Sentiment Analysis of the Most Popular Song Lyrics

Motivation

Music is a powerful form of expression that easily resonates with human emotions. Almost everyone enjoys some genre or type of music, making it a near universal aspect of human culture. As streaming platforms and social media amplify the spread of music, analyzing the emotional tone or sentiment behind popular songs can provide valuable insights into social moods, cultural trends, and listener preferences.

The primary goal of this project is to apply a sentiment analysis model to the lyrics of the most popular songs from the last few years. By doing so, we aim to uncover patterns in how sentiment correlates with song popularity, genre, or temporal trends.

Data Engineering

To achieve this, we first identify which songs were most popular in recent years. In this case, we web scrape the Wikipedia entries for the "*Billboard Year-End Hot 100 singles*" from the past 20 years, selecting the top 25 songs from each year. We chose Wikipedia over Billboard's official website because Wikipedia is a static site, making it easier to scrape, and both sources contain the same core information. In addition, we apply the same scraping mechanism for the most-streamed songs on Spotify. This allows us to have another point of view of popularity based on streaming behavior rather than just chart rankings.

Next, we collect the lyrics and genres of the songs. We retrieve lyrics by web scraping Genius.com, a comprehensive resource for song lyrics. After scraping, we clean the lyrics by removing structural elements such as chorus and intro labels, ensuring better performance in downstream sentiment analysis model. To determine the genre of each song, we use the MusicBrainz API, a reliable open-source database for music metadata.
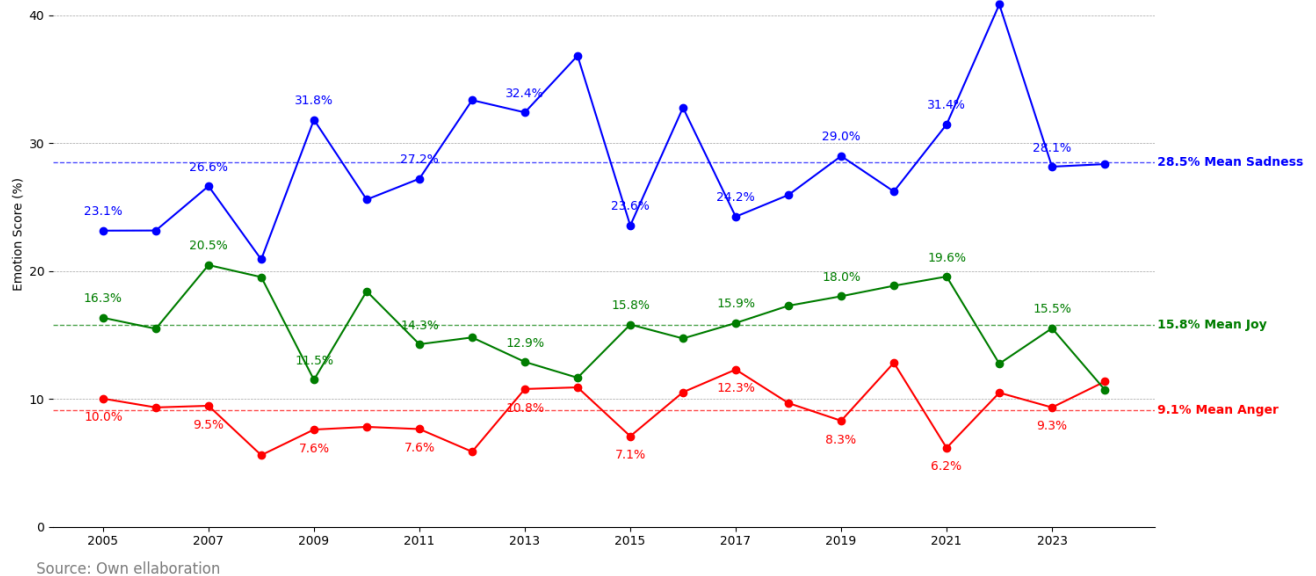
Finally,   we   apply   a   sentimental   analysis   model   from   HuggingFace, j-hartmann/emotion-english-distilroberta-base. This model classifies English text into seven emotion scores, each one of them with a rating from 0 to 1, so we have filtered previously the songs from other languages. And for the scope of this project, we focus on three key emotions: **'Joy'**, **'Sadness'**, and **'Anger'**.
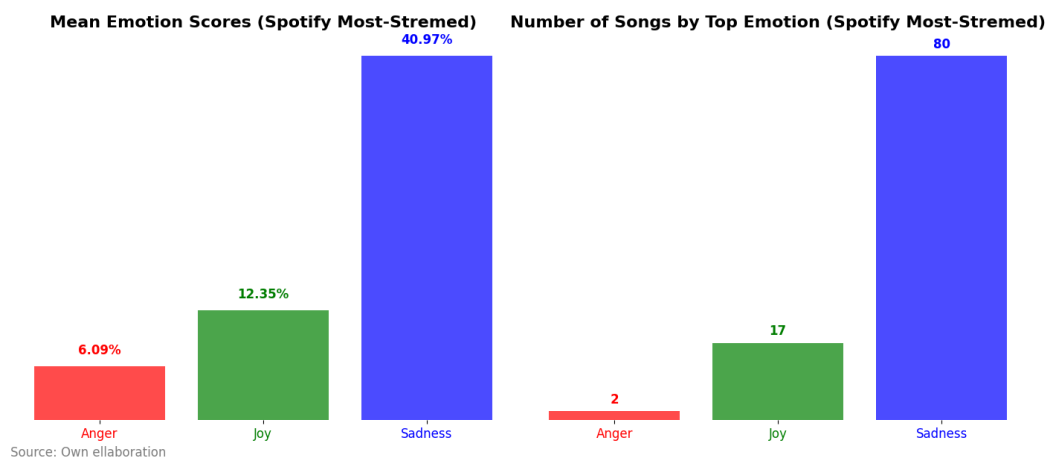
Main Insights

