Data Science Project Proposal
By Lili Barsky, Vikram Kharvi, and Josh Cheema (Team Hungry Hornets)

Theme:
We would like to explore the relationship between officer compliance with department medical policy including but not limited to incidents of drug and alcohol abuse, with allegations of police misconduct, officer salary and rank, use of force by police, location and type of police work conducted, and distribution of awards given by the department. We will explore these relationships with other aspects of health as the data allows for.

Here are the specific questions we will aim to address at each checkpoint:

Relational Analytics:

1. What percentage of officers has at least one drug/alcohol abuse allegation?
2. Are drug/alcohol abuse allegations more likely to be made against on or off duty officers?
3. What percentage of officers are involved in medical policy violation allegations? These are categorized differently than drug/alcohol abuse.
4. Are there differences in officer salary and/or distribution of awards among police with and without drug/alcohol abuse allegations and other medical allegations?
5. Among officers with drug/alcohol abuse allegations, what is the average amount of time they have been on the force?

Visualization:

1. Are officers with drug/alcohol abuse allegations more or less likely to be suspended as a result of the fallout of the allegation, compared with officers with all other kinds of misconduct allegations? We will visualize this with two pie graphs next to each other, showing percentage suspended versus not suspended in each group.
2. Is the frequency of drug/alcohol abuse allegations versus all other allegations changing over time? We will visualize this with a connected scatterplot.
3. What is the relationship between drug/alcohol abuse allegations and adverse events, namely those that are proxies for referral to the Traumatic Incident Stress Management Program? We will treat drug/alcohol abuse allegations as a categorical variable (yes or no) and the adverse events as numerical, visualizing the two categories with adjacent bar graphs. We are attempting to get this information through a FOIA request and will provide this visualization if successful.

Interactive visualization:

1. Do drug/alcohol abuse allegations occur more frequently in certain neighborhood? If so, this could represent a function of heightened stress.We will explore the locations of these allegations with an interactive map that shows dots overlying the location of these

allegations. Hovering over these dots will provide additional information, and the graphic will allow users to zoom in or zoom out.

2. Within the broad category of drug and alcohol abuse, what is the breakdown of allegations filed under specific sub-divisions, including opioid abuse, cocaine use, and/or DUI? We will visualize this with an word cloud, changing the size/color of words related to the frequencies of these, and have an interactive feature where additional information will be provided on each sub-division based on user request.

Graph Analytics:

1. Is there a relationship between awards or lack thereof and propensity towards drug/alcohol abuse allegations, and how has that changed over the years? We will analyze this with a bar graph and investigate change over time.
2. How does the annual salary for officers with drug/alcohol abuse allegations vary over time?

Natural Language Processing:

1. Can we create a natural language model that allows us to ask questions directly of the database, and returns answer accordingly? For example, can we create a model than can answer the question “what is the most frequent allegation name for those that have allegations filed under drug and alcohol abuse?”