**CEO vs. Worker Pay in Top 3000 US Companies**

**I. Introduction, Business Questions, and Related Hypotheses**

The CEO-worker pay ratio is a pivotal measure that sheds light on the internal compensation dynamics of companies and plays a crucial role in understanding the broader socio-economic landscape. The data correlates the median worker salary to the CEO salary and allows many relationships to be studied. This study focuses on exploring the intricacies of CEO-worker pay ratios within the top 3000 companies in the United States, considering diverse factors such as gender, industry affiliation, and corporate rankings.

The data set aims to explore and visualize the relationship between CEO pay and worker pay, compare different industries and companies, and uncover patterns and trends in the data. This dataset is an important resource for anyone interested in promoting pay equality and advocating for fair wages.

*Preliminary Hypothesis and Intuition*

*Business Questions*

1. How does gender impact CEO-worker pay ratios, and what are the overarching patterns of gender-based compensation disparities across different sectors within the top 3000 US companies?
2. To what extent does the ranking within the Russell 3000 influence CEO-worker salary ratios in financial companies, and how does this relationship compare across industries?
3. How do sector-specific factors, including median worker pay and gender representation, contribute to the overall CEO-worker salary ratios, and what sector-wide patterns emerge within the top 3000 US companies?

*Business Hypotheses*

**Hypothesis #1**

**H0 (Null Hypothesis)**

Male CEOs in the Information Technology have a higher pay ratio than Female CEOs in the Information Technology. They are inversely correlated, there is a relationship.

Analyzed using Categorical Nominal: “M” and “F”

**Hypothesis #2**

**H0 (Null Hypothesis)**

Financial companies with a higher Rusell 3000 Rank, there will be a higher CEO-worker salary ratio.

Analyzed using categorical ordinal data, where Rusell 3000 ranking is highest at the top (#1) and lowest at 3000.

R-value: -0.05698869 (weak negative correlation)

**Hypothesis #3**

**H1 (Alternative Hypothesis)**

Healthcare companies have the highest median worker pay, and the lowest CEO-worker salary ratio. They are inversely correlated.

Type of Plot: Continuous Interval Data

R-Value: -0.3608818 (strong negative correlation)

**Hypothesis #4**

**H0 (Null Hypothesis)**

Female CEOs in the Top 100 companies have higher CEO-worker salary ratio than the Male CEO’s in the top 100. Companies with Male CEO’s have a 3:1 ratio and have a higher median work pay.

Type of Plot: Continuous Discrete Plot

R-Value: 0.0008506226 (no correlation)

**Hypothesis #5**

**H1 (Alternative Hypothesis)**

Industrial companies with a higher pay ratio indicate in general a higher CEO-worker salary. This shows a positive correlation relationship. Additionally, CEOS in the industrial sector are heavily male with only 6% of CEOS represented in the dataset being female.

Type of Plot: Continuous Interval Plot

R-value: 0.4349431 - strong positive correlation

**II. Data Source and Description**

The data source was a Kaggle dataset for CEO vs Worker Pay in Top 3000 US Companies. This dataset included salaries from Russell 3000 and S&P 500 companies. It is a public data set available for use.

The industries under consideration are: "Health Care," "Materials," "Industrials," "Consumer Discretionary," "Information Technology," "Real Estate," "Financials," "Consumer Staples," "Communication Services," "Utilities," and "Energy”. Consumer discretionary sector are companies that produce non-essential goods and services like Ford, Netflix and Tesla. Consumer staples are companies that sell products that can be considered essential, for example, Walmart, and Proctor&Gamble.

Using functions like “dim” and “colnames” the data dimensions and column names can be defined. The data had 3000 companies and 8 columns, but after dropping some rows the data ended up with 2,175 unique values. The eight columns were the index, ticker, company name, median worker pay, pay ratio, CEO name, salary, and industry. This is further described in the preprocessing steps.

