SIADS 696 Milestone II Project Report

Tracking Shifts in Historic Recession-Related Sentiments in Music Lyrics to Serve as a Non-Economic Predictor of Recession Time Periods

Contributors: Stephanie Maciejewski, Erin Mettler, Ruchi Patil

Introduction

Recent reports by J.P. Morgan have stated that there is an estimated 40% chance of a global recession beginning in 2025 (J.P. Morgan, 2024) [1]. Economists typically track and predict recession probability using financial indicators including gross domestic product (GDP), unemployment rates, and fluctuations in the stock market. However, non-economic indicators of economic recessions and expansions also exist, and many of these indicators are expressed through cultural and artistic outlets and sentiments. For example, the “lipstick index” proposes that women purchase more lipstick and small cosmetic items during times of economic downturns, rather than opting for purchasing more expensive consumer products (Smith, 2025) [2]. Similarly, there is growing discourse and interest in the prevalence of “recession pop” music during the 2008 financial crisis, and its relevance as an indicator of shifting cultural sentiment and values during this time period.

Using supervised and unsupervised learning techniques, we examined whether clear patterns arise in the lyrics and metadata of music of all genres during and surrounding recession time periods - specifically the years surrounding, and inclusive of, 2008 and 2020 (The Conference Board, 2022) [3]. If these lyrical patterns and sentiment trends exist and can be defined, then this analysis has the potential to include modern-day music as another non-economic indicator of a current or impending recession. While similar projects have analyzed music lyrics to track overall societal sentiment during specific socioeconomic trends, none have included quite as much data as ours has, and none has put forth the efficacy of its analysis as a future recession indicator.

For supervised learning, we utilized Support Vector Machine (SVM) to predict the sentiment of song lyrics. Feature engineering included cleaning, tokenization of song metadata, vectorizing lyrics, and application of hyperparameter tuning to optimize model performance.
For unsupervised learning, we began with Latent Dirichlet Allocation (LDA) to uncover dominant themes in the lyrics of thousands of songs across recession time periods. We also implemented methods including k-means clustering to reveal distinct clusters of song moods, Gaussian mixture models to capture more probabilistic and overlapping cluster memberships, and Density-Based Spatial Clustering of Applications with Noise to identify non-linear cluster structures and outliers and isolate which music genres were most popular during economic downturns.

Main findings:

- Our supervised models were able to classify recession-era music with promising accuracy, but were limited slightly by class imbalance and label uncertainties. Overall, Support Vector Machine (SVM) performed the best due to its ability to handle both linearly separable and non-linearly separable data.
- Our unsupervised models revealed interesting topic clusters and sentiment/emotional trends that indicate strong feelings and sentiments, both positive and negative, in the time periods 2008 and 2020, supporting our hypothesis that there are meaningful shifts in lyrical sentiments during times of global recession.
- The feature engineering, cluster analysis and machine learning all demonstrated the complex and layered nature of music, regardless of expression. To more fully capture the connections between features, a multimodal or layered approach such as a neural network may give more clear insights in the future.

Related work

Our analysis builds off of past sentiment analysis research, as well as a previous passion-project by our team member Ruchi. In her past work, Ruchi was trying to utilize music data to see if anthropological patterns could be revealed from analyzing shifts in music lyrics over time. We have refined this initial project to focus only on recession time periods and their impacts on sentiments found within song lyrics, and this work was heavily inspired by several published papers which we further detail and explain below.

In 2009, a sentiment analysis of musical lyrics was performed to determine if a link can be found between the sentiments found in *Billboard* number 1 songs for the years 1955-2003 and threatening social and economic conditions (Pettijohn, et al., 2009) [4]. This study fo1und that there are more meaningful themes and content found in top songs each year when there are threatening social and economic conditions when compared to years when these conditions do not exist. Our proposed music and lyric sentiment analysis differs because we are looking at *thousands* of songs that were released specifically around recession time periods, to see if this cultural sentiment shift goes deeper than just the music that reaches the number one chart spot (which is often music that is most widely promoted by record labels and music industry giants).

Text-based emotion prediction has many connections to the sentiment analysis of music lyrics we implemented in our project, and this was previously studied in 2005 in the context of children's fairy tales (Alm, et al., 2005) [5]. This study aimed to aid text-to-speech applications in effectively determining if a piece of text holds an emotional tone or not, and did so by utilizing a SNoW learning architecture framework with 30 features to classify passages from 22 popular children's stories as containing specific emotions. Similarly, in our analysis, we identified specific sentiments that were exuded by song lyrics and categorized them into different topics before measuring their prevalence in songs both during and surrounding world recession time periods.

Another example of multimodal sentiment analysis comes from a study in 2018 where researchers developed a convolutional multiple kernel learning (MLK) model to improve upon emotional sentiment recognition across audio, visual, and text-based data sources (Poria, et al., 2018) [6]. This approach exemplified the power of integrating multiple modalities to improve the assessment of what emotions are being expressed. Our analysis focuses solely on text-based features in lyrics, but we share a similar goal to this research of uncovering deeper emotional patterns in expressive content.

Data Sources

For our analysis, we initially looked at 4 music data sources: Discogs (available at: https://www.discogs.com/developers/), Musicbrainz (available at: https://musicbrainz.org/doc/MusicBrainz_API), Acousticbrainz (available at: https://acousticbrainz.org/data), and Genius (available at: https://docs.genius.com/). All are public-access API hosts for musical databases which are updated and added to by users, and the return format for all sources is JSON. Our original API calls returned these results directly, to allow us to further investigate usable features, but we modified our requests to return CSVs containing the features that we determined to be the most useful. Each source contained different information, such as country, genre, style, title, artist (Discogs), album, title, artist, release date (Acousticbrainz), metadata tags, mood, tonality, (Acousticbrainz), and lyrics (Genius). The total number of records retrieved was 2,981,584. In accordance with our hypothesis, we focused our requests to the same years, and primarily downloaded data from these APIs for the years 2007, 2008, 2009, 2020, and 2025. Acousticbrainz was the exception to this, as it required a download of the complete database and data for all years contained within it. Around twenty millions records were pulled, in total, from the Acousticbrainz database. Detailed data schemas for all sources are located in *Appendix B*.

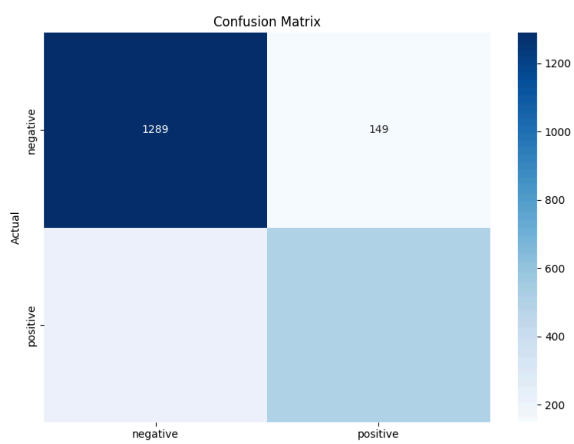
Feature engineering

First, to prepare our API data downloads for unsupervised and supervised learning, we performed numerous pre-processing tasks to combine our Acousicbrainz, Musicbrainz, Discogs, and Genius data sets. We began pre-processing by cleaning our artist and track name columns to enable fuzzy matching of data sets on matching tracks to create a new dataframe. We then used fuzzy string matching via fuzzywuzzy to link records from Discogs, Musicbrainz, and Genius. Next, we dropped entries that were missing lyrics or important metadata (including genre) from our dataframe. Finally, we performed language filtering to ensure that we only had songs with primarily English lyrics in order to simplify our lyric sentiment analysis and visualizations. The data that was pulled from Acousticbrainz, had more than 300 columns, which were dropped according to their relevance to a data analysis done through unsupervised modelling, leaving around 60 columns, of which the probabilistic numerical feature columns were mainly used for the analysis. The records were filtered according to the years of importance, which were the three recessions post 2000, and the window around them, leaving around 10-20 thousand records in each of the 20 csv extracts, which would be further sampled for the study. After manual inspection of the returns for basic requests, we identified several potentially useful features from each dataset (see *Appendix B*). Overlapping data (title, artist, etc) enabled our fuzzy matching and merging, but many more features were unique to each database.

