Data Wrangling:

To be able to analyze executive pay's impact on firm performance, we need to first find a way to gauge firm performance. To gauge economic performance, I decided to take a look at the Return on Assets (ROA). To gauge the financial performance, I looked at Earnings per Share (EPS) and Return on Equity (ROE). To gauge stock market performance, I looked at P/E Ratio and yearly high stock price. To retrieve this data, I utilized the Wharton Research Data Servies. Specifically, I retrieved my desired data from their Compustat database. While working with the data, I realized that there were companies like L Brands Inc and Cabot Oil & Gas that were no longer being identified by the same stock ticker given in the assignment dataset. Because of this inconsistency, I decided to use the Global Company Key (gvkey) to identify each company because gvkey is commonly known as a permanent identifier that doesn't change. To do this, I used the Pandas .unique function to return an array of all the unique gvkeys in the given data. I then used the Numpy function, .savetxt, to save it to my local drive. From Compustat, I retrieved each company's Global Company Key, Data Year, Assets - Total, Stockholders' Equity, Earnings Per Share (Basic), Net Income, and Price High - Annual. After reading in the new file and dropping the trivial columns, I performed a left merge that merged on each gvkey and year pair to ensure each executive gets associated with their corresponding company's stock information. After the merge, I renamed some columns for legibility, then created new columns for ROA, ROE, P/E Ratio. ROA was calculated by dividing the NI column by the AT column, ROE by dividing NI by SEQ, P/E Ratio by dividing the stock price by EPS, and bonuses by subtracting Total Compensation by Salary. To conclude, I exported the newly merged and cleaned dataset. I believe an analyst could utilize these performance indicators to perform several regression analyses to discover how various aspects of executive pay impact firm performance.

Data Visualization:

To prepare the data for visualization, I had to first use the Pandas .get_dummies() function on the categorical variables in the dataset. In this case, the Gender variable was the only one that needed to be converted. I then dropped the columns that will not aid in visualizing the data and renamed the remaining variables for easier comprehension. I started by using pd.groupby() to group the data by year and take the value of every variable per year. Using a regression plot, I found that the percentage of women in executive positions is increasing linearly per year and has more than doubled since 2010. I also found that that average compensation, salary, and bonuses from this dataset have a strong positive correlation with year and have steadily increased since 2010. I then explored the interaction between rank and salary between men and women. I plotted the average salary of men and women in 2010 and in 2020. I found that while both averages have increased substantially, the difference in salary between men and women remains between 11 and 12 percent indicating no significant change in this salary difference. Using the financial data that I wrangled, I wanted to explore the impact of salary and compensation on these firm performance indicators. Utilizing KDE and regression joint plots, I found that there is little to no correlation between executive pay and firm performance. Through the visualizations, it can be observed that all these aspects of executive pay appear to resemble the normal distribution, indicating that the majority of executives' pay will remain close to the mean regardless of firm performance.. To summarize my findings, I plotted a correlation heatmap. Here, you can see the small and insignificant correlations between performance and executive pay. Although, the visual reiterates my earlier findings that rank is negatively correlated with salary, compensation, and bonuses.