The limitations from this data set are that there may lack of data completeness, resulting in NULL values. Additionally, while the data was collected in 2023, there is no indication that the median pay for each worker is the latest average, this could result in inaccurate data quality. There also may be some biases, from the selection process of the companies into the top 3000 companies by Russell. There may be a bias towards companies that were nominated in the selection process.

**III. Data Preprocessing**

The preprocessing of the data from the dataset was done methodically. The first step was seeing the parameters of the data. Starting with the “summary” function to determine the dataset. This was a quick way to identify if there were any NULL values in the data. There were 28 NULL values in some of the rows, so those rows were dropped. This step was helpful in seeing the max and min salaries of the CEO and the median worker pay. The median worker pay is interval continuous.

The second step was cleaning the data, the data had 28 NULL values for the salaries column. The team decided that the NULL values should be discarded, since the values would not show up in the analysis. This resulted in 2175 unique values for extrapolation.

The next step was the normalization of the data. For the salary, it was in the currency format for USD. The “$” sign and other symbols on the salary was removed and converted to an integer using “as.numeric” and “gsub” function to ensure that the data can be used for analysis. In order to use the data properly, the data was converted to integer, and additional columns were added with “\_int” added to the column name to signify it has been converted to a whole number. This was done to be consistent with the data. Additionally, the pay ratio had to be converted to a whole number. Since the dataset was Number:1, the last two characters “:1” were removed, and the assumption was to add that it was a ratio to one. This normalized the pay ratio column.

To ensure the rows contained categorical nominal and categorical ordinal data for analysis, several functions had to be run. For instance, for categorical nominal data, “Male”, “Female” of each CEO was added to the Gender column. Then, the median worker pay was categorized in categories "Below 10k", "10-20k", "20-30k” and 10k increments till “Above 80k”.

A few data columns were added, including company size, ESG risk levels and categorization. This would have added to the richness of the data, since ESG has been of high interest in recent years. However, due to numerous null values, the columns could not be used.

**IV. Specific Approach/Analysis**

Before proceeding with the specific approach, we took a big-picture snapshot of the data and extrapolated the average and max CEO salary and average and max worker median pay. Additionally, a pie chart was created to ensure that all the top 100 companies – with the highest CEO-worker pay ratio were identified. The highest count was in the “Consumer Discretionary” industry and “Health Care” had the highest count in the lowest 100 companies, indicating that the “Consumer Discretionary” had the highest pay ratio between CEO and workers, with “Health Care” having the lowest pay ratio.

**Table 4-1. Overview Numbers on CEO Salary and Worker Median Pay by S&P 500 and R 3000**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **S&P 500 ($)** | **R3000 ($)** | **Combined ($)** |
| **Average CEO Salary** | 18,582,399 | 8,316,591 | 10,639,856 |
| **Max CEO Salary** | 296,247,749 | 834,959,367 | 834,959,367 |
| **Average Worker Median Pay** | 84,866 | 95,180 | 92,846 |
| **Max Worker Median Pay** | 295,884 | 753,041 | 753,041 |

**V. Characteristics of the Study Frame**

The data consisted of 3000 companies from the Russell index, and the breakdown of the number of companies for the year 2023 are as below. In Figure 5-1 and Appendix A, there is the breakdown of the largest pay-ratio gap ranked, by Top 100 largest and Bottom 100 smallest ratio.

**Table 5-1. Count of companies in separate industries**

|  |  |
| --- | --- |
| Financials​  Industrials​  Healthcare​  Information Technology​  Consumer Discretionary​  Real Estate​  Materials​  Energy​  Consumer Staples​  Communication\*​  Utilities​ | 374  331  288  286  284  145  114  103  97  89  62 |

*\* Alphabet and Meta are considered in ‘Communication’ industry*

**Figure 5-1 Pie Charts of Top 100 Companies and Bottom 100 Companies**

|  |  |
| --- | --- |
| **Top 100 Companies by Industry** A pie chart with different colored circles  Description automatically generated  **Top 2 largest gap:**  Consumer Discretionary, 490.4:1  Communication Services: 297.2:1 | **Bottom 100 Companies by Industry**  **A colorful pie chart with text  Description automatically generated**  **Bottom 2 most equal gap:**  Energy, 80.5:1  Utilities, 78.4:1 |

