**🎵 Lyrics Sentiment**

**V/S**

**Song Popularity**📈

**Over Decades**

**Project Report**

**By Varun Kumar  
(Data Analyst)**

**28th June 2025**

**Contents**

**1.** Introduction

**2.** Dataset Description and Data Collection Process

**3.** Methodology & Tools

**4.** Analysis & Results  
 4.1 Sentiment vs Popularity Correlation  
 4.2 Average Sentiment by Release Year  
 4.3 Popularity Trend Over Time  
 4.4 Which Artists Have the Most Positive/Negative Songs  
 4.5 Word Cloud & Text-based Insights

**5.** Conclusion

**6.** References

**Introduction**

In today’s digital era, music streaming platforms provide listeners access to millions of songs across various genres and languages. While beats, rhythms, and melodies play a crucial role in a song’s appeal, **the emotions conveyed through lyrics often determine how deeply a track resonates with its audience**. As a data analytics enthusiast, I was curious to explore whether the emotions in song lyrics — whether positive, negative, or neutral — have any measurable relationship with a song’s popularity.

**Sentiment analysis** in music refers to the use of Natural Language Processing (NLP) techniques to computationally evaluate the emotional tone of song lyrics. By assigning sentiment scores to lyrics, we can classify them as positive, negative, or neutral and then study how these sentiments affect listener preferences and popularity trends over time.

The goal of this project was to perform a **comprehensive analysis of lyrical sentiment and its relationship with song popularity trends over decades**. I wanted to identify patterns such as whether happier songs are more popular, how the emotional tone of lyrics has evolved over the years, and which artists consistently produce positive or negative songs.

To answer these, I structured my project around **five core research questions**:

* How does sentiment correlate with song popularity?
* How have average sentiment scores varied over the years?
* Are newer songs generally more popular than older classics?
* Which artists produce the most positive or negative songs?
* What insights can be drawn from word frequency patterns in song lyrics?

This self-driven project not only helped me apply my Python, data analysis, and NLP skills but also offered fascinating insights into the evolving emotional landscape of popular music.

**Dataset Description &**

**Data Collection Process**

For this project, I created a custom dataset of 30 popular English songs by combining data from the Spotify Web API and Genius Lyrics API. The objective was to collect each song’s title, artist, popularity score, release date, and full lyrics. This consolidated dataset served as the foundation for my sentiment analysis and popularity trend study.

The final dataset consisted of 30 records (rows) and 6 columns:

* Song Title
* Artist
* Popularity
* Release Date
* Lyrics URL
* Lyrics

**API Extraction Logic:**

I built a base Python script to automate data extraction:

* First, I authenticated using my Spotify API client ID and secret key to generate an access token.
* Then, for each of the 30 song titles, I fetched details like artist name, popularity score, and release date from Spotify’s API.
* Parallelly, I used the Genius Lyrics API to retrieve the corresponding lyrics page URL for each song.
* Finally, I scraped the lyrics content directly from the Genius webpage using BeautifulSoup and stored it in my dataset.
* This entire extraction process was run inside a for loop iterating over a list of 30 selected songs, with a short delay between requests to avoid hitting API limits.

Here’s the exact code structure I used for extraction:

python

def get\_spotify\_data(song\_title, token):

...

def get\_lyrics\_url(song\_title, genius\_token):

...

def scrape\_lyrics(url):

Once all data was fetched, it was consolidated into a CSV file named raw\_song\_dataset.csv.

**Data Processing & Cleaning:**

After collecting the raw data, I processed it using pandas:

* Removed any missing or incomplete records (though in my case, all 30 records were retained).
* Converted release dates to a consistent datetime format for analysis.
* Calculated sentiment polarity scores for each song’s lyrics using the TextBlob library.
* Created a new column called Sentiment Polarity in the dataset to store these scores.

Finally, I saved the cleaned and sentiment-enhanced dataset as final\_song\_dataset\_with\_sentiment.csv.  
This processed file was then used for all further analysis across the five Python scripts in this project.

**Methodology & Tools**

To perform this project efficiently, I divided the entire workflow into **data extraction**, **data processing**, **sentiment analysis**, and **insight generation** phases. I modularized my analysis using **five Python scripts**, each dedicated to a specific set of research questions, ensuring clarity and easy debugging.

**Libraries & Tools Used:**

I used several Python libraries for data manipulation, visualization, natural language processing, and API interaction:

* **Pandas —** for data manipulation, cleaning, and CSV handling.
* **Requests —** to fetch data from the Spotify Web API and Genius API.
* **BeautifulSoup —** for scraping lyrics content from Genius lyrics webpages.
* **TextBlob —** to perform sentiment analysis and assign polarity scores to lyrics.
* **Seaborn & Matplotlib —** for creating various visualizations like bar plots, line plots, and violin plots.
* **WordCloud —** to visualize the most frequently occurring words in positive and negative song lyrics.

**Sentiment Polarity Calculation:**

To determine the emotional tone of each song, I used TextBlob, a Python library for simple natural language processing tasks.  
I passed the lyrics of each song to the TextBlob sentiment.polarity function, which assigned a polarity score between -1.0 (extremely negative) and +1.0 (extremely positive).

This sentiment polarity value was then stored in a new column named Sentiment Polarity within the dataset final\_song\_dataset\_with\_sentiment.csv.

**Correlation Calculation:**

To analyse the relationship between lyrical sentiment and song popularity, I computed the Pearson correlation coefficient between the Sentiment Polarity and Popularity columns using pandas.corr() function.  
This helped identify whether higher positivity or negativity in lyrics had any notable effect on a song's popularity score.

**Types of Visualizations Made:**

I used a variety of plots to effectively represent the trends and patterns in the dataset:

* **Bar Plots —** to compare average popularity across sentiment categories and artist-wise sentiment distribution.
* **Line Plots —** to show trends in average sentiment and song popularity over the years.
* **Violin/Box Plots —** to visualize the distribution of sentiment scores among different artists.
* **Word Clouds —** to highlight the most frequent words in positive and negative songs.
* **Histograms —** to display the overall sentiment score distribution across all songs.

These visualizations provided a clear and interactive way to interpret the insights derived from the dataset.

**Threshold Values & Categorization Rules:**

For categorizing songs based on sentiment polarity, I applied the following threshold values:

* **Positive:** if Sentiment Polarity > 0.2
* **Negative:** if Sentiment Polarity < -0.2
* **Neutral:** if Sentiment Polarity lies between -0.2 and 0.2

This rule-based classification helped simplify analysis like calculating average popularity for each sentiment category and comparing sentiment trends over time.

**Analysis & Results**

**4.1 Sentiment vs Popularity Correlation:**

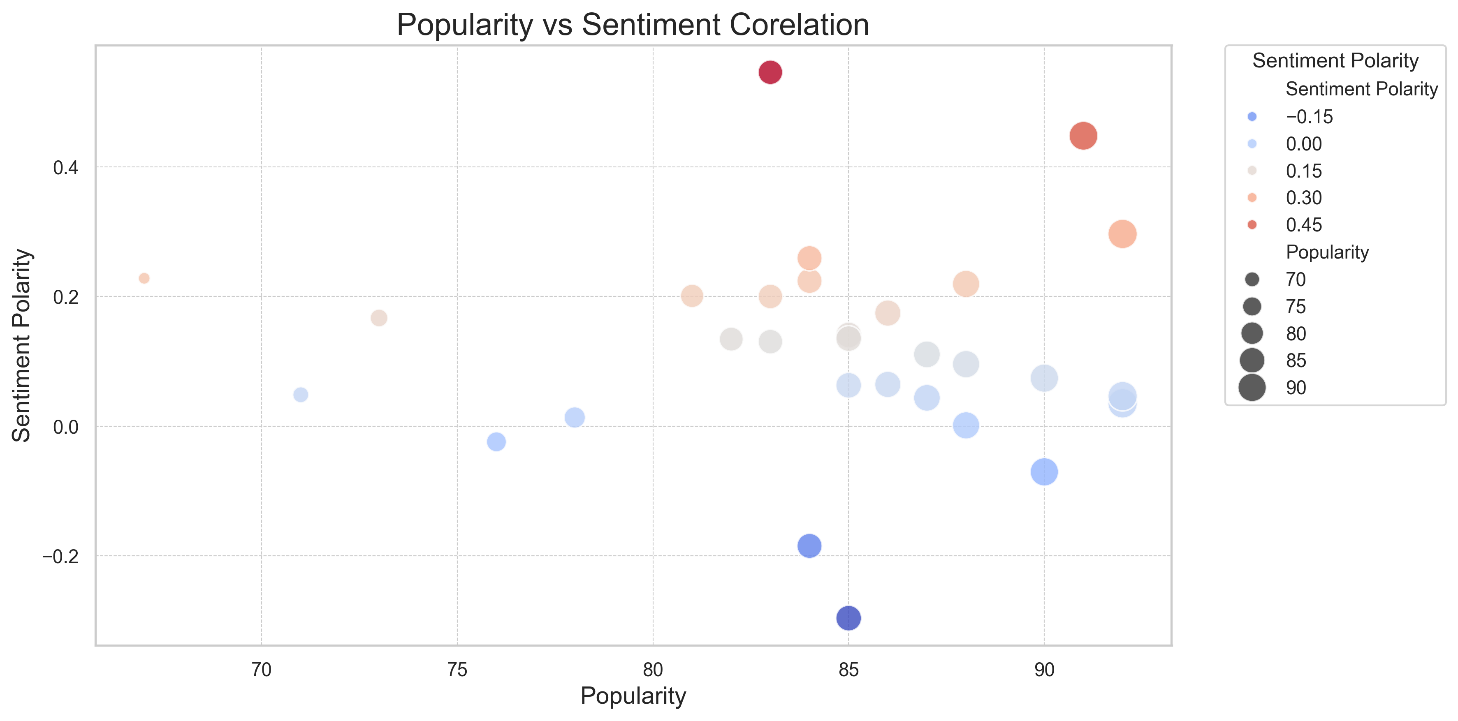
In this part of the project, I examined whether the **emotional tone of a song’s lyrics** influences its popularity score on Spotify. Several mini-analyses were performed to understand different angles of this relationship:

**a) 📊 Correlation Value**

I calculated the Pearson correlation coefficient between Sentiment Polarity and Popularity.

