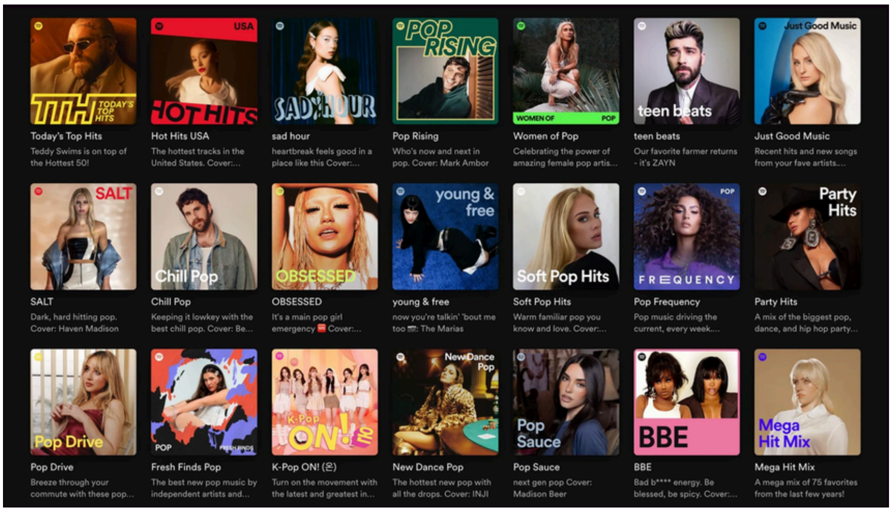
The Invisible Playlist Problem:

Why Spotify's Editorial Sound Doesn't Represent Real Curation

---



Introduction

Music streaming platforms have reshaped not only how we listen to music, but how music is organized and made visible. At the center of this ecosystem sits the playlist — the primary interface through which listeners encounter new artists, moods, and genres. For most people, playlists are personal objects shaped by memory, emotion, and routine. For a small fraction, however, playlists become public-facing products optimized for reach and discoverability.

For aspiring curators, this raises an uncomfortable question: if playlists are everywhere, why do only a handful gain visibility?
Most large-scale analyses focus almost exclusively on highly popular or editorial playlists — those already amplified by platforms. This creates a distorted picture of what "successful" curation looks like, while obscuring a basic reality: most playlists are expressive and never meant to scale.

This project asks a question that matters directly to anyone trying to break into the music curation scene:

> *Do the playlists that rise to the top reflect how people actually curate music — or are they selected for a narrow, platform-friendly template?*

By comparing playlist structure, title language, emotional signals, and audio features across popularity levels, this analysis shows that as playlists gain visibility, expressive diversity narrows — helping explain why distinctive playlists struggle to surface, and the tradeoffs curators face in today's editorial landscape.

Data Sources and Methodology

Sources:

- Spotify Million Playlist Dataset (MPD): 1,000,000 user-generated playlists, ~66 million playlist-track interactions.
- Audio Features: Stratified downsampling yielded ~1,040 playlists for audio analysis. Track previews were retrieved from YouTube, resampled, truncated, and normalized before feature extraction with librosa.

Popularity Buckets:

Playlists were categorized by follower count:

- Niche/Personal: < 10 followers
- Moderate: 10–100 followers
- Popular: 100–10,000 followers
- Very Popular: > 10,000 followers

Features Extracted:

- Rhythm: Tempo, zero-crossing rate
- Energy: RMS intensity
- Brightness: Spectral centroid
- Timbre: MFCCs (1–5)

Limitations:

- Deprecation of Spotify Web API audio-feature endpoints limited reliable retrieval of low-level audio descriptors (e.g., tempo, energy).
- Preview Bias: YouTube clips may omit full-track dynamics.
- Genre Skew: Availability favors mainstream tracks.
- Bucket Thresholds: While defined, follower counts are imperfect proxies for cultural reach.

Despite these constraints, stratified sampling and standardized preprocessing ensure comparability across buckets.

How Does Playlist Structure Change as Playlists Gain Visibility?

Before looking at sound or language, it's useful to ask a basic question: do popular playlists even look like personal ones? We compared playlist lengths across popularity buckets, measuring how many tracks appear in personal versus popular playlists.

