Gender bias analysis in the Black Saber Software

Exploring gender bias in salary, promotion, and hiring systems statistically

Report prepared for Black Saber Software by Sunflower Co.

2021-04-21

Contents

Executive summary

Background and Aim

The topic aims to explore whether gender bias exists in the Black Saber wage system, promotion mechanism, and recruit behavior in their company.

This analyst will conduct research from three aspects. Firstly, whether men's wages in Black Saber Software are higher than women's; secondly, whether employees of different genders have bias in promotion; thirdly, whether there is gender bias in the AI recruitment system. Furthermore, this analyst will integrate the comprehensive data of Black Saber Current Employees to conduct research. For the first two problems of salary and promotion, we will consider gender, role seniority, leadership for level, team, and productivity as our research factors. We will analyze the graphs, compare models, analyze the data, and finally draw reliable conclusions. For the third question, there are a total of three phases in the recruitment.Phase1 is completed by the AI system. We will consider the cover letter, cv, GPA, gender, extracurriculars, work_experience, and team_applied_for as our research factors. Phase2 is also carried out by the AI system. We will use technical skills, writing skills, leadership presence, speaking skills, and gender as our research factors. For these two phases, we will conduct chart analysis, build models, and conduct data analysis to reach the final conclusion. Phase3 is two interviewers scoring the interviewers. This phase lacks data support. We will conduct simple research.

The expectation for this analyst is that gender bias exists in the salary and promotion system, but does not exist in the AI recruitment system.

Key Findings

- There is evidence of gender bias in the current company worker's salary. If the gender of the worker is female, she is likely to receive a lower wage than her male co-workers.
- From Table 1.1, the average difference in salary we predicted between males and females is 1834.92.
- There is evidence of gender bias in the current company worker's promotion process. From Table 1.1, the difference in promotion numbers between males and females we predicted is 0.267 on average.
- The experiment did not detect any evidence of gender bias in the AI service of selecting candidates. (see Table 1.2)
- Surprisingly, we find that productivity has a negative relationship with salary for both genders. In other words, being more productive at work leads to a lower wage rate, which is worth to

be noticed.

- Female workers face a disadvantage in the promotion process. Compared to males, more females tend to receive "needs improvement" in the evaluation of leadership presence, which makes it harder for female workers to receive a promotion.
- The gender variable does not seem to have any impact on whether the applicant would pass the first two phases of the interview–which are operated by AI service. (see Table 1.2)

limitation

- There are some limitations in the data set of current employees since it only contains eight variables. Our model fits for salary and promotion were only based on these eight variables. If the data provided more variables such as the highest academic degree obtained, performance in teamwork, etc., our result could be more precise.
- There might be some limitations in our analysis of gender bias in salary. The data used for model fitting is only from the second quarter of 2020. If we could fit models for more recent quarters, we might raise confidence in the results, or have new discoveries.
- In our analysis of gender bias in promotion, the model fit had limitations. There are many unpromoted employees. However, the model cannot distinguish whether they are new employees or not. Data of these unpromoted new employees were used in our analysis, which may lead to somewhat biased conclusions.
- The data of the AI hiring system is not large enough, with only 602 applicants expressing their gender. Reliable research and analysis require a large sample to support its credibility. Thereby, Black Saber Software needs more samples, and the channels accessing them should be wide enough.

Table 1.1: Gender effects in Salary and Promotions

```
##              Question         Gender.effects Significant
## 1         Salary ($)   $1835 lower in Woman         Yes
## 2 Promotion (number)   0.267 lower in Woman         Yes
```

Table 1.2: Gender Proportion of entering each phase

```
##   Phase    Man  Woman
## 1     I 48.34% 51.66%
## 2    II 48.82% 51.18%
## 3   III  68.18% 31.82%
```

Technical report

Introduction

The aim of the present research was to examine the raising concern about potential bias in Black Saber Software's hiring and remuneration process. Based on the fact that workers of the company have complained about the issue of gender bias, the study set out to investigate whether the company's hiring, promotion, and salary processes were all fair. In other words, whether all processes were only based on people's talent and value to the company. Moreover, the study focused on Phase I and Phase II of the hiring process, where the company used AI service to screen applications. The study aimed to determine if the AI process of selecting candidates was fair and without any bias. Furthermore, the data was given based on current employees of the company. Meanwhile, the data was not collected on ethnicity/race.

Research questions

We were interested in the following questions:

- *Whether the salary paid to the workers was fair and only based on talent and value to the company?*
- *Whether the promotion process of the company workers was fair and only based on talent and value to the company?*
- *Whether the AI service of selecting applicants had no bias?*

1. Whether the salary paid to the workers was fair and only based on talent and value to the company?

Data

We utilized the data about the promotion and salary of the employee provided by the Black Saber Software. This data set provides the gender of the employee, and during the tenure, their leadership performance, position records, and the team to which they belong. There are 6906 employee records in total. The data is recorded quarterly, from the second quarter of 2013 to the end of 2020.

The purpose is to find out whether there is gender bias in the salary of the company, thereby we needed the data about each employee's salary and the variables which could affect his/her salary. Then, we choose the recent data from last year for our analysis. Since we only need one record per employee, we randomly select the data in the second quarter of 2020.

Next, we changed the response variable of salary from a character to a numerical variable, in order to better visualize the data. Then, we used the ggplot package in R to create figures about the dataset.

Below are figures about the distribution of gender and salary in Black Saber Software.

