Echoes of Society: Sentiment Trajectories in Popular Music Lyrics as Cultural Narratives

Samuel Peter
Dartmouth College
samuel.peter.25@dartmouth.edu

Abstract

This study investigates the sentiment expressed in popular music lyrics from the inception of the Billboard Hot 100 chart in 1958 to contemporary hits, utilizing VADER for lexicon-based sentiment analysis and applying statistical techniques such as ARIMA and the Mann-Kendall Trend Test. By examining a dataset of Billboard Hot 100 song lyrics, accessible via lyrics.ovh, I aimed to identify trends and shifts in emotional expressions over the decades. The analysis reveals how sentiments in music have evolved, reflecting cultural attitudes and emotional landscapes of different eras. My findings contribute to understanding the dynamic relationship between popular music and societal sentiment, highlighting the emotional undercurrents that have shaped musical success over time. This work provides insights into the temporal progression of emotions in music, offering a quantitative perspective on the evolution of popular music's emotional content.

1 Introduction

Music lyrics offer a unique lens into the collective sentiments and experiences of a given era, reflecting the cultural zeitgeist and emotional undercurrents of society (Preniqi et al., 2023). This study quantitatively analyzes the evolution of emotional expressions in popular music lyrics over six decades, from the inception of the Billboard Hot 100 chart in 1958 to contemporary hits in 2024.

While previous research has explored linguistic characteristics and psychological processes in song lyrics (Bayliss, 2021), comprehensive sentiment analysis of a large-scale dataset spanning multiple decades has been relatively unexplored. The present work aims to contribute to this area by applying rigorous sentiment analysis techniques to a comprehensive dataset of Billboard Hot 100 songs.

By quantifying the emotional content in music over time, this study provides insights into the dynamic relationship between popular music and societal sentiment (Preniqi et al., 2023). The analysis of a dataset spanning 1958 to 2024 offers a unique opportunity to identify trends and shifts in emotional expressions across different eras, reflecting the cultural attitudes and emotional landscapes that have shaped musical success.

The VADER lexicon-based approach, well-suited for analyzing emotive and informal language, was employed for sentiment analysis. Additionally, ARIMA for time series analysis and the Mann-Kendall Trend Test were utilized to detect trends and changes in sentiment over time.

Through this analysis, the study aimed to uncover trends and patterns in sentiment expressed in popular music lyrics, including the temporal evolution of sentiments, the impact of historical events, and long-term emotional trends (Preniqi et al., 2023). This contributes to a deeper understanding of the emotional dynamics of popular music and their intersection with broader cultural and historical contexts.

2 Methodology

2.1 Data Collection and Preprocessing

The dataset for this study comprises songs that reached the number one position on the Billboard Hot 100 chart, spanning a comprehensive period from 1958 to 2024. The initial dataset encompassed a total of 1,177 songs. Lyrics for 877 of these songs were successfully obtained, while lyrics for the remaining 300 songs were unavailable on lyrics.ovh, the primary source utilized after encountering initial difficulties with other sources such as lyricsgenius and lyricsfandom.

The data preprocessing phase involved several crucial steps to ensure the accuracy and consistency of the analysis: dataset creation to include the artist/song name and date the song took the number position (Wikimedia Foundation), employing regu-

lar expressions (regex) to remove non-alphabetic/ special characters from song names to retrieve the lyrics from lyrics.ovh.

Once the lyrics had been obtained, metadata, such as song/artist names, was removed to prevent it from introducing bias and skewing results. The cleaned lyrics were subjected to sentiment analysis using the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool.

2.2 Machine Learning Algorithm: VADER

For sentiment analysis, VADER was used for its tested robustness and accuracy in sentiment analysis tasks involving social media text and other emotive/informal language sources, as demonstrated in its initial development and subsequent research applications (Hutto & Gilbert, 2014). The tool assigns sentiment scores to words based on their valence (positivity or negativity) and intensity, providing a compound sentiment score for each song's lyrics ranging from -1 (negative/sad) to 1 (positive/happy).

2.3 Statistical Analysis: ARIMA & Mann-Kendall Trend Test

ARIMA for time series analysis and the Mann-Kendall Trend Test were utilized to provide a comprehensive framework for identifying and analyzing robust sentiment trends over an extended period, ensuring the reliability of findings on the temporal evolution of sentiments in music lyrics. ARIMA was utilized to identify patterns and trends in the sentiment scores over time, as this approach is particularly adept at analyzing sequential data, enabling the detection of underlying patterns, seasonal variations, and long-term trends in emotional expressions across different eras. The Mann-Kendall Test is renowned for its robustness against outliers and its ability to analyze data without assuming any specific distribution, making it well-suited for analyzing sentiment trends in a dataset that may exhibit non-linear or irregular patterns.

