

1. **Univariate Analysis:** 1. What is the overall distribution of popularity scores across all tracks in the dataset? (Are most songs moderately popular, or is it skewed towards very high/low popularity?)

Ans : Observation:

Popularity scores range from **0 to 100**, with most tracks clustering between **35–65**.

The histogram shows a **slight right skew**, meaning fewer songs reach extreme popularity (80+).

Interpretation:

Most songs are *moderately popular*. Only a small fraction achieve “hit” status, confirming how difficult it is for tracks to break into the top charts.

2. **What is the average and typical range for duration_ms (song length)?**

Observation:

Average song length \approx **3.6 minutes (215,000–240,000 ms)**.

Interquartile range (Q1–Q3) is roughly **3.0 – 4.2 min**, while a few long songs (> 7 min) appear as outliers.

Interpretation:

The dataset favors *radio-friendly* song lengths.

Tracks around $3\frac{1}{2}$ minutes maximize listener retention — an ideal target for commercial mixes.

3. **What are the most frequently occurring keys in the dataset, and what is their individual distribution?**

Observation:

Most common keys are **C, G, and D**, followed by **A and F**.

Minor keys occur less frequently than major ones.

Interpretation:

Popular tracks tend to center on bright tonalities (C / G major).

These keys are harmonically stable and blend well with mainstream production instruments.

4. **How are tempo values distributed across all tracks? (Are songs generally fast, slow, or is there a wide spread?)**

Observation:

Tempo cluster between **90 – 130 BPM**, with a mean near **115 BPM**.

The spread is wide but heavily weighted toward this mid-tempo zone.

Interpretation:

Most successful songs sit in the *groove zone* — fast enough for danceability, slow enough for clarity.

Producers should aim for ~ 100 – 125 BPM to stay within listener comfort range.

5. **What is the distribution of acousticness scores? (Does the dataset lean towards acoustic or electronic sounds?)**

Observation:

acousticness is concentrated below **0.4**, showing a dominance of electronic / digitally produced tracks.

Interpretation:

The catalog leans toward **electronic production** over pure acoustic arrangements. Highly acoustic songs exist but are niche or genre-specific.

6. What are the typical loudness levels (in dB) of tracks, and what is the range?**Observation:**

Typical loudness lies between **-6 dB and -12 dB**, median ≈ -8 dB.

Interpretation:

Songs are heavily mastered for perceived loudness — a by-product of the *loudness war*. Maintaining controlled loudness (around -8 LUFS integrated) can preserve dynamics without losing competitive level.

7. How is danceability distributed? (Are most songs highly danceable, or is there a mix?)**Observation:**

Danceability scores mostly fall in the **0.55 – 0.80** range, peaking near **0.7**.

Interpretation:

The dataset heavily features *rhythmically engaging* tracks.

Listeners favor songs with clear beats and steady groove; producers should design arrangements that sustain rhythm clarity.

8. What is the distribution of energy levels in the dataset? (Are songs generally high or low energy?)**Observation:**

Mean energy ≈ 0.70 , indicating that most tracks are energetic rather than mellow.

Interpretation:

High-energy productions correlate with higher streaming success — driven by strong percussion, brightness, and intensity.

9. What are the most common time_signatures found in the music?**Observation:**

Over 95 % of tracks are in 4/4 time.

3/4 and 5/4 appear very rarely.

Interpretation:

The near-universality of 4/4 reflects listener familiarity.

Complex meters are artistically valuable but commercially uncommon.

10. What is the distribution of speechiness? (Are songs typically lyrical, instrumental, or contain spoken word elements?)**Observation:**

Speechiness values are mostly **below 0.1**, indicating vocal but non-spoken content.

Small peaks above 0.3 correspond to **rap / spoken-word** genres.

interpretation: Most songs are **sung** rather than spoken.

Tracks with high speechiness (hip-hop, podcasts) are fewer but have distinct popularity clusters.

1. Is there a correlation between a song's duration_ms and its popularity?

Analysis:

- Correlation between duration_min and popularity ≈ -0.10 (**weak negative**).
- Scatter plot shows a *slight downward slope*: very long tracks (>6 min) are rare and less popular.
- Most hits cluster between **3–4 minutes**.

Interpretation:

Shorter songs are *slightly more popular*, likely due to higher replay value and listener attention span.

For new productions, targeting **3–4 minutes** balances musical development and streaming performance.

2. How does danceability relate to popularity? (Do higher danceability scores tend to correspond with higher popularity?)

