
Gender Parity in Black Saber Software Company

Gender Parity in Hiring, Salary and Promotion

Report prepared for Black Saber Software by Diamond
Hand

2021-04-21

Contents

Executive summary	3
Background and Aim	3
Key Findings	3
Technical report	5
Introduction	5
Method	5
Result	7
Discussion	14
Consultant information	15
Consultant profiles	15
Code of ethical conduct	15

Executive summary

Background and Aim

Although AI technology can bring convenience to the hiring process, it may be biased towards a certain population. Black Saber Software wants to ensure that their hiring pipeline only chooses their interview candidates based on their ability and value that they can provide to the company. Also, this rule is applied to the promotion and remuneration process within the company.

Key Findings

- In all phases of the AI recruitment pipeline, gender does not affect the result of the interview, as shown in figure 1.
- In the first phase of the interview, incomplete submission of a cover letter or cv would directly lead to a failure of passing the interview.
- In phase two, if an applicant achieved the average score of technical skill, writing skill, leadership presence and speaking skill, the probability of a male candidate being passed is 0.1%.
- We discovered the value for productivity would not affect a person's chance of promotion. The median productivity within the company is at about 50 which is a satisfactory level of effort. This means even if you work extremely hard in the Black Saber Software company, the chance of getting yourself promoted will not increase.
- If you are a male in the company, it will grant you a higher chance of promotion compared to female. We do not have evidence showing people prefer not to say their gender has a higher or lower promotion chance due to the small sample size.
- The median of women's salary is decreasing annually; it dropped from about \$60,000 in 2013 Q4 to below \$40,000 in 2020 Q4.
- Although role seniority directly influences the amount of salary of an employee, gender also played a significant role in it.

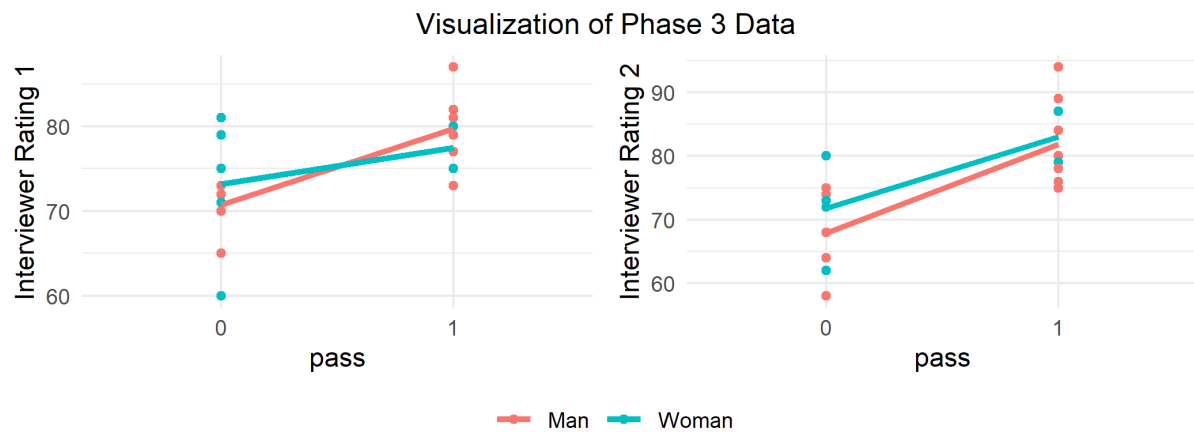


Figure 1. Black Saber Phase 3 Interview Rating and Results.
 Dots on 0 and 1 vertical axis represent not pass and pass the phase.
 Red and blue lines connect mean passing and not passing interviewer ratings.
 Male applicants have higher rating 1 in passing the interview on average.
 Female applicants have higher rating 2 in passing the interview on average.