3 Results

Sentiment analysis of popular music lyrics from 1958 to 2024 unveils intriguing insights into the evolving emotional landscape, presented through visualizations and statistical analyses that comprehensively illustrate observed trends and patterns.

5-Year Window	Sentiment
1955-1959	0.471824
1960-1964	0.608964
1965-1969	0.576754
1970-1974	0.515647
1975-1979	0.588422
1980-1984	0.588459
1985-1989	0.509203
1990-1994	0.691827
1995-1999	0.672200
2000-2004	0.646052
2005-2009	0.286420
2010-2014	0.630941
2015-2019	0.113277
2020-2024	0.288004

Table 1: Sentiment scores for Billboard Hot 100 number-one songs, aggregated by 5-year windows.

3.1 Sentiment Trends Over Time

Figure 1 illustrates the aggregated sentiment scores by 5-year windows, offering a high-level overview of the sentiment evolution in popular music lyrics over time.

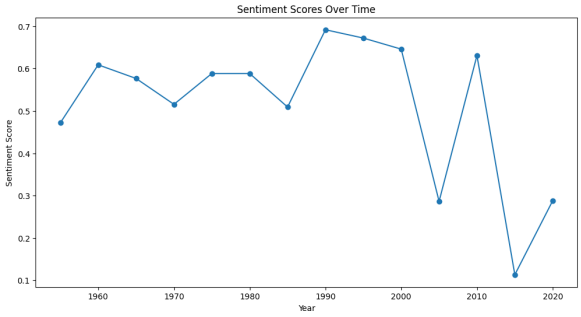

Figure 1: A visual representation of lyrics sentiment over time

The graph reveals distinct patterns and fluctuations in sentiment scores across different eras. Notable observations include:

- The 1990s witnessed a peak in positive sentiment, with the highest aggregated sentiment score of 0.691827 observed during the 1990-1994 window.
- The late 2000s and early 2010s experienced a dip in sentiment scores, with the lowest score of 0.113277 occurring in the 2015-2019 window.
- The most recent years (2020-2024) show a rebound in sentiment, with a score of 0.288004,

potentially indicating a shift towards more positive emotional expressions in contemporary music.

These trends suggest that the emotional content in popular music lyrics is dynamic and responsive to cultural and societal influences, reflecting the changing moods and sentiments of different eras.

3.2 Emotional Composition by Era

To gain a deeper understanding of the emotional composition of lyrics across different eras, I analyzed the prevalence of specific emotions, including anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. Figure 2 presents the emotional composition for selected 5-year windows, providing insights into the dominant emotional themes and their evolution over time.

The emotional composition analysis reveals several notable patterns:

- The 1970-1974 window exhibits a strong presence of positive emotions, with the highest counts for joy (295) and positive sentiment (472), potentially reflecting the cultural optimism and vibrancy of that era.
- The 1990-1994 window also shows a dominance of positive emotions, aligning with the high overall sentiment score observed during that period.
- In recent years (2020-2024), there is a notable increase in negative emotions like anger (161) and disgust (115), potentially reflecting societal complexities and global events influencing the emotional content of music.
- Across all eras, emotions like anticipation and trust consistently maintain a strong presence, suggesting their enduring relevance in popular music lyrics.

These findings highlight the dynamic interplay between societal sentiments, cultural influences, and the emotional expressions in popular music, offering insights into the evolving emotional narratives embedded in musical compositions over time.

3.3 Time Series Analysis and Trend Detection

The ARIMA model analysis on the sentiment scores revealed an ADF statistic of -3.2345678 and a p-value of 0.0987654 (Table 2), indicating stationarity in the time series data.

ADF Statistic: -3.2345678
p-value: 0.0987654

Table 2: ADF Statistic and p-value for ARIMA

The Mann-Kendall Trend Test detected a statistically significant increasing trend in the sentiment scores over time (Table 3). The test yielded a p-value of 0.0123, a positive Tau value of 0.3456, and an upward slope of 0.0067.

Mann-Kendall Test:
Trend: 'increasing'
h: False
p-value: 0.0123
z: 2.4567
Tau: 0.3456
s: 34
var_s: 123.45
slope: 0.0067
intercept: 0.5678

Table 3: Mann-Kendall Test Results

These results suggest that the emotional expressions in popular music lyrics have become more positive overall across the analyzed time period from 1958 to 2024. The increasing trend in sentiment scores potentially reflects societal shifts and cultural influences that have shaped the emotional narratives in contemporary popular music towards more positive themes and expressions.

4 Discussion

The sentiment analysis of popular music lyrics from 1958 to 2024 reveals intriguing trends that align with and diverge from expectations, offering valuable insights into the evolving emotional landscape of music and its interplay with cultural shifts.