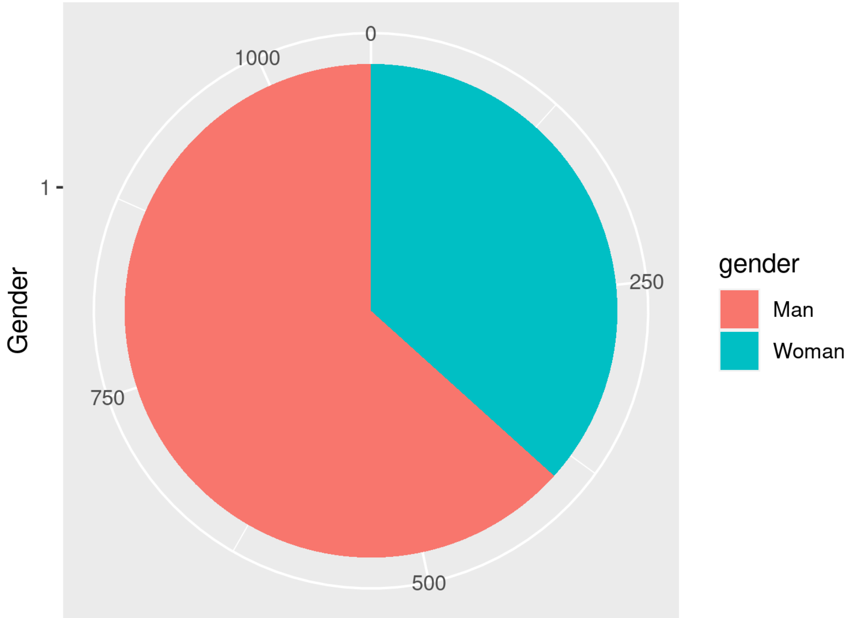

Figure 1: Gender distribution in Black Saber Software

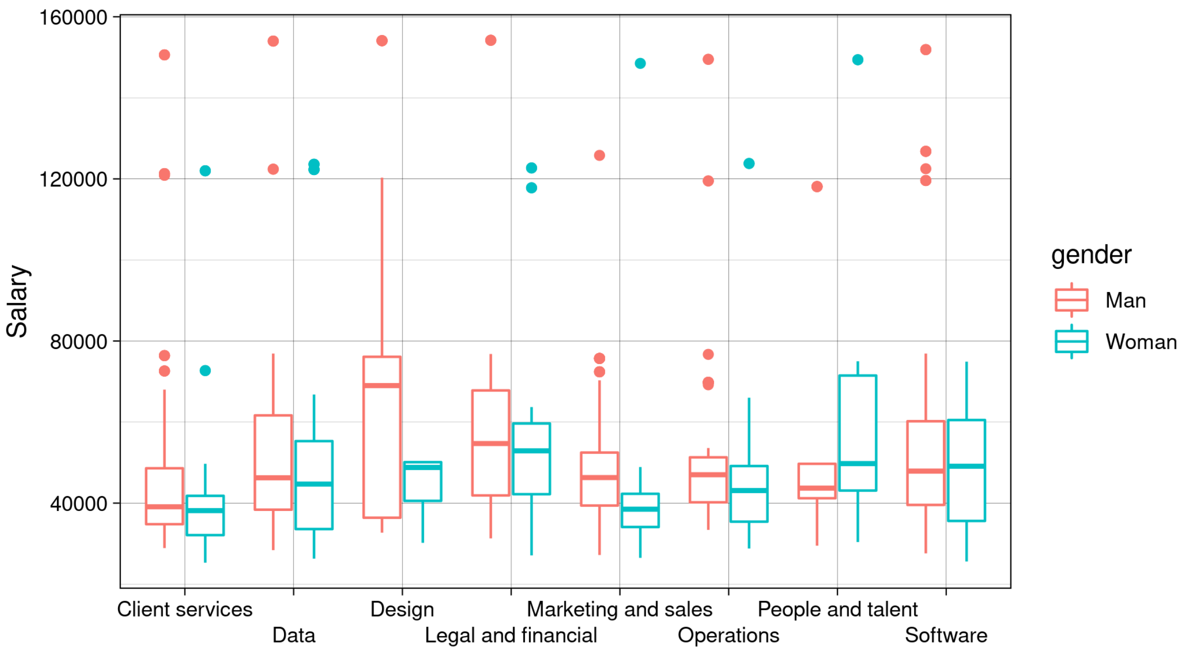

Figure 2: Distribution of team and salary in Black Saber Software

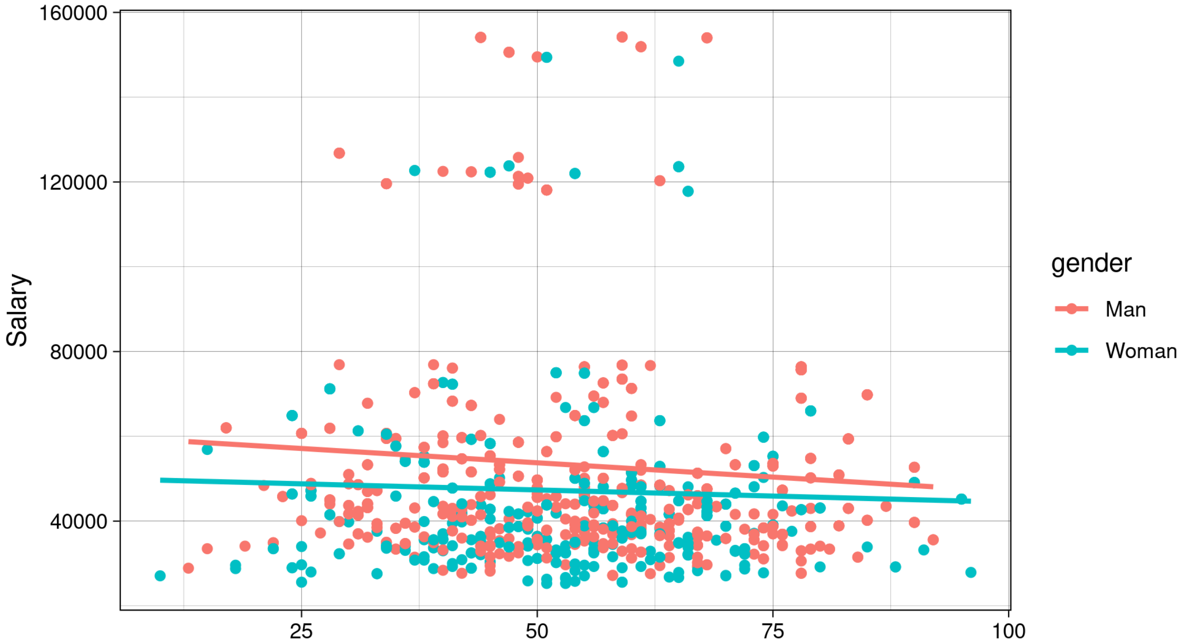

Figure 3: Distribution of productivity and salary in Black Saber

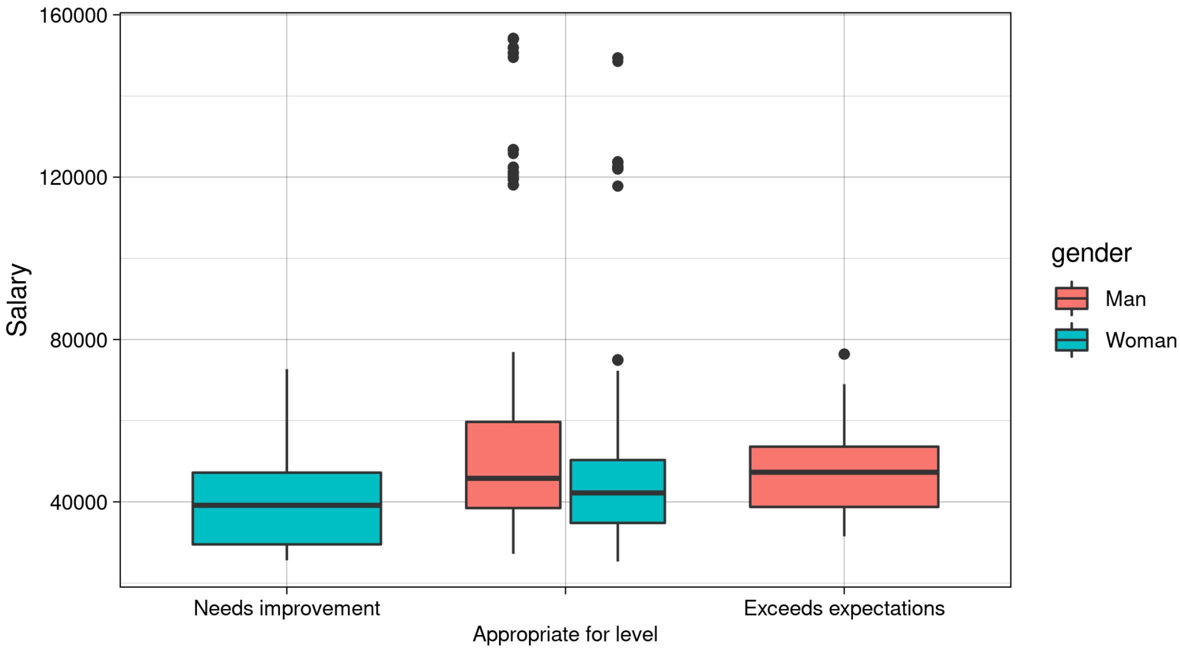

Figure 4: Distribution of leadership and salary in Black Saber Software

Figure 5: Distribution of position and salary in Black Saber Software

As can be seen in Figure 1, in general, more than half of the company's employees are males. From figure 2, consider all the teams in the company, the median salary of males is much higher than that of females in the design team. In client services, data, legal and financial, marking and sales, and operations teams, the median salary of men is a bit higher than that of women. While the median salary of females is slightly higher than that of males in people and talent, and software teams. Moreover, figure 3 is a scatterplot showing the distribution of productivity and salary, and we obtained two best-fitted lines of the data for each gender. It is obvious that, for both genders, productivity is reversely proportional to salary, and male employees averagely receive more salaries than female employees. Besides, from figure 5, we can see that males are paid more than females in almost all positions except the entry-level.

Method

Regression analysis is a statistical method investigating the functional relationship between dependent variables and the response. The simple linear regression can be used to find the statistical linear relationship between the dependent variables and the response. However, if different measurements of one observation are correlated, we prefer a linear mixed model. Compared with the simple linear model, the linear mixed model allows random effects. This could describe the correlation between different measurements of the same observation.

As stated above, the purpose is to find out whether there is gender bias in the company's salary, the response variable is the salary and it is a numeric variable. The explanatory variables include the employee's gender, the role seniority, and the level of leadership, and the team the employee works in. We plan to find out which factors could have significant effects on salary, and whether gender is included in these significant factors.

Thus, we considered the linear regression analysis. Considering that the salary would vary between teams, the data are not actually independent. We treated the team as a grouping unit. Thereby, rather than the simple linear regression, we planned to fit a linear mixed model with a random effect of the team.

We first fitted a linear mixed model by R. The fixed effects are the employee's gender, the role seniority, and the level of leadership, while the random effect is the employee's team. After fitting a full model with all variables, we used the likelihood ratio test to detect whether certain factors would be removed.

The 95% confidence interval provides a range of possible values for the estimated coefficients, which indicates that we have a 95% chance that the true value is included in this interval. If the confident interval of a coefficient contains 0, then the variable is not significant to be included in the model. We checked the confidence intervals of the coefficients of fixed effects. If the confidence interval for a given estimated coefficient contains 0, it means such a fixed effect is not significant. Therefore, we could detect whether it can be removed from the model.

