
Research on Potential Gender Discrimination in Black Saber Software

Focusing on the Salary Level, Promotions and Hiring

Report prepared for Black Saber Software by YYDS
Limited Co.

April 21, 2021

Contents

Executive summary	3
Introduction	3
Target Questions	3
Investigation	3
Conclusion	4
Technical report	5
Introduction	5
Research Question 1: Are the salary processes fair in the company? If so, what factors will affect the salary level?	6
Research Question 2: Whether there is any gender discrimination in the employees' promotion, and whether the level of work performance corresponds to the employee promotes	15
Research Question 3: Finding factors that influence applicants' selection and whether AI-automated selection is fair to gender	22
Discussion	28
Consultant Information	30
Consultant Profiles	30
Code of Ethical Conduct	30
Reference	31

Executive summary

Introduction

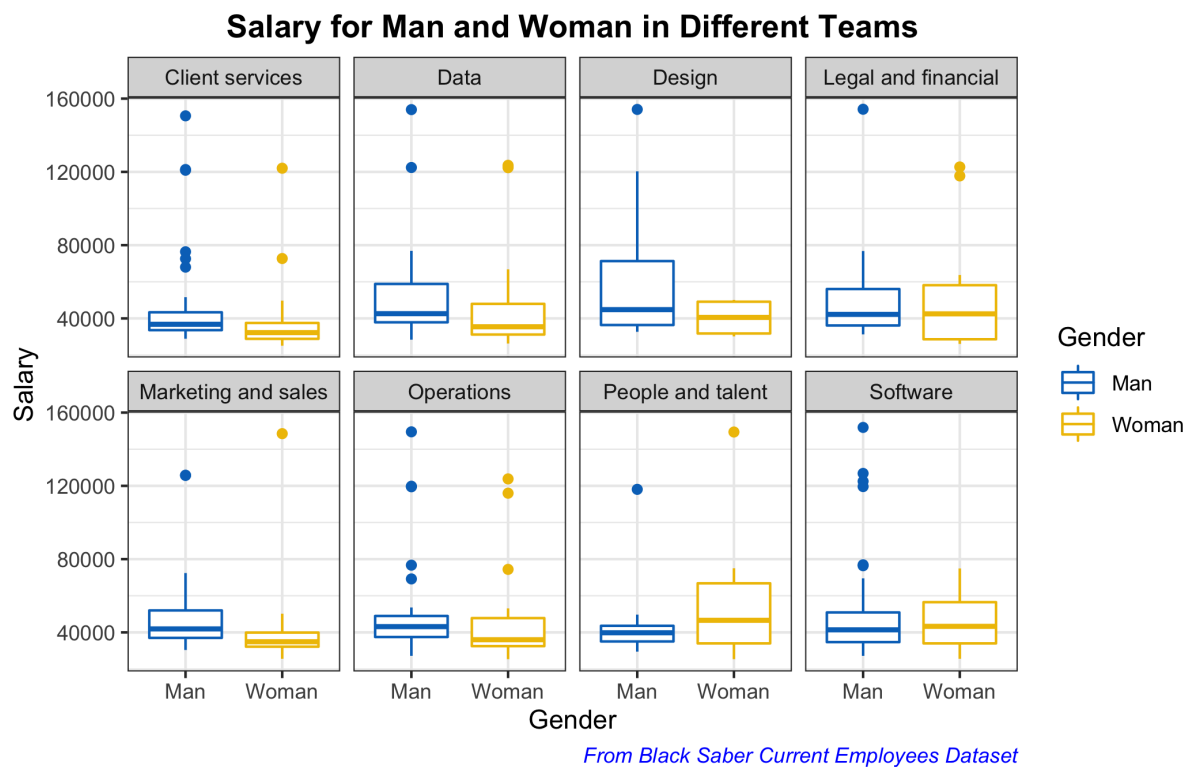
Regarding the concerns about gender parity in the hiring, salary, and promotion processes in Black Saber Software, a statistical analysis is made to investigate the potential issues. The more detailed technical analysis is attached below in the section of the technical report. In this report, target questions are listed, critical visualizations and tables are provided to illustrate the result of our final investigation.

Target Questions

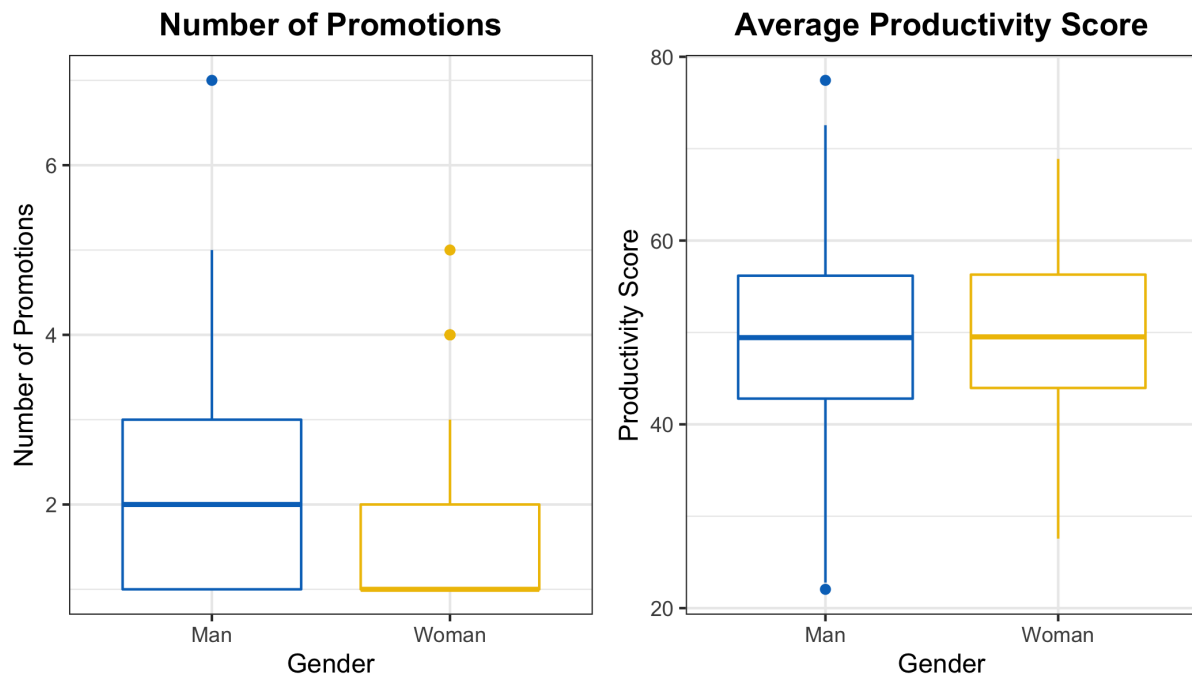
- Is the salary process fair in the company?
- Is the promotion process fair in the company?
- Is the hiring process fair in the company?

Investigation

After digging into the data, a few data visualizations and tables are created.



In the above Figure 1 located at the first research question section of the technical report, we can see that for most teams, the boxplots for man have slightly higher median values than the boxplots for woman, which means that generally man's salary is higher than woman's salary. This implies the unfairness may exist in the salary process of the company.



The above figure is combined by Figure 3 and Figure 4 in the second research question section of the technical report. The median number of promotions for man is about 2, while the median number of promotions for woman is 0. But their average productivity scores are nearly the same since the boxes are almost identical. This means while there is no difference in productivity between men and women, more men get promoted than women do. As it suggests, this is pretty unfair.

