How Sad Are We?: Global Sentiment Analysis of Music

Zi Yan Wu
MIT
ziyanwu@mit.edu

Anna Kwon
MIT
annakwon@mit.edu

Sneha Sriram
Wellesley
ss135@mit.edu

ABSTRACT

Have you ever listened to a song in a language you didn't understand and asked: would I still like this song if I knew the words? The world is not only separated by its borders, but by extreme cultural differences and the impacts they faced from major historical events, whether locally or globally. With those differences comes radical shifts in mindsets from country to country. This project aims to analyze that difference as it relates to pop culture, specifically, the top 100 songs across the world.

It is a well known fact that what some cultures consider important are not as such for other cultures. We also know that attitudes among populations can have a high correlation with the current events that are happening around them, such as during high economic downtown, large civil unrest, natural disasters, or heightened political tensions. Our approach to doing this project takes these into consideration as we try to understand the relationship between people and their choices for entertainment.

There have been many studies on popular culture across time periods, but our hypothesis is that even within the same timeframe, cultural mindsets are best exposed through what media people choose to consume.

1 INTRODUCTION

The intersection of culture, emotion, and economics is a rich field of study, revealing much about societal values and reactions to global and local events. Music, a universal form of expression and a language that trancends cultural barriers, offers a unique lens to explore these intersections, particularly through the emotional content of lyrics. This paper delves into the question: how do the emotional expressions in music lyrics correlate with economic indicators across different cultures?

Our project is motivated by the desire to understand the relationship between music lyrics, emotional content, and economic factors. By analyzing the top 1 and 100 songs from various countries, we aim to uncover patterns in emotional expression and correlate these patterns with economic indicators. This approach not only highlights cultural differences in emotional expression but also explores how economic conditions may influence the emotional landscape of societies.

The design of our solution involves several key steps. Initially, we collect data from various sources, including song lyrics datasets and GDP data for different countries. We preprocess the lyrics data and analyze emotional content using a pre-trained BERT-based model for multi-label emotion classification. This model enables us to represent each song as a probability distribution vector across 28 different emotions, providing a standardized measurement of emotional prevalence within each song.

Furthermore, we evaluate the relationship between emotional expression in music and GDP by correlating the emotional vectors with economic indicators for each country. Our evaluation highlights intriguing findings, including the prevalence of certain emotions across different countries and the correlation between emotional expression in music and a country's GDP.

Our study reveals discernible patterns in emotional expressions among countries, highlighting variations tied to economic disparities. Through this study, we found that the music of nations with lower GDPs often emit specific emotions, admiration and approval. Future directions include automating the retrieval of top songs per country and addressing limitations in translation tools. Additionally, exploring alternative evaluation metrics and refining emotion classification methodologies could enhance the depth of analysis. Moreover, there's a need to consider cultural nuances beyond GDP in future comparative studies.

2 BACKGROUND

Understanding the relationship between cultural expression, emotions, and economic indicators is a complex endeavor with far-reaching implications. Music, as a universal medium of human expression, encapsulates a myriad of emotions and cultural nuances, making it a ground for exploration that we hope to utilize in this project. In the past, studies have been conducted on negative-positive sentiment scores on popular music from around the globe, as well as how sentiment in music interacts with stock prices. In this project, we aim to explore the intersection between emotional expression and economic indicators through the lens of GDP.

3 DATA USED

3.1 Dataset Selection and Motivation

For this study, we have selected the following datasets:

3.1.1 *Kworb Spotify Dataset.* **Source:** Kworb.net

Description: The Kworb Spotify dataset provides detailed information on Spotify charts, including song rankings, streams, and artist details. It covers both global and regional charts, updated daily.
Motivation: This dataset offers a comprehensive view of music trends and popularity over time. By analyzing this dataset, we can uncover patterns in music consumption, popularity dynamics, and the impact of various factors on a song's success. It is particularly useful for studying temporal trends and geographical differences in music preferences.

3.1.2 *Genius Lyrics Dataset.* **Source:** Genius.com

Description: This dataset contains lyrics for a wide range of songs available on Genius.com, a popular platform for song lyrics and annotations.
Motivation: Lyrics are a crucial aspect of a song's appeal and meaning. Analyzing the Genius lyrics dataset allows us to explore linguistic patterns, themes, and sentiment in popular music. This can provide insights into the relationship between lyrical content and song popularity, as well as cultural and social trends reflected in music.

3.1.3 *Kaggle National GDP Per Capita Dataset.* **Source:** Kaggle

Description: This dataset includes GDP per capita information for various countries over time, offering an economic context that can be correlated with other socio-economic data.
Motivation: GDP per capita is a fundamental economic indicator influencing various societal aspects, including cultural consumption like music. By incorporating this dataset, we can analyze how economic factors correlate with music trends and popularity, providing a broader socio-economic perspective.

