

Quantifying the Complex Relationship between Lyrics, Chord Progression, and Emotion Stimulation

Examine the Song Features using Natural Language Processing and Clustering Techniques

1. Background

Music is a potent force that transcends mere auditory enjoyment; it wields the power to stir deep emotions, memories, and collective feelings within listeners. Within songs that exude a similar ambiance or vibe, the delicate interplay between lyrics and chord progression stands out as a pivotal factor. This distinction isn't merely academic; it holds profound implications for record labels in the arduous process of scouting and recruiting bands that can resonate with broader audiences. Central to this investigation is the association between lyrics and the emotions they elicit. Can a song's lyrics, coupled with its musical arrangement, consistently evoke emotions within diverse communities? This question is especially pertinent in the context of social movements. Historically, political songs have served as influential vehicles to engage vast swathes of the population. They act as conduits to disseminate missions, galvanize support, and forge unifying identities amidst diverse individuals. But beneath the surface, is there a discernible pattern amid the seemingly chaotic fusion of lyrics and musical structures? Or does the magic of music lie in its unpredictability? This project endeavors to unravel these intricate threads, shedding light on the science behind the art.

2. Literature Review

Songs are predominantly composed of two fundamental elements: tune (defined by chord progression and key) and lyrical content, characterized by specific word choices and recognizable symbols. While the interplay between lyrics, tune, and emotion may initially appear elusive, a myriad of studies has ventured to quantify the impact of songs on human emotions. For instance, a study by Napier and Shamir (2018) employed digital humanities and data science methodologies to scrutinize the evolution of lyrics from the 1950s to the present, revealing significant upticks in expressions of anger, disgust, fear, sadness, and conscientiousness, contrasted by declining representations of joy, confidence, and openness. Further, Kolchinsky et al. (2017) proposed a novel technique to explore the relationships between chord types and emotional valence, investigating the correlations among chord categories (e.g., Major and Minor) and their association with sentiments and lexicons across various genres, regions, and time periods. Additionally, text similarities in lyrics could serve as a new dimension for clustering music and devising recommendation algorithms through the application of Natural Language Processing (NLP) and the K-Nearest Neighbor (KNN) algorithm (Choi, Song, and Kim, 2018).

3. Research Question

In the realm of music, how do lyrics, melodies, and emotions interweave to produce the songs that resonate with listeners? This investigation seeks to answer this central question using a robust dataset containing 2735 distinctive music pieces from 651 songs. With the application of both unsupervised and supervised machine learning methods, this study aims to decode the intricate relationship between a song's lyrics, its musical tune,

and the emotion it evokes. The preliminary phase employs Word2Vec and K-means clustering to probe for any inherent commonalities in chords and keys across songs bearing similar lyrical messages. This study then turns the focus to the emotional resonance of these songs, employing sentiment analysis coupled with K-means clustering. Delving deeper, Latent Dirichlet Allocation (LDA) is used to explore the thematic underpinnings of various songs. A subsequent step involves the use of Principal Component Analysis (PCA) to discern whether a song's lyrics or its chord progression holds more sway in determining thematic similarity. Building on this, advanced algorithms like the Random Forest and Gradient Boosting Machine will predict song labels by synthesizing a range of musical attributes alongside textual insights obtained from Natural Language Processing (NLP). In the latter stages, this study would discern whether certain repetitive phrases in songs are emblematic of broader social movements or perhaps echo collective societal memories. Through this multifaceted approach, this study aspires to present a holistic understanding of the symphony of elements that come together in the creation of memorable music.

4. Data

The study amalgamates two publicly accessible datasets. The first (Hamer, 2019), derived from Hooktheory and Spotify API, offers chord progression annotations based on song prevalence within its database. The second (Shrirang, 2022), Spotify's Million Song Dataset available on Kaggle, provides a compendium of song names, artist details, links, and lyrics, facilitating song recommendations, classifications, and cluster analyses. Following the merger of these datasets, 2735 distinct observations were identified as the unit of analysis. However, some inherent limitations exist, such as the assignment of lyrics to the same 'artist' and 'song', leading to multiple pieces (e.g., verses and chorus) sharing identical lyrics. To mitigate this, Python was utilized to segregate chorus and verses, assigning them to corresponding sections. Furthermore, the chord progression was simplified to 29 chords, potentially overlooking more obscure and complex chords.

5; Preliminary Data Analysis

Upon preliminary data examination, a bar chart was constructed (refer to Appendix 1) showcasing song prevalence per chord progression, revealing dominant patterns, primarily the 1,5,6,4,1 sequence and their variations. An analysis of song keys highlighted C, A, and C#/Db as predominant (refer to Appendix 2). The metric of valence, ranging from 0 to 1, represents musical positiveness and can be accessed via the Spotify API. Filtering for chord progressions with more than five counts showed significant variations in valence across them and keys. Remarkably, V-commencing progressions and songs in the keys of A and G#/Ab had a higher valence score (see Appendix 3 and 4). Additionally, the correlation matrix (refer to attachment 5) showcased a direct positive relationship between energy and loudness, juxtaposed by a negative correlation between acousticness and energy. From the genre frequency data (refer to Appendix 6), pop and punk emerged as the most prevalent genres, suggesting popular taste leans towards these styles. This preliminary analysis provides profound insights into song structures, emotional resonance, and audience preferences.

