Spotify Dataset Exploratory Data Analysis

Uncovering patterns, trends, and insights hidden within millions of tracks to understand what makes music resonate with listeners worldwide
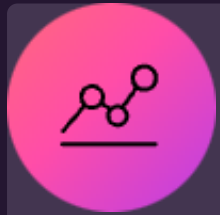
Dataset Overview & Scope

Key Data Points

- Song titles and artist information

- Popularity metrics

- Audio features: energy, danceability, valence

- Technical attributes: tempo, loudness, speechiness

- Acoustic characteristics and liveness scores

Our Goal :

- Observing Changes in Different Music Features With Decades

- What are The preferences in songs in 2020's decade

- Most popular artist in recent decades in several languages

- valence and characteristics

# Observing Changes in Different Music Features With Decades

Change In Valence Score in Decades

- This suggests most songs music has generally become less cheerful and more subdued or melancholic over the last five decades.
- While music with all types of moods still exists, the "average" song is becoming less overtly joyful. This could be due to changes in production style, genre popularity (e.g., the rise of darker electronic music or introspective hip-hop/pop), or a reflection of broader cultural moods.
- This suggests the shift away from the peak cheerfulness of the 70's/80's primarily happened in the 90's, and then the average mood settled into a more neutral or slightly positive-to-neutral range for a few decades before the latest dip
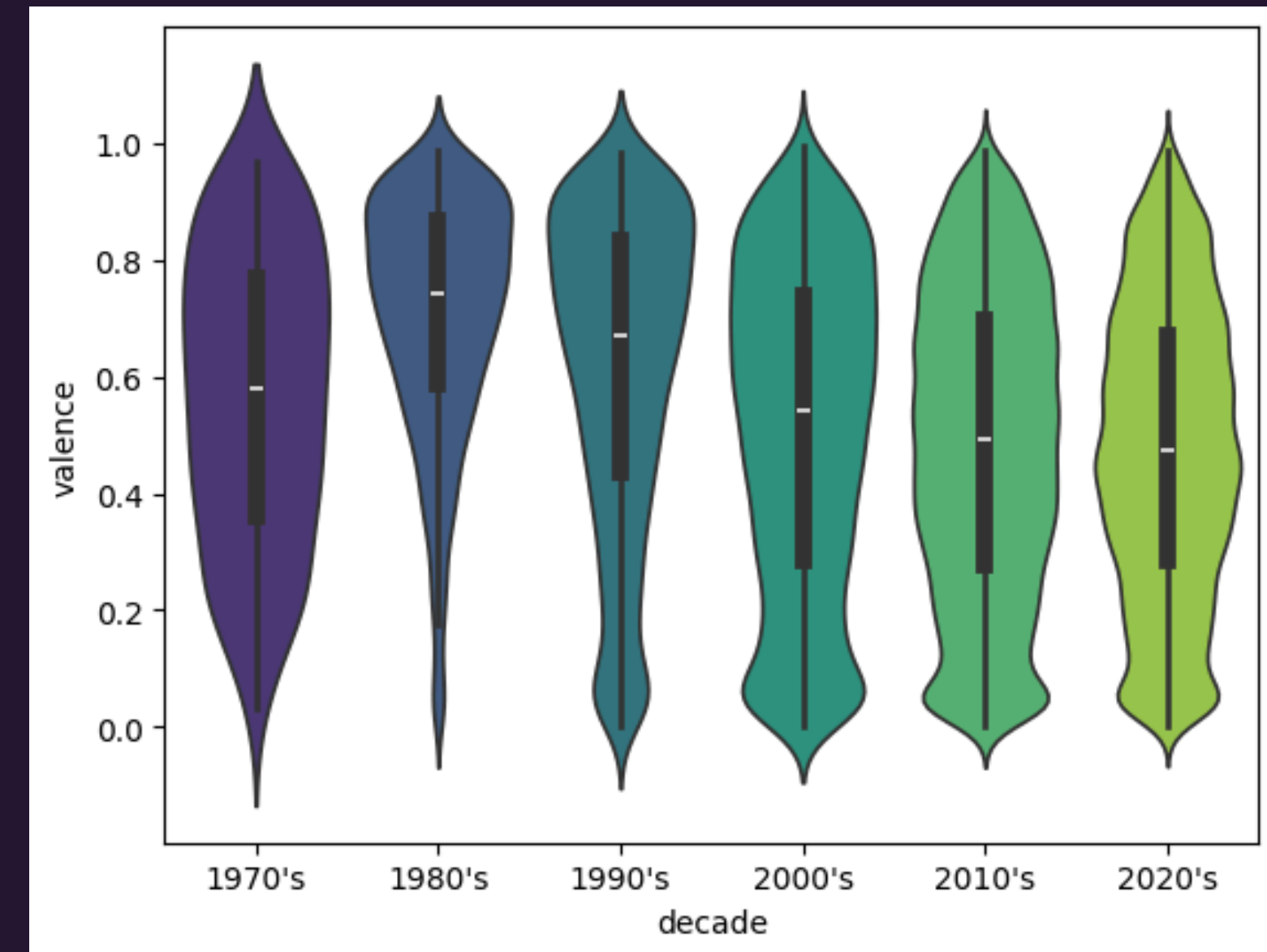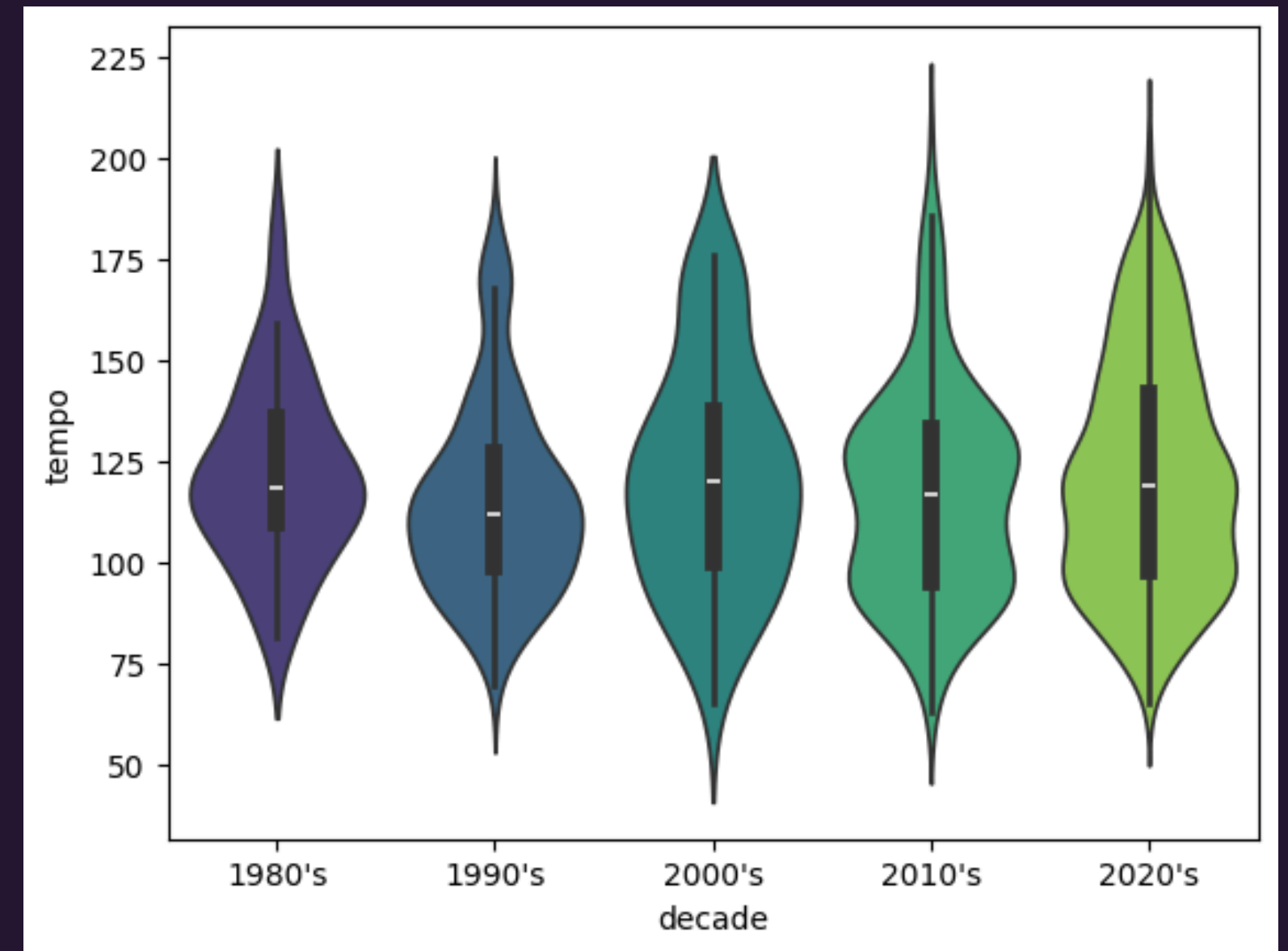
Change In Valence Score in Decades

