PRESENTATION

SPOTIFY DATA ANALYSIS

SAYAN HASSAN
ROLL:GCECTB-R24-3030
DEPT:CSE[3RD SEM]

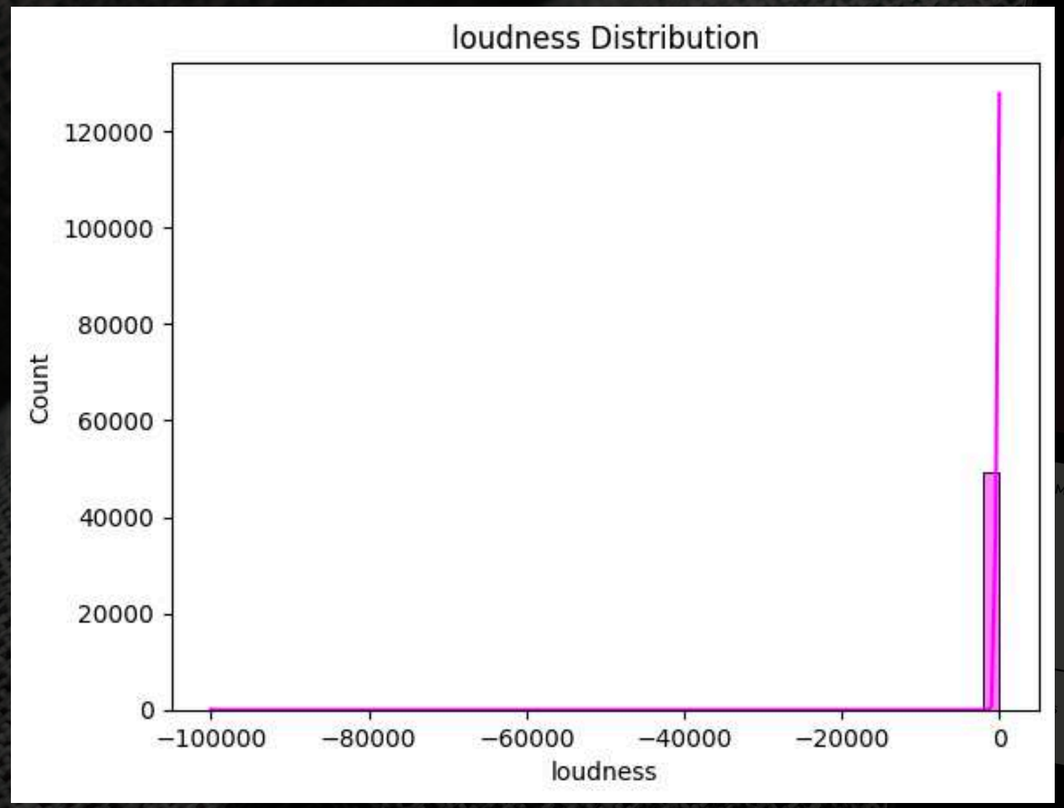
5 OCTOBER 2025

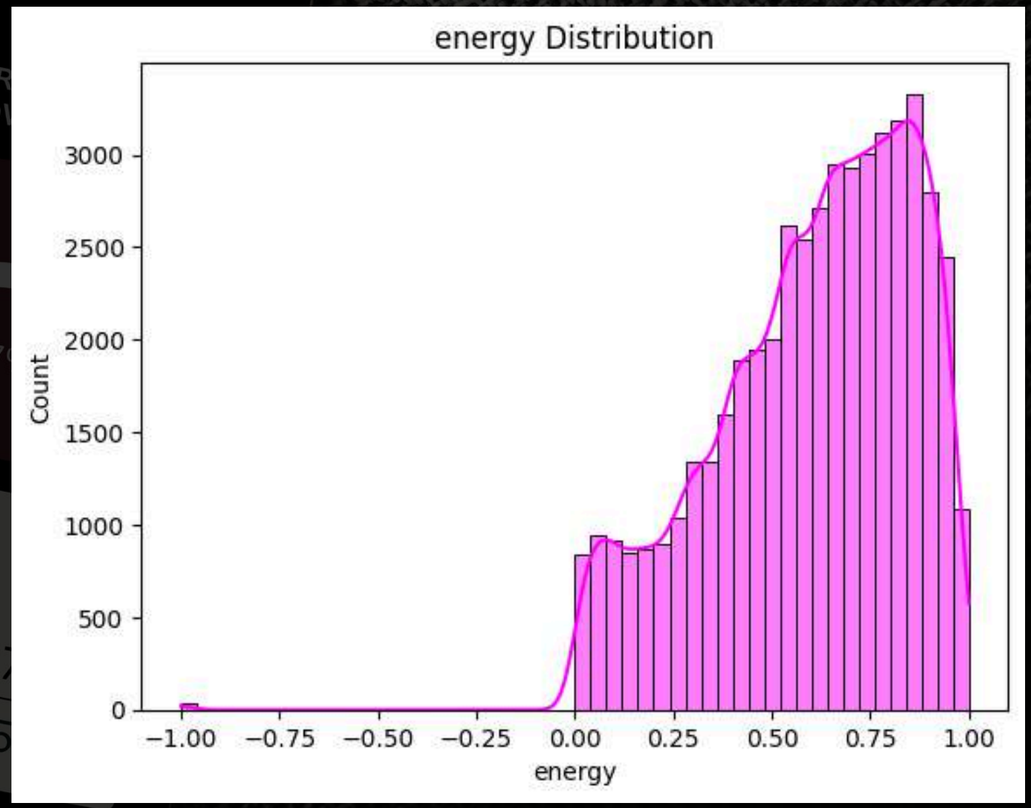
DECODING MUSIC: DATA CONTEXT & OBJECTIVE

KEY FEATURES ANALYZED (THE DATA'S DNA)

FEATURE 1: INTENSITY

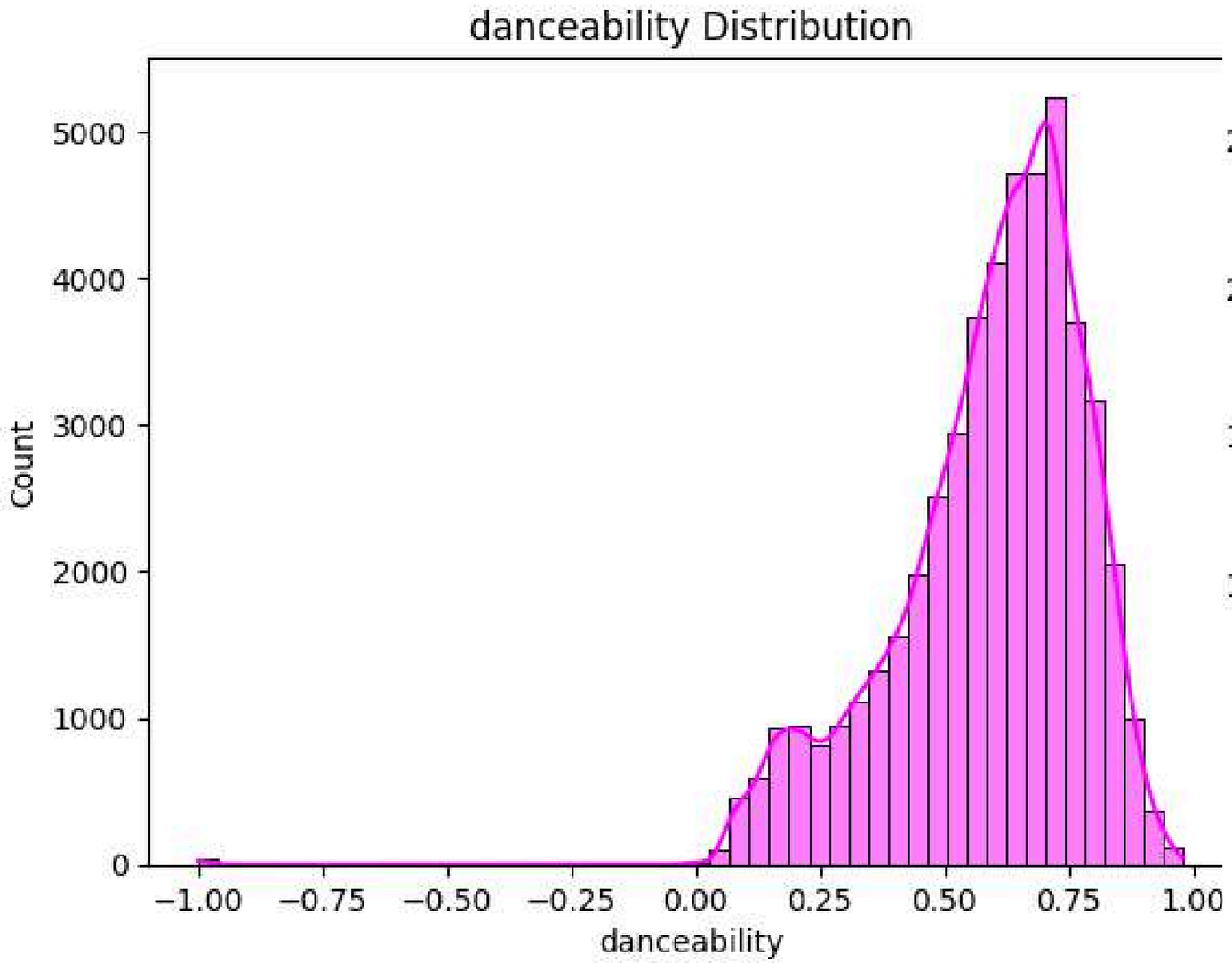
ENERGY & LOUDNESS (MEASURES THE INTENSITY, ACTIVITY LEVEL, AND VOLUME OF THE TRACK.)

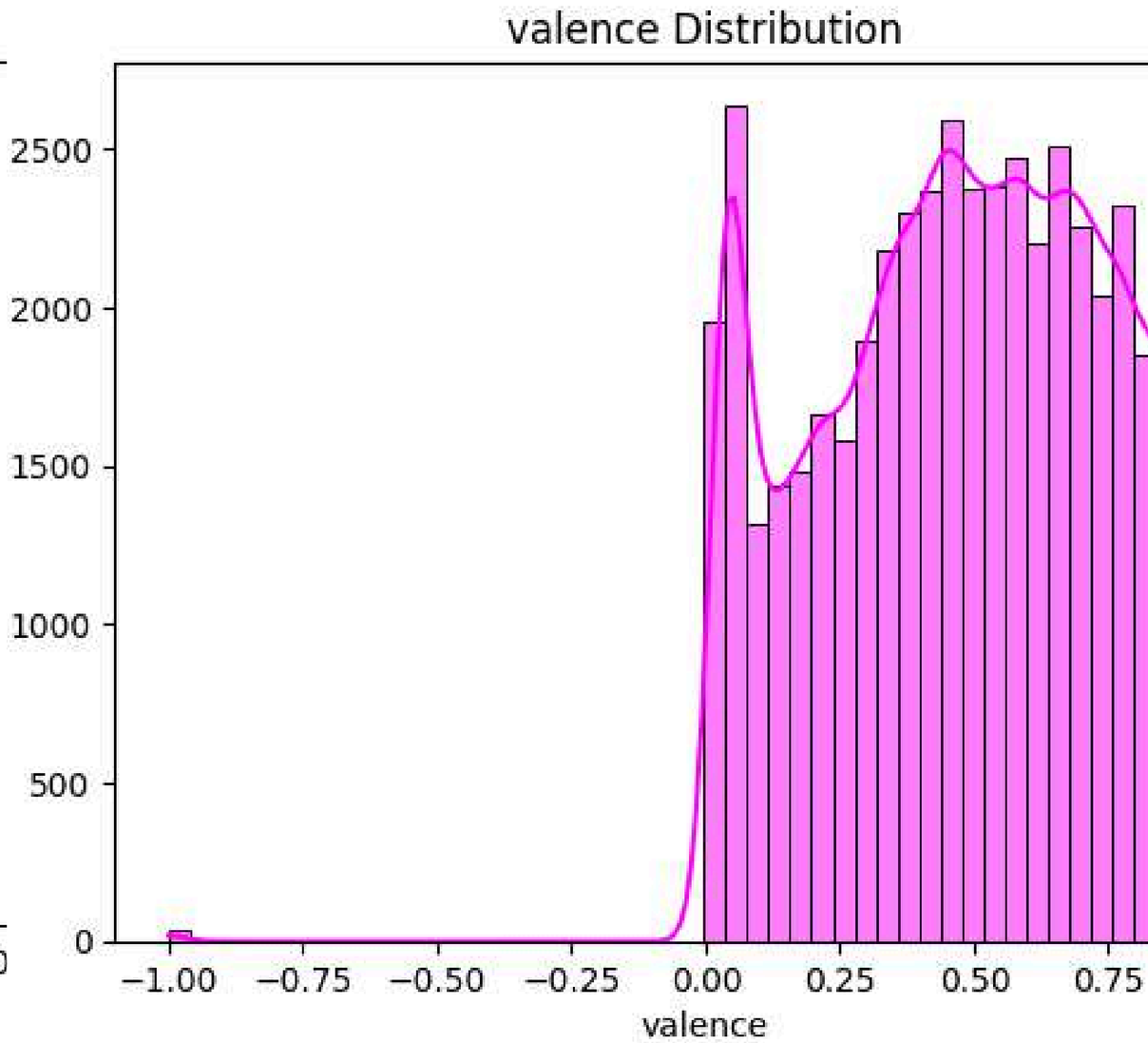




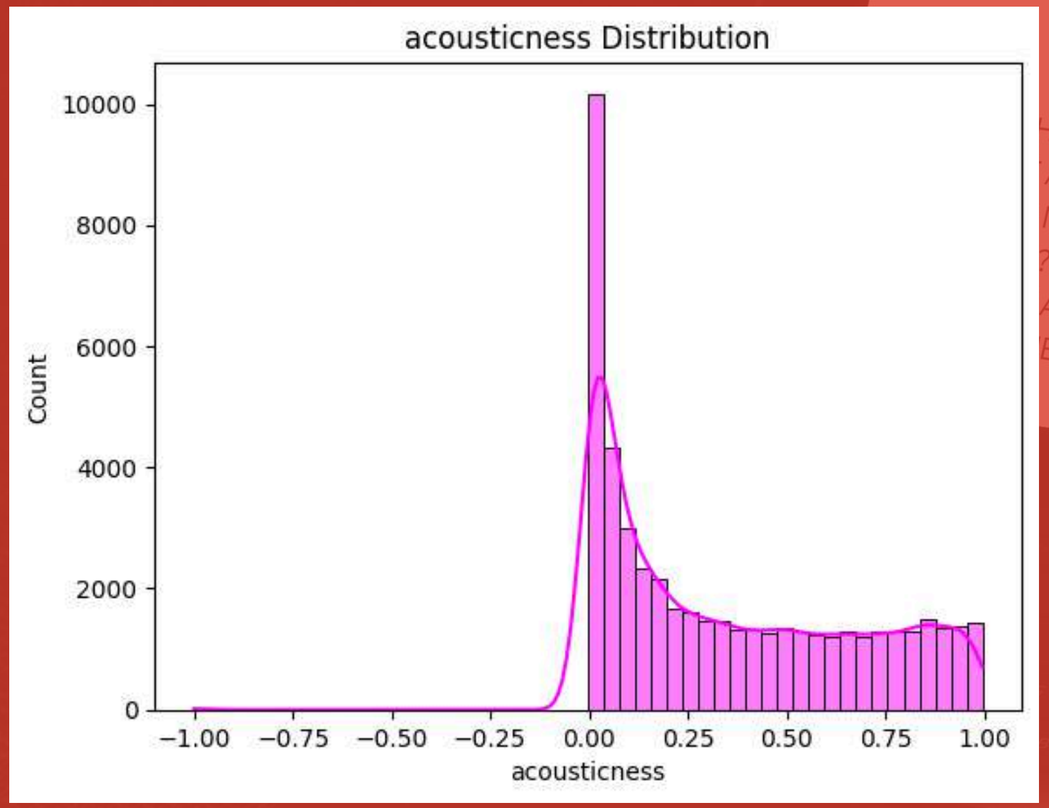
FEATURE 2: MOOD & RHYTHM

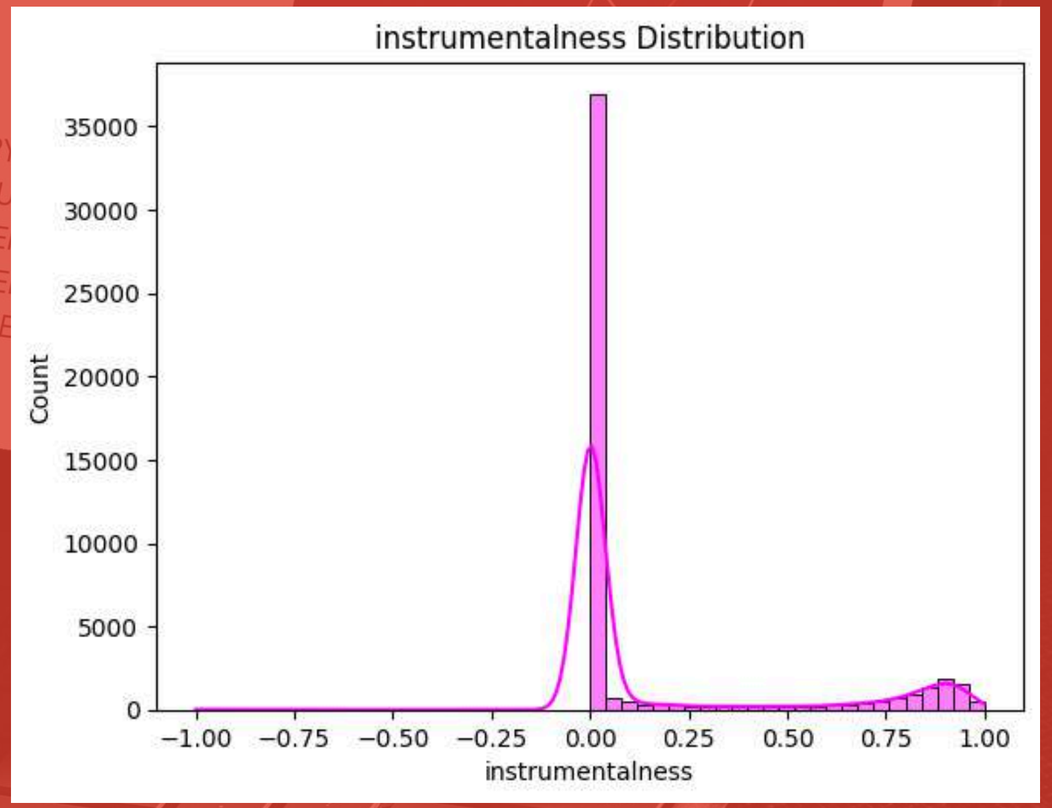
DANCEABILITY & VALENCE (MEASURES THE SUITABILITY FOR DANCING AND THE MUSICAL POSITIVITY, OR "MOOD.")





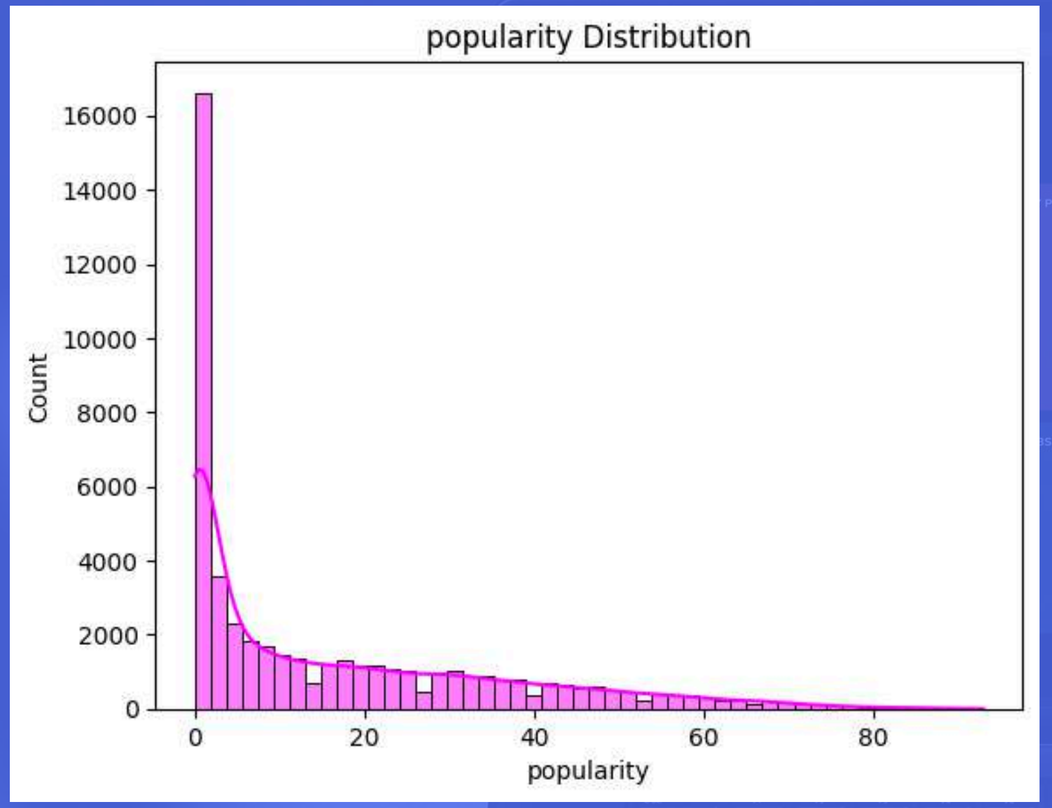
FEATURE 3: PRODUCTION STYLEACOUSTICNESS & INSTRUMENTALNESS (MEASURES THE PRESENCE OF NON-ELECTRONIC SOUNDS AND VOCAL ABSENCE.)





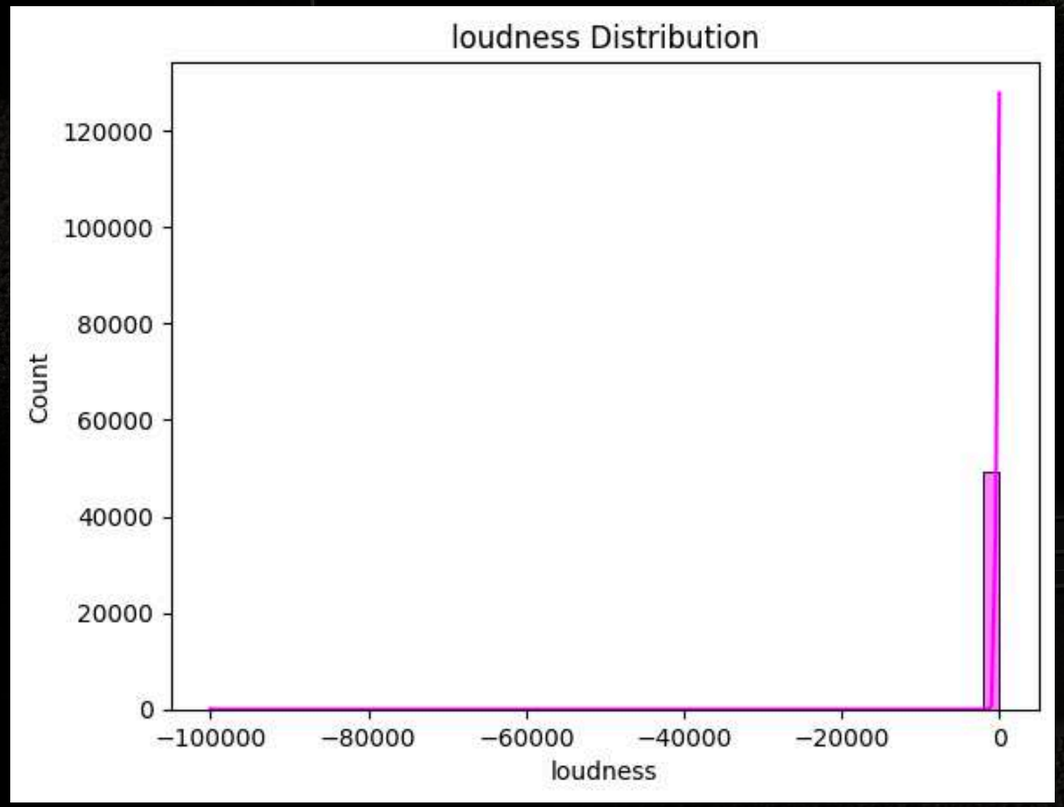
TARGET VARIABLE

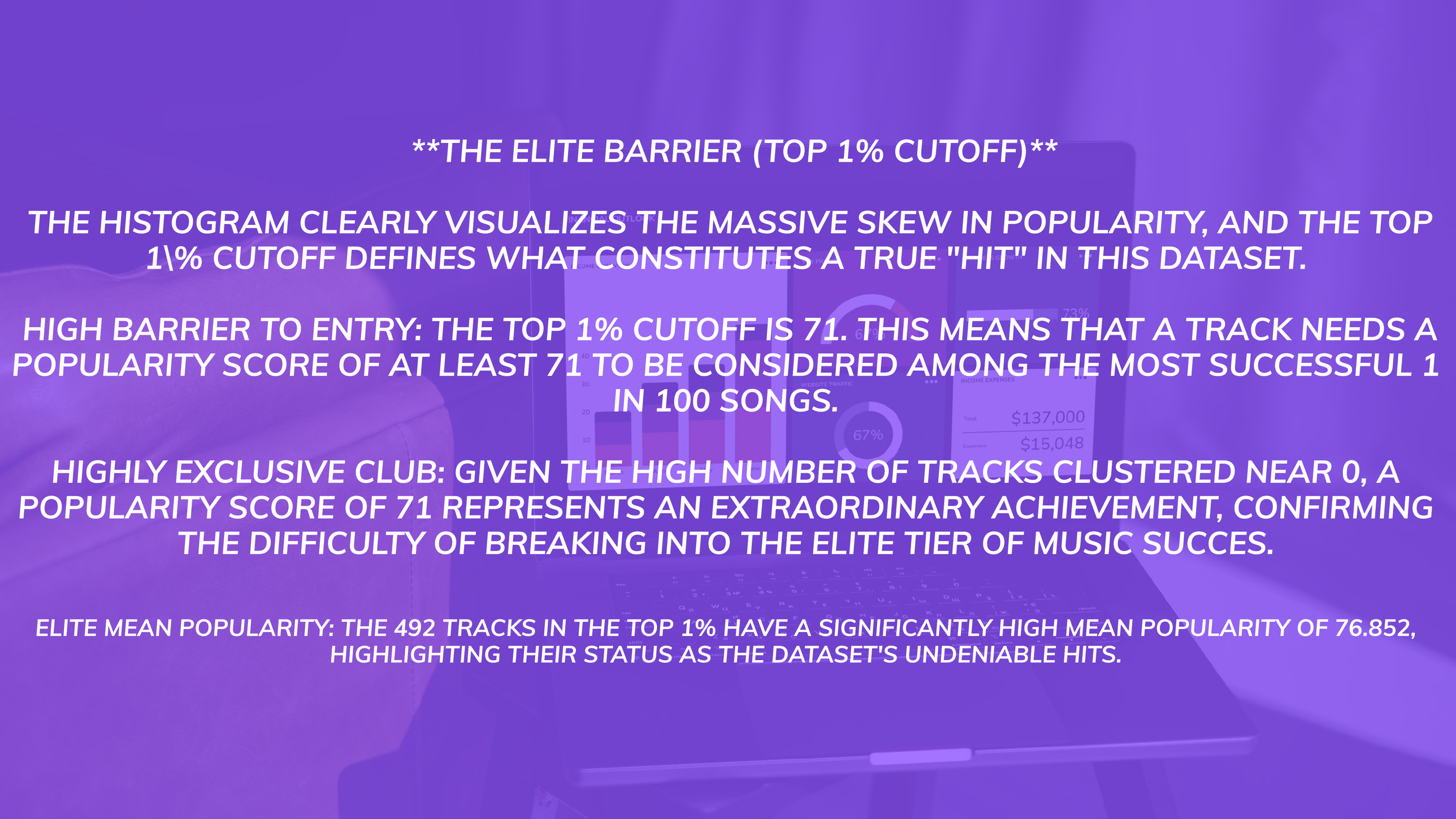
POPULARITY (THE SCORE (0-100) WE ARE TRYING TO PREDICT AND EXPLAIN.)



TREND 1: THE LOUDER, THE BETTER

CONTINUOUS GROWTH (1950S – 2010S): BOTH THE AVERAGE ENERGY (INTENSITY/ACTIVITY) AND LOUDNESS (DECIBEL LEVEL) OF TRACKS HAVE STEADILY AND CONTINUOUSLY INCREASED OVER THE DECADES.





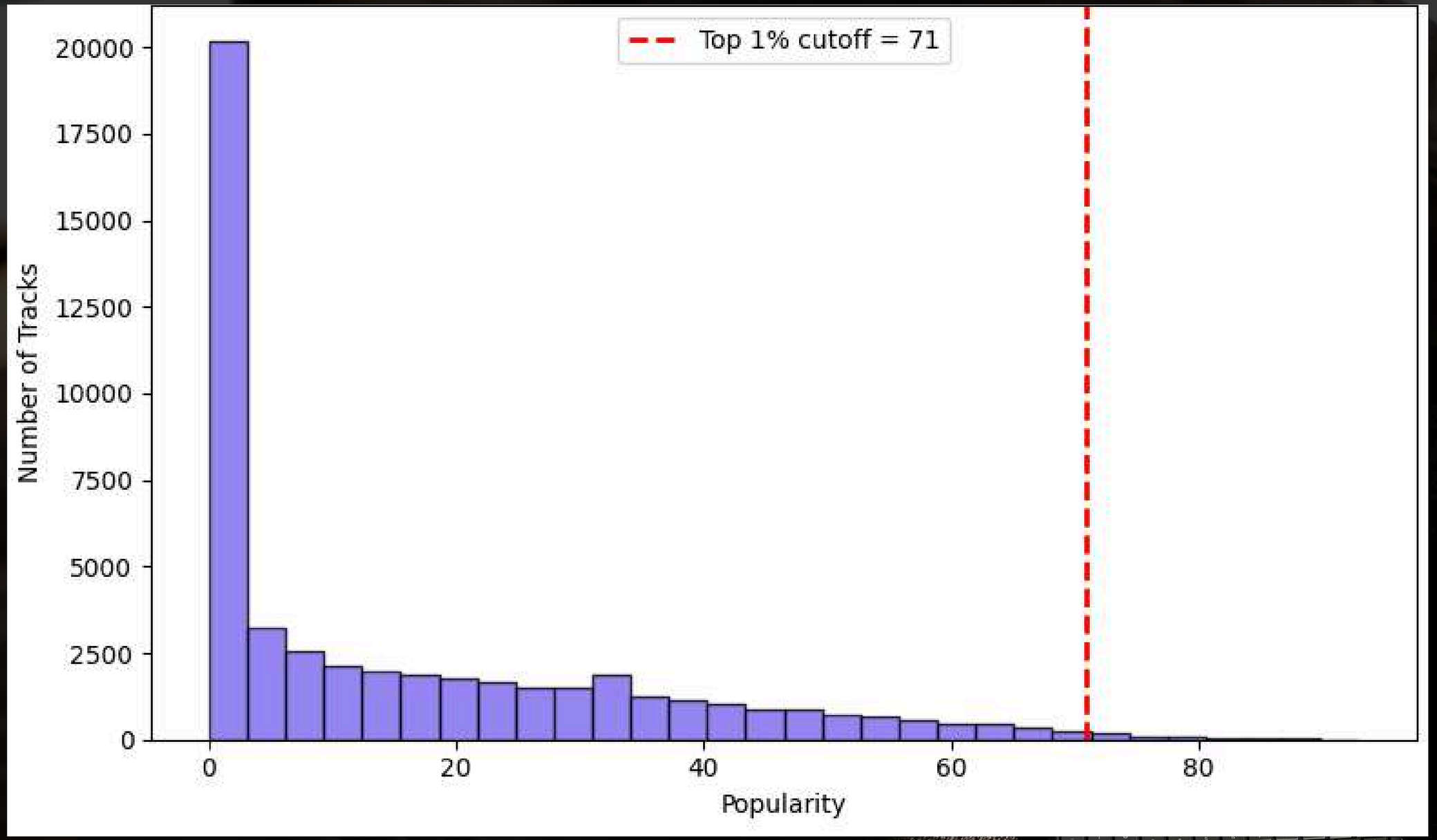
THE ELITE BARRIER (TOP 1% CUTOFF)

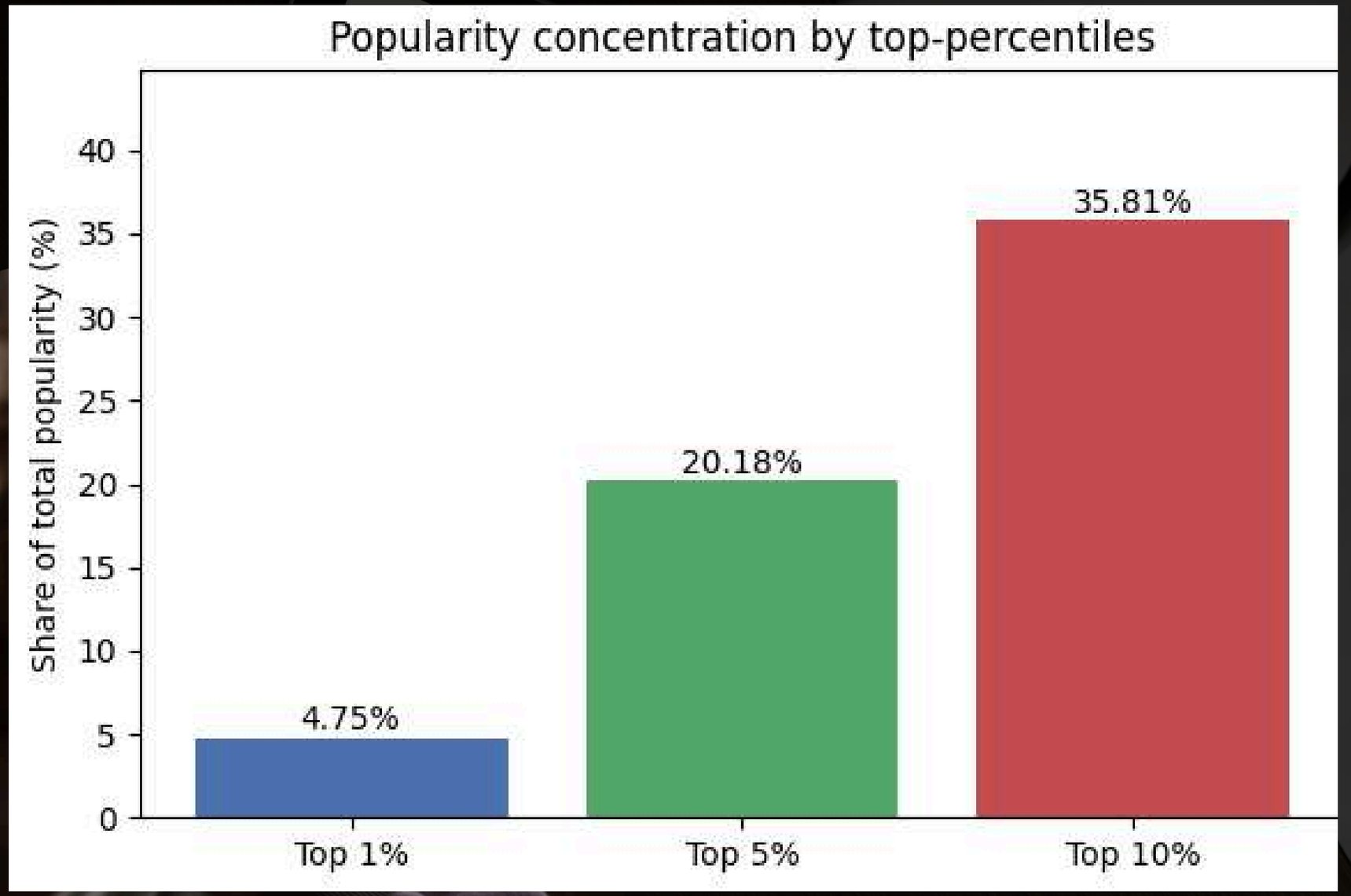
THE HISTOGRAM CLEARLY VISUALIZES THE MASSIVE SKEW IN POPULARITY, AND THE TOP 1\% CUTOFF DEFINES WHAT CONSTITUTES A TRUE "HIT" IN THIS DATASET.

HIGH BARRIER TO ENTRY: THE TOP 1% CUTOFF IS 71. THIS MEANS THAT A TRACK NEEDS A POPULARITY SCORE OF AT LEAST 71 TO BE CONSIDERED AMONG THE MOST SUCCESSFUL 1 IN 100 SONGS.

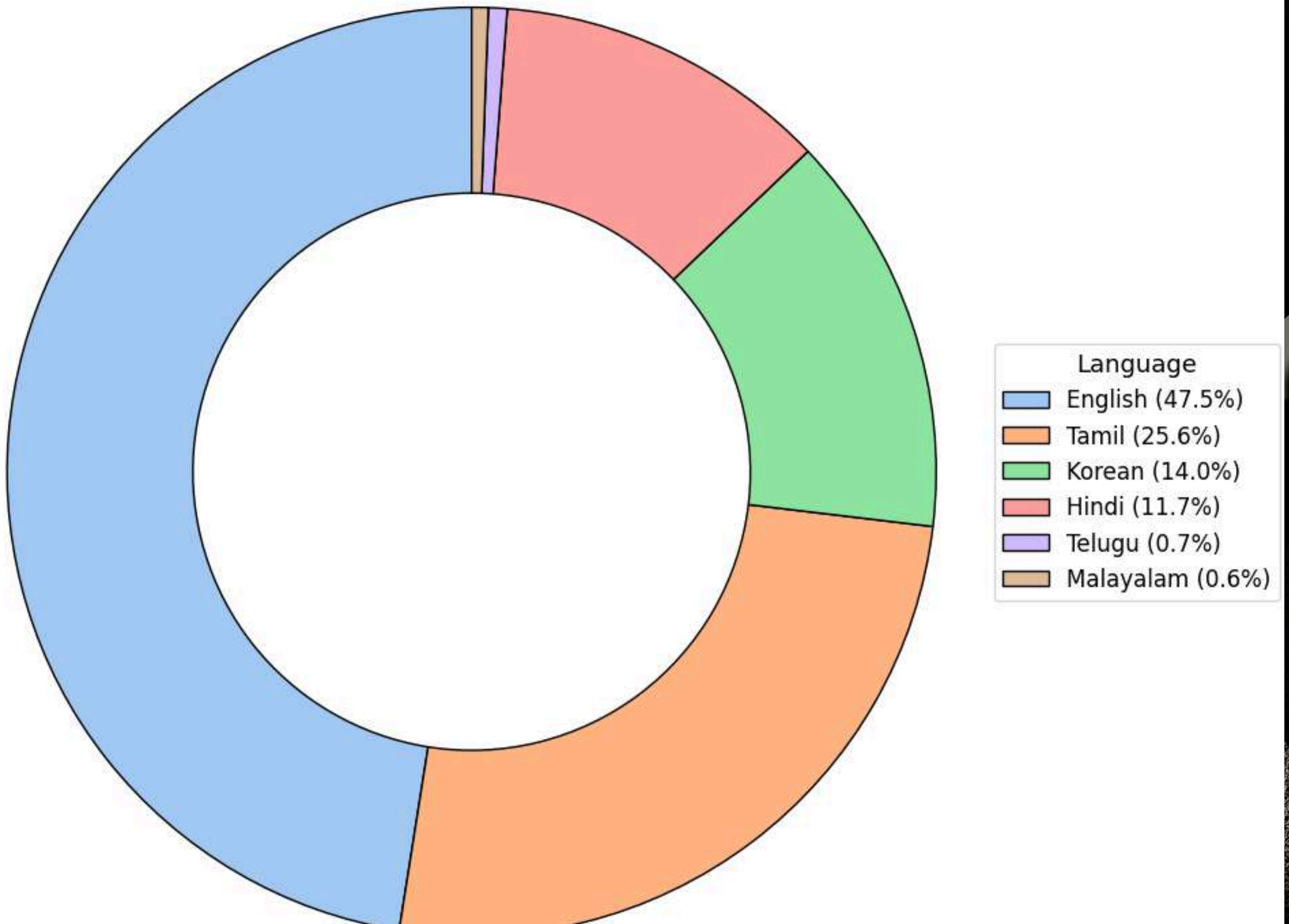
HIGHLY EXCLUSIVE CLUB: GIVEN THE HIGH NUMBER OF TRACKS CLUSTERED NEAR 0, A POPULARITY SCORE OF 71 REPRESENTS AN EXTRAORDINARY ACHIEVEMENT, CONFIRMING THE DIFFICULTY OF BREAKING INTO THE ELITE TIER OF MUSIC SUCCES.

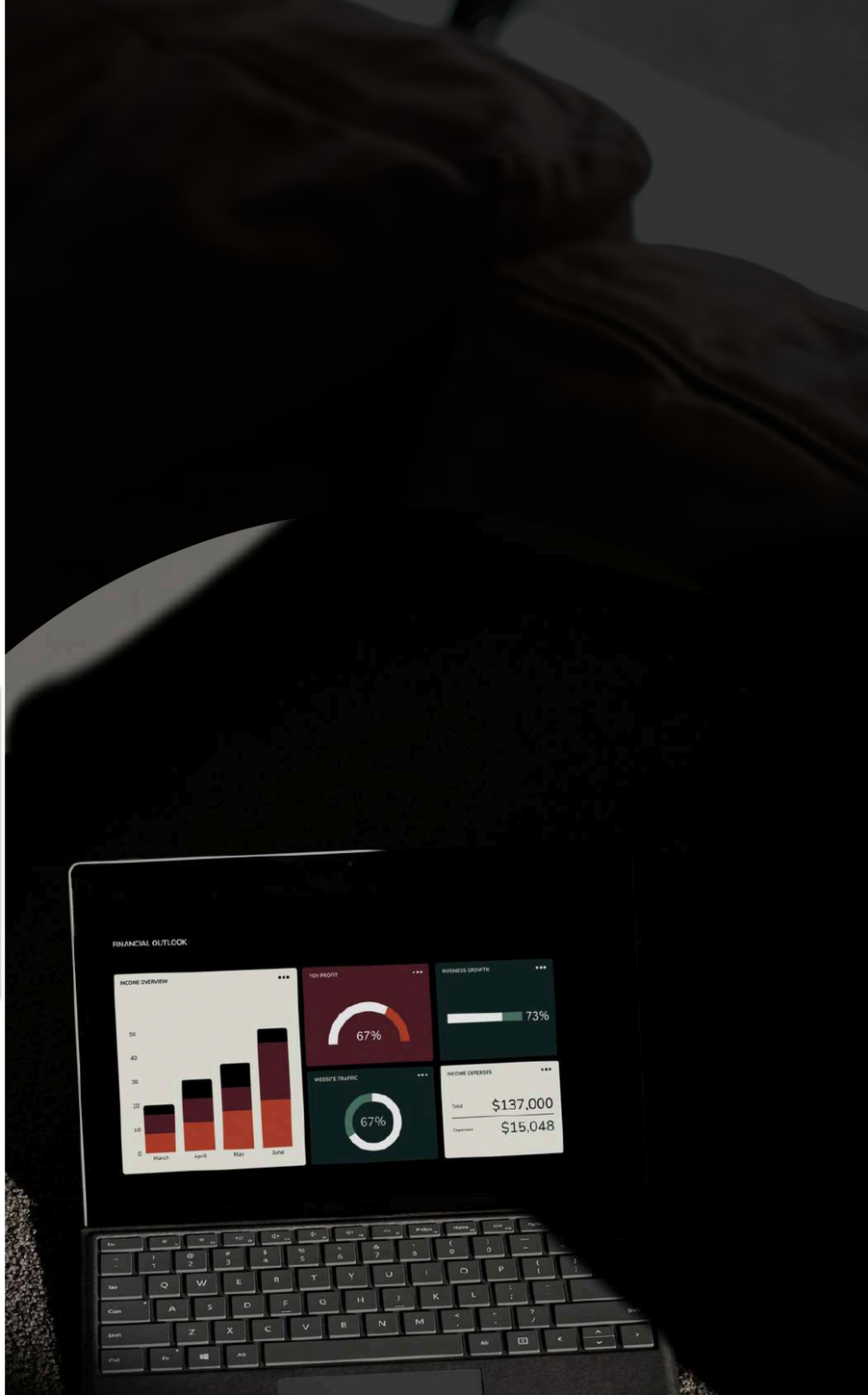
ELITE MEAN POPULARITY: THE 492 TRACKS IN THE TOP 1% HAVE A SIGNIFICANTLY HIGH MEAN POPULARITY OF 76.852, HIGHLIGHTING THEIR STATUS AS THE DATASET'S UNDENIABLE HITS.