The likelihood ratio test could be applied to compare different models (the model and its reduced model). The null hypothesis is that we should choose the simpler model, while the alternative hypothesis is to choose the full model. Given that the null hypothesis is true, the P-value is the probability of obtaining at least or more extreme results than the observed value. If the p-value we get is less than 0.05, it means we have strong evidence against the null hypothesis and should accept the alternative hypothesis. Therefore, the factors can not be reduced and the full model is better. Otherwise, a p-value greater than 0.05 indicates the simple model is better. Using likelihood ratio test could remove useless terms from the model which are not significant on the salary. Therefore, we can build an accurate linear mixed model which determines the factors that have significant effects on the company's salary.

If gender is included in the final model, and the confidence interval of its estimated coefficient does not contain 0, then it implies that gender is a significant effect on the company's salary. This means, there is a gender bias in the company's salary.

Using R as the tool, our final linear mixed model for salary:

$$lmer(salary \sim (1|team) + leadership\_for\_level + role\_seniority + productivity)$$

Since gender has two levels: "Man" and "Woman", it is a categorical variable. 1 represents male and 0 represents female in the model. The estimated coefficient of gender in the final model represents the average difference in salary between males and females.

Result

We obtained the final model in R.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## salary ~ gender + (1 | team) + leadership_for_level + role_seniority +
##     productivity
##    Data: datasalary
##
## REML criterion at convergence: 20343.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.8888 -0.7721  0.0616  0.7458  2.2863
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  team     (Intercept)  5461879 2337
##  Residual             11796271 3435
## Number of obs: 1072, groups:  team, 8
##
## Fixed effects:
##                                             Estimate Std. Error  t value
## (Intercept)                               122239.574   1198.746  101.973
## genderWoman                                -1834.915    243.429   -7.538
## leadership_for_levelAppropriate for level    804.798    679.227    1.185
## leadership_for_levelExceeds expectations    1269.129    860.625    1.475
## role_seniorityEntry-level                 -89851.911    621.351 -144.607
## role_seniorityJunior I                    -84970.330    595.369 -142.719
## role_seniorityJunior II                   -82402.288    622.924 -132.283
## role_seniorityManager                     -48713.311    654.194  -74.463
## role_senioritySenior I                    -76877.180    600.456 -128.031
## role_senioritySenior II                   -71158.538    602.619 -118.082
## role_senioritySenior III                  -65919.737    603.270 -109.271
## role_seniorityVice president               30533.953    913.702   33.418
```

