Woman in Workplace *

Faustine Fan

April 20, 2023

Abstract

Newspaper(Will add after I finished this paper:)

Table of contents

1 Intoduction **2**

2 Data **2**
2.1 Source and Methodology 2
2.2 Weakness and Strength 3

3 Model **4**
3.1 Linear Regression 4
3.2 Summary 5

4 Results **6**
4.1 Woman Fortune 500 CEOs 6
4.2 Woman University Presidents in US 7
4.3 Hihg-Education Gender Comparison in US 8
4.4 Gender Wage Gap in US 9

5 Discussion **10**
5.1 Woman Fortune 500 CEOs 10
5.2 Woman University Presidents in US 10
5.3 Hihg-Education Gender Comparison in US 10
5.4 Gender Wage Gap in US 10
5.5 Limitations and Next steps 10

Appendix **11**
Datasheet for the Dataset 11

References **16**

*Code and data are available at:https://github.com/Faustine123/Newspapers-in-Times-of-Low-Advertising-Revenues

1 Intoduction

Gender inequality in the workplace is a pervasive issue that has been a topic of concern for many years. Despite significant progress in recent years, women continue to face a range of challenges in the workplace, including barriers to entry, lower wages, reduced opportunities for advancement, and gender-based discrimination. These issues can have a profound impact on women's lives, affecting their financial security, professional development and overall well-being. The lack of workplace equality between genders can be attributed to factors such as long-standing patriarchal sociocultural narratives, differences in education levels, unconscious employer bias and outdated cultural narratives that continue to limit women's potential in the workplace.

Although progress has been made on gender inequality in the workplace in recent decades with the awakening and resistance of female self-awareness, gender gaps still exist and women continue to face challenges specific to their gender. It is therefore important to continue to explore gender inequality in the workplace, identify the root causes of the problem, and develop strategies to address it. One of the key factors contributing to gender inequality in the workplace is the cultural narrative of patriarchy that is deeply rooted in our society. In addition, women often face unconscious bias from employers and co-workers, and women's marital and reproductive status can be included in invisible lists of discrimination, all of which without question further limit their access to career advancement and earning potential.

Another key manifestation of gender inequality in the workplace is the lack of female representation in leadership positions. Despite the progress made in recent years, women still make up only a small percentage of executives, including CEOs and university presidents. This lack of representation can limit women's access to advancement opportunities and can perpetuate a culture of gender-based discrimination within organizations. This lack of representation can stem from society and from women themselves. Additionally, the gender pay gap is an important factor in gender inequality in the workplace. The most cited we have seen is that for every $1 an adult male earns, a female can only earn $0.70. To earn the same amount, women need to work harder (Kidd and Shannon 1996). The causes of the pay gap are complex and multifaceted, but they are sometimes rooted in discrimination and bias based on unconscious gender.

This paper will explore gender inequality in the workplace through education levels, the percentage of female CEOs in Fortune 500 companies, the percentage of female university presidents, and pay gap data.By examining these factors, we can better understand the root causes of gender inequality in the workplace and look for ways to potentially eliminate these deep-rooted biases and discrimination.

2 Data

2.1 Source and Methodology

This article is based on data collected in data.world. Women Fortune 500 CEOs dataset from Fortune 500 and Catalyst. The Gender Wage Gap and Adult Education Level datasets are from OECD.org.

The Fortune 500 is an annual list compiled by Fortune magazine that ranks the top 500 companies in the United States based on their total revenue. The companies on this list are leaders in their respective industries, and their financial success is seen as a reflection of their overall management and competitiveness. Catalyst, on the other hand, is a global nonprofit organization dedicated to promoting gender equality in the workplace.Catalyst's research and initiatives have made significant progress in gender diversity and equality in corporate leadership, particularly at Fortune

500 companies. Their work highlights the benefits of having a diverse leadership team, including increased innovation, better decision-making and improved financial performance.

The American Council on Education (ACE) is a membership organization that represents the interests of U.S. colleges and universities. Founded in 1918, the organization advocates for higher education policies that support access, equity and student success. In addition to its policy work, ACE provides a range of services to its members, including research and data analysis. The American College Presidents Study (ACPS) is a research project conducted by ACE to examine the demographics, experiences and attitudes of college and university presidents in the United States. The study provides a comprehensive overview of the challenges and opportunities facing higher education leaders, including issues related to gender and diversity. Through its research and advocacy efforts, ACE is playing an important role in shaping the future of higher education in the United States.

