

CPDB Project - Checkpoint 05

Devyani Gauri, James Wilkinson, Kaleem Ahmed

Data Science 339
Prof. Jennie Rogers
November 29, 2021

Questions

Q.1: Does this code compile and run without significant effort on the part of the reviewer?

Ans: Yes, please refer to the README file and the Google Colab URL attached in the submission folder.

Q.2, 3: Quantitative, Qualitative Analysis. Does the code answer the questions from the proposal? Does the analysis described in the write-up adequately analyze the problem we are investigating? Was it thoughtful and thorough? Does it reveal any interesting open questions? (last one not required)

Ans: Our code answers the following question from the proposal:

Do complaints against officers get more negative over time during their career?

NOTE: We recognize that the current dataset only has 1926 entries and our analysis is based on the small dataset we have, however, our pipeline could produce different results given more data, and that would be a great step towards future research for this project.

We tackle this question by performing sentiment analysis on unstructured text from the summary column of the data_allegation table and dividing officers into groups based on their time in force at the time of the incident. We then perform comparative analysis using aggregates of sentiment scores received for complaints for these officer groups. Since the summary text from data_allegation doesn't have pre-defined labels, we performed unsupervised sentiment analysis on the text using the nltk VADER module that comes with a sentiment intensity analyzer. Since it doesn't require a large amount of text pre-processing and uses capitalization of certain words and their intonations within the written text to produce the sentiment scores, it is helpful in analyzing text in its natural form.

The VADER module produces compound scores by normalizing the positive, negative and neutral sentiments it identifies in any given text, and those compound scores act as our sentiment scores. Since we do not have enough data for each officer, we group them based on their time in force at the time of the incidents that have been reported in the data_allegation table. We use the pandas qcut method to find quartiles that divide the data approximately equally as follows:

```
bin_group.compound.count()
```

```
bin_code
0      482
1      481
2      481
3      482
Name: compound, dtype: int64
```

Here, the 4 quartiles are based on the following ranges for time in force:

0: 5.999 to 2047.75 days ~ up to 5 years 7 months

1: 2047.75 to 3918.0 days ~ between 5 years 7 months and 10 years 9 months

2: 3918.0 to 6032.5 days ~ between 10 years 9 months and 16 years 6 months

3: 6032.5 to 14423.0 days ~ between 16 years 6 months and 39 years 5 months

An interesting observation that pops up almost immediately upon looking at the data is that the last quartile bins officers with time in force between 16 years all the way up to 39 years. We also created a group of rookie officers (<1 year of employment) from the dataset and it only had 24 entries. Both of these indicate the sampling problem in the data_allegation table. With a well balanced dataset, our analysis could show more significant results. Here, we also note that we noticed a singular outlier that had a negative time of employment:

```
df[df['time_of_employment']<0]
```

officer_id		summary	incident_date	appointed_date	time_of_employment	cleaned_summary	scores	compound
1710	30806	In an incident involving an on-duty CPD Probat...	2004-12-08	2014-04-28	-3428	In an incident involving an on duty CPD Probat...	{'neg': 0.085, 'neu': 0.848, 'pos': 0.067, 'co...	-0.128

After discussion, we decided to drop this row, however, we believe that this could be due to a data entry error where the year in incident date was entered as 2004 instead of 2014.

With our dataset of 1926 entries, we used the following aggregates to statistically compare the various bins that the officers were divided into:

bin_group.compound.mean()	bin_group.compound.median()	bin_group.compound.std()
bin_code	bin_code	bin_code
0 -0.763613	0 -0.95205	0 0.425798
1 -0.800647	1 -0.95490	1 0.358609
2 -0.781778	2 -0.93200	2 0.362523
3 -0.737321	3 -0.89410	3 0.401559
Name: compound, dtype: float64	Name: compound, dtype: float64	Name: compound, dtype: float64

From the means and the medians for the quartiles, we can clearly see that the sentiment scores stay basically consistent over the time that officers spend in the force. We do notice marginally more negative sentiment for officers in the second bin (bin number "1", representing 5yrs 7months - 10yrs 9 months), which aligns with our prior analysis around use of force from prior checkpoints, which indicated an increased uses of force in this cohort.

Overall, the small amount of variation in sentiment between the four bins could be due to a few reasons:

We don't know how the complaints are summarized .i.e. Whether it is based on a form filled out by civilians (since the complaints seem to be similar in formats) or whether it is

free text. If officers are responsible for summarizing complaints, they could take liberties in wording the incident the way they want and bring up unconscious biases.

If the text is written directly by civilians, this could either mean that an officer's time in force does not impact how complaints against them are worded, or it could be due to the way this data is sampled, with a huge chunk of data coming from officers that are neither too new to force nor too old. We believe that with more data to work with, our pipeline could produce statistically significant results, whether it reinforces the results we have here, or by showing a certain group of officers having more negative sentiment scores.

In conclusion, our analysis of unstructured text data within complaints gives limited insight into police officers' learned behaviour patterns. Whilst any differences in sentiment are marginal across the different officer cohorts, there is an indication that complaints against officers that have been in the force between 5 years 7 months and 10 years 9 months give rise to the most negative complaints. This complements our prior analyses which found that officers within this same cohort are more prone to having TRRs attributed to more severe uses of force compared to other cohorts (in particular, the more junior cohorts). While this NLP analysis is limited, we see a small increase in evidence that officers' misconduct is a learned behavior that exacerbates in severity as officers progress from junior, to more senior officers, with a peak between the 5-10 year period of time-employed in the CPD.