Checkpoint1 Finding

I also write in README.md that you can execute all the SQL statements all in one by typing :

> *psql cpdb < code.sql(remember you need to be in the src/ directory to do this!)*

But there are specific comments for each line of code which make the code more readable. You may refer to these comments.

First I'd like to give answers from my proposal questions:

1. what's the relation between average number of complaints received per year and the age of a police officer? older - >more complaints or something else?

   the young and too old police officers have much lower frequency of receiving complaints, but for those who are born between 1982 and 1941 have average **0.8** complaint receiving rate per officer and per year.

   which suggests those police officers who just start their career are good cops and devote to their work. As time pass by, they could be affected by others and start receiving complaints. As for those born before 1940, I guess because of insufficient recording of the complaints, their records of complaints are lower. Or simply in the past, Chicago PD is far better than now.

2. Do those who receive no complaints get more awards?

   Sad results. 2/3 of the Chicago police officers have received complaints, but they receive more than 90% of the awards (656116/(40073 + 656116)).

3. what's the average number of complaints received per year and officer?

   Chicago PD has received 244455 complaints in total. They have 22813 + 10858 = 33671 officers. And the dataset records the past 50 years complaints data. so the average number of complaints received per year and officer is : 244455 / (33671 * 50) = 0.145

4. what's the average number of complaints from different racial officers per year?

RACE	OFFICER_COUNT	NUMBER_OFFICER_COMPLAINED	NUMBER_COMPLAINT_COUNT	RESIDENT_COUNT	COMPLAINT_COUNT_FROM_RESIDENT
white	20671	13373	134710	1729593	17961
black	7630	5697	66854	1712474	50322
hispanic	4579	3290	38342	1569483	8494
asian	539	385	3808	305590	712
native american	62	53	648	4097	109

   from the table we can conclude easily that there are racial discrimination when police officers deal with black residents. Hispanic and Asians may not want to spend much time on complaining police. Their complaints/population ratio is relatively

lower. But for our question, it is clear that police of different races have similar ratio of being complained. White officers have the lowest ratio of being complained(less than 0.7).

5. what's the average income of those officers who have received complaints and the average salary of the whole Chicago police officers?

   Sadly like the awards question, the average salary of those who ever received complaints are even a bit higher than the total average. I guess the Chicago PD won't give salary deduction as punishment for those who receive complaints. The Chicago PD may not take those complaints seriously.

Extra thoughts for checkpoint-1:

1. The most common locations of allegation are Public way, private residence or police buildings
2. Among all ranks of officers, they have similar frequency of being complained(0.8), which suggests that the rank does not affect their work attitude? For those who rank high, the variation is because of too few samples.
3. I try to get some statistics of those who receive more than 50 complaints. The distribution is like the distribution of ranks in all officers. However, the director of caps also receives more than 50 complaints, which is interesting.
4. For my final goal of predicting settlement amount for each case, we need more feature engineering. From checkpoint 1, it seems many features are not that useful like officer's ran and salary. I need to integrate with the misconduct data for the details of misconduct to make prediction.

Because my teammate dropped MSAI339 and I am working on my own, I do not have much time to make more extra explorations. During data visualization part, more interesting points may be found out.