**Introduction**

For this project, I decided to analyze the lyrics of British new wave band, "Depeche Mode." They have changed their sound numerous times over their 40+ year career, but most of their songs with themes such as politics, religion, love, and lust to varying degrees of ambiguity. They have also been able to make many chart-topping hits and successful albums despite (or even because of) their dark tone and remain one of my favorite bands to this day.

Through this analysis, I hope to see if the sentiment of Depeche Mode lyrics falls in line with my perception of the band, that they make music with dark, underlying meanings. I also want to see if the tone of their songs affects the popularity of their albums.

This report will seek to answer the following questions:

1. What is the overall sentiment of Depeche Modes discography?
2. Does how does the sentiment change from album to album?
3. Does the sentiment of the album correlate with the popularity of that album?
4. What are the most frequent words in each album and do the results reflect the sentiment of the album?

# Data Preparation

## **Scraping the Lyrics**

The main data set used for this project was the song lyrics from Depeche Mode's 14 studio albums (released from 1980 to 2017). I used the original track listing for each album (not including instrumentals) for a total of 145 songs. I decided to focus of the albums rather than the songs themselves for two reasons:

1. Like most bands, songs are often re-released or remastered on live/compilation albums, and I did not want any duplicates.
2. Also like most bands with a worldwide following, track listings can vary wildly for country to country.

By focusing solely on the original track listings, I hope to better capture the sentiment of each album.

To obtain the lyrics, I scraped them from the website "lyrics.com", which has the lyrics of every Depeche Mode song (including re-releases). For each album, I created a list of URLs and used a self-made function to iterate through the list, scraping the lyrics from each URL with BeautifulSoup.

## **Chart Website Scraping**

To answer one of the data questions, we need to find the chart performance of each Depeche Mode album. For this, I scraped album data from another website:

<http://archives.depechemode.com/discography/>

Each URL contains the top chart position for the album in both the UK and the US. I created a function (like the one I used for the lyrics) to scrape the album name, the UK chart position, and the US chart position for the website and append them to a corresponding list.

One of the albums (Construction Time Again) did not chart in the US, so I had to create a separate unique function that would only scrape the album name and the UK chart position. I then inserted those values into the lists mentioned above (placing a "Nan" value for the US chart position).

## **Sentiment Analysis**

The most important component of this report is the sentiment analysis done on the lyrics. NLTK's Sentiment Intensity Analyzer was used for this report. Given a list of sentences, this analyzer will return four values:

Negative Score (0 to 1): Indicates the percentage of the sentence that has a negative sentiment.

Neutral Score (0 to 1): Indicates the percentage of the sentence that has a neutral sentiment.

Positive Score (0 to 1): Indicates the percentage of the sentence that has a positive sentiment.

Compound Score (-1 to 1): Indicates the overall sentiment of the sentence with anything less than 0 being negative and anything greater than zero being positive.

One of the limitations of performing sentiment analysis on song lyrics is that they are not structured like most text. Like poetry, songs often lack the punctual and grammatical structure of something like a novel. Therefore, rather than trying to put each song into a sentence format, I will treat the lyrics of the entire song as one sentence so each song will have a sentiment score in the four categories listed above.

Using the Sentiment Analyzer, I will create eight lists, two for each of the score types with one containing all 145 songs and the other containing the average score for each album.

## **Creating a Pandas Dataframe for Analysis**

To help in analyzing the data, I took all the data thus far and put it into a Pandas data frame. The columns (not including the index) will be:

Album: The name of the album

UK: The top position of the album on the UK Charts

USA: The top position of the album on the USA Charts

Negative: The negative sentiment score of each album

Neutral: The neutral sentiment score of each album

Positive: The positive sentiment score of each album

Compound: The compound sentiment score of each album

Table

Description automatically generated

## **Tokenization**

Since most of the examples of sensitivity analysis we have seen in class were untokenized, I decided to leave the lyrics mostly unaltered up until this point. But to answer my final analysis question, some tokenization needs to be done. For each song, I wanted to tokenize the lyrics, remove all stop words, and remove any non-alphabetic characters. I created a function called "tokenize" that would accomplish all these steps and called it for each album.

One of the problems I ran into when trying to use nltk's tokenization function was that it did not deal with apostrophes that well. For example, in the song "Just Can't Get Enough" from the first album, the stop word "can't" is repeated multiple times. The tokenizer would treat "can" and "'t" as two separate words, and while "can”, and "t" are both recognized stop words the apostrophe was not removed from the t. So, the word "'t" kept appearing in the tokenized word list.

After some research (sources listed at the end of the paper) I found a tokenization package called "gensim" that was able to tokenize words with apostrophes more accurately.

# Data Questions

## **1. What is the overall sentiment of Depeche Mode's discography?**

During the sentiment analysis portion of the data preparation, I mentioned I created four lists, each containing a respective sentiment score for each song. By averaging these lists, we can find the overall sentiment of their music.

