

# EXPLORATORY DATA ANALYSIS (SPOTIFY DATASET)

PRESENTED BY

SHANTANU SOREN

# PROBLEM STATEMENT

CONSIDER YOURSELF AS A MUSIC DIRECTOR OR MIXING ENGINEER WORKING IN A FAST-PACED, DATA-DRIVEN MUSIC INDUSTRY WHERE MILLIONS OF TRACKS COMPETE DAILY FOR LISTENER ATTENTION.

YOUR GOAL IS TO OPTIMIZE NEW SONG FOR MAXIMUM POPULARITY BUT THE CHALLENGE LIES IN UNDERSTANDING WHAT EXACTLY MAKES A TRACK STAND OUT.



The dataset is a collection of Spotify tracks with various features used in this analysis.

## Column Name

## Description

• popularity	A measure of how popular a track is, ranging from 0 to 100.
• acousticness	A confidence measure indicating whether track is acoustic, ranging from -1.0 to 1.0.
• danceability	A measure of how suitable a track is for dancing, ranging from -1.0 to 1.0.
• duration_ms	The duration of the track in milliseconds.
• energy	A perceptual measure of intensity and activity, ranging from -1.0 to 1.0.
• instrumentalness	Predicts whether a track contains no vocal content, ranging from -1.0 to 1.0.
• key	The key the track is in, represented as an integer (e.g., 0 = C, 1 = C#, etc.).
• liveness	Detects the presence of an audience in the recording, ranging from -1.0 to 1.0.
• loudness	The overall loudness of a track in decibels (dB).
• mode	Indicates the modality (major or minor) of a track (0 for minor, 1 for major).
• speechiness	A measure detecting the presence of spoken words in a track.
• tempo	The overall estimated tempo of a track in beats per minute (BPM).
• time_signature	An estimated overall time signature of a track.
• valence	A measure from -1.0 to 1.0 describing the musical positiveness conveyed by a track.

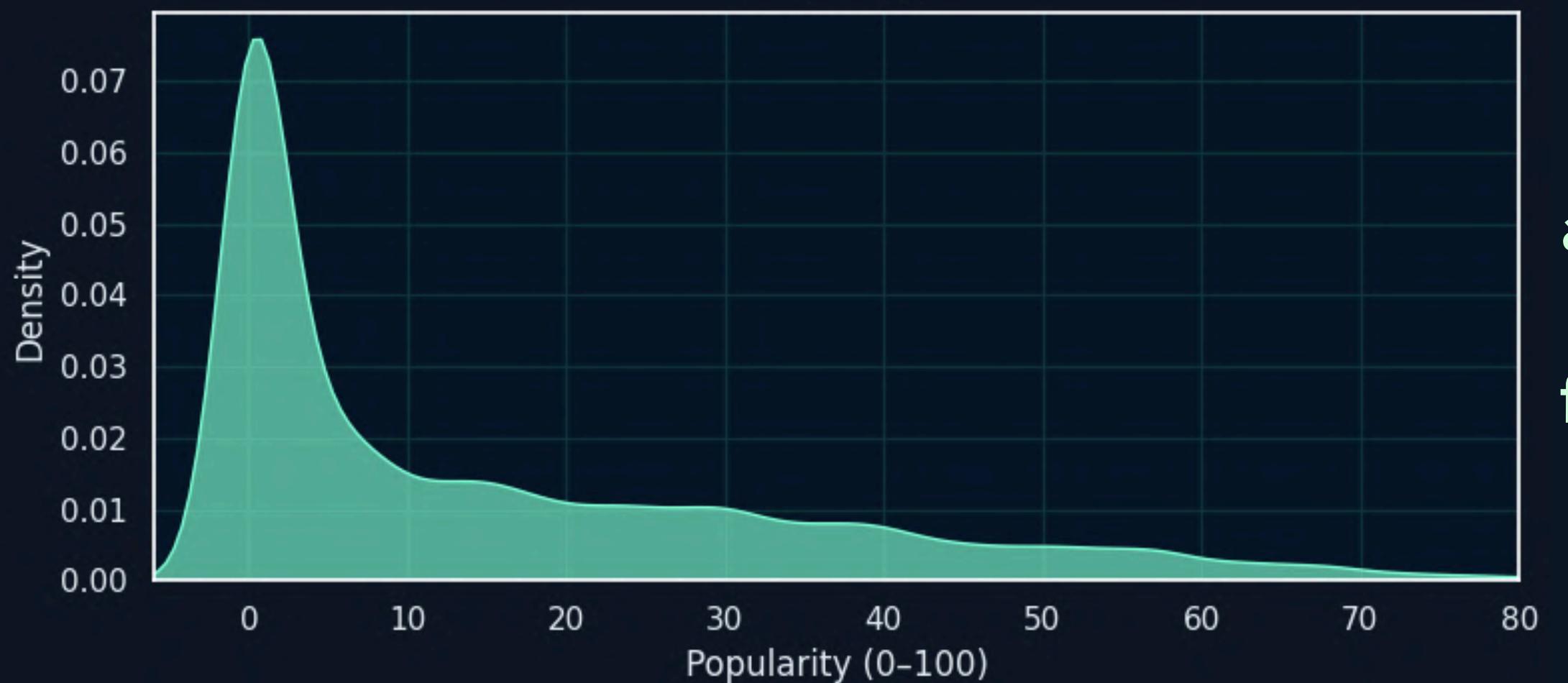
# Spotify Track Analysis



## CONTENTS

1. UNIVARIATE ANALYSIS
2. BIVARIATE ANALYSIS
3. MULTIVARIATE ANALYSIS
4. TIME-SERIES ANALYSIS
5. INSIGHTS/ANALYSIS

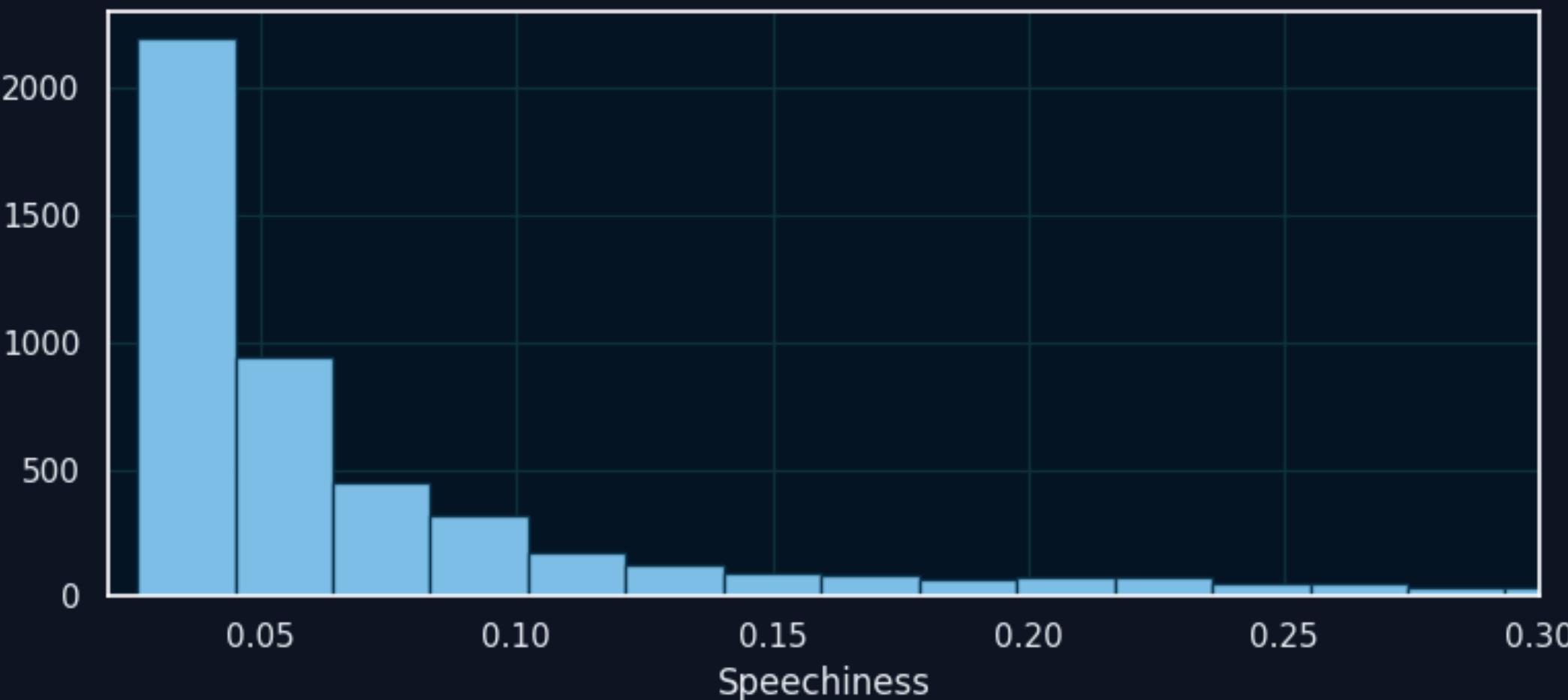
Popularity



Most songs are moderately popular or low-rated, with only a small fraction achieving “hit” status while the majority receive limited attention. Popularity follows a power-law distribution pattern, put effort on high-potential songs that show strong energy or catchiness.

