Faraaz Beyabani, Andrew Vagliano, Harkirat Gill
November 25, 2020

Checkpoint 5: Natural Language Processing

For this checkpoint, our aim was to continue our examination of policing per area and population. We started out by accumulating allegation data for each community, including median income, complainant race and gender, and the text summaries included for each allegation. With the latter, we used the Python NLTK library to calculate the sentiment of each summary, or in other words, the positive, negative, or neutral feelings associated with a piece of text. Having done this, we attempted to predict the strictly negative sentiment values with the rest of our features using linear regression, in the hopes of discovering relationships between sentiment and complainant race, gender, certain areas, and area median income.

These are the questions that we aimed to answer with this assignment:

- How is the sentiment of an allegation summary affected by the complainant's race or gender?
- Is there a correlation between the negativity of a sentiment and various geographic and complainant features?
- Is the difference in sentiment of complaint texts across race and gender statistically significant?
- Which types of features are more effective at predicting the sentiment of a given text?

Data Processing

The data used for this assignment was pulled from the data_allegation, data_area, and data_complainant tables to form feature vectors for each allegation. These features were allegation ID, "section name", "column name", text content, area name, median income, complainant gender, complainant race, and complainant birth year. We chose these features in particular because we were curious as to how both geographical and complainant-specific demographics would affect the sentiment of a given text content.

Before performing any sort of large-scale analysis, such as with sentiment analysis or linear regression, we first cleaned the data by removing allegation entries that were related to outcomes and findings, rather than filings. We could find these entries by their texts of "NOT SUSTAINED" and "NOMISCONDUCT", for some examples. We also removed any duplicate entries and cleaned the remaining allegations' texts by removing final periods and new-line characters. Below is a small sample of the data we accumulated.

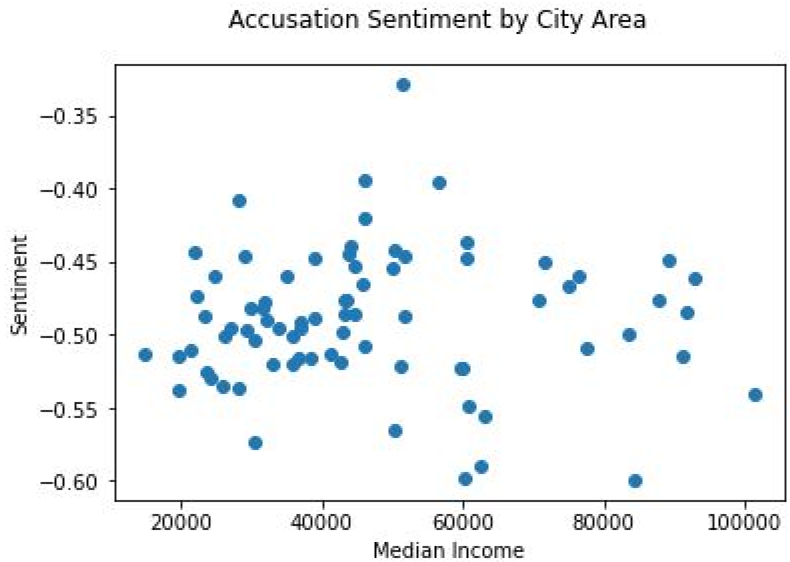
	Allegation ID	Section Name	Column Name	Text Content	Area Name	Median Income	Complainant Gender	Complainant Race	Complainant Birth Year
0	1062924	Accused Members	Initial / Intake Allegation	the complainant stated that she was involved in traffic accident with another ewvillan the complainant alleged that the accused officer incorrectly and inacurrately completed the traffic crash accident report the complainant did not have her case report number	Hermosa	$39,057	F	Hispanic	1971.0
1	1050875	Accused Members	Initial / Intake Allegation	the complainant alleges that the accused failed to arrest a female offender who threw a bottle at her and refused to leave the premises, the complainant alleges that the accused failed to arrest a female offender who threw a bottle at her and refused to leave the premises	Chicago Lawn	$32,944	F	Black	1977.0
3	1060347	Accused Members	Initial / Intake Allegation	and that she was afraid that the might come back. the reporting alleges that the accused was rude and and stated in an aggressive "maybe we can look at the tape. if so scared, change your locks. i'm not to do anything for you."	North Park	$51,589	F	White	1970.0

Sentiment & Area Income

Our first analysis attempt was to calculate the sentiment of each text and compare it with the income of its originating area. There was a very weak connection between these two concepts, but we were curious to see how well our findings from previous checkpoints involving median income and allegation count would fit in with our current efforts. After calculating the scores for every allegation, we averaged them over each area. Below is a table of our results for 10 areas.