As we can appreciate in the next plot, over the past two decades Billboard's Top 25 songs have consistently leaned toward a more **Sadness** tone, which dominates the lyrical sentiment in this timeline. Beginning at 23.1% in 2005, it steadily rises, peaking at around 41% in 2022. With an average of 28.5%, **Sadness** clearly leads the emotional charge in mainstream music, suggesting that popular songs increasingly reflect themes of vulnerability and heartbreak, especially in recent years. On the other hand, **Joy** remains relatively stable

but secondary, averaging 15.8%. It experienced highs like 20.5% in 2007 and 19.6% in 2021, yet dips notably in 2009. **Anger** stays the least expressed emotion, with an average of 9.1%.

Average Emotion Scores in Billboard Top 25 Songs by Year
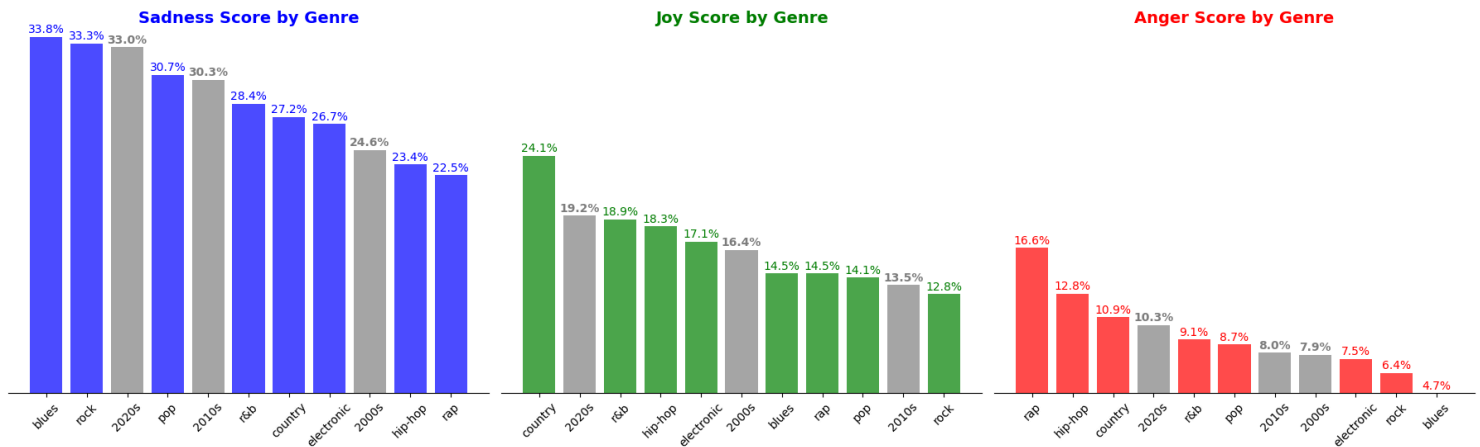

Source: Own ellaboration

This emotional dominance of **Sadness** isn't just only to Billboard's yearly charts—it becomes more pronounced when examining the most-streamed songs of all time on Spotify. As shown in the next plot, **Sadness** significantly dominates, representing an average of 40.97% of the emotion detected in lyrics- surpassing by far **Joy** (12.35%) and **Anger** (6.09%). The second chart supports this pattern, showing that 80 songs from the dataset were classified with **Sadness** as their top emotion, compared to just 17 with **Joy** and only 2 with **Anger**.

