My documentation and output has been reported on the 223.ipynb file. I have been working on the project on my own at a lab, and planning to submit the paper similar to this pipeline. This is a small snapshot of my overall project.

The project analyzes user comments on nicotine pouches on reddit through VADER sentiment analysis and topic LDA analysis. The issues I faced are trying to find the best LDA topics, filtering out very common but niche words that are not in the stop words array, and the duplicates on posts. I spent the most time filtering data. The other issue is setting the threshold of positive, neutral, and negative sentiment, where most literature support the cutoff of -0.05 and 0.05 as the best representation of comments. The word cloud explains the most common words from each topic, whereas sentiment distribution is done to see whether the comments are generally positive or not.