MGT 6203 Group Project

A Statistical Analysis of the Gender Pay Gap

GitHub: MGT-6203-Fall-2023-Canvas/Team-26

Team 26

Marissa Robinson

Jamie Hernandez Kluesner

William D'Onofrio

Sanjay Lindsay

Emmett Drake

Background

As we commemorate the 19th Amendment's 103rd anniversary, we reflect on the substantial progress made in women's rights in the United States and acknowledge the unfinished journey toward full gender equality. Despite historic strides—from the 19th amendment to the second wave of feminism in the 1960s, and the surge of women in higher education—the gender pay gap persists, marking a significant challenge in achieving equal compensation [4]. This analysis aims to dissect the factors contributing to this wage discrepancy through the lens of various theoretical frameworks and legal statutes.

From a business standpoint, bridging the gender pay gap transcends ethical imperatives; it is strategically imperative in the marketplace. Equal pay for equal work is not only a matter of compliance with anti-discrimination laws, such as the Equal Pay Act of 1963 and the Lilly Ledbetter Fair Pay Act but is also fundamental to fostering a diverse and dynamic workforce [3,7]. Drawing from Labor Market Segmentation Theory and Human Capital Theory, our analysis considers how educational attainment, occupational segregation, and seniority impact salary outcomes. Additionally, feminist economics provides a critical perspective on the systemic undervaluing of work traditionally done by women, challenging the economic structures that perpetuate gender disparities [5,8]. To provide further context, definitions of the theoretical frameworks and legal statutes are as follows:

- **19th Amendment:** a historic U.S. constitutional amendment ratified in 1920 that granted women the right to vote, symbolizing a significant milestone in the women's suffrage movement and representing an early step toward gender equality.
- **Equal Pay Act of 1963:** a U.S. federal law aimed at abolishing wage disparity based on sex. It mandates equal pay for equal work by men and women.
- **Lilly Ledbetter Fair Pay Act:** an act signed into law in 2009, which restores and extends the statute of limitations for filing an equal-pay lawsuit regarding pay discrimination.
- **Labor Market Segmentation Theory:** this theory posits that the labor market is divided into separate segments or sectors, each with its own norms and rules, which can lead to and perpetuate wage disparities.
- **Human Capital Theory:** this theory suggests that economic development depends on the quality of human capital, which includes the skills, knowledge, and experience of the workforce, and implies that investment in human capital will lead to better salary outcomes.
- **Feminist Economics:** a branch of economics that critiques traditional economics for biases against women and examines the systematic undervaluation of work traditionally performed by women.

In essence, this analysis aims not only to identify factors linked to the pay gap but also to inform the implementation of comprehensive strategies. Aligned with the spirit of historic legislative efforts and contemporary economic theories, the goal is to support organizations in offering competitive financial incentives and upholding the principles of gender equality.

This approach not only attracts and retains top talent but also contributes to sustained organizational success.

Initial hypothesis

Several hypotheses will guide our investigation of the gender pay gap. First, we hypothesize that factors such as performance evaluation scores, education level, and seniority (years of experience) will be highly correlated with the pay disparity between men and women. Specifically, we expect that women with lower performance scores, lower education levels, and fewer years of experience will experience a greater pay disparity compared to men. Turning our attention to role-specific disparities, we hypothesize certain departments and job titles—potentially male-dominated ones like engineering or management—to exhibit heightened gender pay gaps relative to roles where gender representation is more balanced. Additionally, we hypothesize that advanced education acts as a buffer against the gender pay gap. That is, women equipped with advanced degrees, such as master's or PhDs, will experience a lesser pay gap in contrast to those with just high school diplomas or bachelor's degrees. Lastly, in examining age-related variances, we hypothesize that older age brackets, such as those aged 40 and above, will experience increased pay gaps. This is based on the notion that historical gender biases have left their mark on these cohorts, while younger age groups, molded by evolving societal norms, may present a narrowing gap.