Created by: Diamond Hand for Black Saber Software

Source: phase3-new-grad-applicants-2020

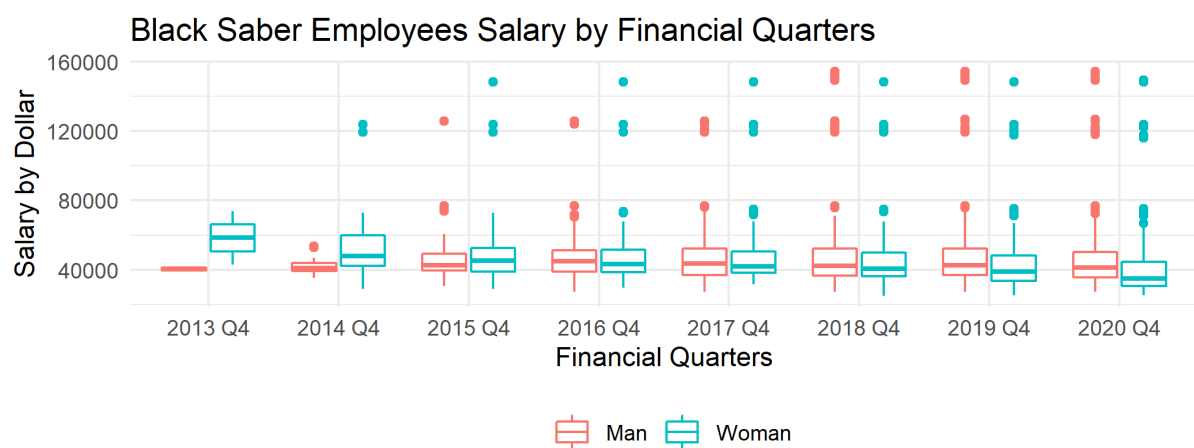


Figure 2. Distribution of the salary across every fourth financial quarter from 2013 to 2020. The horizontal line in the box is the median also known as the middle number which separates the top 50% and bottom 50%. The bottom line of the box is the first quartile, which is located at the 25 percent of the data from lowest to highest. The top line of the box is the third quartile, which denotes the value at the 75% of the data from lowest to highest.

Created by: Diamond Hand for Black Saber Software

Source: Black Saber current employees

Technical report

Introduction

With the advancement of technology, many companies opt to use AI technology to filter and select interview candidates for them. Although this process saves plenty of time and human resources, it brought to attention some of the machine learning agents can have bias in their performance. Specifically, discrimination against women in workplaces reported in the press brings worry that this issue is potentially being replicated in the AI methods. Besides, there is also increasing attention to whether women are treated equal as men in aspects such as promotion and salary. Our client Black Saber Software is also concerned with this issue as they only value an individual's ability and contribution to the company, and they want to insure that this principle is strictly followed in their hiring, promotion and salary process. They provided us the hiring data for the new grad in which the selection pipeline is AI-automated up to the final interview and the data of their current employees from the second financial quarter in 2013 to the last financial quarter in 2020. We have information about the applicants' id, gender, and evaluation at their job relevant skills such as technical skills and speaking skills. In the promotion and salary data, we have information about the employees' id, gender, team, role, and their leadership rating, productivity and salary for each financial quarter. With these data, we will look into fairness only through gender; whether hiring, promotion and salary are only determined by their ratings and productivity.

Research questions

- Is gender a determining factor in the AI recruitment pipeline?
- What are the variables that would affect the number of promotion times?
- Is gender a significant factor that determines employee remuneration?

Method

Data Wrangling and Models Selection

The hiring data provided by Black Saber consists of four datasets. There are three phases in the hiring process, each narrowing down the field of applicants. Each applicant is assigned a unique ID, this ID is consistent across phases, and for applicants successfully passed the first phase of interview, their corresponding ID will be recorded in the second phase dataset, the same algorithm applied to the second and third phase of interview. The first change we did to the datasets is to create a new variable 'pass' in every phase of datasets, this is a binary variable

