A look into Chicago's Police Officers

Sarah Ahmad, Vikram Kohli, Junhao Li

Theme: Learning about the officers involved in complaints

Topic: Importance of officer life

Interactions b/w cops and civilians are shaped not just by actions but a variety of discrepancies in social location including

- Race
- Gender
- Socioeconomic location
- Age

A lot of the time, studies focus on the **standpoint of the accuser**; e.g. the disproportionate police brutality against black people but very little research is done with respect to officers

Our goal was to figure out how these various factors, particularly **relating to the officer**, implicated # of allegations, intensity of settlements, length of appointment and other factors.

Relational Analytics

(1.) **Clear** differences between types of allegations certain groups are accused for.
- Black women officers are an **outlier**
- **Racial profiling** interesting results

(2.) Disparity in allegations based on race;
- 59% of allegations were **from** black complainants; only 23% were **against** black officers
- Only 21% of allegations were **from** white complainants, but 61% were **against** white officers
- This may just be because of

(3.) TRR counts by race and gender **mostly matched** distribution of complaints by race and gender; with the exception of **black men** officers.

Data Integration – approach

For data integration, we looked at integrating:

- CPDB
- Settlement database
- Arrest database.

Joining and cleaning in Trifacta. We joined the CPDB and settlement database using officer names and joined the CPDB and arrest database by beat.

We cleaned the databases (e.g. translating names to uppercase) and computed pivot tables for answering our questions

Data Integration – findings

(2) Cost to city in settlements / # of complaints vs. race and gender:

- Ratio of cost to city and number of complaints did not vary much across race and gender
- At less than 25 complaints, female officers surprisingly had some of the highest cost to city sums

(4) # of allegations before 1st settlement vs # of allegations after:

- Officers have about the same counts of allegations after 1st settlement

Spark Analytics – approach
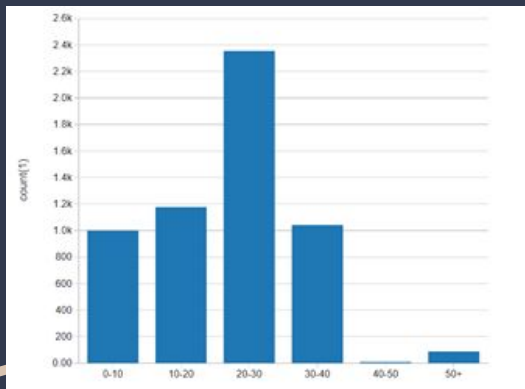
Questions:

- Is there a correlation between race and gender vs. length of appointment?
- What are the race and gender of the officers with the most allegations by beat?

We prepared our tables through Postgres and Trifacta. The tables were loaded into Databricks and explored the through queries.

Visualization. We used the readily available visualizations in Databrick to understand the data..

Spark Analytics – findings

(1) Is there a correlation between race and gender vs. length of appointment?

- No major differences in the distribution of length of appointments across gender or race

- Matches with the similar rates of discipline after allegations across gender we found previously

(2) What are the race and gender of the officers with the most allegations by beat?

- Black officers made up 25% of officers with the highest allegation counts in their beat despite being 15% of all officers

Machine Learning – approach

Is it possible to predict whether an officer will be the subject of a settlement using the details of their complaint history?

- Machine learning model: decision tree
- Attributes: frequency of complaints and nature of complaints
- Used 2009 as proxy to simulate predicting whether officer would be involved in a query after 2009
- 70% training set, 30% testing set

Machine Learning – findings

- Decision Tree accuracy: 72.4%
- Cases where officers would be involved in settlement:
    - If the number of allegations was less than 4.5 and the officer was involved in bribery/official corruption
    - If the number of allegations was greater than 4.5 but less than 10.5
    - If the number of allegations was greater than 24.5, all types of complaints would result in a settlement except use of force, conduct unbecoming, and first amendment.
- From checkpoint 2: wasn't a correlation between the amount that an officer owed in a settlement and the number of complaints an officer was involved in
- how much an officer owes in a settlement and whether an officer will be involved in a settlement are not related by number of allegations

Neural Network – Approach

Our **goal** was to build a predictor to determine, based on the **text of summaries** from previous allegations, whether an officer involved in a complaint would receive more complaints in the future.

Data formatting: Similar to the previous section, we used 2014 as a proxy in order to simulate predicting whether an officer would be involved in a query in the future.

Building the network:
https://github.com/huggingface/pytorch-openai-transformer-lm/blob/master/model_pytorch.py

Neural Network – Findings

Validation accuracy: reached a maximum of **78%**.
    - Relatively high accuracy rate; does significantly better than ZeroR, and is able to determine the correct answer around ¾ of the time.

Test accuracy: only around **62%.**
    - Notably pretty bad since this was <u>worse than ZeroR</u>; running more iterations seemed to improve this, but only very slowly.

Due to the classifier being prohibitively slow and opaque, was difficult to detect false positives, improve performance, etc.

Visualization – approach

Initial question: Relationship between the race of the complainant and the nature of the complaint.

- Split bar chart in Tableau using race of complainant and complaint category as dimensions and number of records as measure
- Decided to change our question to look at the race of the officer in order to better fit our theme, but would file our findings on complainant race for later
- Compared different visualizations such as side-by-side bar charts, split bar charts, and packed bubbles.

Visualization – findings
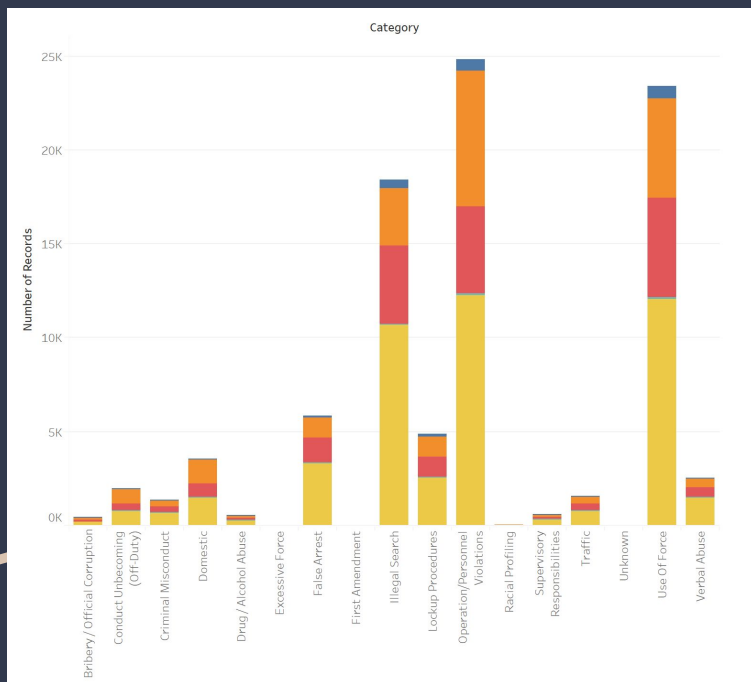
- Overall fairly similar distribution over each nature of complaint
- Two interesting cases: Conduct Unbecoming (Off-Duty) and Domestic
- Interesting outside finding from ABC 7 News: "Most Chicago police accused of domestic violence go undisciplined"

Future Research

- Doing more normalization on our findings using the demographics of Chicago's police officers

- Comparing our findings on the CPDB to other cities
    - Are there unique characteristics to the Chicago force?

- Looking at the frequency and distribution of allegations across a police officer's career (and the implications of social location on these)

- Experimenting with adding more features to our neural net besides summary (such as race and age)