A look into the version history of my sentiment analysis code, or the files I completely scrapped and started over, will quickly reveal a learning curve of many trials and errors. Still, each time I elicited absurd predictions for my collected tweets or simply failed to understand the bugs in my code, I reminded myself of the creative possibilities for what I could do with the data I extracted from the public's tweets.

The idea of sentiment analysis first came to me after taking a coursera course titled "Big Data and Language." Instead of using corpora I learned about, I wanted to analyze people's opinions on aiding Ukraine as expressed on Twitter, and see if these opinions aligned with the civilian casualties in the nation.

With this idea in mind, I sought out Professor Adam Lee to mentor me. I first tested the Python library TextBlob for my analysis. Once I was able to extract tweets from the Twitter API, he taught me what it meant to sanitize, stem, and vectorize my data for machine learning. Yet, after I worked on these edits with him and independently, I still noticed many inconsistencies and errors in the way TextBlob classified the polarity (positive/negative sentiment) and subjectivity.

Through some Internet browsing, I realized the reason for this was that TextBlob was rules-based and entirely relied on a process of counting and multiplying conditional probabilities. I wanted to instead use a pre-trained transformer (a state-of-the-art Natural Language Processing technique) that could be bolstered with even more data. I coded this new model which encoded tweets with the pre-trained transformer, then decoded this embedding with a MLP (multi layer perceptron with three fully-connected neural network layers), then used a final softmax layer to predict each tweet's polarity. I lastly downloaded a dataset of 1.6 millions pre-labelled tweets from Kaggle to train my model to around 82% accuracy.

After extracting Tweets each day in a month, I could finally average the results from each day and apply the same statistical analysis I used in my Tinsley project to compare public opinion on foreign aid to Ukrainian war casualties each day. While my data analysis did not yield any statistically significant results, I still had lots of faith for the applications of sentiment analysis. Whether it's creating more humane human-computer interactions or helping autistic children navigate emotion, I know I can be exceedingly creative through Natural Language Processing.