Potential Gender Bias in Black Saber Software

Evidence that in 2020 male employees are more likely to get promotions and more salaries than women even if the hiring process is unbiased

Report Prepared for Black Saber Software by Data Protect Corporation

2021-04-21

Contents

Executive summary

Background & Aim:

In early April, our company Data Protect received a request letter from the chief people officer of Black Saber Software about checking if their hiring and remuneration processes are all fair. He mentioned that several companies in their area have recently received negative comments from their employees because these companies seem to have a potential gender bias in the processes of recruiting. The chief people officer sent us the hiring data for their new grad program and the dataset including the promotion and salary of their entire staff. It should be noted that Black Saber Software is using an AI-automated selection pipeline for the first two phases of their hiring process.

We decide to solve Black Saber Software's request by answering the following three questions using the given two datasets: Whether gender is a key factor that would influence AI's hiring decisions? Would female employees have a lower chance to get promoted than males? Would female employees have a lower salary compared to males?

After fitting the data to several statistical models, we conclude that their hiring processes are all fair, but there exists gender inequality in employees' salary and promotion.

Key findings:

- AI's decisions do not have any bias on gender when evaluating these applicants based on gender and applicants' talents and values at phase 1 and phase 2.
- There is no evidence showing that interviewer ratings and gender have effects on final hiring decisions made by humans.
- Gender distributed unequally in each seniority level as shown in figure 2.
- A woman employee gets 0.748 promotions of a man of the same team, average productivity and salary raise.
- As shown in figure 3, there are 65% male and 34% female in the employee group with salary higher than the average $44,945.3, but 52.1% male and 46.3% female in the employee group with salary lower than the average.
- As shown in figure 4, there are 52.1% male and 46.3% female in the employee group with salary lower than the average.
- For a male and a female to have the same productivity, quality of demonstrated leadership, role seniority, and in the same team, male's salary will be $2,173.654 higher than female.

Limitations:

- We do not have the most recent data for the first quarter of 2021, so the analysis result does not accurately reflect the company's current state.
- Approximately 2% of employees involved in this study prefer not to tell their gender, so that 2% of data does not contribute to our final conclusion.
- The data collection is limited, and it would be better if we can collect more information from employees to build the prediction models.

Key graphs:

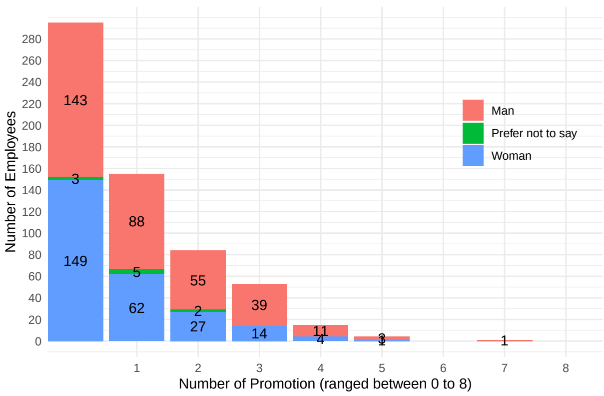


Figure1: Number of promotion distribution by gender.

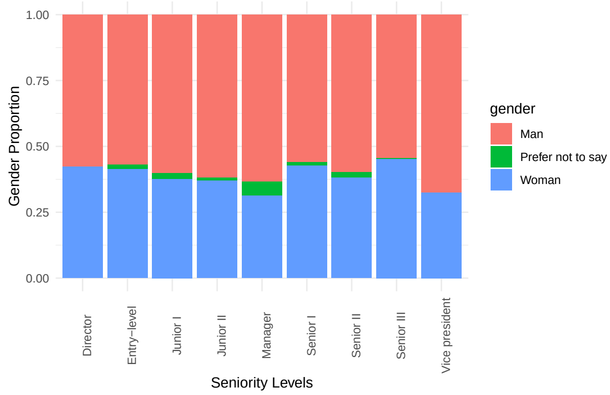
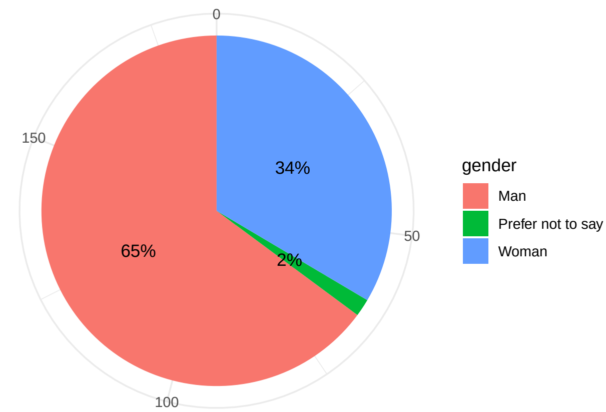


Figure2: Gender proportion in each seniority Level.



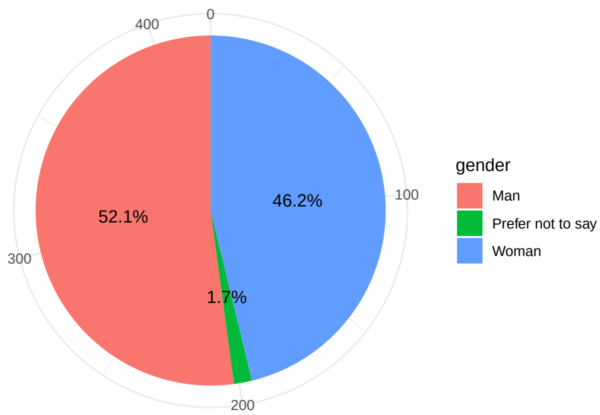


Figure3: Gender proportion for high salary employees.

Figure4: gender proportion for low salary employees.

Technical report

Introduction

There is a long history of gender inequality. Even though the prejudice issue had been noticed by recruiters, the inequality still exists in the workplace. To eliminate the injustice of sexuality, many corporations had applied AI programs to supervise the hiring process. Since the algorithms written for AIs are invisible, algorithmic bias can still be created from proxy variables. According to *Women in the Workplace 2016*, women hardly reach the top jobs because they have less probability to be promoted to middle positions relatively to men(Spiggle, 2019). Thus, there are less women candidates in the pool of top jobs. Although the gender pay gap has been reduced a lot for these years, the pay gap remains. Women and men do different jobs, and jobs done by women are often undervalued. A large portion of high-paid jobs are held by men. Less than 20% of women's upstream jobs are in industrial and goods sectors (Agriculture, Utilities, Construction, Manufacturing, and Mining), compared to 45% for men (Ward, 2018). For example, most engineers are male, and most of the service industry is done by women. Even if women and men are in the same jobs, a higher percentage of women tend to be paid lower. One of the reasons why there is such a big gender wage gap is because of motherhood. After women married and had the birth of their child, they could not concentrate on work, and they had more pressure and workload from their family.

In early April, our company received a request from Black Saber Software, and we aim to find out whether there is potential bias in the company's hiring and remuneration processes. Specifically, we are interested in three questions: Whether gender is a determinant factor that would affect AI's decision when hiring? Would the number of promotions employees at Black Saber Software received depend on their gender? Would female employees receive lower salary compared to males at Black Saber Software?

Research questions

- *Whether AI's decisions have a bias on gender during the hiring process?*
- *Whether there is a potential sexual prejudice issue in the promotion process?*
- *Whether there exists salary inequality between male and female exmployees?*

Analysis of potential gender bias in the hiring process

Data wrangling and description

The research question focuses on whether AI's decisions have bias on gender. We could also explore what kind of values and talents of applicants have effects on AI's decisions in hiring employees. For Black Saber Software, the first two hiring phases are evaluated by AI and the last phase is evaluated by interviewers. All phases are used to decide if the company will hire the applicants.

