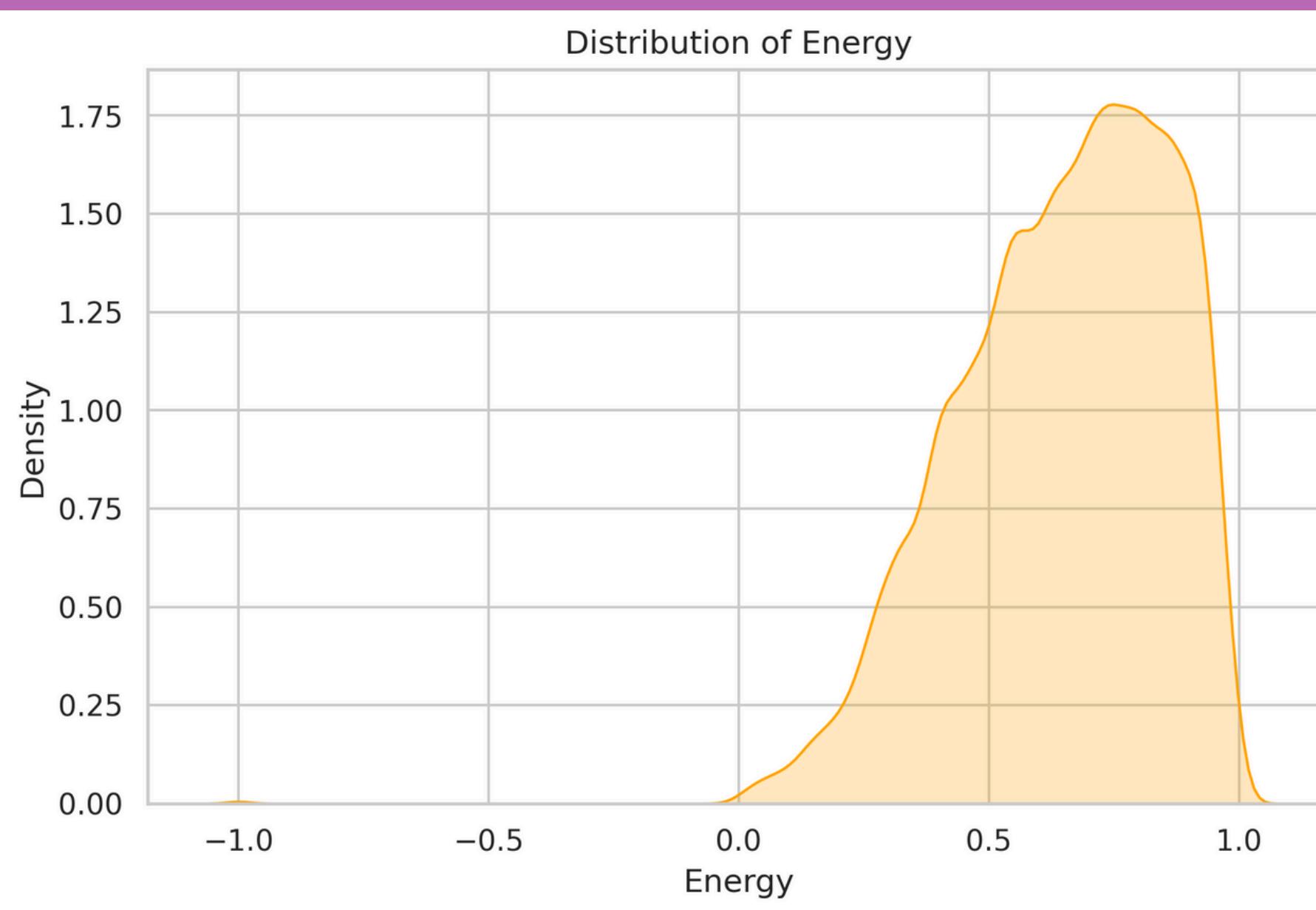
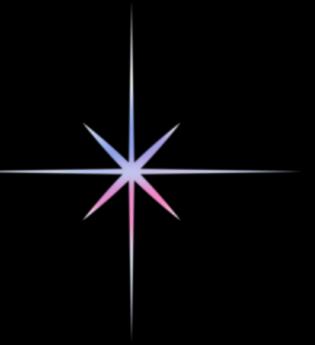
- **Highly Skewed Distribution:**
 - The distribution is right-skewed (long tail towards the right), meaning most tracks have very low popularity scores, while only a small number achieve high popularity.
- **Concentration of Low-Popularity Tracks:**
 - A large spike near the 0–5 popularity range indicates that the majority of songs are relatively unknown or rarely listened to.
- **Few Highly Popular Tracks Dominate:**
 - As popularity increases, the frequency drops sharply – showing that a small fraction of songs capture the majority of listener attention, a pattern typical in music streaming data (the “hit song” phenomenon).

Danceability Distribution



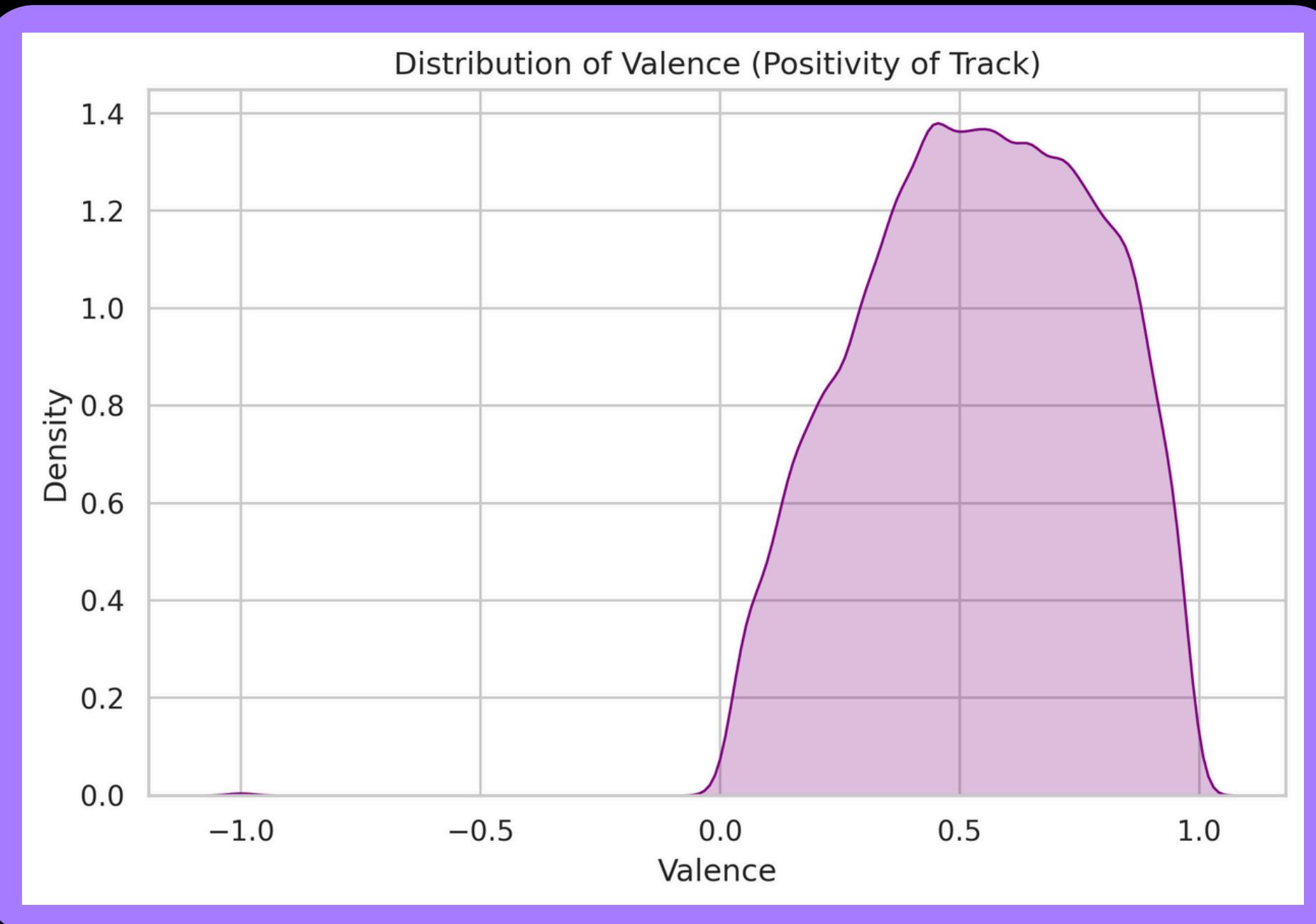
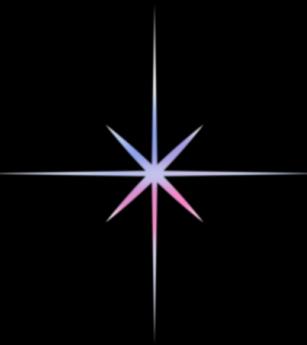
- **Right-Skewed Towards Higher Danceability:**
 - The majority of songs have danceability scores between 0.5 and 0.8, indicating that most tracks in the dataset are relatively easy to dance to.
- **Peak Around 0.7:**
 - The distribution peaks near 0.7, implying that a large number of tracks are moderately to highly danceable – likely suitable for general listening and party playlists.

