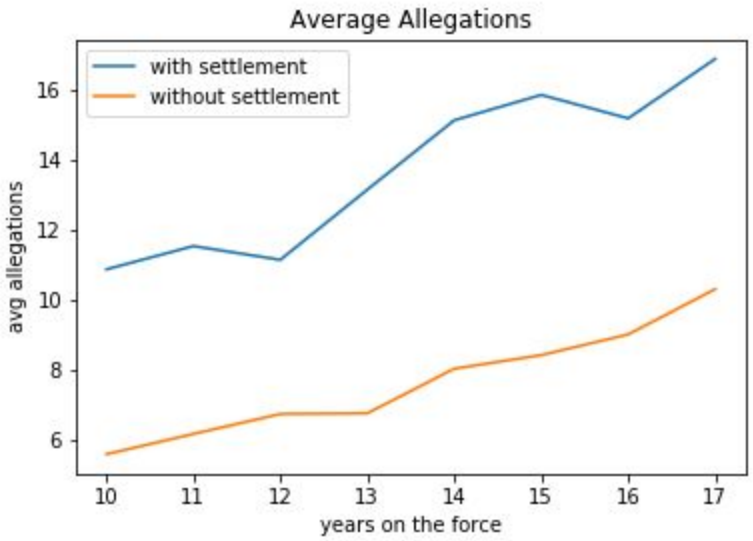
# Checkpoint 3: Data Integration

The Wise Lobsters

## Question 1

**What is the average number of complaints against an officer with any type of settlement compared to officers with no settlements by years on the force?**




**Overview**

For this question the subset of officers from checkpoint 1 was split into two groups, those with a settlement payment and those without. For those without a settlement, officer with no allegations are excluded. Then officers were grouped by years on the force and averaged allegation counts. Payments records were used to indicate a settlement so that we were only looking at settlements with an outcome. Not necessarily paid out, but with an amount settled upon. For reference, the subset of officers is currently active officers that started between 1/1/2000 and 12/31/2007. That time frame is used to have officers with at least 10 years experience and a consistent view of the early years of their career.

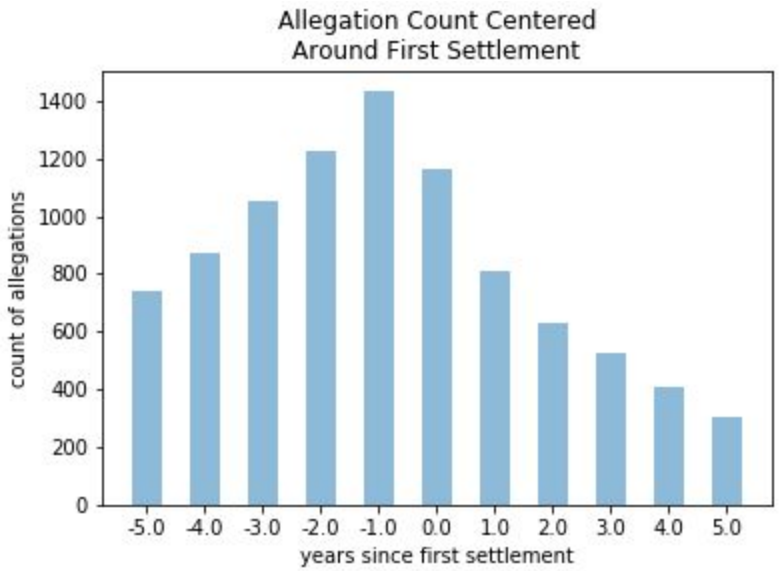
**Analysis**

The trend of this graph follows an expected pattern, officers with settlement have a much higher average allegation count than those without any settlements. It also shows that there is a slight increase in the number of average allegations over time for officers with no settlements. With settlements has a range from 10.89 to 16.89 (6 allegation spread, or 35% increase) compared no settlements with to 5.83 - 10.43 (4.6 allegation spread or 44% increase).

**Relating to Theme**
The goal of this question is to determine how having a settlement will affect the likelihood for an officer to increase allegations over time. While the findings do not show a large difference in the trend over the years on the force there is a slight difference. For our final analysis of exploring events in early officers careers that lead to repeaters this could be a small predictor. To bring this together with previous analysis we can start to look and see if the trend changes for certain types of allegations.