- **This means that most songs that are produced are getting less cheerful but on the other hand in popularity scatterplot we observe a flat regression line . Which means popularity is not influenced by cheerfulness or sadness of the songs. This also indicates versatile music taste of song listeners. so valence is not a factor for popularity**


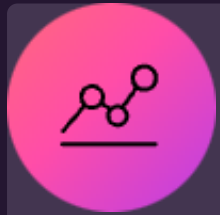valence and popularity scatterplot
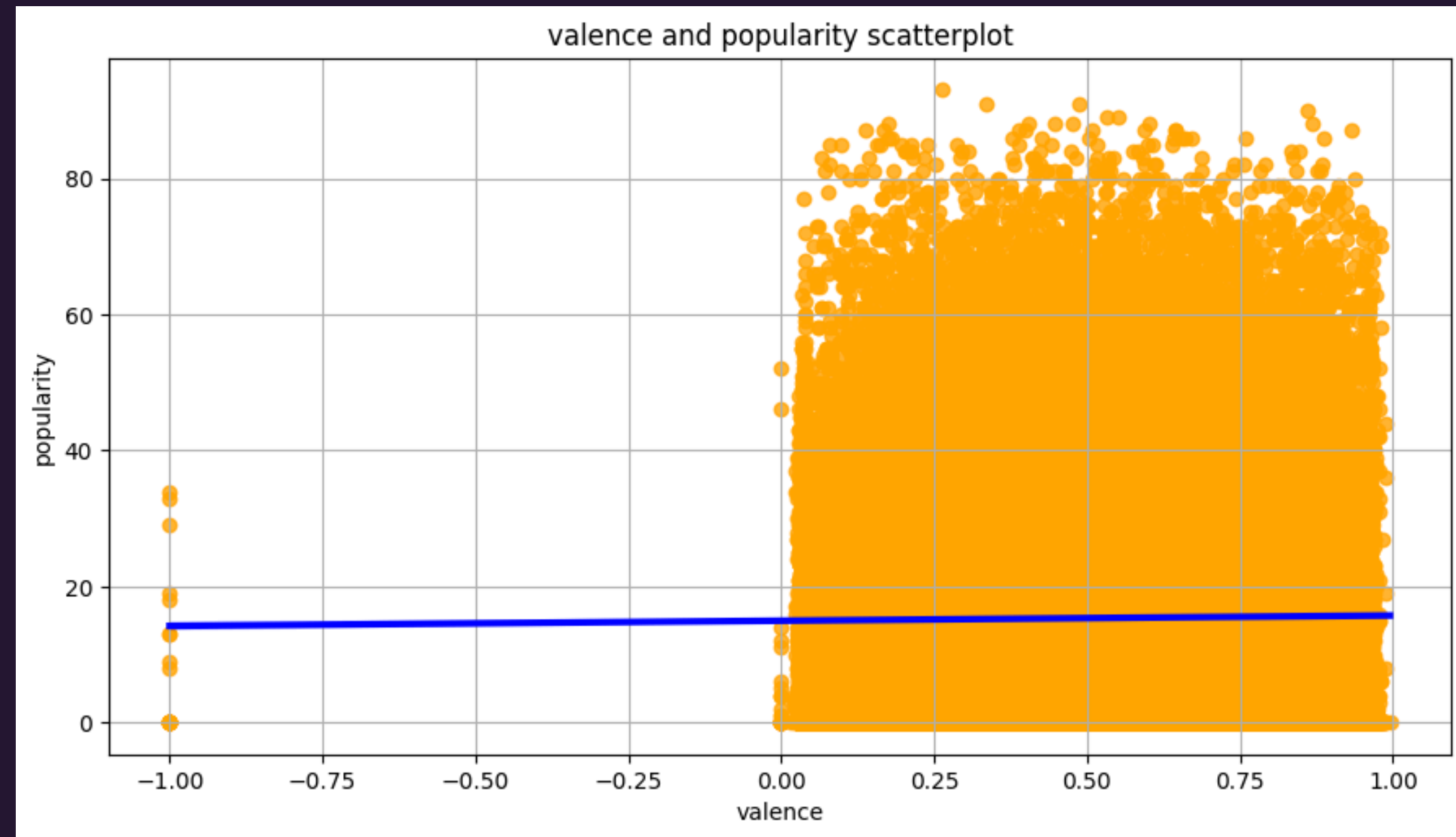
Change In Tempo Score in Decades

- **Consistent Median: The median tempo has hovered around 110 to 130 BPM for all five decades, suggesting that this range remains the most common speed for music.**
- **Increased Range/Outliers: The vertical length of the "violin" (the range of tempos) is dramatically longer for the 2010's and especially the 2020's. This indicates that music produced recently covers a much wider spectrum of speeds, from very slow to very fast.**

Change In Danceability Score in Decades

- **Extreme Diversity in the 1990s and 2020s:** These two decades show the longest tails, indicating that they produced the widest variety of music, from completely non-danceable (scores near 0.0) to maximum danceability (scores near 1.0).
- **Increased Focus on Mid-Range (2000s):** The 2000s distribution is notable for avoiding the low end, suggesting that most music produced was at least moderately danceable.
- **Stability of Peak Danceability:** For all decades, the music frequently reaches the highest danceability scores (near 1.0), which suggests the presence of highly dance-focused genres (e.g., electronic, pop, hip-hop) has been consistent over the 40-year period.
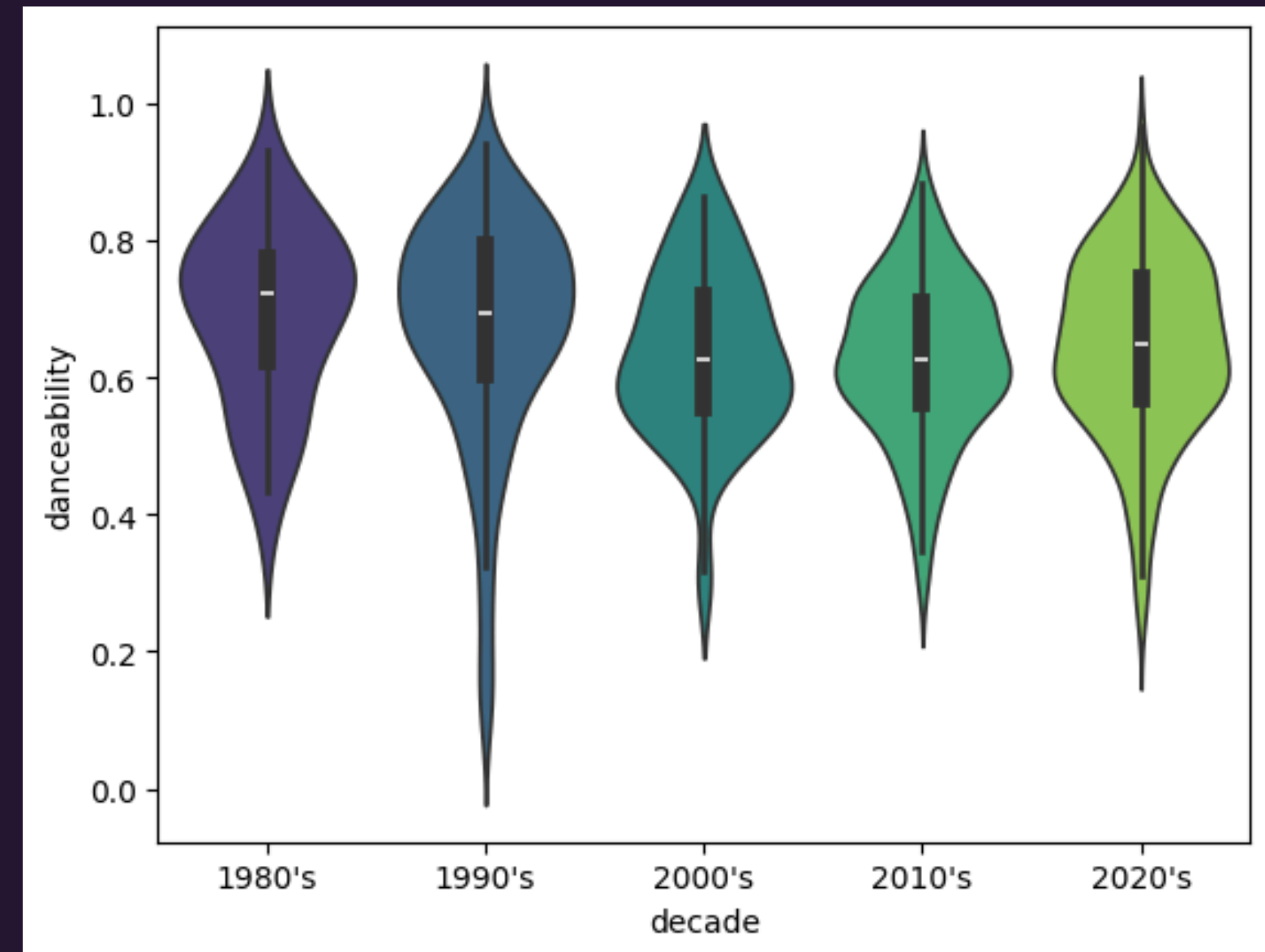
What are The preferences in songs in 2020's decade

