An investigation on Black Saber Software's potential systematic gender bias

A statistical analysis on unfair Hiring selection, Promotion opportunities and Salary increase in Black Saber Software

Dataable Inc.

2021-04-21

Contents

Executive summary **3**
- Background & Aim . . . 3
- Key findings . . . 3
- Limitations . . . 3
- Graphs . . . 4

Technical report **5**
- Introduction . . . 5
- Potential Gender Bias in Hiring Phase . . . 6
- Fewer promotion opportunities for women: an investigation . . . 13
- Potential Gender Bias in Salary . . . 20
- Discussion and Conclusion . . . 22
- Limitations . . . 23

Consultant information **24**
- Consultant profiles . . . 24
- Code of ethical conduct . . . 24

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

Executive summary

Background & Aim

Internal concerns on the potential bias in the hiring and remuneration process in the Black Saber Softwares has been raised by its current employees and board members. In response to the board, current and future employees, this research is carried out by the Dataable Inc. to study whether the decisions of candidates who receive promotion on position, raising of salary and final offers are entirely based on their talent and value to the company without interruption of gender bias.

Key findings

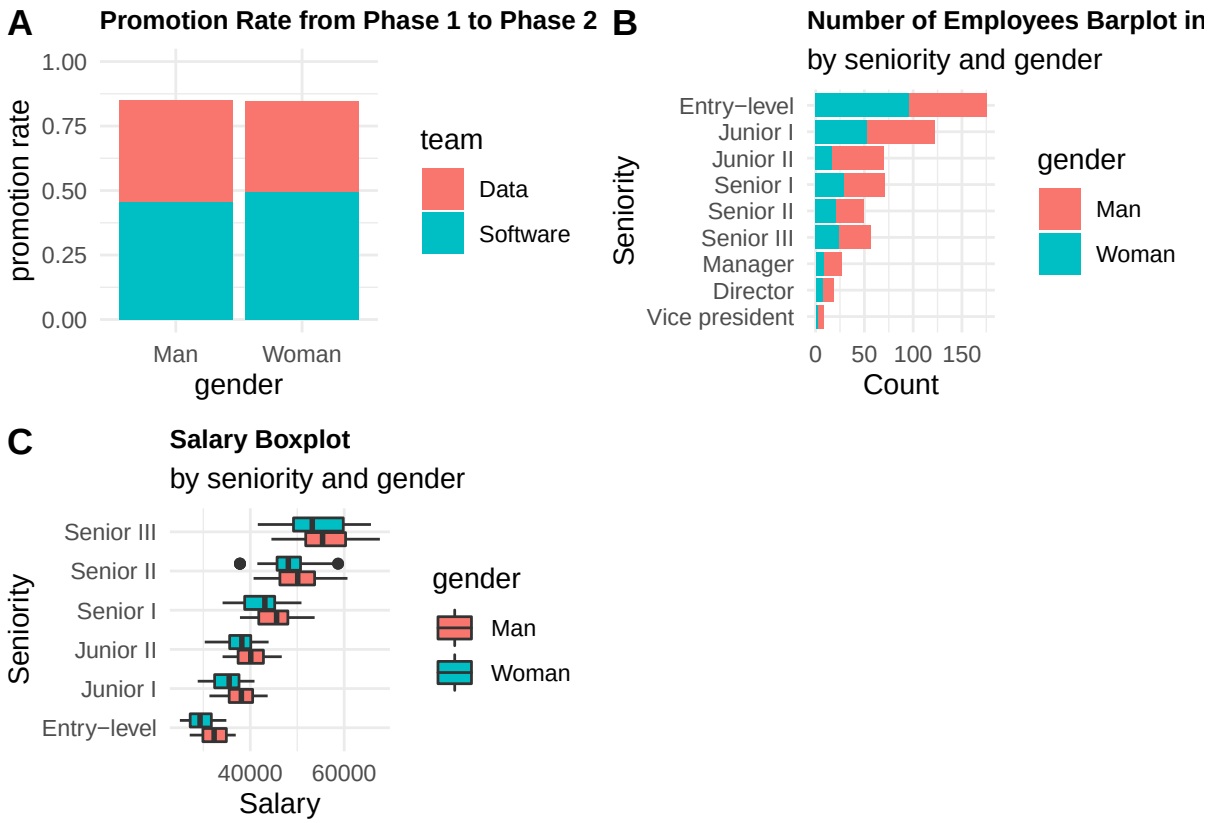
- The data team prefer males than females in phase 1 during the hiring process while software team are more fair with gender in this phase.
- No signal of gender discrimination is identified in phase 2 and phase 3 in both data and software teams.
- For the leadership level rating, only men receive positive feedback i.e. Exceeds expectations and only women receive lower feedback i.e. needs improvement
- Gender bias exists in the promotion scheme and differs by team, but most of their magnitudes are moderate, mostly less than 10% in terms of average counts of promotion.
- After accounting for the effect of the department that one employee is working in, years they have been working (this is just the representation of repeated measurement of employees), the females' average salary is 4854.75 less than what of male, and male's average salary is approximately 45016.94.