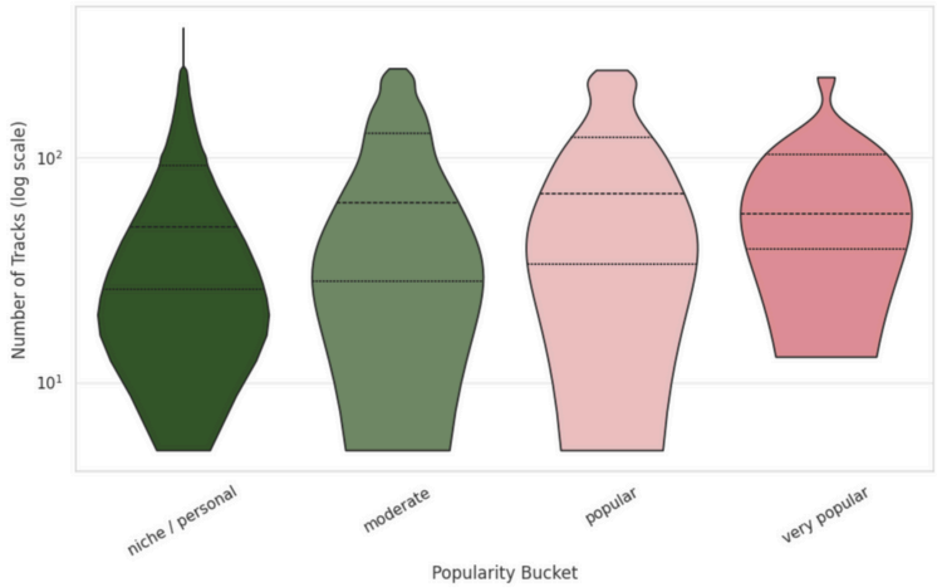


Figure 1: Playlist length distribution

Personal playlists are long and uneven, often growing over time. Popular playlists cluster tightly around shorter lengths. Personal playlists are longer and more variable, often exceeding 100 tracks. Popular playlists cluster tightly around 20–50 tracks. This reflects personal playlists as archives versus popular playlists as optimized products.
The log-scaled y-axis highlights this divergence: extreme values are far more common among personal playlists, indicating a tolerance for length and accumulation that does not persist at scale.

> *Once playlists are structurally constrained, how do curators change the way they present them?*

How Do Playlist Titles Shift from Personal Expression to Public Signaling?

2.1 Do Titles Become More Standardized as Playlists Grow?

Titles are one of the few explicitly human-authored elements of playlists. They reveal whether a playlist is meant for oneself or for discovery by others. We measured title length and looked for generic keywords commonly used in search-optimized playlists.

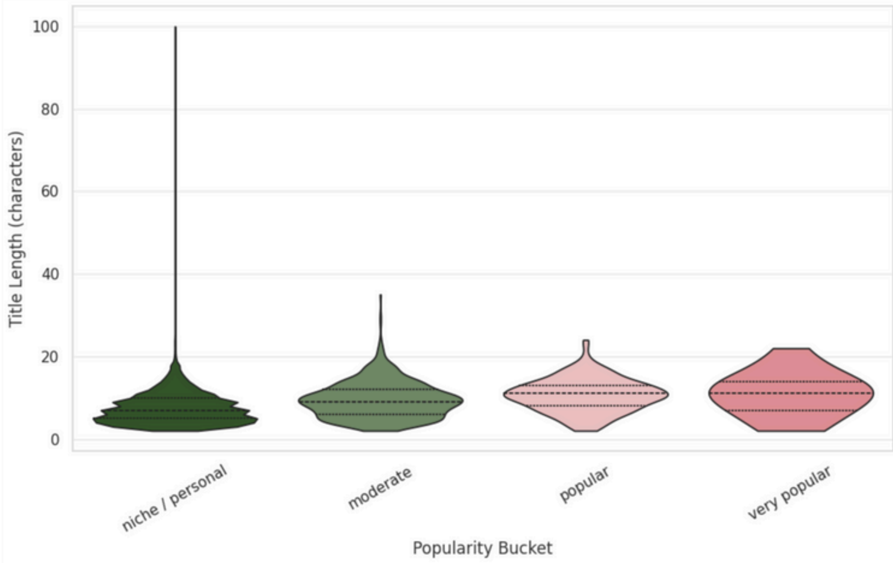


Figure 2: Title length by popularity

Playlist title length varies systematically with popularity, but not in a strictly monotonic direction. Personal and niche playlists tend to have the shortest median titles, reflecting informal and expressive naming practices, while also exhibiting substantial variability, including a long tail of unusually long titles.