**Result:**  
The correlation value was **-0.0087**, indicating virtually no relationship between the sentiment of lyrics and a song’s popularity score. This suggests that popularity isn’t necessarily driven by how positive or negative a song's lyrics are.

*A scatter plot with a color gradient based on sentiment polarity was created to visualize this, showing a random, non-linear spread.*



**b) 📊 Average Popularity per Sentiment Category**

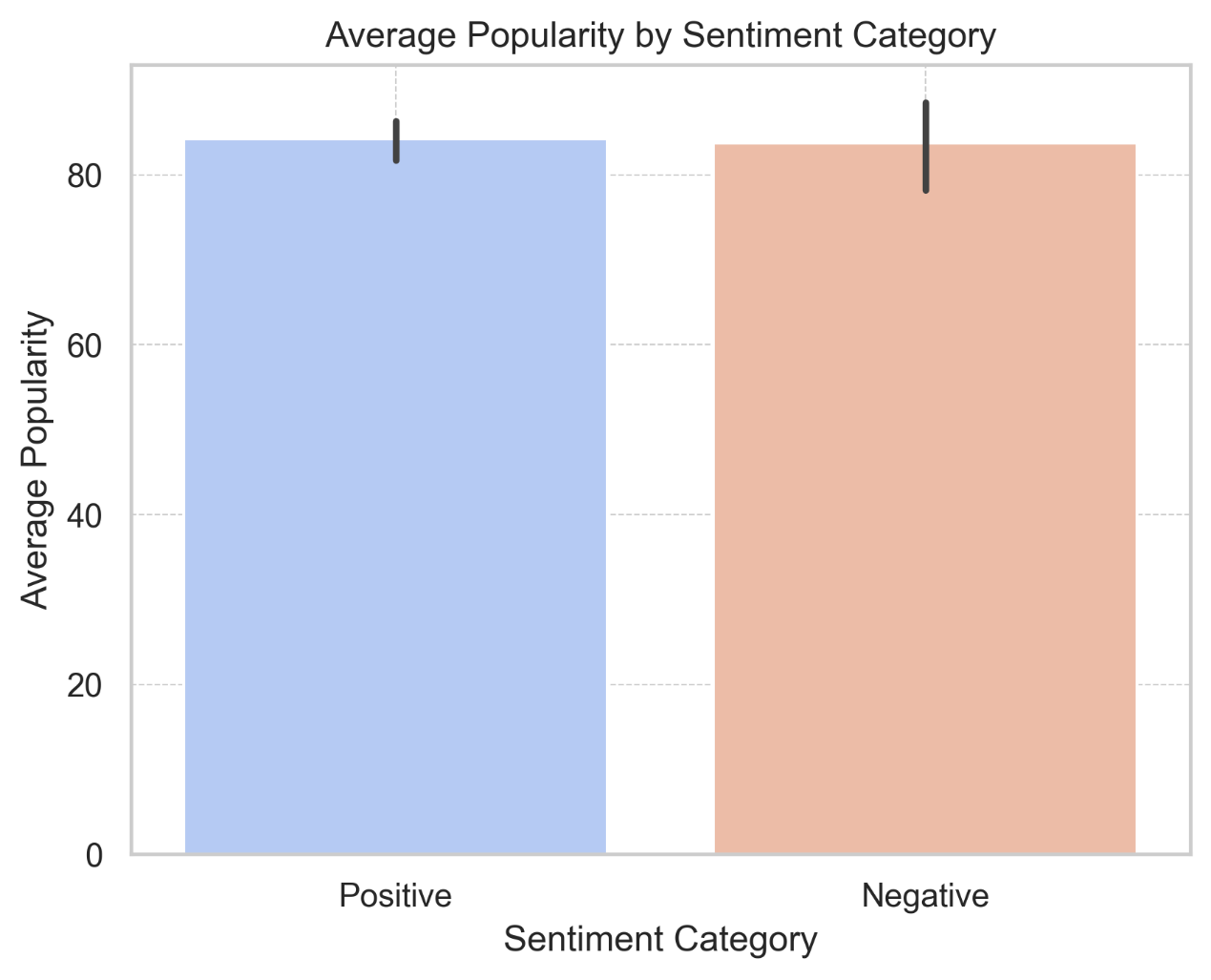
I categorized songs as:

* **Positive**: Sentiment Polarity > 0
* **Negative**: Sentiment Polarity < 0
* **Neutral**: Sentiment Polarity = 0

Then, I calculated the **average popularity for each category**.

**Result:**  
Positive songs had a slightly higher average popularity (**84.27**) compared to Negative songs (**83.75**) — but the difference was **negligible**.

A bar plot clearly reflected this marginal difference, indicating no major popularity bias based on sentiment category.

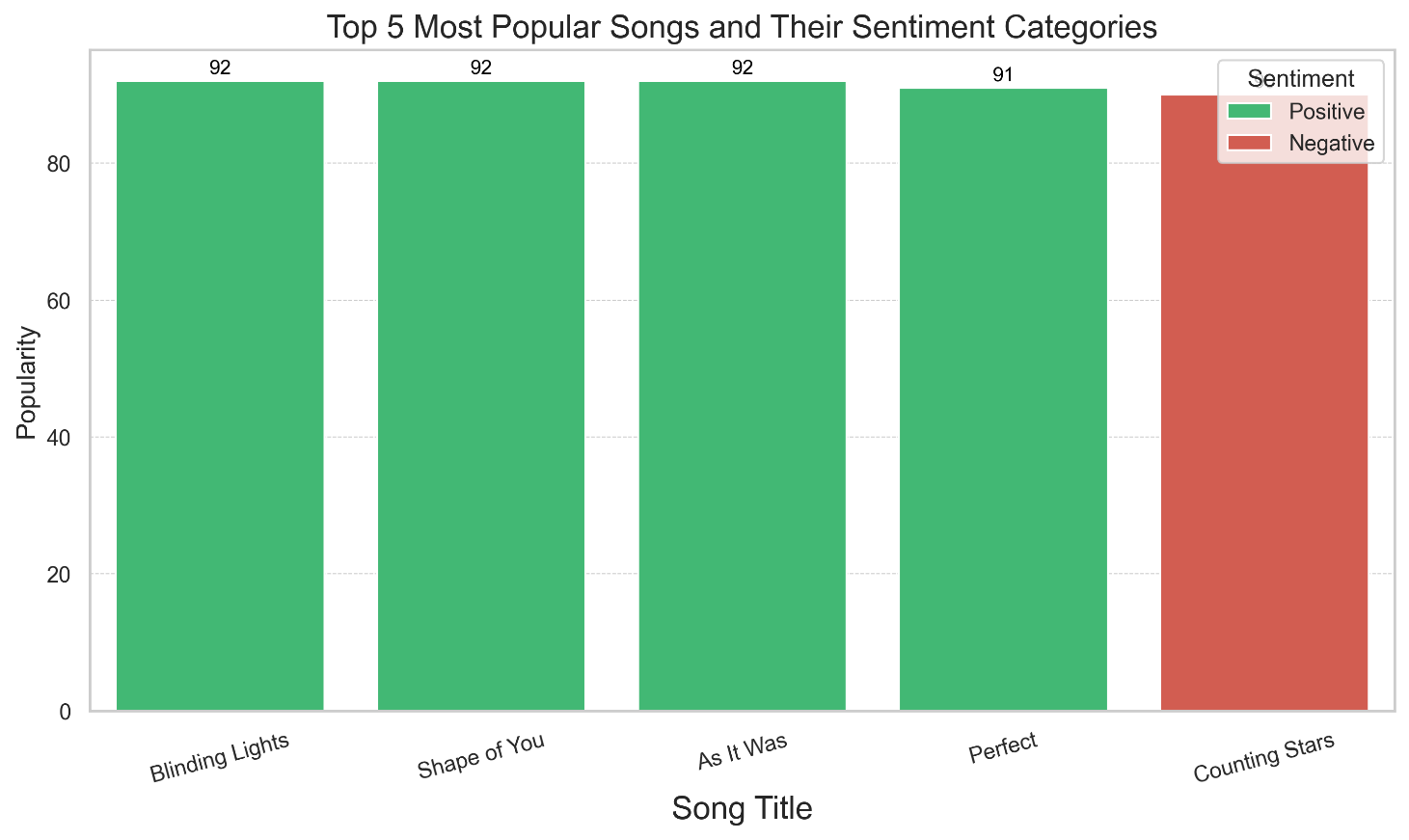


**c) 📊 Sentiment Profile of Top 5 Most Popular Songs**

I identified the **top 5 songs by popularity** and analyzed their sentiment categories.

**Result:**Among these, **4 were categorized as Positive** and **1 as Negative**.  
This indicates that while positive sentiment tends to appear more in popular songs, **popularity is not strictly sentiment-driven** — other factors likely play a stronger role.

A color-coded bar plot visualized the top 5 songs, showing both their popularity and sentiment category.



**d) 📊 Do Extreme Sentiments Affect Popularity?**

I filtered songs with **extreme sentiments** (Polarity > 0.4 or < -0.2) and compared their average popularity to the overall average.

**Result:**  
Songs with extreme sentiments had an average popularity of **86.33**, which was **slightly higher** than the overall average of **84.2**.  
Though based on a small sample of 3 songs, this suggests that **strongly emotional songs might attract marginally more attention**.