Music Preferences of 2020's
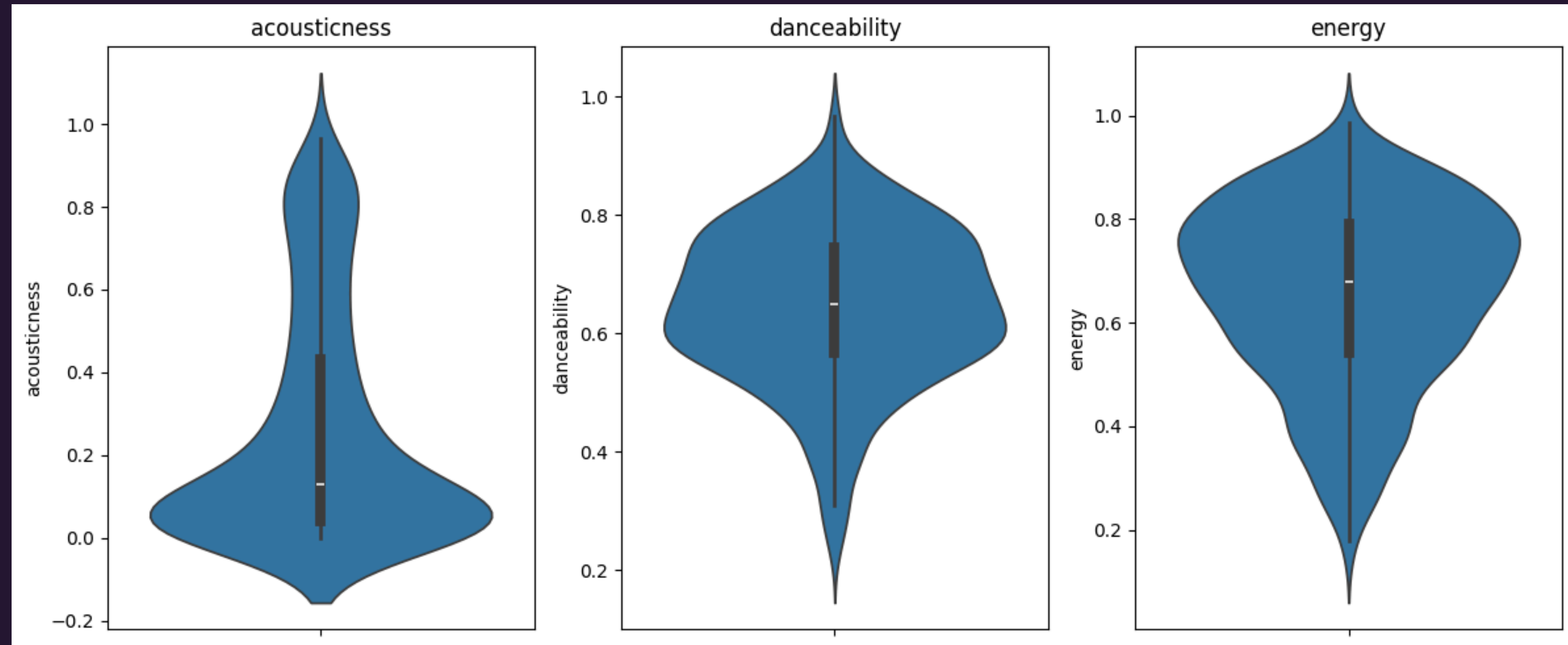
By plotting most popular songs of 2020's we get this type of violon plots :

- In most popular songs of 2020's (till 2024) very less amount of song have high acoustic score and also most of the songs have very low acoustic score.
- Also in 2nd violin plot we can see that most of the songs have high and moderate danceability score. As upper half(above median) has more volumn that lower half(lower median)
- In 3rd plot music in the dataset generally has high energy. The violin is widest in the upper range, indicating that a significant portion of the music has energy scores between 0.6 and 0.9. The distribution is much more even than acousticness, meaning the dataset contains a healthy mix of both lower-energy and higher-energy tracks, but the lean is towards the high end.

Music Preferences of 2020's
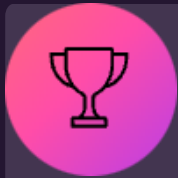
- So we can say that songs which are getting popular most of them have low acoustics , high to moderate danceability and are highly energetic.

- So for popularity one artist may focus on energetic ,less acoustic , danceable songs which are getting popular among people.
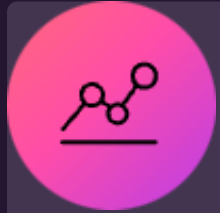
Most popular artists in recent decades in several languages

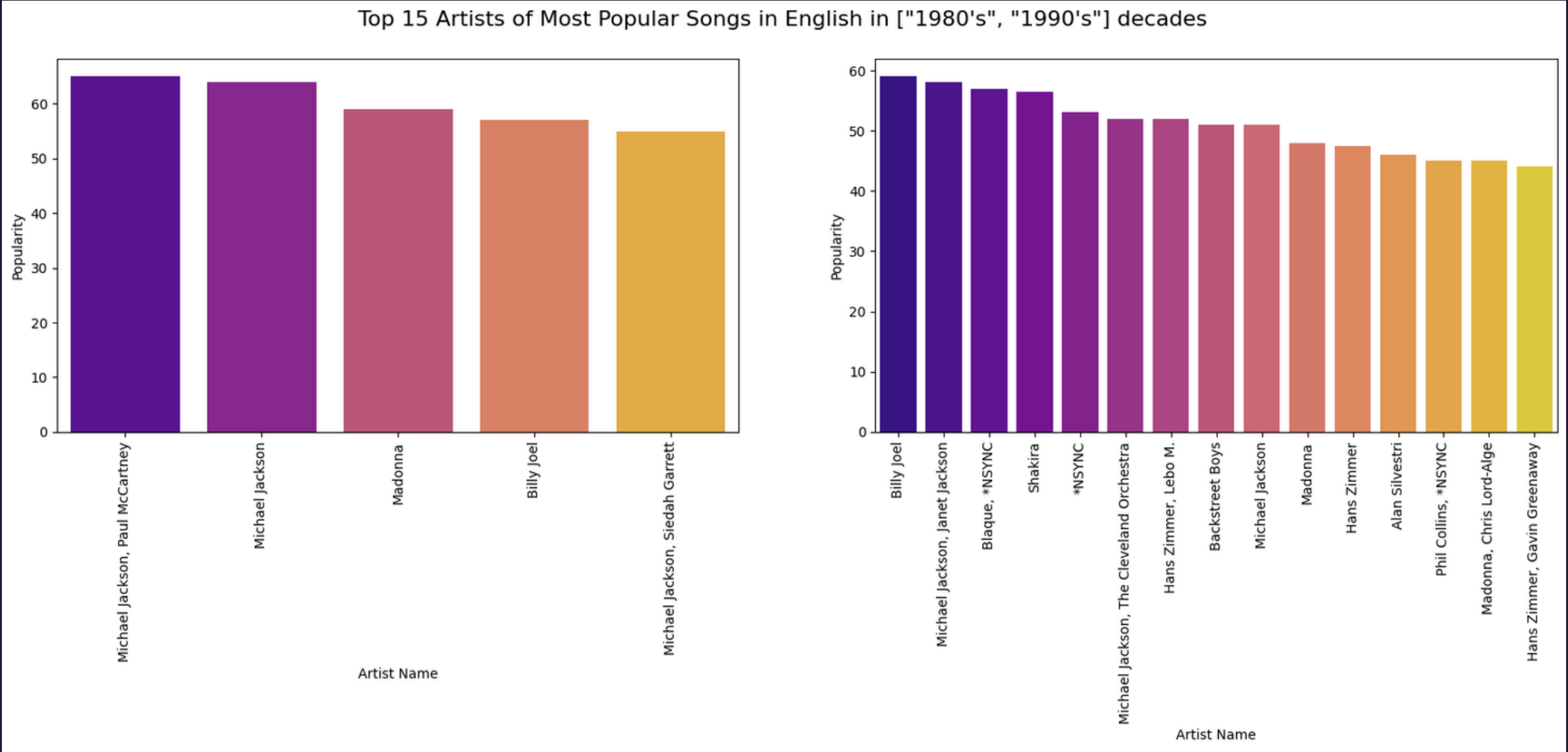# Popular Artist in English Language 1980's and 1990's

