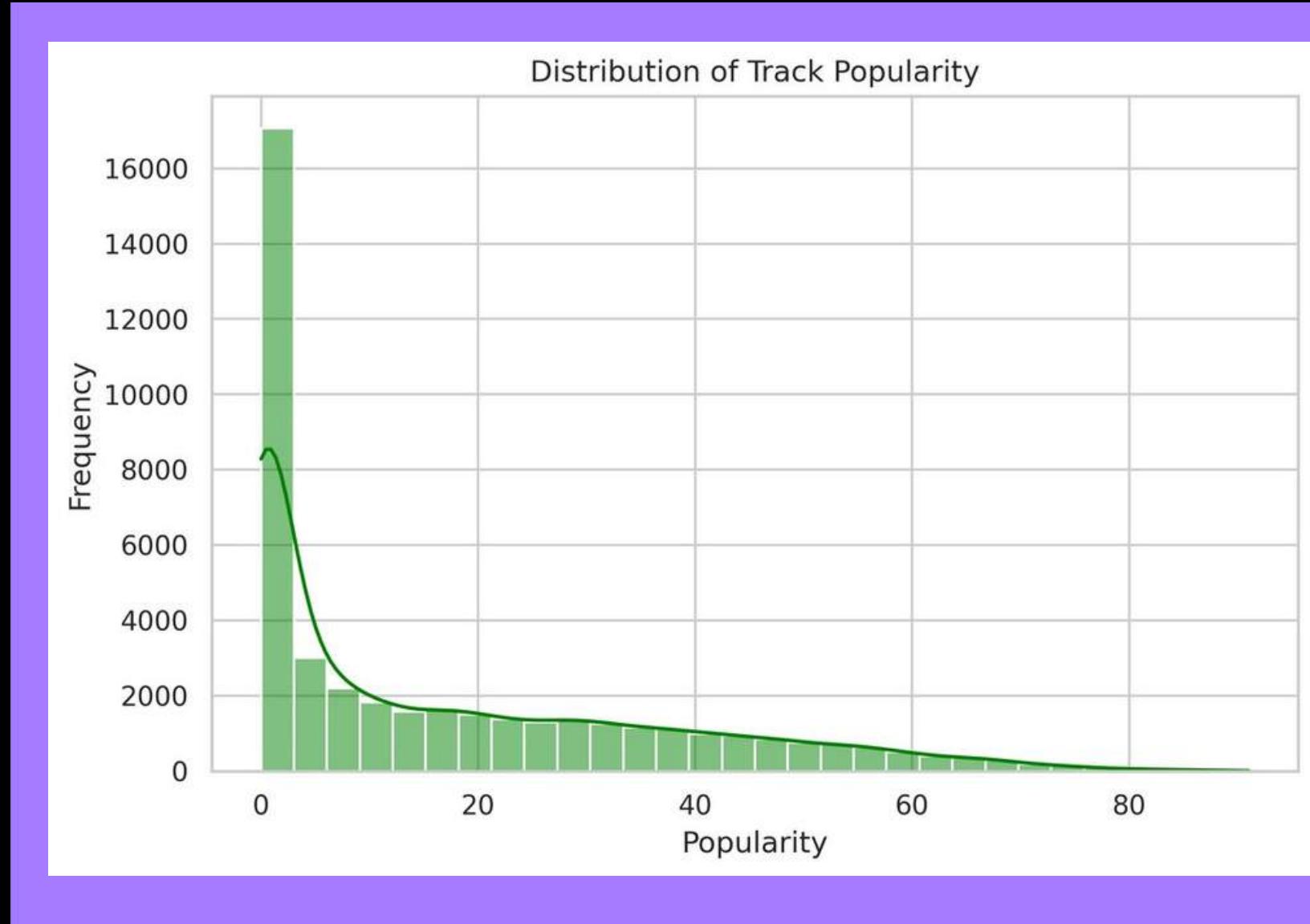
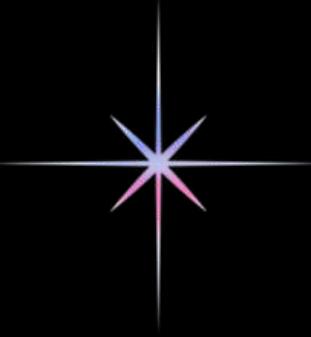


# **EXPLORATORY DATA ANALYSIS OF SPORTIFY DATA TRACKS**

Prepared By - **SRINJAN BHAKAT (IT)**  
**GCECTB - R24 - 2039**

Next Slide

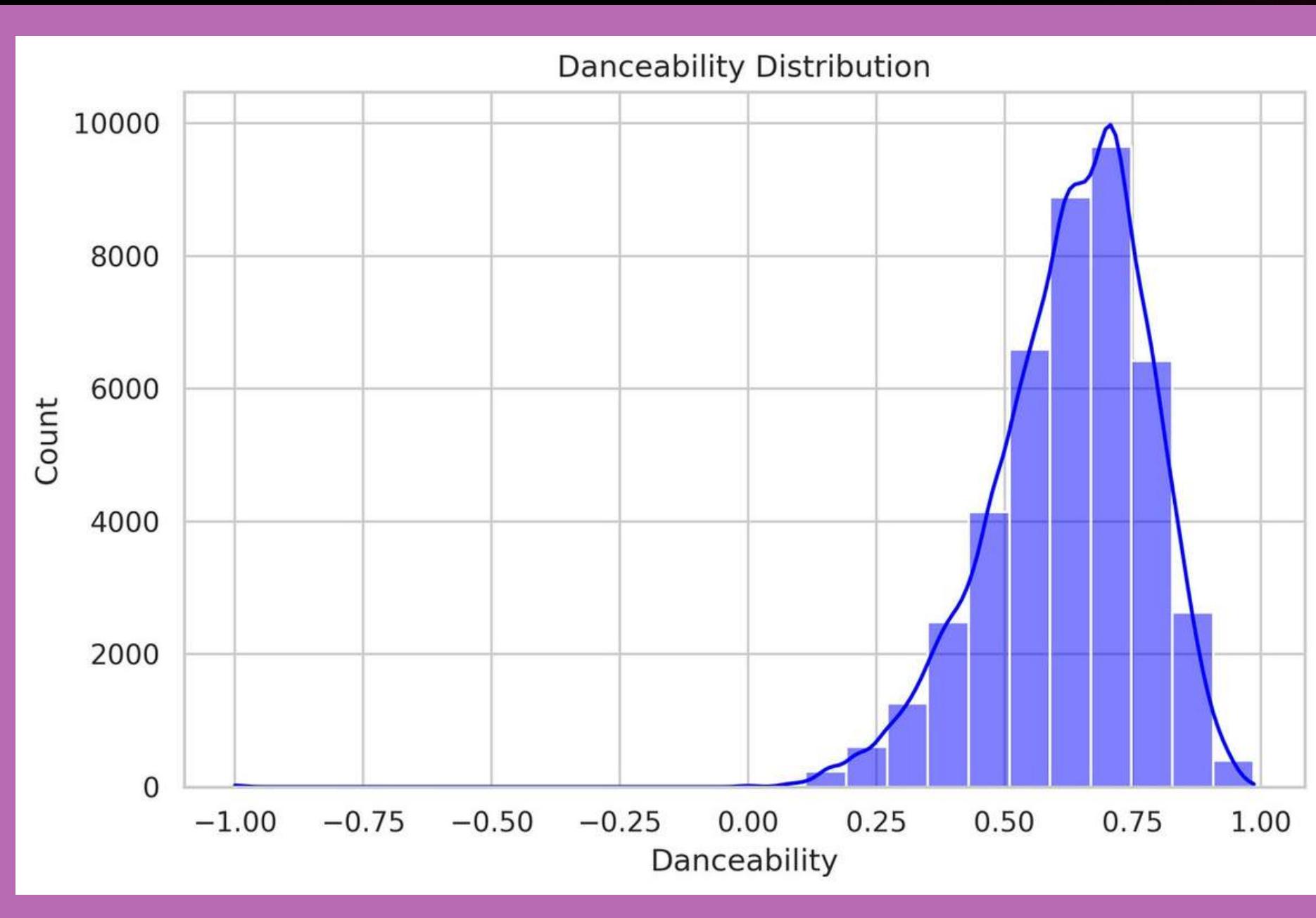
# Popularity Distribution



- **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).

Next Slide

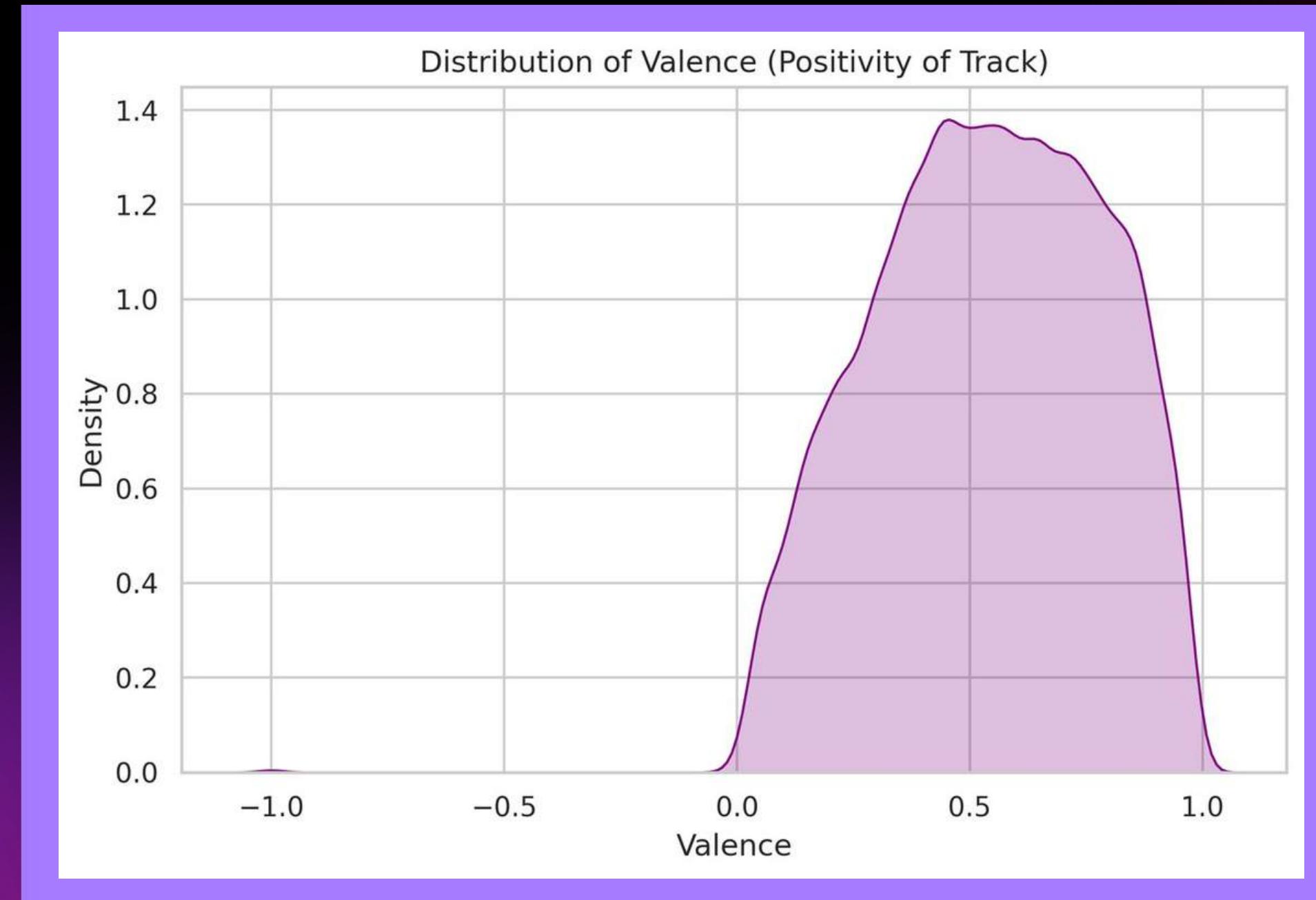
# Danceability Distribution



- **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.

