Comparing Full Song and Chorus Sentiment Through Vectorization and Clustering

Introduction

Lyrics have long offered insight into cultural moods, social undercurrents, and emotional intensity. This project explores how lyric structure shapes patterns in sentiment and language, and whether the emotional and thematic weight of a song is retained in the chorus compared to the full composition. Utilizing vectorization, clustering, and classification techniques, the analysis compares full song lyrics and choruses across three influential rock artists: Green Day, Tom Petty, and R.E.M. The goal is to evaluate whether choruses carry distinct linguistic and emotional signatures and to test the effectiveness of sentiment tools and vector based models in revealing those differences.

In addition to identifying thematic trends, this work compares cluster assignments between full songs and choruses, quantifies sentiment deltas, and tests the separability of text samples using TF-IDF and logistic regression. Rather than focusing solely on genre or artist style, where the analysis began, the project ultimately centered on the chorus as a lyrical substructure to assess whether it carries more emotional intensity or reflects the tone of the full song. The results offer insight into how different sections contribute to a song's emotional profile and demonstrate that structural segmentation can serve as a powerful analytical lens in text mining.

Method

Data Collection and Structure: Lyrics for 30 total unique songs, 10 from each artist, were collected, and the chorus was manually segmented from the full song. These were stored in a DataFrame containing song title, artist, song, and chorus.

To build a focused and meaningful corpus for classification, exploratory analysis of multiple artists to ensure lyrical diversity and clarity in the chorus. After evaluating several options, Tom Petty, Green Day, and R.E.M. were chosen for their well defined choruses, emotionally expressive content, and the availability of clean, structured lyrics. Full songs and chorus sections were collected separately to support comparison across song segments, resulting in a curated dataset suitable for text vectorization and classification analysis.

Sentiment Analysis: Sentiment scoring was performed using VADER, a lexicon based tool optimized for social media and stylized text, making it suited for analyzing song lyrics. Compound sentiment scores were calculated separately for both full songs and chorus sections. These scores were then used in three ways: to compare sentiment distributions by artist, explore the relationship between sentiment and song length, and compute average sentiment by cluster. The difference in sentiment between full songs and their choruses was also calculated as a delta score, offering a direct measure of emotional shift across structural segments.

Vectorization and Preprocessing: To isolate meaningful language patterns in the lyrics, CountVectorizer was applied using a custom stop word list built on top of NLTK's standard English stopwords. Additional filters were added to exclude song specific filler terms that are

common in music but offer little analytical value. To maintain interpretability and avoid sparsity, max_features=50 was set to retain only the most prominent unigrams. Vectorization was performed separately on the full song texts and their corresponding choruses to preserve structural distinctions between the two segments.

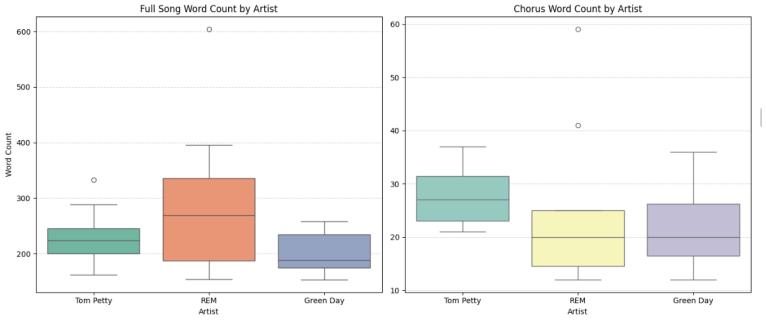
For clustering, the resulting document term matrices or DTMs were converted into NumPy arrays using the .values attribute. This step ensured compatibility with unsupervised learning models, which require numeric array input. The dense matrices maintained unigram frequency information for each entry, serving as the input for downstream projection and clustering techniques.

In addition to raw counts, TF-IDF vectorization was implemented using a custom tokenizer that applied stemming and removed punctuation. These TF-IDF matrices, generated for both full songs and choruses, were incorporated into clustering and classification workflows to assess whether weighting rare terms improved structural or emotional signal separation.

