An Analysis of the Fairness in Hiring, Promoting, Salary Distributing Process

A Special Focus on Potential Gender Bias

Report prepared for Black Saber Software by KoCad

2021-04-21

Contents

Executive summary

Background & Aim

Current research was conducted by KoCad to investigate potential biases as well as the gender disparity issues that may exist in Black Saber's recruitment, wage, and promotion schemes. To study the potential gender bias in the company's salary and promoting systems, we evaluated whether employees at the company were being paid and promoted fairly and that those processes were not influenced by discriminatory factors such as gender. We also investigated whether or not any kind of bias was included in each hiring phase that uses either the AI service or the score ratings of interviewers who evaluated applicants.

Key findings

- Male employees' average salary was approximately $1785.6 dollars more than that of female.
- Among those who have never been promoted, the proportion of female employees was slightly greater than that of male employees.
- Unlike the proportion of male employees that increased with the number of promotions, the proportion of female employees decreased as the number of promotions increased.
- The promotion rate per financial quarter for the male employees was about 1.38 times greater than female employees.
- In the first and second hiring phases assessed by AI, the proportion of applicants who got approved to next phase was approximately evenly split between male and female.
- In the third hiring phase, only 7% of the applicants were invited to an interview and among them, the averages of interview scores between male and female applicants were similar.

Limitations

The lack of information on salary determination methods, promotion policies, and evaluation criteria in the hiring process has limited our models' ability to produce more accurate results.

Conclusion

There was a gender bias in both wage and promotion systems in Black Saber Software but not in the hiring process.

Figures

The key findings of the analysis are summarized in the following figures:

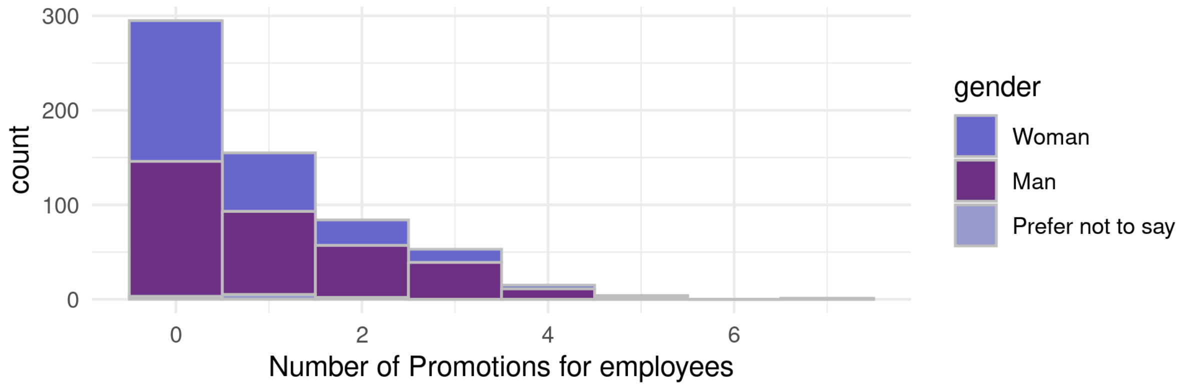

Figure1: Distribution of employees' number of promotion, showing genders as different colour. Each bar's height represents the count of employees with each number of promotions.

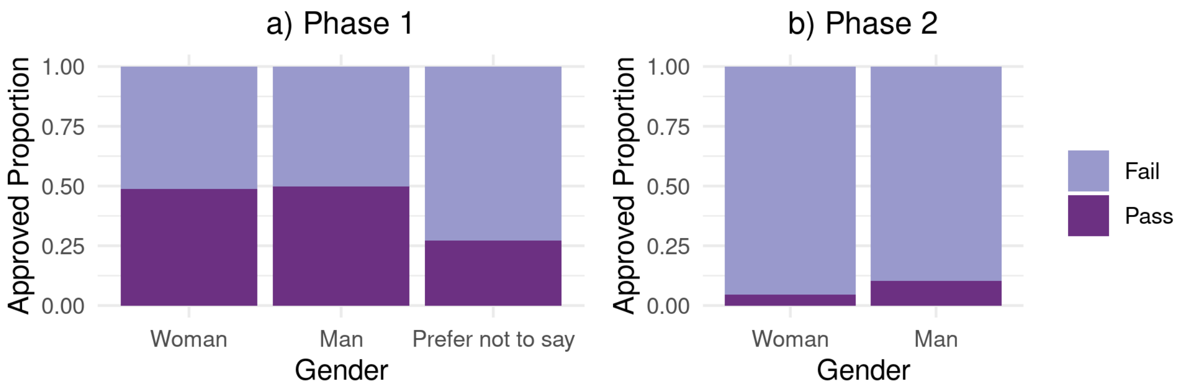

Figure 2: Proportion of Applicants who Passed/Failed in (a) Phase 1 and (b) Phase 2 by Gender