Limitations

- Very few data is available in analyzing the potential gender bias in the third phase of hiring process thus results are not completely accurate.
- Alogorithms for AI grading scheme is not provided thus no conclusions can be generate about whether the grading system is genderly biased.
- The group of employees who prefer not to say their gender is excluded from the research because very few people is found to be in this group. Addtioanlly, since the group with gender prefer not to say may include males, females and LGBT group, thus including them in the research is not scientific and we encourage a more specific category about these groups to be established so prejudice on sexual minorities can be studied in depth.

Graphs

Here are some figures of summary statistics:



Technical report

Introduction

In the following technical report, we mainly presented all the statistical details of all the answers to the three research questions listed below. We have excluded the gender "Prefer not to say" since the sample size for this gender is too small for all the three research questions we have studied. In the first research question, we discussed the potential gender discriminations, i.e., whether males and females were facing a different probability of getting recruited into the Black Saber Company. In the second research question, this analysis explores any gender bias in job promotion in the Black Saber Company. Last, we studied whether there is potential gender bias in salary among the current employees while considering the team difference. For all the three research questions, we first presented the statistical methods used in each of the research questions thoroughly, including what they are, why they are used, how they are used, and why they are appropriate for each of the research questions. Second, we showed all the statistical results, such as tables and figures produced by exploratory data analysis, and some statistical summaries like values of coefficients, p-values and confidence intervals. In the last part of the report, we discussed the limitations of our analysis and some potential improvements that can be done in future study as well.

Research Questions

The research questions are listed as following:

- Based on team applied for, whether there exists any bias on women for each hiring phase.
- Based on the current employees promotion data, whether or not there exists any gender bias in job promotion in the Black Saber Software Company.
- Based on the current employees salary data, whether or not there exists any gender bias in salary in the Black Saber Software Company.

For the first research question listed above, we have defined bias as different probability in getting into the next phase between male and female while setting other factors like GPA being the same, and in the second research question, we defined the bias as different opportunities in promotion between male and female while other factors being the same, such as comparing the promotion times in unit of working times, i.e., to set off the effects of different working times. Here a metric of promotion procedure can be the number of times an employee gets promoted in their career, and the scope of gender bias can be a significant counts gap by gender. Also, if such bias exists, does it differ by which team group one belongs to? We would like to use the data

available with a valid model to estimate such gender differences. For the third research question, we wanted to know whether there exists salary difference between male and female after taking into consideration of different teams.

Potential Gender Bias in Hiring Phase

Methods

- Study Design
  - A comprehensive study focusing on the potential gender discrimination was conducted to compare its effect on the Blake Saber Software's hiring process.
- Participants
  - A total of 613 participants involved in the New Grad program who went through the application process with personal information uploaded were included in this research.
- Data Wrangling
  - Considering the fact that Blake Saber Software's hiring process involved 3 separate stages where only qualified appliers will be moved forward, the analysis would also be separated to discover the potential sexual discrimination and thus the data provided were reconstructed for a better visualization.
  - Data with the gender "Prefer not to say" will not be considered in the model since the sample size is too small.
  - The new created variable Promotion in all three phases was measured according to the presence in the next stage with 0 if not pushed forward, 1 being selected for the next stage.
  - Stage one data involves basic information first underwent a wrangling process that appliers with absence in the submitting of cover letter and resume are removed because of the fact that no appliers with no resume or cover letter were moved to phase2. A new column named promotion was added to indicate whether the appliers were moved forward. Considering the fact that the data team and development team have different idea appliers, the table was divided by the department that appliers apply for.
  - Similar to stage one, the data for stage two which collects AI-auto graded skills was also divided into two tables according to the department they applied for and the column named promotion indicating whether appliers were moved to third phase was created.

  - The stage three data which contains interview scores from two interviewers was also granted the new column promotion indicating whether the appliers were selected for the formal employees. Due to the discovery that two scores are similar in magnitude, a new variable named "interviewer rating" was added to identify the overall satisfaction from the company.