Next Slide

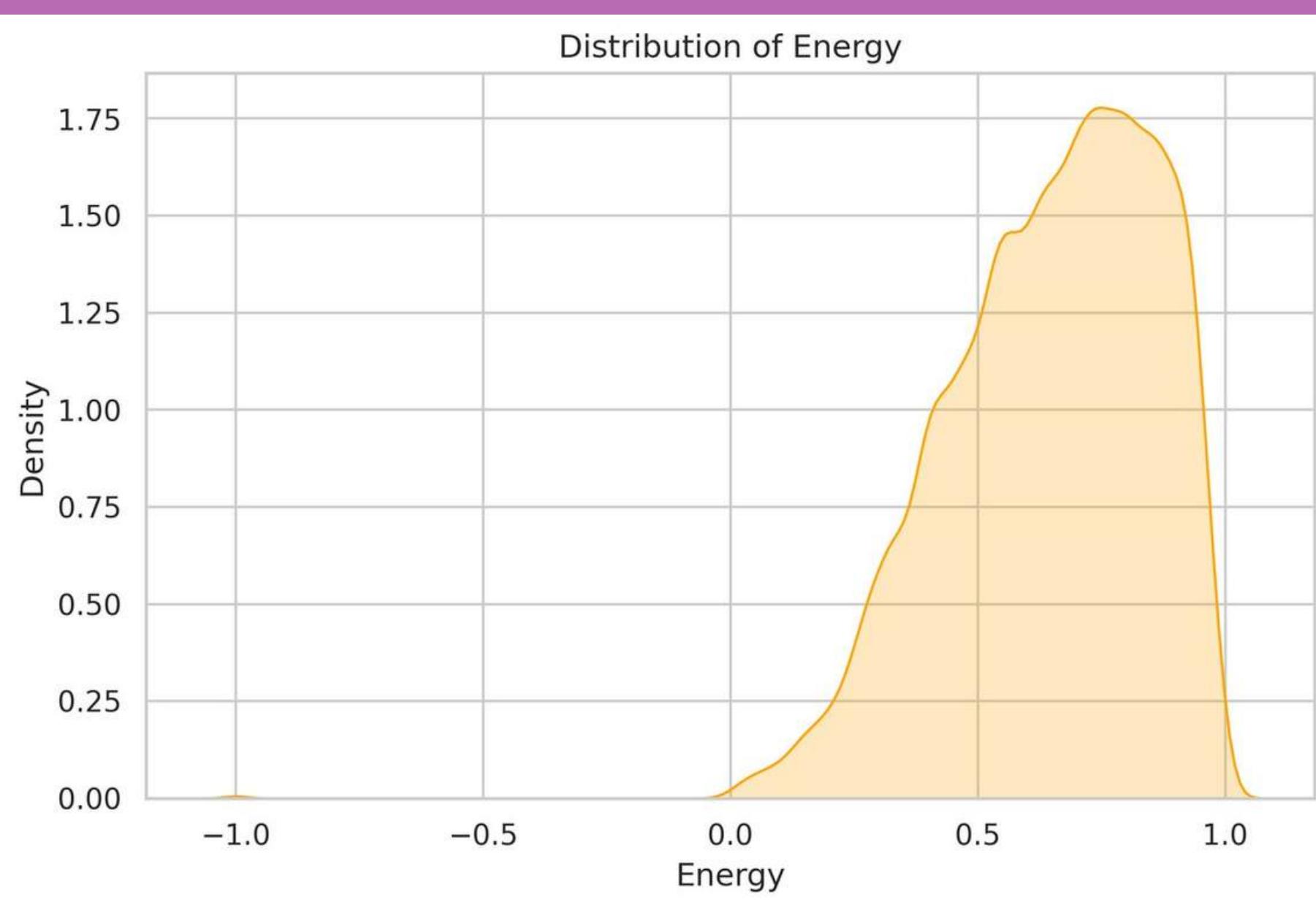
# Valence Distribution



- **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.

Next Slide

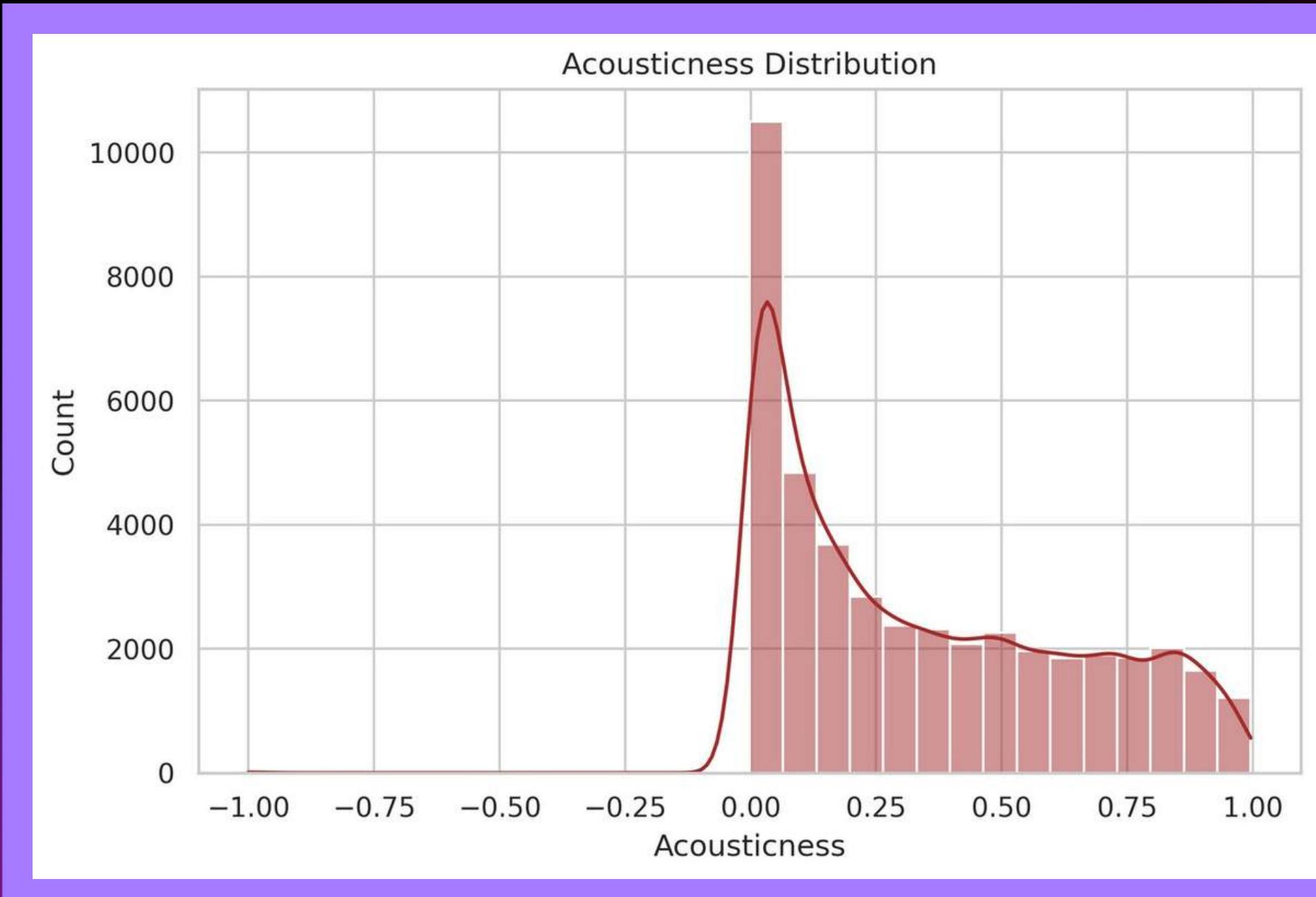
# Energy Distribution



- **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.

Next Slide

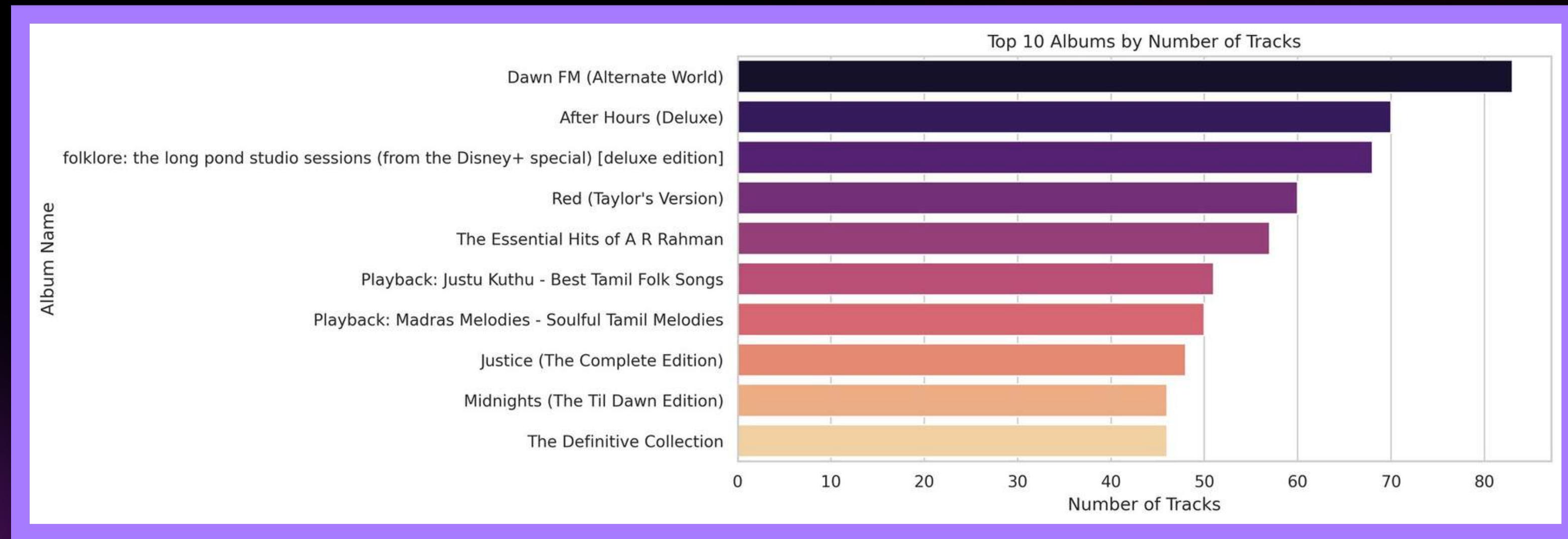
# Acousticness Distribution



- **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.

Next Slide

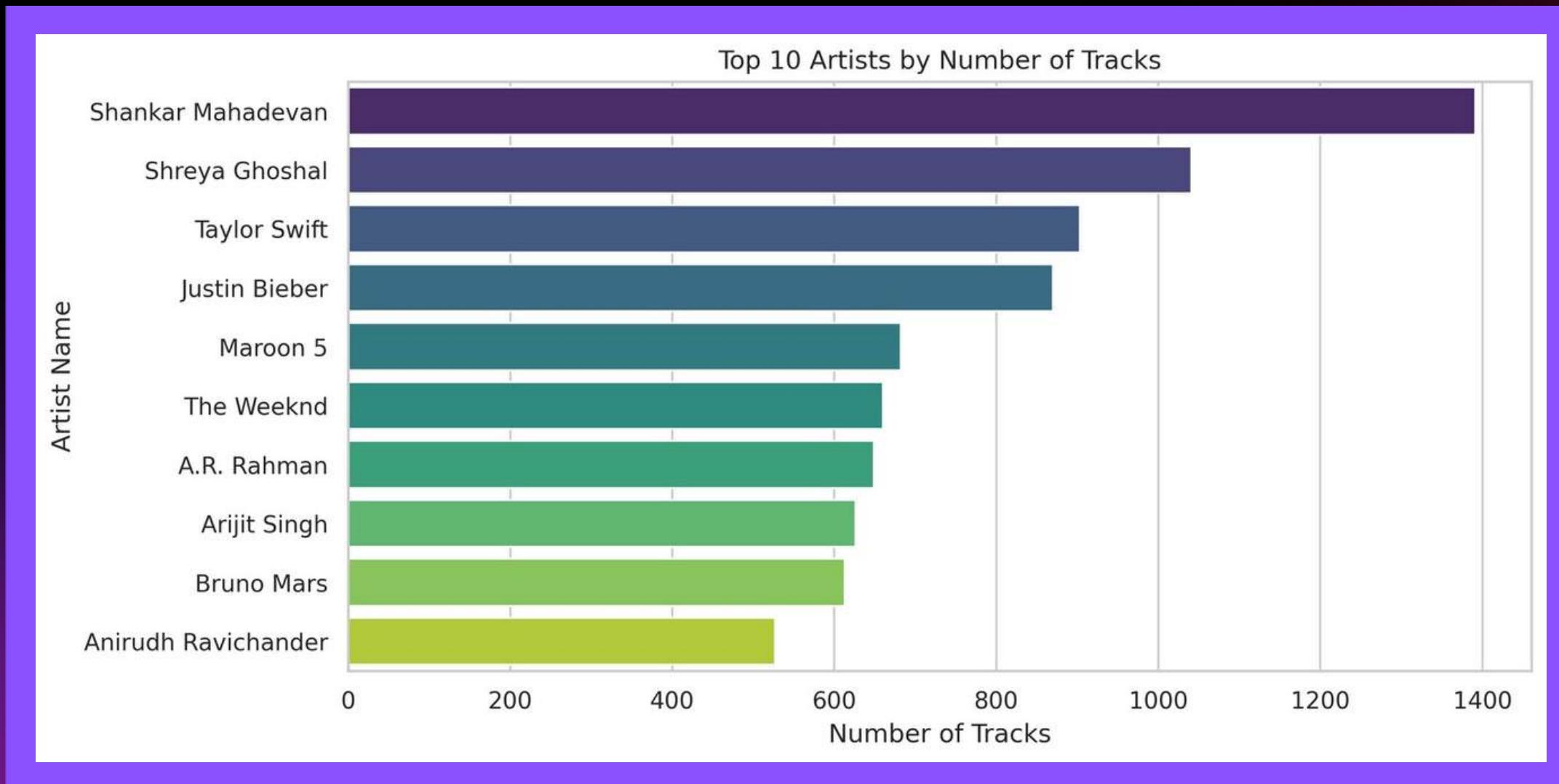
# Top 10 Albums by Number of Tracks



- **‘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.