Top 1% cutoff = 71
Number of Tracks
20000
17500
15000
12500
10000
7500
5000
2500
0
0
20
40
60
80
Popularity


Popularity concentration by top-percentiles
Share of total popularity (%)
0
5
10
15
20
25
30
35
40
4.75%
20.18%
35.81%
Top 1%
Top 5%
Top 10%

Proportional Distribution of Top 6 language Categories





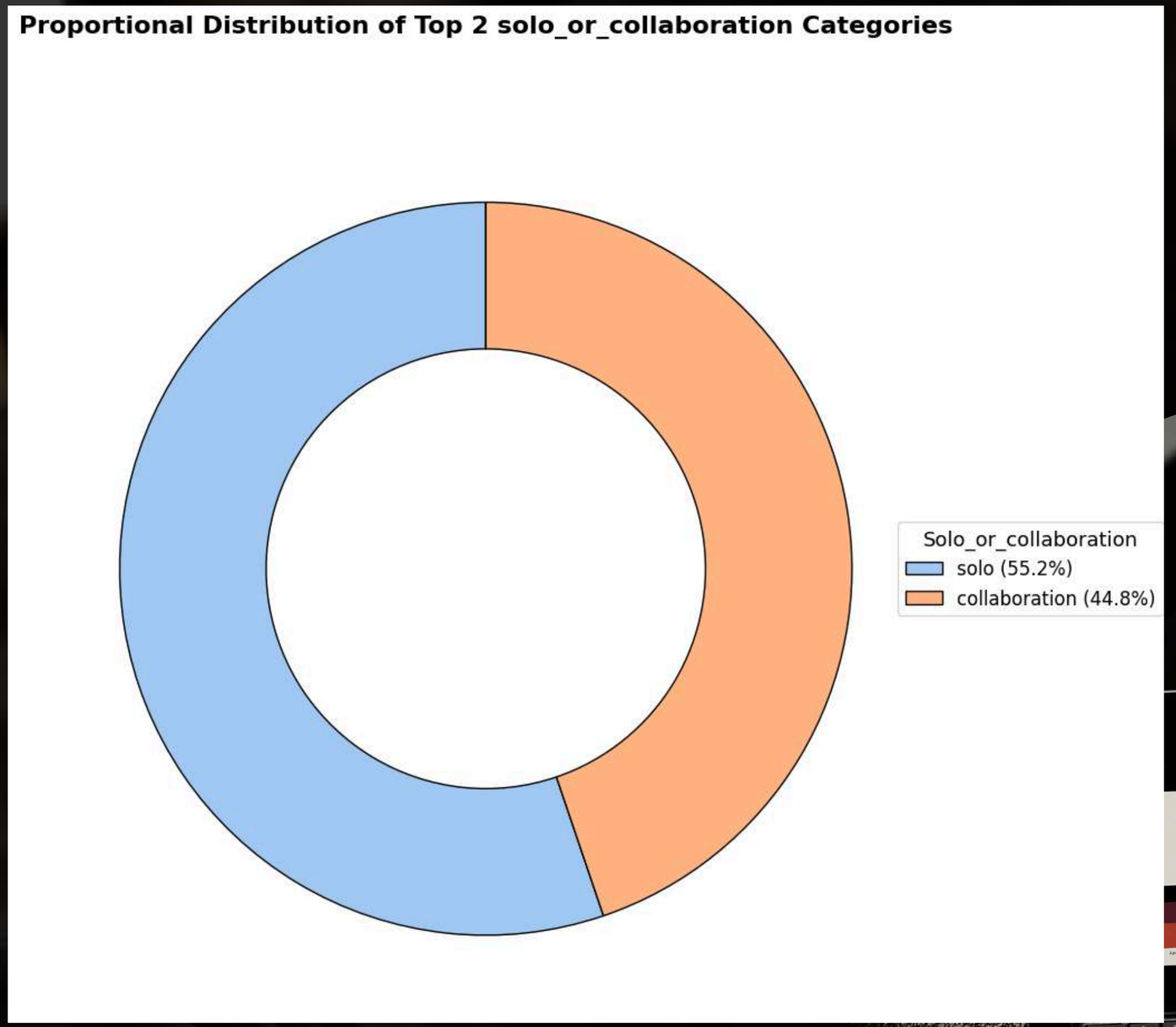
KEY INSIGHTS:

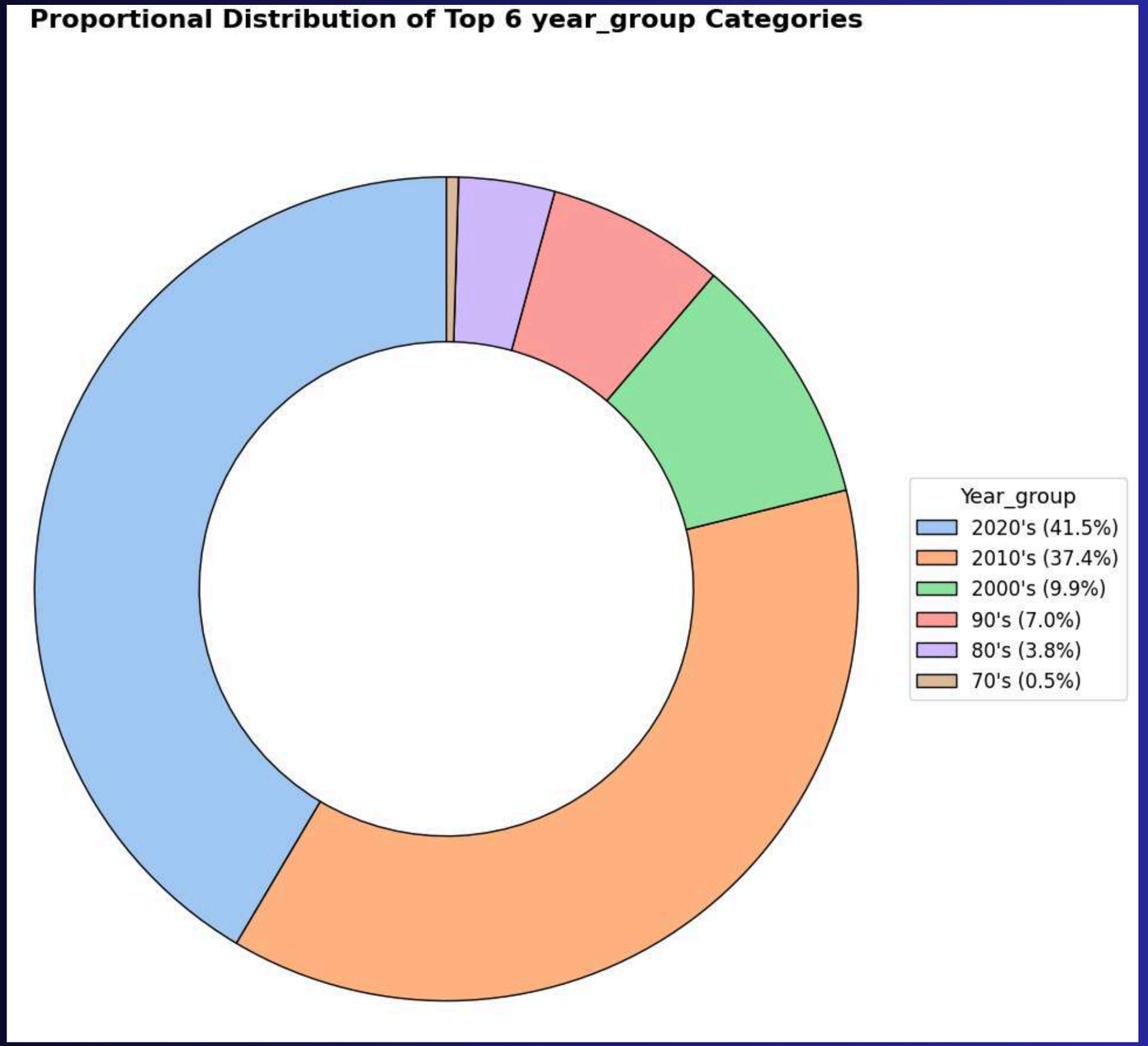
ENGLISH DOMINANCE: ENGLISH IS THE PRIMARY LANGUAGE, REPRESENTING ALMOST HALF (47.5%) OF THE TOTAL TRACKS, CONFIRMING ITS STATUS AS THE MOST WIDELY REPRESENTED LANGUAGE IN THIS SPECIFIC DATASET.

STRONG SOUTH ASIAN PRESENCE: THE COMBINATION OF TAMIL (25.6%) AND HINDI (11.7%) SHOWS A SIGNIFICANT REPRESENTATION OF SOUTH ASIAN MUSIC, COLLECTIVELY ACCOUNTING FOR OVER ONE-THIRD (37.3%) OF THE TRACKS. THIS SUGGESTS DATASET MAY BE DRAWN FROM A REGION OR PLATFORM WITH HIGH INDIAN LANGUAGE MUSIC CONSUMPTION.

GLOBAL POPULARITY: KOREAN MUSIC ALSO HOLDS A NOTABLE SHARE AT 14.0%, LIKELY REFLECTING THE GLOBAL INFLUENCE OF K-POP.

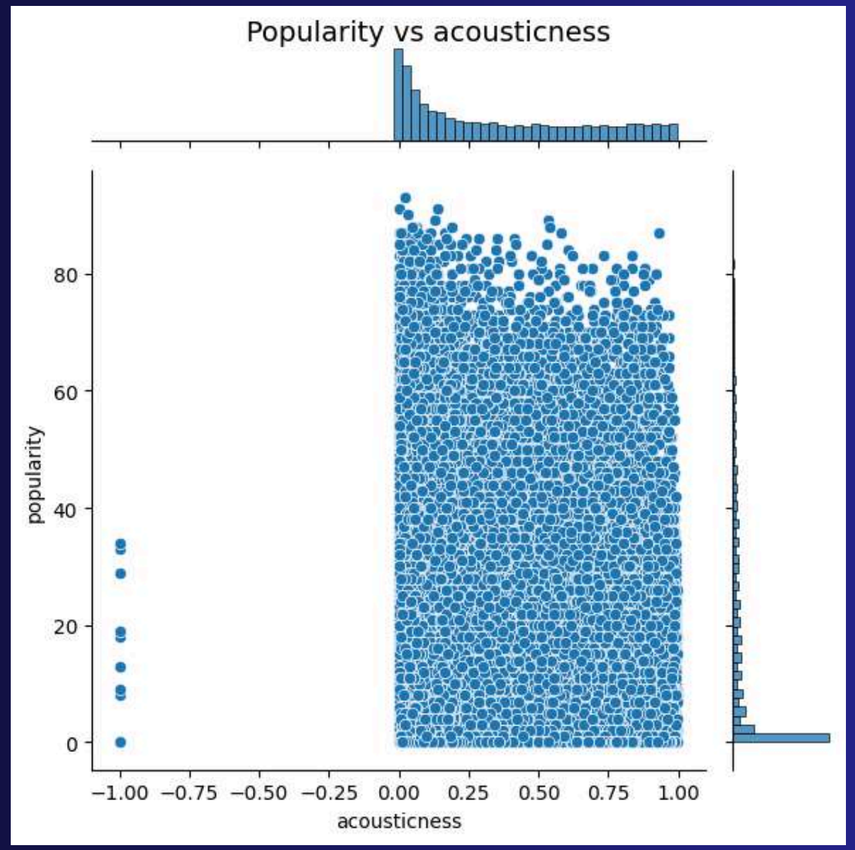
MINORITY LANGUAGES: LANGUAGES LIKE TELUGU (0.7%) AND MALAYALAM (0.6%) REPRESENT A MINIMAL FRACTION OF THE DATASET.


Proportional Distribution of Top 2 solo_or_collaboration Categories
Solo_or_collaboration
solo (55.2%)
collaboration (44.8%)


Proportional Distribution of Top 6 year_group Categories
Year_group
2020's (41.5%)
2010's (37.4%)
2000's (9.9%)
90's (7.0%)
80's (3.8%)
70's (0.5%)

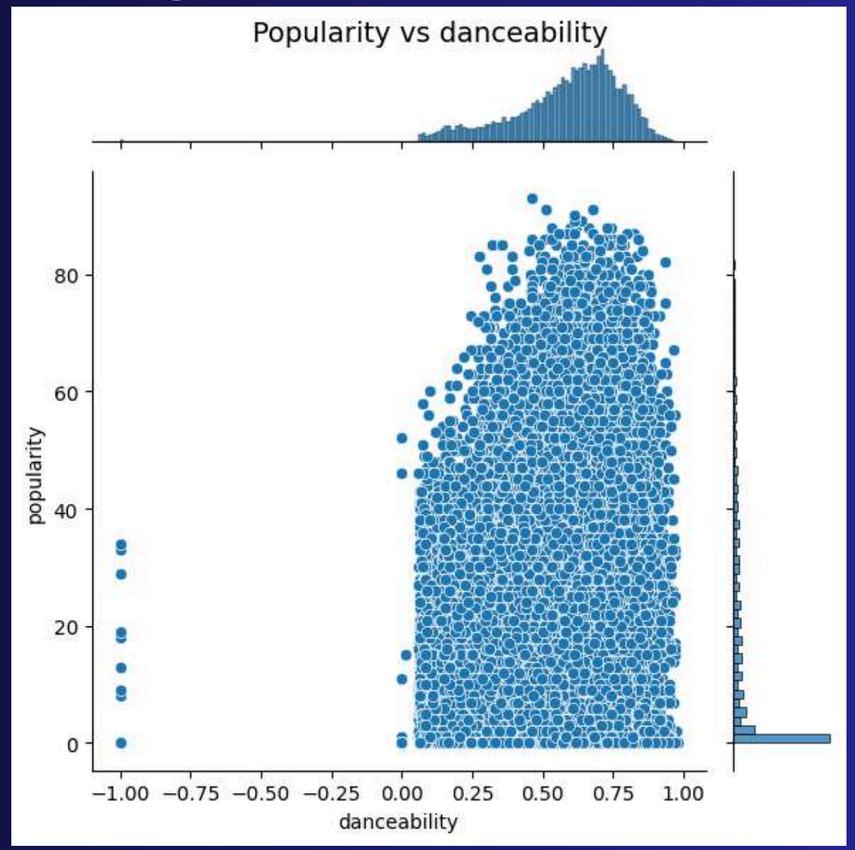
1. Popularity vs. Acousticness :**

The data shows no clear correlation with popularity. High popularity scores (80-90) are achieved across the entire acousticness spectrum (0.0 to 1.0). This means that both highly digital, non-acoustic tracks and purely acoustic tracks have the potential to become major hits, indicating that acousticness is genre-dependent rather than a universal barrier or driver of popularity.



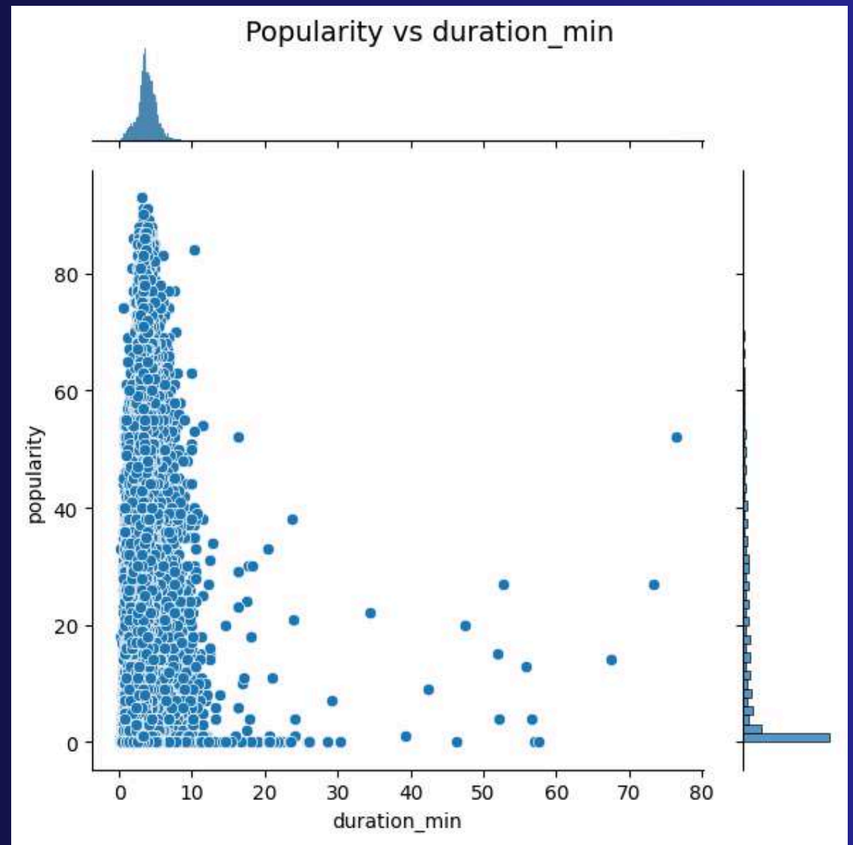
2. Popularity vs. Danceability :

There is a strong visual prerequisite for danceability. Tracks with popularity scores above 40 rarely have a danceability score below 0.25. The highest popularity scores (85-90) are concentrated in the high danceability range (0.50 to 0.90). This suggests high danceability is a necessary ingredient for achieving breakout success.



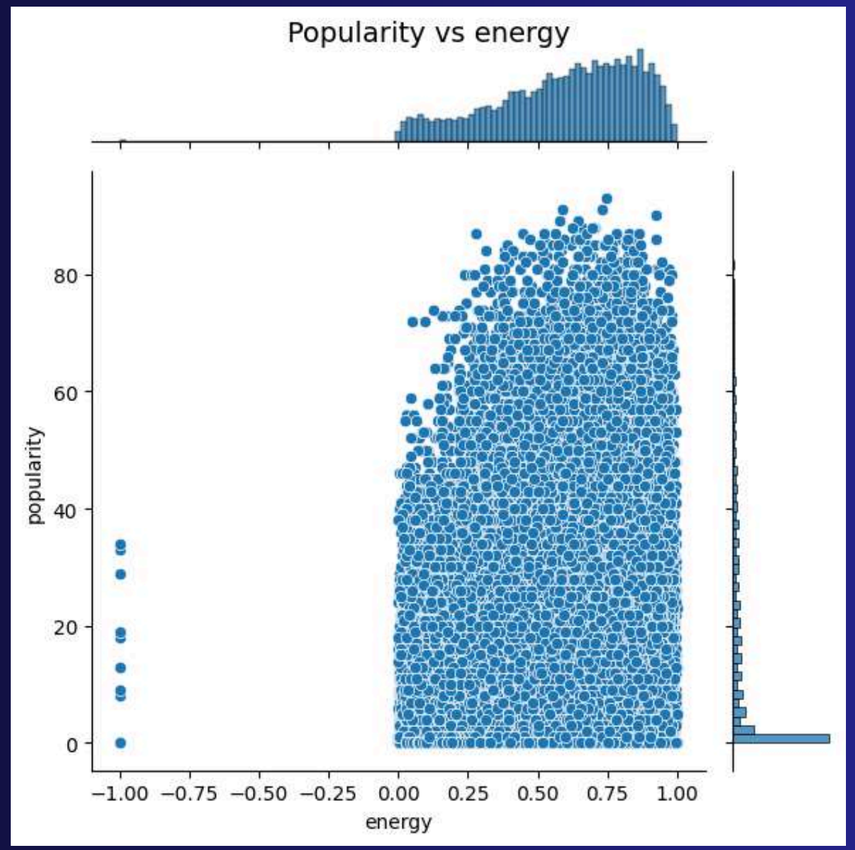
3. Popularity vs. Duration (min) :

There is a clear optimal duration window for popularity. The vast majority of popular tracks (popularity > 60) are tightly clustered under 6 minutes. Tracks longer than 10 minutes almost never exceed a popularity score of 25, demonstrating a strong listener preference for short-form content in popular music.



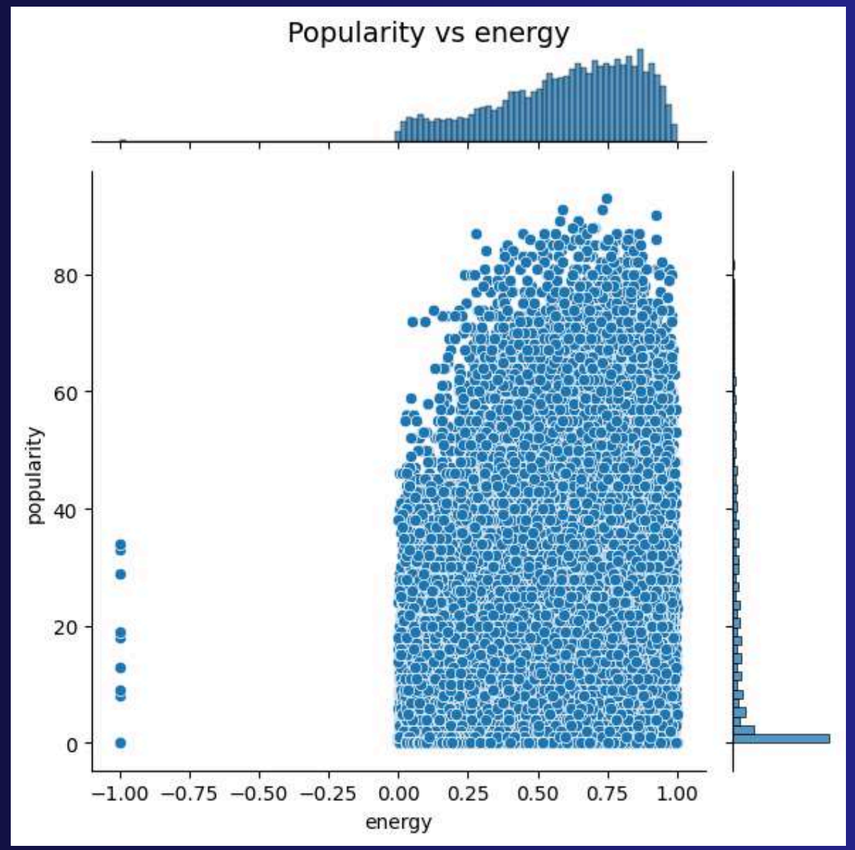
4. Popularity vs. Energy :

Similar to danceability, high energy is strongly correlated with maximum popularity. Tracks with the highest popularity scores (85-90) almost exclusively have an energy score above 0.25. Below 0.25, it is nearly impossible to achieve a popularity score above 60. High energy is a key feature for modern hits.



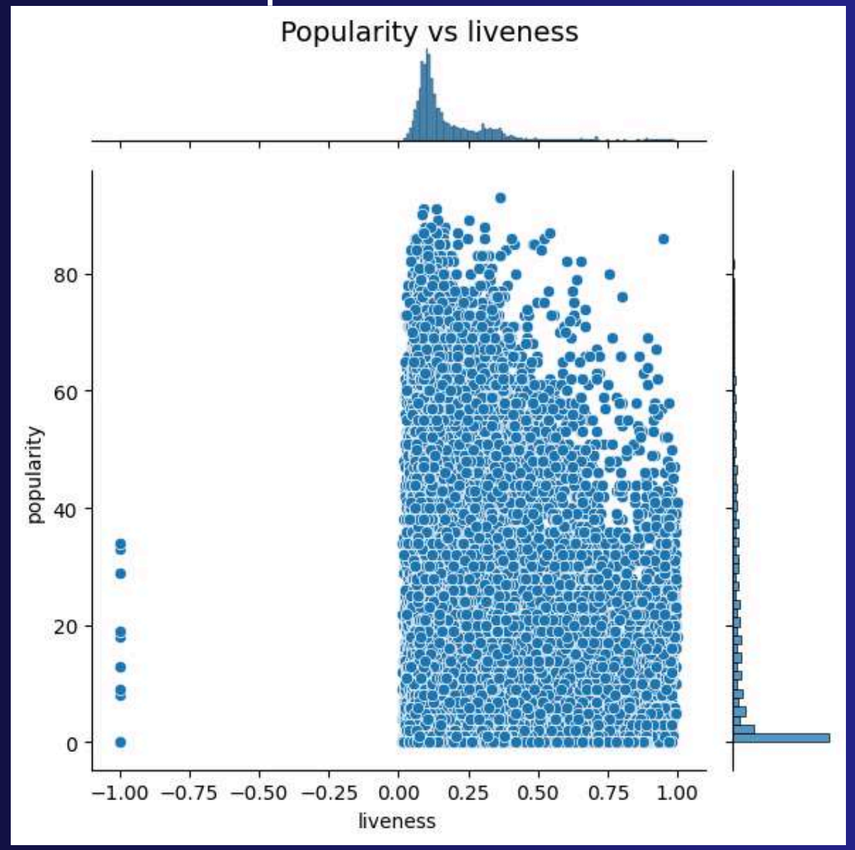
5. Popularity vs. Instrumentalness:

Instrumentalness is a strong negative factor for maximum popularity. The highest popularity scores (70+) are almost entirely found among less-instrumental tracks (0.0 to 0.25). This suggests that vocals are a necessary component to reach the elite level of success.



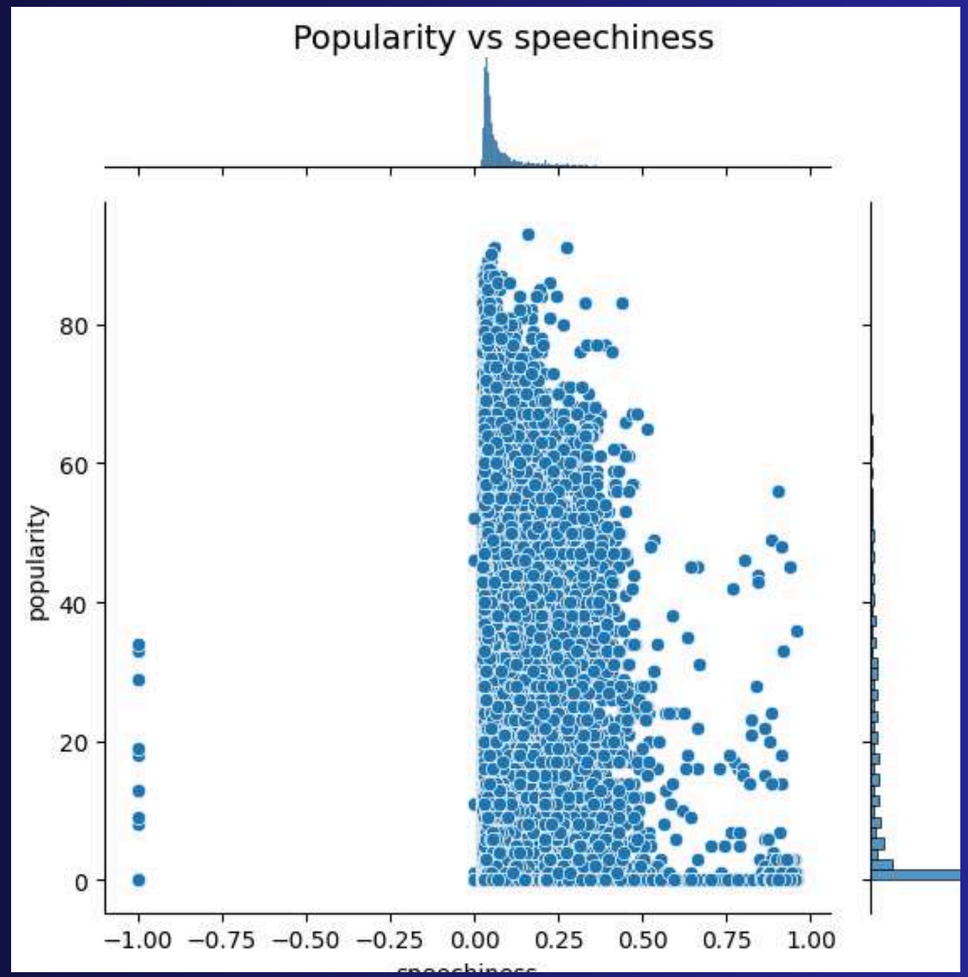
6. Popularity vs. Liveness:

The highest popularity scores (80) are predominantly found among tracks with a low liveness score (below 0.30), indicating they are studio recordings. While some moderately popular tracks exist at high liveness scores (0.80), live recordings face a lower ceiling for mass popularity compared to polished, studio-produced music.



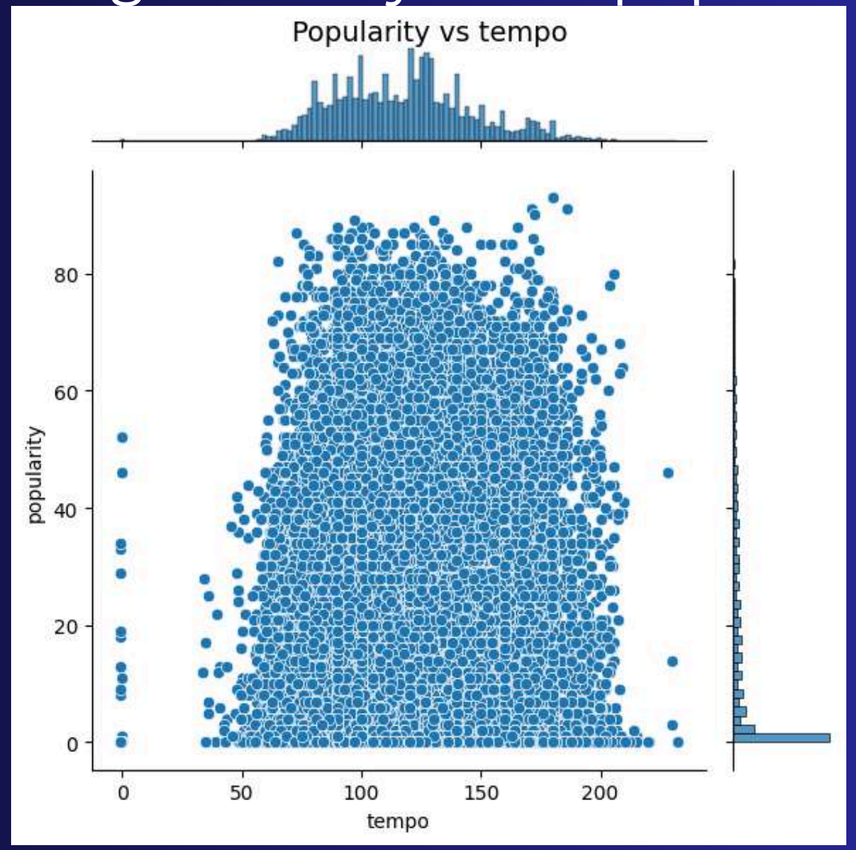
7. Popularity vs. Speechiness:

There is a clear popularity ceiling for highly spoken-word tracks. While the highest popularity tracks are concentrated at the low end (0.0 to 0.40), tracks with a speechiness score above 0.50 (likely rap, spoken word, or podcast clips) rarely exceed a popularity score of 50 or 60. Sung music is overwhelmingly favored for maximum popularity.



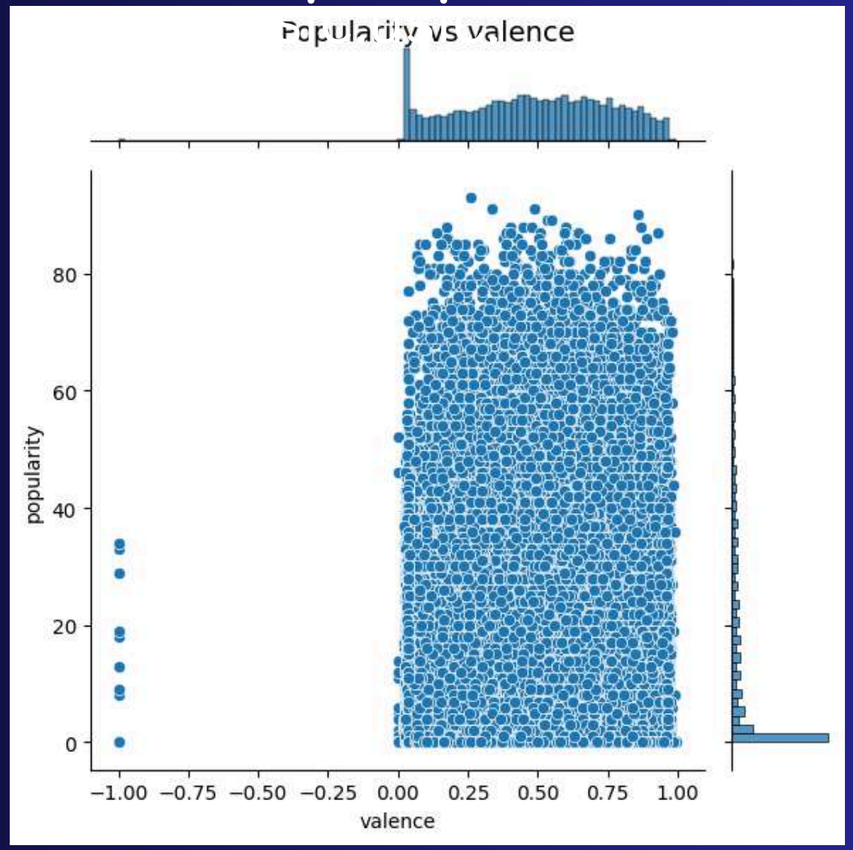
8. Popularity vs. Tempo (BPM) :

Popularity is achieved across a wide but limited tempo range. The highest popularity scores (80-90) are concentrated between approximately 80 and 160 BPM, with a slightly denser concentration between 100 and 140 BPM. Tracks with extremely slow (below 50 BPM) or extremely fast (above 200 BPM) tempos have a significantly lower popularity ceiling.



9. Popularity vs. Valence (Mood/Positivity) :

Popularity is generally achievable across the entire valence spectrum (0.0 to 1.0). However, the data cloud is slightly denser and taller in the high valence (positive mood) range (0.50 to 1.00). This indicates that while sad or dark tracks can be hits, the market may have a slight bias or greater capacity for cheerful, positive



The correlation heatmap is crucial for understanding what drives (or kills) popularity. Here is the short and to-the-point version:

Key Correlation Insights

1. Popularity Drivers & Barriers (vs. Popularity) These correlations, though weak, indicate linear trends for success:

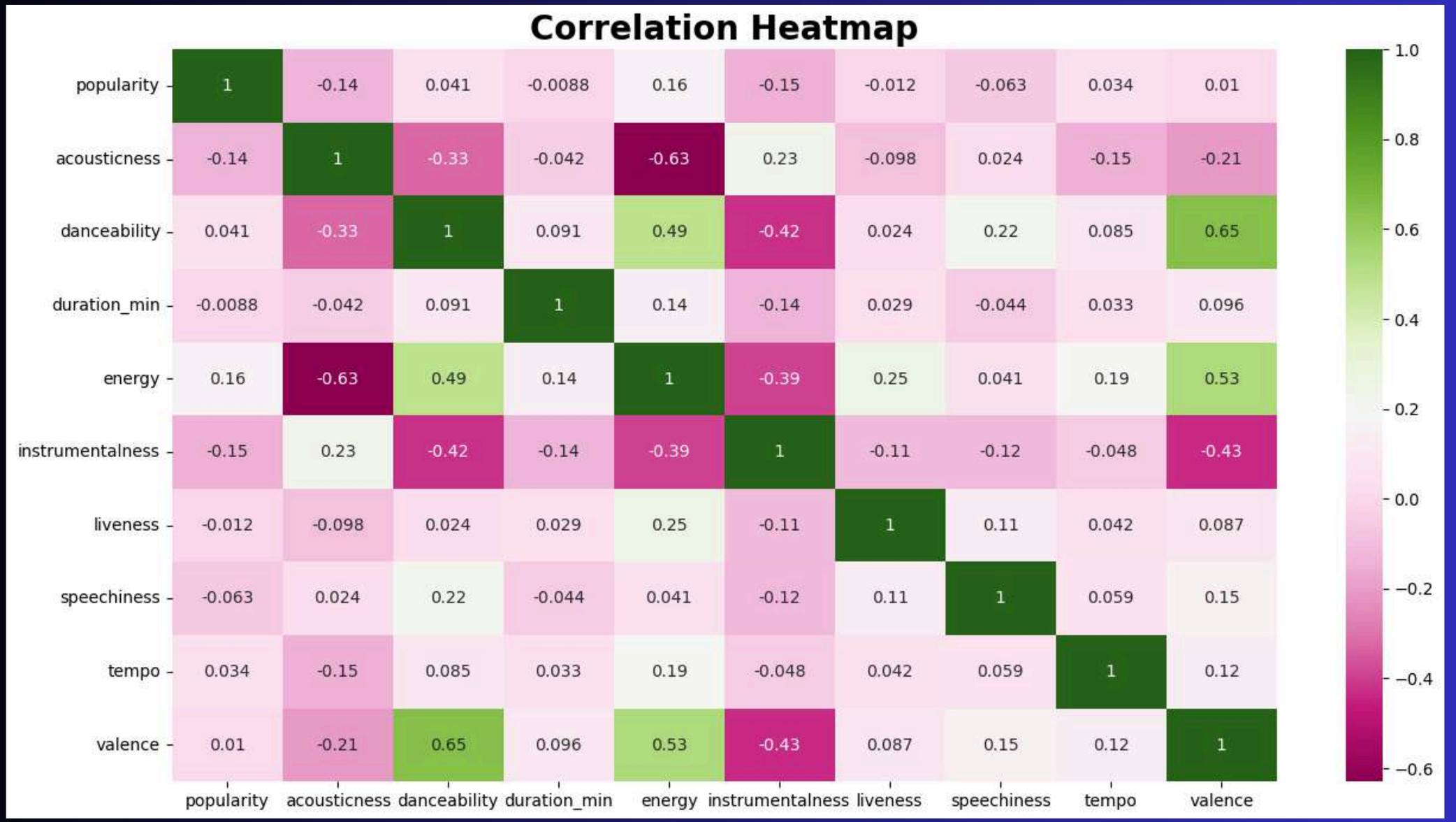
Primary Driver: Energy (r=0.16) is the strongest positive predictor. Upbeat, intense tracks are slightly more likely to be popular.

Primary Barrier: Instrumentalness (r=-0.15) and Acousticness (r=-0.14) are the strongest negative predictors. Music without vocals and organic, non-produced sounds are less likely to be highly popular.

Neutral Factors: Valence (r=0.01) and Duration (r \approx 0.0) have virtually no linear relationship with popularity; hits can be happy or sad, and length is a poor predictor.

2. Strongest Feature Relationships (Feature vs. Feature) These show how song characteristics naturally group together:

Dance & Mood Go Together: Danceability and Valence (positivity) have the strongest positive link (r=0.65). Tracks that are easy to dance to are overwhelmingly positive in mood. The Production Trade-Off: Energy and Acousticness have the strongest negative link (r=-0.63). High-energy music is rarely acoustic, confirming the divide between modern production and organic sound. Instrumental is Low-Mood: Instrumentalness is moderately negative with both Valence (r=-0.43) and Danceability (r=-0.42), suggesting instrumental music tends to be less cheerful or rhythmically simple.


Correlation Heatmap
popularity 1 -0.14 0.041 -0.0088 0.16 -0.15 -0.012 -0.063 0.034 0.01
acousticness -0.14 1 -0.33 -0.042 -0.63 0.23 -0.098 0.024 -0.15 -0.21
danceability 0.041 -0.33 1 0.091 0.49 -0.42 0.024 0.22 0.085 0.65
duration_min -0.0088 -0.042 0.091 1 0.14 -0.14 0.029 -0.044 0.033 0.096
energy 0.16 -0.63 0.49 0.14 1 -0.39 0.25 0.041 0.19 0.53
instrumentalness -0.15 0.23 -0.42 -0.14 -0.39 1 -0.11 -0.12 -0.048 -0.43
liveness -0.012 -0.098 0.024 0.029 0.25 -0.11 1 0.11 0.042 0.087
speechiness -0.063 0.024 0.22 -0.044 0.041 -0.12 0.11 1 0.059 0.15
tempo 0.034 -0.15 0.085 0.033 0.19 -0.048 0.042 0.059 1 0.12
valence 0.01 -0.21 0.65 0.096 0.53 -0.43 0.087 0.15 0.12 1
popularity acousticness danceability duration_min energy instrumentalness liveness speechiness tempo valence
1.0 0.8 0.6 0.4 0.2 0.0 −0.2 −0.4 −0.6

Cluster Insights:

Tempo vs. Popularity, The analysis effectively splits the dataset based on both Tempo (Slow/Fast, likely split around 120-130 BPM) and Popularity (High/Low, likely split around a median/quartile cutoff).

1. High Popularity Clusters (The Hits) Tracks with high popularity are generally found in both slow and fast tempos, but the split is revealing: Slow/HighPop (Blue Cluster): This cluster represents tracks that achieve success despite having a slower tempo (approximately 60 to 125 BPM). These hits often include genres like ballads, mid-tempo pop, or hip-hop that emphasize rhythm over speed.

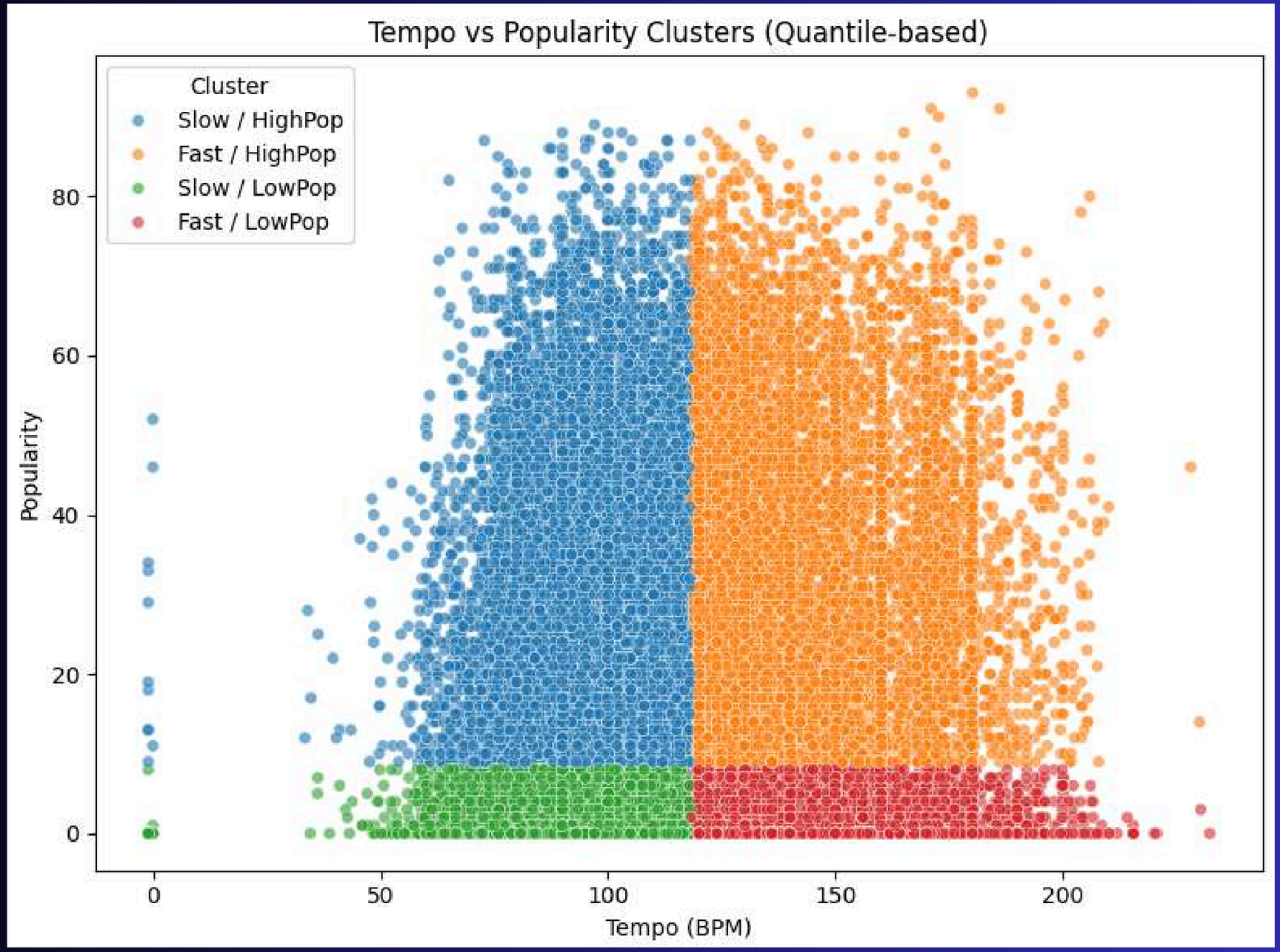
Fast/HighPop (No Cluster): Crucially, there is no designated "Fast/HighPop" cluster in this image. The high popularity scores (Popularity > 10) in the faster tempo range (Tempo > 125 BPM) are absorbed into the Fast/LowPop (Orange Cluster). This means that while fast songs can be hits (points exist high up in the orange area), a much larger proportion of the fast songs do not achieve high popularity compared to the slow songs.

2. Low Popularity Clusters (The Majority) Low popularity is concentrated at both ends of the tempo spectrum:

Fast/LowPop (Orange Cluster): This is the largest cluster in terms of count (volume of data points), occupying the upper range of tempo (approximately 125 to 200 BPM). The density of points is high throughout the low-popularity range (Popularity approx 0-10). This suggests a large volume of fast music is produced that fails to gain traction.

Slow/LowPop (Green Cluster): This cluster represents slow songs that are unpopular. It is clearly the smallest cluster in terms of area/count, occupying a small space at the bottom-left of the plot (Tempo approx 60-125 BPM). This visually suggests that songs in the slow tempo range have a higher overall chance of being popular than fast songs.

Conclusion The distribution suggests that low-tempo tracks are a more "efficient" path to popularity in this dataset. While there are popular tracks in both the slow and fast ranges, the Slow/HighPop cluster is more clearly defined and dense than its fast counterpart, and the volume of successful slow songs appears visually balanced against the unsuccessful ones. Conversely, the Fast/LowPop cluster is enormous, showing that high-tempo production is extremely saturated and highly competitive.

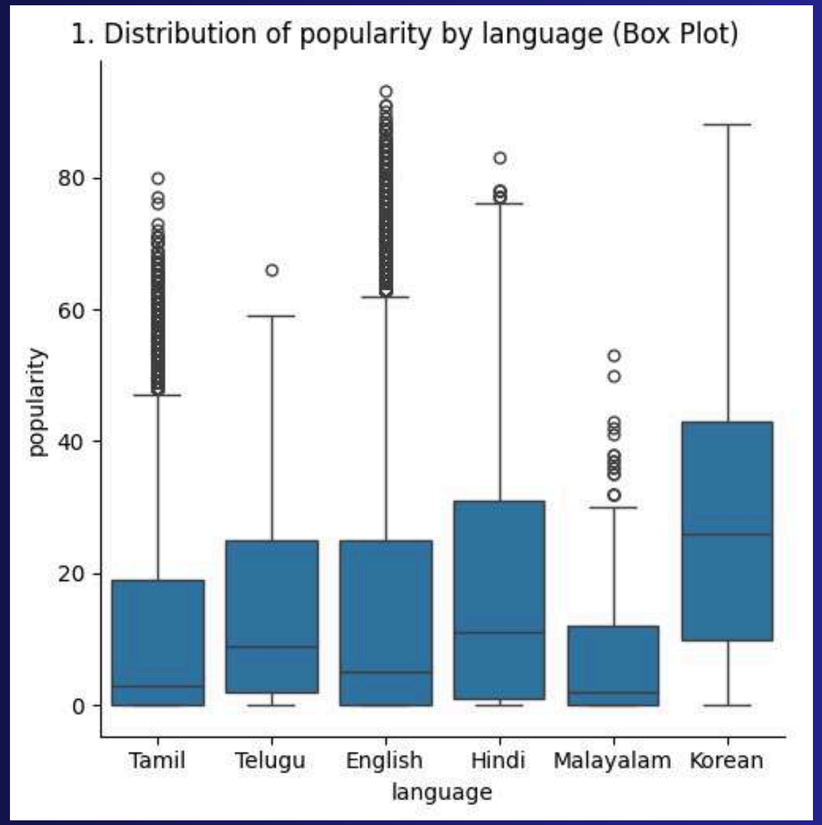

Tempo vs Popularity Clusters (Quantile-based)
Cluster
Slow / HighPop
Fast / HighPop
Slow / LowPop
Fast / LowPop
Popularity
0
20
40
60
80
0
50
100
150
200
Tempo (BPM)

1. Popularity by Language :

English tracks show the highest potential for extreme popularity, as evidenced by the tallest box plot and the highest density of top outliers (approximately 90).

