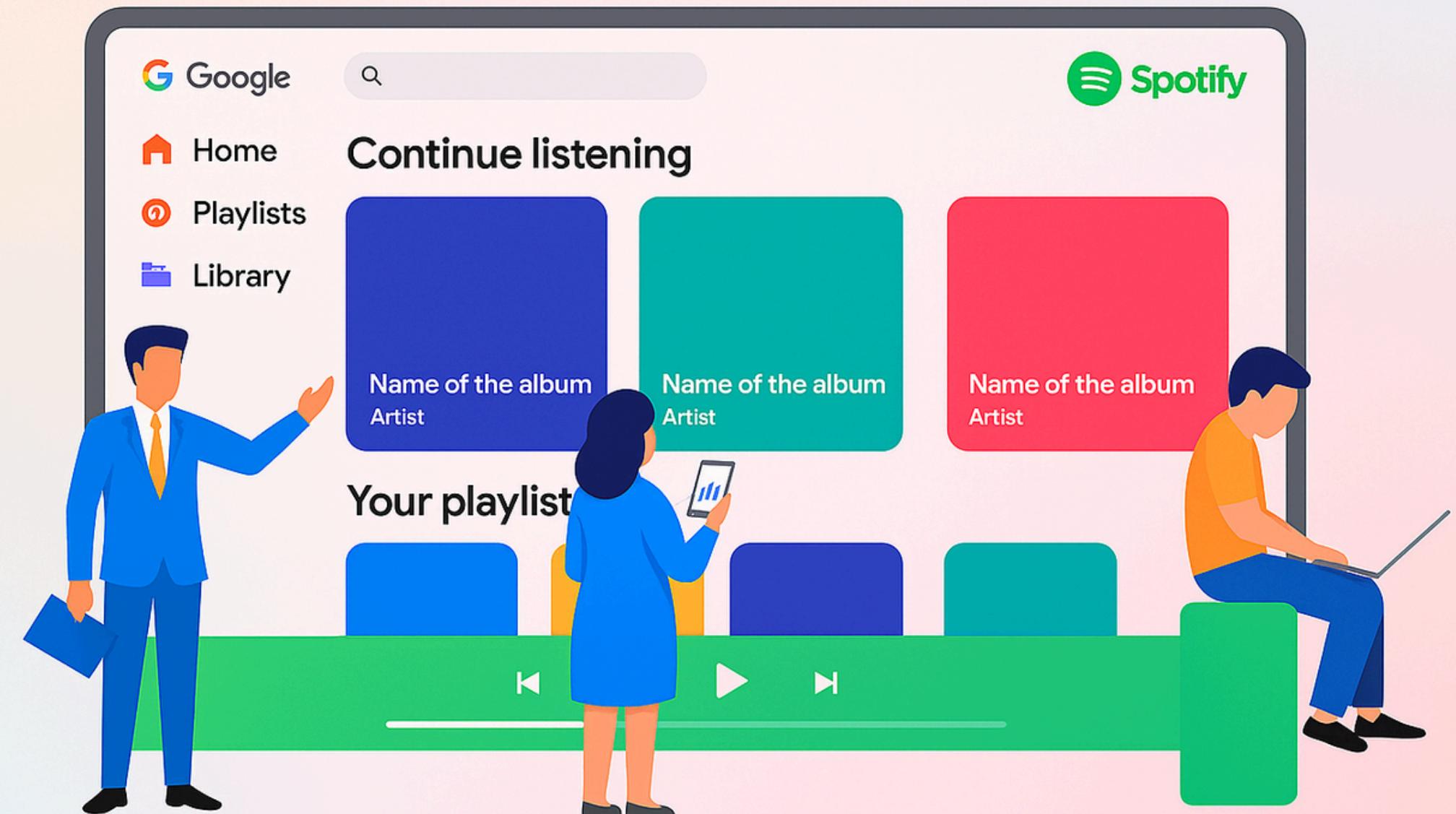


Project 1 on Exploratory Data Analysis (Spotify Dataset)

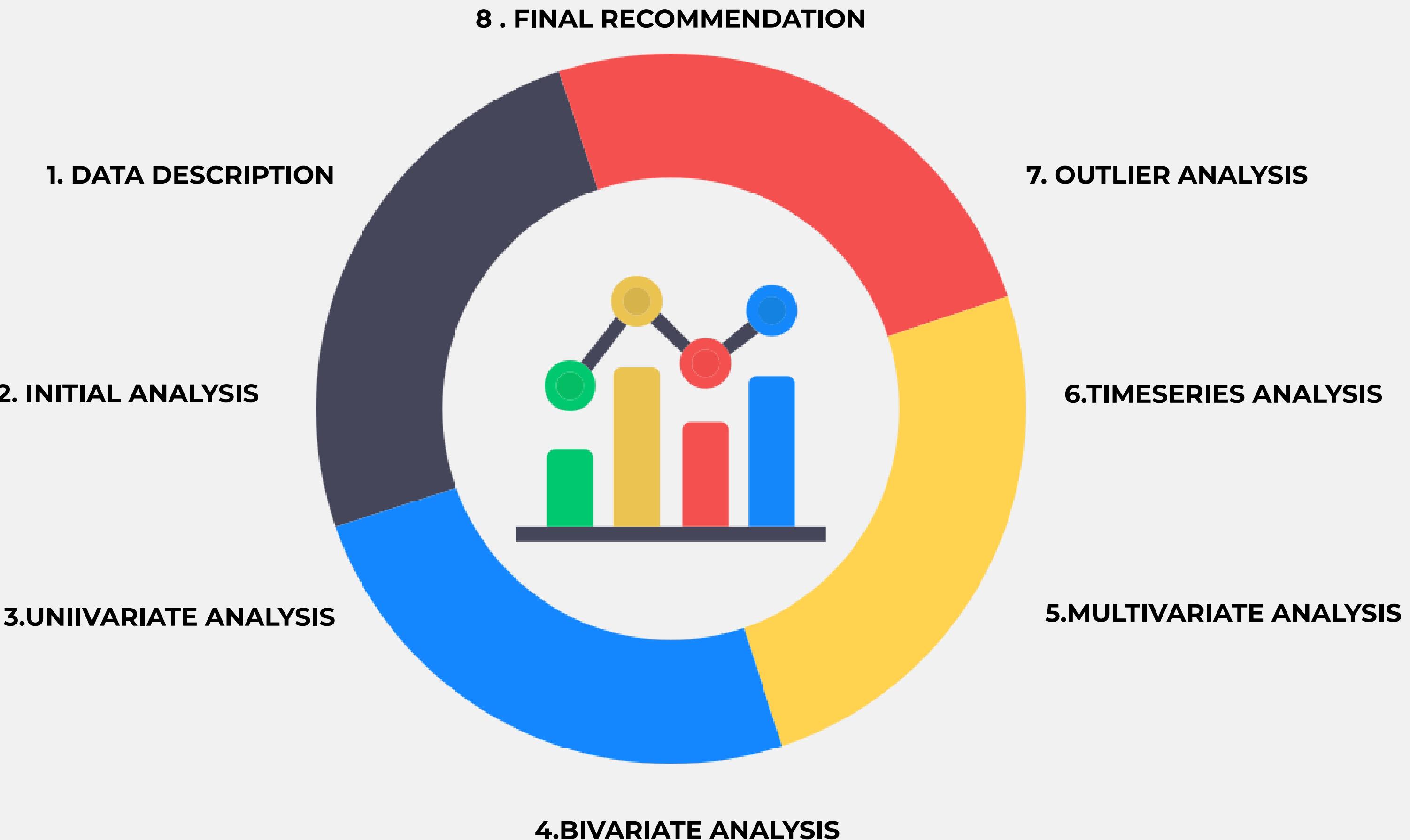
Problem Statement Summary:

As a Music Director or Mixing Engineer, the goal is to optimize new songs for popularity.

Using the Spotify dataset, which includes features like danceability, energy, loudness, and duration, we aim to perform Exploratory Data Analysis to uncover insights that guide music production for better audience appeal.



Presented by Subham Singh



Initial Analysis and Data Description :

Data columns (total 22 columns):

#	Column	Non-Null Count	Dtype
0	track_id	49312	non-null object
1	track_name	49312	non-null object
2	artist_name	49312	non-null object
3	year	49312	non-null int64
4	popularity	49312	non-null int64
5	artwork_url	49312	non-null object
6	album_name	49312	non-null object
7	acousticness	49312	non-null float64
8	danceability	49312	non-null float64
9	duration_ms	49312	non-null float64
10	energy	49312	non-null float64
11	instrumentalness	49312	non-null float64
12	key	49312	non-null float64
13	liveness	49312	non-null float64
14	loudness	49312	non-null float64
15	mode	49312	non-null float64
16	speechiness	49312	non-null float64
17	tempo	49312	non-null float64
18	time_signature	49312	non-null float64
19	valence	49312	non-null float64
20	track_url	49312	non-null object
21	language	49312	non-null object

dtypes: float64(13), int64(2), object(7)

Numerical Columns		Categorical Columns									
		year popularity acousticness danceability duration_ms energy instrumentalness key liveness loudness mode speechiness tempo time_signature valence									

Univariate Analysis: Statistical analysis of numerical variables

```
→ count    49312.000000  
mean      16.175333  
std       19.072456  
min       0.000000  
25%      0.000000  
50%      8.000000  
75%      28.000000  
max      93.000000  
Name: popularity, dtype: float64
```

```
→ count    49312.000000  
mean      3.889885  
std       1.708285  
min       0.083333  
25%      3.147546  
50%      3.811933  
75%      4.610929  
max      76.358050  
Name: duration_min, dtype: float64
```

```
→ count    49312.000000  
mean      -79.953488  
std       2662.992208  
min      -100000.000000  
25%      -10.817250  
50%      -7.315000  
75%      -5.290750  
max      0.894000  
Name: loudness, dtype: float64
```

```
→ count    49312.000000  
mean      0.352142  
std       0.319976  
min      -1.000000  
25%      0.055000  
50%      0.260000  
75%      0.624000  
max      0.996000  
Name: acousticness, dtype: float64
```

```
→ count    49312.000000  
mean      117.995563  
std       28.602312  
min      -1.000000  
25%      95.999000  
50%      118.197000  
75%      135.029000  
max      232.198000  
Name: tempo, dtype: float64
```

```
→ count    49312.000000  
mean      0.601348  
std       0.252853  
min      -1.000000  
25%      0.434000  
50%      0.645000  
75%      0.807000  
max      1.000000  
Name: energy, dtype: float64
```