Source: Own ellaboration

Regarding the emotion scores per genre, **Sadness** is most prominent in blues, rock, and pop (all above the 30% threshold), with the **2020s** showing a particularly high sadness score of 33%—the highest among decades. Meanwhile, **Joy** is most evident in country music, followed by the 2020s and R&B, indicating a secondary presence of optimism and resilience. **Anger** is most prominent in rap and hip-hop, aligning with themes of resistance and confrontation often explored in these genres. Across decades, the **2000s**, **2010s**, and

2020s show relatively stable **Joy** and **Anger** levels, but the increase in **Sadness** is somewhat noticeable, as it has been increasing over time.
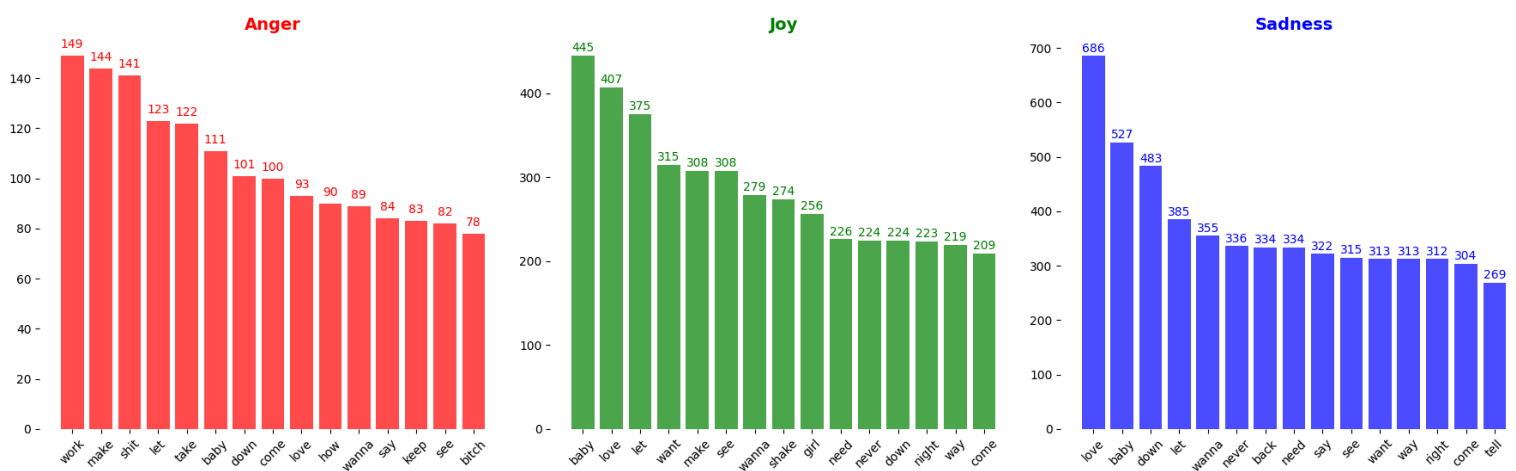
Emotion Scores per Genre (Billboard Top 25)