Energy Distribution



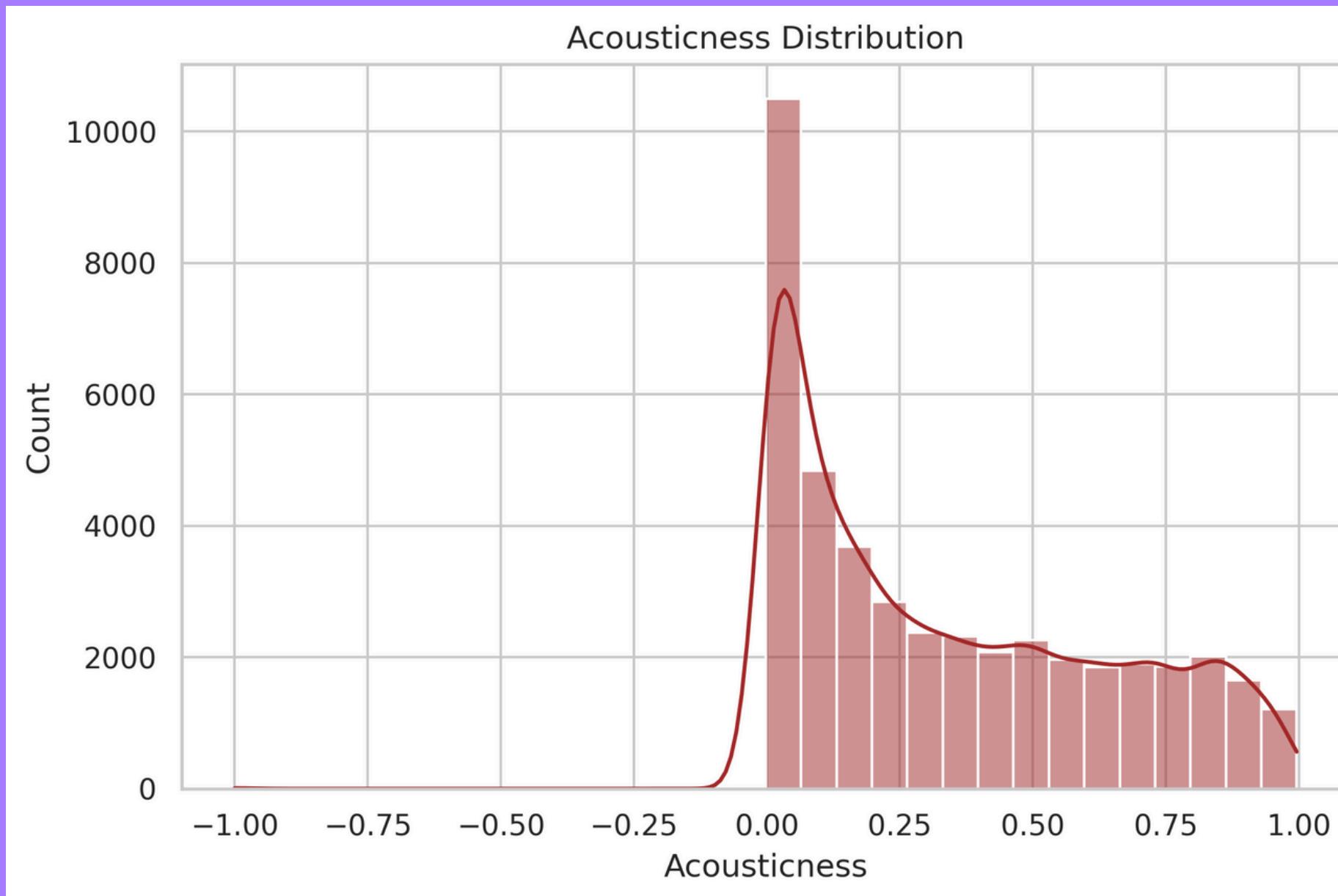
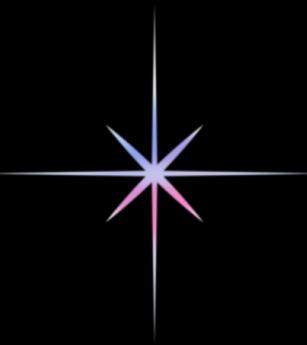
- **Moderately High Energy Dominance:**
 - Most tracks have energy values between 0.5 and 0.9, indicating that the dataset primarily contains energetic and lively songs rather than calm or mellow ones.
- **Left-Skewed Distribution:**
 - The distribution is skewed towards higher energy values, meaning low-energy tracks are relatively rare, suggesting that energetic music tends to be more common or popular in this dataset.

Valence Distribution



- **Moderate to High Valence Dominance:**
 - Most tracks have valence values between 0.4 and 0.8, indicating that the majority of songs in the dataset tend to have positive, happy, or cheerful tones.
- **Fewer Low-Valence Tracks:**
 - The distribution is skewed towards higher valence values, suggesting that sad or melancholic tracks are relatively less common compared to more upbeat ones.