that explains whether a candidate passes the interview. For example, if applicant's ID is in the final phase datasets, then in the third phase data, a value of 1 would show under variable "pass", 0 otherwise. The data on the first phase of interview contains information on applicants' ID, applied team, submission of cover letter, submission of cv, gpa, gender, relevance or skills building of work experience and extracurricular to the team they applied for. After some data exploring, we find that applicants who do not submit cover letter or curriculum vitae are not passing phase 1 interview; applicants' scored 0 in their extracurricular session are all failed to pass the first round of interview as well. We removed observations with any one of the conditions: not submitting cover letter; not submitting curriculum vitae; extracurricular autorated score being 0, so the model is more effective on predicting the effect of other explanatory variables. Moreover, variable "work_experience" contains three levels, we collapse two of the levels into one level so it is now a binary variable. We collapse internship experience with score 1 and 2 into 1 as observations with a score of 2 are sparse. Phase 3 dataset only has information on applicants' ID, and two interviewer ratings. To apply a regression model investigating the effect of gender, we need information on gender, so we merge phase 3 dataset with phase 2 dataset by applicants' ID. For every phase of the interview, we fitted a generalized linear model as our outcome of interest is "pass", a binary variable; we do not have repeated measures in our data, so cases are independent. In phase 1, we fitted a model with predictive variables gpa, gender and work experience and examined their effect on passing an interview. In phase 2, we measured the effects of gender along with technical skills, writing skills, speaking skills, and leadership presence in interview results. In phase 3, we fitted two generalized linear models, one with two interviewer ratings and gender, the other model did not use gender as an explanatory variable. Then we applied a likelihood ratio test to see if adding gender in our model helps explain the data. Lastly, for quantitative variables like the four skills, gpa, and interview rating, we centered these covariates in the models so the baseline intercepts in the output make more sense.

For the promotion research question, we made some minor changes to our original dataset. For the original data, it only contains 8 different columns that correspond to 8 different categories that we might want to consider in our research question. The goal for this research question is to look for fairness in promotion. We added 2 more columns for our data manipulation, where one is called promotion times and the other one is the work duration. Promotion time counts the number of promotions each person has received in total. Work duration measures how long each person worked in this company. We multiply each quarter time to 0.25 to get the time duration in double. Lastly, we changed the role_seniority variable from different ranks to numbers, where 1 is the lowest rank that corresponds to Entry-level and 9 corresponds to vice president. The reason why we convert this variable into integer is because we are also interested in the relationship of whether the ranks would affect the number of promotions or not. The model we decided to use is the general linear mixed model. We have repeated measurement for the same person, and

this is why we choose a mixed model. In order to measure the promotion, we want to look for a duration of time which we think a poisson distribution would be appropriate. These are the main reasons why we decided to choose this model.

In order to analyze salary, we simply removed the dollar sign in that column. Therefore, we can treat it as a numerical variable. To examine the fairness in remuneration between male and female employees, we first plotted a series of boxplot to compare the median of salary in different dimensions. When dealing with data like salary, it is essential to use medians to do comparison because the median cannot be affected by outliers such as the vice presidents who earn much higher than regular workers. By plotting the data first, we observed that the imbalance occurs in the latest financial quarter, and each team is composed of a different number of males and females. Hence, we decided to filter the data to only look into the fourth financial quarter in 2020. Also, it is best to treat teams as a random effect to explain the variance because each team would have a different salary standard. Moreover, we chose a linear mixed model to fit the data because we have a continuous variable as the response and a random effect in the predictors.

Result

Hiring:

From fitting three generalized linear models, the effect of gender in all three phases of interview is statistically non-significant.

Table 1: MLEs of baseline odds and odds ratios, with 95% confidence interval and p-value in phase 3

	est	2.5 %	97.5 %	p-value
Baseline	2.658	0.897	4.782	0.007
gpac	7.363	4.822	10.780	0.000
genderPrefer not to say	-0.533	-4.512	3.945	0.829
genderWoman	0.392	-0.881	1.743	0.552
as.factor(work_experience)1	5.331	3.594	7.670	0.000

From the table above we see the confidence interval of gender contains 1, which suggests that gender is not statistically significant. The effects of the predictor variables women and prefer not to say have p-values of 0.55 and 0.83. The odds of an applicant being passed when GPA is 3.0

and is a male with working experience not relevant to the job at all is 14.267, which corresponds to a probability of 93.4%. In addition, for each unit increase in GPA, the odds of passing phase 1 increase in percent by 157607.8; for one unit increase working experience score, the odds of passing phase 1 increase by 20561.32 in percent.