3.2 Potential Shortcomings

While these datasets are rich and provide significant insights, they also present certain limitations. At a high level, the countries that we have chosen to explore in this project are fundamentally different in terms of data availability: some have widely available data, and some do not. For ease of collection, we chose to use a popular songs database (Kworb Spotify Dataset) that collects results only from Spotify, however we recognize that Spotify may not be the top music streaming service for every country that we investigate. In future iterations of this project, we hope to consolidate data from numerous sources according to streaming service popularity, to make our analysis more accurate.

4 MOTIVATION

The primary motivation behind this project is to explore the relationship between music lyrics, emotional content, and economic indicators such as GDP. By analyzing the emotional content of lyrics from songs across different countries, we aim to gain insights into cultural and societal differences in emotional expression. Additionally, correlating these emotional expressions with economic indicators like GDP allows for a deeper understanding of how economic factors may influence or reflect cultural sentiments. This analysis could have implications in various fields, including sociology, psychology, and economics, providing valuable insights into cross-cultural emotional dynamics and their connection to broader societal trends.

5 DESIGN

The solution involves several steps:

- Data Collection and Preprocessing: Initially, data is collected from various sources, including song lyrics datasets and GDP data for different countries. The lyrics data is preprocessed to remove translation credits and empty lines, ensuring clean input for analysis. For the purposes of sentiment analysis, Google Translate API was used for translation.
- Emotion Analysis: Using a pre-trained BERT-based model for multi-label emotion classification, the emotional content of lyrics is analyzed sentence by sentence. The model is a PyTorch implementation of GoEmotions, which utilizes the Hugging Face Transformers library. Hugging Face Transformers is a popular open-source library for natural language processing (NLP) tasks. It provides a wide range of pre-trained models and tools for building, fine-tuning, and deploying state-of-the-art NLP models. The library simplifies the process of working with transformer-based architectures. Emotions are associated with sentences, allowing for a granular understanding of emotional expression within songs.
- Model Architecture: First, the BertTokenizer from the transformers library is used to tokenize the input text. Tokenization involves breaking down the text into individual tokens (words or subwords) and converting them into numerical representations that can be processed by the model. The BertForMultiLabelClassification model is a variant of BERT specifically designed for multi-label classification tasks. It consists of a pre-trained BERT model followed by a classification layer. The pre-trained BERT model comprises multiple layers of transformer blocks, each containing self-attention mechanisms and feedforward neural networks. These

layers enable the model to capture contextual information from the input text effectively. A list containing the mapping of label IDs to emotion labels is created (28 emotion categories). Each index corresponds to a specific emotion label. Then, a dictionary is initialized to map emotion labels to their corresponding IDs for later use. The code iterates over lyrics from different countries. For each country's lyrics:
  - Each sentence in the lyrics is tokenized using the tokenizer.
  - The tokenized input is fed into the BERT model, and the model's outputs are obtained.
  - The outputs are processed to obtain scores for each emotion label using a sigmoid function. These scores represent the model's confidence in predicting each emotion label for the input sentence.
  - Emotion labels with scores above a certain threshold (defined by threshold) are considered as predicted labels for the sentence.
  - The predicted labels and their corresponding scores are stored in a list for each sentence.
  - These lists of predicted labels and scores are aggregated for all sentences in a country's lyrics and stored in the emotions dictionary, with each country mapped to its list of predicted emotions for each sentence.
- Normalization and Mapping to GDP: The emotion vectors for each country are normalized based on the proportion of each emotion in the lyrics. Countries are then mapped to their corresponding GDP data for a specific year, allowing for the correlation between emotional expression and economic indicators.
- Visualization: The correlation between GDP and emotional expression is visualized through scatter plots, showing the relationship between GDP and the proportion of each emotion for each country. Additionally, a similarity matrix is generated using cosine similarity between emotion vectors, providing insights into the similarity of emotional expression across different countries.

The design of this solution allows for a comprehensive analysis of the relationship between music lyrics, emotional expression, and economic indicators. By leveraging advanced natural language processing techniques and data visualization methods, the solution provides valuable insights into cross-cultural emotional dynamics and their connection to economic factors. While existing work may focus on individual aspects such as sentiment analysis or economic analysis in isolation, this integrated approach offers a more holistic understanding of the complex interplay between culture, emotion, and economics.