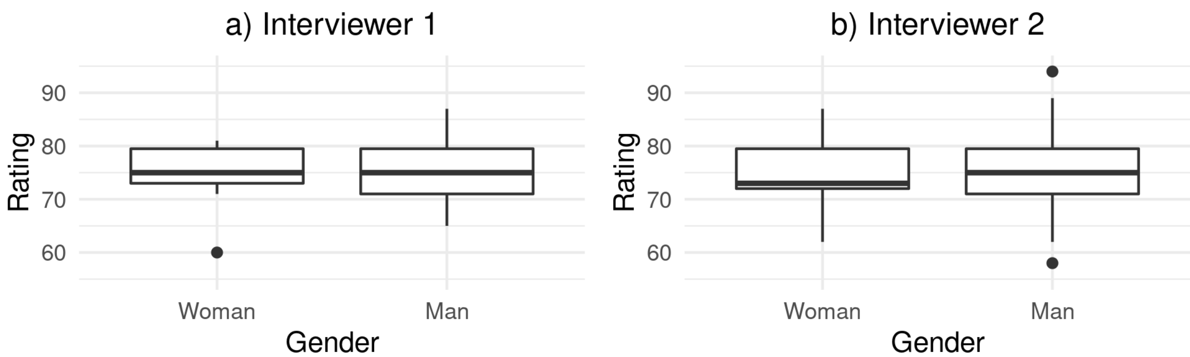

Figure 3: Distribution of Interview Scores evaluated by (a) Interviewer 1 and (b) Interviewer 2, by Gender

The horizontal line in the middle of each box represents the point separating the highest 50% interview scores from the lowest 50%. The horizontal lines at the top and bottom of each box represent the lowest 75% and 25% of the scores, respectively. The vertical lines and dots in the plot give the full range of values, with some extreme values further away from the middle of the box.

Technical report

Introduction

Society has been discriminating against women. They received unequal opportunities to get jobs and earn money based on their assets. With increasing concern about EDI, companies are moving toward implementing EDI initiatives. Considering this social atmosphere, our client, Black Saber Software also requested to launch an investigation on potential bias in their hiring, promoting and salary distributing processes. Using the hiring data for their new grad program and data about promotion and salary for their entire staffs, the present report investigated whether employees of Black Saber Software were hired, promoted, and paid based on their value to the company, not on discriminatory factors such as gender. First, we explored whether the employees were being paid fairly and that salary was not affected by one's gender, using a linear mixed effects model. Then, we studied the relationship between promotion frequency and some potential factors, particularly focusing on identifying the presence of gender bias in the promotion process. Here, we used a generalized linear model with the number of promotions as the response. In addition, with a simple linear regression and generalized mixed effects models, we explored the fairness of the recruitment process by investigating whether the implementation of AI algorithms and the interviewers who evaluated job applicants had gender bias. Finally, we discussed the limitations and strengths of our methods and suggested a future direction for improvement.

Research questions

Throughout the analysis, we focused on assessing the fairness in the hiring, promoting, and salary distributing processes at Black Saber Software. Below are our questions of interest:

Salary

- What factors, such as gender, team, role, leadership level, and productivity, are related to increased rate of salary?
- Does gender have significant influence on the salary?

Promotion

- What are some potential factors affecting the number of promotions of employees?
- Is the common assumption that male employees are more likely to be promoted than female employees also true at the Black Saber Software?

Hiring process

- Are there gender biases throughout the recruitment phase, especially phase 1 and phase 2 in which AI service was implemented?

A Potential Gender Bias in Salary

Data Wrangling and Visualizations

For data wrangling, we first changed the salary variable in the current employee data from characters to doubles so that we can fit a regression model with the salary variable as a response. We removed the dollar sign in the front and a comma sign that separated the thousands value. Then, we reordered the factor level of the employee's seniority of role such that it matches the actual hierarchy in the company rather than being in an alphabetical order. We have also set woman as the baseline for gender variable so that we can make comparisons of each of salary and promotion by gender easily.

Based on the current employee data, leadership level and productivity of an employee seemed to be the variables that show the employee's value to the company. However, we also suspected that team and seniority of role may also have fixed effect on salary since based on those variables, the employee havs differential value to the company. For instance, some teams like software and data team, might be more valuable to the company because Black Saber is a software company. Also, for seniority of role, the higher position an employee takes, the more challenging work they have to do and thus more valuable to the company.

To figure out whether such speculation is true, we fitted a boxplot.

