

---

# **Gender Parity in Black Saber Software**

Evidence of Gender Bias in Hiring, Promotion and Salary Assignment Processes

Report prepared for Black Saber Software by Zeusolutions

2021-04-21

## Contents

<b>Executive summary</b>	<b>3</b>
Background and Aim . . . . .	3
Key findings . . . . .	3
Limitations . . . . .	3
<b>Technical report</b>	<b>5</b>
Introduction . . . . .	5
Fairness of the AI Recruitment Pipeline Manager . . . . .	5
Fairness of Promotion process . . . . .	15
Fairness of Salary assignment for current employees at Black Saber . . . . .	18
Discussion . . . . .	20
<b>Consultant information</b>	<b>22</b>
Consultant profiles . . . . .	22
Code of ethical conduct . . . . .	22

## Executive summary

### Background and Aim

Black Saber Software has been implementing an AI service in its application screening process. The AI system is responsible for rating candidates from a pre-recorded video and assessing both a timed technical coding task and a writing sample that are submitted by the candidates. The final interview in Phase 3 is conducted and scored by people while the final hiring decision is made by the HR team. No data on ethnicity or race is collected, thus the AI system does not take those factors into account. In addition, the company also collected data on the salaries and promotions of their current employees.

The aim of the present research is to examine factors that contribute to hiring, promotion and salary processes of Black Saber Software and determine whether they are fair. In particular, the hiring pipeline using AI is of most interest. Factors that may be of interest are gpa, gender, number of extracurricular activity, number of previous work experience, presence of CV, presence of cover letter, etc.

### Key findings

- The AI recruitment pipeline is unbiased in its scoring/rating, but gender biased in those who proceed from one phase to another.
- In Figure 2, one can see the gender disposition between each phase and the final hires, where in a “Woman” dominated pool of applicants, “Man” dominated the final hires.
- Average productivity and average leadership for level do not have a great effect on promotion.
- Women on average received 0.33 less promotions for every promotion received by their male counterparts.
- In Figure 1, we see that within the same roles, males tend to be paid more than their female counterparts across almost all positions.
- We see that women, on average, are paid \$1998 less than men.

### Limitations

- The scope of the findings are based current and provided data and cannot be used to infer about the future Black Saber Software.
- There may be confounds such as race and ethnicity that could account for the bias in Black Saber Software, however data for such factors were not provided.