**e) 📊 Sentiment Polarity by Popularity Tiers**

I divided songs into **Low (≤75)**, **Medium (76-85)**, and **High (>85)** popularity tiers and calculated their average sentiment polarity.

**Result:**  
The average sentiment polarity across all tiers was **nearly identical**.  
This confirmed that **sentiment does not significantly influence whether a song falls into a high, medium, or low popularity group**.

*A table summarizing average sentiment by tier showed minimal variation, and hence no further visualization was necessary.*

**Output Table:**

|  |  |  |
| --- | --- | --- |
| **Popularity** | **Tier** | **Sentiment Polarity** |
| 0 | High | 0.118315 |
| 1 | Low | 0.147739 |
| 2 | Medium | 0.110069 |

**Final Insight for 4.1**

While lyrical sentiment adds emotional flavour to a track, **popularity on platforms like Spotify appears to be influenced more by other factors** such as production quality, artist fanbase, marketing, trends, and genre appeal. Sentiment alone doesn’t drive a song’s success.

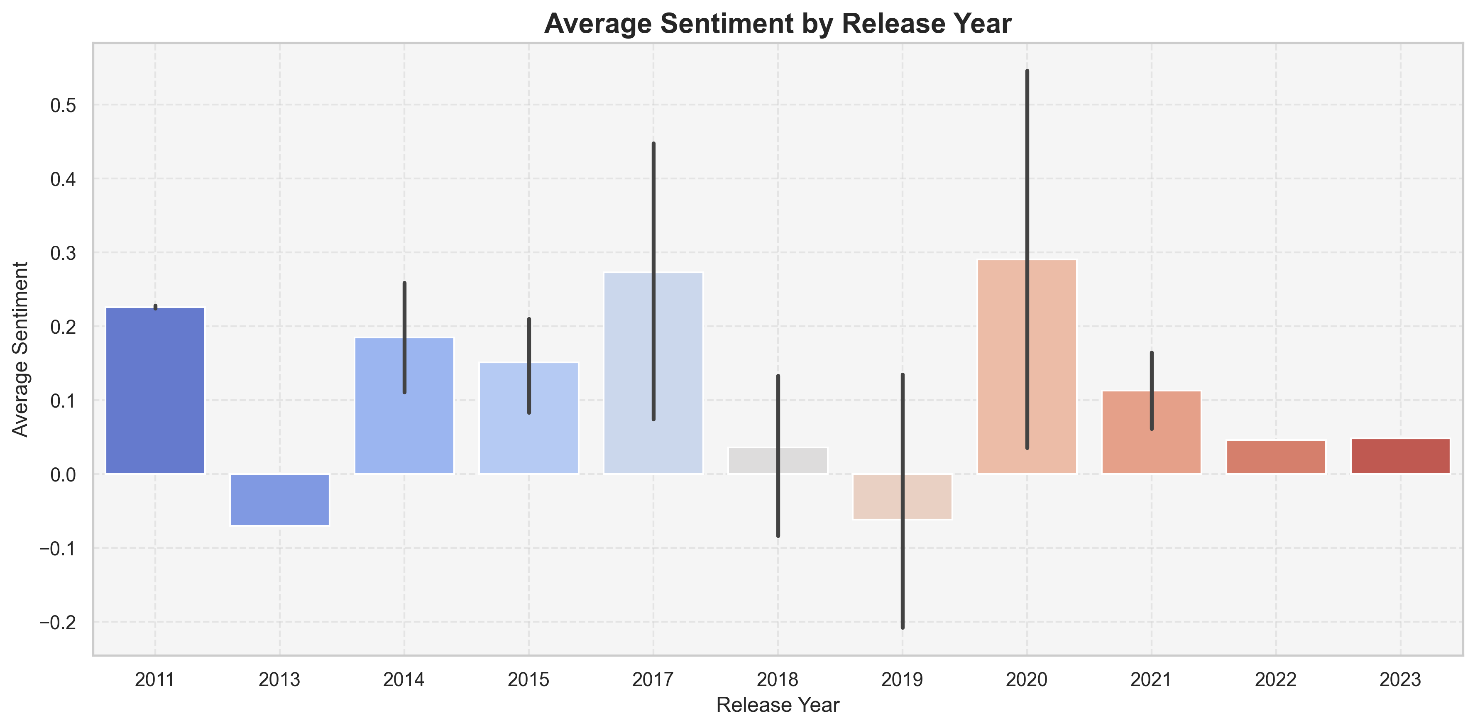
**4.2 Average Sentiment by Release Year:**

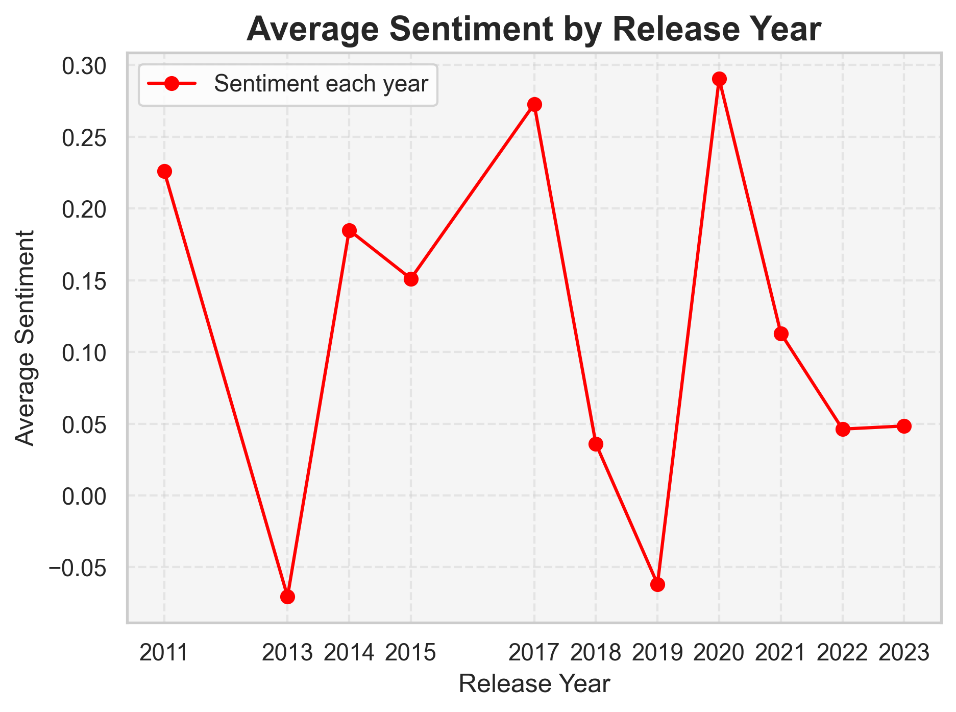
In this analysis, I examined how the **average sentiment polarity of song lyrics** has changed over different release years. The objective was to observe whether music has grown more positive, negative, or neutral over time, and whether any years stood out in terms of emotional tone.

**a) 📊 Sentiment Trend Over Time**

I calculated the **average sentiment polarity per release year** using a group-by operation.

**Result:**  
The **line plot and bar plot** indicated that lyrical sentiments have fluctuated over the years, with minor ups and downs but generally stayed within a neutral to mildly positive range.



****

**Question:** Are songs becoming sadder, happier, or more neutral over time?  
**Answer:** According to the analysis, **songs over time are gradually becoming more neutral in sentiment**. While certain years like **2020** saw a peak in positivity, the overall trend in recent years shows a **noticeable decline in average sentiment polarity**, moving closer to neutral values.

This suggests that while earlier years had slightly more emotionally positive songs, **modern music tends to maintain a more emotionally balanced or neutral lyrical tone**, with neither very positive nor very negative sentiments dominating.

**b) 📊 Years with Highest/Lowest Average Sentiment**

Using the grouped averages:

* 📈 **Highest average sentiment**: **2020**
* 📉 **Lowest average sentiment**: **2013**

**Insight:**  
In this dataset, songs released in **2020** were the most positive on average, while **2013** witnessed the least positive (or most neutral/negative) songs.

**c) 📊 Trend Stability/Volatility**

To measure trend stability, I computed the **absolute year-on-year changes in average sentiment** and then calculated their mean.

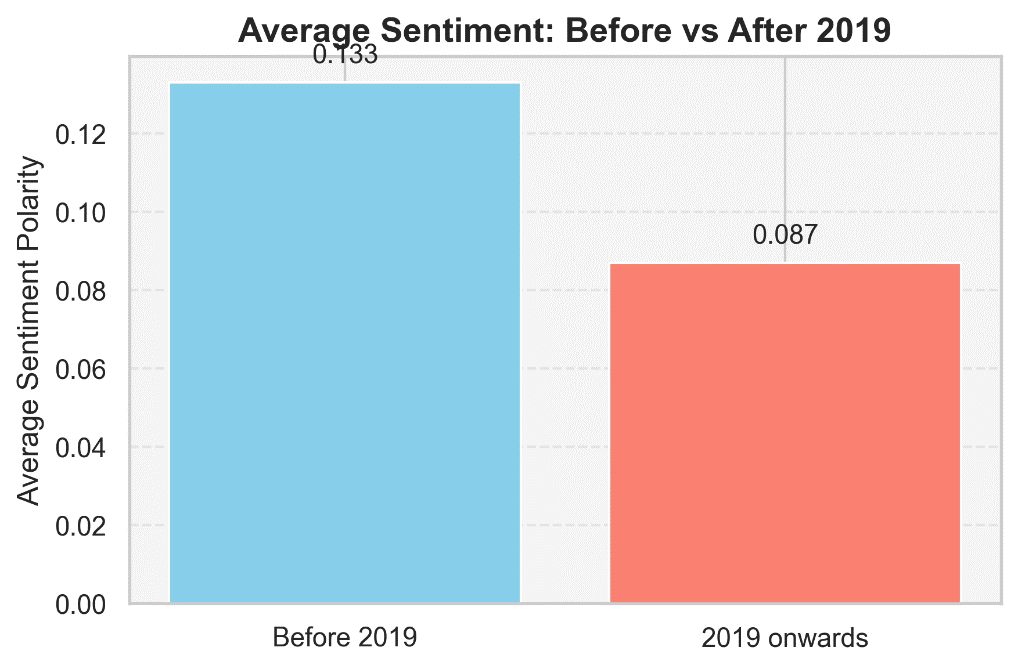
**Result:**  
The **average year-on-year change was 0.102**, which is considered relatively high for sentiment polarity (since the range is from -1 to +1).