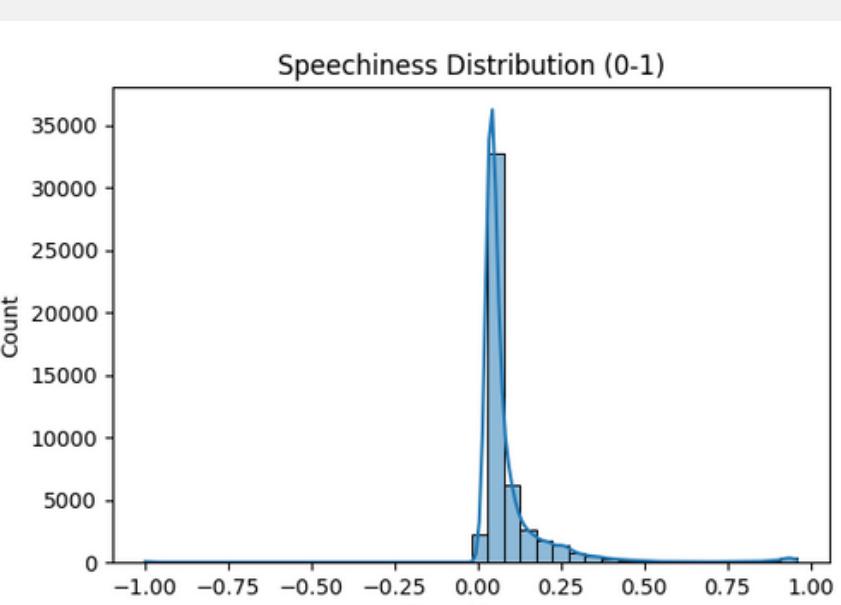
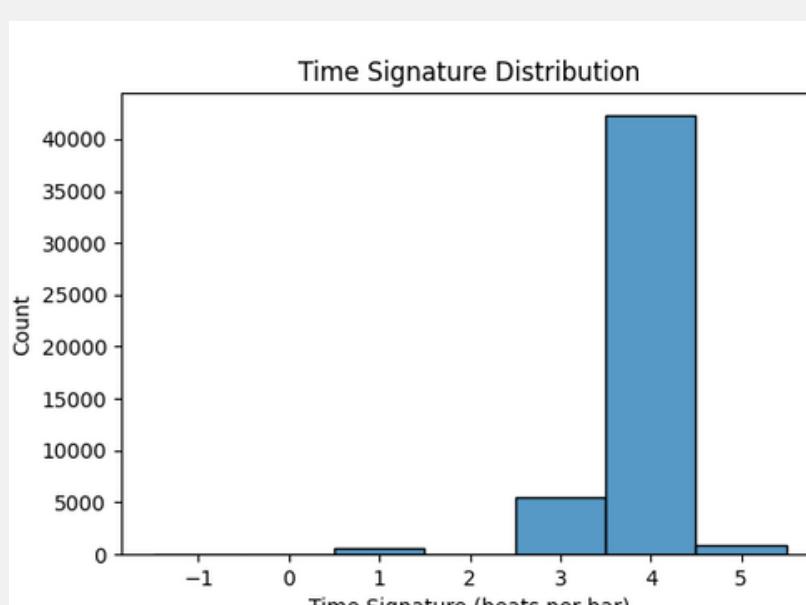
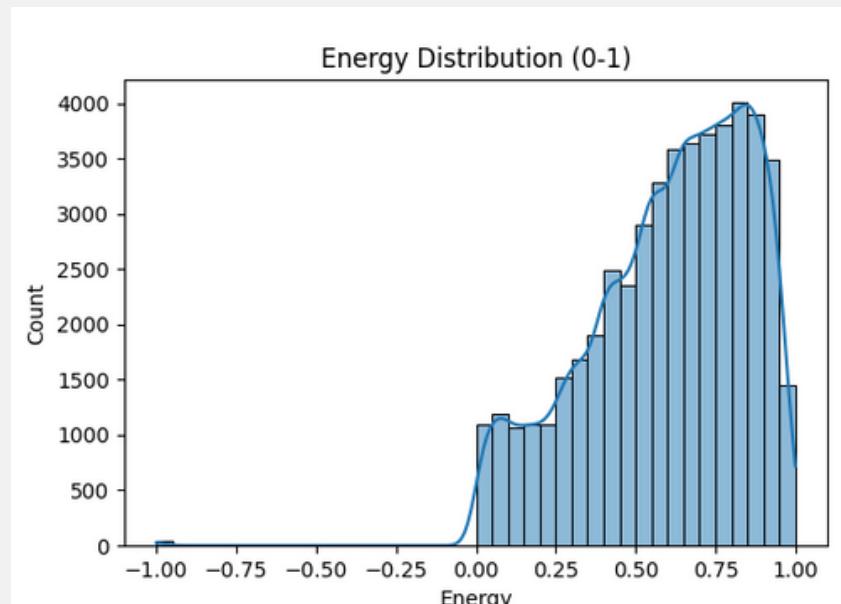
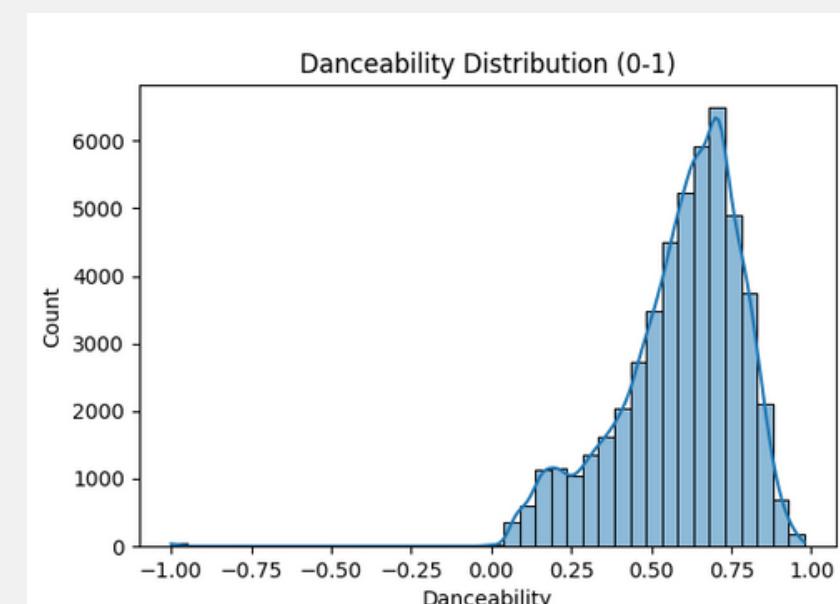
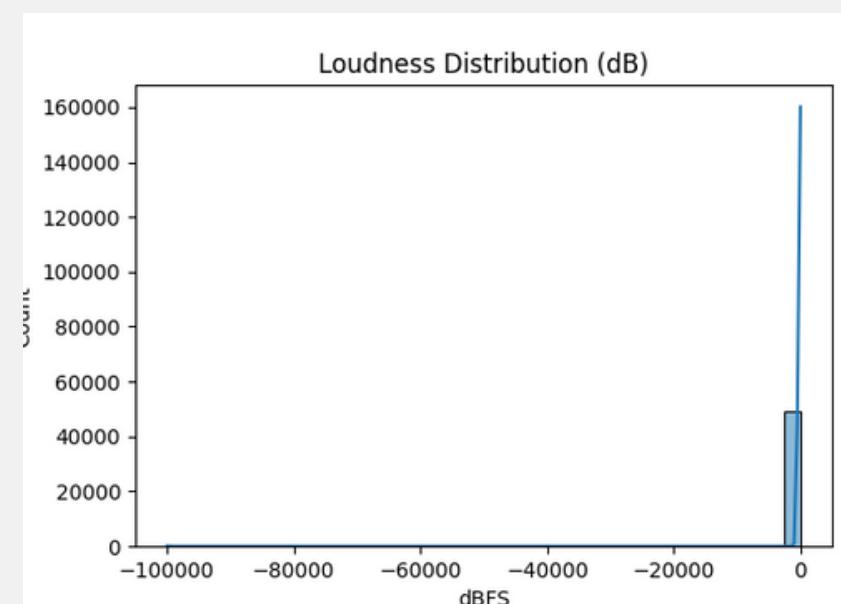
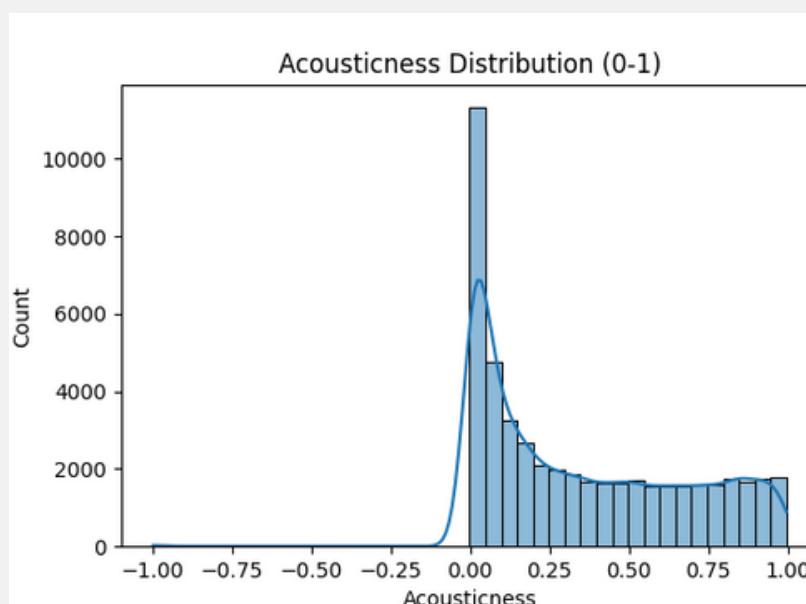
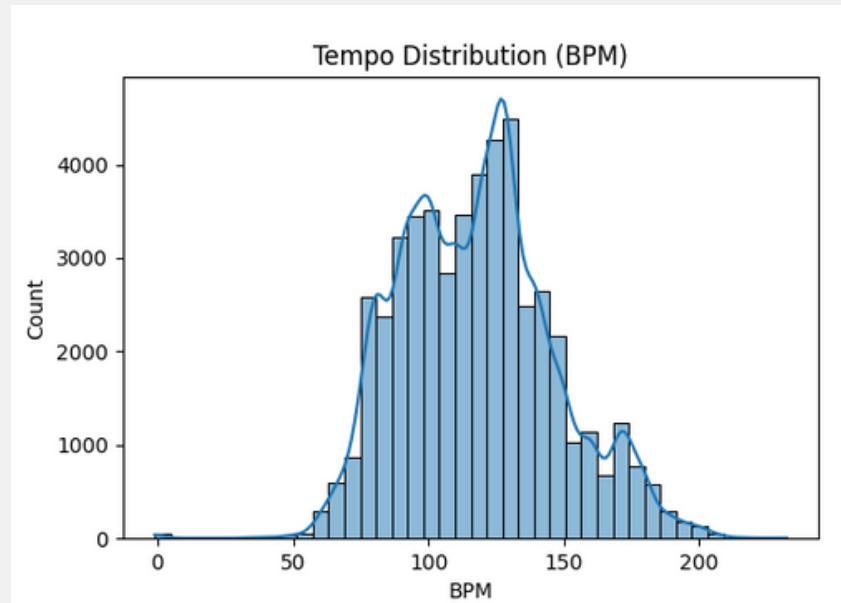
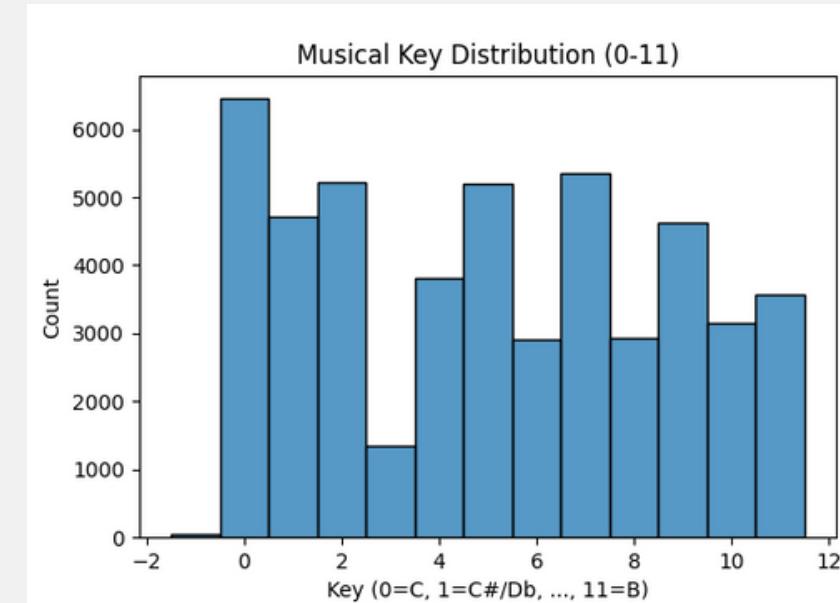
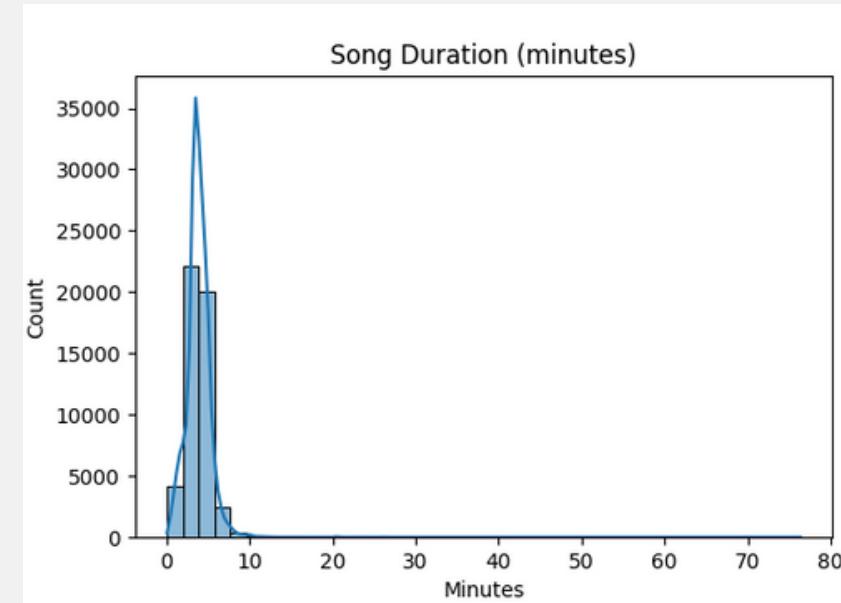
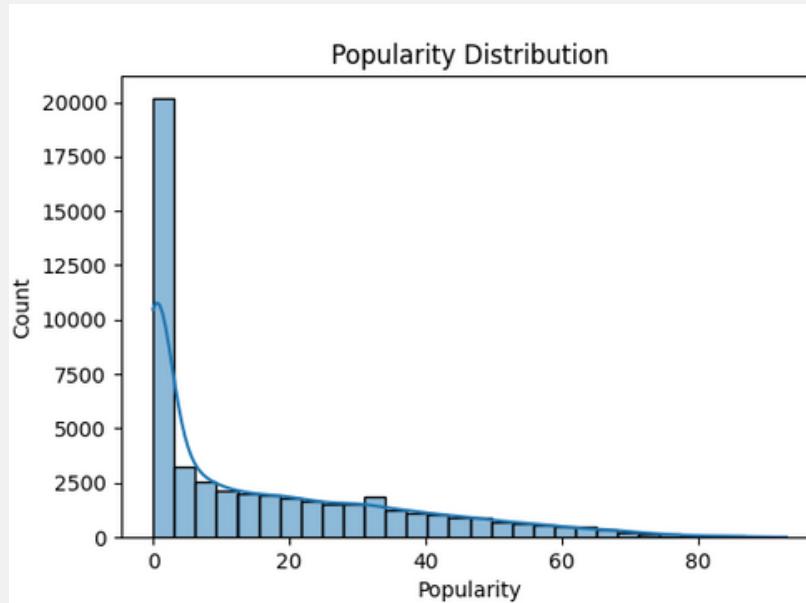
```
→ count    49312.000000  
mean      0.590508  
std       0.190893  
min      -1.000000  
25%      0.488000  
50%      0.627000  
75%      0.726000  
max      0.979000  
Name: danceability, dtype: float64
```

```
→ key  
-1      35  
0       6463  
1       4707  
2       5219  
3       1347  
4       3809  
5       5197  
6       2898  
7       5360  
8       2932  
9       4622  
10      3156  
Name: count, dtype: int64
```

```
→ time_signature  
-1      35  
0       16  
1       547  
3       5555  
4       42347  
5       812  
Name: count, dtype: int64
```

```
→ count    49312.000000  
mean      0.089369  
std       0.122405  
min      -1.000000  
25%      0.037000  
50%      0.048600  
75%      0.089100  
max      0.959000  
Name: speechiness, dtype: float64
```