Part A. Supervised Learning

Support Vector Machine (SVM)

Support Vector Machines (SVM) are supervised machine learning algorithms that are used for both classification and regression tasks. These algorithms are especially useful for finding the optimal hyperplane that separates data points into different classes while also maximizing the margins between them. Due to this, SVMs are well suited for high dimensional data sets, like the lyrics and music metadata datasets we are working with, and have been successfully applied to prior sentiment analysis research including Du's study on sentiment analysis in music lyrics (Du, 2024)[9]. Given our datasets complexity and the presence within it of linearly and non linearly separable features, SVM was the ideal choice for classification.




```
Training set size: 8643
Test set size: 2161
Training SVM model...
Test Accuracy: 0.8468

Classification Report:
              precision    recall  f1-score   support

    negative       0.85      0.94      0.89      1438
    positive       0.85      0.66      0.74       723

    accuracy                           0.85      2161
   macro avg       0.85      0.80      0.82      2161
weighted avg       0.85      0.85      0.84      2161
```

Figure 1: Confusion matrix and Classification Metrics for SVM Model Predictions of Recession-Era song Sentiment

Sentiment Analysis

Sentiment analyzers use natural language processing, text analysis, and computational linguistics, to extract, quantify subjective information. Sentiment analysis is widely applied to voice of the customer materials such as: product reviews, social media posts, and marketing materials. Certain models, such as

RoBERTa, allow more difficult data domains to be analyzed, e.g., song lyrics. The Cardiff NLP sentiment model with Twitter Roberta base was compared to a previous Distilbert base. Both models operate as binary classifiers, leading to single labels (1: positive, 0:neutral, and -1: negative), which allows for quantitative evaluation in comparison to emotion-labeling classifiers. Initial use of Distilbert model lead to more balanced proportions of negative & positive sentiment scores (negative: 7191, positive: 3613), which was part of the 'rbf' factored SVM, with a final accuracy of 0.85. However, use of the Cardiff model produced different sentiment scores (neutral: 6902, negative: 2504, positive: 1404), that when used for the SVM model produced a final F1 of 0.93.

Logistic Regression

A probabilistic linear classifier that provides interpretable coefficients and probability estimates, this method was chosen to provide a simple binary classification of the data, based on the sentiment of the lyrics.

Supervised Evaluation

Failure Analysis of Support Vector Machine

To test the model's effectiveness, 3 song samples were chosen from the data subset 'recession_artists', which contains the cleaned and processed lyrics from major releases in our selected year from previously identified pop recession artists. In all 3 examples (the songs 'TiK ToK by Kesha, 'Waking Up In Vegas' by Katy Perry, and 'Just Dance' by Lady Gaga), the model incorrectly identified the song as 'negative'. The initial output of the model structure showed that there was a distinct imbalance in the data, as noted by the sentiment scores ('negative': 7191, 'positive': 3613). LIME analysis results show that the data imbalance may have led to overfitting, resulting in an initially misleading high accuracy (0.8468) result. This kind of systematic error could be corrected with SMOTE application or possibly adjusting the threshold sensitivity to improve the recall/precision balance. Although LIME was used primarily for interpretability, it provided indirect sensitivity analysis by revealing feature reliance across examples. For instance, misclassifications were often tied to overemphasis on single emotionally ambiguous words (e.g., "cry").

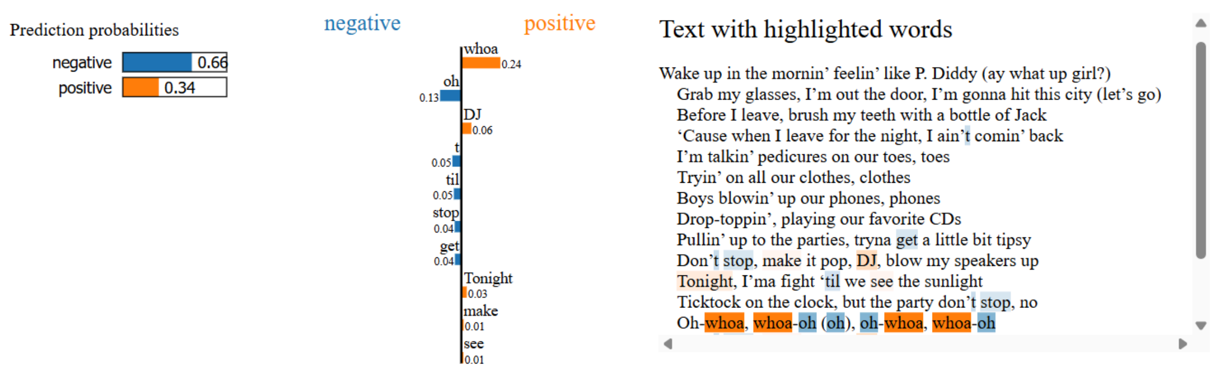


Figure 2: Confusion Matrix and Metrics Revealing Imbalances in SVM Sentiment Classification Performance

Feature Importance & Ablation Analysis of Logistic Regression

Using the learned weights from Logistic Regression, we identified the top positive and negative features contributing to sentiment predictions, (love +1.43, party +1.27, hate –1.31, alone –1.18). To gain insight into which features contributed most to classification accuracy, we conducted an ablation analysis, evaluating performance after systematically removing key feature groups and comparing the resulting accuracy to the full model baseline.

Removed Feature Group	# Features Removed	CV Accuracy	Accuracy Drop
Baseline (0)	0	0.8863	0
Unigrams	3140	0.8191	-0.0672
Bigrams	1860	0.8720	-0.0143
(+) Word list	6	0.8738	-0.0125
(-) Word list	4	0.8863	~0.00

Table : *Ablation failure analysis of Logistic Regression classifier*

Part B. Unsupervised Learning

Unsupervised Learning Workflow and Methodology

To explore underlying structure in song lyrics without labeled data, we implemented K-means clustering (a centroid-based, non-probabilistic method) and Latent Dirichlet Allocation (LDA) (a probabilistic, generative topic model). These approaches were chosen to complement each other and provide different perspectives on the latent structure of the lyrics dataset.

The lyrics were preprocessed with standard NLP techniques (lowercasing, tokenization, stopword removal). For K-means, lyrics were transformed into high-dimensional vectors using TF-IDF (Term Frequency-Inverse Document Frequency), to balance word frequency with inverse document commonality. K-means was selected for its simplicity and effectiveness in high-dimensional vector spaces like TF-IDF. It provides a direct way to group documents based on vector similarity. This representation is effective for K-means, as it captures surface-level semantic information while maintaining sparse, interpretable vectors. Dimensionality was reduced via Principal Component Analysis (PCA) for visualization, although clustering was performed in full TF-IDF space.
K-means was used to partition the lyrics into k distinct groups based on cosine similarity in TF-IDF space. We evaluated k from 2 to 10. To identify the optimal number of clusters, we applied:

- Elbow Method: Analyzed inertia (within-cluster sum of squares) to identify the point of diminishing returns.