6 EVALUATION

Our first metric of evaluation that we used is by representing each song as a probability distribution vector across 28 different emotions with the emotions being: "admiration", "amusement", "anger", "annoyance", "approval", "caring", "confusion", "curiosity", "desire", "disappointment", "disapproval", "disgust", "embarrassment", "excitement", "fear", "gratitude", "grief", "joy", "love", "nervousness", "optimism", "pride", "realization", "relief", "remorse", "sadness", "surprise", and "neutral".

We do this to provide a standardized measurement of the prevalence of each emotion within each song to compare across the different songs. Instead of keeping raw counts which would create bias from the lengths of the songs, by taking the proportion, we can see how pervasive an emotion is relative to that specific song.

One important note is that the model we used is able to assign more than one emotion for a given sentence leading to the probability distribution not being a true distribution, as a sentence can be placed in two different buckets at the same time.

Our second metric of evaluation used was using the GDP of the different countries we looked at. We used this measure as a way to have a standard measure of comparison between the different countries. Although GDP itself doesn't fully represent the uniqueness of each country, especially among cultural divisions, GDP is nonetheless still a useful measurement in categorizing countries, as countries with similar GDPs generally also have a similar quality of life for the people who reside there. For the countries with a better quality of life, we would expect that the sentiments among the people living in those countries to be more positive. The question that we're trying to answer is whether or not those sentiments are also reflected through the sentiments of the popular songs within that country.

For our first results, we started by representing each country solely based on their top one song listened to within that country. From this, we were able to generate a similarity heat map across the different countries, as shown in Figure 1.

One noticeable result was how similar most of the countries were to one another. Using cosine similarity, we can see that a majority of countries are mostly at least 0.8 in terms of similarity. When looking into this, we see an observation that potentially biases the similarity metric. When looking into the sentence categorizations themselves for each sentence, we notice that the emotion "neutral" was exceedingly prevalent for every song. Among all the 74 different countries, the average proportion of "neutral" sentences for each song was 75%. From this, we can see another interesting result from the countries of Cyprus and Saudi Arabia having clearer

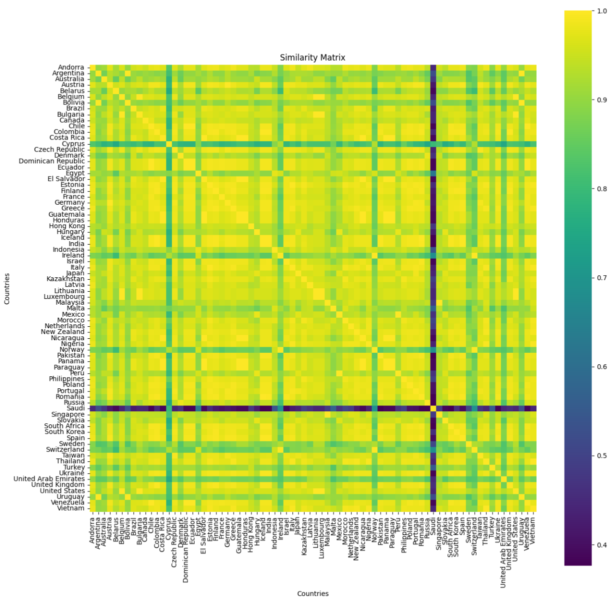

Figure 1: Simlarity Heatmap

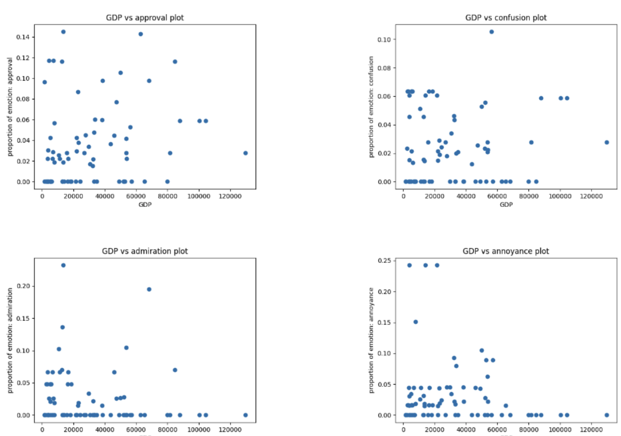

Figure 2: Emotion Plots (Top 1 Song)

dissimilarities between the other countries. The reason for this is due to their songs having less “neutral” tones, with only 51% and 21% respectively. For Saudi Arabia specifically, its highest proportion of emotion was “sadness” with 41%, which explains the deep dissimilarities as most of the other songs don’t have as much as this emotion.

Next, we plotted, for each emotion, the proportion of that emotion in the country’s top song (y-axis) to its GDP (x-axis).