The Data

Data Preprocessing & Exploration:

The dataset under examination, designated as 'Glassdoor Gender Pay Gap.csv,' was procured from the Kaggle platform [2]. Notably, Kaggle has indicated that the original compilation of this dataset was conducted through Glassdoor. It is vital to recognize that Glassdoor functions as an opt-in platform where users voluntarily report their data. Consequently, the platform does not undertake verification procedures to ascertain the accuracy of the user-submitted data. This scenario inherently introduces the possibility of response bias. Such bias may manifest as the dataset predominantly reflecting the experiences of those individuals who have chosen to participate in the data submission process, potentially skewing the data away from a representation of the broader demographic spectrum. However, the specific method employed for the data's extraction and compilation by Glassdoor remains unclear, rendering the extent of selection bias indeterminate but nevertheless a pertinent factor for consideration in the analysis.

The dataset contains information such as Job Title, Gender, Age, Performance Evaluation, Education, Department, Seniority, Base Pay, and Bonus. From the summary statistics, we observed that our dataset has:

- Ages ranging from 18 to 65, with a median age of 41.
- Performance evaluations scores between 1 to 5, averaging around 3.

- Base salaries ranging from $34,208 to $179,726 with a median base salary of $93,328.
- Bonuses ranging from $1,703 to $11,293, with a median bonus of $6,507.
- Furthermore, regarding feature engineering, we created a Salary variable by summing the Base Pay and Bonus for each record.

Exploratory Data Analysis (EDA):

A key aspect of the EDA involved creating a new variable, Salary, which combines base pay with bonuses to reflect total earnings. The data cleaning process was found to be minimal, as no missing values were detected. This streamlined the analysis, allowing for an uninterrupted focus on interpreting the data.

In terms of outliers, a methodical approach was taken where the interquartile range (IQR) was calculated, and boundaries were set to identify salaries that fell outside this range. Only three potential outliers were identified and all three were identified to be high earning males. Further analysis revealed that no female employees matched the outlier criteria set based on the high-earning male outliers, indicating a potential gender-based discrepancy in high-earning roles. Therefore, these values were retained for subsequent analysis to preserve the integrity of the dataset and avoid potential loss of valuable information. Additional feature engineering was also employed by categorizing the continuous age variable into discrete groups, facilitating the analysis of salary across age categories and enabling the examination of the interaction between age and gender in relation to pay.

In conclusion, the EDA was integral in shaping the analytical framework. It provided crucial insights into the gender pay gap and brought to light the impact of various socio-demographic factors on salaries. The process of creating new variables and refining existing ones proved successful, setting the stage for a more nuanced and targeted investigation into the complexities of salary structures.

Hypothesis 1: The Impact of Job Performance, Education Level, and Seniority on Pay Disparity is Moderated by Gender

To deepen our understanding of the factors influencing salary, particularly in terms of gender equality in the workplace, a multiple linear regression analysis was conducted. This method was chosen to assess how a combination of variables—performance evaluation, education, seniority, and gender—predicts salary variations. The model also explored if the impact of these factors on salary varies with gender, using interaction terms.

The regression analysis resulted in a statistically significant model with a 46.15 F-statistic and corresponding $2.2e^{-16}$ p-value and yielded several crucial insights. Foremost, gender stood out as a significant salary determinant, a finding that is visually underscored in Figure 2 "Salary vs. Performance Evaluation by Gender", where males are consistently positioned at higher salary brackets across all performance evaluation scores. The analysis revealed that being male correlates with higher salaries at the 0.01 alpha significance level, even when

factoring in performance evaluation, education, and seniority. This finding resonates with historical evidence of the gender pay gap and is graphically represented by the separation between the genders in the Figure 2 scatter plots, highlighting persistent gender-based salary disparities and aligning with theories of labor market discrimination [5].

Seniority was another 0.01 alpha level significant factor in salary determination, as shown in Figure 2 "Salary vs. Seniority by Gender" where both genders exhibit an upward trend in salary with increasing seniority. However, its interaction with gender did not significantly affect salaries, suggesting a uniform effect of seniority on salaries across genders, which aligns with human capital theory's emphasis on experience and tenure [3]. This is visually evidenced by the parallel progression of salary increases for both genders as seniority levels rise.

In terms of education, Figure 2 "Average Salary by Education Level and Gender" bar chart reveals that higher educational achievements, particularly a PhD, were linked to higher salaries, supporting human capital theory. Yet, it also visually conveys that within each educational level, a gender pay gap exists, with males receiving higher average salaries than females, emphasizing the gender-neutral impact of education on salary.