- Silhouette Score Analysis: Measured average intra-cluster cohesion vs. inter-cluster separation.

Both methods suggested k=3 as the optimal value, though the maximum silhouette score (~0.08) indicated weak structure, suggesting that lyrics vary along a continuum more than discrete categories. For the feature based unsupervised modelling, 2 main aspects of clustering were considered; algorithm and sample sized. K-Means would focus on the even distribution of groups, defining the description of each cluster (general song style). A Gaussian mixture model would be expected to focus on how each cluster's behavior changes over the year, giving year based comparisons that may yield observations.DBSCAN tends to suit data with noise, so the success would determine that our data had a lot of inconsistency. Sample sizes to be considered were 100, 1000, 10 000. The Acoustic Brainz data had to have its dimensionality reduced; in this case, a reduction was favorable, so around 30 numerical probability based feature columns were to be considered; no encoding had to be done. In total 20 CSV files were pulled from the Acoustic Brains, from a hash sharded set of zipped files of json data, in total resulting in information about around 20 million songs, from which the years around the last 3 recessions (2001, 2008, 2020) were filtered in to be analysed.

Hyperparameter Tuning and Exploration

The values of k were varied and evaluated via inertia and silhouette score, to guide selection of the final model. For LDA, a bag-of-words representation was used, and the data converted into a document-term matrix suitable for probabilistic topic modeling. This was to allow the model to infer abstract thematic distributions rather than surface similarity. LDA was used to uncover latent themes or topics across the corpus. We experimented with topic counts ranging from 2 to 20, evaluating model coherence using the UMass coherence score. An 8-topic model produced the most coherent and interpretable results. LDA offered valuable semantic groupings (e.g., themes around love, violence, or partying), complementing the less interpretable K-means clusters.Provides a probabilistic, semantic layer of interpretation not captured by K-means. It infers latent structure in terms of thematic composition, which is essential for text data. The combination of a geometric clustering method and a probabilistic topic model ensures a diverse analysis of both form and content in the lyrics. The variation in the number of topics and an examination of both topic coherence and manual inspection of top words per topic for interpretability helped determine the final model.

In the feature based hyperparameter tuning, the method of clustering and sample sizes were the main hyperparameters around the clustering that were considered. It was expected that a 100 sample size will provide a status check in the pipeline, serving as a quicker method to debug in iterations. A 10 000 sample size was expected to provide the general story over the years, without stressing the computational resources, while a 1000 sample size would be providing a magnification of the general story and help pinpoint the insightful observations.

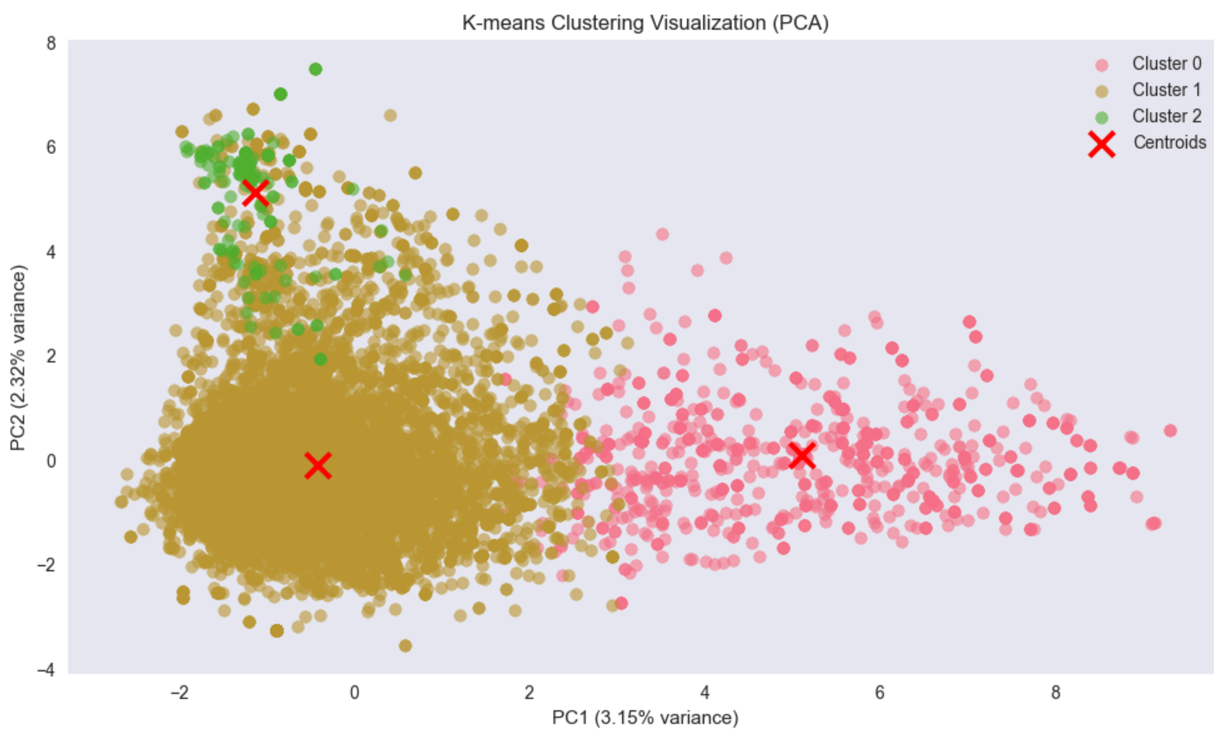


Figure 3: *PCA Visualization of K-Means Clustering Results*

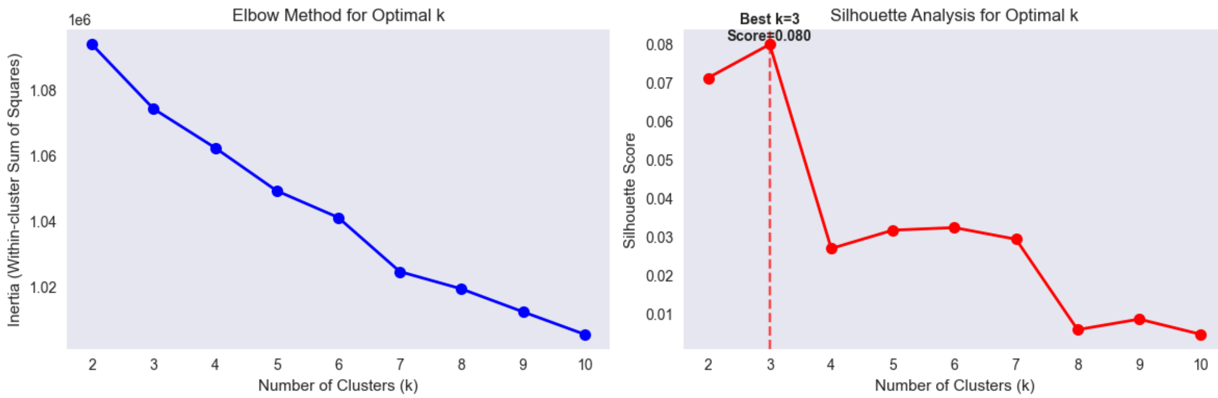


Figure 4*: Elbow and Silhouette Evaluation for Optimal Number of Clusters (k)*

The two principal components chart in Figure 3 shows three clusters with some structure, although not as spherical or clearly defined as desired. Cluster 0 in pink is relatively well-separated to the right, while Cluster 1 in Brown is the most populous and spread out, overlapping both other clusters. Cluster 2 in green is more compact and clearly defined, in the top-left. The centroid positions align with the visual centers of the point groupings, which is expected. Cluster 1 is very dense and overlaps others, indicating many lyrics may be thematically or semantically ambiguous or diverse. Cluster 2 is the most distinct, possibly representing a niche or stylistically unique group of songs (e.g., extreme sentiment, genre, or vocabulary). The elbow plot in Figure 4 again shows a lack of strong, obvious natural clusters; this aligns with the earlier struggle to confine genres to a single component, as music is often multi-layered and drawing from other themes. The peak silhouette score shown in Figure 4 at k=3 with a Silhouette Score of ~0.08 is very low, suggesting again that the clustering structure is weak, especially since the scores drop off dramatically beyond k=3 and remain low.

