
Gender Parity Report

On Salary, Promotion and Hiring

Report prepared for Black Saber Software by Lagrange
Company

2021-04-21

Contents

Executive summary	3
Background & Aim	3
Key Findings	3
Limitations	4
Technical report	5
Introduction	5
Gender Parity Study on Salary	5
Gender Parity Study on Promotion	9
Gender Parity Study on Hiring	11
Discussion	15
Consultant information	17
Consultant profiles	17
Code of ethical conduct	17

```
## Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
## had status 1
```

Executive summary

Background & Aim

Internally, employees have been concerned about possible bias in the compensation and hiring processes of Black Saber Software company. This report is prepared to examine whether salary, promotion, and hiring processes are fair between men and women within Black Saber Software company, and to analyze whether gender bias exists in these three aspects. There are three research questions to study in this project: whether gender affects salary in the company; whether gender affects promotion in the company; and whether there is a tendency to hire more men than women in hiring processes, especially in AI processed hiring phases. The first two research questions were studied based on the current employees dataset, and the third research question were studied based on the hiring data sets. In each research question, we did operations on data, drew plots for visualization, and did statistical analysis.

Key Findings

Gender Parity Study on Salary

- *Gender bias exists in employees' salary in Black Saber Software company. There is a negative effect of gender on salary when gender is woman instead of man.*
- *According to Figure 1, in 7 of the total 8 teams in Black Saber Software company, male employees' average salary is higher than female employees in 2020 Quarter 4 of current employees data.*

Gender Parity Study on Promotion

- *Gender bias exists in employees' promotion in Black Saber Software company. There is a negative effect of gender on promotion when gender is woman instead of man.*
- *According to Figure 2, in 7 of the total 8 teams in the company, male employees' average promotion opportunity is higher than female employees based on all the given current employees data.*

Gender Parity Study on Hiring

- *There is no gender bias in all three hiring phases. Gender has no effect on hiring.*
- *Skills, GPA and professional experiences play important roles in the first two hiring phases which were both processed by AI.*

Limitations

- *When AI deals with hiring processes, there may be no fixed criteria and scoring standards, which may lead to eliminating applicants who actually fit the company's roles.*
- *Sample size in phase3 is too small, so the data and findings are not accurate and reliable.*