Overall, this analysis supports the hypothesis of gender's significant role in salary disparities and highlights the need for continued efforts to understand and address the gender pay gap's complex dynamics, considering both theoretical frameworks and legislative contexts. The visualizations provide a stark representation of these disparities, emphasizing the necessity for a nuanced examination of the factors at play.

```
Call:
lm(formula = Salary ~ PerfEval + Education + Seniority + Gender +
    PerfEval:Gender + Education:Gender + Seniority:Gender, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-55127 -14447   -751  13238  66683

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     62487.39    3503.19  17.837  < 2e-16 ***
PerfEval                           52.72     666.93   0.079 0.937012
EducationHigh School             -638.90    2586.53  -0.247 0.804950
EducationMasters                 4829.85    2720.51   1.775 0.076148 .
EducationPhD                    10104.63    2726.06   3.707 0.000222 ***
Seniority                       10143.97     685.89  14.789  < 2e-16 ***
GenderMale                      13784.28    4853.39   2.840 0.004602 **
PerfEval:GenderMale               165.49     920.21   0.180 0.857319
EducationHigh School:GenderMale -2956.40    3676.36  -0.804 0.421494
EducationMasters:GenderMale      -442.92    3719.14  -0.119 0.905226
EducationPhD:GenderMale         -5323.50    3772.44  -1.411 0.158513
Seniority:GenderMale            -1130.03     939.09  -1.203 0.229140
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20560 on 988 degrees of freedom
Multiple R-squared:  0.3394,    Adjusted R-squared:  0.3321
F-statistic: 46.15 on 11 and 988 DF,  p-value: < 2.2e-16
```

Figure 1. Linear Regression Model Summary

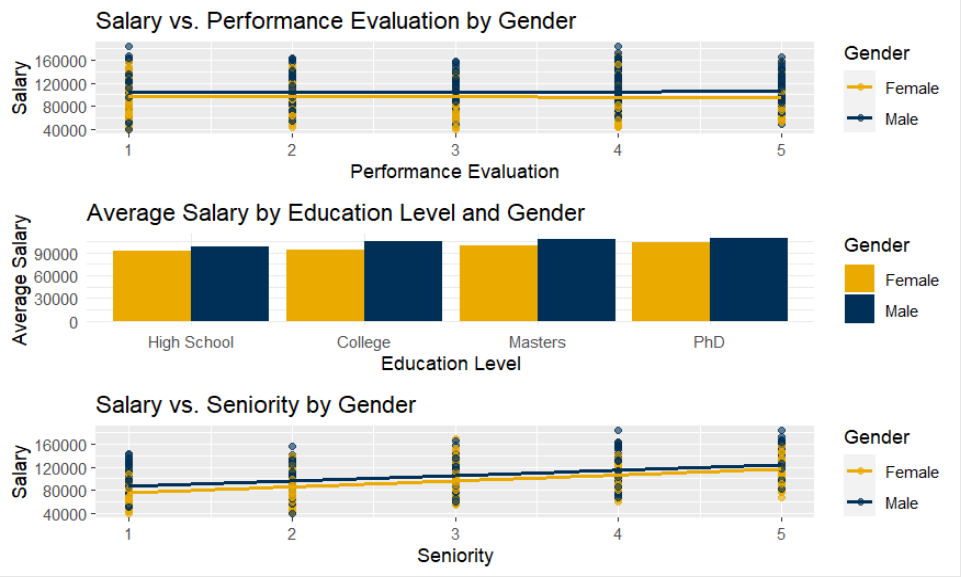

Figure 2. Salary vs Performance Evaluation, Education, & Seniority by Gender

Hypothesis 2: Gender Disparities in Different Departments and Roles

To further explore gender disparities in salary, specifically across different departments and job titles, a second hypothesis was tested using an Analysis of Variance (ANOVA) approach. This statistical method is particularly useful for comparing means across multiple groups, making it ideal for examining whether salary disparities exist based on the combination of Department, Job Title, and Gender. The ANOVA model was structured to assess the interaction effect of these three variables on salary. The model aimed to determine if the average salaries differed significantly when considering these factors in conjunction. Following ANOVA, a Tukey's Honestly Significant Difference (HSD) test was conducted to identify specific group pairings with significant mean differences.