As playlists become more popular, title lengths converge toward a narrower and more standardized range, suggesting increasing constraints driven by discoverability and readability.

2.2 What Happens to Linguistic Uniqueness at Scale?

Length alone doesn't capture creativity. A long title can still be generic.
We measured lexical rarity — how uncommon title words are relative to the full dataset.

Playlist titles were evaluated for generic, template-style terms (e.g., "top," "hits," "mix," "playlist," "music"). Generic naming is rare in niche and personal playlists (≈ 6.6%) and remains low in moderately popular ones (≈ 8.4%), but rises sharply among very popular playlists, where about 40% contain at least one generic keyword.

Personal playlists show the widest dispersion in lexical rarity, including extreme outliers such as "4x4," "Belters," and "#ViNTAGE." As popularity increases, lexical distributions compress: moderate and popular playlists rely on more common vocabulary with reduced variability. Very popular playlists form a distinct regime—despite often having longer titles, they exhibit the lowest mean lexical rarity

> *If language becomes more neutral and standardized, does the music itself follow the same pattern?*

2.3 Does Emotional Expression Disappear from Playlist Titles?

Emotion is central to how people use music privately. If it vanishes at scale, that signals a deeper transformation. We applied a transformer-based emotion classifier to playlist titles and aggregated results by popularity bucket.

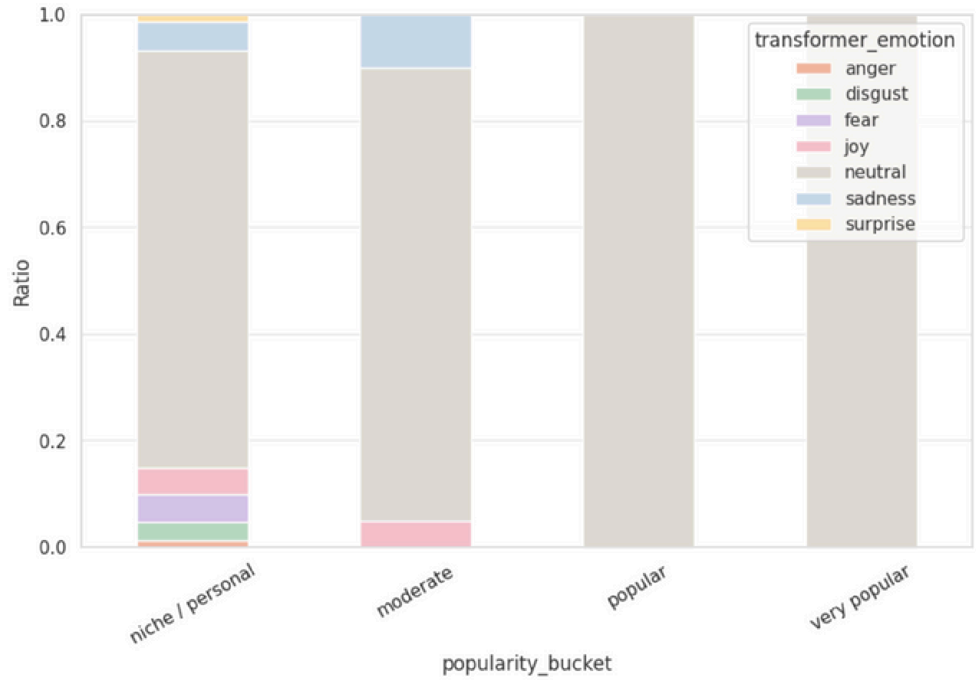


Figure 4: Transformer-Based Emotion Composition of Playlist Titles by Popularity

Niche and personal playlists display a broader range of emotions. As popularity increases, emotionally marked titles become increasingly rare, and popular and very popular playlists are almost entirely neutral.