Total Negative Sentiment Score: 0.10508275862068966

Total Neutral Sentiment Score: 0.7531172413793104

Total Positive Sentiment Score: 0.14184137931034482

Total Compound Sentiment Score: 0.24148551724137932

The overall sentiment of Depeche Modes discography is slightly positive with a compound sentiment score of 0.24, which goes against my initial assumption of the band's material. Through answering the other questions, I hope to shed light on this result.

## **2. How does the sentiment change from album to album?**

The following plots graph the mean scores for each of the sentiment categories by album.

Chart, line chart

Description automatically generated

Except for the third album, there is a gradual increase in the negative sentiment score across the albums. The album with the highest score is the 10th album (Exciter) at around 14% while the lowest was the first album (Speak and Spell) at around 5%.

Chart, line chart

Description automatically generated

The positive sentiment percentage seems to be more random than the negative percentage, fluctuating between 10 to 20 percent of the album. "Some Great Reward" had the highest percentage (around 19%) while "Construction Time Again" had the lowest (around 9%).

To get a better idea of the difference in positive and negative sentiment for each album, I have created a scatterplot where the positive score is the x-axis, and the negative score is the y-axis.

Chart, scatter chart

Description automatically generated

This graph shows that there is some correlation between the positive and negative scores, but four albums don't fit the trend. Album's 1, 5, and 7 have low percentages in both categories while album 10 has a relatively high percentage in both categories.

Chart

Description automatically generated

Although the neutral sentiment of the albums is not too important, but it’s worth mentioning that this makes up at least 65% of all the album's content.

Chart, line chart

Description automatically generated

This chart confirms that the majority of Depeche Mode's discography has a positive sentiment. Only three albums (Construction Time Again, Playing the Angel, and Spirit) have a negative compound score.

## **3. Does the sentiment of the album reflect the popularity of the album?**

To answer this question, I created two scatterplots (one for the UK and one for the US) where the x-axis is the compound sentiment score, and the y-axis is the album's chart position. The lower the dot is, the higher that album charted.

Chart, scatter chart

Description automatically generated

Give the United Kingdom is their home country, there isn't much surprise that every one of their albums cracked the top 10. But there seems to be no correlation between the sentiment of the album and where it charted. The randomness of this data suggests that the chart position is based on other factors not captured in the data available.

Chart, scatter chart

Description automatically generated

Like the UK plot, there seems to be no discernable correlation between the sentiment of the album and its popularity. Rather, there is more correlation with time, as the later albums chart higher than the earlier albums.

## **4. What are the most frequent words in each album?**

For each album, I ran the songs through a word frequency function I created that return a list of the most frequent words and their frequency (in this case, I used the top 10). I also used a function to create a word cloud for each album.

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text, letter

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated with low confidence

Text

Description automatically generated

Text

Description automatically generated

Looking at the frequency of words for the albums, it's a little easier to see why the sentiment scores are positive for most of the albums. The word "love" appears in the most frequent words list in 8 of the 14 albums (with it being the most frequent word in 3 of them). Love is a common theme in Depeche Mode songs, but that is not a guarantee that it is always in a positive context. In most of these cases, these words appear in the chorus of the song and are repeated multiple times.

# Conclusion and Next Steps

I began this report expecting the following results:

1. The overall sentiment of Depeche Modes lyrics would be negative.
2. The more negative albums would be more popular as many Depeche Mode fans (including myself) are drawn to the dark themes of their music.

Neither of these hypotheses were supported by the data, which leads me to believe that the sentiment of a song does not always convey the message of a song and that doing analysis on song lyrics is much different than for a novel or a speech. There are many elements that the lyrics alone can’t capture, such as the tempo and key of the song, that play the vital role in the songs overall theme.

This report opens the door for future research. Only one sentiment analysis method was used for simplicity, but it would be a good idea to try out different sentiment methods and compare the results. There is also room to dive deeper into the discography. Not only doing a more targeted analysis on the songs themselves, but comparing the sentiment of singles to non-singles, as well as the comparing the choruses to the verses.

Finally, more experimentation can be done with word tokenization. We were barley introduced to the concept of n-grams in this class, but they might be useful in tokenizing things like poems and songs where words and phrases are repeated. Given that this is the case in many Depeche Mode songs, incorporating n-grams into the tokenization process might reveal new insights into the content of their lyrics.

# Sources and References

Main sources of data:

<https://www.lyrics.com/>

<http://archives.depechemode.com/discography/>

Referenced for my sentiment analysis code and some of my plots:

<https://towardsdatascience.com/how-to-analyze-emotions-and-words-of-the-lyrics-from-your-favorite-music-artist-bbca10411283>

Referenced to help create the word clouds:

<https://www.analyticsvidhya.com/blog/2021/05/how-to-build-word-cloud-in-python/>

Helped with some of the plot features:

<https://www.delftstack.com/howto/matplotlib/matplotlib-label-scatter-plot-points/>

Resources for the "gensim" module:

<https://radimrehurek.com/gensim/>

<https://towardsdatascience.com/5-simple-ways-to-tokenize-text-in-python-92c6804edfc4>