# Checkpoint 5:

- Sentiment analysis of misconduct allegations plotted against disciplinary action taken

I used tokenization and transformations to break up the allegation notes and then performed sentiment analysis to get the sentiment scores used throughout this checkpoint. Since there were so few allegations with reports that resulted in discipline, I also tried to see if there was a correlation between sentiment score and allegation count since it may be reasonable to believe that complainants would be more upset with officers that tend to have more allegations.
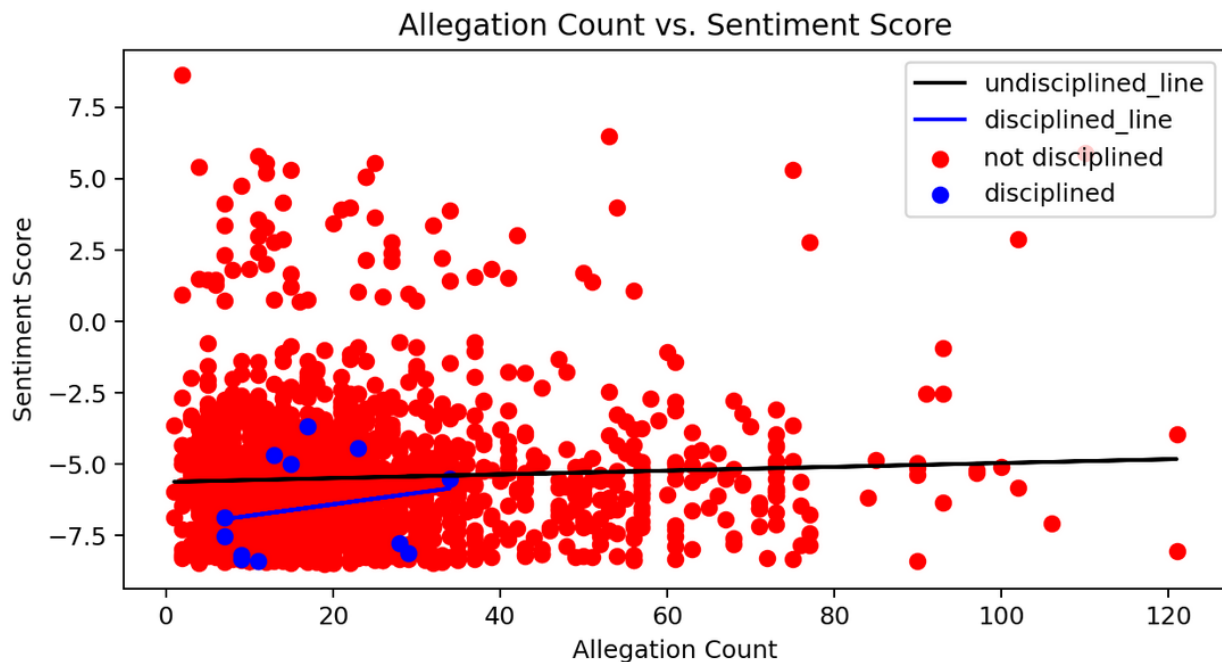

```
R Squared Not Disciplined: 0.002467671233613637
Pearson Correlation Coefficient Not Disciplined 0.04967566037420778
Spearman Correlation Coefficient Not Disciplined 0.03011874668301905


R Squared Disciplined: 0.045865923107537274
Pearson Correlation Coefficient Disciplined 0.2141633094335659
Spearman Correlation CoefficientDisciplined 0.30175624350367664


KS Test: Ks_2sampResult(statistic=0.30701754385964913, pvalue=0.17147898551166485)
AVERAGE DISCIPLINED SENTIMENT: -6.54
AVERAGE NOT DISCIPLINED SENTIMENT: -5.47

DISCIPLINED STD SENTIMENT: 1.6799528763232219
NOT DISCIPLINED STD SENTIMENT: 2.3838242942844228
```

The above graph shows the relationship between allegation count and sentiment analysis for different allegations. As can be seen, the relationship between these two variables is very noisy.