In the first phase, 613 applicants provide their own information about the team applied for, cover letter, CV, GPA, gender, extracurriculars and internship experience, and they are evaluated based on these factors. After phase1, some applicants are rejected, and the remaining 300 applicants are accepted into phase2. In the second phase, applicants need to submit a pre-recorded video, a timed technical coding task and a writing sample for being evaluated by AI. Based on those files, AI would give scores of technical skills, writing skills, leadership presence and speaking skills and evaluate applicants. After phase2, some applicants are rejected, and the remaining 22 applicants are accepted into phase3. In the third phase, applicants are evaluated based on the first interviewer rating and the second interviewer rating. After phase3, the remaining 10 applicants are finally hired. Therefore, we obtained four data sets, including phase1, phase2, phase3 and final hire. In the phase3 dataset, there are only two ratings from the interviewers. The final hire data set only contains applicant id.

Then we do data processing. To get the gender of applicants and other information about applicants in phase3 and final hiring results, we integrate the four data sets together into one data set, which further helps extract information and draw plots.

Then we add a new variable called "progress". Specifically, if an applicant is in phase1, but he/she is rejected in phase2, then we define "progress" of this applicant as "rejected at phase1". If an applicant is in phase2, but is rejected at the phase3, we define the "progress" of this applicant as "rejected at phase2". If an applicant is in phase3, but is rejected in the final hiring, we define the progress of this applicant as "rejected at phase3". Then if an applicant is accepted in the final hiring, we define "progress" of this applicant as "hired". Next, we create a binary variable called "result", which only contains two results, "yes" and "no". Specifically, "yes" means an applicant is accepted into the next phase or he/she is hired. "No" means an applicant is rejected at one phase. The new variable "result" would be convenient for us to fit a logistic model. Besides, when fitting models at phase1 and phase2, we filter the data of men and women to analyze, without the gender of preferring not to say.

How score at phase2 rated by AI is distributed by gender should be focused on, which is significant to see whether AI's decisions have a bias on gender based on AI's rating score. The following graphs can help to illustrate our findings.

Tables and graphs

From figure 1.1, we could observe that nearly the same number of men and women are rejected at phase1 and phase2. It seems that AI provides nearly the same number of opportunities for men and women. Thus, in this case, bias is not obviously shown.

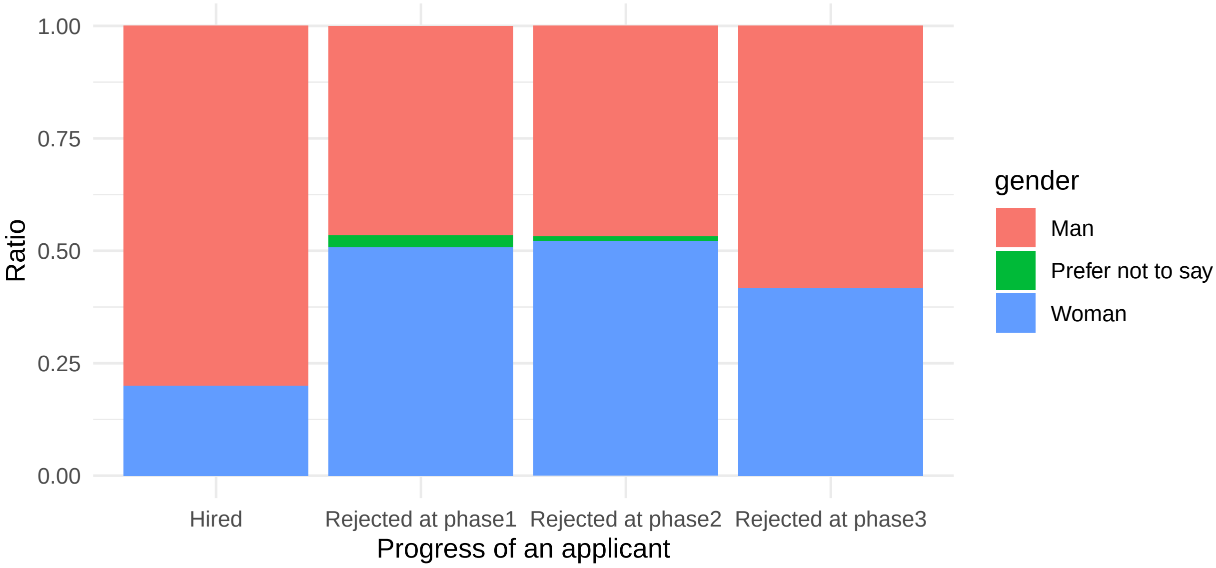


Figure1.1: Gender distirbution based on applicants' progress.

From figure 1.2, we can see the number of men with high scores of speaking skills is larger than the number of women, which seems more beneficial to men, so there exists a rating bias for speaking skills scores between men and women.

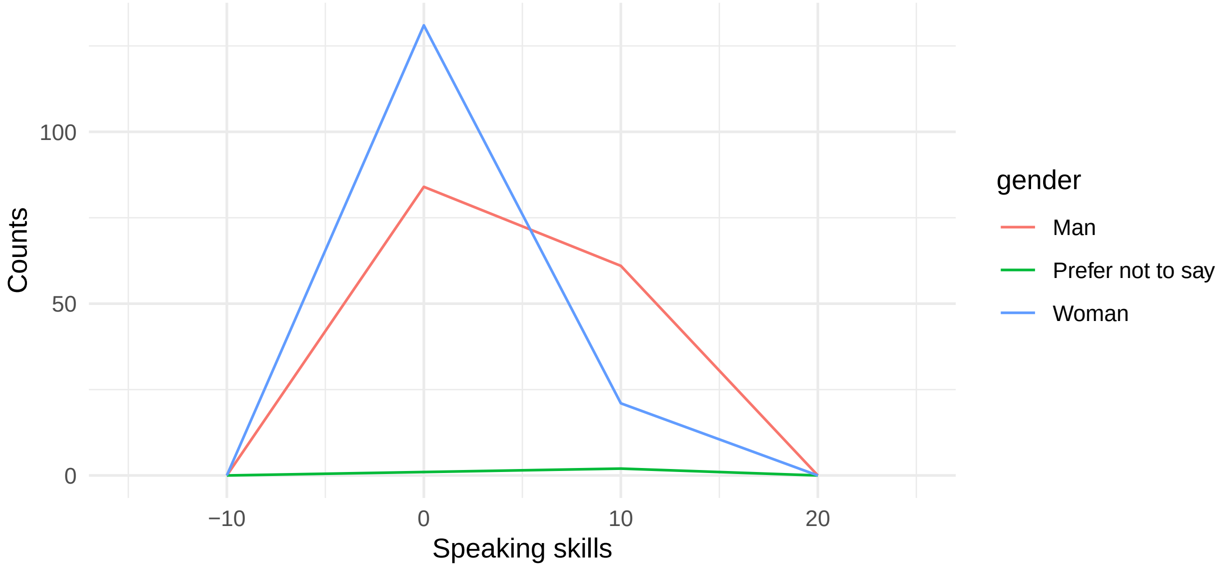


Figure1.2: Speaking skills on gender.

From figure 1.3, we could observe that the number of women with high scores of writing skills is larger than the number of men, which seems more beneficial to women, so there might exist a rating bias for writing skills scores between men and women.