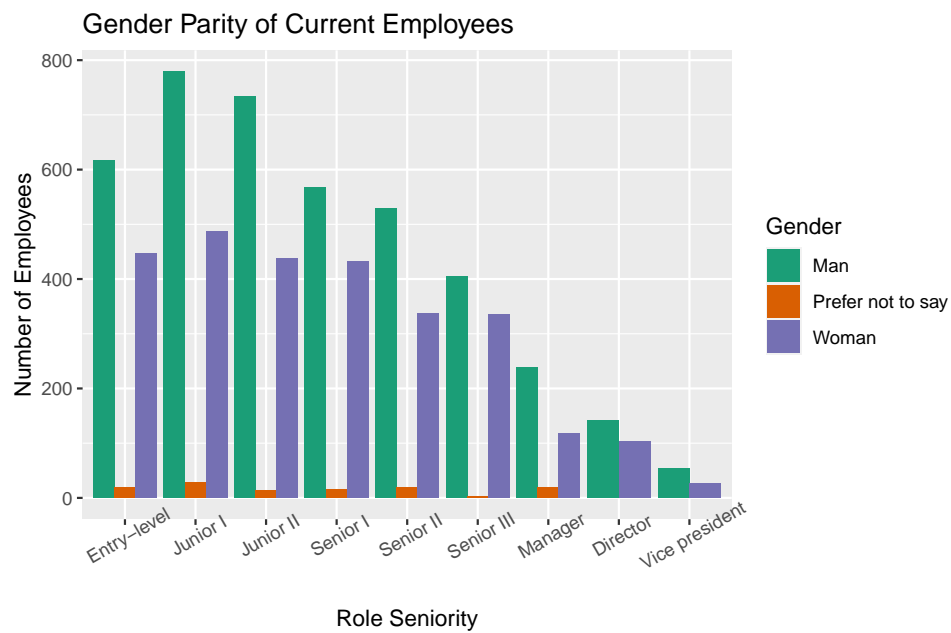


Figure 1. Bar chart for the number of employees by different genders

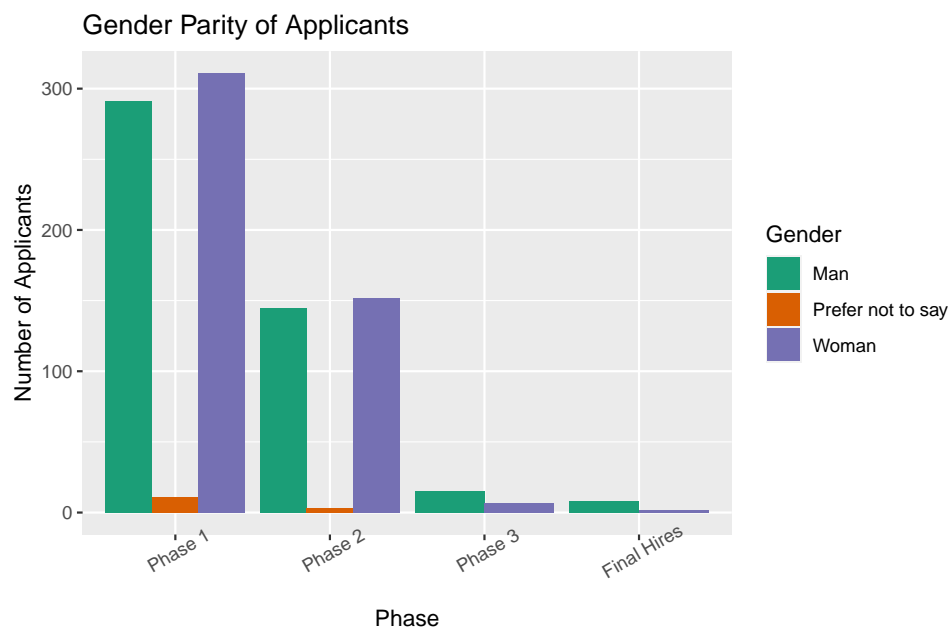


Figure 2. Bar chart for the number of applicants by different genders

## Technical report

### Introduction

The following report will be utilizing statistical concepts to properly investigate the research questions revolving around the fairness of hiring, promotion and salary assignment processes within Black Saber. Concepts revolving around data wrangling and explanatory variable analysis will be demonstrated throughout the technical report and be performed through R programming language. Binomial and Poisson distributions will be involved in selecting appropriate models to investigate the factors of interest. Generalized linear mixed models and linear mixed models will be used to model these factors. In order to investigate the general distributions of genders across different areas of interests, marginal probabilities and odds ratio will also be present.

### Research questions

The following research questions are going to be addressed within this report.

- What factors play a role in being hired through Black Saber's recruitment pipeline manager?
- What factors contribute to promotions of Black Saber's current employee's?
- What factors dictate the salary assignment of current employees at Black Saber?

### Fairness of the AI Recruitment Pipeline Manager

The company Black Saber has a new AI recruitment pipeline manager that has been trialing for the Data and Software teams. The board has concerns about the process and questions whether or not it is fair and based on value and talent of each applicant. We have been given the data for each phase of the process, in which there are a total of three. Through each phase there is certain data collected through question and tasks in which the AI uses to narrow down the field of applicants from one phase to another. We want to determine what factors play a major role in being hired through Black Saber's recruitment pipeline manager.

### Defining Features of the Application Form

The first phase of the hiring pipeline is the submission of an application form, cover letter and CV. Each line of `phase1-new-grad-applicants-2020.csv` contains the information of each applicant in phase 1:

- `applicant_id` = A unique ID given to each applicant in Phase 1

- `team_applied_for` = Software or Data
- `cover_letter` = 1 if present, 0 if not.
- `cv` = 1 if present, 0 if not
- `gpa` = 0.0 to 4.0
- `gender` = “Man”, “Woman”, or “Prefer not to say” as the only options
- `extracurriculars` = 0, 1 or 2. Description in which is accessed automatically by keywords where 2 indicates high, 1 indicates some and 0 indicates no relevant skill building extracurriculars.
- `work_experience` = 0, 1 or 2. Description in which is accessed automatically by keywords where 2 indicates high, 1 indicates some and 0 indicates no relevant work experiences.

**Table 1:** The first five applicants from phase 1 of the hiring pipeline

<i>applicant_id</i>	<i>team_applied_for</i>	<i>cover_letter</i>	<i>cv</i>	<i>gpa</i>	<i>gender</i>	<i>extracurriculars</i>	<i>work_experience</i>
1010	Software	0	1	1.3	Man	1	1
1020	Software	0	1	3.4	Woman	2	1
1030	Data	1	1	2.4	Woman	2	0
1040	Software	0	1	2.7	Man	1	1
1050	Data	1	0	2.1	Prefer not to say	0	1

Similarly, we are also given the list of applicants who have made it to phase 2 along with the factors assessed in phase 2. However, in this section we are only interested in which applicants made it to phase 2. For this reason we will create a variable called `passed_1` and set it to 1, if the applicant made it from phase 1 to phase 2 and 0, if not. This will be the response variable we are interested in for developing our model. The other variables will be fixed effects besides team that the applicant applied for. We believe that team that the applicant applied for should be added as a random effect because each team has a different set of required skills and/or experiences. Other than that, since each row corresponds to a unique applicant, this suggests that the observations are independent, hence our assumptions are not violated. We want to determine which is the most impactful component of the application form that allows the applicants to proceed to phase 2 of the hiring process.

Given the binomial response variable and random effect, we will be constructing a generalized linear mixed model. Before constructing the model, we want to first consider the gender distribution for those who passed phase 1 and those who did not. In addition, consider also the marginal probabilities and odd ratios.