This pattern indicates a shift from expressive naming to functional labelling as playlists scale. Titles aimed at broad audiences avoid emotional specificity, favouring neutrality to maximise clarity and reach. While absolute emotion labels should be treated cautiously due to sentiment-model limits on short text, the consistent differences across popularity buckets reinforce the conclusion that emotional expressiveness in playlist titles declines with scale.

How Does Playlist Sound Change as Popularity Increases?

Structure and language can be optimized intentionally. Sound is harder to fake. If convergence appears here too, it suggests deeper selection pressures. We extracted audio features related to rhythm, energy, brightness, and timbre from track previews and compared their distributions across popularity levels.

3.1 Do Popular Playlists Avoid Rhythmic Extremes?

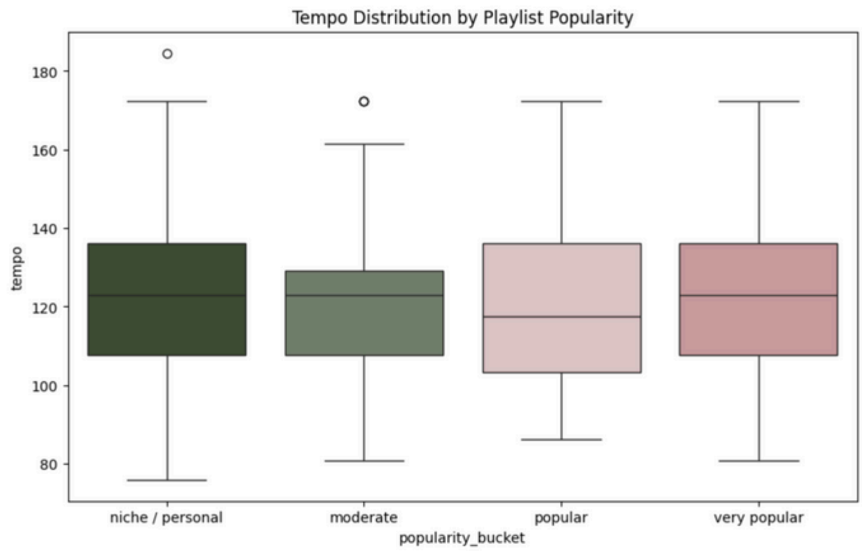


Figure 5. Tempo distribution by playlist popularity

Personal playlists span wide tempo ranges; popular playlists cluster narrowly around mid-tempo (100–120 BPM).

3.2 Does Energy and Brightness Converge with Popularity?

RMS energy and spectral centroid were used as proxies for intensity and brightness. While median values remain similar, variability drops with popularity. Popular playlists avoid both low- and high-energy extremes.

