Project 1 Plan 3510

For this project, I would like to answer the question: "What are the main recurring themes, emotional patterns, and linguistic characteristics present in The Smiths' lyrics, and how do these evolve and perform over their discography?" I hope that by doing this, I will gain a deeper understanding of the overall themes and lyrical patterns within their catalog and how these elements change throughout their discography.

To begin the project, I will collect a comprehensive dictionary of songs and their lyrics using the Spotify API and Genius Lyric API. I will then need to clean the data by removing unnecessary characters and converting all the text to lowercase for easier comparison. ChatGPT made me aware of the NLTK library, which will hopefully let me tokenize the lyrics for analysis. If that approach isn't fruitful, I will likely use regex to handle the cleanup and any text manipulation. Once the data is processed, I will move on to identifying the main themes in their lyrics. By finding the most commonly used words and removing stop words, I will be able to categorize and group the words into categories such as love and conflict. This will give me a clear idea of what The Smiths wrote about and will hopefully allow me to create a dataframe containing the songs and their characteristics, making it possible to compare songs and timeframes to one another. This will also allow me to see which tones and sentiments were most prevalent at different stages of their career.

Next, I will attempt to gauge the sentiment of each song, ideally using the TextBlob library. I will determine the sentiment based on positive or negative key phrases and words. TextBlob will then help me quickly establish an overall sentiment score for each song. The results of this sentiment analysis will be visualized to illustrate how the emotional tone of The Smiths' lyrics changes over time.

Visualizations are a key component of the project, and I plan to use Altair and WordCloud to represent the data. I will use word frequency data to create word maps that highlight the most common words in The Smiths' lyrics, allowing for easy interpretation. Using Altair, I will also create graphs to track sentiment changes over time, showing how the emotional content of their music evolved across different albums.

One possible roadblock I might encounter is the human factor behind the lyrics. Specifically, with The Smiths, the tone might not always match the findings of an analysis like this. However, this is also why I chose them, as they often create songs that sound happy or, at least, upbeat and fast, while having darker lyrics. I believe that manually reviewing the findings and listening to their music will mitigate any possible discrepancies.

This project aims to provide an effective and insightful analysis to uncover the deeper undertones in The Smiths' discography. Through careful data preparation, theme and sentiment analysis, and thoughtful visualization, I hope to provide a deeper understanding of what makes their work so powerful and memorable.