The Organization for Economic Cooperation and Development (OECD) is an international organization that promotes economic development, prosperity and sustainable growth. It was founded in 1961 and is comprised of 37 member countries, including the United States, Canada, Australia, and several European and Asian countries. The OECD is committed to providing governments with data, analysis and policy recommendations to address a wide range of economic and social issues. One of the OECD's key initiatives in recent years has been the Gender Initiative, which aims to promote gender equality and women's empowerment in a range of policy areas. Through extensive data collection and analysis, the OECD has become a key resource for policy makers and researchers around the world seeking to understand and address the challenges facing modern economies.

To conduct this analysis, I utilized the R programming language (R Core Team 2022) and various software packages for data cleaning, exploration, and visualization. The raw data was obtained in the form of csv. files. To clean and analyze the data, I used the packages tidyverse (Wickham and Girlich 2022) and dplyr (Wickham et al. 2022). To visualize the data, I utilized the ggplot2 (Wickham 2016) package, scales (Wickham and Seidel 2020) package. And I used the lubridate (Grolemund and Wickham 2011) package for the automatically generated dates.

2.2 Weakness and Strength

The strength of these datasets is that they are based on a large representative sample of the U.S. population and provide empirical evidence of gender inequality in the workplace across years. These datasets are the basis for action to address gender inequality in the workplace. By providing specific evidence of gender disparities, policymakers and organizations can design targeted interventions to promote gender equality.

The disadvantage of these datasets is that the data are not updated in real time and we cannot track the most recent data. Second, because Women University Presidents only has data for the U.S., the rest of the datasets used in this paper are filtered for the U.S., which may lead to some ambiguity and bias.

3 Model

3.1 Linear Regression

Building a model to analyze gender trends in various areas such as Fortune 500 CEOs, university presidents, higher education degree completion rates, and the U.S. gender wage gap requires careful consideration of several factors. First, selecting the appropriate variables is critical to creating a model that accurately reflects the relationships between the variables being analyzed. Second, data cleaning and pre-processing are critical to ensure the accuracy of the data and to eliminate any potential bias. Third, choosing the appropriate regression method is important to ensure that the model accurately reflects the relationships between the variables. In this case, a linear regression model may be appropriate because it can show the strength and direction of the relationship between two variables.

Linear regression is a statistical method that is typically used to analyze trends in data over time. In the case of the female Fortune 500 CEO dataset, linear regression can be used to examine trends in the representation of women in CEO positions over the years. By fitting a regression line to the data, it is possible to quantify the rate of change in female CEO representation and identify any significant changes or shifts in trends over time. The equation for the simple linear regression model can be written as y = mx + b, where y denotes the dependent variable, x denotes the independent variable, m denotes the slope of the line, and b denotes the y-intercept. In the data set analyzed, the linear regression model can be used to examine the percentage of female CEOs in the Fortune 500, in addition to the percentage of female presidents of U.S. universities, the gender distribution of U.S. higher education degree completion rates, and the gender wage gap in the U.S. in relation to the independent variable of time. By analyzing the regression coefficients, the direction and strength of the relationship between the variables can be determined and the information used to make predictions or draw conclusions about the data. Linear regression can be a powerful tool for identifying trends and making predictions, but it is important to carefully consider the data and assumptions made when constructing the model.

However, since I used four datasets with different variables, I created their own linear regression model for each dataset to examine the relationship between the two variables. The general hypothesis behind considering this model is that the percentage of female ceo's, female principals, and women completing higher education all increase over time, and the wage gap between the two genders decreases over time.However, in general, the relationships between these variables are reflective of discrimination against women in the workplace, which may be slowly improving.

3.2 Summary

Table 1: Model Summary of

	(1)	(2)	(3)	(4)
(Intercept)	-488.28	-1306.49	-362.12	361.07
	(37.52)	(90.72)	(44.30)	(80.18)
year	0.24	0.66		
	(0.02)	(0.05)		
TIME			0.20	-0.17
			(0.02)	(0.04)
Num.Obs.	24	6	7330	946
R2	0.930	0.988	0.012	0.014
R2 Adj.	0.927	0.985	0.012	0.013
AIC	37.3	18.9	61621.8	7589.7
BIC	40.8	18.3	61642.5	7604.2
RMSE	0.46	0.71	16.18	13.32

A summary plot shown in Table 1 is generated using the modelsummary package (Arel-Bundock 2022) in R (R Core Team 2022).

