Mijente: Police disciplinary action (Spring 2021)	
Team Members	Rachel Peng, Tina Wang, Shelli Gorokhovsky, Maya Webb, Daniel Delijani
Project Description	The client would like to understand how Boston Police Department officers under disciplinary investigation contribute politically to the Boston city council and Mayor. The analysis will overlay the Suffolk County DA office's LEAD dataset that has identified 54 BPD officers confirmed for misconduct. Central research question: How are BPD officers attempting to influence the internal investigations process through political contributions?
Data Sets	Data sets include: • 10 year database of Boston Police disciplinary action • Employee Earnings Report 2011-2019 (search for police) and payroll definitions • Campaign Finance Data (2010 - 2020) • Suffolk County DA Police Watch List • Race/Ethnicity data for the BPD personnel as of 09/04/2020
Suggested Steps	**Step one:** Obtain the 10 year database of Boston police disciplinary action from the Globe. The features in the collected dataset should include the following: Type of misconduct, rank, race/ethnicity, name, year, unique case ID, outcome of the investigation. **Step two:** Merge police officer names from the BPD disciplinary action dataset to the campaign finance data. For names that are not unique, verify them through the Race/Ethnicity BPD personnel dataset and if that's not possible flag them and exclude them from analysis. The output should be a merged table pulled from the BPD disciplinary action dataset and the campaign finance data with the name of the police officers as the key ID. **Step three:** Overlay the LEAD police blacklist officers onto the merged dataset using officer name, this should add another 54 BPD officers. **Step four:** Compute campaign contribution totals between all BPD officers and BPD officers under investigation. First find the total campaign contributions from Boston police officers to Boston city councilors and the mayor from 2010 to 2020. Define BPD officers using their employer and occupation fields in the campaign finance data, for unidentifiable officers mark the percentage that are missing and flag them for later research unless they can be matched to the BPD personnel dataset. Then find the total campaign contributions from BPD officers under investigation, what's the difference in campaign contributions between all BPD officers and BPD officers under investigation?

	Step four: Analyze discrepancies and leading features that cause officers under investigation to contribute politically. Setup a multi-linear 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. **Step five:** Complete visualizations for each of the features: race, rank, severity of misconduct of BPD officers under investigation and how they contributed politically. What are the differences in how officers of different ranks, races/ethnicities, severity of misconduct contribute?
Strategic questions	1. 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? 2. Do officers under investigation contribute more in total and more frequently? 3. What are the discrepancies between officer rank in all BPD officers and BPD officers under investigation contributions? 4. How are very severe allegations like rape and shooting affecting officers political contributions?
Issues and questions (List of limitations)	Does it make sense to include state legislative gifts as well? In my experience, when someone has a special treatment issue, they call whoever might help you, which is often a State Rep or Senator and not city councilor. Even without that, note that a political donation may not coincide in time with the internal investigation. Another potential issue that could arise is that we might find significant predictors or results with our multiple linear regression, so we might need to try different types of modeling.
Additional Information	**Tools & Methods** *Data scraping:* Selenium Webdriver, Scrapy, Beautiful soup, urllib2 for scraping and parsing data from the web. *Data pre-processing:* Pandas, NumPy for processing and cleaning the data. *Data Visualization:* Matplotlib, Seaborn, Tableau for all kinds of interactive visualizations
Meeting Schedules	Bi-Weekly meeting with PM and client: Wednesday 3PM EST Weekly meetings with only group members: Thursday 5PM EST