Deliverable 3

By: Shelli Gorokhovsky, Rachel Peng, Maya Webb, Tina Wang, Daniel Delijani

- Early Insights Presentation
- Final Report Draft

For this deliverable, we focused on completing step 3 of our suggested steps. However, another look at the merged contributions and disciplinary action dataset showed that there was an issue with the preprocessing and merge that was completed for deliverable 2. There was a small bug in the preprocessing of our Contributions dataset that didn't correctly account for entries that only contained first and last names. Because of this, during the fuzzy merge, we were actually trying to match the Disciplinary Action's last name's first character with the Contribution's first name's first character. Once we fixed the issue in the preprocessing, we re-ran the fuzzy merge and re-filtered for people who worked for the Boston Police. Our final filtered and merged dataset can be found here.

After locating our errors and fixing the bugs in our code, we were able to retrieve a final merged dataset containing the misconduct data along with the contributions. With this cleaned data, we redid the overlay of the LEAD blacklist data, which added 54 new names onto our dataset. Finally, we were able to get information about employee earnings from an earnings report made public by the Boston city government. After filtering out Boston police officers' earnings, we conducted another round of fuzzy matching to merge the updated dataset we created with earnings data.

Finally, we began conducting preliminary analysis on the finalized dataset. We computed the contributions to political campaigns made by all Boston police officers and all indicted Boston police officers respectively. This gave us some initial insight regarding the patterns of officers who make criminal offenses and their inclination to donate money to campaigns. We found that in our dataset, roughly a third of total contributions were made by indicted officers, which composed over half of the total officers from the set. To further find correlations between these two variables, we plan to conduct logistic regression and more analysis going forward.

Next Steps:

- Analyze discrepancies and leading features that cause officers under investigation to contribute politically. Setup a logistic regression to find what predicts an officer likely to politically contribute - rank, race, severity of misconduct. The severity of misconduct will be ranked with input from the client.
- Complete visualizations for each of the features: race, rank, severity of misconduct of BPD officers under investigation and how they contributed politically.
- Answer some questions:

- Do BPD officers under investigation attempt to influence the process via political campaign contributions to city lawmakers and decision makers such as city councilors and the Mayor?
- Do officers under investigation contribute more in total and more frequently?
- What are the discrepancies between officer rank in all BPD officers and BPD officers under investigation contributions?