

PROJECT REPORT

Data Science Seminar

The Royal Zombies
Shivangi Sharma
Shalin Parikh
Shraddha Bangad

Theme:

Our project aims to investigate the reasons behind why some complaints filed against police officers are sustained while others aren't. We aim to dig into the data to analyze how the police are held accountable for their actions.

In 2017, the Civilian Office of Police Accountability (COPA) came into existence, replacing the Independent Police Review Authority (IPRA). This project aims to research whether the police have been punished for their misconducts over the years, whether the recommended punishment suggested by the disciplinary committees matched the actual punishment given to them (i.e. whether the complaints are sustained or not sustained) and how did the factors like **gender, race, age, rank, salary** manipulated their decisions and what effect did the punishments (if any), have on their behavior/career ahead.

Checkpoint 1: Relational Analytics:

To begin with our research. In checkpoint one, we performed a few SQL queries to acquire some demographic findings from the database. The first set of data we sought to gather was the percent of total complaints which were sustained depending on the allegation category. This helped us understand the various types of allegation categories against police officers in the database(refer fig.1.1).

To further analyze if there could be a relation between the types of complaints and their sustainment, We explored the nature of the complaints. We defined “severe consequences” as any of the following categories: ‘Administrative termination’, ‘Resigned’, ‘Resigned - not served’ or ‘Suspended indefinitely’. We found that ‘Operation/personnel violations’ had the highest number of severe consequences, followed by ‘Drug/Alcohol Abuse’ and ‘Criminal Misconduct’. Surprisingly, ‘illegal

search' and 'false arrest' are the two least reprimanded complaint categories. (refer fig.1.2) We then went out to decipher which of the allegation categories led to the least severe punishments (refer fig.1.3). Here too, we found that 'Operation/Personnel violations' was the category with the highest number of least severe punishments, followed by 'Use of force' and 'illegal search'. This analysis tells us that no matter how severe the punishment is, drug and alcohol abuse and violations of any kind are taken seriously by the police department and there are always some actions taken against them.

Further comparison of complaints also shows us that the police department is more lenient with the complaints from the civilians and does a stringent check for the complaints made by the police officer (refer fig.1.4).

Before our analysis, we predicted that sustained allegations would be more detrimental to an officer's likelihood of being promoted, and it appears this trend holds in some years. Interestingly, however, in 2005, 2006, 2010, 2011, 2012, the category of officers most likely to be promoted in these years were officers with one or more sustained allegations against them. Although the case was very different for the officers with two or more sustained allegations in a given year. In most years, none of them were promoted (refer fig.1.5). Performing an analysis on whether police officers with 1+,2+ allegations against them were awarded showed us that the more allegations an officer had against him in a given year, the more likely he was to be awarded that year too!

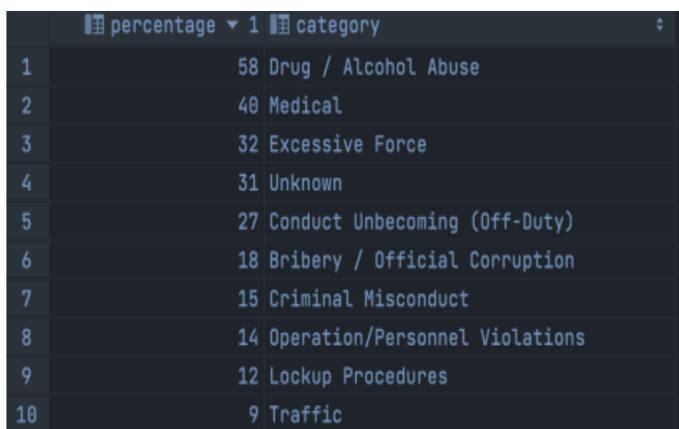


Fig.1.1 Percent of total complaints which were sustained depending on the allegation category.

