

**Team Hungry Hornets Project Proposal**  
**By Lili Barsky, Josh Cheema, Vikram Kharvi**

**Project 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:

**Checkpoint 1: 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 (+\$1012.55) and/or distribution of awards (1 less honorable mention) 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?

**Checkpoint 2: Visualization**

1. What is the frequency of each kind of drug/alcohol abuse and medical allegation made by officers as compared with civilians? We will visualize this with two adjacent word clouds, showing the relative frequency with which each allegation was made by officers and civilians, respectively.
2. Is the frequency of drug/alcohol abuse and medical allegations versus other allegations changing over time? We will visualize this with a connected scatterplot.

**Checkpoint 3: Interactive Visualization**

1. Does assignment toward a particular neighborhood yield higher drug/alcohol abuse and medical allegations among officers? This could represent a function of heightened stress. Alternatively, are officers with drug/alcohol abuse allegations more likely to work in certain areas? How do the demographics of each neighborhood correlate with the frequency of drug & alcohol abuse and medical allegations? We will explore this question by creating an interactive map in D3, in which viewers can hover over each neighborhood in order to see the frequency of drug & alcohol abuse and medical allegations.
2. How is the frequency of each type of drug/alcohol and medical abuse allegation varying over time? We will visualize this with an interactive word cloud built in D3, in which viewers can use a scrollbar to see how the word/phrase sizes change over time.

**Checkpoint 4: Graph Analytics**

1. Among officers who have drug and alcohol abuse or medical allegations against them, how often are they co-accused? How does this co-accusal pattern vary based on whether officers earn below or above the average annual salary across the full department? We will compare this using network analytics, specifically the triangle count and PageRank score.
2. Among officers who have drug & alcohol abuse and medical allegations against them, how does the co-accusal pattern compare between more and less decorated officers? Specifically, how does the co-accusal pattern compare between those with more and less than the average number of award types won, respectively? We will compare this using network analytics, specifically the PageRank score.

**Checkpoint 5: Natural Language Processing**

1. Can we build a natural language processing model that allows users to ask questions directly to it, in order to retrieve information on officers who have drug and alcohol abuse and medical allegations against them? The aim is to create a tool that would allow novice users to find out more information about these individuals without coding experience. We will do this by adapting the TAPAS language model created by Google and available publicly. Due to constraints in computational power with the model itself, we will limit our analysis to officers with more than 2 allegations of drug and alcohol abuse, of whom have had at least 1 sustained allegation and have a numerical value listed for their current salary. We will measure the success of our approach by asking the following questions to the model across every single officer and testing the overall accuracy per question. Here, [X] represents each individual officer in our table.
  - a. What is the race of [X]?
  - b. What is the gender of [X]?
  - c. What is the birth year of [X]?
  - d. what is the allegation count of [X]?
  - e. What is the sustained count of [X]?
  - f. What is current salary of [X]?