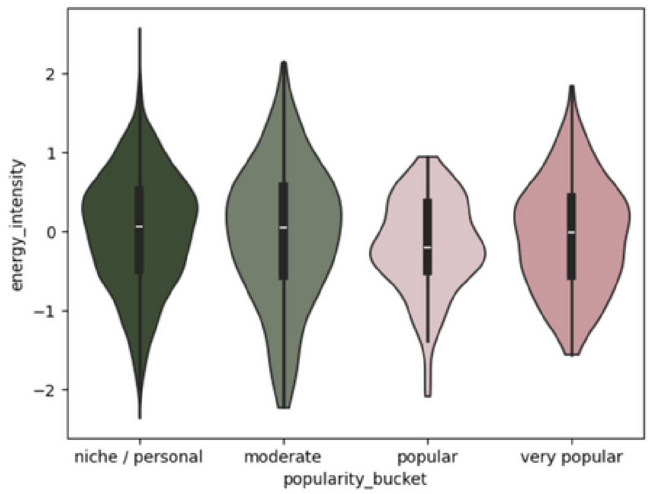


Figure 6(a). RMS energy distribution by popularity bucket

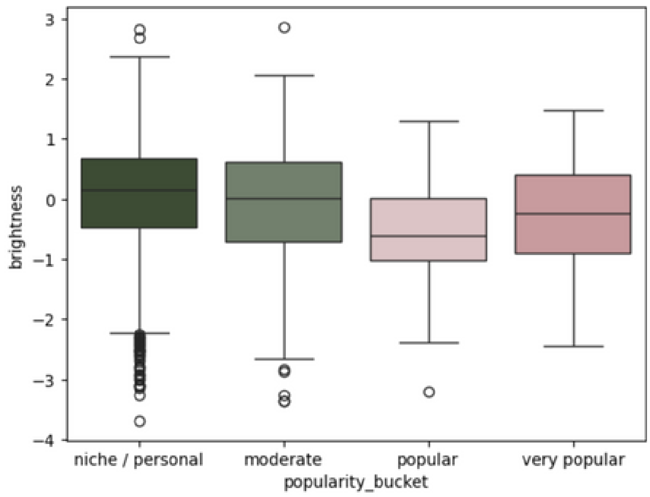


Figure 6(b). Spectral centroid by playlist popularity

Is Variability the Hidden Factor That Separates Personal from Popular Playlists?

4.1 Do Popular Playlists Allow Less Sonic Variety Within a Playlist?

Within-playlist variability shows how much range a playlist allows: high variability signals freedom, low variability signals constraint. We measured differences in tempo, energy (RMS), brightness (spectral centroid), and timbre (MFCCs), then compared variance across popularity buckets.

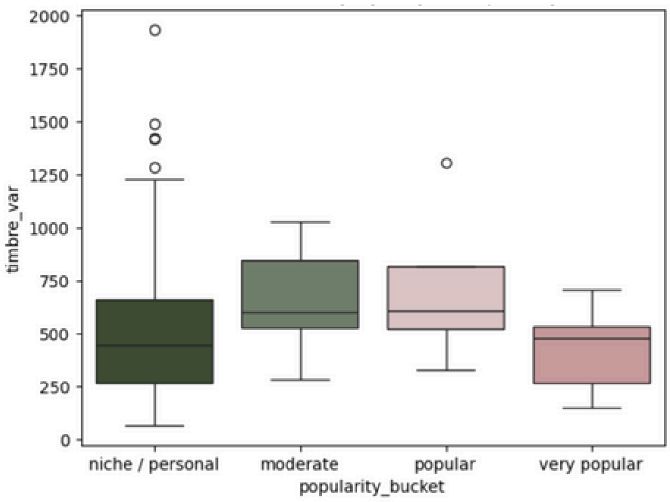


Figure 7(a). MFCC variance by popularity bucket

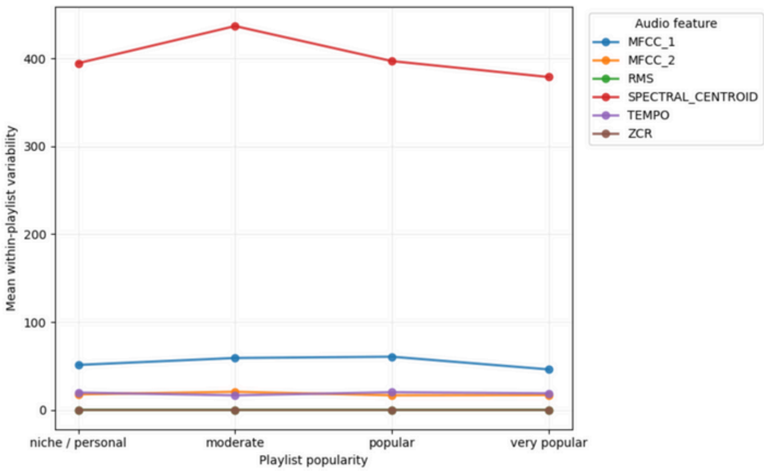


Figure 7(b). Sonic Variability Decreases with Playlist Popularity figure.

Niche playlists show the highest timbral diversity, while very popular playlists are the most uniform in sound. Across features, tempo remains the most variable, even in mainstream playlists, whereas timbre, loudness, and tonal qualities stay consistent. Overall, popularity correlates with sonic cohesion, but controlled variability—especially in tempo—helps maintain listener interest.

