# Fairness of Black Saber Software's Hiring, Promotion, and Salary Processes

Potential Bias Against Women In AI Hiring Algorithm and Salary Processes

Report prepared for Black Saber Software by Emimi Co

2021-04-21

# Contents

# Executive summary

## Background and Aim

This report is prepared by Emimi Co for Black Saber Software to address the concerns regarding potential bias Black Saber Software's hiring and renumeration process and determine if there are any potential issues that Black Saber Software should be aware of. This report aims to determine if there are any gender biases in the 3 phases of Black Saber Softwares hiring process that is conducted with the help of AI and salary and promotion processes.

## Key Findings

The key findings of this

- There is no evidence of gender biases in the selection of which applicants proceed to the next phase of the hiring process.
- There is no evidence of the gender biases in the interviewers ratings in phase 3 of hiring process. However, there is potential gender bias in the algorithm employed by the AI when scoring an applicants leadership presence and speaking skills. Specifically, the AI tends to give women lower leadership presence and speaking skill scores than men. This can be visualized in Figure 1.
- The is evidence of gender biases in salary and promotion processes. This can be visualized in Figure 2.

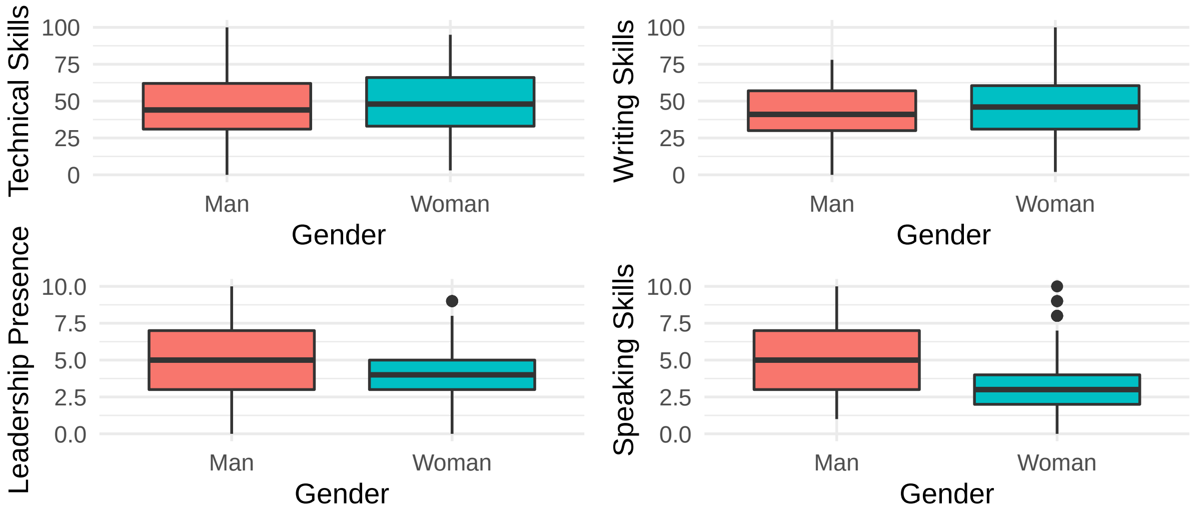

Figure 1: Boxplots of Skill Scores

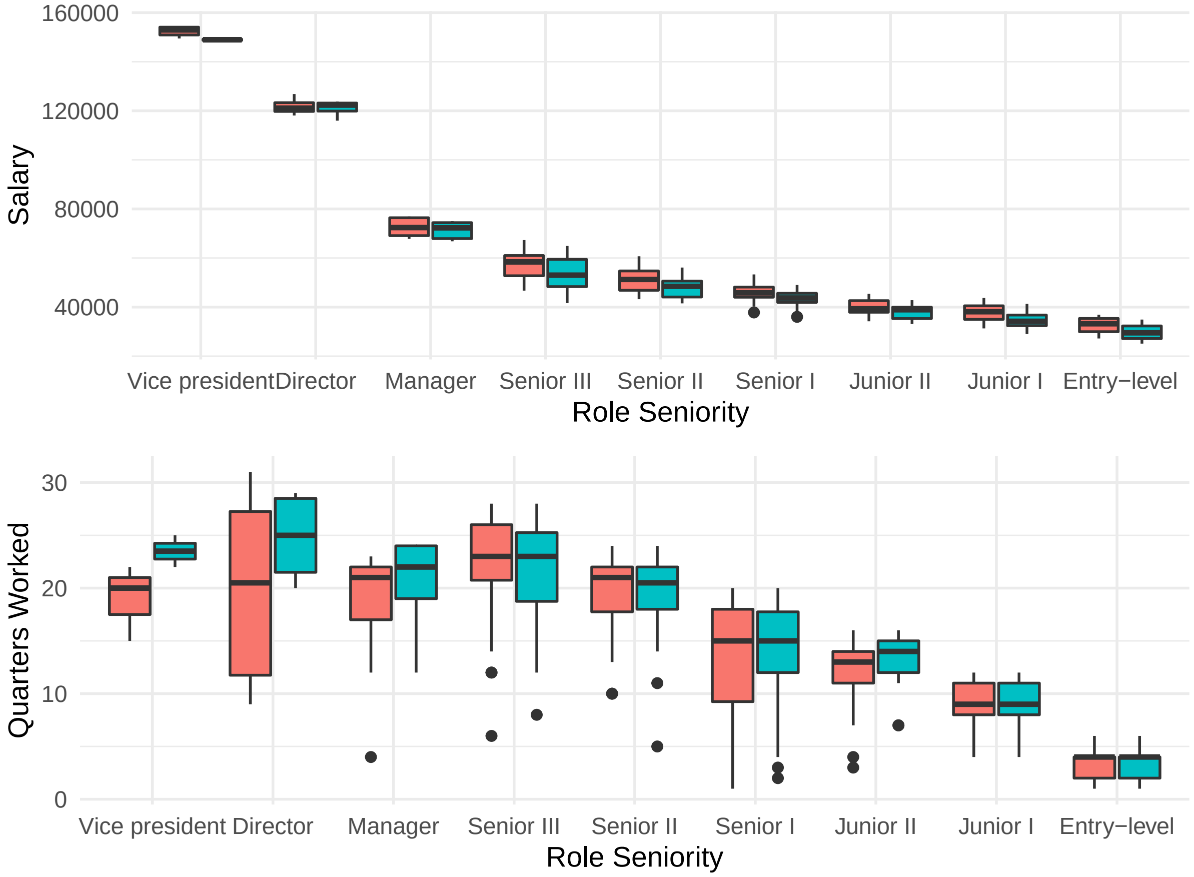

Figure 2: Boxplots of Salary and Quarters Worked

## Limitations

Some limitations include:

- Due to the fact that ethnicity and race data was not collected, whether or not there are potential biases due to ethinicity/race in the hiring and renumeration process was unable to be investigated.
- Whether or not there is gender bias in the salary and promotion process was only investigated for the lastest financial quarter, so claims cannot be made for the previous financial quarters.
- This report warns of there potentially being a gender bias in the promotion process but there is not solid evidence for or against this claim.