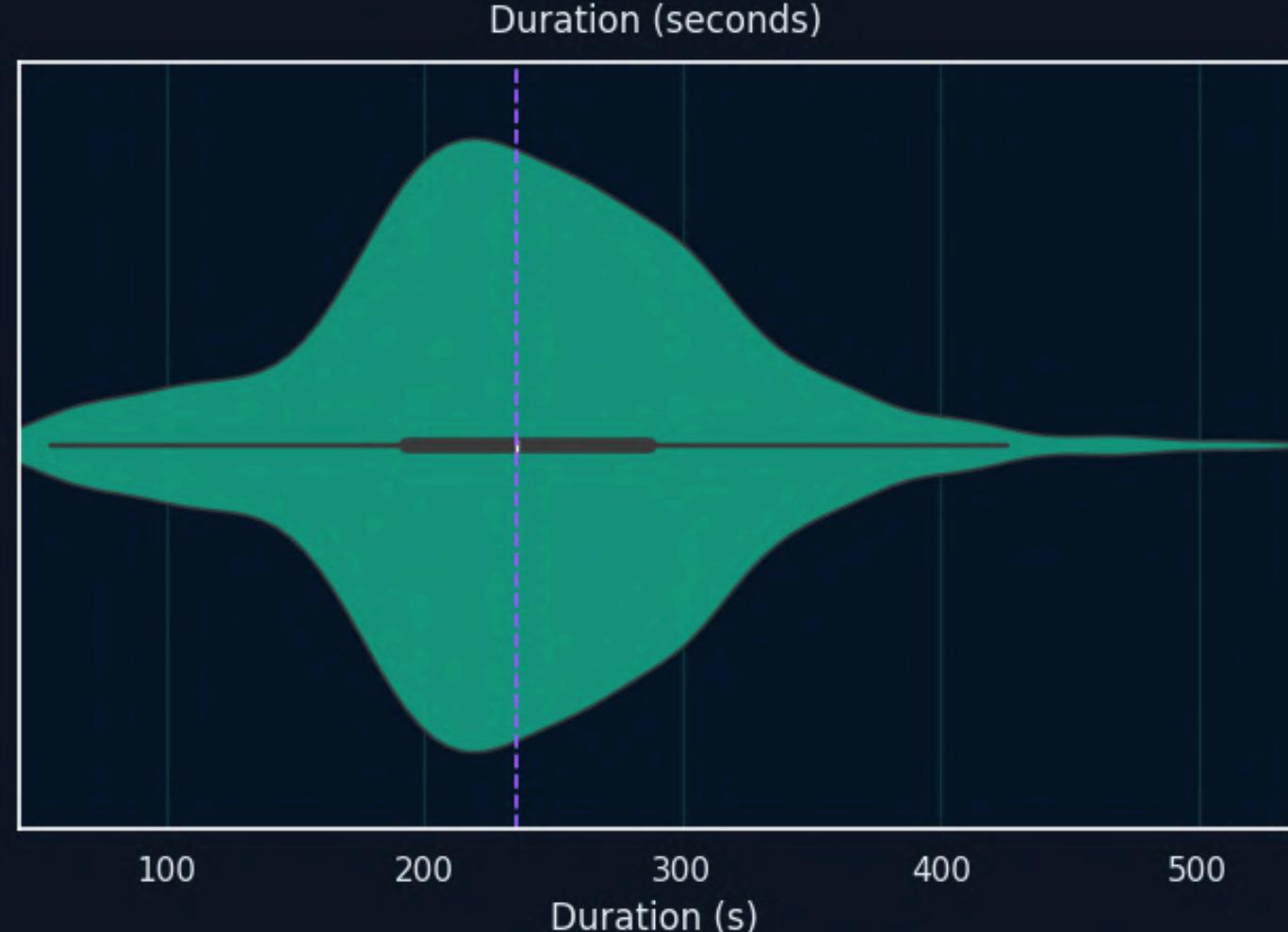
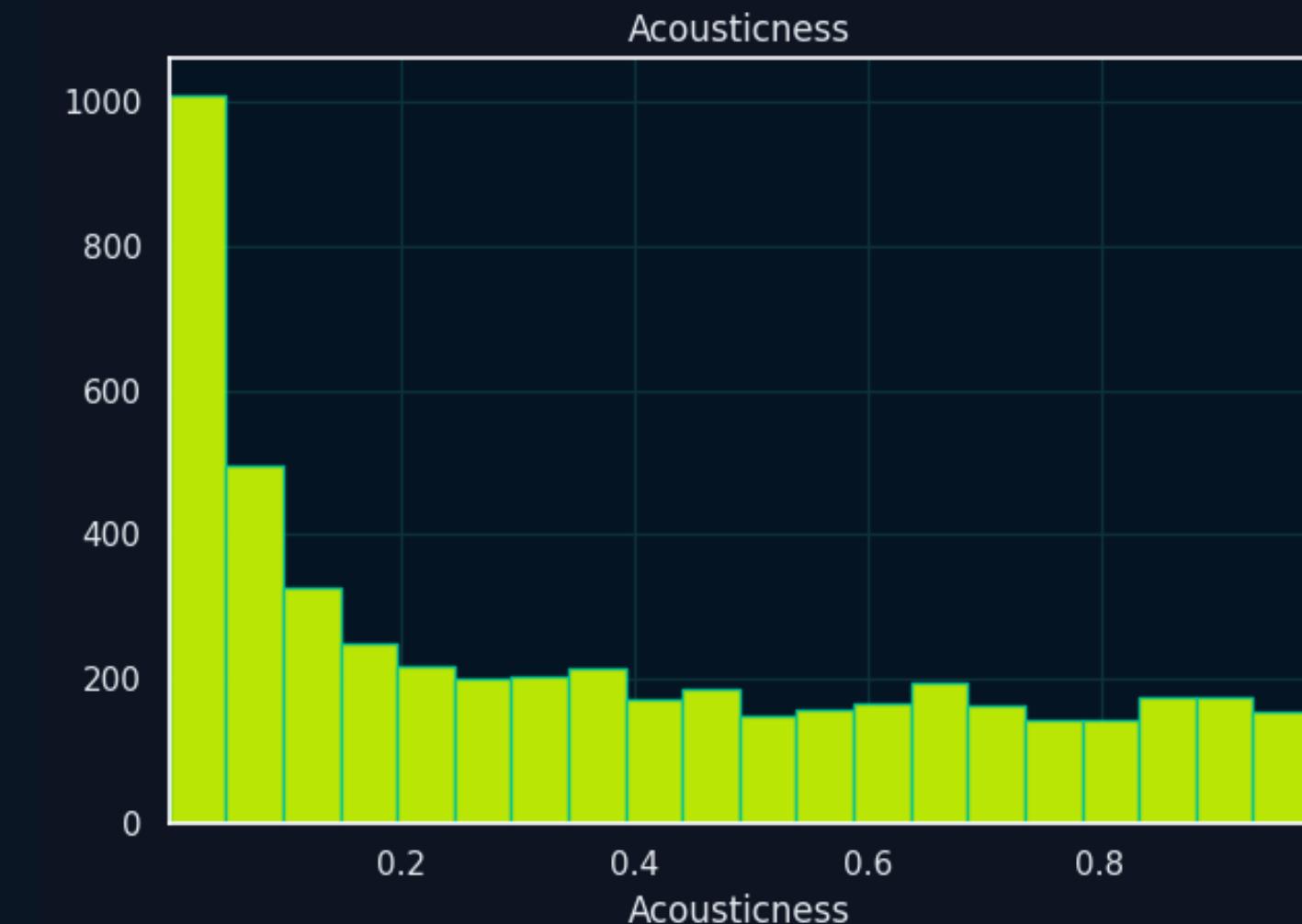
Speechiness is very low for most tracks indicating that melodic singing dominates while Spoken or rap-heavy content exists but is limited to specific genres. Keep vocals melodic for pop, or mix spoken phrasing with melodic hooks in rap/urban for commercial crossover.

Speechiness Distribution



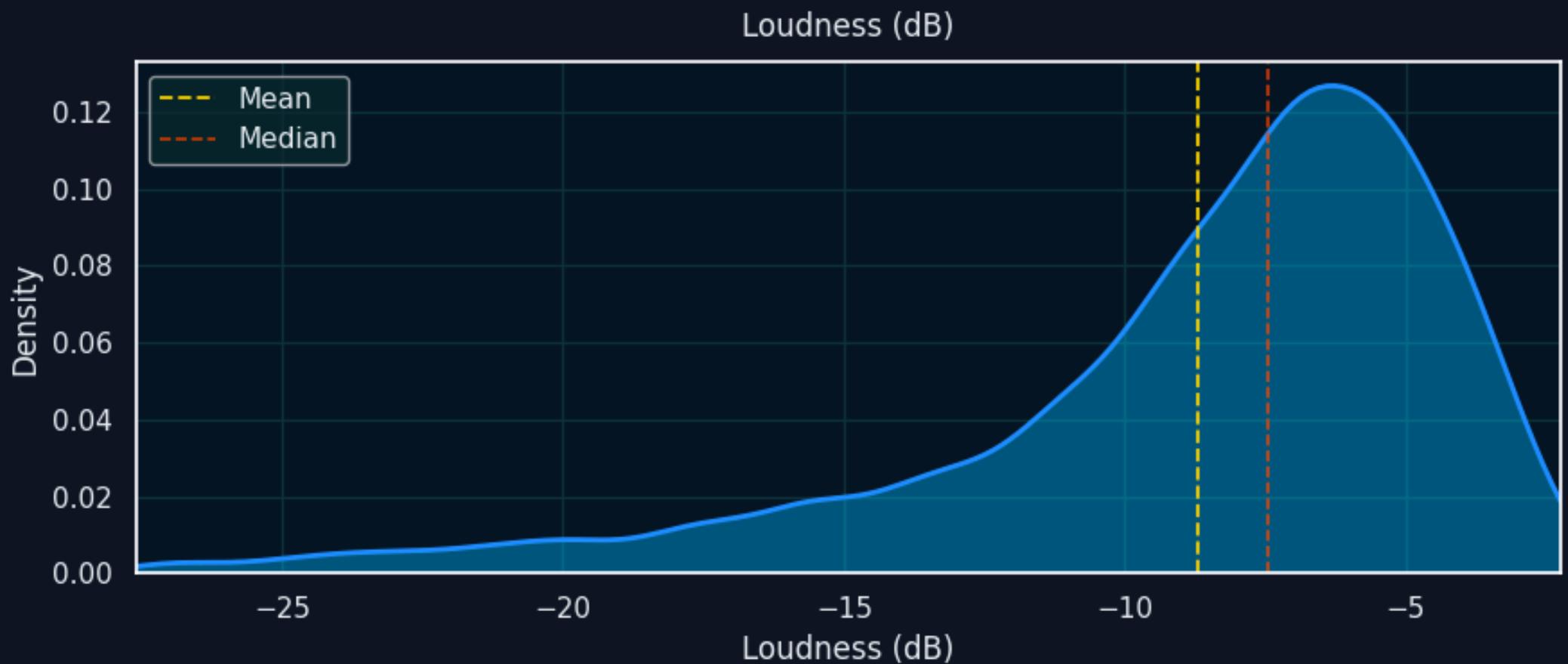
Most tracks lean toward low-to-moderate acousticness, implying dominance of electronic, hybrid, or studio processed production. With relatively few fully acoustic tracks, most songs are digitally produced with subtle organic layers.

Using acousticness as supportive element within modern mixes (as offsets or layers) is recommended.



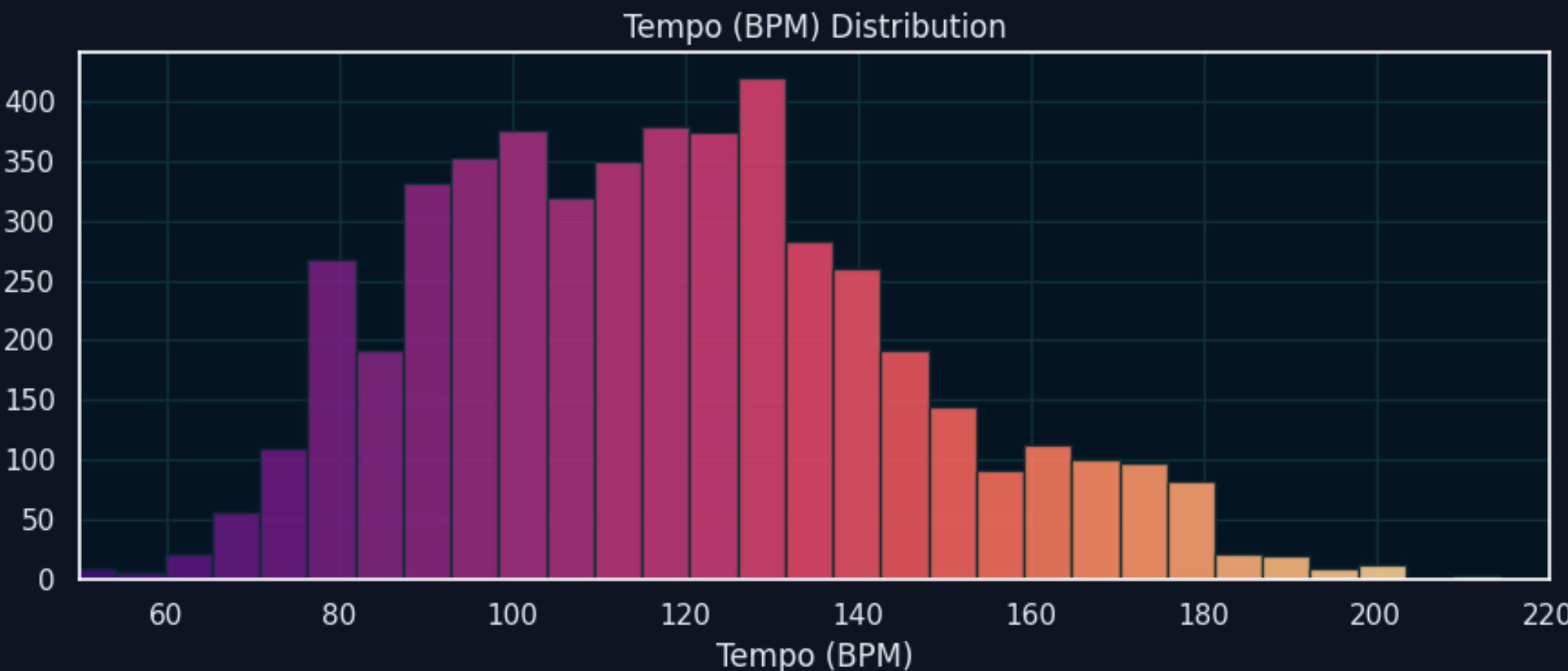
The average song duration is around 4 minutes, with most tracks ranging between 3–4.5 minutes, aligning with current mainstream pop standards and current streaming optimizations. There are very few very short or overly long songs.

Keep total length around 3.5 to 4.0 mins, hit the first hook within 30–40 seconds & avoid long intros or outros.