Next Slide

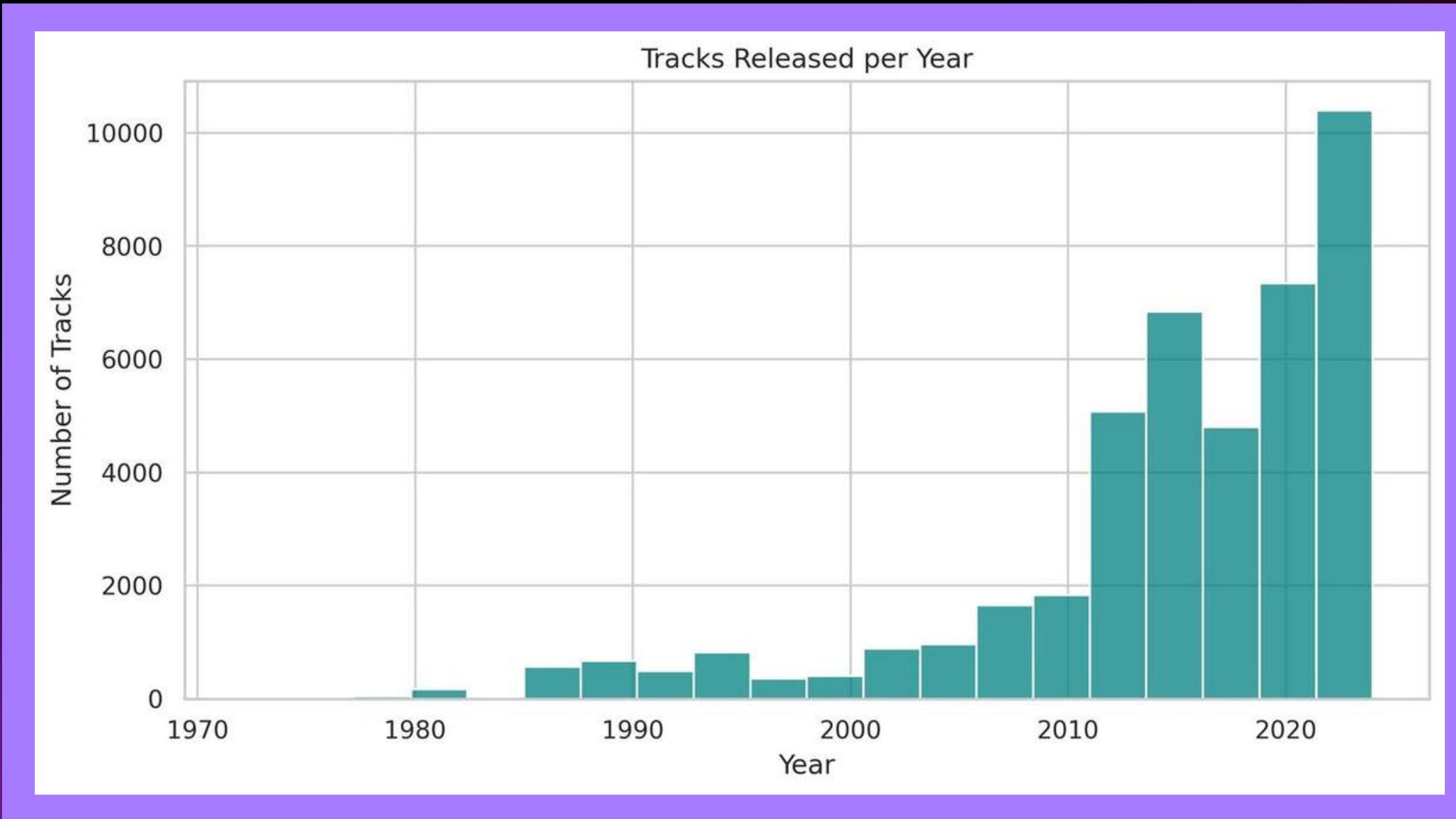
# Top 10 Artists by Number of Tracks



- **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.

Next Slide

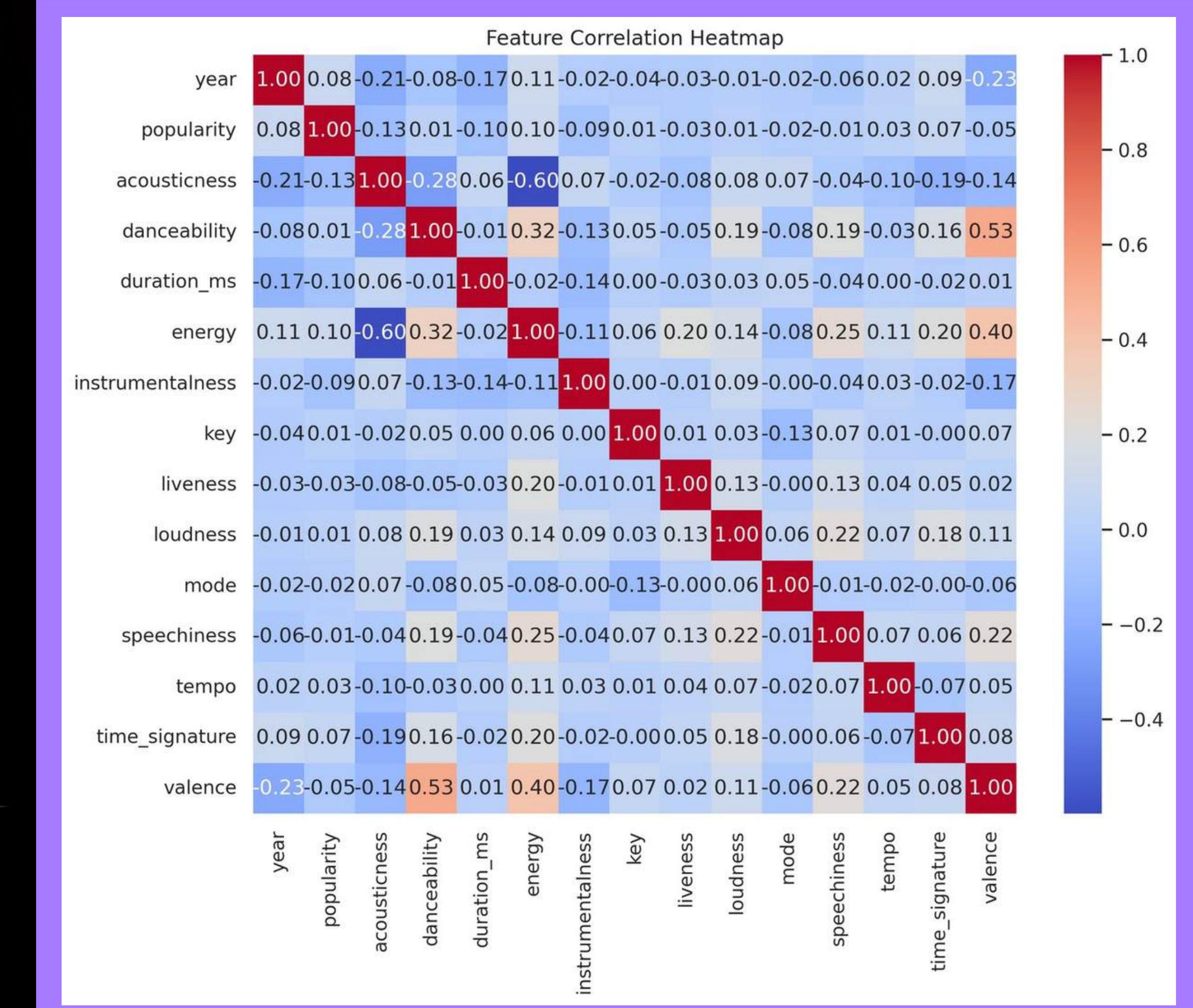
# Tracks Released Per Year



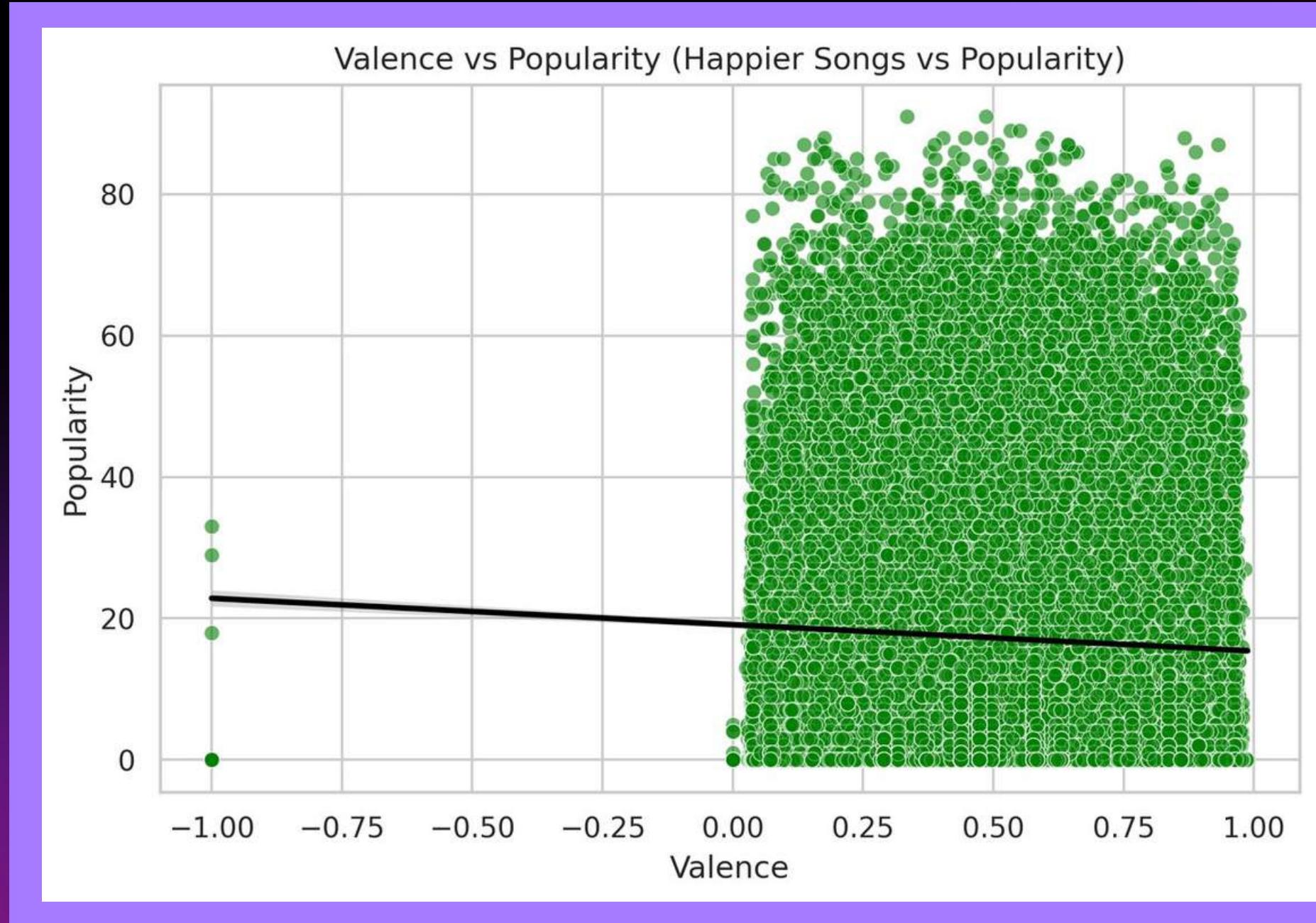
- **Significant Growth After 2010:**
  - The number of tracks released increased sharply after 2010, indicating a boom in music production – likely due to the rise of digital platforms, streaming services, and easier access to music distribution tools.
- **Peak Around 2020s:** The highest number of tracks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