1	641	Operation/Personnel Violations
2	270	Drug / Alcohol Abuse
3	249	Criminal Misconduct
4	198	Conduct Unbecoming (Off-Duty)
5	109	Use Of Force
6	67	Bribery / Official Corruption
7	46	Domestic
8	43	Lockup Procedures
9	34	Supervisory Responsibilities
10	16	Traffic
11	12	Illegal Search
12	10	Verbal Abuse
13	2	False Arrest

Fig.1.2 Categories of complaints are most/least likely to have severe consequences

count	category
68217	Operation/Personnel Violations
53793	Use Of Force
35982	Illegal Search
13229	Verbal Abuse
12620	Lockup Procedures
7868	False Arrest
7480	Traffic
6430	Conduct Unbecoming (Off-Duty)
5943	Criminal Misconduct
5847	Domestic
4313	Supervisory Responsibilities
705	Bribery / Official Corruption
671	Drug / Alcohol Abuse
75	First Amendment
43	Racial Profiling

Fig.1.3 Allegation categories led to the least severe punishments

year	officer_count	perc_adm
2003	6681	1.2572968118545127
2004	5779	1.2978023879563938
2005	4428	1.1065943992773262
2006	5398	1.5561319007039645
2007	6527	1.8231959552627548
2008	6221	0.9805497508439157
2009	5792	0.7424033149171271
2010	5235	0.3247373447946514
2011	4778	0.041858518208455424
2012	4361	1.1006649850951618
2013	4066	1.0821446138711266
2014	3968	2.04133064516129
2015	3770	1.1006649850951618

Fig.1.4 Proportion of officers with 1 or more allegations and 0,1+,2+ sustained allegations against them get promoted each year

percentofsustainedcomplaints	is_officer_complaint
1	false
2	true

Fig.1.5 Percentage of sustained complaints from civilians and officers

Checkpoint 2:Data Exploration:

For this checkpoint, our group continued to look for the relations between sustained complaints and factors affecting them. The visual analysis gave us a lot of insight into the working of the Chicago police department. The most interesting outcome that we encountered here was that complaints against White police officers were more likely to be sustained than their other counterparts. Their sustainment rate was over 50%, with Black police officers coming second with a sustainment rate of over 35% (refer fig.2.1). This somewhat contradicts the notion that White police officers do not get reprimanded for their actions.

Another interesting insight was that the sustainment rate for the allegation category ‘Use of force’ has gone up significantly in the past decade.(refer fig.2.2.1 and 2.2.2) This can easily be correlated with the articles about the Chicago police department that we have read in class. The fact that the number of complaints in this category has also increased in the last decade goes on to show why citizens have started losing faith in the Police department.

Despite this, fig 2.3 shows us that there are still a large number of complaints in this category ('Use of force') that are taken lightly, and police officers are not reprimanded as severely as they should be for the same(refer fig. 2.3). 'Illegal search' is another allegation category that has popped up in the last decade, which police officers still go mostly unpunished for.