Univariate Analysis: Visualization of Variables using Histogram



Key Insights and Implementations:-

Popularity

- Insight: Most tracks land in the middle—not many big hits.
- Do: Combine multiple strong traits (catchy hook, energy, modern mix) and push releases strategically (playlists, timing, marketing).
- Song popularity follows a **power-law distribution** — a tiny fraction of tracks get most of the attention, while the majority receive very little.
- Instead of spreading effort evenly across all songs, focus production, marketing, and promotion on the small group of high-potential tracks. These are the ones most likely to enter the “hit” category and drive most of the impact.

Duration

- Insight: Most songs are ~1.6–3.6 minutes.
- Do: Aim for 2:20–3:10. Get to the hook within 30–40 seconds. Avoid long intros/outros.

Key

- Insight: A few tonal centers dominate.
- Do: **Use vocalist-friendly keys** (E ♭ /E/F/G major and minors). Add color with parallel major/minor changes.

Tempo

- Insight: Clusters around 110–170 BPM (esp. 120–130 & 160–170 double-time).
- Do:
 - 118–128 BPM = dance/pop sweet spot
 - 90–105 or 70–85 halftime = hip-hop/R&B
 - Use halftime feel if going fast.

Acousticness

- Insight: Most tracks aren't fully acoustic.
- Do: Use acoustic elements as layers. For acoustic-led songs, offset with modern drums, hooks, and processing.

Loudness

- Insight: Tracks sit around -6 to -10 LUFS.
- Do: Master at -9 to -7 LUFS, keep true peak under -1 dBTP. Use saturation and parallel compression smartly.

Danceability

- Insight: Mid–high danceability is common.
- Do: Tighten kick/bass, use sidechain, and build rhythmic hooks. Refresh sections every 8–16 bars.

Energy

- Insight: Most songs sit at moderate–high energy.
- Do: Build energy arcs—low intro, lift in pre, max at chorus, drop/post-chorus payoff.

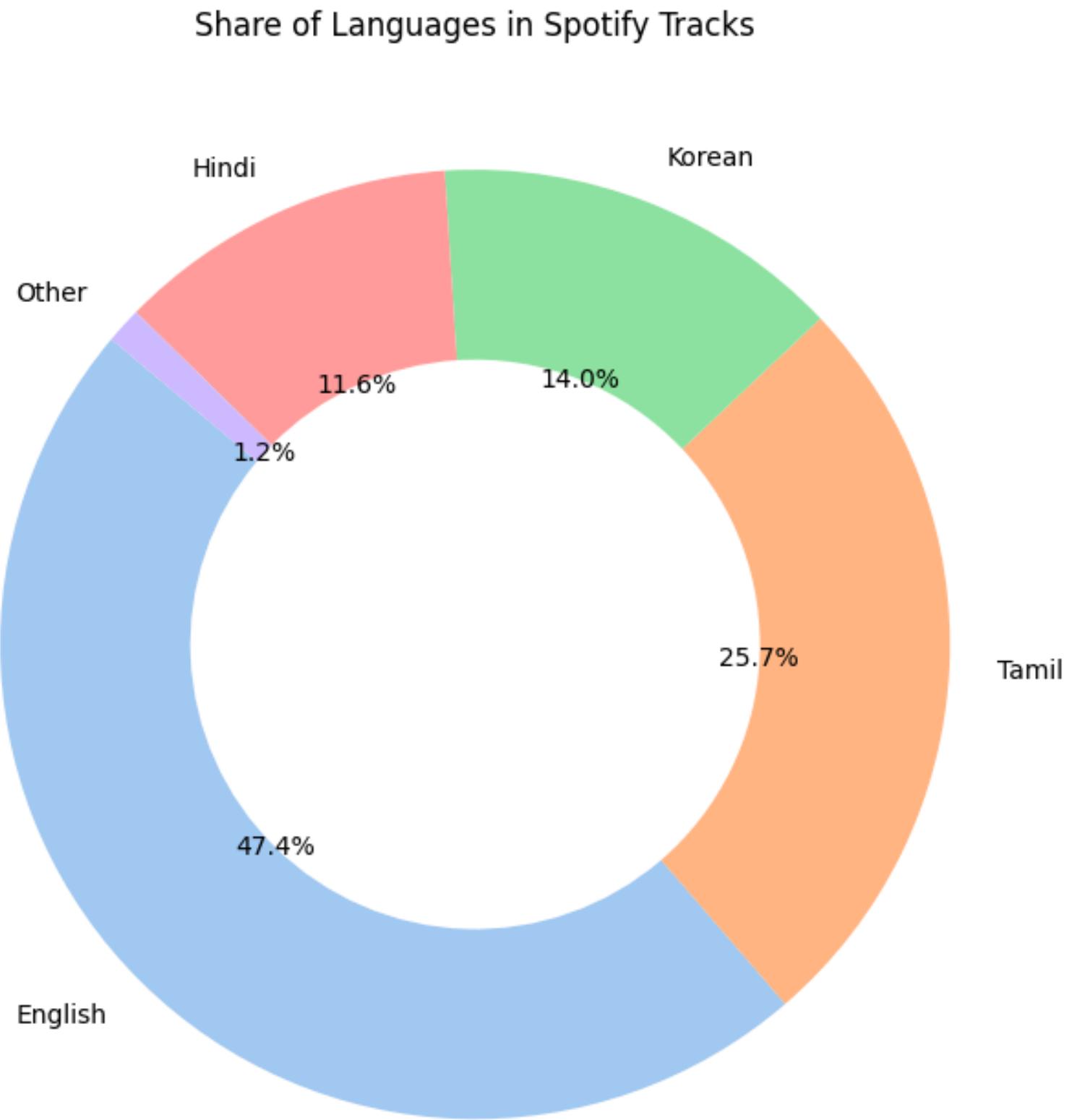
Time Signature

- Insight: Mostly 4/4.
- Do: Stick to 4/4 for mainstream appeal. If experimenting, keep core sections in 4/4.

Speechiness

- Insight: Mostly low–moderate; some rap elements present.
- Do: For pop, keep vocals melodic. For rap-heavy tracks, add memorable sung hooks.

Univariate Analysis: Categorical Variables



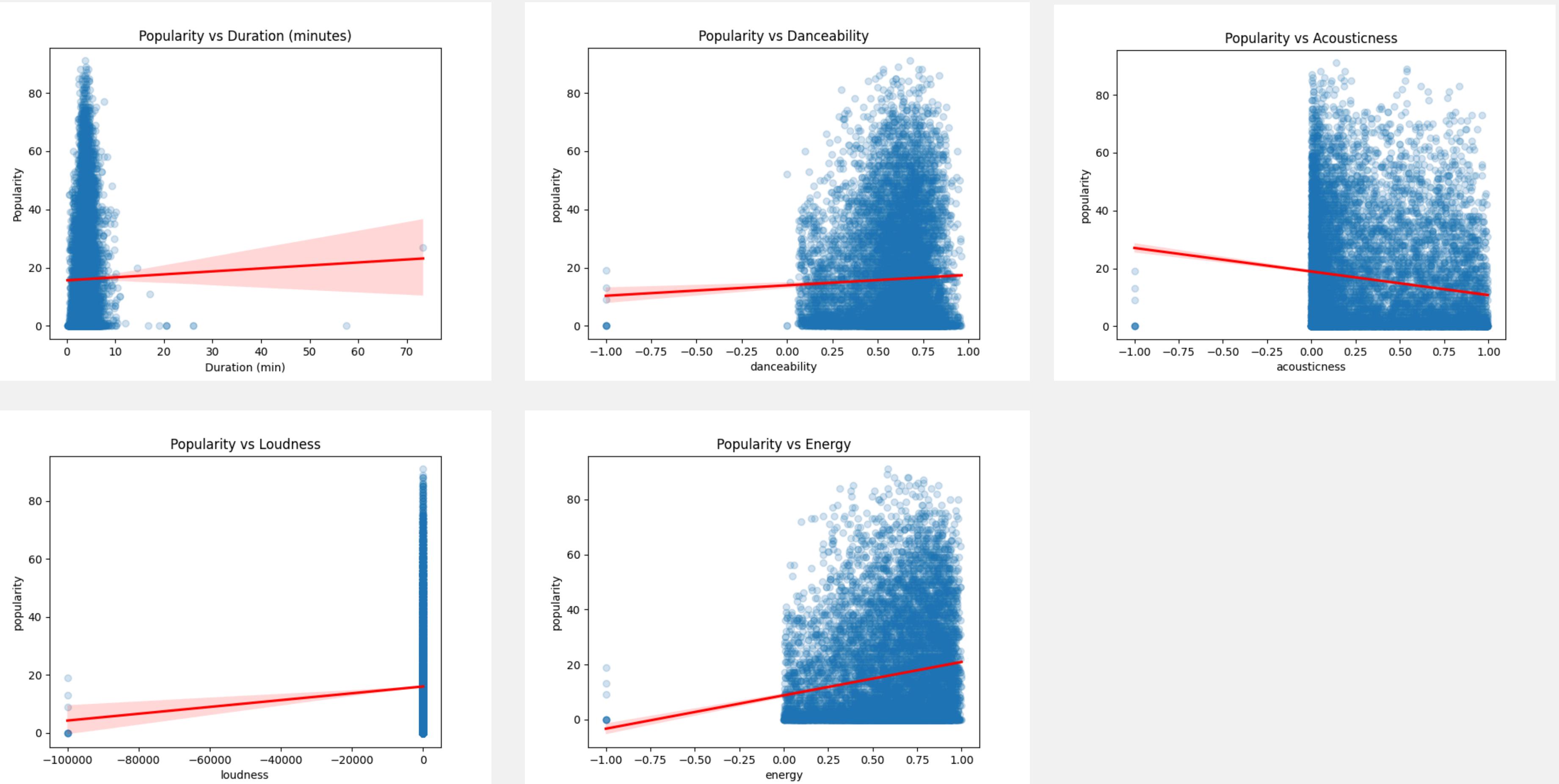
Key Insights:

- **English dominates** (47.4% of tracks), indicating a saturated market.
- Tamil (25.7%) and Hindi (11.6%) together make up 37%+, showing **strong regional content coverage**.
- Korean songs (14%) have the **highest mean popularity**, suggesting high engagement despite lower volume.
- Other languages are rare (1.2%).