**Table 2:** The number of applicants that passed phase 1 based on gender

	Did Not Pass Phase 1	Passed Phase 1
Man	146	145
Prefer not to say	8	3
Woman	159	152

**Table 3:** The probability of an applicant's gender given that they pass phase 1

	Did Not Pass Phase 1	Passed Phase 1
Man	0.4664537	0.4833333
Prefer not to say	0.0255591	0.0100000
Woman	0.5079872	0.5066667

Hence from Table 3, given that the applicant passes phase 1, the probability that their gender is a “Man” is 48%, whereas a “Woman” is 51%. Also, the odds ratio of Woman vs Man who pass phase 1 is approximately 0.96. That is, the odds of a “Woman” passing phase 1 is less than that of a “Man”. This odds ratio is relatively close to 1, hence there is no reason to suspect that there are any gender biases in those who passed phase 1.

Then, based on the response variable and random effect we will be considering a generalized linear mixed model.

**Table 4:** Generalized linear mixed model of applicants who passed Phase 1

Characteristic	log(OR)	95% CI	p-value
cover_letter	63	-24,344, 24,471	>0.9
cv	52	-41,425, 41,528	>0.9
gpa	13	6.1, 19	<0.001
gender			
Man	—	—	
Prefer not to say	0.63	-25, 26	>0.9
Woman	0.91	-1.0, 2.8	0.4
extracurriculars	10.0	5.1, 15	<0.001
work_experience	12	6.1, 17	<0.001

OR = Odds Ratio, CI = Confidence Interval

In the model seen in Table 4, it is evident that GPA, relevant extracurriculars and relevant work experience have the most significant effect on passing Phase 1, based on their log odds. We

cannot interpret the odds ratio of GPA because there is no one with a 0.0 GPA, however we can interpret the odds ratio of the coefficients for relevant skills from extracurriculars and work experience to an extent. Since the variables for relevant skills from extracurriculars and work experience are given scores of either 0, 1 or 2, the odds ratio is an average of scores 1 and 2. Nonetheless, we find that the odds of passing Phase 1 with some to high relevant skills from extracurriculars is 21267.4 times that of those without and the odds of passing Phase 1 with some to high relevant work experience is 107591 times that of those without. Overall, we can conclude that the most impactful component of the application form that allows the applicants to proceed to phase 2 of the hiring process in order of is GPA, followed by relevant skills from extracurriculars and work experience.

## AI Grading Influences

The second phase of the hiring pipeline invites applicants to do a timed technical coding task, writing sample and submit a pre-recorded video. Each line of `phase2-new-grad-applicants-2020.csv` contains the AI graded scores for skills relevant to each task:

- `applicant_id` = A unique ID given to each applicant in Phase 1
- `technical_skills` = AI graded score from 0 to 100 for the timed technical task
- `writing_skills` = AI graded score from 0 to 100 for the timed writing task
- `speaking_skills` = AI graded rating from 0 to 10 for the speaking ability in the pre-recorded video
- `leadership_presence` = AI graded rating from 0 to 10 for “leadership presence” in the pre-recorded video

In this phase, we are interested in what factors from the application form affect the AI scoring/rating of technical, writing, and speaking skills and leadership presence. Thus we will be using the data from `phase1-new-grad-applicants-2020.csv`, which has been explained in the previous section. We will have 4 models with the response variables being the scores for technical and writing task, and rating for the speaking skills and leadership presence. For the fixed effects and random effects, we will be using the exact same found in the final model presented for phase 1 except for the variables for cover letter and CV. The reason for this is because every applicant who made it to phase 2 had a cover letter and CV present with their application form. Based on the response variables, we will be using a linear mixed model for each.

From the linear mixed model seen in Table 5, it follows that the most influential fixed effect is GPA followed by relevant skills from extracurriculars. Moreover, each 1.0 increase in GPA increases the score by 8.8 on average. Those with some relevant skills from extracurriculars have a higher score than those who do not by an average of 4.6 and doubled for those with



**Table 5:** Linear mixed model of the  
Technical Skills Score

Characteristic	Beta	95% CI
gpa	8.8	3.9, 14
gender		
Man	—	—
Prefer not to say	23	-1.2, 47
Woman	1.0	-3.8, 5.8
extracurriculars	4.6	-0.39, 9.6
work_experience	-0.62	-6.6, 5.4

CI = Confidence Interval

high relevant skills. We also notice that those who put “Prefer not to say” for their gender have higher score than those who put “Man” by an average of 23. This is not concerning because the sample of those who put “Prefer not to say” is very small and thus can be seen as outliers to the overall average. Besides that, those who put “Woman” for their gender only have a higher score than those who put “Man” by an average of 1.