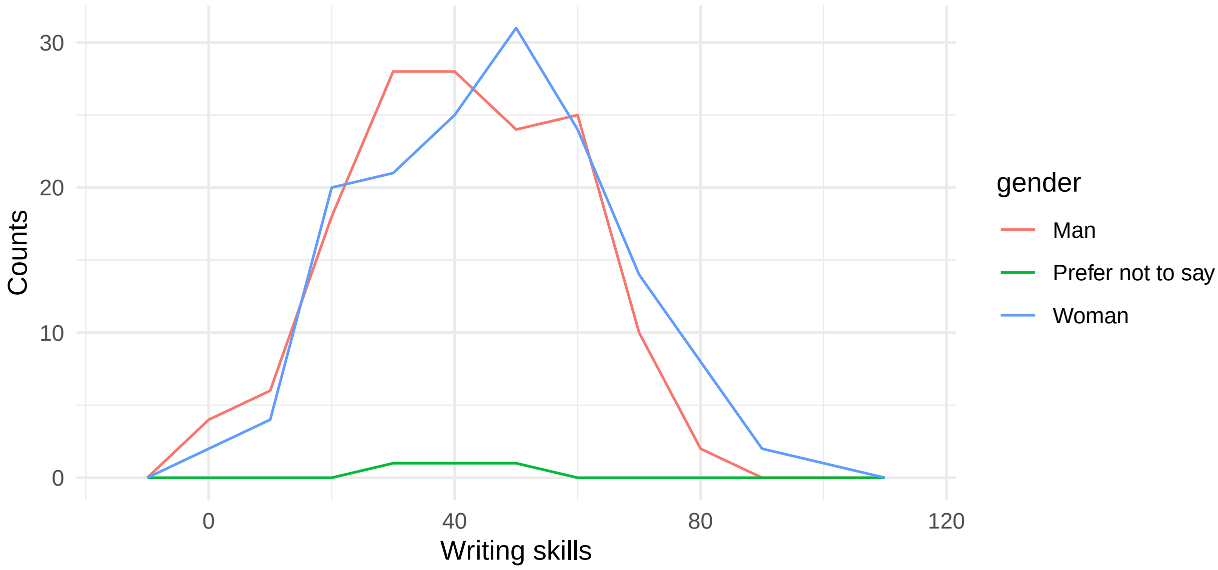


Figure1.3: Writing skills on gender

From figure 1.4, we could see that the ratio of men with high scores of writing skills is larger than the ratio of women, which seems more beneficial to men, so there might exist a rating bias for leadership presence scores between men and women.

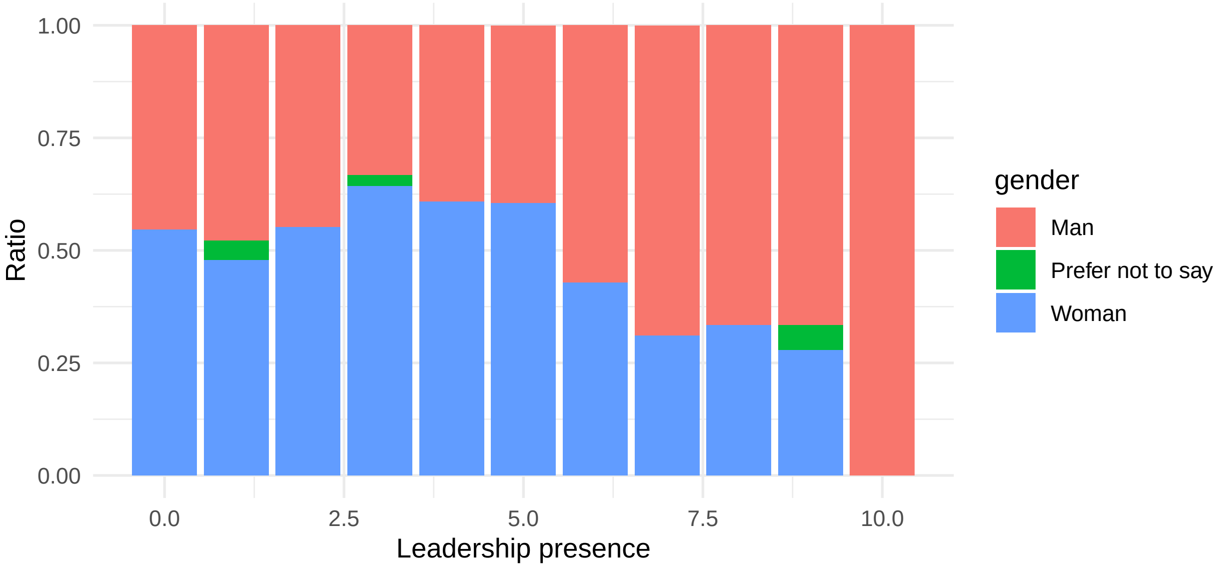


Figure1.4: Gender distribution based on leadership presence

Model

We focus on the first two phases. Generalized linear models (GLM) are implemented to investigate whether there exists bias of AI decisions on gender. Before fitting GLMs in the following phases, we would check the assumptions of GLM first. The data for GLM are from each applicants' unique progress and each case is independent, so they are independent. The response variable for all three phases is called "result", and it is created in the data wrangling step, which is binary. The response variable is from a binomial family. Besides, we assume a linear relationship between the transformed response and the explanatory variables.

Firstly, we need to choose a model to fit in the first phase. The response variable is result and we choose gpa and gender as explanatory variables to fit the model. Since we want to find the best fitted model, we constantly add variables, such as extracurriculars, work experience, cover letter, cv and team. After that, we could get six models. The Akaike Information Criterion(AIC) evaluates how well a model fits the data(Bevans, 2020). Then we compare these six models with AIC to find the best-fit model that explains the greatest amount of variation using the fewest variables. Next, we could get a best generalized linear fit, whose response variable is result and explanatory variables are gender, gpa, extracurriculars, work experience, cover letter and cv. The equation for this model is $E(Result)= e^{\beta_0+\beta_1 gender+\beta_2 gpa+\beta_3 extracurriculars+\beta_4 workexperience+\beta_5 coverletter+\beta_6 cv+\epsilon}$, where ϵ is the residual error.

The summary for phase1 model:

Table 1.1: coefficients, p-values and confidence interval for phase1 model

	Estimate	95% confidence interval	p-value
(Intercept)	-157.70	(-2620.81,-4289.72)	0.98
gender: Women	0.90	(-0.92,3.01)	0.35
gpa	12.61	(7.32,20.64)	0.00
extracurriculars	9.95	(6.02,16.04)	0.00
work experience	11.56	(7.09,18.33)	0.00
cover_letter	-49.79	(-126.79,1870.54)	0.99
cv	61.30	(862.37,1228.94)	0.98

According to the summary of the best fitted model, the p-value for gender of women is greater than 0.05, so we can conclude that there is no evidence of showing that AI has prejudice against gender. P-values of extracurriculars, gpa and work experience are less than 0.05, so there is enough evidence to show extracurriculars, gpa and work experience have significant effects on AI's decisions. By calculating the confidence interval, 95% CI for the gender of women is (-0.92,3.01), which includes zero. This means the gender of women is not significant in AI making decisions. 95% CI for extracurriculars is (6.02,16.04), and 95% CI for gpa is (7.32,20.64) and 95% CI for work experience is (7.09,18.33), which do not include zero. This means extracurriculars, gpa and work experience are significant in AI making decisions. The conclusion obtained from the confidence interval fit with the conclusion we got from p-values.

In phase2, we start to build the model whose response variable is result and explanatory variables are gpa and gender. Since we want to find the best fitted model, we constantly add variables extracurriculars, work experience, technical skills, leadership presence, speaking skills and team into this model. Then we get seven models. After we compare these seven models with AIC, we can get a best generalized linear fit, whose response variable is result and explanatory variables are gender, gpa, extracurriculars, work experience, technical skills, leadership presence, speaking skills.