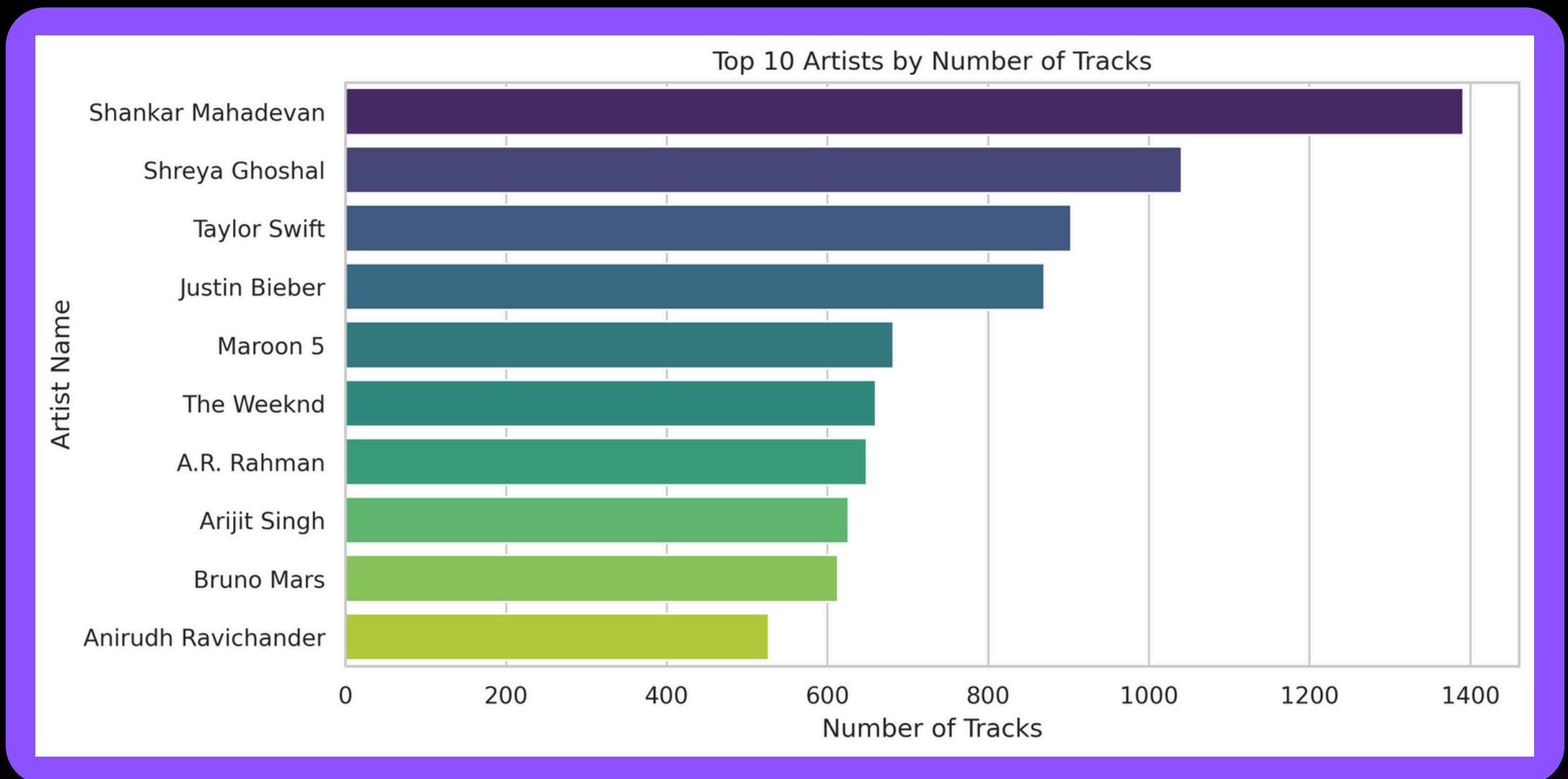
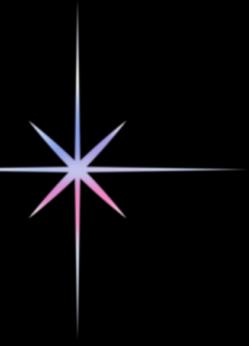
Acousticness Distribution



- **Dominance of Low Acousticness Tracks:**
- Most tracks have acousticness values close to 0, indicating that the majority of songs in the dataset are heavily produced or electronic rather than purely acoustic.

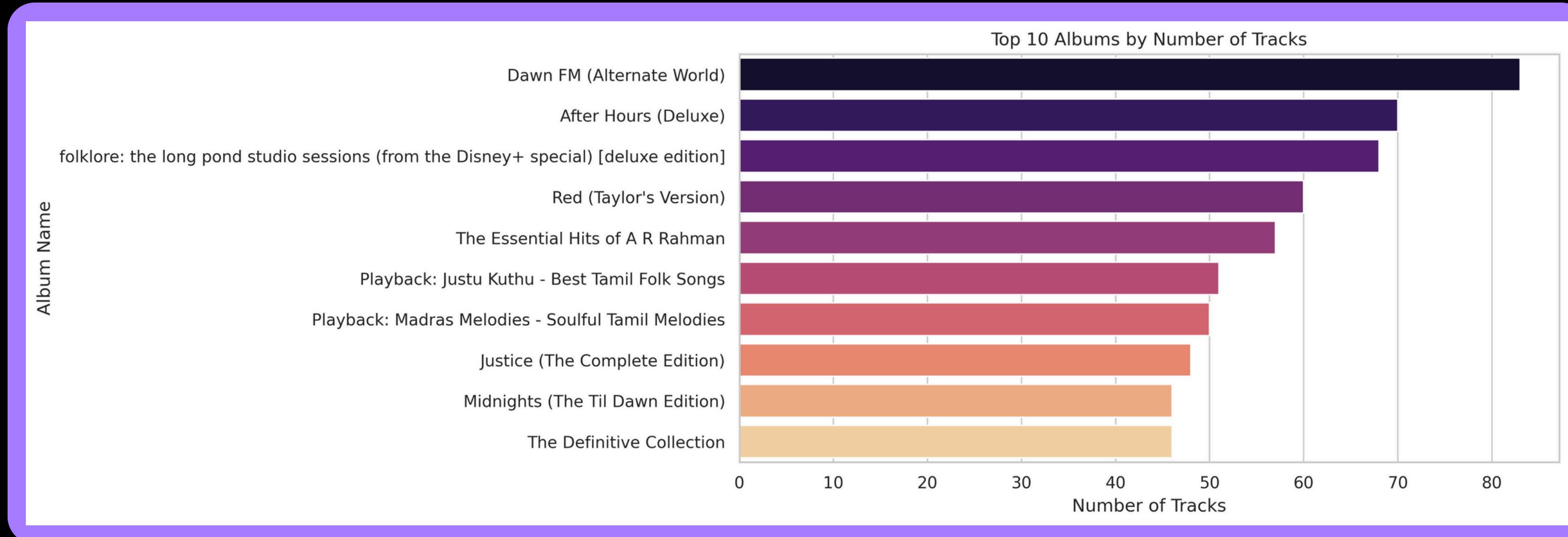
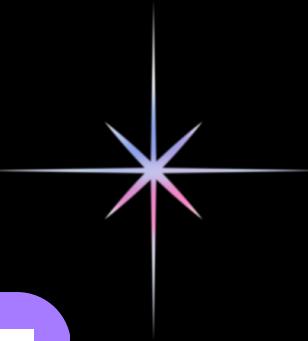
- **Gradual Decline Toward High Acousticness:**
- As acousticness increases, the number of tracks gradually decreases, suggesting that fully acoustic or unplugged songs are relatively less common in the dataset.

