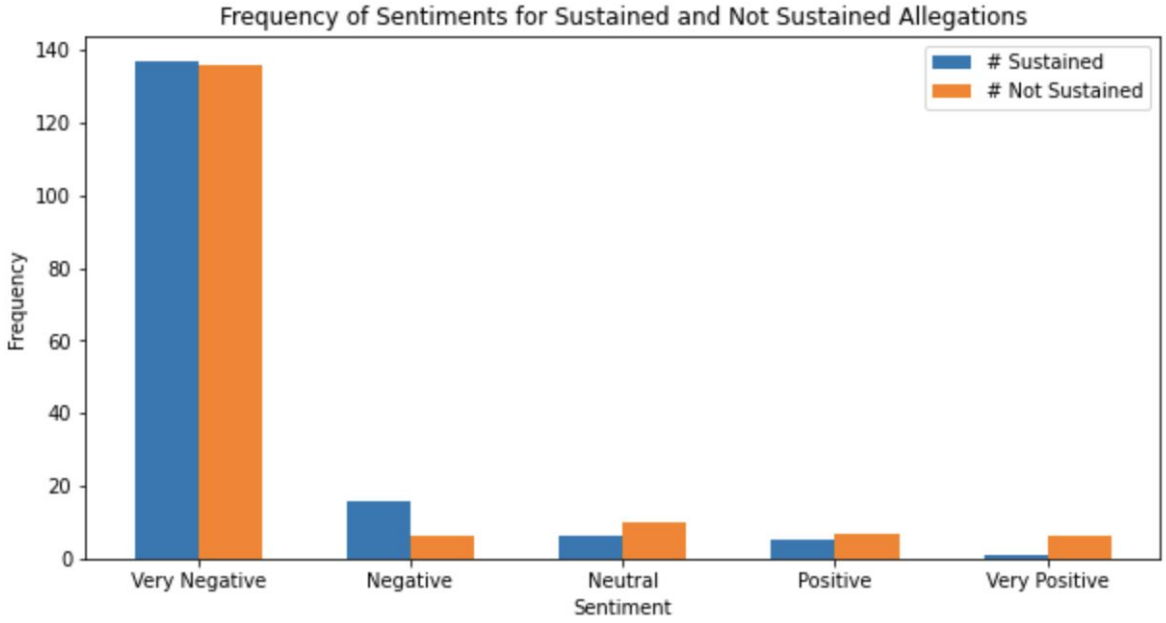
William Ansehl, Krish Seth, Nicholas Easton
Data Science Seminar - CP 5 Findings

Our goal is to determine if all allegations are listened to the same. This statement implies that all allegations made regardless of race, gender, location, etc. receive the proper due diligence. We aim to determine this via exploring allegations that result in sustained filings vs allegations that result in not sustained. For Checkpoint 5, we sought to analyze the distribution of case summary sentiment for complaints that resulted in “sustained” and compre this distribution to that of complaints that resulted in “not sustained.” In sum, Are the summary reports written pertaining to complaints that later became sustained, differ from the summary reports written for complaints that result in “not sustained” on a distribution basis. Our hope was that in analyzing the data, we would also find distributional differences between sustained vs not sustained cases where the complainant was a female, the complainant was a male, the officer was a female, the officer was a male, and by beat.

In querying our data, we resulted in a dataframe with 719 rows and 4 columns. Each row has an allegation id, a summary id, the beat id of the allegation and the final finding (SU or NS). All null values were removed and duplicates were removed (last duplicate was kept). We proceeded to split our data into two groups - cases where the final finding was sustained, and cases where the final finding was not sustained. In total, there were 554 not sustained complaints and 165 sustained complaints. The not sustained complaints were down sampled to promote a balanced dataset for sentiment analysis.