Implementations for Mixing Engineer/Music Producer:

- English: Focus on niche or standout tracks to compete in a crowded market.
- Tamil & Hindi: **Maintain high-quality output** to serve a large, active regional audience.
- Korean: **Adopt global production standards** to improve appeal across all languages.
- Strategic Volume: Limit high-volume English releases for impact, maintain robust supply of Tamil tracks

Bivariate Analysis: Numerical vs Numerical Variable



• Key Insights

1. Duration → Popularity : No meaningful effect. Focus on feel and arrangement, not length.
2. Danceability → Popularity: Very weak positive link. **Groove matters more than duration.**
3. Energy → Popularity: **Strongest positive relationship** among all features.
4. Loudness → Popularity: Little to no impact once loudness is in modern standards.
5. Acousticness → Popularity: Light negative trend. **More electronic/synthetic elements perform better.**

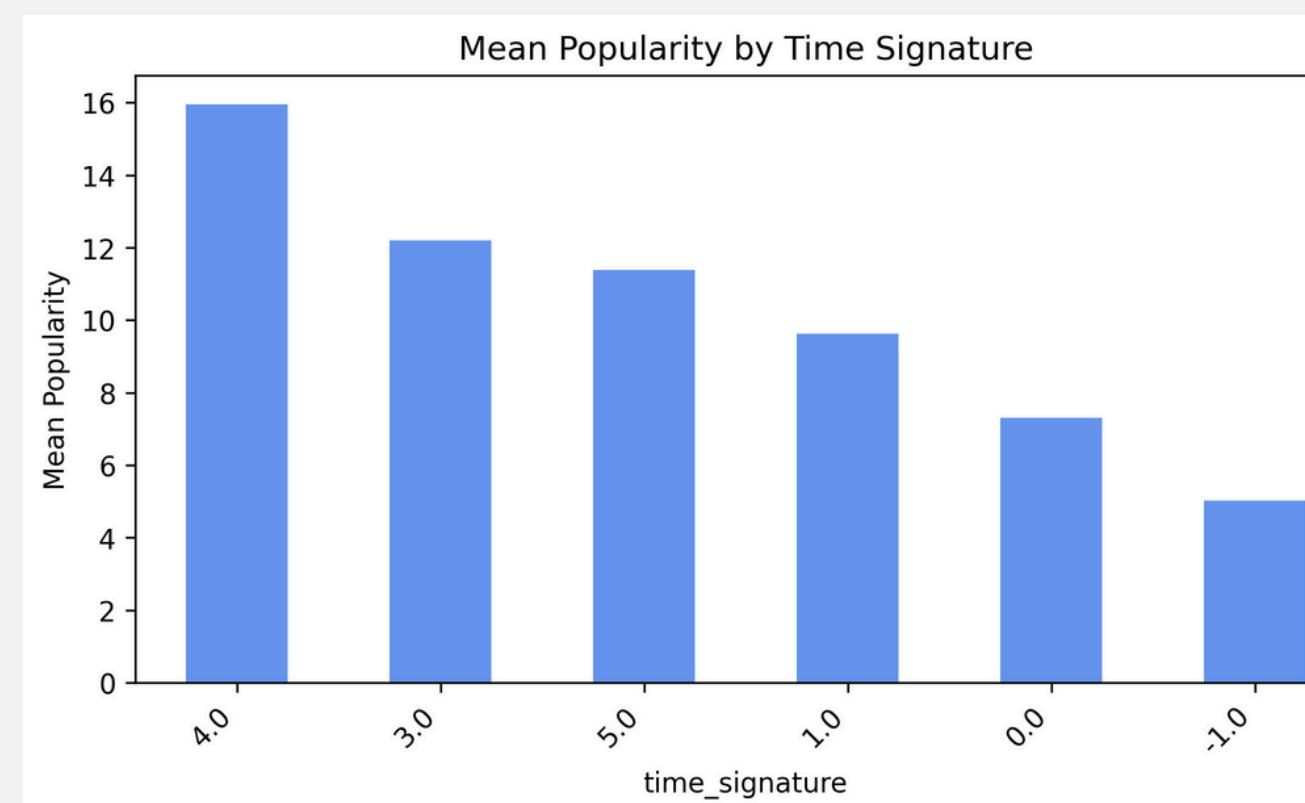
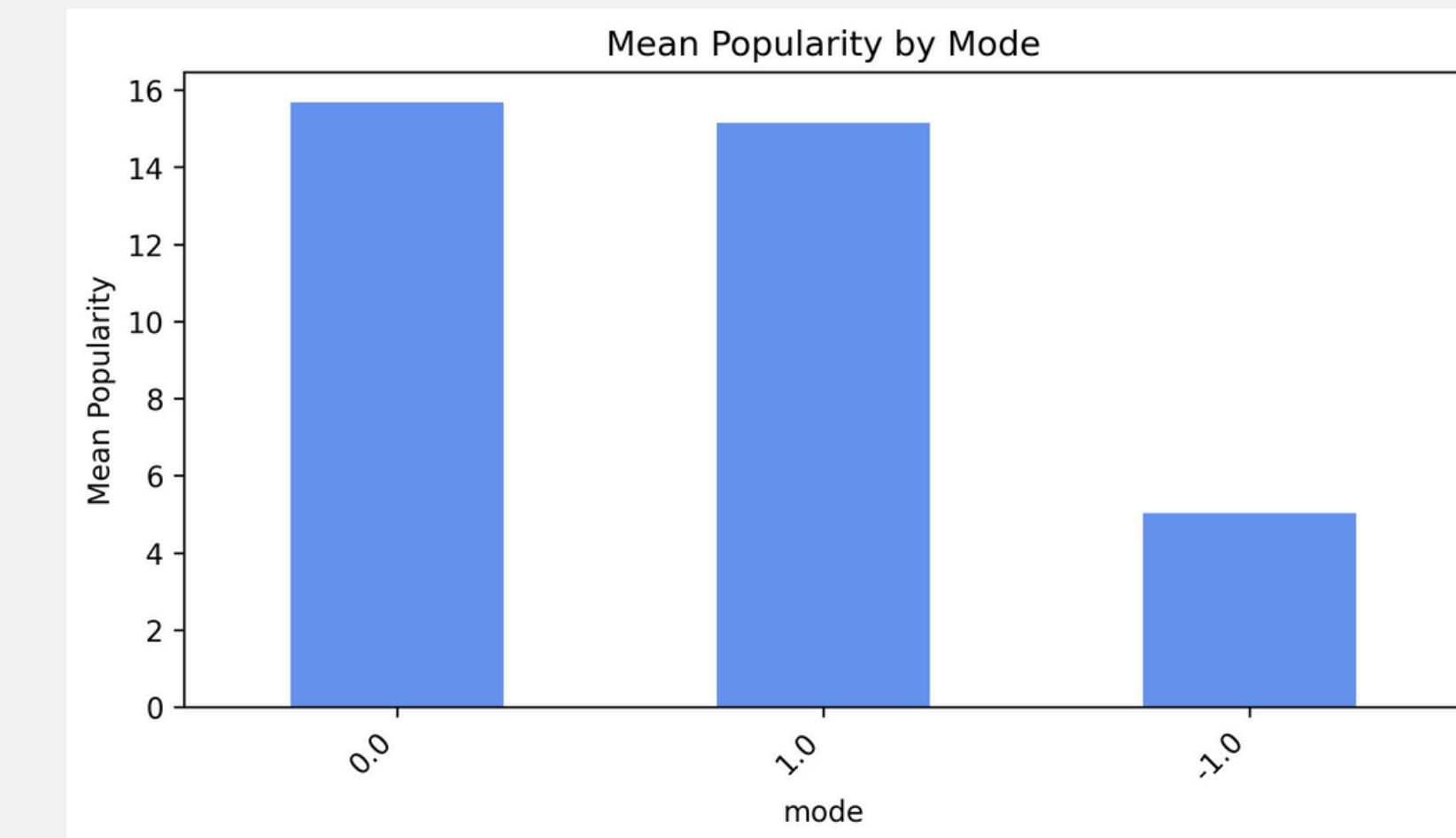
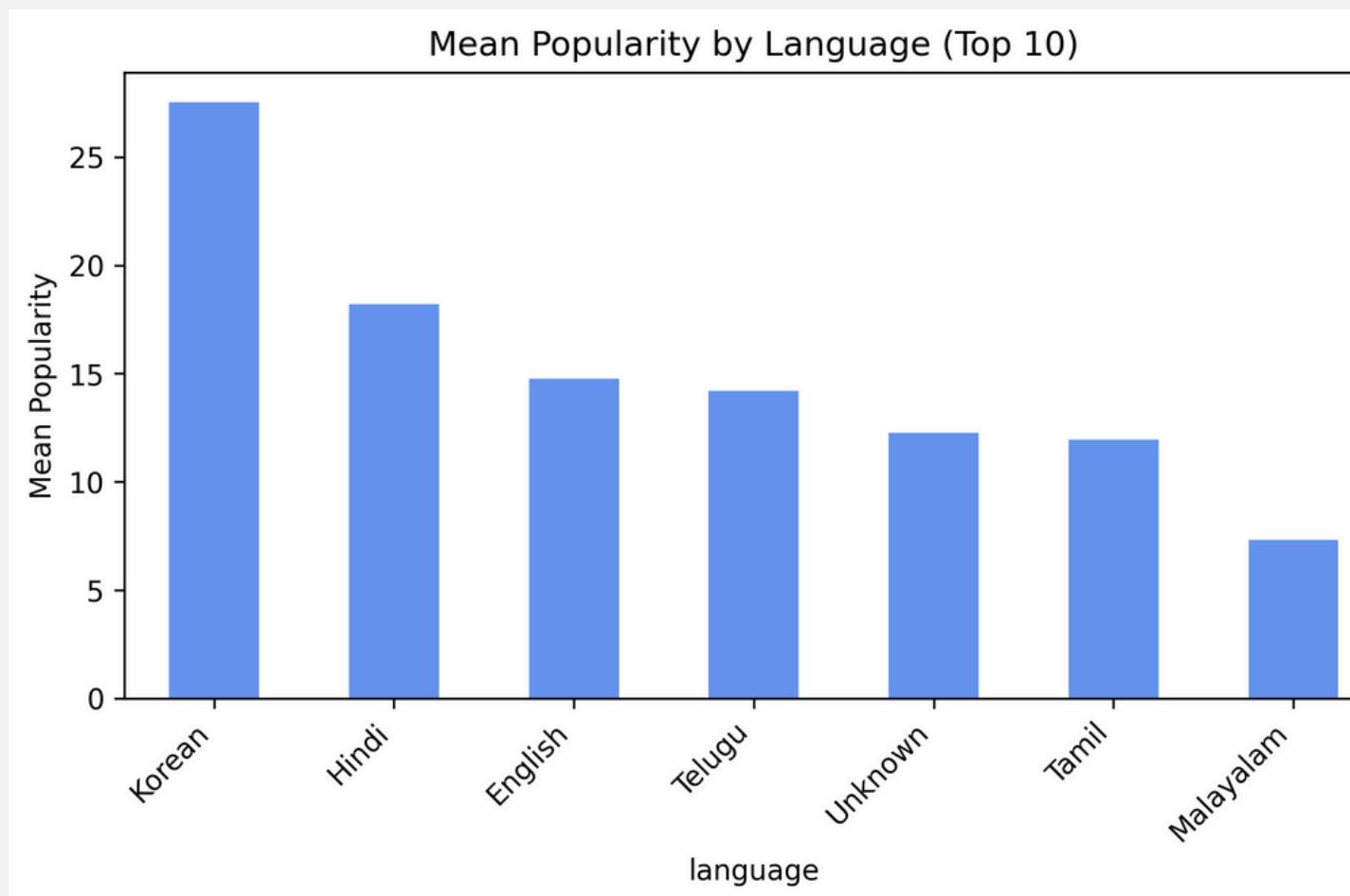
• Section-Level Mixing Blueprint:

- Verses – Moderate energy, clear groove, defined kick–bass.
- Pre-Chorus – Build with risers, extra percussion, widening.
- Chorus – Maximum punch: parallel drums, stacked vocals, strong low-end control.
- Bridge/Break – Contrast, then return bigger for final chorus.