Next Slide

# Correlation Heatmap Of Various Features



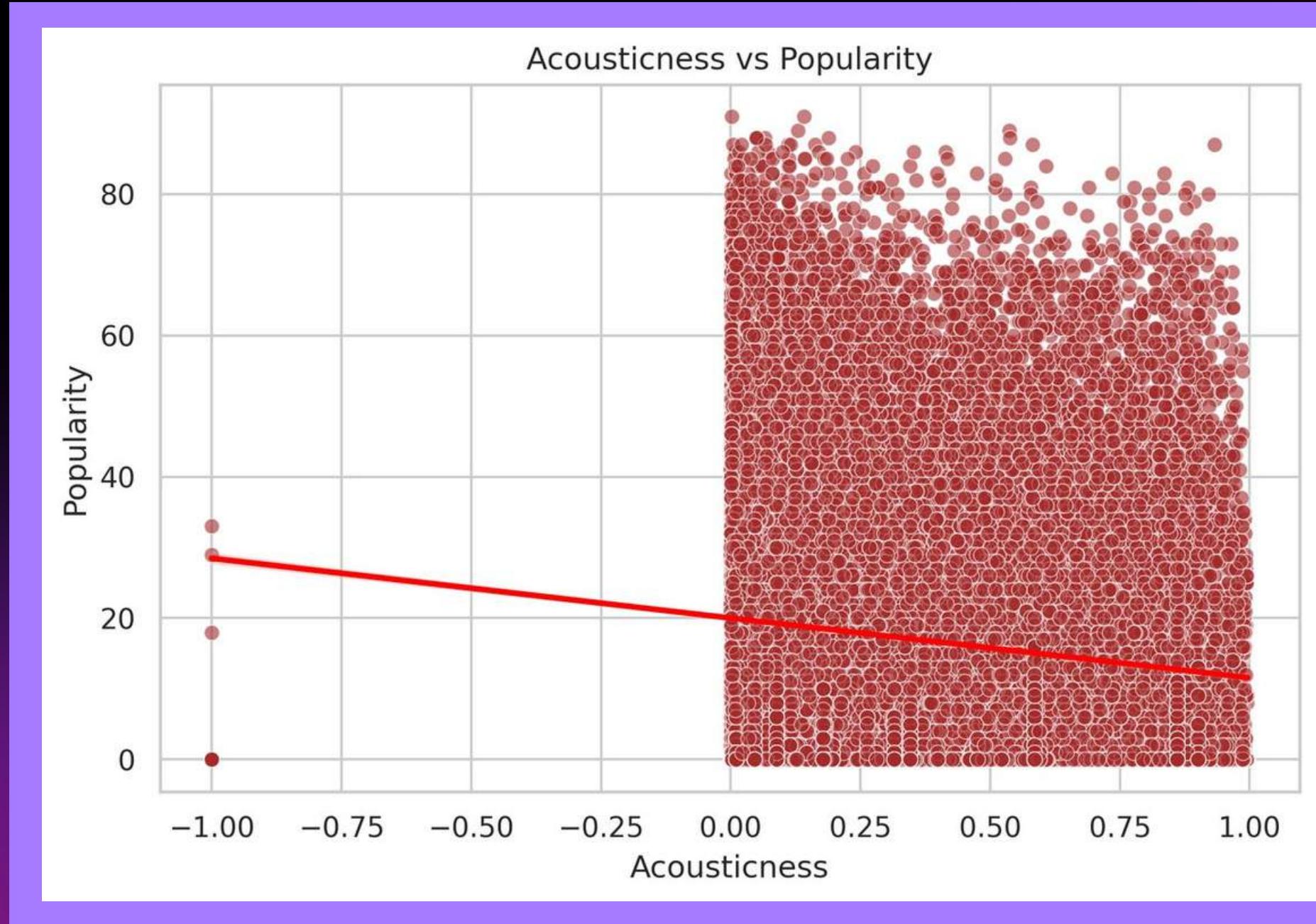
# Popularity VS Valence



- **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.

Next Slide

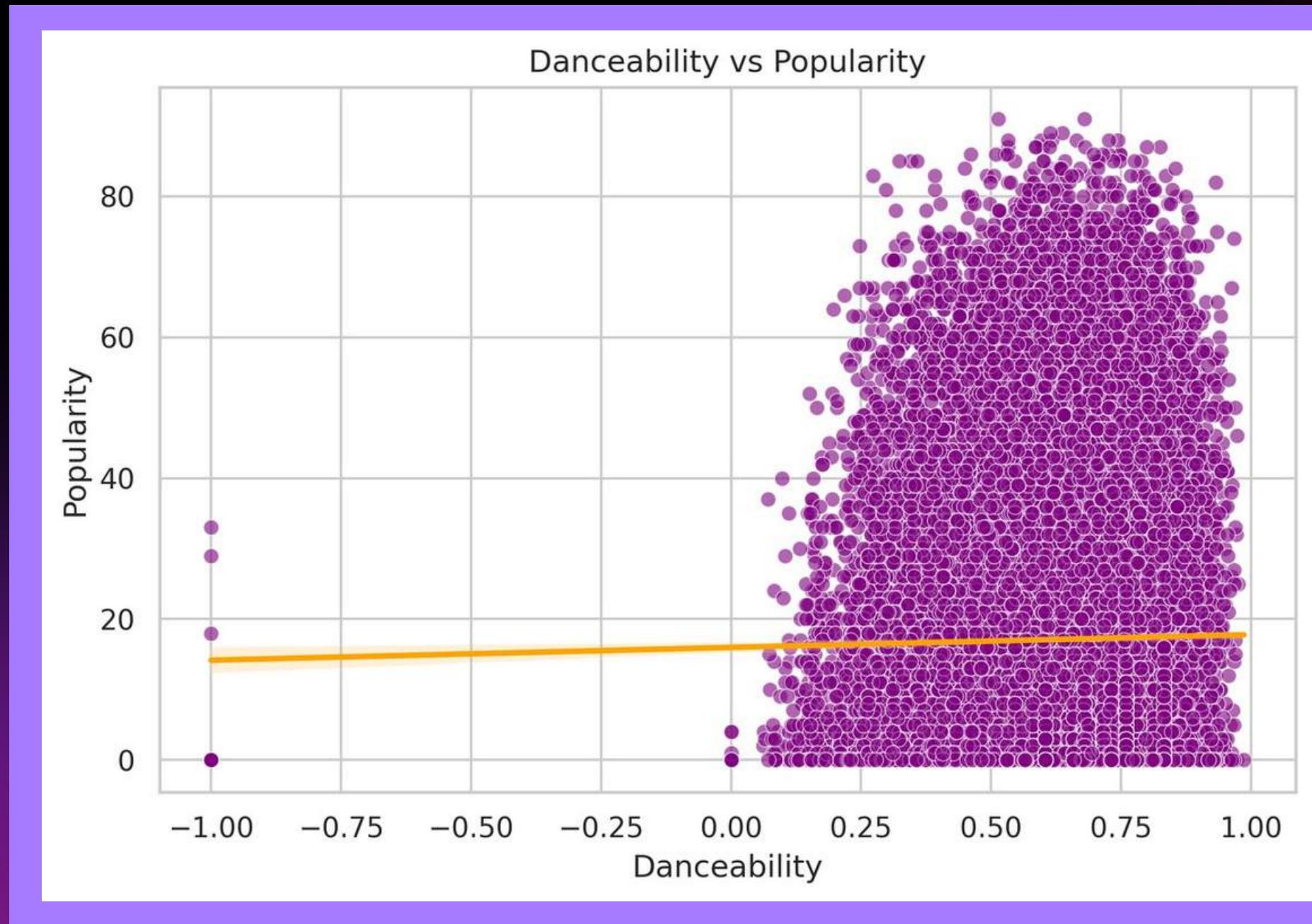
# Popularity VS Acousticness



- **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.

Next Slide

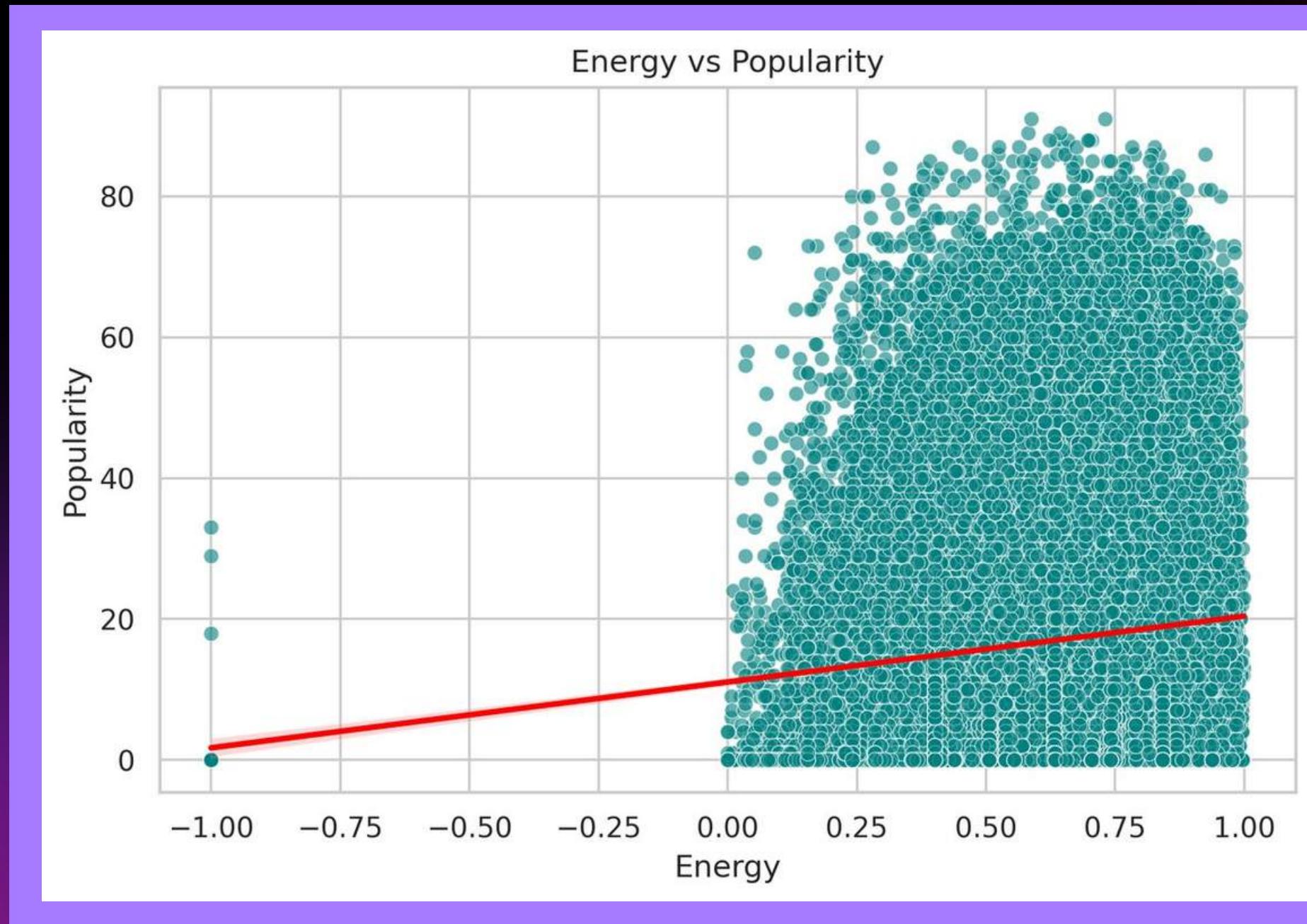
# Popularity VS Danceability



- **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.