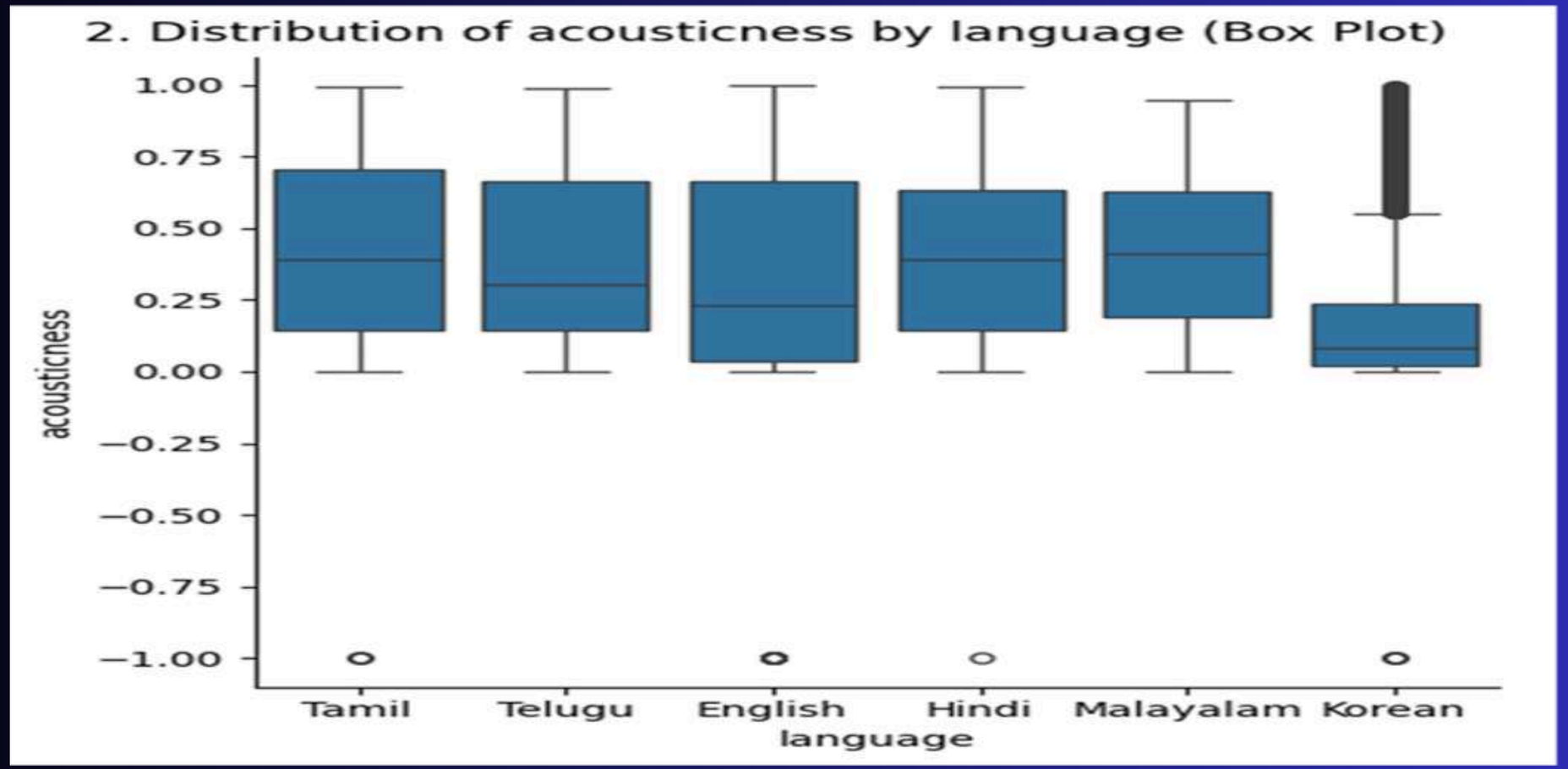
Comparative Median Success: Korean and Hindi tracks have the highest median popularity, indicating that their typical (50th percentile) track is more popular than the median tracks in other languages.

Low Success Floor: Telugu and Malayalam tracks have the lowest popularity floor (25th percentile) and lowest overall median, suggesting tracks in these languages are less likely to achieve high popularity in this dataset.



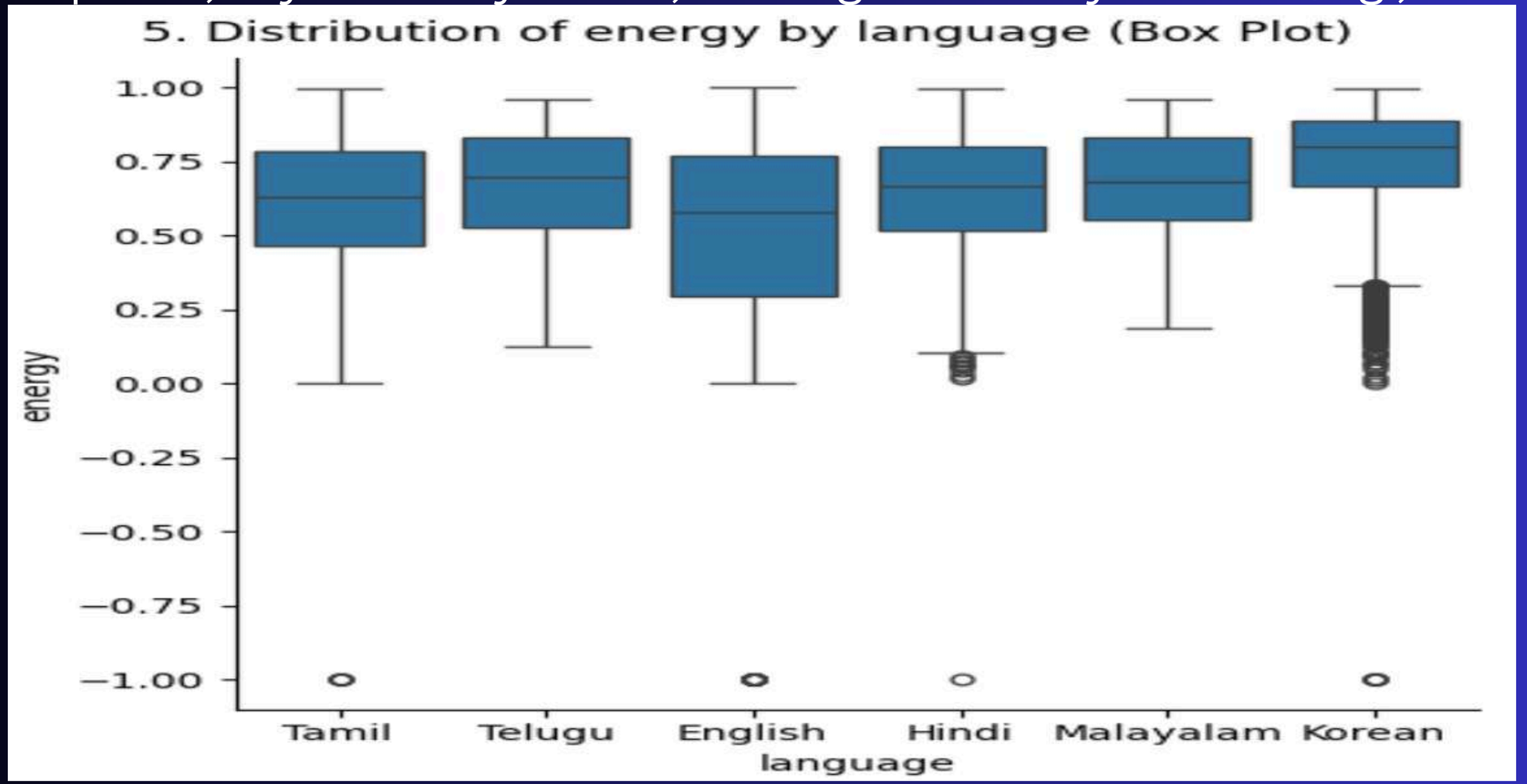
2. Acousticness by Language :

Korean music is significantly less acoustic than all other languages. The median acousticness for Korean is near 0.0, and the box is very compressed at the low end. This suggests Korean music in this dataset is overwhelmingly digital and electronically produced. Highest Acoustic Scores: Hindi, Malayalam, and Tamil have higher median acousticness (0.35 to 0.40), indicating a greater presence of acoustic instrumentation or less heavily produced sound styles.



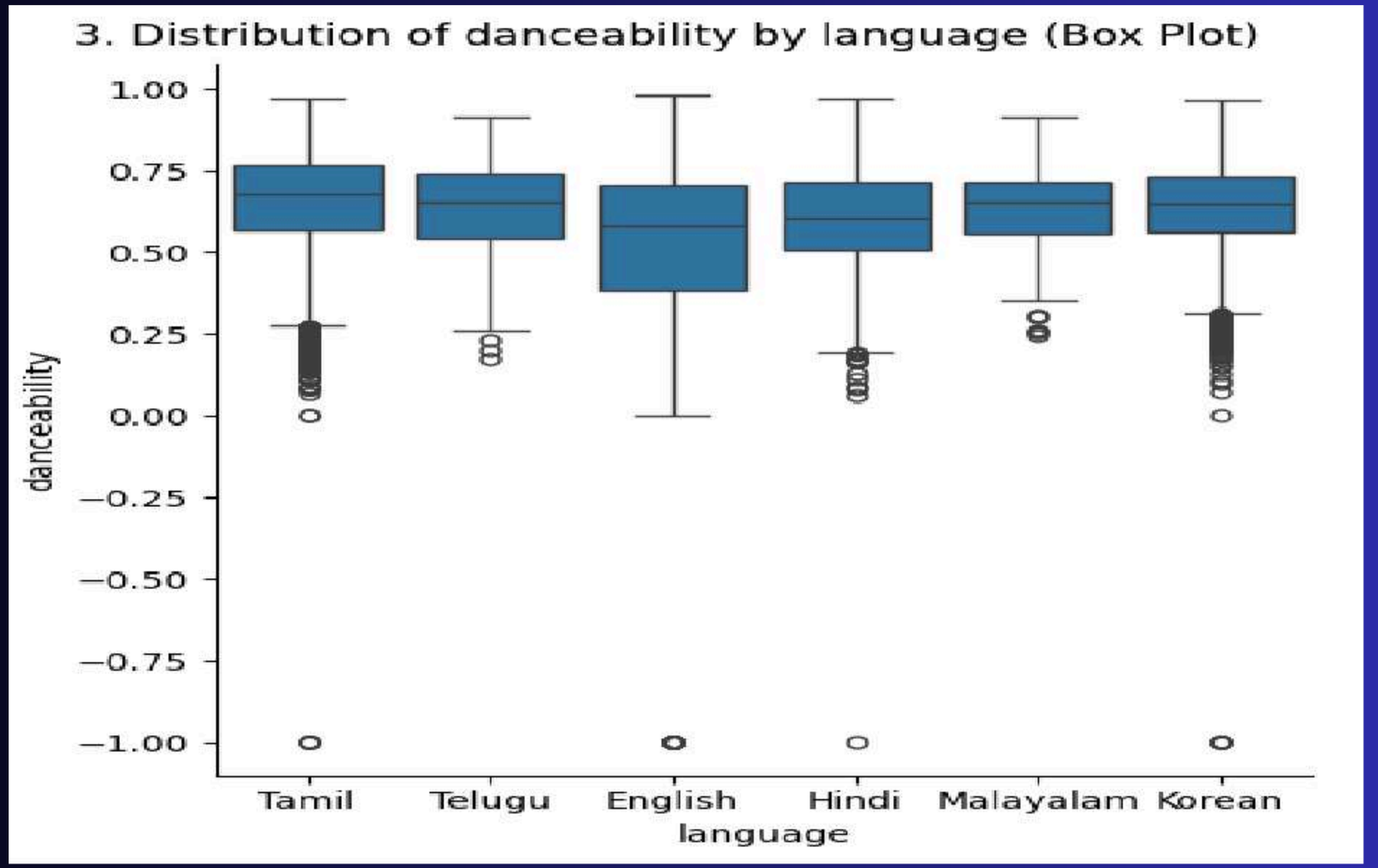
3. and 5. Danceability & Energy by Language:

Korean : Korean tracks are characterized by the second highest median Danceability and the highest median Energy. This aligns with the previous finding and reinforces that the Korean music in this sample is strongly biased toward upbeat, rhythmically active, and high-intensity sounds (e.g., K-Pop).



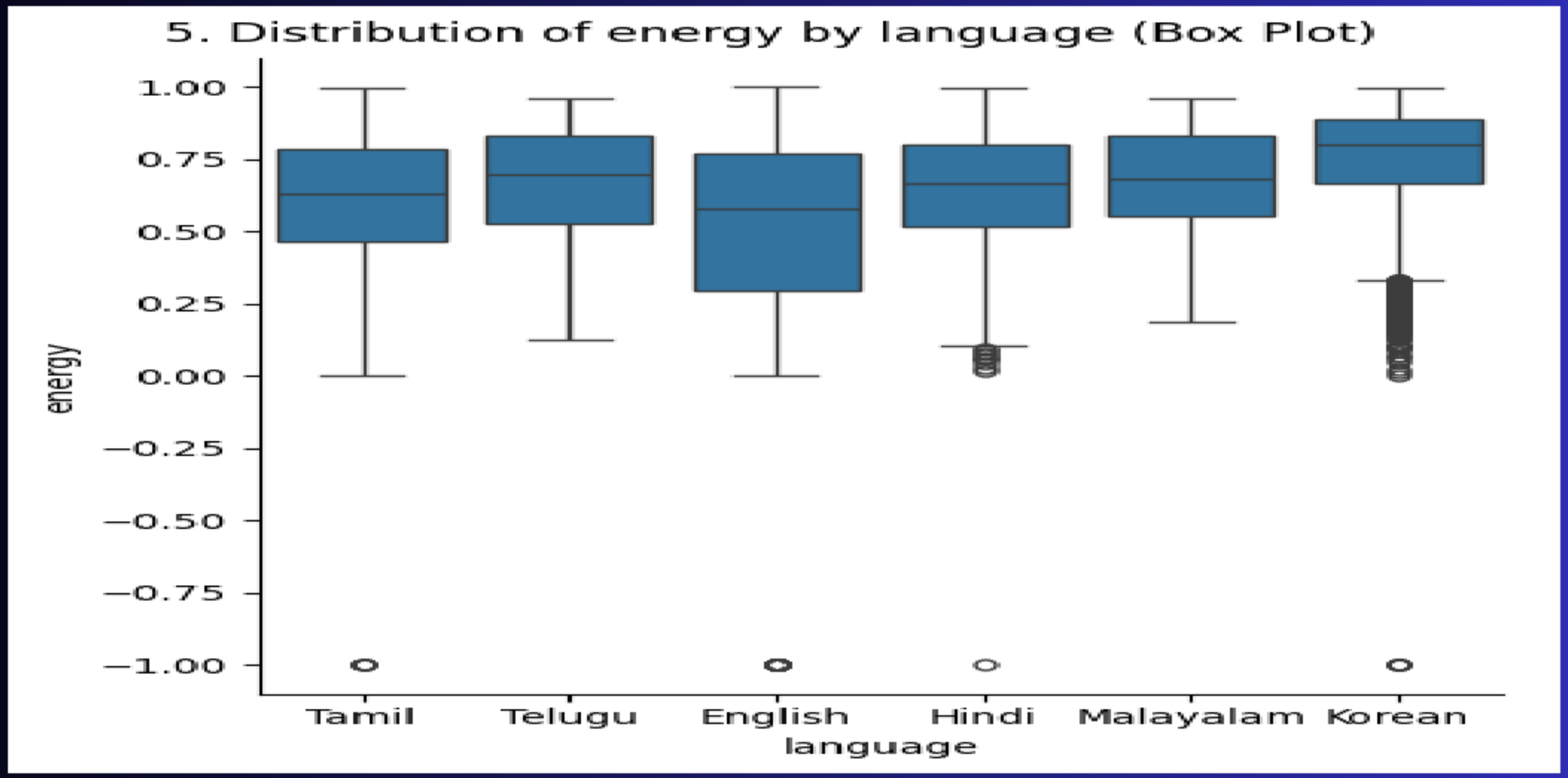
4. Duration by Language :

The core duration (Median) is remarkably consistent across all languages (4 minutes), confirming the universal standard song length. Hindi and English have the most extreme long outliers (up to approximately 75 minutes), likely due to non-standard music content or genre variations.



6. Instrumentalness by Language :

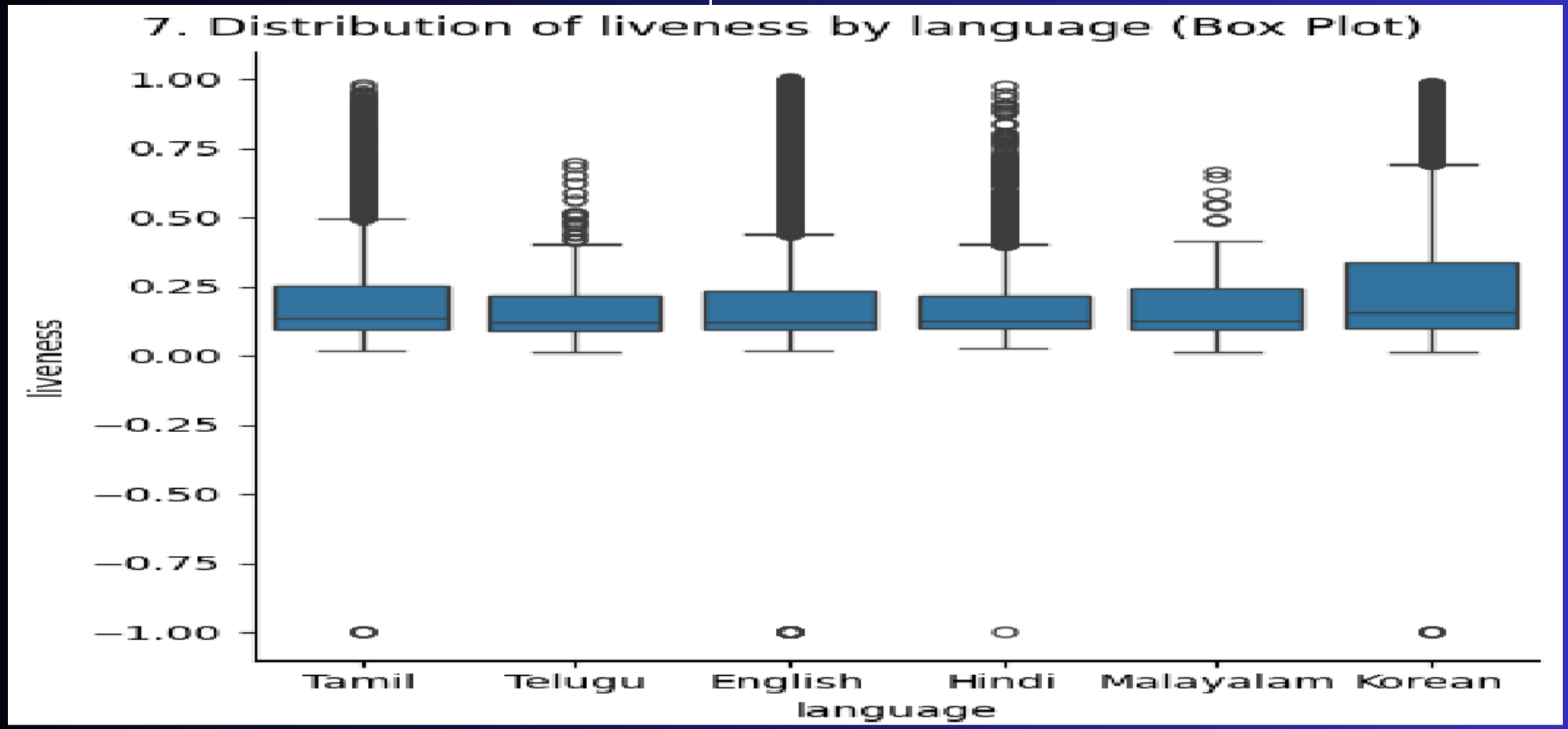
English shows the greatest spread of instrumentalness and has the highest presence of tracks in the upper quartiles. This suggests that the English group contains a wider variety of genres, including more instrumental-heavy subgenres (e.g., electronic, ambient, jazz) compared to the other languages, which are mostly clustered near 0.0 (vocal-centric).



7. Liveness by Language :

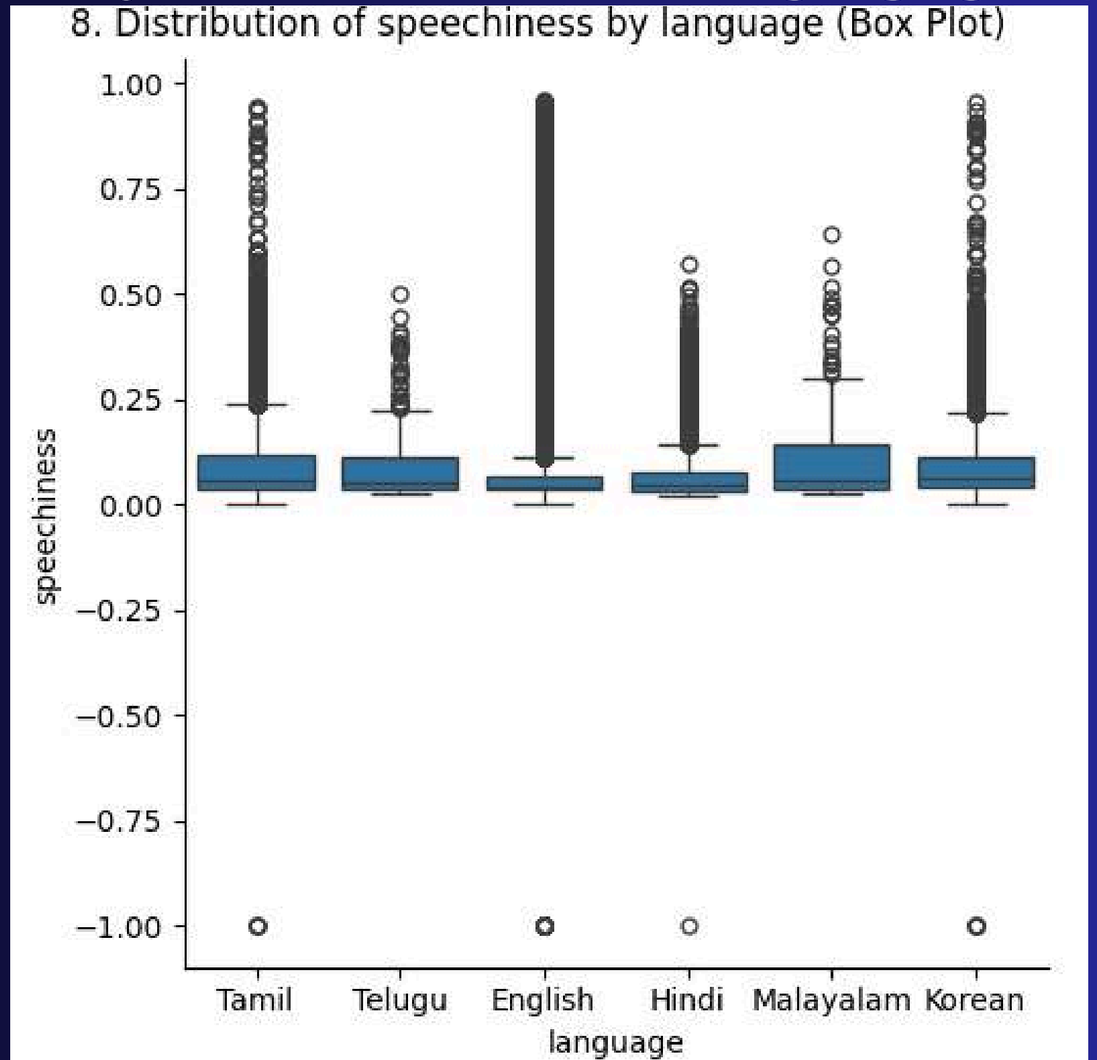
Tamil and Korean tracks tend to have slightly higher median and upper quartile liveness scores (0.25 to 0.30) compared to the other languages. This may indicate a slightly higher frequency of live recordings or performances captured within those language groups.

General Low Liveness: Overall, Liveness remains low across all languages, confirming the general dominance of studio production.



8. Speechiness by Language :

While all languages have a low median speechiness (0.05), Tamil, English and Korean show the tallest outliers reaching 1.0 and the highest density of outliers above 0.50. This suggests that tracks with substantial spoken-word content (e.g., rap, podcasts, dialogue) are most commonly found in these two language groups.

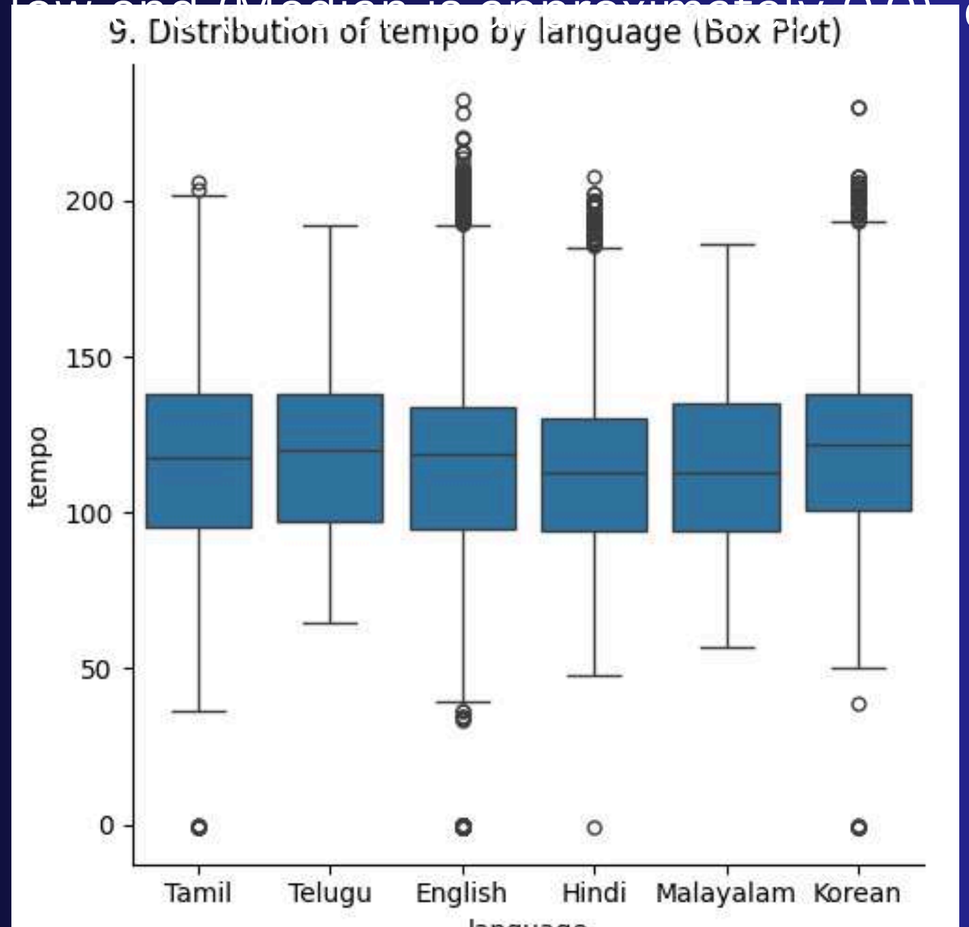


9.Tempo by Language:

Tempo is highly consistent across all languages, with the median falling between 100 and 120 BPM. Hindi is slightly slower overall (lower median), while English shows the largest range of high-tempo outliers

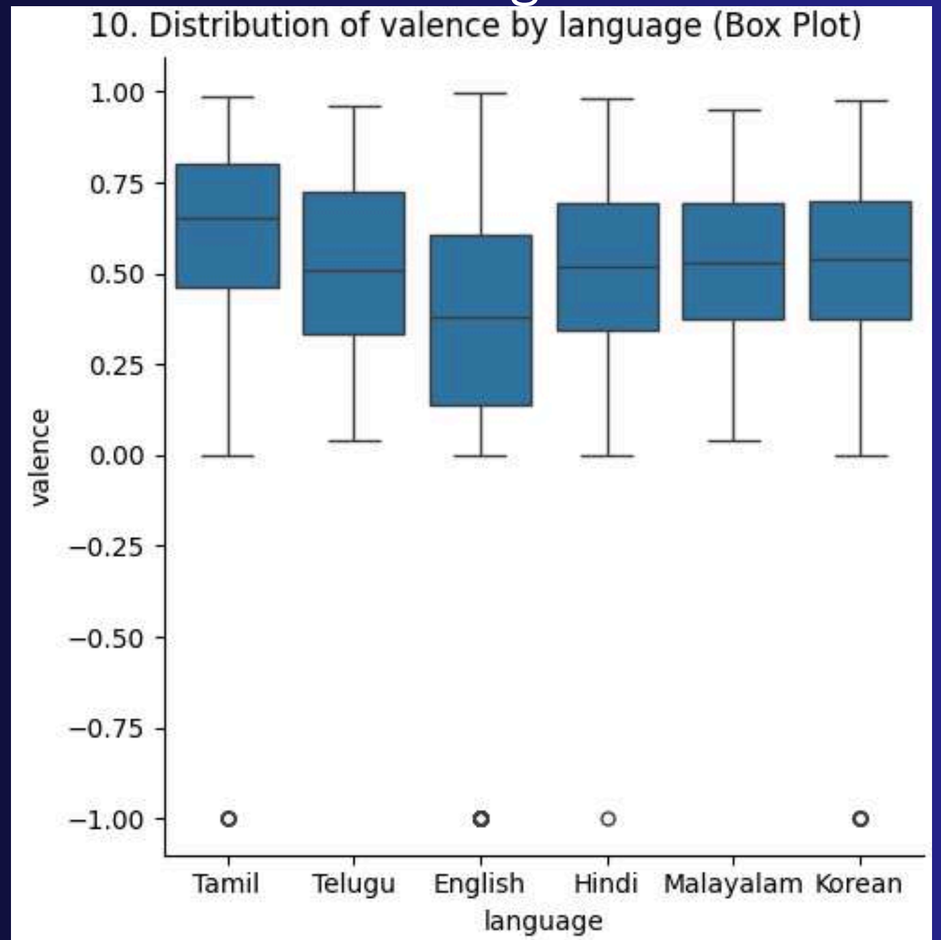
General Consistency: All language groups maintain a generally high level of median danceability (0.60) and energy (0.60), but Korean consistently sits at the top quartile

Vocal Dominance: For all South Asian languages (Tamil, Telugu, Hindi, Malayalam), instrumentalness is tightly clustered at the extreme low end (Median is approximately 0.0), emphasizing their focus on sung vocals.



10. Valence (Mood) by Language :

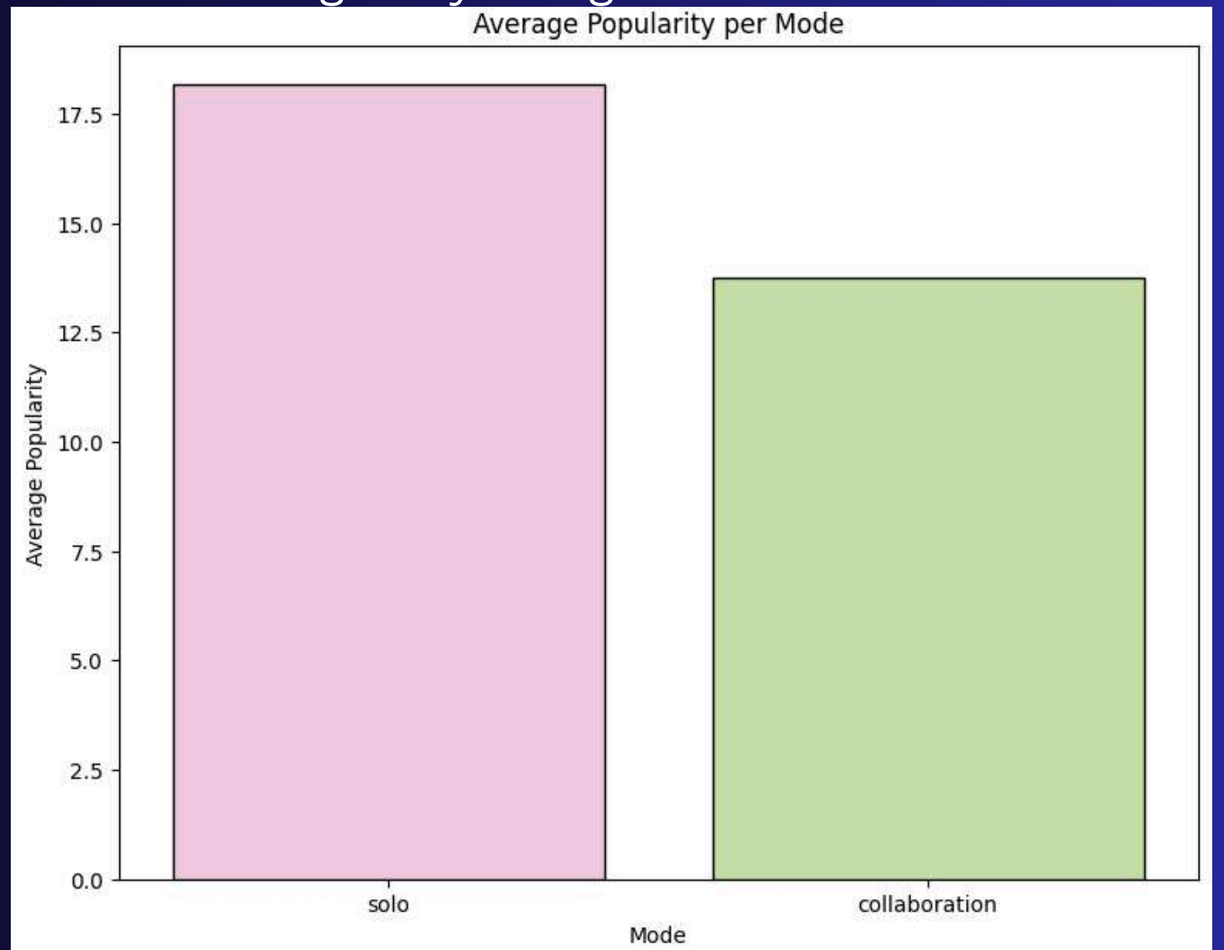
Tamil tracks exhibit the highest median and upper quartile valence, suggesting a stronger tendency toward positive, cheerful, or happy moods compared to the other languages. Neutral-to-Positive: Most other languages have a median valence near 0.50, indicating an overall balance between positive and negative moods, but all show a box extending well into the positive range. But English tracks show a noticeable dip in the lower quartile, indicating a considerable presence of negative songs.



This bar chart compares the average popularity of Solo tracks versus Collaboration tracks. Solo Tracks are More Popular: The average popularity of Solo tracks is visually higher (approx 17.8) than the average popularity of Collaboration tracks (approx 13.8).

Contradictory Finding: This contradicts the common industry assumption that collaborations always boost popularity by combining fanbases. In this dataset, tracks by a single artist are, on average, more successful.

Hypothesis: This result suggests that the inclusion of collaborations in this specific dataset may involve many less-popular, niche, or unknown team-ups, bringing the average down. The highly popular tracks are often successful studio singles by a single established artist.



This bar chart compares the average popularity of tracks categorized by their musical 'mode' (key: Major, Minor, or Undetected).

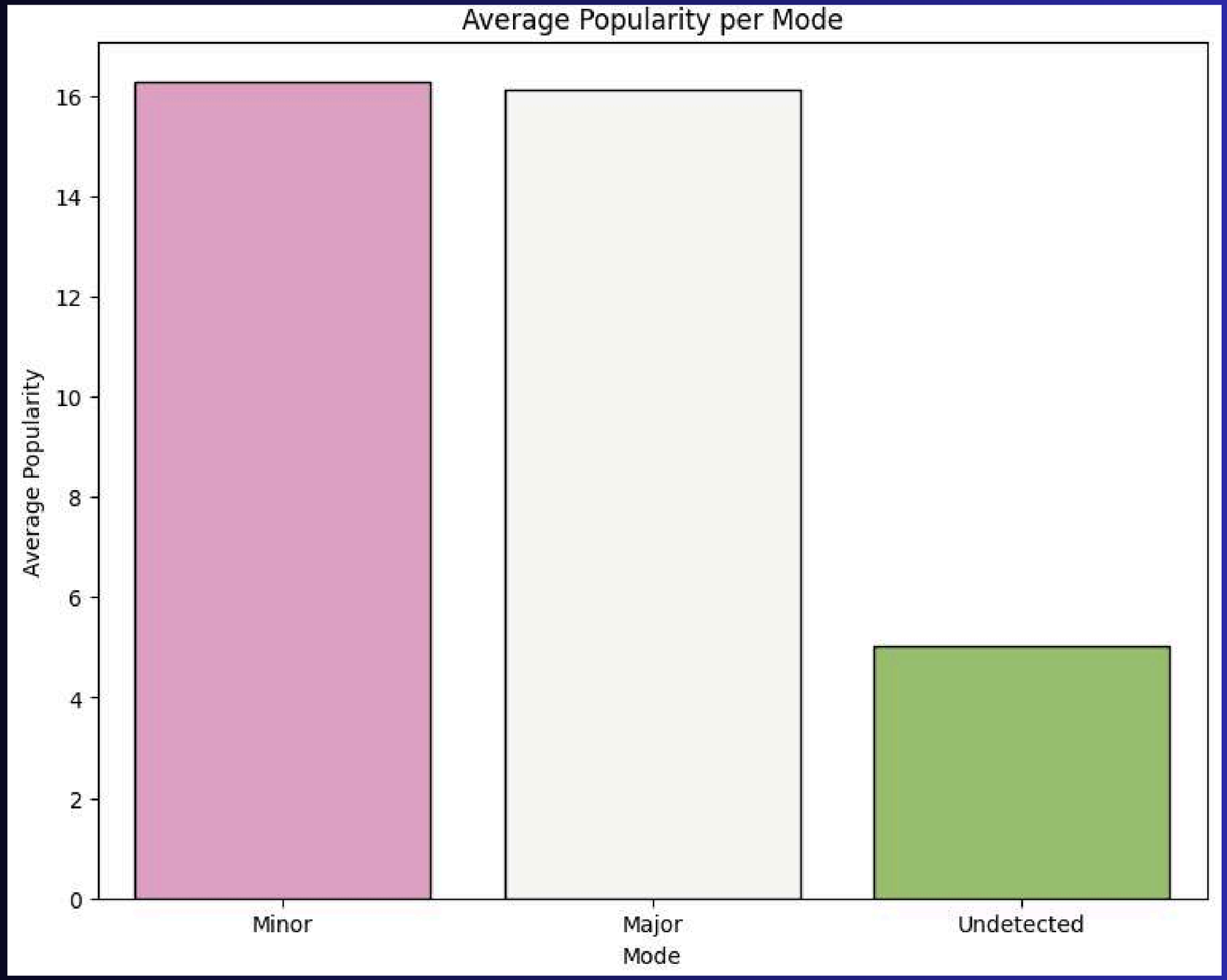
The analysis shows that there is no significant difference in average popularity between Major and Minor key tracks, while tracks whose mode could not be detected are much less popular.

Equal Success for Major and Minor: Tracks in a Minor key have the highest average popularity (approx 16.2). Tracks in a Major key are almost equally popular (approx 16.0).

Insight: This suggests that the emotional context of the key—whether happy/bright (Major) or sad/dark (Minor)—does not determine average commercial success. Both moods are equally capable of achieving widespread popularity.

Low Popularity for Undetected Mode: Tracks with an Undetected mode show a drastically lower average popularity (approx 5.0).

Insight: "Undetected" mode often points to music that is highly experimental, extremely ambient, lacks a clear harmonic center (atonality), or has poor data quality. This finding indicates that music with unconventional or ambiguous harmonic structure is overwhelmingly unpopular.

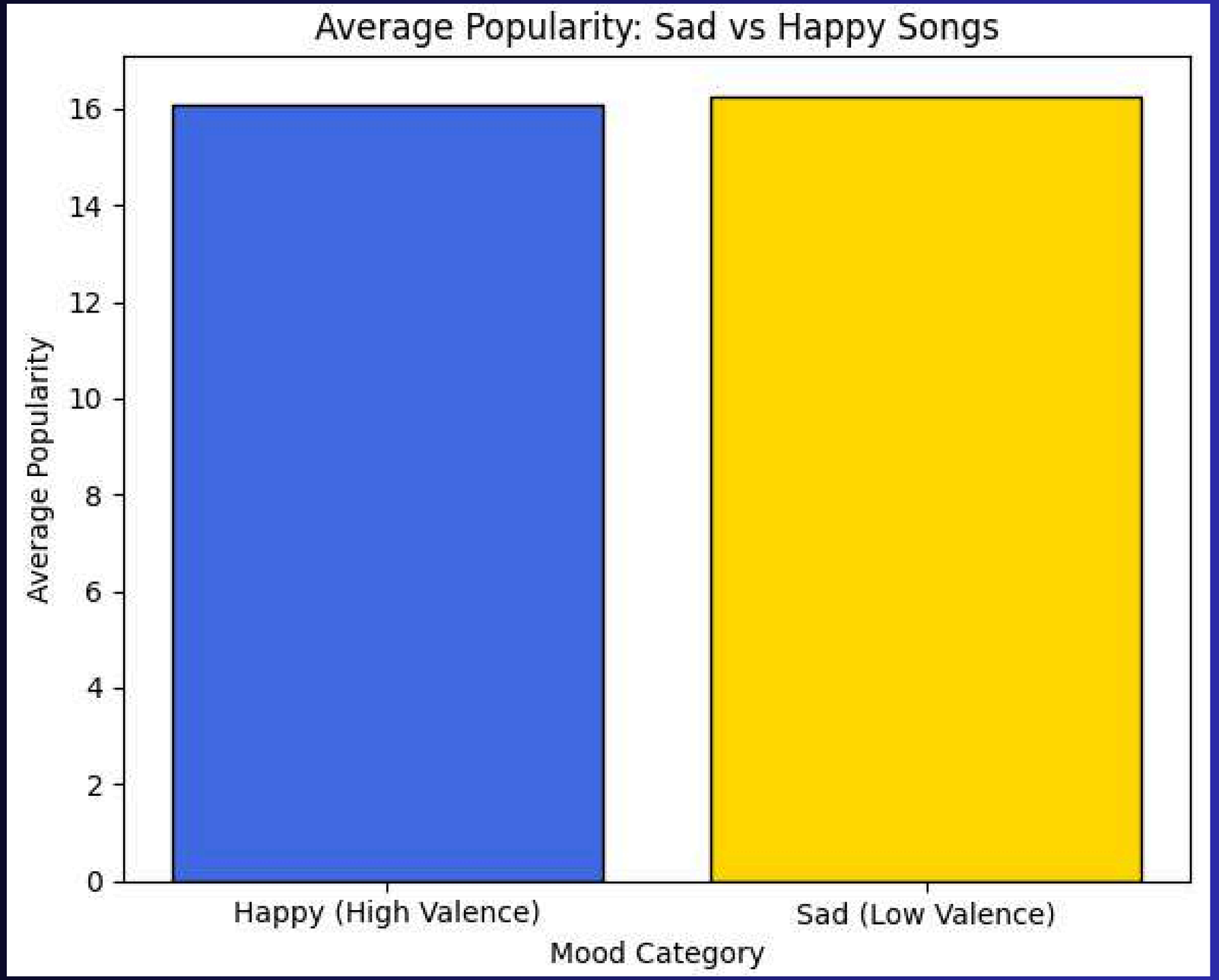

Average Popularity per Mode
Average Popularity
0
2
4
6
8
10
12
14
16
Minor
Major
Undetected
Mode

This bar chart compares the average popularity of songs categorized by their 'valence' (mood) into 'Happy' (High Valence) and 'Sad' (Low Valence).

Insights on Popularity by Song Mood The analysis shows a remarkable finding regarding song mood and success: Mood is Not a Predictor of Average Popularity: Tracks categorized as Happy (High Valence) have an average popularity of 16.087, while tracks categorized as Sad (Low Valence) have an average popularity of 16.262.

Equal Success: The average popularity is virtually identical across both mood categories. The minuscule difference (Sad songs being slightly, but negligibly, higher) confirms the finding from the correlation heatmap: the emotional tone of a track (positive or negative) does not, on average, determine its level of commercial success.

Conclusion: This indicates that the market has an equal appetite for both happy/upbeat and sad/melancholic music; both moods are equally capable of achieving widespread average popularity.