**Insight:**  
This indicates a **moderately volatile trend** in average sentiment polarity across the years in this dataset. Certain years experienced noticeable emotional tone shifts.

**d) 📊 Difference Between Last 5 Years vs Earlier**

I compared the average sentiment:

* 📊 **Before 2019**: 0.133
* 📊 **2019 onwards**: 0.087

****

**Insight:**  
Songs released from **2019 onwards** have been **less positive on average** than those released before 2019.  
This suggests a possible shift towards more **emotionally balanced or slightly neutral/negative themes in recent mainstream songs**.

**Final Insight for 4.2**

The data shows that while there are fluctuations in lyrical sentiment over time, **popular music generally remains within a mild, balanced emotional range**.  
**2020 stood out as a notably positive year**, but the last 5 years hint at a soft decline in positivity.  
A larger dataset could help confirm whether this is an emerging industry trend or just a pattern within this sample.

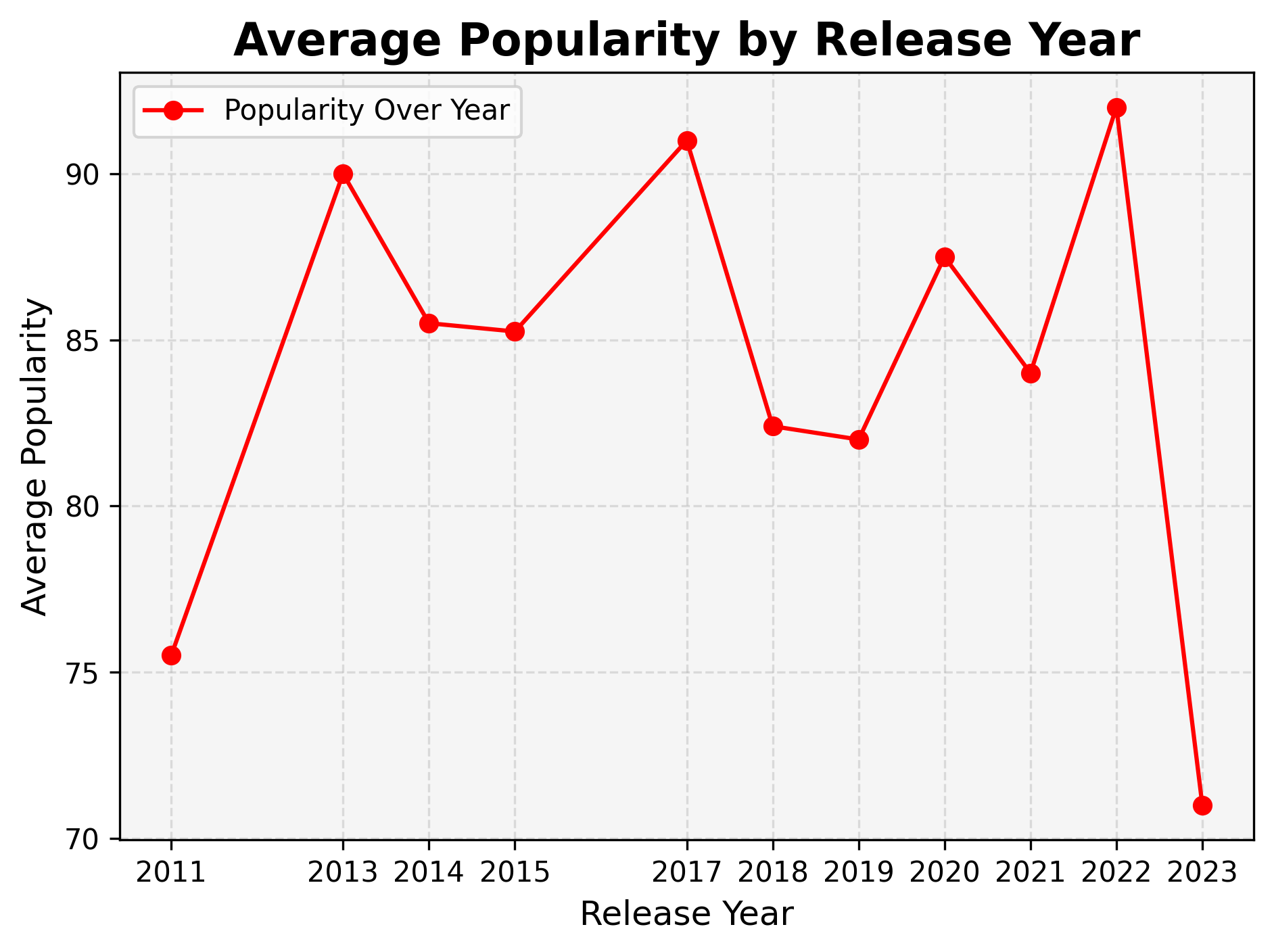
**4.3 Popularity Trend Over Time:**

This analysis aimed to understand how **the average popularity of songs has changed across different release years**. The focus was on observing whether **newer songs are generally more popular than older classics**, and how consistent that trend is over time.

**a) 📊 Are Newer Songs More Popular or Older Classics?**

I calculated the **average popularity for each release year**.

**Result:**  
The **line plot of average popularity by release year** clearly showed that **newer songs tend to be more popular than older classics** in this dataset.

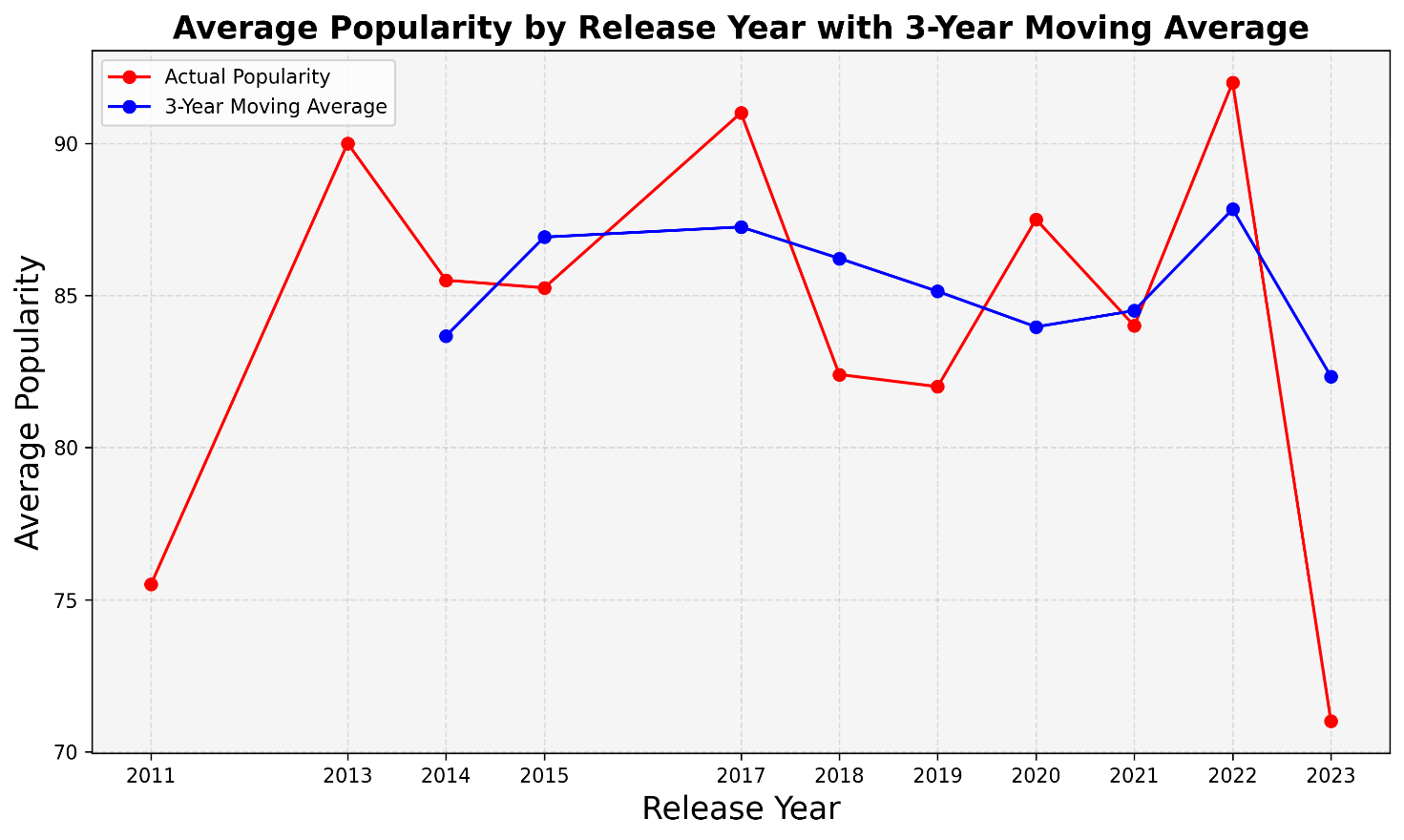


**Insight:**  
Songs released in recent years (up to 2022) generally had higher average popularity scores on Spotify compared to those from earlier years

**b) 📊 Is the Popularity of Songs Increasing or Decreasing Over the Years?**

To check for a consistent trend, I visualized the **3-year moving average of popularity scores**.