- Statistical methods

  - Graph describes the promotion rate from phase 1 to phase 2 based on gender will be presented. Teams will be distinguished on the graph by using difference colour.
  - The conditional table was applied when an overview of the distribution of successful application conditional on each gender are needed from phase 2 to phase 3. By establishing such table, we are ignoring the potential difference between groups i.e. we are assuming holding a similar background between each gender groups, how are the gender affecting the success in each state.
  - A dot plot of promotion versus interviewer rating is fitted to see if there is any relationship between interviewer rating and promotion from phase 3 to final hiring.
  - Generalized linear models and linear models are introduced and fitted to solve the question on gender bias in each phase.
    * For phase 1, two generalized linear models are fitted:

$$promotion_{di} \sim Bernoulli(\mu_{di})$$

$$promotion_{si} \sim Bernoulli(\mu_{si})$$

$$log\frac{\mu_{di}}{1-\mu_{di}} = \beta_{d1}gpa_{di} + \beta_{d2}gender_{di} + \beta_{d3}extracurriculars_{di}$$

$$log\frac{\mu_{si}}{1-\mu_{si}} = \beta_{s1}gpa_{si} + \beta_{s2}gender_{si} + \beta_{s3}extracurriculars_{si}$$

      The subscript d represents the model for data team, and the subscript s represents the model for software team.

      There are two categories for gender, man and woman.

      There are three levels for extracurriculars: 0, 1 and 2. 0 means there is no related extracurriculars, 2 means that there are several high level skills extracurriculars. 1 is the moderate level.

      Generalized linear models are fitted since the response variable promotion has only two values 0 and 1, which follows a bernoulli distribution. Besides, I assume there

is no violation in the independence assumption since gpa, gender, extracurriculars are considered to be independent.

* For phase 2, two generalized linear models are fitted:

$$promotion_{di} \sim Bernoulli(\mu_{di})$$

$$promotion_{si} \sim Bernoulli(\mu_{si})$$

$$log\frac{\mu_{di}}{1-\mu_{di}} = \beta_{d1}techinical\ skills_{di} + \beta_{d2}writing\ skills_{di}$$
$$+\beta_{d3}speaking\ skills_{di} + \beta_{d4}leadership\ presence_{di} + \beta_{d5}gender_{di}$$
$$log\frac{\mu_{si}}{1-\mu_{si}} = \beta_{s1}techinical\ skills_{si} + \beta_{s2}writing\ skills_{si} +$$
$$+\beta_{s3}speaking\ skills_{si} + \beta_{s4}leadership\ presence_{si} + \beta_{s5}gender_{si}$$

Each skills have a scope of 0 to 100, AI autograded.

* For phase 3, two generalized linear models are fitted:

$$interviewer\ rating_{di} \sim \mathcal{N}(\mu_{di}, \sigma_{di}^2)$$

$$interviewer\ rating_{di} \sim \mathcal{N}(\mu_{si}, \sigma_{si}^2)$$

$$\mu_{di} = \beta_{d1}techinical\ skills_{di} + \beta_{d2}writing\ skills_{di}$$
$$+\beta_{d3}speaking\ skills_{di} + \beta_{d4}leadership\ presence_{di} + \beta_{d5}gender_{di}$$
$$\mu_{si} = \beta_{s1}techinical\ skills_{si} + \beta_{s2}writing\ skills_{si} +$$
$$+\beta_{s3}speaking\ skills_{si} + \beta_{s4}leadership\ presence_{si} + \beta_{s5}gender_{si}$$

Interviewer rating stands for the average rating of job fit given by the two interviewers on a scale of 0 to 100.

Linear model is fitted since the response variable is interviewer rating, which is assumed to be normally distributed. Besides, techinical skills, leadership prensence, gender, speaking skills, writing skills are assumed to be independent.

Results

Phase 1 to Phase 2

- Figure 1 below describes the promotion rate from phase 1 to phase 2 based on gender. Note that colour represents different teams that applicants apply for. The promotion rate

of woman is higher than man when they apply for the software team. The promotion rate of woman is lower than man when they apply for the software team. The overall promotion rates are similar within gender.