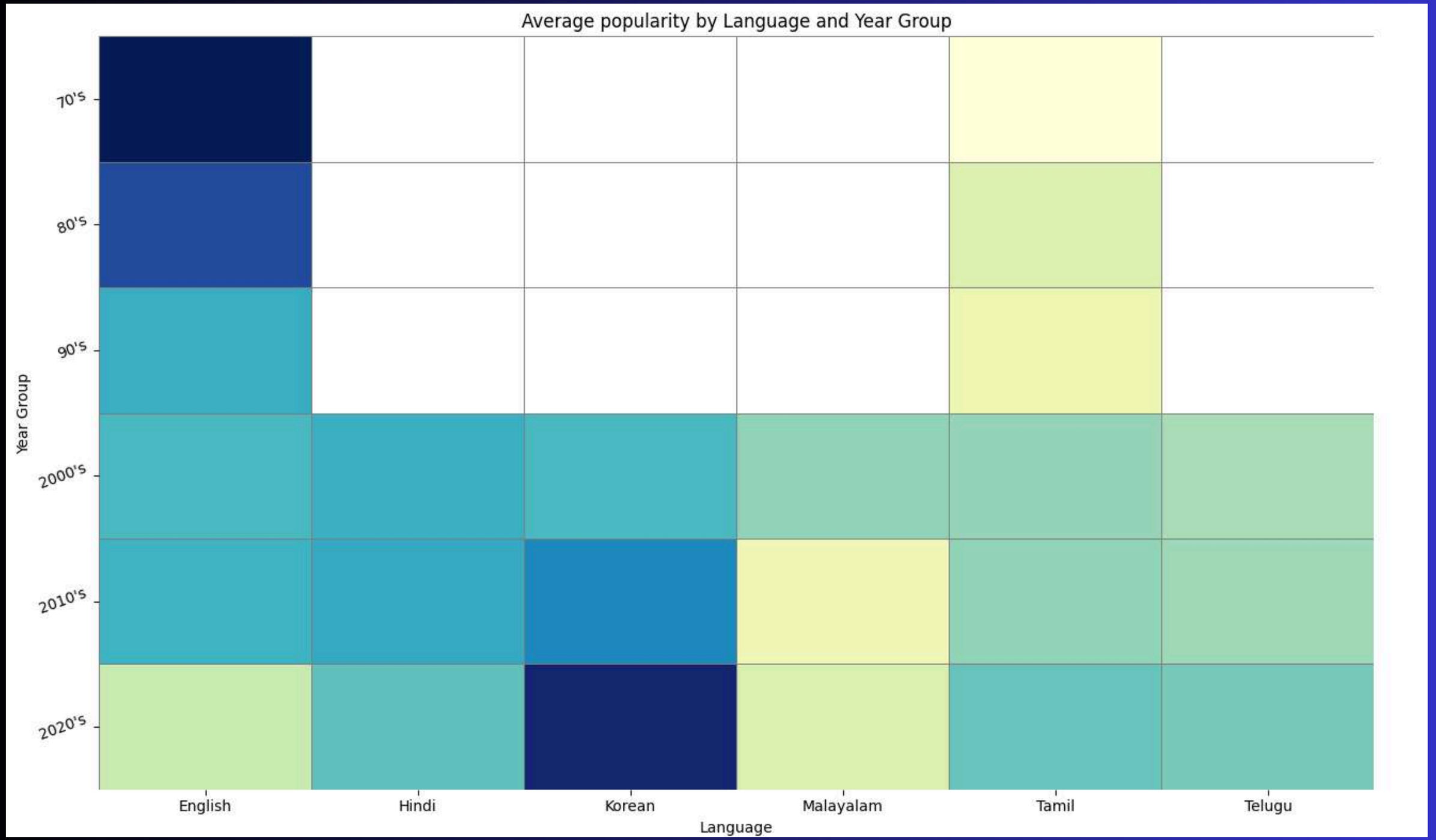

Average Popularity: Sad vs Happy Songs
Average Popularity
0
2
4
6
8
10
12
14
16
Happy (High Valence)
Sad (Low Valence)
Mood Category

1. Average Popularity by Language and Year Group:

Key Insight (Temporal): Older English tracks dominate the historical popularity in this dataset. English music from the 70s, 80s, and 90s shows the highest average popularity scores (darkest blue).

Key Insight (Contemporary): For the most recent decade (2020s), Korean music shows the highest average popularity, while English popularity has comparatively decreased from its historical peaks.

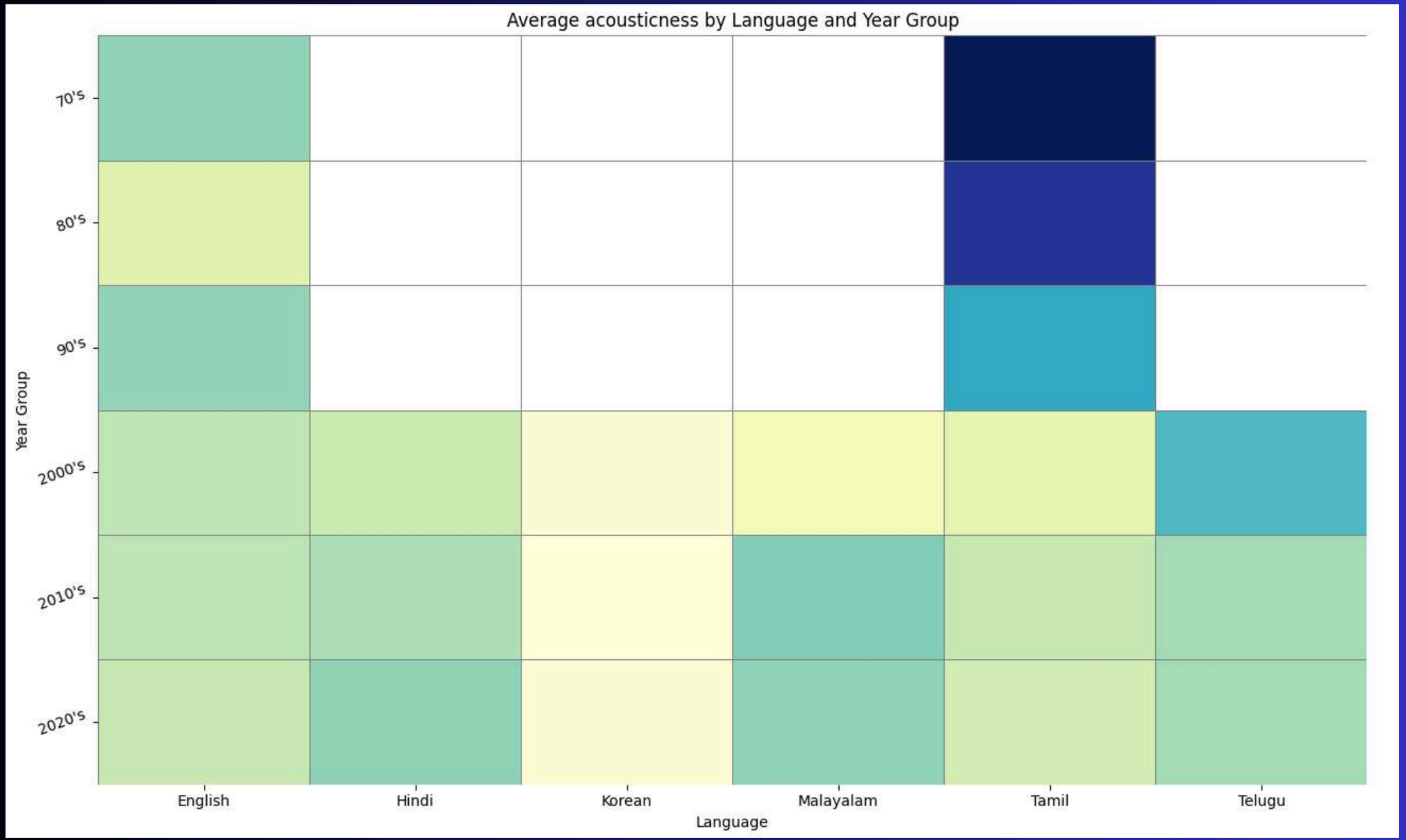
Key Insight (Language Group): Malayalam music consistently shows lower average popularity (lighter colors) across all time periods compared to English and Korean.


Average popularity by Language and Year Group
Year Group
70's
80's
90's
2000's
2010's
2020's
English
Hindi
Korean
Malayalam
Tamil
Telugu
Language

2. Average Acousticness by Language and Year Group:

Key Insight (Temporal Shift): There is a universal trend of decreasing acousticness over time (except in Hindi and Malayalam). For every language that has data across multiple decades (especially English and Tamil), the colors move from dark/medium (high acousticness) in the 70s/90s to light (low acousticness) in the 2010s/2020s. This shows a global shift toward digital, non-acoustic production.

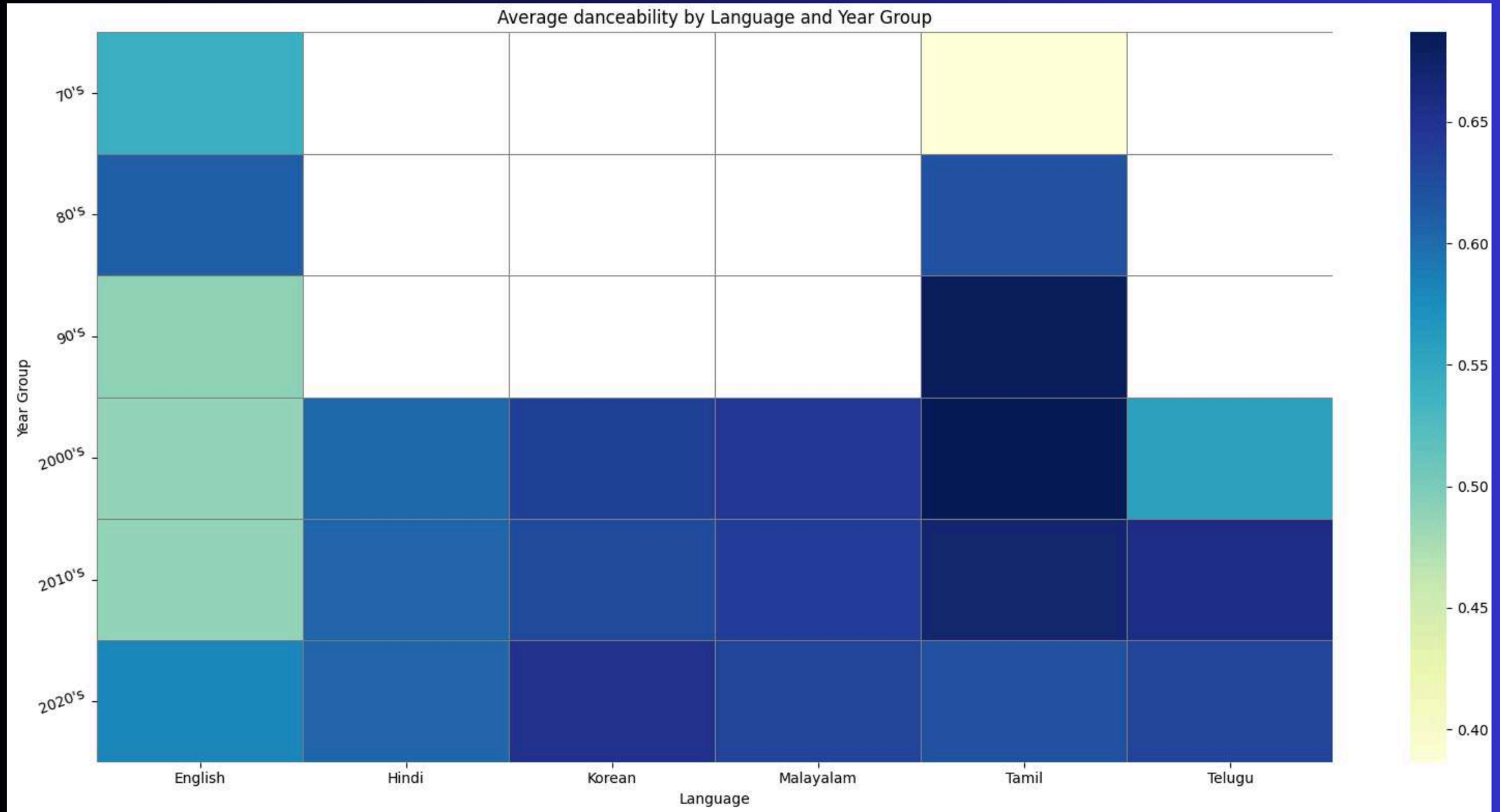
Key Insight (Language Contrast): Tamil music from the 70s and 80s had the highest average acousticness (darkest blue), while contemporary music from all groups (2010s/2020s) is highly non-acoustic.


Average acousticness by Language and Year Group
Year Group
70's
80's
90's
2000's
2010's
2020's
English
Hindi
Korean
Malayalam
Tamil
Telugu
Language

3. Average Danceability by Language and Year Group:

Key Insight (Modern Dance Focus): There is a strong trend toward increased danceability in recent years across several languages. Korean, Malayalam, and Telugu music in the 2010s and 2020s are consistently deep blue, indicating very high average danceability.

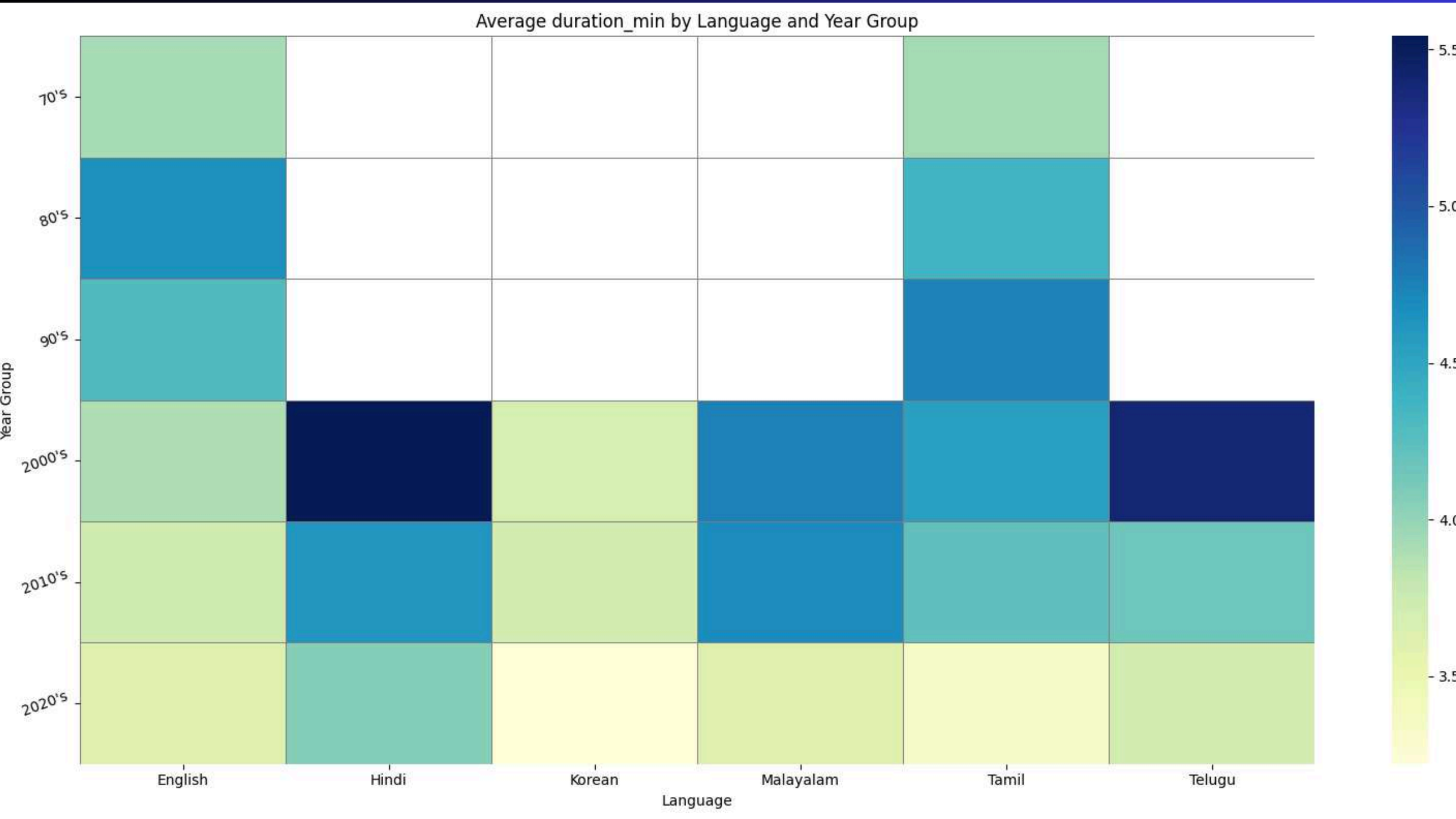
Key Insight (Cultural Difference): Korean, Malayalam, and Telugu consistently have higher modern danceability than English music in the 2010s.


Average danceability by Language and Year Group
Year Group
70's
80's
90's
2000's
2010's
2020's
English
Hindi
Korean
Malayalam
Tamil
Telugu
Language
0.65
0.60
0.55
0.50
0.45
0.40

4. Average Duration (min) by Language and Year Group:

Key Insight (Historical Length): English and Tamil music from the 90s and 80s had the longest average track durations. This reflects the era of longer album tracks and less streaming pressure.
Key Insight (Modern Standardization): Most modern music (2010s/2020s) across all languages shows a standardized, shorter average duration (lighter colors), confirming the pressure to reduce track length for streaming and rapid consumption.

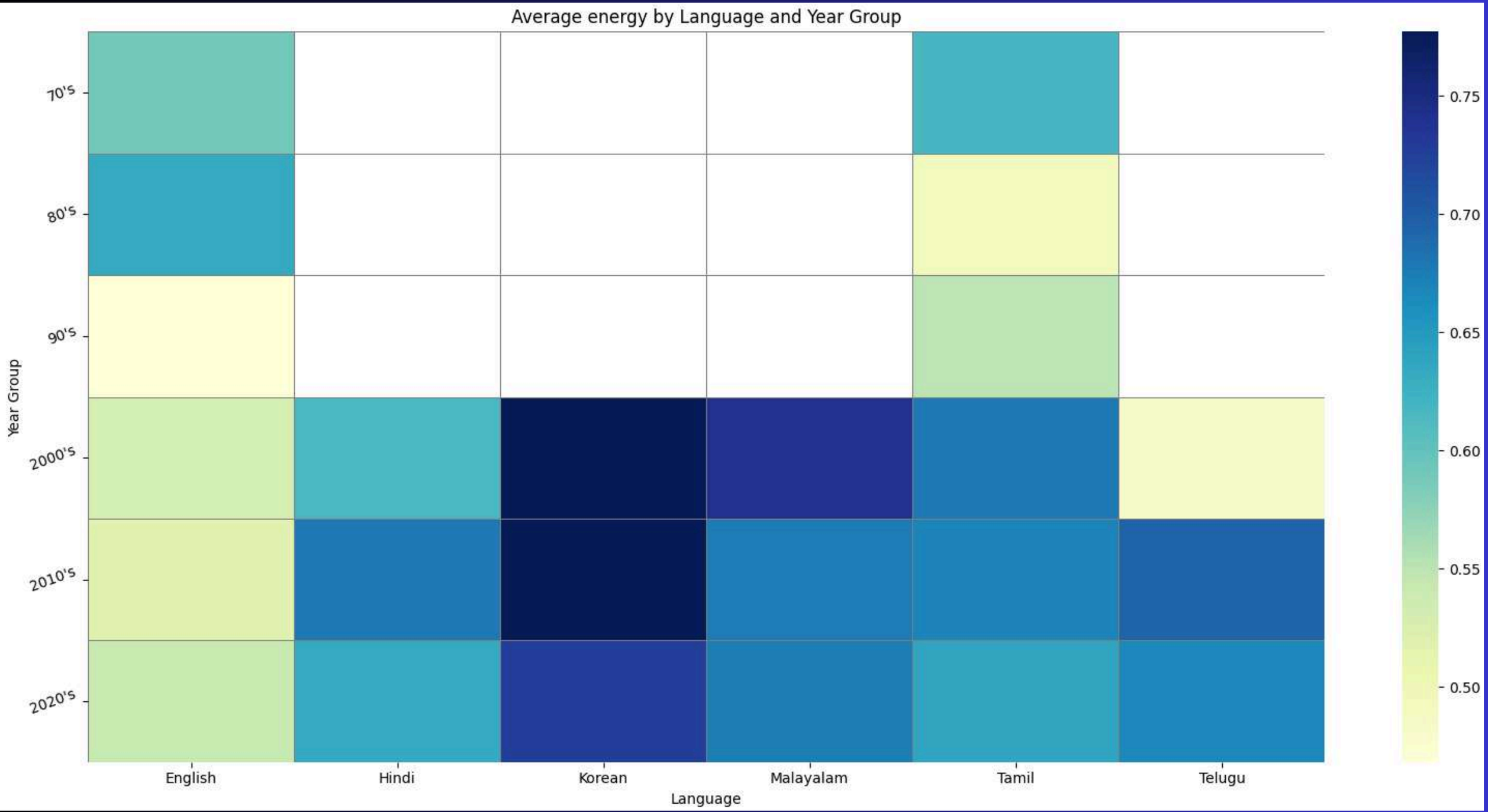
Key Insight (Outlier): Hindi and Telugu music in the 2000s also showed a trend toward longer average tracks (dark squares).


Average duration_min by Language and Year Group
70's
80's
90's
2000's
2010's
2020's
Year Group
English
Hindi
Korean
Malayalam
Tamil
Telugu
Language
5.5
5.0
4.5
4.0
3.5

5. Average Energy by Language and Year Group:

Key Insight (Energy Decline): All languages have experienced decline in energy levels in 2020s. Although some languages (Hindi, Tamil & Telugu) had experienced an intermediate energy surge.

Key Insight (English Trend): English music has been consistently low in energy (yellow/light green) for the past three decades, suggesting its energy levels have been stable, while other languages have experienced periods of rapid energy increase.

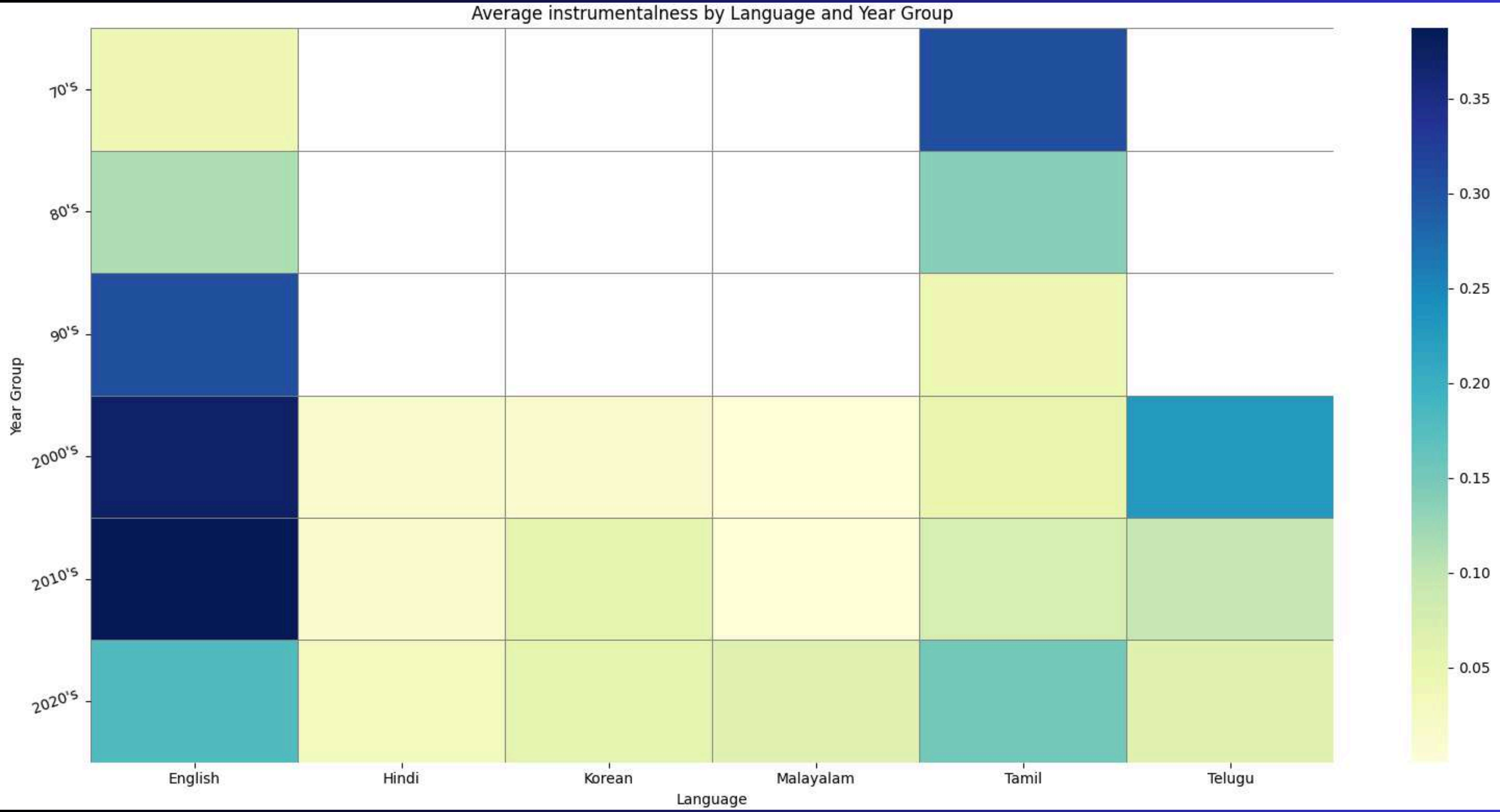

Average energy by Language and Year Group
Year Group
70's
80's
90's
2000's
2010's
2020's
English
Hindi
Korean
Malayalam
Tamil
Telugu
Language
0.75
0.70
0.65
0.60
0.55
0.50

6. Average Instrumentalness by Language and Year Group:

Key Insight (Niche Concentration): Instrumental music has never been a core feature. The darkest cells are scattered: 90s, 2000s and 2010s English and 70sTamil show the highest instrumental presence, likely due to niche subgenres of rock, jazz, or film scores from those periods.

Key Insight (Korean, Malayalam and Hindi): Korean, Malayalam and Hindi music consistently show the lowest instrumentalness (lightest colors) across all recent decades, confirming the finding that they are almost exclusively vocal-centric.

Key Insight (Modern English): English music shows a slight resurgence in instrumentalness in the 2020s (light blue), possibly reflecting the growth of instrumental electronic or ambient genres in the global market.

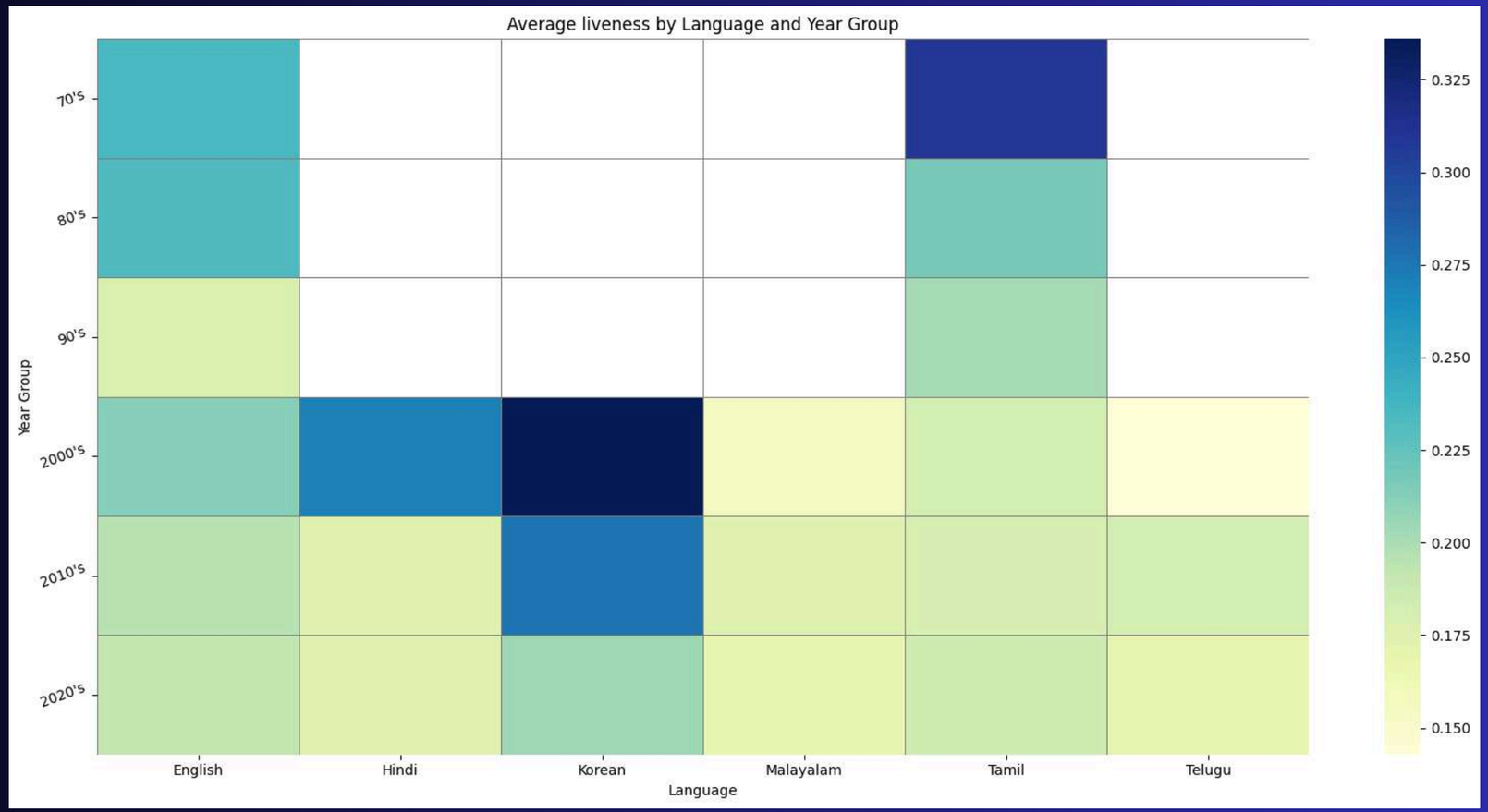

Average instrumentalness by Language and Year Group
Year Group
70's
80's
90's
2000's
2010's
2020's
English
Hindi
Korean
Malayalam
Tamil
Telugu
Language
0.35
0.30
0.25
0.20
0.15
0.10
0.05

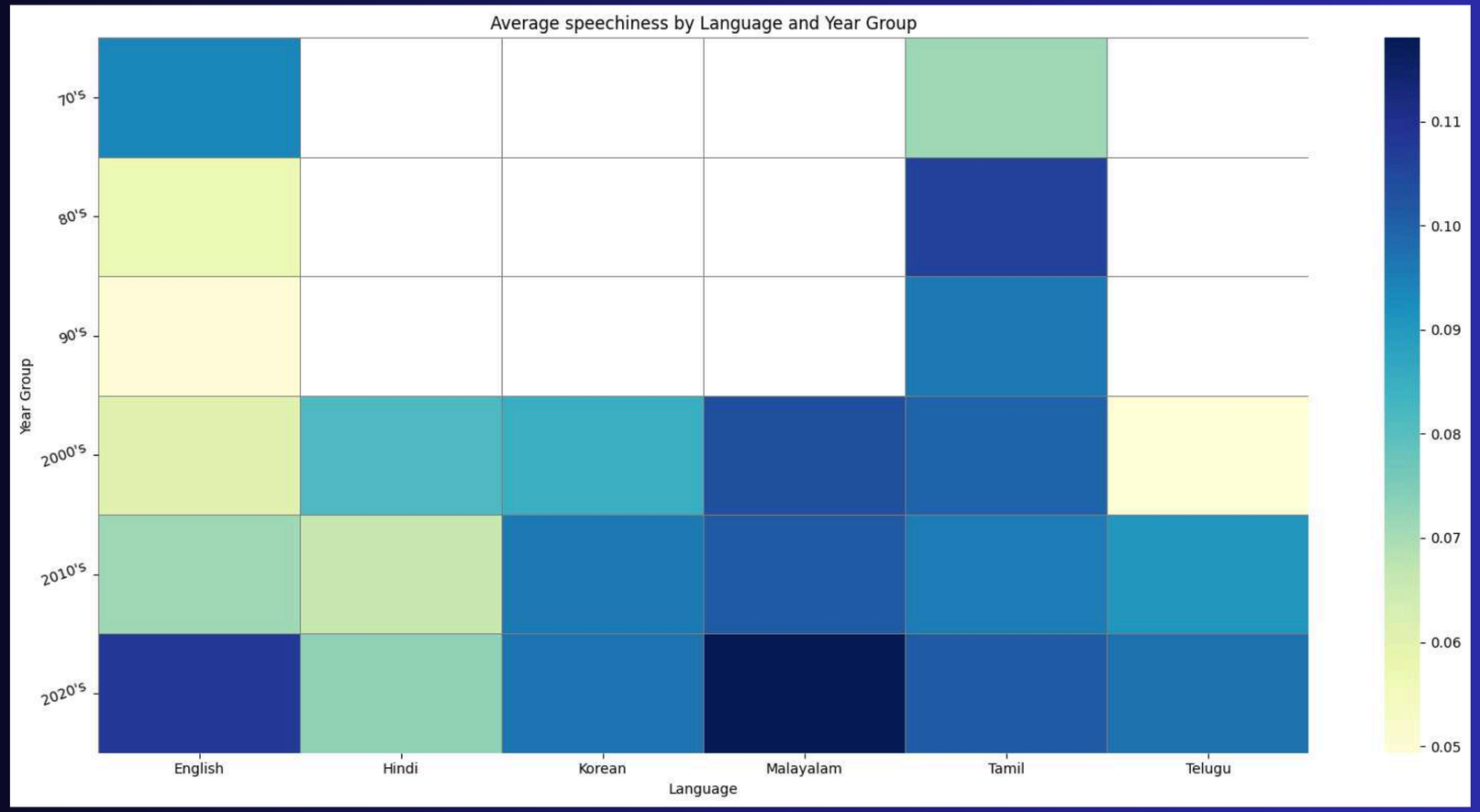
7. Average Liveness by Language and Year Group:

Key Insight (Historical Liveness): The darkest cells are concentrated in theb70s/80s (English and Tamil) and the 2000s (Hindi and Korean). These represent periods or genres where live performance recordings were more common or more frequently included in the dataset.

Key Insight (Modern Decline): Most languages in the 2010s/2020s are consistently very light (low liveness), reinforcing the shift towards polished studio recordings.

Key Insight (Korean Anomaly): Korean music had a period of very high liveness in the 2000s (darkest square), but this dropped significantly by the 2010s, perhaps reflecting a shift in data collection or K-Pop production styles.




Average speechiness by Language and Year Group
Year Group
70's
80's
90's
2000's
2010's
2020's
English
Hindi
Korean
Malayalam
Tamil
Telugu
Language
0.11
0.10
0.09
0.08
0.07
0.06
0.05

9. Average Tempo (BPM) by Language and Year Group:

Key Insight (Historical Speed): 70s/80s English music show the highest average tempo (darkest green/yellow). This may reflect the popularity of fast-paced dance and rock from those decades.
Key Insight (Tempo Slowdown): English music shows a clear slowdown trend, moving from fast (dark green) in the 70s to much slower (light green/yellow) in the 2020s.

Key Insight (2000s Spike): Malayalam music shows a distinct, localized spike in tempo in the 2000s (darkest blue/black), indicating a specific genre trend for that period. Most languages have stabilized in the moderate tempo range (110-125 BPM).

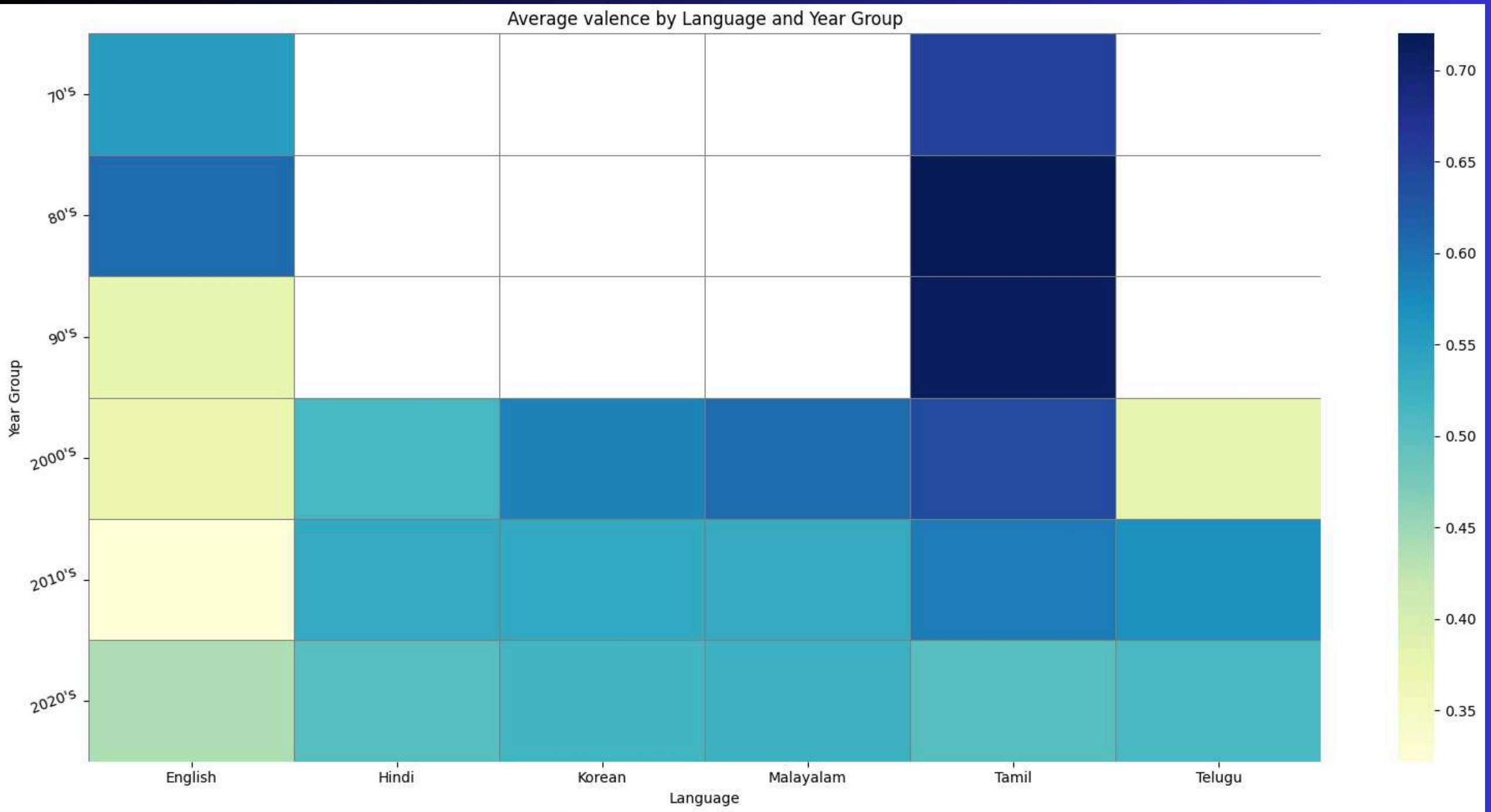

Average tempo by Language and Year Group
Year Group
70's
80's
90's
2000's
2010's
2020's
English
Hindi
Korean
Malayalam
Tamil
Telugu
Language

10. Average Valence (Mood) by Language and Year Group:

Key Insight (Historical Mood): 70s/80s English and 90s/80s Tamil music had the most positive average valence (darkest blue/black).

Key Insight (Universal Decline in Positivity): There is a near-universal trend toward decreasing valence (more neutral/sad moods) across all languages over time. The colors become universally lighter/bluer (lower valence) in the 2010s/2020s.

Key Insight (English and Tamil): English and Tamil music in 2020s show the lowest average valence (lightest blue), suggesting contemporary music in these languages is, on average, the least cheerful.


Average valence by Language and Year Group
Year Group
70's
80's
90's
2000's
2010's
2020's
English
Hindi
Korean
Malayalam
Tamil
Telugu
Language
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35

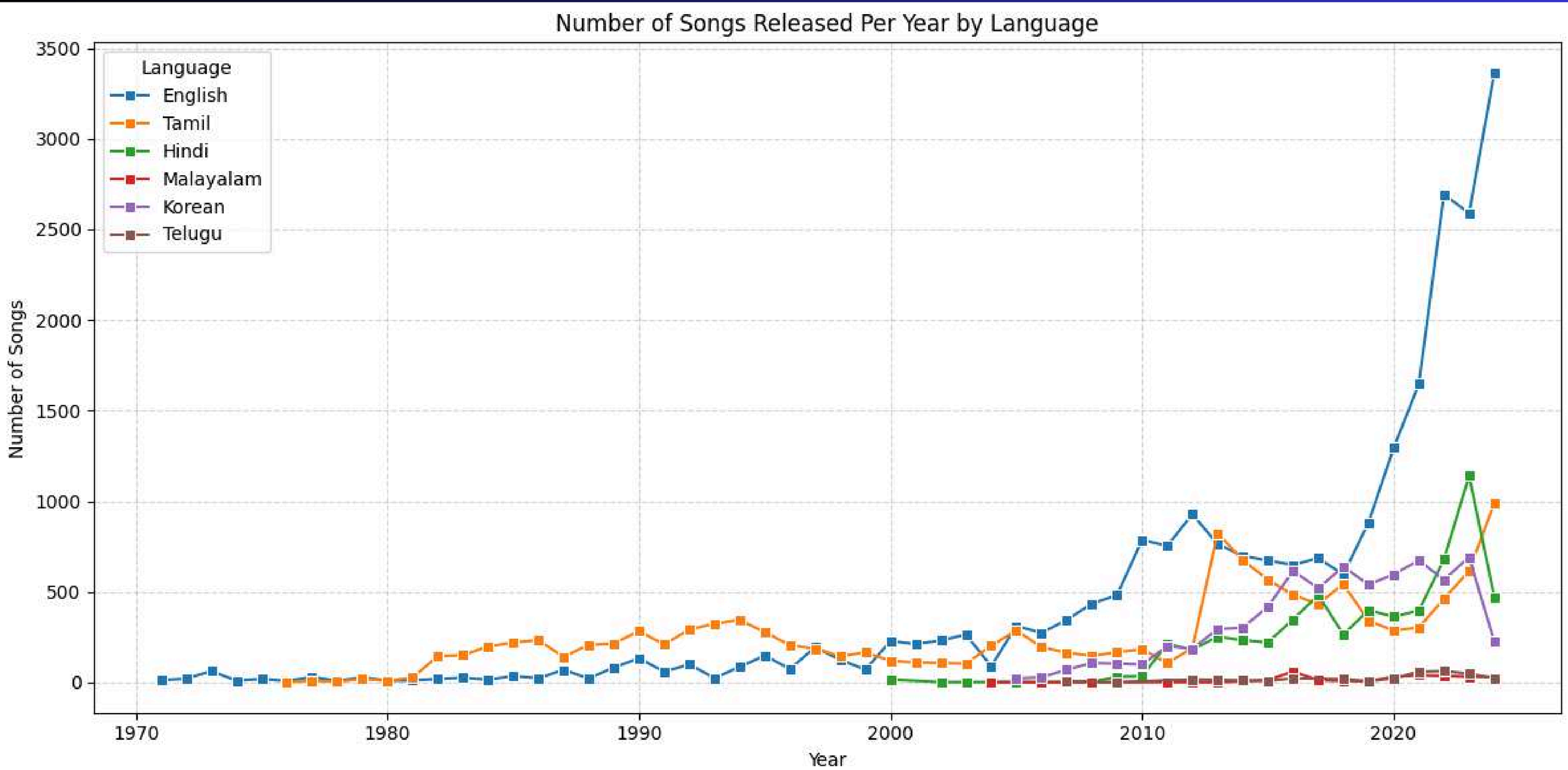
This line chart, showing the Number of Songs Released Per Year by Language, is crucial for understanding the volume and growth dynamics within the dataset.