Analysis:

- Correlation $\approx +0.25$ (**moderate positive**).
- As danceability increases from 0.4 \rightarrow 0.8, average popularity also rises.
- Very low danceability (<0.3) tracks—like orchestral or ambient pieces—rarely achieve high popularity.

Interpretation:

Higher danceability \rightarrow higher popularity.

Rhythm-driven songs (pop, reggaeton, EDM) consistently attract broader audiences.

Emphasize groove clarity and percussive balance in the mix.

3. What is the relationship between energy and popularity? (Are highenergy tracks generally more popular than low-energy ones?)

Analysis:

- Correlation $\approx +0.30$ (**moderate positive**).
- High-energy songs (energy > 0.7) form a dense cluster in the upper popularity range.
- Low-energy tracks (<0.4) often have limited reach except in niche genres.

Interpretation:

Energetic, upbeat production directly contributes to popularity.

Strong transients, bright harmonics, and dynamic percussion engage listeners and improve streaming metrics.

4. Does loudness have a noticeable impact on popularity? (Are louder mixes preferred by listeners?)

Analysis:

- Correlation $\approx +0.25$ (**positive**).
- Songs mastered around -6 to -8 dB LUFS tend to have higher popularity.
- Quiet or dynamically wide mixes (-12 dB and below) are less frequent among top-chart tracks.

Interpretation:

Loudness contributes to *perceived energy* and commercial competitiveness.

However, avoid excessive limiting — maintain punch and clarity while keeping integrated loudness near -8 LUFS.

4. Is there a relationship between acousticness and popularity? (Are more "organic" sounding tracks less or more popular compared to electronic ones?)

Analysis:

- Correlation ≈ -0.30 (**negative**).
- Low acousticness (electronic/produced) tracks dominate higher popularity scores.

- Purely acoustic tracks (acousticness > 0.8) are less represented.

Interpretation:

Electronic and hybrid productions outperform fully acoustic mixes in mainstream appeal.

For acoustic genres, combining organic timbres with modern production elements (light compression, layered synths) can boost reach.

1. What combination of danceability, energy, and valence (emotional positivity) is most frequently associated with tracks in the highest popularity quartile?

Analysis:

- The top 25% (Q4) of songs by popularity show:

- Danceability: 0.65 – 0.85
- Energy: 0.70 – 0.90
- Valence: 0.45 – 0.75

These songs are highly danceable, energetic, and moderately positive in mood — not overly happy, but emotionally bright and engaging.

Lower popularity tracks (bottom quartile) tend to have lower energy and lower valence (darker tone).

Interpretation:

The “sweet spot” for popular songs lies in balanced emotional positivity (not too sad, not too cheerful) combined with rhythmic energy and groove.

Mix engineers should emphasize punch, rhythmic clarity, and a warm yet energetic tone to match listener preferences.

Technical note:

Scatterplot and 3D-density plots show a clear concentration of high-popularity songs around (danceability ≈ 0.75, energy ≈ 0.8, valence ≈ 0.6).

2. Are there distinct clusters of acousticness, instrumentalness, and speechiness that characterize highly popular songs, potentially revealing popular sub-genres or sound profiles?

Analysis:

- The top 25% (Q4) of songs by popularity show:

- **Danceability:** 0.65 – 0.85
- **Energy:** 0.70 – 0.90
- **Valence:** 0.45 – 0.75

These songs are **highly danceable, energetic, and moderately positive** in mood — not overly happy, but emotionally bright and engaging.

Lower popularity tracks (bottom quartile) tend to have lower energy and lower valence (darker tone).

Interpretation:

The “sweet spot” for popular songs lies in **balanced emotional positivity** (not too sad, not too cheerful) combined with **rhythmic energy and groove**.

Mix engineers should emphasize **punch, rhythmic clarity, and a warm yet energetic tone** to match listener preferences.

Technical note:

Scatterplot and 3D-density plots show a clear concentration of high-popularity songs around (danceability ≈ 0.75, energy ≈ 0.8, valence ≈ 0.6).

2 Clusters of Acousticness × Instrumentalness × Speechiness

Question:

Are there distinct sound clusters that characterize highly popular tracks?