**Table 6:** Linear mixed model of the  
Writing Skills Score

Characteristic	Beta	95% CI
gpa	9.9	5.8, 14
gender		
Man	—	—
Prefer not to say	5.2	-15, 26
Woman	3.8	-0.23, 7.8
extracurriculars	-1.5	-5.6, 2.7
work_experience	4.1	-0.88, 9.2

CI = Confidence Interval

The linear mixed model in Table 6 shows similar results to the one for the Technical Skill Score with the most influential fixed effect being GPA. Each 1.0 increase in GPA here increases the score by 9.9 on average. Unlike the Technical Skill Score, the second most influential fixed effect for the of the Writing Skill Score is from relevant work experience. Those with some relevant work experience have a higher score than those who do not by an average of 4.1 and doubled for those with high relevant experiences. We also find that those who put “Woman” for their gender only have a higher score than those who put “Man” by an average of 3.8.

The linear mixed model in Table 7 is significantly different from the previous models. The

**Table 7:** Linear mixed model of the Speaking Skills Rating

Characteristic	Beta	95% CI
gpa	-0.30	-0.81, 0.21
gender		
Man	—	—
Prefer not to say	1.6	-0.90, 4.1
Woman	-1.8	-2.3, -1.3
extracurriculars	0.47	-0.05, 1.0
work_experience	0.93	0.31, 1.6

CI = Confidence Interval

model shows that the most influential factor of the rating is from relevant work experience. More specifically, those with some relevant work experience have a higher rating than those who do not by an average of 0.93 and doubled for those with high relevant experiences. The second most influential factor of the rating is relevant skills from extracurriculars, showing a higher rating on average of 0.47 for those who have some relevant skills from extracurriculars compared to those who do not and doubled for those with high relevant skills. Surprisingly, there is a negative coefficient for GPA, showing that those with higher GPAs have a lower rating on average. We also see a negative coefficient for those who put “Woman” for their gender as they have a lower rating compared to those who put “Man” by an average of 1.8.

**Table 8:** Linear mixed model of the Leadership Presence Rating

Characteristic	Beta	95% CI
gpa	0.22	-0.32, 0.75
gender		
Man	—	—
Prefer not to say	-0.44	-3.1, 2.2
Woman	-0.91	-1.4, -0.39
extracurriculars	0.76	0.22, 1.3
work_experience	0.16	-0.48, 0.81

CI = Confidence Interval

The linear mixed model in Table 8 shows that the most influential factor is no other than from relevant skills from extracurriculars. This is followed by GPA and then relevant skills from work experiences. By interpreting the coefficients of the relevant skills from extracurriculars and work experiences, we find that those who have some of the relevant skills for each, compared to those

who do not, have a higher rating on average of 0.76 and 0.16, respectively, and doubled if they have higher relevant skills and/or experiences. As for GPA, each 1.0 increase in GPA increases the score by an average of 0.22. Similar to the Speaking Skills Rating, we also see a negative coefficient for those who put “Woman” for their gender as they have a lower rating compared to those who put “Man” by an average of 0.91.

After analyzing all of the possible factors that can affect the scores and rating given by the AI, we conclude that the top most influential factors are GPA, and relevant skills from extracurriculars and work experiences. We also find that those who put “Woman” as their gender, score higher than those who put “Man” on average for the Technical and Writing Task. On the other hand, those who put “Man” as their gender, were rated higher than those who put “Woman” on average for the Speaking Skills and Leadership Presence. Ultimately the differences between the genders are small relative to the other factors, hence there is no concern about bias in gender parity for the AI scores and ratings.

### Human Rating Differences

The third phase of the hiring pipeline is the final interview. Each line of `phase3-new-grad-applicants-2020.csv` contains the overall job fit rating given by the interviewers:

- `applicant_id` = A unique ID given to each applicant in Phase 1
- `interviewer_rating_1` = First interviewer’s rating from 0 to 100 for overall job fit
- `interviewer_rating_2` = Second interviewer’s rating from 0 to 100 for overall job fit

In this phase, we are interested in what factors from the application form and the AI scoring/rating affect the overall job rating given by each interviewer. Thus we will be using the data from `phase1-new-grad-applicants-2020.csv` and `phase2-new-grad-applicants-2020.csv`, which has been explained in the previous sections. We will have 2 models with the response variables being the overall job fit ratings given to the applicants by the first and second interviewer. For the fixed effects and random effects, we will be using the exact same found in the models presented for phase 2 with the inclusion of the 4 variables which were the response variables. That is, the AI scores for the technical and writing tasks and rating for speaking and leadership presence will be added as fixed effects. We will also scale up the ratings for speaking and leadership presence by 10, so that they are on the same scale as the scores for the technical and writing tasks. Based on the response variables, the model will be similar to that of phase 2, hence we will be using a linear mixed model for each.

**Table 9:** Linear mixed model of Interviewer 1's Overall Job Ratings

Characteristic	Beta	95% CI
gpa	0.62	0.34, 0.90
gender		
Man	—	—
Woman	0.24	-0.08, 0.56
extracurriculars	0.43	0.08, 0.79
work_experience	0.66	0.38, 0.94
technical_skills	0.29	0.28, 0.30
writing_skills	0.27	0.26, 0.29
speaking_skills	0.29	0.28, 0.30
leadership_presence	0.29	0.28, 0.30

CI = Confidence Interval

Looking at the model in Table 9 for the ratings given by Interviewer 1, we only need to look at each coefficient relative to each other. It is evident that the least influential factor is in fact gender. It is seen that on average, those of the gender “Women” have a score that is 0.24 higher than that of those who are the gender “Man”.