Unsupervised Evaluation

Failure Analysis of Latent Dirichlet Allocation Model

As is common with these models, the perplexity and coherence scores were obtained as performance metrics of the model. The extremely high perplexity (420) suggests poor predictive performance, as does the lower log-likelihood (-750780). The top topics identified also showed incoherence and non-informative words. However, these issues, along with the concept of repetition, are to be expected when dealing with song lyrics, as many newer songs (particularly of the ‘recession pop’ variety) contain these same traits.

Failure Analysis of Feature Based Clustering
Several constraints were experienced when performing the unsupervised-based analysis of the song features. The lack of supervised labels (play count, listener count, other success metrics) limited the study to unsupervised clustering. Attempts to fetch success metrics from Last.fm were slowed down by strict API rate limits (~5 requests/sec), resulting in large-scale data pulls being impractical within the time frame of the overall study. Computational costs also restricted scalability; higher-core VMs on clouds such as Azure were exceeding budget constraints. Thread-based concurrent code helped to mitigate the computational and time related issues.

Discussion

The main challenge that we encountered while completing our analysis was related to the massive amount of data that was needed to effectively find insights via supervised learning. Initially, we found that downloading this data via API calls was extremely time-consuming and computationally expensive and would be infeasible without employing a high performance computing (HPC) cluster to assist us. We utilized the Great Lakes Cluster available through the University of Michigan IT services, which helped us a lot with overcoming this challenge, but still was a very time-consuming process. This challenge was again faced when trying to combine the output data and perform data pre-processing and feature analysis tasks, as the large data output files output by the four separate APIs needed to be combined and evaluated together. Given these limitations with dataset merging, the Acousticbrainz song metadata was analyzed separately from other lyrical data found in the Musicbrainz data files.

With more time and computational resources, we would be able to greatly expand the accuracy of our analysis and gain additional insights into recession-era trends and patterns in emotion and sentiment that are expressed through artistic outlets such as music lyrics. Additionally, more time and computational resources would allow the team to begin looking outside of English-only lyrics and see if similar trends in sentiment are faced by other countries and cultures during global recessions.

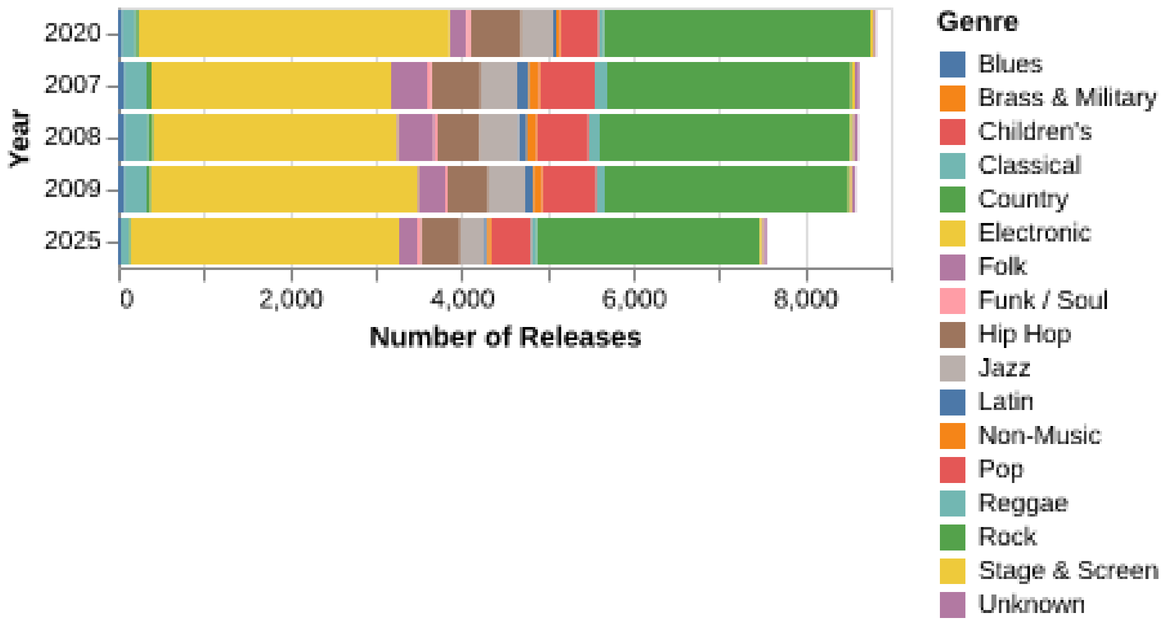


Figure 5: *Distribution of Total Music Releases by Genre in Selected Years (2007, 2008, 2009, 2020, 2025)*

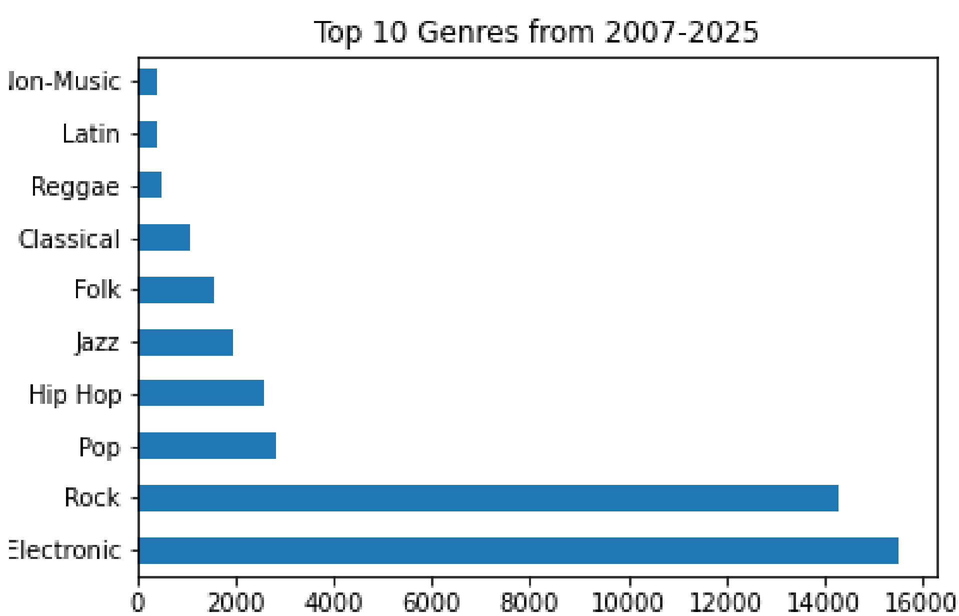


Figure 6*: Top 10 Music Genres of Tracks Released from 2007-2025*

As can be seen in Figure 5, approximately the same number of songs were released per year; 2025 showing less due to the year being in progress. The largest percentage of songs for each year are ‘Rock’ and ‘Electronic’. This does not necessarily correlate with popularity, but rather suggests that these songs are widely produced. In the case of ‘Electronic’ music, this may be due to the ease with which anyone can produce the music. It is also possible that the large number of ‘Rock’ and ‘Electronic’ songs is due to the genres being catch-all terms for less popular sub-genres such as “pirate metal”, which is often included under the ‘Rock’ umbrella, and “Vaporwave”, which is included under the ‘Electronic’ umbrella. Figure 6 shows that across all years collected, the number of ‘Rock’ and ‘Electronic’ songs is far greater than any other genre, by over 500%.