• Practical Takeaways for Mixing Engineer:

1. Duration → Focus on Engagement
 - Don't aim for a specific runtime.
 - Use early hooks, minimal dead space, tight intros/outros.
2. Danceability → Groove First
 - Tempo: ~100–130 BPM (or 70–90 halftime).
 - Clear kick–bass balance and rhythmic clarity.
 - Mix tip: Tight low end, transient clarity, reduce muddy low-mids.
3. Energy → Controlled Intensity
 - Use bright leads, layered drums, parallel compression.
 - Add energy lifts through arrangement (fills, risers, pre-chorus builds).
 - Keep vocals present (2–6 kHz), use saturation tastefully.
4. Loudness → Competitive, Not Crushed
 - Avoid over-limiting for volume.
 - Aim for -9 to -12 LUFS, true peaks under -1 dBTP.
 - Punch and clarity > "loud."
5. Acousticness → Hybrid or Synthetic Edge
 - For mainstream appeal, use electronic or layered textures.
 - In acoustic genres, add polish: light synths, electronic drums, subtle saturation.

Bivariate Analysis: Numerical vs Categorical Variable

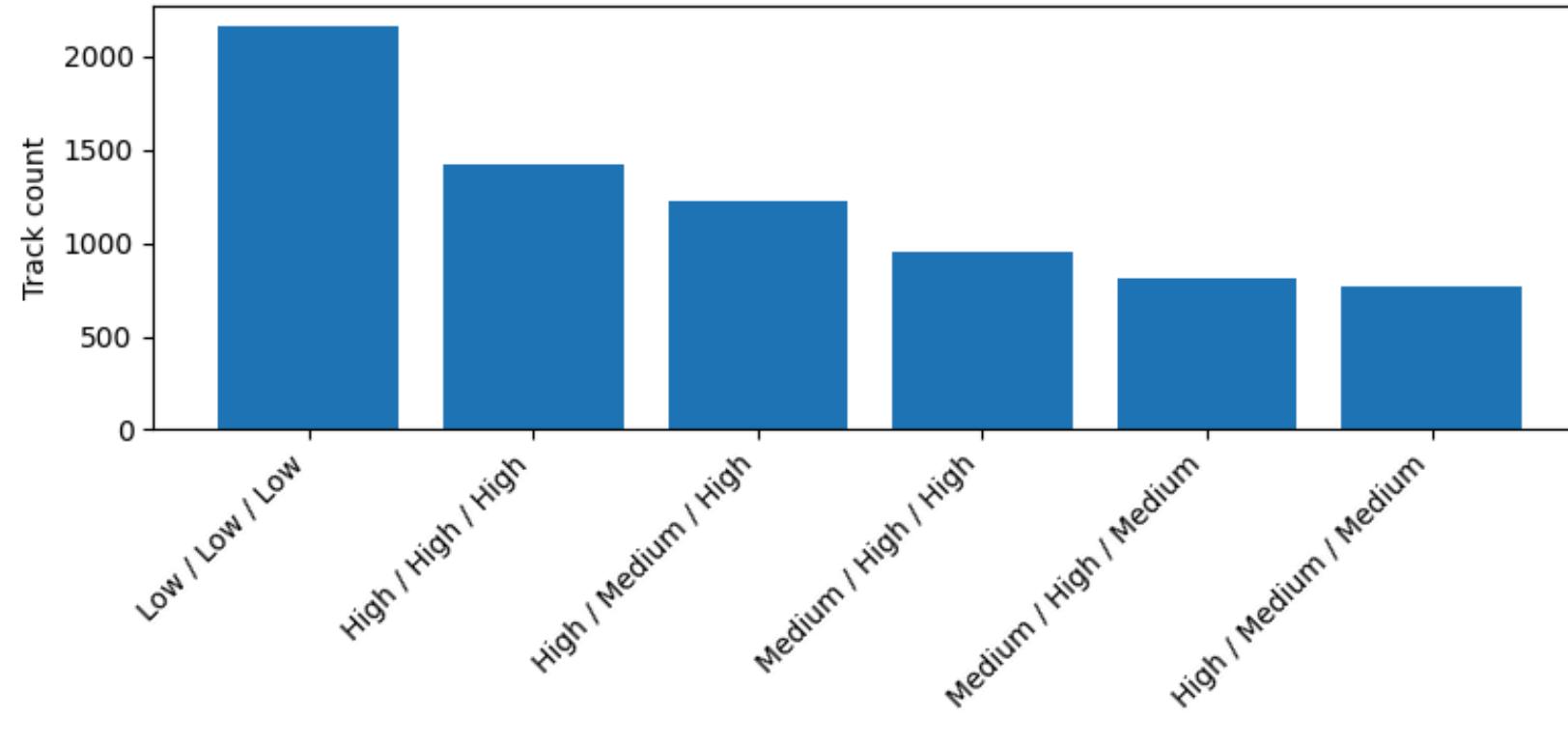


Key Insights and Implementations for Mixing Engineer:-

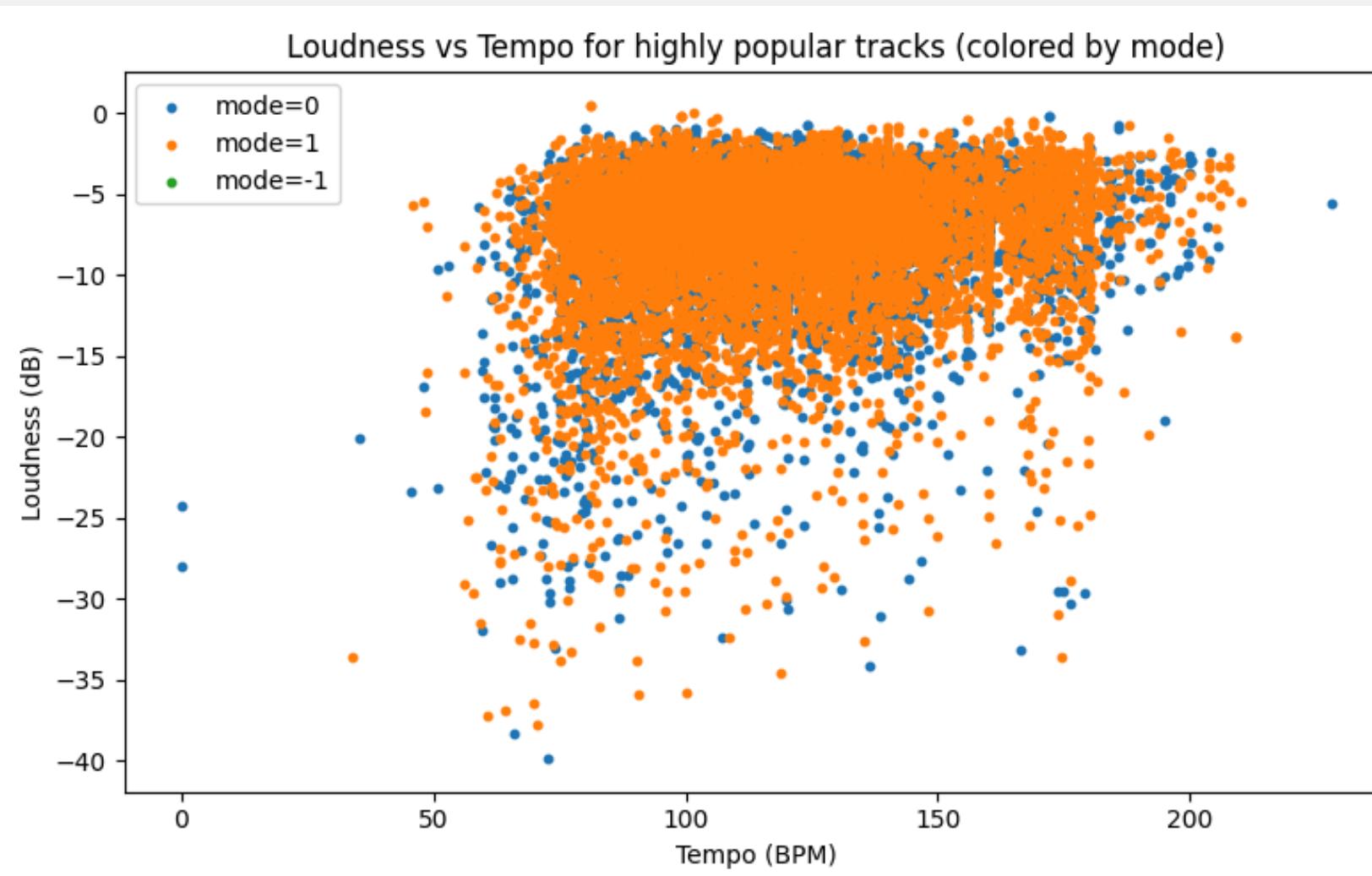
- The 4.0 (4/4) time signature is significantly more popular than all others.
- **Prioritize 4/4 Rhythm:** Focus production efforts almost exclusively on 4/4 time to maximize broad listener appeal and familiarity.
- Korean music shows the highest **mean popularity**, indicating a powerful global trend (K-Pop). Hindi, Telugu, and Tamil show high mean popularity, often above English.
- **Study K-Pop Production:** Analyze the sound design, mixing, and mastering of popular Korean tracks. Apply those global-standard techniques to your new releases.
- 0.0 (Minor) and 1.0 (Major) modes have nearly identical, high mean popularity.
- Mode is Secondary: **Do not let mode constrain your creative choice.** Focus on the emotional narrative and energy; the Major/Minor distinction is not a primary popularity driver.

Multivariate Analysis:

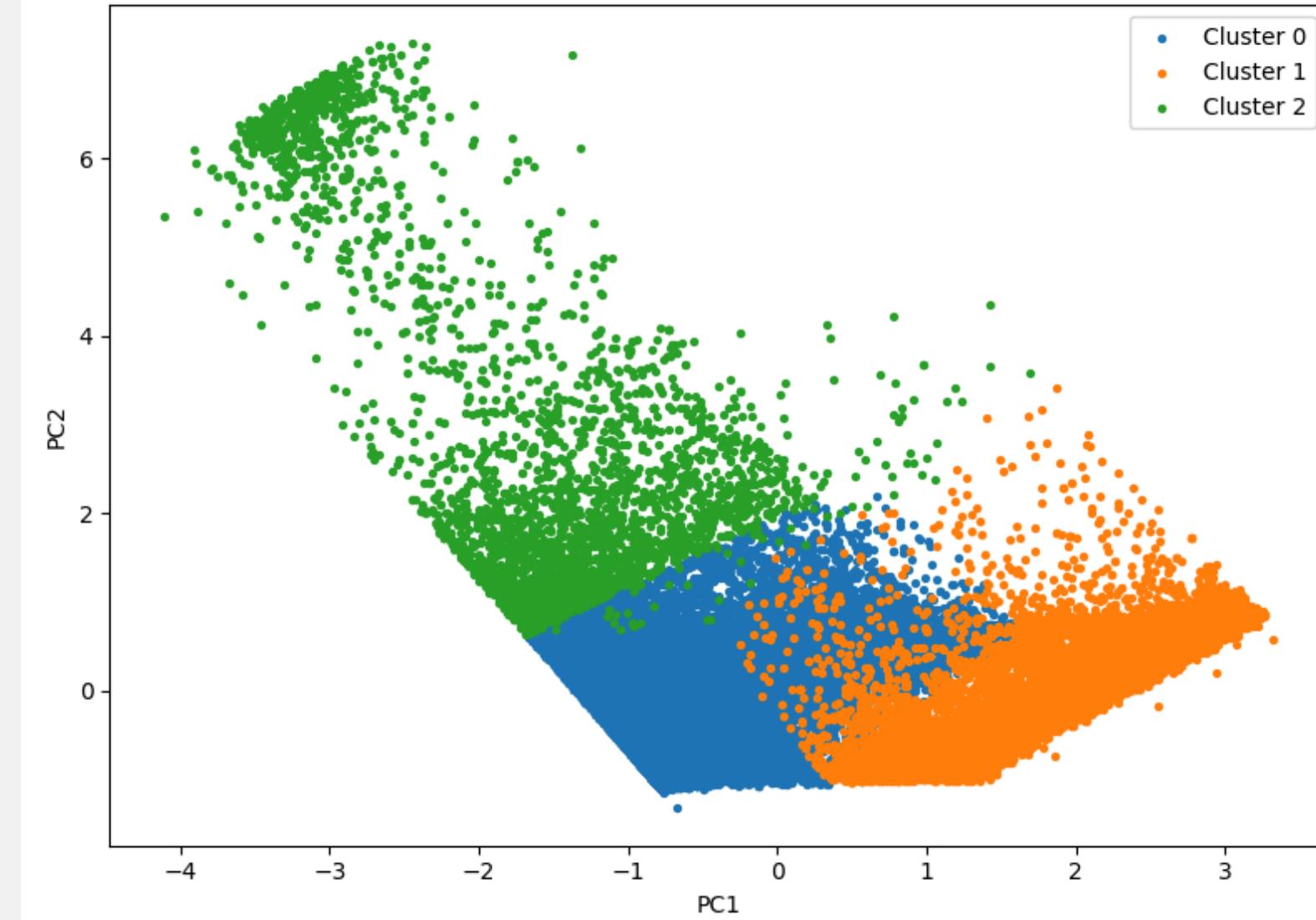
Top 6 danceability-energy-valence combos in top popularity quartile



Loudness vs Tempo for highly popular tracks (colored by mode)



PCA projection of acoustic/instrumental/speechiness clusters



Key Insights and Implementations for Mixing Engineer:-

1. Danceability + Energy + Valence

- Insight: Popular songs lean toward medium–high danceability, high energy, moderate positivity.
- Implementation: **Focus on groovy, rhythmic beats.** Keep energy high with strong percussion, bright synths, and layered instruments. Maintain emotional balance: not overly happy, but not too dark either.