# Question 2

**After a settlement, does the average number of allegations decrease?**




**Overview**
This question requires getting the date of an officer's first settlement. Settlement file date was used for this question (`cases_case.date_filed`). Settlements tend to drag on for a long period of time and assuming an officer would know when a settlement is filed, or at least relatively close to that time. Once a first settlement date is identified, all allegations (`data_officerallegations`) for an officer can be stamped with a relative time difference from that date. Years were used to simplify the analysis. 0 is the same year as the first settlement, -1 is within the year before the settlement and 1 is the year after the settlement. The graph is limited to 10 year so that the counts are consistent with the subset of officers.

**Analysis**
This shows a very interesting trend. From this analysis, officers do tend to decrease the number of allegations after their first settlement is filed. Note that these officers are active at the time of the last data pull, so there are no cases of officers begin fired as a result of an allegation or settlement. Allegation counts steadily increase until the year a settlement is filed, then decrease at a greater rate than the increase (48% increase, 78% decrease from the high point). It's

assumed that year 0 follows the downward trend given that the trend started around the time the settlement is filed, which is sometime within year 0.

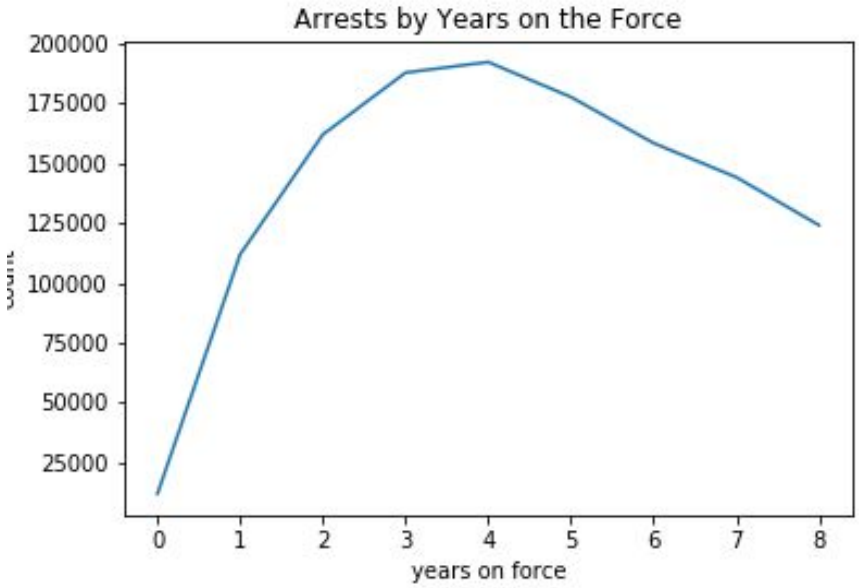
**Relating to Theme**
This is a strong trend that can be used in our final analysis to indicate events that trigger behavioral changes in officers on the path to a repeater. This analysis alone is a decent indication that a settlement decreases allegations and the lack of a settlement tends to increase. Although, settlements really can only serve as an indication. Settlements are not a form of punishment and obviously something that the city wants to avoid at all cost. Perhaps a better metric is the likelihood of a settlement. Further research into type of allegations that lead to a settlement or tracking to promotions or awards around the same time could refine this trend.

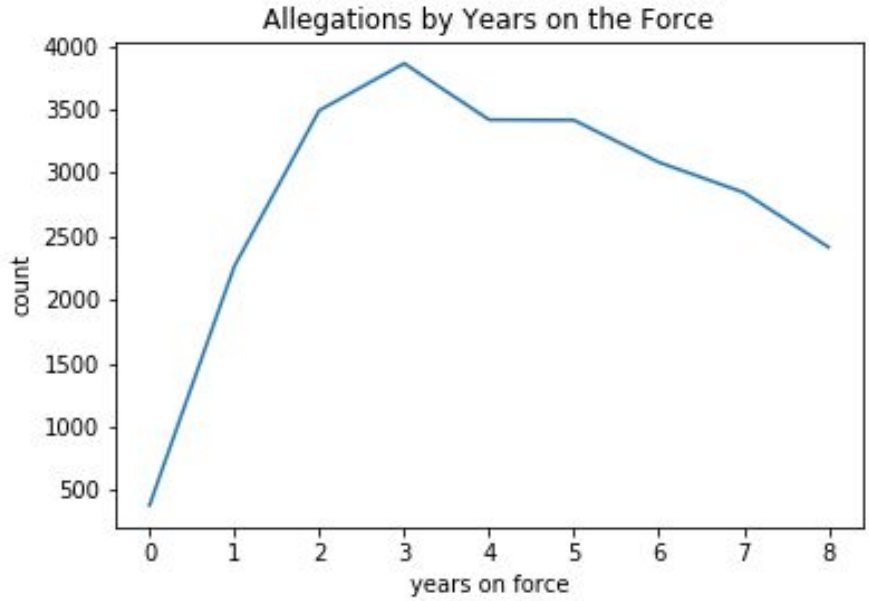
# Question 3

**What is the average number of arrests per officer over years on the force and how many allegation occurred with an arrest?**

**Overview**
By using the officer_subset from checkpoint1, the question has been addressed for determining the count for number of allegations and the arrests against a particular officer. Utilizing a temp-view for the (`officer_subset`), two views have been created for (`officerarrest_years`) and the (`officeralleagtions_years`) accordingly.