By looking at the final hiring decision, we can see in the hiring process, 8 men and 2 women are hired. In this case, gender seems to be the issue here, but we also find out the relevant scores are very similar between men and women. After sufficient statistical analysis, it turns out that gender doesn't actually affect the hiring decision. Therefore, the AI-automated selection in hiring process is fair for different gender.

Conclusion

In Black Saber Software, the salary and promotion processes are unfair to gender, men have more advantages than women. However, the hiring process led by AI-automated selection is relatively fair, it is purely based on employees' talents, skills, and values to the company.

Technical report

Introduction

Black Saber Software provides us five sets of data. The data of current Black Saber software employees contains personal information such as employee's unique ID and gender. It also includes employees' positions indicating by the team they are located in and their seniorities, and financial information is shown as the salary of each financial quarter, productivity scores, and leadership levels present employees' work performance in the company. The other four sets of data include the hiring information of each phase until the final round. There are a total of 3 phases. The given datasets include applicants' information about unique ID, gender, GPA, which team they apply for, whether they have a cover letter and CV, work experience, and what level of extracurricular activities they have. The data of phase 2 adds scores of applicants' technical, writing, speaking, and leadership levels. Phase 3 data contains each interview rating for applicants. The final hiring data only shows applicants' unique IDs.

In this reseaeching report, hiring and current employees data from Black Saber Software will be analyzed, and models would be built based on the collected data. Investigations on the topic of gender parity in the hiring, wages, and promotion process will be made in this technical report. In this report, summary tables and informative figures are generated to convince the general finding. Statistical inferences are also used to provide evidence of hypotheses.

Research questions

1. Is the salary process fair in the company?
 - What factors will affect the salary level?
 - Is gender unrelated to salaries?
2. Is the promotion process fair in the company?
 - Is there any gender discrimination in the employees' promotion?
 - Is work performance/productivity taken into account?
3. Is the hiring process fair in the company?
 - What factors will influence Applicants' Selection?
 - Is AI-automated Selection Fair to Gender?

Research Question 1: Are the salary processes fair in the company? If so, what factors will affect the salary level?**Data Manipulation**

After observing the data of Black Saber Software's current employees, we find that the variable *gender* contains three values: "Man", "Woman" and "Prefer not to say". We decide to filter out "Prefer not to say" since it is not representative of gender. We also find that the values in the *salary* variable are characteristics with some characters such as "\$" and ",". Since we want to study the relationship between salary and other factors, we remove these special symbols and convert the characteristics values into numeric values which are suitable for our analysis. We also discover that the number of observations for each unique employee is unequal. For example, the employee with *employee_id* 24596 has four observations, whereas the employee with *employee_id* 24381 has eight observations. To solve this, we decide to only use the data for the latest financial quarter – Q4 of the year 2020. In this way, each employee will have exactly one observation, and the observations will be independent of one another. Also, the *2020 Q4* data represents the latest information for the company, and it will be the most supportive and sufficient information for our analysis. What's more, we look at the *role_seniority* variable and notice that different levels are not ranked from least senior to most senior. It might cause some misunderstandings in the results. Thus, we rearrange the levels in *role_seniority* so that it will be shown from the least senior to the most senior.

Exploratory Plots and Summary Tables

The core of our research question is to figure out whether the salary process is fair in the company. We make an initial guess that the salary process might be unfair. If this is true, then some factors affect the salary level to make the processes unfair. Then we start to explore what factors could affect the salary level. After observing the data of current employees in Black Saber Software, we think about *gender* and *seniority* might be effective to the salary level. Then we draw two box plots to test our guesses. The box plots are shown below:



Figure 1 is a boxplot with salary as the y-axis and gender as the x-axis with respect to different teams. A boxplot is a standardized way of displaying the dataset based on a five-number summary: the minimum, the maximum, the sample median, and the first and third quartiles. The box is drawn from the first quantile to the third quantile with a horizontal line drawn in the middle to denote the median. Figure 1 shows that, on average, a man's salary is a little bit higher than a woman's salary across different teams since the median in the boxplots of man is slightly higher than the median in the boxplots for woman.

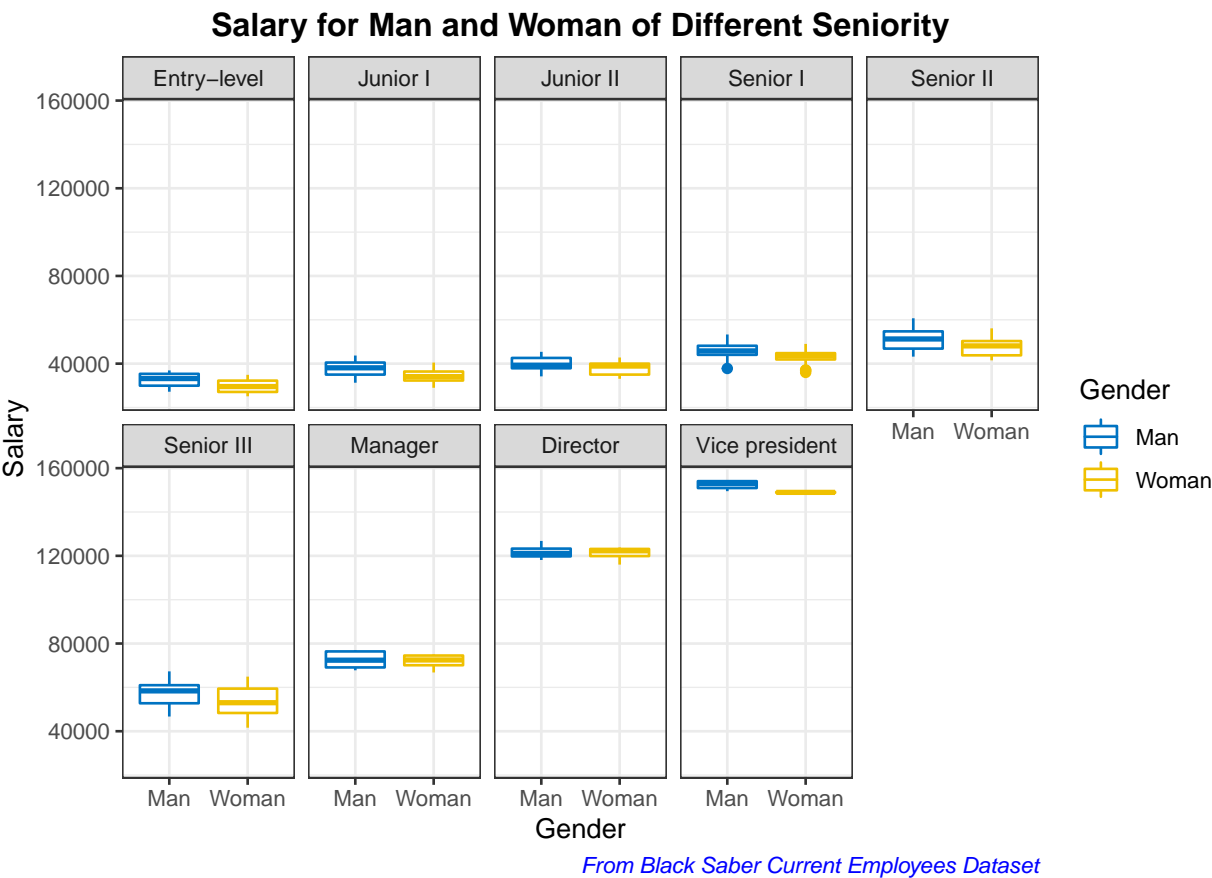


Figure 2: Salary for Man and Woman of Different Seniority

Figure 2 is a boxplot with salary as the y-axis and gender as the x-axis with respect to different seniority. It shows that, on average, a man’s salary is a little bit higher than a woman’s salary across different seniority. It seems that the more senior the employee, the higher the salary they get paid.