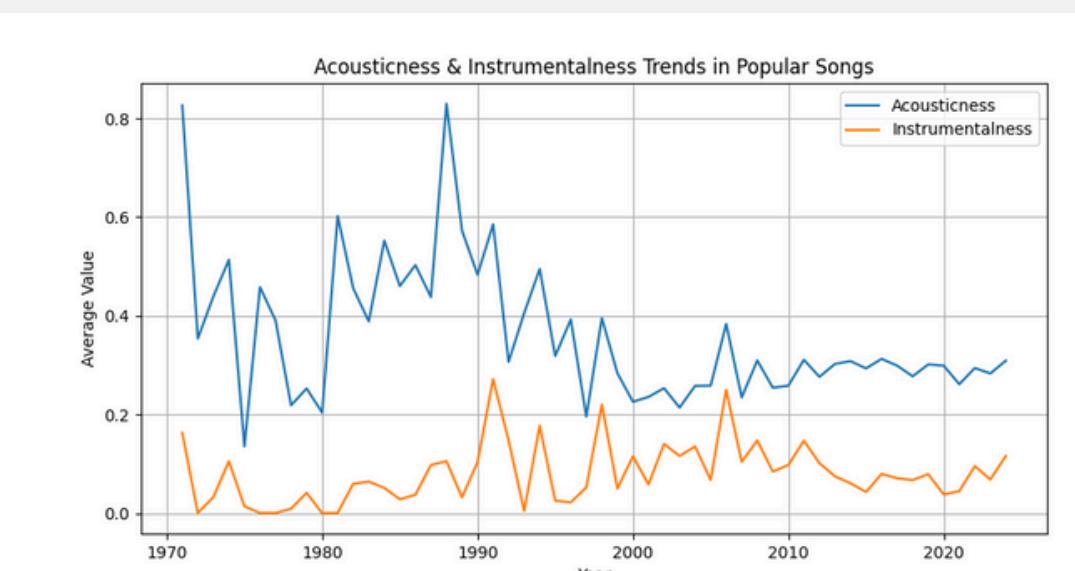
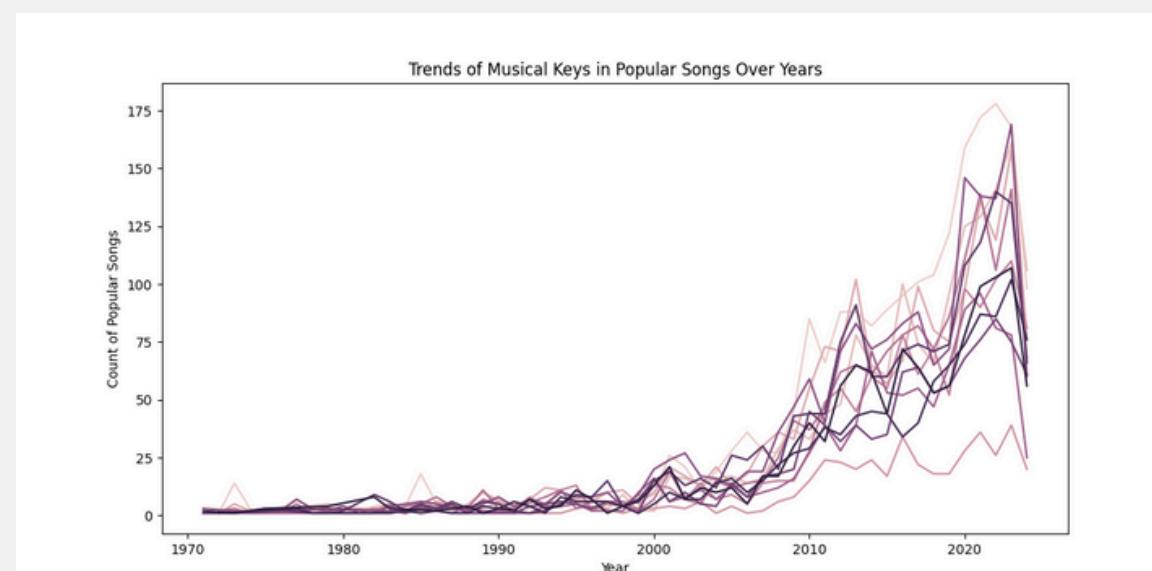
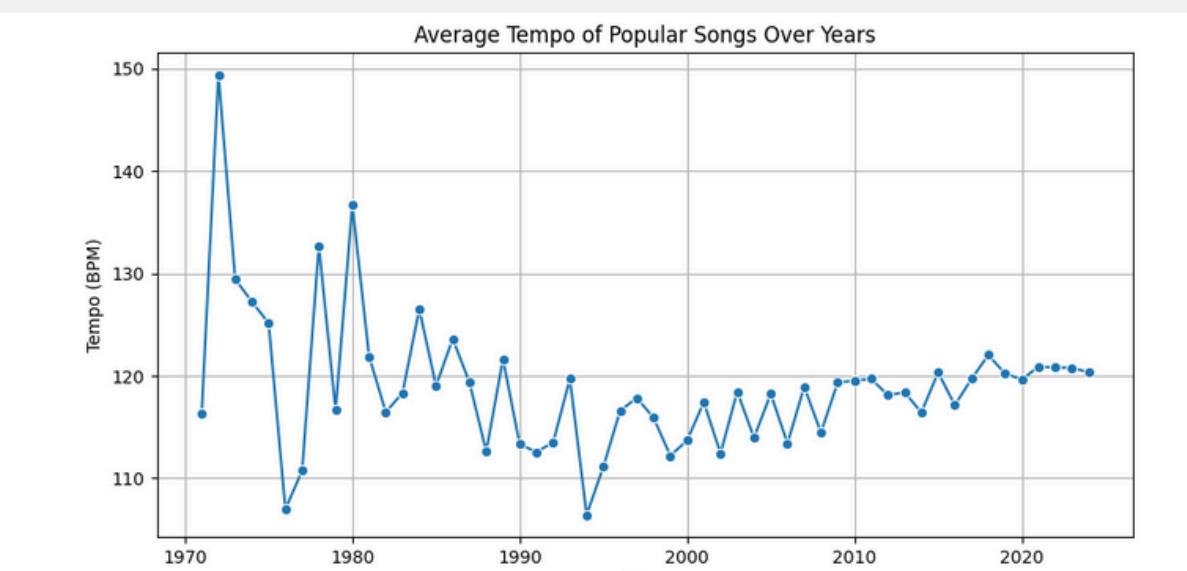
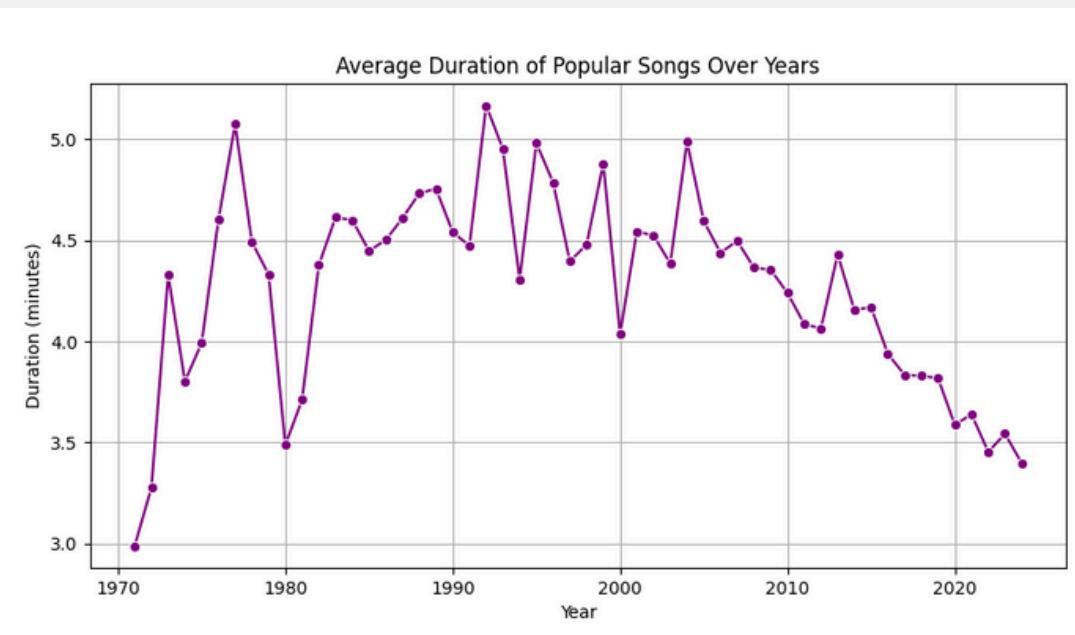
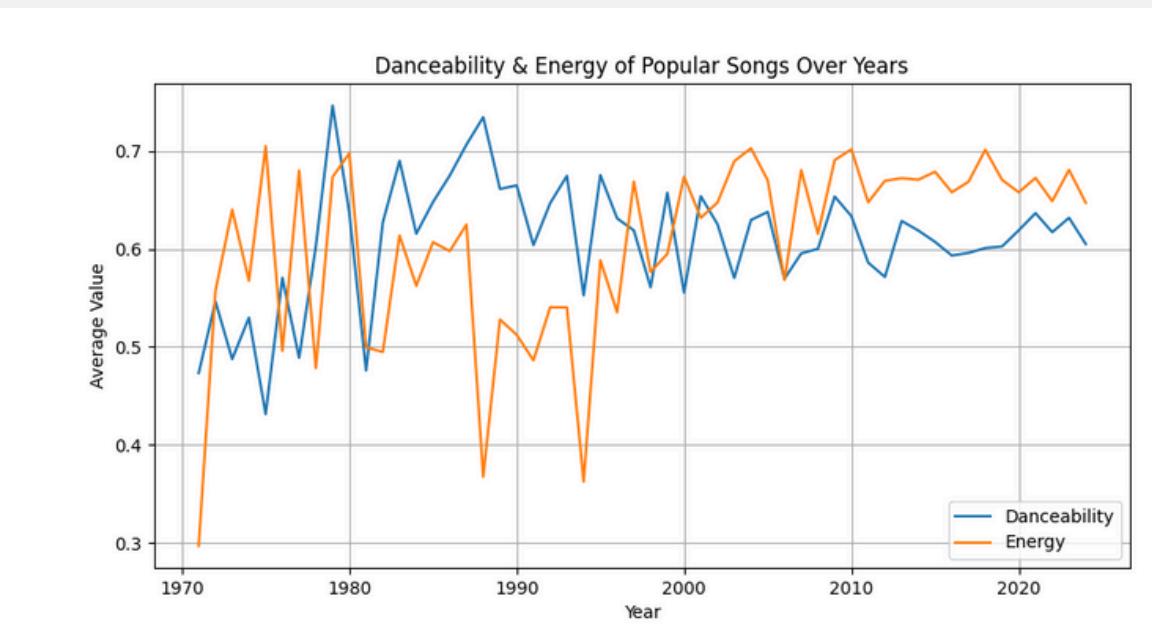
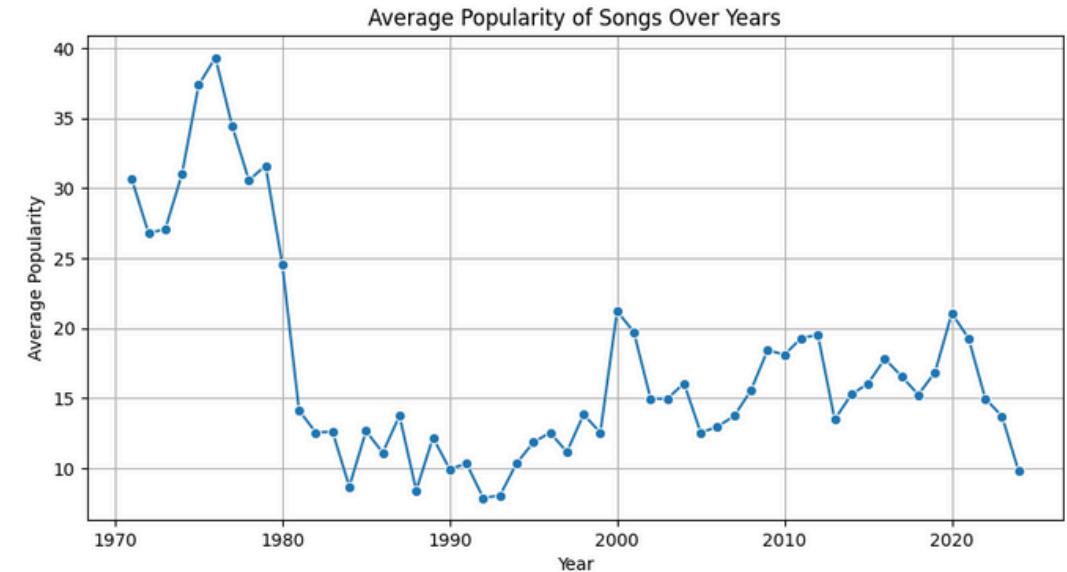
2. Sound Profiles (Acousticness, Instrumentalness, Speechiness)

- Insight: Popular tracks cluster around urban/hip-hop (high speechiness) and EDM/electronic (low acousticness, high instrumental layers).
- Implementation: Incorporate rap-style vocals, rhythmic spoken elements, or vocal chops. Use EDM textures (drops, synth layers) to appeal to mainstream listeners. Avoid purely acoustic/stripped-down mixes unless aiming for niche/indie appeal.