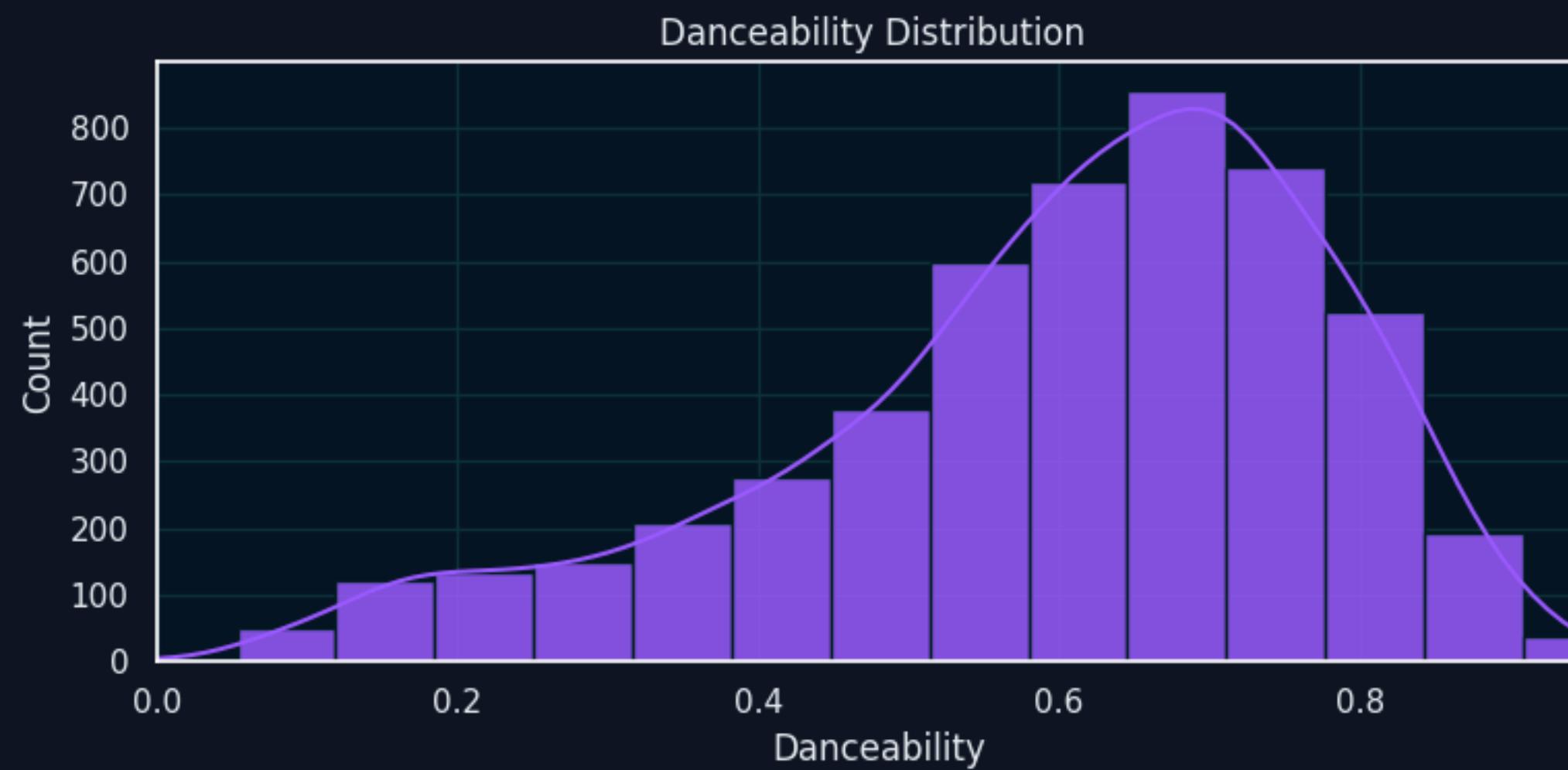
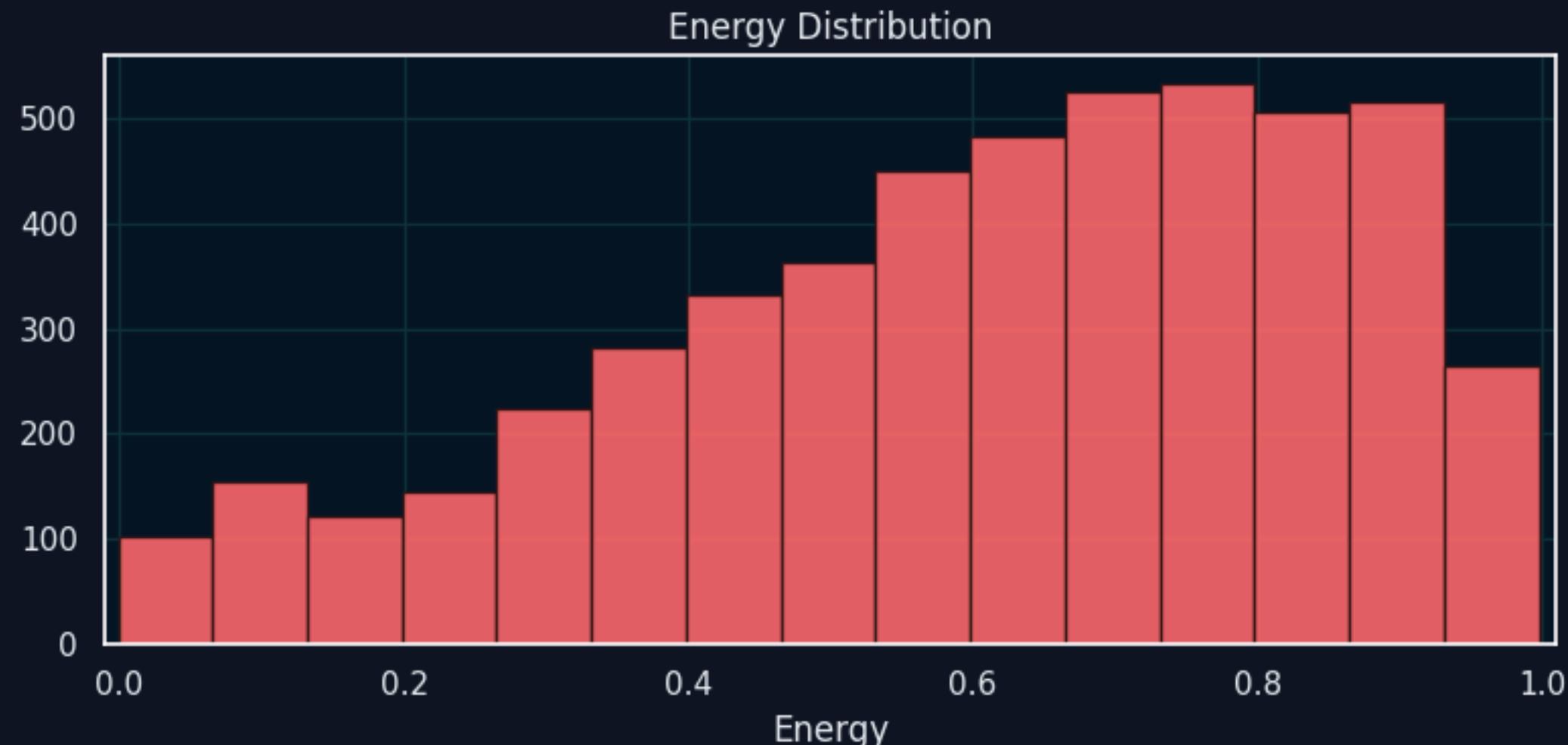


Most tracks maintain loudness between -6 to -10 dB, which is moderately loud with a tight dynamic range, consistent with streaming-mastered pop and EDM. Target -9 to -8 LUFS on average, keeping true peak around < -1 dBTP.

Songs are generally mid-tempo, clustering around 110–125 BPM, showing wide but controlled spread with fewer extremely fast or slow tracks. For mass appeal, aiming for 118–128 BPM (pop/dance) or 70–90 BPM (hip-hop) is recommended.

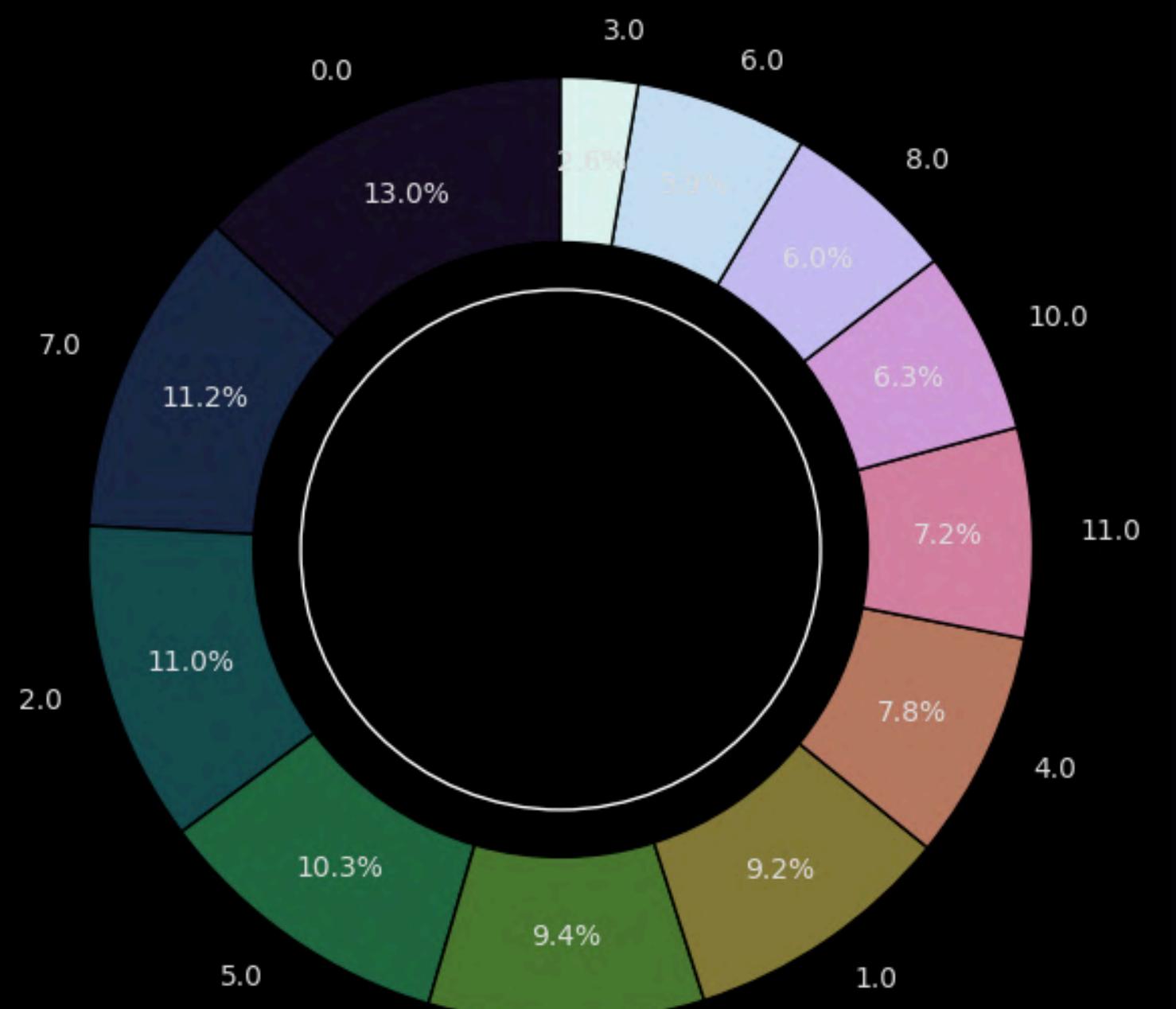


Most songs are highly danceable, with a fair mix of chill and mid-tempo tracks. The majority fall between 0.5–0.75 on the danceability scale. Focus on rhythmic structure and clean low-end mixing to enhance groove.



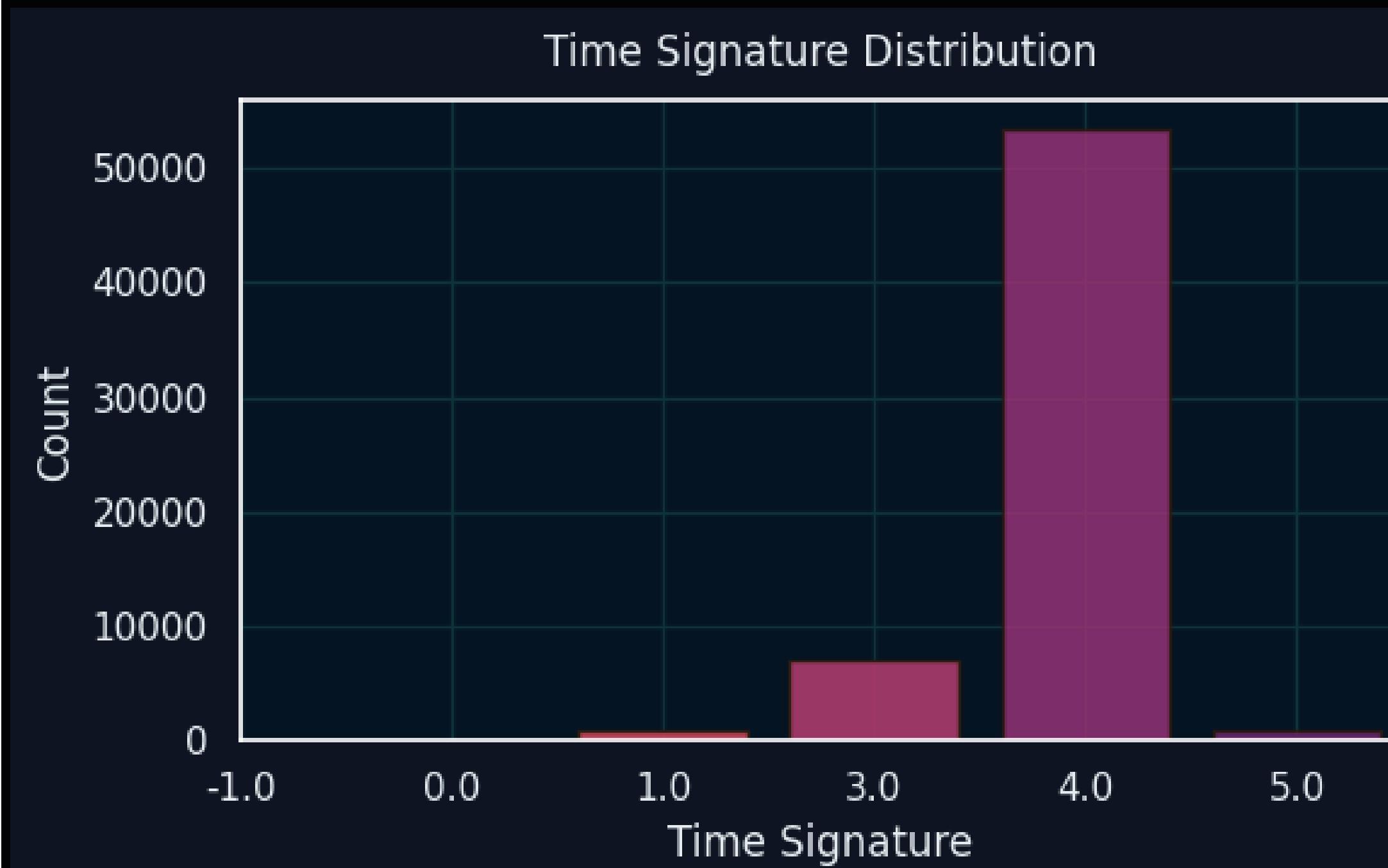
Most songs show medium-to-high energy around 0.6–0.8, implying lively and vibrant production trends. Try building dynamic progressions that lift energy in pre-chorus, maximize at chorus, drop post-chorus for contrast.