**Result:**  
The moving average plot revealed that while **newer songs are generally more popular**, the popularity does **not increase consistently year after year**.  
There are noticeable fluctuations between some years.

****

**Insight:**  
This suggests that while modern songs have an edge in popularity, **other factors such as artist fanbase, platform trends, and viral reach play a significant role**, preventing a strictly upward trend.

**c) 📊 Which Year Had the Highest and Lowest Average Popularity?**

**Result:**

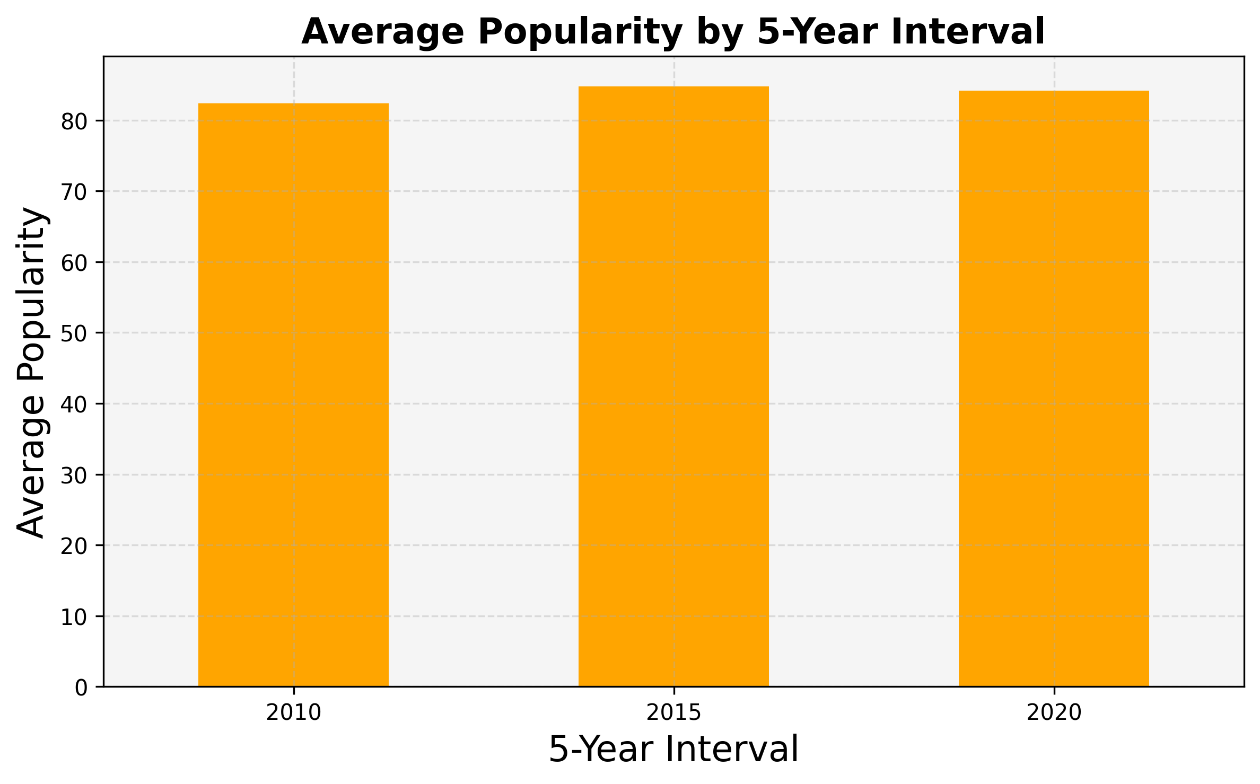
* 📈 **Highest Average Popularity:** **2022**
* 📉 **Lowest Average Popularity:** **2013**

**Insight:**  
This reflects how certain years can outperform others in terms of popular releases, likely influenced by breakthrough hits, album drops, and global music trends.

**d) 📊 Grouped Popularity Analysis (5-Year Intervals)**

To smooth out year-by-year noise, I grouped songs into **5-year intervals** and calculated their average popularity.

**Result:**  
The **5-Year Average Popularity bar chart** shows that the interval with the **highest average popularity was 2015**, followed closely by **2020**, with **2010** slightly lower.



**Insight:**  
This indicates that **recent years generally have higher popularity averages**, though the difference between intervals is not huge — confirming that while newer music enjoys good popularity, it fluctuates between periods based on individual hit releases and audience preferences.

**Final Insight for 4.3**

The data suggests that **newer songs are generally more popular on Spotify than older ones** in this dataset. However, the **popularity trend is not strictly increasing year after year**, indicating that factors like viral marketing, streaming culture, and artist fandom heavily influence song success beyond just the release date.

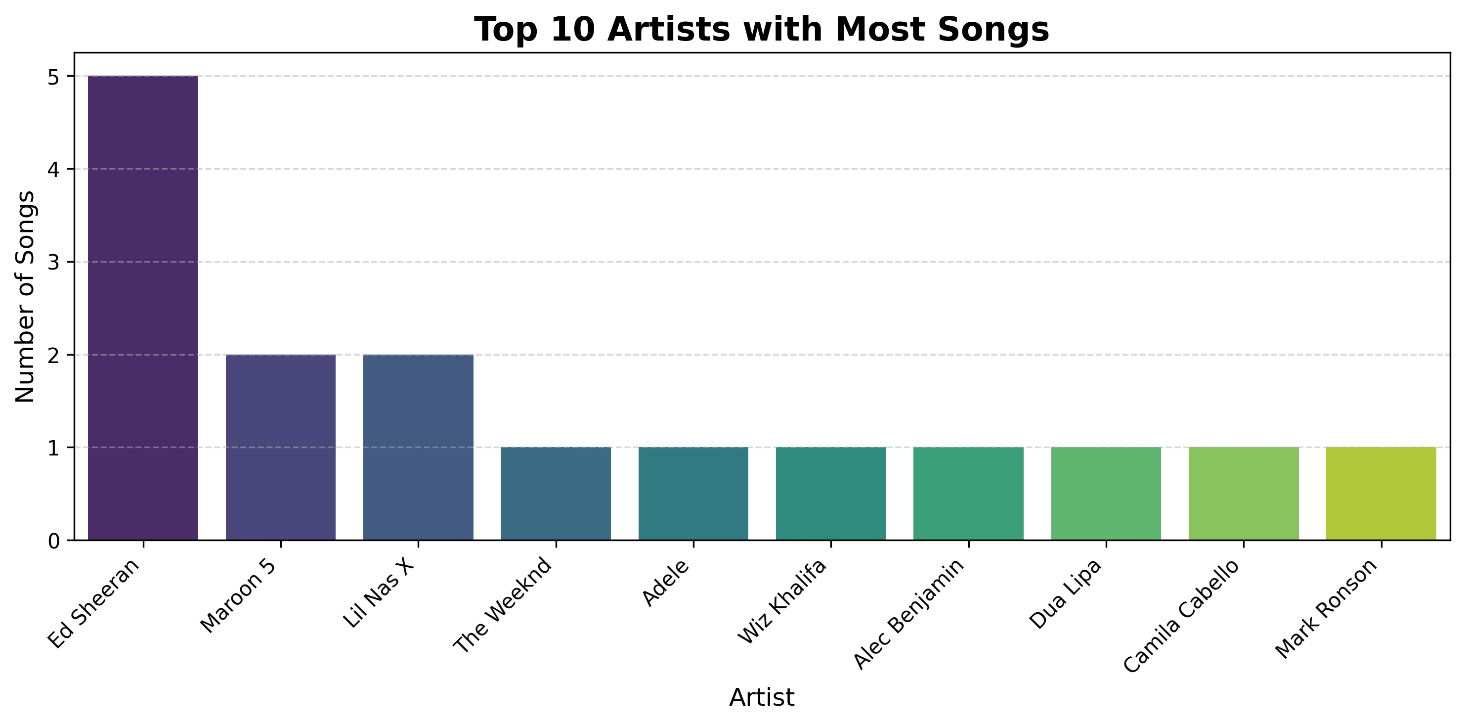
**4.4 Artist Sentiment Profiles**

This analysis explored **which artists contributed the most songs to the dataset**, as well as their **average sentiment profiles** and how their songs’ sentiments are distributed. It also identified **which artists produced the most positive and negative songs** overall.

**a) 📊 Top 10 Artists with Most Songs**

I calculated the **count of songs for each artist** and plotted the top 10 contributors.

**Result:**  
The **bar chart of song counts per artist** revealed that **Ed Sheeran** topped the list with **5 songs**, followed by **Maroon 5** and **Lil Nas X** with **2 songs each**.