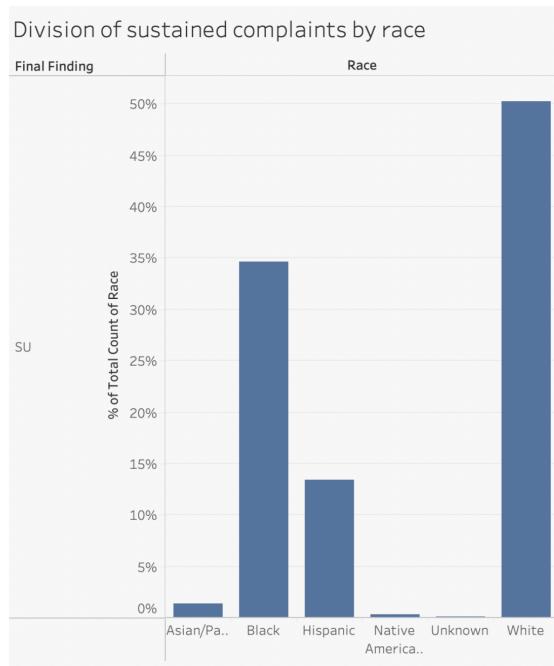


Fig.2.1 Sustained complaints Count by race

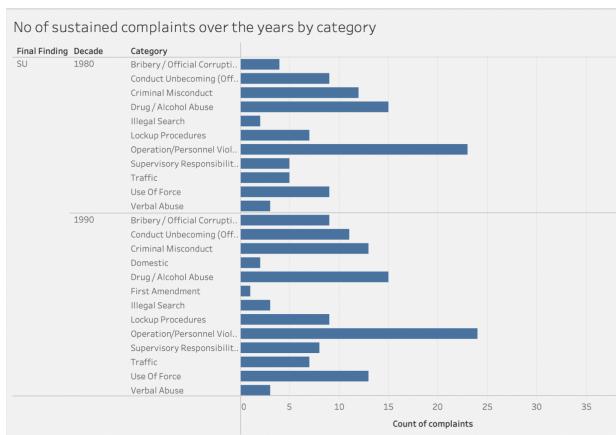


Fig 2.2.1 Trend of sustained complaints by category over the decades

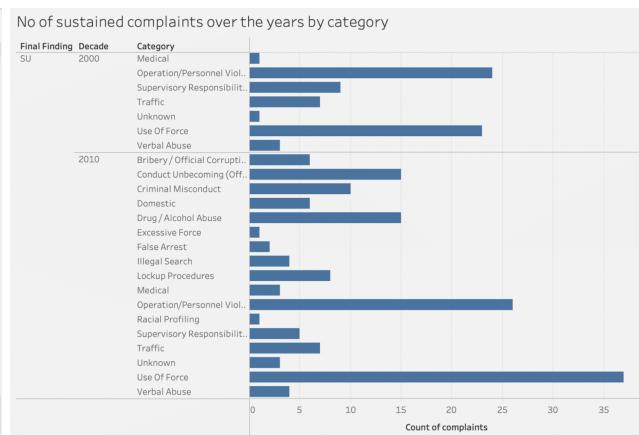


Fig. 2.2.2 Trend of sustained complaints by category over the decades

Category of complaints with less severe consequences

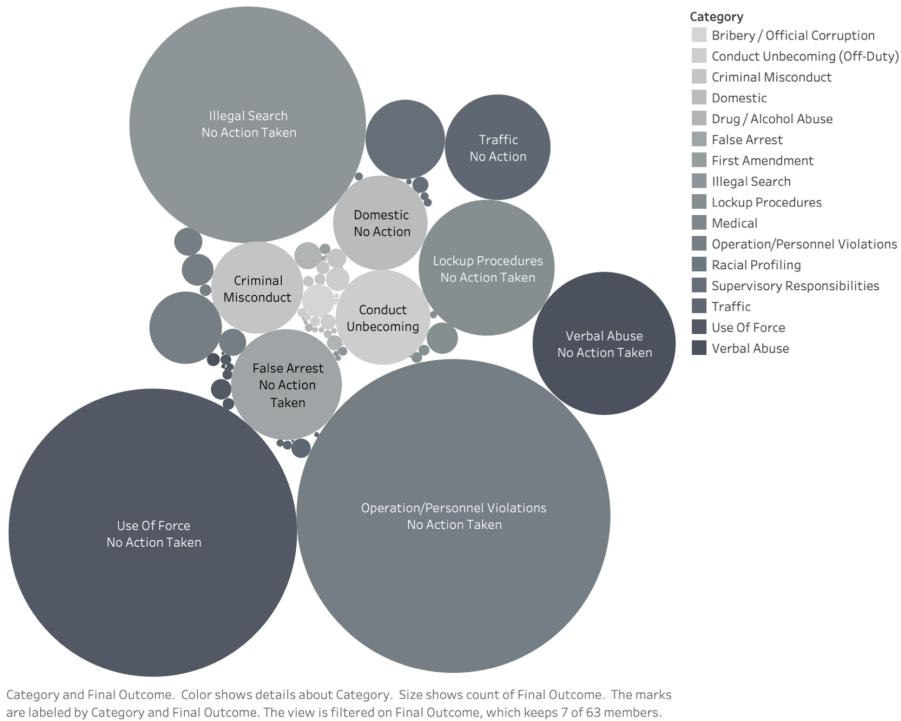


Fig. 2.3 Least severe consequences count by category

Checkpoint 3: Interactive Visualization:

For this checkpoint, we divided the police officers into various groups based on their age and salary, after which we tried to correlate the data with other metrics including their race, gender, and rank. The salaries of the majority of police officers lie between the range of 80-130k per year.

The analysis of the police officers' salaries shows that the highest salary recorded is around 170,000 USD per year (refer fig. 3.1) However, there are some discrepancies when it comes to the correlation between age and salaries of the officers. We'd expected the salaries to increase with age, but some police officers within the age range of 30-50 still receive a salary of less than <100000 USD per year. We observed that there are a large number of complaints against such police officers, with a sustained rate of only 5%. It could thus be hypothesized that even though such police officers are not being 'punished' per se, there has been withholding in their salary increment levels.