# Technical report

## Introduction

This report will address the concerns about Black Saber Software's hiring and renumeration process, especially whether or not these processses are free of gender bias. This report uses hiring data from Black Saber Software's new grad program and data about the promotion and salary of their current staff. The data will be used to assess whether or not Black Saber Software's hiring, promotion, and salary processes are based on talent and value to the company, and free of gender biases.

### Research questions

In particular, the questions that will be addressed are:

- Is Black Saber Software's AI used in the hiring processes free of gender biases when chosing which applicants proceed to the next phase?
- Is Black Saber Software's AI used in the hiring process free of gender biases when scoring applicants? Are the interviewers in phase 3 of the hiring process free of gender bias?
- Is Black Saber Software's promotion and salary processes based on talent and value to the company as assessed by how long an employee worked for the company, and their productivity and leadership skills? Is this process free of gender bias?

## Description of Datasets

The hiring data consists of 4 datasets:

- Phase 1 data contains data about the applicant (ie. an unique ID, the position they applied for, gender, gpa), whether or not the submitted a cover letter or cv (1 if submitted and 0 if did not submit), and a extracurricular score and work experience score using an AI that compares proprietary key terms to a phrase bank (takes on values 0, 1, and 2).
- Phase 2 data contains the same data as phase 1 except only for applicants that passed phase 1, along with a score for their technical skills, writing skills, leadership presence and speaking skills as rated by an AI.
- Phase 3 data contains the applicant ID of the candidates that moved on from phase 2 of the hiring process and 2 ratings on job fit given by two different interviewers.
- The data about the final hires only contains the applicant IDs of the successful candidtaes.