Figure 1: Promotion Rate from Phase 1 to Phase 2 Based on Gender

- Two GLM models have been fitted. The estimation, p-values and bounds for confidence intervals of each variable from phase 1 are summarized in Table 1 and Table 2 (all values except p-values are exponentiated):

Table 1: Summary Statistics of Generalized Linear model of Phase 1 Data Group

	Estimate	p.value	Lower.Bound	Upper.Bound
(Intercept)	0.000000e+00	0.9889	NA	6.545002e+42
gpa	8.136700e+01	0.0011	9.993	2.405151e+03
genderWoman	1.040000e-01	0.0122	0.014	5.350000e-01
extracurriculars	2.503286e+08	0.9921	0.000	NA

Table 2: Summary Statistics of Generalized Linear model of Phase 1 Software Group

	Estimate	p.value	Lower.Bound	Upper.Bound
(Intercept)	0.000	0.0000	0.000	0.000
gpa	92.974	0.0000	19.457	724.374
genderWoman	2.392	0.1713	0.710	8.954
extracurriculars	21.423	0.0045	3.837	406.336

- In our model with categorical variables, the coefficients can be interpreted as log odd ratios, and after exponentiation, $\frac{\frac{\mu_{female}}{1-\mu_{female}}}{\frac{\mu_{male}}{1-\mu_{male}}} = e^{\beta_{gender}}$ denotes the odd ratio i.e. the percent change in the response for a categorical change in X.
- In table 1, our major variable of interest "gender" as well as the constant term are statistically significant with at the 0.05 level. The gender estimate($\beta_1 \approx 0.694 < 1$) implies that the odds that female applicant who apply for the data team get promoted into next phase are approximately 0.9 times that of male applicants on average while holding other variables fixed. It shows a potential gender discrimination to female in terms of the applicants of data team from phase 1 to 2.
- In table 2, gender is not statistically significant with at the 0.05 level. So we fail to reject that is no evidence of gender discrimination in terms of the applicants of software team from phase 2 to 3.

Phase 2 to Phase 3

- Table 3 is the conditional table of the distribution of successful application conditional on each gender. The table indicates that female applicants have a lower promotion rate than male applicants.

Table 3: Summary of Promotion Rate from phase 2 to phase 3 Conditional on Gender

	not promoted	promoted
Man	0.897	0.103
Woman	0.954	0.046

- Two GLM models have been fitted. The estimation, p-values and bounds for confidence intervals of each variable from phase 2 are summarized in Table 4 and Table 5 (all values except p-values are exponentiated):

Table 4: Summary Statistics of Generalized Linear model of Phase 2 Data Group

	Estimate	p.value	Lower.Bound	Upper.Bound
(Intercept)	0.000	0.0002	0.000	0.000
technical_skills	1.060	0.0258	1.014	1.128
writing_skills	1.080	0.0081	1.027	1.153
leadership_presence	2.173	0.0016	1.431	3.863
speaking_skills	1.721	0.0166	1.158	2.917
genderWoman	0.609	0.6257	0.070	4.245

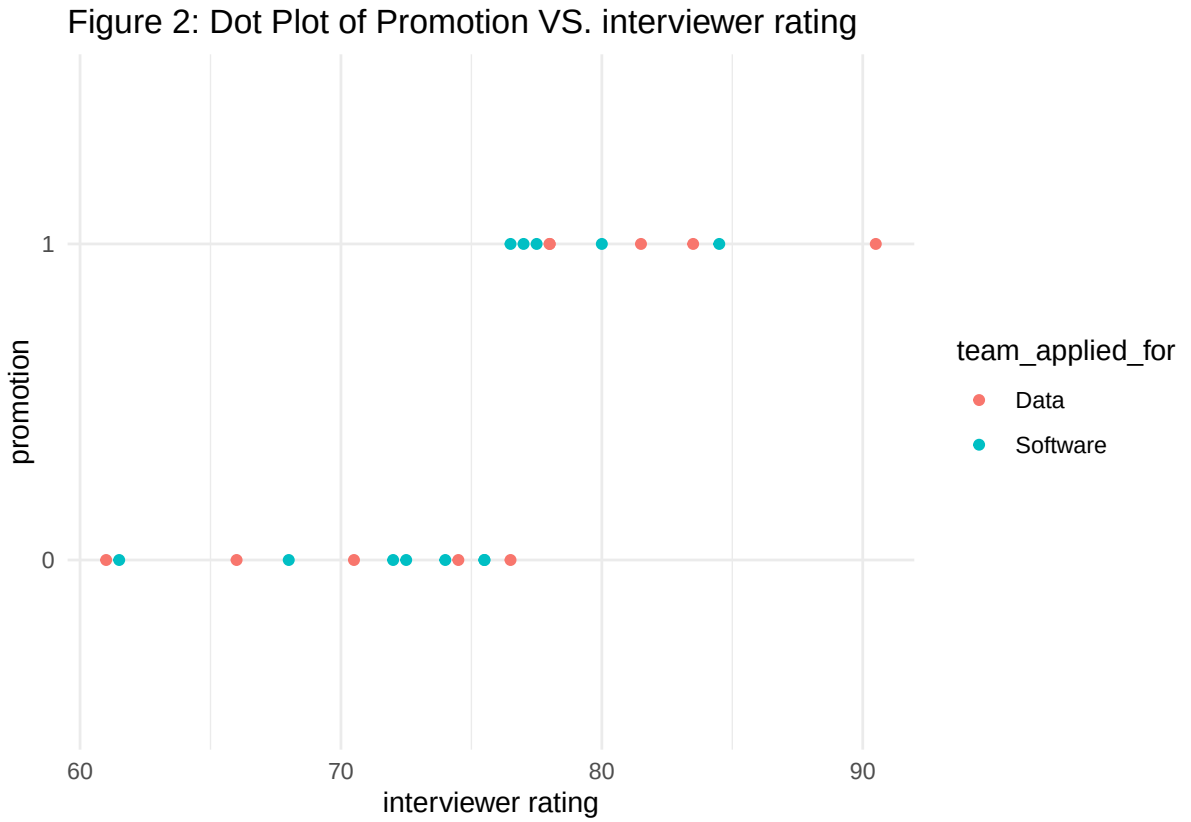
Table 5: Summary Statistics of Generalized Linear model of Phase 2 Software Group