Table 2: MLEs of baseline odds and odds ratios, with 95% confidence interval and p-value in phase 2

	est	2.5 %	97.5 %	p-value
Baseline	0.001	0.000	0.009	0.000
genderWoman	0.567	0.128	2.276	0.433
technical_skillsc	1.084	1.047	1.135	0.000
writing_skillsc	1.097	1.051	1.156	0.000
leadership_presencec	2.450	1.724	3.896	0.000
speaking_skillsc	2.045	1.515	2.954	0.000

In phase 2, as summarized in the table, gender is not statistically significant in deciding if a candidate passes the interview as its 95% confidence interval contains 1, and it has a p-value of 0.433, greater than the threshold. The odds of a male candidate being passed when the score of technical skill is 44, writing skill is 46, leadership presence and speaking skill is 4 is 0.001, which corresponds to a probability of 0.1%. In addition, for each score increase in technical skill, the odds of passing phase 2 increase in percent by 8.44; for one score increase in writing skill, the odds of passing phase 2 increase in percent by 9.661; for one score increase in leadership, the odds of passing phase 2 increase in percent by 144.962; for one score increase in speaking skill, the odds of passing phase 2 increase by 104.533 in percent.

Table 3: Likelihood ratio test between model with and without gender

#Df	LogLik	Df	Chisq	Pr(>Chisq)
3	0	NA	NA	NA
4	0	1	0	0.9999591

In the last phase, gender does not show a significant effect on the decision of choosing applicants.

Likelihood ratio test is performed on two generalized linear models, one with gender as an explanatory variable and one does not include gender as an explanatory variable. A p-value of 1 in table indicates that adding gender as an explanatory variable does not help explaining the data, and thus there is no evidence that gender is significantly associated with the AI recruitment system.

Promotion:

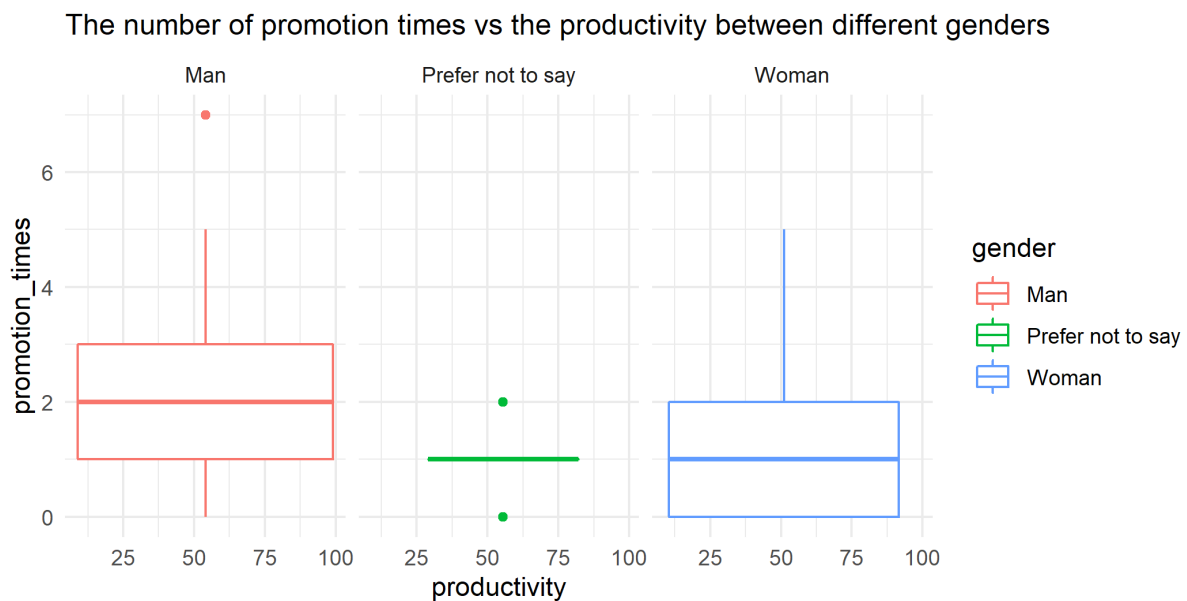


Figure 3.
Created by: Diamond Hand for Black Saber Software
Source: Black Saber current employees

