Hungry Hornets Project Proposal: Updated
Lili Barsky, Josh Cheem, Vikram Kharvi

Checkpoint 1:

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?

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?

Checkpoint 2:

1. What is the frequency of each kind of drug/alcohol abuse and medical allegation made by officers as compared with civilians?

2. Is the frequency of drug/alcohol abuse and medical allegations versus other allegations changing over time?

Checkpoint 3:

1. Is the frequency of drug/alcohol abuse and medical allegations versus other allegations changing over time?

2. How is the frequency of each type of drug/alcohol and medical abuse allegation varying over time?

Checkpoint 4:

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?

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?

Checkpoint 5:

Approach: We would like to build an NLP algorithm that performs relational analytics on a consolidated table of data allegations, to answer the following queries:
1. 1- column analysis: Among officers with drug and alcohol abuse and medical allegations, who has the highest salary?
2. 2-column analysis: What are the names of officers who have 5+ drug and alcohol abuse and medical allegations?

Evaluation Method: We will compare how the NLP algorithm performs, when created using SQL vs. on Colab. The evaluation will be performed using a confusion matrix.