**VI. Analyzing the Relationship between CEO Gender in the IT Industry**

**Figure 5-1. Violin Plot of CEO-Worker Salary Ratio Across IT by Gender**

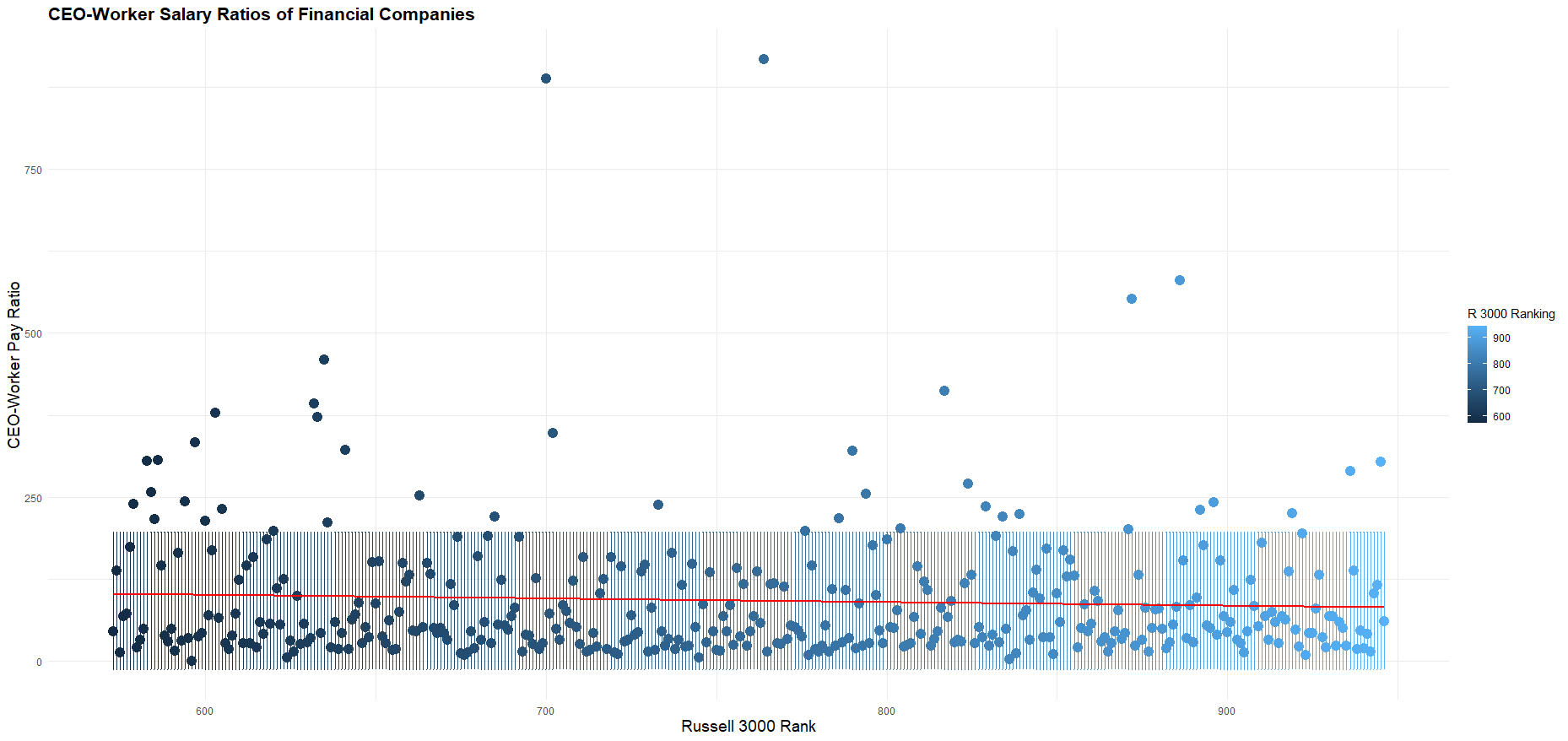
**A graph of a person and person

Description automatically generated**To determine the relationship between CEO-worker Salary Ratio in the Information Technology (IT) industry, the plot was graphed on R using a violin plot. Since the data was categorical nominal data, this plot was chosen to study the relationship. The relationship found was that Male CEOs do have the highest CEO-worker pay ratio, but the majority of Male CEOs and Female CEOs have an average ratio of 234:1­­­­­.

