Checkpoint 4 - Graph Analytics

INTRODUCTION

Chicago PD offers rewards to their police officers to recognize their efforts on the job. Civilians file complaints against police officers based on their encounters. We compute a metric which we call ‘Complaint Severity Score’ or CSS based on the average annual complaints of an officer multiplied by custom defined weights based on the complaint category. The weights are defined based on research by our previous cohort and ranges between 0.0 - 43.0. With our analysis on whether CSS is correlated with the awards police officers receive, we hope to understand whether ‘good’ police officers i.e. officers with low CSS scores are rewarded by the Chicago PD more as compared to cops with higher CSS scores.

MOTIVATION

We want to know whether the Chicago PD rewards its good police officers and whether their first degree connections share similar qualities. We have observed that police officers with a high number of complaints filed against them have been rewarded by the Chicago PD. It would be interesting to observe whether good police officers are also rewarded. With graph analytics, we aim to observe the CSS and rewards for first degree connections of police officers.

ANALYSIS

We categorize the police officers under buckets based on their CSS percentile. We build a graph network by selecting the nodes as police officers and their edges (relationship) defined by both officers being co-accused by the civilian. Column ‘allegation_id’ in the table ‘data_officerallegation’ is used to create the edge. Therefore our nodes dataframe looks like this -

TABLE 1

id	officer_name	allegation_count	CSS	award_count
32312	Randall Ryan	10.0	4.2757	35
32358	Kevin Stoll	24.0	9.8512	102
31804	Diana Anderson	1.0	0.252	13
8175	Christina Fabian	11.0	2.3819	6
9534	Gerardo Garcia	10.0	1.5739	8
25219	Linda Salustro	8.0	1.52	4
15636	Marie Lane	8.0	1.9359	8
3613	Patricia Cain	4.0	0.7031	7
31793	Adam Aleszczyk	3.0	1.1367	62
4022	Michael Carroll	6.0	1.8722	11
32005	Tony Green	11.0	4.0107	17
32080	Eric Johnson	5.0	1.323	18
32188	Elmore Metcalfe	27.0	8.619	17
30482	Evelyn White	5.0	0.8569	5
32308	Giselle Ruiz	2.0	0.2719	40
27908	Flora Suttle	5.0	1.1471	5
2642	Leatheia Brady-Rh...	3.0	0.7727	14
32434	Steven Yee	2.0	0.3359	25
32263	Jeffrey Pineda	1.0	0.8333	21
2735	Charles Breckenridge	10.0	2.7615	6

We divide ‘src’ police officers under different buckets based on CSS percentile
TABLE 2 represents the same for 10th, 25th, 50th, 75th, 95th and 99th percentile-

TABLE 2

percentile_approx(CSS, 0.1, 10000)
0.4326

percentile_approx(CSS, 0.25, 10000)
1.1363

percentile_approx(CSS, 0.5, 10000)
2.7191

percentile_approx(CSS, 0.75, 10000)
5.7014

percentile_approx(CSS, 0.95, 10000)
13.7504

percentile_approx(CSS, 0.99, 10000)
22.7343

Unexpectedly, we observe that as CSS of ‘dst’ police officers (i.e. first degree connection of ‘src’ police officer) increases, the average reward count for that bucket also increases.

TABLE 3

avg_first_degree_awards_10_pct
21.860159803955916

avg_first_degree_awards_10_to_20_pct
25.46627998993069

avg_first_degree_awards_20_to_30_pct
30.047711027528198

avg_first_degree_awards_30_to_40_pct
32.95962149742117

avg_first_degree_awards_40_to_50_pct
36.791168060871605

avg_first_degree_awards_50_to_60_pct
39.66095448761085

avg_first_degree_awards_60_to_70_pct
44.538497022665915

avg_first_degree_awards_70_to_80_pct
49.864908735240974

avg_first_degree_awards_80_to_90_pct
55.292877540050966

avg_first_degree_awards_90_to_95_pct
59.63980966451585

avg_first_degree_awards_95_to_99_pct
65.3961662025186

avg_first_degree_awards_99_above_pct
67.85431130525008

TABLE 3 shows that officers with high CSS have received progressively more rewards.

Furthermore, we use PageRank GraphX algorithm to observe the most connected (i.e. they have been listed along with other officers under complaints filed by civilians) police officers having a high pagerank score and observe their average CSS. We carry out the same steps for the least connected police officers.