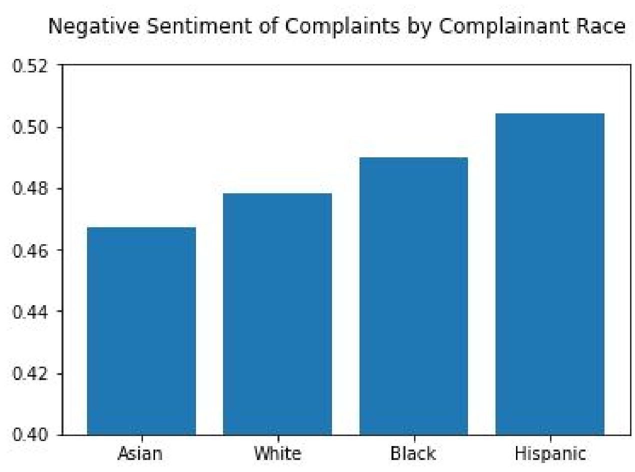
	Area	Score	Income
0	Edison Park	-0.599305	$84,331
1	Jefferson Park	-0.597945	$60,384
2	Lincoln Square	-0.589593	$62,427
3	South Lawndale	-0.573953	$30,603
4	Hegewisch	-0.565878	$50,252
...	...	...	...
72	O'Hare	-0.420273	$46,065
73	Oakland	-0.407486	$28,269
74	Portage Park	-0.395742	$56,649
75	Edgewater	-0.394468	$46,103
76	North Park	-0.328843	$51,589

Having done this, we converted median income to a numeric value and created a scatter plot of each area's overall sentiment and median income, which is shown below.



As can be seen, the results for this analysis were very inconclusive. We could not find a concrete relationship between these two variables, as there is a wide range of sentiment values regardless of the median income. However, since there are much fewer areas with income greater than $60,000, our data is skewed toward the lower end. Perhaps with more data spanning multiple cities, we would potentially be able to map out a relationship.

Sentiment & Complainant Race



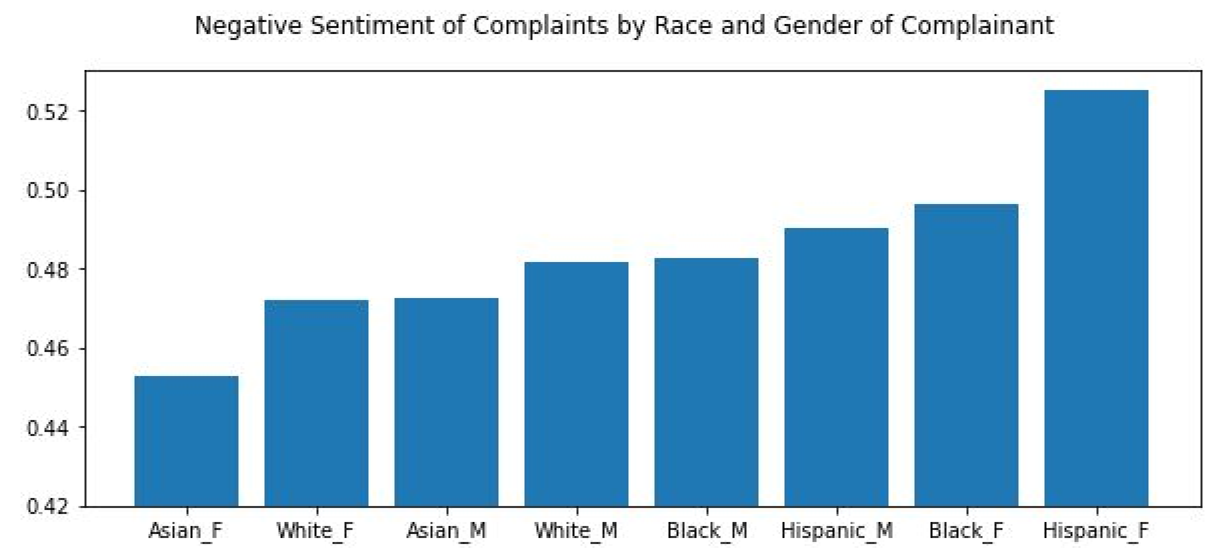
For a more direct analysis, we decided to try comparing sentiment directly to complainant race. While this provides less of a geographical view, we were more confident about this comparison, as it would potentially highlight bias or even racism against certain races. To do this, we performed the same process as we did for median income. In order to look only at the most relevant and appropriate parts of our data, we decided to ignore allegations related to Native