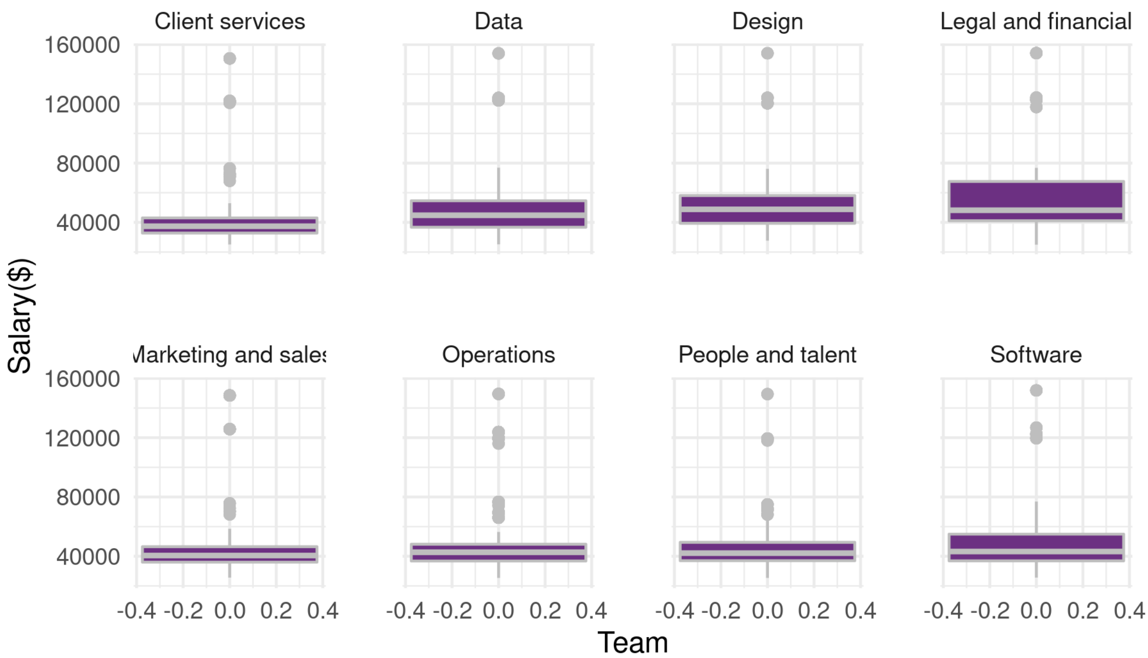

Figure1: Boxplots for Salary by Team

Indeed, the software and data teams generally received higher salary than other teams as shown in Figure 1. Thus, we also included team and seniority of role as predictors to a regression model.

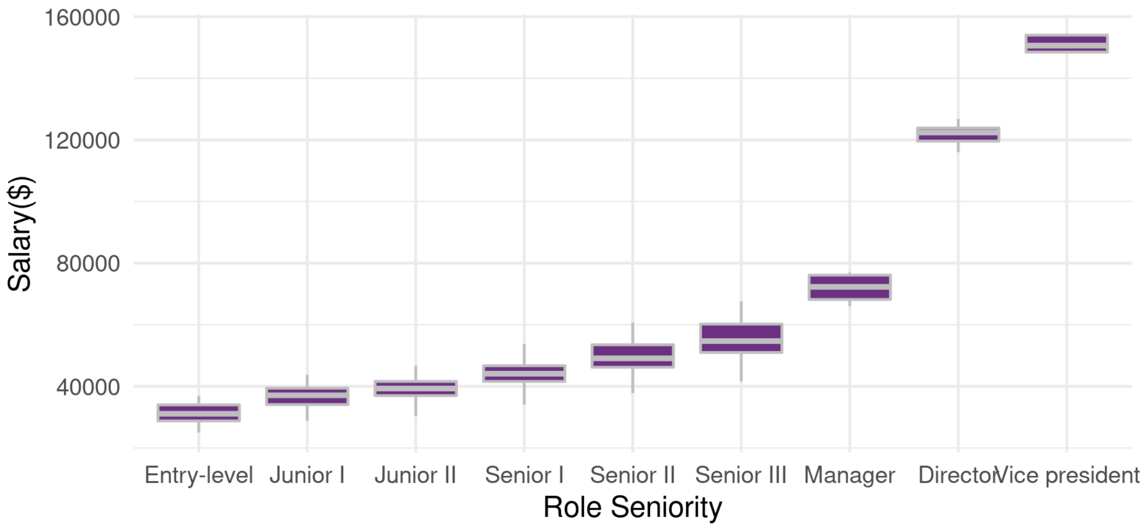

Figure2: Boxplots for Salary by Seniority Role

Also, we saw that salary significantly depended on the person's role seniority level. In particular, as shown in Figure 2, manager, director and vice president received salaries significantly different from other positions.

We also looked at a boxplot for salary by gender to see whether there was a gender difference in salary (see Appendix Figure A.1). Although gender did not seem to create sysemtatic difference, the one-way anova of gender as a predictor and salary as a response suggested a significant difference. However, we needed to see this along with other variables related to an employee's value to the company. Thus, we investigated whether a regression model with and without gender as a predictor significantly differs after we fit each model and make comparison.