6. Unsupervised learning

6.1. Bag of words method

After cleansing the lyrics with regular expressions, converting to lowercase, tokenizing, and lemmatizing, we obtain a purified corpus, free from punctuation, numbers, and stop words. This refined data set serves as the basis for generating a word cloud, where the prominence of each word directly corresponds to its frequency within the lyrics. The word cloud (See Appendix 7), a vibrant visual depiction, instantly conveys the most dominant themes and expressions in the dataset, setting the stage for deeper analysis. In this word cloud, we can see words generally related to the description of personal feelings like ‘love’, ‘cause’, ‘wannn’, ‘kiss’ and ‘go’ are prevalent in the lyrics.

The project advances by establishing a document frequency list and a count document-term matrix (DTM), which facilitate a granular understanding of the textual data. Additionally, we juxtapose various forms of the term frequency-inverse document frequency (TF-IDF) matrix: the standard form, the smooth form, and the L2 normalized form. To discern inherent structures within the text, the project employs K-Nearest Neighbors (KNN) clustering on the TF-IDF matrix, grouping documents by similarity. The optimization of cluster quantity is informed by the silhouette score, a metric indicative of the fit of data points within their assigned clusters relative to neighboring clusters. The silhouette analysis graph (See Appendix 8) guides the selection of $K=20$, a point beyond which the score stabilizes, suggesting a balance between cluster cohesion and separation. This graph is a crucial piece of the analytical puzzle, as it underpins the decision-making process for the number of clusters to use. With $K=20$ clusters established, the project generates individual word clouds for each cluster, visually representing the most frequent and characteristic terms within each group (See Appendix 9). Additionally, the project explores hierarchical clustering, presenting an alternative word cloud visualization (See Appendix 10) for each cluster identified through this method.

6.2. Sentimental analysis and LDA for each sentiment class

Integrating the rule-based processing approach, we observe the sentiment class frequencies as dictated by the Vader lexicon scoring. This visualization serves as a testament to the sentiment distribution within our corpus. The bar graph elucidates a pronounced inclination towards positive sentiment, with the highest frequency observed in the tenth sentiment class (See Appendix 11). This reveals a trend of positivity prevailing within the lyrics, a pattern that could potentially be attributed to the nature of the content or the predisposition of the users. Meanwhile, the presence of a moderate number of lyrics in the middle spectrum underscores a state of neutrality. The lower sentiment classes, though less represented, are crucial for a balanced analysis, as they highlight the instances of negative sentiment.

Upon document-term matrix (DTM), we fit the LDA model, specifying the desired number of topics to discern latent thematic structures that pervade the lyrics within each sentiment class. The LDA model unveils the distribution of topics, which we then distill into the most significant words per topic, crafting a list that embodies the essence of each thematic cluster. Specifically for sentiment class 10, which represents the most positive lyrics, the LDA analysis reveals topics abundant with positively connotated words. The word clouds

blossom with optimism, showcasing the uplifting language that characterizes this segment of our lyrical universe (See Appendix 12). The stark contrast in topic composition between this class and others may underscore the diverse emotional spectrum captured by different sentiment classes, illuminating the multifaceted nature of sentiment in lyrics.

6.3 Word2Vec and KNN

The raw lyrical content is first subjected to a cleaning process that strips it of extraneous characters and unnecessary punctuation, thus distilling the text down to its most basic and meaningful form. Words and phrases are then tokenized—converted into discrete, analyzable components—while a careful removal of stop words, digits, and brief tokens ensures that only significant elements are retained for analysis. Further refining the textual data, we enrich our preprocessing with a custom set of stop words.

With the lyrics cleansed and curated, we proceed to the vectorization stage, wherein the Word2Vec model transforms the processed text into numerical representations—vectors that encapsulate the semantic properties of the words. (See Appendix 14) These vectors are the key to unlocking the patterns within the lyrics, as they serve as inputs to the clustering algorithm. By applying a K-means clustering approach, the lyrics are grouped into distinct sentiment classes, each cluster representing a different emotional nuance as reflected in the lyrics. The silhouette score graph shows that average score is higher than the bag of words method. And here we choose $k = 5$ for the further study. Then we also show the top terms for each cluster (See Appendix 16), however, this project found that there are highly frequently overlapping words in each cluster.

6.4. Supervised Learning

The predictive power of these song features, including chord progression, is then scrutinized through supervised learning algorithms. By framing the clusters as the target variable, we aim to understand whether these diverse song features can accurately predict the sentiment-driven clusters previously identified. The features, spanning from danceability to chord progression, form the independent variables 'X' in our modeling effort. To prepare for the application of machine learning techniques, all categorical variables within 'X' are transformed into numerical form using label encoding. This transformation is crucial as it enables the models to process the categorical data effectively. Once the data is prepped, we fit it to both a Decision Tree and a Random Forest model (See Appendix 17 and 18) These models serve as our analytical lenses, providing insights into the relationships between song features and lyrical sentiment clusters. Initial validation scores suggest that the models, while not yet optimal, do reveal the chord progression ('cp') as a standout predictor. This feature's importance indicates a potential correlation between the harmonic structure of a song and the sentiment of its lyrics.