American complainants, while coalescing Asian and Asian/Pacific Islander data into a single category.

From the results shown above, we can indeed see that there is a difference in the negative sentiment between different race groups. Hispanic complainants saw an average of 0.5 negative sentiment (on a scale from 0 to 1), while their Asian equivalents only saw an average of around 0.465. However, it is difficult to tell whether this is a meaningful observation, considering the small range of the data. Furthermore, it is also difficult to attribute the cause of this small discrepancy. It could be the result of fewer Asian and White people engaging with police, or it could be bias on the part of the official filing this complaint (as many of the text summaries read as if written by a third party).

Sentiment & Complainant Race + Gender

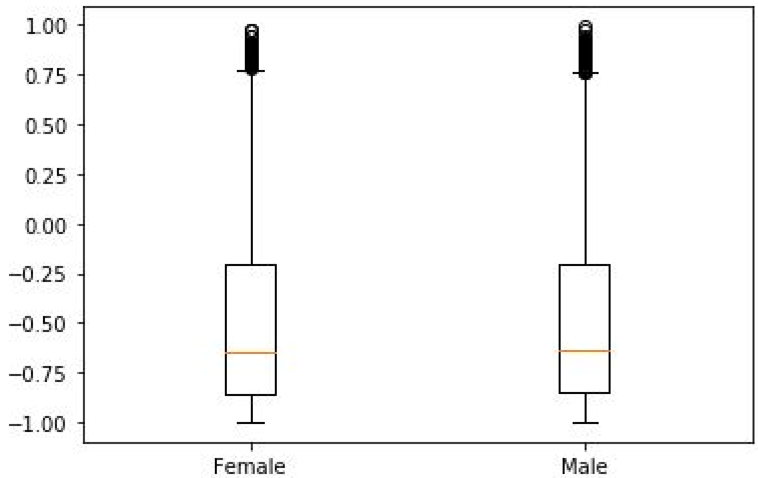
To see if these results changed when looking at a smaller granularity, we performed this same experiment over race and gender combinations, rather than just race.



Our findings were not too dissimilar from our previous analysis. Asian and White complaints had somewhat lower sentiment values as compared to their Black and Hispanic counterparts. Somewhat surprisingly, Asian and White women had lower average negative sentiment to their male equivalents, while Black and Hispanic women were the opposite. While we can't immediately explain this discrepancy between race groups, it was significant enough to imply some sort of underlying prejudice toward the groups with higher negative sentiment.

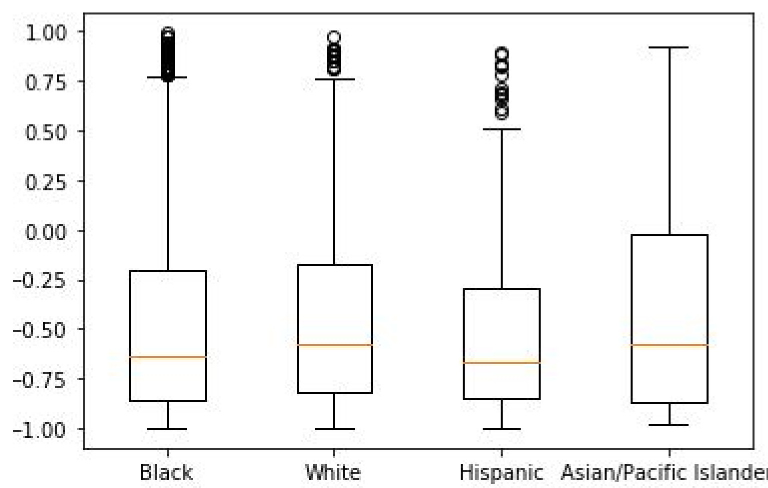
Testing Significance of Difference in Sentiments

