***Chicago Police Department Project Proposal - Checkpoint 2***

**Introduction**

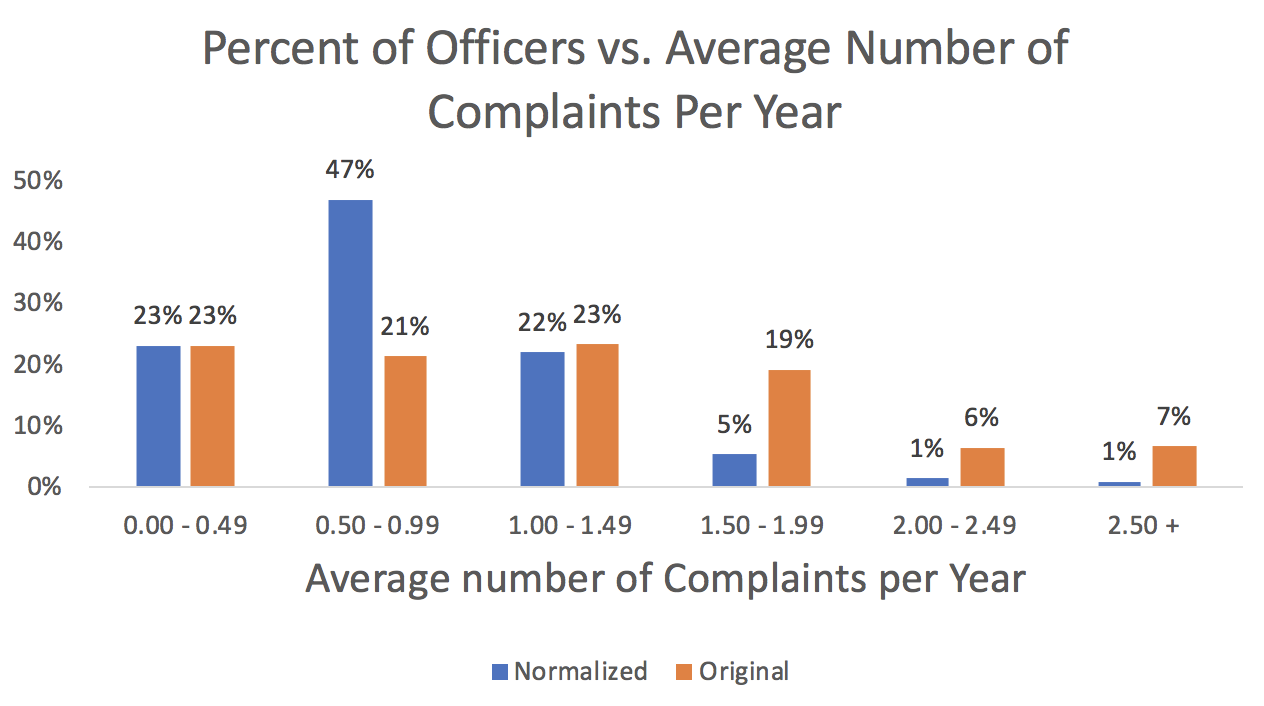
In this checkpoint, our team focused on integrating the crime and arrest data to normalize the information that we previously discovered about claims and officers. In order to accomplish this task we used the “All Sworn in Officers” snapshot as a base of this initial analysis. This data set, which we had not originally thought to integrate, allowed us to group officers by district with simplicity - and use our previously computed statistic of average number of complaints per year (at an officer level) very well. Our queries file details this process in more granularity, along with the steps that we took to clean the data with Trifecta.

**Core Questions**

***Q #5****- How does normalizing the number of claims for patrol area affect the statistics computed above in the context of all officers (including officers without claims against them)?*

We first computed the district average number of complaints per officer per year. We completed this task in Trifecta using our previously computed statistic of average number of complaints at an officer level in checkpoint one. From there, for each officer we divided their current average number of complaints by their district average. Hence, if there was a higher district average, an individual officers number of complaints should go down. Below is a histogram of the normalized results in conjunction with the original unnormalized analysis.