Top 12 Musical Keys



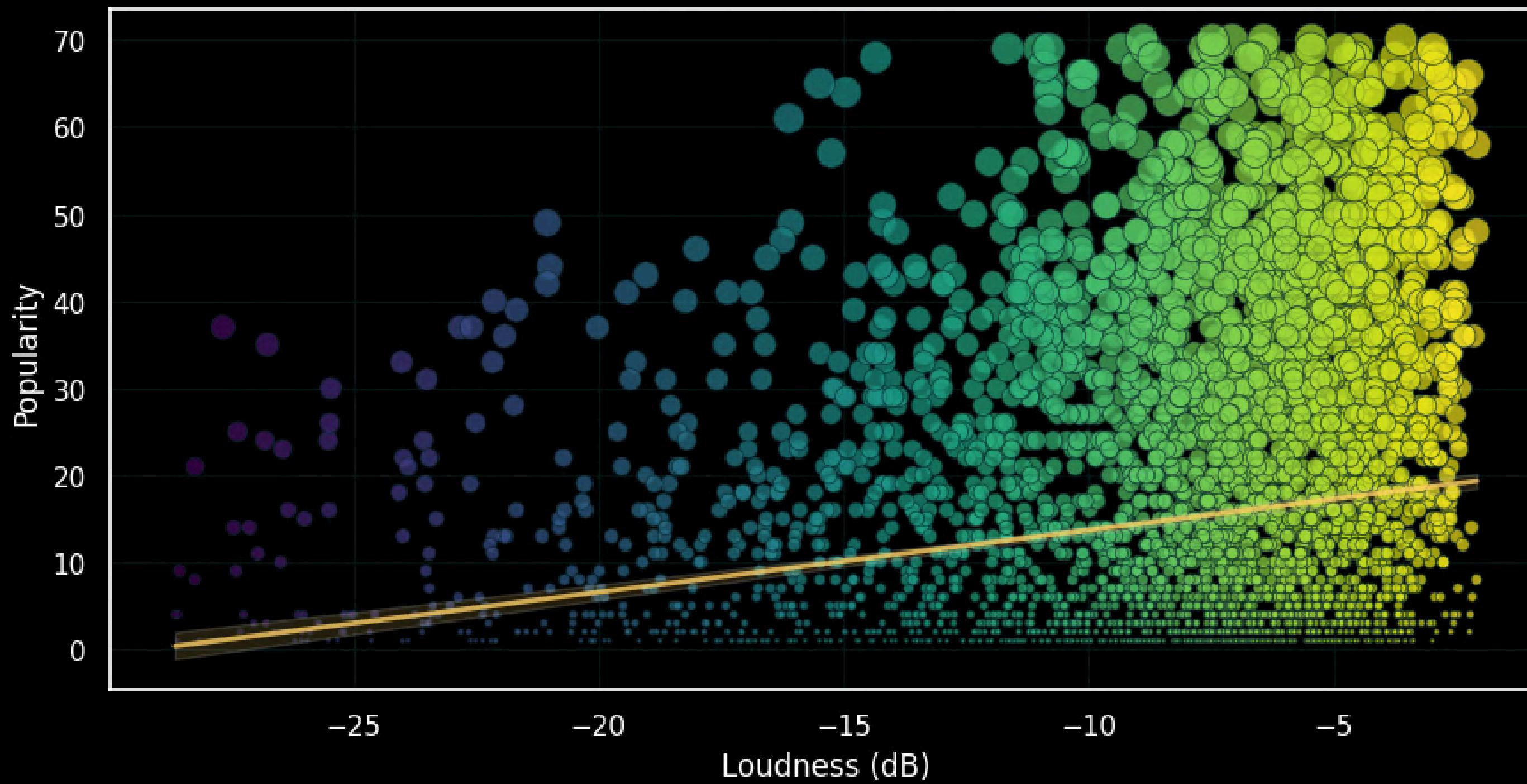
The C, G & A minor are the most common keys. Roughly 60% of songs use major keys and rest use minor keys, showing a balance of uplifting and emotional tones. Use major keys for broader appeal; incorporate minor passages to add depth and mood contrast.

4/4 is overwhelmingly dominant, reaffirming its role as the universal standard for rhythmic familiarity in pop, rock and dance. The most common time signature is 4/4 ( $\approx 86\%$ ), with 3/4 and 5/4 being rare exceptions found in certain niche or experimental music. Use 4/4 for mainstream reach; non-4/4 works best for stylistic experimentation.

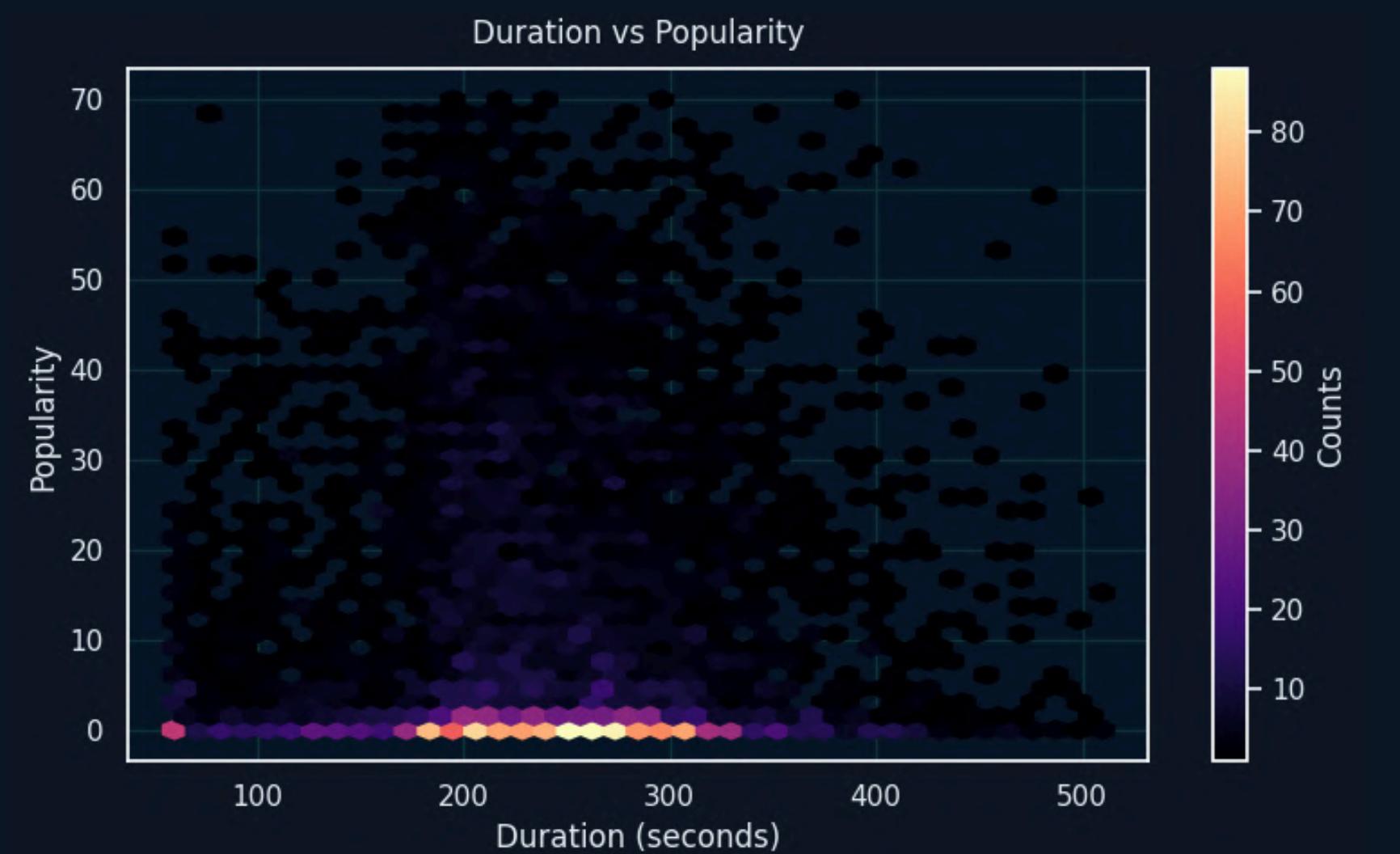


# BIVARIATE ANALYSIS

Loudness vs Popularity



Slight positive association found, modern masters are consistently loud, but loudness alone doesn't make a song popular. There is no significant impact beyond meeting modern norms. Once a song is competitively loud, mix quality and energy for success. Master around  $-9$  LUFS and focus on clarity and transient impact rather than relying on loudness alone.