	Estimate	p.value	Lower.Bound	Upper.Bound
(Intercept)	0.000	0.0017	0.000	0.000
technical_skills	1.164	0.0019	1.079	1.317
writing_skills	1.173	0.0185	1.057	1.389
leadership_presence	4.585	0.0056	2.034	19.421
speaking_skills	4.362	0.0024	2.072	14.842
genderWoman	0.220	0.2754	0.010	2.842

- In table 4, our major variable of interest "gender" is not statistically significant with at the 0.05 level. So we fail to reject that is no evidence of gender discrimination in terms of the applicants of data team from phase 2 to 3.
- In table 5, gender is not statistically significant with at the 0.05 level. So we fail to reject that is no evidence of gender discrimination in terms of the applicants of software team from phase 2 to 3.

Phase 3 to Final Phase

- Figure 2 is the dot plot of promotion versus interviewer rating. A positive relationship has been shown between promotion and interviewer rating. Students who get a score over 76.5 are hired by the company, and those who has a score lower than 76.5 are not hired. Moreover, out of two applicants with a score of 76.5, one is hired and the other is not.



- Two linear models have been fitted. The estimation, p-values and bounds for confidence intervals of each variable are summarized in Table 6 and Table 7:

Table 6: Summary Statistics of Generalized Linear model of Phase 3 Data Group

	Estimate	p.value	Lower.Bound	Upper.Bound
(Intercept)	1.265	0.0002	0.000	1.053968e+18
technical_skills	1.512	0.0258	1.112	2.056000e+00
writing_skills	1.469	0.0081	1.230	1.755000e+00
leadership_presence	8.002	0.0016	0.754	8.490000e+01
speaking_skills	6.198	0.0166	0.313	1.226570e+02
genderWoman	1.880	0.6257	0.002	1.573349e+03

Table 7: Summary Statistics of Generalized Linear model of Phase 3 Software Group

	Estimate	p.value	Lower.Bound	Upper.Bound
(Intercept)	0.001	0.7340	0.000	8.401699e+18
writing_skills	1.473	0.0571	0.984	2.205000e+00
technical_skills	1.490	0.0124	1.130	1.965000e+00
leadership_presence	18.651	0.0538	0.937	3.713390e+02
speaking_skills	11.588	0.0906	0.591	2.272710e+02
genderWoman	1.202	0.9452	0.002	6.328090e+02

- In table 6 and 7, technical skills, writing skills, leadership presence, speaking skills are statistically significant at the 0.05 level. While gender is not statistically significant. So holding other variables constant, there is no evidence that the average rating from two interviewers differs from man to woman. So there is no evidence that there is gender discrimination in interviewer rating.
- By the positive relationship between the interviewer rating and promotion, there is no evidence that there is gender bias from phase 3 to the final hire phase.