**Table 10:** Linear mixed model of Interviewer 2's Overall Job Ratings

Characteristic	Beta	95% CI
gpa	-0.63	-6.3, 5.0
gender		
Man	—	—
Woman	0.02	-6.4, 6.5
extracurriculars	3.7	-3.4, 11
work_experience	-4.9	-11, 0.66
technical_skills	0.42	0.20, 0.64
writing_skills	0.49	0.18, 0.80
speaking_skills	0.14	-0.08, 0.35
leadership_presence	0.12	-0.07, 0.31

CI = Confidence Interval

The same results are evident in Table 10 for the ratings given by Interviewer 2. In fact, the result is most apparent here where on average, those of the gender “Women” have a score that is only 0.02 higher than that of those who are the gender “Man”. The only concerning thing

here is that on average, those with higher GPA and/or more relevant work experience have a lower rating for the second interviewer. Since being listed as “first” or “second” interviewer is arbitrary, there only reason for this is possibly outliers in the data.

All in all, there are no strong evidence that suggest applicants are not judged on their talents and value to the company.

### The Quintessential Element of a Job Offer

Since there are only three phases of the recruitment pipeline, in this section we will be looking at those who were sent an offer letter. Each line of `data/final-hires-newgrad_2020.csv` contains the unique IDs of those who were sent an offer letter at the end of the recruitment pipeline. We are only interested in which applicants got an offer letter after phase 3. For this reason we will create a variable called `offered` and set it to 1, if the applicant got an offer letter after phase 3 and 0, if not. This will be the response variable we are interested in for developing our model. As for the fixed effects, we will be using gender and the two interview ratings. The reason for only using the interview ratings is because we find that the interview rating encompasses all of the data collected from each applicant from phases 1 and 2. Furthermore, the interview is usually the make or break situation for the applicants to prove themselves to the employer that they are worthy of a job offer. Then similar to the previous models, we will add the team the applicant applied for as a random effect.

Analogous to the analysis in phase 1, we will be constructing a generalized linear mixed model, but before that we want to first consider the gender distribution for those who got an offered letter to those who did not. In addition, consider also the marginal probabilities and odd ratios.

**Table 11:** The number of applicants that got an offer letter based on gender

	Did not get an offer letter	Got an offer letter
Man	7	8
Woman	5	2

**Table 12:** The probability of an applicant’s gender given that they got an offer letter

	Did not get an offer letter	Got an offer letter
Man	0.5833333	0.8
Woman	0.4166667	0.2

**Table 13:** The probability of an applicant getting an offer letter given gender

	Did not get an offer letter	Got an offer letter
Man	0.4666667	0.5333333
Woman	0.7142857	0.2857143

Hence from Table 12, given that the applicant got an offer letter, the probability that their gender is a “Man” is 80%, whereas a “Woman” is 20%. We also see that the probability of getting an offer letter for an applicant whose gender is “Man” is 53%, whereas a “Woman” is 29%, from Table 13. Also, the computed odds ratio of Woman vs Man who pass phase 1 is approximately 0.35. That is, the odds of a “Woman” passing phase 1 is less than that of a “Man”. This odds ratio is relatively close to 0, hence there is reason to suspect that there are possible gender biases in those who received an offer letter. Thus lets now look at the model and see if there is a possible case of bias between those who received and not received an offer letter.

**Table 14:** Generalized linear mixed model of applicants who got an offer

Characteristic	log(OR)	95% CI	p-value
gender			
Man	—	—	
Woman	-81	-45,150,252, 45,150,089	>0.9
interviewer_rating_1	25	-5,175,516, 5,175,566	>0.9
interviewer_rating_2	31	-3,694,418, 3,694,480	>0.9

OR = Odds Ratio, CI = Confidence Interval

From the model seen in Table 14, we can see that the coefficients and p-values are absurdly large, however it is still useful to interpret. We see that the coefficient of “Women” is a large negative, where as the ones for the interview ratings are only somewhat large positives. Hence, the odds of the gender, “Women” are essentially 0 compared to “Man”. However, this does not change the fact that the overall job fit ratings from the interview increases the odds of getting an offer letter significantly. This is knowing that the overall job fit ratings given by the interviewers are unbiased based on the analysis in the previous section. In either case, for whatever reason, those who are the gender “Woman”, do not have the a chance at getting an offer letter compared to those who are of the gender, “Man”. Though the process of the recruitment pipeline seemed fair, there seems to be a gender disparity in the selection of those who were sent an offer letter in the end.

## Fairness of Promotion process

Let us now shift focus from the hiring pipeline to the current employees at the company. Black Saber aims to promote an equitable and inclusive work environment based on merit. The board wants to assure that its promotion processes are not only fair, but strictly based on talent to the company. We have been given the data for the company's current employees and their key performance indicators throughout their time at the company marked by financial quarters. We wish to determine which factors contribute to employees moving up the ranks of Black Saber seniority.