The equation for this model is: $E(Result)=$ $e^{\beta_0+\beta_1 gpa+\beta_2 gender+\beta_3 extracurriculars+\beta_4 workexperience+\beta_5 technicalskills+\beta_6 leadershippresence+\beta_7 speakingskills+\epsilon}$, where ϵ is the residual error.

The summary for phase1 mode2:

Table 1.2: coefficients, p-values and confidence interval for phase2 model

	Estimate	95% confidence interval	p-value
(Intercept)	-12.80	(-18.02,-8.52)	0.00
gpa	0.16	(-1.04,1.39)	0.79
gender: Women	0.21	(-1.09,1.53)	0.75
extracurriculars	-0.15	(-1.41,-1.09)	0.81
work experience	-0.06	(-1.36,1.17)	0.93
technical skills	0.06	(0.03,0.10)	0.00
leadership presence	0.67	(0.39,1.01)	0.00
speaking skills	0.51	(0.26,0.80)	0.00

According to the summary of this model, the p-value for gender is greater than 0.05, which means that there is no evidence of showing that AI has prejudice against gender. P-values of technical skills, eadership presence and speaking skills are less than 0.05, so there is enough evidence to show technical skills leadership presence and speaking skills have significant effects on AI's decisions. By calculating the confidence interval, 95% CI for the gender of women is (-1.09,1.53), which includes zero. This means the gender of women is not significant in AI making decisions. 95% CI for technical skills is (0.03, 0.10), and 95% CI for leadership presence is (0.39,1.01) and 95% CI for speaking skills is (0.26,0.80), which do not include zero. This means technical skills, leadership presence and speaking skills are significant in AI making decisions. The results obtained from the confidence interval correspond to the results we got from p-values.

Phase3 is the phase where rating scores were made by interviewers. We fit a GLM with response variable "result" and explanatory variables: the first interviewer rating, the second interviewer rating and gender.

The equation for phase3 model is: $E(Result) = e^{\beta_0+\beta_1 gender+\beta_2 interviewerrating1+\beta_3 interviewerrating2+\epsilon}$, where ϵ is the residual error.

`Table 1.3: summaries of coefficients and p-values for phase3 model`

	Estimate	p-value
(Intercept)	-2557.10	1
gender: Women	-48.30	1
first interviewer rating	15.06	1
second interviewer rating	18.94	1

The summary of this model indicates that all the p-values for gender, the first interviewer rating and the second interviewer rating are greater than 0.05, which means there is no evidence showing that interviewer ratings and gender have effects on final decisions made by people. Furthermore, it means that it is a random event where the applicants enter final hiring from phase3.

Results

We fitted two GLMs to see whether there is any bias on gender based on AI's decisions at phase1 and phase2. Through data analysis, the p-value of gender of women is greater than 0.05, which means there is no evidence showing there exist differences of being admitted between men and women. Therefore, AI's decisions do not have any bias on gender when evaluating these applicants based on gender and applicants' talents and values at phase1 and phase2.

In phase1's model, p-values of gpa and work experience are less than 0.05, so there is enough evidence to show gpa and work experience have significant effects on AI's decisions.

In phase2's model, p-values of technical skills, leadership presence and speaking skills are less than 0.05, so there is enough evidence to show technical skills, leadership presence and speaking skills have significant effects on AI's decisions.

To see whether there is bias on gender in people's making decisions in the final hiring, we fit one GLM to see whether there is any bias on gender based on interviewer ratings. By data analysis, the p-values of all variables, including gender of women, are greater than 0.05, which means there is no evidence showing that interviewer ratings and gender have effects on final decisions made by people. Therefore, applicants are randomly hired from phase3 and no factors affect decisions of final hire.

Analysis of gender differences in promotion

Data wrangling and description

The current employee dataset records all employees' information quarterly for the whole duration of their employment. We are going to use the current employee dataset to solve the potential sexual prejudice issue in Black Saber Software. Employee ID, gender, average productivity, salary raise, leadership, and a newly created variable counting the times promoted for each employee are the key variables in this research question.

Each employee is uniquely identified by the 5 digits Employee ID. Later, we only keep each employee's basic information in 2020 Q4 once to eliminate the random effect brought by repeated measurements.

Each employee can identify himself/herself as "Man" or "Women", or one can "Prefer not to say" about gender. We want to know the distribution of gender in different teams, seniority levels, and in each promotion time.

The average productivity is calculated to represent the general productivity for each employee based on the number of quarters he/she worked at Black Saber Software. The salary raise is defined as the maximum salary minus the minimum salary that one employee had ever got in the company.

The variable leadership for level demonstrated whether the leadership quality is appropriate for an employee's seniority level in each quarter of the year. The working year is calculated by summing the number of quarters each employee had worked then divided by four.

The variable called number promoted counts the number of promotions one employee received during the time in Black Saber Company. From "Entry-level" to "Vice president", there are a total of 9 seniority levels. The count is the total number of levels each employee had been working at minus one. For example, one employee was at "entry-level" in the beginning, and during his/her career in the Black Saber Software, he/she had become "Senior I", "Manager" and "Director". Then, his/her total number of levels was 4, and the number of promotions he/she got is 4 minus 1 equals 3 times. Thus, the number of promotions is ranged between 0 to 8. If the number of promotions is zero, meaning this employee never received a promotion.

Since only the general representations of each employee's productivity and salary are considered, we updated the dataset to keep employee id once based on the newest 2020 Q4 data. Thus, only 607 observations are kept from the 6906 observations.

Tables and graphs

Table 2.1 lists the number of employees under each gender category. There were 340 males, 257 females, and 10 employees who preferred not to say about their gender.

Table 2.1: Number of Employees under Each Gender Category

Gender	Count
Man	340
Woman	257
Perfer not to say	10

Table 2.2 shows how many males, females, people who preferred not to identify the gender are under each promotion count. There is no employee who gets 6 or 8 promotions, so these two columns are empty.

Table 2.2: Gender Counts under Each Number of Promotions

	0	1	2	3	4	5	6	7	8
Man	143	88	340	55	11	3	0	1	0
Perfer not to say	3	5	10	2	0	0	0	0	0
Woman	149	62	257	27	4	1	0	0	0

Figure 2.1 illustrates the gender distribution in each number of promotions counted. We can observe that within the 295 people who did not receive any promotion, 149 of them are females, and 143 of them are males. Gender distributed relatively equally for non-promoted employees. However, among the 53 employees who got 3 promotions, only 14 of them are female while 39 of them are male. The only one who received 7 promotions is a male.