These two boxplots strengthen our guesses that gender and seniority could be effective factors influencing the salary level. To make our guesses more convincing, we calculate the average salary for both man and woman in different seniorities. Then we summarize the results in the table below:

Table 1: Average Salary for Man and Woman of Different Seniority

Role Seniority	Average Salary for Man	Average Salary for Woman
Entry-level	32739.24	29797.92
Junior I	37834.78	34394.34
Junior II	39837.74	38135.29
Senior I	45973.81	43075.86
Senior II	51117.86	48071.43
Senior III	57421.88	53900.00
Manager	72436.84	71762.50
Director	121891.67	121171.43
Vice president	152383.33	148950.00

Table 1 summarizes the average salary for men and women of different seniorities. Within each seniority, a man has a higher salary than a woman. The more senior the employee, the higher the salary they will get. This corresponds to the observational results from Figure 2.

From Figure 1, Figure 2, and Table 1, we are convinced that the salary process could be unfair. Women seem to be discriminated against on their salary level, while seniority seems to be a strong factor that affects salary level. However, without appropriate statistical analysis, it is possible for our conclusion being wrong by just observing the data. Also, there might be other factors that affect the salary level. Thus, a proper statistical procedure is carried out to find out what factors are significantly affecting employees' salary.

Method

The linear mixed model is an extension of a simple linear model to allow both fixed and random effects and is particularly used when there is non-independence in the data. Since our data contains repeated employee ID within the same team, the observations are dependent. In this case, a linear mixed model is chosen.

Our goal is to find out the factors that affect salary, then an appropriate way to do this is to use regression analysis. The data we used is *2020 Q4* data with 8 variables: *employee_id*, *gender*, *team*, *role_seniority*, *leadership_for_level*, *productivity*, *financial_q* and *salary*. Across these eight variables, since we only use *2020 Q4* data, *financial_q* variable should be neglected. Also, the *employee_id* variable is not predictive for salary, and we excluded this one as well. We believe that the *team* variable should be considered a random effect since the average salary might differ across teams. Therefore, we would fit a linear mixed model with salary as the response variable and team as a random intercept.

There are many possibilities to build models that include different fixed effects. Since we do not know which model is the best, we start from the simplest model with only productivity as fixed effects and team as a random intercept. We add a new variable whenever we build a new model and compare it with the old one using likelihood ratio tests. Using the likelihood ratio test, we can determine which fixed effect is significant so that it helps us choose a better model. After several rounds of comparison, we get the best predictive model. After obtaining the best model from the likelihood ratio test, we use *summary()* function to get the estimated coefficient for our fix effects. Then we can interpret the results. To further check which fixed effect is significant in affecting the salary level, we obtain the 95% confidence interval for their estimated coefficients. If the 95% confidence interval includes 0, then it is not significant. If it excludes 0, then it is significant. Note that the 95% confidence interval results could be different from the likelihood ratio test; this is because the likelihood ratio test treats all variables as a whole while the confidence interval checks the significance individually. The likelihood ratio test is more reliable at this point.

Model Equation

- *Model 1: salary* = $\beta_0 + \beta_1 X_{productivity} + U_i$
- *Model 2: salary* = $\beta_0 + \beta_1 X_{leadership_for_levelExceeds\ expectations} + \beta_2 X_{leadership_for_levelNeeds\ improvement} + \beta_3 X_{productivity} + U_i$
- *Model 3: salary* = $\beta_0 + \beta_1 X_{genderWoman} + \beta_2 X_{leadership_for_levelExceeds\ expectations} + \beta_3 X_{leadership_for_levelNeeds\ improvement} + \beta_4 X_{productivity} + U_i$
- *Model 4: salary* = $\beta_0 + \beta_1 X_{role_seniorityJunior\ I} + \beta_2 X_{role_seniorityJunior\ II} + \beta_3 X_{role_senioritySenior\ I} + \beta_4 X_{role_senioritySenior\ II} + \beta_5 X_{role_senioritySenior\ III} + \beta_6 X_{role_seniorityManager} + \beta_7 X_{role_seniorityDirector} + \beta_8 X_{role_seniorityVice\ president} + \beta_9 X_{genderWoman} + \beta_{10} X_{leadership_for_levelExceeds\ expectations} + \beta_{11} X_{leadership_for_levelNeeds\ improvement} + \beta_{12} X_{productivity} + U_i$

Where U_i represents random effect for $team_i$. The reference level for $role_seniority$ is “Entry-level”. The reference level for $leadership_for_level$ is “Appropriate for level”. The reference level for gender is “Man”.

Result

In the likelihood ratio test, the chi-square statistics are used to test the significance. Model 1 and Model 2 are compared in the first likelihood ratio test, and the corresponding p-value is 9.057×10^{-10} . Since it is less than the significance level of 0.05, we have evidence to address that the $leadership_for_level$ variable is significant. We have to include this in our model. Model 2 is better than Model 1. Thus we prefer model 2. In the second likelihood ratio test, Model 2 and Model 3 are compared. From the above model equation, we can see that Model 3 is more complicated than Model 2 by adding an additional independent variable $gender$. The corresponding p-value here is $2.56 \times 10^{-7} < 0.05$, which means $gender$ is statistically significant to the model and necessary to be included. Model 3 is better than Model 2; Model 3 is preferred. In the last likelihood ratio test, Model 3 and Model 4 are compared. Model 4 is more complicated than Model 3 by adding an additional independent variable $role_seniority$. Resulting in a 2.2×10^{-16} p-value indicating the more complicated model is better than the relatively simple one. Therefore, $role_seniority$ is significant, and Model 4 is preferred. Factors of $leadership_for_level$, $gender$, $role_seniority$ and $productivity$ should all be taken into account. Overall, after these comparisons, Model 4 is chosen as the final model for further analysis in the first research question.

Therefore, the best model is given by $leadership_for_level$, $gender$, $role_seniority$, and $productivity$ as fixed effects and a random intercept for $team$. The 95% confidence interval for the estimated

coefficient of *genderWoman* is (-2761.51857,-1574.77563), which excludes 0 and is negative. This corresponds to the likelihood ratio tests that *gender* is significant in affecting the salary level. The estimated coefficient for *genderWoman* is -2164.541 with a t-value of -7.096. This means that when keeping all other variables the same, under the same team, a woman's salary is lower than a man's salary by 2164.541.