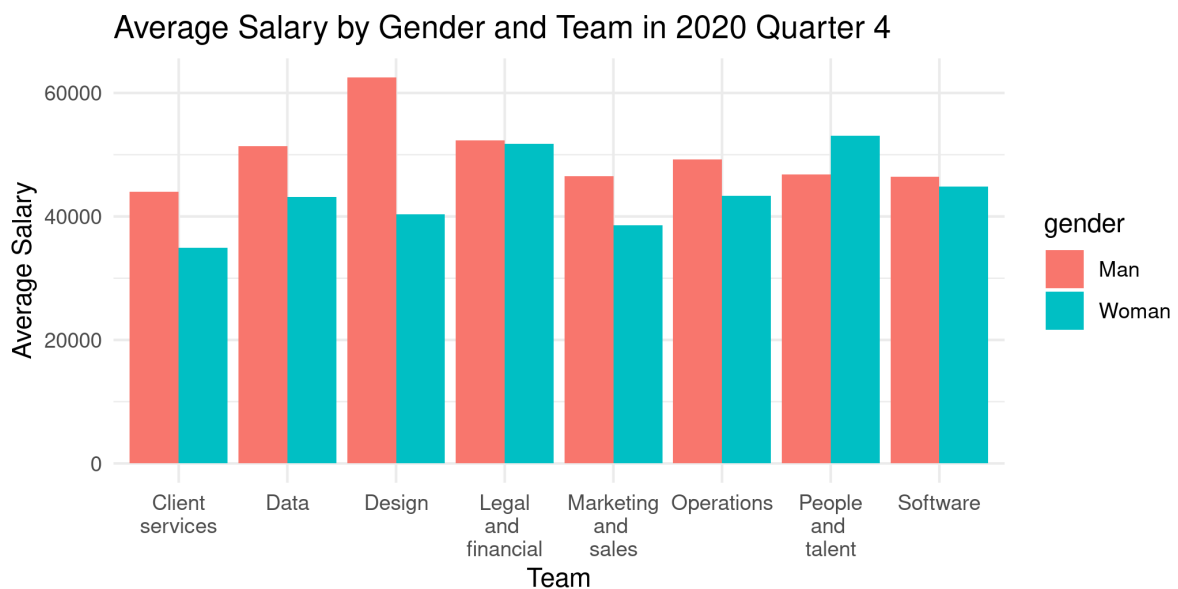


Figure 1. Average Salary by Gender and Team in 2020 Quarter 4

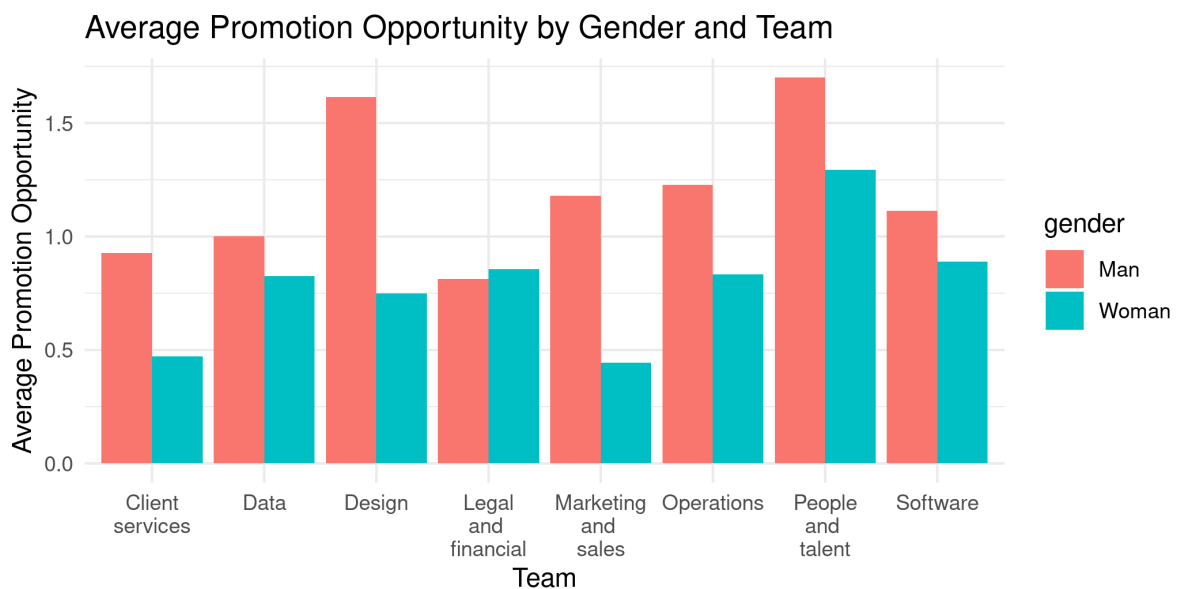


Figure 2. Average Promotion Opportunity by Gender and Team

Technical report

Introduction

In this project, we aim to address our client's needs to determine whether salary, promotion and hiring processes are fair between man and woman, and based on talent and value to the company. We used different datasets, methods and models to study whether there exists gender bias in employees' salaries, promotions, and hiring processes. Before addressing each research question, we excluded employees who preferred not to say their gender from all the datasets, because this group of employees were not within our interest of research.

First of all, to study the effect of gender on employees' salaries, two linear mixed-effects models were fitted for the fourth quarters of 2019 and 2020 in the current employees dataset. Then, to study the effect of gender on employees' promotions, a generalized linear model based on poisson regression was fitted for all data from the current employees dataset. Finally, we studied the effect of gender on three hiring phases respectively by fitting generalized linear models based on binomial regression, using hiring data sets. We also drew boxplots to visualize differences in average salaries, promotion opportunities, and passing rate in hiring processes between different genders.

According to all of our model summary tables and plots, if there were evidence of the effect of gender on salaries, promotions, or hiring processes, it would indicate gender bias on employees in complementary aspects. If there were no evidence of the effect of gender, it would be a good sign for gender parity in this company.

Research Questions

- *Does the gender of employees affect their salaries?*
- *Does the gender of employees affect their promotions?*
- *Does the gender of employees affect the results of hiring processes?*

Gender Parity Study on Salary

For the first research question of gender effect on salary, we planned to manipulate data, draw two sets of boxplots for straightforward visualization and then fit two linear mixed-effects models, based on the fourth quarters of each of 2019 and 2020.

Data Manipulation:

For the convenience of our analysis in this part, we firstly manipulated the data by converting salary variable to a numerical variable, filtering out employees who preferred not to say their gender, and extracted two data sets of each fourth quarter of years 2019 and 2020 from the current employees dataset. The reason why we used the fourth quarters of 2019 and 2020 is that it could be representative of the entire dataset and situations in most recent years. Each of the two new datasets contains variables about employee id which identifies each employee, gender of employee which is either man or woman, which one of the 8 team the employee works in, financial quarter and year, role seniority level, quality of leadership based on role level, productivity rated on a 0-100 scale, and salary at the given financial quarter and year.

Visualization

Based on the new datasets, we used ggplot method to draw two sets of boxplots of salary by team and gender, in order to visualize the difference of salaries between male employees and female employees in each team. We firstly drew the boxplots for the fourth quarter of 2020. The Figure 3 boxplots demonstrate a clear comparison of male and female salaries in different job positions because five of the boxplots corresponding to five of the eight teams show that the median salary of men is higher than that of women. We could draw similar conclusions from the following Figure 4 boxplots of the fourth quarter of 2019. Therefore, the boxplots provide some evidence in visualization that there exists gender bias in salary.