Fewer promotion opportunities for women: an investigation

Methods

- Study Design
  - We conducted a retrospective analysis to find if there is any gender bias in promotion in terms of the number of times employees achieve higher role advancement.
- Participants
  - A total of 607 Black Saber Software current employees identified with unique ID were included, but those who preferred not to report their gender type were excluded for better male/female comparison. The staff data was kindly provided by Mr. Gideon Blake, the chief people officer of the client company. There are 8 teams with 9 different seniority roles, and the performance of employees is evaluated by two aspects, leadership level and work productivity, using a three-level ranking and a 0-100 numerical scale rating, respectively.
- Data wrangling

– Two new measurements called promotion time and working time, were created in terms of the number of financial quarters. The aim of these two variables is to gauge how many times an employee got promoted and how long they have been in this company in their career, separately. The steps of data manipulation are as follows:

1. Change the names of each variable to make it clearer for the general audience
2. Remove employees with gender type being "Prefer not to say" for better comparison purpose
3. Create two new variables called "promotion time" and "working time"
4. Remove "Q" in the financial quarter column as well as the space in front of "Q", and make it numeric for ordering purpose. Also remove special signs such as "$", "," in the salary column and make it numeric.
5. Join the auxiliary tables together, only left with variables of interest for promotion research question.
6. Remove employees with gender type being "Prefer not to say" for better comparison purpose

– Table 8 and Table 9 are samples of observations before and after wrangling, respectively:

Table 8: The first five observations in the initial dataset

employee_id	gender	team	financial_q	role_seniority	leadership_for_level	productivity	salary
24600	Man	Client services	2020 Q4	Entry-level	Appropriate for level	43	$36,800
24600	Man	Client services	2020 Q3	Entry-level	Appropriate for level	42	$36,800
24599	Man	Client services	2020 Q4	Entry-level	Appropriate for level	61	$29,600
24599	Man	Client services	2020 Q3	Entry-level	Appropriate for level	60	$29,600
24598	Woman	Client services	2020 Q4	Entry-level	Appropriate for level	73	$28,700

Table 9: The first five observations in the dataset after manipulation

employee_id	gender	financial_q	role_seniority	team	promotion_times	working_time
23994	Man	20162	Manager	Software	2	19
23995	Man	20141	Senior II	Software	3	28
23996	Man	20153	Senior II	Software	3	22
23997	Man	20184	Director	Software	0	9
23998	Man	20171	Senior II	Software	2	16

- Statistical Methods

– To compare the differences of the counts of promotion by gender, we decided to estimate a negative binomial generalized linear model, after a long model selection process which is discussed later. The expression of the model is the following:

$$log(\frac{\lambda}{working\ time}) = \beta_0 + \beta_1(gender) + \beta_2(team)$$

which is equivalent to

$$log(\lambda) = \beta_0 + \beta_1(gender) + \beta_2(team) + log(working\ time)$$

where λ is the mean count of promotion, *working time* is the total number of financial quarters one has stayed in this company, and other variables have been introduced previously. The negative binomial model can be considered as a Poisson model where λ is random and follows a Gamma distribution.

– We took *team* as a fixed effect due to the fact that every employee always gets promoted within their initial team that hired them and that different team groups are likely to have different promotion schemes.

– The model included an offset term $log(working\ times)$ since different employees may have worked for this company for different years. For example, a vice president has been promoted 6 times over his 10-year career in this company, whereas an entry-level worker hasn't got any promotion just because he is newly employed yesterday! $\frac{\lambda}{working\ time}$denotes the promotion time counts in terms of the rate per financial quarter for each employee, which accounts for the differences in working time.

– Since the negative binomial model is indeed a Poisson model whose parameter λ is random, it suffices to check the Poisson model assumptions with an adjustment in the model=variance assumption:

  * Poisson Response. The response variable is a count of promotion times per quarter, so it is a valid Poisson response.
  * Independence. Work performance and productivity of employees are the main metrics of promotion. Some may argue that promotion chances differ by team, but we will show via Likelihood Ratio Test that such team grouping effects are not statistically significant by a generalized linear mixed model. As a result, it is plausible to assume that promotion counts are independent of each other.
  * Linearity. Linearity with respect to $log(\lambda)$ is difficult to discern without continuous predictor variables, thus not the main concern here.
  * Mean = Variance. Check if the empirical means and variances of the count of promotion times are approximately equal for the overall data(Table 10) as well as for each team(Table 11):

Table 10: Table C: Overall mean and variance

overall.mean	overall.variance
0.938	1.374

Table 11: Mean and variance across team

team	mean	variance
Client services	0.6666667	1.0677083
Data	0.9078947	1.0980702
Design	1.4117647	1.5073529
Legal and financial	0.8333333	0.8333333
Marketing and sales	0.8679245	1.5252471
Operations	1.0500000	1.4151899
People and talent	1.4444444	2.0256410
Software	1.0681818	1.5449341

We can see from Table 11 that most of the variances are moderately larger than means and there are different variability among team groups, which is an evidence of the violation against the mean=variance assumption. This is also the purpose/reason why we apply the negative binomial model to deal with over-dispersion.