The 95% confidence interval for estimated coefficient of *role_seniority* is listed below in a table:

Table 2: 95% Confidence Interval for Estimated Coefficient of *role_seniority*

Level of <i>role_seniority</i>	Confidence Interval for Estimated Coefficients
Junior I	(4057.02044, 5621.04540)
Junior II	(6645.57295, 8569.07859)
Senior I	(12120.75820, 13996.84512)
Senior II	(17285.00518, 19448.21312)
Senior III	(23201.67336, 25251.39193)
Manager	(39316.78674, 42093.65519)
Director	(88360.83717, 91561.03870)
Vice President	(117478.82465, 122268.60141)

Table 2 contains the results of the 95% confidence interval for estimated coefficient of *role_seniority*. Based on Table 2, The 95% confidence interval for estimated coefficients of *role_seniority* all exclude 0 and is all positive. This corresponds to the likelihood ratio tests that *role_seniority* is significant in affecting the salary level. The estimated coefficients and t-values of *role_seniority* is given in Table 3 on the next page.

Table 3: Estimated Coefficients and T-values of *role_seniority*

Level of <i>role_seniority</i>	Estimated Coefficients	t-value
Junior I	4839.461	12.024
Junior II	7608.033	15.370
Senior I	13059.020	27.049
Senior II	18369.527	33.007
Senior III	24228.082	45.936
Manager	40712.872	57.025
Director	89961.779	109.243
Vice President	119879.345	97.276

Table 3 can be interpreted as: compared with *Entry-level* seniority, when keeping all other variables the same and under the same team, the higher the seniority, the higher the salary. For example, fix all other variables the same, *Manager* has a salary of 40712.872, which is more than an *Entry-level* employee does, whereas *Vice president* has a salary of 89961.779, which is more than an *Entry-level* employee does under the same team.

Table 4: 95% Confidence Interval for Estimated Coefficient of *leadership_for_level*

Level of Leadership	Confidence Interval for Estimated Coefficients
Exceeds expectations	(-2007.04356, 2307.38228)
Needs improvement	(-3438.11853, 736.63934)

Table 4 shows the 95% confidence interval for estimated coefficient of *leadership_for_level*. The 95% confidence interval for estimated coefficients of *leadership_for_level* all include 0, which means it is not significant. This contradicts to likelihood ratio tests that *leadership_for_level* is significant in affecting salary level. The estimated coefficients for the level “*Exceeds expectations*” is 173.248 with a t-value 0.154, whereas the estimated coefficients for the level “*Needs improvement*” is -1356.226 with a t-value -1.263. This means that when keeping all other variables the same and under the same team, compared with “*Appropriate for level*” in *leadership_for_level*, “*Exceeds expectations*” employee has a salary of 173.248 higher than “*Needs improvement*” employee who has a salary of 1356.226.

The 95% confidence interval for estimated coefficient of *productivity* is (-27.15337, 11.95991) which includes 0. This contradicts to likelihood ratio tests that productivity is significant in affecting salary level. The estimated coefficient for productivity is -7.531 with a t-value of -0.749. This means that when keeping all other variables the same and under the same team, when productivity increases by 1 unit, salary decreases by 7.531 on average. The reason for a negative estimated coefficient for productivity might be: the standard for productivity rating will be higher when the seniority is higher. However, directors and vice presidents have a much higher salary than other employees. Thus the model gives us a negative estimate for productivity.

Although the likelihood ratio test contradicts the 95% confidence interval for the significance of *productivity* and *leadership_for_level*, these two fixed effects should still be considered significant since the likelihood ratio test is more reliable.

Research Question 2: Whether there is any gender discrimination in the employees' promotion, and whether the level of work performance corresponds to the employee promotes**Data Description**

By looking at the data from Black Saber Software's current employees, we can see that for some employees, and they have the different role of seniority, the number of changes in the titles indicates the number of the employee's promotions. For example, the *employee_id* 24537 had a role of seniority as "*Entry-level*" on 2018 Q2 and Q3, starting from 2018 Q4 he promoted as "*Junior I*" until the newest quarter, which means from 2018 to 2020, he only promoted once. Then we count the number of unique titles each employee has and subtract it by 1, and the result is the number of promotions for each employee. We also calculate the average year of service and average productivity score of each employee to see if they are factors affecting the employee's promotion. What's more, we notice that some employees have comments "*Exceeds expectations*" on their leadership level, and they will promote on the next quarter. Since we are interested in whether the employees' work performance is effective to their promotions, we count how often each employee comments "*Exceeds expectations*" as a factor. Save the above factors into a new dataset called *promotion_data*.

Additionally, the central core of the question is to determine if gender is influential to employees' promotion, so that we have to include the variable gender in our dataset. We left to join the data of black-saber current employees with *promotion_data* by *employee_id*.

Finally, we create a new dataset called *promotion_data_2*, which only has factors we will explore. Note that we filter out those employees who have the number of promotions as 0; since they have not yet been promoted, they are not our object of study, and we want to avoid the effect on the excess of zero counts.

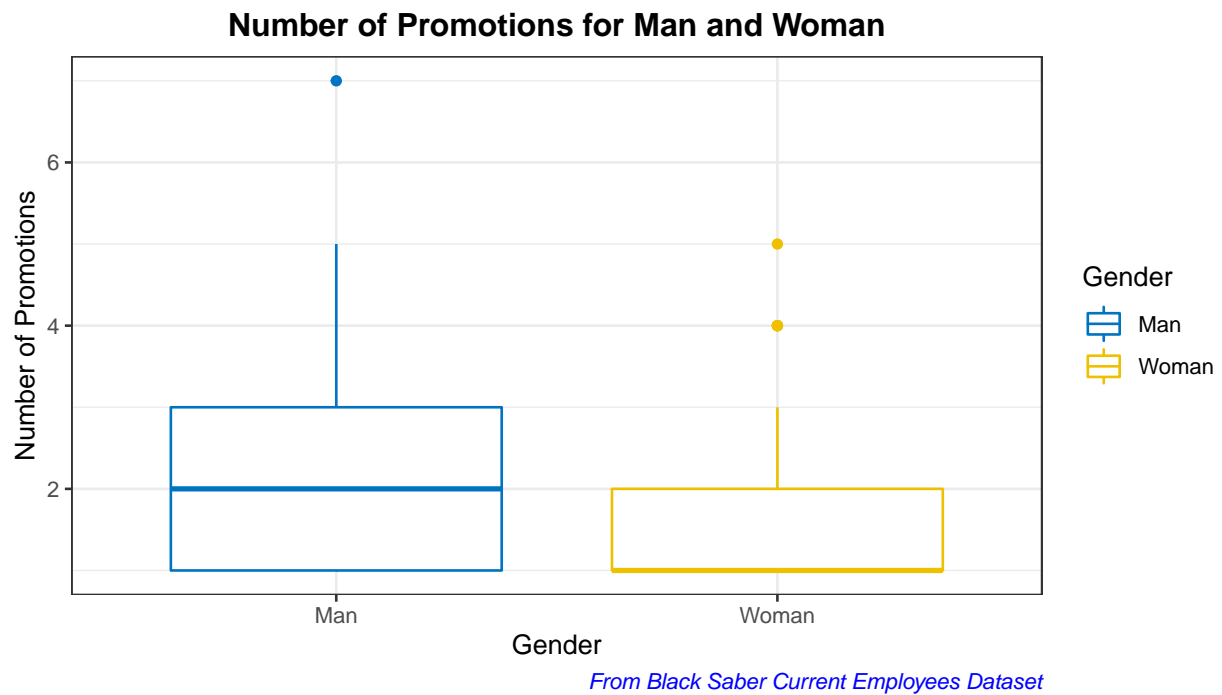
Exploratory Plots**Figure 3:** Number of Promotions for Man and Woman

Figure 3 shows how the number of promotions is different in employees' gender. The x-axis is represented by gender, and the y-axis is represented by the number of promotions each employee has throughout all their service years. Notice the significant difference in means of the two boxplots. Man's average number of promotions is much higher than woman's. The third quantile of man is relatively higher as well, which means there is about 25% of the data for the man having more than 3 promotions, while the 25% of the data for woman has more than only 2 promotions.