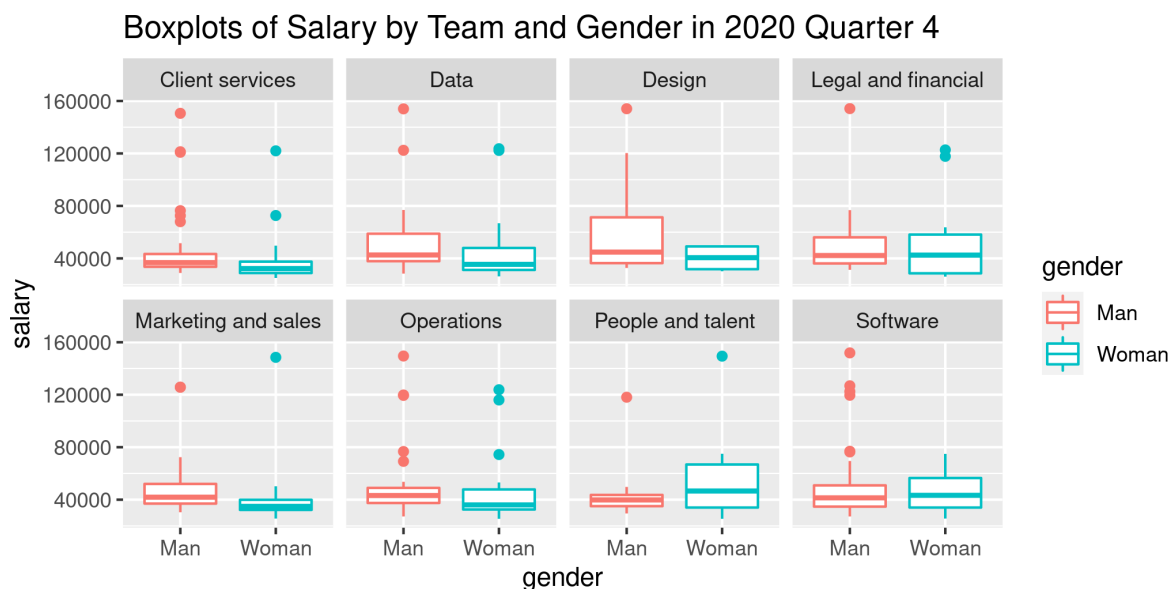


Figure 3. Boxplots of Salary by Team and Gender in 2020 Quarter 4

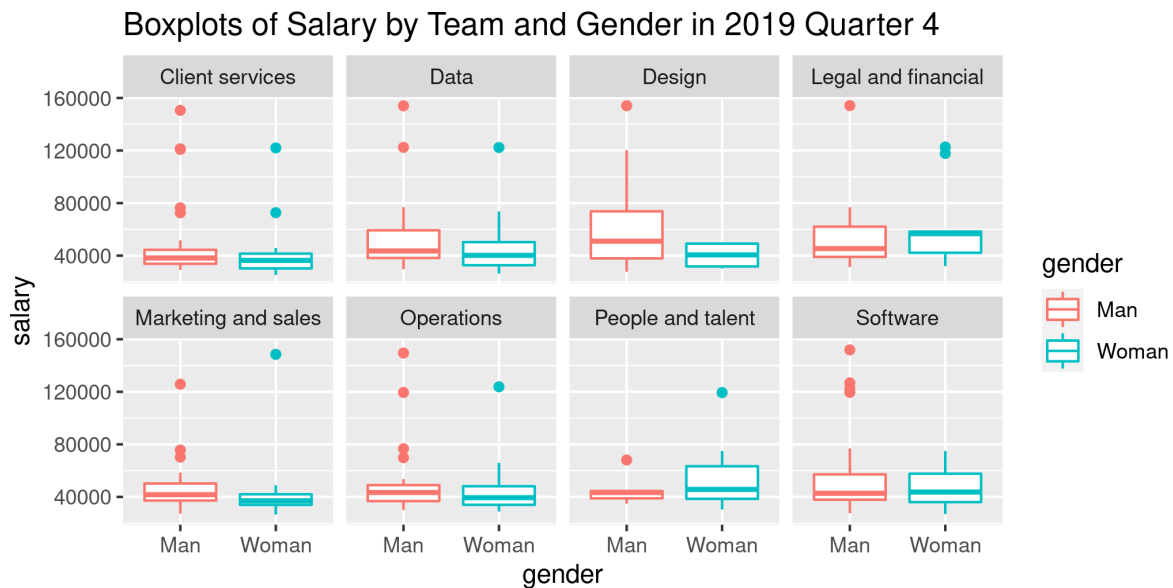


Figure 4. Boxplots of Salary by Team and Gender in 2019 Quarter 4

Methods and Models

We fitted two linear mixed-effects models based on these two new datasets and made summary tables for further analysis. The main reason why we used linear mixed-effects models is that we need to analyze both fixed effects and random effects. In each model, the response variable is salary, and four fixed effect variables we used were gender, quality of leadership based on role level, productivity and role seniority level. Here, we considered the team variable as a grouping unit that might have random effects on salary since it is reasonable that salaries are different among different teams. Two models are similar to each other, including the same variables and the same fourth quarter, except for the only difference of a year.