Next Slide

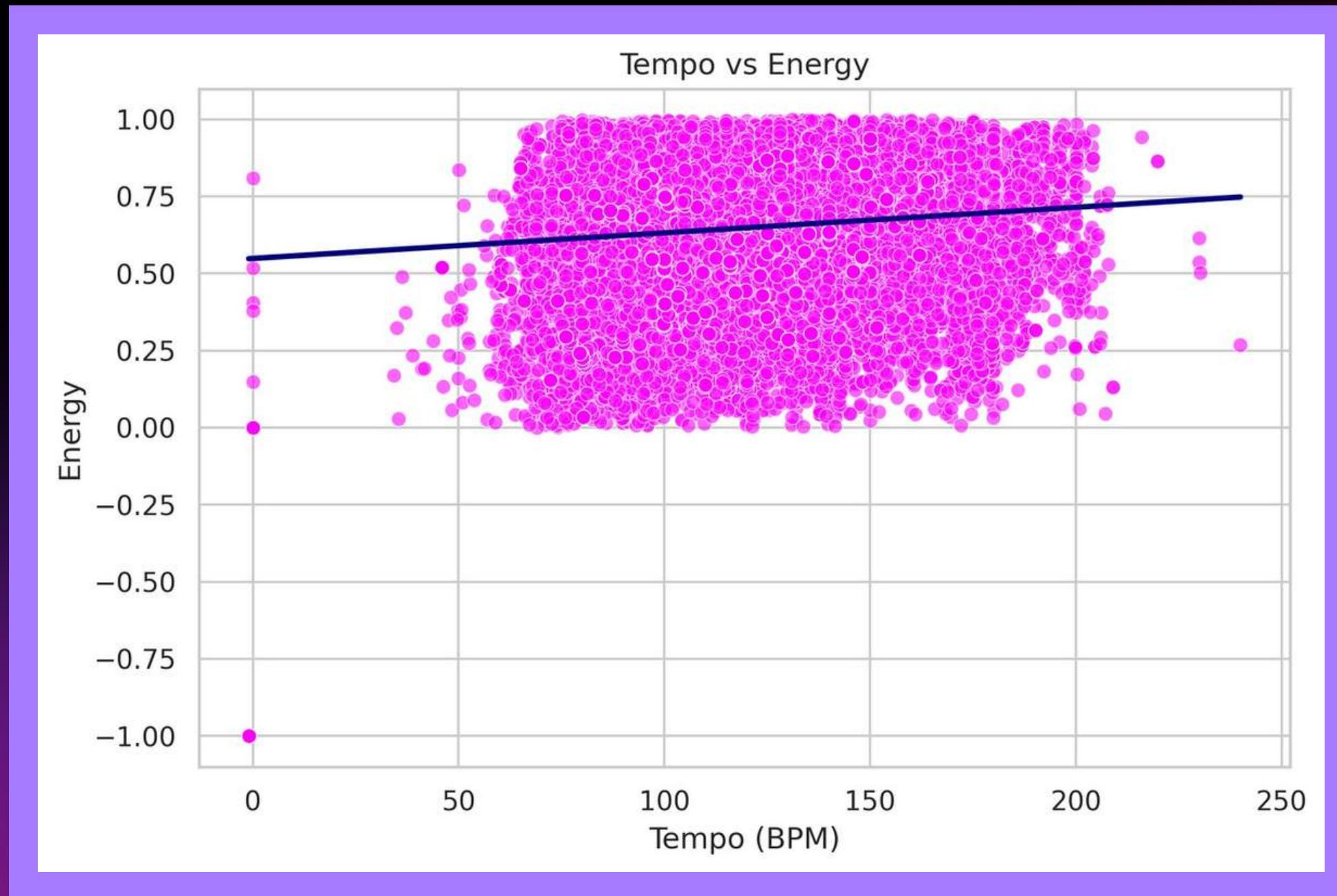
# Popularity VS Energy



- **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.

Next Slide

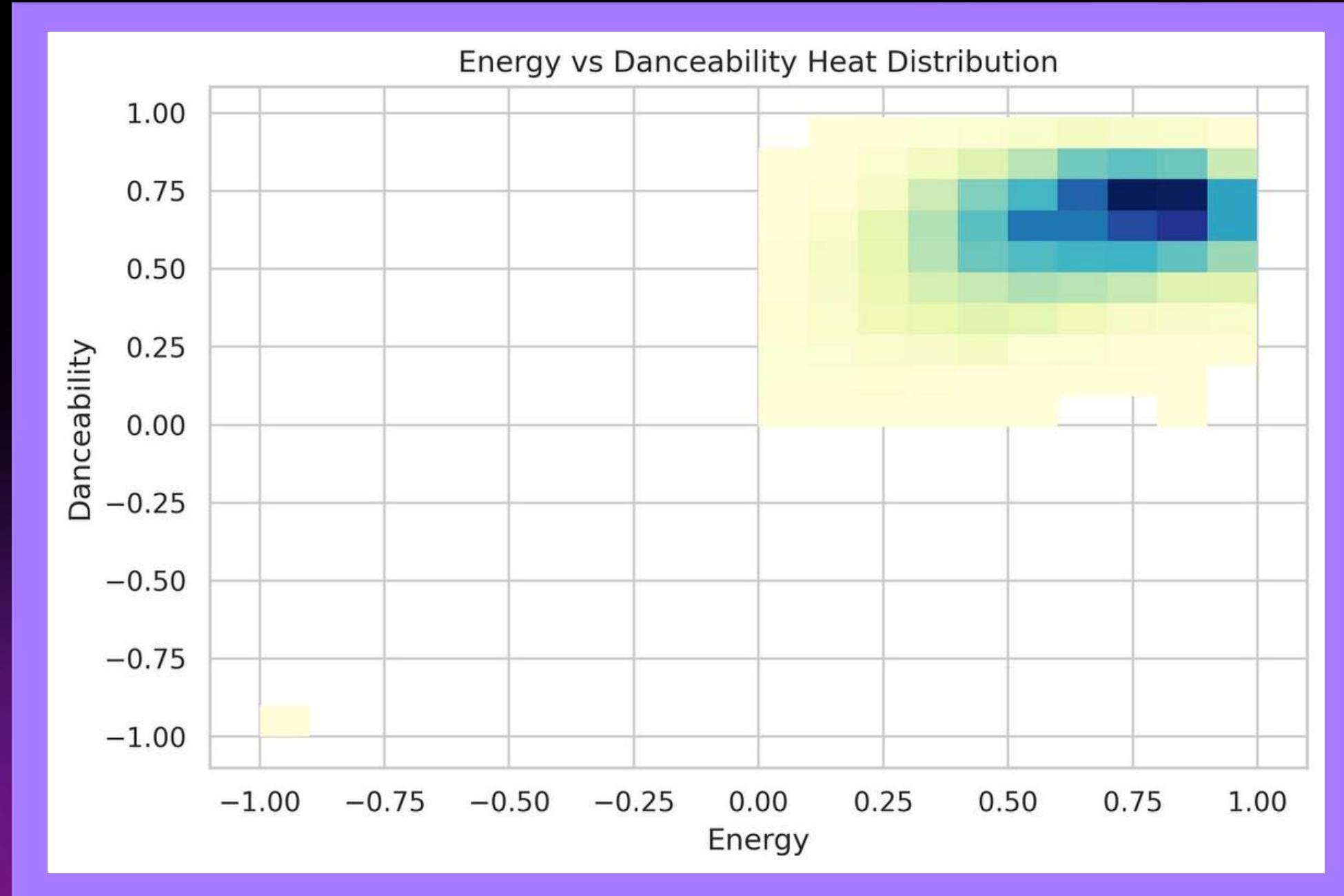
# Energy VS Tempo



- **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.

Next Slide

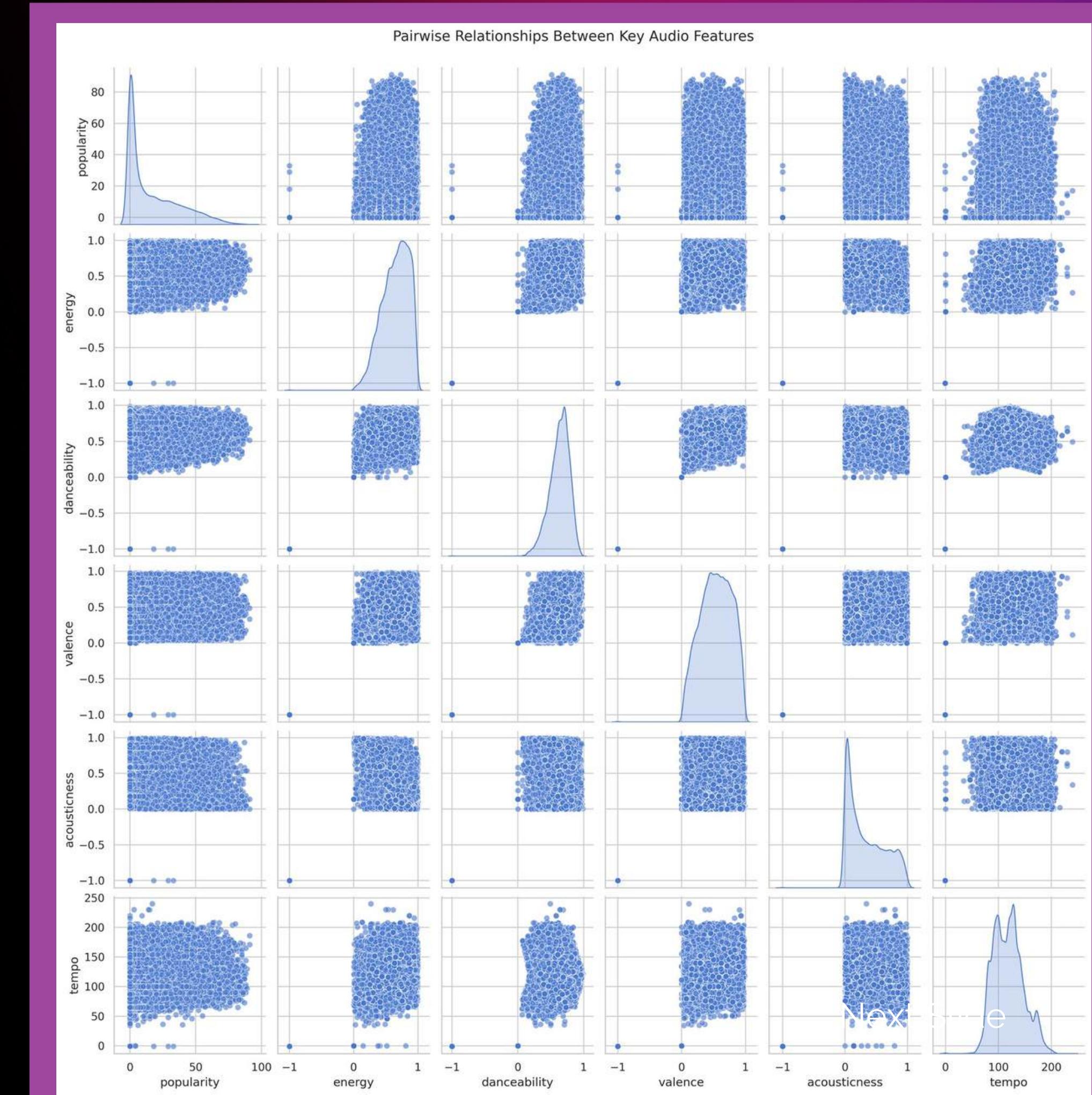
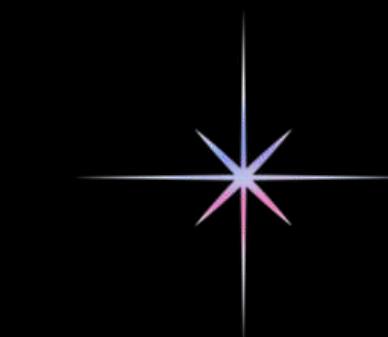
# Energy VS Danceability Heat Distribution



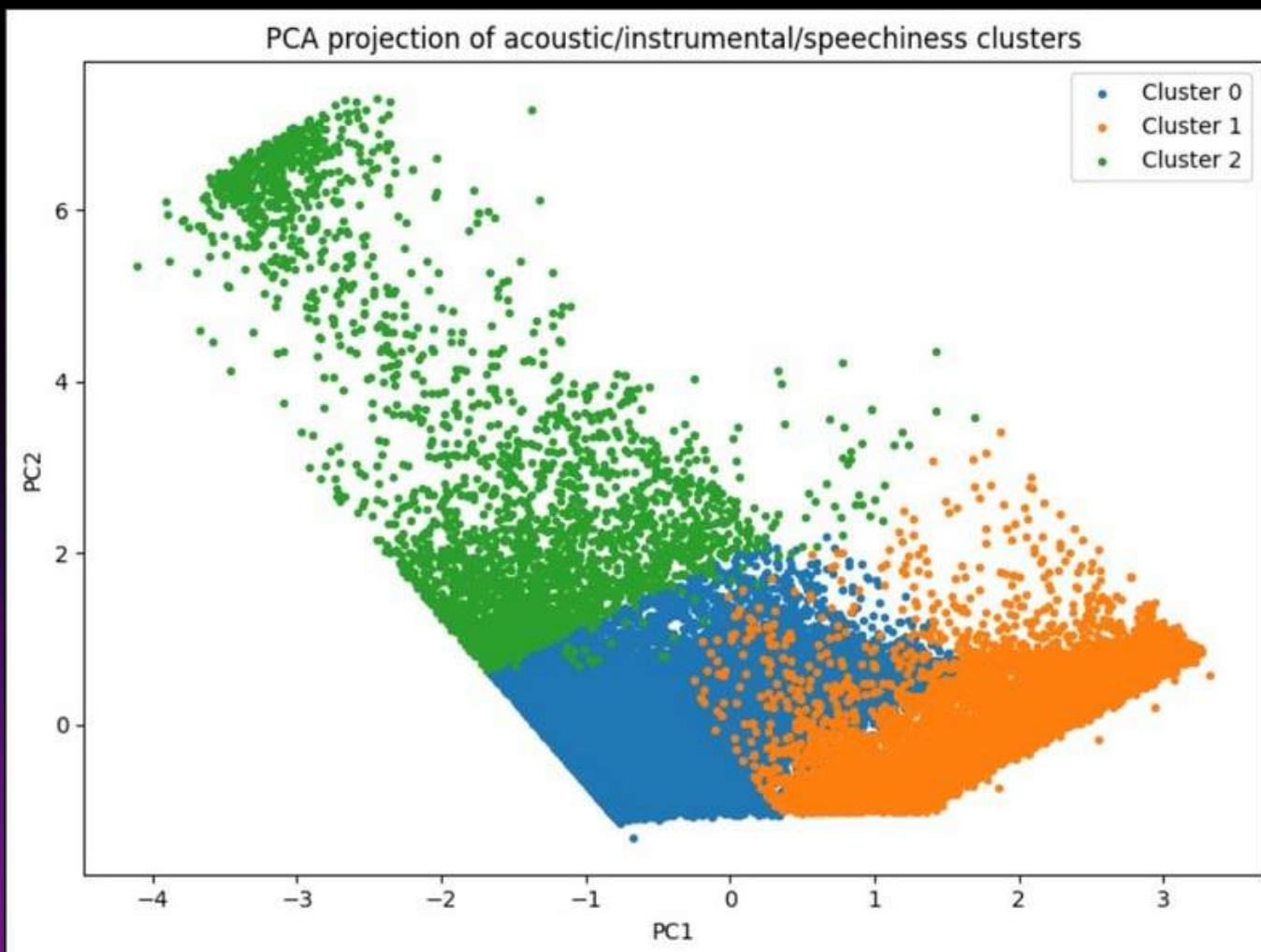
- **Potential Positive Correlation:** The data suggests a likely positive correlation between tempo and energy. As the tempo (BPM) increases along the x-axis the scale for energy on the y-axis also increases.
- **Focus on Mid-Range Tempo for High Energy:** The energy scale includes negative values, but 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.