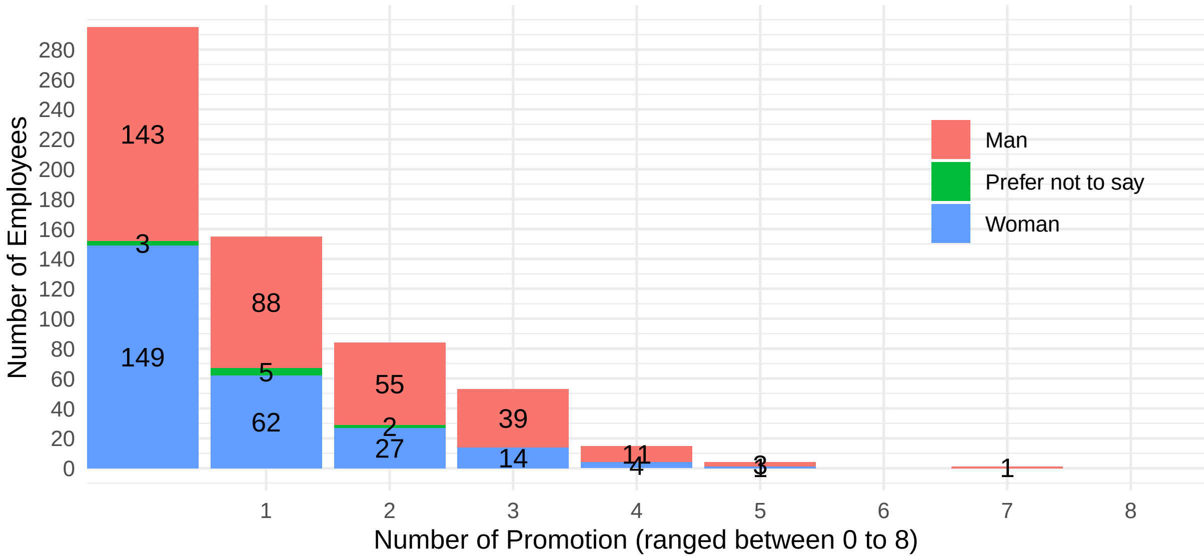


Figure2.1: Number of Promotion Distribution by Gender in Black Saber Software

Figure2.2 shows the gender proportion for each promotion count. From the bars, we observe that as the number of promotions increases, the male distribution increases. Only around one-quarter of females received 5 promotions while the other three quarters are males.

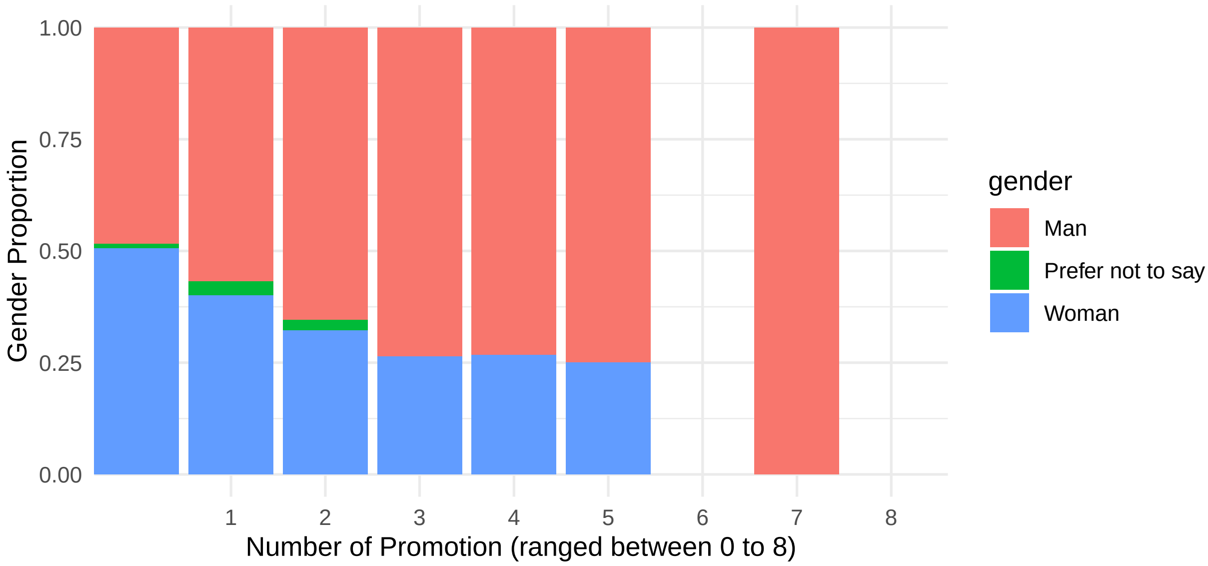


Figure2.2: Gender Proportion in Each Number of Promotion in Black Saber Software

We are interested in the historical gender distribution at each seniority level, so we created figure2.3. From the bars, there are more males for all seniority levels since the proportion taken by men all exceeds 50%. From the second senior table, we can see that the highest proportion of women was about 45.2% in the Senior III level, which is still less than 50%. And the lowest proportion of women is approximately 31.3% at the manager level.

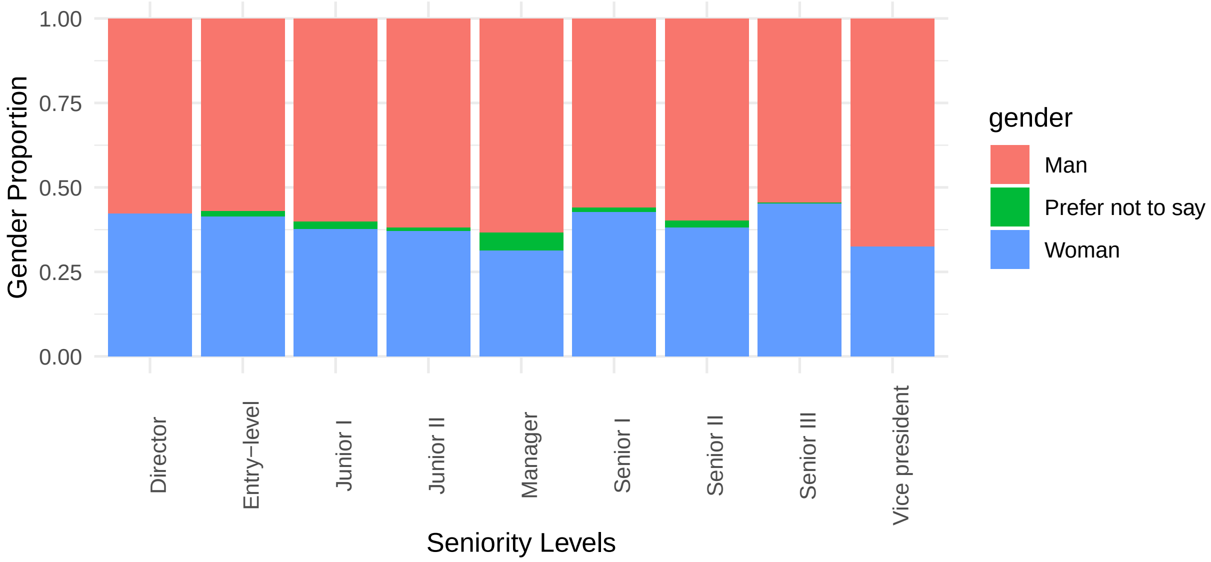


Figure2.3: Gender Proportion in Each Seniority Level in Black Saber Software

Model

We want to know what factors will influence the number of promotions received, so we treat the number of promotions as the response variable. This variable counts the number of promotions ranging from 0 to 8, which satisfies the characteristic of a Poisson distribution. We verify that the mean of the Poisson random variable number promoted is almost equal to its variance (1.513 and 1.689). Before fitting the generalized linear models (GLMs), we want to check the assumptions of GLM first. Since each employee has his/her own promotion times received, cases are independent in the data. Furthermore, the response variable number promoted has a linear relationship with the explanatory variables. We use GLMs to fit the following model.

The dataset used to fit the GLMs only contains each employee's information in 2020 Q4. The repeated measurements are excluded, so there is no need to use a mixed model.

However, each employee's average productivity and salary increase are related to the number of years they had worked at Black Saber Software. For example, some experienced employees may have worked for over ten years while some just entered the company. We solve this problem by taking log on the working year as an offset to standardize the number of years worked.

As to whether gender bias exists, the first GLM included gender as the explanatory variable and year working as an offset. We consider this model to be the base model, and more variables will be added to the base model. For the second model, a variable team is added. For the third model, average productivity is included. For the fourth model, salary raise is added and leadership for level is added to the last model.

