Discussions

For this project, we are investigating the factors that would affect a job's salary, including the job's location and the company's rating and revenue. The following are what we conclude after digging into the datasets.

For Question 1, we would see it's not necessarily the case that the salary has a strong positive relationship with the rating of that company. Since it is impossible to plot a scatter plot in SQL to see the correlation between company's revenue and employee's salary, we would use an alternative way to see the relationship between revenue and salary. Here we again category the rating of the corresponding company and the salary for each job as either 'high' or 'low' according to whether they are greater than the average which is showing in the VIEW Category. Then we are going to count the percentage of the jobs where their salary and rating category match. Eventually we could see from the table Q1, there are only 45.7% jobs which have the matched category. Since it is not even over 50%, we can see the higher rating of the company does not necessarily offer more salary. As the positive relationship between salary and the rating fails, this leads to Q3 where we are digging into another factor, company's revenue.

In terms of if there is any city with much higher average salary than others(Question 2), we find the city with the highest job salary is Marin City, which has the average salary of 126K and it's almost twice as the average salary across all cities, which is around 72.6K.

In general, we would expect companies with higher revenues to provide their employees with more salaries.(Question 3) So we used this dataset to see whether this is true and here the methods are quite similar as Q1. We categorized each company's revenue into 'High' and 'Low' based on whether its revenue is greater than the mean revenue of all companies in the dataset or not. And salary is also categorized into 'High' and 'Low' based on whether the salary is greater than the mean salary of all jobs in the dataset or not. Based on this, we further categorized each company into four categories 'highSalaryHighRevenue', 'lowSalaryHighRevenue', 'lowSalaryLowRevenue', 'highSalaryLowRevenue'. From table Q3, we can see that 404 companies which is 50.8% of all companies follow our assumption that there is a positive relationship between revenue and salary, that is companies with higher revenue would provide higher salary, while companies with lower revenue would provide lower salary. And among all companies, 49.2% of them do not follow our assumption. It indicates that the number of companies which follow our assumption and number of companies which do not follow our assumption are nearly equal, so it is not true that companies with higher revenues will give employees more salaries, and vice versa. But there is some limitation on this, first we did not control over other variables, for example, we did not compare the salary among the same job in different companies, which may influence salary as well. In addition, our classification of revenue and salary is too easy.