Now getting deeper into the lyrics themselves, the most frequently used words linked to each emotion can provide us insights into the emotional vocabulary of the music. For **Anger**, words like **"work," "make," "shit,"** and **"let"** are on the top—suggesting themes of confrontation or/and frustration. The presence of bad words like **"bitch"** reinforces the raw and often aggressive tone, especially in genres like rap and hip-hop that we have seen in the last plot.
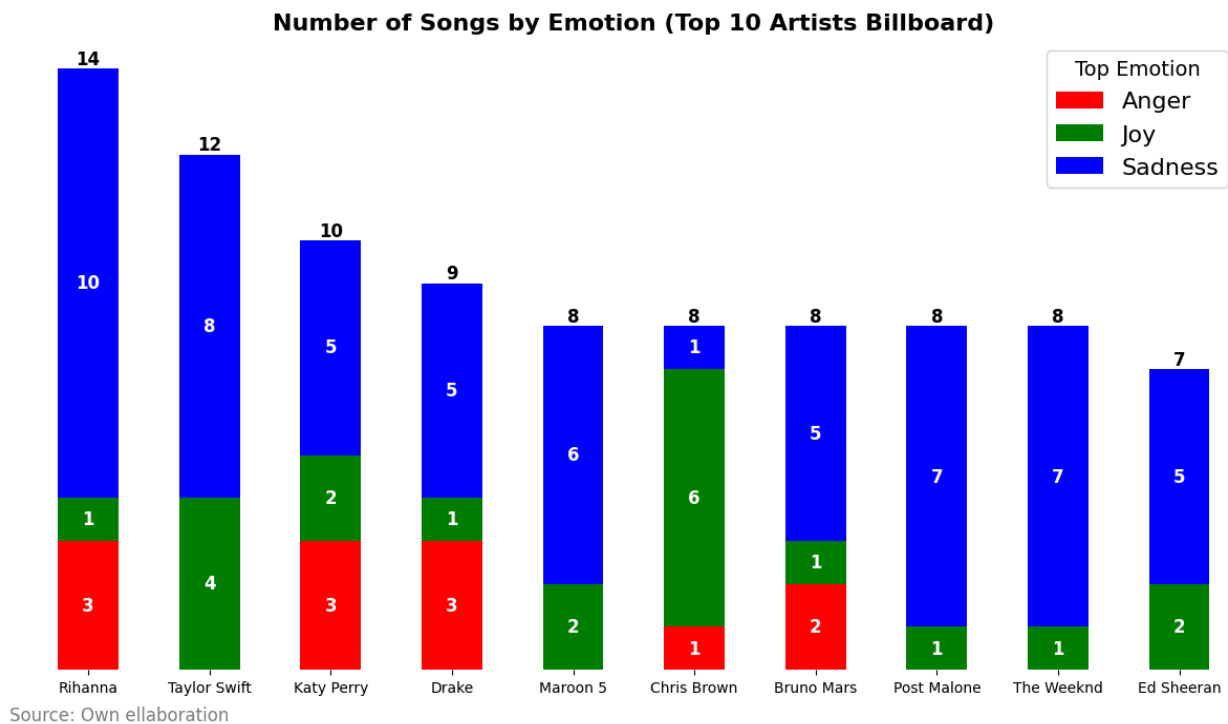
In contrast, **Joy** lyrics are saturated with affectionate language. Words like "baby," "love," "let," and "want" are frequent, reflecting themes of connection and desire energy. On the other hand, **Sadness** is marked by emotionally heavy and introspective vocabulary. Words like **"love," "baby," "down," "let,"** and **"never"** are on the of the list, highlighting loss and emotional struggle. Also, it is very interesting the significant overlap across these last emotions— words such as "baby," "love," "let," and "come"—suggest that context and tone play an important role in how these words contribute to an overall emotional perception, where the same vocabulary can provoke different feelings depending on the context.
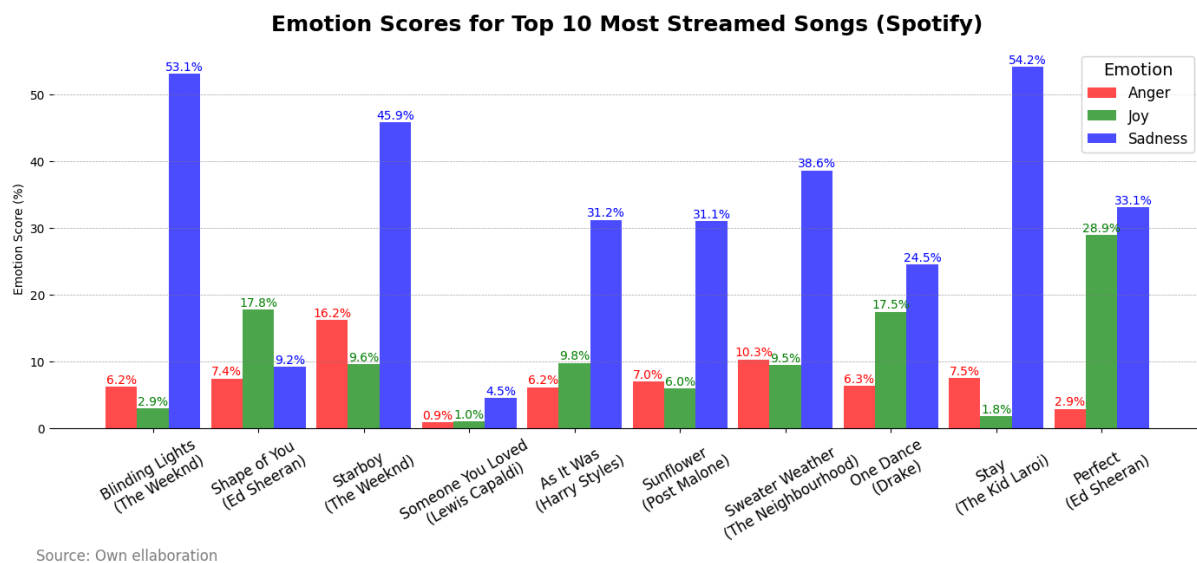
Most Common Words per Emotion