3. Mix & Technical Recipe (Loudness, Tempo, Mode)

- Insight: Loudness: ~-6 to -7 dB RMS.
Tempo: 100–130 BPM is most common.
Mode: Majority in major key.
- Implementation: Mix to -6 to -7 dB for streaming competitiveness (master with controlled limiting). Write/produce tracks in the 110–125 BPM range for crossover between pop, EDM, and hip-hop. Favor major keys for broad appeal, but sprinkle minor sections for mood contrast.

Timeseries Analysis:



Key Insights and Implementations for Mixing Engineer:-

1. Popularity Over Time

- *Newer songs (streaming era) are consistently more popular.
- Implementation: Focus on current streaming trends when producing – optimize for playlists and viral potential.

2. Danceability & Energy

- Danceability has increased, energy stays high.
- Implementation: Prioritize groovy beats and energetic arrangements for mainstream appeal.
-

3. Keys & Tempo

- Mid-tempo (100–130 BPM) dominates; keys like C, G, D remain common.
- Implementation: Write tracks in dance/pop-friendly tempos and stick to familiar keys for accessibility.

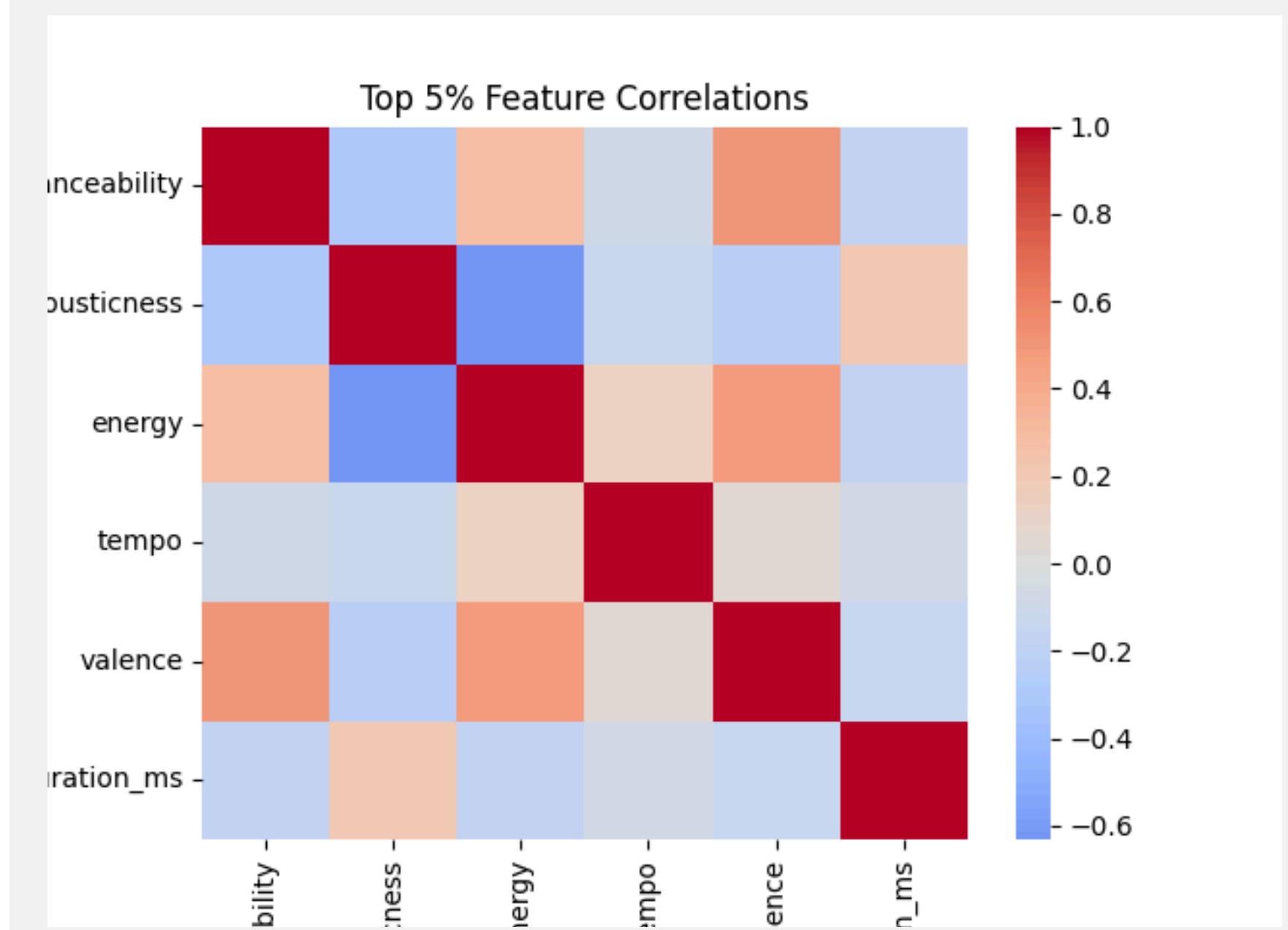
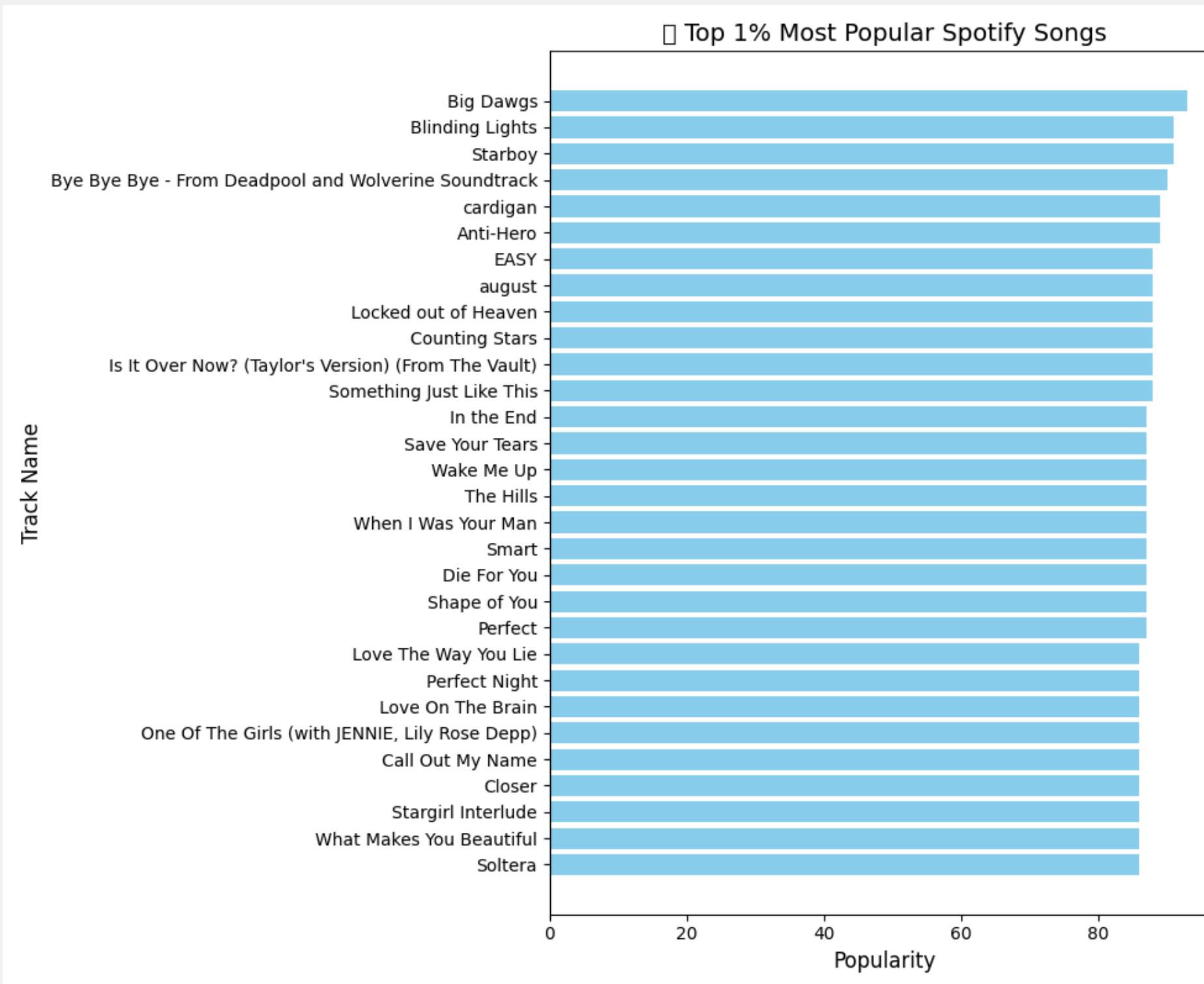
4. Duration

- Songs are getting shorter (~3 min now vs 4–5 min before).
- Implementation: Keep songs short, punchy, and replayable to maximize streams.
-

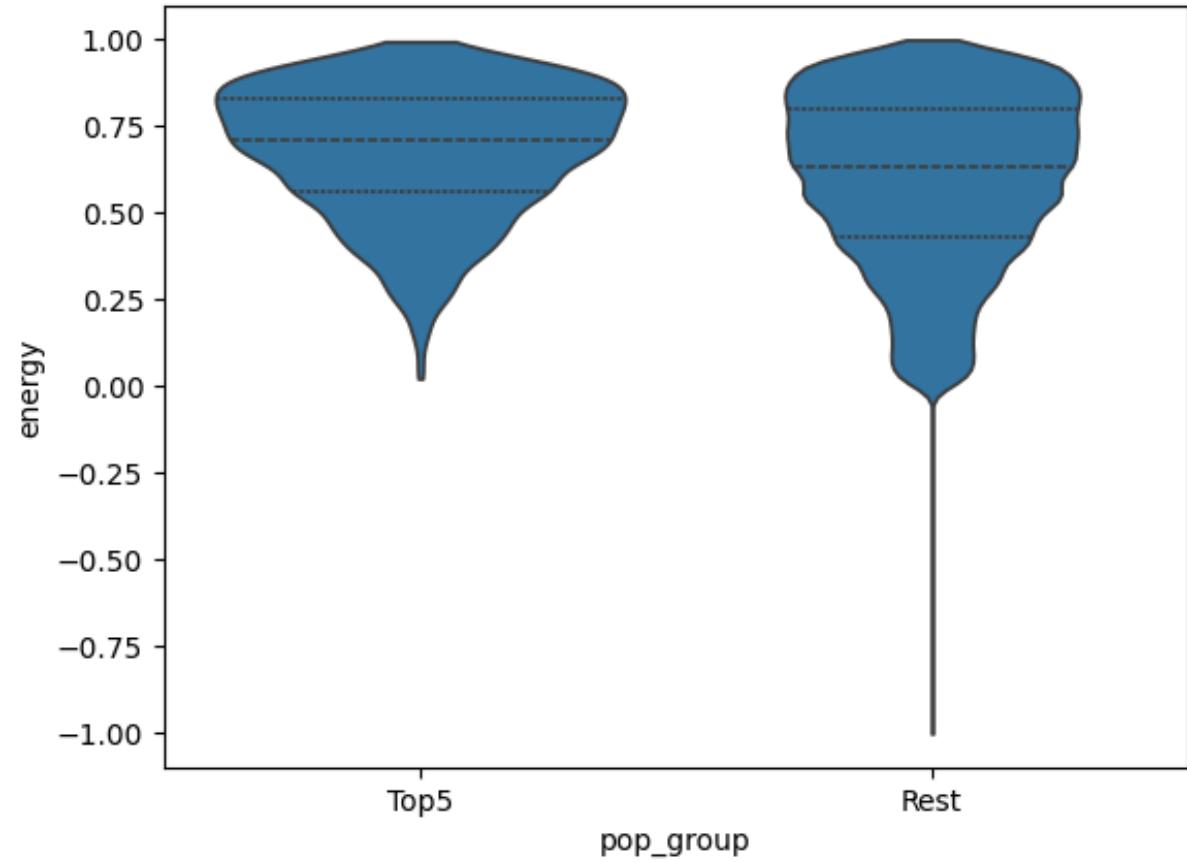
5. Acousticness & Instrumentalness

- Acoustic/Instrumental songs have declined; vocals dominate.
- Implementation: Make vocals central and strong, with modern electronic/urban production layers.