The data about Black Saber Software's employees data contains data about the current employees for the entire duration of their employment. It contains data about an employee's ID, their position within the company, gender, leadership, productivity and salary, across all financial quarters for which they worked.

## Fairness of Selection of which Applicants Move onto the Next Phase

### Fairness in Judging which Applicants Move onto Phase 2 of the Hiring Process

Phase 1 to Phase 2:

To judge the fairness of the AI in selecting which applicants moved onto phase 2 of the hiring process, the data from phase 1 was compared to the data from phase 2 to determine the conditions which results in a candidates rejection. This was done by getting a summary of the data from phase 1 and phase 2.

**Table 1:** Phase 1 Summary

| | Cover Letter | CV | GPA | Extracurriculars | Work Experience |
|---|---|---|---|---|---|
| Min. | 0.00 | 0.00 | 1.20 | 0.00 | 0.00 |
| 1st Qu. | 0.00 | 1.00 | 2.20 | 1.00 | 1.00 |
| Median | 1.00 | 1.00 | 2.70 | 1.00 | 1.00 |
| Mean | 0.64 | 0.88 | 2.71 | 1.22 | 0.96 |
| 3rd Qu. | 1.00 | 1.00 | 3.20 | 2.00 | 1.00 |
| Max. | 1.00 | 1.00 | 4.00 | 2.00 | 2.00 |

**Table 2:** Phase 2 Summary

| | Cover Letter | CV | GPA | Extracurriculars | Work Experience |
|---|---|---|---|---|---|
| Min. | 1 | 1 | 2.00 | 1.0 | 0.00 |
| 1st Qu. | 1 | 1 | 2.70 | 1.0 | 1.00 |
| Median | 1 | 1 | 3.20 | 1.0 | 1.00 |
| Mean | 1 | 1 | 3.11 | 1.4 | 1.15 |

| | Cover Letter | CV | GPA | Extracurriculars | Work Experience |
|---|---|---|---|---|---|
| 3rd Qu. | 1 | 1 | 3.52 | 2.0 | 1.00 |
| Max. | 1 | 1 | 4.00 | 2.0 | 2.00 |

By comparing the results from table 1 and table 2, we see that the minimum value for both cover letter and cv in phase 2 are 1, which means that any applicants that fail to submit both a cover letter and cv are automatically rejected. Furthermore, the minimum gpa of phase 2 is 2.00, so applicants with less than a 2.00 are eliminated. Lastly, the minimum extracurriculars score of phase 2 is 1, so any applicants with a score less than 1 are also rejected.

Next to determine if there where any other conditions that results in an applicants rejection, the hiring data from phase 1 and phase 2 were full joined using the applicant's ID. Then the data was filtered for all the applicants that meet the requirements above (ie. submitted cover letter and cv, has GPA greater or equal to 2, and extracurricular score greater or equal to 1) but did not move onto the second phase as determined by no technical skill score. By looking at the filtered results, it is shown that applicants with an extracurricular score of 1 and work experience score of 0 were also rejected.

Therefore, the process of determining which applicants moved on from phase 1 to phase 2 were determined by objective criteria, so this part of the hiring process is fair as all applicants are subject to the same criteria.

Phase 2 to Phase 3:

To assess the fairness in selecting which applicants moved onto phase 2 of the hiring process, the data from phase 2 was left joined to the data from phase 3. Then a new variable was created to indicated whether or not an applicant moved onto phase 3 (1 if passed, 0 otherwise). Furthermore, the gender of the applicants were manipulated such that any applicants that did not give a gender were considered to be women. This was chosen based on suggestions on what to do if a person provides a gender that is not male or female.

Next, a logistic regression model was built to determine the factors that result in whether or not an applicant moves onto phase 3. The model models whether or not an applicant passed based on their gpa, extracurriculars, work_experience, technical_skills, writing_skills, leadership_presence, speaking_skills and gender. Although gender shouldn't determine whether or not an applicant moves onto the next phase in an unbiased process, it was included because if the p-value of it's regression coefficient is less that 0.05 there is an indication that gender affects the hiring process, which indicates that thier is evidence that gender affects this part of the hiring process. The other variables were included because they could be factors that are looked at when comparing applicants in real life.