When examining in the below plot the top artists that have more songs on the chart, it's clear that **Sadness**, without any surprise, dominates the emotional tone of biggest hits. Artists like Rihanna, Taylor Swift, and Maroon 5 have more than a half of their songs classified under this emotion, with Rihanna leading the group at 10 tracks in this category. This prevalence could suggest to us an emotional connection between melancholy themes and commercial success in pop music. Even performers known for having overall energetic styles—such as The Weeknd or Bruno Mars have the majority of their most popular songs associated with **Sadness**.

Still, some artists break the trend. For instance, Chris Brown stands out with six **Joy** songs, the highest in this category, while Katy Perry and Drake show more emotional diversity, each contributing songs to all three emotional groups. This emotional breakdown shows us the complexity of modern music, where the popularity of a song is not only based on the upbeat or feel-good themes but also deeply rooted in the vulnerability, heartache, and personal struggle.

Number of Songs by Emotion (Top 10 Artists Billboard)

Source: Own ellaboration

Lastly, while we look closer at the top 10 most-streamed songs of all time on Spotify, again the emotional tilt of **Sadness** remains the most prominent emotion in 8 out of the 10 tracks, with striking examples like *Blinding Lights* by The Weeknd (53.1%) and *Stay* by The Kid LAROI (54.2%). These songs that are widely perceived as upbeat, radio-friendly or that can appeal to a happy feeling with their melody and beat, but at the end they have a lyric that is somewhat sad, as it appeals to a feeling of love or personal loss.

Notably, Ed Sheeran's *Perfect* stands out as an emotional outlier, with **Joy** (28.9%) slightly surpassing **Sadness** (33.1%), offering one of the rare moments of emotional balance. In contrast, *Starboy* by The Weeknd carries a heavier **Anger** score (16.2%), the highest among this selection. This individual breakdown reveals how even the most globally beloved hits aren't necessarily "happy" songs—instead, their massive appeal might be rooted in a deeper emotional resonance that mixes vulnerability, heartbreak, and introspection.

Emotion Scores for Top 10 Most Streamed Songs (Spotify)

Source: Own ellaboration

Conclusions

To sum up, the way emotions show up in popular music has clearly changed over time, with **Sadness** becoming more common in lyrics. While **Joy** and **Anger** still appear, they tend to pop up in certain moments—often reflecting cultural or social events—whereas **Sadness** has slowly taken over as the main emotional tone in many top songs. This shift lines up with a broader cultural move toward being more open about emotional struggles and mental health, showing how music is adapting to reflect what people are feeling and going through.

Listeners don't seem to be choosing sad songs by accident—they're actively connecting with them, as we have seen on the most-streamed song Spotify plots. The steady rise of sad music, especially on streaming platforms, suggests that people find something relatable or comforting in these songs. Whether it's the raw honesty in the lyrics or the overall vibe of modern pop and other melancholic styles, it's clear that **Sadness** has become central to what people want to hear. This fits with the idea that we use music not just for fun, but to process emotions, feel understood, or just sit with our feelings.

Looking at genres helps back this up—**Sadness** shows up strongly in pop, rock, and even blues, while **Joy** leads more in country music, and **Anger** mostly appears in rap and hip-hop. At the same time, the words used in these songs often overlap across emotions, proving that it's not just about the lyrics themselves, but how they're used and which it's the purpose.

References

- Billboard Year-End Hot 100 singles - https://en.wikipedia.org/wiki/Billboard_Year-End_Hot_100_singles_of_2024
- List of Spotify streaming records - https://en.wikipedia.org/wiki/List_of_Spotify_streaming_records
- Genius.com
- MusicBrainz API - https://musicbrainz.org/doc/MusicBrainz_API
- Hugging Face Model - https://huggingface.co/j-hartmann/emotion-english-distilroberta-base