

Executive Summary

This Analysis set out to find answers to a number of questions pertaining to educational spending and high paying financial sector jobs using data acquired from public sources, including the Census Bureau. The three highest paying industries within the finance and insurance sector are investment banking and securities dealing (\$223,619 average salary), securities and commodity exchanges (\$205,300 average salary), and securities and commodity contracts intermediation and brokerage (\$160,014 average salary). The states with the most jobs in the aforementioned industries include New York, California, Illinois, Florida, and Texas. A Multivariate Machine Learning Regression was created in order to predict the payroll for each combination of Industry at Level 6 Granularity for our financial sector jobs and State. The regression showed that there were only two significant independent variables for this prediction. The most significant variable was the number of employees and the less significant factor was revenue. This regression had an R-squared of .806, thus it explains most of the data well. While analyzing and exploring the Census data, it was observed that data was being lost almost every time the Census went into greater granularity. An analysis and graph were created to show how many employees' data was lost for each industry at level 4 granularity when it went down to level 5 granularity. Some industries were discovered with high losses, some industries with no losses, and some industries with small gains. The conclusion made for this is: "As companies categorize themselves in the Census data, sometimes they are unsure where they belong in more granular breakdowns of industry, so they leave that question blank. Thus data is lost. Sometimes an employer will also resume categorizing themselves after skipping a particular level of granularity, thus data is regained."

The states that have the highest total expenditure on primary and secondary education are California, New York, Texas, and Illinois. When expenditure is broken down to a per pupil level, the list of states changes to Alaska, New York, Washington D.C., and New Jersey. After generating a correlation matrix between each state's financial sector's annual payroll and each state's fourth and fifth graders' NAEP Math and Reading scores, the correlation coefficient between payroll and each academic performance metric was observed to be between 0 and 0.05. This implies that there is a very weak positive correlation between how well a state's students perform in school and the state's financial/insurance sector's payroll. After separate scatter plots between payroll and each of the academic performance metrics were created, it was evident that as academic performance metrics increase, payroll sporadically changes with no regard to any academic performance metric. This confirms the correlation coefficients observed in our matrix. After generating a correlation matrix between each state's financial sector's annual payroll and each state's academic expenditure, the correlation coefficient between payroll and each academic expenditure type was observed to be about 0.3. This implies that there is a positive correlation between how much a state spends on education and the state's financial/insurance sector's payroll. After separate scatter plots between payroll and each academic expenditure type were created, it can be seen that as academic expenditure increases, payroll also increases.

For areas of future study, obtaining data on college tuitions, retention rates, graduation rates, and percentages of graduates who find jobs would allow for further analysis into the relationship between higher education and those states and industries that have the highest salaries and the most jobs.