This table shows a summary of four linear regression models with different variables. Each model is summarized with its intercept, coefficients, and goodness-of-fit measures (e.g., R2, R2Adj., AIC, BIC, and RMSE). The R2 values in the model summaries indicate how well the regression line fits the data, with values ranging from 0 to 1. Model 1 is associated with the Female Fortune 500 CEO dataset, which has an R2 value of 0.93, indicating that the model explains 93% of the variation in the data. Model 2 is associated with U.S. women's college presidents with an R2 value of 0.99, indicating a strong relationship between the independent and dependent variables. Model 3 is related to high education gender comparisons in the United States with a fairly low R2 value of 0.012, indicating that the model does not explain well the variation in the data. Model 4 is related to the gender wage gap in the United States with a negative coefficient on the year variable (-0.17), indicating a decreasing trend in the gender wage gap over time. Overall, these linear regression models provide insight into the relationships between different variables and help make data-driven decisions.

Moreover,the intercept and annual coefficient indicate the starting point and rate of change of the dependent variable over time, respectively. For the female Fortune 500 CEO dataset, the intercept is -488.28 and the annual coefficient is 0.24, which indicates a positive linear relationship between the number of female CEOs and time. Similarly, the U.S. female university presidents dataset shows a positive linear relationship between the number of female presidents and time with an intercept of -1306.49 and an annual coefficient of 0.66.

In addition, the AIC and BIC values represent the quality of the model, with lower values indicating a better fit. Root mean squared error (RMSE) values provide information about the predictive accuracy of the model, with lower values indicating higher accuracy. Overall, the linear regression models in these four datasets provide useful insights into trends in gender diversity in leadership positions.

4 Results

4.1 Woman Fortune 500 CEOs

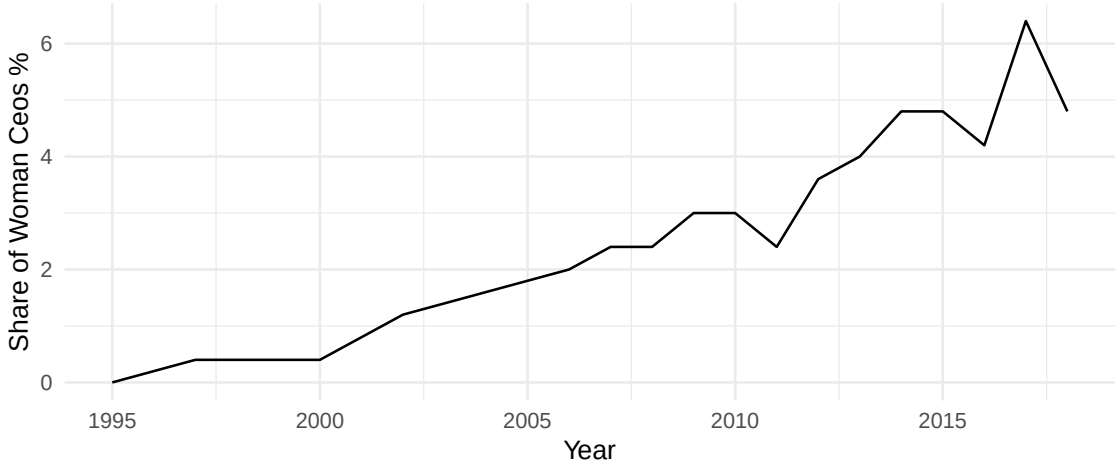

Figure 1: Woman Fortune 500 CEOs

According to the Figure 1, it is clear that the percentage of female CEOs has generally trended upward over the years. in 2000, the percentage of female CEOs in Fortune 500 companies was only 0.4%. Since then, however, there has been a consistent increase, with the percentage reaching 4.8% in 2018. In 2017, the figure reached its highest in 24 years, 6.4%. This shows that more and more women are proving their ability to take charge of the discourse. Despite this growing trend, it is clear that there is still a long way to go before true gender parity is achieved in top leadership positions. The remaining 90 percent or so exclude LGBT people (a group that in some sociocultural narratives may have a harder time getting promoted in companies than women, despite their abilities), thus showing that the percentage of male leaders is a huge majority.

The percentage of female CEOs in the Fortune 500 remains quite low, suggesting that women continue to face significant barriers to entry and advancement in the corporate world. Overall, the chart of Fortune 500 female CEOs shows the progress and challenges in achieving gender equality in the corporate world. While the percentage of female CEOs has increased over time, there is still much work to be done to break down the barriers that prevent women from reaching top leadership positions.