```
## productivity                                   -22.442      7.672    -2.925
```

Table 1: Estimated Coefficients of the salary model

Fixed effects	Estimate	Std. Error	t value
(Intercept)	122239.574	1198.746	101.973
genderWoman	-1834.915	243.429	-7.538
leadership_for_levelAppropriate for level	804.798	679.227	1.185
leadership_for_levelExceeds expectations	1269.129	860.625	1.475
role_seniorityEntry-level	-89851.911	621.351	-144.607
role_seniorityJunior I	-84970.330	595.369	-142.719
role_seniorityJunior II	-82402.288	622.924	-132.283
role_seniorityManager	-48713.311	654.194	-74.463
role_senioritySenior I	-76877.180	600.456	-128.031
role_senioritySenior II	-71158.538	602.619	-118.082
role_senioritySenior III	-65919.737	603.270	-109.271
role_seniorityVice president	30533.953	913.702	33.418
productivity	-22.442	7.672	-2.925

Table 2: 95% Confidence Interval of the salary model

Fixed effects	Lower bound	Upper bound
	2.5 %	97.5 %
.sig01	1404.24795	3953.409754
.sigma	3275.04024	3565.567812
(Intercept)	119900.61830	124578.900724
genderWoman	-2310.44067	-1360.747726
leadership_for_levelAppropriate for level	-517.09408	2133.219560
leadership_for_levelExceeds expectations	-404.33994	2954.479828
role_seniorityEntry-level	-91062.90576	-88638.834965
role_seniorityJunior I	-86131.40464	-83808.750979
role_seniorityJunior II	-83616.84660	-81186.675723
role_seniorityManager	-49991.22175	-47438.893258
role_senioritySenior I	-78047.67620	-75705.162165
role_senioritySenior II	-72334.82294	-69983.815697
role_senioritySenior III	-67095.99568	-64742.483975
role_seniorityVice president	28751.12999	32315.626481
productivity	-37.45488	-7.517441

From Table 1, by observing the model and its coefficients, the estimated coefficients of genderWoman and productivity are -1834.92 and -22.44 respectively. This means the average difference in salary between males and females is 1834.92. Also, for every unit increase in the productivity of the employee, the salary decreases by 22.4 on average.

From Table 2, the confident interval of the estimated coefficient for genderWoman is from -2310.44 to -1360.747726 which does not include 0. This means gender is a significant fixed effect on the salary. Also, Table 2 states that the confidence intervals of the estimated coefficients for other factors do not include 0 also. This means the employee's gender, level of leadership, role seniority, and productivity are all significant assessing salary.

Overall, the average salary of males is higher than that of females in this company. Gender bias does exist.

2. Whether the promotion process of the company workers was fair and only based on talent and value to the company?

Data

We have discussed the information that the dataset provided before. To determine whether there is a gender bias on the promotion, we considered the response as the promotion numbers of each employee. The main exploratory variable is gender. We wanted to detect whether gender is significant in the promotion numbers. Therefore, we focus on the variable "role seniority", which describes each employee's position in each quarter of his/her tenure. We first calculated the number of positions per employee, then one less than the number of positions is the number of promotions.

Below are figures about the distribution of gender and promotions.

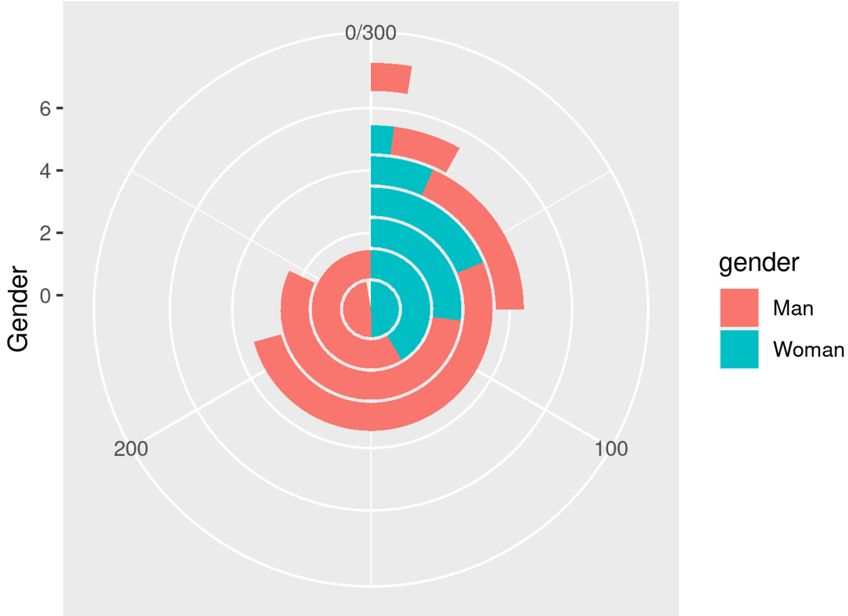

Figure 6: Distribution of promotion numbers in Black Saber Software

Figure 7: Distribution of position and promotion in Black Saber Software

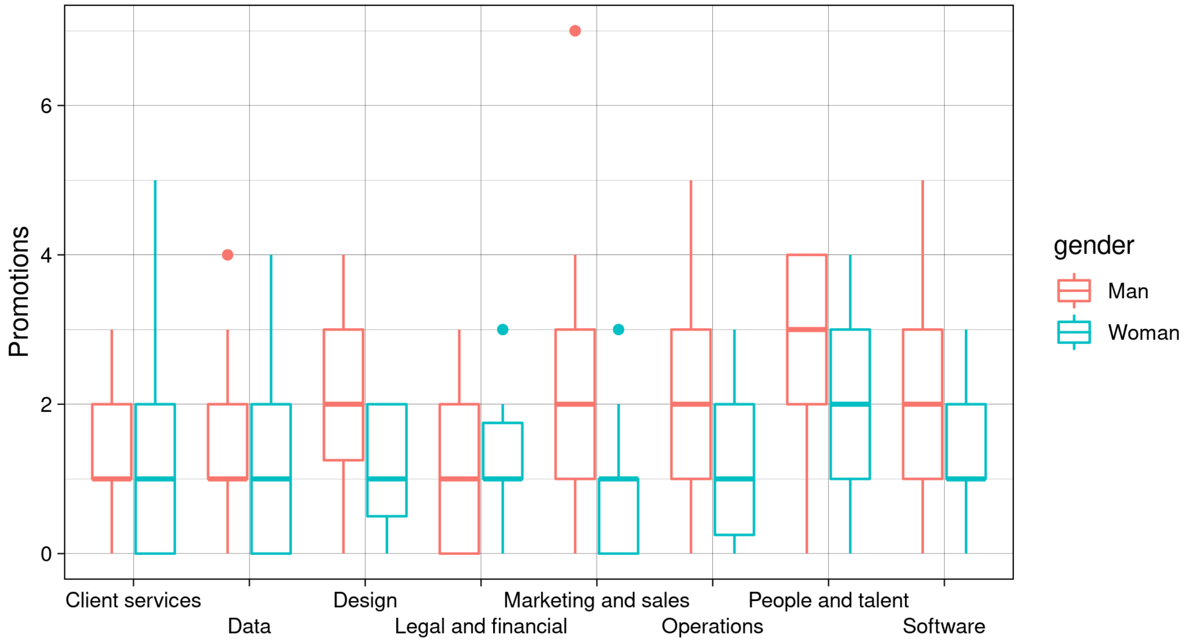

Figure 8: Distribution of team and number of promotions in Black Saber Software

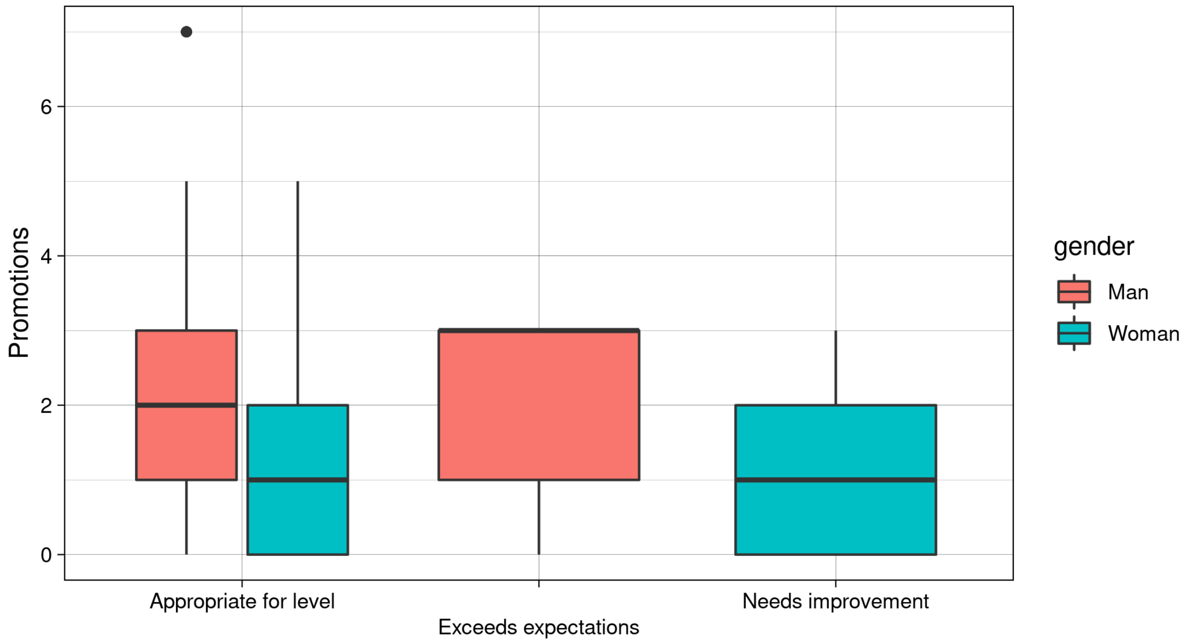

Figure 9: Distribution of leadership and number of promotions in Black Saber Software