Next Slide

# Pairwise Relationship Between Key Audio Features

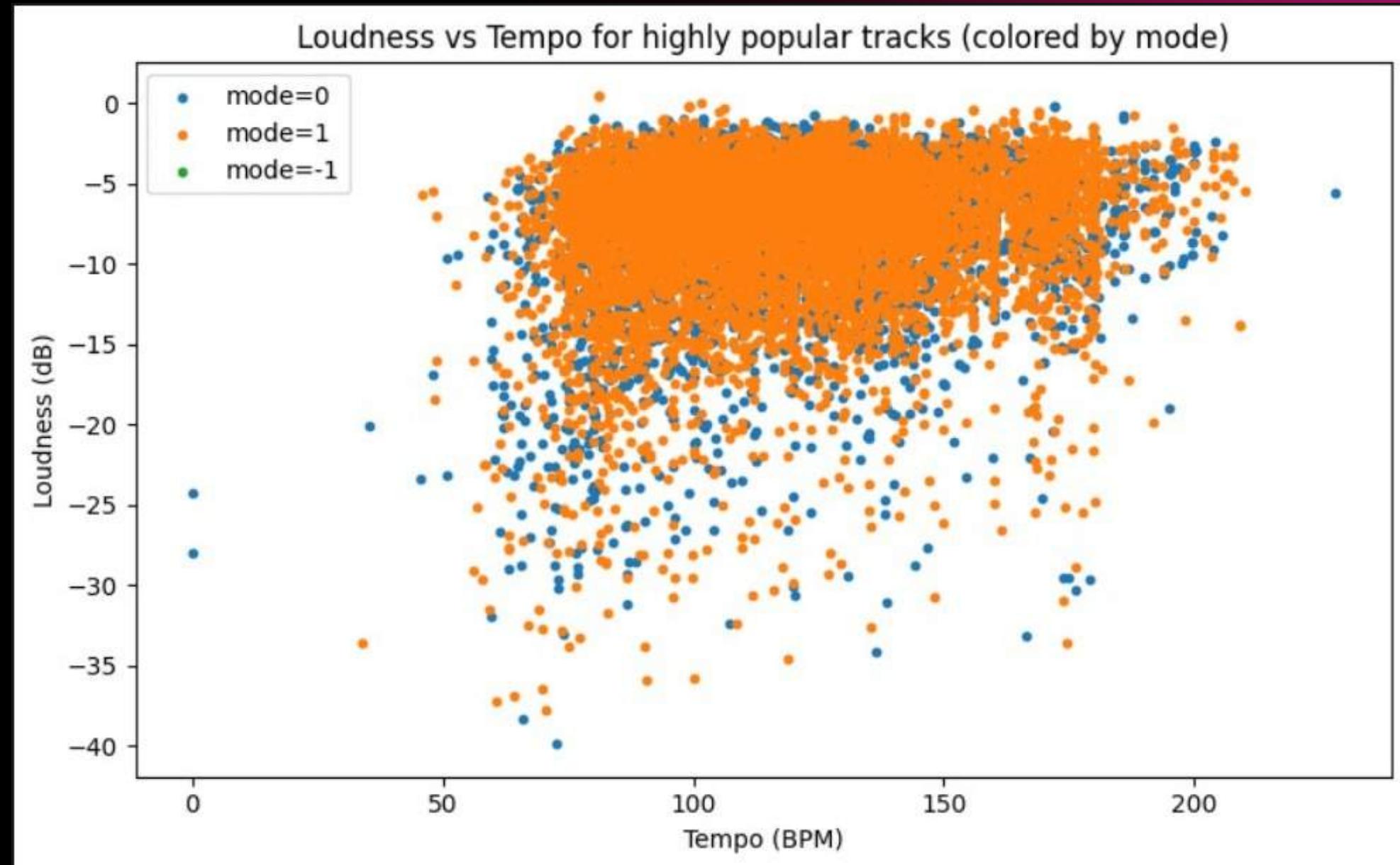


# Multivariate Analysis



- **Distinct clustering pattern:**
- The PCA plot clearly separates songs into three clusters, indicating meaningful differences in acousticness, instrumentalness, and speechiness.
- **Cluster 2 dominates a specific feature range:**
- The large green region suggests many tracks share high acoustic or speech-related characteristics, forming a dense cluster.

Next Slide



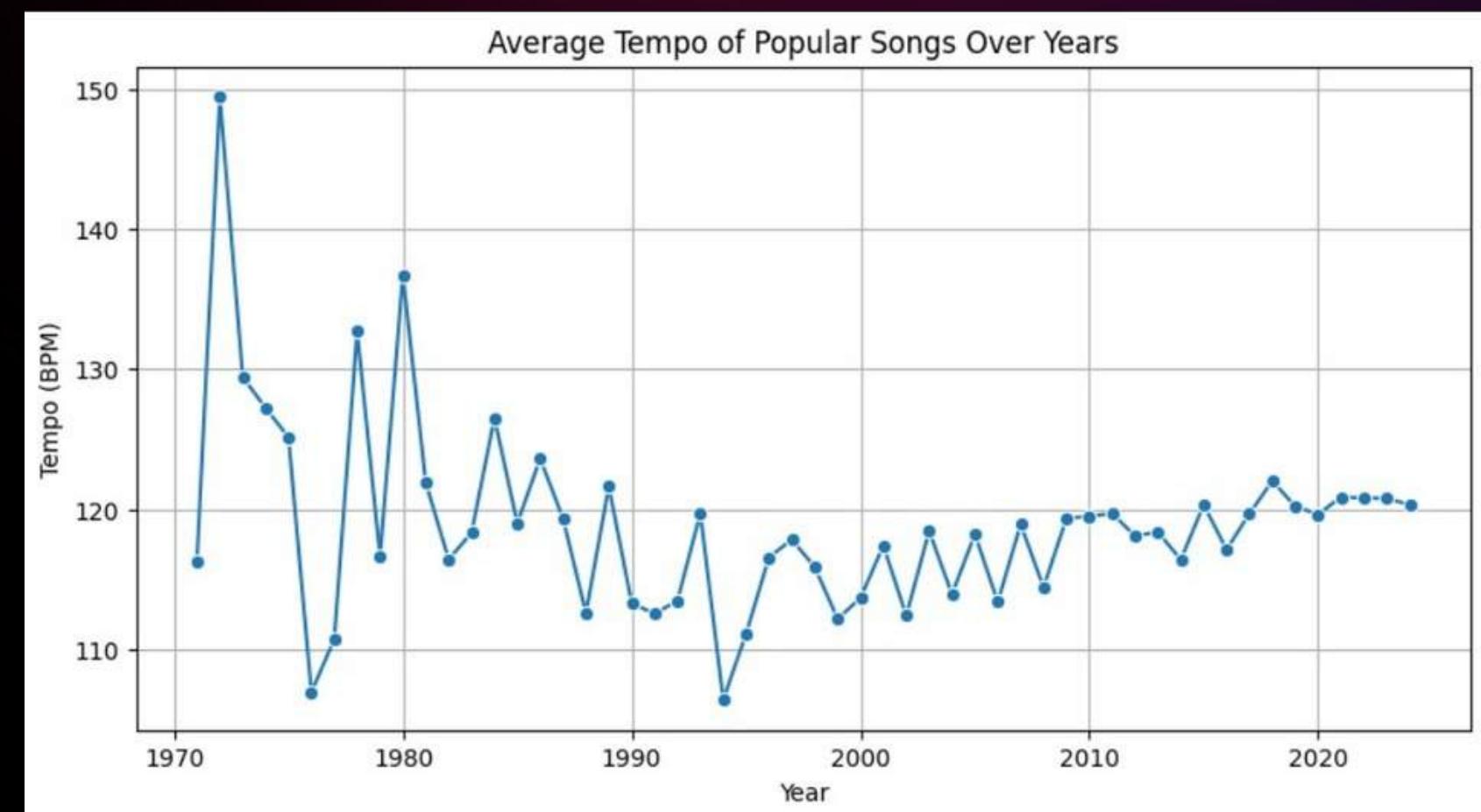
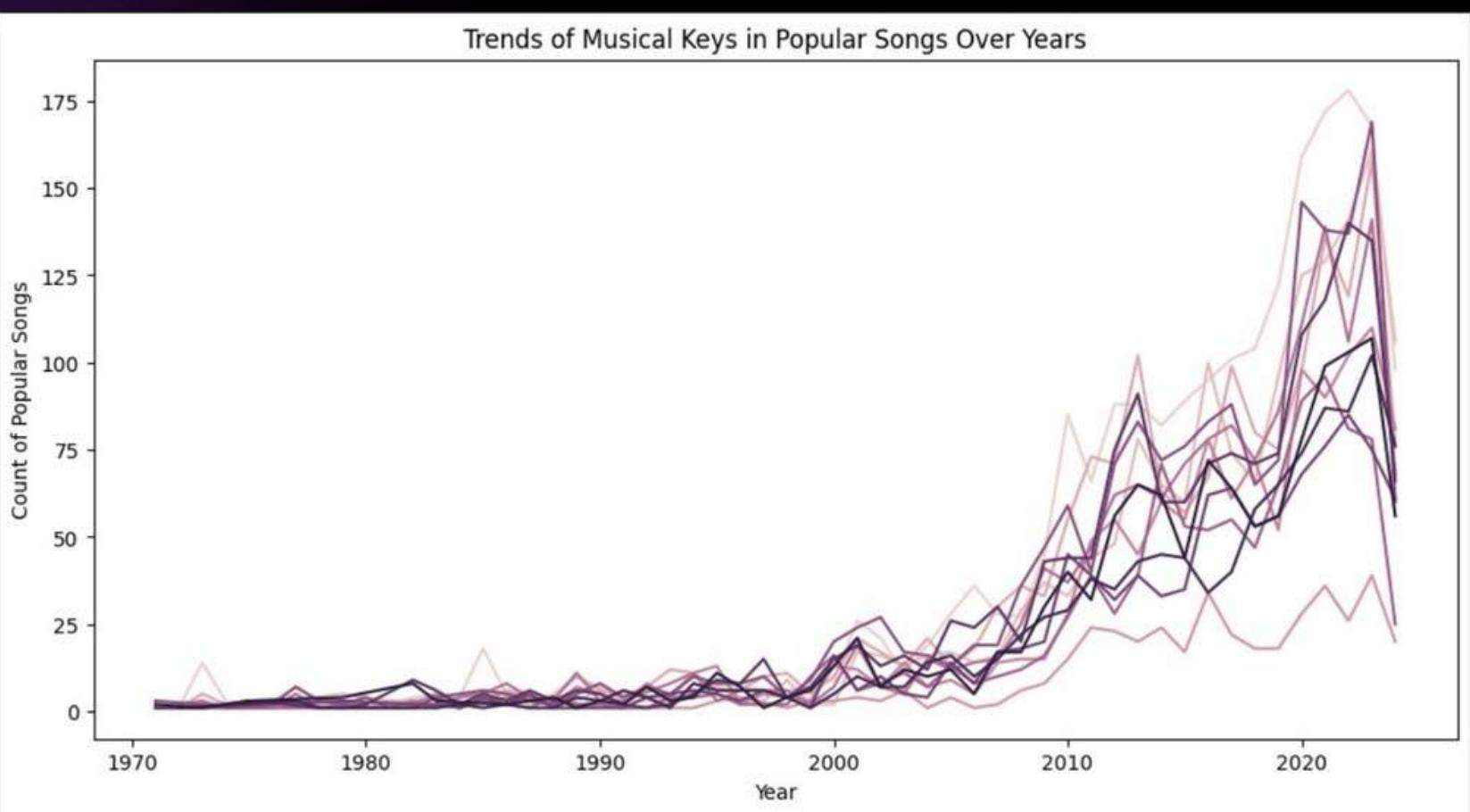
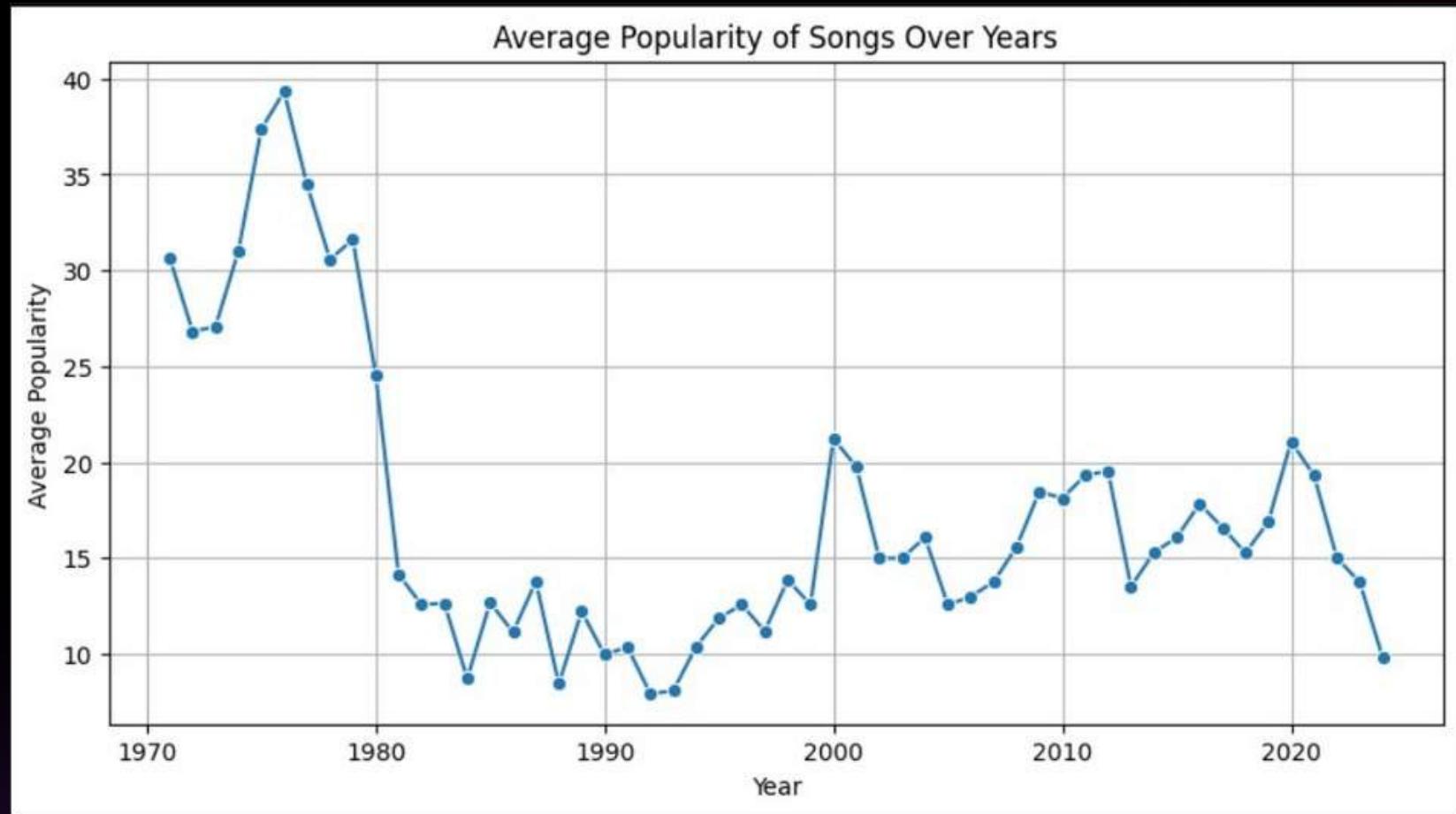
- Popular songs are generally louder:
  - Most highly popular tracks fall between -5 to -10 dB, indicating that louder songs tend to perform better.
- Tempo mostly stays in a moderate range:
  - Popular songs commonly lie between 80-140 BPM, showing that medium tempo is preferred over extremely slow or fast tracks.