Results and Conclusions:

The boxplots provide some evidence in visualization that men tend to have higher average wages than women in this company. However, only graphs were not enough to support our conclusion, so we also fitted two linear mixed-effects models and then made their summary tables and calculated 95% confidence intervals as follows.

Table 1. Summary Table of Linear Mixed-Effects Model of Salary 2020 Quarter 4

##	Estimate	t value	95% CI
----	----------	---------	--------

## Woman	-2164.541	-7.096	(-2761.51857, -1574.77563)
## Exceeds expectations	173.248	0.154	(-2007.64356, 2367.38228)
## Needs improvement	-1356.226	-1.263	(-3438.11853, 736.63934)
## Productivity	-7.531	-0.749	(-27.15337, 11.95991)
## Entry-level	-89961.779	-109.243	(-91561.03870, -88360.83717)
## Junior I	-85122.318	-101.711	(-86748.05086, -83495.75941)
## Junior II	-82353.746	-93.691	(-84061.55536, -80645.67478)
## Manager	-49248.907	-48.47	(-51230.86573, -47280.74814)
## Senior I	-76902.759	-87.739	(-78605.21018, -75199.06157)
## Senior II	-71592.252	-78.058	(-73376.57233, -69812.10816)
## Senior III	-65733.697	-73.035	(-67483.22738, -63985.59278)
## Vice President	29917.566	20.95	(27137.64457, 32687.85915)

According to the results in the above Table 1 based on data from 2020 quarter 4, the effect size of gender on salary is 2164.541, and the effect direction is negative when gender is woman instead of man. It indicates evidence showing that if the employee is a woman, the salary would be \$2164.541 lower on average, compared to man. Also, the 95% confidence interval for the effect of gender is (-2761.51857, -1574.77563), which does not include 0 thus indicates a significant effect of gender on salary.

Table 2. Summary Table of Linear Mixed-Effects Model of Salary 2019 Quarter 4

##	Estimate	t value	95% CI
## Woman	-2002.1	-5.749	(-2680.70806, -1330.27716)
## Exceeds expectations	352.8	0.431	(-1233.63951, 1935.85638)
## Needs improvement	1449.69	1.008	(-1342.40712, 4229.32515)
## Productivity	-10.48	-0.901	(-33.06506, 11.99809)
## Entry-level	-90331.82	-93.493	(-92203.03454, -88459.96453)
## Junior I	-84927.37	-91.217	(-86732.09739, -83125.10834)
## Junior II	-83043.07	-85.848	(-84913.86898, -81166.02040)
## Manager	-48870.89	-43.429	(-51057.42174, -46696.49518)
## Senior I	-77361.1	-78.627	(-79268.79589, -75457.03603)
## Senior II	-71483.89	-73.052	(-73381.97683, -69590.73423)
## Senior III	-66533.91	-66.622	(-68468.44456, -64599.43725)
## Vice President	29529.6	19.164	(26545.60397, 32515.38049)

According to the results in the above Table 2 based on data from 2019 quarter 4, the effect size of gender on salary is 2002.10, and the effect direction is negative when gender is woman instead

of man. It indicates evidence showing that if the employee is a woman, the salary would be \$2002.10 lower on average, compared to man. Also, the 95% confidence interval for the effect of gender is $(-2680.70806, -1330.27716)$, which does not include 0 thus indicates a significant impact of gender on salary.

Results from 2019 quarter 4 and 2020 quarter 4 are similar, both suggesting a significant effect of gender on salary. All plots and tables above support the conclusion that female employees tend to get lower salaries than male employees in this company. Therefore, gender bias exists in employees' salaries.

Gender Parity Study on Promotion

For the second research question of gender effect on promotion, we planned to manipulate data, fit a generalized linear model based on Poisson regression. We made a barplot and a summary table to visualize and analyze our results, using all data from the current employees dataset.

Data Manipulation:

After filtering out employees who preferred not to say their gender from the current employees dataset, we generated a new dataset from it by grouping all data by employee id, keeping gender variable and team variable, and extracting distinct rows in case of duplicated employee data. Also, in this new dataset, we created one variable representing the length of work in years for each employee and the other variable showing the number of promotions for each employee. We used all the data because it usually takes a long time to track employees' promotions. Generally, this new dataset contains variables about employee id, gender of the employee, which one of the eight teams the employee works in, length of work in years, and the number of promotions.