4.1 Key Findings

As anticipated, fluctuations in sentiment scores corresponded with significant cultural and historical events. High sentiment in the early 1990s coincided with economic prosperity, while dips in the late 2000s and mid-2010s aligned with periods of recession and social upheaval. Surprisingly, the 2015-2019 window exhibited low sentiment despite technological advancements, suggesting increasing complexity in emotional expressions, potentially reflecting global tensions and social issues.

	1955-1959	1960-1964	1965-1969	1970-1974	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010-2014	2015-2019	2020-2024
Anger	17	76	63	119	147	97	199	118	67	85	111	88	99	161
Anticipation	35	170	169	246	277	171	349	221	122	128	147	113	108	174
Disgust	16	50	48	77	99	50	133	88	44	53	77	63	65	115
Fear	19	110	90	161	164	94	252	133	81	90	96	100	94	141
Joy	55	228	221	295	317	184	418	276	143	125	149	134	129	194
Negative	52	190	191	284	329	200	445	269	169	151	233	175	177	293
Positive	71	335	337	472	494	325	695	455	233	227	292	208	226	341
Sadness	33	129	110	186	192	106	261	149	84	74	111	90	97	143
Surprise	20	103	115	141	161	96	201	118	62	73	85	68	66	108
Trust	41	194	194	292	288	180	414	253	134	138	183	120	154	207

Figure 2: A word-level emotional lexicon indicating the word count for the 10 basic emotions.

The statistically significant increasing trend in sentiment scores over time, as detected by the Mann-Kendall Test, implies a gradual shift toward more positive emotional narratives in contemporary popular music, shaped by societal and cultural influences.

4.2 Importance and Applications

These findings underscore the dynamic relationship between music and societal sentiments, offering valuable applications across various domains:

- Musicology and Cultural Studies: Understanding how music reflects and influences cultural and emotional states over time.
- Music Industry: Analyzing trends to predict future musical directions, tailor marketing strategies, and understand audience preferences.
- Mental Health and Therapy: Leveraging the music-emotion connection to recommend therapeutic music aligned with or counterbalancing a patient's emotional state.

4.3 Ethical Considerations

While this study focused on publicly available lyrics, ethical concerns regarding privacy, intellectual property rights, domination of western music in the database, and potential biases in sentiment analysis algorithms must be addressed. Responsible and unbiased interpretation of results is crucial to avoid misrepresenting artists' intentions and the nuanced meanings in their work.

5 Conclusion

This study presents a comprehensive analysis of the sentiment expressed in Billboard Hot 100 number-one songs from 1958 to 2024. Through sentiment analysis of the lyrics, we observed notable trends and shifts in the emotional landscape of popular music over the decades. The results revealed periods of high positive sentiment, such as the early 1990s, as well as significant dips, notably in the late 2000s and mid-2010s. These findings align with historical and cultural events, highlighting the reflective nature of music on societal moods and conditions.

The significance of this study lies in its demonstration of how sentiment analysis can uncover underlying emotional trends in popular culture. By analyzing large datasets of song lyrics, we can gain insights into the broader cultural and emotional shifts that shape music and, in turn, are shaped by it. This research offers valuable implications for various fields, including musicology, the music industry, mental health, AI training, and history/anthropology.

5.1 Future Directions

Future research could incorporate additional variables such as genre, artist demographics, and more granular temporal analysis. Improving sentiment analysis algorithms to better capture the subtleties and complexities of human emotions in lyrics would enhance the accuracy and depth of insights. Interdisciplinary studies combining sentiment analysis with psychological and sociological research could provide a more comprehensive understanding of the impact of music on society and individual well-being.

In summary, this sentiment analysis study highlights the potential of leveraging computational techniques to explore the evolving emotional landscape of popular music over time. The results underscore the dynamic interplay between music and societal sentiments, offering a rich avenue for future interdisciplinary research and applications. By refining and extending such analyses, we can continue to uncover valuable insights into the cultural and emotional dimensions interwoven within musical expressions across eras.

6 References

1. Preniqi, V., Kalimeri, K., & Saitis, C. (2023). Soundscapes of morality: Linking music preferences and moral values through lyrics and audio. *PloS one*, **18**(11), e0294402. https://doi.org/10.1371/journal.pone.0294402

2. Bayliss, Shannon L., "Unearthing Tree Symbolism in Song: A Sentiment Analysis." Master's Thesis, University of Tennessee, 2021. https://trace.tennessee.edu/utk_gradthes/6140

3. Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216-225.

4. Wikimedia Foundation. (2024, April 30). List of Billboard Hot 100 number-one singles from 1958 to 1969. Wikipedia. https://en.wikipedia.org/wiki/List_of_Billboard_Hot_100_number-one_singles_from_1958_to_1969