The ANOVA results were significant with a 3.881 F-value and $2e^{-16}$ p-value, and as depicted in Figure 3, demonstrated some variation in salaries across departments and job titles when viewed collectively with gender. The bar chart illustrates a complex pattern of salary distributions, where within some job titles, like Data Scientist, females tend to make as much as or more than males on average. This trend is also observed across most departments for Graphic Designer as well. However, for Software Engineer, this trend inverts, with females making less than males on average, suggesting a less equitable salary distribution. Then for other roles, such as IT or Driver, there is less of a discrepancy between male and female average salaries across each department.

The subsequent Tukey HSD test, however, focused on pairings within the same department and job title across genders, found no significant gender-based salary differences. This is visually corroborated by the close clustering of salary ranges for males and females within the same job titles and departments, suggesting compliance with the principles of the Equal

Pay Act on a role-specific level. Yet, this raises questions about the broader enforcement across different sectors, as the graph shows different salary trends across departments.

This pattern reflects the complexities posited by Intersectionality Theory, which suggests that disparities cannot be fully understood without considering the intersecting influences of job role and departmental context [1]. The lack of significant gender disparities within the same roles in the Tukey HSD test points to a need for more refined analyses that take into account systemic gender biases, which may manifest differently across various labor market sub-sectors.

In summary, the second hypothesis aimed at dissecting gender disparities in salary across departments and job titles revealed that while gender, job title, and department interact to affect salaries, gender differences within the same job title and department are not statistically significant. This nuanced outcome, supported by the detailed salary distributions in the provided graph, calls for a deeper investigation into other potential factors influencing salary disparities and underscores the complexity of addressing the gender pay gap.