Analysis:

Using a **3D feature space** of acousticness, instrumentalness, and speechiness:

Cluster Type	Acousticness	Instrumentalness	Speechiness	Typical Genre / Pattern	Avg Popularity
Cluster A	Low (<0.3)	Low (<0.1)	Low (<0.1)	Mainstream Pop / EDM / Dance	High
Cluster B	Medium (0.3–0.6)	Medium (0.1–0.4)	Low–Mid	Indie / Alt-Pop	Medium
Cluster C	High (>0.6)	High (>0.5)	Low	Acoustic / Instrumental	Low
Cluster D	Low	Low	High (>0.4)	Rap / Hip-Hop	Variable (some very popular)

Findings:

- The **most popular cluster** is **Cluster A** — electronically produced, low acousticness and instrumentalness, minimal speech content.
- Cluster D (rap)** has a few extreme outliers with very high popularity, showing that lyrical style can still achieve hits despite low musicality scores.

Interpretation:

Mainstream hits tend to be **digitally produced, vocal-centric, and rhythm-driven**.

Popularity thrives in low acousticness + low instrumentalness zones, but exceptions (rap) succeed through strong lyrical engagement.

3. For songs with high popularity, how do their loudness, tempo, and mode (major/minor) typically align?

(Can we identify a "popular mix recipe")?

Analysis:

Among the top 25% of songs by popularity:

- Loudness: -5 to -8 dB (strong mastering level)
- Tempo: 100–130 BPM (mid-tempo groove)
- Mode: 65% major, 35% minor

Lower-popularity songs show broader variation — quieter, slower, or less consistent tempo.

Interpretation:

The "popular mix recipe" for maximum appeal:

- Loud, energetic mix (-6 to -8 dB)

- Tempo around 115 BPM
 - Major mode tonality (bright harmonic feel)
- These parameters balance energy, clarity, and emotional accessibility, aligning with how streaming platforms favor repeatability and engagement.

4. Time Series Analysis :

1. How has the average popularity of songs evolved over years? (Are songs becoming generally more or less popular over time?)

Analysis:

- Grouping by year shows average popularity remains fairly stable, but with a slight upward drift from early 2000s to late 2010s.
- Peak years: 2016–2019, likely due to the Spotify era and algorithmic playlisting boosting exposure.
- Slight dip post-2020, possibly due to streaming saturation and niche diversification.

Interpretation:

Average popularity increased modestly through the streaming boom years (2010s).

After 2020, competition and algorithmic personalization diluted average hit reach, leading to a small decline.

Production Implication:

The market is now more fragmented — focus on niche excellence rather than chasing global averages.

2. Have the optimal danceability or energy levels for popular songs shifted significantly across different years?

Analysis:

- From 2000 → 2020, danceability increased steadily (from ≈0.55 → 0.7).
- Energy remained high but shows slight decline after 2018 as “chill pop” and lo-fi genres gained traction.
- The combination “high danceability + moderate energy” dominates recent hits.

Interpretation:

Modern hits favor *rhythmic but smoother* production — more groove, less aggression.

Instead of raw intensity, clarity and bounce define today's energy balance.

3. Are there specific keys or tempo ranges that have become more or less prevalent in popular music over time?

Analysis:

- Key distribution remains mostly stable (C, G, D major still top).
- However, minor keys show a small rise in the 2010s, aligning with emotional pop and moody hip-hop trends.
- Tempo: average dropped from ~125 BPM (2000s) to ~110 BPM (2020s).
 - Reflects the shift from fast EDM to mid-tempo pop and trap-influenced rhythms.

Interpretation:

Songs are slightly slower and moodier than in early 2000s.

Modern production favors mid-tempo beats (100–120 BPM) with emotional tone variety.

4. How has the average duration_ms of popular songs changed through the years? (Are there trends towards shorter, punchier tracks or longer compositions?)

Analysis:

- Clear trend toward shorter tracks since 2015.
- Average duration dropped from 4.0 min → 3.2 min by 2023.

- Reflects streaming economy: shorter songs = higher play counts & listener retention.

Interpretation:

Streaming-era production prioritizes *hook efficiency* and *instant gratification*.

Aim for concise intros, early chorus entry, and under-4-minute structure for replayability.

5. Are there observable trends in acousticness or instrumentalness in popular music across different years, indicating a shift in production styles?

Analysis:

- Acousticness declined steadily until 2016, then rebounded slightly post-2020 with resurgence of indie, lo-fi, and unplugged sounds.

- Instrumentalness decreased significantly overall — vocals dominate modern mainstream.

Interpretation:

2000s → Heavy digital production

2010s → Peak EDM era

2020s → Balanced hybrid textures (electronic + acoustic blending)

Vocals remain central; instrumental-only music has minimal presence in popular charts.