

The Glorious Gorillas  
Data Science Seminar  
Checkpoint 1 - Relational analytics

We are interested in investigating the trends between the seniority of an investigator in terms of their age and rank, and their history of complaints. The focus of this investigation is twofold: First, how does the history of complaints correlate with each investigator over time? For example, as an investigator rises in the ranks, does the rate of complaints increase for this officer? Similarly, is there a trend in the type of complaint that the investigator receives? Second, can we understand potential biases that the officers might hold against their victims?

Completing Checkpoint 1 helped our group to find concrete data surrounding aspects of our theme which can be further analyzed for insights.

The first set of data we sought to gather was the total amount of complaints an officer had over their career. This helps to showcase a timeline of complaints as measured over an entire officer's career. This is supplemented by the number of complaints per officer every year, over an entire career. Both of these values can be cross referenced and compared between officers, as well as against their own timeline of complaints. The query below gives the count of allegations by officer id. Here we can see that officer 8562 has recorded the highest amount of complaints versus all other officers. This information can warrant a deeper investigation into these particular officers, and so thus we can narrow down why the volume of complaints are so high at the officer level. The one below it addresses the count of complaints for that officer per year that they worked, so that we can investigate even deeper into a per-year basis.

	officer_id ↕	count ↕
1	8562	175
2	21837	137
3	17816	136
4	8138	132
5	21468	125
6	28805	123
7	31631	117
8	29033	114
9	32166	109
10	4807	109
11	31119	104
12	13788	104
13	32164	103
14	10890	102
15	32265	101
16	3897	100
17	13095	98
18	11634	97
19	2015	96
20	23265	95

	officer_id ↕	year ↕	count ↕
1	13095	1996	28
2	32166	2000	28
3	13788	2000	26
4	12074	2003	26
5	13169	2000	25
6	13788	2001	25
7	29033	2000	25
8	21974	2000	25
9	8562	2004	24
10	8562	2000	24
11	4807	2000	23
12	12825	2003	23
13	13330	2000	22
14	20713	1991	22
15	32265	2002	22
16	28805	1994	21
17	29033	1997	21
18	23207	1996	21
19	12479	1991	21
20	12478	2001	21

We also wanted to understand the nature of complaints for these officers. We created a query to find the most targeted demographic groups per officer. For each officer, one can see the most commonly victimized demographic group along race, gender, and (eventually) age groups. This helps us understand if an officer has any biases against certain demographic groups. The sample output below shows the amount of complaints each officer had, race-wise. Here we see officer 8138 again, and the race of the people that filed under those complaints. This can help

us investigate potential causes for racism or bias throughout these officer's careers.

	frequency ↕	officer_id ↕	race ↕
1	107	8138	Black
2	97	13937	Black
3	95	25898	Black
4	92	31847	Black
5	91	32164	Black
6	88	21098	Black
7	87	4118	Black
8	87	32159	Black
9	86	16551	Black
10	85	32016	Black
11	85	11847	Black
12	84	3897	Black
13	83	32383	Black
14	81	21124	Black
15	81	17647	Black
16	80	16385	Black
17	80	27415	Black
18	79	14442	Black
19	77	17613	Black
20	76	32116	Black

The next query is to answer the same question, except on the gender of the complainant:

	frequency ↕	officer_id ↕	gender ↕
1	86	3764	M
2	82	13937	M
3	78	3605	M
4	78	14442	M
5	75	25898	M
6	72	3897	M
7	72	32159	M
8	72	16385	M
9	72	29670	M
10	72	31847	M
11	69	16551	M
12	69	21124	M
13	68	32383	M
14	66	32164	F
15	66	8138	M
16	63	21098	M
17	60	3056	M
18	58	7603	M
19	57	17613	M
20	56	8668	M

Lastly, we are trying to ultimately understand the severity of complaints against an officer over a career timeline. In Checkpoint 1, we pulled data for the most common accusation category for each officer over their entire career (much like demographic bias testing). We aim to consult

with the Invisible Institute to create a thorough ranking system to append severity of offenses to the database so it can be further analyzed with queries. Aside from understanding severity, we are also able to see which offense these officers are most frequently getting flagged for, which might give useful information into the personalities and habits of these officers. That is, since officer 8562 has been cited for illegal search 72 times, perhaps there is more to the officer's behaviour than just coincidentally being reported for the same offense across multiple people.

	frequency ↕	officer_id ↕	category ↕
1	72	8562	Illegal Search
2	57	8138	Use Of Force
3	55	31119	Illegal Search
4	54	29033	Use Of Force
5	51	17816	Illegal Search
6	50	17041	Operation/Personnel Violations
7	49	32164	Illegal Search
8	49	21468	Use Of Force
9	47	28805	Illegal Search
10	47	32265	Use Of Force
11	45	32336	Illegal Search
12	45	31631	Use Of Force
13	45	23841	Illegal Search
14	44	12478	Illegal Search
15	43	15760	Operation/Personnel Violations
16	43	27778	Illegal Search
17	42	32166	Illegal Search
18	42	4807	Traffic
19	42	10890	Use Of Force
20	42	13391	Use Of Force

Additionally, investigating this topic may uncover related trends that may not necessarily be confined to the scope of this topic, but are interesting and important nonetheless. For example, trends between officers that rise in rank and their complaint volume might reveal some information about how the Chicago PD operates as an institution. That is, an officer may receive complaints from the public after an incident, but may be awarded or promoted internally but the institution. We will keep a close eye as to what the data is telling us and try to uncover analytics that we might not specifically be looking for.