Methods

We chose a linear mixed effects model because we suspected that financial quarter could also produce sysematic difference in salary - for instance, in financial quarters where Black Saber earns a lot of money, employees will receive bonuses. However, we were not interested in this variable because we only cared about the employee's value to the company. Further, statistically speaking, employees' salaries were sampled across various financial quarter and so this would violate the independence assumption for observations. Thus, it was important to add financial quarter as a random intercept rather than a fixed effect.

Result of the maximum likelihood test comparing models with and without gender as predictor while including seniority of role, team, leadership, and productivity as other predictors was significant. The significant p-value showed that we have no evidence against the claim that simpler model is better. This means that we should go with the complex model, which is the model that includes gender as predictor, to have better predictive ability for salary. This result implicates that gender caused notable difference in salary and therefore there was a gender bias in salary.

Note that we also tested whether we should include team as predictor for salary becuase it might be ambiguous of whether team had fixed effect on salary based on Figure 1 as compared to seniority of role. The result showed that we should also include team.

Results

This table shows the coefficient estimates for Gender variable, calculated from the final model:

Table 1: Partial Table of Coefficient Estimates for Salary Model Including Gender

	Estimate
(Intercept)	27851.5437
genderMan	1785.6056
genderPrefer not to say	499.8409

Table 1 shows that for our final model predicting salary, man received salary approximately 1785.6 dollars more than woman, on average, while controlling for variables related to employee's value to the company. Similarly, on average, people who did not prefer to say their gender also received salary higher than woman by approximately 500 dollars[1].

In conclusion, there was a gender bias in salary since gender improved salary prediction above and beyond inclusion of other variables, which were the factors that showed the employee's value to Black Saber Software. Moreover, such factors not only included employee's productivity and leadership, but also the team they are in and the role they take in the company.

A potential Gender Bias in Promoting Process

Data Wrangling and Visualizations

For this particular topic, employee data cleaned in the previous section was used to study several potential factors that affected the company's promoting process. Since the main purpose of this research question was to see a potential bias presented in the promotion process, we created a promotion dataframe that includes the information about employees collected on the last

[1] For full information about coefficient estimates for the final model for salary, please see Table A.1 in the Appendix.

quarter of 2020, a work period variable that accumulates all financial quarters an employee has experienced, and a new variable that stores each employee's roles numerically rather than categorically. The last variable allowed us to track changes in their roles more easily and thus helped us determine the initial and current roles of each employee and the difference between them. Such difference in roles gave us the number of times each employee was promoted.

For this analysis, the response variable was the number of promotions per employee received. Since there may be some employees who have been promoted more than once, we thought that having a Bernoulli variable[2] as our response would not take into account the variability in different number of promotions each employee received. Therefore, we chose our response to be the number of promotions.

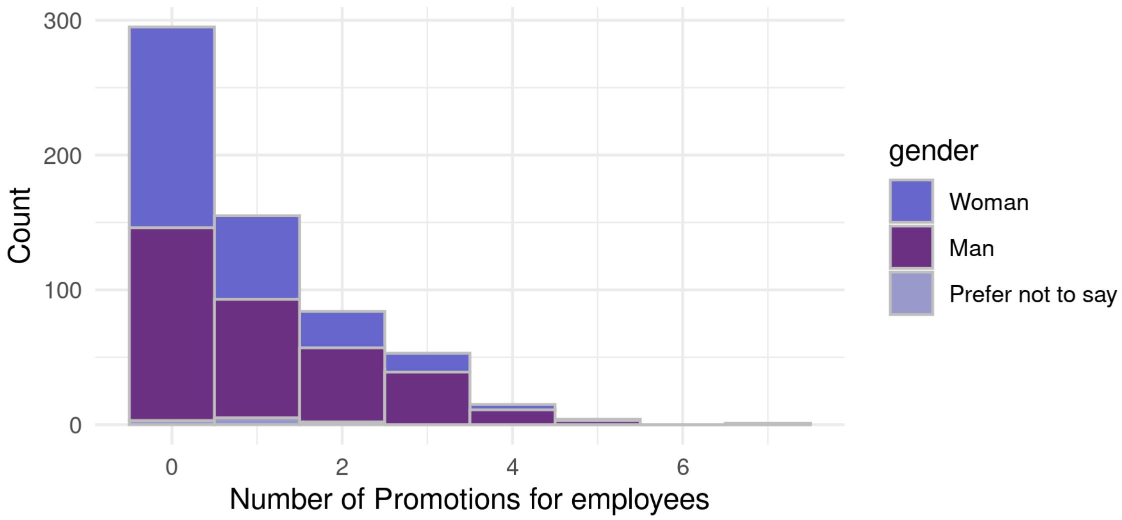

Figure3: Distribution of Employees' Number of Promotions by Gender

According to Figure 3, among employees who have never been promoted, there were more women than men. This trend of women receiving less promotion than men was becoming more pronounced as the number of promotions increased: Among those who were promoted at least once, male employees accounted for more than women. Thus, it was important to identify the potential gender bias present in the promotion process.