Next Slide

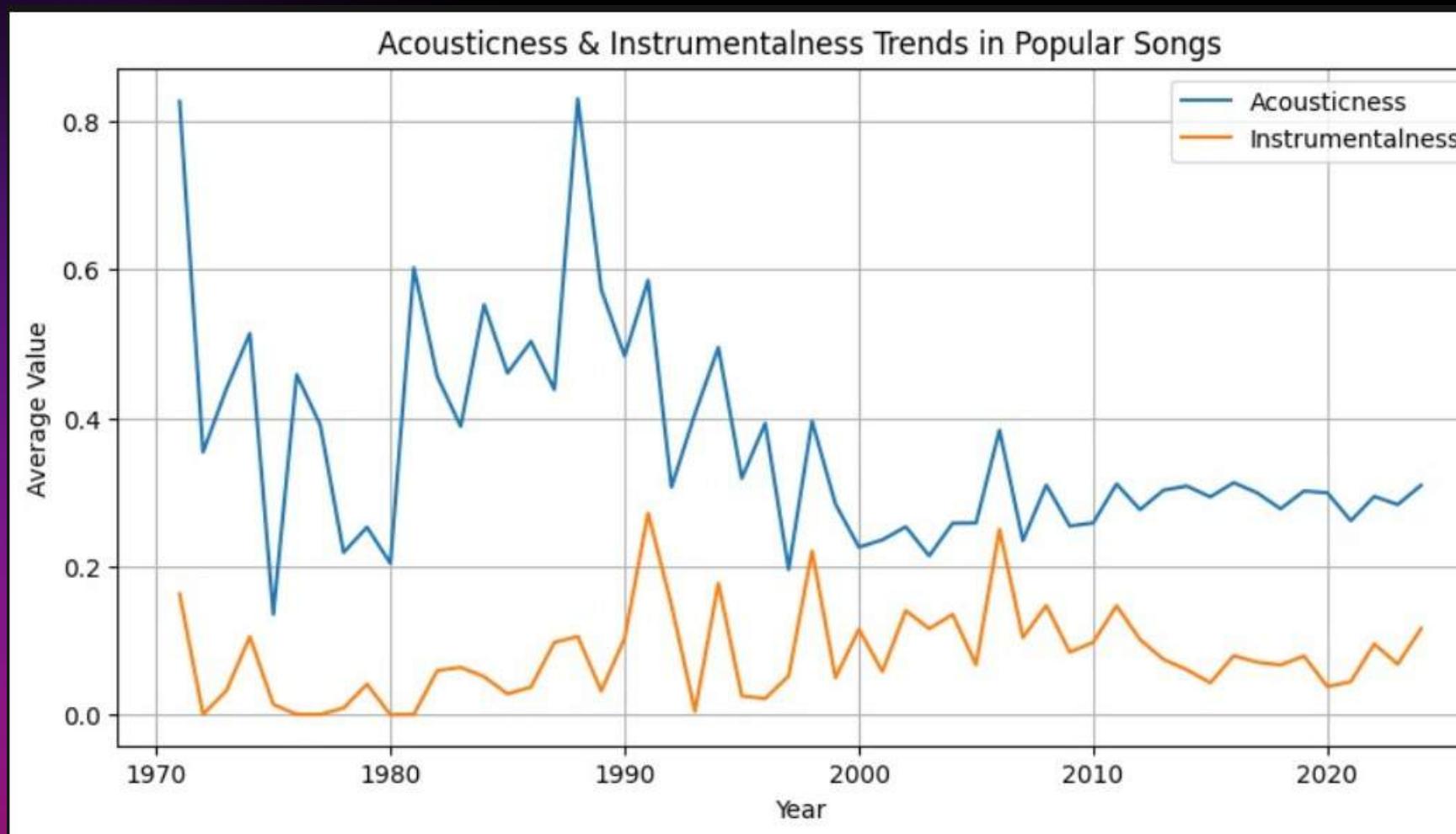
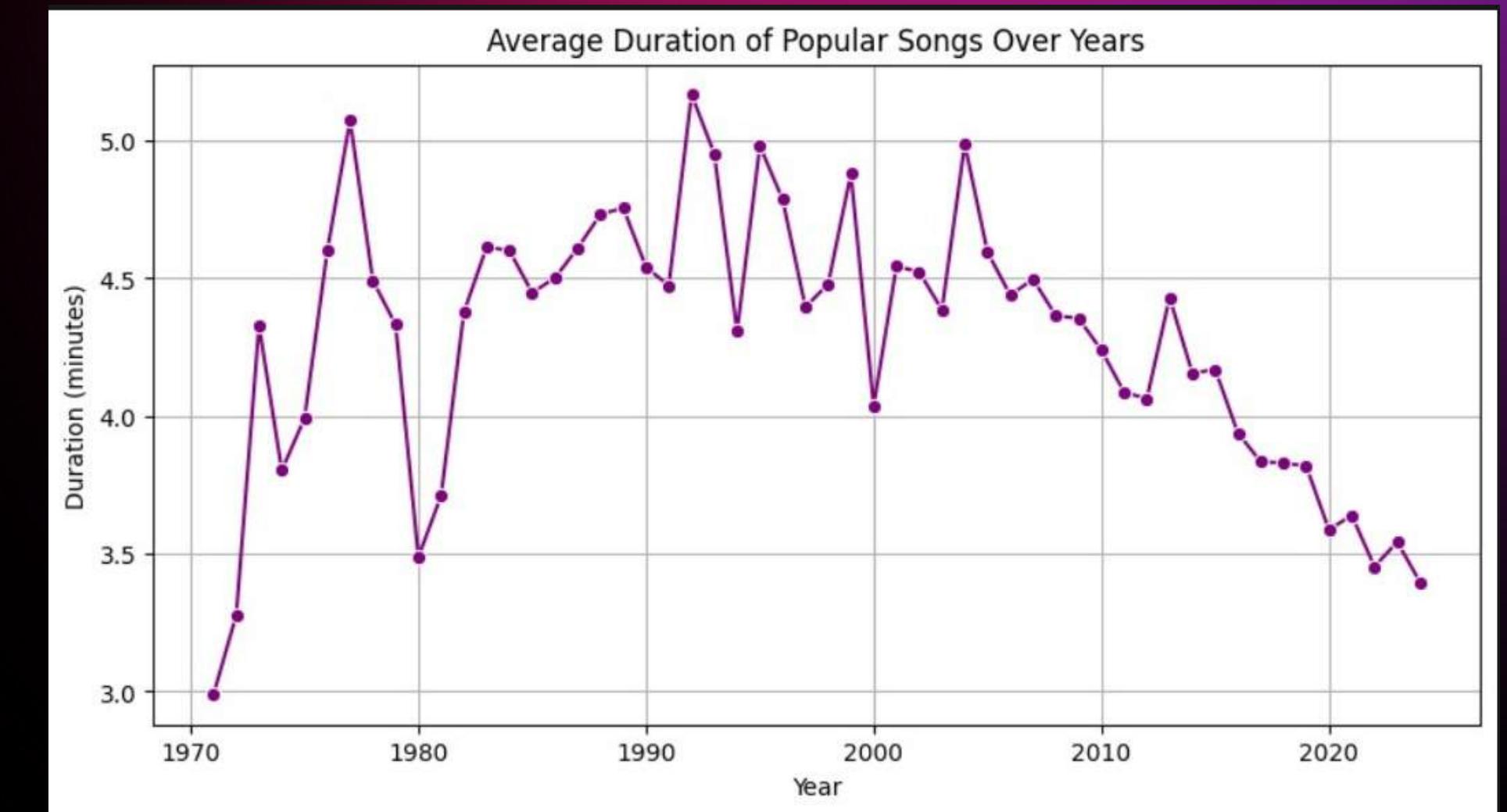


# Time series Analysis:

Next Slide



Next Slide



Next Slide

# Key Insights of Timeseries Analysis

## Acousticness & Instrumentalness Trends

- Acousticness has decreased over time, showing a shift from acoustic-heavy music (1970s–1990s) to more electronic-produced styles in recent decades.
- Instrumentalness remains very low overall, meaning vocals dominate popular songs, with only occasional spikes in instrumental tracks.

## Average Duration of Popular Songs Over Years

- Song duration peaked during the 1980s–2000s (mostly 4.5–5 minutes).
- In recent years, song lengths have dropped to around 3–3.5 minutes, indicating a trend toward shorter, more concise tracks (possibly due to streaming culture).

## Average Tempo of Popular Songs Over Years

- Tempo has gradually decreased since the 1970s, stabilizing around 115–120 BPM in recent decades.
- This suggests that modern popular music prefers mid-tempo beats instead of the fast-paced styles common earlier.