Overlap in Top Artists:
- **Michael Jackson (both solo and in collaborations) appears multiple times and is highly ranked on both charts, underscoring his lasting impact across both decades.**
- **Madonna also appears in the top tier of both charts, confirming her status as one of the most popular artists of the era.**

Insights from the 1980's :
- **Close Competition at the Top: The top two entries, Michael Jackson, Paul McCartney and Michael Jackson (solo), have virtually identical, highest popularity scores (both over 60), highlighting their immense and nearly equal popularity in the aggregated timeframe.**

Top 15 Artists of Most Popular Songs in English in ["1980's", "1990's"] decades

Insights from the 1990's :
- **Dominance of Early-to-Mid 90s Artists: Artists like Billy Joel, Michael Jackson, Janet Jackson, Shakira, and Blaque are highly ranked.**
- **Boy Band and Pop Presence: The inclusion of *NSYNC (multiple times, once solo and once with Blaque), and Backstreet Boys confirms the significant popularity of pop music and boy bands in the 1990s.**
- **Soundtrack/Classical Influence: The appearance of The Cleveland Orchestra, Lebo M, and multiple entries for Hans Zimmer points to a high popularity for songs associated with soundtracks (likely The Lion King given Lebo M's inclusion and the time frame). This indicates that non-traditional pop artists, particularly those from film music, secured top spots.**

Popular Artist in English Language 2010's and 2020's

Insights from the "2010's" Decades Chart (Left):

Electronic/Dance-Pop Focus:: The presence of The Chainsmokers (topping the chart), Calvin Harris, and Daft Punk highlights the significant popularity of Electronic Dance Music (EDM) and dance-pop in the 2010s.

Strong Collaborations: Many top entries are collaborations, showing the importance of joint projects in achieving peak popularity:

- **The Chainsmokers, Halsey**
- **Eminem, Rihanna**
- **Rihanna, Calvin Harris**
- **Justin Bieber, Ludacris**
- **The Weeknd, Daft Punk**
- **Rihanna, Kanye West, Paul McCartney**

Top 15 Artists of Most Popular Songs in English in ["2010's", "2020's"] decades
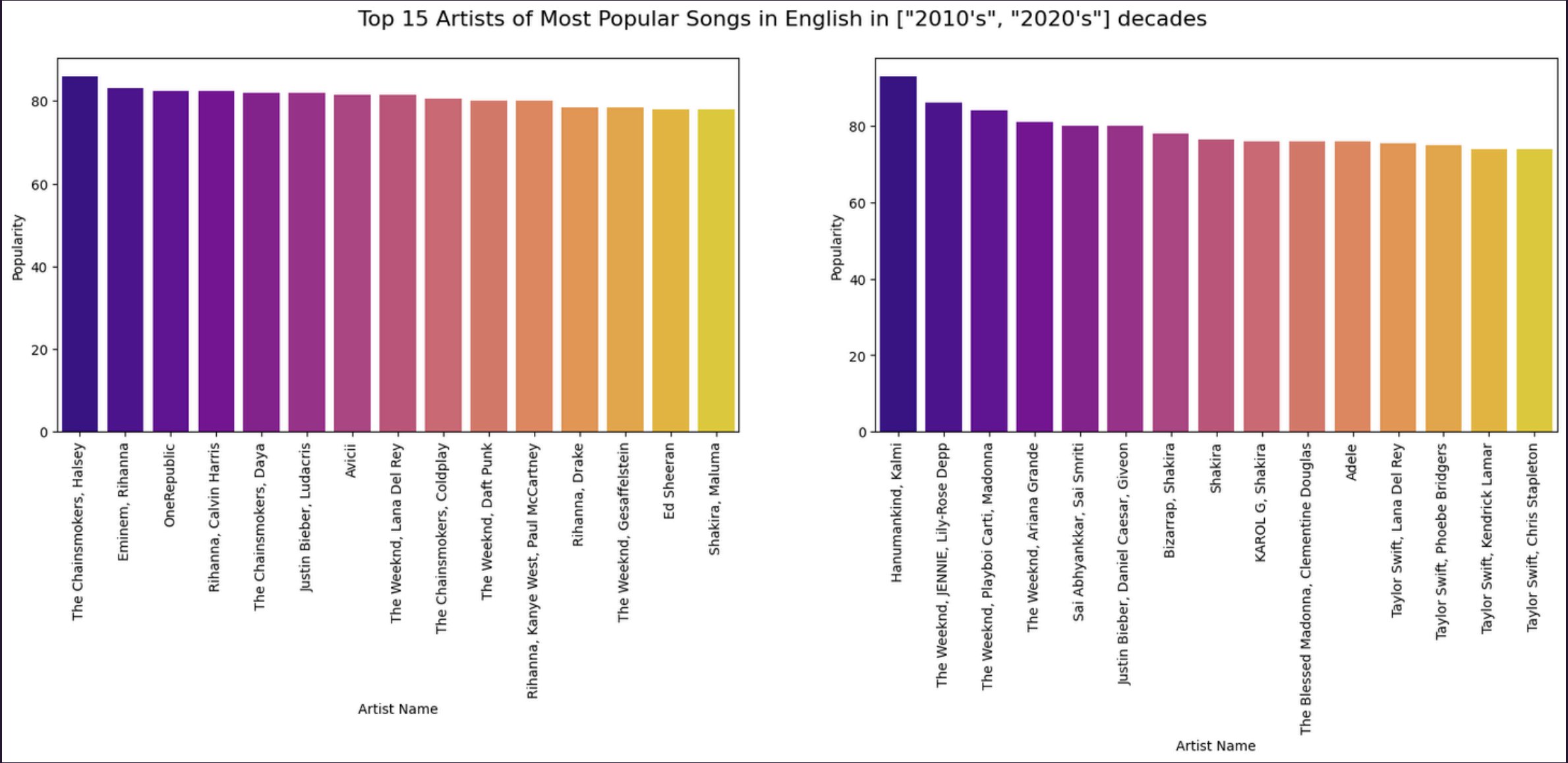
Insights from the "2020's" Decades Chart (Right):

New Global and Diverse Artists: The chart features artists like HanumanKind, Kalmi, Sai Abhyankar, and Sal Smriti, suggesting a greater influence of global, possibly non-traditional English-language markets, or new digital-first artists achieving high popularity.

- **K-Pop and Global Pop Influence:** The appearance of JENNIE (with The Weeknd and Lily-Rose Depp) and KAROL G (with Shakira) indicates the growing and mainstream success of artists from K-Pop and Latin Pop scenes in English-language music popularity.
- **Continuation of Pop/R&B:** Artists like Ariana Grande, Giveon, and Shakira (who also appeared in the 90s chart) maintain strong relevance in the 2020s.