4.2 Woman University Presidents in US

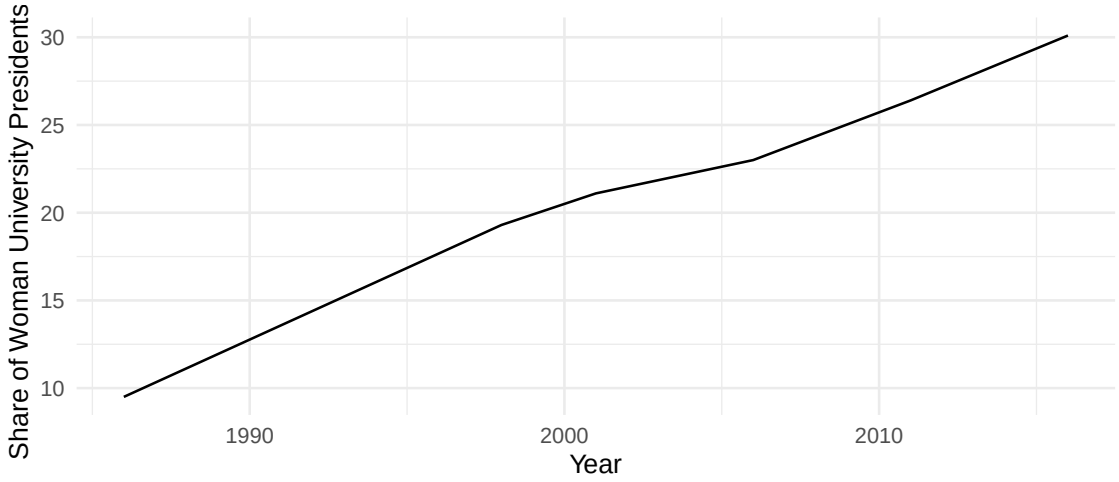

Figure 2: Woman University Presidents in US

The Figure 2 shows the percentage of female presidents within U.S. universities from 1986-2016. The data shows that the percentage of female presidents has continued to rise over three decades, improving by 20.6%. From 1986, when only 9.5% of university presidents were women, it rose to 30.1% in 2016. While this may seem like progress, it is important to note that this increase occurred over a thirty-year period, indicating that university leadership has been slow to make progress toward gender parity. The percentages in this data are based on accredited U.S. degree-granting institutions, which means the data are relatively credible. Overall, the graph highlights the continued underrepresentation of women in university leadership positions and the need for additional efforts to break down barriers and promote gender equity. Universities must actively work to achieve gender diversity in leadership to promote the inclusion and representation of women in higher education.

4.3 Hihg-Education Gender Comparison in US

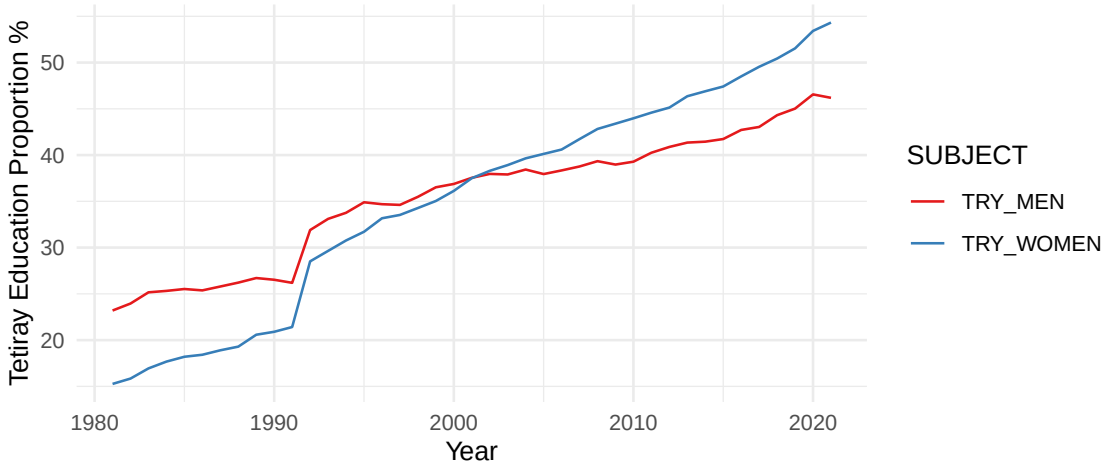

Figure 3: Hihg-Education Gender Comparison in US

Figure 3 shows a comparison of the gender distribution of higher education degree completions for the U.S. population aged 25-64 years from 1981 to 2021. The data show that participation in higher education completion has trended upward dramatically over the past four decades for both women and men. From an initial rate of approximately 19%, the number of people with a higher education degree in 2021 has reached 50.32% of the total population. This trend is a testament to the growing importance of education and the overall positive social development.