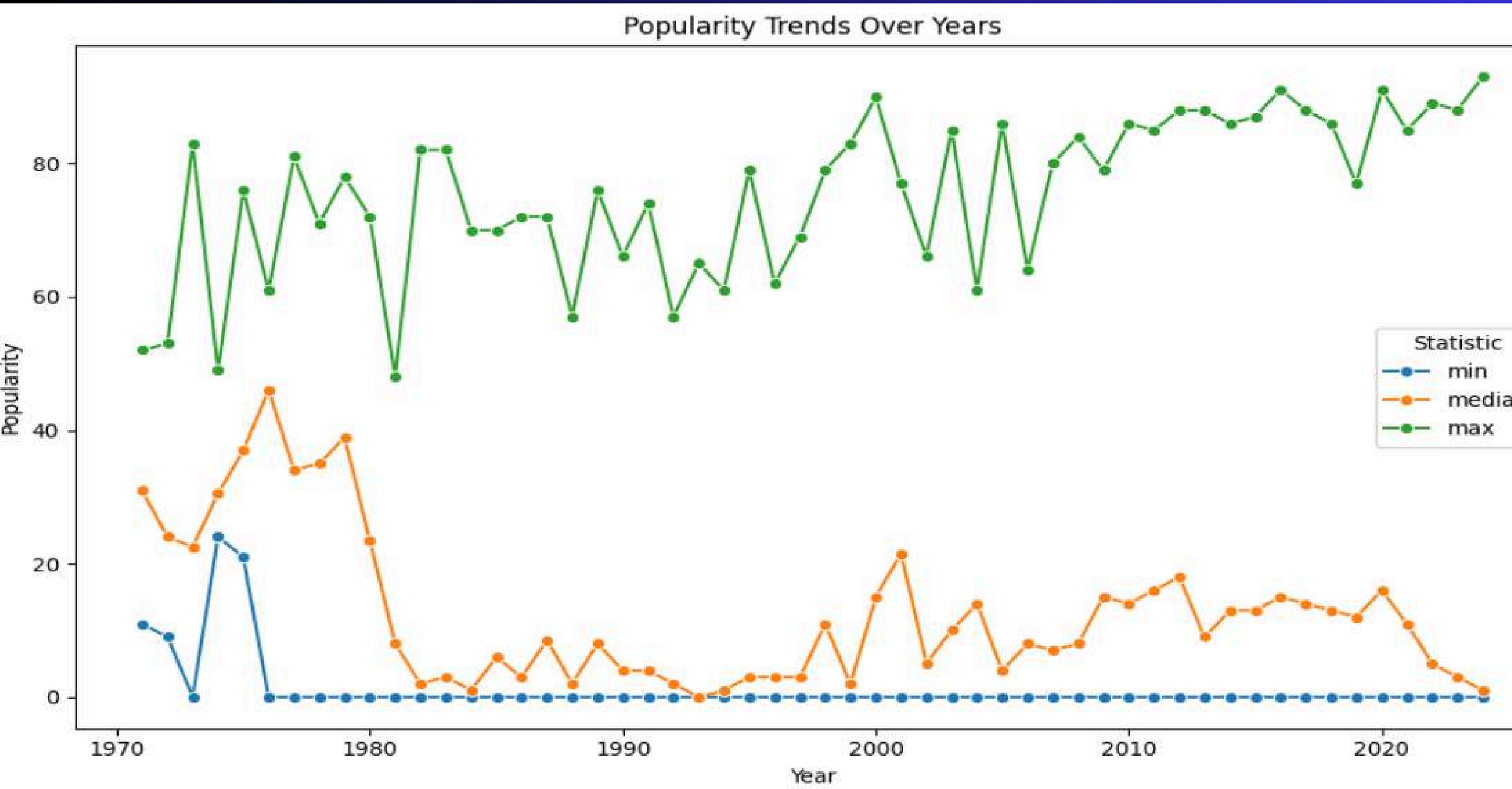
Insights on Song Release Volume Over Time

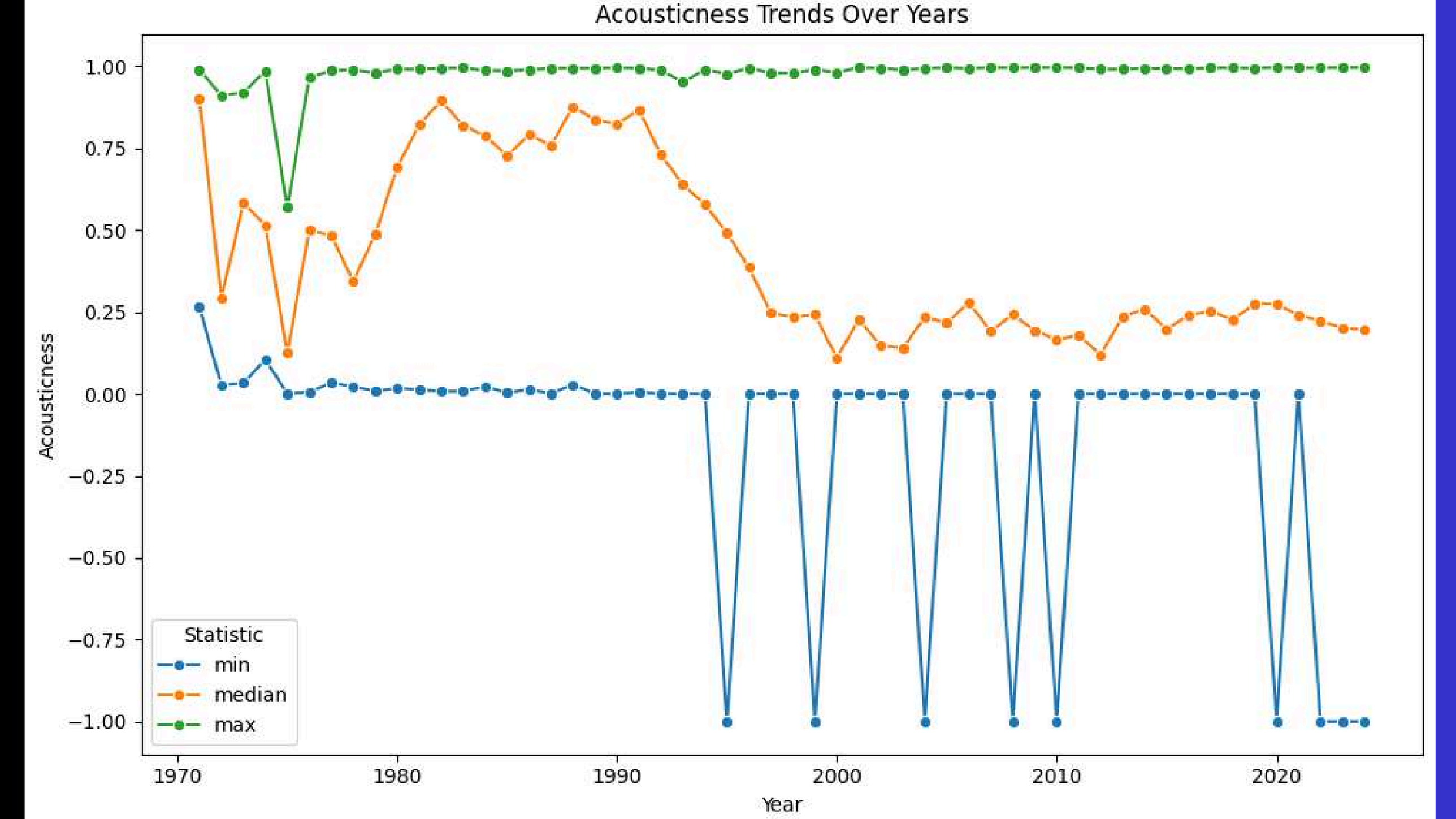
1. The Global Explosion in English Releases Dominant Growth: The English language line (blue) shows the most dramatic and consistent growth, particularly from around 2010 onwards. Recent Spike: English tracks show an unprecedented, massive spike in the final years of the dataset (post-2020), indicating a huge surge in the availability and inclusion of English tracks. This is likely driven by the global expansion of music streaming platforms.

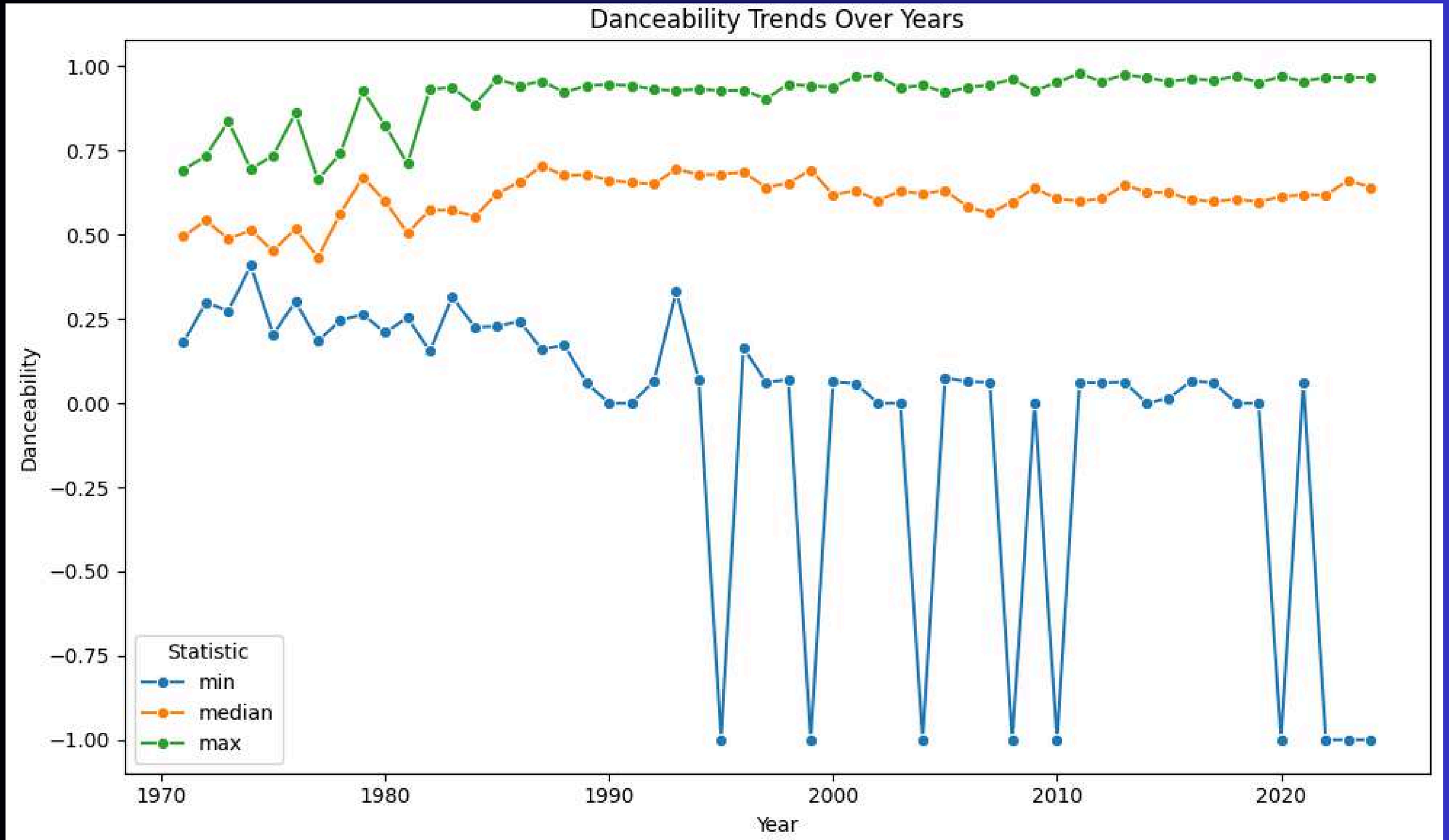
2. High Volume and Consistency (Tamil) Long-Term Presence: Tamil (orange) shows a remarkable early and sustained presence, with a relatively high volume of releases starting as early as the late 1980s and remaining consistent through the 1990s and 2000s, often exceeding English volumes during that period. Modern Competition: While Tamil volumes remain high, they are overshadowed by the recent English spike.

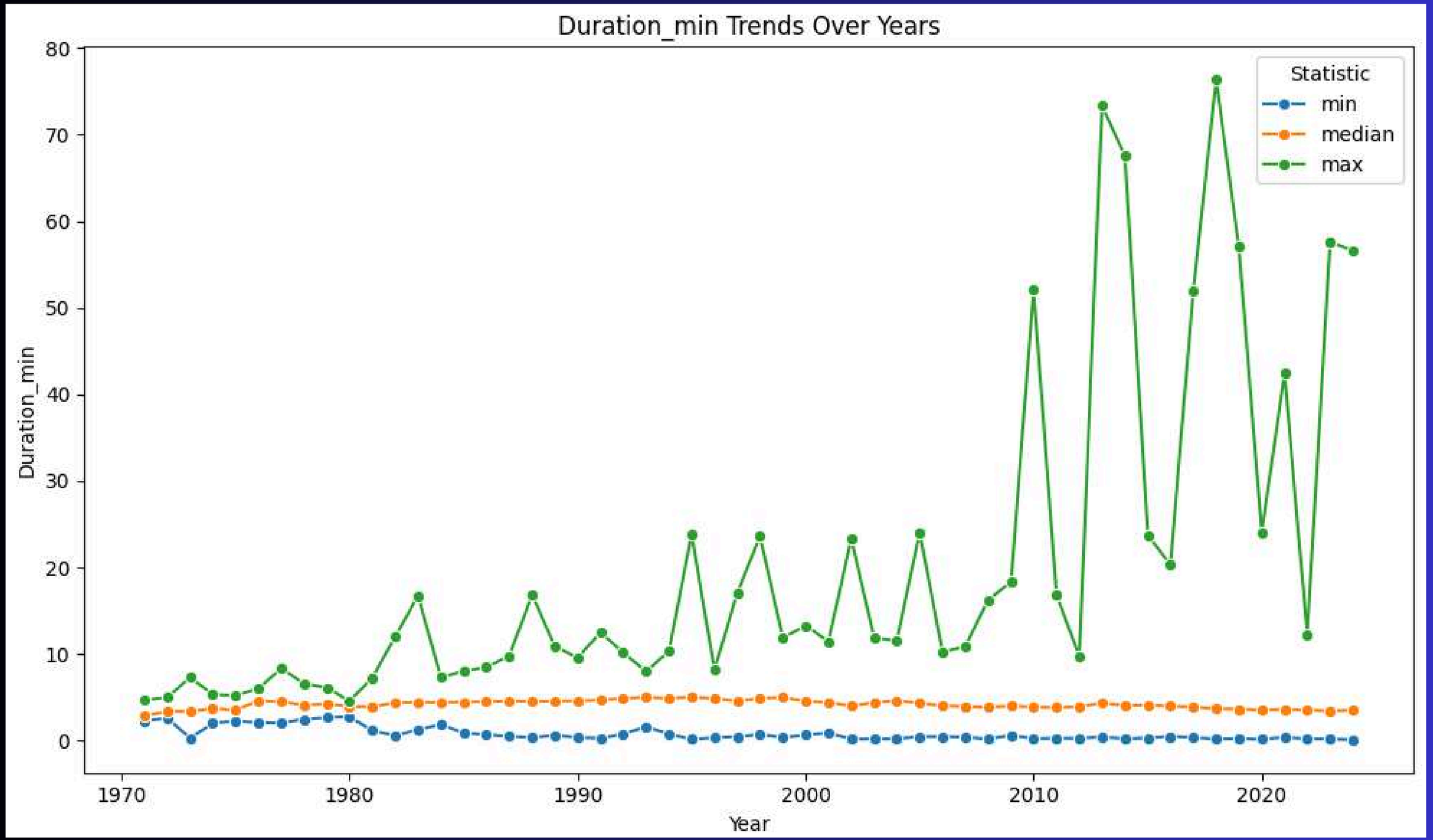
3. Rapid Emergence of Korean and Hindi Korean Boom (Purple): Korean tracks show a sharp and relatively recent acceleration, with volumes surging significantly around 2015 and accelerating rapidly into the final years. This visually represents the rise and global distribution of K-Pop. Hindi Growth (Green): Hindi tracks also show a strong period of growth, peaking around 2015 before slightly stabilizing. This confirms a substantial increase in Indian music content in the dataset over the last decad
4. Low Volume and Stability (Malayalam and Telugu) Minimal Growth: Malayalam (dark red) and Telugu (dark green/gray) consistently show the lowest number of annual releases throughout the entire period. Their lines are nearly flat, suggesting limited growth in volume compared to the other languages.

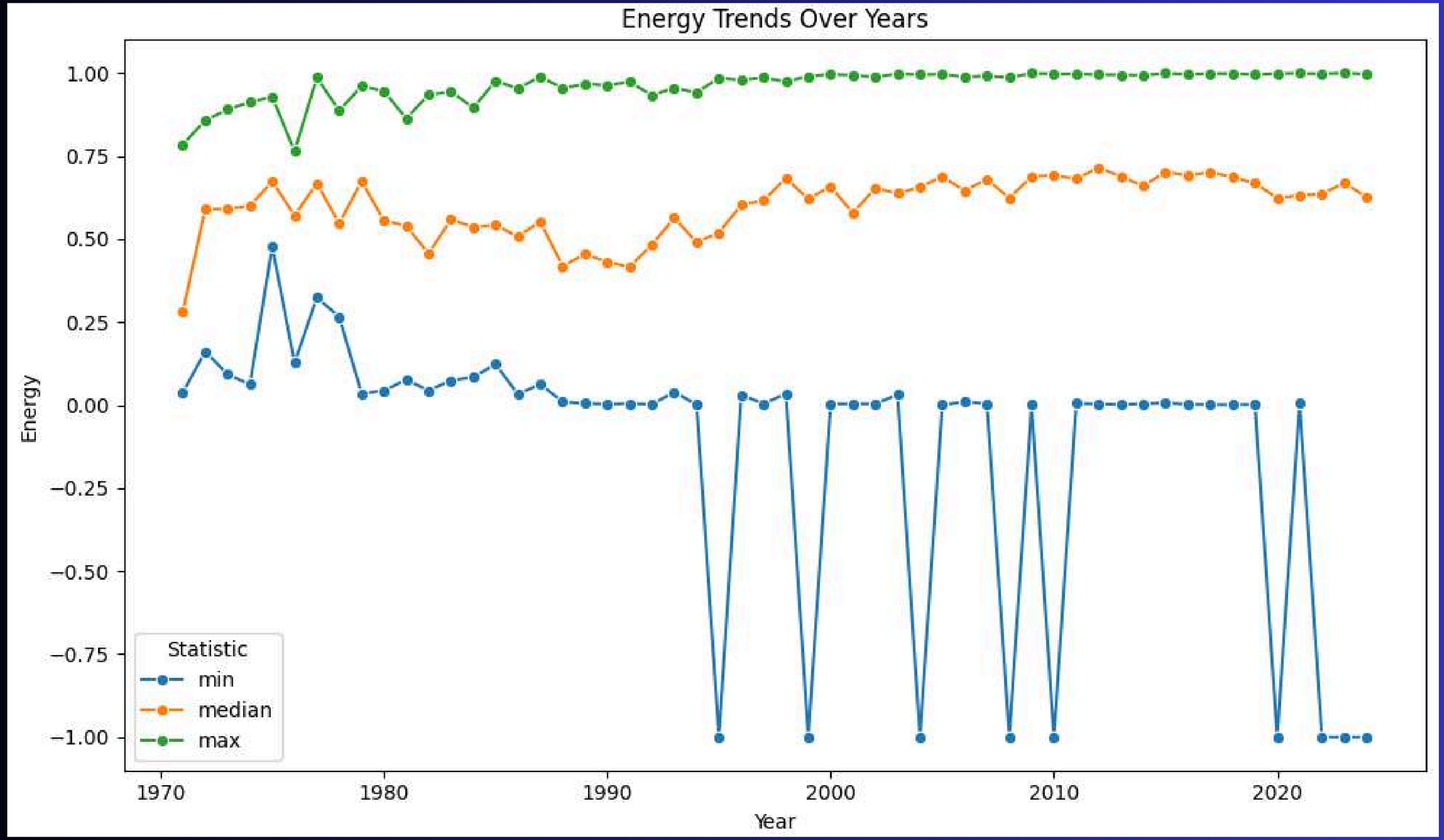

Number of Songs Released Per Year by Language
Language
English
Tamil
Hindi
Malayalam
Korean
Telugu
Number of Songs
Year
0
500
1000
1500
2000
2500
3000
3500
1970
1980
1990
2000
2010
2020

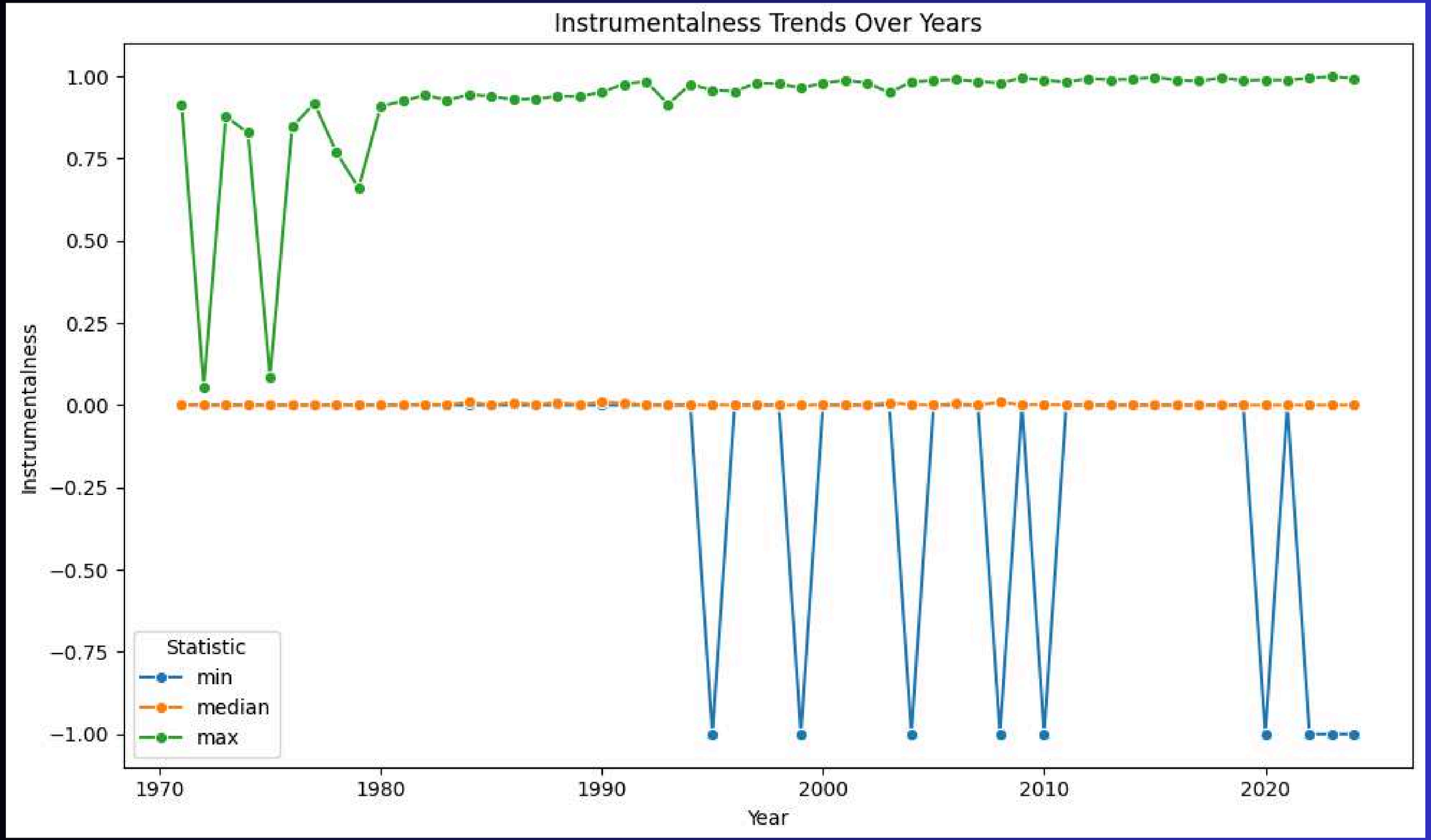

Popularity Trends Over Years
Popularity
Year
0
20
40
60
80
1970
1980
1990
2000
2010
2020
Statistic
min
media
max

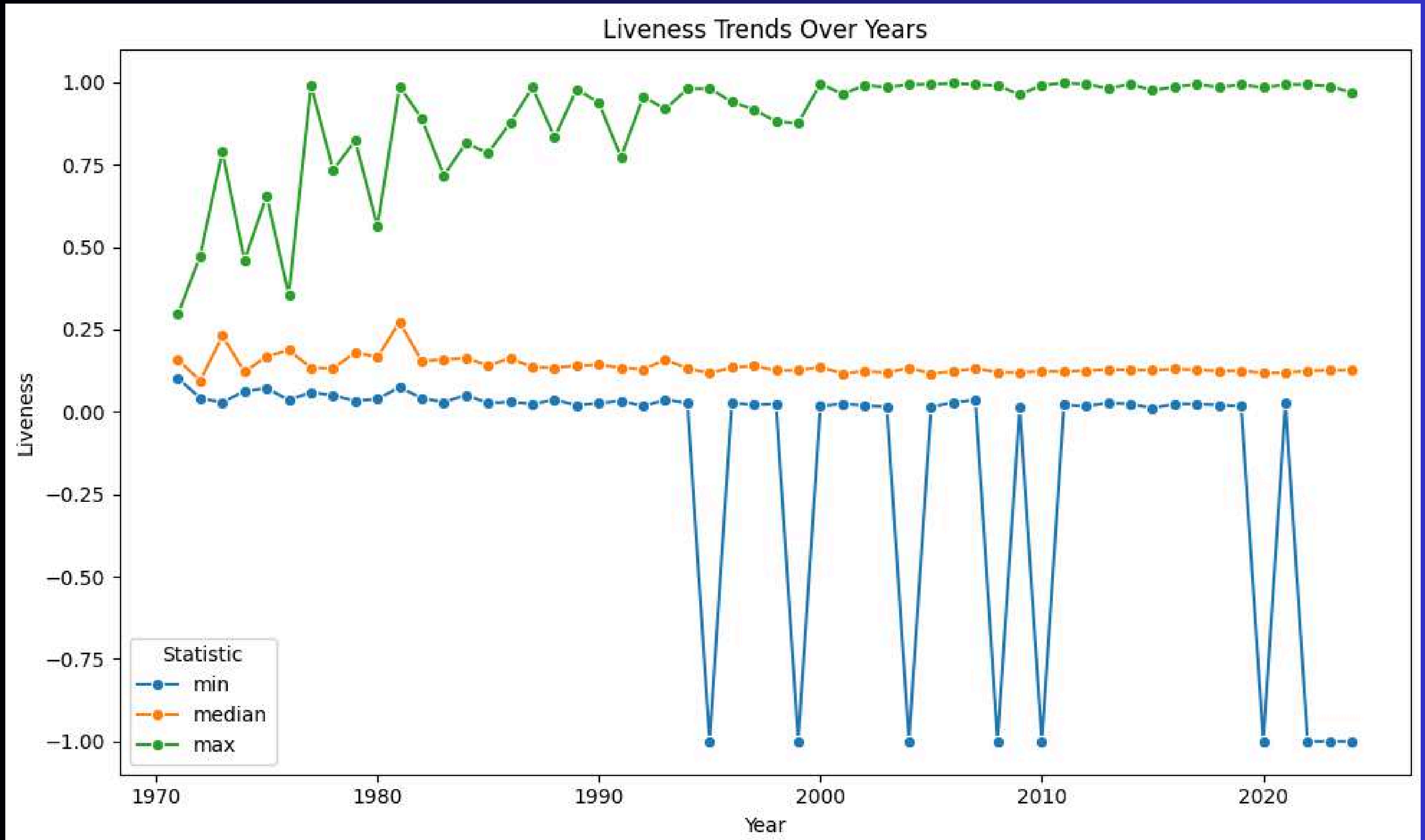

Acousticness Trends Over Years
Acousticness
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
1970
1980
1990
2000
2010
2020
Year
Statistic
min
median
max

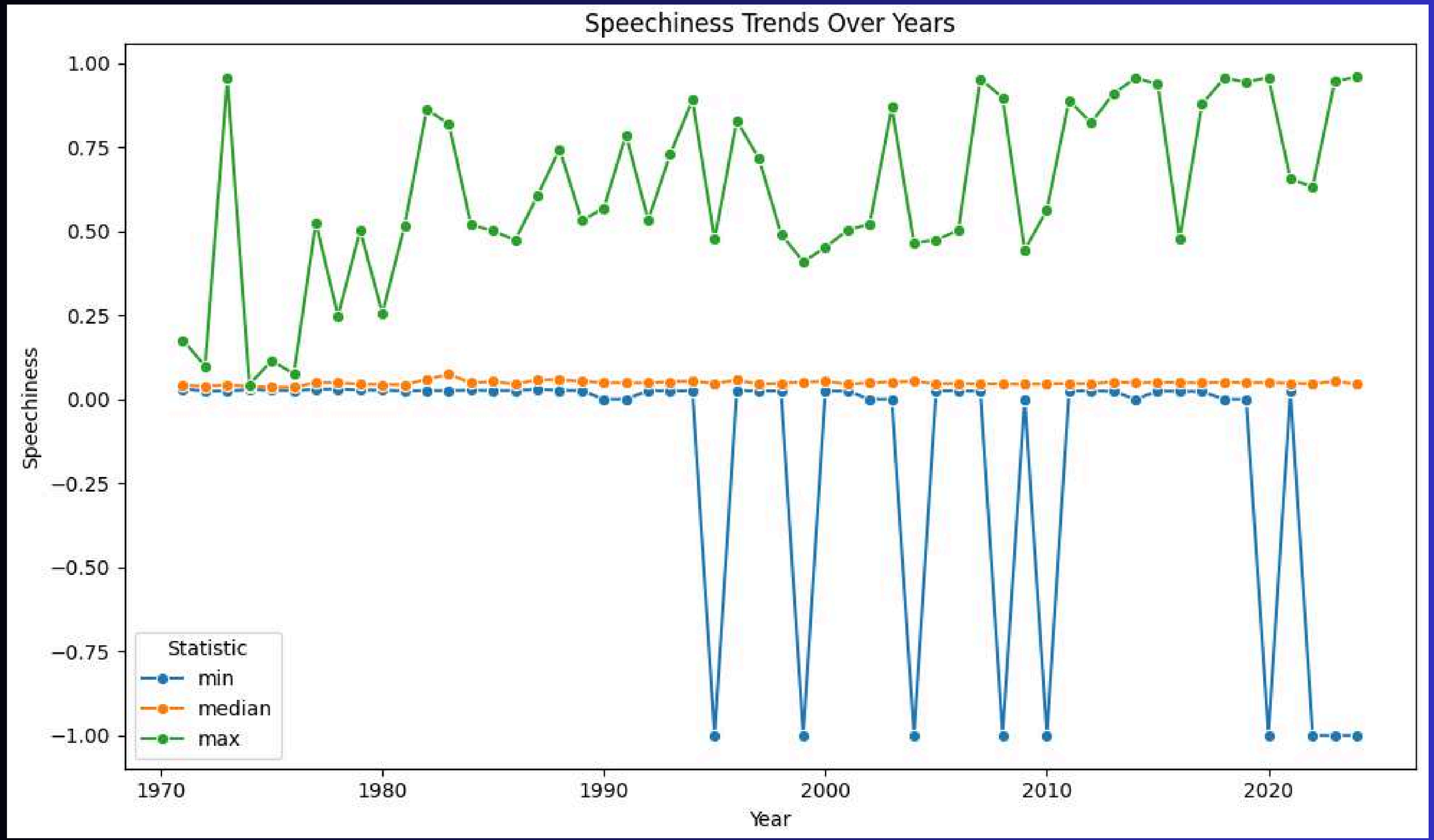

Danceability Trends Over Years
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Danceability
1970
1980
1990
2000
2010
2020
Year
Statistic
min
median
max

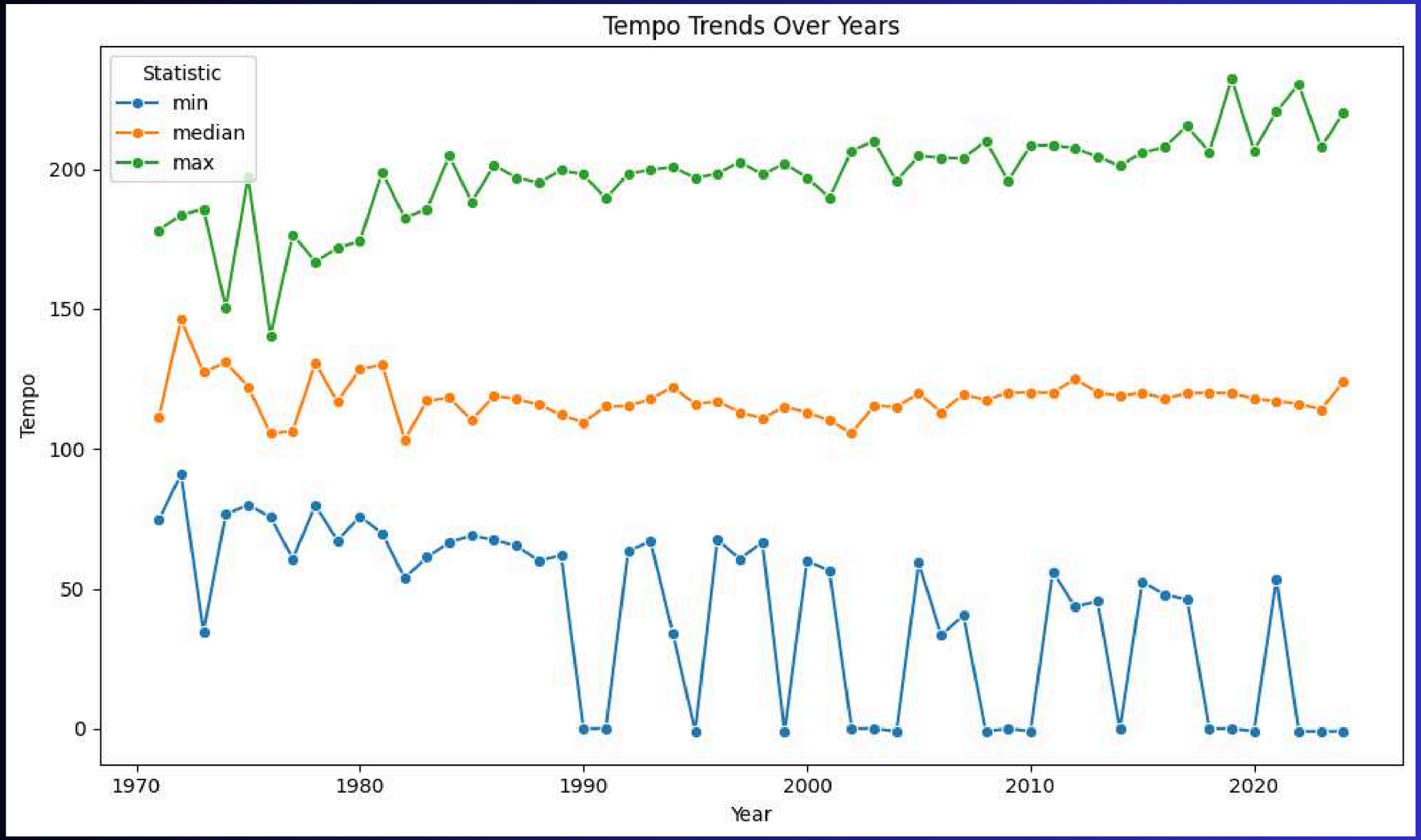

Duration_min Trends Over Years
Statistic
min
median
max
Duration_min
Year
0
10
20
30
40
50
60
70
80
1970
1980
1990
2000
2010
2020

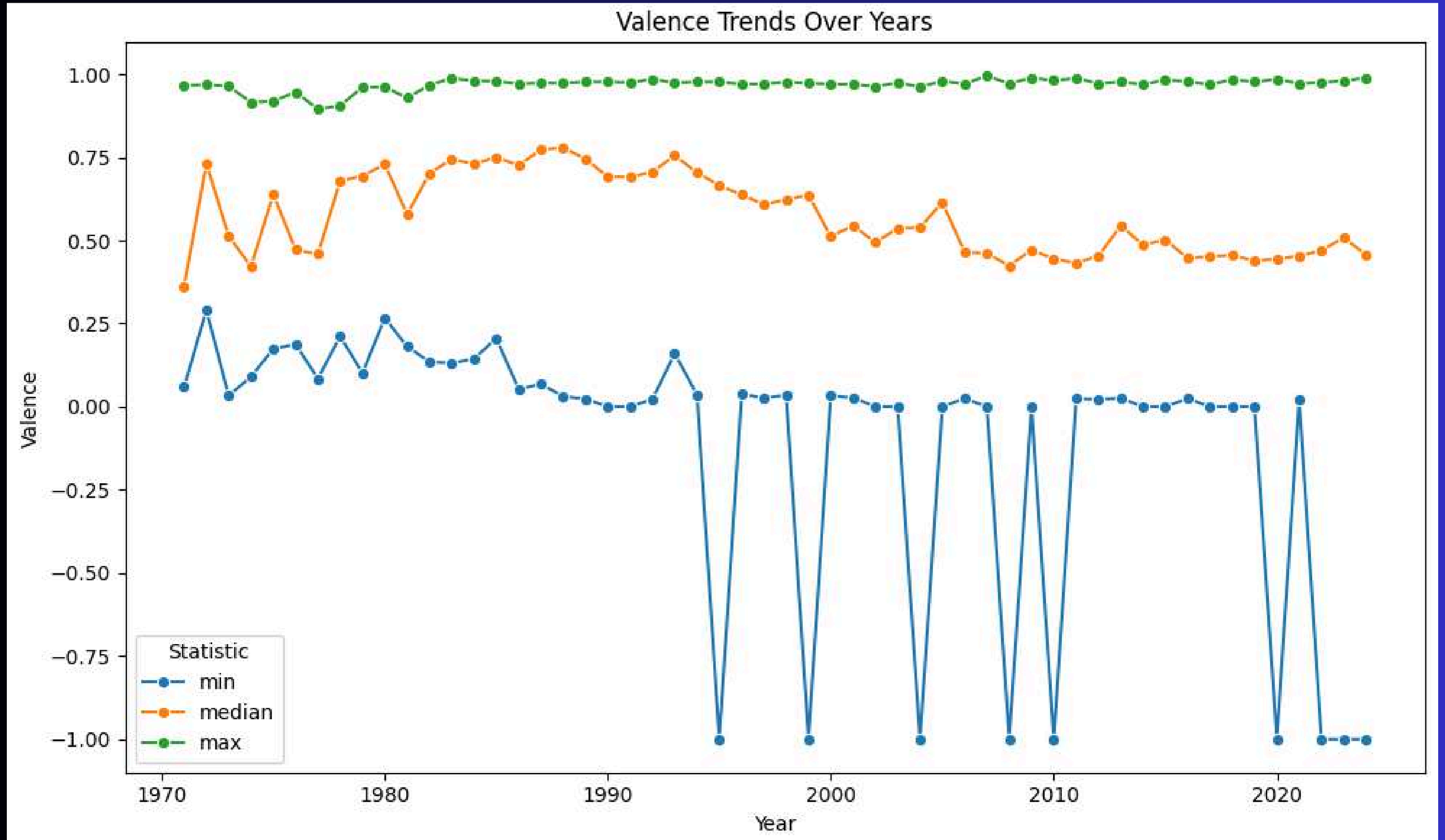

Energy Trends Over Years
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Energy
1970
1980
1990
2000
2010
2020
Year
Statistic
min
median
max

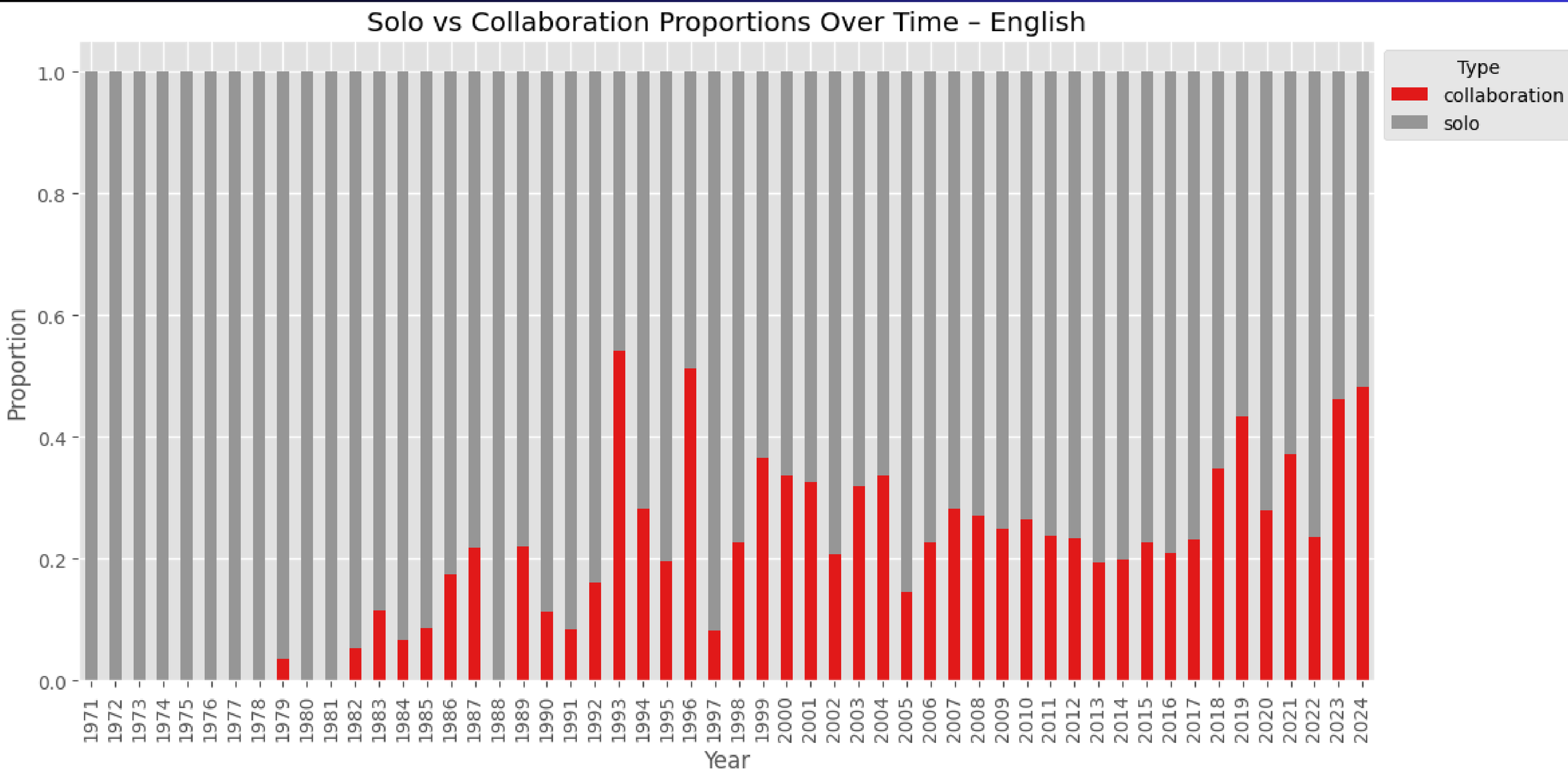

Instrumentalness Trends Over Years
Instrumentalness
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
1970
1980
1990
2000
2010
2020
Year
Statistic
min
median
max

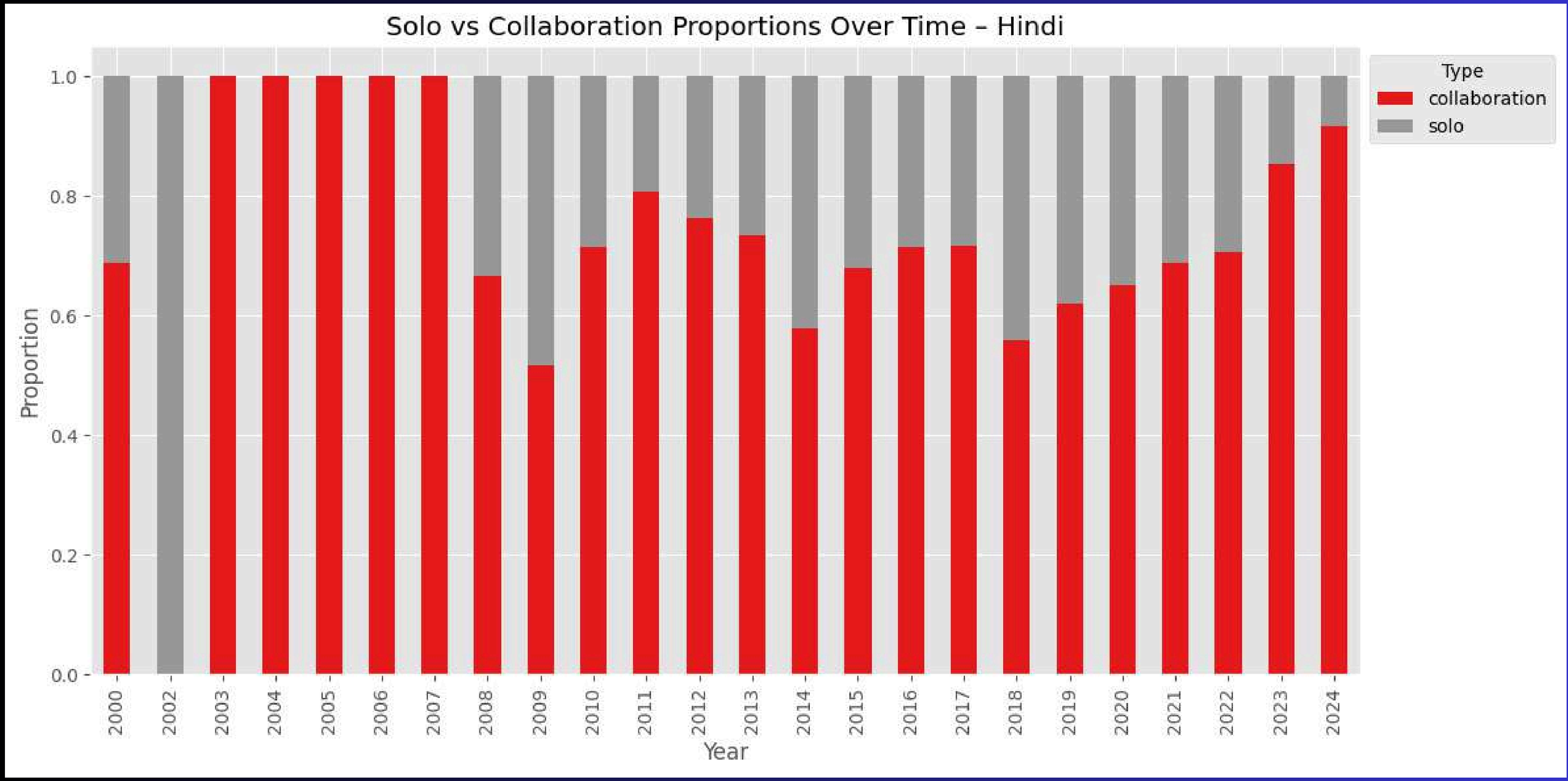

Liveness Trends Over Years
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Liveness
1970
1980
1990
2000
2010
2020
Year
Statistic
min
median
max

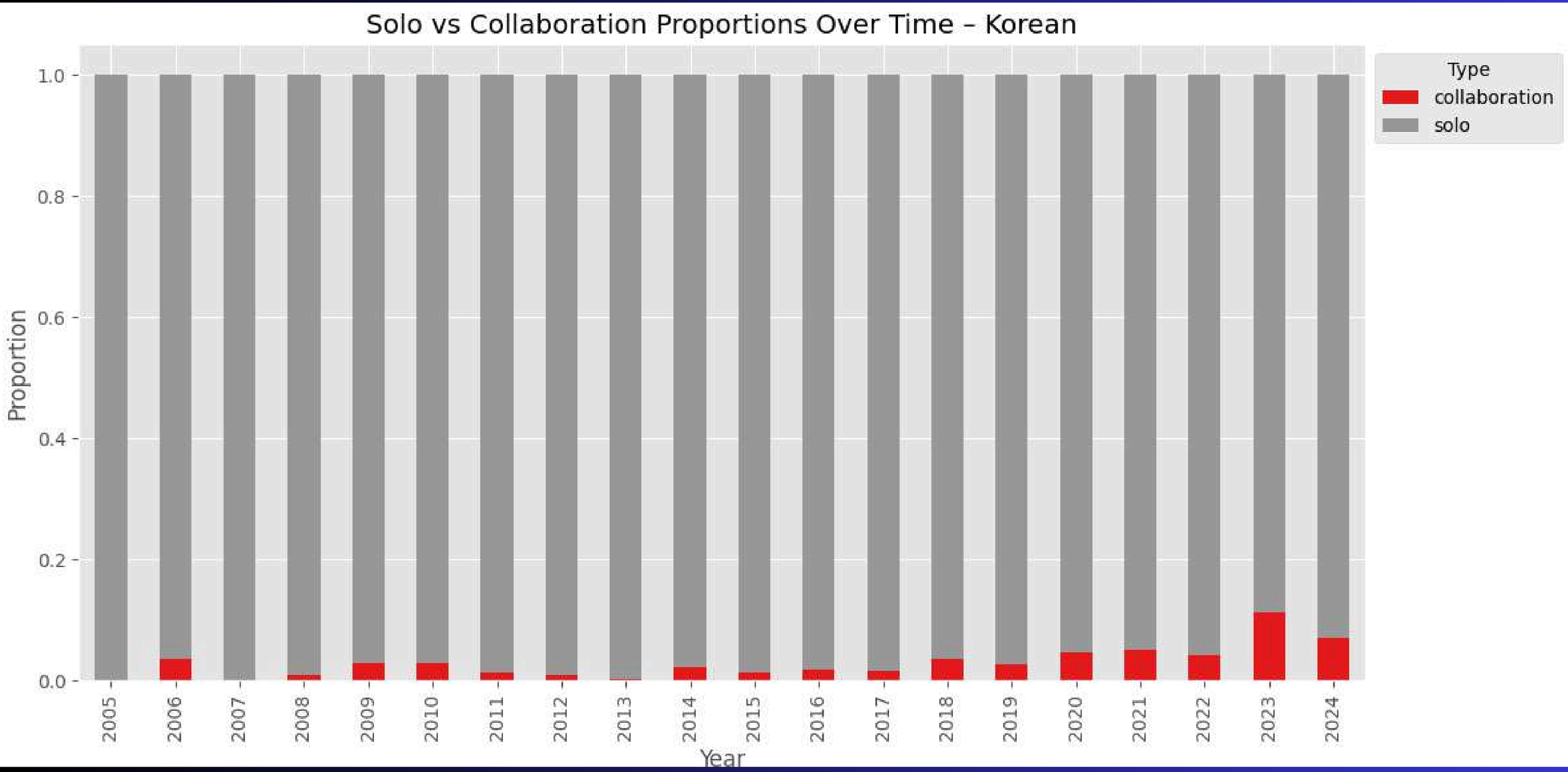

Speechiness Trends Over Years
Speechiness
Year
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
1970
1980
1990
2000
2010
2020
Statistic
min
median
max