In conclusion, the integration of machine learning techniques with lyrical and song feature analysis paves the way for fascinating insights into the sentiments conveyed by music. The iterative process of model validation and refinement promises not only enhanced predictive power but also a richer understanding of the intricate

tapestry that is music sentiment analysis. As we continue to refine our models and interpret their results, we delve deeper into the art of translating the abstract language of music into the concrete terms of data science.

7. Similarity score calculator

The project's next innovative step is the development of a similarity score calculator, a tool that enables a multifaceted comparison of musical acts (See Appendix 19). By constructing radar graphs, this feature visually contrasts two bands that appear similar or the same band across different eras. The comparison spans several dimensions, including lyrical themes identified through topic modeling, revealing shifts or consistencies over time. Musical elements such as chord progression, tempo, and key are analyzed to discern patterns in creative output. Moreover, attributes like danceability and valence provide a holistic assessment of a band's evolution or the distinctiveness of separate bands. This tool not only serves as a novel mechanism for fan engagement, offering a machine-generated perspective on a band's biography but also acts as a valuable resource for festival planners aiming to curate stages that resonate with specific audience preferences. Furthermore, it can substantiate claims of imitation by highlighting which musical aspects are most closely mirrored between different acts.

8. Future work

The ambition is to augment the current system, which allows for the random selection of song indices, to assess similarity based on quantifiable features such as tempo and key, alongside lyrical and chord progression likenesses. Users will be empowered to tweak the significance of each feature, observing the impact on the similarity score. The integration with Spotify's API is anticipated, facilitating direct song input for similarity scoring. This enhancement lays the groundwork for a tailored recommendation algorithm, capable of matching songs based on user-defined preferences and hierarchical clustering techniques. Such a tool would be invaluable to record labels seeking artists with a particular sound, to musicians in search of inspiration, or to anyone interested in exploring music with specific characteristics. Beyond individual utility, this tool could contribute to AI's creative process, supplying a curated content pool for the generation of targeted musical compositions, such as a bespoke post-punk song, by analyzing and synthesizing commonalities in lyrics and other song features.

(Word Count: 2495)

Bibliography

Choi, J., Song, J. H., & Kim, Y. (2018, June). An analysis of music lyrics by measuring the distance of emotion and sentiment. In 2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) (pp. 176-181). IEEE.

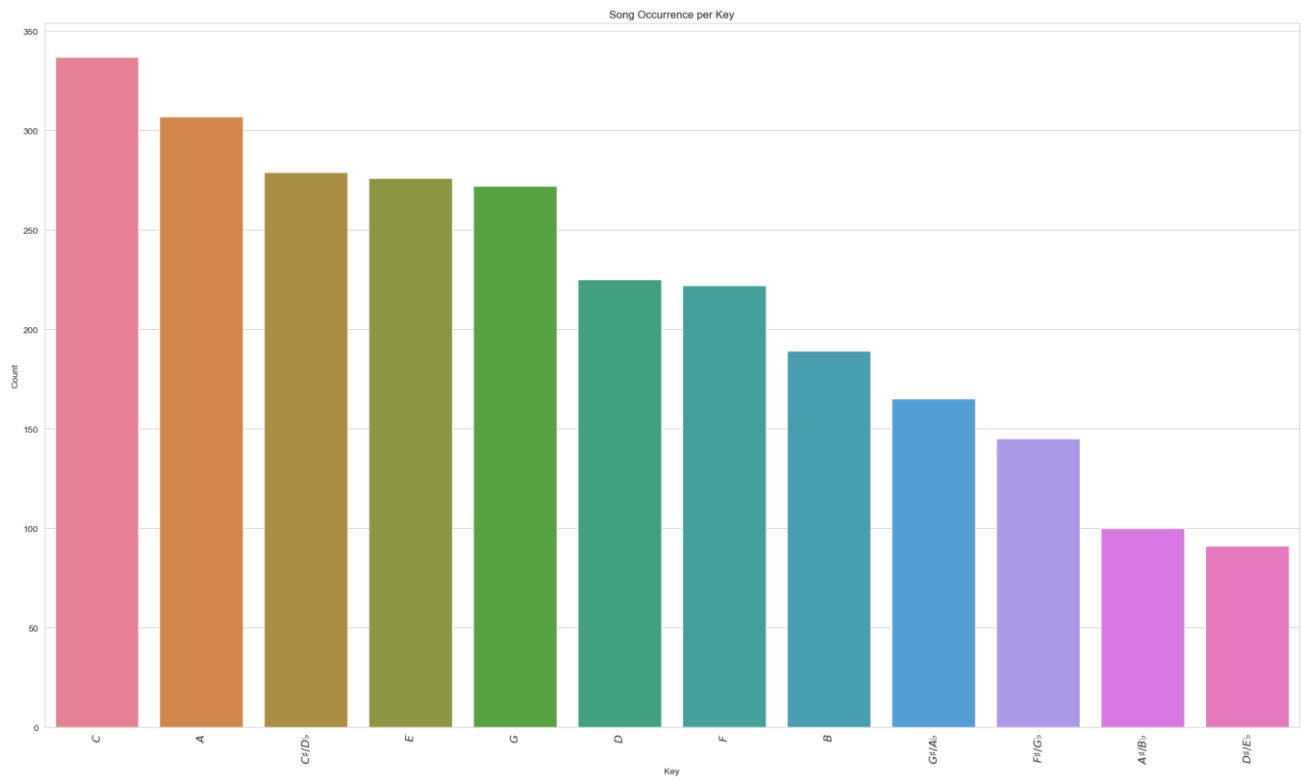
Hamer, J. (2019). Chord Progression Assistant. Retrieved from https://github.com/jhamer90811/chord_progression_assistant