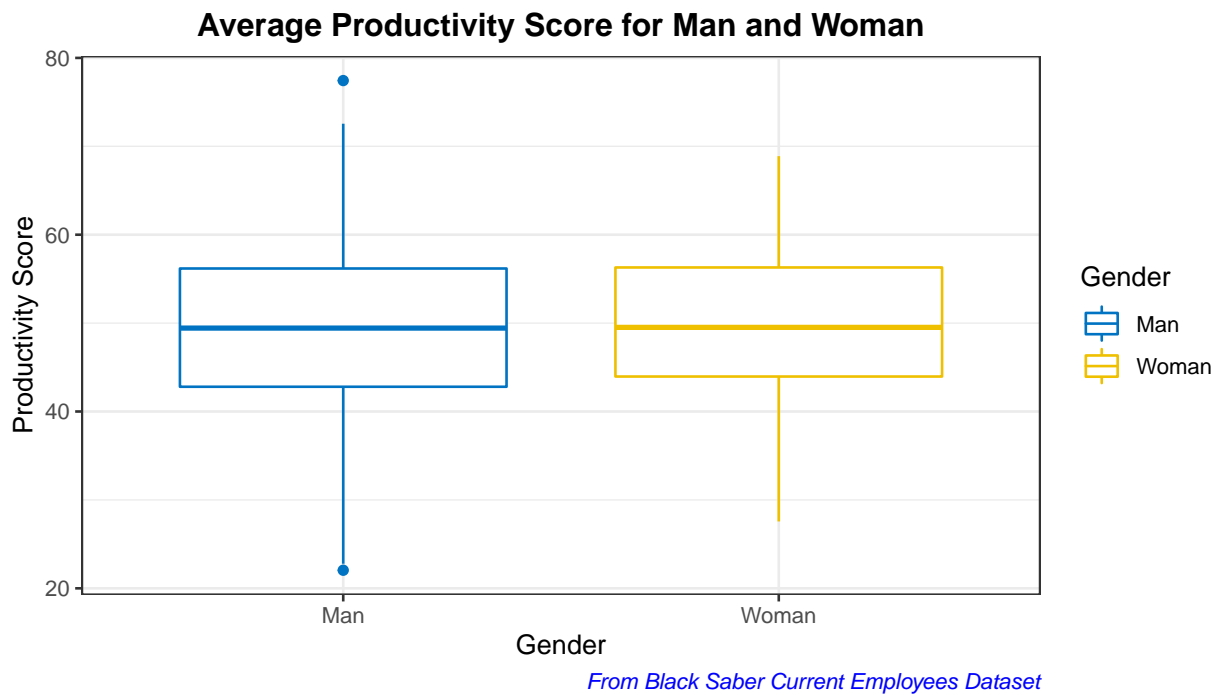


Figure 4: Average Productivity Score for Man and Woman

Figure 4 shows how the average productivity score is different in employees' gender. The x-axis is represented by gender, and the y-axis is represented by the average productivity score each employee has. The two boxplots seem to be very similar, and this implies that there is no much difference in productivity between man and woman.

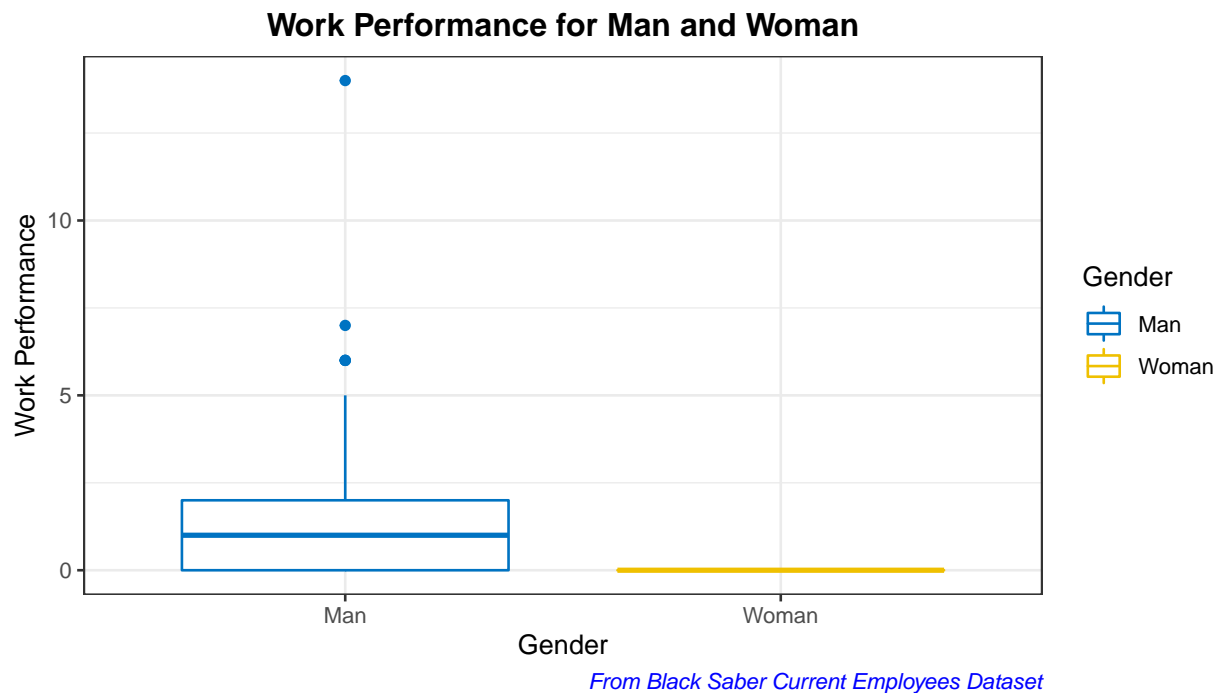


Figure 5: Work Performance for Man and Woman

Figure 5 shows how the work performance is different in employees' gender. The level of work performance is indicated as the leadership level in our data. The more times employees get the level "*Exceeds expectations*", the better the work performance they are considered to have. The x-axis is represented by gender, and the y-axis is represented by the number of times each employee gets the level "*Exceeds expectations*". By observing Figure 5, we can easily see that man does a significantly great job in work performance than a woman does.

In the above figures, even though we can only see gender affecting the employees' promotions, we explore how men and women perform in the company. By comparison, more male employees get "*Exceeds expectations*" as their level of leadership, but we have to keep in mind that man and woman do not have much difference in productivities, which means we can not conclude man does better than a woman in the company. Therefore, by the condition of man and woman having a similar level of work performance, the higher number of promotions male employees get can be caused by the possible gender discrimination. To further analyze this research question, we are going to introduce the statistical method used to study relative factors and conclude more precise results.

Method

The generalized linear model (GLM) is a flexible generalization of ordinary linear regression that allows for response variables with error distribution models other than a normal distribution. A generalized linear model does not assume a linear relationship between the response variable and the explanatory variables, but it does assume a linear relationship between the explanatory variables. Furthermore, a generalized linear model expresses the relationship between the influencing variables and the response variable in a linear and additive way even though the underlying relationships may be neither linear nor additive.