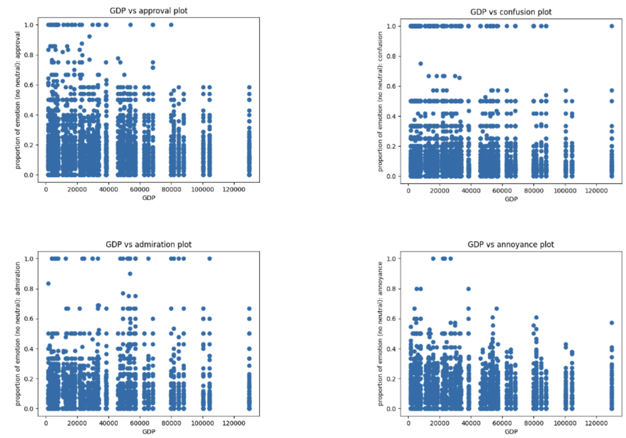

Figure 3: Emotion Plots (Top 100 Songs)

A subset of these plots are shown in Figure 2. For many of the plots, like in GDP vs “approval” and GDP vs “confusion” above, we often see no correlations, however, for some of them we do see a semblance of a pattern, like for GDP vs “admiration” and GDP vs “annoyance”. One issue that we noticed during this step, however, is how we would often see the same song representing multiple countries, which makes sense as popularity isn’t defined or restricted by any geographical boundaries.

To resolve these issues, we removed the “neutral” emotion from consideration and renormalized the distribution vector to better highlight the other emotions. We also chose to expand the number of top songs per country from the top 1 to the top 100 to add more variation between different countries, to get better generalizations and less bias. From there, we plotted the emotion vs GDP graphs in Figure 3, with the same emotions as Figure 2, and we noticed a clearer trend that wasn’t visible when only looking at the top one songs. For the GDP vs “approval” and GDP vs “confusion” plots, we can see that there is a slight correlation between higher proportions of those emotions and the countries’ GDPs and for the GDP vs “admiration” plot, we can see that there isn’t any correlations between the two measures, both results of which show a different story as compared to using only the top 1 songs, like in Figure 2. With these results, we can see that the GDP of a country does have an effect on the different emotions invoked between the popular songs of that country.

7 RELATED WORK

One piece of prior work is *GoEmotions: A Dataset of Fine-Grained Emotions by Demszky et al* [2]. In this paper, it describes the dataset that our model for classifying emotions is based off of. In this paper, it talks about curating a dataset of Reddit comments to use as training for emotion classification, designing the set of emotions used to classify the different comments, and how it uses a BERT model to classify emotions among new text. In our work, we build on the findings of this paper by extending the analysis and model across songs, which are closer to literary works than Internet forum comments, and across different languages, through translation.

Another piece of prior work is *Self-report captures 27 distinct categories of emotion bridged by continuous gradients* [1] by Cowen and Keltner. In this paper, it describes the different range of emotions and how there are 27 unique varieties that they can be boiled down to. This paper examined many different forms of short videos and developed a framework to categorize specific emotions through 27 distinct values versus the classic dimensions such as valence versus arousal. Our work builds on this by using this model to represent the emotions of the different songs we encounter. Instead of categorizing each song by a specific emotion, we break down the song into its distinct emotion proportion values to capture better subtleties within each song.

8 CONCLUSION

The main takeaways of our work is that there is a pattern among the emotions evoked and the differences between countries. By developing a pipeline that allows us to take the currently best charting songs, extract their lyrics, and analyze the emotions within the lyrics, we were able to analyze the top 100 songs per country across 28 different emotions, highlighting that certain emotions are more prevalent for countries with lower GDPs, such as admiration and approval.

For future lines of work, some systems changes that can be done is by incorporating a way to automatically and periodically get the top songs of each country, such as through the use of a webscaper, to have a completely self-functioning process. Another system change is the use of a local word translation tool, as currently our system is bottlenecked by the amount of API calls Google Translate allows us per day.

For the evaluation, other metrics should also be explored as our representation might not reflect all possible nuances of the data. For example, by adjusting the chunking of the songs (e.g. sentences versus stanzas), we might achieve a different emotion classification based on the available context. Another thing is by developing a way to better tease out the subtleties among the "neutral" emotion, as since this was a highly common classification found in our data, differentiating them would allow us better insights. Finally, future work can also expand on the measurement used to compare the different countries, as even though GDP is a good indicator of certain aspects of a country, things like cultural differences aren't able to be represented by such a number, so exploring a way to take that into account would prove to be immensely useful.

REFERENCES

[1] Alan Cowen and Dacher Keltner. 2017. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences of the United States of America* 114 (09 2017). https://doi.org/10.1073/pnas.1702247114

[2] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. *arXiv preprint arXiv:2005.00547* (2020).