Figure 7: *Word clouds of most prominent topics by year.*

As can be seen in the topic word clouds in Figure 7, there are several notable thematic patterns that can be found in song lyrics for songs released in the years during and surrounding recession time periods.

Words including “love”, “like”, “heart”, and “feel” appeared in the word clouds for numerous years, confirming that songs and their lyrics act as an outlet for the emotions of songwriters and music listeners alike. The ‘Recession Artists’ and ‘All Years’ word clouds provide further insights into the sentiment differences between Recession-specific years and non-recession years. In the word clouds from Recession Artists, there are a lot of topics indicating longing and desire, including “want”, “get”, and “like” which support the theory that recession pop contains sentiments and emotional expression for longing for more at a higher rate than is seen in the aggregated “All years” word clouds where the topics appear to be words that come up more in natural conversation regardless of the underlying emotion or subject matter.

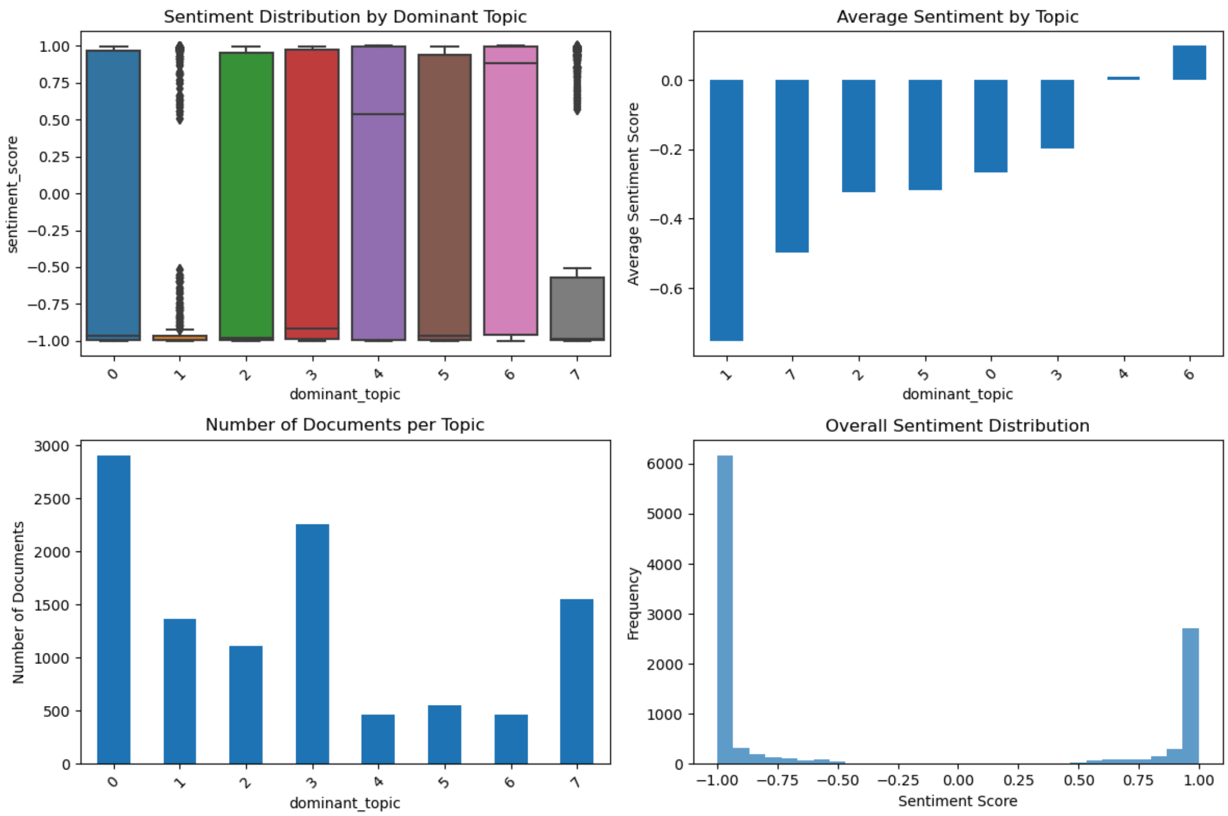


Figure 8: *Topic Modeling and Sentiment Analysis Across Lyrics:* ***(a)*** *Sentiment Distribution by Dominant Topic,* ***(b)*** *Average Sentiment Score by Topic,* ***(c)*** *Number of Songs by Topic, and (****d****) Overall Sentiment Score Distribution*

As shown in Figure 8(a) all dominant topics, except topics 1 and 7, show usage in songs with all rankings of sentiment, suggesting that they are words commonly used in lyrics, as well as likely more generally used words in conversation. Figure 8(b) exemplifies that most topics, except 4 and 6, are predominantly negative in sentiment; this suggests an overall dissatisfaction that people are expressing through music (across all 5 recession and recession-adjacent time periods). Figure 8(d) shows that, in general, the sentiment of nearly all songs is either strongly negative or strongly positive, with very few songs that are neutral in emotion. This is expected, as music is an artistic medium, and therefore naturally associated with intense emotional states that promote expression through the medium, but also could potentially be linked to heightened emotional experiences being expressed during times of economic downturn and recession. Even for much smaller genres, such as military, non-song, and children’s music, there is still emotional connection and expressiveness.

Feature Based Clustering

In the clustering analysis of the feature based track data from AcousticBrainz, the time window spanned 1999-2020, including the boundary years for each recession year in the last two decades (2001, 2008, 2020). K-Means and gaussian mixture models were run on thousand and ten thousand random samples from the year filtered dataset. The feature set included probabilistic estimates of mood, genre, timbre, tonality, and voice instrumentation.

The trend most pertinent to our study appeared in the behavior of cluster 0, which is composed of high-energy, party electronic music. The counts for this group seem to be dropping during the recession years in question, 2001, 2008, 2020; this is most observable in figure 9(c). This feature analysis was focused on the behaviour of track-releasing artists, as the track counts were being observed. So the dips in high energy party music aligned with economic downfalls, may be suggesting that artists were pulling away from creating clubbing songs. THese dips might have reflected the subjectively sober temperaments of the public, or clubbing activity plummeting due to club closures of clubbing affordability.