Kolchinsky, A., Dhande, N., Park, K., & Ahn, Y. Y. (2017). The minor fall, the major lift: inferring emotional valence of musical chords through lyrics. *Royal Society open science*, 4(11), 170952.

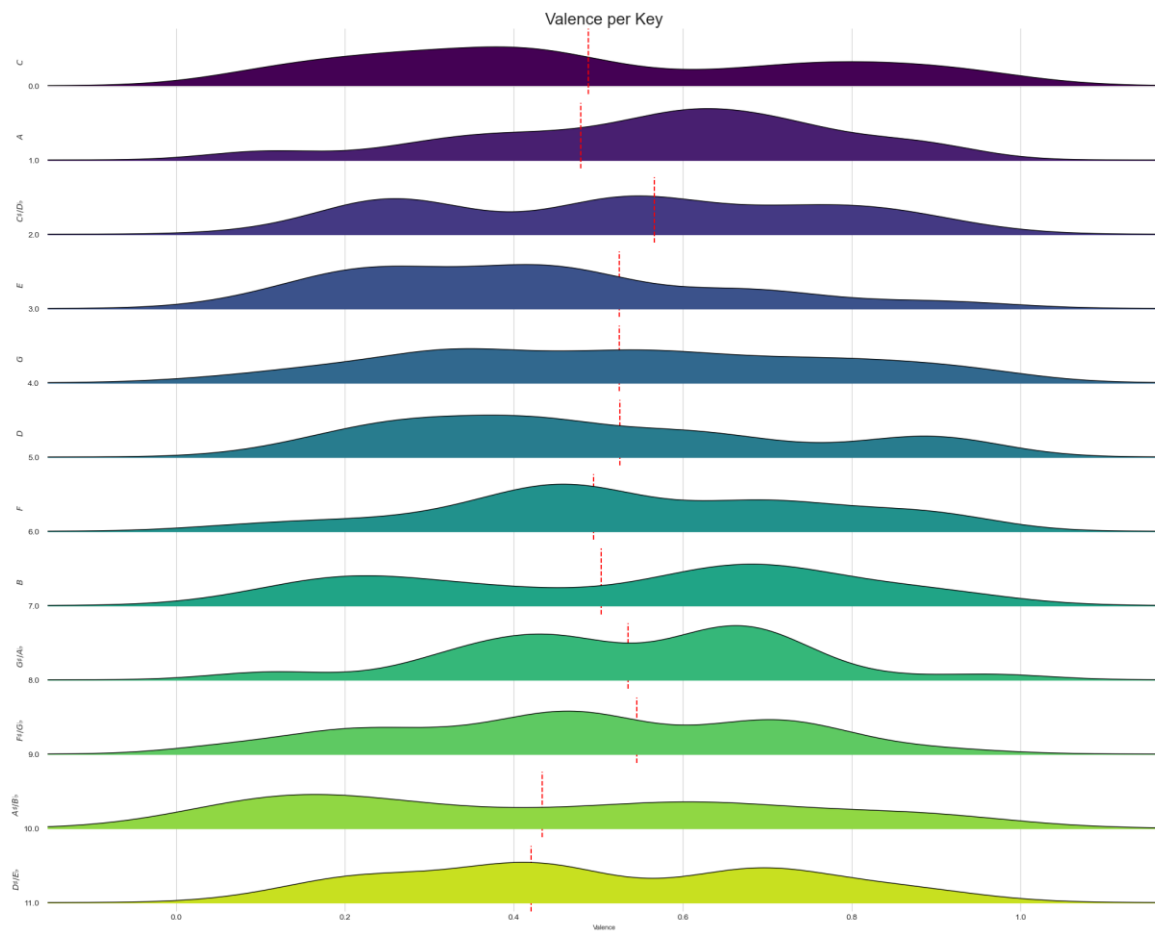
Napier, K., & Shamir, L. (2018). Quantitative sentiment analysis of lyrics in popular music. *Journal of Popular Music Studies*, 30(4), 161-176.