Instead of linking arrest data to allegations, these views have been utilized to address the above question. There was not a clear way to determine with a level of confidence if an arrest corresponds to an allegation which was our original goal. Instead of the direct link, a comparison of arrest and alligation by years on the force is used. Two plots are shown below showing the trend of allegation counts and arrests count.




Also, another view has been developed for the count of arrest.id and to generate an average column for the total arrest, per year on the force. This can be observed in the analysis section.

**Analysis**

1–500 of 501+

| | first_name | last_name | officer_id | count | years_on_force | avg_arrest_over_years |
|---|---|---|---|---|---|---|
| 1 | John | Lipka | 16302 | 1782 | 11 | 162 |
| 2 | Stephen | Gregory | 10740 | 2633 | 18 | 146 |
| 3 | Kimberly | Oppedisano | 21350 | 1992 | 14 | 142 |
| 4 | David | Brown | 3056 | 2196 | 16 | 137 |
| 5 | Ted | Jozefczak | 13909 | 1750 | 13 | 134 |
| 6 | Jeffrey | Curia | 5930 | 1463 | 11 | 133 |
| 7 | Richard | Pellerano | 22150 | 2122 | 16 | 132 |
| 8 | Angelo | Travlos | 28871 | 2251 | 17 | 132 |
| 9 | Michael | Durkin | 7603 | 1566 | 12 | 130 |
| 10 | Jason | Cloherty | 4860 | 1939 | 15 | 129 |
| 11 | Eugene | Sledge | 26616 | 1927 | 15 | 128 |
| 12 | Peter | Roehl | 24475 | 2289 | 18 | 127 |
| 13 | Anthony | Siriscevich | 26513 | 2270 | 18 | 126 |
| 14 | Alberto | Zayas | 31598 | 1605 | 13 | 123 |
| 15 | Ricky | Rivera | 24090 | 1462 | 12 | 121 |
| 16 | Matthew | Bouch | 2470 | 1931 | 16 | 120 |
| 17 | Jason | Landrum | 15631 | 1589 | 14 | 113 |
| 18 | Rolando | Ruiz | 24905 | 1351 | 12 | 112 |
| 19 | Karl | Kruger | 15273 | 1336 | 12 | 111 |
| 20 | Peter | Babich | 1021 | 1552 | 14 | 110 |
| 21 | Matthew | Graf | 10529 | 1869 | 17 | 109 |
| 22 | Guy | Dailey | 6087 | 1539 | 14 | 109 |
| 23 | Christoph | Larson | 15727 | 1311 | 12 | 109 |
| 24 | Brian | Ustaszewski | 29186 | 1950 | 18 | 108 |
| 25 | Justin | Conner | 5179 | 1192 | 11 | 108 |
| 26 | Juan | Duran | 7585 | 1260 | 12 | 105 |
| 27 | Shawn | Alonzo | 442 | 1376 | 13 | 105 |
| 28 | Nick | Beckman | 1647 | 1458 | 14 | 104 |
| 29 | Hitesh | Patel | 21955 | 1775 | 17 | 104 |
| 30 | Lawrence | Darko | 6223 | 1559 | 15 | 103 |
| 31 | Victor | Razo | 23537 | 1135 | 11 | 103 |
| 32 | Adrian | Vivanco | 29670 | 1455 | 14 | 103 |
| 33 | Matthew | Tegtmeier | 28301 | 1142 | 11 | 103 |
| 34 | John | Beluso | 1756 | 1340 | 13 | 103 |

The view generated above shows a map of the total count of allegations against a particular officer. This can show some trends in regard to average count of the arrest as well, over the years for these officers. The trend shows a large number of these officers having an average count of above 100, in the average values for arrest.

The plots and the views generated for the count against years on the force, be it against the arrests or allegations for the officers have been mapped for mostly  a zero to nine year span since the (`appointed-date`). The insights from this can be drawn from the plots generated.