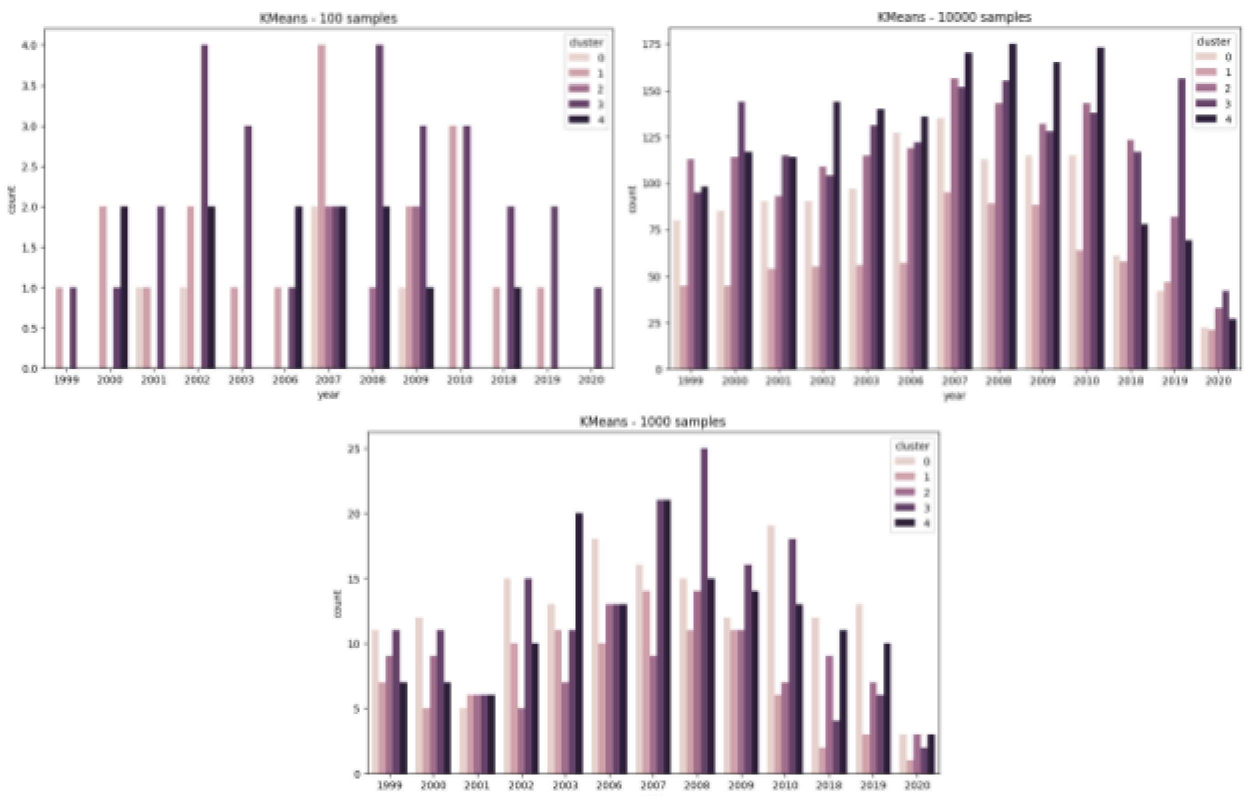


Figure 9*: K-Means Cluster on Probability Based Features: (a) One hundred Samples, (b) Ten Thousand Samples, (c) One Thousand Samples*

Cluster 2, was defined by its high acoustics and very low aggressive moods, suggestive of soft, introspective folk or indie songs. This group seems to maintain consistency in how much it dominates the market, which is inline with the ‘niche’ public image, indie music often garners. It can be observed, especially in figure 10(b), that this sombre indie cluster is significantly peaking in 2008, and then dipping. This suggests to us, many artists, may have sought out independent studios as opposed to major record labels as the economy worsened, and they may have also opted for unplugged and acoustics productions as a cheaper option. It is interesting to consider, we do not see a similar uprising as we approach the 2020 recession. It may be worth investigating how quick, voluminous TikTok or Reel dominated content that dictates song virality might have been influencing what artists decide to pursue.

In the GMM 1000 sample, which can be viewed in Figure 10(c), it is very observable that cluster 4 is peaking in 2008 very strongly, contrasting itself from the other groups observed during recessions. This cluster is defined by its strong party, relaxed and electronic moods and genre. This is a more rhythm-centric and emotionally suppressed observed group, leaning away from the euphoric dancing music of cluster 0. It suggests a rising prominence in electronic music that was more integrated into an individual's daily life; we are looking at emotionally toned down music such as synthpop and soft electric such as Bastille and Owl City.

The clustering patterns suggest that artists may shift away from euphoric club tracks, and perhaps, towards introspective, compelling and lower energy tracks, in times of economic crisis. The data around 2008, which has been considered the Great Recession, seems to highlight these patterns very prominently. One thing to consider here is that, the sparsity in data around 2020, due to not being logged enough as the database has died down, may have caused the patterns around 2020 not to appear as strongly as they could. A study could also be conducted in the last decades and current years to identify what factors such as media consumption are impacting music artist behaviour, in order to offset the results for recession patterns going forward.

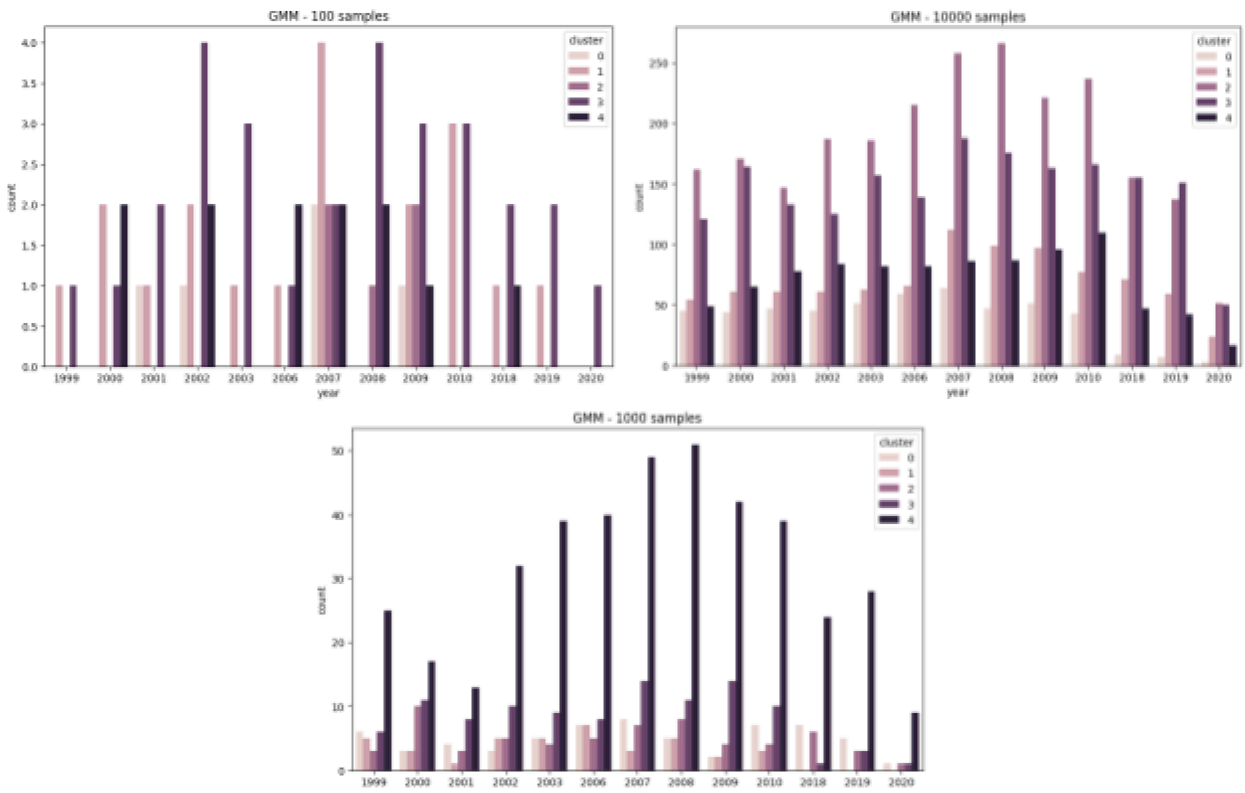


Figure 10: *GMM Cluster on Probability Based Features: (a) One Hundred Samples, (b) Ten Thousand Samples, (c) One Thousand Samples*