Figure 6 indicates an overall view of promotions and gender. The number of promotions among males is higher than that of females in general. As can be seen in figure 7, within each position, males are generally paid more than females. From figure 8, in the design, software, people and talent, operations marking, and sales teams, the median salary for males is higher than that of females. In legal and data teams, the median salaries for males and females are almost the same.

Method

Regression analysis is a statistical method that determines the relationships between explanatory variables and the response. When the response is a count variable and follows the Poisson distribution, a generalized linear model with Poisson regression could be applied.

Our purpose is to find out whether there is gender bias in the company's promotions, so we want to detect whether gender is significantly related to the promotion numbers. We plan to perform the regression analysis. We wanted to build an accurate statistical model which determines the factors that have significant effects on the company's promotions.

We know that the response is the promotion number of each employee, while the explanatory variables are employees' gender, their leadership levels, and working teams. Considering that promotion number is a count variable, we fitted a generalized linear model with poisson regression.

However, employees who have worked longer are likely to receive more promotions, while employees who have been recently hired will be promoted less often. Working time could be a strong confounder of the relationship between gender and promotions. Thus, we need to correct this issue by adding an offset term for the working time in the generalized linear model.

We first fitted a full model using the function GLM in R with all variables: each employee's salary, gender, which team the employee works on, the employee's level of leadership, and his/her productivity and we treated working time as an offset term.

Next, we found out the 95% confidence intervals of the coefficients. The 95% confidence interval provides a range of possible values for the estimated coefficients, indicating that we have a 95% chance of including the true value. If 0 is included in the confident interval, then the variable is not significant. If it contains 0, it indicates that such a factor is not significant. Thereby, we wanted to detect whether it can be removed from the model.

We compared it with the reduced model by the likelihood ratio test. The null hypothesis is that the simpler model is better, while the alternative hypothesis is that the full model is better. A p-value that is larger than 0.05 means we do not have strong evidence against the alternative hypothesis. Thus, we should use the simple model. Otherwise, a p-value that is smaller than 0.05 indicates the full model is better.

If gender is included in the final model and the confidence interval of its estimated coefficient does not contain 0, then gender is significant on the salary. Besides, we can also compare the p-value of its estimated coefficient. The null hypothesis is that the estimated coefficient of gender is significantly 0, while the alternative hypothesis is the estimated coefficient of gender is not significantly 0. The p-value is the probability of obtaining at least or more extreme results than we observed when the null hypothesis is true. If it is smaller than 0.05, then we have strong evidence against the null hypothesis, which means, the estimated coefficient of gender is not significantly 0. There is gender bias in the company's promotions.

Using R as the tool, our generalized linear model for promotion is:

$$glm(promotion \sim gender + salary, offset = log(workingtime), family = poisson)$$

Since gender has two levels: "Man" and "Woman", it is a categorical variable. 1 represents male and 0 represents female in the model. The estimated coefficient of gender in the final model is the average difference in promotion numbers between males and females.

Result

The final model is:

$$\tilde{promotion} = -2.26 - 2.67 \times 10^{-1} genderWoman + 2.53 \times 10^{-6} salary$$