Visualization

To draw a bar plot showing the average promotion opportunity of different genders in different teams, all applicants that have once worked in Black Saber Software are selected with their number of promotion opportunities. The average promotion opportunity of male employees is calculated by the total number of male employees' promotion opportunities divided by the total number of male employees. The average promotion opportunity of female employees is calculated using the same method. Then a bar plot of the average promotion opportunity of different genders in different teams can be drawn (Figure 5).

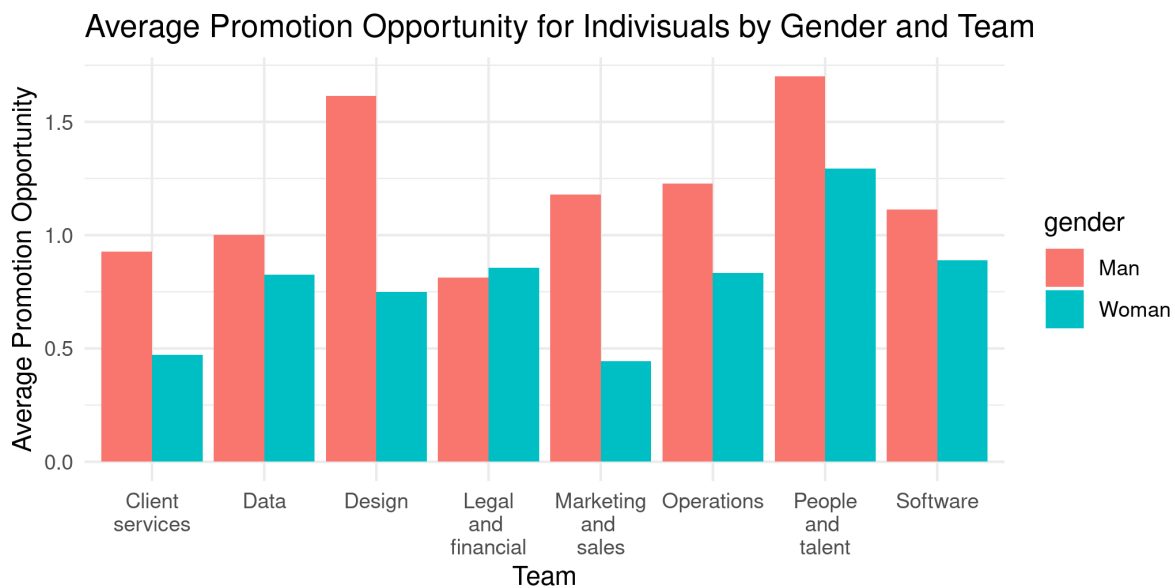


Figure 5. Average Promotion Opportunity for Individuals by Gender and Team

Figure 5 shows that except for the legal and financial team, the average promotion opportunities of male employees are higher than that of female employees, meaning that there is a clear indication of gender bias in giving promotion opportunities in Black Saber Software.

Methods and Model:

Based on the new dataset, we fitted a generalized linear model based on Poisson regression and made a summary table to further analyze. The main reason why we used a generalized linear model is that our response variable is discrete, following Poisson distribution. In the model, the response variable is the number of promotions, and two explanatory variables are gender and team. We considered the log of length of work as the offset term of the model. The reason is that different employees may have worked for different lengths of time, which might affect their opportunities for promotion. Thus, length of work should be controlled for us to analyze the effect of gender on promotion appropriately and accurately.

Results and Conclusions:

The following summary table 3 is made from the model summary and 95% confidence intervals of our generalized linear model described above. According to Table 3, The effect size of gender on promotion is approximately 0.3522, with a negative direction when gender is woman. The p-value for the estimate of gender effect on promotion is 0.000167, much smaller than 0.05, which

means gender is a significant factor. Therefore, there exists strong evidence of the effect of gender on promotion. It indicates that if the employee is a woman, the number of promotions would be 0.3522 lower on average, compared to a man. Also, the 95% confidence interval for the effect of gender is (-0.5374246, -0.1704522), which does not include 0 thus indicates a significant effect of gender on promotion.