Some Insights :
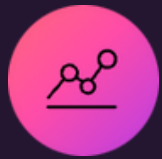
Comparison of the Decades (2010s vs. 2020s) Shifting Dominance:
- **While The Chainsmokers and Rihanna dominate the top of the combined "2010's" chart, the "2020's" chart is led by HanumanKind and Kalmi, followed closely by The Weeknd (with JENNIE and Lily-Rose Depp).**
- **The Weeknd's Longevity: The Weeknd is a dominant artist across both charts, appearing in multiple entries in the "2010's" chart (solo, with Lana Del Rey, and with Daft Punk) and even more prominently in the "2020's" chart (in three of the top ten entries). This demonstrates his sustained popularity across the two decades.**
- **Rihanna's 2010s Success: Rihanna is highly ranked in the "2010's" chart, appearing three times in the top 15 (solo, with Eminem, and with Calvin Harris), but her presence is notably absent from the very top of the "2020's" chart.**
- **Taylor Swift's Ascendance (or Late 2020s Success): Taylor Swift appears multiple times at the lower end of the "2020's" top 15 (solo, with Phoebe Bridgers, and with Kendrick Lamar), indicating a strong wave of popularity in the early-to-mid 2020s, despite not reaching the absolute top position. She is absent from the "2010's" top 15 shown.**
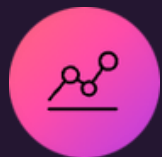
Some Observatios : In 1990's and 1980's the median popularity score of artists are nearly < 60 or slightly high . But in 2010's and 2020's in recent decades median popularity score of artists are nearly 80 or > 80 . This also indicates:
- **Internet accessibility and influence of social media platfroms causes songs of different to be reached between large number of people.**
- **Also increase in populatin can be a reason.**

Popular Artist in English Language 2010's and 2020's

Dominant Trends Across Both Decades (2010s & 2020s):

1. **Pritam's Undisputed Reign: Pritam is the most dominant figure across both decades. He appears in the number one or number two position in both charts and is featured in a vast majority of the top 15 entries for both the "2010s" (left chart) and the "2020s" (right chart). This indicates his status as the most popular and prolific composer/artist in Hindi film music for this entire period.**

2. **Arijit Singh's Consistency: Arijit Singh is the most prominent male vocalist. He appears multiple times in both charts, often with Pritam (e.g., "Pritam, Arijit Singh"), confirming his position as the leading voice of the era.**

3. **The Importance of Duos and Collaborations: The charts overwhelmingly feature collaborations, often listing the composer first, followed by vocalists or other collaborators. For example:**

- **Pritam, Tanishk Bagchi**
- **Sachin-Jigar, Shreya Ghoshal, Sachin-Jigar**
- **Vishal-Shekhar, Shilpa Rao, Caralisa Monteiro, Vishal Dadlani**

Top 15 Artists of Most Popular Songs in Hindi in ["2010's", "2020's"] decades
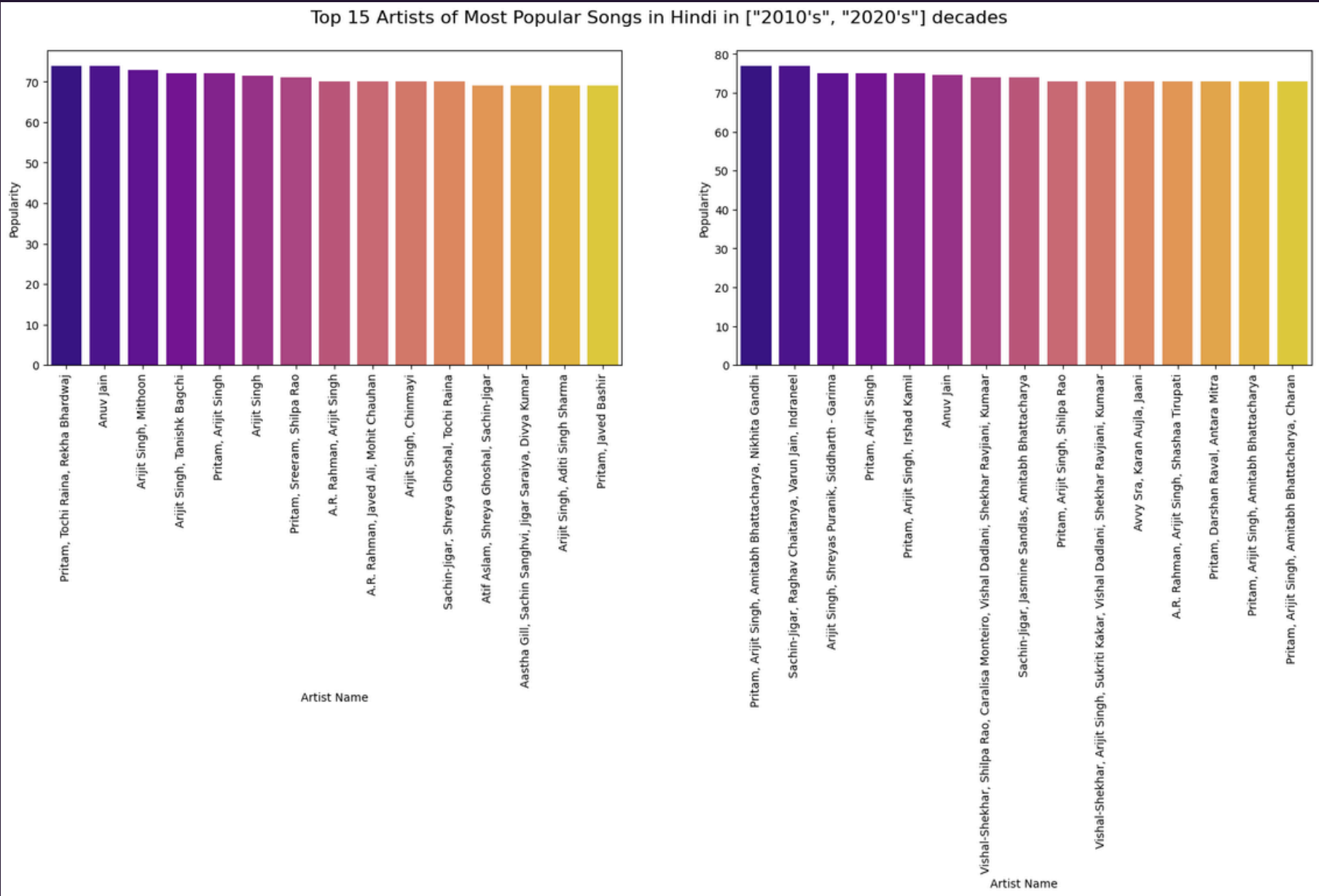
Observation:

- **This structure confirms that modern Hindi music popularity is driven by composer-vocalist-lyricist teams, rather than single solo artists.**
- **Composer-Singer Pairs: The frequent pairing of Pritam with Arijit Singh and Sachin-Jigar with various vocalists suggests powerful and successful hit-making partnerships.**
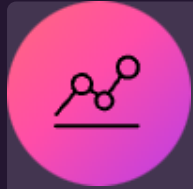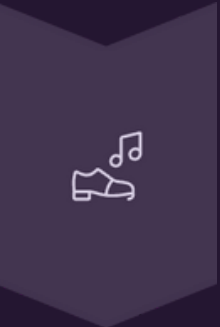
valence and characteristics

Valence Analysis :

Cheracteristic Of Valency in dataset

- 25% = 0.292 → 25% of tracks are rather low valence (sad/serious).
- 50% = 0.507 → Median is neutral, slightly positive.
- 75% = 0.710 → 75% of songs are at least moderately happy.
- max = 0.995 → The happiest songs have near-maximum positivity.