**Table 3:** Regression Results for Passing Phase 2

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | -0.393 | 0.090 | -4.351 | 0.000 |
| gpa | 0.004 | 0.029 | 0.130 | 0.897 |
| extracurriculars | 0.008 | 0.029 | 0.266 | 0.790 |
| work_experience | -0.030 | 0.034 | -0.867 | 0.387 |
| technical_skills | 0.003 | 0.001 | 3.934 | 0.000 |
| writing_skills | 0.003 | 0.001 | 3.540 | 0.000 |
| leadership_presence | 0.032 | 0.006 | 5.377 | 0.000 |
| speaking_skills | 0.023 | 0.006 | 3.739 | 0.000 |
| as.factor(gender)Woman | -0.008 | 0.030 | -0.272 | 0.786 |

Table 3 shows that the p-value for the regression coeffient for work experience, technical skills, writing skills, leadership presence, and speaking skills are 0, which means that there is evidence that these factors are considered during the hiring process. Furthermore, the p-value for gpa, extracurriculars, and gender are greater that 0.05 which means that there is no evidence that these factors affect the phase 2 of the hiring process. In particular, since the p-value for the regression coeffient for gender is greater than 0.05 there is no evidence of gender bias in this part of the hiring process.

Phase 3 to Final Hires:

The fairness of the process in chosing the final hires was done by comparing the phase 3 applicants to the final hires. In particular, the average interviewer rating was calculated for each applicant. Then by comparing the average interview ratings of the applicants it was found that the 10 applicants with the highest interview rating were hired. Therefore, this selection is also fair as it is based off an objective criteria.

### Gender Biases of the AI in Scoring Applicant Skills and the Interviewers

Gender Biases in the AI:

An applicants technical skills, writing skills, leadership presence, and speaking skills were scored by AI using an unknown algorithm which means that there is a chance that this algorithm is biased. Box plots can be used to visualise whether or not the algorithm is biased against a specific gender for a specific skill.