In terms of model selection, we first considered a negative binomial model for which $log(\frac{\lambda}{working\ time})$is only linear in gender. Both the gender and constant terms are significantly different from zero, which is an indication of gender difference in promotion counts. Next, we added the "team" term to account for group difference in promotion and applied the likelihood ratio test for model comparison. The <0.05 p-value indicates that it is statistically significant to add the variable "team" as a fixed effect.

```
##                Estimate Std. Error   z value     Pr(>|z|)
## (Intercept)  0.3553328 0.03436913 10.338721 4.707793e-25
## genderWoman -0.3693147 0.05903704 -6.255643 3.958803e-10
```

```
## Likelihood ratio test
##
## Model 1: promotion_times ~ gender
## Model 2: promotion_times ~ gender + team
##   #Df  LogLik Df  Chisq Pr(>Chisq)
```

```
## 1   3 -1896.2
## 2  10 -1878.0  7 36.543  5.724e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results

First, let's do some explanatory data analysis to get an intuition of our promotion research question: Is there any existing gender bias against women in career promotion?

Table 12: Number of employees

gender	count
Man	340
Woman	257

From Table 12, the gender ratio of employees between men and women is about 3:2. There are ~ 100 more male employees and we should take that into account when we think about gender ratios of other variables of interest. To have a further breakdown of promotion regarding different teams,

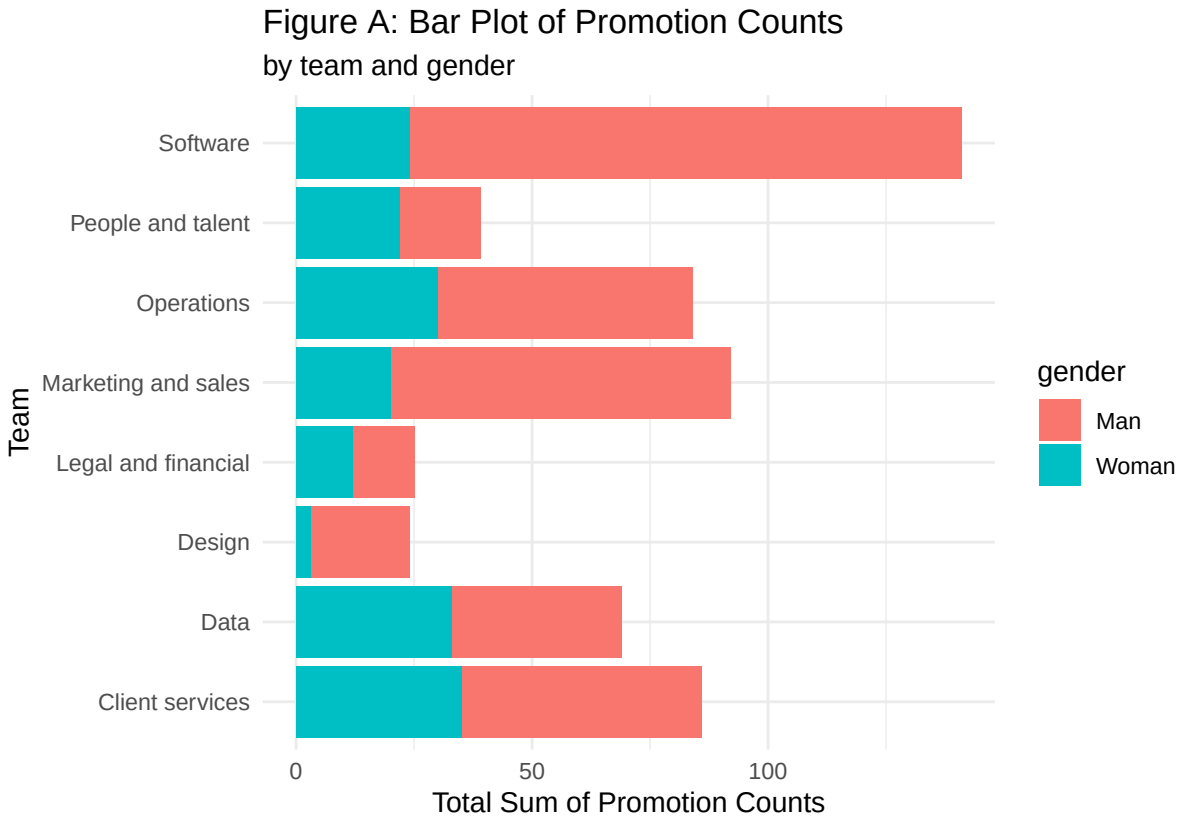


Figure A illustrates the total number of promotion all of the team members have experienced across team. The ratio patterns are quite different across team, which possibly implies different gender bias situations in different teams.

So we fit the negative binomial GLM model as proposed in the Methods section. The estimation, p-values and bounds for confidence intervals of each variable are summarized in Table 13 (all values except p-values are exponentiated):

Table 13: Summary statistics of Negative Binomial Model

	Estimate	p.value	Lower.Bound	Upper.Bound
(Intercept)	1.192	0.0192	1.027	1.378
genderWoman	0.694	0.0000	0.616	0.781
teamData	1.130	0.2415	0.920	1.387
teamDesign	1.533	0.0053	1.130	2.061
teamLegal and financial	1.127	0.4118	0.842	1.495
teamMarketing and sales	1.023	0.8184	0.844	1.239
teamOperations	1.287	0.0111	1.059	1.564
teamPeople and talent	1.910	0.0000	1.487	2.442
teamSoftware	1.285	0.0056	1.077	1.536

In our model with categorical variables, the coefficients can be interpreted as log risk ratios i.e. $\beta_1 = log(\frac{\lambda_{female}}{\lambda_{male}})$, and after exponentiation, $\frac{\lambda_{female}}{\lambda_{male}} = e^{\beta_1}$ denotes the risk ratio i.e. the percent change in the response for a categorical change in X. Note that gender=0 for male and gender=1 for female, and there is an indicator function for each team. Also notice that the underlying gender is men and the default team is Client services, so the exponential of estimates can be deemed the percentage change in the mean number of promotion times relative to a male employee in the Client services team.