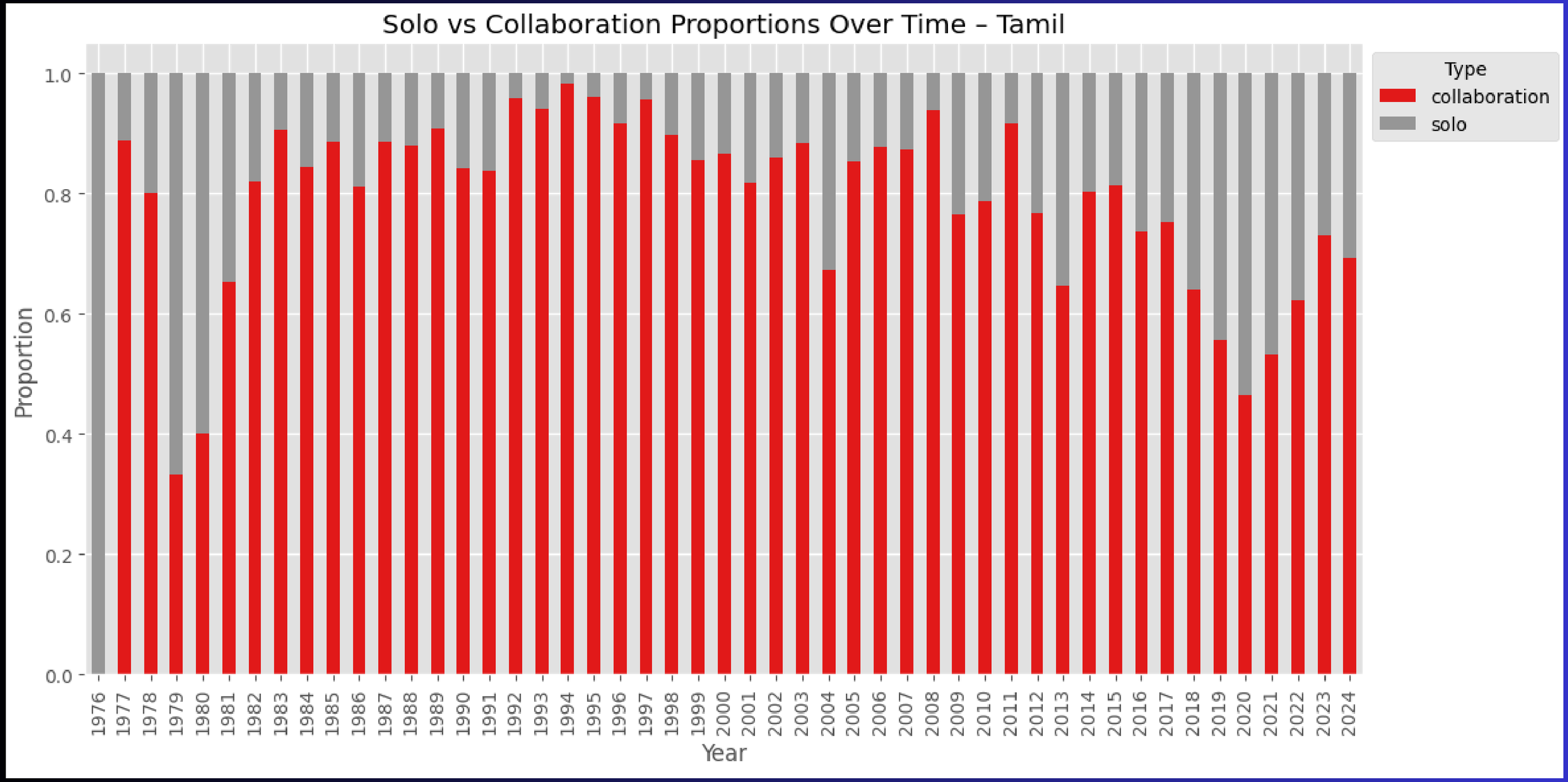

Tempo Trends Over Years
Statistic
min
median
max
Tempo
200
150
100
50
0
1970
1980
1990
2000
2010
2020
Year

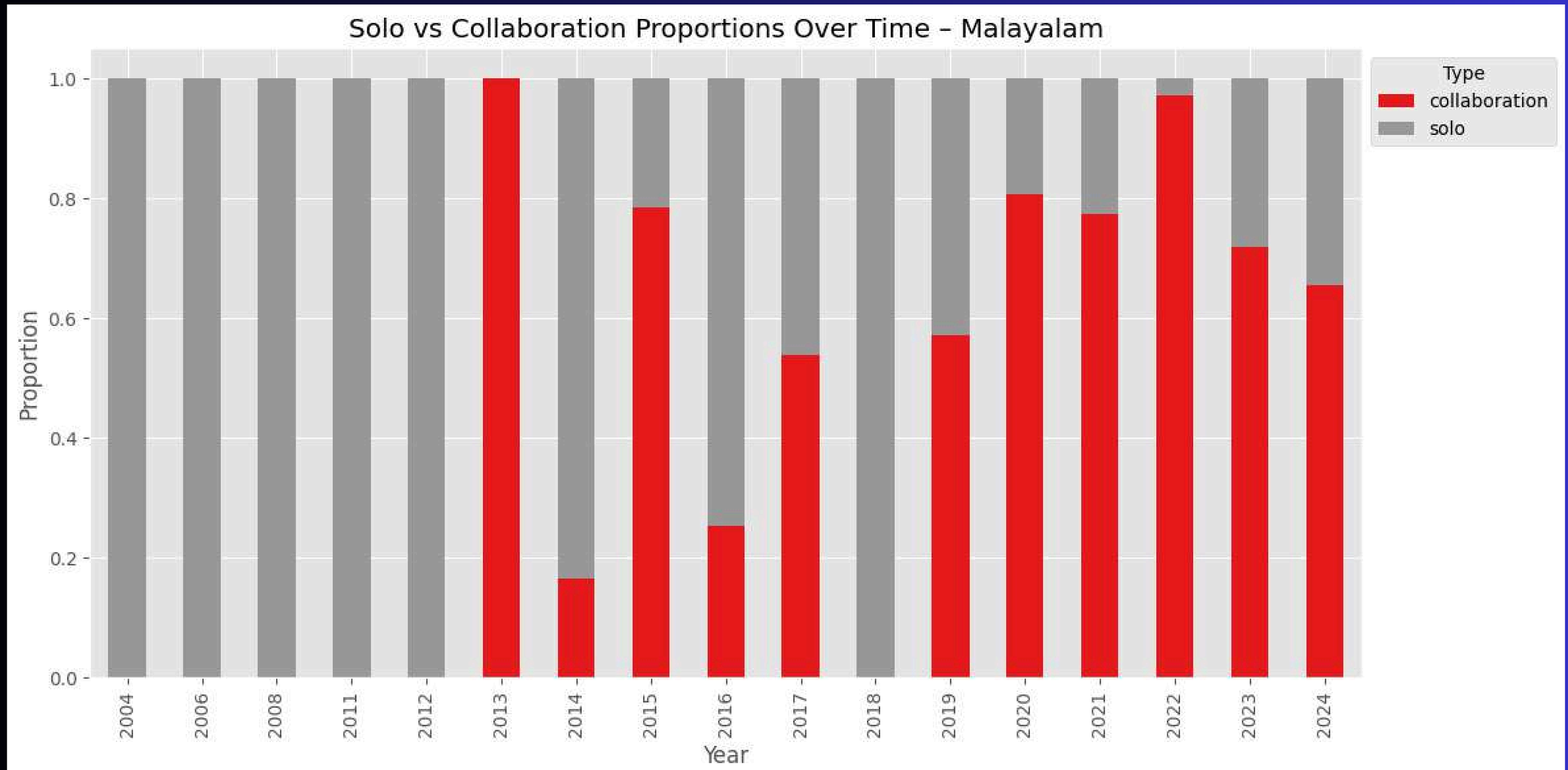

Valence Trends Over Years
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Valence
1970
1980
1990
2000
2010
2020
Year
Statistic
min
median
max

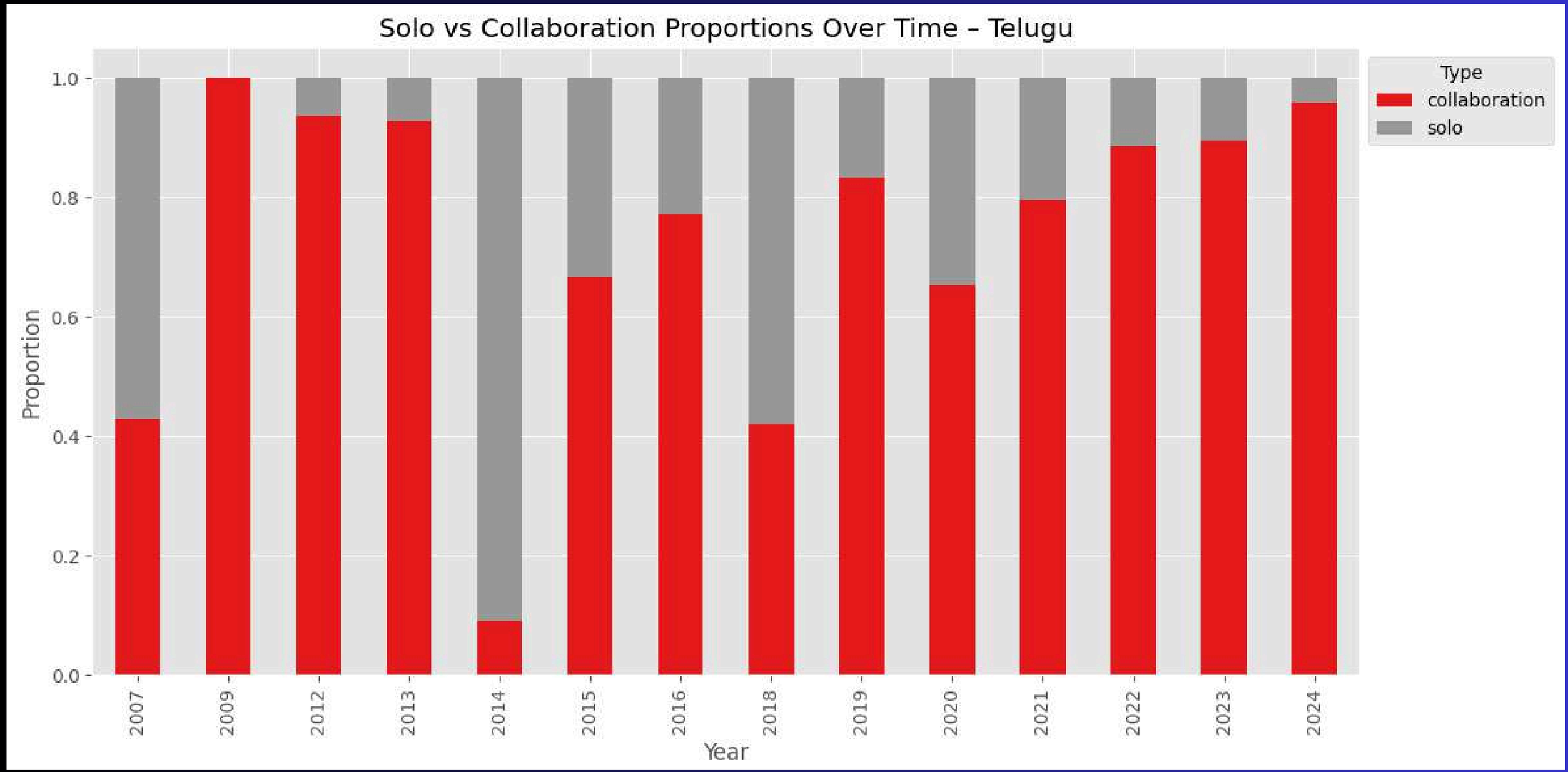

Solo vs Collaboration Proportions Over Time – English
Type
collaboration
solo
Proportion
1.0
0.8
0.6
0.4
0.2
0.0
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
Year

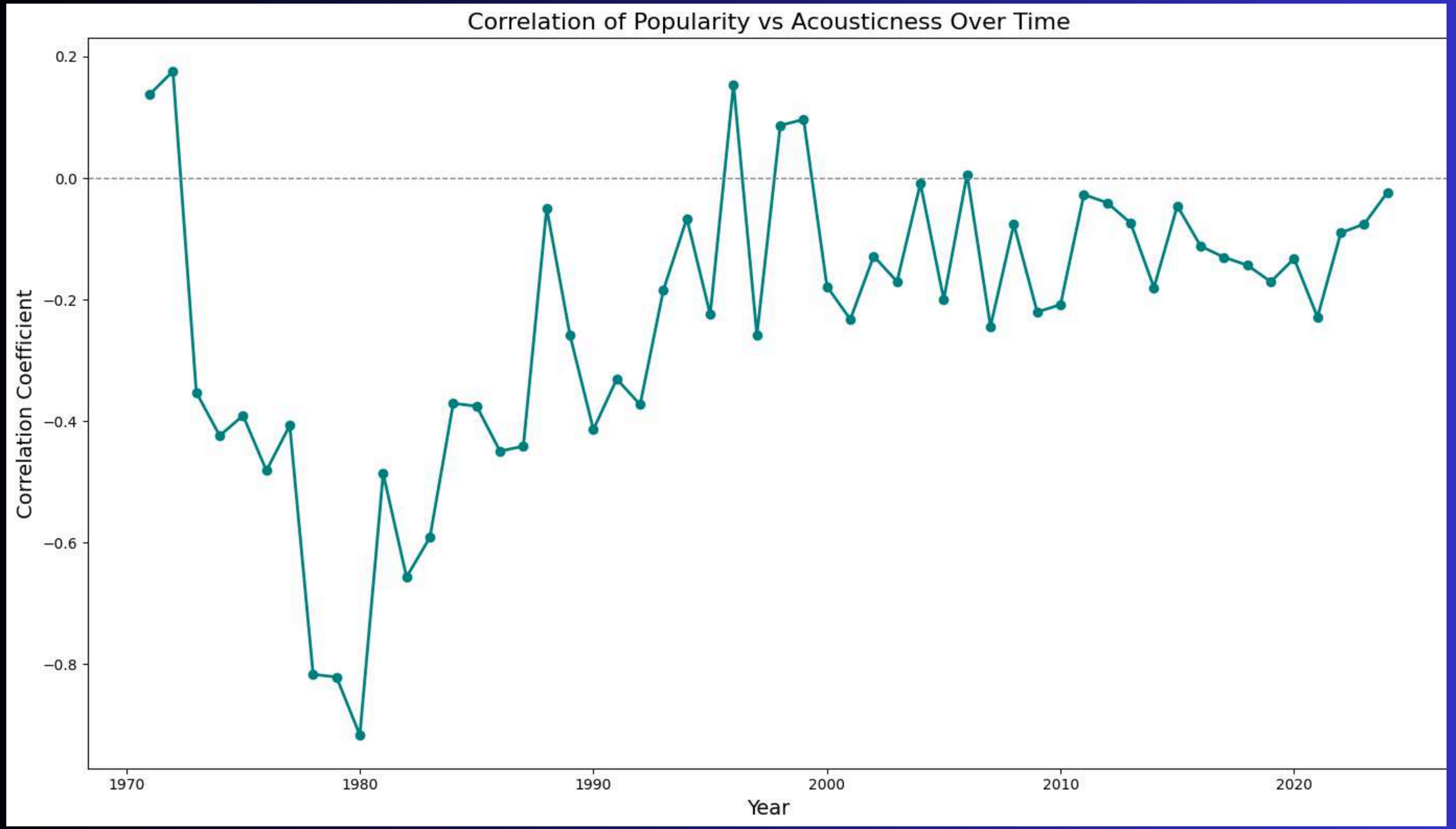

Solo vs Collaboration Proportions Over Time – Hindi
Type
collaboration
solo
Proportion
1.0
0.8
0.6
0.4
0.2
0.0
2000
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
Year

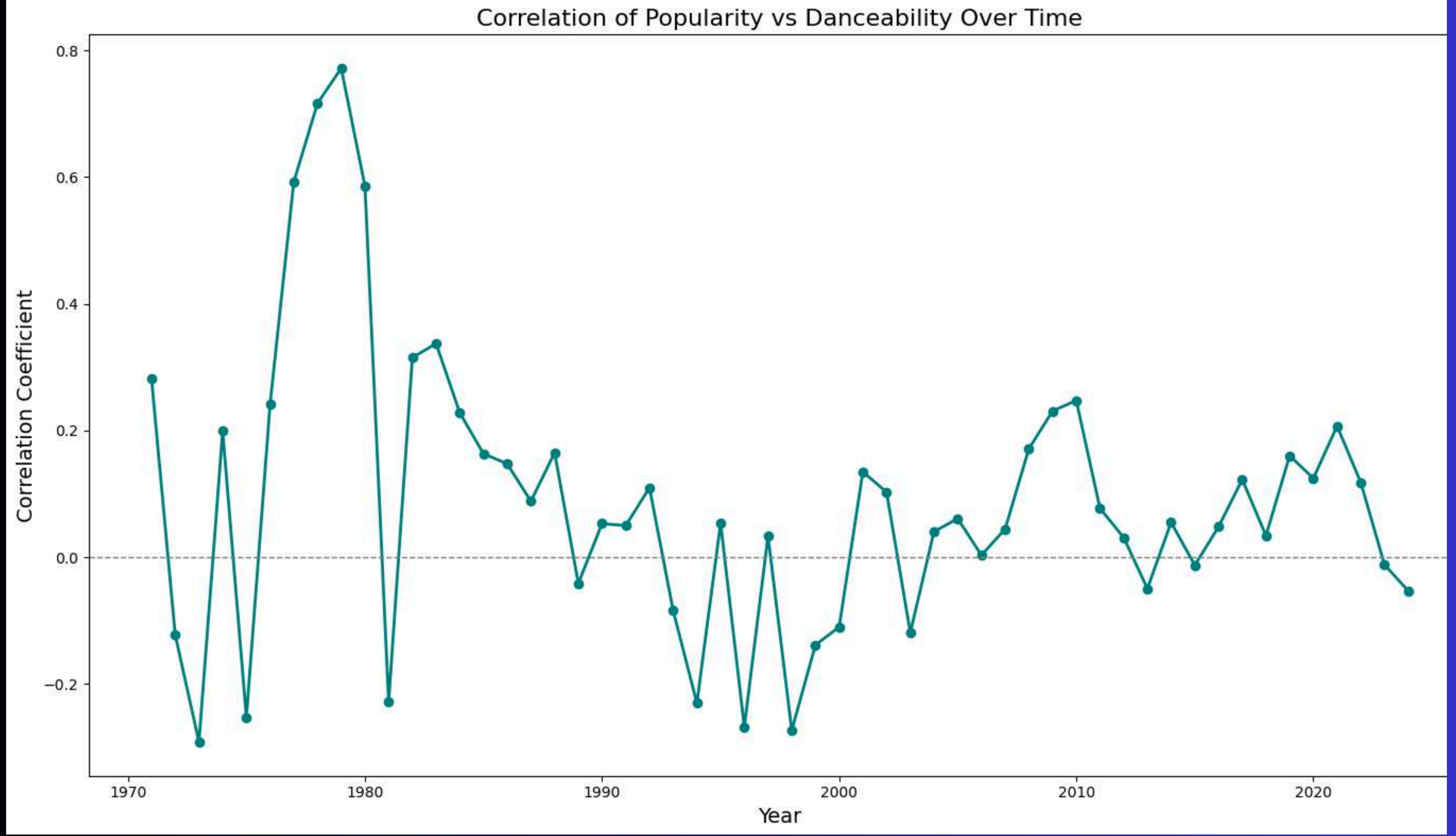

Solo vs Collaboration Proportions Over Time – Korean
Type
collaboration
solo
Proportion
1.0
0.8
0.6
0.4
0.2
0.0
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
Year

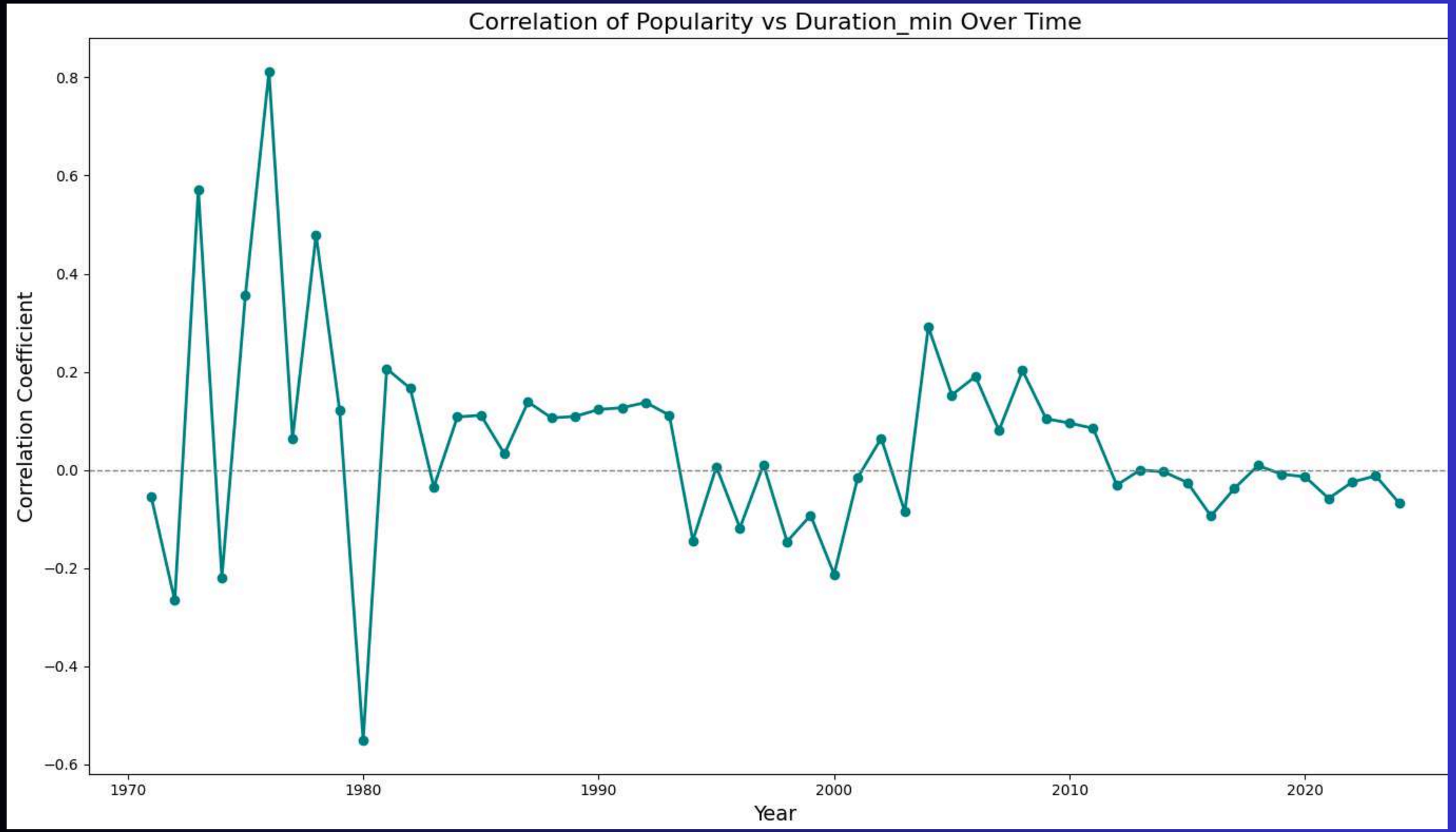

Solo vs Collaboration Proportions Over Time – Tamil
Type
collaboration
solo
Proportion
1.0
0.8
0.6
0.4
0.2
0.0
Year
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024

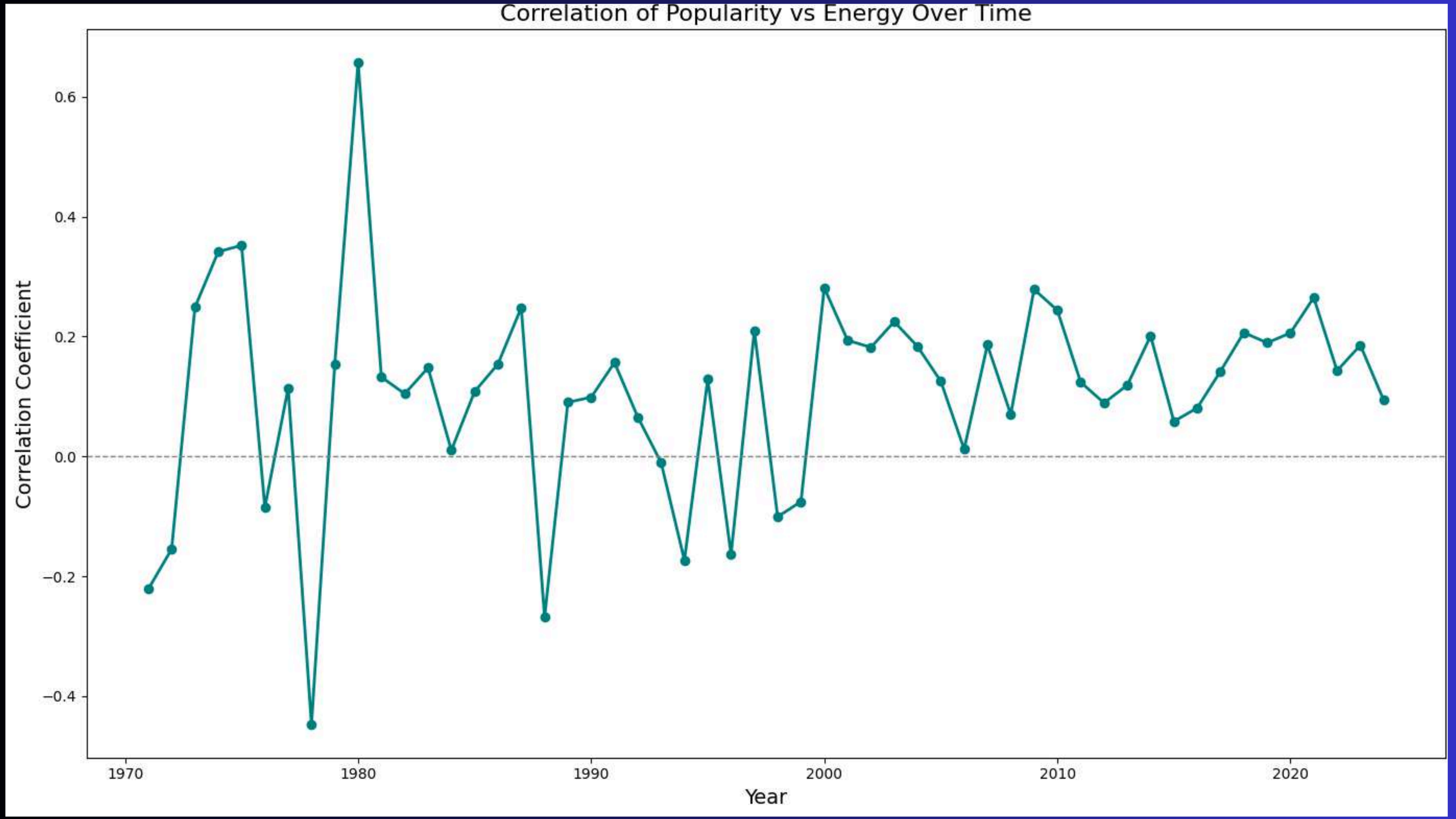

Solo vs Collaboration Proportions Over Time – Malayalam
Type
collaboration
solo
Proportion
1.0
0.8
0.6
0.4
0.2
0.0
2004
2006
2008
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
Year

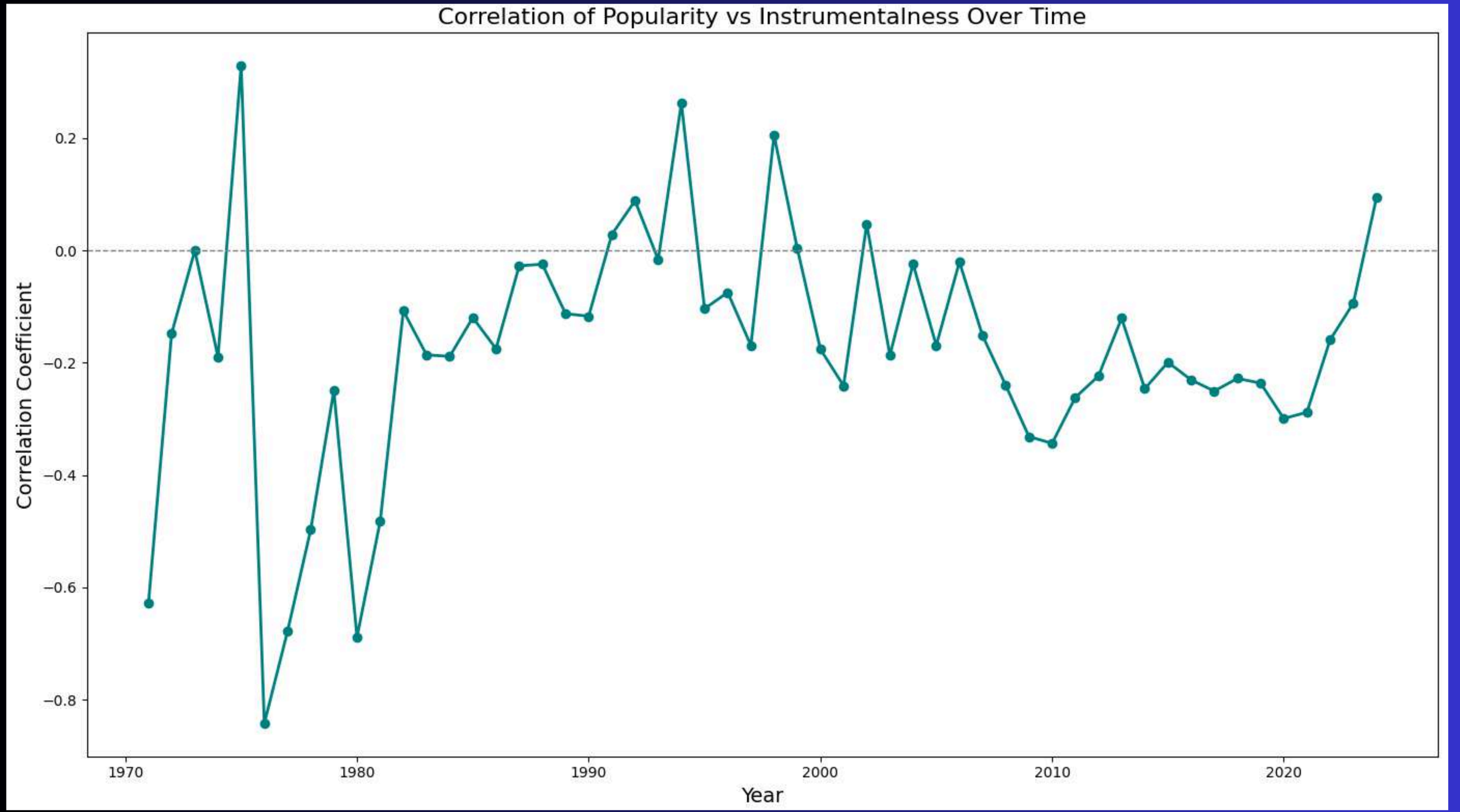

Solo vs Collaboration Proportions Over Time – Telugu
Type
collaboration
solo
Proportion
1.0
0.8
0.6
0.4
0.2
0.0
2007
2009
2012
2013
2014
2015
2016
2018
2019
2020
2021
2022
2023
2024
Year

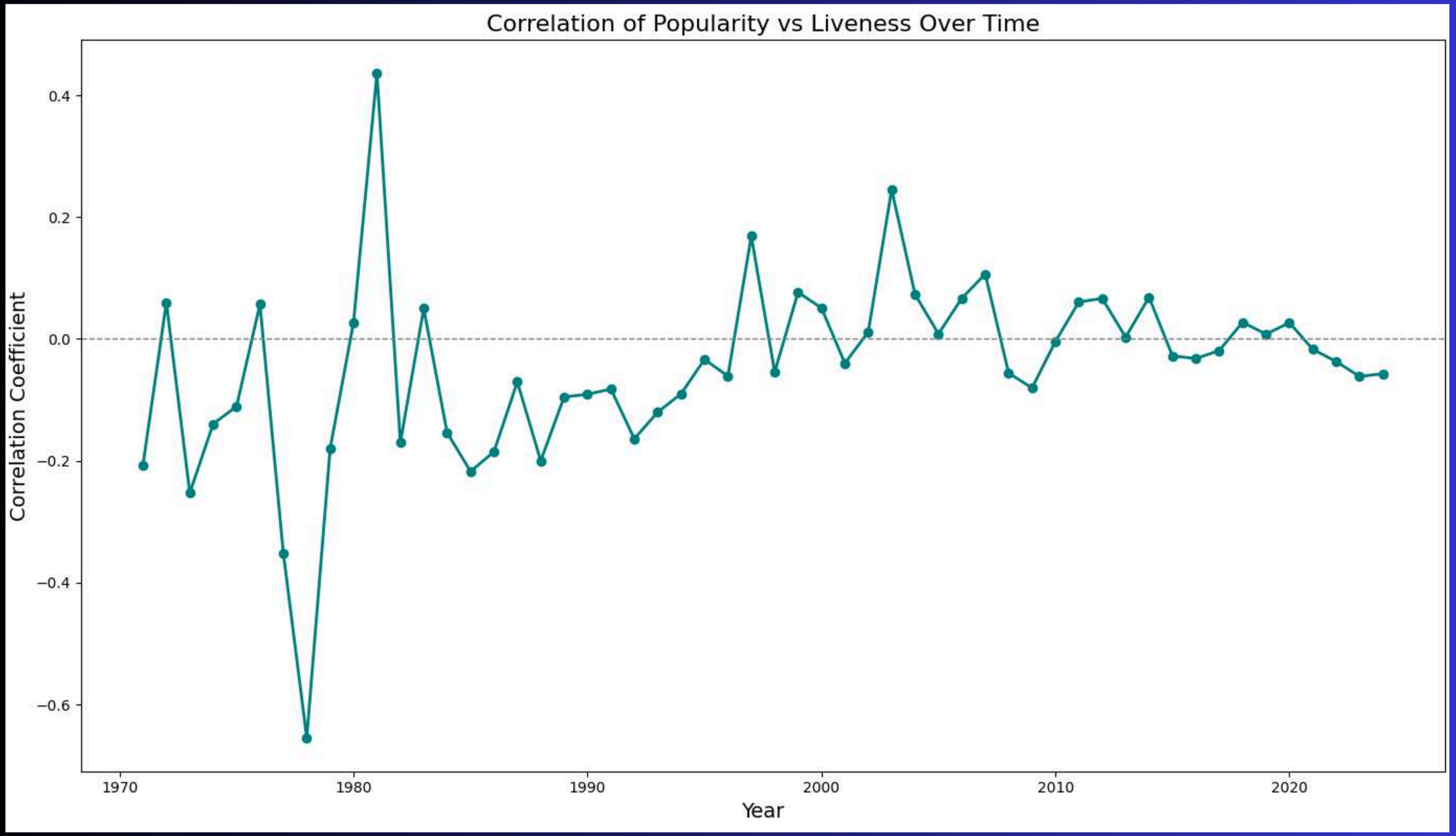

Correlation of Popularity vs Acousticness Over Time
Correlation Coefficient
0.2
0.0
−0.2
−0.4
−0.6
−0.8
1970
1980
1990
2000
2010
2020
Year

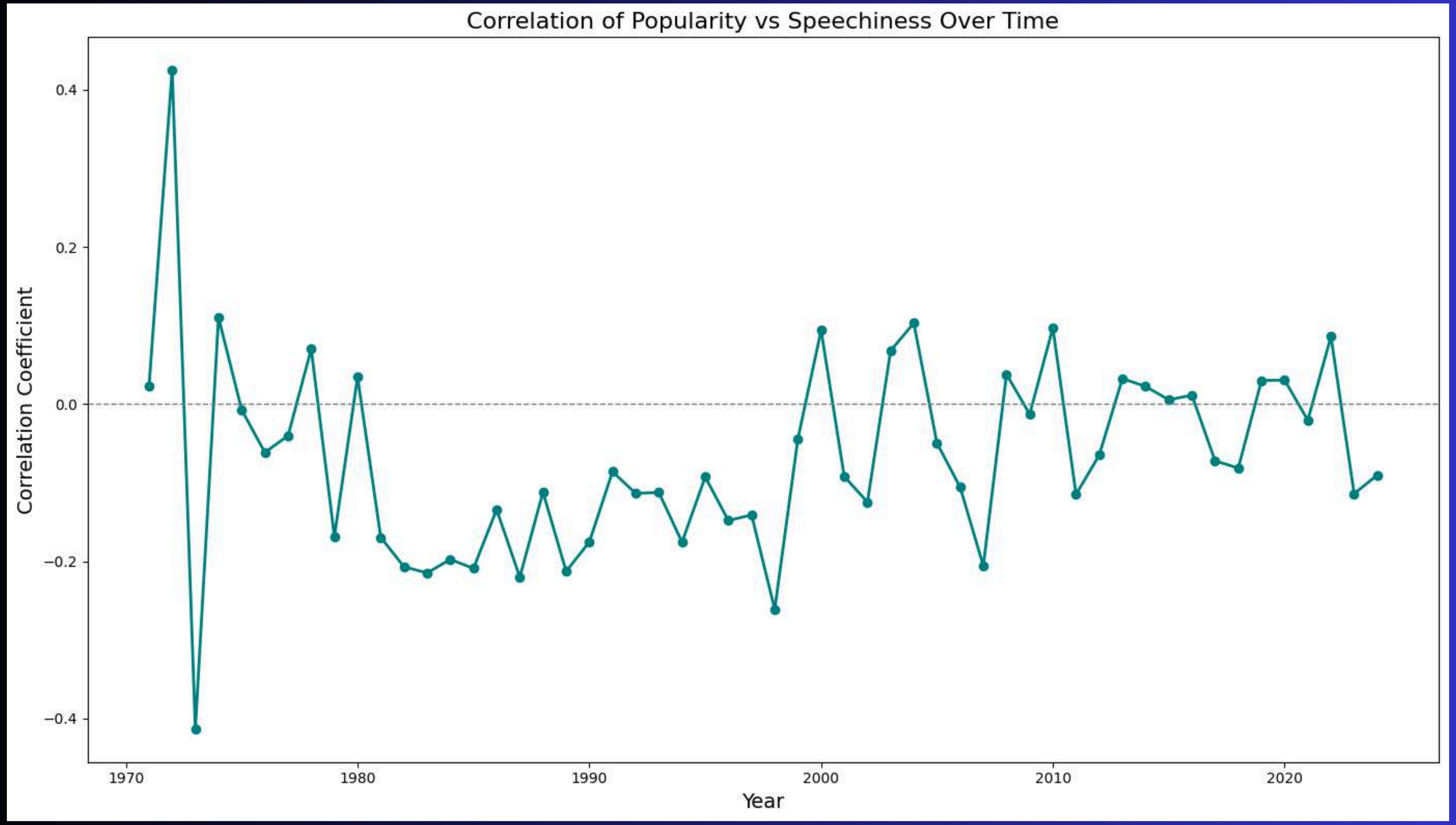

Correlation of Popularity vs Danceability Over Time
Correlation Coefficient
0.8
0.6
0.4
0.2
0.0
−0.2
1970
1980
1990
2000
2010
2020
Year

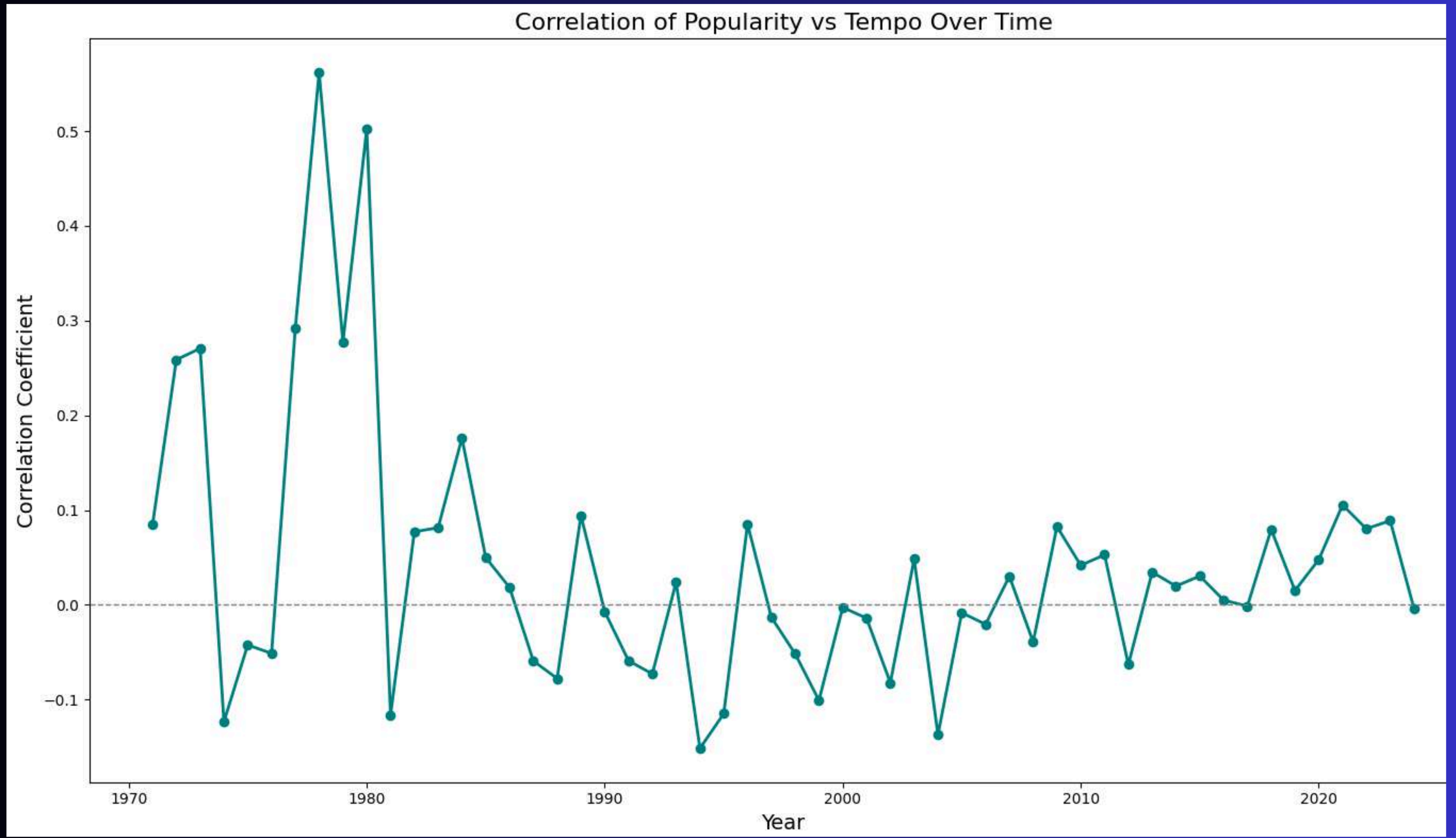

Correlation of Popularity vs Duration_min Over Time
Correlation Coefficient
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
1970
1980
1990
2000
2010
2020
Year

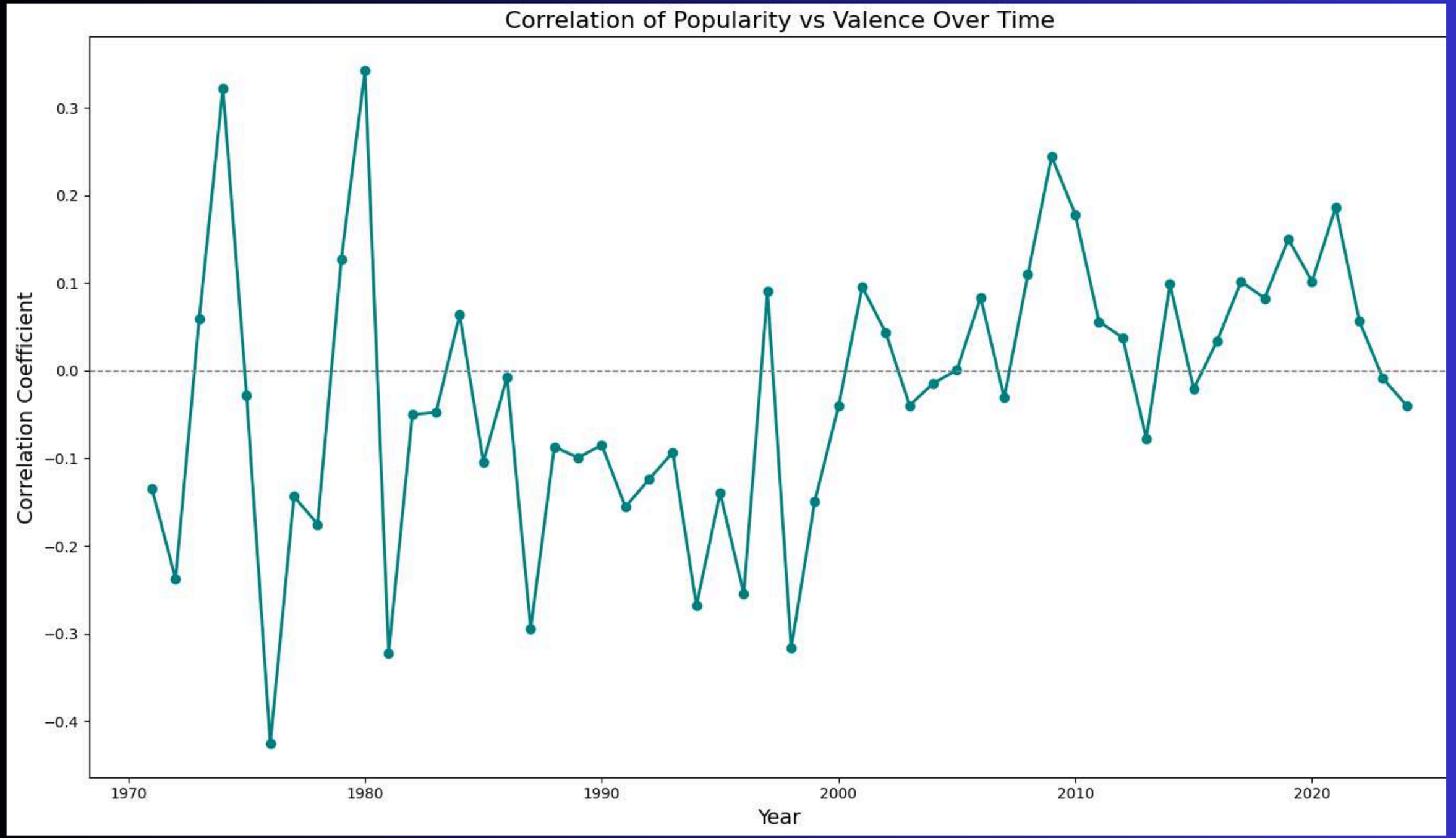

Correlation of Popularity vs Energy Over Time
Correlation Coefficient
0.6
0.4
0.2
0.0
−0.2
−0.4
1970
1980
1990
2000
2010
2020
Year