4.2 Are Popular Playlists More Sonically Cohesive Overall?

Variance measures spread feature by feature. Cohesion asks a broader question: how tightly do tracks within a playlist cluster when considered together? We applied Principal Component Analysis (PCA) to the audio features and calculated each track's distance from its playlist's centroid in feature space. Smaller distances indicate greater internal cohesion.

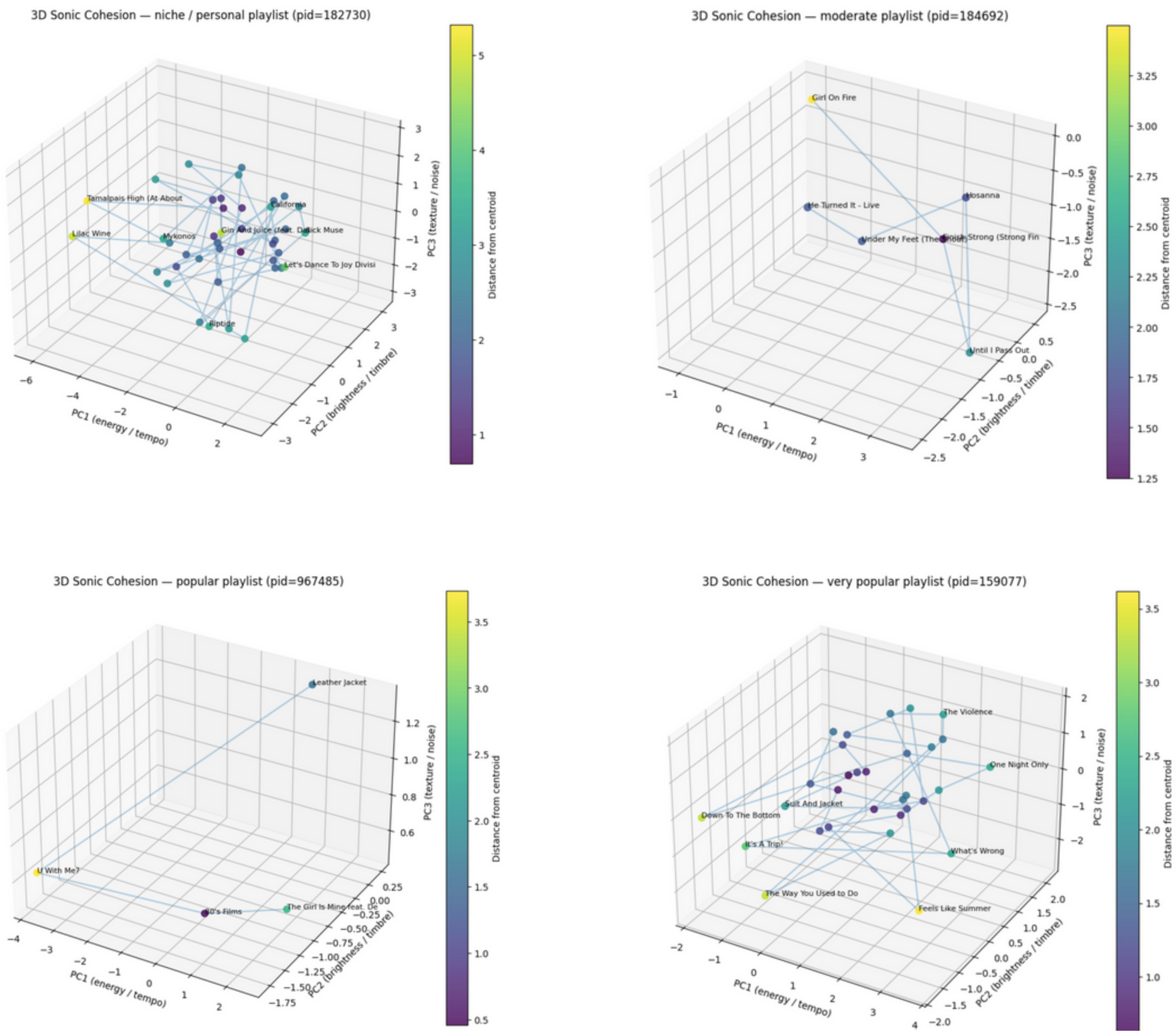


Figure 8. PCA projections of audio features for representative playlists across popularity buckets.