Moreover, the gap between the percentage of women and men with a higher education degree has been narrowing since 1990, and in 2002, the percentage of women with a higher education degree was 38.30% compared to 37.96% of men. From 2002-2021, the percentage of women with higher education degrees has been higher than that of men. These data suggest that women are increasingly seeking higher education opportunities, which may reflect a shift toward greater gender equity in educational attainment. It is important to note that while the data show significant progress in educational attainment, it does not necessarily translate into equal representation or opportunity in the workforce.

4.4 Gender Wage Gap in US

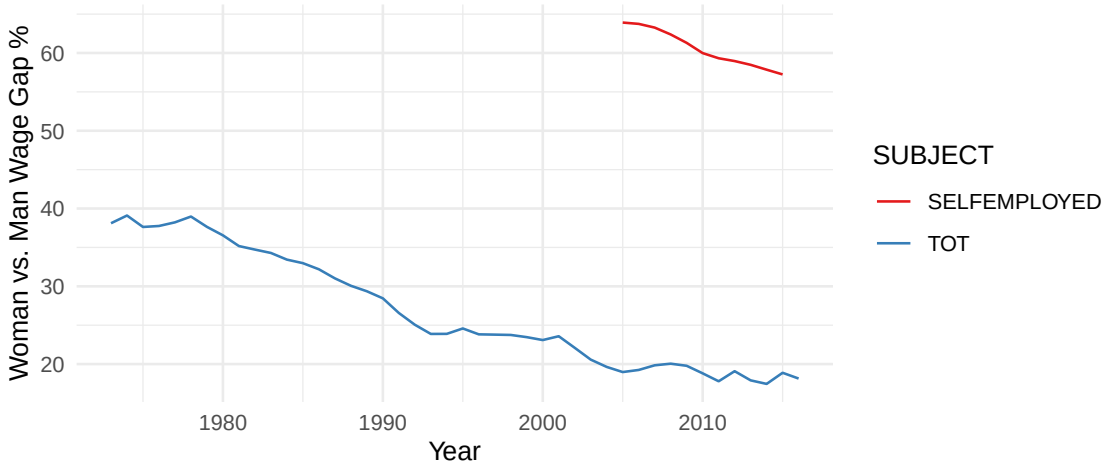

Figure 4: Gender Wage Gap in US

The Figure 4 shows the gender wage gap for full-time employees in the U.S. from 1973 to 2016 and the gender wage gap for self-employed workers in the U.S. from 2005 to 2015. Overall, the wage gap between the genders of full-time employees has been gradually decreasing from about 40% initially to about 20% in 2016. the wage gap decreased faster around 1980 to around 1990 however, since then, the rate of decline has slowed significantly. in 2016, men earned only 82 cents for every dollar earned by women. The gender earnings gap for the self-employed declined from 63.91% in 2005 to 57.24% in 2015. Despite the declining numbers, this large wage gap is sobering.

While the persistence of the wage gap can be attributed to a variety of factors, it is of concern that men hold more of society's resources, which I believe is the primary cause of these gaps. While there have been efforts to address the wage gap, such as equal pay legislation and initiatives to promote pay transparency, much more needs to be done to ensure that women are paid fairly for their work. The gender wage gap not only affects women's economic well-being, but also has broader implications for gender equality and social justice.

5 Discussion

5.1 Woman Fortune 500 CEOs

5.2 Woman University Presidents in US

5.3 Hihg-Education Gender Comparison in US

5.4 Gender Wage Gap in US

5.5 Limitations and Next steps

Appendix

Datasheet for the Dataset

Extract of the questions from (Gebru et al. 2021)

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*
   - This dataset was created to analyze discrimination and bias against women in the workplace. Incomplete data is needed for incomplete gaps.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
   - The survey used to collect the dataset was created by Yixin Fan, a undergraduate student at the University of Toronto and the author of this paper.
3. *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*
   - No organization funded the creation of the dataset.
4. *Any other comments?*
   - None.