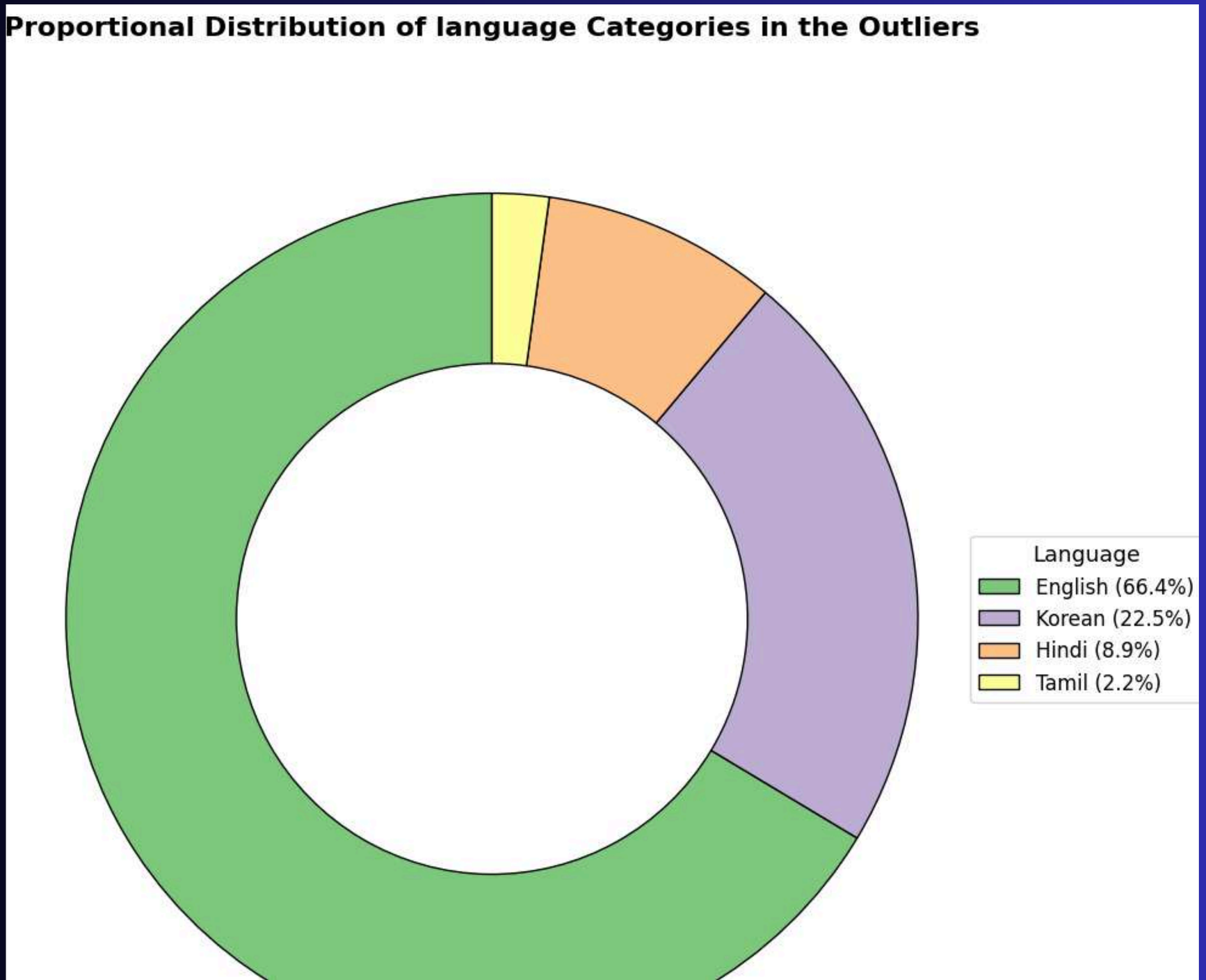

Correlation of Popularity vs Instrumentalness Over Time
Correlation Coefficient
0.2
0.0
−0.2
−0.4
−0.6
−0.8
1970
1980
1990
2000
2010
2020
Year

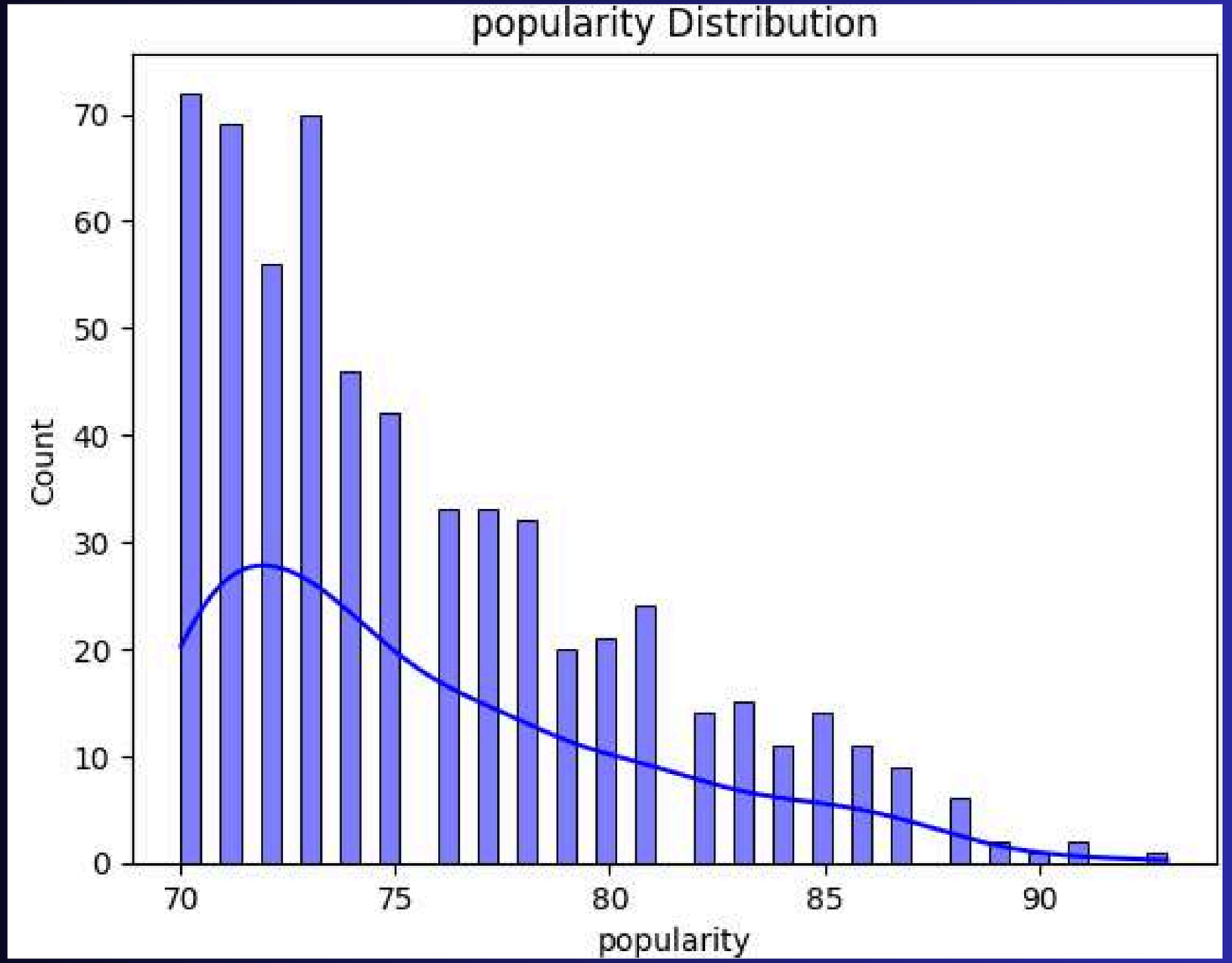

Correlation of Popularity vs Liveness Over Time
Correlation Coefficient
0.4
0.2
0.0
−0.2
−0.4
−0.6
1970
1980
1990
2000
2010
2020
Year

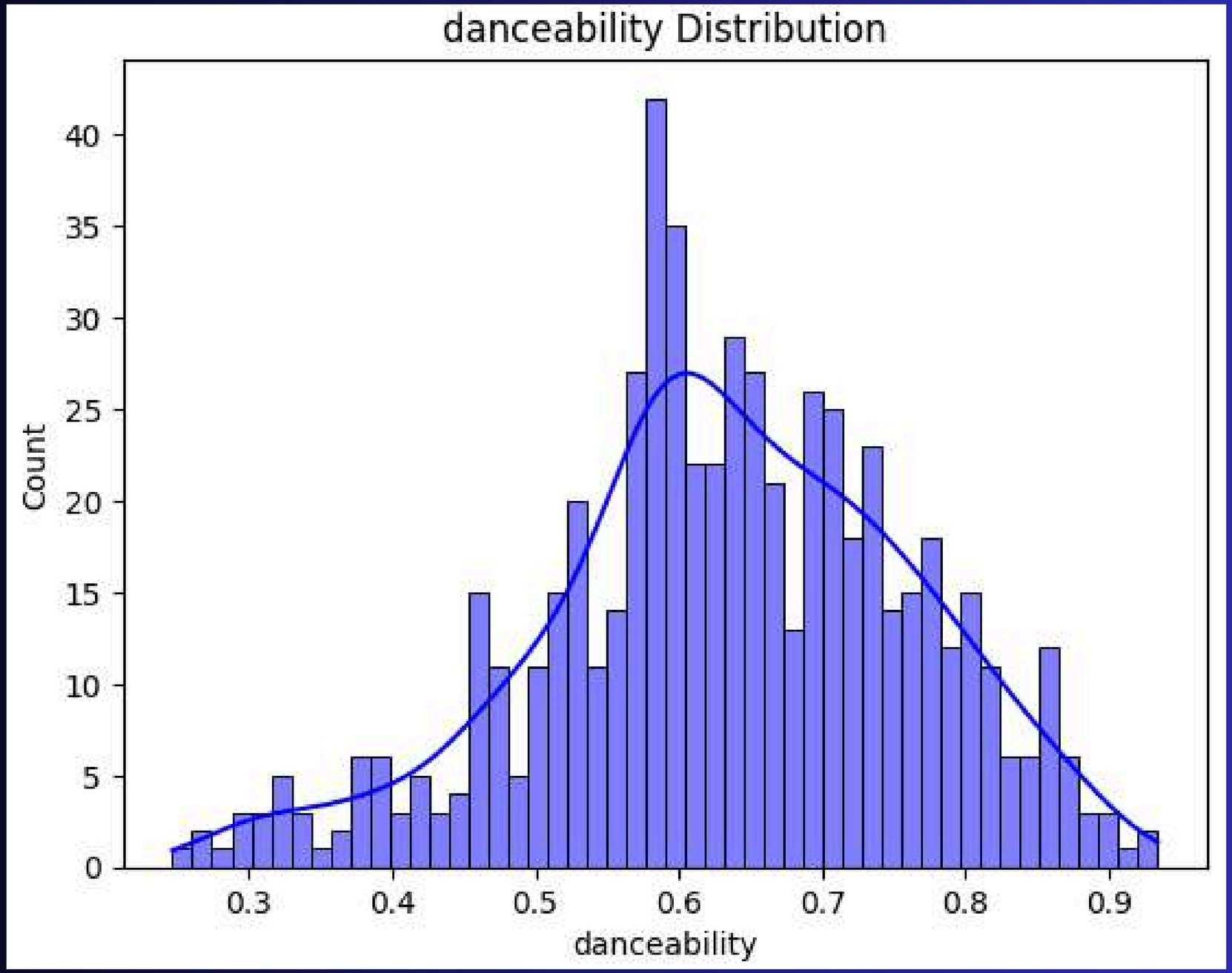

Correlation of Popularity vs Speechiness Over Time
Correlation Coefficient
0.4
0.2
0.0
−0.2
−0.4
1970
1980
1990
2000
2010
2020
Year

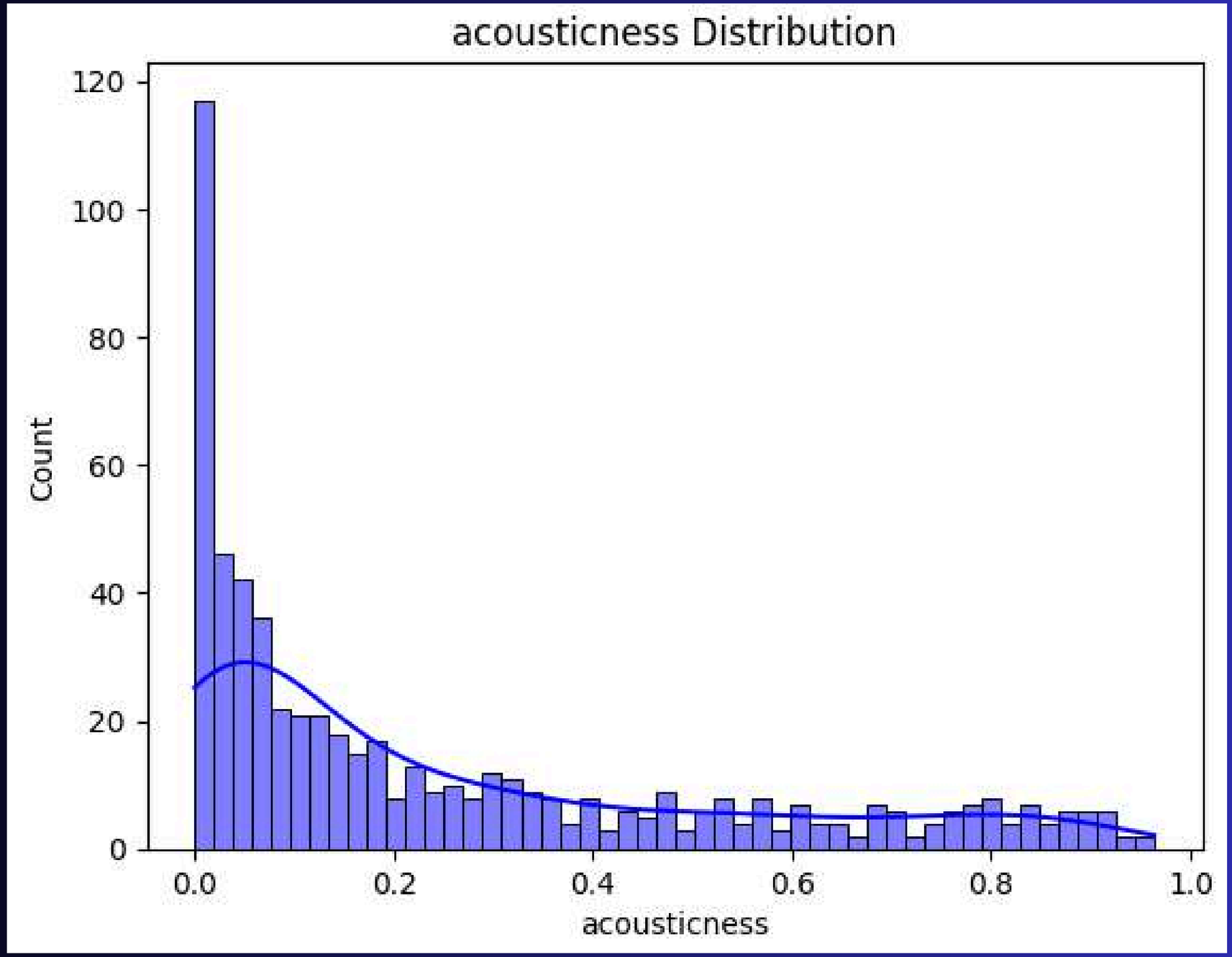

Correlation of Popularity vs Tempo Over Time
Correlation Coefficient
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
1970
1980
1990
2000
2010
2020
Year

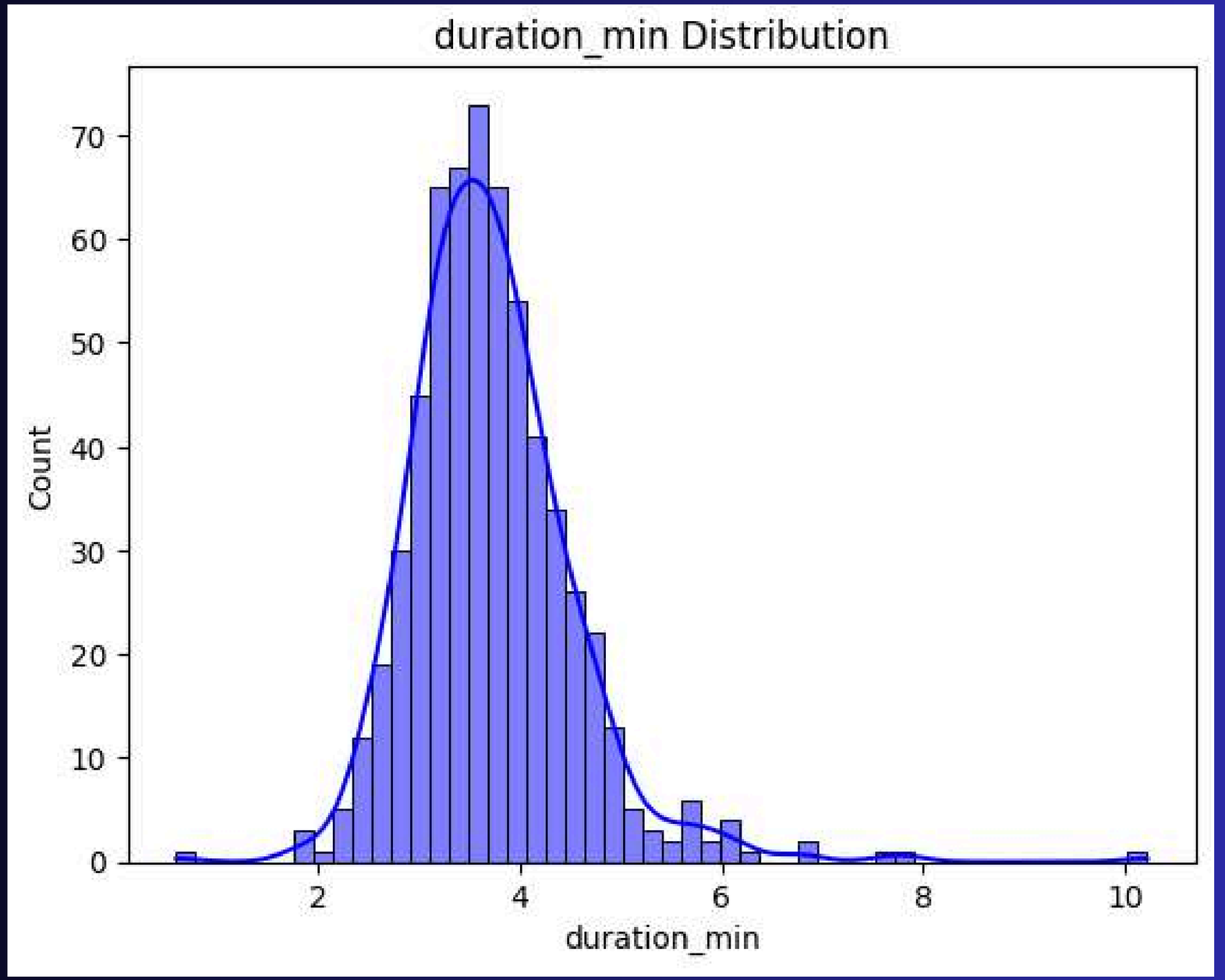

Correlation of Popularity vs Valence Over Time
Correlation Coefficient
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
1970
1980
1990
2000
2010
2020
Year

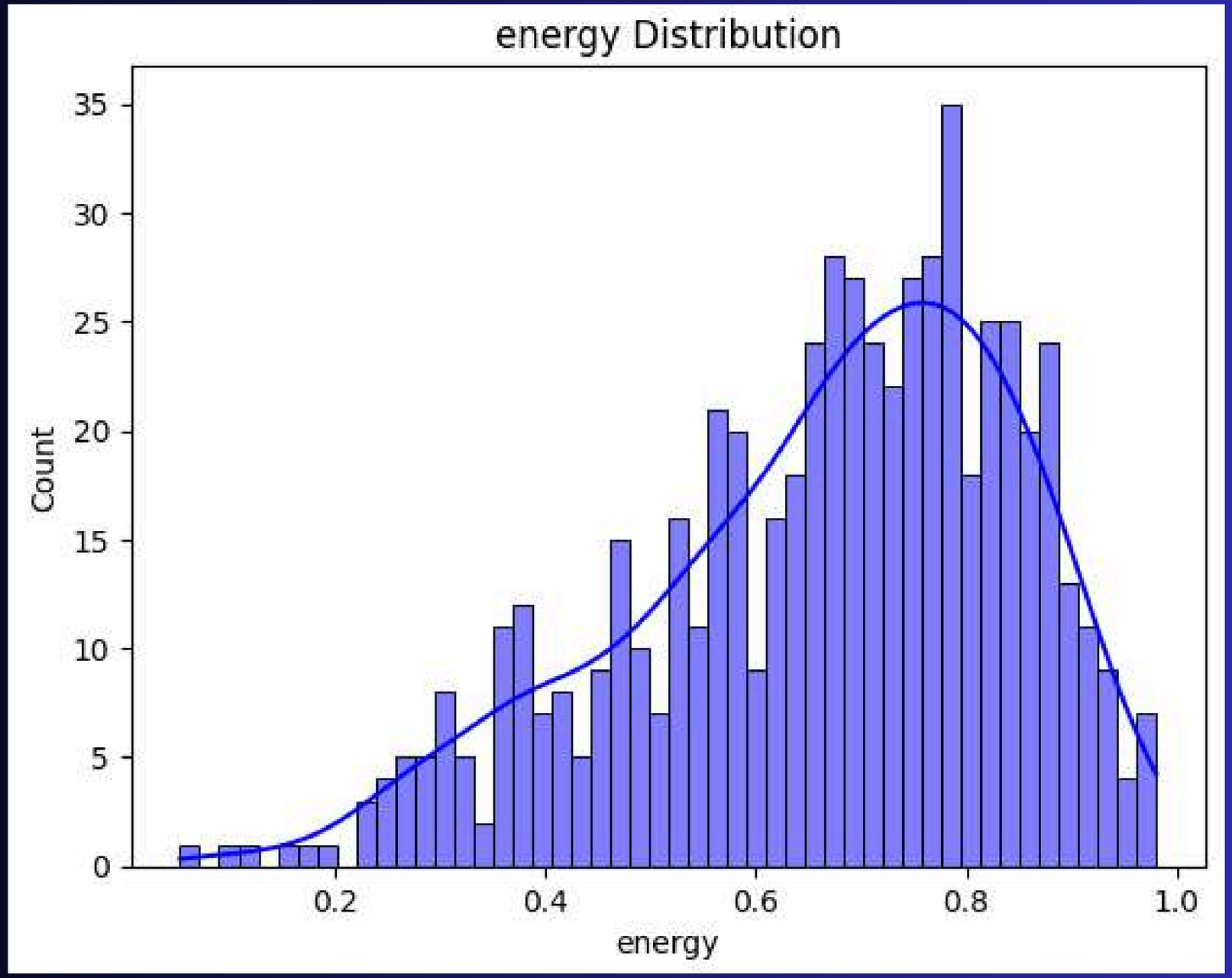

Proportional Distribution of language Categories in the Outliers
Language
English (66.4%)
Korean (22.5%)
Hindi (8.9%)
Tamil (2.2%)

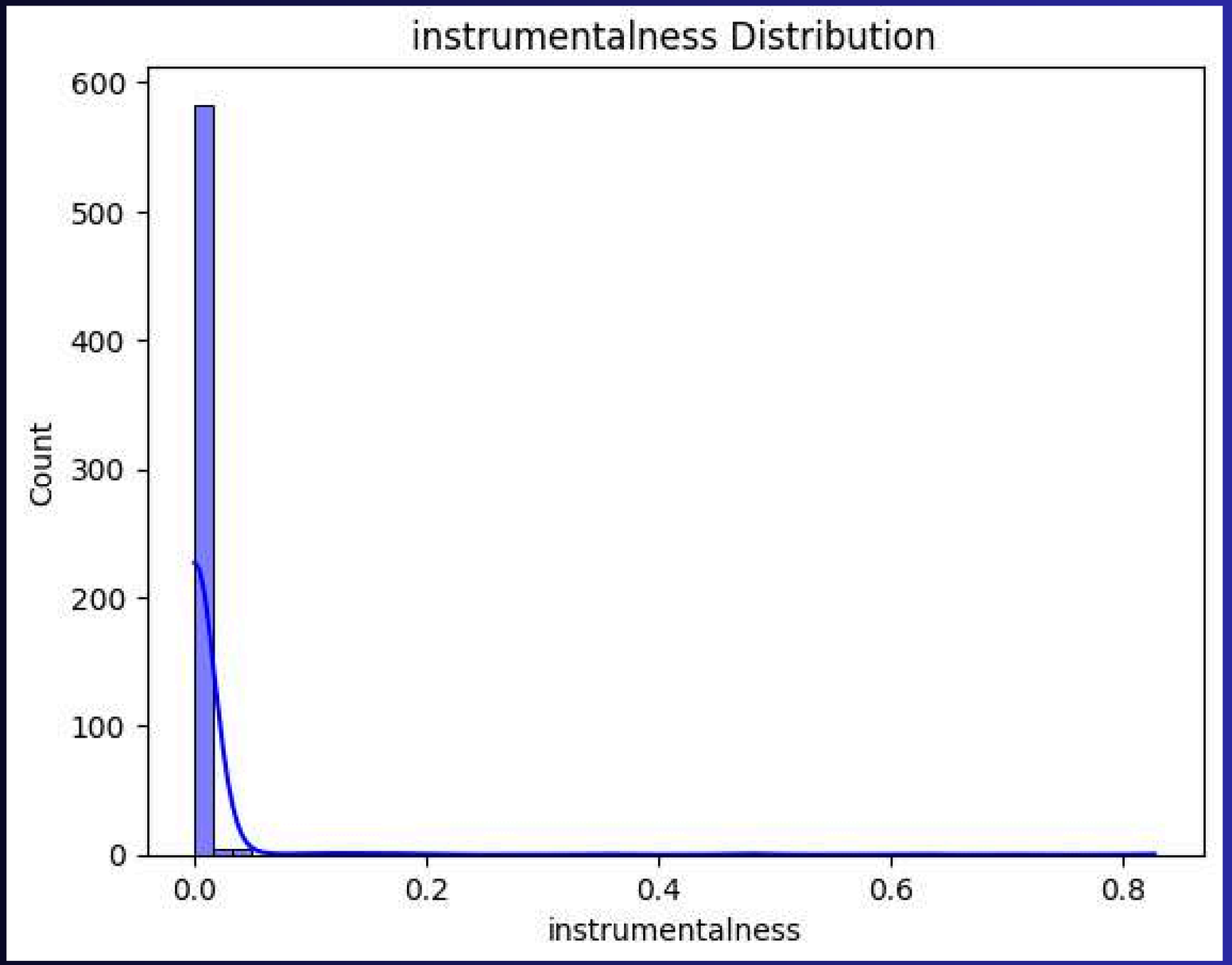

popularity Distribution
Count
0
10
20
30
40
50
60
70
70
75
80
85
90
popularity

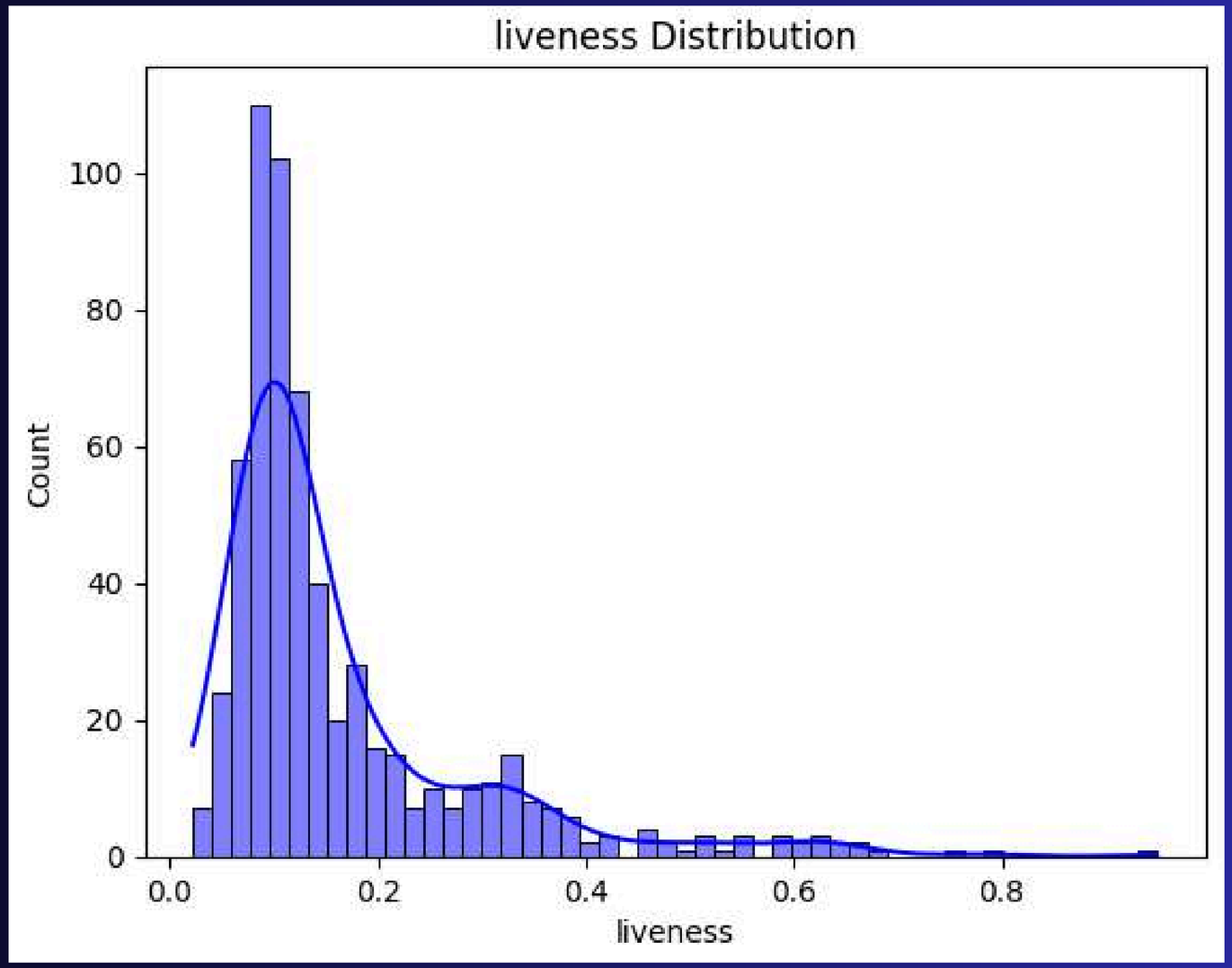

danceability Distribution
Count
0
5
10
15
20
25
30
35
40
0.3
0.4
0.5
0.6
0.7
0.8
0.9
danceability

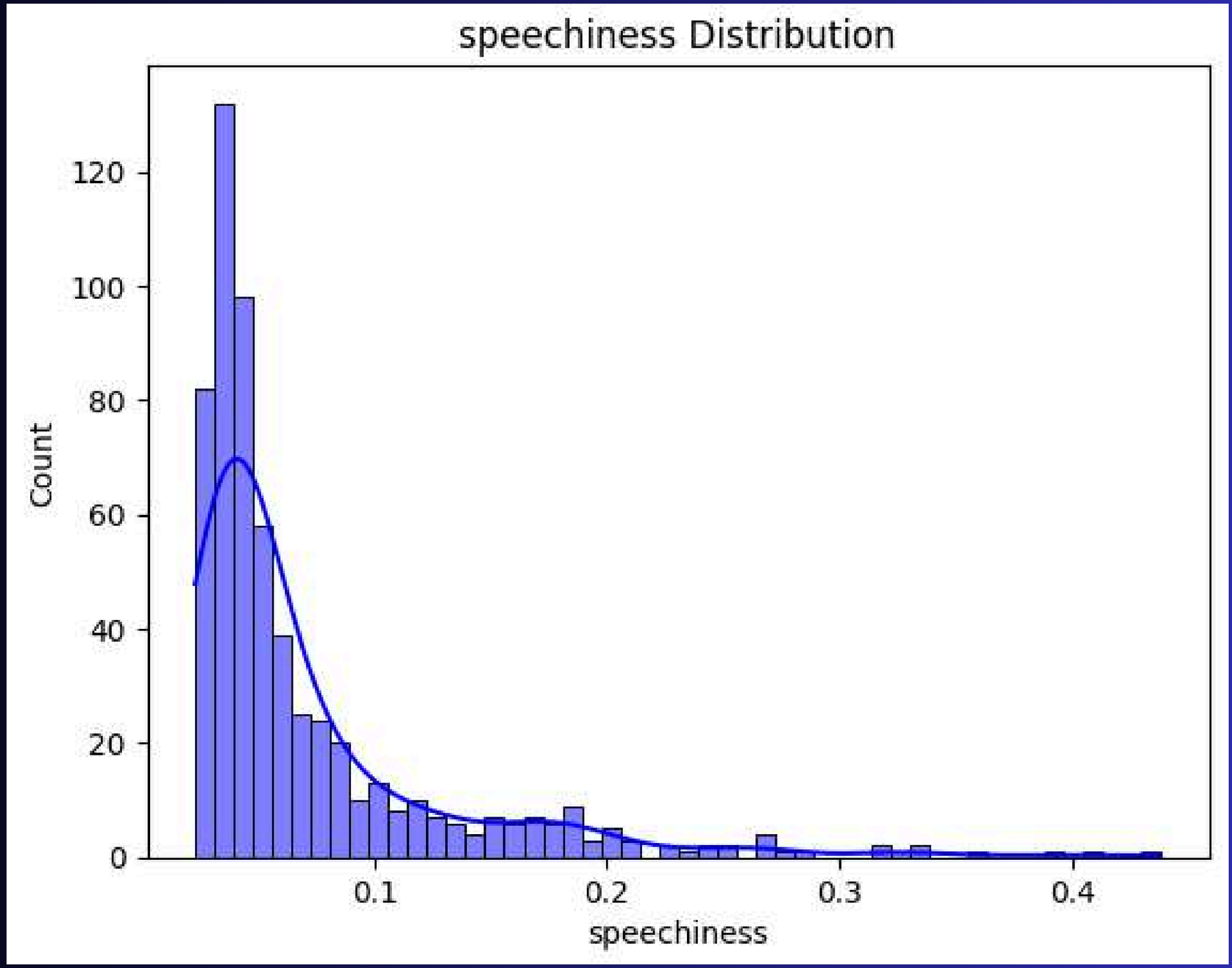

acousticness Distribution
Count
120
100
80
60
40
20
0
0.0
0.2
0.4
0.6
0.8
1.0
acousticness

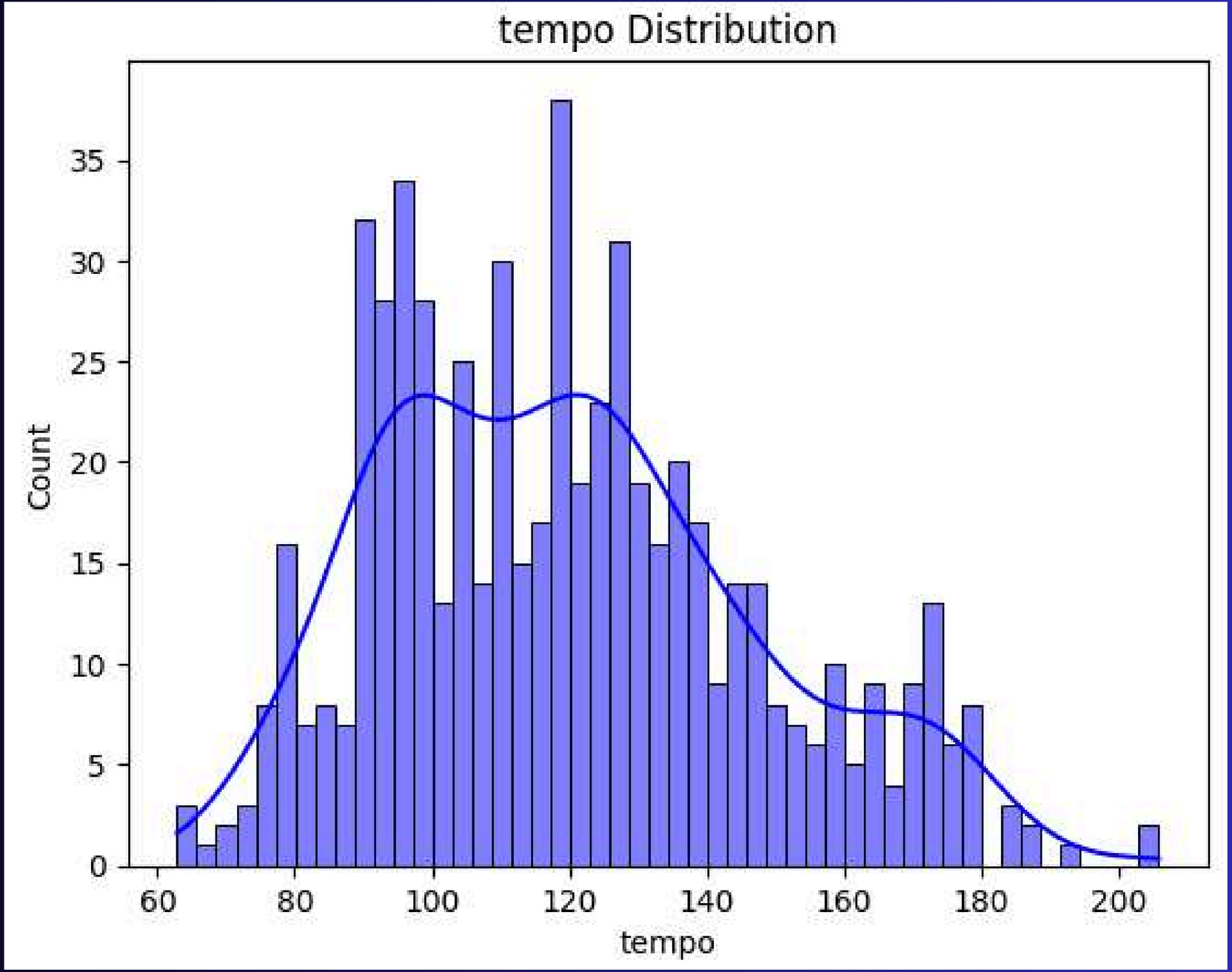

duration_min Distribution
Count
0
10
20
30
40
50
60
70
2
4
6
8
10
duration_min

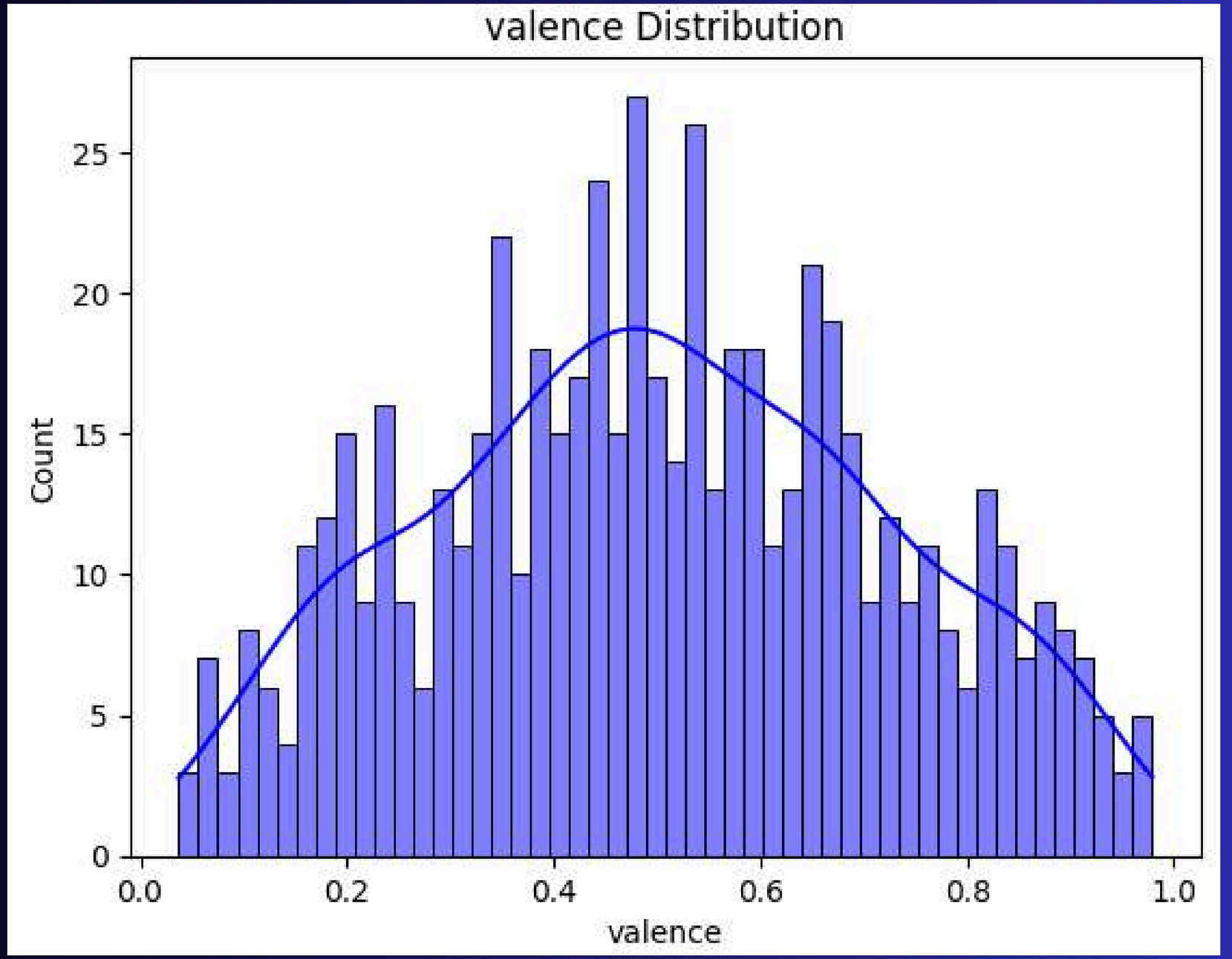

energy Distribution
Count
energy
0
5
10
15
20
25
30
35
0.2
0.4
0.6
0.8
1.0

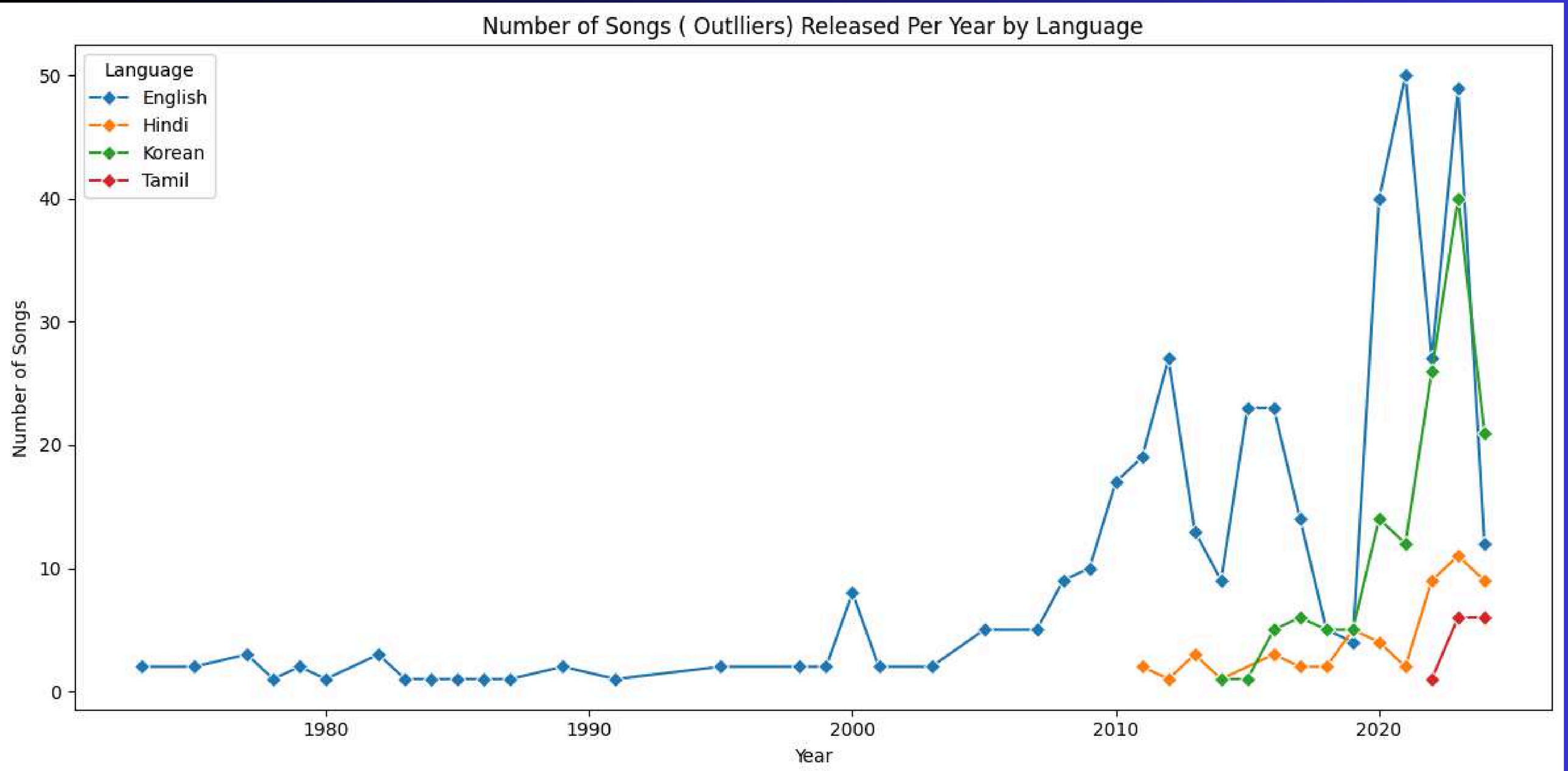

instrumentalness Distribution
600
500
400
300
200
100
0
Count
0.0
0.2
0.4
0.6
0.8
instrumentalness