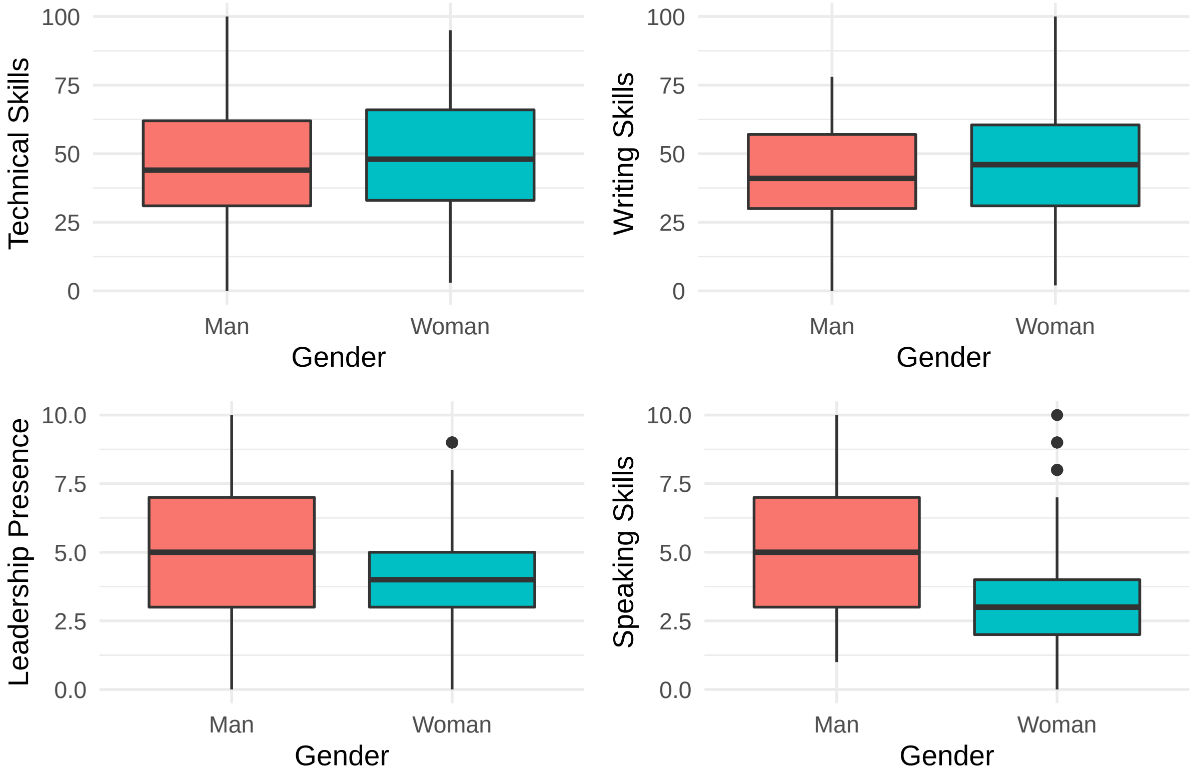

Figure 1: Boxplots of Skill Scores

From these boxplots, we can see that the average speaking skill women is way lower than the average speaking skill of mean. Furthermore, the average leadership presence of men appears to slightly greater than of women. However, the average writing skills of women appears to be greater than that of mean.

To get a more detailed picture, the average score of each skill based on gender was calculated.

**Table 4:** Average Ratings of Skills Based on Gender

| Gender | Avg Technical Skills | Avg Writing Skills | Avg Leadership Presence | Avg Speaking Skills |
|---|---|---|---|---|
| Man | 46.379 | 41.779 | 4.924 | 5.076 |
| Woman | 48.329 | 46.116 | 4.071 | 3.374 |

Table 4 shows that the average technical and writing skill of women were greater than men, but the average leadership presence and speaking skills of men were greater than of women. However, both technical skill and writing skill were assessed using a timed written format, so gender has no influence on the AI autograder. However, leadership presence and speaking skills were assessed using a pre-recorded video, which means that the gender could have an influence on the AI autograder. For example, the qualities of a man's voice and a women's voice are different which means that the algorithm could be favouring the qualities of a certain gender's voice as good leadership presence and speaking skills with leads to bias. Or if the algorithm is comparing the applicants videos to videos of current company employees that are in the higher roles which could results in bias if the ratio between males and females in the higher roles are biased.

To determine whether there is a bias between men and women, confidence intervals and p-values for the differences in means were calculated.

**Table 5:** 95% Confidence Intervals for Differences in Leadership and Speaking Skills Between Genders

| Skill | Male Avg | Female Avg | P-value | Lower Bound | Upper Bound |
|---|---|---|---|---|---|
| Leadership Presence | 4.924 | 4.071 | 0.002 | 0.324 | 1.382 |
| Speaking Skills | 5.076 | 3.374 | 0.000 | 1.189 | 2.214 |

Table 5 shows that the p-value for whether or not there is a difference between the average score for leadership presence is 0.002 and the confidence interval is (0.324, 1.382) which does not contain 0, which means that there is evidence of a difference between the average leadership presence of males and females. Furthermore, the p-value for whether or not there is a difference between the average score for speaking skills is 0.000 and the confidence interval is (1.1189, 2.214) which does not contain 0, which means that there is evidence of a difference between the average speaking skills of males and females. This means that the differences of the scores is not due to chance, and indicates a potential bias in the algorithm employed by the AI. Specifically, there is evidence that the AI is giving lower leadership presence and speaking skill scores to women than to mean.

Gender Biases of Interviewers:

**Table 6:** 95% Confidence Intervals for Differences in Interview Rating Between Genders

| Male Avg | Female Avg | P-Value | Lower Bound | Upper Bound |
|---:|---:|---:|---:|---:|
| 75.433 | 74.714 | 0.826 | -6.196 | 7.634 |

Table 6 shows that the p-value for whether or not there is a difference between average interview ratings for males and females is 0.826 and the confidence interval is (-6.196, 7.634) which contains 0, which means that there is no evidence of a difference between the average interview ratings of males and females. This means that there is no evidence that the interviewers have gender biases.

