

## My analysis

As I examine the data, I had a creeping suspicion that the dataset is fake. I surmised that my boss handed me spurious data in order to test the data engineering skills of a new employee. To confirm my hunch, I decided to take the following steps to generate a visualization of the data to confirm this.

First, I created a histogram to visualize the most common salaries range for employees. Based on histogram, it shown that there were almost 150000 employees' salaries are between \$40k and \$45k.

Secondly, I broke it down into showing Salary by Employee's Job Title and the histogram further confirmed my suspicion. The histogram shown that the Assistant Engineer, Engineer, Senior Engineer and Technique Leader made the same amount of money. This is highly impossible as Assistant Engineer generally make lesser as they are more junior in the job. Similarly, Senior Engineers' salaries are higher than Assistant Engineer and Engineer as they have more experiences or been in the job longer. Furthermore, the histogram also shown that Senior Staff and Staff were sharing the same salary number. Lastly, Technique Leader supposed to be making more than Senior Staff, however, the histogram shown that they were making the same money as an Assistant Engineer. Therefore, I concluded that this data is a fake dataset.

Lastly, I tried looking up my ID 499942 from the data frame to check for my information and found that I was tricked as the data shows my name as 'April Fool'.