Interestingly, for female CEOs, there are two additional groups, both of which have a higher ratio, approximately 600:1 and 1800:1, Therefore, it can be concluded that the hypothesis is true that male CEOs in the Information Technology sector have a higher pay ratio compared to female CEOs. The correlation can be analyzed using Cramer’s V, but since it is a categorical nominal value, the violin plot shows the relationship with the frequency.

**VII. Analyzing R3000 Ranking with CEO-Worker Ratio in Financial Companies**

**Figure 6-1. Error bar of CEO-Worker Salary Ratio Across Financial Companies by R3000 Rank**



Financial companies with a higher S&P Rank have a positive correlation with the ratio of CEO’s salary to worker’s salary in the that company. This scatterplot illustrates the relationship between the Russell 3000 (R3000) rank and the CEO-worker salary ratio within financial companies. The x-axis represents the categorical ordinal data of R3000 rank, with a higher rank indicating a lower numerical value. The y-axis portrays the CEO-worker pay ratio.

Upon analysis, the computed correlation coefficient is approximately -0.057, indicating a weak negative correlation. Specifically, the interpretation suggests that as the company's R3000 numerical number increases (moving towards a higher numerical value), the CEO-worker pay ratio tends to decrease. In simpler terms, a higher numbered company corresponds to a lower pay ratio meaning that the companies with lower rank also have a lower pay ratio. This observation aligns with our initial hypothesis, predicting that companies with a high Russell 3000 rank (top 10) exhibit a higher CEO-worker pay ratio.

However, it is crucial to approach this conclusion cautiously. Despite the alignment with our hypothesis, the correlation is notably weak, almost negligible. Consequently, we refrain from making strong assertions and treat this finding as inconclusive. In statistical terms, we consider the null hypothesis (H0) that there is no significant correlation between R3000 rank and the CEO-worker pay ratio. This cautious approach is grounded in the understanding that the observed correlation is close to zero, emphasizing the importance of prudence when interpreting the results.

**VIII: Analyzing CEO-Worker Salary Ratios of Healthcare Companies**

**Figure 7-1. CEO-Worker Salary Ratios of Healthcare Companies**

**A graph of a salary rate

Description automatically generated with medium confidence**

**Figure 7-2. CEO-Worker Salary Ratios with Classifications**

|  |  |
| --- | --- |
| **Table 7-1 Median Worker Pay Categories** | |
| **Classifications** | **Frequency of Companies in Healthcare Industry** |
| Below 10k  10-20k  20-30k  30-40k  40-50k  50-60k  60-70k  70-80k  Above 80k | 2  4  3  16  17  21  30  21  174 |

**A graph of a salary rate

Description automatically generated with medium confidence**

The hypothesis here is proven to show that the data is rightly skewed, giving us more clarification regarding the salary distribution in the healthcare sector. In our graph, while the salary gap is smaller, there exists a larger concentration of healthcare companies with lower CEO-worker salary ratios compared to the median. Within our dataset, healthcare companies are distinct from other industries in offering higher median worker pay while maintaining a notably lower CEO-worker salary ratio. Thus, on average, employees working for healthcare companies receive relatively higher interval wages compared to workers in other industries in our dataset.