```
##
## Call:
## glm(formula = promotion ~ gender + salary, family = poisson,
##     data = dataQ2, offset = log(working_time))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5864  -0.6774  -0.1014   0.3431   1.5821
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.259e+00  5.586e-02 -40.430  < 2e-16 ***
## genderWoman -2.670e-01  5.039e-02  -5.300 1.16e-07 ***
## salary       2.534e-06  7.832e-07   3.236  0.00121 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 702.71  on 1156  degrees of freedom
## Residual deviance: 660.24  on 1154  degrees of freedom
```

```
## AIC: 2902.4
##
## Number of Fisher Scoring iterations: 5
```

Table 3: Estimated Coefficients of the promotion model

	Estimated coefficient	P-value
(Intercept)	-2.259e+00	< 2e-16
genderWoman	-2.670e-01	1.16e-07
salary	2.534e-06	0.00121

Table 4: 95% Confidence Interval of the promotion model

Variable	Lower bound	Upper bound
	2.5 %	97.5 %
(Intercept)	-2.368118e+00	-2.149097e+00
genderWoman	-3.664822e-01	-1.689174e-01
salary	9.785413e-07	4.049412e-06

As seen in Table 3, the estimated coefficient of genderWoman in our model is -0.267. This implies that the difference in promotion numbers between males and females is -0.267 on average. Besides, from Table 4, the confidence interval of the estimated coefficient for the genderWoman ranges from -0.366 to -0.169, which excludes 0. It indicates that gender is a significant variable affecting promotion.

Thereby, we conclude that on average, males are promoted more often than females in this company. Gender bias does exist in promotions.

3. Whether the AI service of selecting applicants had no bias?

Predict the Final result by general gender proportion

In the initial time of the phase 1, contains a total of 602 interviewees. Through the general pie chart, we can see that 311 women and 291 men were selected to enter phase 1, accounting for 51.66% and 48.34% respectively. From this data, we can assume that there is no gender bias throughout , the female and male proportion in every phases should be similar to this figure.

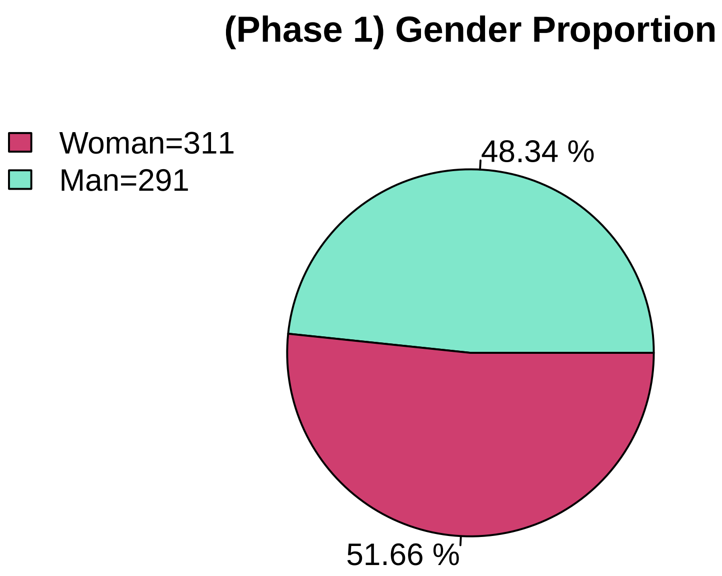

Figure 10: (Phase 1) Gender Proportion

Predict the Phase1 result by Phase2 Gender Proportion

The original phase1.csv file contains a total of 602 interviewees, and the 8 explanatory variables are, applican_id, cover_letter, cv, gpa, gender, extracurriculars, work_experience, team_applied_for. Each row contains data provided by the interviewer to Black Saber Software. In this analysis, the response of interest is that the AI system's selection and filtering of interviewees has nothing to do with gender, but focuses on students' gpa, extracurriculars, and work_experience.

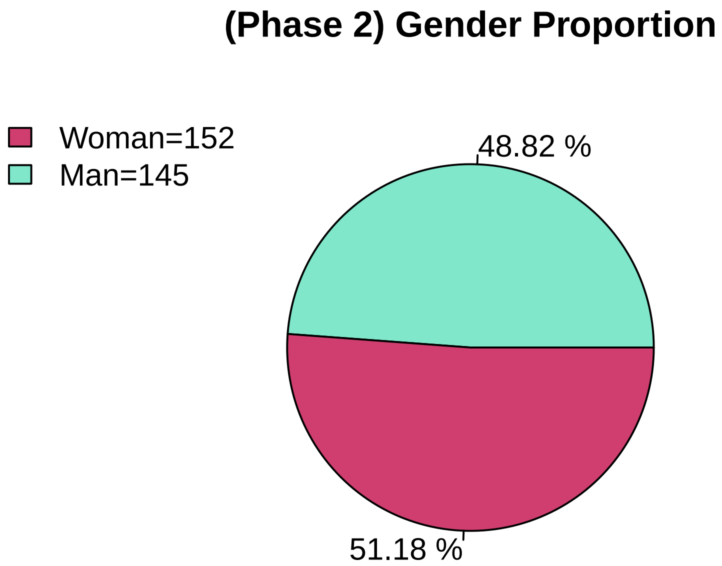

Figure 11: (Phase 2) Gender Proportion

We use one of the most direct methods to compare the proportion of males and females selected by AI to enter phase2 through a pie chart, and we can predict whether there is gender bias in the AI system. Through the chart, we can see that 152 women and 145 men were selected to enter phase 2, accounting for 51.18% and 48.82% respectively. From this data, we can predict that there is no gender bias in phase1. But this method is not scientific and rigorous enough. We still need to consider cover_letter, cv, gpa, gender, extracurriculars, work_experience, team_applied_for in phase1. So we have established a logistic regression model to help us better analyze the results of phase1.

Fit Phase1 GLM Model

We use the applicant_id of phase2 as a reference. If the ID appears in phase2, we will always mark it as "1" in phase1 which means they pass the phase1, otherwise "0". This also allows us to add the variable "pass" in phase1. And we filter the applicant in phase1, the interviewers who do not want to reveal their gender will be removed. Base on the variables that phase1 provided and the variable "pass", we fit a generalized linear model, the distribution is binomial. Base on this model, we can study whether cover_letter, cv, gpa, gender, extracurriculars, work_experience, and team_applied_for is affected the AI determination, which is the pass. After a summary of our model, we can observe the P-value of each variable obtained from the r-output, and it can be found that in phase1, only gpa, extracurriculars and work_experience are significant, their P-value are respectively 0.000214, 5.97e-05, 7.07e-05. There is strong evidence to show that they have an impact on the pass. The p-value of gender is 0.293, which means it is not significant and has no effect on the AI system.