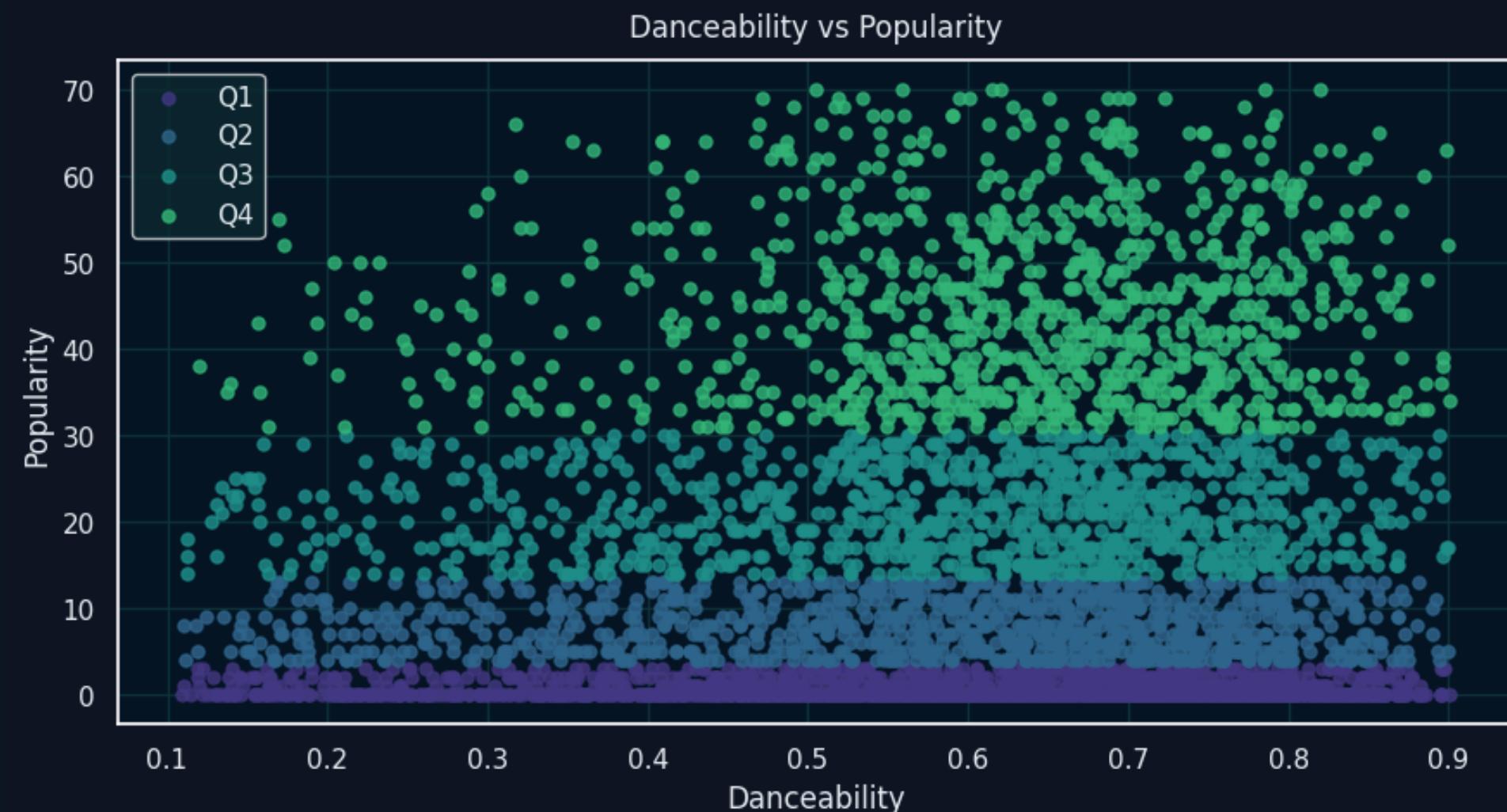


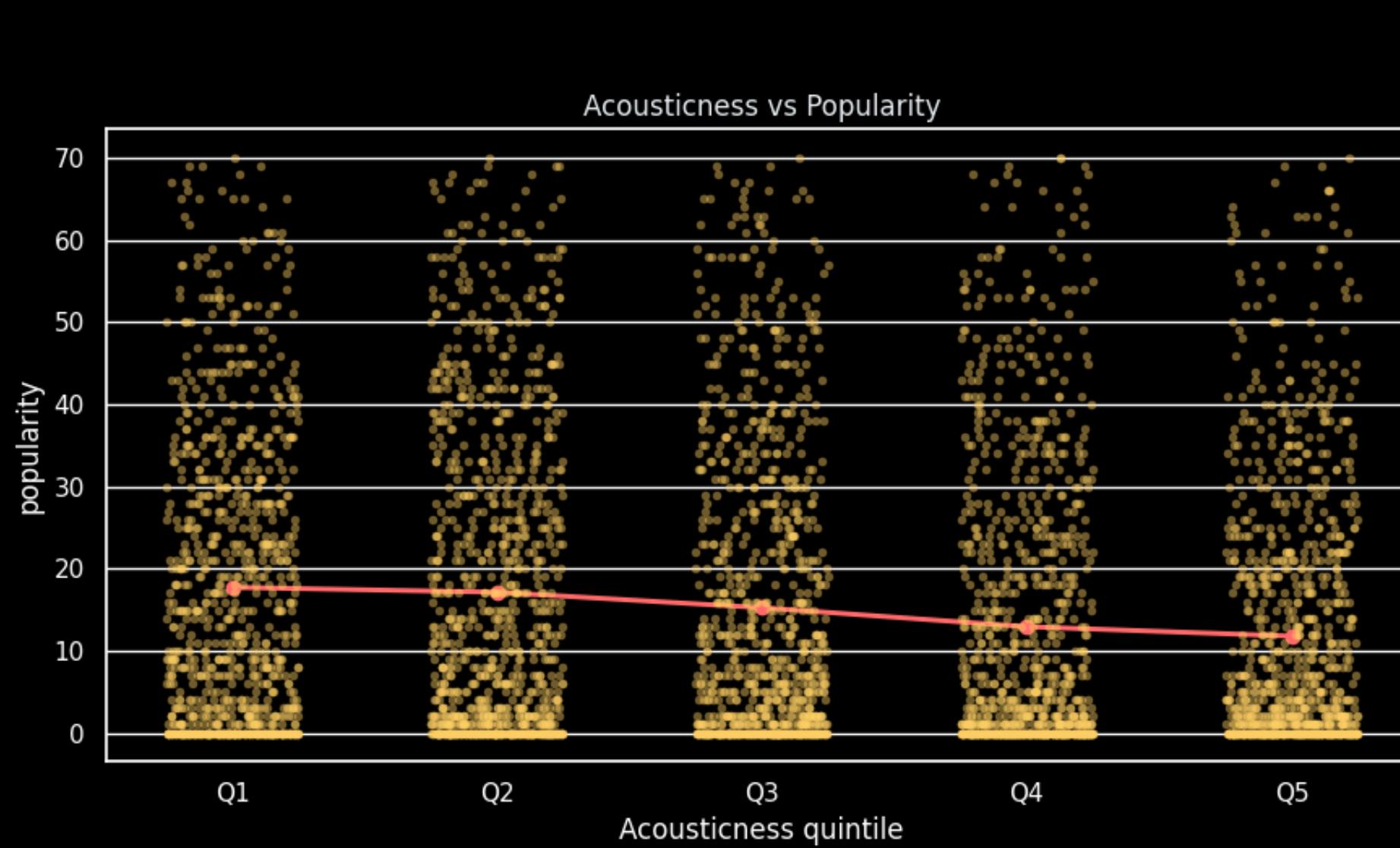
**Song length does not significantly affect popularity. Tracks both short and long can succeed if the content and hook are strong. There is no meaningful correlation between duration and popularity. Shorter songs aren't necessarily more popular, though the trend toward shorter average lengths likely improves replay rates.**

**Focus on song structure and hook timing, not total runtime; engagement quality matters more than length.**

**Higher danceability slightly increases the chance of popularity, but it's not the sole driver. Other production and vocal factors play stronger roles. There is a weak positive relationship tracks with moderate-to-high danceability tend to perform a bit better on average, but the effect is not dominant.**

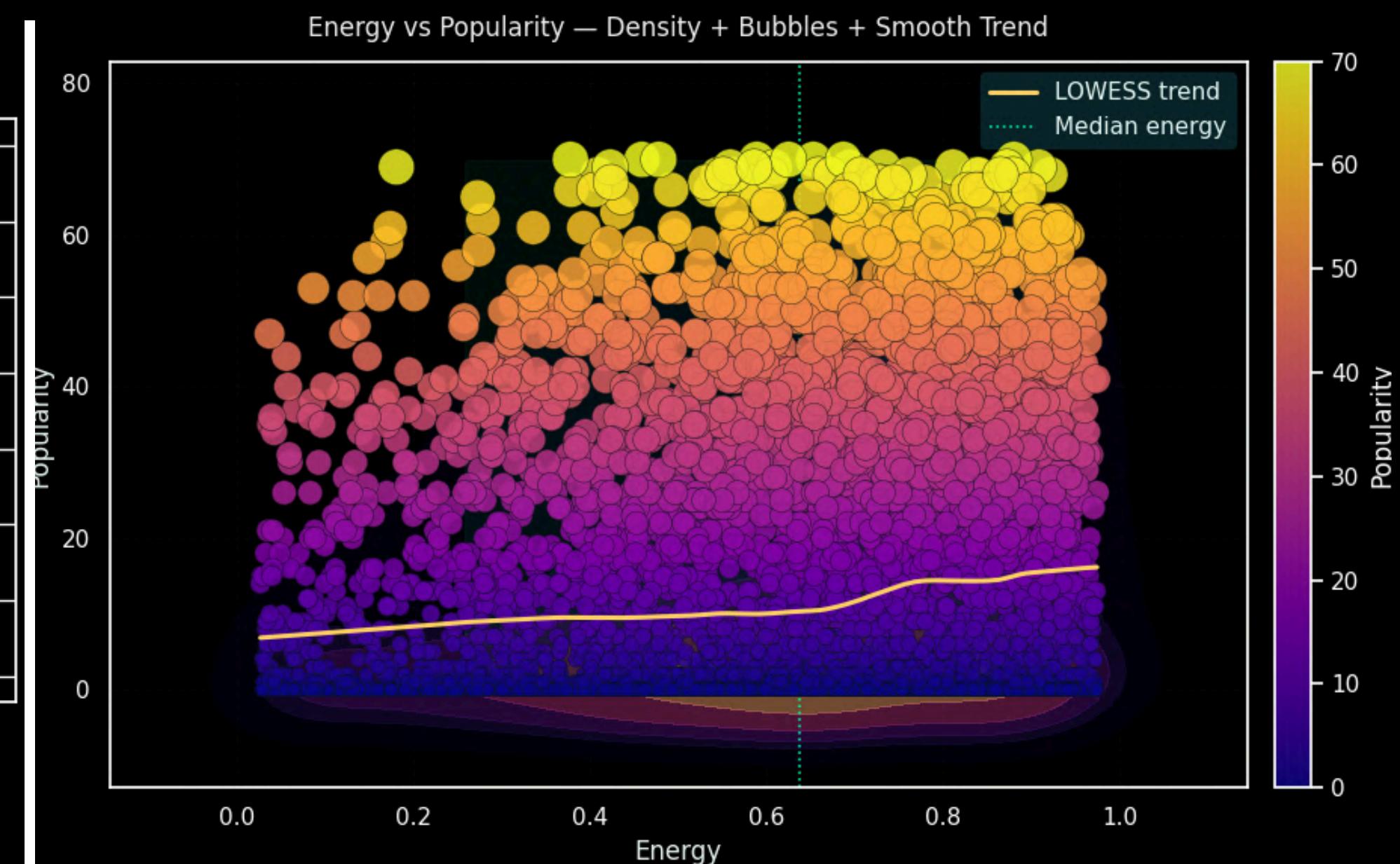
**Ensure rhythmic clarity and groove; it supports engagement but doesn't guarantee popularity.**





**More electronic and hybrid tracks tend to be slightly more popular than purely acoustic ones. There is a light negative relationship; organic tracks underperform compared to digitally produced or hybrid songs.**

**For mainstream appeal, combine natural tone with modern digital polish rather than relying purely on acoustic production.**

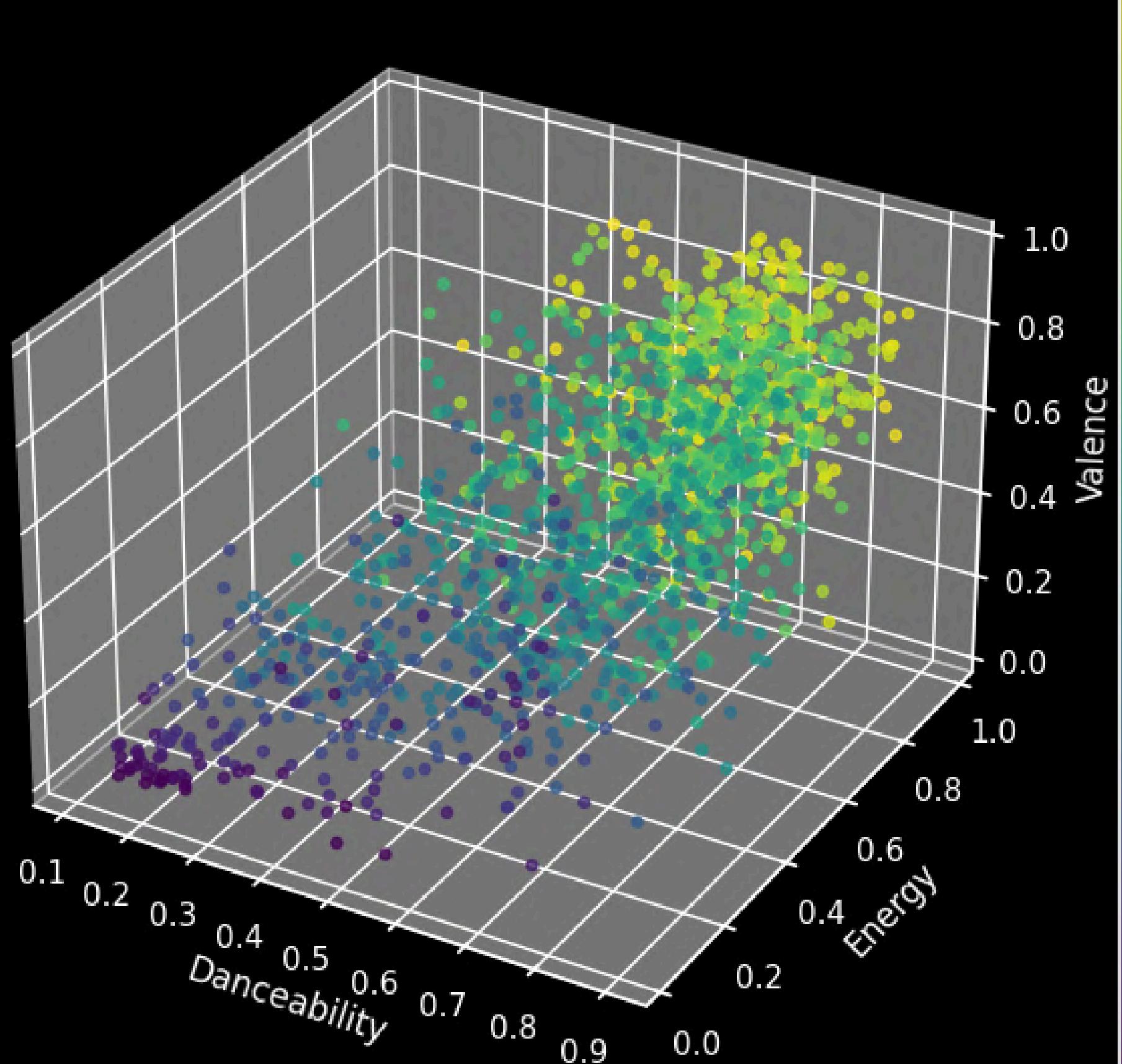


**Energy shows the strongest positive link with popularity among audio features. Energetic, vibrant tracks draw more listeners and repeat plays. High-energy songs are generally more popular than low-energy ones, indicating listener preference for dynamic and intense productions.**

**Prioritize energetic arrangements and bright, full-range mixes; energy intensity is a core popularity predictor.**

# MULTIVARIATE ANALYSIS

3D Feature Space – Danceability, Energy, Valence



Highly popular tracks are moderately high in all three dimensions groovy (danceable), energetic and emotionally positive. The combination forms a “sweet spot” between rhythmic drive and emotional brightness, typical of mainstream pop and dance crossover hits.