Furthermore, the compensation gap between CEO and worker pay is smaller. The observed negative correlation reported of -0.36 between median worker pay and the CEO-worker salary ratio supports this relationship. Moreover, as median worker pay is typically higher in healthcare companies, the CEO-worker salary ratio is lower, validating the inverse relationship. Our data and graphical representation show healthcare companies prioritize providing higher wages to their employees while minimizing the compensation gap between employees and CEOs when compared to the industries studied in our dataset. This key takeaway highlights the uniqueness of the healthcare industry in its compensation practices compared to other sectors.

**IX: Analyzing CEO-Worker Salary Ratios of Healthcare Companies**

**Figure 8-1 and Figure 8-2. Top 100 CEO-Worker Salary by Ratio, colored by Gender and Industry**

**A graph of a graph with a red and blue line

Description automatically generatedA graph of a salary ratio

Description automatically generated with medium confidence**

Female CEOs in the Top 100 companies have higher CEO-worker salary ratio than the Male CEO’s in the top 100. Companies with Male CEO’s have a 3:1 ratio and have a higher median work pay.

The investigation sought to substantiate the hypothesis that female CEOs in the Top 100 companies would exhibit a higher CEO-worker salary ratio than their male counterparts. Through a meticulous analysis of salary data and the construction of dual-axis plots, our study aimed to uncover distinct patterns within each gender category. Contrary to the initial assumption, the graphical representation revealed that companies with male CEOs consistently maintained a 3:1 CEO-worker salary ratio, and the median worker pay was higher compared to those led by female CEOs. The statistical analysis, encapsulated by an R-Value of 0.0008506226 denoting no correlation, dispels the notion that CEO gender serves as a significant predictor for variations in salary structures. In essence, our findings challenge the presumed link between CEO gender and salary ratios, underscoring the need for a nuanced understanding of the multifaceted factors influencing corporate compensation practices.

**X. Analyzing Industrial Companies Pay Ratio with Gender**

**Figure 9-1. Pay Ratio vs CEO-Worker Salary in Industrial Companies**

**A graph with a line and a point

Description automatically generated**Industrial companies with a higher pay ratio indicate in general a higher CEO-worker salary. This shows a positive correlation relationship. Additionally, CEOS in the industrial sector are heavily male with only 6% of CEOS represented in the dataset being female.

A scatter plot was created to show the relationship of Pay Ratio to Median Worker Salary in the Industrial sector of our data. The pay ratio falls under 250 for most of the data with a positive correlation being present and the data interval continuous. For the industrial sector, we saw that tendency present for as the pay ratio increases, there being a tendency for median worker salary to also increase. The conclusion can be made based on that that companies with higher pay ratios in the industrial sector tend to have higher median worker salaries. The industrial sector among the top companies we found in our dataset is severely lacking female representation due to there only being a handful of CEOs who are female. The conclusion can be made that diversity and inclusion in the industrial sector is greatly lacking. The data shows the R-value of 0.4349431 which is a strong positive correlation indicating that the hypothesis rings generally true.

**XI. Specific Outcomes and Key Learning**

Our analysis of the data reveals potential gender bias in CEO compensation across various sectors. Within the Information Technology sector, male CEOs exhibit a higher pay ratio compared to their female counterparts. This disparity is further exemplified in the industrial sector, where only 6% of CEOs are female. Companies should consider implementing measures to promote gender diversity at the executive level to address these imbalances.

Financial companies demonstrate a weak negative correlation between S&P Rank and CEO-worker pay ratio, indicating that higher ranked companies might have higher pay ratios. Similarly, healthcare companies exhibit the highest median worker pay and the lowest CEO-worker salary ratio, suggesting a more equitable compensation structure in this sector.

Data preprocessing revealed missing data points for crucial criteria like salary. We implemented strategies to ensure the reliability of our findings despite these limitations.

While our analysis encompasses data from roughly 4,000 companies only, which could be negligible in comparison to the total number of the US companies (33.2 million) as it remains uninvestigated, potentially limiting the scope of our insights. This gap is particularly relevant for private companies, as over 86% of companies in the U.S. with 500 or more employees are privately held and operate outside the scope of our data.