K-Means 10k sample AVG means							
cluster	danceability	gender	genre_dortmund	genre_electronic	genre_rosamerica	genre_tzanetakis	ismir04_rhythm
0	0.6	-0.72	0.55	-0.28	0.09	-0.37	-0.24
1	0.45	-0.19	-0.14	0.68	0.42	2.74	0.32
2	-0.08	0.24	0.11	-0.66	0.21	-0.3	-0.2
3	-0.77	0.36	-0.11	-0.26	-0.53	-0.27	-0.03
4	0.22	0.08	-0.4	0.89	0.13	-0.31	0.31
	mood_acoustic	mood_aggressive	mood_electronic	mood_happy	mood_party	mood_relaxed	mood_sad
0	0.5	0.48	0.89	0.2	0.61	-0.16	0.45
1	0.12	0.34	0.08	0.37	0.69	0.63	-0.08
2	0.55	-1.07	-0.31	-0.24	-0.88	-0.31	0.43
3	-1.16	0.06	-0.51	-0.18	-0.53	-0.33	-0.85
4	0.29	0.41	0.06	0.09	0.63	0.54	0.17
	moods_mirex	timbre	tonal_atonal	voice_instrumental			
0	0.9	0.5	0.65	0.3			
1	0.18	-0.1	0.04	0.28			
2	-0.17	-0.21	-0.21	-0.29			
3	-0.6	-0.01	-0.16	-0.2			
4	-0.03	-0.18	-0.21	0.12			

Figure 11: *K-means 10k Sample Cluster Centroids*

Ethical Considerations

We conclude our report with a list of ethical considerations for our analysis, and ways that ethical concerns with our processes can be mitigated in future iterations of our report.

Potential ethical considerations include:

- Limiting to english-only lyrics may limit the view into cultural effects of recession felt by other groups. Additionally, some lyrics may be misinterpreted by the model during the sentiment analysis if they include lyrics from artists of marginalized backgrounds who may have songs in different English dialects, such as African American Vernacular English. This could also have the added unintended effect of skewing our results towards only representing music that is more mainstream or pushed by well-known record labels. These concerns could be addressed in future iterations of this project by using more diverse training data and ensuring that the model has similar performance across various subgroups.
- The sentiment expressed through an artist's lyrics may be misinterpreted by the model, leading to unintended interpretation of what the artist intended to say in their songs. Our analysis relies heavily on sentiment classification models that could have the ability to oversimplify phrases or lyrics that contain more nuance. To mitigate this, future iterations of this work could involve more sentiment analysis using additional tools that have been trained specifically on lyrical inputs, as well as include a human element to the analysis by incorporating artist intent, such as artist notes for songs that are occasionally provided on the Genius website or in the liner notes of physical vinyl records.
- Reducing our genres prior to our clustering analysis potentially raises ethical concerns surrounding misclassification of more specific sub-genres of music. Sub-genres of music are occasionally representative of political and cultural subgroups that are already marginalized, and reducing these sub-genres into broader more mainstream genres can risk the erasure of diversity in artistic expression. To solve this ethical concern, future iterations of this project should preserve sub-genres of music (where feasible and as time and computational power allow).
- Presenting our findings and the results of our analysis also requires attention to ethical considerations, as we need to ensure that we are properly emphasizing that the relationships between lyrical sentiment and recession time periods is correlational and not causational. For example, we need to ensure that our analysis does not suggest that certain song lyrics and topics appearing more frequently means that it is certain that a recession is looming, but rather that they

can be another potential non-economic indicator of a recession. To solve this ethical conundrum, we need to make any public publishing our usage of our analysis clearly state that this is only a potential non-economic indicator and not a deterministic forecasting tool for recessions.

Statement of Work

Stephanie Maciejewski	Erin Mettler	Ruchi Patil
Musicbrainz API Data Pulls, Discogs API Data Pulls, Genius API data Pulls, Data Preprocessing, Lyrics Sentiment Analysis, Data Visualization, Related Work Research, Supervised Evaluation, Failure Analysis, Report Writing	Data cleaning and concatenation, Data preprocessing, Related Work Research, Ethical analysis and research, Data Schema and Notebook Cataloging, Results Synthesis, Report Writing	Initial Project Conceptualization, Acousticbrainz API data pulls, Feature Engineering, Clustering for Unsupervised Learning

Appendix A - References

[1] J.P. Morgan. (2024). *Recession probability: What the odds of a downturn look like now*. https://www.jpmorgan.com/insights/global-research/economy/recession-probability

[2] Smith, J. (2025, March 28). How accurate are recession indicators based on culture? Marketplace. https://www.marketplace.org/story/2025/03/28/recession-indicators-internet-culture-trend

[3] The Conference Board. (2022, September 27). What is a global recession, and what can trigger it? https://www.conference-board.org/publications/what-is-a-global-recession-and-what-can-trigger-it

[4] Pettijohn, T. F., & Sacco, D. F. (2009). The language of lyrics: An analysis of popular Billboard songs across conditions of social and economic threat. Journal of Language and Social Psychology, 28(3), 297–311. https://doi.org/10.1177/0261927X09335259

[5] R. Alm and R. Sproat, "Emotions from text: machine learning for text-based emotion prediction," in Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2005, pp. 579–586.

[6] S. Poria et al., "Convolutional MKL based multimodal emotion recognition and sentiment analysis," in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018, pp. 225–234.

[7] Nattagh, H. (2018, June 15). Hugo-Nattagh/2017-hip-Hop: 2017 Rap Albums' Text Mining and sentiment analysis. GitHub. https://github.com/Hugo-Nattagh/2017-Hip-Hop

[8] Tsandefer, T. (2019, May 14). Tsandefer/dsi_capstone_3: Neural machine translation: Explaining the meaning behind lyrics (DSI final capstone project). GitHub. https://github.com/tsandefer/dsi_capstone_3/tree/master

[9] Du, J. (2024). Sentiment analysis and lyrics theme recognition of music lyrics based on natural language processing. Journal of Electrical Systems, 20(9), 315-321. Retrieved from https://proxy.lib.umich.edu/login?url=https://www.proquest.com/scholarly-journals/sentiment-analysis-lyrics-theme-recognition-music/docview/3081430818/se-2

Appendix B - Data Schema

Data Source: example info for one track API call
https://github.com/ruchithelamp/music/blob/main/MusicBrainz_results

Musicbrainz Data Schema

Column Name	Description	Data Type
title	The title of the track	string
track_length_ms	Length of the track in milliseconds	float
album	Title of the album that the track is on	string
release_date	The date that the track was released	date
artist	The artist who released the track	string
MBID	MusicBrainz unique identifier	string
source	Source of the metadata (Musicbrainz)	string
lyrics	Lyrics of the track (if available)	string
lyrics_found	Whether or not lyrics were successfully found for the track	string