Table 3. Summary Table of Generalized Linear Model of Promotion

##	Estimate	p value	95% CI
## Woman	-0.352175	0.000167	(-0.5374246, -0.1704522)
## Data Team	-0.005289	0.973892	(-0.3245791, 0.3103171)
## Design Team	0.197778	0.394235	(-0.2784422, 0.6357812)
## Legal and Financial Team	-0.072103	0.750994	(-0.5377263, 0.3570024)
## Marketing and Sales Team	-0.053225	0.724086	(-0.3487803, 0.2433069)
## Operations Team	0.018637	0.90338	(-0.2830105, 0.3199289)
## People and Talent Team	0.319081	0.098856	(-0.0700150, 0.6902005)
## Software Team	0.113323	0.414157	(-0.1561968, 0.3885763)

Gender Parity Study on Hiring

For the last research question of gender effect on hiring processes, we planned to use hiring data sets, including phase1 dataset, phase2 dataset, phase3 dataset, and final hires dataset. Since there are three phases of hiring, with the first two phases screened by AI and the last phase rated by two interviewers, we manipulated hiring data sets. We fitted three generalized linear models based on the binomial distribution.

Visualization

To draw a bar plot showing the hiring rate of different genders applying for different teams from phase 1 to 2, all applicants that successfully entered phase 2 are selected. The hiring rate of male applicants from phase 1 to 2 is calculated by the number of male applicants entering phase 2 divided by the total number of male applicants. The hiring rate of female applicants from phase 1 to 2 is calculated using the same method. Then a bar plot of the hiring rate of different genders applying for different teams from phase 1 to 2 can be drawn (Figure 6). The phase 2 to 3 bar plot (Figure 7) can be drawn using the same method.