Top 10 Artists by Number of Tracks



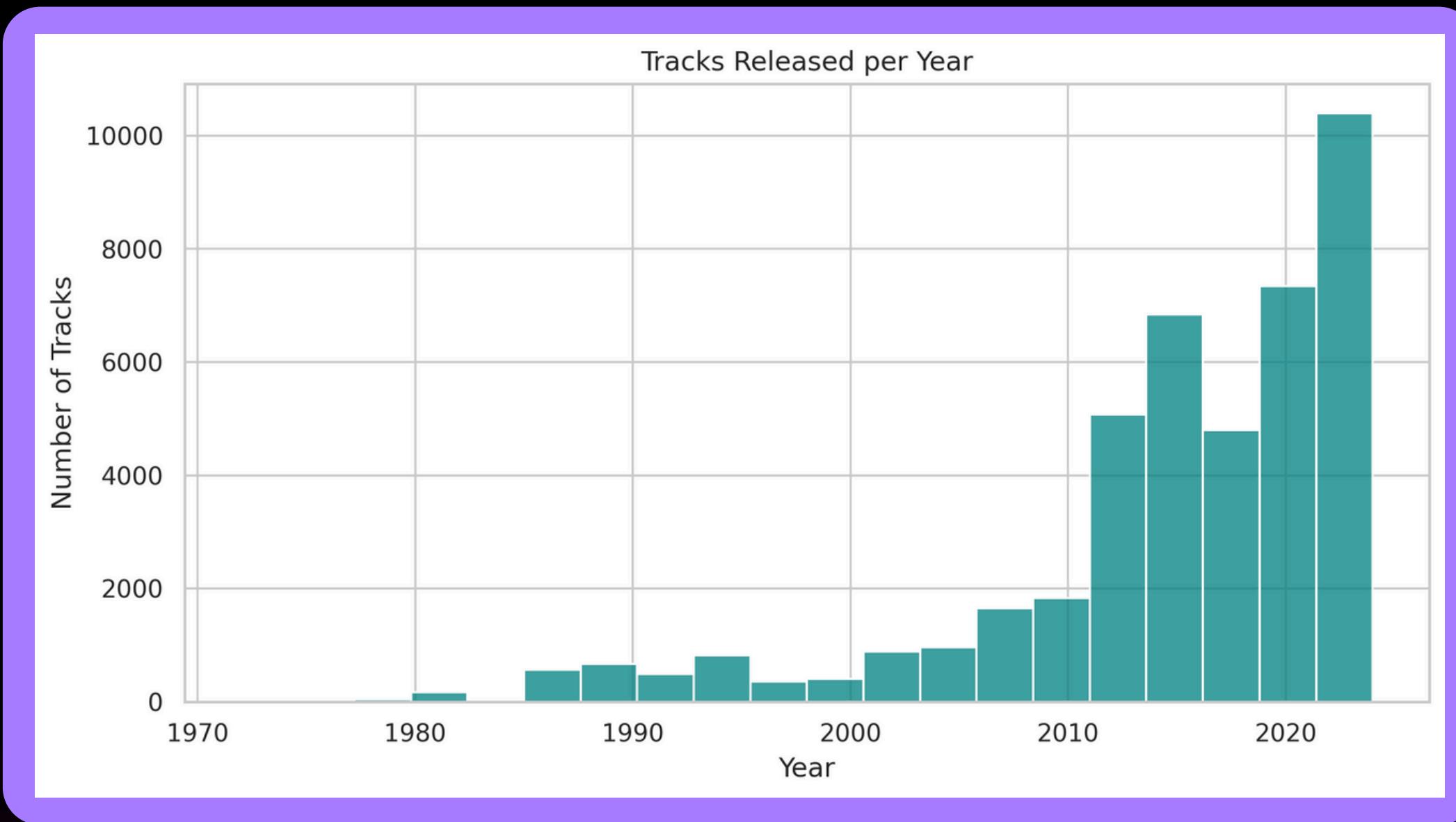
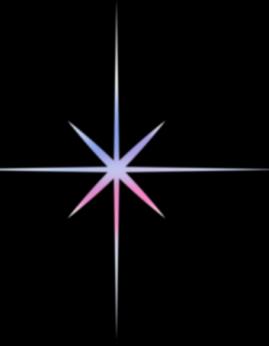
- **Shankar Mahadevan Leads in Track Count:**
 - Shankar Mahadevan has the highest number of tracks – significantly more than others – indicating his extensive music catalog in the dataset.
- **Mix of Indian and International Artists:**
 - The chart includes both Indian artists (like Shreya Ghoshal, A.R. Rahman, Arijit Singh) and global artists suggesting that the dataset contains a diverse range of music across different regions and genres.

Top 10 Albums by Number of Tracks



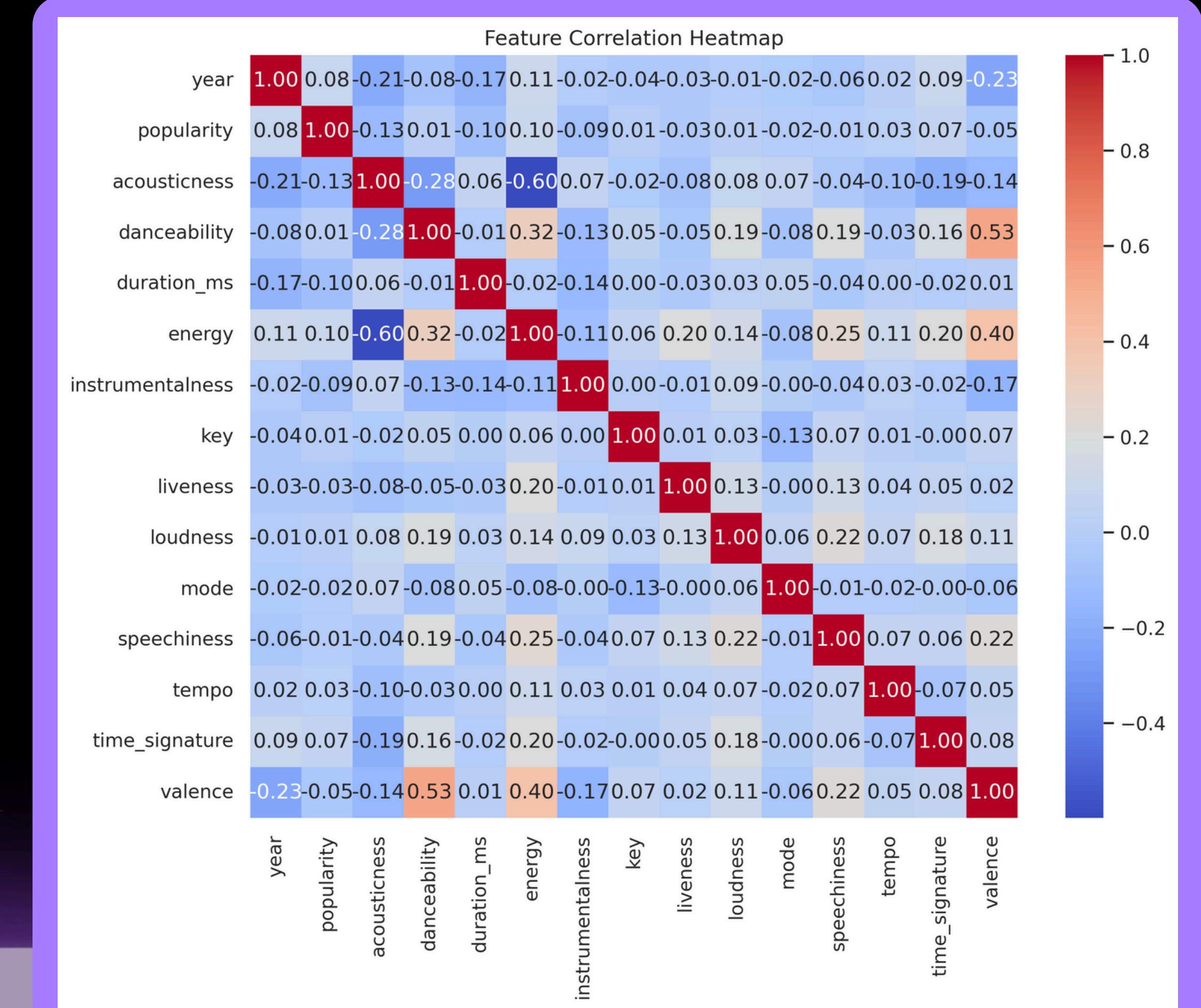
- **‘Dawn FM (Alternate World)’ Leads with the Most Tracks:**
- This album has the highest number of tracks (around 85), indicating it’s a comprehensive or deluxe edition with multiple versions or extended content compared to other albums.
- **Mix of International and Indian Albums:**
- The list includes both global albums (like Taylor Swift’s and The Weeknd’s) and Indian collections (like A.R. Rahman’s and Tamil folk playlists), showing a diverse dataset combining different music cultures and genres.