Data Source: example info for one track API call
https://github.com/ruchithelamp/music/blob/main/Discogs_results

Discogs Data Schema

Column Name	Description	Data Type
country	Country where release was published	string
year	Year of release	string
format	Music format type (CD, vinyl, EP, etc.O)	list
label	List of labels/publishers of the music	list
type	Type of item being cataloged	string
genre	Genre of the item	list
style	Musical style of the item	list
id	Unique Discogs release ID	int
barcode	Barcode identifiers list	list
user_data	Dictionary of user's collection	int

master_url	URL to masters release	string
uri	Lyrics of the track (if available)	string
catno	Catalog number as provided by the label	string
title	Title of the release (artist + album/EP title)	string
thumb	URL to thumbnail of cover image	string
cover_image	URL linking to cover image	string
resource_url	API endpoint for these release	string
community	Dictionary with want and have counts	dict
format_quantity	Number of units in the package (i.e. 1 vinyl, 2CDs)	int
formats	List of dictionaries for each format	list

Data Source: example info for one track API call
https://github.com/ruchithelamp/music/blob/main/AcousticBrainz_results

Acousticbrainz Data Schema

Column Name	Description	Data Type
danceability_probability	Probability that the track is danceable	float64
danceability_value	Label interpretation of if the track is danceable or not	string
gender_probability	Probability that the gender of the vocals on the track are male or female	float64
gender_value	Predicted gender of the track vocals	string
genre_dortmund_probability	Probability of the track being classified into the Dortmund genre	float64
genre_dortmund_value	Label from Dortmund genre classification probability	string
genre_electronic_probability	Probability of the track being classified into the electronic sub-genre	float64
genre_electronic_value	Predicted electronic sub-genre label	string
ismir04_rhythm_probability	Probability that the track fits into a ISMIR2004 Genre classification category	float64
ismir04_rhythm_value	Predicted ISMIR2004 Genre classification	string
mood_acoustic_probability	Probability that a track is acoustic or not	float64
Mood_acoustic_value	Categorical label indicating if the track is acoustic	string

mood_aggressive_probability	Probability that a track is “aggressive” or not	float64
mood_aggressive_value	Categorical label indicating if the track is “aggressive”	string
mood_electronic_probability	Probability that a track can be defined as electronic in genre	float64
mood_electronic_value	Categorical label indicating that a track is electronic in genre	string
mood_happy_probability	Probability that a track is “happy” or not	Integer
mood_happy_value	Categorical label indicating if the track is “happy”	string
mood_party_probability	Probability that a track can be defined as a “party” track or not	float64
mood_party_value	Categorical label defining the track as a “party” track	string
mood_relaxed_probability	Probability that the track could be considered to be “relaxed”	float64
mood_relaxed_value	Categorical label indicating if the track is “relaxed”	string
mood_sad_probability	Probability that a track is “sad” or not	float64
mood_sad_value	Categorical label indicating if the track is “sad”	string
timbre_probability	Probability estimate for features that define timbre	float64
timbre_value	Categorical label of the tracks timbre (perceived sound)	string
tonal_atonal_probability	Probability that the track is tonal or atonal	float64
tonal_atonal_value	Label indicating if the track is tonal or atonal	string
voice_instrumental_probability	Probability that a track is purely instrumental	float64
voice_instrumental_value	Label showing if the track is purely instrumental or not	string
metadata_tags_album	Album name associated with the track	string
metadata_tags_artist	Artist name associated with the track	string
metadata_tags_date	Release year	date
metadata_tags_musicbrainz_recordingid	Unique ID from the Musicbrainz database	string
metadata_tags_title	Title of the track	string

Appendix C - Notebook Catalog

All notebooks can be accessed at Github with this link: https://github.com/ruchithelamp/music

Below table lists the notebook names and their functionalities

Notebook Category	Notebook Description	Notebook Name
Data Acquisition	Used to pull music data from the Musicbrainz flexible API at https://musicbrainz.org/ws/2/	API_Updated.ipynb
Data Acquisition	Used to pull music data from the Discogs API at https://api.discogs.com/database/search and music lyrics from the Genius API at https://genius.com/api-clients	API_Updated2.ipynb
Data Acquisition	Exploratory data pulls from LastFM, Genius, and Musicbrainz	Milestone_ELT_API.ipynb
Data Acquisition	Exploratory data pulls from Discogs and Genius APIs	simple_music_collector.py
Data Cleaning	Cleans the Acousticbrainz data files including extracting years from the dates, assigning binary values to music metadata such as “danceable” and “happy”, and exporting results to a csv	Acousticbrainz_cleaning.ipynb
Data Cleaning	Concatenate all Discogs files into one merged dataframe	Complete_merger.ipynb
Data Cleaning	Cleans the Acousticbrainz data sets, concatenates them into a large dataframe, then fuzzy-matches the rows to matching songs in the already combined Musicbrainz and Discogs English music dataframe	Cleaning_AB_and_combining_with_MB_Discogs.ipynb
Data Cleaning	Concatenates all Musicbrainz and discogs with lyrics files into one csv	Concat_all_lyrics.ipynb
Data Cleaning	Concatenates all available discogs lyrics into one csv	Concat_discogs_lyrics.ipynb
Data Cleaning	Concatenates all Discogs files into one large dataframe	Concat_discogs_releases.ipynb

Data Cleaning	Concatenates all available Musicbrainz lyrics into one csv	Concat_musicbrainz_lyrics.ipynb
Data Cleaning	Concatenates all Musicbrainz files into one large dataframe	Concat_musicbrainz_releases.ipynb
Data Cleaning	Checkpoint/sample test code to explore data cleaning and visualization	Data_cleaning_M2.ipynb
Data Cleaning	Merging musicbrainz dataframe with lyrics with discogs data frame with lyrics	Dataframe_merging.ipynb
Data Cleaning	Converts raw JSON files to CSV for dataframe analysis	JSON_converter.py
Data Cleaning	Converts raw JSON files to CSV for dataframe analysis	JSON_converter_m.py
Data Cleaning	Filters for Years, from AcousticBrainz CSV files	YearFilter.ipynb
Data Cleaning	Dropping unrequired feature columns and other metadata from AcousticBrainz CSV files	AB DimReduction
Data Analysis	Performs Latent Dirichlet Allocation sentiment analysis on song lyrics	Latent_Dirichlet_Allocation.ipynb
Data Analysis	Preforms k means clustering, Elbow and Silhouette Evaluation for Optimal Number of Clusters (k) for Acousticbrainz data	Pycaret_acousticbrainz.ipynb
Data Analysis	Creates and runs a support Vector Machine recession era songs/sentiment classifier	SVM_Classifier.ipynb
Data Analysis	Preforms k means clustering, Elbow and Silhouette Evaluation for Optimal Number of Clusters (k) for all lyrics data	Silhouette_kmeans.ipynb
Data Analysis	Preforms tokenization of song metadata including genre and sentiment topics	Tokenization_prep.ipynb
Data Analysis	Runs clustering analysis on n-sized samples from AcousticBrainz, and plots	FullClusteringPipline.ipnyb
Visuals	Violin and bar plot outputs of our topic and sentiment analysis for all lyrics	topics_by_sentiment.png

Visuals	Wordcloud outputs of our topic and sentiment analysis for all lyrics	wordclouds.png