## Trends of Musical Keys Over Years

- Use of musical keys has diversified significantly after 2000, with more songs across different keys becoming popular.
- This suggests greater musical experimentation and variety in modern productions compared to earlier decades.

## Average Popularity of Songs Over Years

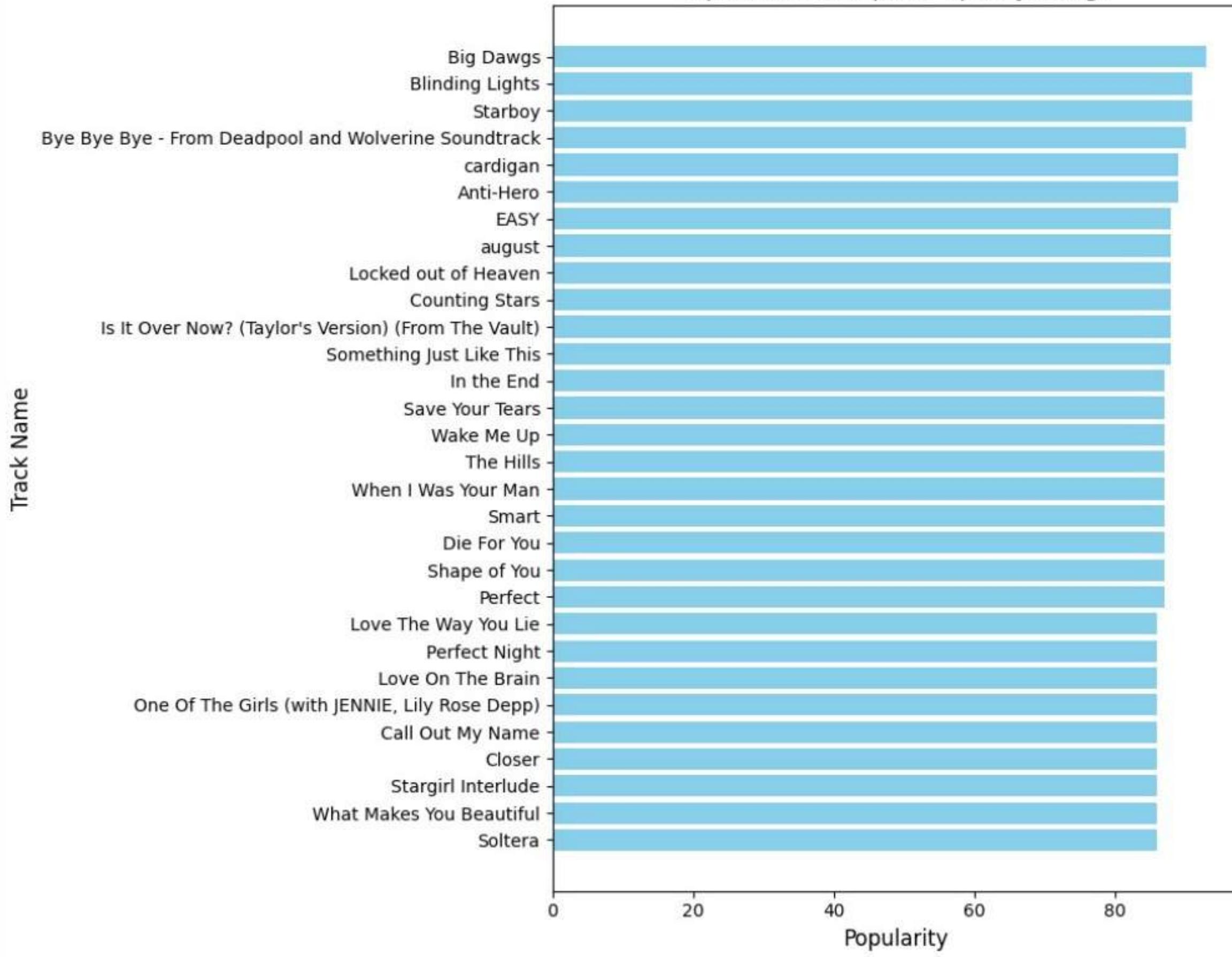
- Popularity peaked in the 1970s–early 1980s, followed by a major decline in the late 1980s–1990s.
- In the 2000s–2020s, popularity shows small fluctuations but stays relatively stable, with no major upward trend.



# Outlier Analysis:

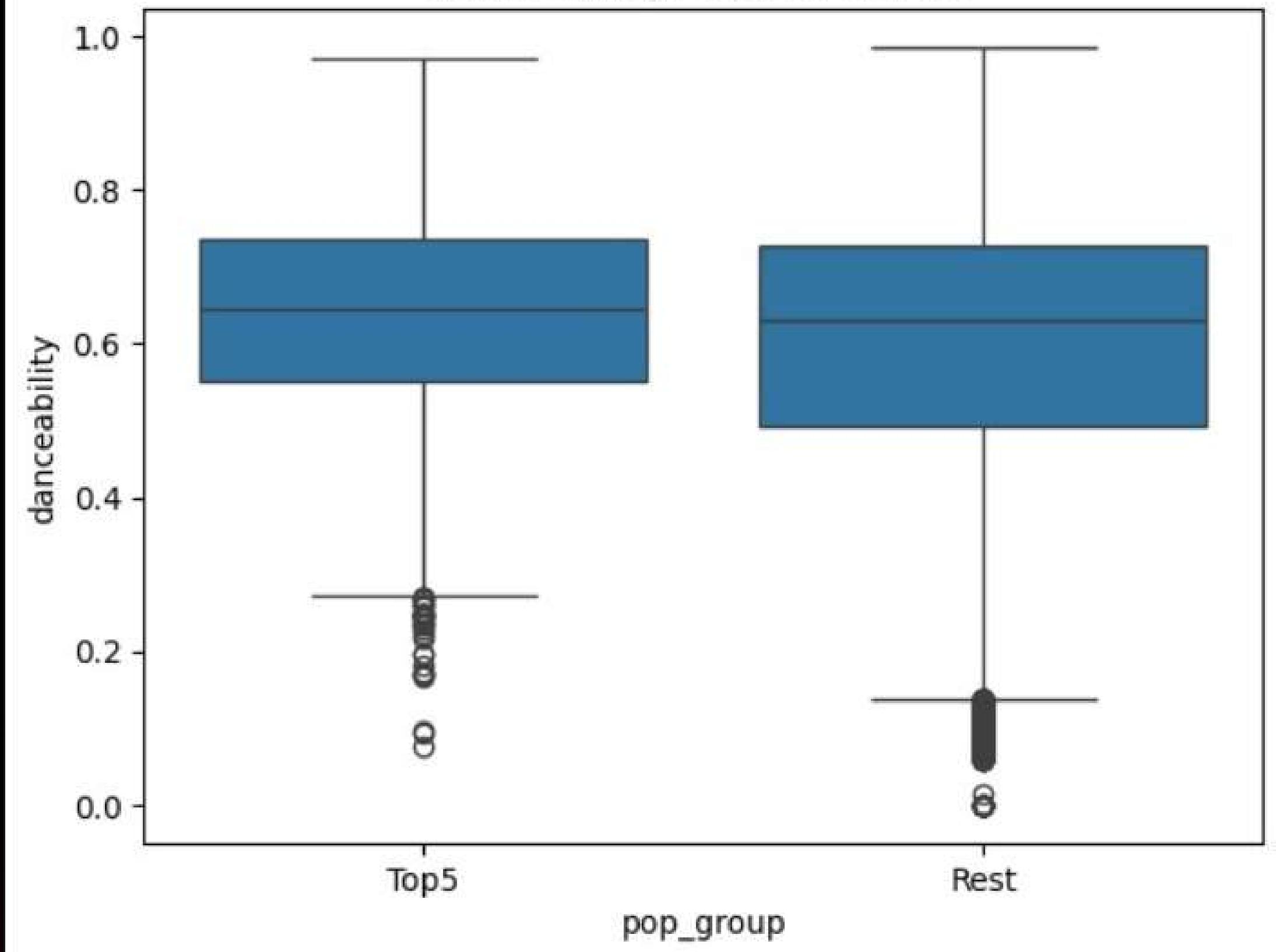
Next Slide

## Top 1% Most Popular Spotify Songs



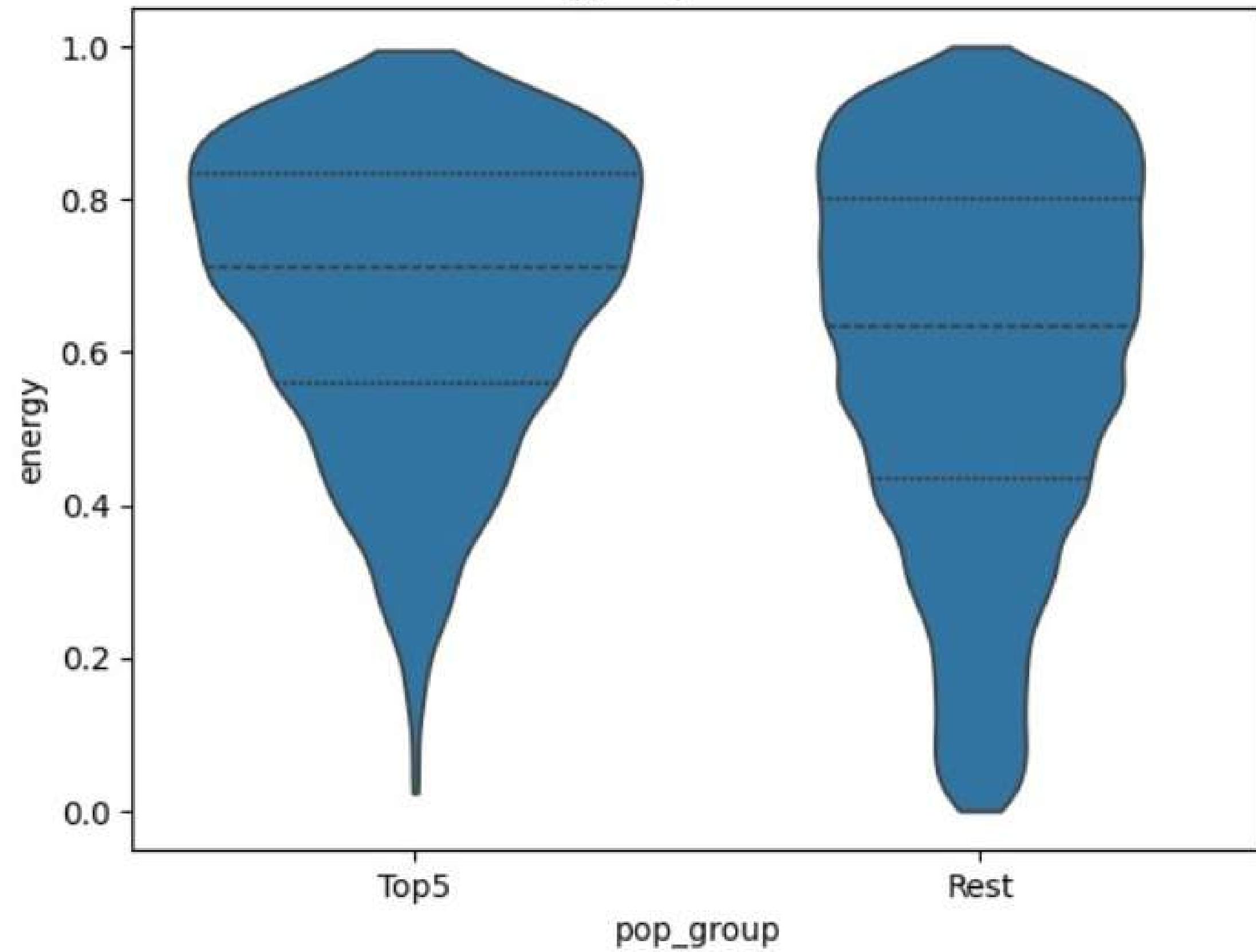
Next Slide

## Danceability: Top 5% vs Rest



Next Slide

### Energy: Top 5% vs Rest



Next Slide

# Key Insights of Outlier Analysis

## 1. Top 1% Most Popular Spotify Songs

- Popular songs consistently have very high popularity scores, showing strong listener engagement across multiple well-known tracks.
- Many of the top songs come from globally recognized artists, suggesting artist popularity strongly influences track popularity.
- 

## 2. Danceability: Top 5% vs Rest

- Both groups show similar median danceability, indicating danceability alone does not strongly separate top songs from others.
- The top 5% has slightly fewer very low-danceability songs, implying extremely low danceability may reduce popularity.
- 

## 3. Energy: Top 5% vs Rest

- The top 5% songs show slightly higher concentration of high energy values, meaning high-impact songs tend to perform better.
- However, the distributions still overlap, showing energy is important but not the only factor driving popularity.

Next Slide

# 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).
- 
- 

Next Slide

# THANK YOU!