## Features Recorded of Black Saber's Current Employees

While employed at Black Saber Software, employee performance indicators are recorded every financial quarter. Each line of `black-saber-current-employees.csv` contains the information of each employee at the financial quarters they were employed:

- `employee_id` = A unique ID given to each employee of Black Saber
- `gender` = "Man", "Woman", or "Prefer not to say" as the only options
- `team` = Divisions of Black Saber staff categorized into: "Software", "People and talent", "Operations", "Marketing and sales", "Legal and financial", "Design", "Data", and "Client services".
- `financial_q` = Financial quarter where following features were recorded.
- `role_seniority` = Hierarchy of employees within each team. Rank order from least senior to most senior: "Entry-level", "Junior I", "Junior II", "Senior I", "Senior II", "Senior III", "Manager", "Director", "Vice president".
- `leadership_for_level` = Quality of demonstrated leadership, taking into account role level. (i.e. "Appropriate for level" requires much less for entry-level employees than for a manager). Categorized into "Needs Improvement", "Appropriate for Level", and "Exceeds Expectations".
- `productivity` = Work output in relation to job description, rated on a 0-100 scale with 50 being satisfactory and above 50 indicating better than expected productivity.
- `salary` = Salary at the given financial quarter (note: these are effective yearly values for the current wage, but don't take in to account previous salary steps in the same year, etc.)

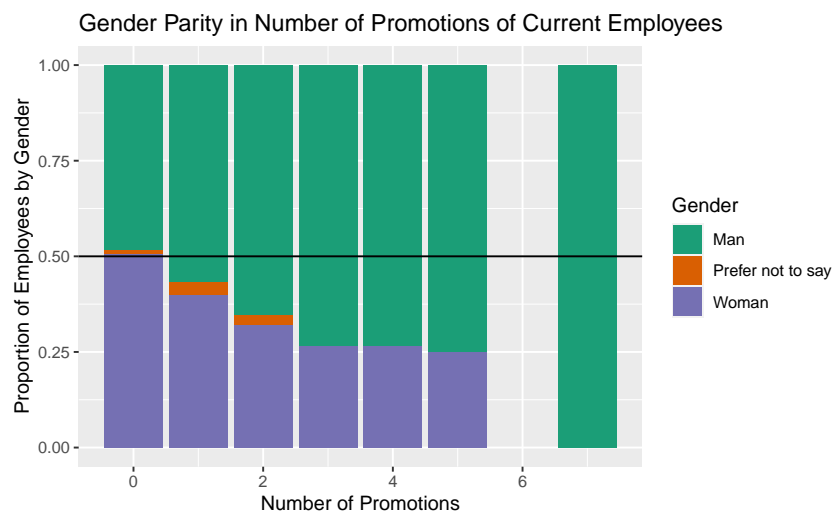
For this question, we are interested in how many promotions an employee received while working at Black Saber Software. For this reason, we will create a variable called `number_of_promotions` counting the change in `role_seniority` of each employee. This will be the response variable we are interested in for developing our model.

**Table 15:** Five Anonymous Employees at Black Saber Software

<i>employee_id</i>	<i>gender</i>	<i>team</i>	<i>financial_q</i>	<i>role_seniority</i>	<i>leadership_for_level</i>	<i>productivity</i>	<i>salary</i>
24527	Man	Client services	2020 Q4	Junior I	Appropriate for level	17	\$33,500
24351	Woman	Legal and financial	2020 Q4	Senior III	Appropriate for level	35	\$57,700
24368	Man	Legal and financial	2020 Q4	Entry-level	Appropriate for level	59	\$34,100
24030	Man	Software	2020 Q4	Senior I	Appropriate for level	65	\$48,500
24016	Woman	Software	2020 Q4	Senior III	Appropriate for level	16	\$56,900

Additionally we will group employees by their ID and each role seniority they have had to create variables averaging productivity, leadership levels as a numeric, and counting number of quarters an employee has worked since they were hired.

Given the response variable is a count we will be consider a Poisson Regression model. Before constructing the model, we want to first consider the gender distribution in number of promotions of current employees. In addition, consider also the joint probabilities.



**Figure 3. Gender distribution in the number of promotions of employees.**  
The line in the middle assumes equal proportion of promotions in both genders.



**Table 16:** The number of employee promotions based on gender

	0	1	2	3	4	5	7
Man	143	176	165	156	55	18	8
Prefer not to say	3	10	6	0	0	0	0
Woman	149	124	81	56	20	6	0

**Table 17:** The probability of an employee's gender given number of promotions

	0	1	2	3	4	5	7
Man	0.4847458	0.5677419	0.6547619	0.7358491	0.7333333	0.75	1
Prefer not to say	0.0101695	0.0322581	0.0238095	0.0000000	0.0000000	0.00	0
Woman	0.5050847	0.4000000	0.3214286	0.2641509	0.2666667	0.25	0

From the figure Gender Parity in Number of Promotions of Current Employees, we see that the proportion of male to female is only balanced at 0 promotions, becoming more male-dominated as we increase in promotions.