It has been observed that for the `officerarrest_years`, we have an increase in the count for date-range of 3 to 6 years, for the officer. This essentially means that for the officers with varied appointment dates and timelines of service, a large number of arrests have been done in between 3 to 6 years of serving on the force.

It has been observed that for the `officerallegation_years`, we have an increase in the count for date-range of 2 to 7 years, for the officer. Similar to the above trend of arrests, this implies that for the officers with varied appointment dates and timelines of service, a large number of allegations have been done in this range.

**Relating to Theme**

The insights from this topic and question can be worked in more detail, and by linking arrests to allegations data and forming a map against the officers. Interesting patterns can be found in

regard to regions or based on the demographic of officers chosen to study as a subset. Overall arrests follow the same trend as allegation, so there is not much insight to draw from the trend overall.

# Question 4

**Does a high number of allegations with an arrest early in an officer's career lead to a higher average of allegations?**

**Overview :**

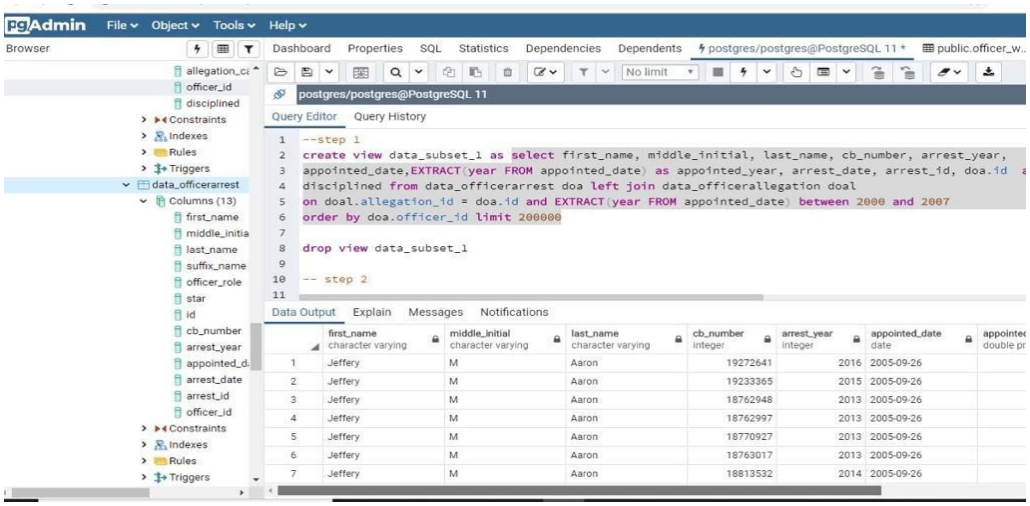
Here we are trying to conclude that if officers are having a higher number of the allegation with an arrest early in an officers' career whether it will be a major factor which leads to higher average allegation count for officer's career. Firstly we found out the officers having appointed between 2000 to 2007. We extracted the year from the appointed_date column and thus we are able to look at the records as per the requirement. Then we simply created views and load the data through a query. Afterward, as we are trying to identify the count for allegation_id as a number of allegations for every officer for the early period. Here we refer early period as the first three years after the officer's appointment. We again loaded that data for the next observation by creating another view that can be used as a reference for further processing. Now as we are trying to observe that whether the high number of allegations with arrest rarely in an officer's career lead to average of allegation thus we need to find out the average for allegation count for officer throughout his career so we simple calculated for each officer the average number of allegation for officer throughout his career . Hereby closely observing different observations for both early and average allegation count, we made out conclusions.