Poisson regression models are generalized linear models (GLM) used to model count data with the logarithm as the link function. The poisson regression model assumes the response variable has a Poisson distribution which has only one parameter that stands for both mean and standard deviation.

When collecting count data, larger or smaller counts may be observed based on the size of observational units or the amount of time spent measuring a particular unit. In our data, the number of observations for each employee varies by the amount of service year. Ideally, each employee would have the same duration for their employment. However, if the durations of employment are various, then we would expect to observe larger number of promotions for employees observed for longer periods of time. In this case, it may be more appropriate to model the rate of promotions observed per unit of service year rather than the number of promotions observed for each employee. Using an offset gives us a convenient way to model rates per service year, instead of just modeling the raw counts.

To explore factors affecting employees' promotion, it is appropriate to apply a regression analysis on count data with the number of promotions for each employee being the response variable. Therefore, a Poisson generalized linear model (GLM) is fitted with offset for each employee's year of service. The model studies factors including gender and work performance such as average productivity score and the number of "*Exceeds expectations*" on the leadership level.

Model Equation

$$\log(\lambda_{\text{number_of_promotion}}) = \beta_0 + \beta_1 X_{\text{genderWoman}} + \beta_2 X_{\text{avg_productivity}} + \beta_3 X_{\text{num_exceeds}} + \log(\text{service_year}) + \epsilon$$

Where $\lambda_{\text{number_of_promotion}} = E(\text{numer_of_promotion})$, $\log(\text{service_year})$ is the offset term.

Result

Table 5: Summary Table of The Poisson Regression Model

term	estimate	std.error	statistic	p.value
(Intercept)	-1.054	0.234	-4.507	0.000
genderWoman	-0.237	0.102	-2.328	0.020
avg_productivity	0.007	0.005	1.328	0.184
num_exceeds	0.008	0.030	0.257	0.797

Table 5 shows the summary table of the model results, including coefficients, standard errors, statistics, and p-values. From the table, we observe that being a woman in gender will reduce the number of promotions, have a higher average of productivity score, and have more comments of “*Exceeds expectations*” on leadership level can increase the number of promotions. Moreover, only the factor gender has a p-value smaller than the significance level of 0.05, which means we have strong evidence to reject the null hypotheses such that gender does not have an effect on the number of promotions for employees. Thus, gender has a significant effect on the number of promotions. Gender discrimination exists in the promotions for employees of black-saber. In addition, although better work performance represented by average productivity score and the number of “*Exceeds expectation*” comments in leadership level can result in an increase in the number of promotions, their p-values are both greater than 0.05, then we will keep the null hypotheses such that the two factors do not affect the number of promotions for employees of Black Saber Software.

Table 6: 95% Confidence Interval for Estimated Coefficient of The Poisson Model

Factors	Confidence Interval for Estimated Coefficients
genderWoman	(-0.438579983, -0.03892403)
avg_productivity	(-0.003108127, 0.01626319)
num_exceeds	(-0.053771254, 0.06426183)

Table 6 describes the 95% confidence interval of estimated coefficients in the Poisson regression model generated in the model section. Given the result that *gender* has a significant effect on employees' promotion since its 95% confidence interval excludes 0. Both confidence intervals of average productivity score and the number of times receiving the level "*Exceeds expectations*" include 0; this implies that the two factors do not significantly affect our response variable. This result further strengthens the fact that gender discrimination in employees' promotion and work performance is not considered important consideration of the promotion process.

Research Question 3: Finding factors that influence applicants' selection and whether AI-automated selection is fair to gender

Data Cleaning

After loading in the datasets of *phase1*, *phase2*, *phase3* and *final hires*, a new variable *pass* is created based on whether the applicant is selected to the next phase or not with value *1* represents the applicant is selected to the next phase and value *0* indicates the applicant is not selected. For instance, now the *phase1* dataset includes a column called *pass* indicating whether the corresponding applicants have successfully moved in phase2 or not. *phase2* dataset also includes a column called *pass* indicating whether the corresponding applicants have successfully moved in phase3 or not. The original *phase3* dataset does not include enough information for further modeling, so it is merged by the unique *applicant_id* from *phase2* dataset, then a dissimilar column *pass* is created indicating whether the corresponding applicants have successfully moved to the final interview. As mentioned earlier, for variable *gender*, the “Prefer not to say” category is filtered out since it is not representative for gender.

Exploratory Plots and Summary Tables

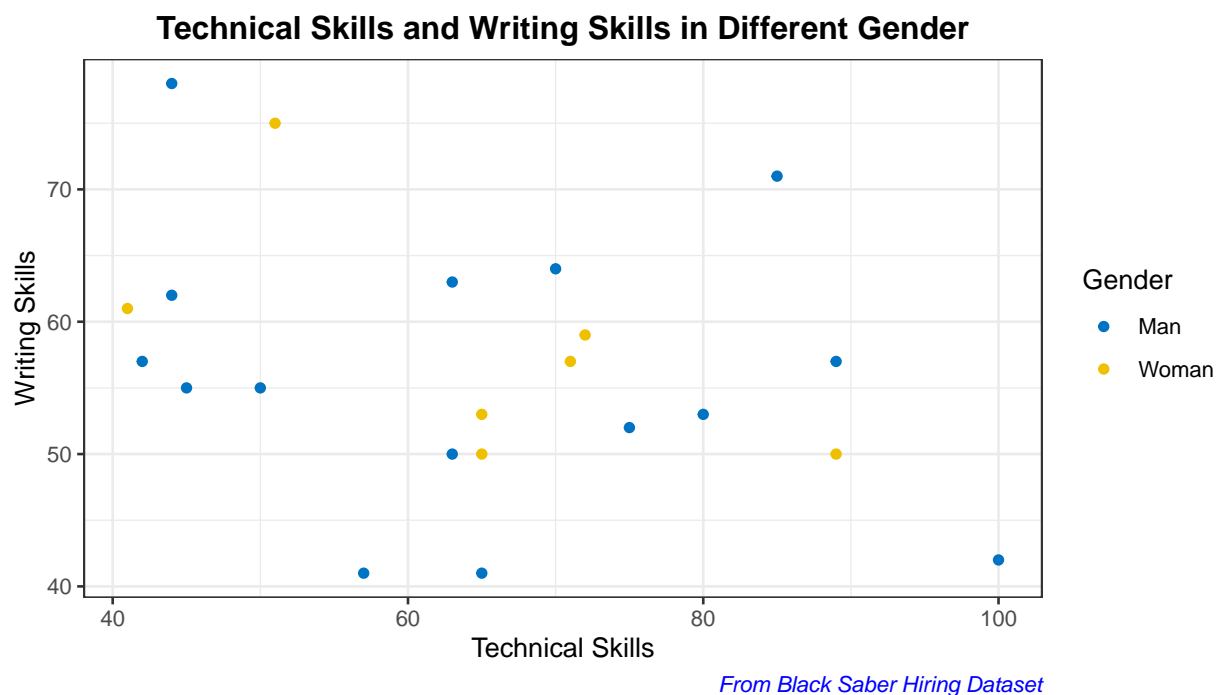


Figure 6: Technical Skills and Writing Skills of Phase 3 Applicants (n = 22)

Figure 6 shows the relationship between technical skills and writing skills for those 22 applicants who successfully passed the first two phases: AI-automated selection. The gender difference is distinguished through different colors. The data points in blue are for men, and in yellow is for women. From Figure 6, there is no such a clear pattern for the scores, and all scores are above 40. Moreover, the scores seem to be similar between different gender.