```
##
## Call:
## glm(formula = pass ~ team_applied_for + cover_letter + cv + gpa +
##     gender + +extracurriculars + work_experience, family = binomial,
##     data = phase1)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -4.045   0.000   0.000   0.000   1.284
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -158.8808  5679.4915  -0.028 0.977683
## team_applied_forSoftware   -1.0528     0.9302  -1.132 0.257735
## cover_letter               61.7582  2870.4600   0.022 0.982835
## cv                         50.3709  4900.6475   0.010 0.991799
## gpa                        12.7695     3.4498   3.702 0.000214 ***
## genderWoman                 1.0901     1.0361   1.052 0.292756
## extracurriculars            9.9372     2.4756   4.014 5.97e-05 ***
## work_experience            11.8482     2.9814   3.974 7.07e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 834.443  on 601  degrees of freedom
## Residual deviance:  33.551  on 594  degrees of freedom
## AIC: 49.551
##
## Number of Fisher Scoring iterations: 22
```

Predict the Phase2 result by Phase3 Gender Proportion

Again, we first use a pie chart to determine whether the AI service of selecting applicants is fair. Based on the pie chart, we see that there are 68.18% of male candidates and 31.82% of female candidates passed the second round of the AI selecting process. The percentages of males and females who passed Phase II and entered Phase III are not equal. The chart seems to demonstrate a gender bias in the AI selecting process.

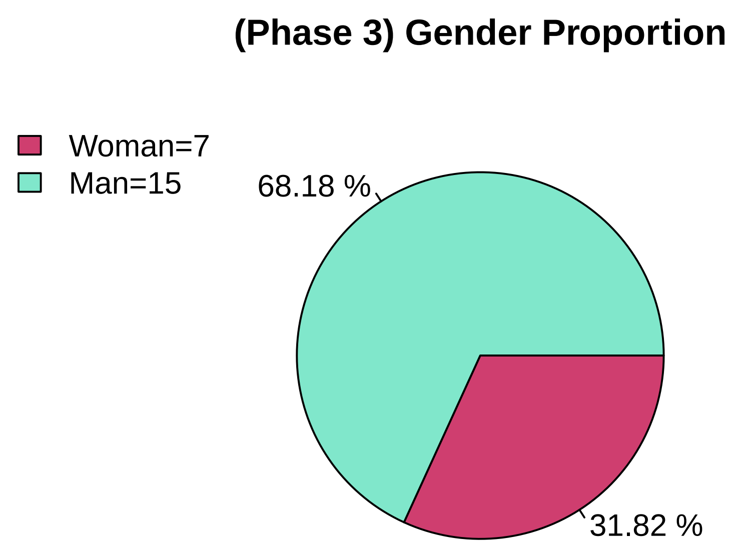

Figure 12: (Phase 3) Gender Proportion

Fit Phase2 GLM Model

In order to investigate if our assumption is correct, we construct a generalized linear model with a binomial distribution. Except for the seven variables being examined in phase I, we add four additional terms to the data of Phase II for evaluation-technical_skills, writing_skills, leadership_presence, and speaking_skills. Meanwhile, we build a GLM model using "passing Phase II or not" as the response variable. At the same time, using technical_skills, writing_skills, leadership_presence, speaking_skills, and gender as the explanatory variables.

Moreover, P-value for each variable is obtained from summarizing the model using R. From the R outputs, we see that P-values for all variables except gender are less than 0.05, while the P-value for gender is 0.43291. In other words, all variables are significant except gender. There is no evidence that the gender variable has an impact on whether the applicant would pass Phase II or not. The result contradicts the result obtained from the pie chart. There is no gender bias in Phase II of the AI selecting process.

```
##
## Call:
## glm(formula = pass ~ technical_skills + writing_skills + leadership_presence +
##     speaking_skills + gender, family = binomial, data = phase2)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.04382  -0.16974  -0.04687  -0.01151   2.94839
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -20.77341    3.75121  -5.538 3.06e-08 ***
## technical_skills      0.08106    0.02035   3.984 6.76e-05 ***
## writing_skills        0.09222    0.02384   3.868  0.00011 ***
## leadership_presence   0.89593    0.20439   4.383 1.17e-05 ***
## speaking_skills       0.71556    0.16781   4.264 2.01e-05 ***
## genderWoman          -0.56658    0.72247  -0.784  0.43291
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 156.847  on 296  degrees of freedom
## Residual deviance:  69.247  on 291  degrees of freedom
## AIC: 81.247
##
## Number of Fisher Scoring iterations: 8
```

Fit Phase3 GLM Model

However, we are also curious about whether the last round of the interview-which is conducted by two human interviewers-is fair or not. Again, we construct a generalized linear model with a binomial distribution. This time, the response variable for the GLM model is "passing Phase III or not". Meanwhile, we use the first interviewer's rating, the second interviewer's rating, the team applied for, and gender as the explanatory variables.

From the R output, we see that P-values for all variables are greater than 0.05. In other words, all variables are not significant. The result falls under our expectation since the ratings given by the interviewers are random and should not have any impact on whether the applicant passes Phase III or not.