**Figure 1**: Percent of Officers vs. Average Number of Complaints Per Year



Here we see some interesting, but largely expected results. We see that holistically, the number of officers with a large amount of normalized complaints shifts downwards. However, the shift created a sort of skewed left distribution. We were fairly surprised to see this fact, and specifically that there was still a significant amount of officers with over 1 complaint per year on average even though we normalized for the area. Intrigued, we to continued to dig further.

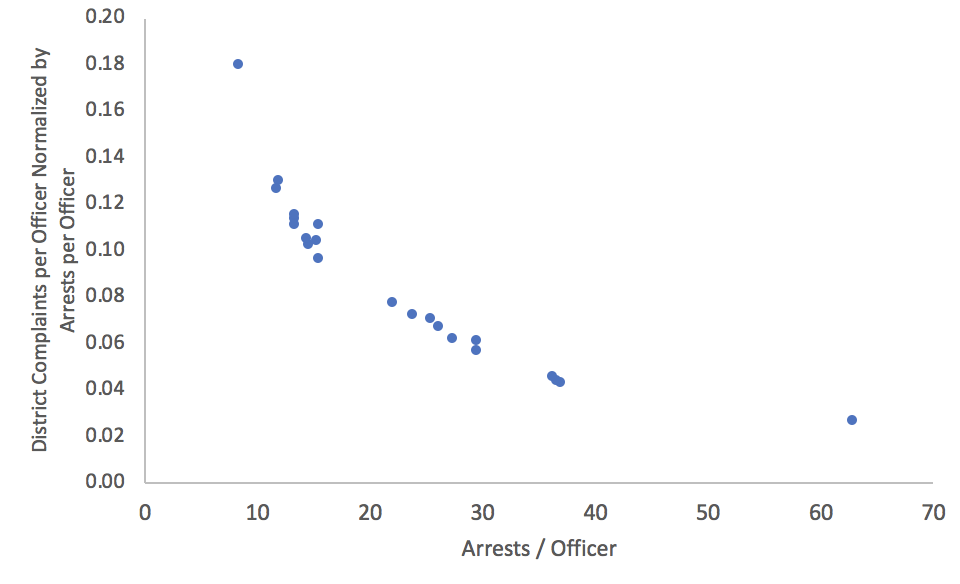
**Q #6**- *Are officers in more crime ridden areas more likely to have more complaints filed against them?*

Our next step was to normalize the number of complaints for each officer by the crime and arrest rate in each district. To do this, we calculated the following for each district:

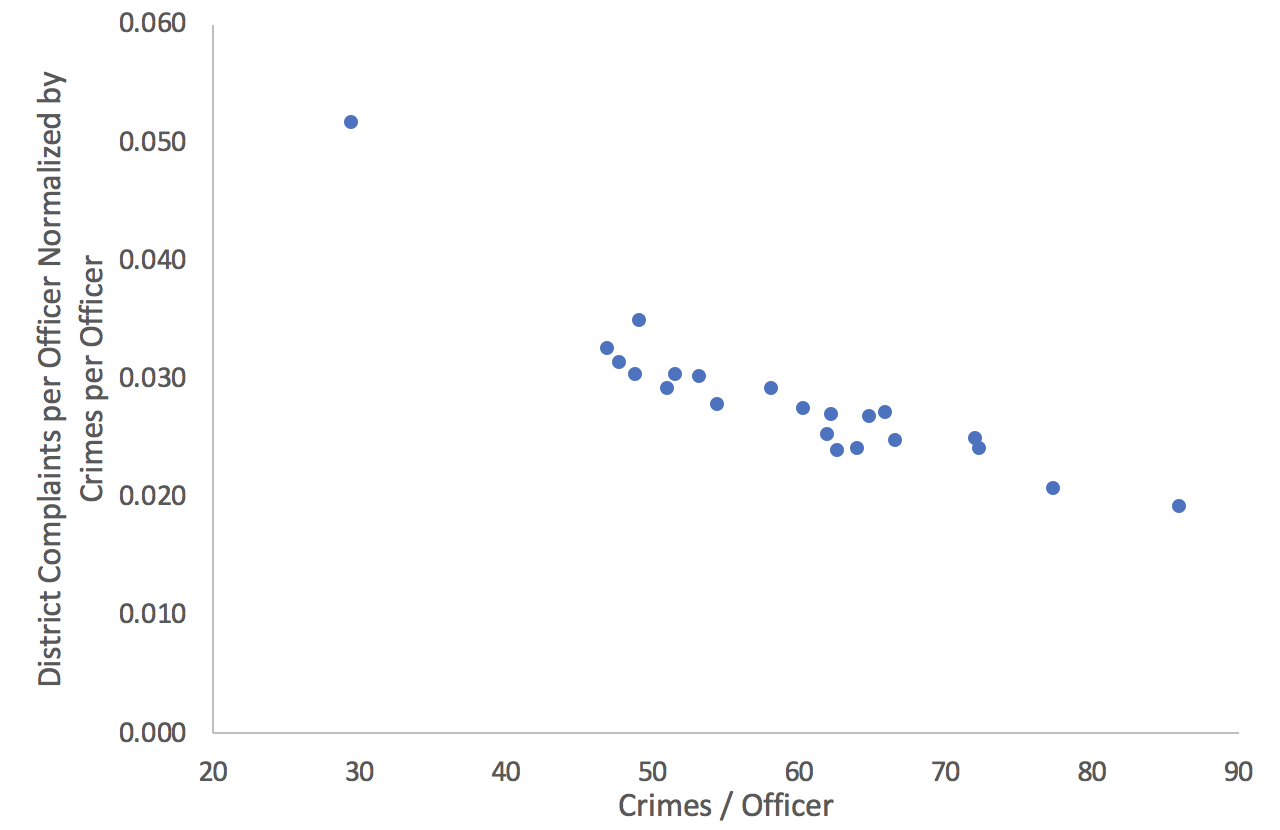
* The total number of complaints
* The total number of officers
* The total number of arrests
* The total number of instances of crime

We then created two graphs, which are shown below. Each point on both graphs represents a separate district. On the y axis of each graph, is the District complaints per Officer normalized by the number of crimes or number of arrests per officer. Then, on the x axis is the number of arrests or crimes per officer. This gives us a sense of how having more crimes or arrests per officer (an indicator of likelihood to use force) correlates to how many complaints are filed against associated officers on a per capita basis.

**Figure 2**: District Complaints Normalized by Arrests



**Figure 3**: District Complaints Normalized by Crime Data



One interesting finding that we were not expecting, was the trend of relative complaints as crime and arrest rates go up in the districts. We were expecting there to be a higher proportion of complaints for the districts that were in areas that had high levels of arrests and crime. But from our analysis, there is actually a negative correlation.

As we dug into question 7, which is shown below, we realized that finding an accurate solution to this problem is not possible given the dataset we have on arrests. There was not enough detail in the arrest dataset to link instances of arrests to the arresting officer, so we could not do this arrest analysis at the officer level.

**Q #7**- *For each officer, what is the relationship between claims and actually arresting civilians? Is there a pattern here for repeat offenders targeting specific populations? More explicitly, what are the demographics of the victims of officers who we see are exhibiting traits of being repeat offenders vs. their victims demographic information?*

Conclusion

One overarching takeaway from this analysis is the importance of normalizing when dealing with large datasets spanning large areas and groups of individuals. We found that rates of crime varied across the districts, so it would be unfair to compare the rate of complaints without taking their relative rates of crime into consideration. And after normalizing, we found that districts with high crime rates actually have a lower rate of complaints relative to their crime rate.