Shrirang, N. (2022). Spotify Million Song Dataset. Retrieved from <https://www.kaggle.com/datasets/notshrirang/spotify-million-song-dataset>

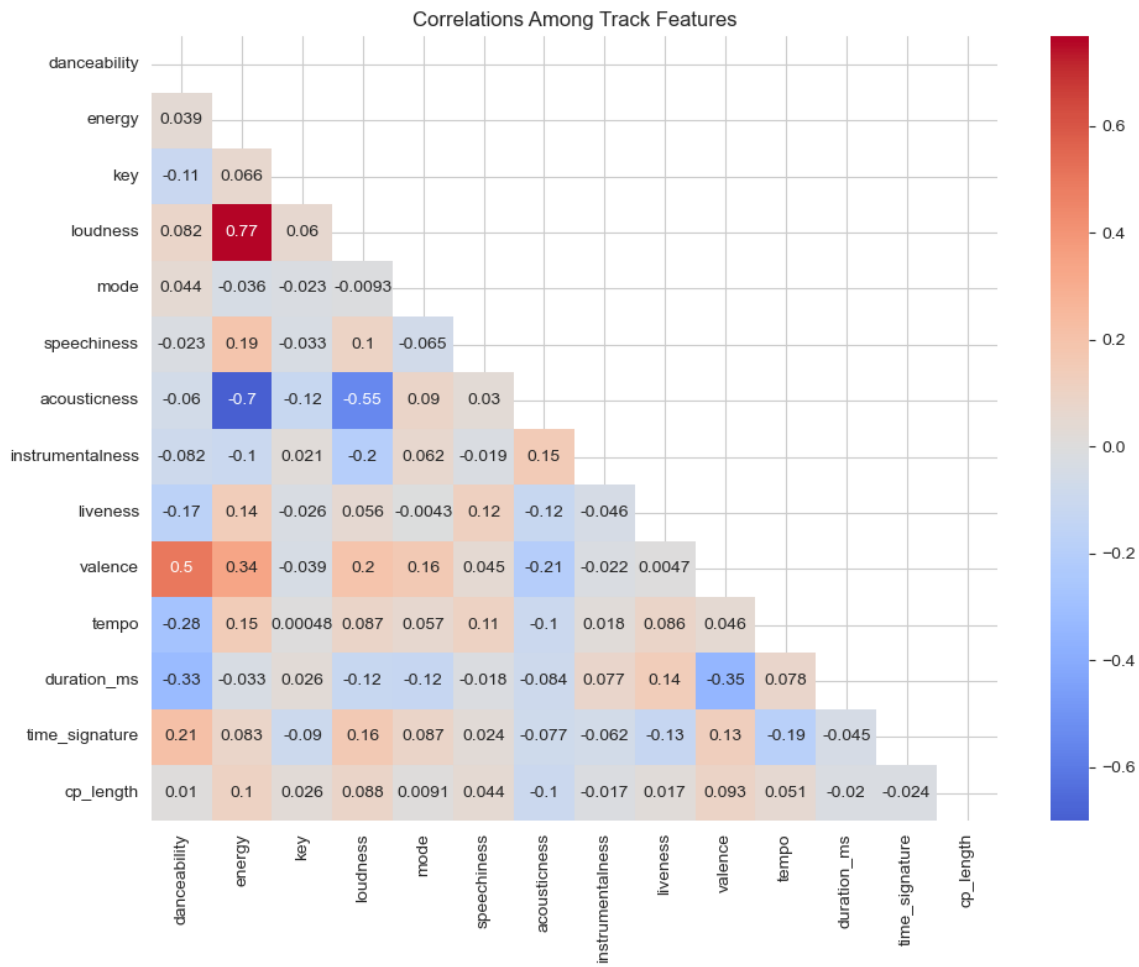
Appendix 3



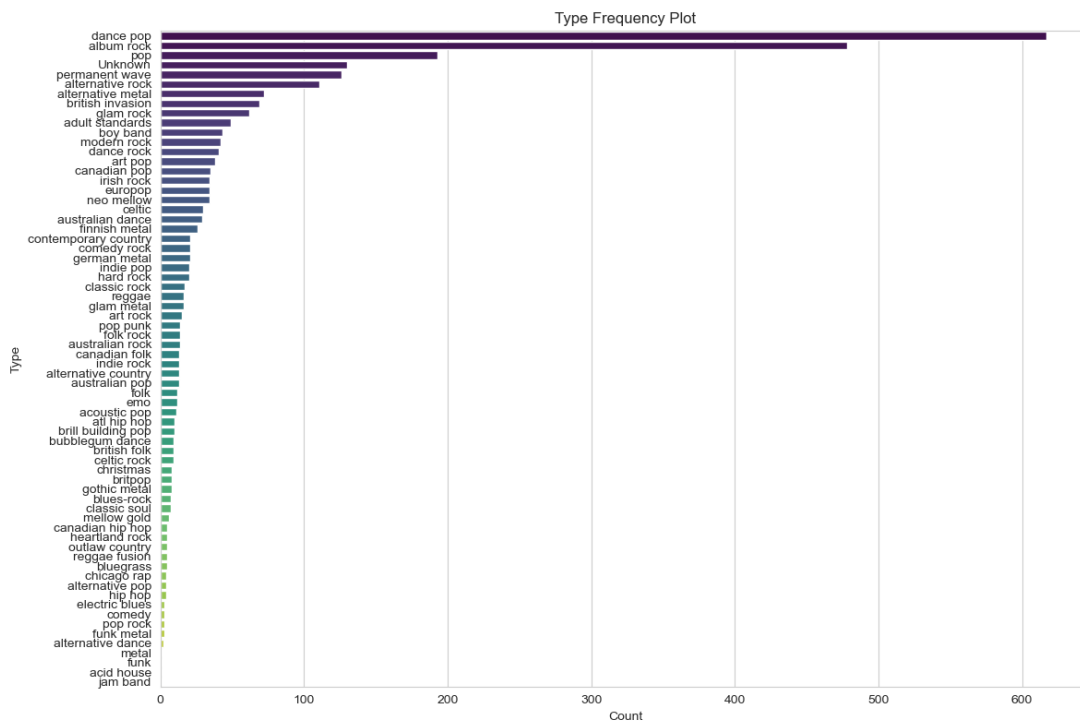
Appendix 4