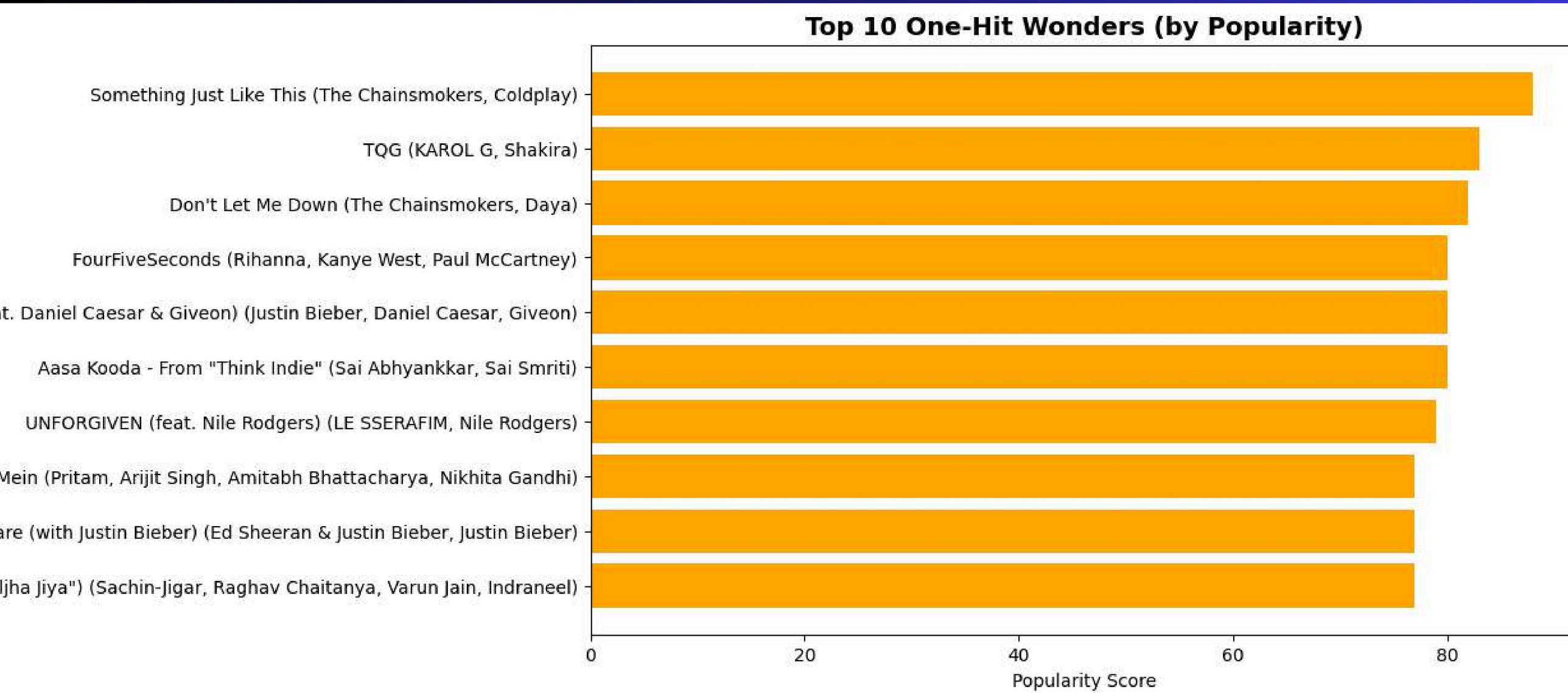

liveness Distribution
Count
100
80
60
40
20
0
0.0
0.2
0.4
0.6
0.8
liveness

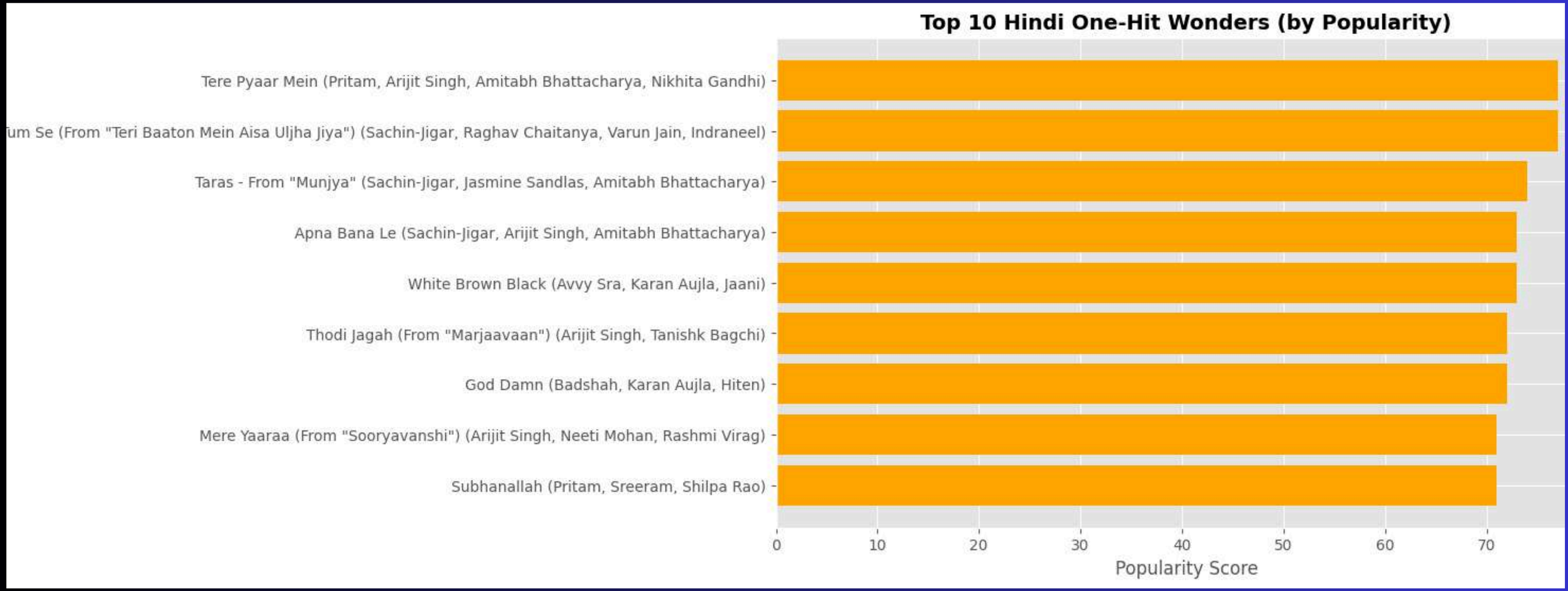

speechiness Distribution
Count
120
100
80
60
40
20
0
0.1
0.2
0.3
0.4
speechiness

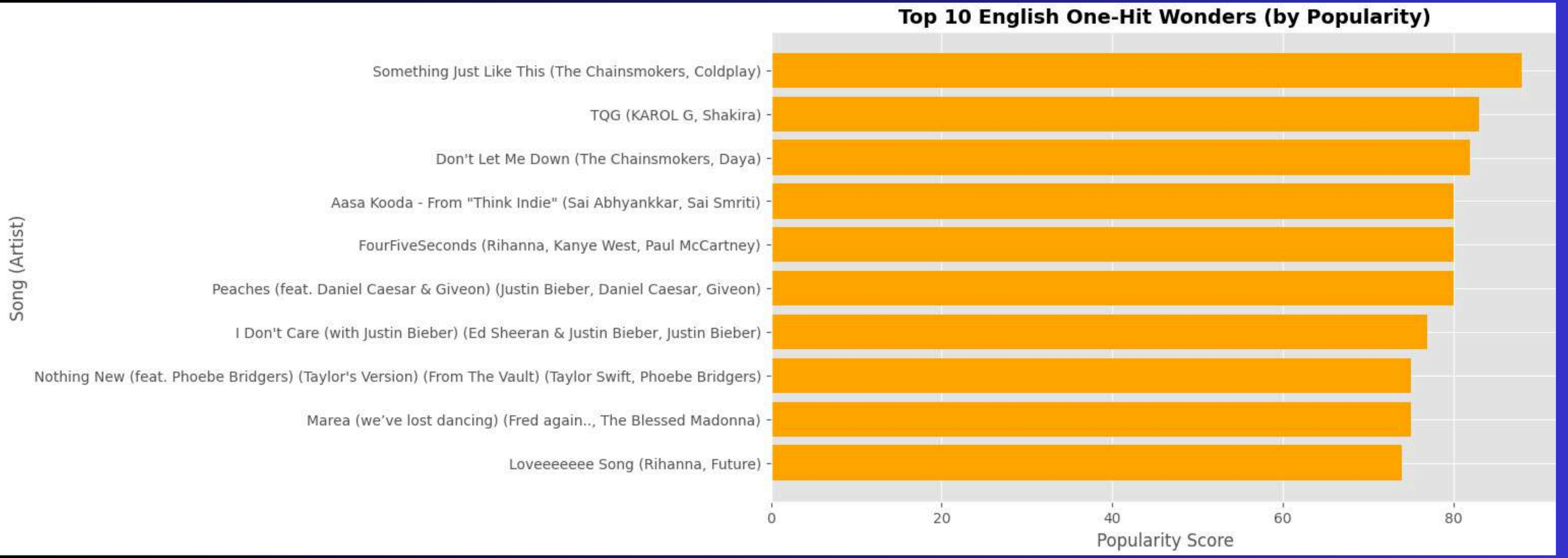

tempo Distribution
Count
0
5
10
15
20
25
30
35
60
80
100
120
140
160
180
200
tempo

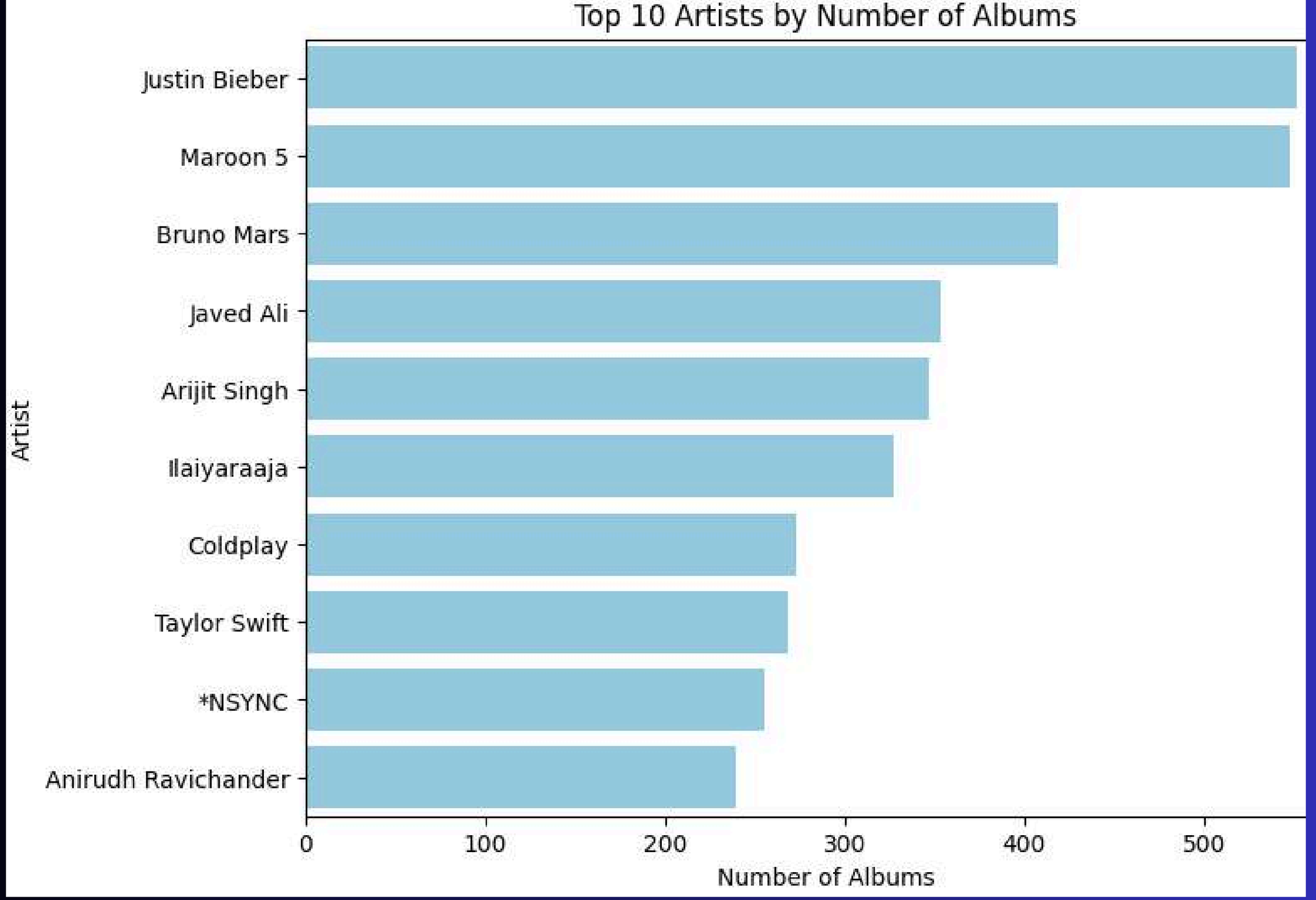

valence Distribution
Count
0
5
10
15
20
25
0.0
0.2
0.4
0.6
0.8
1.0
valence

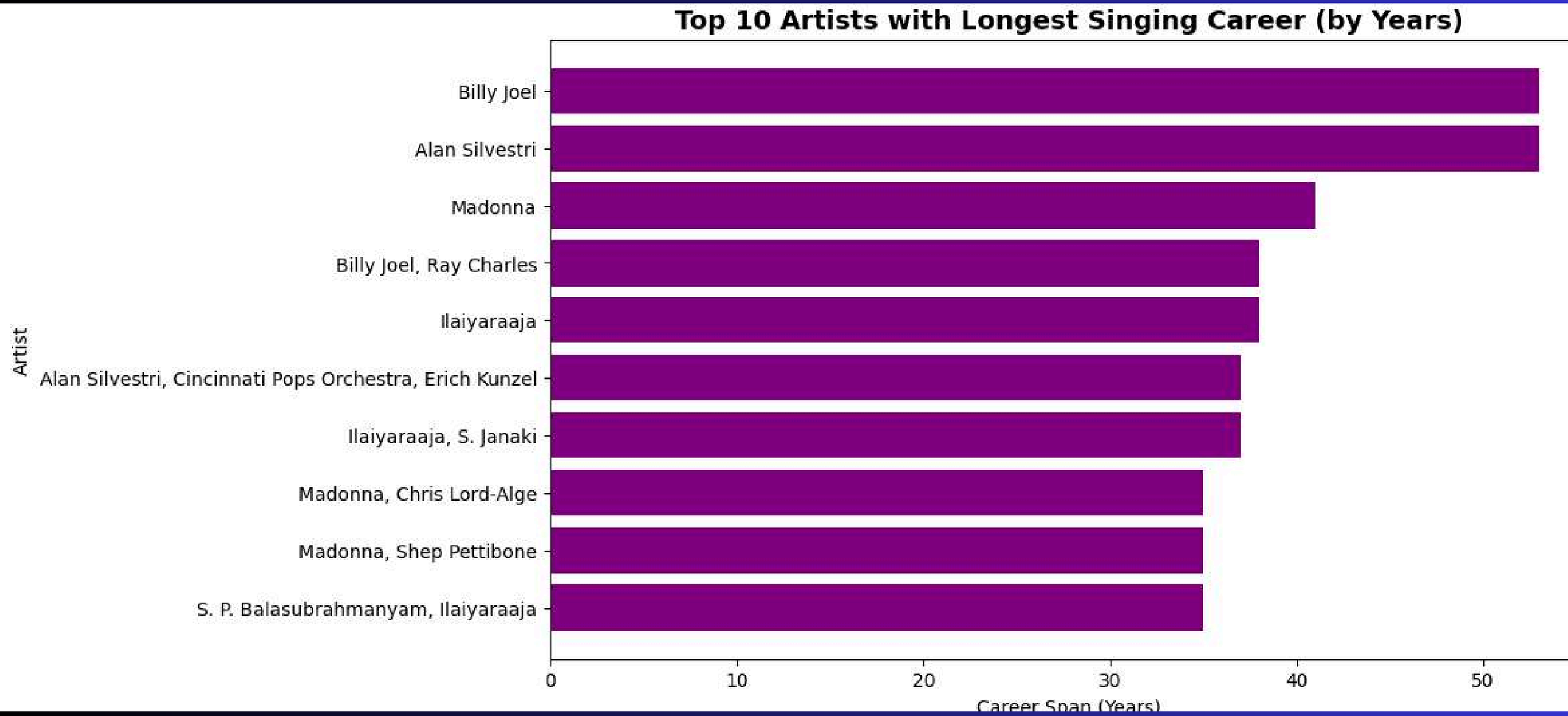

Number of Songs (Outlliers) Released Per Year by Language
Language
English
Hindi
Korean
Tamil
Number of Songs
0
10
20
30
40
50
1980
1990
2000
2010
2020
Year

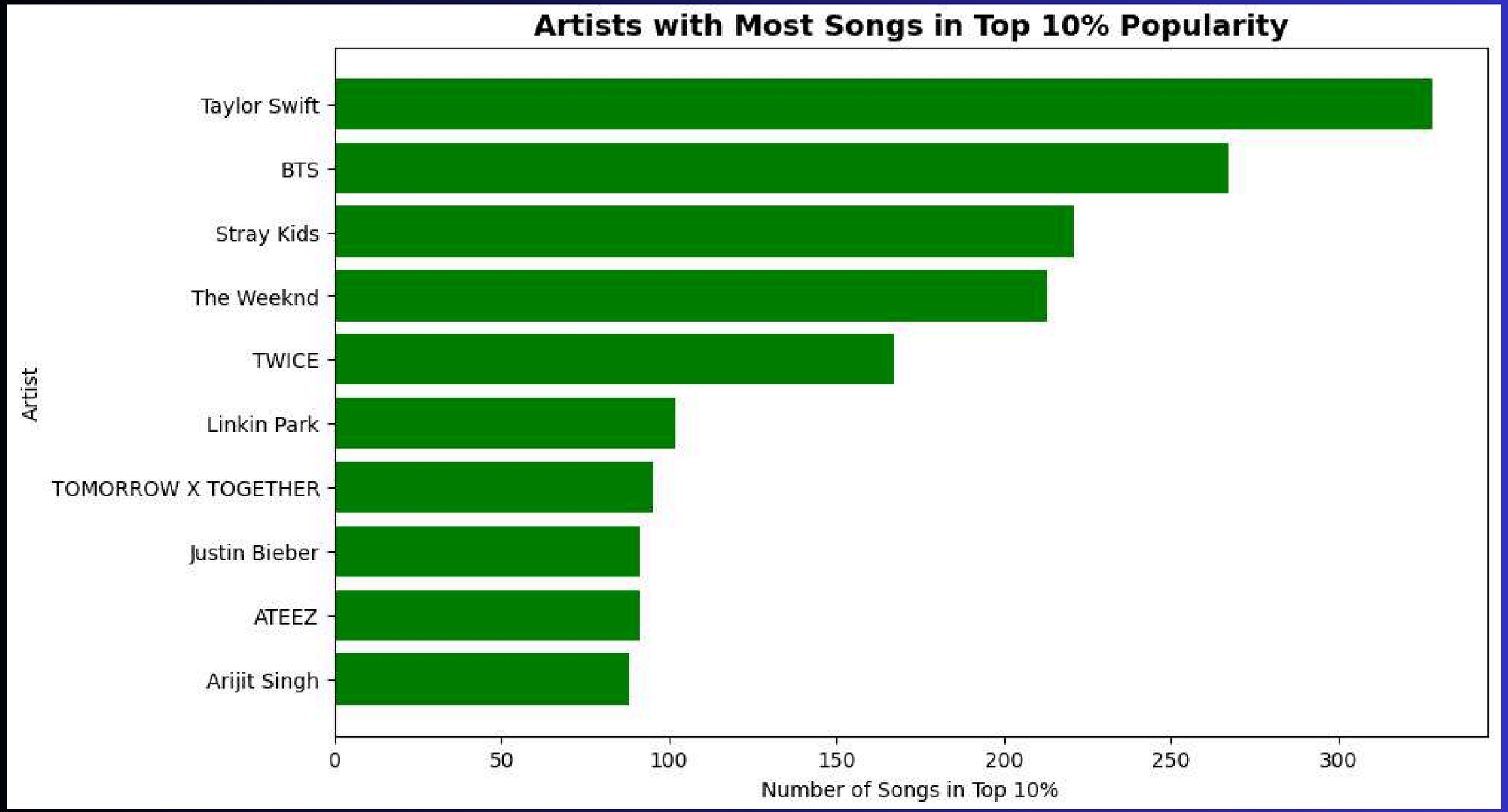

Top 10 One-Hit Wonders (by Popularity)
Something Just Like This (The Chainsmokers, Coldplay)
TQG (KAROL G, Shakira)
Don't Let Me Down (The Chainsmokers, Daya)
FourFiveSeconds (Rihanna, Kanye West, Paul McCartney)
t. Daniel Caesar & Giveon) (Justin Bieber, Daniel Caesar, Giveon)
Aasa Kooda - From "Think Indie" (Sai Abhyankkar, Sai Smriti)
UNFORGIVEN (feat. Nile Rodgers) (LE SSERAFIM, Nile Rodgers)
Mein (Pritam, Arijit Singh, Amitabh Bhattacharya, Nikhita Gandhi)
are (with Justin Bieber) (Ed Sheeran & Justin Bieber, Justin Bieber)
ljha Jiya") (Sachin-Jigar, Raghav Chaitanya, Varun Jain, Indraneel)
0
20
40
60
80
Popularity Score

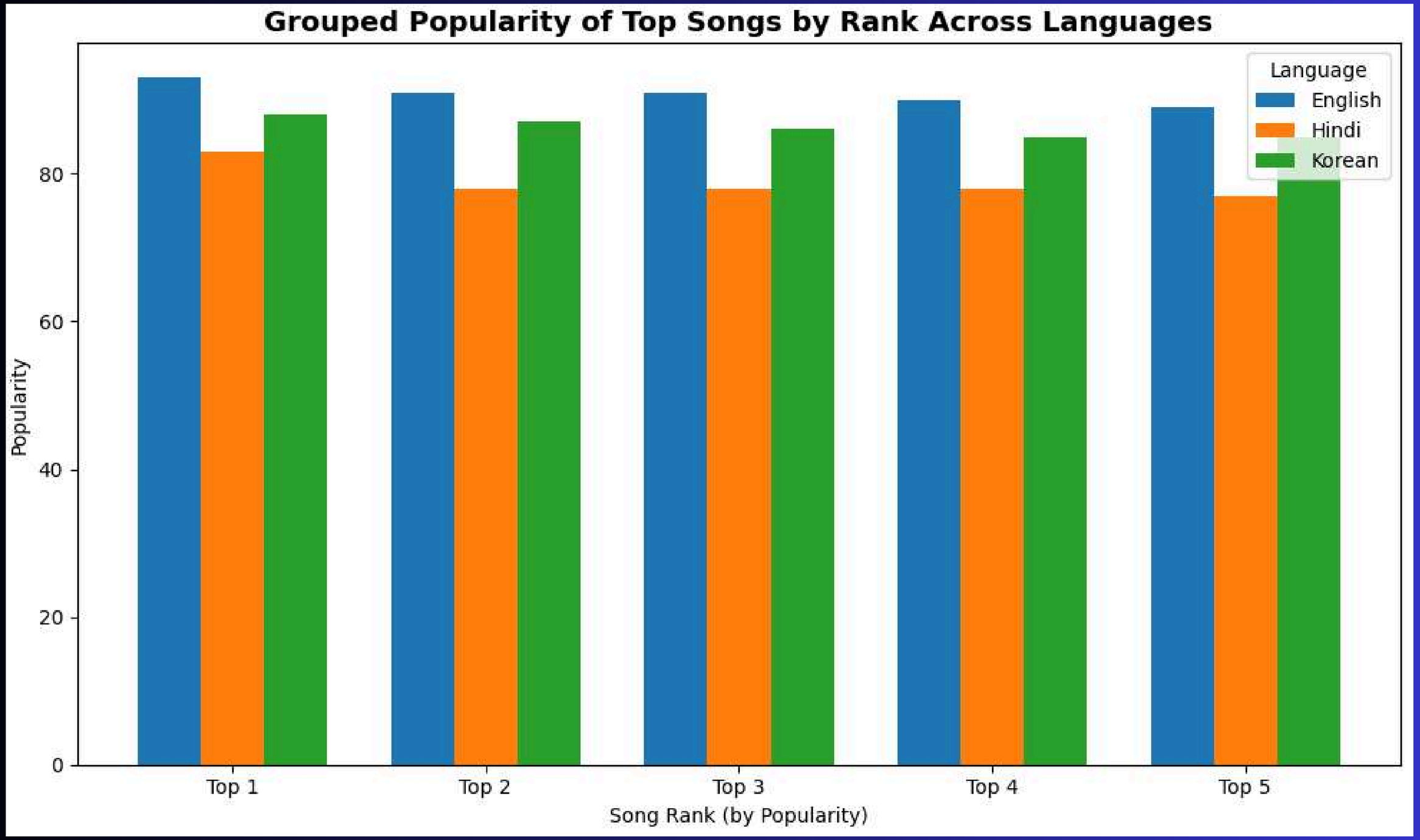

Top 10 Hindi One-Hit Wonders (by Popularity)
Tere Pyaar Mein (Pritam, Arijit Singh, Amitabh Bhattacharya, Nikhita Gandhi)
um Se (From "Teri Baaton Mein Aisa Uljha Jiya") (Sachin-Jigar, Raghav Chaitanya, Varun Jain, Indraneel)
Taras - From "Munjya" (Sachin-Jigar, Jasmine Sandlas, Amitabh Bhattacharya)
Apna Bana Le (Sachin-Jigar, Arijit Singh, Amitabh Bhattacharya)
White Brown Black (Avvy Sra, Karan Aujla, Jaani)
Thodi Jagah (From "Marjaavaan") (Arijit Singh, Tanishk Bagchi)
God Damn (Badshah, Karan Aujla, Hiten)
Mere Yaaraa (From "Sooryavanshi") (Arijit Singh, Neeti Mohan, Rashmi Virag)
Subhanallah (Pritam, Sreeram, Shilpa Rao)
0
10
20
30
40
50
60
70
Popularity Score

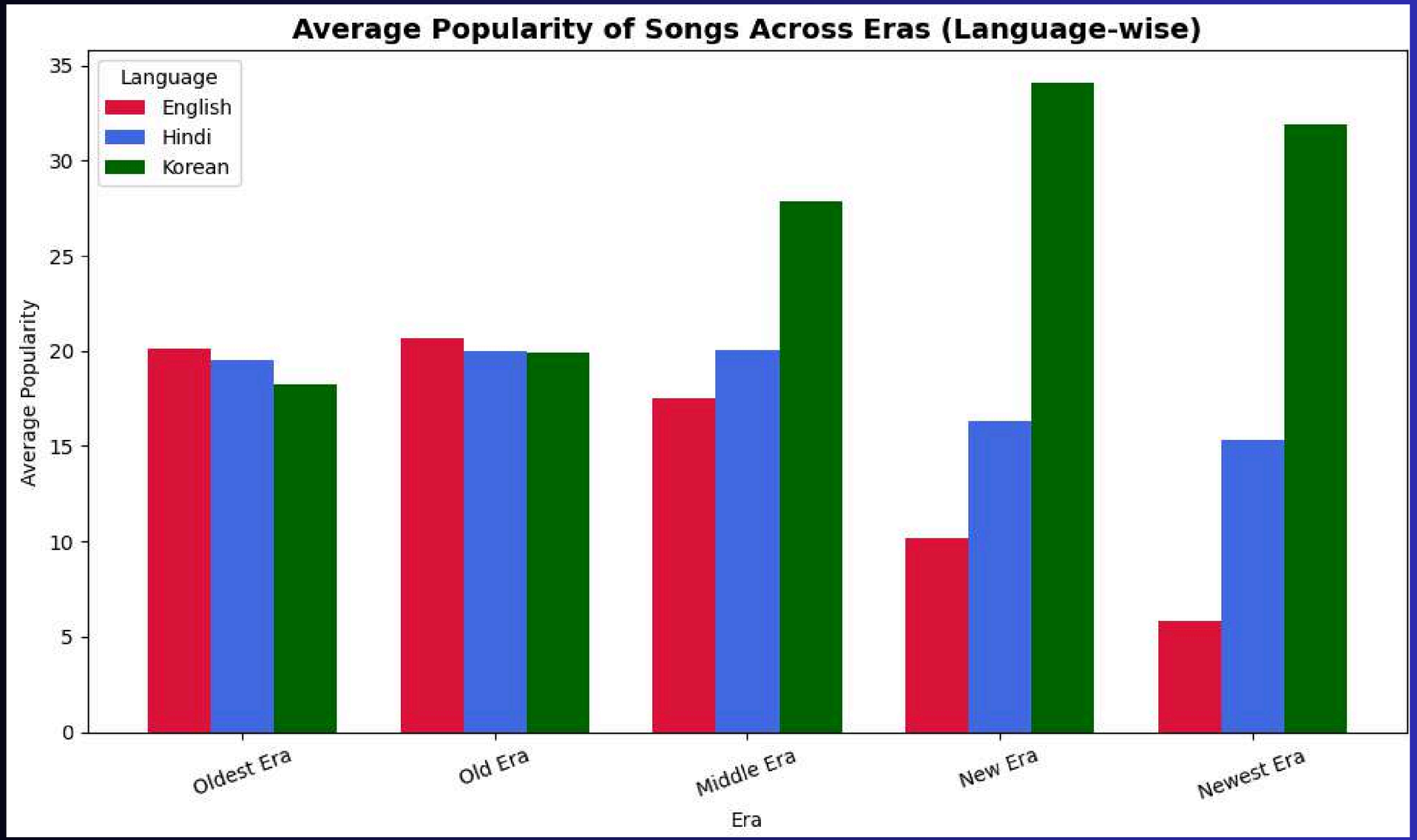

Top 10 English One-Hit Wonders (by Popularity)
Song (Artist)
Something Just Like This (The Chainsmokers, Coldplay)
TQG (KAROL G, Shakira)
Don't Let Me Down (The Chainsmokers, Daya)
Aasa Kooda - From "Think Indie" (Sai Abhyankkar, Sai Smriti)
FourFiveSeconds (Rihanna, Kanye West, Paul McCartney)
Peaches (feat. Daniel Caesar & Giveon) (Justin Bieber, Daniel Caesar, Giveon)
I Don't Care (with Justin Bieber) (Ed Sheeran & Justin Bieber, Justin Bieber)
Nothing New (feat. Phoebe Bridgers) (Taylor's Version) (From The Vault) (Taylor Swift, Phoebe Bridgers)
Marea (we've lost dancing) (Fred again.., The Blessed Madonna)
Loveeeeeee Song (Rihanna, Future)
0
20
40
60
80
Popularity Score

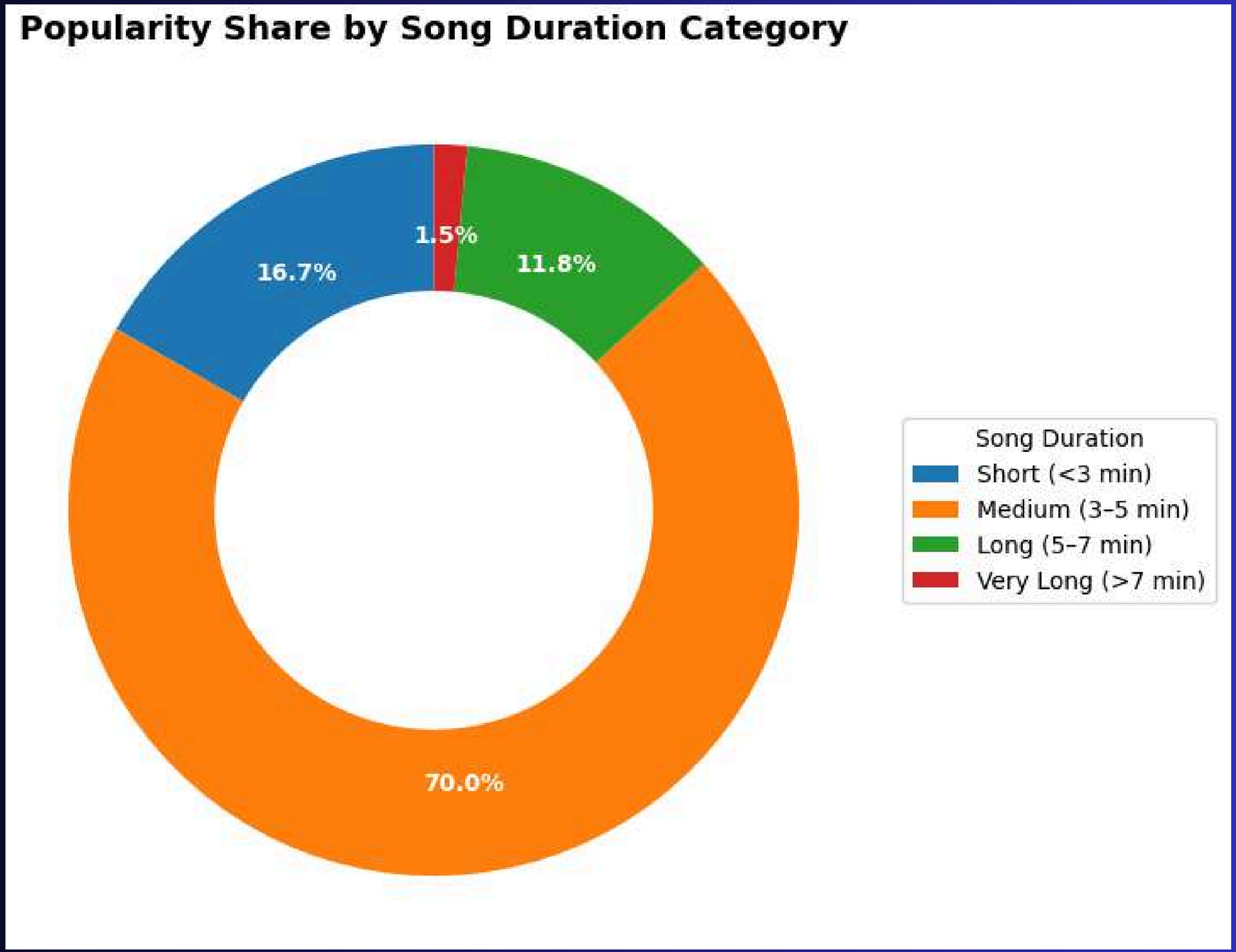

Top 10 Artists by Number of Albums
Justin Bieber
Maroon 5
Bruno Mars
Javed Ali
Arijit Singh
Ilaiyaraaja
Coldplay
Taylor Swift
*NSYNC
Anirudh Ravichander
Artist
0
100
200
300
400
500
Number of Albums

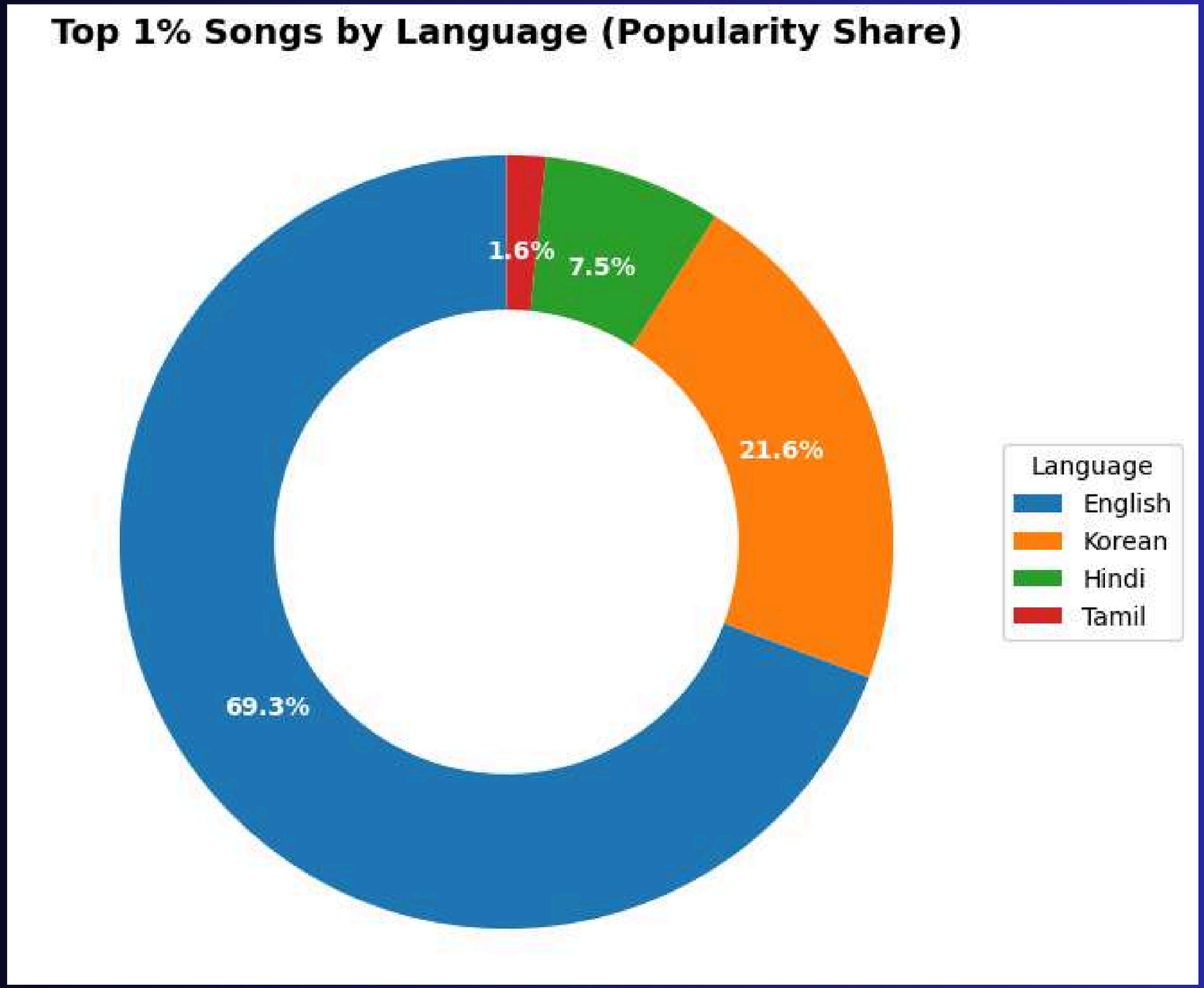

Top 10 Artists with Longest Singing Career (by Years)
Artist
Billy Joel
Alan Silvestri
Madonna
Billy Joel, Ray Charles
Ilaiyaraaja
Alan Silvestri, Cincinnati Pops Orchestra, Erich Kunzel
Ilaiyaraaja, S. Janaki
Madonna, Chris Lord-Alge
Madonna, Shep Pettibone
S. P. Balasubrahmanyam, Ilaiyaraaja
0
10
20
30
40
50
Career Span (Years)


Artists with Most Songs in Top 10% Popularity
Taylor Swift
BTS
Stray Kids
The Weeknd
TWICE
Linkin Park
TOMORROW X TOGETHER
Justin Bieber
ATEEZ
Arijit Singh
Artist
0
50
100
150
200
250
300
Number of Songs in Top 10%


Grouped Popularity of Top Songs by Rank Across Languages
Language
English
Hindi
Korean
Popularity
0
20
40
60
80
Top 1
Top 2
Top 3
Top 4
Top 5
Song Rank (by Popularity)


Average Popularity of Songs Across Eras (Language-wise)
Language
English
Hindi
Korean
Average Popularity
0
5
10
15
20
25
30
35
Oldest Era
Old Era
Middle Era
New Era
Newest Era
Era


Popularity Share by Song Duration Category
16.7%
1.5%
11.8%
70.0%
Song Duration
Short (<3 min)
Medium (3–5 min)
Long (5–7 min)
Very Long (>7 min)


Top 1% Songs by Language (Popularity Share)
1.6%
7.5%
21.6%
69.3%
Language
English
Korean
Hindi
Tamil

1. Core Track Structure

Keep duration in the 3–5 min sweet spot; avoid extremes.
Target tempo range: 80–200 BPM, balancing radio-friendly and energetic pacing.

2. Sound & Energy Profile

Mix for moderate-to-high energy and danceability (0.5–0.8).
Favor digital/electronic textures; minimize excessive acousticness.
Maintain loudness clarity with dynamic control—energy should feel strong but not distorted.

3. Emotional & Creative Factors

Both happy (high valence) and sad (low valence) songs can succeed; focus on mix quality over mood.

Experiment with genre fusions—cross-style collaborations often yield standout tracks.

4. Market & Audience Orientation

English remains dominant, but K-pop and Hindi show rapid growth—adapt mixing aesthetics accordingly.

Solo tracks perform slightly better on average, but international collaborations create global hits.

Market & Language Insights :

Top 1% Songs by Language: English (69.3%) dominates, Korean (21.6%) rising, Hindi (7.5%) and Tamil (1.6%) smaller shares.

Popularity by Language: English highest median (approx 62), others lower (8–30).

Energy by Language: Korean highest (approx 0.8), English lowest median (approx 0.5).

Tempo by Language: Stable medians around 115–125 BPM, Malayalam peaked at ~145 BPM (2000s).

Danceability by Language: All languages stable (approx 0.6–0.7).

Valence by Language: Tamil highest (approx 0.7), English lowest (approx 0.4).

Top-N Analysis (Artists & Tracks) :

Chainsmokers: 2 collabs in global top 3 one-hit wonders cross-collaboration success.
Justin Bieber: Featured in 3 top one-hit wonders - frequent collaborator pattern.
Hindi One-Hit Wonders: 9 of top 10 from films; Sachin-Jigar appear 3 times, Arijit Singh in 4 songs.
Top Artists by Albums: Justin Bieber & Maroon 5 lead with 550+ albums; Bruno Mars follows with 420.
Longest Careers: Billy Joel & Alan Silvestri ~53 years; Madonna 41 years; Ilaiyaraaja multiple appearances.
Most Songs in Top 10% Popularity: Taylor Swift dominates with 330+ songs; 5 K-pop groups in top 10.
Song Duration Popularity (Top 1%): Medium (3–5 min) = 70%, Short = 16.7%, Long = ~13%.

Time Series Insights for Mixing Engineers

1. Track Releases & Growth

English: Explosive rise from ~900 (2019) to 3,300+ songs (2025) - dominant production volume.
Hindi: Minimal for decades, but sharp surge post-2023, reaching 1,100+ songs.
Korean: Grew from near zero (2010) to 600–700 songs (2025) - reflects K-pop’s global boom
Overall: 2019–2020 = inflection point across languages due to digital platforms

2. Popularity Trends Over Time

Maximum Popularity: Climbed from ~52 (1970) to ~93 (2025), but with high volatility (50–90)
Median Popularity: Peaked ~46 (mid-1970s), collapsed below 10 in 1980s, stayed low (5–20) since.
Minimum Popularity: Fell near-zero after 1976, stayed at 0–1 for 50+ years.
Pattern: Rising inequality - few mega-hits dominate, while most songs remain obscure.

3. Music Feature Evolution

Acousticness: Shift from mostly acoustic (1970s) to mostly electronic/digital (2000s onwards).
Danceability: Stable over 50 years (~0.6).
Energy: Significant rise post-2000 → songs became more intense and energetic.

4. Collaboration Trends by Language

Hindi: 100% collaborations (2003–2007), then solos dominated; recently (2023–2024) collabs surged again to 85–92%.
English: Predominantly solos until 1990s (95–100%), then collabs rose to 40–50% in 2019–2024.
Korean: Consistently solo (95–100%), tiny rise to ~10–12% collabs in 2023–2024.
Tamil: Opposite trend—65–100% collabs (1976–2010), then flipped to 70–100% solos post-2015.

5. Changing Correlations with Popularity

Danceability vs Popularity: Strong in 1970s (0.77), now near zero.
Duration vs Popularity: Strong in 1970s (0.81), now near zero.
Energy vs Popularity: Moderate in 1970s (0.67), still slightly positive today.

6. Outliers & Distributions

Danceability: Bell curve centered at ~0.6.
Valence: Normal distribution around ~0.45 (slightly negative lean).
Liveness & Speechiness: Heavily skewed low → most songs are studio-recorded, not speech-heavy.
Popularity: Right-skewed, top cluster ~70–73 → only a few tracks achieve high success.

Future Work Directions:

1. User Listening Behaviors

Study how listeners engage with tracks over time.

2. Recommendation System Development

Build smarter systems using feature relationships.

3. Exploring Lyrics Aspects

Extend analysis beyond audio to lyrical content.

THANKS YOU!