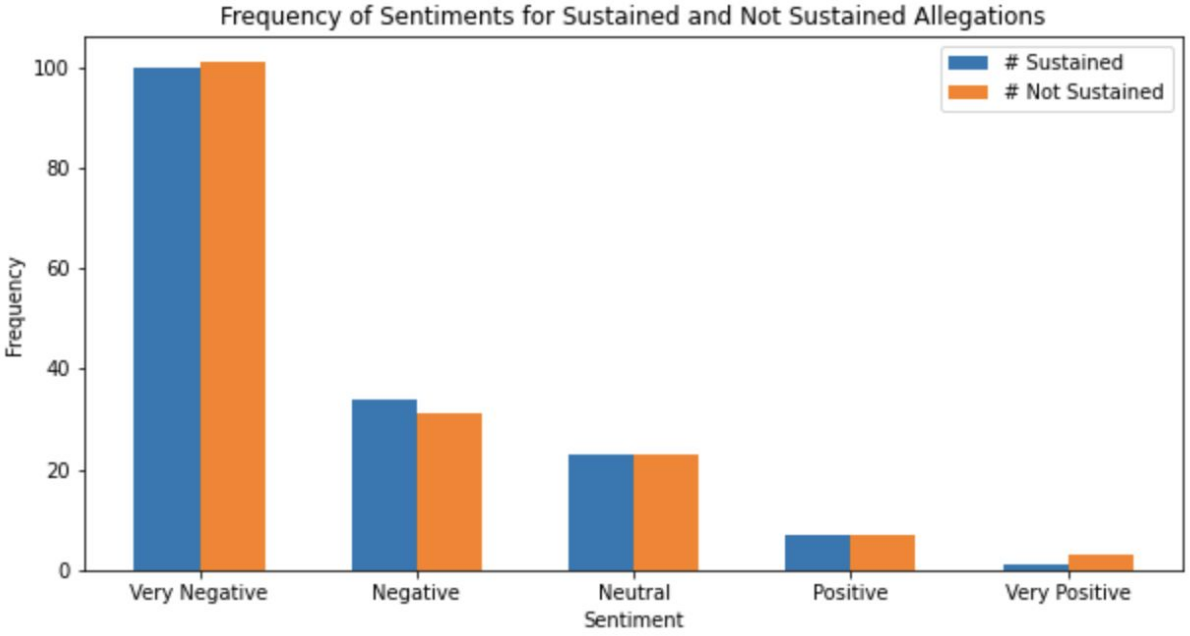
The library used for sentiment analysis was NLTK. Specifically, we only tracked the compound value of the polarity score. Each summary was fed into a sentiment analyzer and the results were plotted. Each label description corresponds to a bin of compound values, with Very Negative corresponding to scores between -1 and -0.6, Negative corresponding to scores between -0.6 and -0.2, Neutral corresponding to scores between -0.2 and 0.2, Positive corresponding to scores between 0.2 and 0.6, and Very Positive corresponding to scores between 0.6 and 1.



In the above plot, we see that the distribution of sentiments for sustained and not sustained complaints are roughly the same. There appear to be more negative sentiments for

sustained complaints and more positive sentiments for not sustained on a grouping by grouping basis. However, the differences are not that drastic and it is hard to say that these results are statistically significant enough to warrant such claims (especially given the low amount of data). Surprisingly, there appeared to be allegation summary reports that were “very positive.” However, such reports are more likely cases where the sentiment analyzer failed to properly grasp an accuracy sentiment metric for the given information.