```
##
## Call:
## glm(formula = pass ~ interviewer_rating_1 + interviewer_rating_2 +
##     gender + team_applied_for, family = binomial, data = phase3)
##
## Deviance Residuals:
##        Min          1Q      Median          3Q         Max
## -3.541e-05  -2.100e-08  -2.100e-08   2.100e-08   4.104e-05
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -2167.54 2465765.65  -0.001    0.999
## interviewer_rating_1         14.06   10020.59   0.001    0.999
## interviewer_rating_2         14.14    9795.35   0.001    0.999
## genderWoman                 -19.62  351903.15   0.000    1.000
## team_applied_forSoftware     37.14 1969321.44   0.000    1.000
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3.0316e+01  on 21  degrees of freedom
## Residual deviance: 3.6008e-09  on 17  degrees of freedom
## AIC: 10
##
## Number of Fisher Scoring iterations: 25
```

Discussion

Our fitted model for assessing salary tells that team is a random effect which means that being in different team affects the salary received, although factors like leadership and role seniority, which indicates talent and value, have positive relations with salary, being a woman in gender is very likely to receive a lower salary. Moreover, for productivity, we usually would regard greater productivity as a sign of receiving more salary, but surprisingly our results indicate that possessing greater productivity would likely lead to a lower salary.

Besides, apart from salary, gender is the only other factor that would significantly affect the promotion, and being a woman in gender could lead to a lower probability of promotion. Gender bias also exists.

Salary and promotion are expected to be assessed only based on an employee's talent and value, for example, factors like productivity and leadership, etc., apart from these, being in different teams might also lead to varied salaries and chances for promotion which is reasonable. However, based on the result of the data analysis, we detect some evidence for gender bias, being a woman in gender could lead to a necessarily lower salary and less chance for promotion. We also find in surprise that for productivity, we usually would regard greater productivity as a sign of receiving more salary, however, our results indicate that possessing greater productivity would likely lead to lower salary for employees.

When we were over-viewing the data provided before analysis, we actually have discovered that all employees receiving "needs improvement" for assessment of their leadership for level are women and all employees receiving "exceeds expectations" are men (Figure 4). Also by plotting the relationship between salary and productivity using linear regression (Figure 3), we found that not only male employees' salaries are averagely higher than that of female employees, but also productivity is in negative relation to salary for both genders. Our results later confirm what we observed from the data overview.

The results are concerning because an employee's gender is not supposed to be associated with their salary and promotion. And in common sense more productive employee is supposed to receive more. There might be some limitations of the results of this analysis, as the data used for model fitting is only from the second quarter of 2020, if we could fit models for more quarters that are more recent, we might be able to raise confidence in the results we found, or have some new discoveries. Despite these limitations, we still recommend you to pay attention to the assessment of salary and promotion, to ensure that all employees can have equal opportunities in their careers and in a very real sense, to receive fewer complaints from employees.

Moreover, the result we expect is that there is no gender bias in AI systems. According to the comparison of various explanatory variables in phase1, we can find that in phase1, the AI system pays more attention to gpa, extracurriculars and work_experience, and we can determine that there is no gender bias in phase1. In phase2, we found that in the "pass" crowd in phase 2, something unexpected happened. That is, the proportion of men who passed phase 2 was 68.18%, while the proportion of women was 31.82%. . This violates the assumption we made for the AI system at the beginning.

We then fitted a GLM model to it and found that this is based on technical_skills, writing_skills, leadership_presence, and speaking_skills as important indicators, not gender. Out of curiosity, we fit a GLM model to phase3. After comparing interviewer_rating_1, interviewer_rating_2, gender, and team_applied_for, we can find from r-output that the result of p-value is equal to or close to 1, which means this is a completely random relationship. And this I think is magical. I think the interview in phase3 is more about the mood or preferences of the interviewer. None of the GLM models we listed for the three phases found any indication of gender as an important indicator, so we can be sure that there is no gender bias in this hiring process.

Strengths and limitations

Under the data that Black Saber Software provided, we have a reliable data set. For the strength of our work, we can provide better strategies for personnel rewards and salaries to Black Saber Software. This report can point out that they have defects in their salary and personnel recruitment mechanism. We can find many things that are bad for the company. Whether in a different team or different role seniority, men's salaries are generally higher than women's salaries. This is a bad phenomenon and will cause anger among employees. Society attaches great importance to the equal relationship between men and women. If the company is found to have such behavior, it will have a huge impact. Surprisingly, employee salary is inversely proportional to productivity. If employees can obtain these data and analyze it, it will be a terrible public relations crisis. Whether in public opinion or within the company, this will cause a major loss. When Black Saber Software obtained our report, it was able to accurately improve the problem and make adjustments in the shortest time.

However, there are some limitations in our analysis.

The results are concerning because an employee's gender is not supposed to be associated with their salary and promotion. And in common sense more productive employee is supposed to receive more. There might be some limitations of the results of this analysis, as the data used for model fitting is only from the second quarter of 2020, if we could fit models for more quarters that are more recent, we might be able to raise confidence in the results we found, or have some new discoveries. Moreover, if the provided data could include more variables such as the highest academic degree obtained, performance in teamwork, etc., our result could be more precise, as observed in the model for promotion, the number of affecting factors is relatively small and there might be other potential factors having an influence. Despite these limitations, we still recommend you to pay attention to the assessment of salary and promotion, to ensure that all employees can have equal opportunities in their careers and in a very real sense, to receive fewer complaints from employees.

The data set contains eight variables. Therefore, the model fit was based on these variables only. This means that significant factors associated with promotion were only selected from these eight variables. But there probably be other factors related to job promotion that are not included in this data set, such as the personal style of the boss in each team or the personality of the employee. In addition, our model has certain limitations. For example, there are many unpromoted employees with a promotion number of 0. However, in our model, we cannot distinguish whether they are new employees or not. Data of these unpromoted new employees were used in our analysis, which may lead to somewhat biased conclusions.

There are also certain limitations in our analysis of AI services. For the strength of our work, we can provide better strategies for personnel rewards and salaries to Black Saber Software. This report can point out that they have defects in their salary and personnel recruitment mechanism. We can find many things that are bad for the company. Whether in a different team or different role seniority, men's salaries are generally higher than women's salaries. This is a bad phenomenon and will cause anger among employees. Society attaches great importance to the equal relationship between men and women. If the company is found to have such behavior, it will have a huge impact. Surprisingly, employee salary is inversely proportional to productivity. If employees can obtain these data and analyze it, it will be a terrible public relations crisis. Whether in public opinion or within the company, this will cause a major loss. When Black Saber Software obtained our report, it was able to accurately improve the problem and make adjustments in the shortest time.

Furthermore, judging from the results, there is no bias in Black Saber Software's AI hiring system. But in terms of data, the size of the sample is not large enough, with only 602 applicants who expressed their true gender. Reliable research and analysis require a large number of samples to support its credibility. Black Saber Software needs more samples, and the channels for obtaining samples should be wide enough. According to the actual situation, most employees in the IT industry are male, and we do not rule out the existence of actual bias. But through the above model, we can actually trust the "Strict Interviewer" (AI) of Black Saber Software, which does not have a gender bias in any phase1 and phase2. However, AI has loopholes in the process of filtering and selecting applicants, and the AI system cannot accurately measure the social communication skills of applicants. This is an important part, whether it is communicating with the team or communicating with the customer, this is an essential skill. Considering that Black Saber Software is a software company, the emphasis on teamwork is necessary.

Consultant information

Consultant profiles

Decheng Luo. Decheng Luo is a fourth-year student at the University of Toronto. He specializes in data analysis and data visualization. Decheng earned his Bachelor of Science this year, Specialist in Statistics, from the University of Toronto in 2021. He wants to travel the world right now.

Yuze Kang. Yuze is a third-year university student. She specializes in reproducible data analysis. Yuze majors in Statistics and Cognitive Sciences and is pursuing a degree of Bachelor of Science in University of Toronto.

Yifan Xu. Yifan is a third-year undergraduate student at the University of Toronto. She specializes in data analysis and majors in Statistics and Math. She is pursuing a degree of Bachelor of Science.

Zixuan (Doris) Zhang. Doris is a fourth-year university student and is planning to graduate this summer. She specializes in data analysis. Doris is pursuing a degree of Bachelor of Science at the University of Toronto. She majors in Statistics and Economics.

Code of ethical conduct

- While using the dataset of current employees at Black Saber, we provided an objective and accurate data visualization. After statistically analyzing the data, we also revealed the limitations in this dataset, the limitations of our analysis process, and the limitations of the results obtained. (1)
- We dedicated to objectively conduct the analysis of data provided by Black Saber Software, the valid information obtained from the analysis adheres to rational public policy and opinion, and we avoided any form of personal prejudice and procedural bias. (1)
- We provided valid information about procedures of processing data from Black Saber Software in a professional manner, and took responsibility for this work, including results and valued information obtained from data analysis and suggestions we gave. (1)

References

(1). Statistical Society of Canada. (n.d.).CODE OF ETHICAL STATISTICAL PRACTICE.https://ssc.ca/sites/default/files/data/Members/public/Accreditation/ethics_e.pdf