The most frequent combination among top-quartile songs is:

- Danceability 0.65–0.75

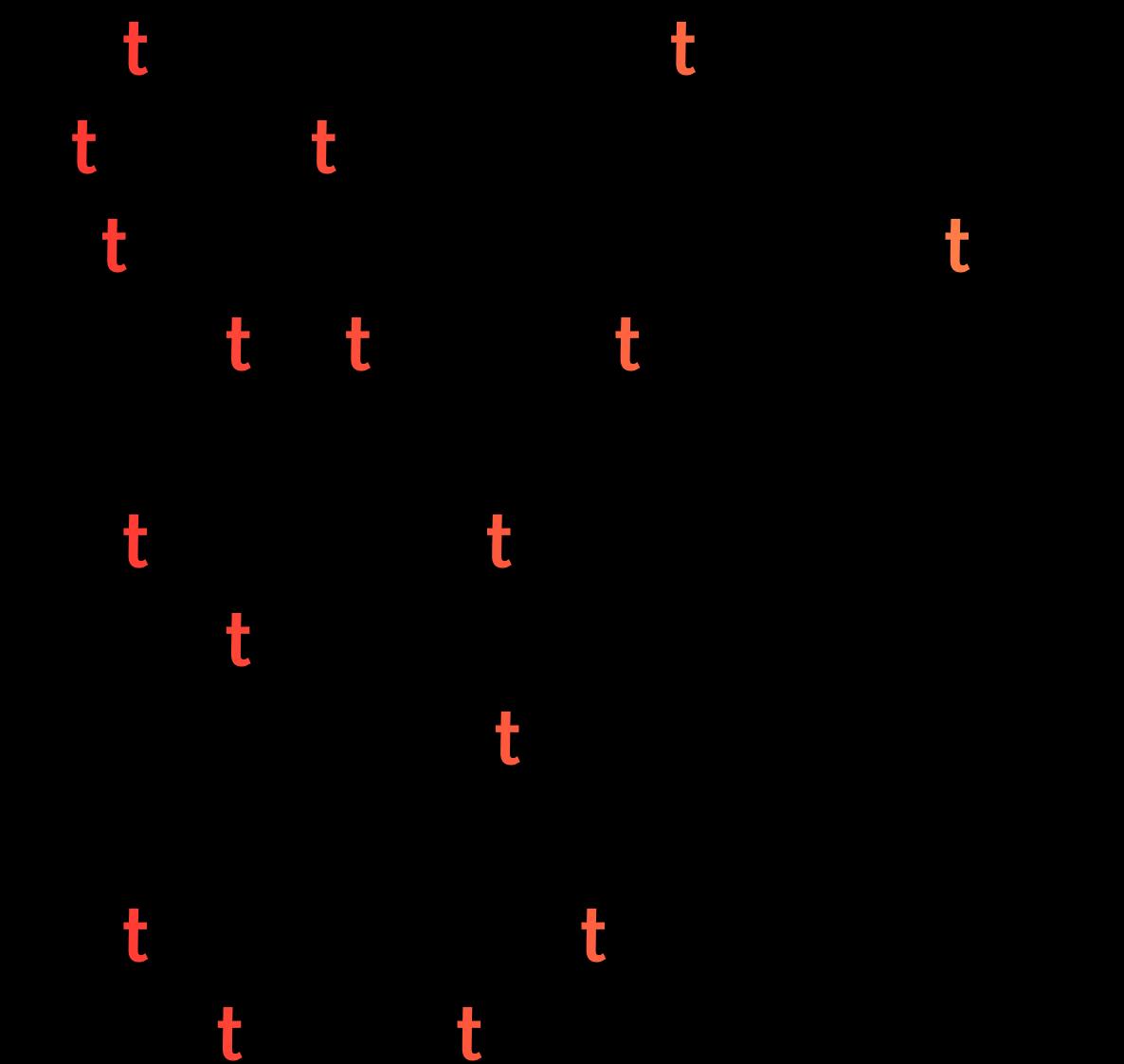
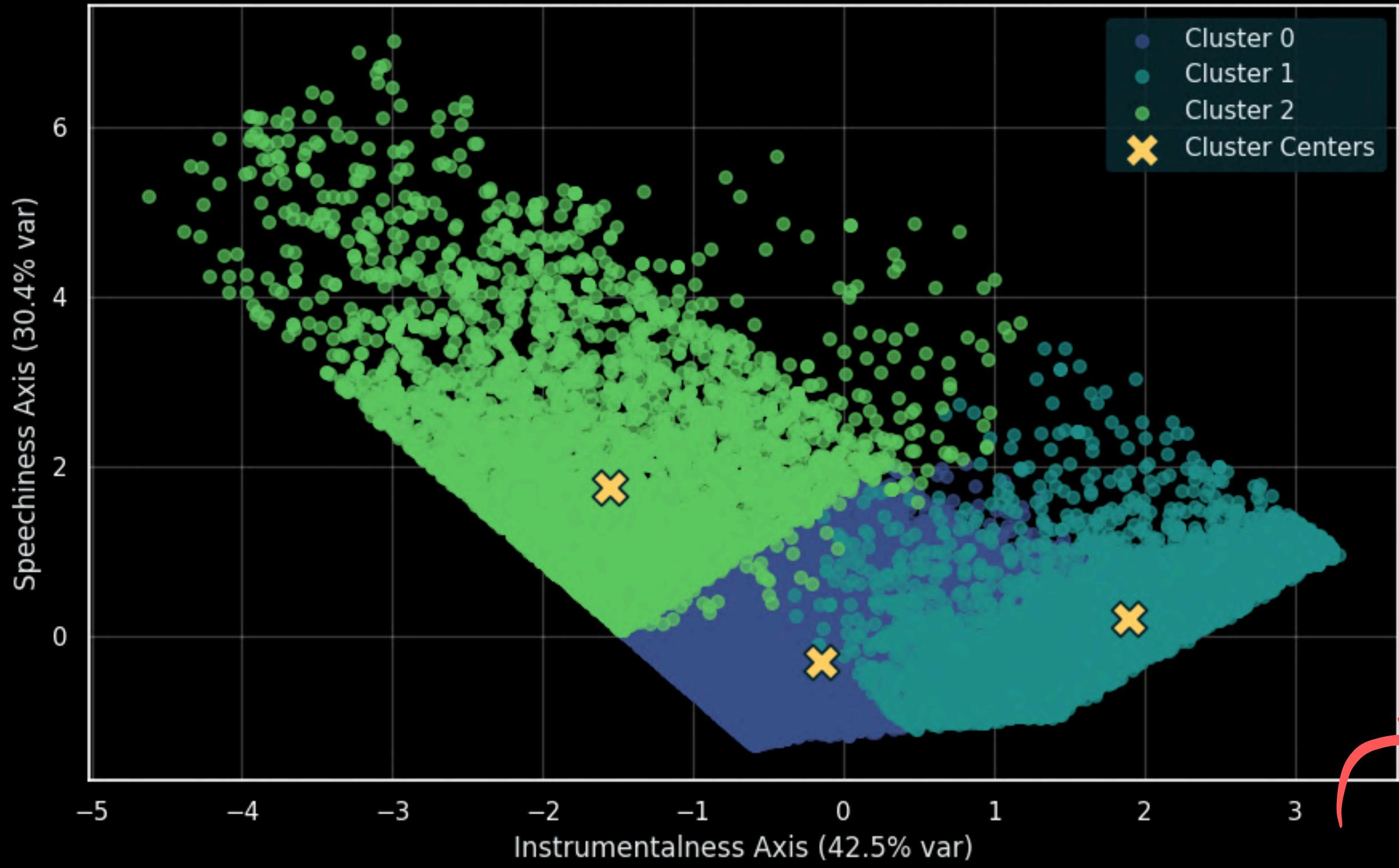
- Energy 0.70–0.80

- Valence 0.55–0.65

This triad represents vibrant, upbeat music with controlled intensity.

Hits thrive where rhythm (danceability), drive (energy), and positivity (valence) converge near the upper-mid range which is cheerful but polished, not frantic or chaotic.

### Clusters — Acousticness, Instrumentalness & Speechiness

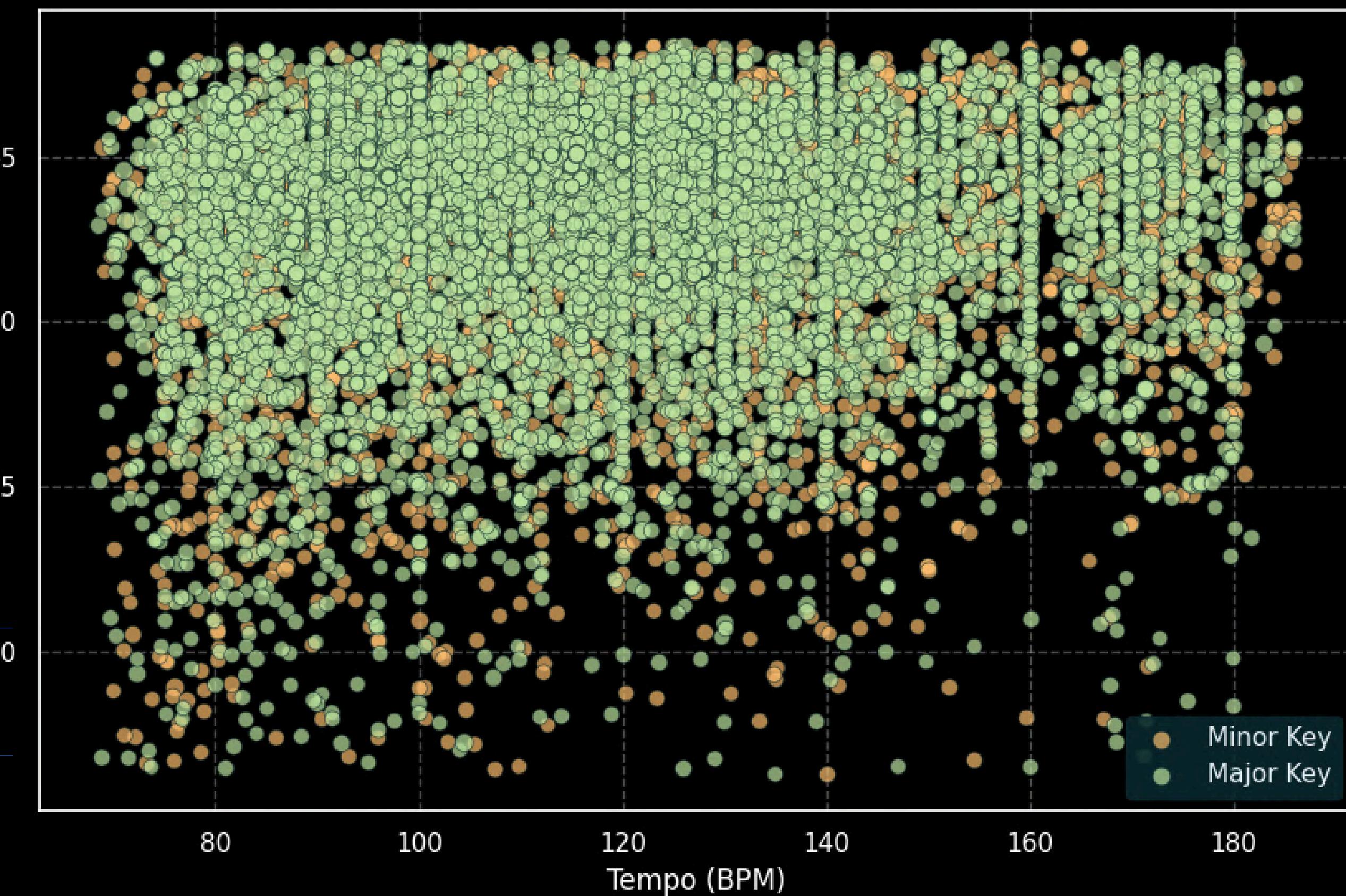


Most popular tracks fall into the Pop/EDM and Urban hybrid clusters, characterized by digital sound design and clear vocals. Purely acoustic tracks remain niche.