**Analysis :**

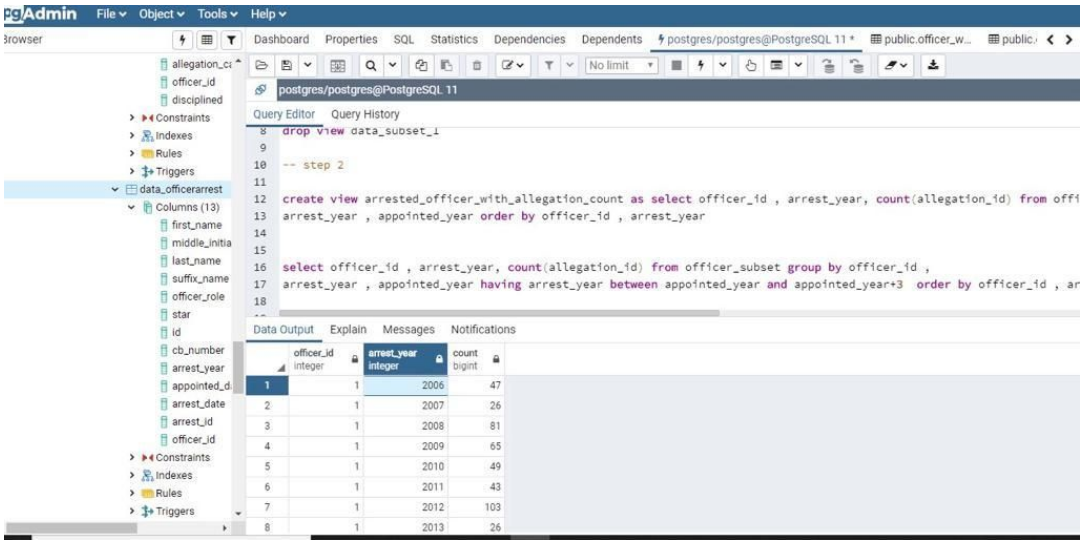
Here we are trying to analyze whether the high number of early allegation count will lead to a higher average of allegations throughout the officer's career. We observe that there are more officers having higher early allegations count as compared to their average allegation count which means that officers having early allegations are far more as compared to average allegation count throughout their career.

- By the Sql queries we were able to find the output step by steps:

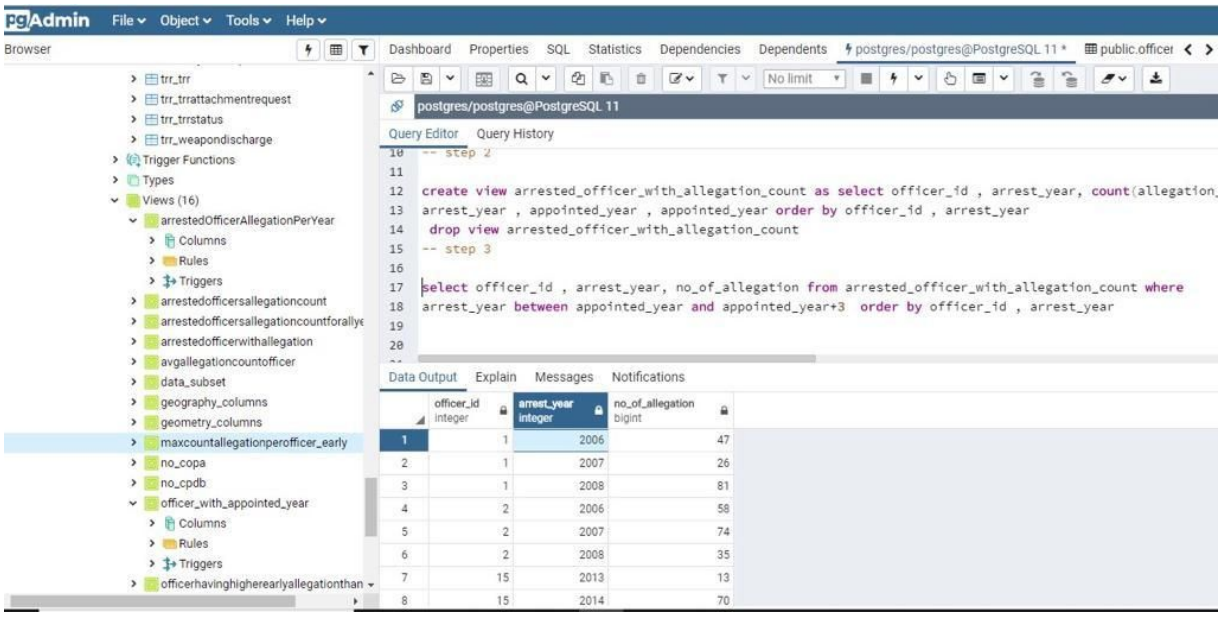
1) First we found a subset of officers who are appointed between years '2000' AND '2007'. We extracted year from each officer's appointed date and added a column as appointed_year by extracting year from appointed date and created view for it. We created view for 200K records.