Valency

Relatively uniform or spread out, with a slight tendency to peak near the moderate to high valence range (around 0.6 to 0.9).
here is a diverse range of emotional tones in the dataset, though slightly more tracks tend to have a positive (higher valence) feel than a negative (lower valence) feel.
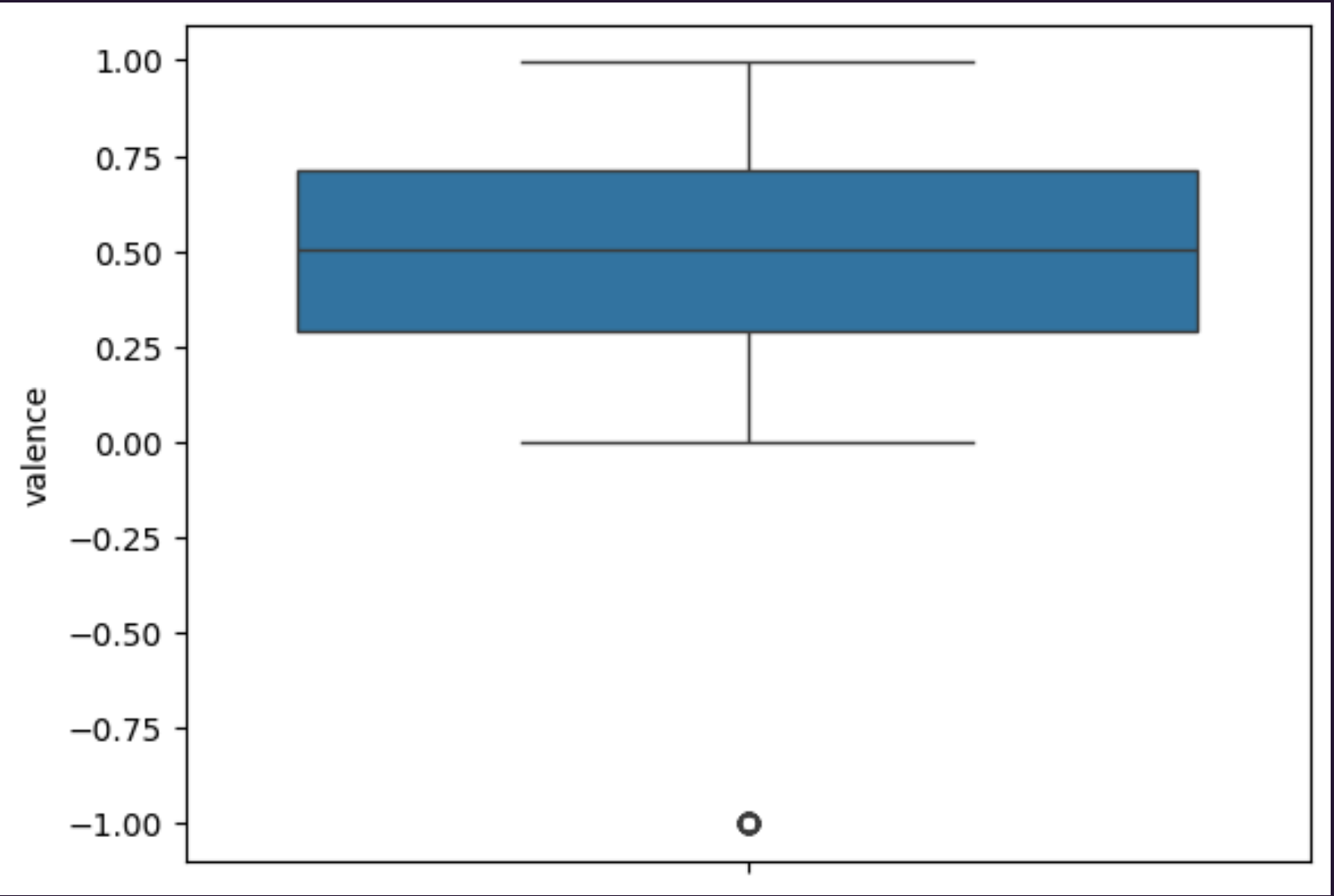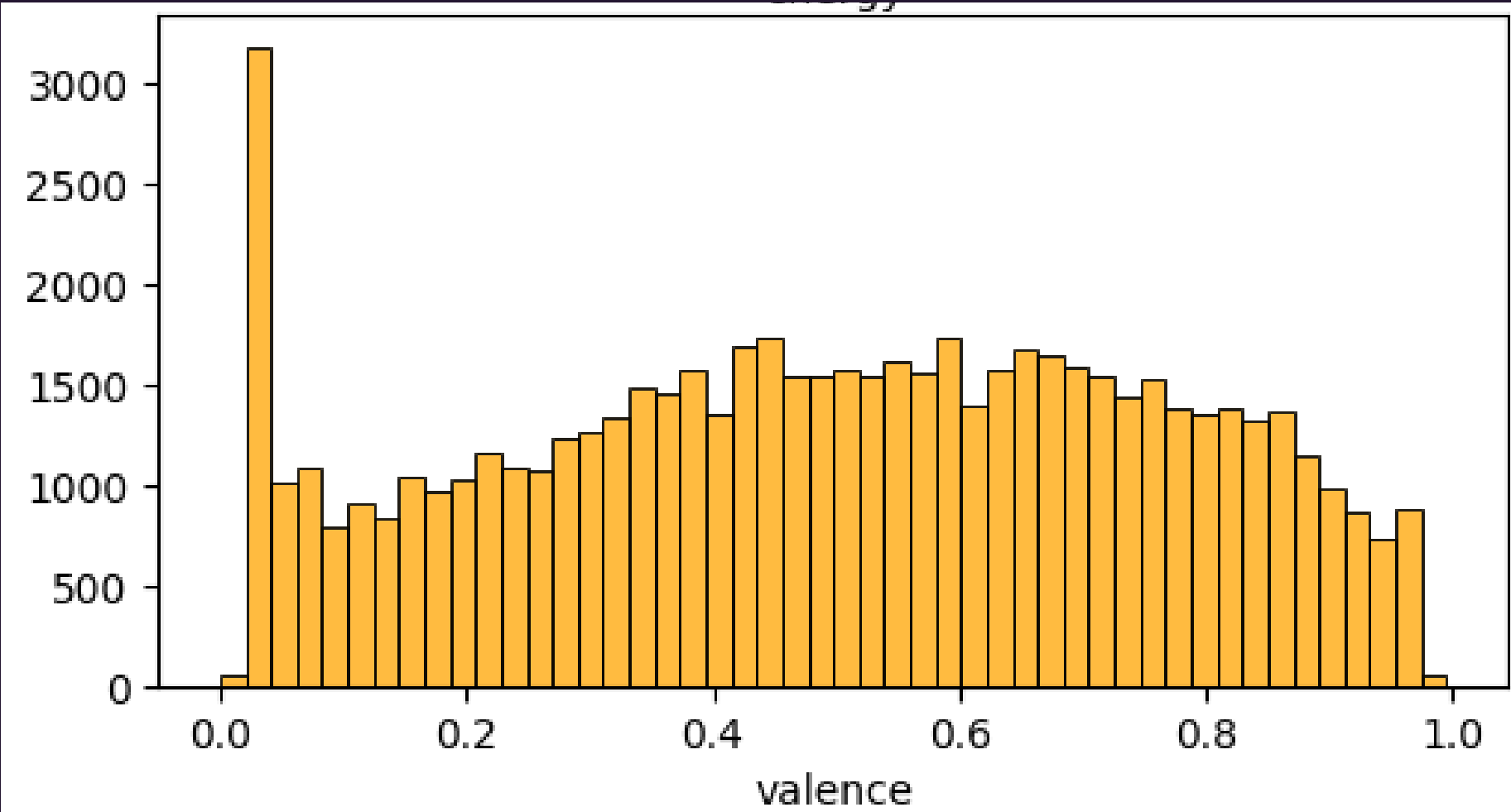
Valence Analysis :
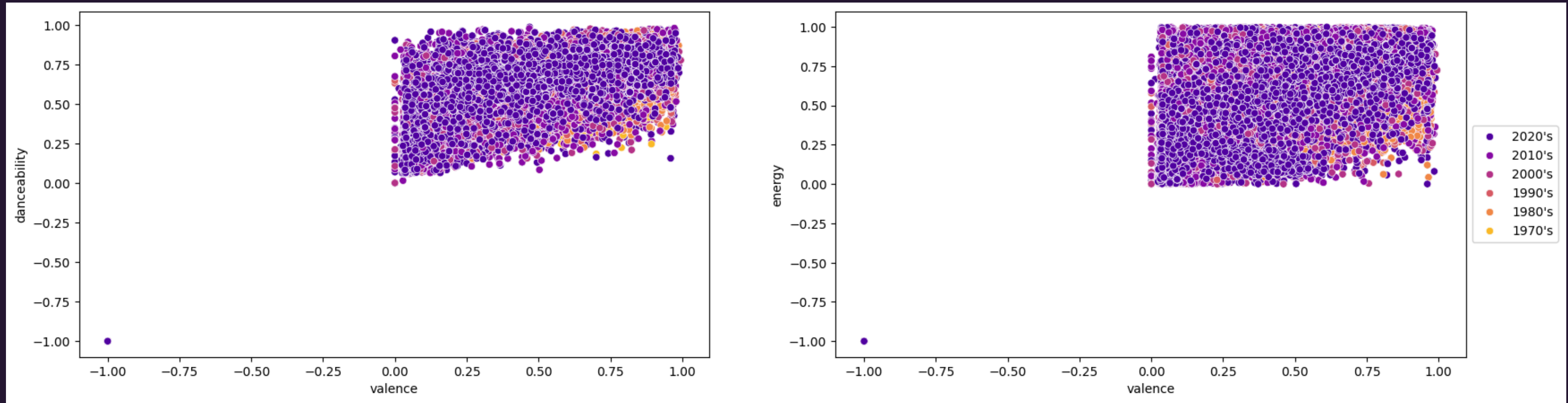
Valence Dependendies on Danceability and Energy :