Tracks Released Per Year

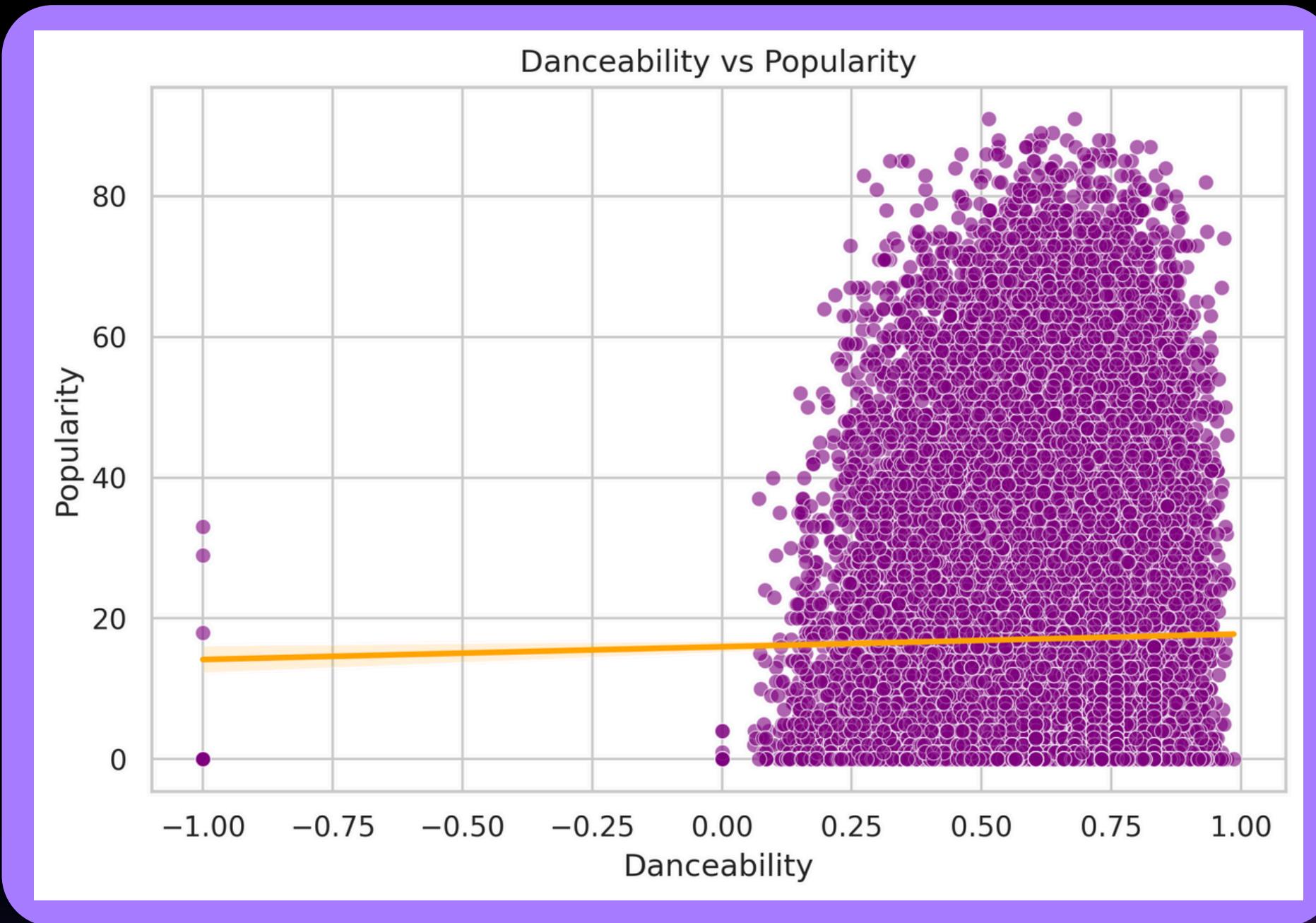
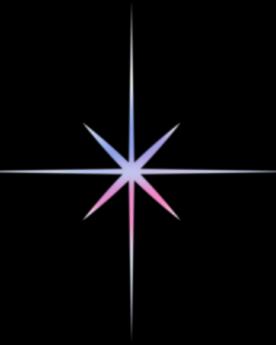


- **Significant Growth After 2010:**
 - The number of tracks released increased sharply after 2010, indicating a boom in music production – likely due to the rise of digital platforms, streaming services, and easier access to music distribution tools.
- **Peak Around 2020s:**
 - The highest number of tracks were released in recent years (around 2020-2023), showing a modern trend of high musical output and continuous growth in the music industry.

Correlation Heatmap Of Various Features

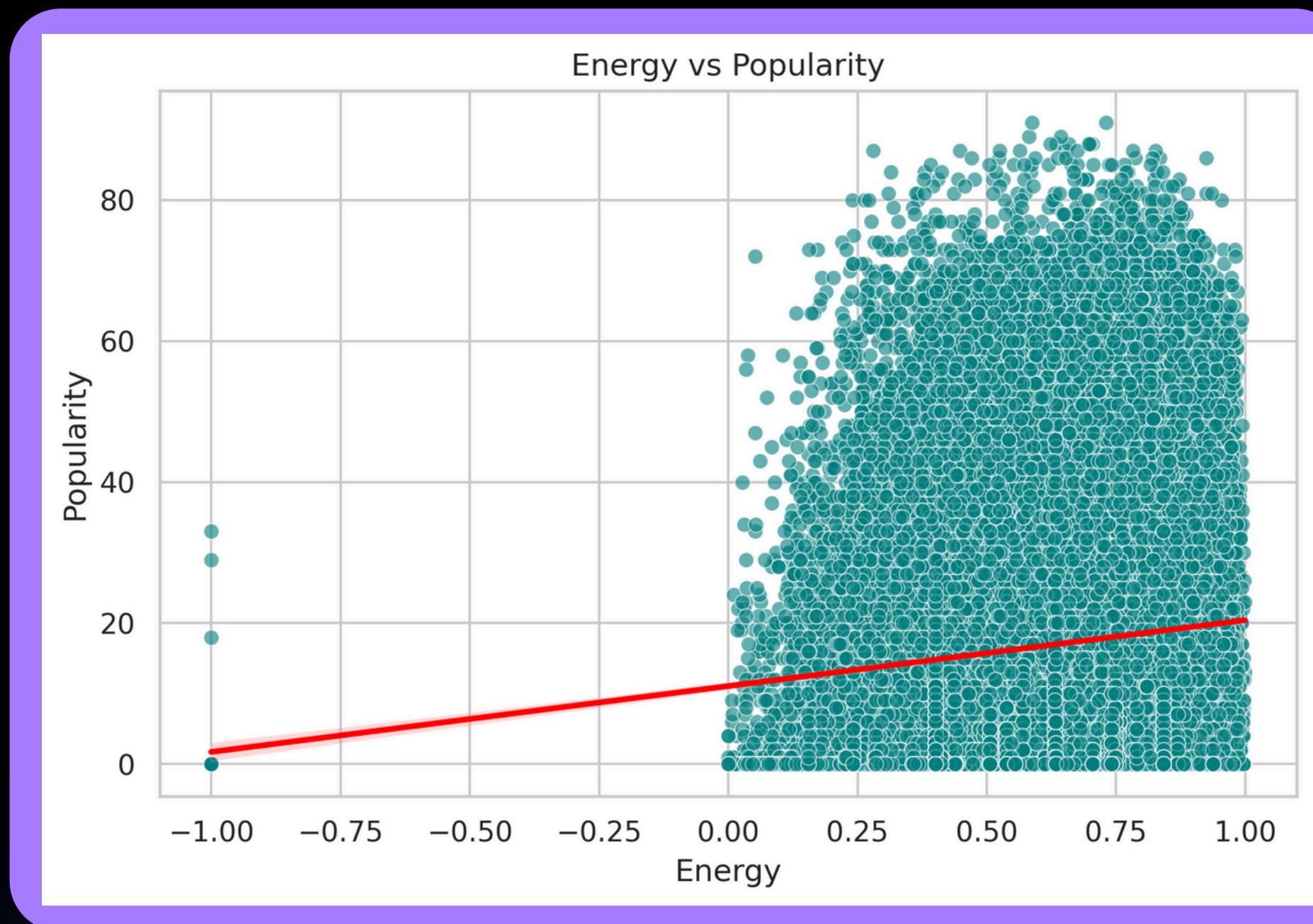
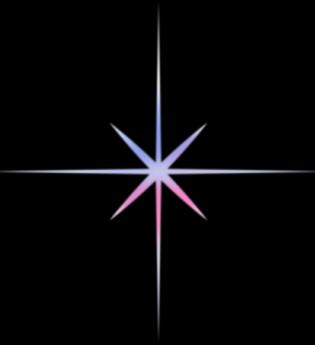