**Popular songs maintain loud, balanced mastering, mid-tempo rhythm, and a slight bias toward major mode for brightness. This combination supports streaming loudness standards and universal rhythmic familiarity.**

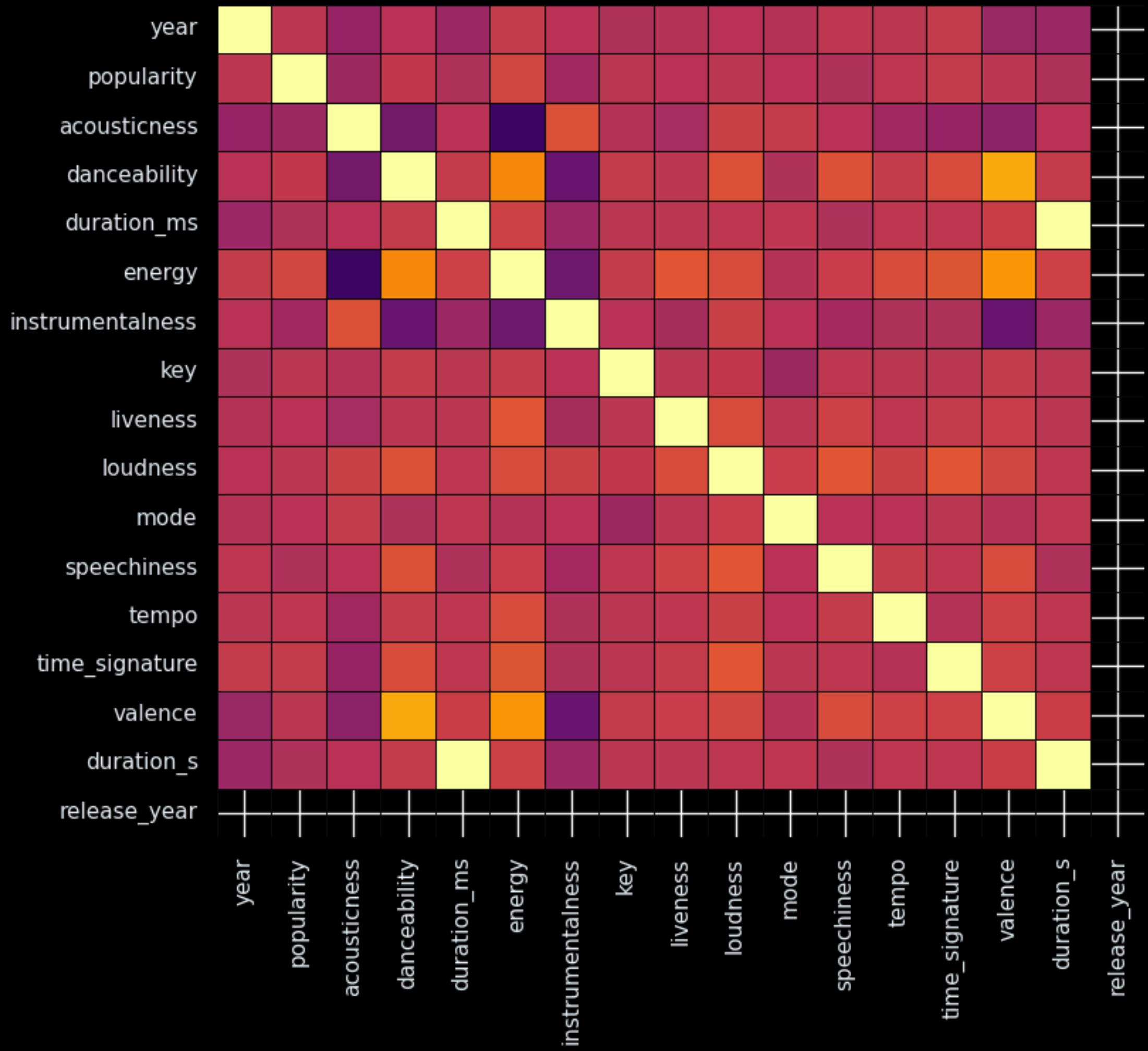
ou  
o o or  
o r or  
ro o r u  
  
ou  
o ro ou  
  
ou

Loudness vs Tempo of highly Popular Tracks by Musical Mode



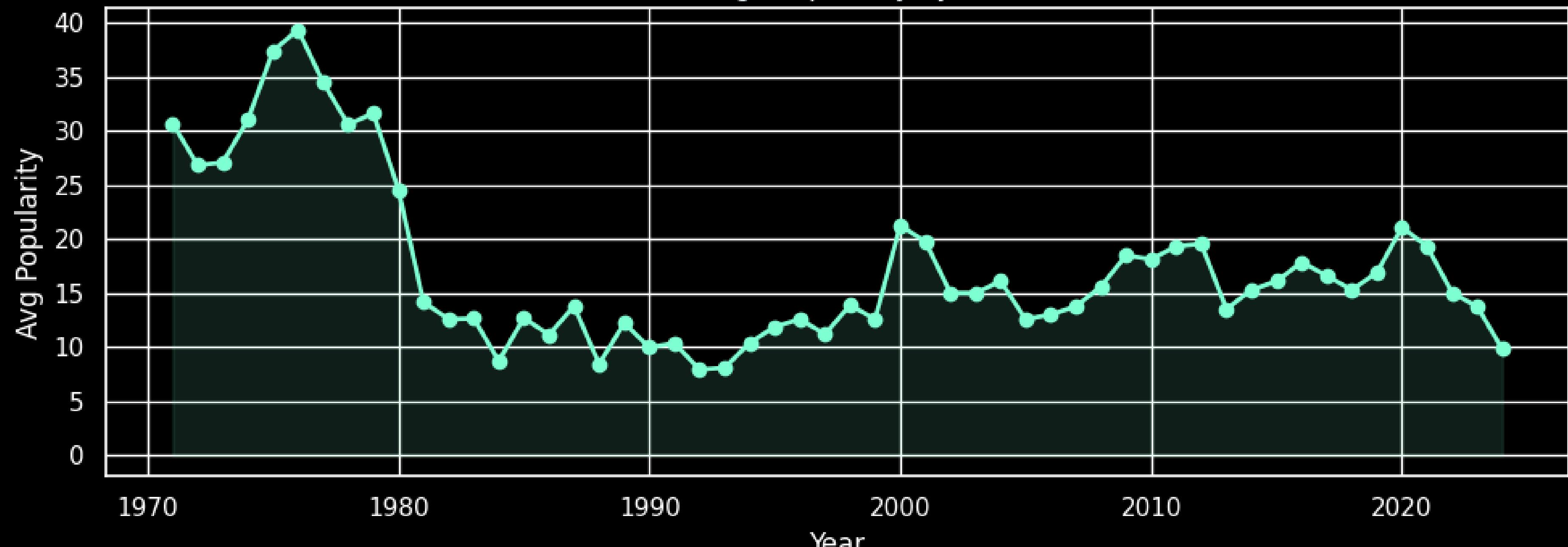
ou u r  
u r r  
or o

Correlation Heatmap



# TIME-SERIES ANALYSIS

## Average Popularity by Year



Mean popularity gradually increases over recent years (post-2015).

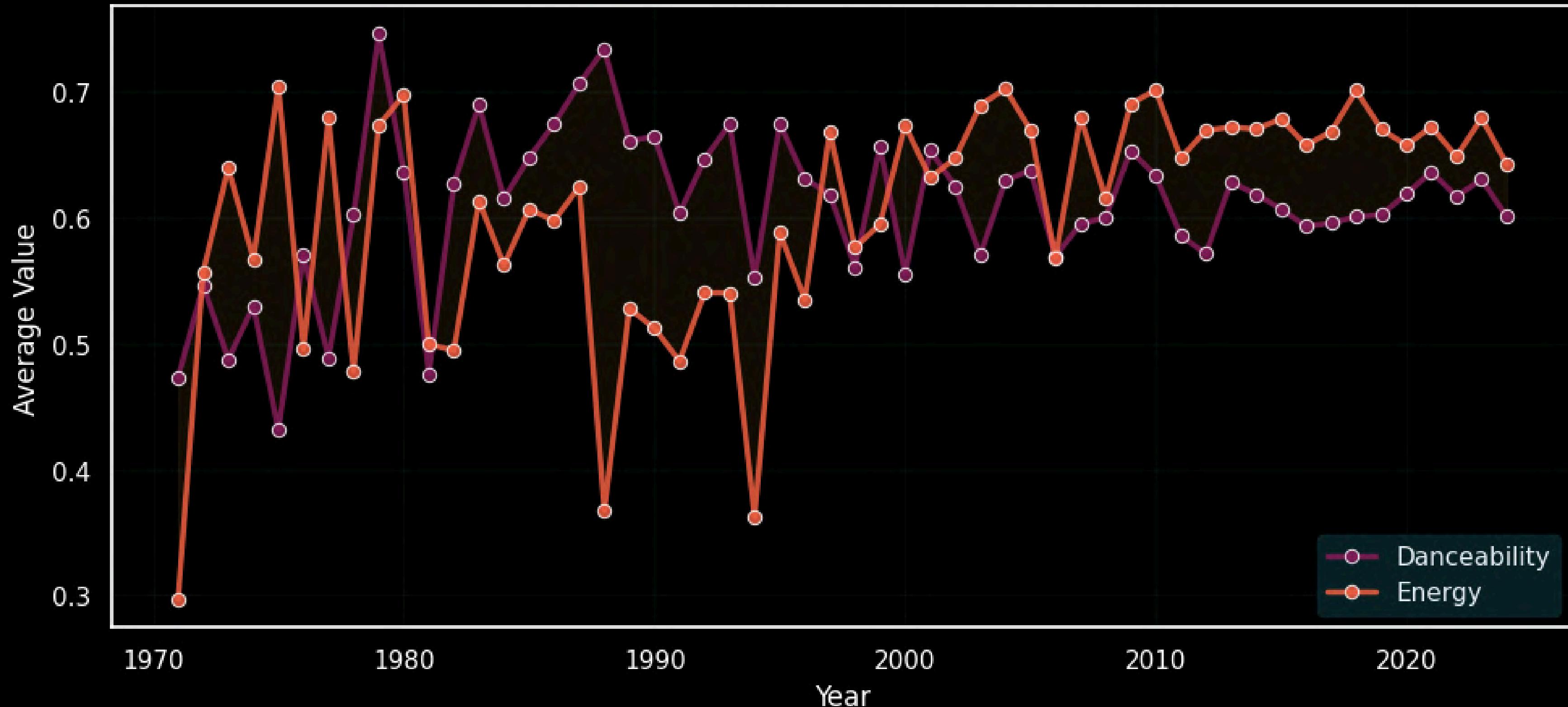
Modern streaming era songs trend ~10–15% higher than early 2000s averages.

Average popularity has risen steadily; songs in streaming era  
are generally more popular and widely spread.

Newer releases achieve higher popularity due to algorithmic promotion, playlist culture & replayability.

Focus production for playlist performance and viral potential as new-age distribution amplifies reach.

## Danceability & Energy of Popular Songs Over Years

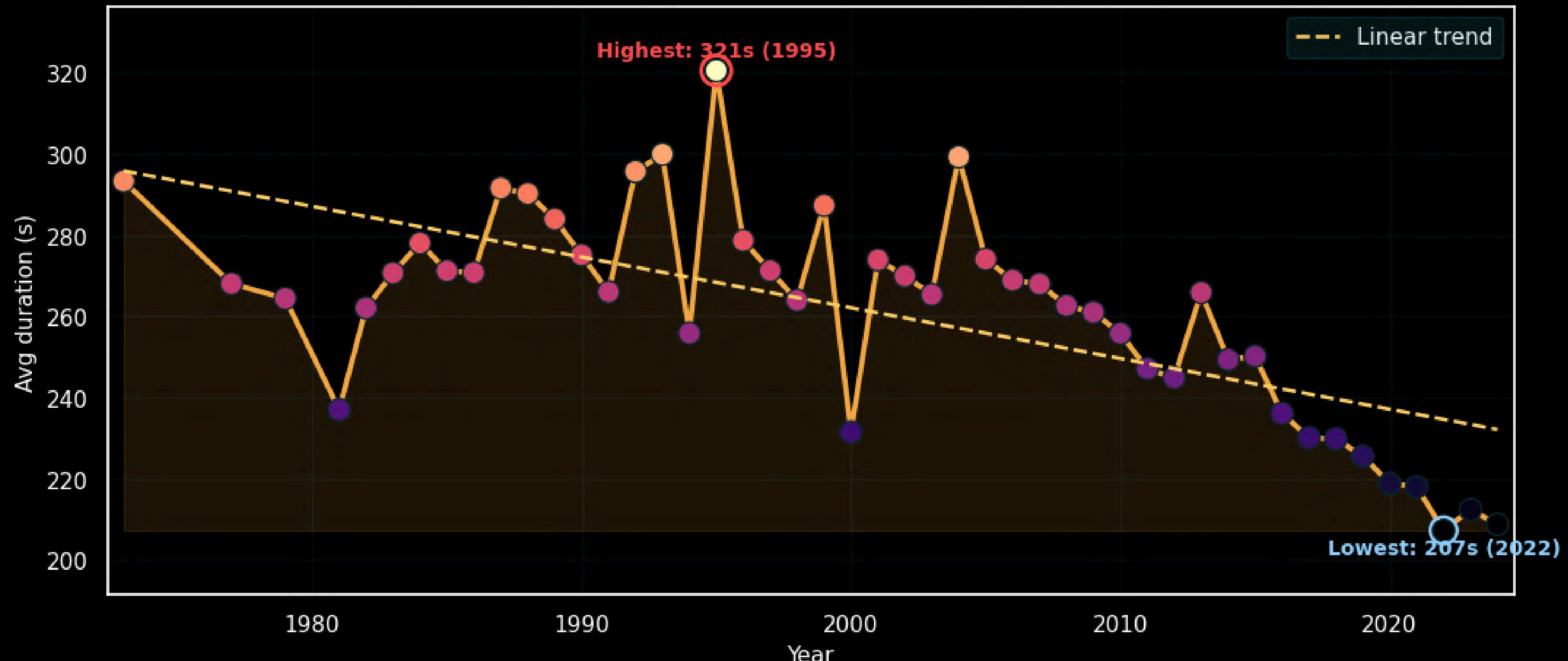


**Danceability shows upward drift over time & Energy remains consistently high but stable.**

*Listeners increasingly favor groovy, rhythmic songs, while high energy has become a baseline expectation. Danceability has increased, while energy levels remain consistently strong across years.*