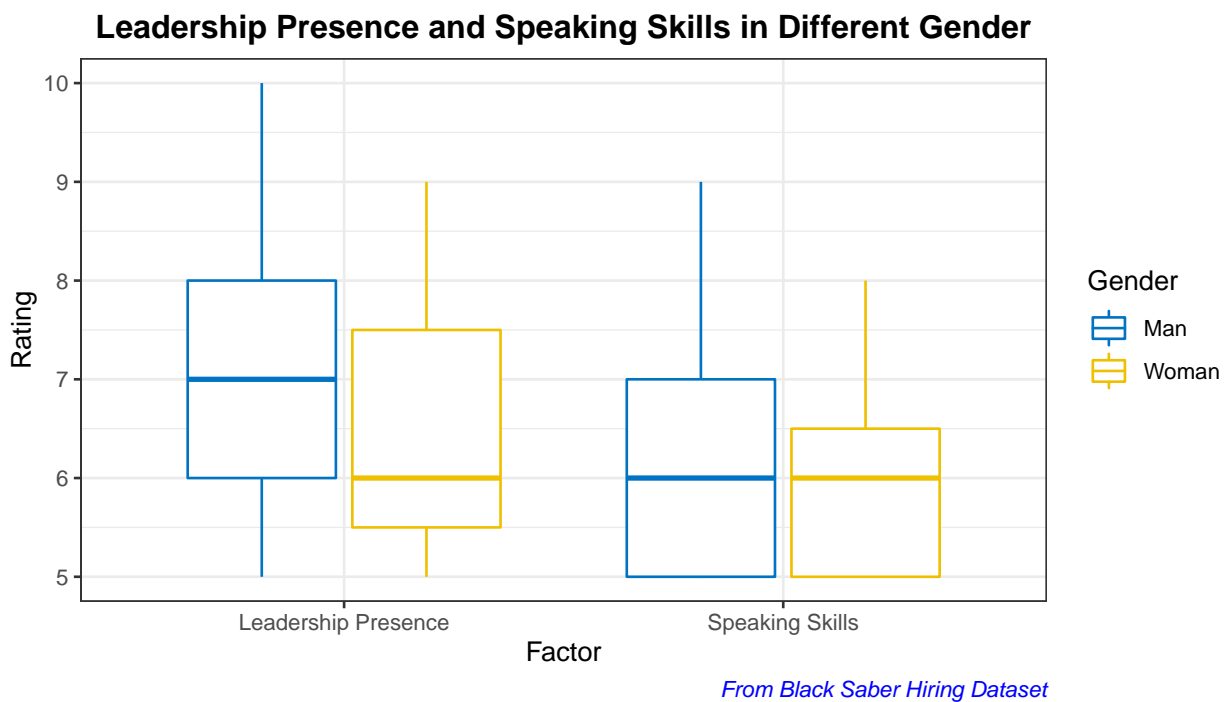


Figure 7: Leadership Presence and Speaking Skills of Phase 3 Applicants ($n = 22$)

Figure 7 shows the rating of leadership presence and speaking skills for the selected 22 applicants to the final interview. The gender difference is distinguished through different colors. The information for men and women is shown in blue boxplots and yellow boxplots, respectively. The rating of speaking skills is similar in men and women while male applicants show a slightly higher rating of leadership presence than women, with the median for men, is about 7 and the median for women is 6.

Table 7: Final Hires' Average Statistics of Various Factors in Different Gender

Gender	Number	Technical Skills	Writing Skills	Leadership Presence	Speaking Skills	Interviewer Rating
Man	8	75.75	56.375	7	6.625	80.812
Woman	2	68.00	55.000	8	5.500	80.250

Table 7 summarizes the cleaned *phase3* data showing the number of applicants that final hire and the average score of technical skills, writing skills, leadership presence, speaking skills, and interviewer rating in different gender. Noticed that there is a big difference in the number of final hires between man and woman. There are eight male applicants hired, while only two female applicants successfully pass the last round. Other than this factor, all other scores are similar in man and woman. Therefore, it is curious that where there is any gender discrimination in the company.

Method

In statistics, logistic regression model is the appropriate regression analysis to conduct when the dependent variable is binary. Logistic regression is a predictive analysis and is used to describe data and to explain the relationship between one dependent binary variable and one or more independent variables.

Since the factors that affect hiring in each phase are interested, it is appropriate to use logistic regression analysis with binary variable *pass* being the response variable of each phase. Therefore, three generalized linear models are fitted from phase to phase to test whether the gender of applicants affects the hiring results and what factors take into account for the selection in each phase. These models are fitted in R using *glm()* function.

Model Equation

- *Model1*: $\text{logit}(\text{pass}_{\text{phase1_to_phase2}}) = \beta_0 + \beta_1 X_{\text{cover_letter}} + \beta_2 X_{\text{cv}} + \beta_3 X_{\text{gpa}} + \beta_4 X_{\text{genderWoman}} + \beta_5 X_{\text{extracurriculars}} + \beta_6 X_{\text{work_experience}} + \beta_7 X_{\text{team_applied_for}} + \epsilon$
- *Model2*: $\text{logit}(\text{pass}_{\text{phase2_to_phase3}}) = \beta_0 + \beta_1 X_{\text{technical_skills}} + \beta_2 X_{\text{writing_skills}} + \beta_3 X_{\text{leadership_presence}} + \beta_4 X_{\text{speaking_skills}} + \beta_5 X_{\text{genderWoman}} + \epsilon$
- *Model3*: $\text{logit}(\text{pass}_{\text{phase3_to_final_hire}}) = \beta_0 + \beta_1 X_{\text{interviewer_rating_1}} + \beta_2 X_{\text{interviewer_rating_2}} + \beta_3 X_{\text{genderWoman}} + \epsilon$

Result

Table 8: Summary Table of Model1

term	estimate	std.error	statistic	p.value
(Intercept)	-158.881	5679.492	-0.028	0.978
cover_letter	61.758	2870.460	0.022	0.983
cv	50.371	4900.648	0.010	0.992
gpa	12.769	3.450	3.702	0.000
genderWoman	1.090	1.036	1.052	0.293
extracurriculars	9.937	2.476	4.014	0.000
work_experience	11.848	2.981	3.974	0.000
team_applied_forSoftware	-1.053	0.930	-1.132	0.258

Table 8 is the summary statistics for *Model1*. Table 8 shows that only GPA, level of extracurricular, and work experience score of applicants are significant factors that affect the result of AI-automated selection from phase 1 to phase 2 since these three factors have a p-value that less than 0.05 significance level indicating the corresponding null hypothesis is rejected. Specifically, gender discrimination does not exist since the factors of gender are not statistically significant (p-value= 0.292756 > 0.05), which means it does not influence the result of passing to phase 2. Moreover, the 95% confidence interval of the estimated coefficient of *Model1* also corresponds to the above results. The 95% confidence interval of *genderWoman* is (-0.8222152, 3.3679995) includes 0, indicating that this factor is not significantly affecting the result of passing. Similarly, for *cover_letter*, and *team_applied_forSoftware*, the 95% confidence interval is (-28.4100330, 1185.5769984) and (-3.0452591, 0.7283961) respectively. Both includes 0, which means cover letter and the team applied in the application do not affect the AI selection result. Thus, AI selection in phase 1 to phase 2 is fair to gender.