Similarly, from Table 17, the only time the probability of an employee being a woman is higher is given the employee has not had any promotions. Conversely, seeing that given a higher number of promotions, the probability of being a man becomes proportionally larger, a gender bias may be suspected however we will withhold any conclusions until we consider our model. Factors such as quarters since an employee was hired may still be a more influential variable as more time at the company could lead to more promotions.

**Table 18:** Likelihood ratio test for Poisson GLMM with ID Random Effect

#Df	LogLik	Df	Chisq	Pr(>Chisq)
8	-1120.880	NA	NA	NA
9	-1113.488	1	14.78515	0.0001205

After creating to variants of a Poisson GLMM model, we conduct a likelihood ratio tests to settle on which model. The results show that we take the Poisson GLMM model that takes gender , average productivity, average leadership and quarters since hiring as fixed effects, and team, role seniority as random effects.

From the model represented in Table 19, we can see gender has an effect on whether or not a employee is promoted. It is evidently the case that women get -0.33 less promotions for every promotion a male employee would receive. This may indicate a gender bias in the promotions of employees at Black Saber.

**Table 19:** Generalized linear model of employee promotions

Characteristic	log(IRR)	95% CI	p-value
gender			
Man	—	—	
Prefer not to say	-0.18	-0.78, 0.42	0.6
Woman	-0.33	-0.49, -0.17	<0.001
avg_productivity	0.00	0.00, 0.01	0.3
avg_leadership_level	-0.54	-1.0, -0.05	0.030
quarters_since_hiring	-0.05	-0.07, -0.03	<0.001

IRR = Incidence Rate Ratio, CI = Confidence Interval

### Fairness of Salary assignment for current employees at Black Saber

Along with the fairness of their promotion practices, the Black Saber board members also want to know that compensation of employees is set by the same standards of merit. Using the same dataset of Black Saber’s current employees, we wish to determine which factors contribute to setting employee salaries.

As with promotion analysis, information of each employee at the financial quarters they were employed will be utilized from `black-saber-current-employees.csv`.

In this section, we are interested in the fairness of salary assignment for current employees at Black Saber. We once again group employees by their ID and each role seniority they have had to create variables averaging productivity and leadership levels as a numeric to see how these factors affect salary assignment.

First, let’s compare gender parity in the salaries of the most recent quarter for every employee. Figure 4 shows, as would make sense, that salary increases as role increases in seniority. It also seems that within the same roles, males tend to be paid more than their female counterparts across almost all positions. However, before we were to reach any conclusions, we should refer to a model.

Comparing nested linear mixed models using likelihood ratio tests, we settled on a model, seen in Table 20, accounting for team and role seniority as random effects, while using gender and average productivity as fixed effects.

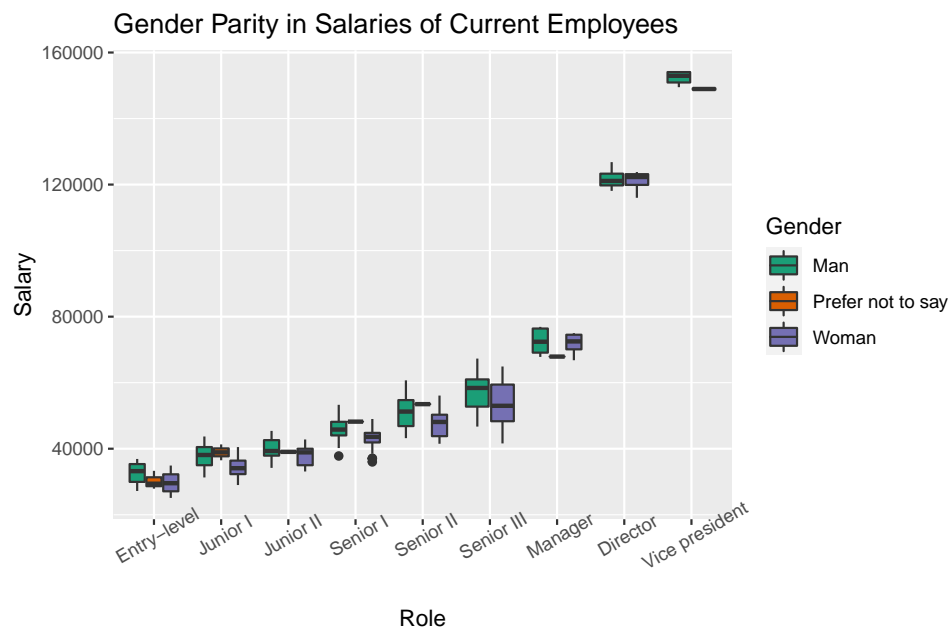


Figure 4. Gender and Salary Distribution based on roles of current employees

Table 20: Linear mixed model of employee salaries

Characteristic	Beta	95% CI
gender		
Man	—	—
Prefer not to say	-1,154	-2,691, 383
Woman	-1,998	-2,412, -1,584
avg_productivity	3.1	-9.8, 16

CI = Confidence Interval

Taking into account the gender and average productivity of a given current employee while randomly selecting based team and role seniority, we see that women are, on average, paid \$1998 less than men. We evidently see that there is a disparity between the salary assignment of current employees as women are being paid less. As a result, we can conclude that the salary processes at Black Saber show evidence of being unfair.