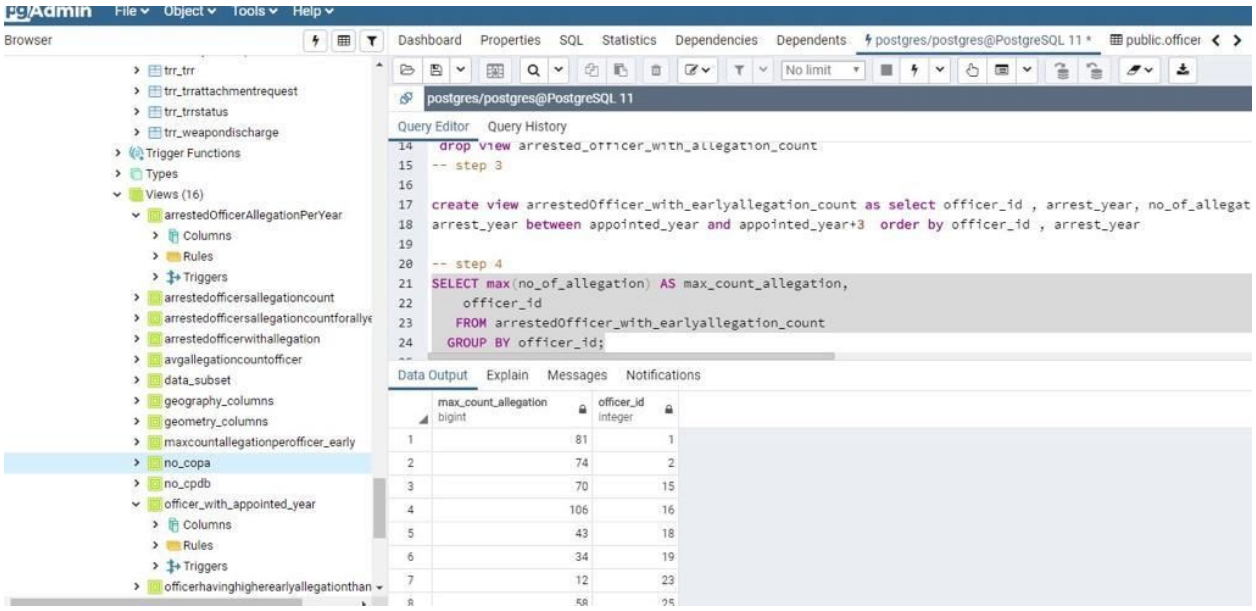

2) Next step, we counted the number of allegations for officers for every arresting year. We created another view accordingly.



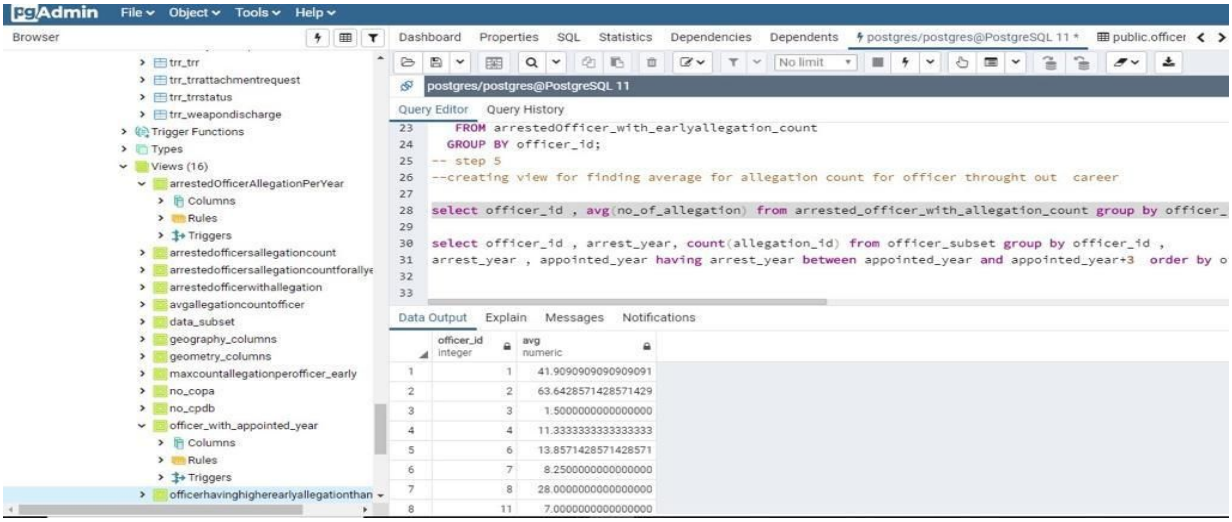

3) Further, we count the number of allegations for first three arresting years of officers as early allegation count. We created another view accordingly.