To find out an appropriate model, we compared the AIC values of the five models. The fourth model that includes gender, team, average productivity, and salary raise as explanatory variables and year working as an offset has the lowest AIC value among all five models. The likelihood ratio tests will compare nested models to check the effect contributions to the model. To double-check the result, likelihood ratio tests are run by first comparing the first model and the second model. When adding the variable team into the base model, the likelihood ratio test outputs a p-value greater than 0.05, indicating the insignificant difference between the two models. Hence, we keep the simpler base model and compare the base model with the third model. Similarly, there is an insignificant difference between the two models. When comparing the base model with the fourth model, the likelihood ratio test outputs a p-value less than 0.05. This p-value indicated that there are significant differences between the two models, so we keep the more complicated model. Finally, when comparing model 4 to model 5, there is no evidence for significant differences. Thus, aligned with the AIC values, we choose model 4 to make further analysis and conclusions.

Results

Model 4 is a generalized linear model with the number of promotions being the response variable, gender, team, average productivity, and salary raise as explanatory variables, and year working is the offset of the model. The equation of the model is:

$\mathbf{E}(numPromoted) = e^{\beta_0+\beta_1 gender+\beta_2 team+\beta_3 avgProductivity+\beta_4 salaryRaise+\epsilon}$,

where ϵ is the residual errors.

We find gender women have a negative estimated intercept value, meaning that as one unit of increase of women, the number of promotions will decrease. The p-value of gender women is 0.00218, which is less than 0.05, indicating gender has significant effects on the number of promotions received by employees. This conclusion is supported when looking at the confidence interval of gender women. The 95% CI of gender women is (-0.477, -0.106) and this interval does not include 0 while all other variables include 0. The observation tells us that we expect that the interval will include the true population mean 95% of the time and in our case, the interval will not include 0 meaning there is some difference showing gender women's effect is not zero. The 95% CI after taking exponential is (0.620, 0.899) and it does not include 1. Again, the observation supports our conclusion that gender women will influence the number of promotions received.

The estimated coefficient of gender women is -0.29. Considering the equation of our model, this means that when team, average productivity, salary raise are held constant, if one employee identifies herself as women, she will get $e^{-0.29(0.748)}$ promotions of a man employee.

The variable measuring the salary raise of an employee also showed a significant effect on the increase of promotions received. This finding is logic-conserved. If one had a huge salary increase, it was likely that he/she got a promotion.

To answer the research question based on our model, there exists some evidence that being a woman will decrease the probability to get a promotion at Black Saber Software. Specifically based on the 2020 Q4 data, a female employee only gets 0.748 promotions of a male employee of the same team, average productivity, and of the same condition on salary raise. However, there are a smaller proportion of women in each team and each seniority level than men. Thus, the next research question we want to address is whether there is sexual prejudice in the hiring process of Black Saber Software.

Analysis of the gender pay gap

Data wrangling and description

The original salary dataset from the Black Saber Software company contains information on all current employees. To explore the relationship between salary and gender, the key variables we decided to use were employees' id, gender, productivity, levels of leadership, salary, seniority level, and team.

For the original salary dataset, rows includes current employee id, and columns includes variables of gender, team, the financial quarter salary, levels of leadership, productivity, and salary. Then we get the data we need to identify possible solutions to our research question.

We select the fourth quarter of 2020 data to work with since that is the latest information. Then the original data is filtered to have each row as a distinct employee. Next, like what we do in the analysis of potential gender bias in the hiring process, we clean the data to make sure variables are suitable for further statistical modeling. We delete useless symbols like '\', '$', and blank space in the variable of salary. After that, we convert the numerical variable of employee id into a categorical variable.

For the research question, we are interested in how the distribution of men and women by the average salary looks like. We calculate the average salary of all employees in the company and create a new variable called salary status to separate employees into two groups. If a random employee's salary is greater than the average salary, then we can conclude he/she is in the "High" salary group. Otherwise, this employee is in the "Low" salary group.

After that, further data visualization and statistical analysis can be carried on the new variables and the existing variables in the original data set.

Tables and graphs

Figure 3.1 illustrates the distribution of employees by men, females, and people who preferred not to say about their gender. Men occupy 56% of the pie chart, while women occupy 42% of the pie chart. There are only 2% of employees who preferred not to say their gender. The pie chart tells that there are more male employees in the Black Saber Software company.

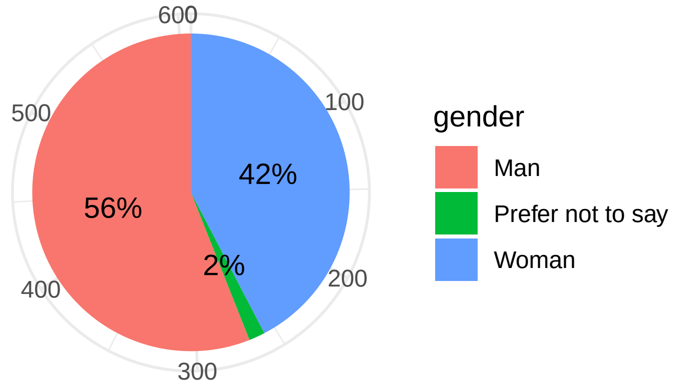


Figure 3.1: Gender proportion of employees in Black Saber

Figure 3.2 compares the proportion of high-paid current employees among different genders. There are 65% of men with high salaries, while there are 34% of women with high salaries, and there are only 2% of employees who preferred not to say their gender has a high salary. This bar chart is similar to figure 1, but the distribution of employees with high salaries by genders is more uneven. This figure suggests that male employees are well-paid employees and are more likely to be a man.

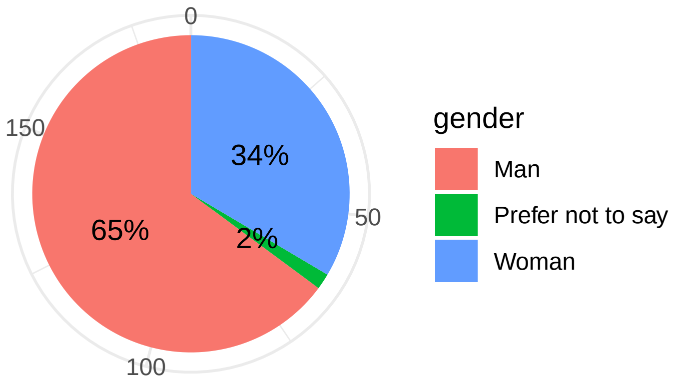


Figure 3.2: Gender proportion for high Salary Employees in

Figure 3.3 compares the proportion of low paid current employees among different genders. There are 52.1% of men with low salary, while there are 46.2% of women with low salary, and there are only 2% of employees who preferred not to say their gender has a low salary. This bar chart is similar to Figure 1, but not as uneven as Figure 2. This distribution shows that men and women look roughly evenly distributed.

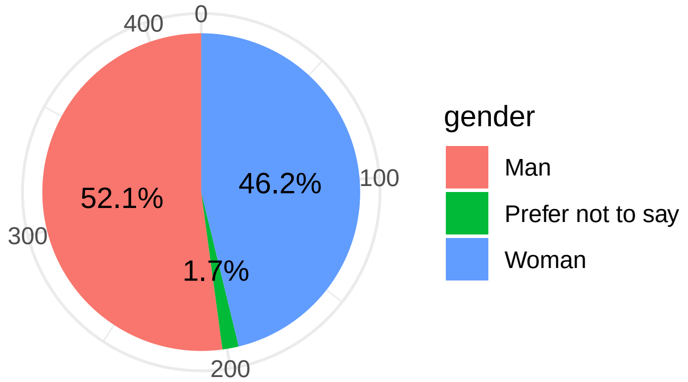


Figure 3.3: Gender proportion for low salary employees in B

Figure 3.4 shows that different teams have different ranges of salary. Thus we could not compare salaries in different teams by genders. This is the reason why we treat the team as a random effect in the following analysis.