In Figure 3 where our histogram indicates the number of promotions between different genders. The first observation we can make is that the averaged productivity for all genders is at 50 which meets the satisfactory bar. However, men have a slightly higher average overall which means they often exceed the expectation. The second observation from the graph is where men have a median of 2 for promotion times that is 1 higher compared to female and unknown gender. Do these findings show that a higher productivity value indicates a higher chance of promotion? To clarify our observations, we decided to use a general linear mixed model for further exploration.

Table 4: Promotion Model Result

	x
(Intercept)	-1.754
productivity	0.000
genderPrefer not to say	0.215
genderWoman	-0.568
leadership_for_levelExceeds expectations	0.021
leadership_for_levelNeeds improvement	0.002
role_seniority	0.015

From the summary table of our general linear mixed model, we can see that genderWomen is significant in the fixed effects where it has a large negative number -0.569. This means women tend to have less number of promotions in the company which they are being treated unfairly. The other variables such as productivity, gender prefer not to say, level exceeds expectations or need improvement and role seniority do not have a significant effect in our model. Since we added the work duration in our model that records how long each person worked in the company, the cases for people who just joined that have a high productivity value is also less likely to get promoted. Therefore, we discovered that there is promotion bias based on gender in our model.

Salary:

Figure 2 shows the company started with women having the highest salary during 2013 and 2014; it later achieved a balance in 2015. However, the trend gradually shifts to the situation where female employees have an overall low income compared with men. Until the fourth financial quarter of 2020, the median salary of female employees is as low as the first quartile salary of male workers, suggesting that women in Black Saber Software generally receive lower income than men.

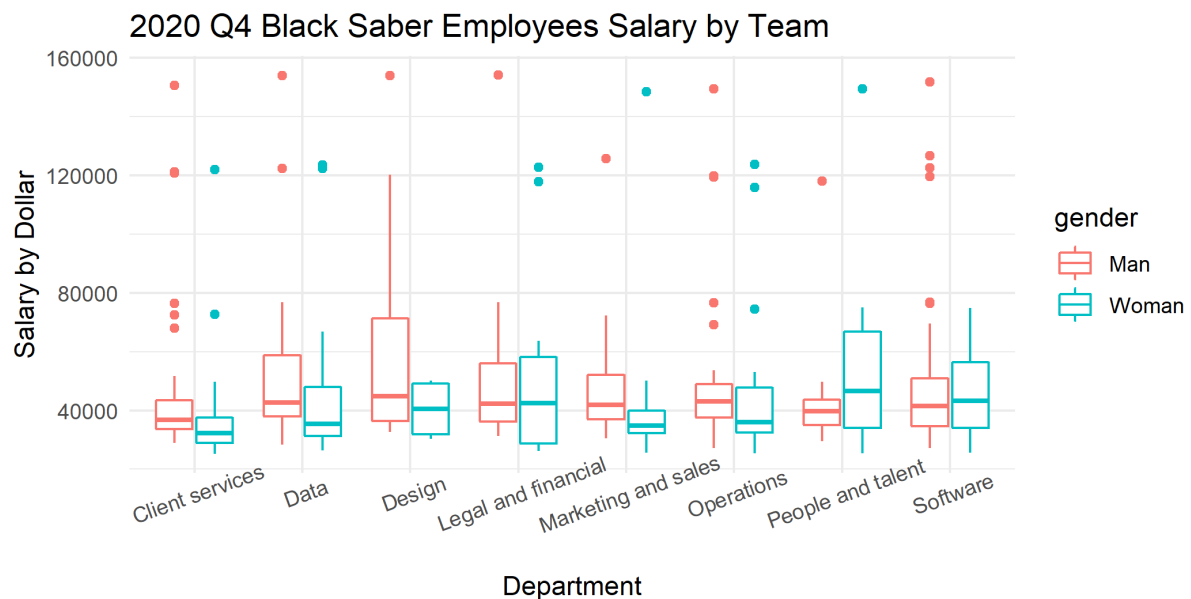


Figure 4.
Created by: Diamond Hand for Black Saber Software
Source: Black Saber current employees