Popularity VS Danceability



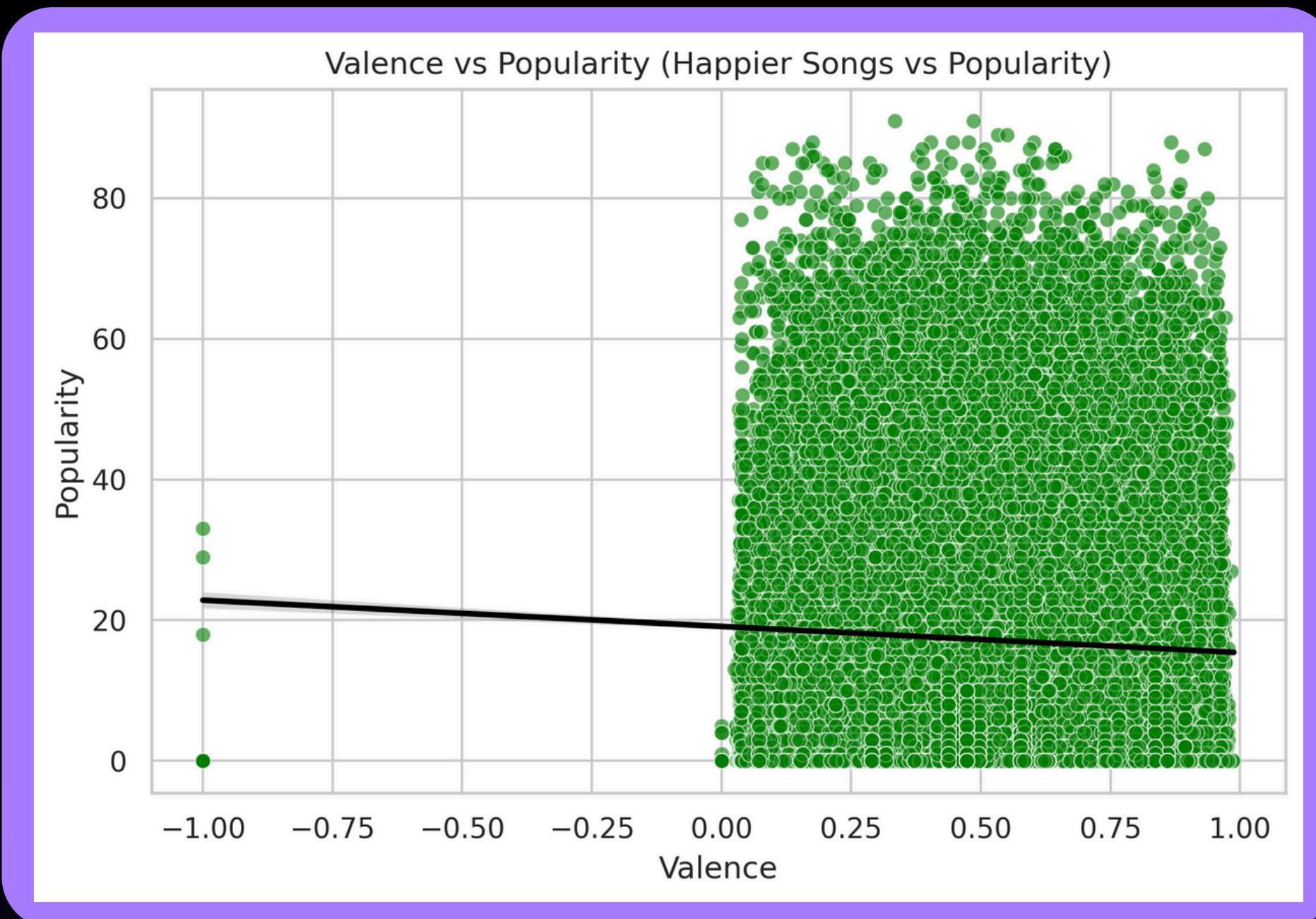
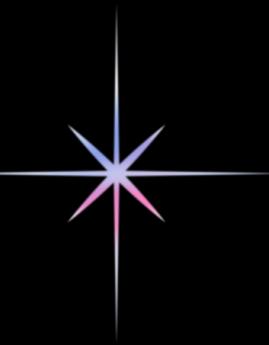
- **Weak Positive Correlation:**
 - There is a slight positive relationship between danceability and popularity – as danceability increases, popularity tends to rise marginally. However, the correlation is not strong, suggesting other factors also influence a song's popularity.
- **Most Popular Tracks Are Fairly Danceable:**
 - The majority of highly popular songs have moderate to high danceability scores (0.5–1.0), indicating that energetic and rhythmic tracks are generally more appealing to listeners.

Popularity VS Energy



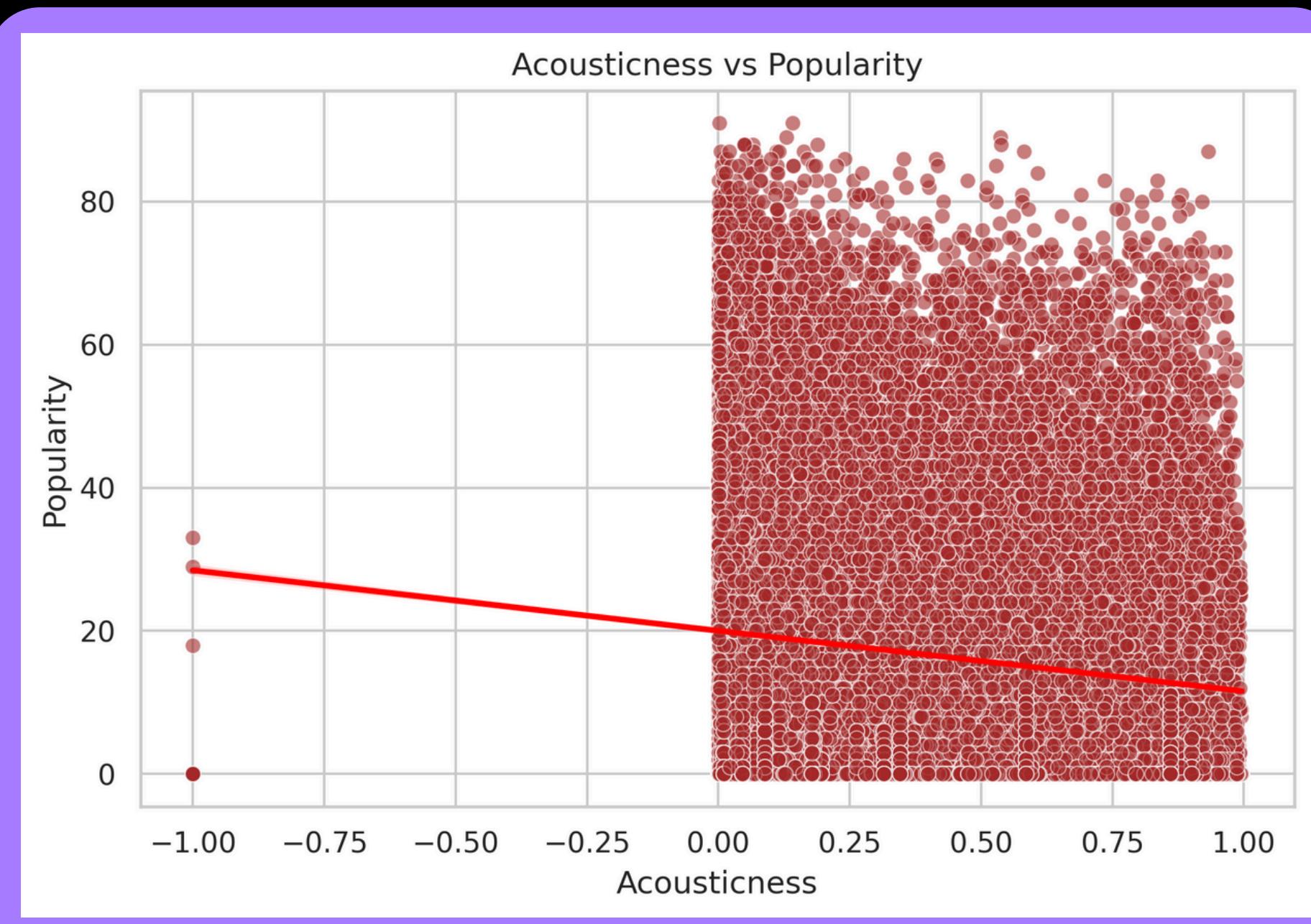
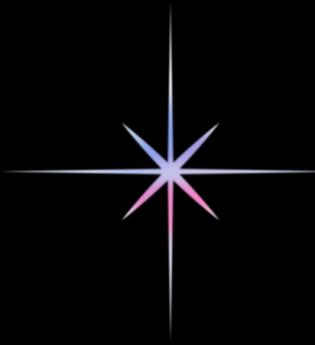
- **Slight Positive Correlation:**
 - There is a weak positive trend between energy and popularity – songs with higher energy levels tend to be slightly more popular, but the relationship is not strong.
- **High Popularity at Moderate to High Energy Levels:**
 - Most of the popular tracks fall within the energy range of 0.5 to 1.0, suggesting that listeners generally prefer energetic and upbeat songs over low-energy tracks.