Next, to test for the statistical significance of differences between sentiments, we compare the p-value between races and genders using scipy's ttest_ind method. We do this first for gender. Looking at the boxplots for gender, there does to seem to be a significant difference.



We get a p-value of 0.461 and so by using a p-value of 0.05 as the cutoff for significance, we can say there isn't a significant difference in sentiments based on gender.

Now looking at the boxplots for race, we see some variation, but it isn't clear if this is significant.



But, looking at the p-values between race pairs, we do find a significant difference between hispanic and white complaint sentiments and hispanic and asian complaint sentiments. However, the difference in complaint sentiments between black and asian and black and white is almost significant with a p-value of 0.056 and 0.08 respectively, but we need more data to say confidently.

p-value	Black	White	Hispanic	Asian/Pacific
Black	1	0.68	0.182	0.056
White	0.068	1	0.023	0.162
Hispanic	0.182	0.023	1	0.020

Asian/Pacific	0.056	0.162	0.020	1

With these results, we can say the difference in complaint sentiments between hispanic complainants and white and asian complainants is likely not due to chance and thus there is a significantly more sentiment in these complaint narratives.

Linear Regression

For further comparison between our features and the resultant negative sentiment values, we opted to use linear regression. We learned a model over a majority of the data and examined the coefficients of the fitted model to see the importance of each feature with regards to the sentiment. In order to do this, we had to one-hot encode many of our categorical variables, including area name, complainant gender, and complainant race. In addition, we also decided to omit certain “organizational” variables such as allegation ID, section name, and column name, as they provided little to no information about the areas or complainants involved with each allegation.

Median Income	0.0000
Complainant Birth Year	0.0000
Area Name_Auburn Gresham	-0.0052
Area Name_Austin	-0.0015
Area Name_Humboldt Park	0.0008
Area Name_Loop	-0.0027
Area Name_Near North Side	-0.0139
Area Name_Near West Side	0.0249
Area Name_North Lawndale	0.0357
Area Name_South Shore	-0.0510
Area Name_West Englewood	0.0001
Area Name_Woodlawn	0.0128
Complainant Gender_F	-0.0409
Complainant Gender_M	-0.0271
Complainant Race_Asian/Pacific Islander	0.1087
Complainant Race_Black	0.0261
Complainant Race_Hispanic	-0.0235
Complainant Race_Native American/Alaskan Native	-0.3174
Complainant Race_White	0.0289

As expected from our analyses, median income and complainant birth year played a negligible role in determining sentiment. The area dummy variables had a slight correlation with sentiment, however a few areas such as Englewood and Humboldt Park had extremely small coefficients, perhaps due to their smaller valid allegation counts. The gender variables had a negative impact on the sentiment as well, with the F variable having a stronger relationship, as we saw in our previous analysis. Lastly, the race variables are varied, with Asian and Native American races having a strong relationship, which is strange, considering the much weaker relationships that the other race variables provide.

Conclusion

Geographic origin of complaints seems to play a negligible role in determining sentiment. This may be due to the fact that complaints are not directly written by the complainant, but rather primarily by an officer. The officers having final say on the written language may minimize severity of reports, skewing complaints away from strong negative sentiment. However, the correlation of race (more negative for black and hispanic complainants, less negative for white and asian complainants) shows that the reports reflect a notable difference in how racial groups are treated, primarily that groups traditionally impacted more by police misconduct are more likely to have more severely written reports.

Future research should focus on improving the sentiment analysis model. While NLTK's Vader model comes with powerful training, customizing the train data to better capture sentiment of this specific form of report would provide deeper insight into the complaints. Expanding the geographic coverage of our data to include more examples in each area would allow for more extensive analysis of how different regions can affect complaints, possibly resulting in more conclusive findings than the current data.