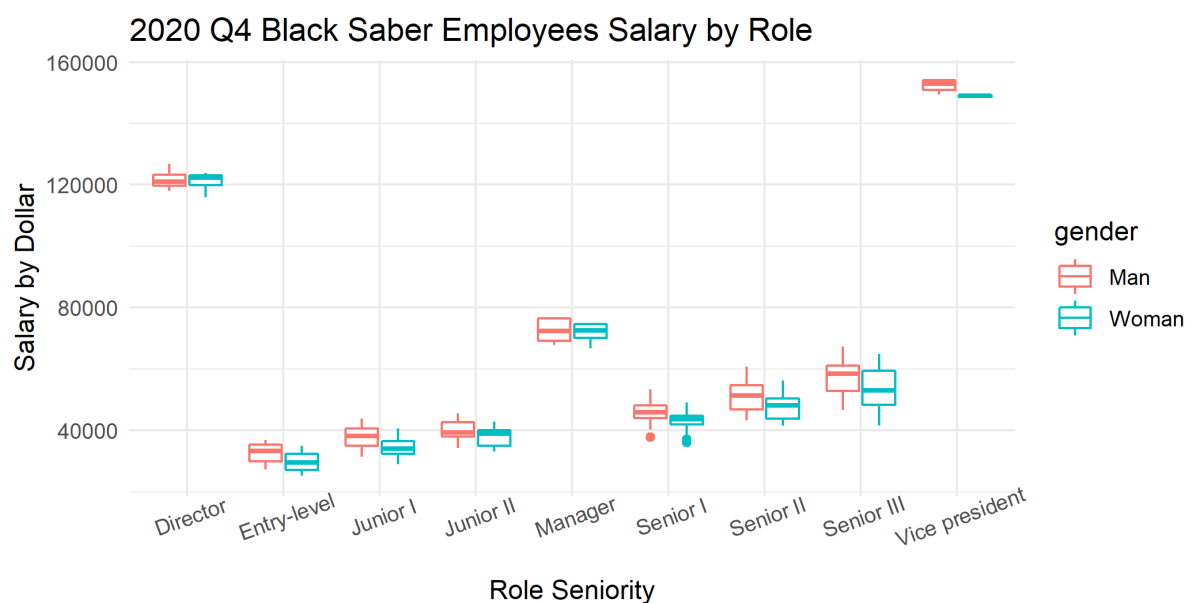


Figure 5.
Created by: Diamond Hand for Black Saber Software
Source: Black Saber current employees

Figure 4 and 5 illustrates a more detailed version of the salary distribution by separating workers based on their team and role. Both charts display women's salary median are on the lower side. Only some rare instances such as female employees in people and talent have a higher median than man, and that all directors and managers are paid equally.

Table 5: Salary Model Result

	x
(Intercept)	122571.236
genderWoman	-2164.541
role_seniorityEntry-level	-89961.779
role_seniorityJunior I	-85122.318
role_seniorityJunior II	-82353.746
role_seniorityManager	-49248.907
role_senioritySenior I	-76902.759
role_senioritySenior II	-71592.252
role_senioritySenior III	-65733.697
role_seniorityVice president	29917.566
productivity	-7.531
leadership_for_levelExceeds expectations	173.248
leadership_for_levelNeeds improvement	-1356.226

Table 6: Model Comparison

#Df	LogLik	Df	Chisq	Pr(>Chisq)
12	-5630.543	NA	NA	NA
15	-5610.608	3	39.86942	1.14e-08

Table 7: 95% CI for Salary Model

	2.5 %	97.5 %
.sig01	902.851	2788.632
.sigma	3160.096	3542.640
(Intercept)	120479.573	124657.667