## Fairness of Black Saber Software Promotion and Salary Processes

To determine whether or not Black Saber Software's promotion and salary processes are free of gender biases, an employees gender, length of employment (in financial quarters), average productivity and leadership across their entire employment duration were compared to their role and salary of the latest financial quarter. In the raw data, productivity is given as a value from 1-100 where 50 indicates a satisfactory productivity for role. Leadership is given in 3 levels (needs improvement, appropriate for level and exceeds expectations), however it was changed to 0, 50, and 100 respectively. This is so an average leadership value can be assigned to each employee and it will take on the same meaning as average productivity, where 50 indicates a satisfactory leadership for the role.

To visualise the distribution of salary and the number of quarters an employee was employed based on their lastest role and their gender, boxplots were used.

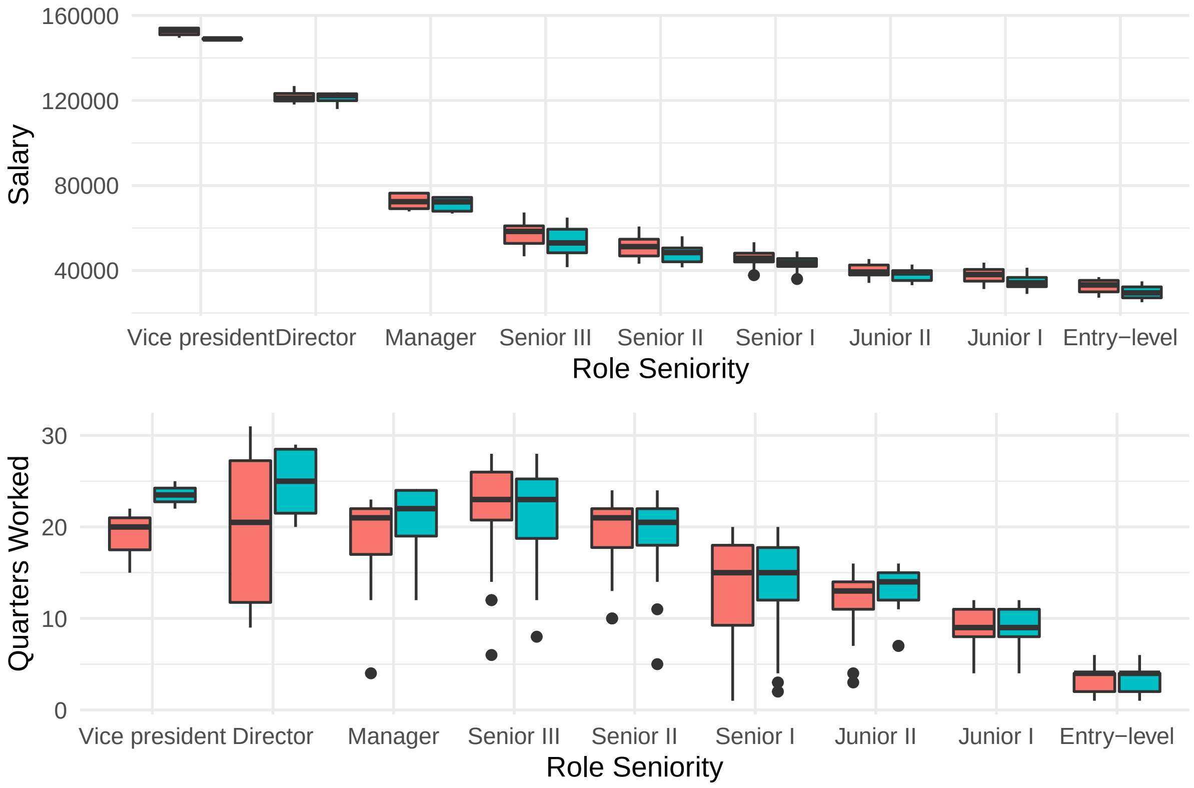

Figure 2: Boxplots of Salary and Quarters Worked

From figure 2, we can see that generally the salary of women is less than the salary of men in the same position which could indicate a gender bias in the salary process. Furthermore, we can see that women in the higher roles tend to have worked more quarters that men in the same role, which could indicate a gender bias in the promotion process.

To determine if the salary process is unfair, salary was modeled against gender, average productivity, average leadership, and quarters worked with role seniority as a random intercept effect because higher roles have a higher salary. After modeling, the confidence intervals for the regression coefficients were calculated.