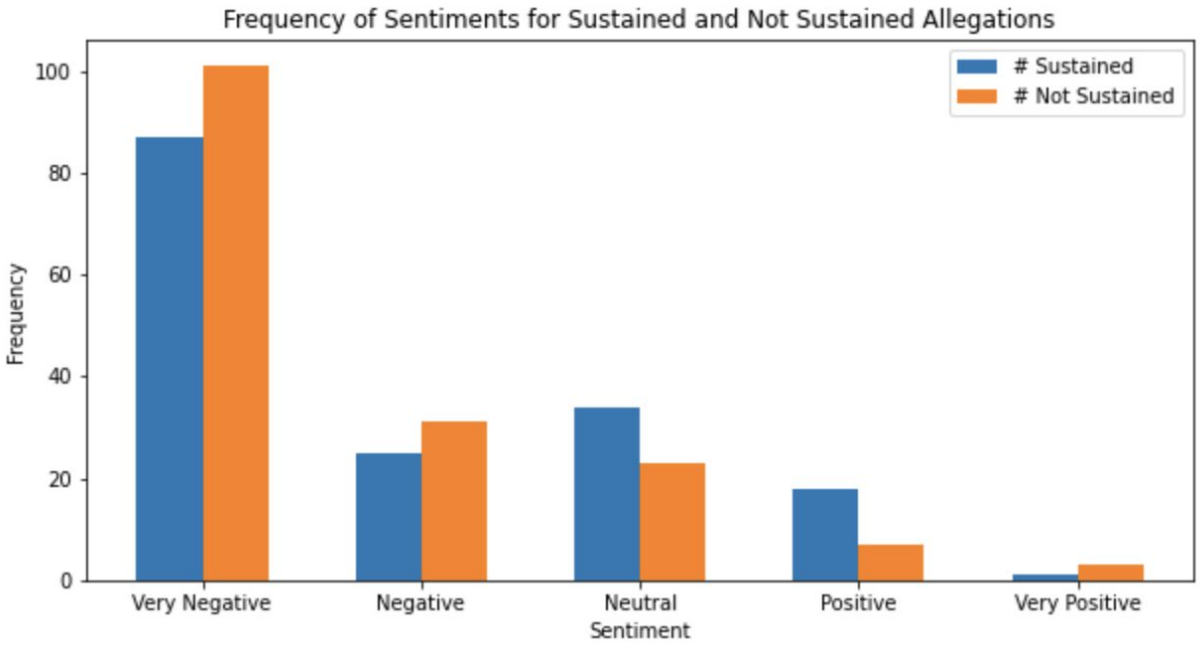
Our group then proceeded with two methods of cleaning the summary reports. We believed the sentiment analyzer would be more accurate with cleaner summaries, resulting in more representative distributions. Our first summary cleaning method implemented syntactic parsing. We filtered each summary and kept only the adjectives and verbs, words we believed that more strongly correlated with the sentiment of the given summary report. This method removed neutral nouns such as “October,” “2000” and also prepositions. The resulting bar chart can be viewed below.



Surprisingly, after syntactic parsing, the distribution of both sustained and not sustained complaints shifted more to the right. More complaints registered as negative, neutral and positive than before. Consequently, fewer complaints for both distributions registered as “very negative.” However, the cases we analyzed as “very positive” still appeared to be rather negative, signifying fault in the sentiment analyzer once again.

Our second method of cleaning the summary reports involved lemmatization. Similar to syntactic parsing, we kept only adjectives and verbs. The results of conducting sentiment analysis on the lemmatized results can be viewed below.

William Ansehl, Krish Seth, Nicholas Easton
Data Science Seminar - CP 5 Findings



This methodology of cleaning seemed to more affect the sustained complaints' distribution rather than the not sustained complaints distribution. The sustained allegations' distribution experienced another right shift towards the direction of more overall positive sentiments and fewer extreme negative sentiments. Regardless, the overall distribution of the two populations are still rather similar. It is still hard to say if there is any significant difference between the two.

Unfortunately, due to a lack of data, it doesn't make sense to further divide the data and analyze the resulting distributions based on the gender of the complainant or officer, or the beat. These actions would further reduce the amount of data we have to work with for analysis.

Ultimately, we found that there was very little distributional difference between the sentiments for sustained vs not sustained allegations. This implies that the sentiment of a summary report is not a strong indicator of whether or not an allegation will result as sustained or not sustained. Further data collection is required to determine how other factors such as race, gender and beat affect these distributions.