Appendix 5



Appendix 6



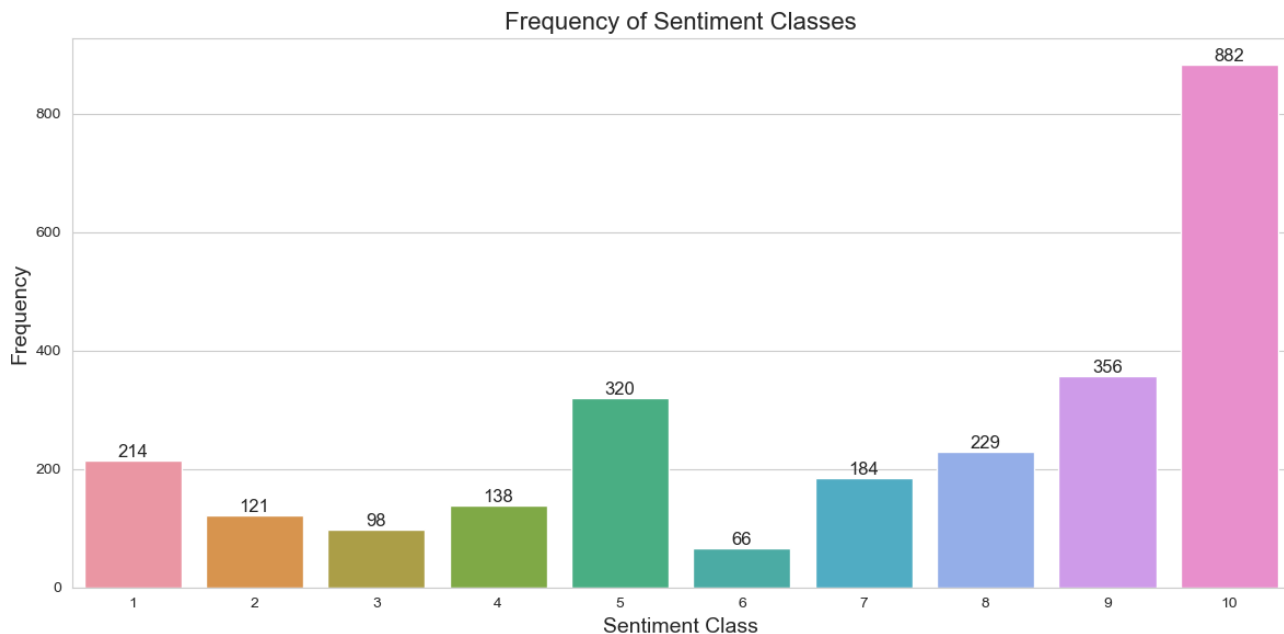
Appendix 9



Appendix 10



Appendix 11



Appendix 12



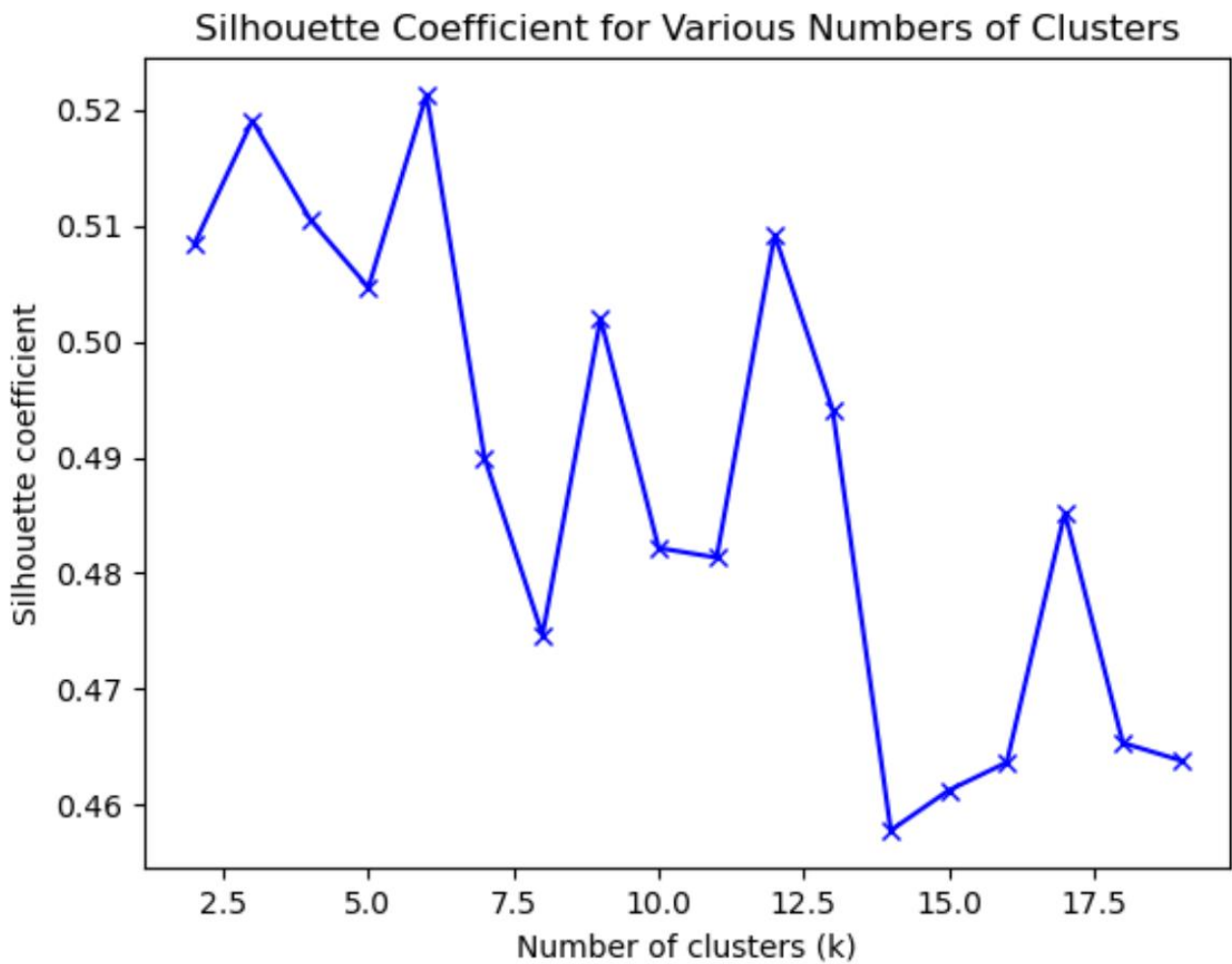
Appendix 13



Appendix 14

```
[('big', 0.9985246062278748),
 ('live', 0.9985194206237793),
 ('shes', 0.9985149502754211),
 ('think', 0.9985094666481018),
 ('much', 0.9985079765319824),
 ('ta', 0.9985046982765198),
 ('face', 0.9985042810440063),
 ('fire', 0.9985026717185974),
 ('man', 0.9985021948814392),
 ('still', 0.9985013008117676)]
```