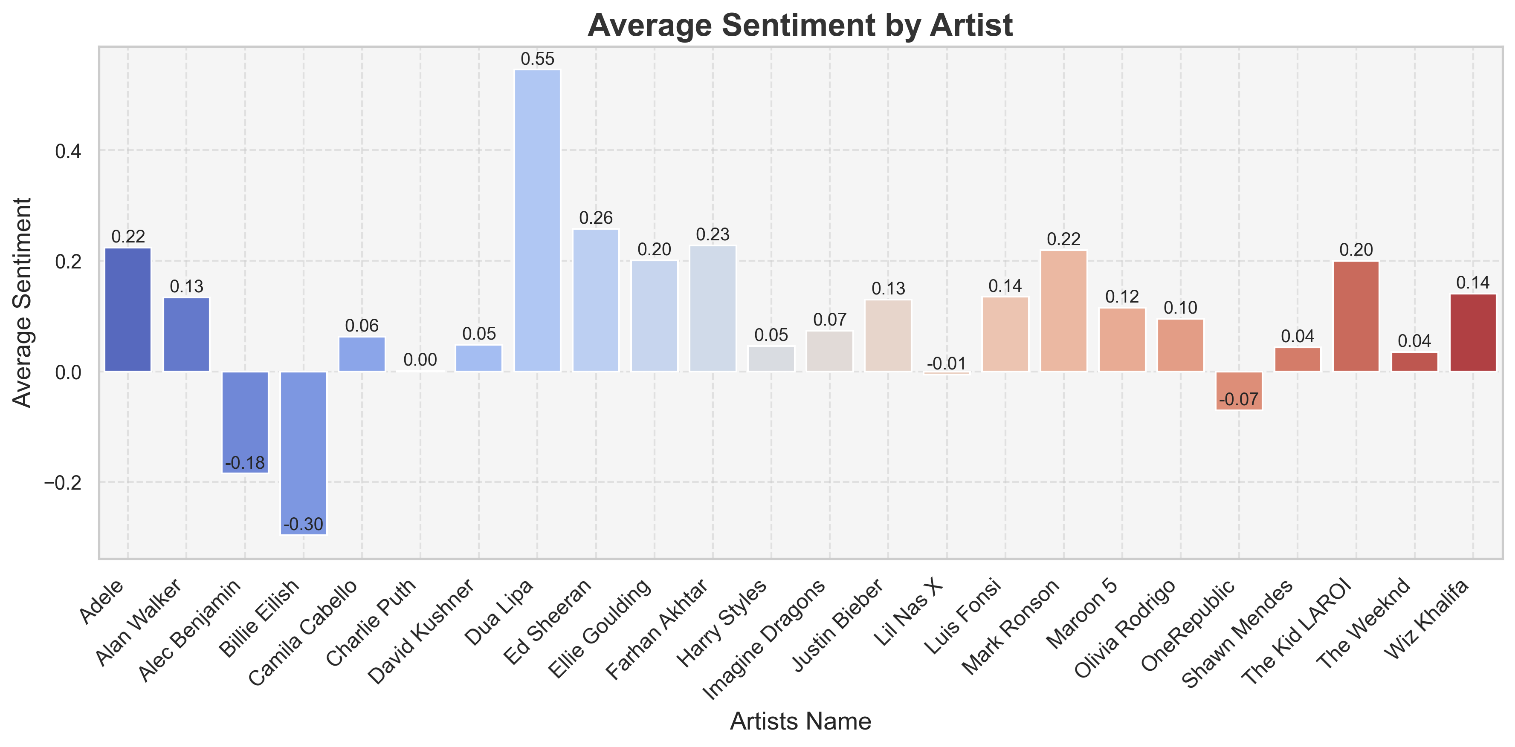
****

**Insight:**  
This indicates a heavier contribution from a few globally popular artists, especially **Ed Sheeran**, within the dataset of 30 tracks.

**b) 📊 Average Sentiment by Artist**

Next, I computed the **average sentiment polarity score for each artist**.

**Result:**  
The **bar chart of average sentiment by artist** showed that **Ed Sheeran** had the highest positive sentiment polarity on average, while artists like **Lil Nas X** leaned toward the negative side.

****

**Insight:**  
This reveals how certain artists tend to create more emotionally positive music, while others express slightly darker or mixed emotions.

**c) 📊 Artist with Most Positive and Most Negative Songs**

By filtering songs based on sentiment polarity thresholds, I identified the artist with the highest count of positive and negative tracks.

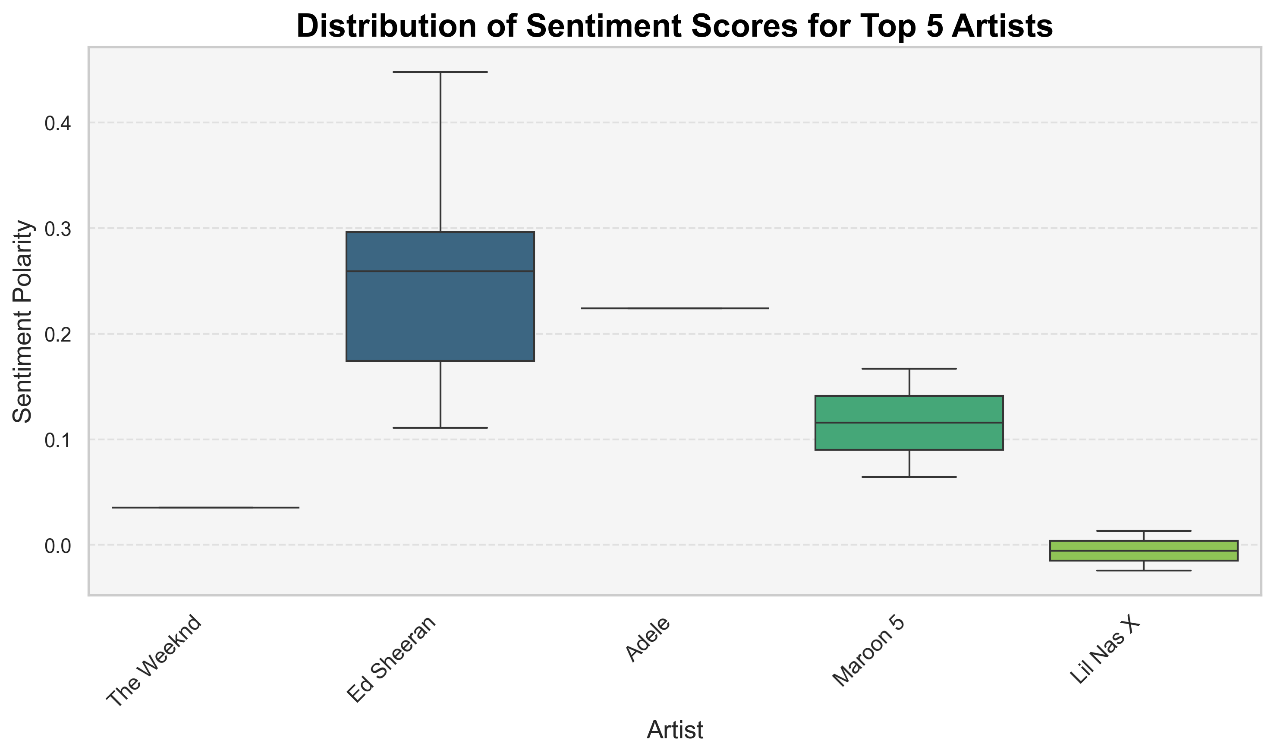
**Result:**  
📈 **Most Positive Songs:** Ed Sheeran  
📉 **Most Negative Songs:** Lil Nas X

**Insight:**  
In this dataset, **Ed Sheeran leads in positive tracks**, whereas **Lil Nas X contributed the most negative tracks**, reflecting differing thematic tones and lyrical moods between these artists.

**d) 📊 Distribution of Sentiment Scores for Top 5 Artists**

I visualized the **distribution of sentiment polarity scores for the top 5 artists** using a boxplot.

**Result:**  
The **boxplot showed that Ed Sheeran’s songs had a wider range of sentiment scores**, mostly positive, while other artists had narrower sentiment ranges clustering around neutral or slightly positive values.

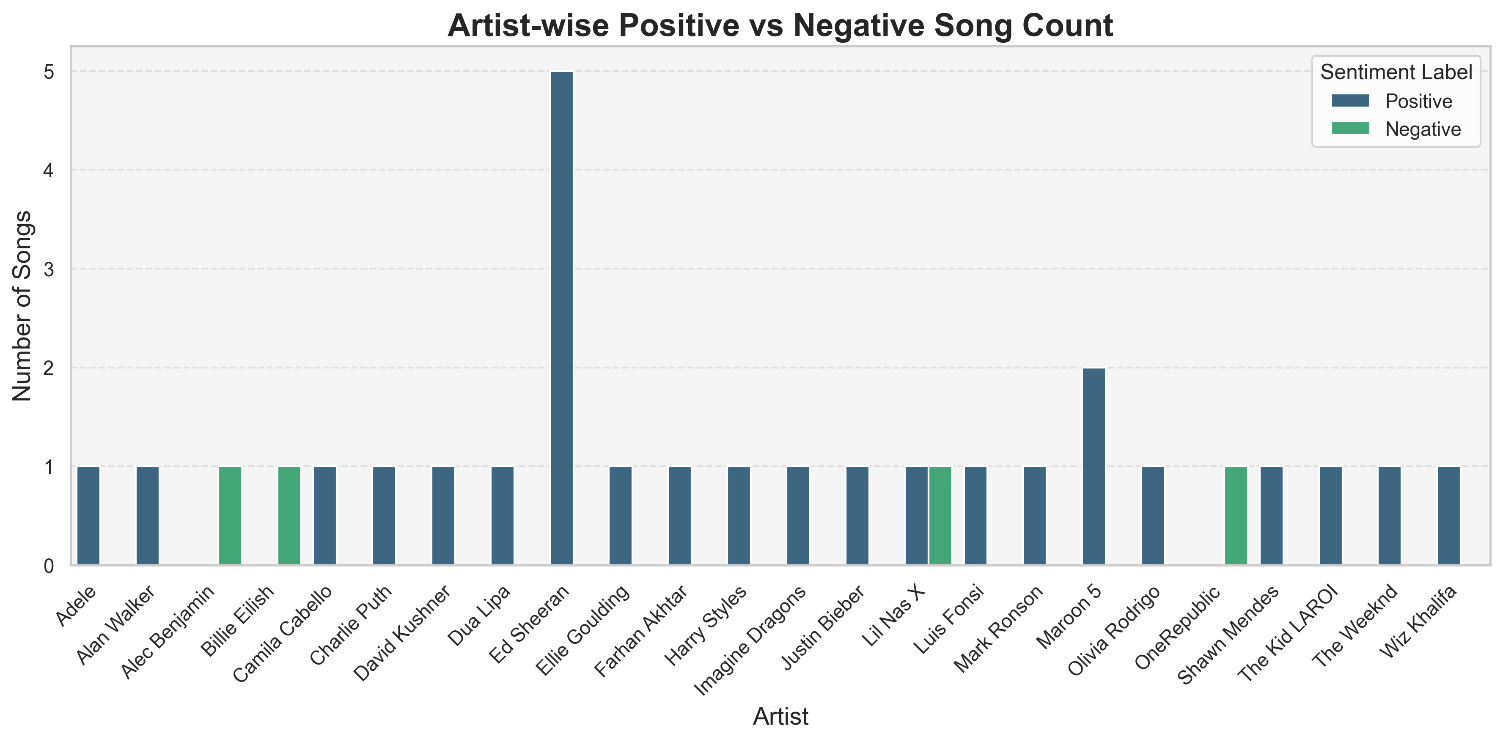


**Insight:**  
This indicates that **Ed Sheeran’s music is both more emotionally expressive and varied**, while other artists in this dataset maintain a more consistent sentiment profile.