Methods

Since the response was a count, a Poisson regression was used to model our data. However, there were several assumptions that we had to check before fitting a Poisson model:

1. **Poisson Response:** Since the response was a count it followed a Poisson model.

[2]This Bernoulli variable indicates whether or not an employee has been promoted at least once

2. **Independence:** This assumption was also met because each observation in the promotion data contained information of different employees.

3. **Mean = Variance:** Accounting for the fact that the number of promotions varied from 0 to 7, the discrepancies between the mean and the variance, ranging from 0.3 to 0.5, were considered small. Therefore, we assumed that the assumption of variability equal to the mean was also satisfied.

4. **Linearity:** Since gender was not a continuous predictor, it was difficult to identify the linearity of $\log(\lambda)$.

As we were interested in studying the potential gender bias in the promoting process of the Black Saber Software, we created two different models and ran a drop-in-deviance test to determine whether including the gender improves our model. In both models, we included an offset, which was the log of the work period, so that the number of promotions could be adjusted to be comparable across employees with different working periods[3].

The initial model contained covariates such as leadership, productivity, salary, and role seniority and this model was compared with another model that also included gender as a predictor, in addition to covariates in the inital model. Based on the small p-value from the drop-in-deviance test, we concluded that adding gender actually improved the model.

Results

This table shows the estimates and 95% confidence interval for Gender variable, calculated from the final model:

Table 2: Partial Table of Coefficient Estimates and 95% Confidence Interval for Gender in the Final Model

	Estimate	2.5 %	97.5 %
(Intercept)	-20.6975	-3004.9657	-3004.9657
genderMan	0.3255	0.1367	0.5180
genderPrefer not to say	0.2164	-0.5394	0.8405

From the summary of the model, only the p-value for male employees was significant[4]. Also,

[3]Employees who have worked longer than others are more likely to be promoted at least once.

[4]The full table of the estimates and 95 % confidence interval for the model is included in the Table A.2 in the Appendix

considering that the 95% confidence interval for male employees did not contain 0, there was an evidence against the hypothesis that there is no difference in promotion frequency between men and women[5]. Specifically, Table 2 suggests that the promotion rate per financial quarter for male employees was nearly 1.38 times greater than that of females, after controlling for other predictors. In other words, a gender bias was actually present in the promoting process at the Black Saber Software.

Fairness in Hiring Process

Data Wrangling and Visualizations

Since our main purpose was to test whether or not gender affected applicants' acceptance, we first made variables telling us whether an applicant passed each hiring phase or not. This variable helped us create a response variable that follows Bernoulli distribution. Phase 3 data for hiring did not have information about applicants' gender, so we made all phases to have gender information by joining data from phase 2. We have also set woman as the baseline for gender variable so that we can make comparison of whether an applicant got hired by gender easily.

In phase 2, none of the applicants who preferred not to say their gender got accepted and therefore, we omitted them from phase 2.

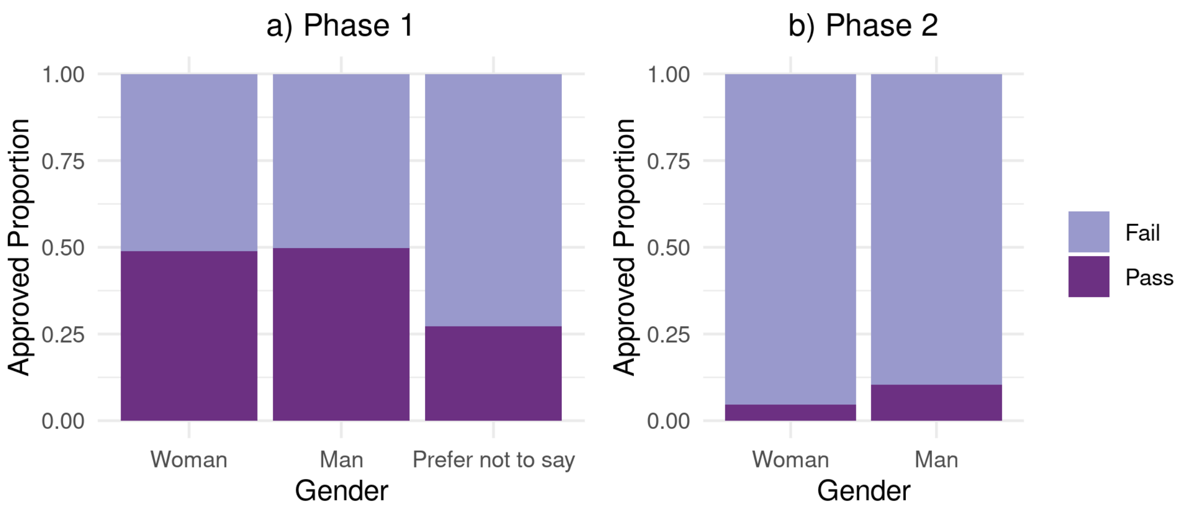

Figure 4: Proportion of Applicants who Passed/Failed in (a) Phase 1 and (b) Phase 2 by Gender

The proportions of female applicants who got accepted in phase 1 was similar to that of male applicants. On the other hand, among those who preferred not to say their gender, there was 25% difference between applicants who got accepted and who did not. However, since only 11 of 613 applicants[6] preferred not to say their gender, having 25% difference among this small group

[5] Woman was our baseline for gender

[6] This is only 2% of all applicants.

of people does not provide much evidence that phase 1 was biased with gender for this group.