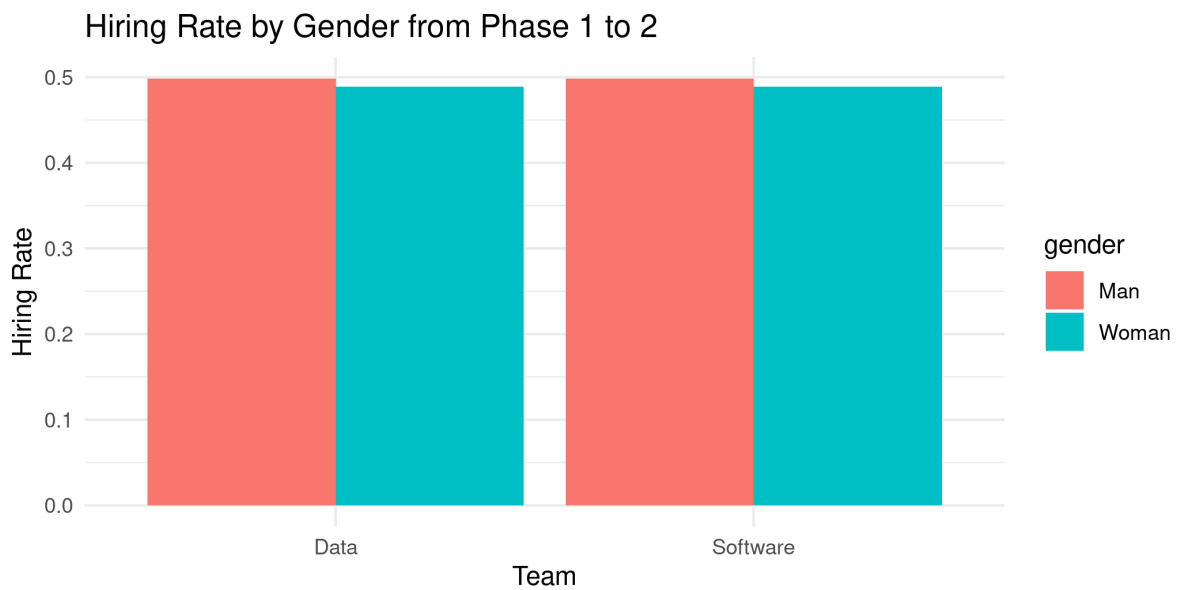


Figure 6. Hiring Rate by Gender from Phase 1 to 2

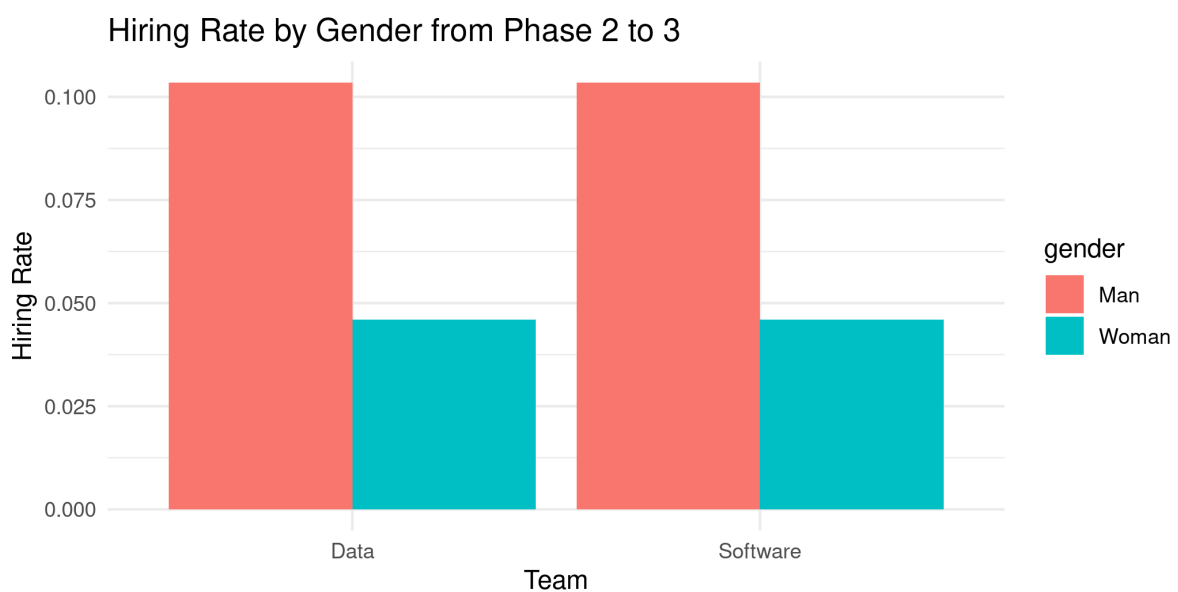


Figure 7. Hiring Rate by Gender from Phase 2 to 3

Figure 6 shows that the hiring rates of male applicants and female applicants applying for the data team and the software team from phase 1 to 2 are almost the same, meaning that there is no clear indication of gender bias in AI selection from phase 1 to 2.

Figure 7 shows that the hiring rates of male applicants applying for the data team and the software team from phase 1 to 2 are much higher than that of female applicants, meaning that

there is clear indication of gender bias in AI selection from phase 2 to 3.

Phase1

After filtering out applicants who preferred not to say their gender from the phase1 dataset, we created a new variable to mark whether each applicant passed phase1 by checking if the applicant id appears in the phase2 dataset. Then we used the new dataset of phase1 to fit a generalized linear model based on the binomial distribution. The reason why we used this model is that our response variable is discrete, following binomial distribution. In the model, the response variable is whether the applicant passed the phase1 or not. All the elements which we were interested in or hiring phase1 took into consideration were added to the model as explanatory variables, including gender, team the applicant applied for, whether the applicant submitted cover letter, whether the applicant submitted CV, GPA, number of extracurriculars and number of work experience.

We made a model summary table as follows. According to Table 4, the effect size of gender on hiring phase1 is 1.09, with a positive direction. However, the p-value is 0.29, which is bigger than 0.05, so gender is not a significant factor. There is no evidence of the effect of gender on the hiring phase1, which AI processes. In contrast, GPA, number of extracurriculars, and number of work experiences are significant factors in the hiring phase1, according to their small p-values.

Table 4. Summary Table of Generalized Linear Model of Hiring Phase1

##	Estimate	p value
## Woman	1.0901000	0.2927560
## Applied for Software Team	-1.0528000	0.2577350
## Submitted Cover Letter or Not	61.7582000	0.9828350
## Submitted CV or Not	50.3709000	0.9917990
## GPA	12.7695000	0.0002140
## Extracurriculars	9.9372000	0.0000597
## Work Experience	11.8482000	0.0000707

Phase2

Similar to phase1 hiring analyzed above, after filtering out applicants who preferred not to say their gender from the phase2 dataset, we created a new variable to mark whether each applicant passed phase2 by checking if the applicant id appears in the phase3 dataset. Then we used the new dataset of phase2 to fit a generalized linear model based on the binomial distribution.

Similar to phase1, we used this model because our response variable is discrete, following binomial distribution. In the model, the response variable is whether the applicant passed the phase2 or not. All the elements, which we were interested in or hiring phase2 took into consideration were added to the model as explanatory variables, including gender, technical skills, writing skills, leadership presence, and speaking skills.

We made a model summary table as follows. According to Table 5, the effect size of gender on hiring phase2 is 0.56658, with a negative direction. However, the p-value is 0.43, which is much bigger than 0.05, so gender is not a significant factor. There is no evidence of the effect of gender on the hiring phase2, which AI processes. In contrast, technical skills, writing skills, leadership presence, and speaking skills are significant factors in the hiring phase2, according to their small p-values.