## Discussion

For the first phase of the hiring pipeline, we fitted a generalized linear model to predict whether a candidate passed phase 1 using six fixed effects and one random effect. We found that the most significant factors that contribute to whether a candidate passes phase 1 are GPA, skills from extracurricular activities, and work experience, in the order of most to least importance.

Next, we fitted four linear mixed models for response variables: score for the technical task, scores for writing task, ratings for speaking skills, and leadership presence. Overall, we found similar results to the first phase where GPA and skills from extracurricular activities, and work experience are the dominant factors. Across four assessed aspects, gender has a generally small effect on final ratings. Hence, we can conclude that there is no evidence for gender bias in the AI system for Phase 2.

For phase 3 in the hiring pipeline, AI is no longer involved. Two interviewers are in charge of rating the candidates. We fitted two linear mixed models for two response variables, which are the job rating by the first interviewer and the second interviewer. We found that work experience and GPA are the two most influential factors in predicting the overall job rating by the first interviewer. However for the second interviewer, those with a higher GPA and more work experience are graded with a lower score than those who have a lower GPA and less relevant work experience. We find that gender has a minimal effect on predicting the job rating score of the first interviewer and nearly 0 for the second. Thus, we can conclude that in phase 3, the interviewers have different preferences for qualities that they value in candidates.

The problem arises clearly when looking at the applicants who received an offer letter. We find that there is a higher probability that an applicant whose gender is a “Man”, receives an offer letter compared to a “Women”. In the end, although the scoring and grading is fair, the final decision on who gets a position is gender bias.

For promotion assignment, we had a count response in the form of number of promotions. Initially, visualizing the data showed an increased proportion of men to women given a larger number of promotions. To continue this further, based on our response, we used a Poisson GLM and Poisson GLMM but decided for the GLMM to account for random effects of repeated employee observations and difference in teams. For our GLMM model, we have the number of promotions as the response variable, team and role seniority as random effects and the rest as fixed. Using our model, we see that quarters since hiring and gender had the most effect on number of promotions. Surprisingly, average productivity and average leadership for level did not contribute much to the response. While quarters since hiring would make sense as the longer an employee worked at a company the more promotions they may receive, gender still plays a negative role when an employee is a woman. Women on average received 0.33 less promotions for every promotion received by their male counterparts.

Evidently we see that salary generally increases as role increases in seniority. Initially it seemed that within the same roles, males tend to be paid more than their female counterparts across almost all positions. When looking at a linear mixed model in which we have Salary as the response and gender and average productivity as fixed effects, we see that women, on average, are paid \$1998 less than men. This is taken into account for randomly selecting team and role seniority.

### **Limitations and Concerns**

Findings from this report analysis on Black Saber's hiring, promotion and salary processes are limited to the data provided. Evidently we recognize that race and ethnicity was not information that was collected or provided by Black Saber. While we attempt to perform analysis on fairness in gender disparity across these processes in Black Saber, we did not consider the race and ethnicity of current and possible hired employees. It is suggested that collecting information on race and ethnicity would be beneficial towards analyzing fairness across race/ethnicity for future considerations.

## Consultant information

### Consultant profiles

**Andy Vu.** Andy is a junior consultant with Zeusolutions. He specializes in statistical modeling. Andy earned his Bachelor of Science, Specializing in Mathematics & Its Applications (Probability/Statistics) and Majoring in Statistics from the University of Toronto in 2021.

**Ethelia Choi.** Ethelia is a junior consultant with Zeusolutions. She specializes in data analytics. Ethelia earned her Bachelor of Science, Majoring in Statistics and Mathematics from the University of Toronto in 2021.

**James F. Kanu.** James is a junior consultant with Zeusolutions. He specializes in data visualization. James earned his Bachelor of Science, Majoring in Mathematics and Minor in Computer Science and Statistics from the University of Toronto in 2021.

**Justin Lee.** Justin is a junior consultant with Zeusolutions. He specializes in data analytics. Justin earned his Bachelor of Science, Majoring in Statistics and Mathematics from the University of Toronto in 2021.

### Code of ethical conduct

Zeusolutions strives to be a forefront in professional statistical work while maintaining the integrity, security and privacy of our customers. We believe in executing our work with proper due diligence and informative research. This is why we believe in the following.

1. Maintaining the trust of our customers as a vital part of our company's asset.
2. Privacy of our customers is essential in all our statistical work. We do not disclose or seek personal gain from any information given to us by our customers. Any possible interest that can be regarded as influencing the results of our statistical work will be clearly communicated.
3. Appropriate security measures and practices take place amongst our employees to ensure sensitive information is kept safe.
4. Support for all employees. All statistical work performed with Zeusolutions is appropriately credited to all employees involved.
5. Ensure the disclosure of all limitations and assumptions made in relevance to data of any particular work.

We recognize these importance's and responsibilities we ought to uphold to our customers.