In phase2, the proportion of male applicants who got accepted was 10%, which was around 6% greater than that of female applicants.

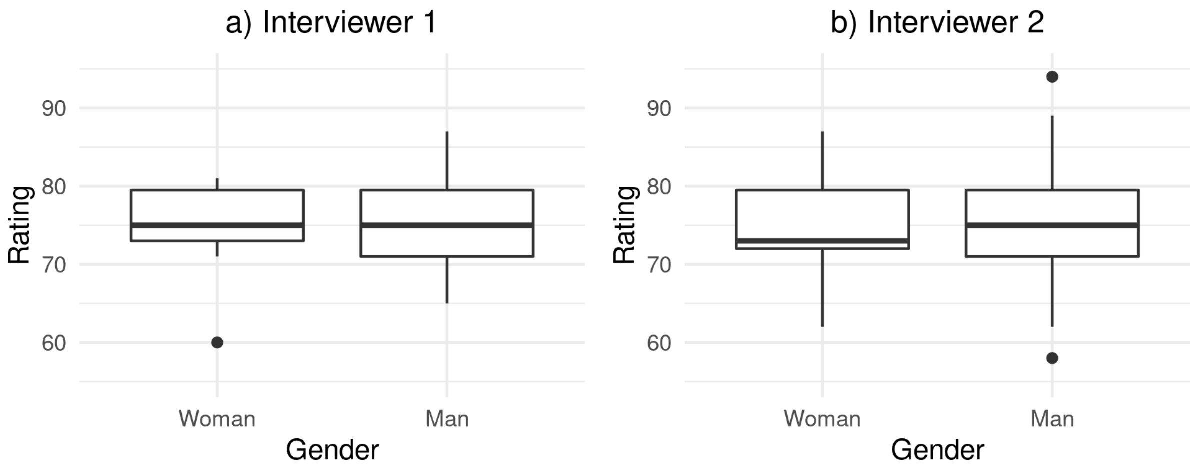

Figure 5: Distribution of Interview Scores evaluated by (a) Interviewer 1 and (b) Interviewer 2 by Gender

In phase 3, the ranges of interview scores given by interviewer 1 and 2 were similar to each other. Although there was a slight difference between mean scores of male and female who were assessed by interviewer 2, it was only a small difference of 2.5 in scores. Therefore, we found some evidence that gender did not influence the interview score.

Methods

Although we did not know how AI evaluated each applicants, we put all variables as fixed effect, thinking that AI will consider all the variables into account. Our response variable followed Bernoulli distribution, which indicated whether or not each applicant passed each phase. Since the logit link was used for Bernoulli response, we could assume that there was a linear relationship between the transformed response and explanatory variables. Also, we assumed that information of each applicants were independent. Since our response variable followed Bernoulli distribution, we used generalized linear models.

In initial model, where gender was our only predictor, confidence interval of all genders included 0. Therefore at the 95% confidence level, it was believable that male and female applicants, and applicants who preferred not to say their gender had the same log odds of being accepted by the AI in phase 1. In addition to this model, we tried to investigate whether the AI was biased with gender and other variables in phase 1 by comparing two models; One with all variables included[7]

[7] This model includes the team that the applicant applied for, whether or not the applicant submitted a cover letter, CV, GPA, gender, the degree of relevance of extracurriculars to the company, and the amount of work experience.

and another with only gender excluded from the previous model.

Lastly, for phase 3, we used a simple linear regression model with a single predictor, gender, for interview ratings. Since no trend was found in the residuals versus fitted values plot, we assumed that the linearity assumtion for the simple linear regression was satisfied. Also, the standardized residuals versus fitted values plot showed that the errors were homoscedastic. Further, the linear trend in the Normal Q-Q plot indicated that the normality assumption was satisfied. Lastly, we assumed that the independence assumption was also satisfied since the data contained interview scores for different applicants.

Results

This table shows the deviances and p-values for model 1 which excluded gender and model 2 which include all variables:

Table 3: Partial Table of Deviances and P-values for Gender in the Phase 1 and 2

	Phase 1		Phase 2	
	Deviance	p-value	Deviance	p-value
Model 1	34.773		69.877	
Model 2	33.562	0.5459	69.247	0.4275

The deviance tests using the two models showed that it is not preferable to add gender as a predictor variable to the model. The deviance difference between models in phase 1 with and without gender was 1.21, which was pretty small. Also, the large p-value supported that adding a new indicator variable, gender, to the model with all the other variables included was not very helpful in predicting the acceptance of an applicant.