Popularity VS Valence



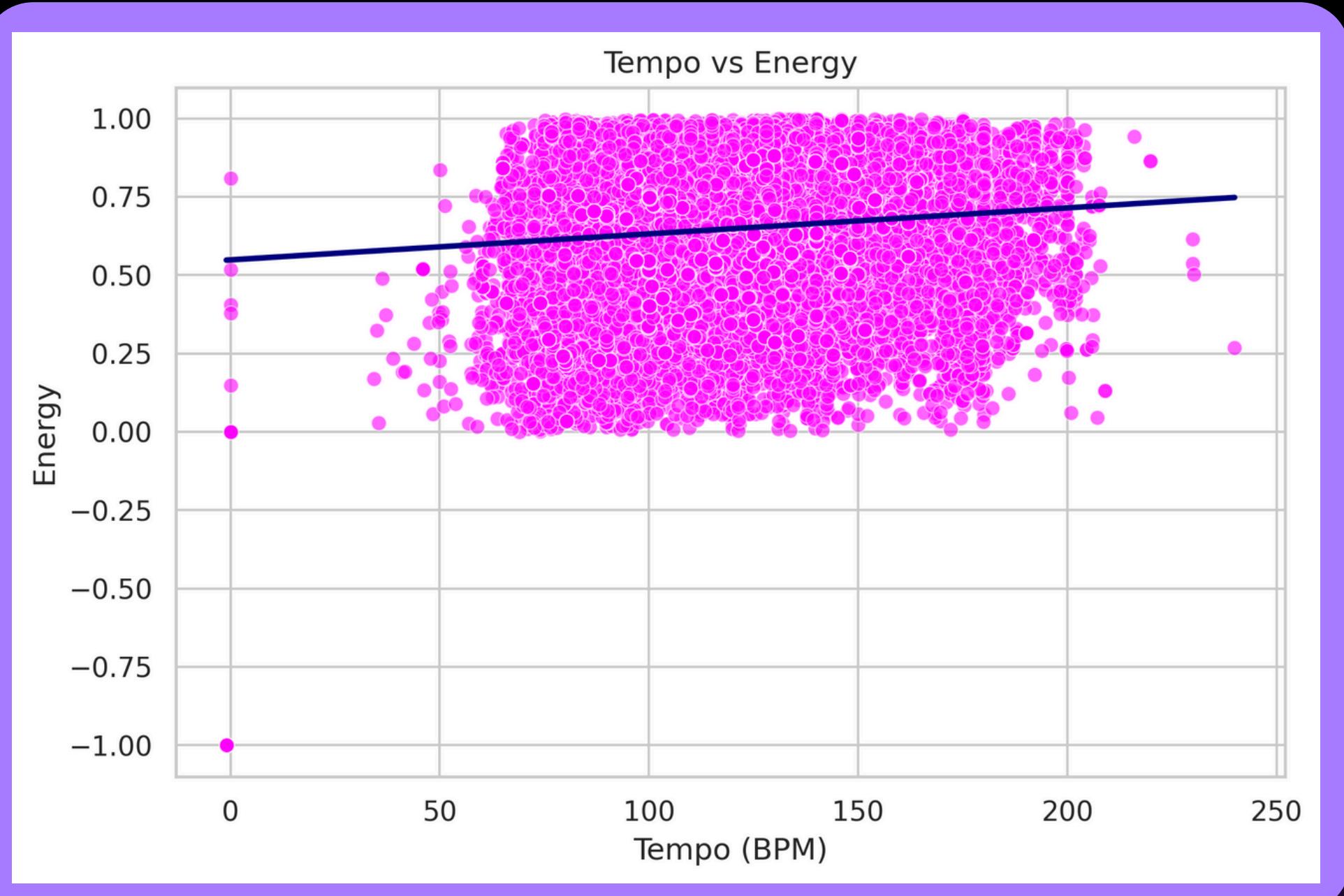
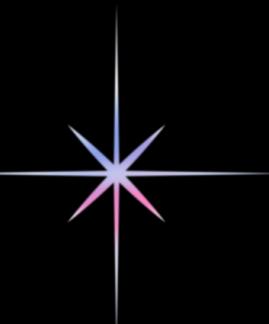
- **Weak Negative Correlation:**
 - There is a slight negative relationship between valence (happiness of a song) and popularity – indicating that happier songs are not necessarily more popular.
- **Popularity Spread Across Valence Range:**
 - Popular tracks are spread across all valence levels, meaning both happy and sad songs can achieve high popularity, suggesting that listeners' preferences are diverse and not solely based on the emotional tone of the music.

Popularity VS Acousticness



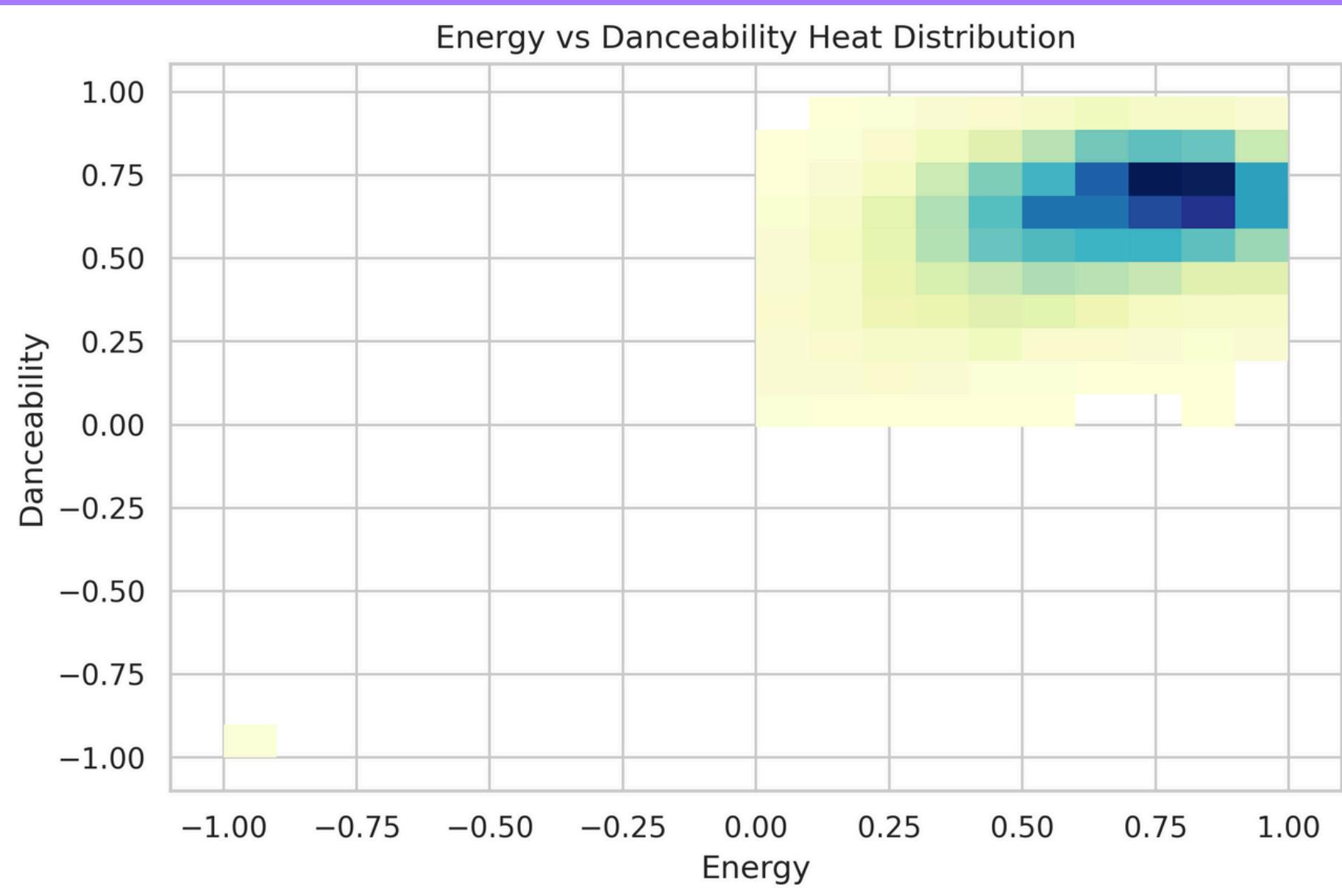
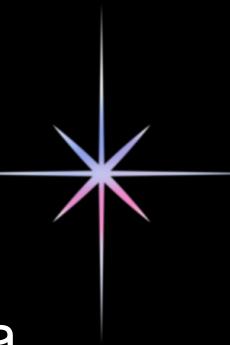
- **Weak Negative Relationship:**
 - There is a slight negative correlation between acousticness and popularity – songs that are more acoustic tend to be slightly less popular overall.
- **Popularity Concentration in Lower Acousticness Range:**
 - Most popular songs cluster around low to moderate acousticness values (0.0–0.4), suggesting that listeners generally prefer tracks with electronic or produced elements over purely acoustic ones.

