

**Members:** Vinny, Alyssa, Tristan, Will, and Miles

**Key performance indicator (KPI):** Paid Wage Per Year (in salary\_data)

Team ANOVA will be analyzing the salary dataset for our DS311 project. The salary dataset displays records from 167,278 Labor Condition Applications (which include H1-B and other specialty visas) and the permanent resident applications from 2008 to 2015. The dataset intends to track who is entering and leaving the US government for work-related reasons. Each applicant provides information based on their work location, paid/prevaling wages, and their selected job position. Within this dataset, there are 25 variables with 2 extra non-meaningful variables. Several goals of the project include comparing prevailing wages and paid wages, comparing wages by job title, comparing wages by location, and more.

### **Notes About the Dataset**

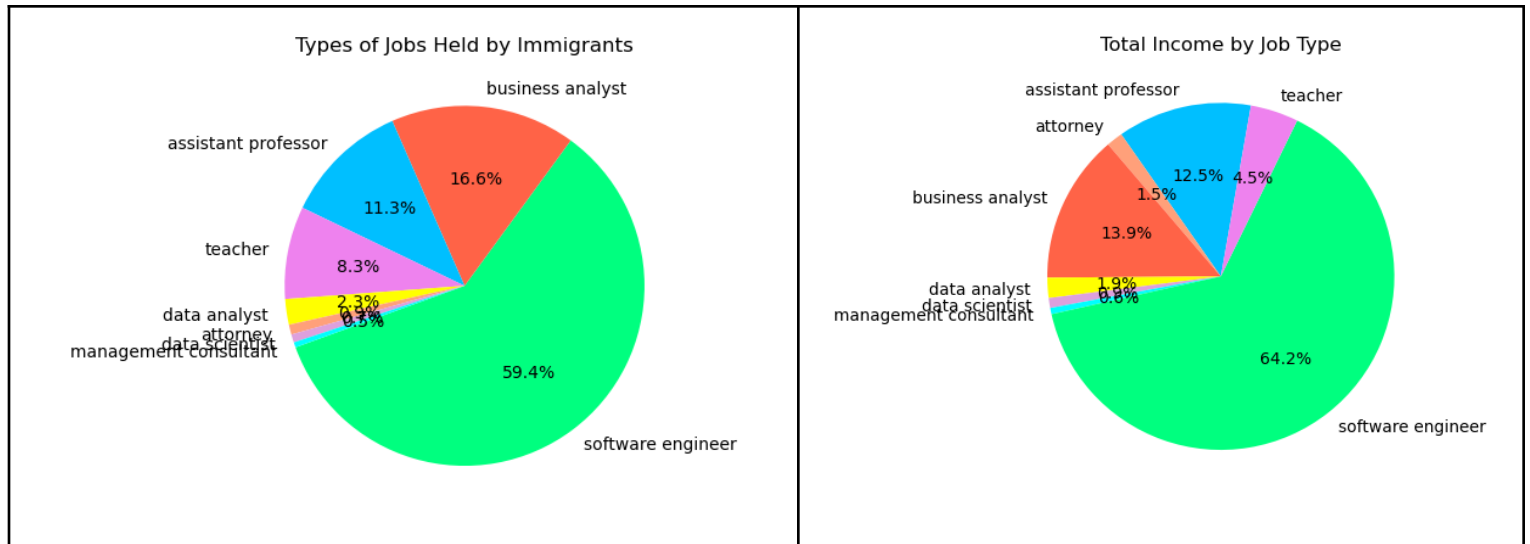
The dataset has 167,278 observations and 27 columns. Two variables are extras or non-meaningful: the initials/codes of U.S. states/territories, and the variable order seems to be an index from a previous uncleaned version of the dataset. Some job titles show up twice in the data. For example, "COMPUTER SCIENCE TEACHER POSTSECONDARY" also shows up as "COMPUTER SCIENCE TEACHER POST SECONDARY."

### **Initial Observations**

The variables we are mainly concerned with are *prevailing wage per year* and *paid wage per year*. The minimum, first quartile, median, third quartile, and maximum prevailing wage per year are as follows: \$10,504, \$56,880, \$70,928, \$90,459, and \$320,000. The mean prevailing wage per year is \$74274.86 with standard deviation \$25356.24. The numerical summary for the paid wage per year is as follows: \$10,500, \$63,000, \$78,600, \$100,006, and \$2,500,000. The mean paid wage per year is \$85,532.77 with standard deviation \$38738.47.

### **Fun Facts**

One fun fact that was discovered during the salary data analysis project was the overwhelming amount of Software Engineers featured in this dataset. The total for Software Engineers came in at a figure of 99,364. This represents a large ratio compared to the total of 167,278 participants included. This job title accounts for 59.4% of the total. The next leading job titles were #2 Business Analyst at 27,811 representing 16.6%, #3 Assistant Professor at 18,866 representing 11.3%, and #4 Teacher at 13,912 representing 8%. From looking at the pie charts, you can see that even though teachers make up 8% of the workforce, they earn 4.5% of all income made by immigrants, which shows there are still immigrants willing to work jobs that pay less than others. The amount of software engineers could be a limitation due to sample size or possibly reflect international interest in the tech-software sector.



A second fact discovered within the frequency analysis of job title subgroups was seeing that Assistant Professor positions are significantly more popular than Data Analyst (2.29%) and Data Scientist (0.73%) positions. This difference could reflect how data-based roles may not be popular on an international scale, as opposed to its popularity in the United States.

A third fun fact we found was the median salaries per job title subgroup. Interestingly, the position with the highest median and mean wage was attorney, with its respective wages being \$85,956.0 and \$91,326.41. The software engineer's yearly median and mean wages are \$80,922.50 and \$82,212.13. Perhaps the difference in yearly wages could be attributed to the education required to pursue an attorney position. While software engineering roles could be achieved without a robust college education, attorneys without this educational background are less common.

A fourth fun fact we found is that the amount of workers in California is 46,782 which equates to 28% of the worker population in the dataset we are looking at and it's the highest out of the rest of the states with the following being Texas at 15,498 which is 9% of the worker population, after that is New York with 11,373 which is 7%, and fourth is New Jersey with 10,198 which is 6%. Meaning that within the dataset California has the largest concentration of workers compared to other states.

Finally, a fifth fun fact is that twenty of the records in the dataset are employees of San Francisco State University! Nineteen are assistant professors and one is a teacher. The assistant professors with the exception of a high outlier made between \$49,716 and \$50,676 per year while the teacher made \$29,544.