Very popular playlists show the lowest mean distances and the tightest distributions, forming compact clusters in feature space. Personal playlists are far more dispersed, indicating looser, more exploratory sequencing. The PCA results confirm that increased popularity is associated with systematic sonic cohesion. This is not the result of a single feature, but a collective narrowing across rhythm, energy, brightness, and timbre.

pid	popularity_bucket	mean_dist
159077	very popular	1.739744
184692	moderate	2.05772
967485	popular	2.087068
182730	niche / personal	2.478965

Table 1 . Mean distance from centroid for playlists in Fig. 11 as a measure of playlist cohesion.

To complement the PCA visualization, Table 2 shows the mean distance from centroid for four playlists, one from each popularity bucket. The hierarchy—lowest for very popular playlists and highest for niche/personal—remained consistent across multiple random sets, supporting our hypothesis that popularity correlates with cohesion.

Beyond these examples, Table 1 reports mean cohesion across all playlists in the very popular (1.76) and niche/personal (1.82) buckets, again confirming that mainstream playlists are more uniform while niche playlists allow greater diversity.

What Changes as Playlists Become Popular — and Why?

Personal playlists function as expressive artifacts rather than optimized products. They tolerate abrupt shifts in mood, tempo, and timbre, reflecting how people actually live with music. Title language is playful and idiosyncratic, and audio features span wide ranges, indicating that variability is a defining characteristic rather than noise.

As playlists gain popularity, this expressive range consistently narrows. Popular playlists emphasize sonic cohesion, neutral language, and predictable emotional tone. PCA and variance analyses show that very popular playlists cluster tightly in feature space, with reduced dispersion across tempo, energy, brightness, and timbre. Emotional markers in titles also decline, converging toward neutrality.

This shift is not incidental. It reflects structural selection pressures: algorithms and editorial norms reward playlists that minimize surprise and maximize legibility at scale. Cohesion improves discoverability, but it does so by filtering out sonic and linguistic risk. What changes with popularity, then, is not just audience size, but the range of expression that survives visibility.

What Does This Mean for Platforms, Curators, and Artists?

The data suggests that today's editorial ecosystem systematically favors predictability over diversity — a design choice with practical consequences.

For platforms, incorporating diversity-aware signals (such as within-playlist variance in tempo or timbre) could counterbalance homogenization without sacrificing usability. Making cohesion–diversity tradeoffs explicit would allow expressive playlists to compete rather than be quietly suppressed.

For curators, growth is not a neutral proxy for quality. Cohesive playlists travel farther, but expressive variation is what differentiates curation. Understanding when to optimize for legibility and when to preserve idiosyncrasy becomes a strategic decision rather than a creative failure.

For artists, the findings clarify why mainstream playlists reward sonic consistency, while personal playlists remain spaces for experimentation. Targeting both contexts — cohesive tracks for scale, diverse work for niche circulation — aligns with how visibility is actually structured.

Reproducibility

- **Data:** Spotify MPD (1M playlists), stratified sample of ~1,040 playlists.
- **Buckets:** Explicit thresholds defined by follower count.
- **Methods:** Preprocessing documented (resampling, truncation, normalization).
- **Limitations:** Preview bias, genre skew acknowledged.

What Gets Lost When Playlists Scale — and Who Decides?

The results show a consistent pattern: personal playlists are diverse, expressive, and emotionally marked, while popular playlists are cohesive, neutral, and optimized for scale. As playlists move from private use into public visibility, variability declines across structure, language, and sound. This does not imply that expressive playlists lack value. Rather, it shows that visibility is governed by constraints that favor a narrow curatorial template. Recognizing this gap reframes the editorial landscape not as a reflection of the best playlists, but as a filtered subset shaped by platform incentives. Understanding what gets lost — and why — allows curators, artists, and listeners to make more informed choices about conformity, experimentation, and the kind of musical culture they want playlists to represent.