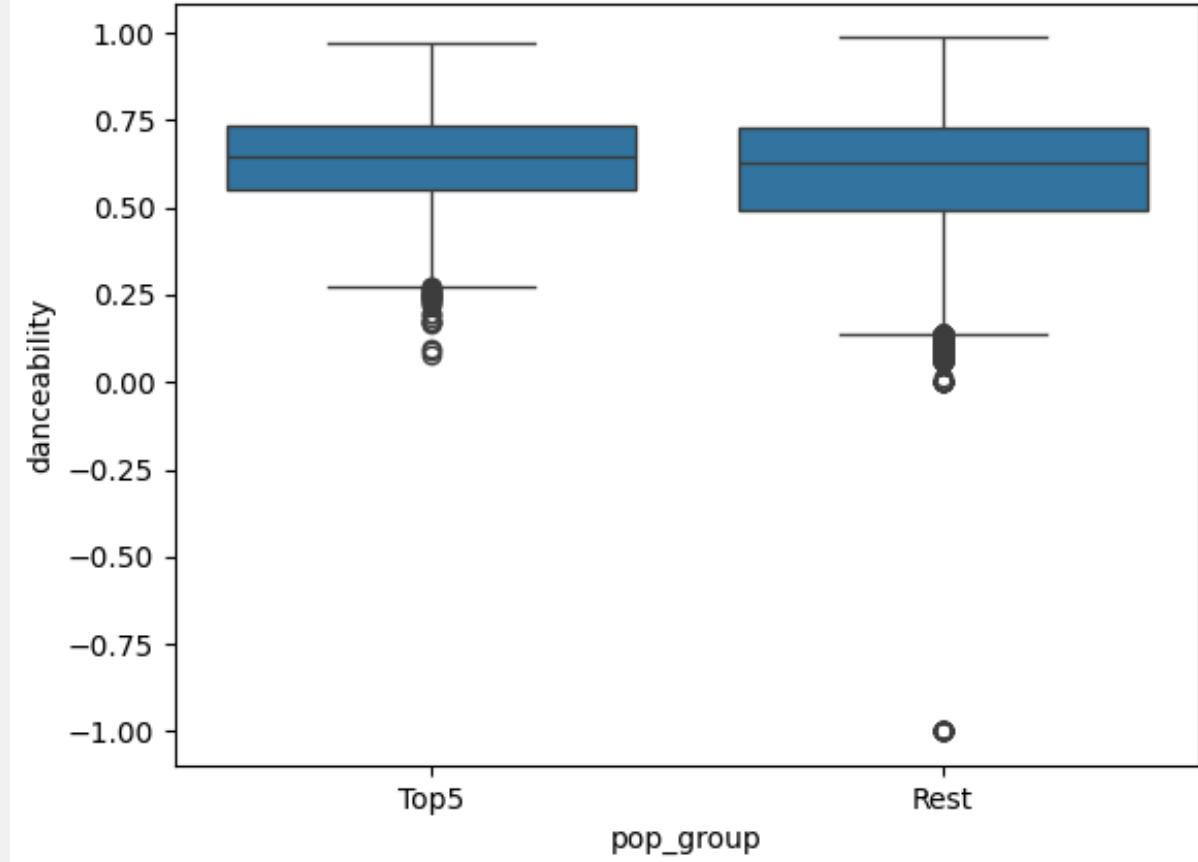
Outlier Analysis:



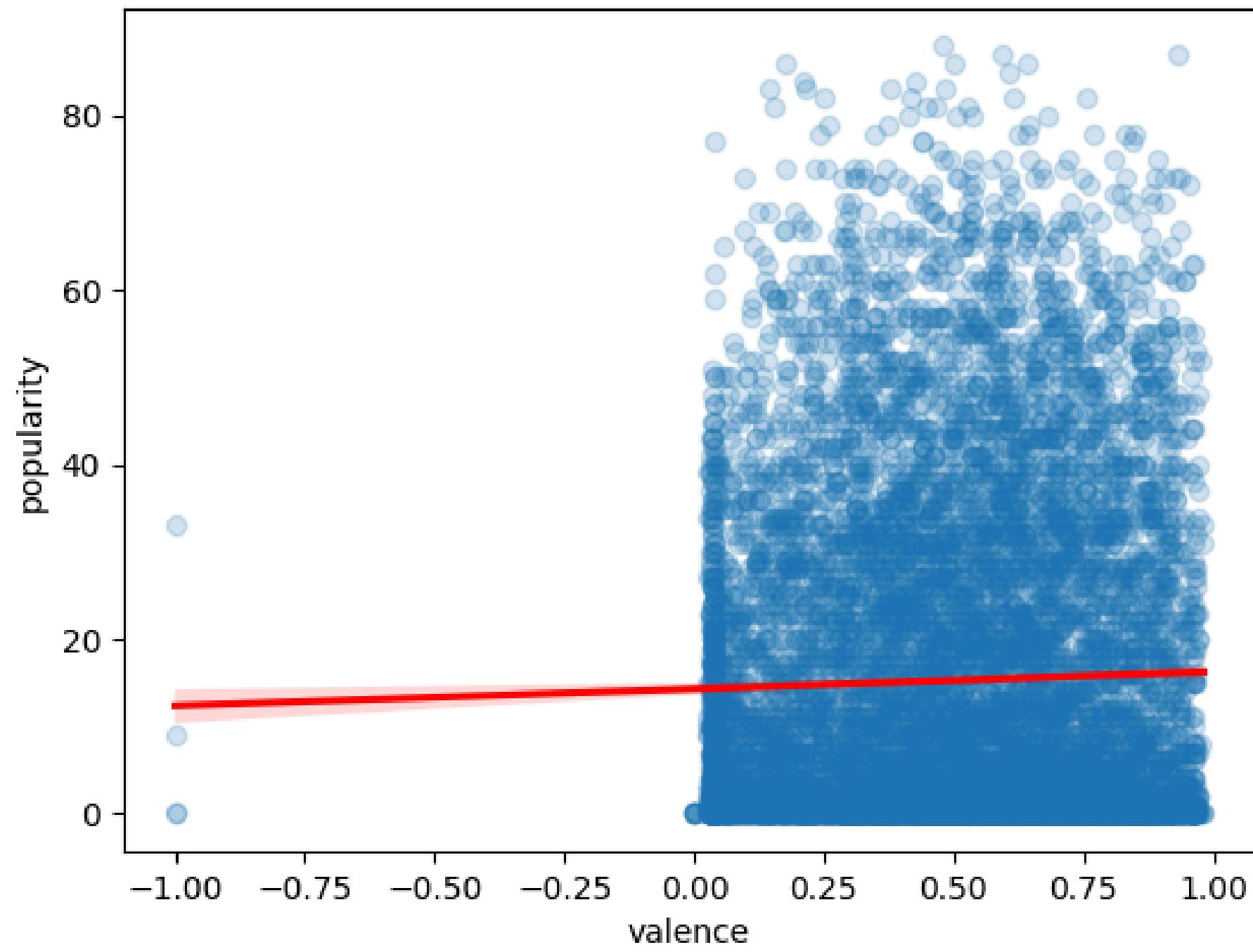
Energy: Top 5% vs Rest



Danceability: Top 5% vs Rest



Popularity vs Valence (sample)



Key Insights and Implementations for Mixing Engineer:-

Danceability: Top 5% vs Rest

- Insight: Top 5% tracks tend to have higher danceability and a tighter spread. Popular tracks cluster into a more consistently danceable range.

Energy: Top 5% vs Rest

- Insight: Highly popular tracks skew toward higher energy with fewer very-low-energy outliers. Energy appears to be a favorable attribute for popularity.
-

Popularity vs Valence (sample + trendline)

- Insight: There's a mild positive trend: happier-sounding songs (higher valence) tend to be slightly more popular, but the effect size is modest—valence alone doesn't explain popularity.

Feature correlations among the Top 5%

- Insight: No single pair dominates. Relationships are moderate at best, suggesting popular tracks are a balanced blend of traits rather than driven by one metric. This supports a multifactor view of popularity.

Practical implementation takeaways

- Push danceability and energy to an above-median band without overextremes. This aligns with where Top 5% tracks cluster.
- Treat valence as a tuning knob, not a cornerstone. Slightly happier vibes help, but don't overfit on "happy."
- Optimize the mix: within top-performing tracks, features don't show extreme correlations, implying that crafting the right combination (rather than maximizing one metric) is more effective.
- For modeling, expect interactions to matter. Consider tree-based methods or interaction terms if you're regressing on popularity.

Final Recommendations for the Mixing Engineer/Music Producer

1. Prioritize High-Potential Tracks

- Insight: Popularity follows a power-law—only a few tracks dominate
- Recommendation: Don't treat every song the same. Allocate your best production, mixing, and marketing efforts to songs with strong hooks, energy, and genre relevance.

2. Mix for Energy and Danceability First

- Insight: Energy is the strongest positive driver; danceability clusters in top tracks
- Recommendation: Use layered drums, punchy percussion, transient clarity, and tight kick–bass relationships. Aim for medium-high danceability and consistent groove.

3. Stay in the Winning Tempo Zone (100–130 BPM)

- Insight: Most popular tracks sit between 110–130 BPM
- Recommendation: Build songs in the 118–128 BPM range for pop/dance, or halftime 70–90 for hip-hop/R&B. This aligns with both streaming and club performance.

4. Use Controlled, Competitive Loudness

- Insight: Popular songs sit around -6 to -10 LUFS; loudness alone doesn't drive popularity
- Recommendation: Master at -9 to -7 LUFS with true peaks under -1 dBTP. Preserve punch with parallel compression and avoid over-limiting.

5. Structure Songs for Fast Engagement

- Insight: Duration doesn't drive popularity, but shorter songs win in streams
- Recommendation: Keep total length around 2:30–3:10. Hit the first hook within 30–40 seconds. Avoid long intros and meandering sections.

6. Favor 4/4 and Modern Production Textures

- Insight: 4/4 dominates; acousticness trends slightly negative
- Recommendation: Stick to 4/4 for accessibility. For acoustic songs, add electronic layers, modern drums, or subtle synth textures to stay relevant.

7. Lean Into High-Energy Mix Arcs

- Insight: Top songs show energy build patterns
- Recommendation:
- Arrange with dynamic lift:
- Verse → moderate
- Pre-chorus → rising
- Chorus → maximum impact
- Bridge → contrast, then re-lift

8. Adopt Global Sound Standards (K-Pop, EDM, Modern Pop)

- Insight: Korean songs have the highest mean popularity
- Recommendation: Incorporate global mixing traits: bright vocal stacks, polished low end, hybrid synth/acoustic blending, punchy master chains.

9. Vocals Front and Center

- Insight: Instrumentalness and acousticness decline over time; speechiness plays a role
- Recommendation:
- For pop: prioritize melodic vocals with clarity (2–6 kHz).
- For rap/urban: use rhythmic vocal phrasing or sung hooks. Avoid muddy midrange.

10. Balance Features Instead of Chasing One Metric

- Insight (Top 5% tracks): No single feature dominates—it's a blend
- Recommendation:
- Design songs with a mix of:
- Medium-high energy
- Moderate positivity
- Groove-focused rhythm
- Clear vocals
- Tight low end

Avoid extremes—synergy matters more than maxing one trait.

Thank You !

ধন্যবাদ

খুম্মা ঘণ্টী