**Table 7:** Confidence Intervals for Salary Model

| | 2.5 % | 97.5 % |
|---|---:|---:|
| sd_(Intercept)\|role_seniority | 26323.32 | 67846.24 |
| sigma | 3451.44 | 3865.84 |
| (Intercept) | 38254.83 | 95896.74 |
| genderWoman | -3336.94 | -1934.55 |

| | 2.5 % | 97.5 % |
|---|---|---|
| avg_prod | -33.43 | 22.12 |
| avg_lead | -41.90 | 84.57 |
| quarters_worked | -75.06 | 90.36 |

From table 7, we see that the confidence interval for average productivity, average leadership, and quarters worked contains 0, which means that these factors do not have an effect on the salary of an employee. This means that differences in salary is not based on an employee's talent or value to the company given the same role. Furthermore, the confidence interval for gender is (-3336.94, -1934.55) which does not contain 0, which means that there is evidence that gender has an effect on the salary of an employee. Women that have the same average productivity, average leadership and length of employment are expected to earn $2635.75 less than men with the same stats.

## Discussion

In conclusion, there is no evidence that the selection of which applicant moves on to the next phase of the hiring process is biased. The selection process from phase 1 to phase 2 is based of of objective criteria (ie. must submit both a cover letter and a cv, has GPA greater or equal to 2, extracurricular score greater or equal to 1, and if work experience score is 0 extracurricular score must be greater than 1). By using a logistic regression model and showing that the regression coefficient for gender has p-value greater than 0.05 there is no evidence that gender effects the selection process from phase 2 to phase 3. Lastly, since the selection process from phase 3 to final hires is based off of the highest average interview score which is an object critria, this part of the selection process is also unbaised.

Furthermore, there is no evidence that the interviewers of phase 3 of the hiring process have gender biases as the difference in means of interview rating depending on gender is given by a p-value greater than 0.05 and its confidence interval contains 0. However, there is evidence that the algorithm employed by the AI when scoring an applicants skills is biased against women as the difference in means of leadership presence and speaking skills depending on gender is given by a p-value less than 0.05 and its confidence interval does not contain 0.

Lastly, there is evidence of gender bias in the salary process, as women in the same position and length of employment with the same average productivity and leadership as a man is expected to earn $2635.75 less based on linear regression modeling. Furthermore, the could be potential bias in the promotion process, especially in the higher roles, as women in the same role as men generally were employed longer.

### Strengths and limitations

One strength of the report is that it is able to pinpoint where the gender bias is in the hiring process (i.e in the algorithm the AI uses to score applicants on leadership and speaking skills).

One main limitation is that due to the fact that ethnicity and race data was not collected, whether or not there are potential biases due to ethinicity/race in the hiring and renumeration process was unable to be investigated. As a result, this report will not be able to address any concerns regarding race and ethinicity which is one of the main talking points when discussing the fairness of hiring and renumeration processes.

Another limitation is that whether or not there is gender bias in the salary and promotion process was only investigated for the lastest financial quarter, so claims cannot be made for the previous financial quarters.

Lastly, this report only warns of there potentially being a gender bias in the promotion process but there is not solid evidence for or against this claim due to the fact that these differences are mainly seen in the higher roles where there are not a lot of data points to deduce an accurate statement.

## Consultant information

### Consultant profile

**Amy Chen**. Amy is a junior consultant with Emimi Co. She specializes in data visualization. Amy earned her Bachelor of Science, Major in Mathematics and Statistics, from the University of Toronto in 2021.

### Code of ethical conduct

Emimi Co strives to follow guidelines regarding ethical statistical consulting as laid out in the Statistical Society of Canada Code of Conduct and the Ethical Guidelines for Statistcal Praction from the American Statistical Society.

- Emimi Co makes sure to adhere to any relevant privacy laws or standards regarding the collection and storage of information and publication of results. Furthermore, Emimi Co does not disclose any information gathered during professional practice without permission, or as directed by law.
- Emimi Co aims to approach statistical consulting in an objective manner free of procedural and personal biases in order to create valid data-based information.
- Emimi Co strives to maintain full disclosure of all relevant assumptions and limitations to the data, analyses, and results, to prevent any misleading summary of data and results.