- slightly upward trend is seen in danceability with valence score of the song . Not a strong correletion is observed but positive correlation is observed .
- In case of energy no trend is observed songs can be sad with high energy but at high valence score more songs cluster above 0.25 energy score .
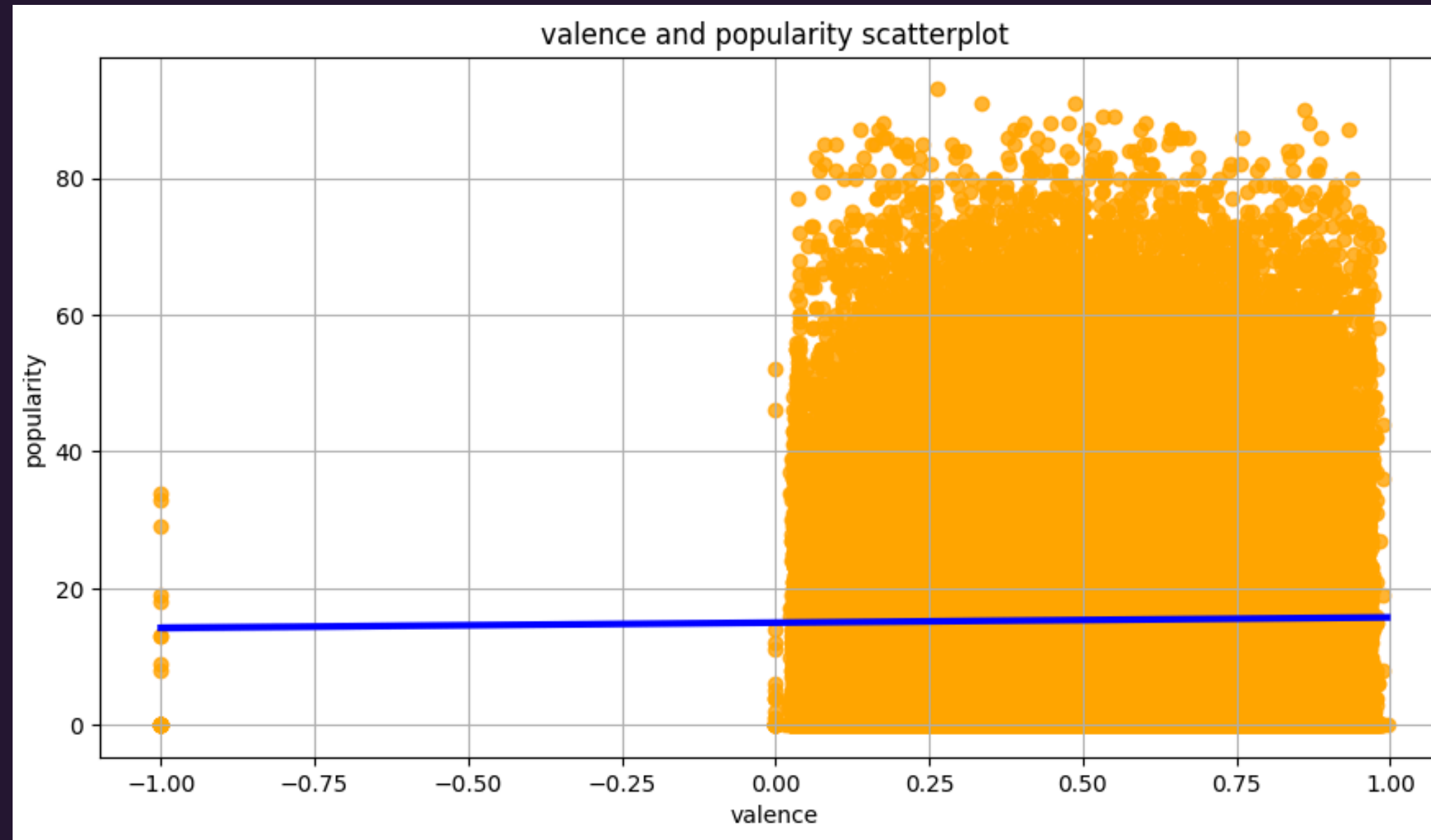
Valence Analysis :
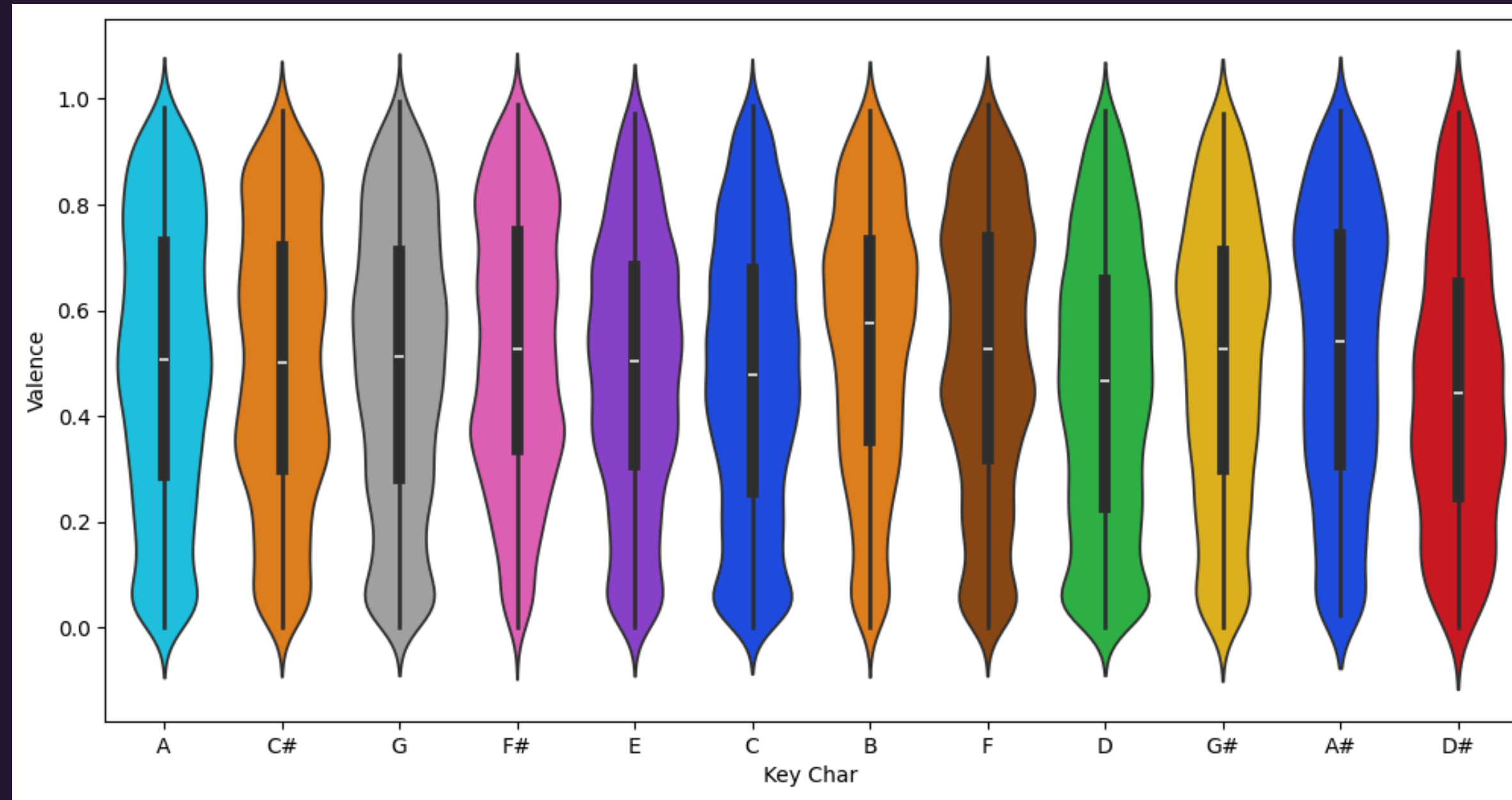
Valence Dependendies on keys :