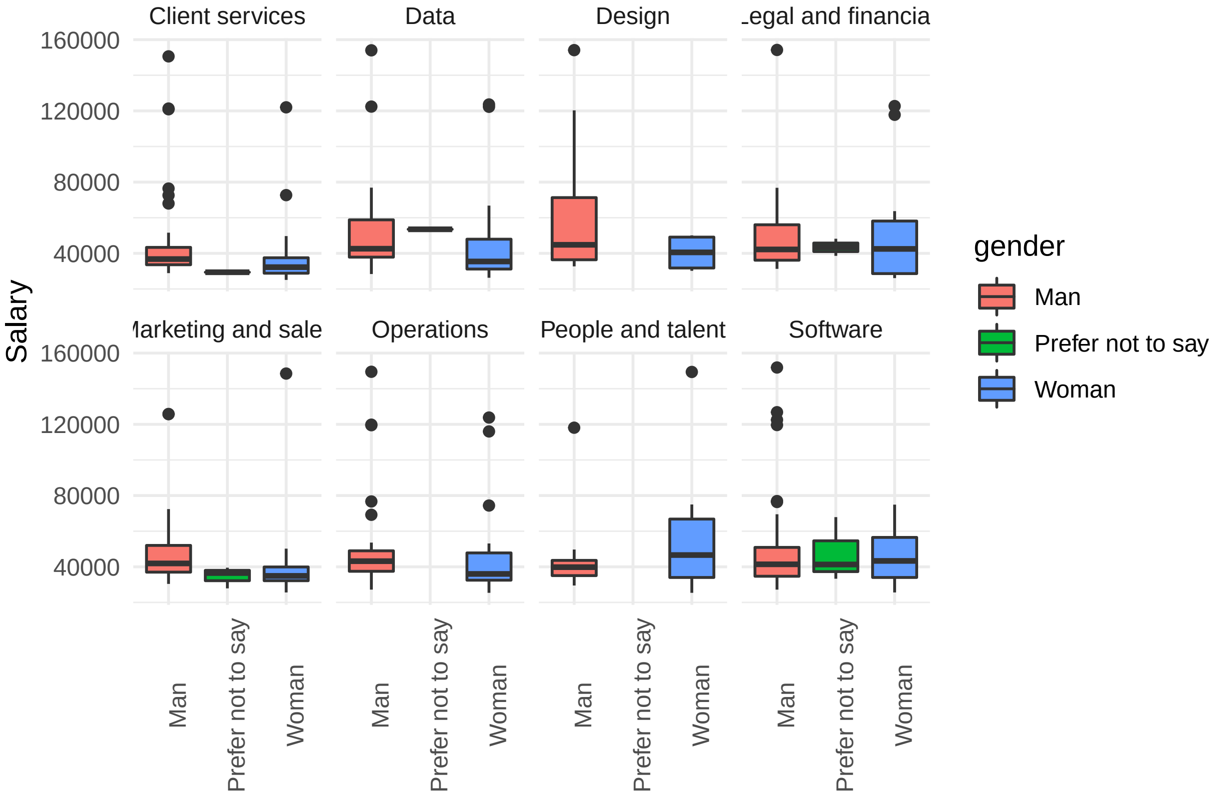


Figure 3.4. Distribution of Salary across 8 teams by gender.

Based on the above analysis of the data visualizations, we doubt that employees' gender would impact their salary. Thus, we build several linear mixed models to solve the puzzle.

Model

For the current employee's dataset, we see there is a dependency structure. The variable of team is the random effect since each department is responsible for different jobs, the salary ranges between departments are different as shown in the figure 3.4. Thus, we create five linear mixed models, and each of the models estimates the effect of one or more explanatory variables on the response which is a continuous variable called salary. The models differ by their different variables used.

Model 1 is the simplest model with only one fixed effect called productivity. Model 2 has both effects of productivity and gender. Through the likelihood ratio test, we get a p-value of 3.554×10^{-11} which is smaller than 0.05, so we have strong evidence against the hypothesis that model 1, with no gender, is as good as our full model. Model 3 has one more fixed effect called levels of leadership than model 2. Through the likelihood ratio test, the p-value is 2.09×10^{-9}, which is smaller than 0.05, so we have strong evidence against the hypothesis that model 2 without levels of leadership is as good as model 3. Thus, adding levels of leadership can fit the model better. Model 4 has one more fixed effect called seniority level than model 3. After the likelihood ratio test, we get a p-value smaller than 0.05, so we have strong evidence against the hypothesis that the simpler model, with no seniority level, is as good as our full model. Model 5 removes one fixed effect of gender from model 4. By the likelihood ratio test, 2.2×10^{-16} is smaller than 0.05, we have strong evidence against the hypothesis that model 5, without gender, is as good as model 4. Thus, adding levels of gender to the model can make the model be better. We conclude that model 4 is better than model 5.

We also compare the AIC value for those 4 models. The AIC result suggests that model 3 is the best since it has the lowest value.

Then we do several diagnostic check for model 3, like check the normal Q-Q plot in figure 3.5, finally we conclude that there is no violations to our original model assumptions.

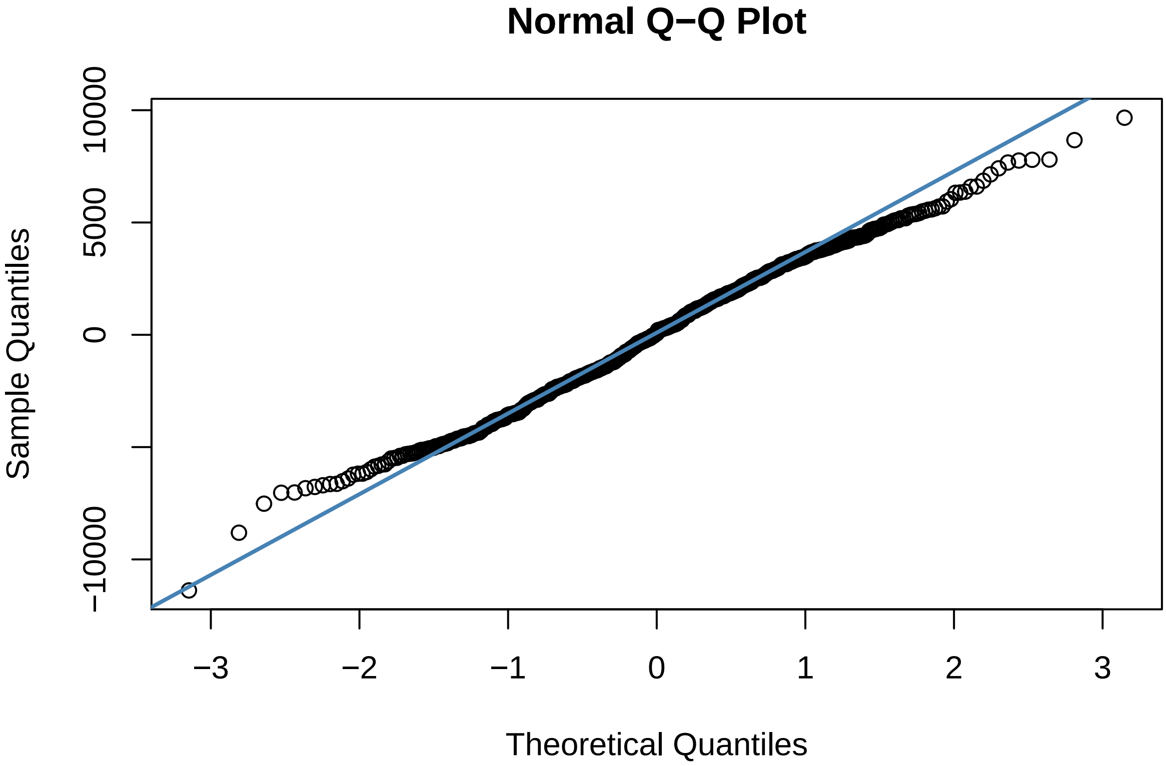


Figure 3.5: Normal QQ Plot For Model 3

The Equation of Model 3:

$Salary = \beta_0 + \beta_1 Gender + \beta_2 Productivity + \beta_3 Role + \beta_4 Leadership + bTeam + \epsilon$

Where b is the random effects for the units, ϵ is the residual errors.