It is interesting to observe that the CSS value increases as we move from least connected to most connected police officers as shown in TABLE 4. Although the reward counts are high for most connected police officers. TABLE 5 shows that most connected officers have a high CSS but also a high award count. TABLE 6 shows the opposite trend where the least connected officers have a low CSS but also a low reward count.

TABLE 4

avg_css_25_pct
2.0275589068029927

avg_css_25_to_50_pct
3.7916747838050373

avg_css_50_to_75_pct
4.51714002751031

avg_css_75_to_90_pct
5.777709070072033

avg_css_90_to_95_pct
6.690802355250243

avg_css_95_above_pct
10.048070461236486

TABLE 5

id	officer_name	allegation_count	CSS	award_count	pagerank
32442	John Zinchuk	23.0	12.8494	97	122.85961890322304
32440	Mark Zawila	34.0	13.4627	25	82.596692708314
32425	Perry Williams	27.0	11.5465	118	72.0279045279787
32410	Joseph Watson	29.0	10.8425	62	68.47949785853883
32430	Michael Wrobel	22.0	16.7592	165	67.9187831349547
32350	Robert Spiegel	20.0	13.0244	183	65.39503775647411
32351	Boonserm Srisuth	25.0	11.0281	75	63.01146602153837
32433	Kenneth Yakes	29.0	9.8227	53	62.051713039664
32284	Mark Reno	76.0	32.1284	142	61.834556455615996
32419	Eric Wier	18.0	7.1819	90	58.028946264273195
32384	Edwin Utreras	47.0	20.8848	43	57.37383713588734
32435	Mohammad Yusuf	22.0	15.2373	148	56.708114377778166
32074	Ronald Jenkins	46.0	13.5129	64	55.47174776302173
32431	Albert Wyroba	15.0	9.888	200	55.44272739297056
32337	Louis Silva	21.0	14.8829	131	55.12138144372036
32413	Carl Weatherspoon	69.0	23.1244	55	54.75312241275796
32289	John Rivera	44.0	12.922	66	53.69168834565139
32401	Joshua Wallace	45.0	22.8745	50	53.6111678756193
32375	James Triantafillo	31.0	18.6742	108	48.42330492775258
32436	Edmund Zablocki	28.0	16.2565	51	46.89728707125255

TABLE 6

id	officer_name	allegation_count	CSS	award_count	pagerank
9427	Robert Gallegos	1.0	0.2248	0	0.323512343251966
30702	Bobby Williams	1.0	0.1727	0	0.323512343251966
812	Patrick Arens	10.0	1.0	6	0.323512343251966
4184	Edward Castellano	1.0	0.3044	1	0.323512343251966
9534	Gerardo Garcia	10.0	1.5739	8	0.323512343251966
16767	Gregory Luszowiak	2.0	0.6833	26	0.323512343251966
32434	Steven Yee	2.0	0.3359	25	0.323512343251966
32263	Jeffrey Pineda	1.0	0.8333	21	0.323512343251966
16144	Don Lewis	2.0	0.4546	2	0.323512343251966
32448	James Martin	5.0	2.0261	0	0.323512343251966
25570	John Scatchell	1.0	0.1827	4	0.323512343251966
3279	Robert Bullock	12.0	2.8877	38	0.323512343251966
178	Richard Aguilar	38.0	8.8356	4	0.323512343251966
545	Alvin Amos	15.0	3.2273	6	0.323512343251966
1640	Sanford Becker	1.0	0.1291	0	0.323512343251966
28534	Jeffery Thompson	4.0	0.5677	3	0.323512343251966
11388	Leon Hardeman	1.0	0.1815	3	0.323512343251966
17954	Garry Mc Carthy	2.0	2.831	2	0.323512343251966
15381	Donald Kumiga	2.0	0.2963	9	0.323512343251966
6021	William Czahor	3.0	0.4473	0	0.323512343251966

CONCLUSION

We observed that ‘good’ cops i.e. those belonging to buckets of low CSS percentile have a lower reward count. The reward count increases as the CSS of cops increases i.e. ‘bad’ cops are more rewarded which is a troubling sign. Furthermore, the most connected officers i.e. those who have multiple counts of being co-accused with other officers have higher rewards as compared to the least connected officers. ‘Bad’ cops stick together and repeat offenses, yet they are heavily rewarded. However we see that ‘good’ cops are not rewarded a lot. The Chicago PD can take CSS into consideration while deciding on award recipients.

Submitted by: Keshav Bhandari
Dhruv Narayan
Akshay Kumar