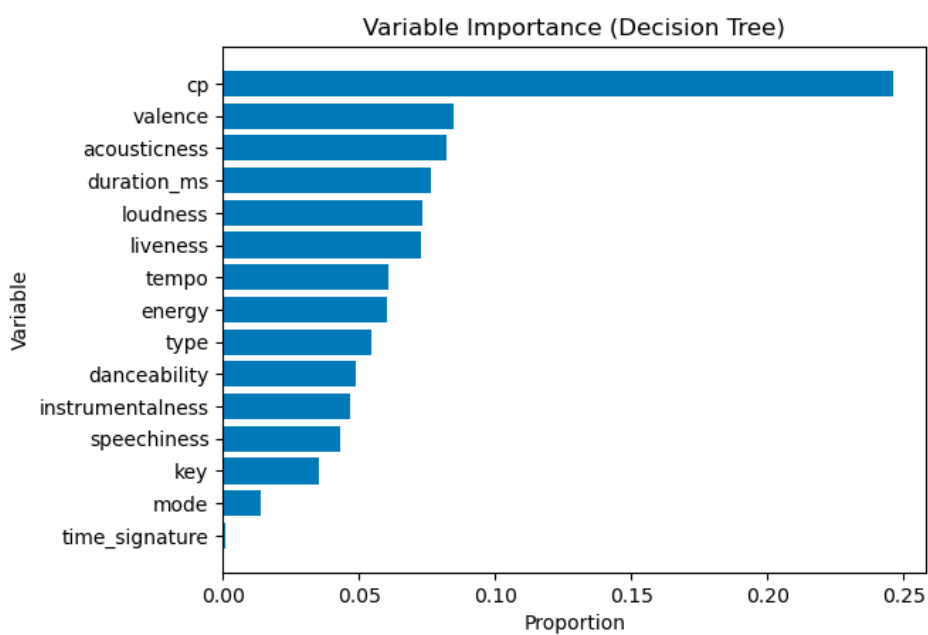

Appendix 15



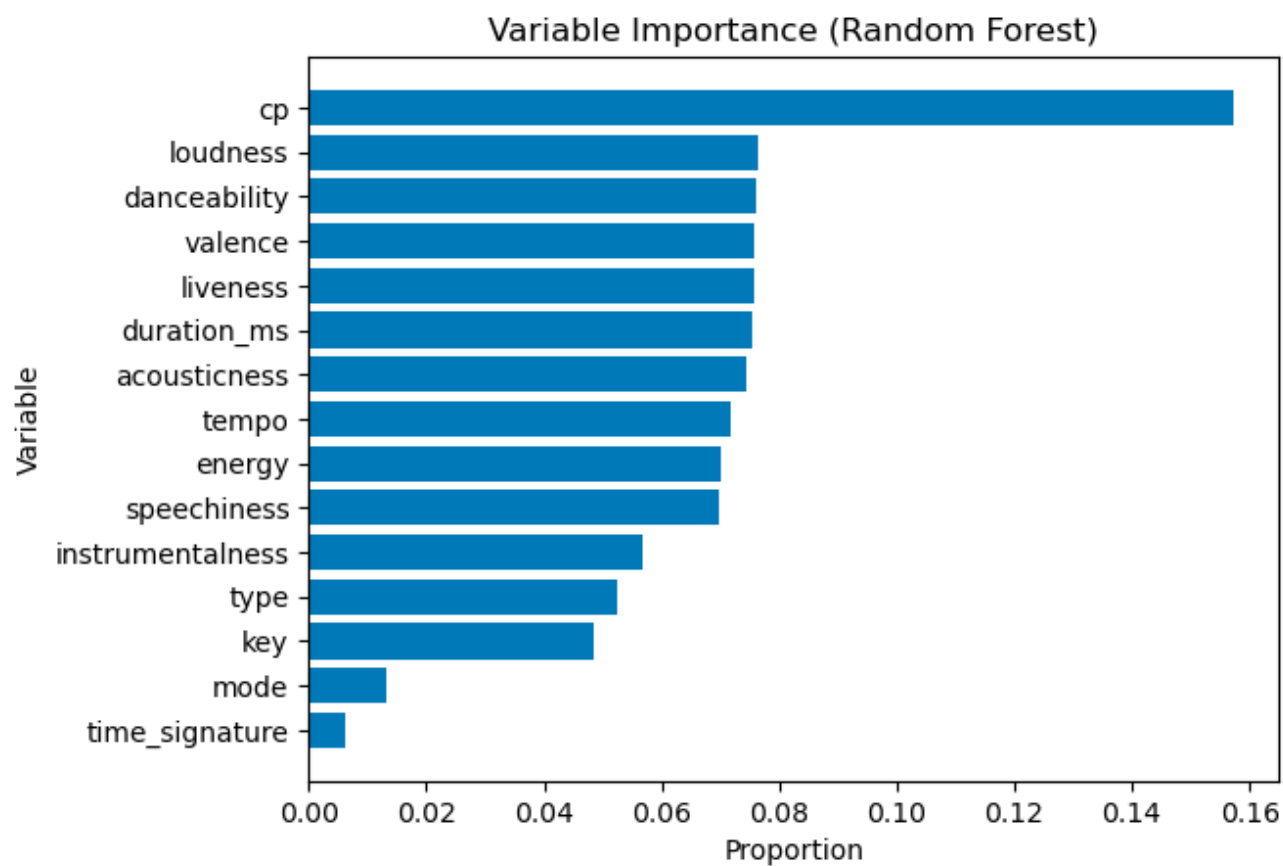
Appendix 16

```
Top terms per cluster (based on centroids):  
Cluster 0: see man get world cause still could one heart ill  
Cluster 1: see get man cause world still could heart one take  
Cluster 2: see tonight everything ill come turn man world little eyes  
Cluster 3: see cause get man heart could world still one ill  
Cluster 4: dream still hes left man see sky song fall would
```

Appendix 17



Appendix 18



Appendix 19

Select Features for Comparison

- danceability x
- energy x
- key x
- loudness x
- speechiness x
- acousticness x
- instrumentalness x
- liveness x
- valence x
- tempo x
- time_signature x

Feature Weights

Lyrics Weight: 0.40

Chord Progression Weight: 0.05

Danceability Weight: 0.05

Energy Weight: 0.05

Key Weight: 0.05

Loudness Weight: 0.05

Song Similarity Calculator

Select Song 1: cryin'

Select Song 2: firework

Calculate Similarity

Similarity Score: 0.5492