Energy VS Tempo



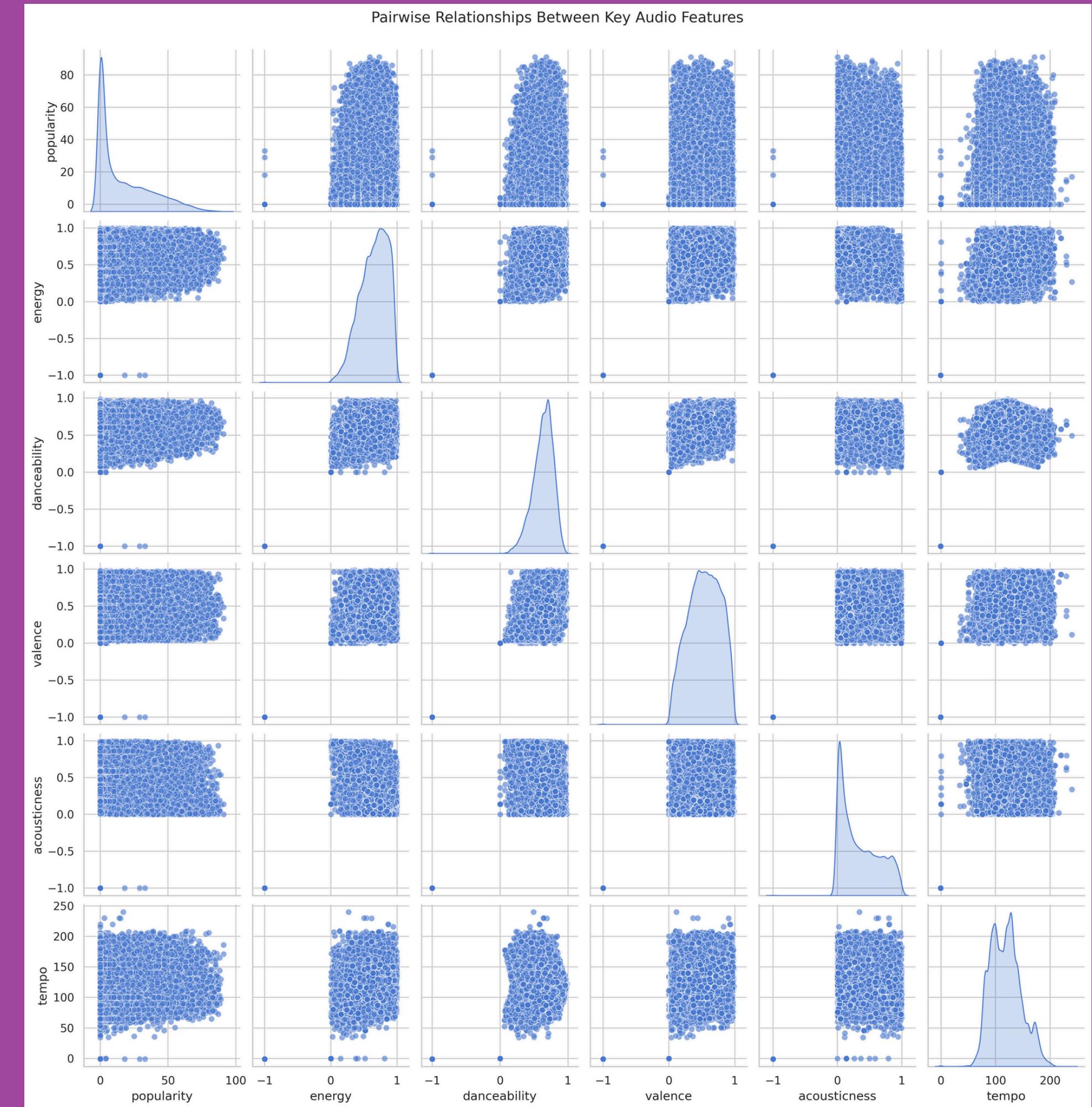
- **Strong Positive Correlation:** The graph shows a clear positive relationship between Tempo (BPM) and Energy. As the tempo increases from 0 to 250 BPM, the energy level also consistently rises.
- **Non-Linear Relationship:** The increase in energy is not constant. The energy rises most sharply at lower tempos and then the rate of increase appears to slow down at higher tempos (e.g., from ~150 to 250 BPM), suggesting a potential plateau effect where adding more BPM has a diminishing return on energy.

Energy VS Danceability Heat Distribution

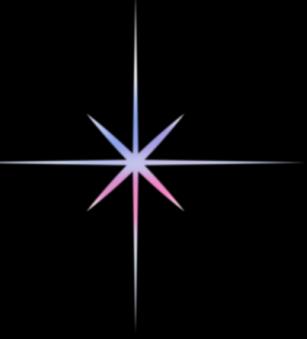


- **Potential Positive Correlation:** The data suggests a likely positive correlation between tempo and energy. As the tempo (BPM) increases along the x-axis the scale for energy on the y-axis also increases.
- **Focus on Mid-Range Tempo for High Energy:** The energy scale includes negative values, but the highest energy ratings (0.75 to 1.00) are likely associated with the mid-to-high tempo range. This indicates that the most energetic tracks are not necessarily the absolute fastest or slowest, but fall within a specific, common "up-tempo" band.

Pairwise Relationship Between Key Audio Features



Conclusion



- **Focus on Popular Audio Traits:**
 - Songs with higher danceability, energy, and moderate valence (positive emotion) tend to achieve greater popularity. Spotify can promote or recommend tracks with these features more prominently in curated playlists.
- **Balance Acoustic Content:**
 - Highly acoustic tracks generally show lower popularity. Spotify could feature these more selectively – for example, in niche playlists (e.g., “Acoustic Moods”, “Relax & Unplugged”) rather than general trending sections.
- **Support Emerging Artists & Albums:**
 - Based on the top artist and album distributions, a few dominate the charts. Spotify might highlight lesser-known artists through personalized recommendations or “Discover New Voices” campaigns.
- **Leverage Temporal Trends:**
 - From the tracks per year data, Spotify can identify growth periods or dips in song releases and align marketing or playlisting strategies (like celebrating throwbacks or yearly trends).

THANK YOU!