Additionally, there were very few examples of allegations that resulted in disciplinary action in which the allegation text was recorded, making it difficult to infer anything predictive about how sentiment analysis can be used to predict which officers will be disciplined. Also of note is that these distributions have a high enough p value so that we cannot reject the null hypothesis that they come from different distributions, but the p value is fairly low. There is also a moderate positive correlation between sentiment score and allegation count for allegations leading to disciplinary action but it is very difficult to make general inferences from 12 data points.

Allegation Count vs. Sentiment Score Partitioned by Gender
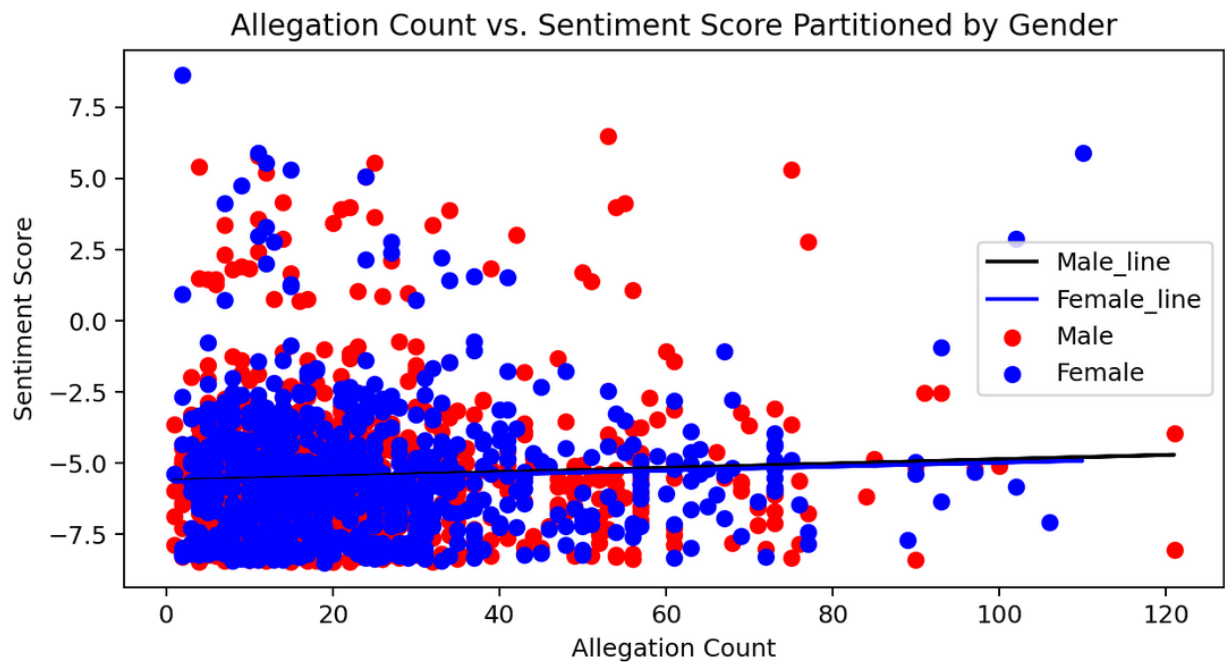
```
R Squared Male: 0.0030686572624313902
Pearson Correlation Coefficient Male 0.055395462471500255
Spearman Correlation Coefficient Male 0.03488430160282707
```

```
R Squared Female: 0.00284025118520724
Pearson Correlation Coefficient Female 0.05329400702900126
Spearman Correlation Coefficient Female 0.05327980859775925
```

```
KS Test: Ks_2sampResult(statistic=0.04139157633662521, pvalue=0.41543990313262324)
AVERAGE Male SENTIMENT: -5.45
AVERAGE Female SENTIMENT: -5.51

Male STD SENTIMENT: 2.416002697500606
Female STD SENTIMENT: 2.2827675238732
```

The above graph shows the allegation count vs. the sentiment score partitioned by gender of the complainant. Both genders show a moderate positive correlation, but neither can be said to be significant. The p value for these distributions is also more than high enough to reject the null hypothesis and the distributions look pretty similar so there does not seem to be a lot of predictive power in this graphic.