The coefficient table of Model 3:

Table 3.1: Estimate Coefficients of Fixed Effects for Model 3

	Estimate		Estimate
Intercept	122607.26	Role of Seniority: Senior I	-76861.57
Gender: Prefer not to say	-1218.42	Role of Seniority: Senior II	-71535.78
Gender: Woman	-2173.65	Role of Seniority: Senior III	-65727.82
Role of Seniority: Entry level	-89968.04	Role of Seniority: Vice president	29917.34
Role of Seniority: Junior I	-85078.65	Leadership for level: Exceeds expectations	196.77
Role of Seniority: Junior II	-82355.64	Leadership for level: Needs improvement	-1370.06
Role of Seniority: Manager	-49465.27	Productivity	-8.128

Result

The best model to explain the data is Model 3, and the possible factors that influence the salary inequality by genders are productivity, leadership levels, salary, seniority level, and team. The 95% CI for gender women is (-2671.73, -1586.38) and this interval does not contain zero, that means the estimate of gender woman is significant.

The model suggests that for a man and a woman with the same average productivity, leadership for level and role seniority, a man's salary is higher than a woman for the whole duration of their employment. The coefficient of gender women is -2173.65 means the average salary of women is lower than the average salary of men by 2173.65 dollars.

Statistical pieces of evidence answer the research question that males have a higher salary than females at the Black Saber Software company. Salary inequality exists, so the company should come up with some policies to deal with injustice.

Discussion

In this analysis, we apply the current Black Saber Software employees' information to build prediction models based on the three research questions. For the first question, we discuss whether AI's hiring algorithm contains gender bias. Based on our model estimation results, we suggest that AI's hiring process does not have any bias on gender, rather its evaluation is based on the employee's talent and skills.

After that, we discuss whether gender bias would influence the number of promotions one can possibly get. The model reveals that if we have two employees with different gender but are in the same team, have the same average productivity, and have the same amount of salary raise, then the number of promotions a female got is 0.748 times of the male's. In other words, we can say that there exists gender inequality in promotion decisions at Black Saber Software.

Finally, we investigate whether females earn less salary than males. We find that when female employees and male employees have the same productivity, role, leadership quality level collected in the fourth quarter of 2020, females would earn 2173.654 dollars less than males. Thus, we can say that gender causes salary inequality at Balck Saber Software.

Limitations and futurn improvements

There are still some limitations in this study. Firstly, we have data missingness in this analysis. Approximately 2% of employees did not tell their gender. If we can have their information, we would do a better estimation on our interested questions. However, it is not ethical to force someone to identify his/her gender. We did not want to exclude the minority group's rights and hence kept the "prefer not to say" group in our analysis. In addition, we are missing the newest data for the first quarter of 2021, so the analysis result does not reflect the company's current state well.

Secondly, we could increase the representativeness of our model prediction by applying all quarters information into the model rather than only apply the fourth quarter data in 2020. For now, we only included the 2020 Q4 data to be focusing on the newest result as one company is improving and we think this is the best representation of the current equity situation at Black Saber Software.

Thirly, the data collection is limited, and it would be better if we can collect more information from employees to build the prediction models. For example, employees' attendance, number of years working, and performance score can be included. Also, it would be better if the dataset includes entry time as additional information. The zero promotions response is a mixture of non-promoted employees and employees who just entered the company. The entry time can distinguish between new employees and those who did not get promotion for a long time. Then, the Zero-Inflated Poisson regression model can be applied to take the new employees into account and hence make our prediction more accurate.

According to our analysis, we find that gender bias exists in promotion and salary paid at Black Saber Software. Therefore, we would recommend the decision makers at Black Saber Software to set up corresponding policies or surveys to help eliminate the inequality. Gender should not be the dominant factor for a company to make decisions, but performance appraisal is.

There exists algorithmic bias and selection bias from AI. The data and information collected from applicants are evaluated by AI at phase1 and phase2. Only factors of technical skills, leadership presence and speaking skills specific factors have significant effects on AI's decisions. So, if applicants' have good performance on those factors, they are more likely to be accepted into next phase, which is not fair for other applicants, who have good performances on remaining factors. Thus, Black Saber Software should also make some adjustments about their hiring process.

Consultant information

Consultant profiles

Our team has five professional consultants working on this case. The team members are:

Changhao Jiang. Changhao is a manager with Data Protect Corporation He specializes in coding and analysis. Changhao earned his Bachelor of Science, Specialist in Statistical Science: Theory and Methods, from the University of Toronto in 2022.

Qing Li. Qing is a senior consultant with Data Protect Corporation. She specializes in statistical communication. Qing earned her Bachelor of Science, Majoring in Economics and Statistics from the University of Toronto in 2022.

Qian Wang. Qian is a senior consultant with Data Protect Corporation. She specializes in reproducible analysis. Qian earned her Bachelor of Science, Double Majoring in Statistics and Mathematics from the University of Toronto in 2022.

Yancheng Sylvia Yu. Yancheng is a junior consultant with Data Protect Corporation. She specializes in statistical analysis and data visualization. Yancheng earned her Bachelor of Science, Double Majoring in Statistics and Mathematics from the University of Toronto in 2022.

Xi Zheng. Xi is a junior consultant with Data Protect Corporation. She specializes in data visualization. Xi earned her Bachelor of Science, Majoring in Mathematics and Statistics from the University of Toronto in 2022.

Code of ethical conduct

Data Protect Corporation promotes high standards in ethical statistical practice. There are five guidelines for all consultants in Data Protect Corporation to follow with, and they are:

- Be honest and show respect to the clients at any time.
- Strictly implement statistical sampling methods and analyze the result with no personal prediction.
- Provide a healthy working environment with no discrimitaion, gender inequality, bullying, and other harassments.
- Protect privacy and the data given by the clients. Sign confidentiality agreements with the clients for every case we undertake.
- Ensures the working procedure and reporting of statistical design and analysis is consistent with local laws.

References

Bevans, R. (2020, March 27). Akaike Information Criterion: When & How to Use It. Scribbr. https://www.scribbr.com/statistics/akaike-information-criterion/.

Spiggle, T. (2019, July 1). Is It Really Harder for Women to Earn Promotions in the Corporate World? The Spiggle Law Firm. https://www.spigglelaw.com/employment-blog/really-harder-women-earn-promotions-corporate-world/.

Ward, D. (2018, June 14). The gender pay gap: What now? Retrieved April 19, 2021, from https://www.kingsfund.org.uk/blog/2018/06/gender-pay-gap-what-now?gclid=EAIaIQobChMIqKuLz_6J8AIVVUNgCh3IcgA9EAAYAyAAEgLdFvD_BwE