- the data shows that Valence is highly variable across all musical keys. While the average valence is fairly consistent across keys, the distribution and range of valence are so broad that one cannot definitively conclude that any single key is significantly "happier" or "sadder" than the others based on this plot.
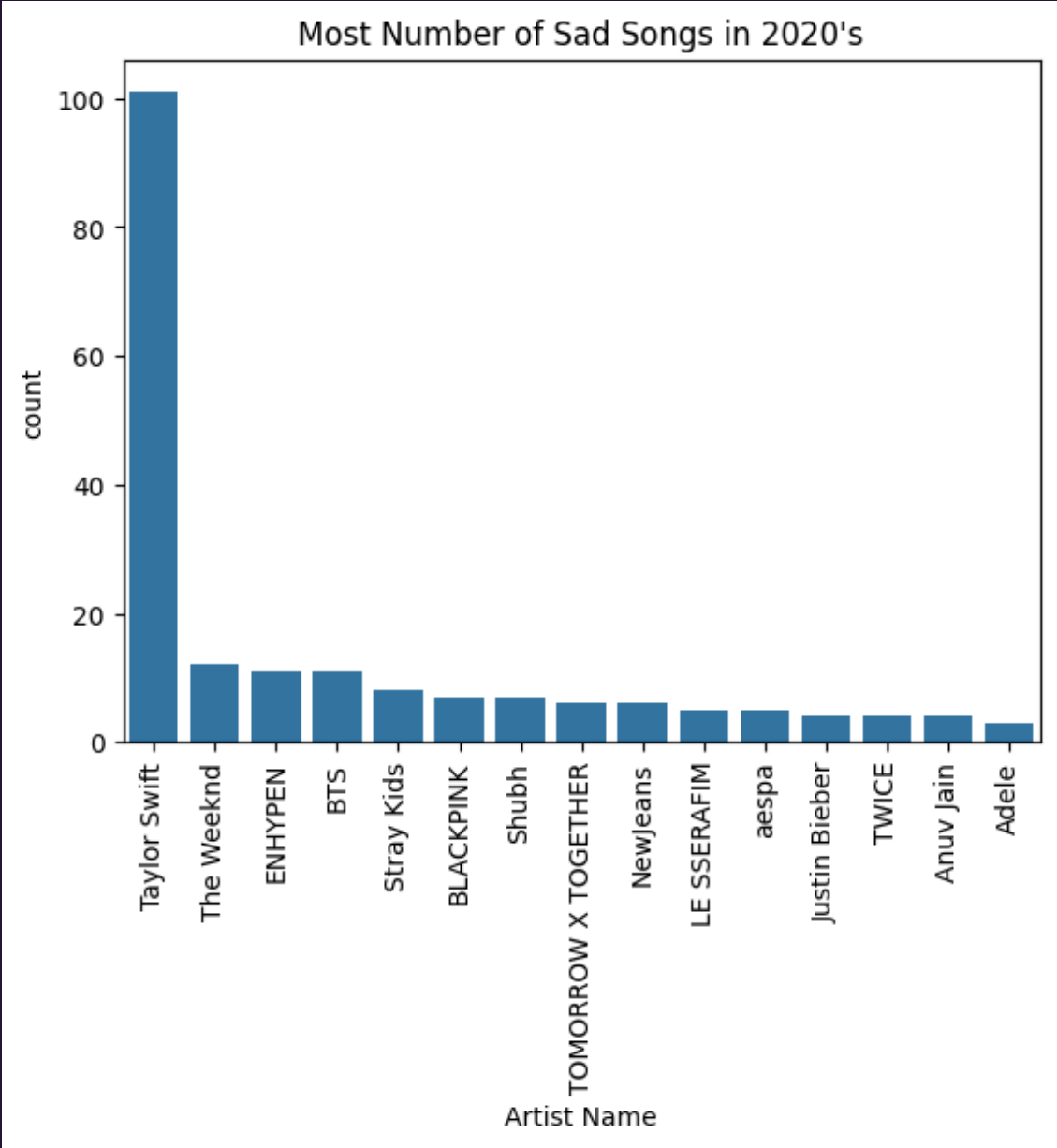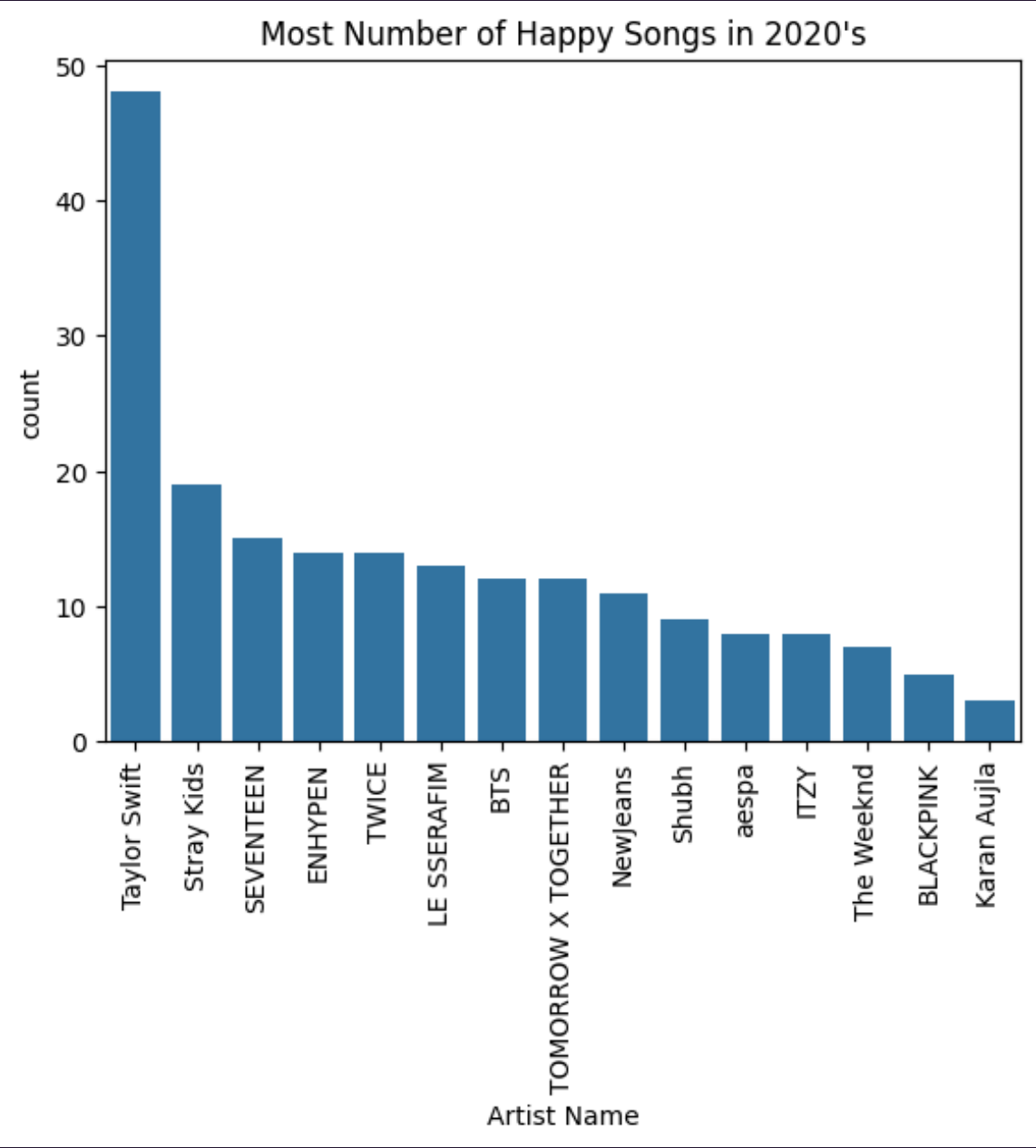
Valence Analysis :

valence and popularity scatterplot

Valence Dependendies on Popularity :

- **Not a perfect correlation is observed songs with high valence as well as low valence are getting same popularity**

Valence Analysis :

Most Number of Happy Songs in 2020's

Most Number of Sad Songs in 2020's

Conclusion:
 The data clearly indicates that Taylor Swift is the most prolific artist for both happy and sad songs among this selection of artists in the 2020s, with a significantly higher volume of sad songs. Beyond her, the distribution suggests:
- There is a greater overall volume of sad songs compared to happy songs produced by the top-ranked artists.
- K-Pop acts tend to dominate the top ranks of the Happy Songs chart, suggesting their output is often classified with a higher valence.
- Artists like The Weeknd and Adele appear to specialize more in sad/low-valence music, as they are ranked high or exclusively on the Sad Songs chart.

Thank You!