Table 9: Summary Table of Model2

term	estimate	std.error	statistic	p.value
(Intercept)	-20.773	3.751	-5.538	0.000
technical_skills	0.081	0.020	3.984	0.000

term	estimate	std.error	statistic	p.value
writing_skills	0.092	0.024	3.868	0.000
leadership_presence	0.896	0.204	4.383	0.000
speaking_skills	0.716	0.168	4.264	0.000
genderWoman	-0.567	0.722	-0.784	0.433

Moving on to phase 2 to phase 3 selection, Table 9 is the summary statistics for *Model2*. Table 9 indicates that the score of technical skills, the score of writing skills, the rating of leadership presence, and speaking skills are essential factors that influence phase 2 to phase 3 AI-automated selection. However, gender does not seem to affect the selection result since the p-value of gender is not significant ($0.43291 > 0.05$), giving us evidence that gender does not relate to the AI selection results. Therefore, AI selection in phase 2 to phase 3 is also fair to gender. The 95% confidence interval results also match with the above outcomes. The 95% confidence interval for *genderWoman* is $(-2.05544971, 0.8225913)$, including 0, justifying that gender is not significantly affecting the selection result from phase 2 to phase 3. The 95% confidence interval for the score of technical skills, the score of writing skills, the rating of leadership presence, and speaking skills do not include 0 indicating that these factors are significant to the AI-selection from phase 2 to phase 3.

Table 10: Summary Table of *Model3*

term	estimate	std.error	statistic	p.value
(Intercept)	-2557.103	1488199.966	-0.002	0.999
interviewer_rating_1	15.064	9039.407	0.002	0.999
interviewer_rating_2	18.939	11079.392	0.002	0.999
genderWoman	-48.296	254384.746	0.000	1.000

Table 10 is the summary statistics for *Model3*. The result of *Model3* shows that the final hiring results are pretty random since the ratings of the first and second interviewer and the factor of gender are not significantly influence the result of the final selection. These three factors all give a p-value that greater than the significance level of 0.05, indicating all of these three factors do not affect the final hiring results, so the final hiring is fair. It is not correlated to the gender of applicants.

Discussion

By applying a linear mixed model using the data from Black Saber Software, we answer the first research question of factors affecting employees' salary while the salary level is unfair in gender. We conclude that *leadership_for_level*, *gender*, *role_seniority* and *productivity* can affect salary, where *gender* and *role_seniority* are more significant than others. By controlling all other factors, the higher the seniority, the more salary employees will get within the same team. Although seniority affects salary level, it is not the causation of gender discrimination. It is very common that high-seniority employees earn more. Thus this is not considered as bias. However, within the same conditions, a woman's salary is significantly lower than a man's salary within the same team. Therefore, the salary process is unfair for men and women in the company.

Next, we address the second research question of any gender discrimination in employees' promotions. After studying the result from the Poisson generalized linear model (GLM), we get the conclusion saying that gender is significantly affecting the promotion of employees in Black Saber Software, then there is possible gender discrimination happening in deciding employees' promotion whereas the work performance is not considered as much as an important factor that influences the promotion.

Finally, we study the third research question of effective factors influencing the AI-automated hiring selection by using logistic regression models. The result shows that gender does not seem to affect the selection result in each phase. The AI-automated selection is fair to gender from phase 1 to phase 2 and from phase 2 to phase 3. Only the GPA, level of extracurricular, and work experience score of applicants are significant factors that affect the result of AI-automated selection from phase 1 to phase 2. Only the score of technical skills, the score of writing skills, the rating of leadership presence, and speaking skills affect the AI-automated selection result from phase 2 to phase 3.

Overall, the potential bias for gender exists in both the salary and promotion process in Black Saber Software, but not in the hiring selection. In general, even though a woman with good working ability can successfully enter Black Saber Software, it is harder for her to promote and get high salaries in the company. Conversely, a man can easily increase his salary and have more chance to promote by just showing a bit of his working ability in the company.

Strengths and Limitations

In this consulting report, three different kinds of models are fitted. This is more appropriate than only fitting one (Maybe a Multivariate Regression model). Because for every process, the potential factors are actually different, in order to get a more accurate result, separate modeling is required. This is our strength in this research. However, there are limitations to the analysis.

First, the data may be biased by itself, decisions of phase 1 and phase 2 in the hiring process are made by AI-automated selection. Notice that there might be algorithmic bias already in these AI systems since they are written by humans. As we don't have the source code, we don't know what standard the algorithm uses. If the standard the program takes is not scientific, the result will also be influenced. For salary, the collected data only records current wage, not total income. If employees get bonuses in each financial quarter, and gender parity kicks in here, only analyzing the salary is not enough. Therefore, more aspects are required for further analysis.

Second, the models that are chosen may not be the best ones, and there are some other models (non-linear models) out there, which may be even more efficient as we check the linear assumptions of data.

For future improvements, if the standard of algorithm is provided, a deeper analysis could be made, which can lead to a better result. And non-linear models could be considered as the assumptions of linear regression (If the data is not linear) are checked.

Consultant Information

Consultant Profiles

Yuchen Cong. Yuchen Cong is a junior consultant with YYDS Limited Co. She specializes in data cleaning, modeling and data analysis . Yuchen earned her Bachelor of Science, Double Major in Mathematics and Statistics, from University of Toronto in 2022.

Jingwen Deng. Jingwen is a senior consultant with YYDS Limited Co. She specializes in data cleaning, data summary and modeling. Jingwen earned his Bachelor of Science, Double Major in Actuarial Science and Statistics, from University of Toronto in 2022.

Yukun Gao. Yukun is a senior consultant with YYDS Limited Co. He specializes in data cleaning and data summary as well as discussions. Yukun earned his Bachelor of Science, Double Major in Economics and Statistics, from University of Toronto in 2022.

Alvin Gong. Alvin is a junior consultant with YYDS Limited Co. He specializes in providing executive word summary, discussions and strength and limitations in the technical report. Alvin earned his Bachelor of Science, Specialist in Statistics Methods and Practice, Economics Major from University of Toronto in 2022.

Code of Ethical Conduct

We are responsible for all the content we provided and we would be willing to have them reviewed by peers.

We will advise clients of any potential or actual conflict between the ethics of statistical analysis and the interests of the client.

We will apply statistical analysis procedures in a professional way without predetermining the outcome.

We will report the limitations of statistical inference and potential sources of error.

We are striving to avoid any procedural and personal bias when possible.

Reference

- Cite R: R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Cite “tidyverse”: Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- Cite “ggpubr”: Alboukadel Kassambara (2020). ggpubr: “ggplot2” Based Publication Ready Plots. R package version 0.4.0. <https://rpkgs.datanovia.com/ggpubr/>
- Cite “kableExtra”: Hao Zhu (2020). kableExtra: Construct Complex Table with “kable” and Pipe Syntax. <http://haozhu233.github.io/kableExtra/>, <https://github.com/haozhu233/kableExtra>.
- Cite “lme4”: Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.
- Cite “lmtest”: Achim Zeileis, Torsten Hothorn (2002). Diagnostic Checking in Regression Relationships. R News 2(3), 7-10. URL <https://CRAN.R-project.org/doc/Rnews/>
- Cite “reshape2”: Hadley Wickham (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL <http://www.jstatsoft.org/v21/i12/>.