Although the data was updated several months ago, subsequent changes in company rankings (e.g., S&P rank, Russell 3000) may render the findings partially inaccurate. Mergers, acquisitions, and other organizational changes could further affect the validity of our observations.

While the data provides valuable insights regarding CEO-worker pay ratios, it neglects important factors like company performance (revenue, service quality) that could significantly influence compensation decisions. Future analyses should consider incorporating these additional dimensions for a more comprehensive understanding of the dynamics at play.

**XII. Summary and Conclusions**

The data revealed a persistent gender pay gap among CEOs in the industries analyzed. The data showed outliers indicating progress among female CEOs, however, further research revealed this gap widened during the pandemic. Female C-suite pay fell to a record-breaking low as a percentage of pay earned by male CEOs that has not been seen since 2012. From 2012 to 2020, male CEOs gained a 27% increase in pay compared to female CEOs, who achieved a 10% increase. By 2020, female CEOs earned 75% of male CEO pay, which is a decrease compared to 88% recorded in 2018 (4).

According the bizjournals, there are 459 male CEOs and only 42 female CEOs in the S&P 500 companies. However, a recent report proves that female CEOs outperform their male counterparts, especially during COVID. And females generally have a lower CEO-worker pay ratio on average as proven in Hypothesis 1. Therefore, companies should priorities gender equality as it is synonymous with company performance.

These insights are derived from proxy statements filed by the S&P500 companies, also revealing that the pay discrepancy could be due to the stock-related awards that men out-earn at a significant margin compared to women. The pay gap could also stem from early careers, where women graduating with bachelor’s degrees typically earn less than men, where they experience biases that influence compensation across varying fields. Additionally, certain career paths are viewed as more appropriate and effective to climbing to the CEO level. Pay transparency has helped to narrow the pay gap within companies for individual employees and managers, but female CEOs are still facing bias. Today, women hold over half of management-level positions and account for half of the U.S. population, however, are underrepresented in CEO roles. Men are outnumbering women in these executive-level leadership positions.

On average, among the Russell 3000 companies, highest 1 CEO Salary = 490.4 Workers Median Pay, lowest, and at best 1 CEO Salary = 78.44. The highest CEO salary is in the Communication Services industry with $834,959,367, The Trade Desk, Inc. Interestingly, the lowest median wage paid companies are companies like Scholastic, and Liberty Latin, mainly in the liberal arts sector.

Through our research we were able to identify top industries where employees earn the highest pay, resulting in lower CEO-to-worker pay ratios. The CEO-worker pay distribution in the Health Care industry can be explained by the company size or by the presence of employees pursuing more specialized roles with advanced degrees. For instance, employees working for research companies in the healthcare industry, such as the top-worker-paid Kodiak Sciences, earn top salaries, with the highest salaries earned in California, New Jersey, and Massachusetts. These high salaries further decrease the CEO-worker pay ratio.

**Table 12-1. Summary of Conclusion to Business Questions**

* How does gender impact CEO-worker pay ratios, and what are the overarching patterns of gender-based compensation disparities across different sectors within the top 3000 US companies?

*Answer: Female CEO’s have lower range of CEO-Worker Pay ratio than Male CEO’s.*

* To what extent does the ranking within the Russell 3000 influence CEO-worker salary ratios in financial companies, and how does this relationship compare across industries?

*Answer: There is no strong correlation with company rank and CEO-worker salary ratio.*

* How do sector-specific factors, including median worker pay and gender representation, contribute to the overall CEO-worker salary ratios, and what sector-wide patterns emerge within the top 3000 US companies?