An analysis of the split of sustained complaints by gender(refer fig. 3.3) shows that male police officers have a higher number of total ($215911/249788 = 86\%$) as well as sustained ($17916/21990 = 81\%$) complaints against them. This could be attributed to the discrepancies in the number of total male and female police officers in the force. A plausible reason for the same could be those female police officers were posted in less violent neighborhoods than their male counterparts.

As far as the correlation of sustainment and race(refer fig. 3.2) of the police officers goes, the sustainment rate is pretty even for all the races, just minorly higher for the black police officers (7%). Since the data is not hard-hitting or large enough to make concrete inferences on the said correlation, we would refrain from making any concrete inferences about the same.

In terms of the rank of the police officers, the maximum number of complaints in the low salary bracket is against the rank ‘police officers’ as expected. When the salary bracket is $>130k$, the maximum complaints are against the captain (27%) and lieutenants (21%), with a surprising finding being that none of those complaints were sustained.

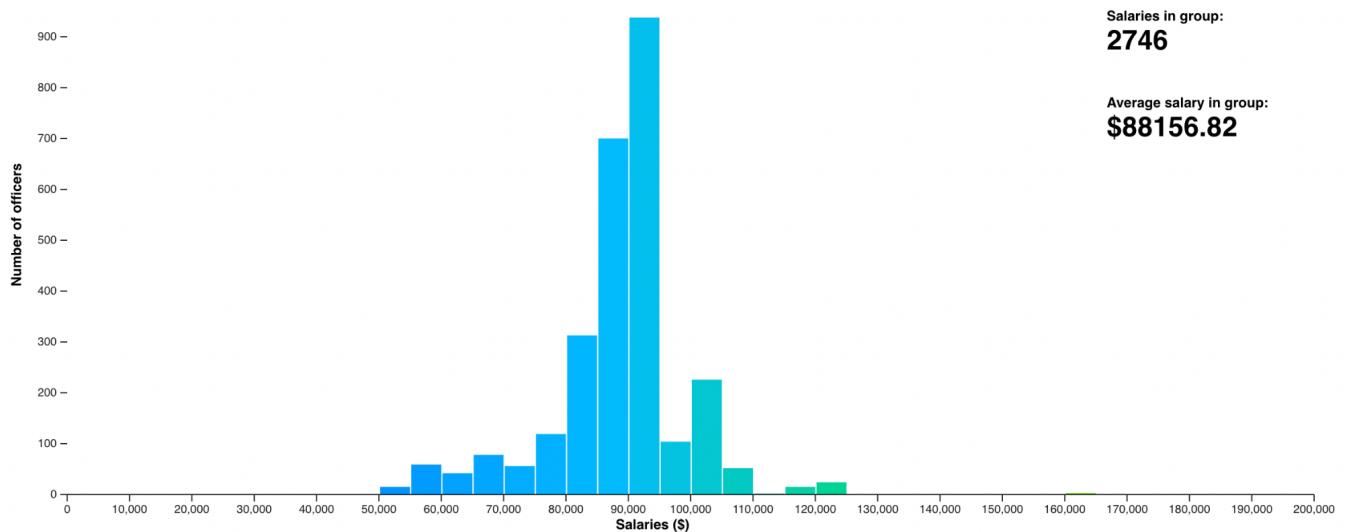


Fig. 3.1 Analysis based on the salaries

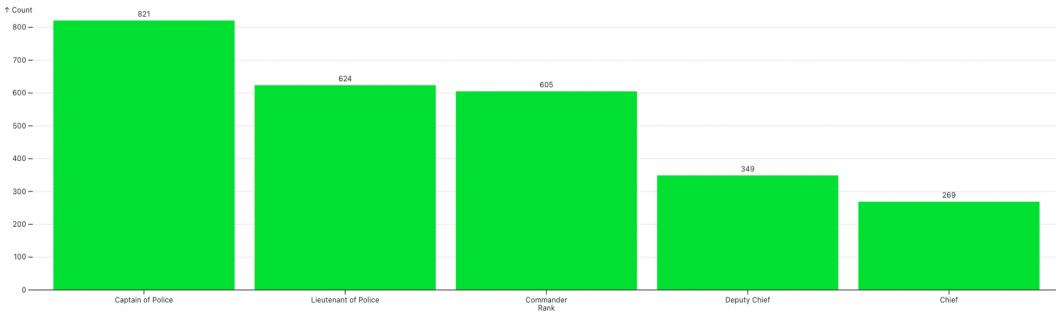


Fig. 3.2 Analysis based on the ranks

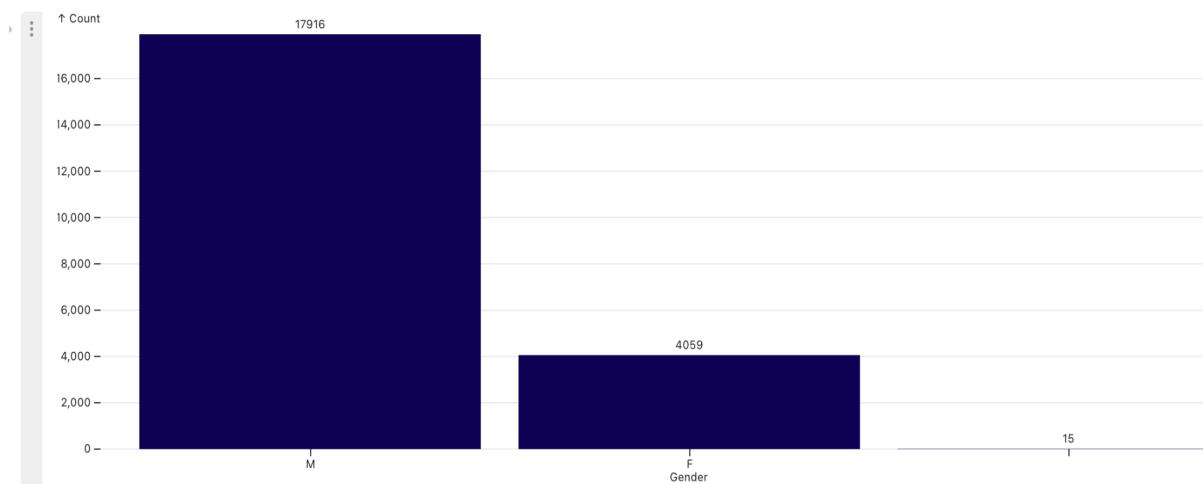


Fig. 3.3 Analysis based on gender

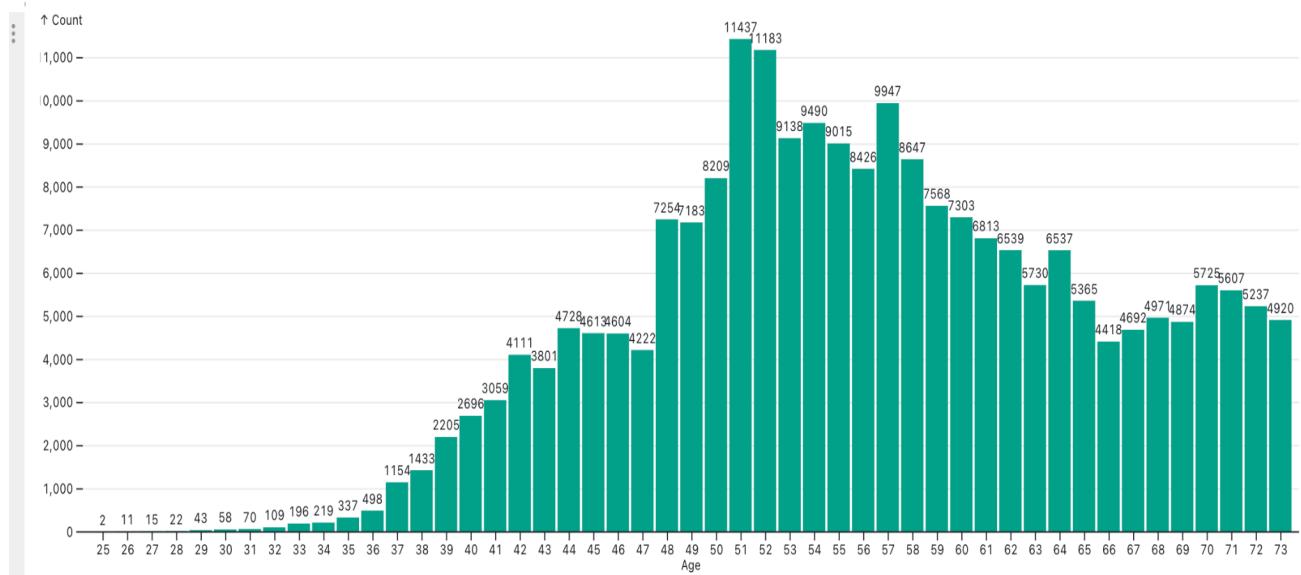


Fig. 3.4 Analysis based on age

Checkpoint 4: Graph Analytics:

Moving forward, for checkpoint 4, we analyzed the rank of Police officers for our analysis as they are the ones with the maximum number of positions in the police department and have the greatest number of allegations and in turn, sustained allegations against them. We implemented graph algorithms like Triangle count, Pagerank, inDegrees, OutDegrees, and Motif finding on our graph implementations to find some interesting observations.

We did our analysis amongst officers who had a rank of ‘Police Officer’ in the police department and tried to find out the frequencies of two officers being co-accused in the same allegation categories. From our previous analysis in Tableau and Observable, we found out that the categories of ‘Operation and Personnel Violations’ and ‘Conduct Unbecoming (Off-Duty)’ are some of the categories with the highest number of allegations and also the highest number of sustained allegations.

The officers who belonged to these ranks are taken as nodes, and the edges represent the relationship between the two officers who were a part of a crime together and hence had a common allegation id. The nodes had the attributes of officer id, officer name, and the count of sustained allegations where the rank is ‘Police Officer’ and the final allegation is ‘Sustained’. The edges have the source and destination of officer ids and the relationship between them is the count of their common allegation ids concerning the aforementioned categories.

After using the various algorithms provided by graph frames, we came to many interesting observations which are as follows:

1. Higher triangle count for a police officer does not ensure that he/she has a higher allegation count. It suggests that the police officer was involved in those sustained allegations, where there were many other police officers involved in each allegation.
2. Higher PageRank for a police officer suggests that a police officer has influenced the other police officers. Higher PageRank suggests that the police officer would influence or be involved in a certain way in an allegation that was sustained for the

two allegation categories we analyzed.

3. There are 4041 communities formed of police officers where they are part of an allegation that was sustained, and the allegation category is ‘Operation and Personnel Violations’.

4. There are 1246 communities formed of police officers where they are part of an allegation that was sustained, and the allegation category is ‘Conduct Unbecoming (Off-Duty)’.

Checkpoint 5: Natural Language Processing:

For this checkpoint, we attempted to determine if there was any correlation between the complaint narrative and the allegation category. We also proceeded to analyze whether the complaints were indicative of the sustainment of complaints against police officers.

We initially started with using BERT for this checkpoint, but the model was taking a long time to infer. We then decided to use LSTMs (Long short term memory) for the same. The model returned an accuracy of 89% on test data and was pretty accurate in predicting allegation categories accurately for any user input.

For the latter part (predicting sustainment of complaints using complaint narrative), we used Scikit-learn (TF-IDF). We changed our model from a Deep learning model to a Machine learning model to deal with the low amount of data in this case. Even after doing this, the model accuracy didn’t increase much. This shows that the complaint narrative was probably not a good indicator of sustainment. It was most probably affected by other factors, such as the rank, race, gender, salary, etc. of the police officers. The complaint narratives also depend on the police officer who is recording the statement, which could also be a contributing factor to sustainment.

Conclusion:

Working with data from the Chicago police database gave us several insights into the internal operations and workings of the police department. A number of these insights correlated and coincided with what we had read about the police in several readings and articles assigned to us in class, but at the same time, a number of them were surprising. We found out that police officers with more sustained complaints against them were more likely to be awarded and promoted. We also found that several senior police officers have gone unpunished for their actions. This could be indicative of senior officers spreading the culture of violence and crime in the entire department. There are also indications that police officers working together as part of a crew are more likely to influence each other to perform criminal activities. One more finding was that the sustainment rate is pretty evenly divided amongst police officers of all races, which contradicts the common conception that white police officers are less likely to be punished for their crimes.

Future Work:

We could expand checkpoint 5 to do a deep dive into the various complaint narratives - including whether the officer writing the report affects the sustainment of the complaint, and whether there is a difference/ similarity between complaint narratives from different regions. The scope of this project could also be expanded to include TRR reports and findings.