Similarly, we could see that there were small deviance of 0.62 and large p-value of 0.43 in deviance test for phase 2 models. Therefore, adding gender to the model did not explain the acceptance of applicants in phase 2.

From the confidence interval in phase 3, we could see that the coefficients of gender contained 0.[8] Therefore, we concluded that interviewers were not biased with gender. In other words, gender did not affect interview scores.

[8]The full table of the 95 % confidence interval for the model is included in the Table A.5 in the Appendix

Discussion

Overall, the results of statistical analysis suggest that there was a gender bias in salary distribution process and promotion process of current employees in Black Saber Software, but not for the hiring process using trialing AI and the human interviewer; For biased processes, men received higher salary and were more likely to get promoted than women, despite controlling for their talents and values to the company.

In particular, we found that salary was affected by gender when other explanatory variables such as productivity, leadership, and team were controlled. Similarly, the number of promotion was affected by gender when other factors such as productivity, leadership, salary and role seniority were fixed. In hiring process, we revealed that both the AI and human interviewers were not biased with gender. Specifically, we found that applicants' acceptance was not affected by gender when assessed by AI. Also, the interview score was not affected by gender when applicants were evaluated by interviewers in phase 3.

Strengths and limitations

Throughout the analysis, we demonstrated high understandings of statistical concepts and the data provided by our client company, Black Saber Software. Instead of taking the superficial details of data, we considered how each variable may play out in real life and thus made appropriate modifications to data and chose the apt model. For example, when we modelled the number of promotions to see if there was a gender bias in the promotion process, we acknowledged that the more a person stays at the company, the more likely they are to get promoted just by the mere fact that there are more opportunities for them to get promoted. Accordingly, we created an offset that accounts for work periods of employees so that our response, the number of promotions, could be adjusted to be comparable across employees with different work periods.

On the other hand, one limitation of our analysis was that the provided data did not contain information on how the company select employees for promotion. Understanding the company's core values or knowing more about employee promotion policies will help us identify the right factors that affect employee promotion. Similarly, we did not know the evaluation criteria for each phase when we made models for hiring process. If we knew how score is being calculated by interviewers, we would be able to explore potential biases in the interview phase more in depth.

For future consideration of our study, we could look into promotion distribution by team, role and financial quarter and identify segments with pronounced biases. This way, we will not only be able to figure out whether or not the company's promotion system is biased but we can also find out specific teams, roles or financial quarters in which the biases are present.

Consultant information

Consultant profiles

Yoon Young Lee. Yoon Young is a senior data analyst with KoCad. She specializes in data visualization and data analysis. Yoon Young earned her Bachelor of Science, Specializing in Psychology and Majoring in Statistics from the University of Toronto in 2022.

Guemin Kim. Guemin is a junior data engineer with KoCad. She specializes in data analysis and machine learning. Guemin earned her Bachelor of Science, Majoring in Actuarial Science and Statistics from the University of Toronto in 2022.

Woolim Kim. Woolim is an junior statistician with KoCad. He specializes in data visualization and statistical communication. Woolim earned his Bachelor of Science, Majoring in Statistics and Economics from the University of Toronto in 2022.

Hojung Kim. Hojung is a junior data analyst with KoCad. He specializes in mathematical statistics and statistical communication. Hojung earned his Bachelor of Science, Majoring in Statistics and Mathematics from the University of Toronto in 2022.

Code of ethical conduct

Responsibility of Statistician - KoCad acknowledges the statisticans' responsibility of manipulating data, analyzing data, and interpreting the result of analysis in a transparent way. KoCad promises not to manipulate data in a way that will introduce bias, in order to create a significant result. KoCad approaches the analysis with the appropriate model for the given data and not by the model that we want. KoCad communicates the results in a manner that considers indiviudal differences in beliefs, opinions, and background.

Responsibility to Clients - KoCad cannot guarantee that the results of the analysis will be exactly aligned with the expectations of the client. KoCad is responsible for retaining full knowledge and understanding of statistical methods, deducing valid conclusions from the data provided, and identifying and explaining any limitations to the conclusions that can be drawn.

Responsibility to other fellow statisticians - KoCad ensures a supportive working environment to fellow statisticians in their professional development. To motivate and inspire fellow professionals, questions and debate on projects are recommended. Avoiding direct criticism of the person, conflicts should be directed and resolved according to procedures. All fellow KoCad consultants are responsible to act with integrity toward others.