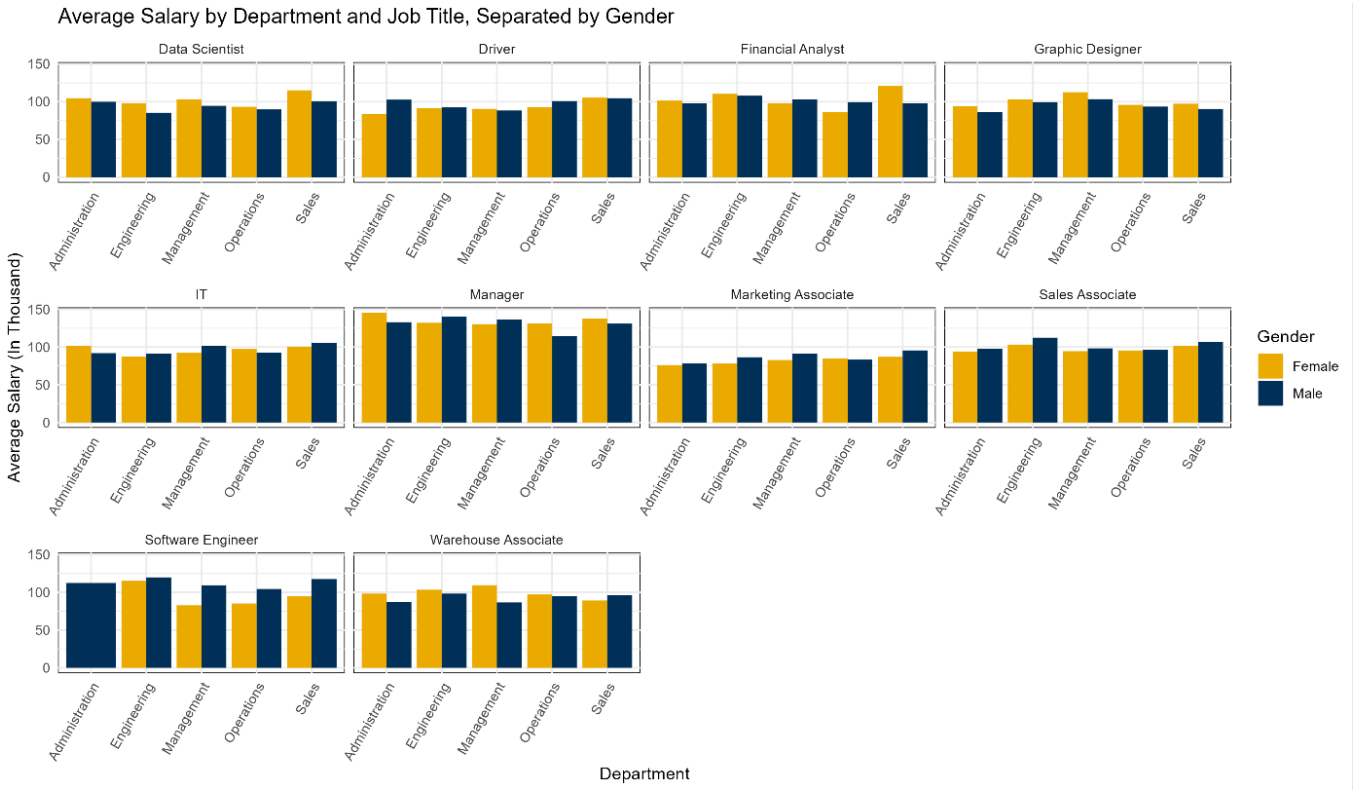

Figure 3. Average Salary by Department & Job Title, separated by Gender

Hypothesis 3: Gender and the Impact of Advanced Education on Pay

The third hypothesis probes the influence of advanced education on salary outcomes, interwoven with the variable of gender and examined through the prism of pertinent legal statutes and economic theories. Utilizing logistic regression, suitable for binary outcomes,

this analysis bifurcated salary into high salary (above median) and not high salary (below median) and explored the predictive potential of education level and gender on this binary salary outcome.

The dichotomization of salary aligns with the binary nature of compliance concerning anti-discrimination laws like the Equal Pay Act of 1963, envisaging equal pay for equal work regardless of gender [7]. Here, education level—spanning High School, College, Masters, and PhD—and gender were predictors in a logistic model estimating the likelihood of earning a high salary.

Statistically significant results from the model underscore that advanced education, particularly at the master's and PhD levels, elevates the odds of high salary echelons, resonating with Human Capital Theory that advocates for the economic value of educational attainment [3]. Concurrently, the finding that males are more likely to earn a high salary invokes the persistent gender disparities that feminist economic theories critique, emphasizing the systemic undervaluing of women's contributions across educational tiers [8].

Figure 4 embodies the interplay of education and gender in salary outcomes, crystallizing the logistic regression analysis. It starkly illustrates the proportion of individuals within each gender and education level crossing the median salary threshold, reinforcing the economic advantage conferred by higher education and the legal imperative to scrutinize and address gender discrepancies in salary.

In sum, the logistic regression corroborates the hypothesis that both advanced education and gender are significant predictors of earning a high salary, furthering the discourse on the value of education in career progression against the backdrop of enduring gender pay disparities. This analysis, therefore, not only reflects upon the strides made since the 19th Amendment and subsequent legislative efforts but also underlines the importance of ongoing strategic interventions to cultivate a workforce landscape where gender equality is realized in practice, not just in principle [7].