Table 5. Summary Table of Generalized Linear Model of Hiring Phase2

##	Estimate	p value
## Woman	-0.5665800	0.4329100
## Technical Skills	0.0810600	0.0000676
## Writing Skills	0.0922200	0.0001100
## Leadership Presence	0.8959300	0.0000117
## Speaking Skills	0.7155600	0.0000201

Phase3

Since the phase3 dataset only contains variables of applicant id, the first interviewer rating and the second interviewer rating, we left joined phase2 dataset by applicant id, in order to get the gender data for each applicant. Then, we filtered out applicants who preferred not to say their gender from the new phase3 dataset. Also, we created a new variable to mark whether each applicant passed phase3 by checking if the applicant id appears in the final hires dataset. Then we used the new dataset of phase3 to fit a generalized linear model based on the binomial distribution. Like the previous two phases, we used this model because our response variable is discrete, following binomial distribution. In the model, the response variable is whether the applicant passed the phase3 or not. All the elements we were interested in or hiring phase3 took into consideration were added to the model as explanatory variables, including gender, interviewer rating 1, and interviewer rating 2.

We made a model summary table as follows. According to Table 6, the effect size of gender on hiring phase3 is 48.3, with a negative direction. However, the p-value is 1, which is much bigger than 0.05, so gender is not a significant factor. There is no evidence of the effect of gender on

the hiring phase3, which the interviewers process. Similarly, two interviewer ratings are both not significant in the hiring phase3 due to large p-values.

Table 6. Summary Table of Generalized Linear Model of Hiring Phase3

##	Estimate	p value
## Woman	-48.300	1.000
## Interviewer Rating 1	15.060	0.999
## Interviewer Rating 2	18.940	0.999

Discussion

The main purpose of this project is to know whether salary, promotion, and hiring of Black Saber Software company are based on skills and value, rather than gender. Therefore, we conducted statistical analyses to study whether there is a problem of gender bias in these three aspects of our interest, i.e., salary, promotion opportunity, and hiring processes. After visualizing several plots and analyzing model summary tables, we summarized the following findings related to our three research questions and discussed pros and cons of our work.

In both salary and promotion aspects, it was found that male employees were often treated better than female employees in this company. In most of the working teams, men tend to have higher salaries and more promotion opportunities than women. We could tell that gender plays a significant role in salary and promotion. However, in the hiring processes, either the first two hiring phases processed by AI or the last hiring phase processed by interviewers did not show any evidence of gender bias. Instead, skills, GPA and professional experiences play important roles in the first two hiring phases.

In conclusion, there exists gender bias in salary and promotion in Black Saber Software company, but the hiring processes are fair between men and women.

Strengths and limitations

Strengths: Firstly, we made appropriate plots for straightforward visualizations of gender bias. Secondly, we used appropriate models based on different effects studied and different data distributions. In detail, we used linear mixed-effects models when we needed to analyze both main effects and random effects, and we used generalized linear models when our response variables were discrete. Last but not least, we did detailed and accurate data analysis on summary tables including effect size and direction, p-values, and confidence intervals.

Limitations: One limitation is that AI technology is not exactly mature or precise, so it might result in inaccurate data of phase1 and phase2 about whether applicants passed these two hiring phases or not. If the data was inaccurate, our analysis could have bias. The other limitation is that the sample size of applicants in the third hiring phase processed by interviewers is too small. This is also a reason that our p-values for this phase are all close to 1. A small sample size could affect the reliability and accuracy of our data analysis results. It would be better if we had bigger sample sizes to analyze gender effect in hiring phase3.

Consultant information

Consultant profiles

Jiayu Chen. Jiayu is a senior consultant with Lagrange Company. She specializes in data analysis with R programming. Jiayu Chen earned her Bachelor of Science, Majoring in Statistics and Environmental Studies from the University of Toronto in 2022.

Yexuan Shen. Yexuan is a junior consultant with Lagrange Company. She specializes in data analysis reports. Yexuan earned her Bachelor of Science, Majoring in Statistics and Economics from the University of Toronto in 2022.

Siyi Huang. Siyi is a senior consultant with Lagrange Company. He specialized in reproducible analysis and statistical communication. Siyi earned his Bachelor of Science, Majoring in Statistics and Mathematics from the University of Toronto in 2022.

Code of ethical conduct

As a professional statistical consulting company, Lagrange has always complied with the Statistical Society of Canada Code of Conduct. For this case of Black Saber Software, our company mainly approaches codes A3, B2 and D4 from CODE OF ETHICAL STATISTICAL PRACTICE.

A3 is mainly concerned with objectivity and maintaining neutrality without creating personal opinions. In our data analysis of the BSS, we extracted data only for men and women to avoid bias and better maintain objectivity. All the analysis methods and charts we used were based on the company's original data, without any artificial reduction in the number of people on either side.

B2 is mainly about not providing the company's information to third parties and not accepting any other benefits leading to the company's loss. We were careful to store the data when we got the complete employee data of BSS. We did not disclose any data to any third party, and we only shared the data with colleagues who were involved in the research analysis.

D4 refers to responsibility for the work and providing objective and reliable information for any review. We are very serious and responsible about our work, and we analyze it in different directions to give our clients satisfactory results. There is no personal feeling and no redaction of any essential data in this process. We can list the results of our analysis in the face of any review.