Appendix

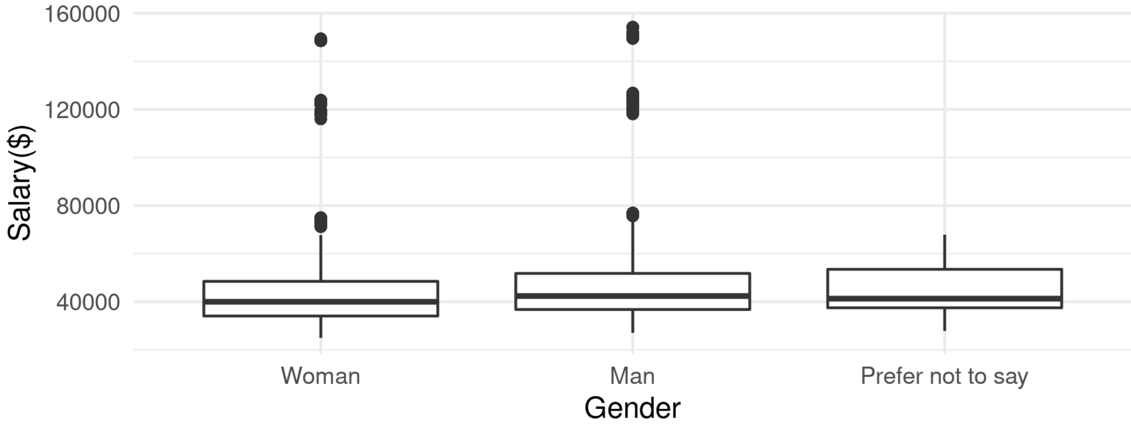

Figure A.1: Boxplots for Salary by Seniority Role

Table A.1: Full Table of Coefficient Estimates for Salary Model

	Estimate
(Intercept)	27851.5437
genderMan	1785.6056
genderPrefer not to say	499.8409
teamData	3796.7121
teamDesign	1180.7819
teamLegal and financial	4545.6853
teamMarketing and sales	1796.9405
teamOperations	1273.5526
teamPeople and talent	-1220.5586
teamSoftware	4839.2788
role_seniorityJunior I	5372.1281
role_seniorityJunior II	7853.7190
role_senioritySenior I	13138.4922
role_senioritySenior II	18614.8519
role_senioritySenior III	24167.9741
role_seniorityManager	40309.9786

	Estimate
role_seniorityDirector	91071.2294
role_seniorityVice president	119589.8496
leadership_for_levelExceeds expectations	-288.2453
leadership_for_levelNeeds improvement	-182.8906
productivity	1.8654

Table A.2: Full Table of Coefficient Estimates with 95% Confidence Intervals for Promotion Model

	Estimate	2.5 %	97.5 %
(Intercept)	-20.6975	-3004.9657	-3004.9657
genderMan	0.3255	0.1367	0.5180
genderPrefer not to say	0.2164	-0.5394	0.8405
leadership_for_levelExceeds expectations	-0.1277	-0.7814	0.4304
leadership_for_levelNeeds improvement	0.1843	-0.6863	0.8855
productivity	0.0017	-0.0045	0.0080
salary	0.0000	0.0000	0.0000
role_seniorityJunior I	17.6121	2172.9166	2172.9166
role_seniorityJunior II	17.7547	3002.0225	3002.0225
role_senioritySenior I	18.1421	2173.4467	2173.4467
role_senioritySenior II	18.0128	2173.3174	2173.3174
role_senioritySenior III	18.1659	2173.4706	2173.4706
role_seniorityManager	18.0216	2173.3266	2173.3266
role_seniorityDirector	18.0293	3002.2998	3002.2998
role_seniorityVice president	17.8445	3002.1172	3002.1172

Table A.3: Full Table of Coefficient Estimates with 95% Confidence Intervals for Hiring Phase 1 Model

	Estimate	2.5 %	97.5 %
(Intercept)	-151.1283	-2449.2960	-3653.7517
team_applied_forSoftware	-0.9071	-2.7961	0.8373
cover_letter	59.0835	31.6094	1184.7225
cv	48.7949	-104.2644	1776.5233
gpa	11.9693	6.9696	19.5409
extracurriculars	9.5009	5.8364	15.2032
work_experience	10.8500	6.7927	17.1193

Table A.4: Full Table of Coefficient Estimates with 95% Confidence Intervals for Hiring Phase 2 Model

	Estimate	2.5 %	97.5 %
(Intercept)	-20.6704	-29.0472	-14.4615
technical_skills	0.0789	0.0444	0.1228
writing_skills	0.0872	0.0473	0.1361
speaking_skills	0.7159	0.4206	1.0787
leadership_presence	0.9111	0.5586	1.3780

Table A.5: Full Table of 95% Confidence Intervals for Hiring Phase 3 Model

	2.5 %	97.5 %
(Intercept_1)	69.4626	79.3945
genderMan_1	-4.9093	7.1188
(Intercept_2)	67.7454	82.2546
genderMan_2	-8.4524	9.1191