

TWA Proposal Outline - Syndicate 4

August 7, 2021

Proposed Topic:

To observe lyrical trends over time via lyrics analysis.

We are interested in how lyrics have changed over time, e.g. what words or phrases have become more or less popular in their usage in songs. Looking at the amount of repetition in lyrics might also reveal some trends. Likewise, the number of unique words in a song might be similarly insightful. It would also be interesting to see if these trends are consistent across genres or if certain genres consistently use similar language across time. Alternatively, it could be observed that there are clusters that suggest there are common themes that are era specific or transcendent across time. One approach to achieving this could be to map out the most popular words for each decade and see if their popularity is consistent across other decades. Note that analysis into all of these topics may not be feasible, given our time and report length constraints.

We plan to get lyrical data from an existing csv if possible (right now we are looking at this [dataset](#)). If this isn't successful, we can attempt to scrape lyrics via Genius and Spotify, likely from popular playlists (e.g. best of 70s, 80s, 90s etc). Either way, we plan to stick to almost exclusively English songs. At this stage we also plan to build a classification model that can tell us what decade a song was released from its lyrics. However, the objective of this classifier is not for prediction, we are only trying to interrogate the model and look at the model parameters and see what insights this can give us pertaining to the patterns of lyrics over time.

This kind of analysis could be applied to how societal changes affect song lyrics or if applied differently, could determine the popularity of a song based on its lyrics (although this is beyond the scope of our proposal).