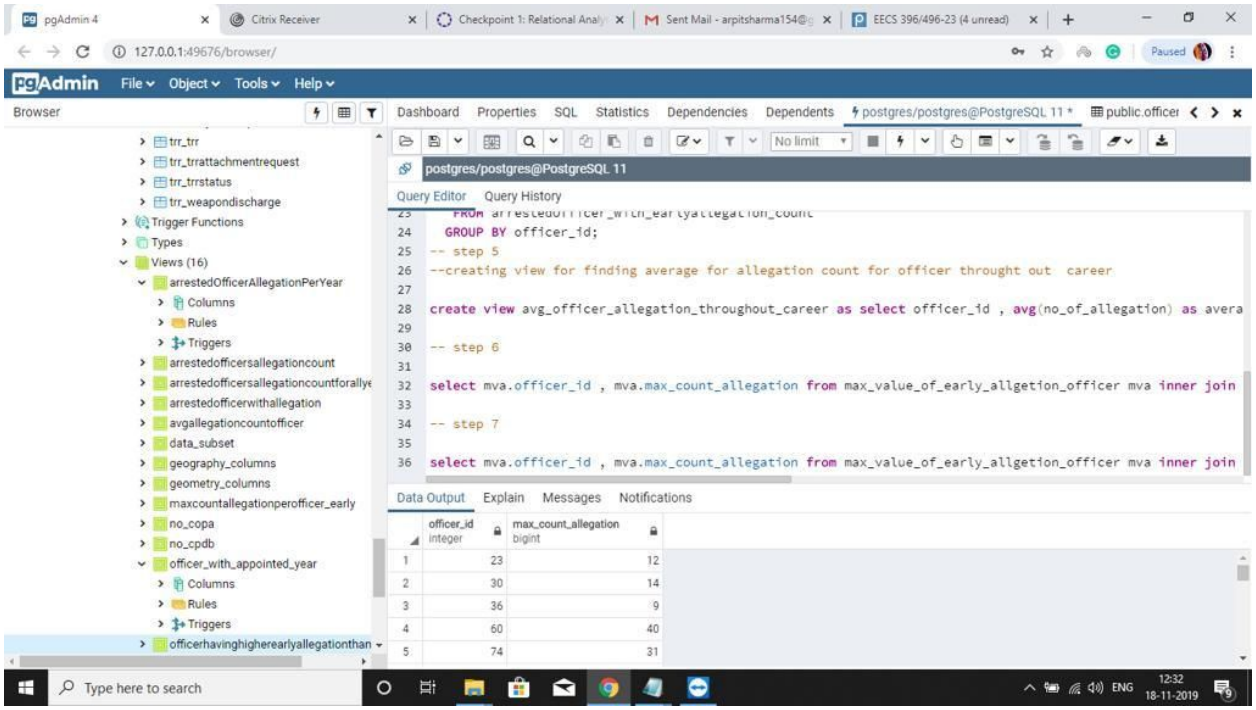

4) Next step, we figure out the maximum count value of officer for early allegation count which will provide us maximum value among the early allegations count.



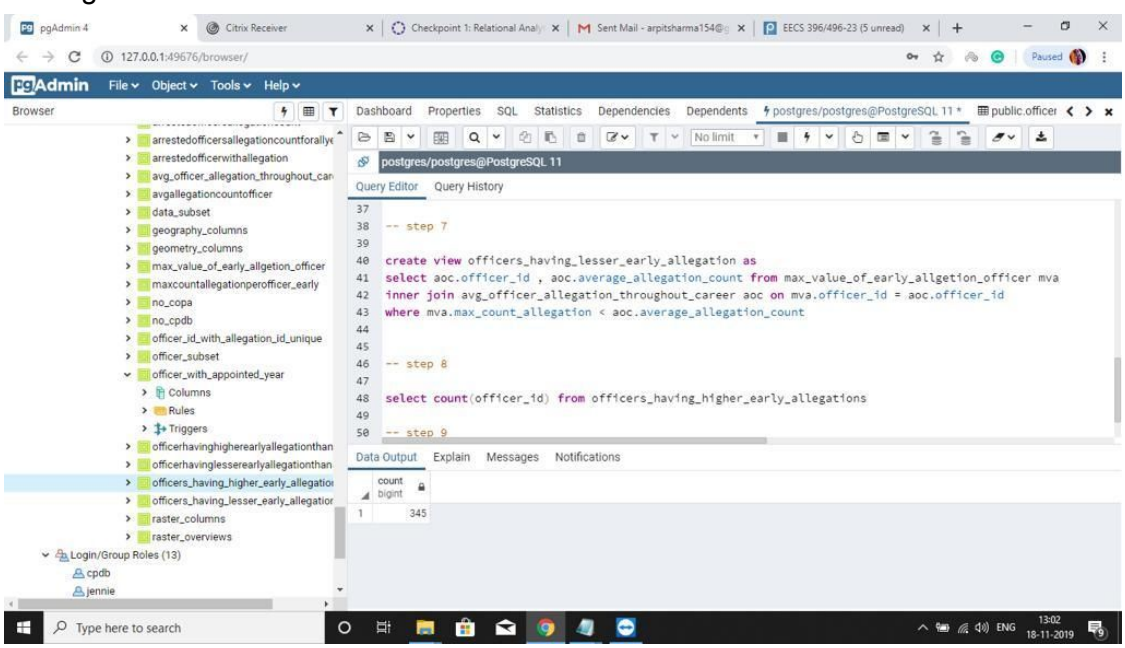
5) Then we find out average allegation count of each officer through-out career here through-out career means considering all arrest years come across during the officer career for finding average for allegation count for officer throughout career.