Allegation Count vs. Sentiment Score Partitioned by Race
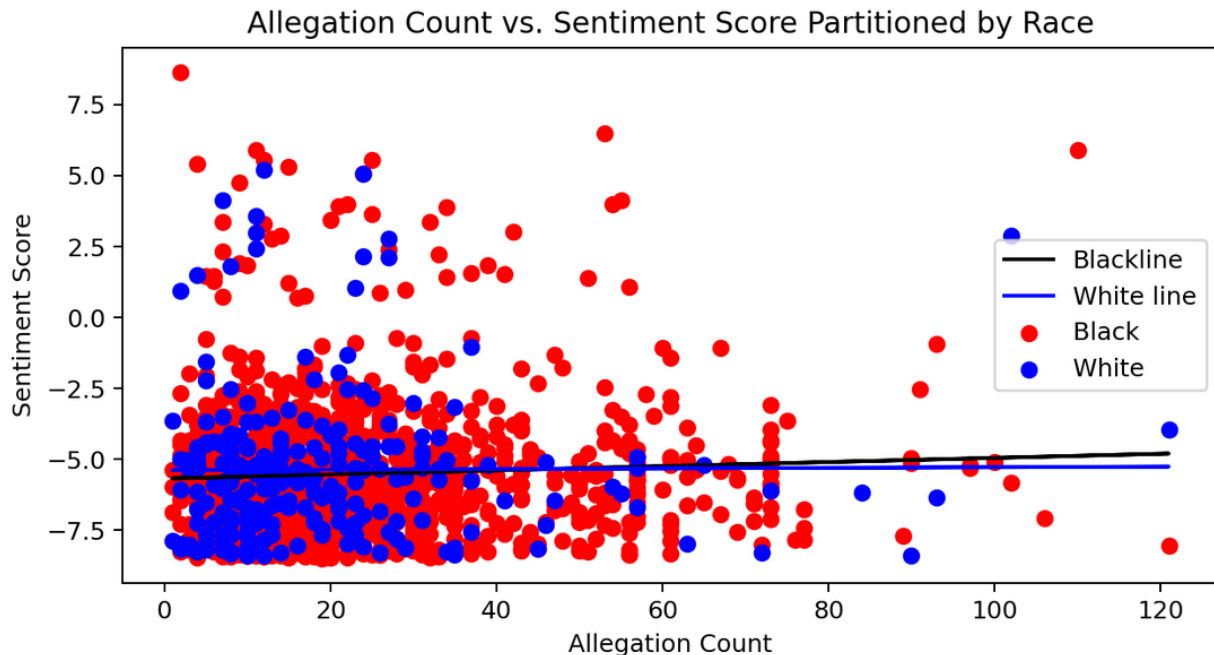

```
R Squared Black: 0.0029748034188291695
Pearson Correlation Coefficient Black 0.05454175848677019
Spearman Correlation Coefficient Black 0.052205647094605394


R Squared Disciplined: 4.943392728077329e-05
Pearson Correlation Coefficient Disciplined 0.007030926488079158
Spearman Correlation CoefficientDisciplined 0.005918997420046354


KS Test: Ks_2sampResult(statistic=0.052904578165262554, pvalue=0.6006255807235266)
AVERAGE black SENTIMENT: -5.51
AVERAGE white SENTIMENT: -5.36

black STD SENTIMENT: 2.3474335798235777
white STD SENTIMENT: 2.571385123518574
```

The above graph shows allegation count plotted against the sentiment score partitioned by race. Once again, we get a high p value for the two distributions, so we reject the null hypothesis. The correlation coefficients were very low for both white and black complainants so no relationship can be inferred between sentiment score and allegation count.

Ultimately, the relationship between allegation count and sentiment score seems to have no definitive correlation and cannot be used to make predictions about which officers are more likely to have allegations against them. Additionally, the lack of data for reports resulting in disciplinary action make it difficult to find any meaningful measure and limits the opportunities for logistic regression and other supervised classification algorithms with sentiment score as a feature vector along with race, gender, and allegation count since there is no way to prevent overfitting a model when one class has three times the magnitude of the other class.