*Answer: Industries that require higher skilled workers like healthcare, tend to have a lower CEO-worker ratio.*

**XIII. Recommendations**

* Consider addressing the limitations associated with data completeness and potential biases in the selection process of the top 3000 companies. Future studies should aim for more comprehensive datasets to enhance the robustness of the analysis.
* Acknowledge the potential impact of changes in company rankings, mergers, acquisitions, and other organizational changes on the validity of the observations. Encourage periodic updates to the dataset to ensure relevance and accuracy over time.
* To provide a more comprehensive understanding of CEO-worker pay ratios, future analyses should consider incorporating additional dimensions such as company performance metrics (revenue, service quality), and possibly stock-related awards that could significantly influence compensation decisions.
* Further explore the relationship between gender diversity in executive roles and company performance. Consider expanding the analysis to explore the impact of diversity and inclusion initiatives on compensation practices and overall organizational success.
* Extend the study beyond the United States to include a global perspective. Comparing executive and worker compensation across different countries and regions can provide valuable insights into the influence of cultural factors on the CEO-worker pay ratio.
* Conduct a more granular analysis within each industry to reveal sector-specific patterns. This can help identify industries that are more prone to pay disparities and those that have adopted more equitable compensation practices.
* Continuously monitor and analyze trends in the gender pay gap, especially considering the recent challenges posed by the pandemic. Investigate the underlying factors contributing to the widening gap and explore potential strategies to address gender-based compensation disparities.
* Emphasize the importance of pay transparency and explore ways to mitigate biases in compensation decisions. Consider recommending strategies that promote fair and unbiased evaluation of executives, irrespective of gender.

**XIV. Limitations and Opportunities for Future Study**

Throughout our analysis, there was a lack of relationship between company performance and the CEO-worker pay ratio. Additionally, we could not see much correlation between female CEO’s and the way a company performs, though gender diversity is proven to improve performance of any team in the corporate context.

A more granular analysis by industry can reveal sector-specific patterns and shed light on whether certain industries are more prone to pay disparities or have adopted more equitable compensation practices. On a global perspective, our study was limited since it did not consider other countries in the dataset. However, for future study, comparing executive and worker compensation across different countries and regions can provide a broader context for understanding the role of cultural factor in the influence of the CEO-worker pay ratio.

**XV. References:**

1. Business Review at Berkeley <https://businessreview.berkeley.edu/why-your-favorite-companies-are-privately-held/>
2. Gender Pay Gap for Executives Grew During Pandemic <https://shrm.org>
3. Kaggle, Top 3000 CEO-Salary Ratio. <https://www.kaggle.com/datasets/salimwid/latest-top-3000-companies-ceo-salary-202223>​
4. Russell 3000 Index, Russell Indices <https://www.barrons.com/market-data/indexes/rua>​
5. You are more likely to have a CEO named David than you are to have a woman at the helm, new report finds <https://www.bizjournals.com/bizwomen/news/latest-news/2023/04/you-are-more-likely-to-have-a-ceo-named-david-than.html>​
6. Women CEOs run 10.4% of Fortune 500 companies. <https://fortune.com/2023/06/05/fortune-500-companies-2023-women-10-percent/>​
7. Gender Pay Gap for Executives Grew During Pandemic ([https://shrm.org](https://shrm.org/))​
8. Dudley, T. (n.d.). Total Number of American Businesses. <https://www.certifiedeo.com/blog-posts/how-many-businesses-are-there-in-america-and-what-does-it-mean-for-employee-ownership>
9. Biery, M. E. (2013). 4 things you don’t know about private companies. Retrieved from https://www.forbes.com/sites/sageworks/2013/05/26/4-things-you-dont-know-about-private-companies/?sh=29287a89291a

**Appendix A:**

Top 100:

|  |  |
| --- | --- |
| **Industry** | **Count** |
| Consumer Discretionary | 54 |
| Information Technology | 18 |
| Consumer Staples | 8 |
| Communication Services | 7 |
| Health Care | 4 |
| Industrials | 4 |
| Materials | 3 |
| Financials | 2 |

Bottom 100:

|  |  |
| --- | --- |
| **Industry** | **Count** |
| Health Care | 29 |
| Financials | 18 |
| Information Technology | 17 |
| Real Estate | 12 |
| Industrials | 6 |
| Consumer Discretionary | 5 |
| Communication Services | 5 |
| Energy | 4 |
| Materials | 2 |
| Consumer Staples | 1 |
| Utilities | 1 |