Clustering and Classification: KMeans clustering was applied to both the CountVectorizer and TF-IDF representations to uncover latent groupings among the lyrics. Cluster assignments were compared between full songs and choruses and visualized using annotated scatterplots. To test whether the language patterns could be used to predict cluster membership, logistic regression was run on the TF-IDF vectors, and classification reports were generated for both sets.

Results

Word Count and Sentiment Distribution: Word count distributions revealed that R.E.M. had the widest range in full song length. Green Day's songs skewed shorter overall, while Tom Petty's lyrics were more consistently mid length, with only moderate deviation. When comparing chorus sections, the distributions compressed across all artists, with most choruses falling between 15 and 32 words. R.E.M. again contained the longest chorus outlier, while Tom Petty's choruses remained more compact. These patterns confirmed that choruses are structurally shorter than full songs and generally exhibit less variation in length, providing a more uniform unit for emotional comparison.

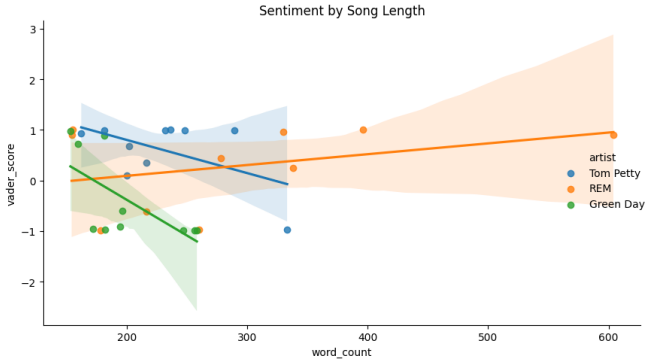


artist	positive	neutral	negative
Green Day	0.3	0	0.7
REM	0.7	0	0.3
Tom Petty	0.9	0	0.1

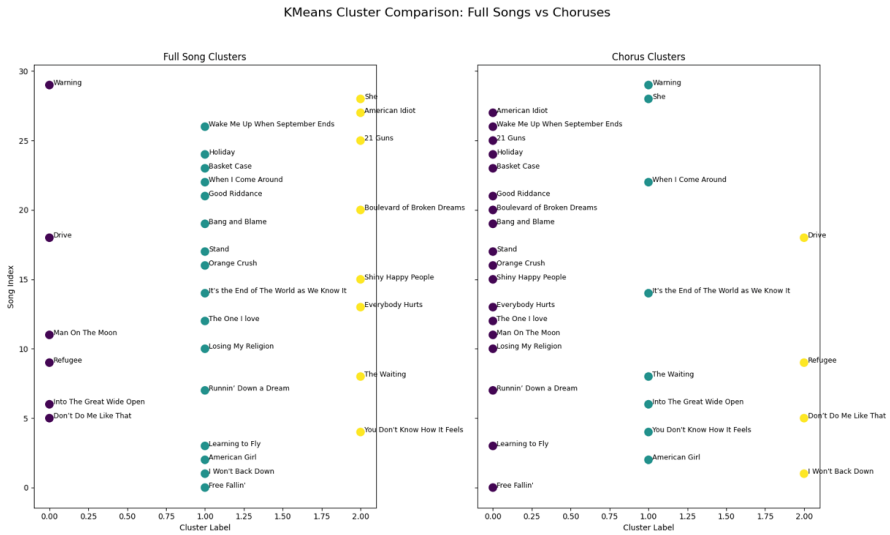
artist	positive	neutral	negative
Green Day	0.6	0.0	0.4
REM	0.4	0.3	0.3
Tom Petty	0.6	0.2	0.2

Tom Petty's full songs show the highest concentration of positive sentiment, followed by R.E.M. Choruses across all three artists are more emotionally balanced, with Green Day showing the most dramatic shift of negative sentiment dropping by half and positive sentiment nearly doubling. R.E.M. and Tom Petty also show increased neutrality and softer polarity in their

choruses. This suggests that choruses tend to moderate emotional tone compared to full lyrics, especially for artists with more polarized songs. When sentiment is viewed by word count, Tom Petty's shorter songs are more positive, while longer ones lean neutral. R.E.M. shows a mild positive slope, though the data is limited. Green Day presents a clear downward trend, with longer songs becoming more negative, likely reflecting a shift toward more critical themes. Overall, emotional tone varies by artist but also changes with structure and length.