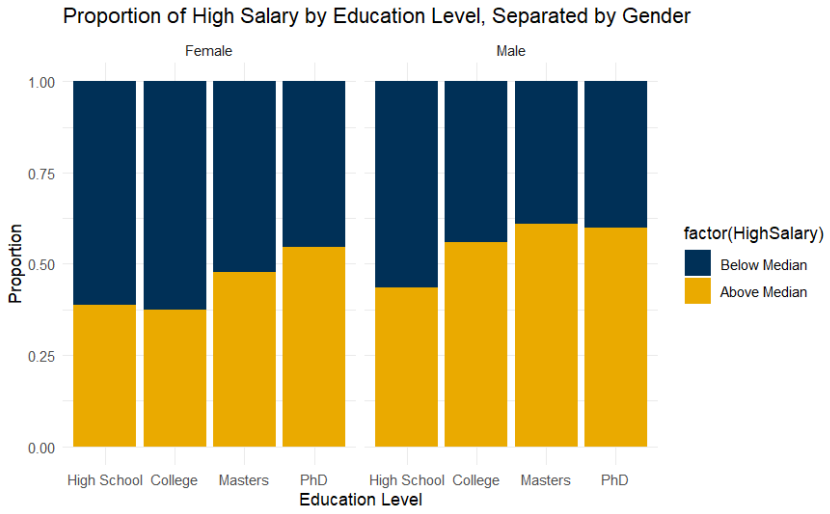

Figure 4. Proportion of High Salary by Education Level, separated by Gender

Hypothesis 4: Gender Pay Gap Variance Across Age Groups

The fourth hypothesis explored the variance in the gender pay gap across different age groups using a multiple linear regression model. This approach is ideal for examining the relationship between a continuous dependent variable (salary) and multiple independent variables, including categorical ones (age groups and gender). To facilitate this analysis, the continuous age variable was categorized into age groups, and an interaction term between age group and gender was included in the model to assess whether the salary disparities between genders varied across different age groups. The results of the regression model indicated that salary does indeed increase with age, which is consistent across all age groups. However, the interaction terms between age groups and gender were not all statistically significant, suggesting that the salary increase with age is generally consistent for both genders, with some variability. A line plot was created to visualize the average salary across age groups by gender, which showed that average salaries tend to increase with age for both genders. However, the plot also revealed that male average salaries were consistently higher than those of females across all age groups.

In summary, the multiple linear regression analysis was inconclusive regarding our initial hypothesis that the gender pay gap worsens starting at age 40. The interaction terms, which did not show a clear impact, suggest that the pay gap does not necessarily worsen with age. However, Figure 5, which shows line plots of average salary versus age groups, indicates an initial and widening gap as age increases, contradicting the model's findings. Nevertheless, these results highlight the complexity of both age and gender when examining salary structures and contribute to the ongoing discussion of a persistent gender pay gap across the lifespan.

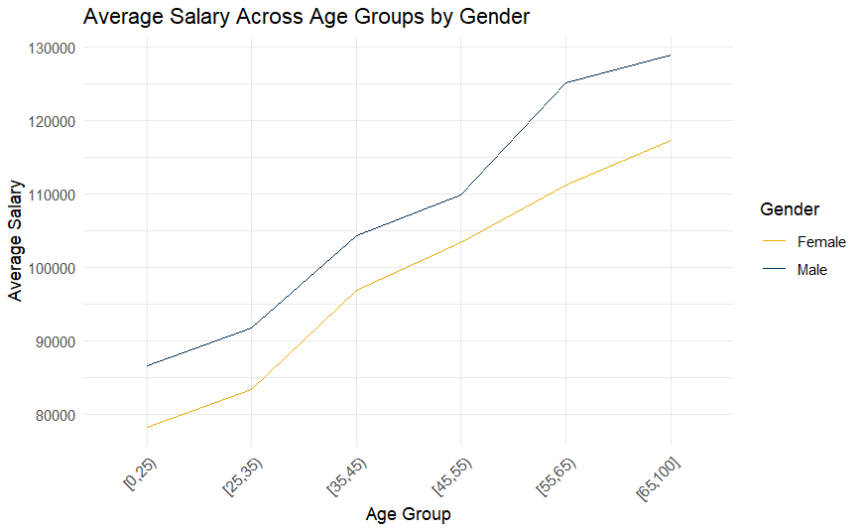

Figure 5. Average Salary across age groups by Gender

Conclusion

In concluding this rigorous study into the persistent gender pay disparities, our suite of statistical models has provided a multifaceted examination consistent with Labor Market Segmentation Theory and Human Capital Theory, while also probing the effectiveness of anti-discrimination laws like the Equal Pay Act of 1963. Through multiple linear regression, logistic regression, and ANOVA analyses, this paper has quantified the extent of salary inequities and revealing statistically significant gender-based disparities that intersect with education, seniority, and age.

Each model—multiple linear regression, ANOVA, and logistic regression—offered unique insights into the dynamics of the gender pay gap, yet their effectiveness varied in terms of explanatory power and specificity. The multiple linear regression models, applied in Hypotheses 1 and 4, exhibited moderate R-Squared values of 0.3394 and 0.3177, respectively. This indicated a reasonable but incomplete capture of the variability in salary, suggesting the presence of other influential factors were not captured in these models. The ANOVA analysis, employed in Hypothesis 2, demonstrated a significant interaction effect of department, job title, and gender on salary, with an F-value (3.881) indicating a moderate effect size. However, its broad categorization might have masked finer nuances that a more granular approach could uncover. The logistic regression in Hypothesis 3 highlighted the influence of education and gender on high salary outcomes, providing a distinct binary perspective on the salary structure. When compared, these models paint a diverse, yet interconnected picture of the gender pay gap, each contributing to a more holistic understanding.