**Modern hits should blend rhythmic feel with sustained energy, mirroring current listener demand.**

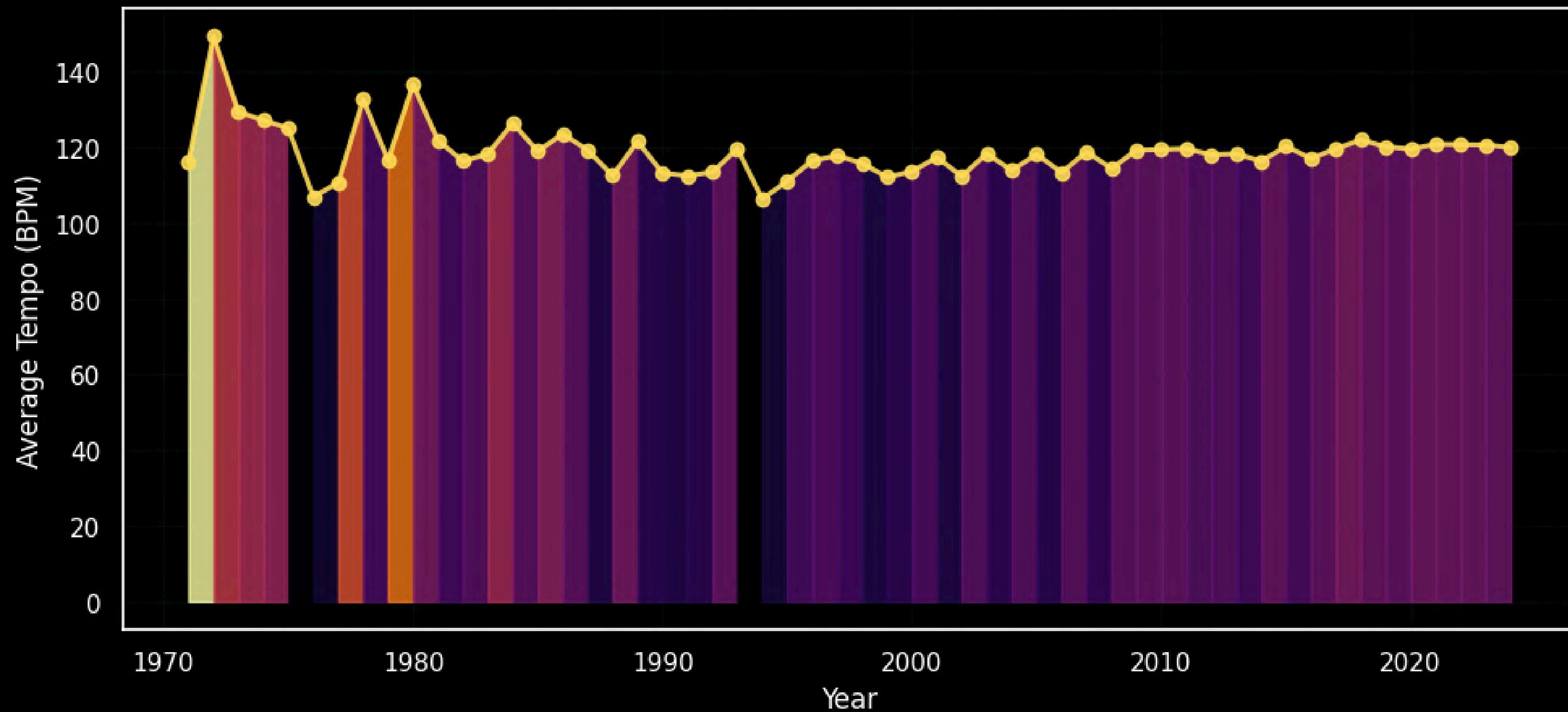
## Avg Duration (s) of Top-Quartile Songs by Year



Songs are getting shorter to favor streaming completion rates and replay loops. Modern tracks are significantly shorter, emphasizing hooks and replayability as Mean duration dropped from ~4.3 min (2000s) to ~3.2 min (recent).

Target 2:45–3:20 min duration for new releases; concise songs perform better in the streaming era.

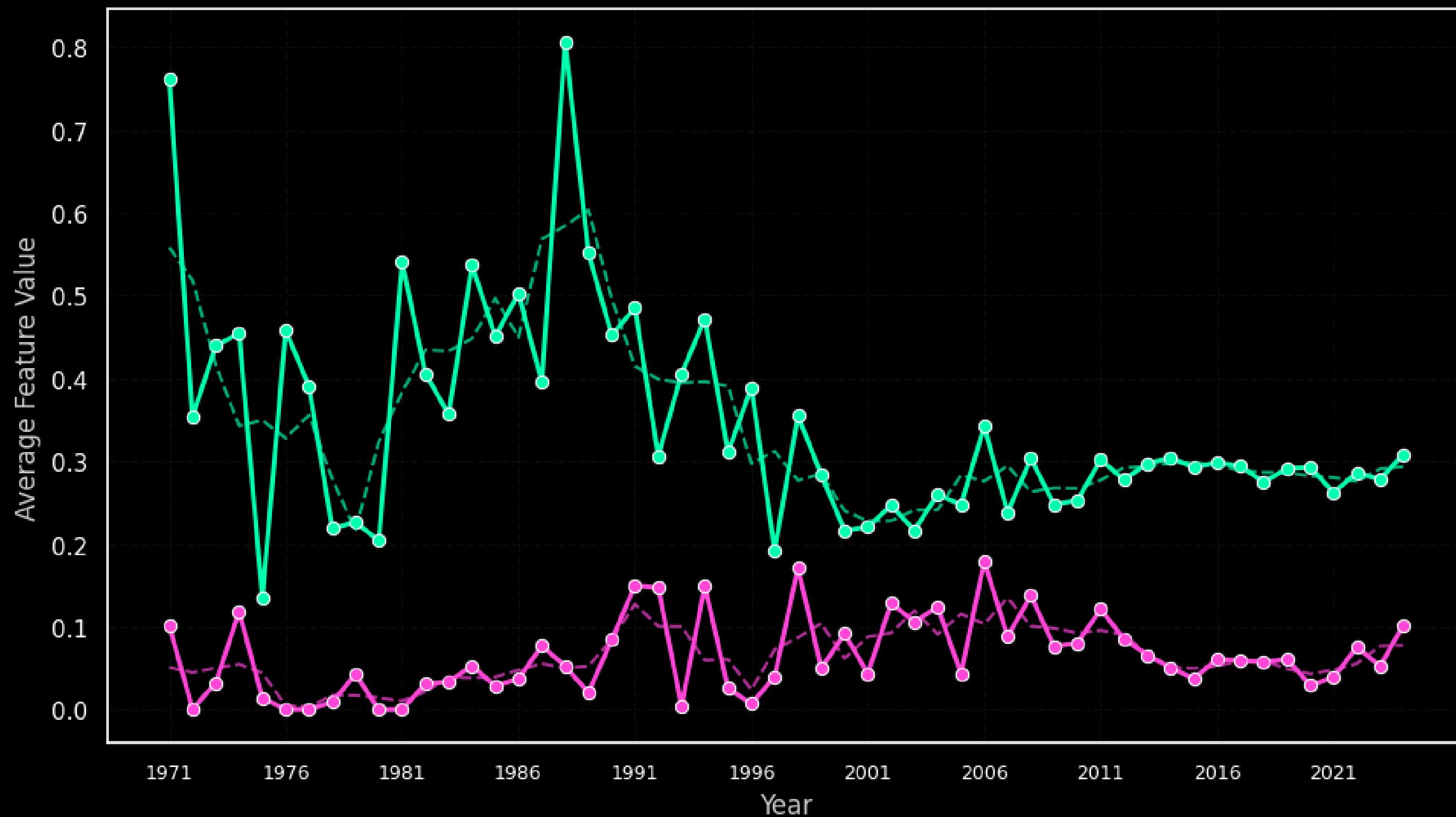
## Average Tempo of Popular Songs Over Years



Common keys (C, G, A minor) remain dominant. Tempo preferences have narrowed to mid-tempo ranges; Average tempo stabilizes around 115–125 BPM. tonal diversity remains, but accessibility drives key choice.

No major change is seen, tempo and key preferences stay stable, confirming a long-standing mainstream blueprint. Try to stick within 110–125 BPM and familiar keys; innovation should come from texture and mix rather than fundamental rhythm.

## Trends in Acousticness & Instrumentalness — Top 25% Popular Songs



**Acousticness and instrumentalness decline steadily post-2010.**

**Vocal-driven, digitally produced tracks dominate.**

**Both metrics have decreased, showing a shift toward electronic, vocal-heavy sound design.**

**Pop music has evolved toward heavily processed, vocal-centric production with minimal pure instrumentation. Modern hits rely on synthetic textures, stacked vocals, and digital layering, signaling a shift away from acoustic purity in production.**

# INSIGHTS SUMMARY

## 1. Keep Song Length Streamlined

- **The average hit track lasts 3–4 minutes ( $\approx 230\text{--}250$  seconds).**
- **Songs beyond 5 minutes lose replayability ; shorter tracks (~3:15) perform better on streaming platforms due to repeat plays.**
- **Focus on hook placement within the first 30–40 seconds to capture attention early.**

## 2. Target the Optimal Tempo and Groove

- **Tempo sweet spot: 115–125 BPM – perfect balance of energy and comfort for pop, EDM, and dance.**
- **4/4 time signature dominates (~86% of tracks) – universally familiar and danceable.**
- **Maintain steady rhythmic consistency; avoid over-complex syncopation that disrupts flow.**

## 3. Balance Energy and Emotion

- **Energy ( $\approx 0.7\text{--}0.8$ ) correlates most strongly with popularity ( $r \approx +0.31$ ).**
- **Hits are energetic but not exhausting – dynamic variation between verses and choruses keeps listeners engaged.**
- **Valence (0.55–0.65) is ideal – emotionally positive but not overly cheerful; “controlled optimism” works best.**

# INSIGHTS SUMMARY

## 4. Groove and Danceability Drive Replay Value

- Danceability 0.65–0.75 is common among top hits.
- Smooth rhythm and predictable groove patterns enhance memorability.
- Emphasize bass–kick synchronization and dynamic rhythm layering for consistent motion.

## 5. Master Loud but Controlled

- Popular tracks cluster around –7 dB loudness, balancing punch and clarity.
- Loudness alone doesn't guarantee success – focus on balanced dynamics and tonal brightness.
- Master between –8 to –9 LUFS, with headroom for streaming compression.

## 6. Focus on Tonal and Key Choices

- Major keys (~58%) slightly outperform minor keys.
- Favor vocally comfortable keys (C, G, A minor, E major).
- Ensure harmonic simplicity – catchy progressions beat complex theory in mass appeal.

## 7. Reduce Acousticness, Embrace Hybrid Production

- Acousticness in top songs is ~35% lower than average.
- Digitally layered production (electronic drums, processed vocals, synth pads) dominates the charts.
- Blend organic warmth + synthetic clarity for a modern, radio-ready sound.

# INSIGHTS SUMMARY

## 8. Keep Vocals Central and Melodic

- Speechiness drops in top hits → melodic content > spoken-word elements.
- Urban genres succeed when rap elements are rhythmically aligned and paired with sung hooks.
- Always ensure vocal clarity (EQ boost around 2–6 kHz) and emotional delivery.

## 9. Follow Evolving Trends

- Danceability has risen 20% and duration has dropped by ~22% in the past decade.
- Acousticness and instrumentalness declined, showing clear preference for processed, vocal-heavy mixes.
- Keep adapting to streaming trends – short intros, early hooks, and high replay value.

## 10. Combine Features – Don't Maximize One

- No single feature ensures success; it's the synergy that matters.
- Most hits cluster around:
  - Danceability ~0.7, Energy ~0.75, Valence ~0.6, Loudness -7 dB, Tempo 120 BPM.
- Maintain equilibrium between rhythm, energy, and emotional tone for broad listener appeal.

*“To make the next hit song, aim for emotional clarity, rhythmic stability, and technical balance. Successful songs are not the loudest or fastest – they are the most replayable, blending catchy rhythm, modern polish, and relatable emotion”.*

**END OF PRESENTATION**

**THANK YOU**