Clustering Analysis: KMeans clustering on the CountVectorizer matrices revealed interesting thematic distinctions. While full songs formed clusters somewhat aligned with artist identity, choruses showed tighter grouping, suggesting more structural similarity. Annotated scatterplots made these patterns visible at the song level. For example, several Green Day choruses clustered together, even when the full songs did not, implying a shared emotional or lexical core across different tracks.



When VADER sentiment scores were averaged by cluster, a noticeable delta between full songs and choruses appeared in some groupings. In one case, choruses were more positive than their corresponding songs by an average of 0.18 sentiment points, supporting the hypothesis that choruses often act as emotional anchors.

song_cluster	song_sentiment	chorus_sentiment	delta
0	0.072	0.253	-0.181
1	0.265	0.194	0.071
2	-0.049	0.084	-0.133

Classification Performance: Logistic regression on the TF-IDF matrices showed moderate success in predicting cluster membership. The classification results reveal a clear contrast in how well the model performed on full songs versus choruses. For the full songs, performance was poor across the board, with an overall accuracy of just 0.44. The model was only able to predict Cluster 1 with any success, and even that came at the cost of completely missing the other two. This may reflect noise or redundancy in the full lyric content, where repeated words or filler dominate the feature set. In contrast, the classification of choruses performed noticeably better, with an accuracy of 0.78 and solid precision and recall for Cluster 0. While the sample size was small, this suggests that the chorus sections carried more distinct or learnable patterns, likely due to their role as thematic and emotional focal points of a song.

Full Song Classification Report

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1
1	0.44	1.00	0.62	4
2	0.00	0.00	0.00	4
accuracy			0.44	9
macro avg	0.15	0.33	0.21	9
weighted avg	0.20	0.44	0.27	9

Chorus Classification Report

	precision	recall	f1-score	support
0	0.75	1.00	0.86	6
1	1.00	0.50	0.67	2
2	0.00	0.00	0.00	1
accuracy			0.78	9
macro avg	0.58	0.50	0.51	9
weighted avg	0.72	0.78	0.72	9

Conclusion

This fun project explored how thematic language and sentiment diverge between full song lyrics and their choruses using a combination of text mining techniques, vectorization, clustering, and classification. The dataset was manually constructed to preserve structural integrity between each song and its corresponding chorus, enabling direct comparison. Through careful preprocessing, custom stopword filtering, and TF-IDF representation, the lyrical content was reduced to its most essential and interpretable terms. KMeans clustering on both full songs and choruses revealed some alignment, but also meaningful variation in how songs are grouped based on the language used in the most emotionally charged segments.

Thematic analysis across the choruses showed a concentration of emotionally potent, action driven language that often differed from the broader narrative arcs presented in full lyrics. This was supported by sentiment scoring, where choruses frequently showed distinct sentiment values relative to their parent songs. These shifts were not random; they aligned with the known function of the chorus as the emotional anchor or hook. Clustering patterns, particularly when visualized together, confirmed that songs with similar full text did not always align in chorus grouping, indicating that artists may use choruses to depart from or intensify their intended narrative.

The final classification task reinforced this distinction. While full song classification suffered from noise and redundancy in language, chorus classification produced stronger predictive results, indicating that the chorus structure carries more consistent and learnable linguistic signals. This analysis highlights how songwriters craft chorus sections to distill emotion and message, often leveraging different linguistic strategies than in the surrounding verses. For text mining applications, especially in music or entertainment analytics, this reinforces the importance of examining structural components independently to capture deeper thematic and emotional variance.

“It's something unpredictable, but in the end it’s right.” --Green Day, Good Riddance