Although the models adeptly captured a moderate proportion of the variance in salaries, the collective evidence they present underscores a systemic issue that transcends mere economic factors, invoking the critical perspectives of feminist economics and the legal

frameworks aimed at gender equity. While these disparities are quantified in the present analysis, their existence calls into question the full realization of the principles enshrined in historic legislation and the need for continued evolution of policies to address these gaps.

The findings herein not only contribute to the existing body of literature on wage inequalities but also serve to guide policy discussions and inform organizational strategies. It is evident that closing the gender pay gap requires a concerted effort that marries theoretical understanding with practical legal enforcement, ensuring that the strides made since the enactment of the 19th Amendment continue to propel us toward a more equitable future in the labor market.

This study has significantly contributed to our understanding of the gender pay gap, yet it also underscores the necessity for continued and more nuanced research in this field. A pivotal area for future investigation is the exploration of intersectionality within wage disparities. It is crucial to examine how these pay gaps intersect across diverse racial, ethnic, and socio-economic backgrounds. Such research requires comprehensive datasets that encompass these various dimensions, offering a more detailed perspective on how intersecting identities impact compensation.

In addition, the changing landscape of work environments, especially the distinction between remote and in-office settings, emerges as a critical domain for future study. The accelerated shift towards remote work, catalyzed by global phenomena like the COVID-19 pandemic, necessitates the development of new datasets that specifically capture these evolving work dynamics. Investigating the gender pay gap within these distinct settings is essential to understanding the influence of workplace models on gender-based compensation.

Moving forward, it is imperative for research in this area to be supported by data that is both expansive and inclusive, capturing the multifaceted nature of the labor market. These enriched datasets will not only facilitate a deeper theoretical understanding but will also be instrumental in guiding policy formulation and organizational strategies. The pursuit of such comprehensive research is vital in ensuring that the progress made towards gender equality is robust and all-encompassing, ultimately leading to a labor market that is equitable and fair for all.

REFERENCES

[1] Atewologun, D. (2018, August 28). *Intersectionality theory and Practice*. Oxford Research Encyclopedia of Business and Management. https://oxfordre.com/business/display/10.1093/acrefore/9780190224851.001.0001/acrefore-9780190224851-e-48

[2] Jauhari, N. (2020, September 12). *Glassdoor- analyze gender pay gap*. Kaggle. https://www.kaggle.com/datasets/nilimajauhari/glassdoor-analyze-gender-pay-gap

[3] Kurt, D. (2023). *Lilly Ledbetter fair pay act: Definition, history, impact*. Investopedia. https://www.investopedia.com/terms/l/lilly-ledbetter-fair-pay-act.asp

[4] National Archives and Records Administration. (n.d.). *19th Amendment to the U.S. Constitution: Women's Right To Vote (1920)*. National Archives and Records Administration. https://www.archives.gov/milestone-documents/19th-amendment

[5] Schieder, J., & Gould, E. (2016, July 20). *"Women's work" and the gender pay gap: How discrimination, societal norms, and other forces affect women's occupational choices-and their pay*. Economic Policy Institute. https://www.epi.org/publication/womens-work-and-the-gender-pay-gap-how-discrimination-societal-norms-and-other-forces-affect-womens-occupational-choices-and-their-pay/?gclid=CjwKCAiA9dGqBhAqEiwAmRpTC825_lMKwpF91F4oPuyqAtDBqhZ9m2auxLTa8hHkDbeKGo5z4B-nqxoCDncQAvD_BwE.

[6] Second wave feminism primary sources & history | gale. (n.d.). https://www.gale.com/primary-sources/womens-studies/collections/second-wave-feminism

[7] The United States Government. (2023, June 21). *On anniversary of equal pay act, signs of progress and remaining challenges for women in the labor market*. The White House. https://www.whitehouse.gov/cea/written-materials/2023/06/21/on-anniversary-of-equal-pay-act-signs-of-progress-and-remaining-challenges-for-women-in-the-labor-market/

[8] Williams, E. (2018, August 29). *What is feminist economics?*. Womens Budget Group. https://wbg.org.uk/blog/what-is-feminist-economics/

[9] Wolfers, J. (2021, November 23). *More women than men are going to college. that may change the economy.* The New York Times. https://www.nytimes.com/2021/11/23/business/dealbook/women-college-economy.html