**e) 📊 Artist-wise Positive vs Negative Song Count**

Finally, I compared the **number of positive and negative songs per artist** in a grouped bar chart.

**Result:**  
The **bar chart revealed that most artists have more positive songs than negative**, though **Lil Nas X was a notable exception** with a higher count of negative tracks.

****

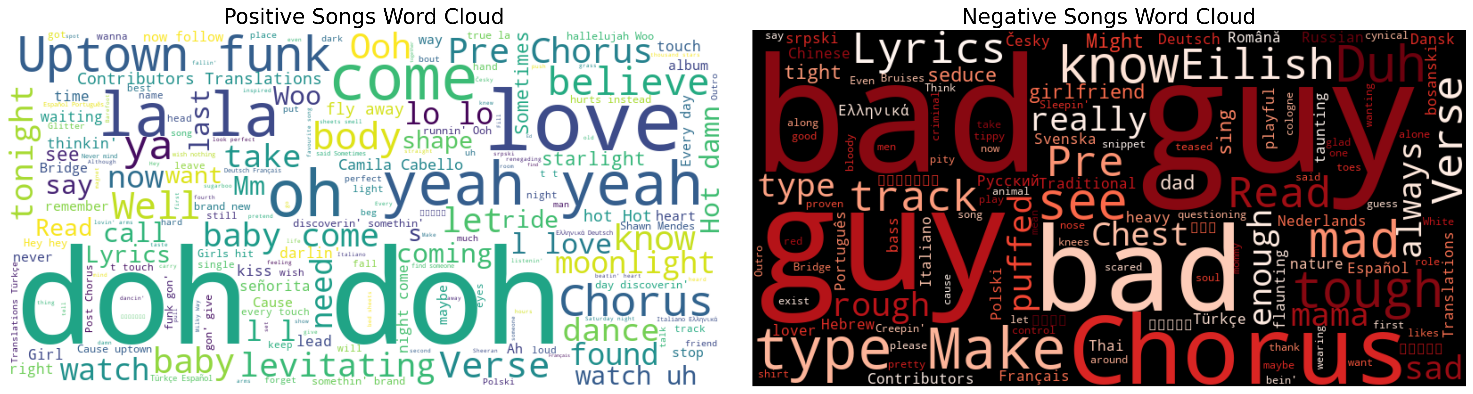
**Insight:**  
The majority of songs in this dataset lean towards positivity, reaffirming the common tendency of mainstream music to evoke uplifting emotions, with a few exceptions reflecting darker lyrical tones.

**Final Insight for 4.4**

The analysis shows that **Ed Sheeran consistently stands out as the most frequent and positively emotive artist in this dataset**. Most other artists contributed fewer songs with sentiment scores clustering around neutral to mildly positive. **Positive songs are clearly more dominant overall**, highlighting the prevalent trend of upbeat and emotionally engaging music in popular culture.

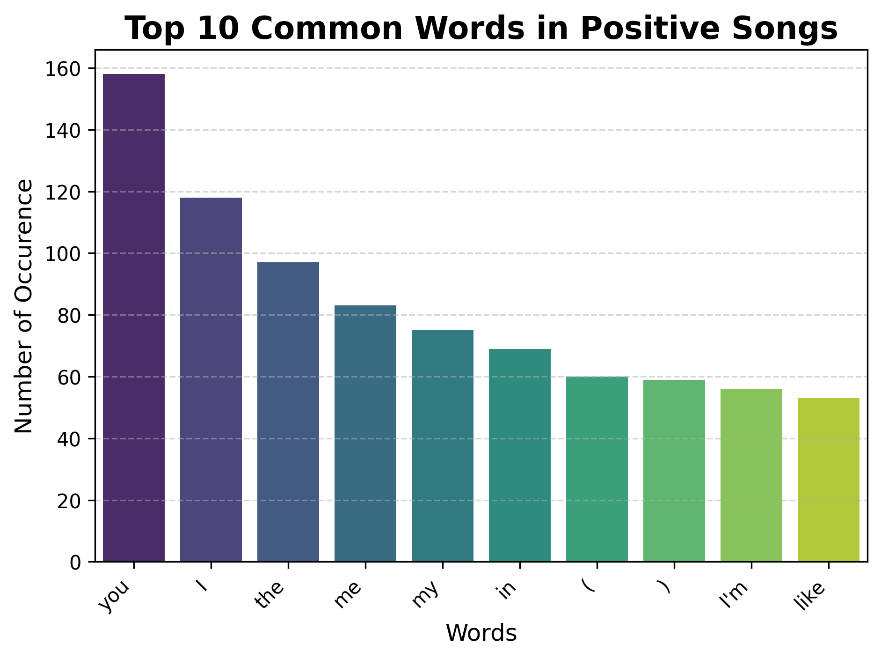
**4.5 Word Cloud & Text-based Insights**

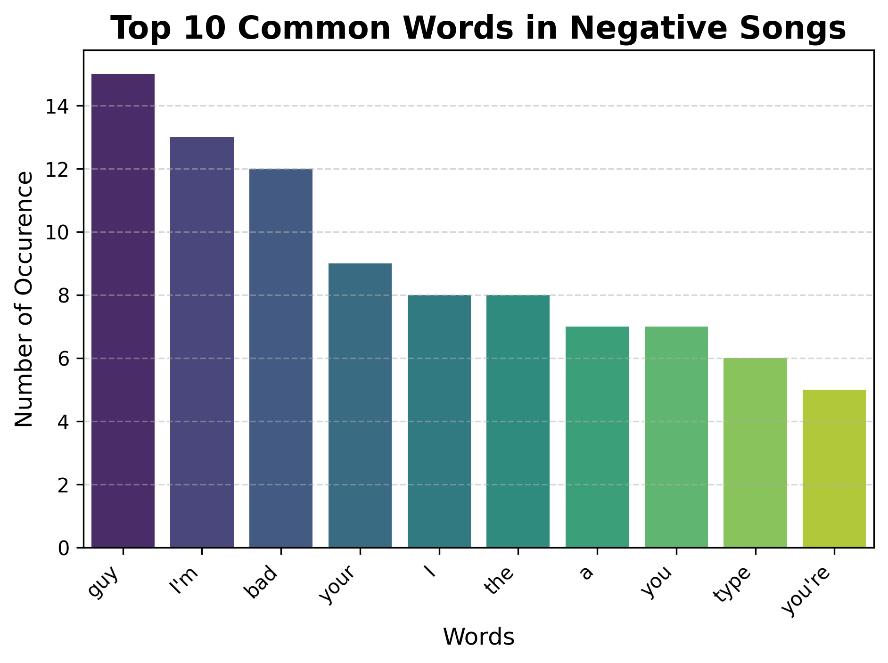
This section explores **the language patterns and sentiment-based content trends in song lyrics** using word clouds, word frequency analysis, word counts, and sentiment distribution.



**a) 📊 Top 10 Words in Positive and Negative Lyrics**

**Result:**  
✅ In **positive songs**, most frequent words were “you”, “I”, “the”, “me”, and “love”.  
✅ In **negative songs**, dominant words included “guy”, “bad”, “you”, and “I'm”.

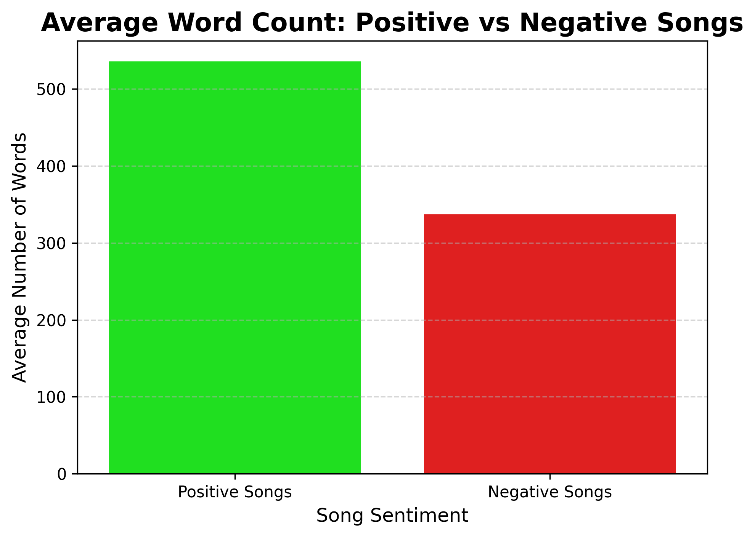




**Insight:**  
Positive lyrics prominently feature words related to **affection, identity, and celebration**, while negative songs emphasize **darker, introspective, or confrontational themes**. This aligns with expected emotional tones in popular music.