	2.5 %	97.5 %
genderWoman	-2761.519	-1574.776
role_seniorityEntry-level	-91561.039	-88360.837
role_seniorityJunior I	-86748.051	-83495.759
role_seniorityJunior II	-84061.555	-80645.675
role_seniorityManager	-51230.866	-47280.748
role_senioritySenior I	-78605.210	-75199.062
role_senioritySenior II	-73376.572	-69812.108
role_senioritySenior III	-67483.227	-63985.593
role_seniorityVice president	27137.645	32687.859
productivity	-27.153	11.960
leadership_for_levelExceeds expectations	-2007.644	2367.382
leadership_for_levelNeeds improvement	-3438.119	736.639

Furthermore, when comparing models with or without gender, the likelihood ratio test produced a p-value close to zero, which suggests including gender as a fixed effect in the linear mixed model does explain the data better. The coefficient of gender indicates that when role seniority, leadership rating and productivity is the same, a female employee receive 2164.54 less than a male employee who is on the same team. Also, the significance of this fixed effect is confirmed in the confidence interval as well. The interval (-2761.52, -1574.78) does not capture zero which suggests gender is needed in the model.

Table 8: Likelihood Ratio Test

#Df	LogLik	Df	Chisq	Pr(>Chisq)
12	-5630.543	NA	NA	NA
15	-5610.608	3	39.86942	1.14e-08

Last but not least, although gender has an impact on one's salary, likelihood ratio test also suggests that productivity and leadership rating is best to be included in the model. However, their confidence intervals include zero; this suggests that they are not as statistically significant

as gender to our model.

Discussion

There is no significant evidence showing that there is a gender bias in hiring, meaning that for both men and women they are being treated equally. Since most of the applicants are being filtered by the AI and gender is not something that the bot values, people are more likely to get hired based on their personal score. However, it is not the case for promotion where men are more likely to get promoted compared to women. We also compared variables such as productivity, leadership for level and role seniority, where none of them shows a significant result that could affect the promotion rate. This means even if you exceed the expectation for productivity, the chance of getting promoted won't increase. higher seniority also doesn't get a privilege of getting more promotions. Likewise, the result also shows that gender is deterministic to an employee's salary in the most recent financial quarter. A female employee receives around 2k less compared to a male employee that is on the same team. This concludes the unfairness between genders in promotion and salary.

Strengths and limitations

The limitations we have involves the number of data we have in later phases for hiring. Most of the applicants are being filtered out by the AI system which the result could be less convincing when the sample size decreases. In both promotion and salary sections, the sample size for gender "prefer not to say" is too small which we cannot determine whether there is a gender bias towards these people or not. We also don't know how the values for productivity count since simple tasks can be achieved easily which could possibly gain a higher rating compared to challenging tasks. Therefore, the results in how productivity would affect promotion and salary is not as reliable as it is. As Hiring data is not enough in later phases of the interview, in further analysis, we would suggest collecting data from multiple interviews to support our results.

Consultant information

Consultant profiles

Deen Xu. Deen is a senior consultant with Diamond Hand. He specializes in reproducible analysis and statistical communication. Deen earned his Bachelor of Science, specialist in Statistics Methods and Practice from the University of Toronto in 2022.

Mingyang Li. Mingyang is a senior consultant with Diamond Hand. He specializes in data visualization. Mingyang earned his Bachelor of Science, majoring in Statistics and Cognitive Science, from the University of Toronto in 2022.

Yinghan Zhang. Yinghan is a senior consultant with Diamond Hand. She specializes in data visualization. Yinghan earned her Bachelor of Science, Specialist in Statistics Methods and Practice, from the University of Toronto in 2022

Zecheng Wu. Zecheng Wu is a senior consultant with Diamond Hand. He specializes in data manipulation. Zecheng earned his Bachelor of Science, Majoring in Statistics and Mathematics, from the University of Toronto in 2022.

Code of ethical conduct

- Our company is providing relevant information to our audience without any bias and misleading information.
- The statistical methods that are used in this project have undergone multiple discussions and considerations which we are providing the most reliable and appropriate models that we can do.
- All the results are relevant to the research questions that our client asked for. Limitations are considered wisely to prevent any misleading conclusion.
- We did not assume nor add any fake data to make our result look better.
- We did not come up with a predetermined outcome before we did our data manipulation.
- It is our responsibility to protect the privacy of our client's company information.