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*
   - Each instance in the dataset represents the occupancy of women in each dataset study.
2. *How many instances are there in total (of each type, if appropriate)?*
   - There are a total of four datasets in this paper, each containing a different type of sample instances and a different number.
3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*
   - The dataset is a sample of instances. It is obtained through statistics from various surveys. The lack of breadth in the collection of the dataset results in bias and infernal limitations.
4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*
   - Each instance includes the occupancy of the instance itself in the dataset, i.e. the percentage of females.
5. *Is there a label or target associated with each instance? If so, please provide a description.*
   - the gender is either male, female, or other.
6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*
   - Data and sample sizes for some years are lacking in the dataset due to insufficient sample size.

7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*
   - There are no relationships between individual instances.
8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*
   - None.
9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*
   - None
10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*
    - The dataset is self-contained.
11. *Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*
    - None.
12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*
    - The dataset does not contain any offensive or insulting data.
13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*
    - The datasets are split by age groups and countries as self-identified by the respondents.
14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*
    - It is entirely impossible to identify individuals from the dataset.
15. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*
    - The datasets do not contain data that might be considered sensitive in any way.
16. *Any other comments?*
    - None.

Collection process 1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.* - The data was collected through data.world. Women Fortune 500 CEOs dataset from Fortune 500 and Catalyst. The Gender Wage Gap and Adult Education Level datasets are from OECD.org. 2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures*

validated? - Survey. 3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?* - None. 4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?* - No one but the author is involved in the data collection process. 5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.* - The data was collected over the span of five days from April 16, 2023 to April 20, 2023. 6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.* - No ethical review processes were conducted. 7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?* - The data was collected through data.world. Women Fortune 500 CEOs dataset from Fortune 500 and Catalyst. The Gender Wage Gap and Adult Education Level datasets are from OECD.org. 8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.* - None. 9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.* - Yes. The dataset was collected long time ago, and by official organizations. 10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).* - None. 11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.* - N/A. 12. *Any other comments?* - None.

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*
    - Cleaned the datasets by filtering for better graphing.
2. *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.*
    - The raw data is accessable in the github link provided in this paper.
3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*
    - statistical programming language R (R Core Team 2022) was used to preprocess the datasets.
4. *Any other comments?*
    - None.

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*
    - Possible, those datasets are representive.

2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*
    - The repository for this paper, which uses the datasets, are available at https://github.com/Faustine123/Newsp in-Times-of-Low-Advertising-Revenues.
3. *What (other) tasks could the dataset be used for?*
    - The datasets could be used to modeling or finding the discrimination of woman in society.
4. *Is there anything about the composition of the dataset or the way it was collected and pre-processed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*
    - None.
5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*
    - None.
6. *Any other comments?*
    - None.

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*
    - Since there dataset is publicly available through the github link, yes.
2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*
    - The dataset is distributed on Faustine Fan's github page, along with this paper: https://github.com/Faustine123/Woman-in-Workplace.git.
3. *When will the dataset be distributed?*

- The datasets used in this paper is distributed in 2017 and 2018.

4. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*
    - The dataset is distributed under American Council on Education, The American College President Study and Fortune 500 and Catalyst and the other 2 datasets without indication.
5. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*
    - No restrictions are placed on the dataset.
6. *Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*
    - None.
7. *Any other comments?*
    - None.

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
    - Yixin Fan is hosting the dataset on the github link at https://github.com/Faustine123/Woman-in-Workplace.git.
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
    - The creator of the datasets, Yixin Fan, can be contacted at yx.fan@mail.utoronto.ca.
3. *Is there an erratum? If so, please provide a link or other access point.*
    - As of writing this datasheet, no explicit erratum have been spotted.
4. *Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?*
    - Any updates will be posted on the github page used to host this paper which include the datasets.
5. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.*
    - N/A.
6. *Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*
    - Older versions of the datasets (if there will be any) will be kept as a record on github.
7. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*
    - One should contact Yixin Fan, the author of the dataset.
8. *Any other comments?*
    - None.

Additional Details{-}

```
- Code and data available at: https://github.com/Faustine123/Woman-in-Workplace.git

- A simulation of the dataset was conducted in scripts/Simulted_data.R which helped formulate plans
```

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

Kidd, Michael P., and Michael Shannon. 1996. "The Gender Wage Gap: A Comparison of Australia and Canada." *ILR Review* 49 (4): 729–46. https://doi.org/10.1177/001979399604900409.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Maximilian Girlich. 2022. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

Wickham, Hadley, and Dana Seidel. 2020. *Scales: Scale Functions for Visualization*. https://CRAN.R-project.org/package=scales.