**b) 📊 Average Word Count in Positive vs Negative Songs**

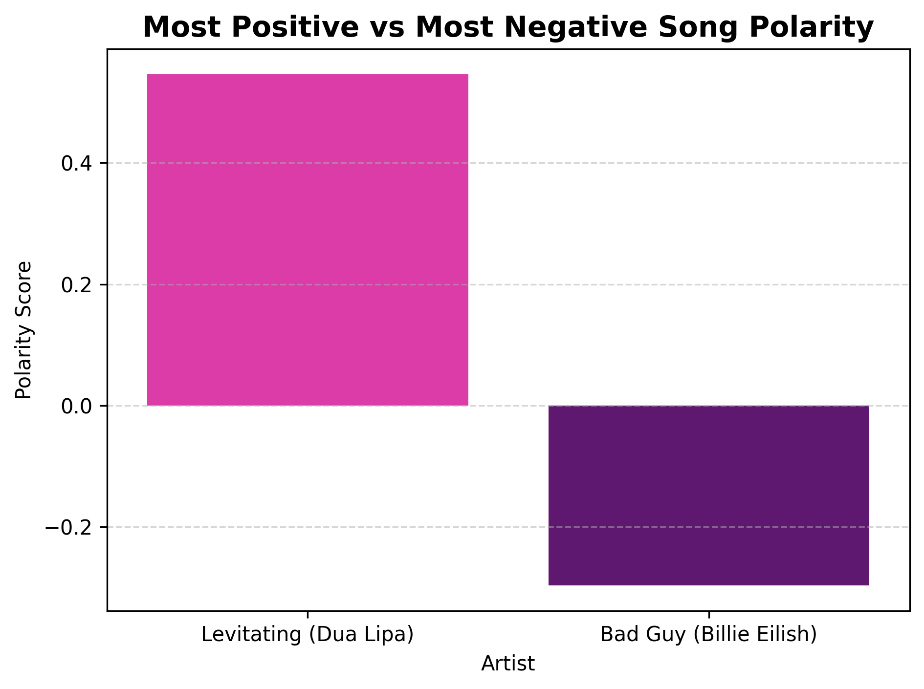
**Result:**  
✅ **Positive songs average:** 535.88 words  
✅ **Negative songs average:** 337.00 words



**Insight:**  
**Positive songs tend to be longer and more descriptively expressive**, while **negative songs are typically shorter and more direct**, possibly reflecting audience preferences for emotionally rich content in upbeat tracks.

**c) 📊 Most Positive & Most Negative Song**

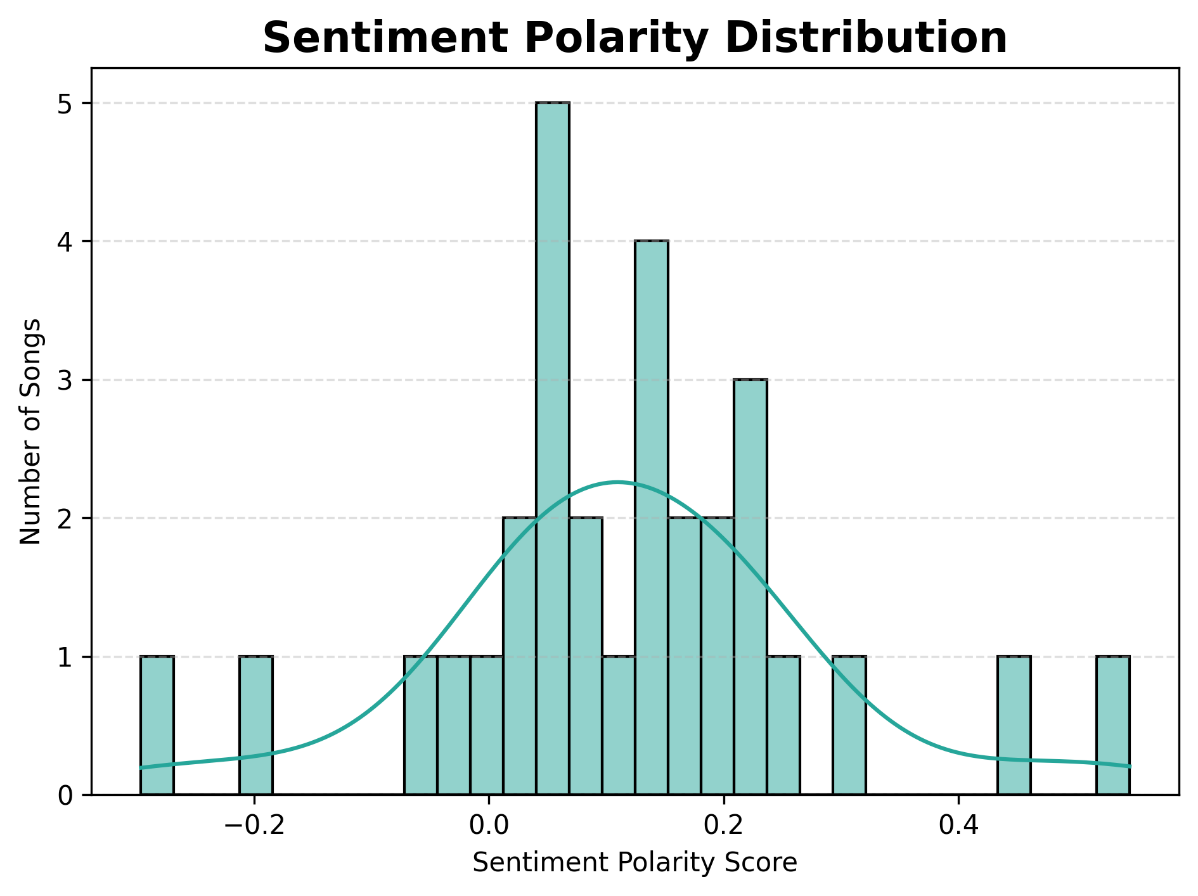
**Result:**  
🎶 **Most Positive Song:** Levitating by Dua Lipa (Polarity: **0.5457**)  
🎶 **Most Negative Song:** Bad Guy by Billie Eilish (Polarity: **-0.2962**)



**Insight:**  
This confirms that even popular mainstream songs can vary greatly in sentiment — with some embracing extremely positive, uplifting lyrics, while others delve into darker or rebellious themes.

**d) 📊 Sentiment Polarity Distribution**

**Result:**  
The histogram reveals most songs have a **sentiment polarity clustered between 0.0 and 0.2**, skewing towards the positive side, with very few songs in strong negative ranges.



**Insight:**  
The lyrical sentiment in this dataset predominantly leans **neutral to mildly positive**, reflecting the pop music industry’s tendency to favour optimistic, feel-good content over darker themes.

**Final Insight for 4.5**

This word and text analysis reaffirmed that **popular song lyrics center around optimistic, personal, and romantic themes**. Positive songs are typically **wordier and lexically richer**, while negative ones are **concise and stark**.  
The overall sentiment distribution also revealed a clear industry bias toward **uplifting, emotionally safe lyrical content**.

**Conclusion**

This project explored the intersection of **lyrics sentiment and popularity trends** in popular music by analyzing a curated dataset of 30 globally popular songs using Python-based sentiment analysis and data visualization techniques.

**Key Takeaways:**

* **Sentiment vs Popularity:**  
  There’s **no significant correlation** between a song's sentiment polarity and its popularity score. Both positive and negative songs can be popular, though extremely polarizing songs tend to attract slightly more attention.
* **Average Sentiment by Release Year:**  
  Songs peaked in positive sentiment around **2020**, but recent years have shown a decline towards **neutrality**, suggesting a shift in the industry toward **emotionally balanced, less exaggerated lyrical content**.
* **Popularity Trend Over Time:**  
  **Newer songs generally outperform older classics** in popularity, though the increase isn’t consistently linear. Popularity trends experience fluctuations, with recent years maintaining higher averages.
* **Artist Sentiment Patterns:**  
  **Ed Sheeran** appeared consistently among top positive artists, while others like **Billie Eilish** dominated negative sentiment tracks. Most popular artists tend to favor **positive lyrical content**.
* **Word & Text-Based Insights:**  
  Positive songs featured **longer lyrics with affectionate and self-referential vocabulary**, while negative songs were **shorter and more direct**. The overall sentiment polarity distribution skewed towards **neutral to mildly positive values**.

**📊 How This Analysis Could Be Useful**

* **For music streaming platforms:** to enhance **personalized playlist recommendations** based on listener mood preferences.
* **For artists and producers:** to gauge **audience sentiment preferences** and craft commercially effective content.
* **For market researchers:** to track **emotional shifts in pop culture music trends** over time.

**⚙️ Limitations & Future Scope**

* **Small sample size (30 songs)** — expanding to a dataset of 1000+ songs would yield deeper, more reliable patterns.
* **Genre and multi-language sentiment variance** wasn’t explored — future work could include **genre classification and multilingual sentiment analysis** using advanced models like **VADER, BERT, or RoBERTa**.

**References**

1. **Spotify Web API**  
   <https://developer.spotify.com/documentation/web-api>
2. **Genius API (Lyrics Data)**  
   https://docs.genius.com/
3. **TextBlob Python Library (Sentiment Analysis)**  
   Loria, S. (2018). TextBlob: Simplified Text Processing. <https://textblob.readthedocs.io/en/dev/>
4. **Matplotlib Python Library (Data Visualization)**  
   Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90-95. DOI:10.1109/MCSE.2007.55
5. **Seaborn Python Library (Statistical Data Visualization)**  
   Waskom, M. et al. (2020). Seaborn: statistical data visualization. https://seaborn.pydata.org/
6. **Pandas Python Library (Data Analysis & Manipulation)**  
   McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, 51-56.
7. **WordCloud Python Library**  
   Mueller, A. (2020). wordcloud: A little word cloud generator in Python. <https://github.com/amueller/word_cloud>