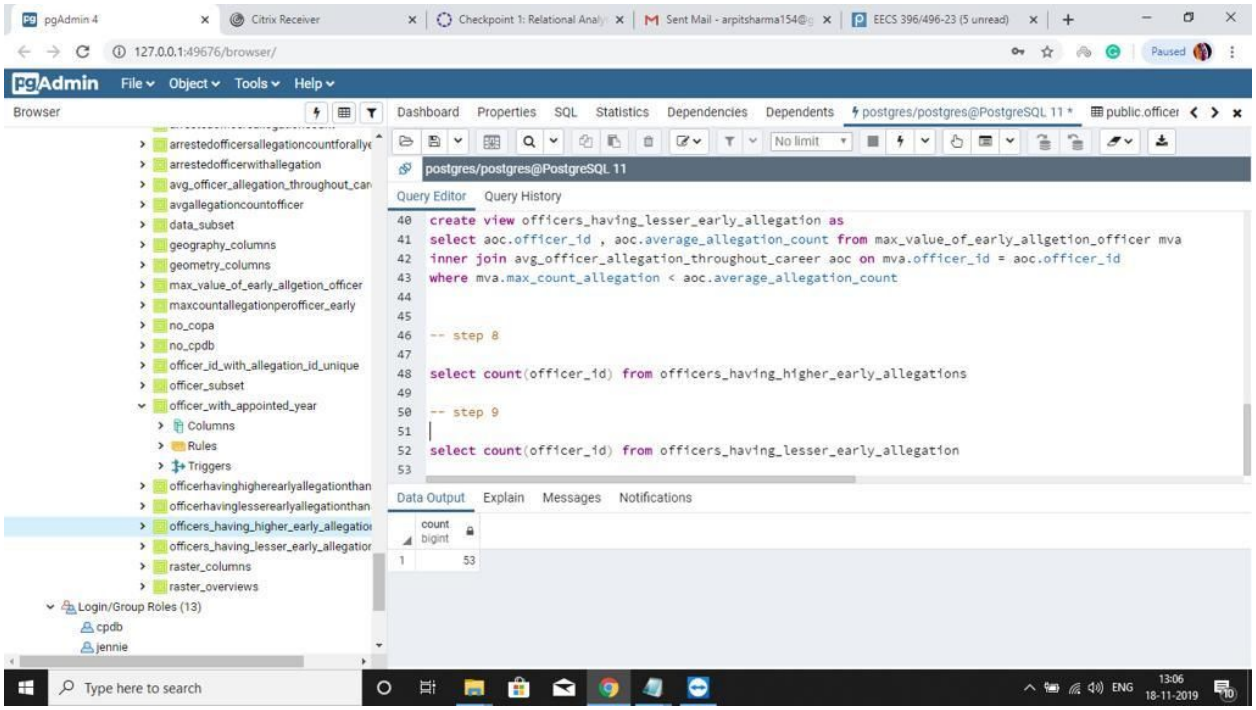

6) We find out the officers having higher number of allegations than the early number of allegations and we find out officers having higher average number of allegation count than the early allegation count.




7) We find the count for officers having higher number of allegation count than the early allegation count.




8) Finally, we find out the count for officers having lesser allegation count than the early allegations count.




9) And we removed all the views.

**Relating to the Theme :**

Our analysis strongly supports the theme as we are trying to identify whether an officer having higher early allegation or a higher number of the early allegations will lead to a higher average allegation number throughout the officer's career. We can identify the factor which is the triggering point for the higher number of allegation counts for an officer or simply which leads bad officers throughout the career.

For future analysis, we will include more number of records as we currently processed 200K records, so that we can conclude more accurate conclusions. Also, we will increase window period for appointed officers as we currently we processed records for appointed dates between 2000 to 2007 which will give us more accurate results.