As for the hypothesis testing, our major variable of interest gender as well as the constant term are statistically significant with at the 0.05 level. The gender estimate($\beta_1 \approx 0.694 < 1$) implies that female employees are expected to have promotion counts approximately 0.7 times that of male employees on average if she belongs to the Client service team. Equivalently, she encounters a 30% difference in promotion counts fewer than men employees.

Moreover, 4 out of 6 team coefficients also report statistical significance in estimation: Design($p = .005$), Operations($p = .011$), People and talent($p = 3.05e^{-7}$) and Software($p = .006$), all of which have higher promotion chances for at least 10%. Such deviations across team confirmed our observations from the Figure A, that is, differences in gender bias in terms of promotion possibilities across team indeed exist in the Black Saber. For example, an female worker may have promotion times $0.694 * 1.285 \approx 0.892$ times that of male workers in the Software team i.e. a nearly 10% gender gap.

As for the confidence intervals, our interpretation results are the same as using p-values. Any interval containing $e^0 = 1$ is an indication of failure to reject the null hypothesis that the corresponding coefficient estimate is zero i.e. no effect on the mean number of promotion counts. The exact interval values are displayed in Table F.

Potential Gender Bias in Salary

Method

- Study Design
  - A comprehensive study focusing on finding whether or not there is unfairness, i.e., bias, in the salary between gender in the Black Saber Software Company.
- Setting
  - This study used the data that contains the employees' salary starting from the date when he or she was recruited by the company, ending in the fourth quarter, 2020.
- Participants:
  - A total of 6789 observations with repeated measurements on the same employees who are currently working for the company, with gender only male and female since the sample size of the employees for Prefer not to say is too small.
- Data Sources
  - The data includes the quarterly salary for each current employee in Black Saber Company, with their id, gender, team, salary's corresponded financial quarter, role seniority for each financial quarter, whether or not appropriate for level in each financial quarter, and productivity scores for each financial quarter
  - To analyze the potential bias, i.e., the difference in salary between male and female while holding other variables fixed, we excluded the gender class "Prefer not to say". Also, the salary term has been turned into the numerical form.
- Statistical method
  - From the data set we can see that, for salary variable, we can consider it as a continuous variable, and since there are repeated measurements for a single person and within each team, the salary tends to be quite similar, we could not regard salary as independent, and therefore, we introduced two random effects into our model, one is employee_id for dependence within the same employee, and the other is team for dependence within a team. Also, since we were interested in salary difference between male and female, we would only contain the gender as our fixed effects. Therefore, we have proposed a linear mixed model as following:

$$Y_{salary,ij} = \beta_0 + \beta_1 X_{gender,i} + b_{employee_{id},i} + b_{team,i} + \epsilon_{ij}$$

– For the model assumptions check, we can see that from the residual vs. fitted values plot, the linearity and constant variance of residuals hold quite good, and so such a linear mixed model is appropriate for the data set.

Results

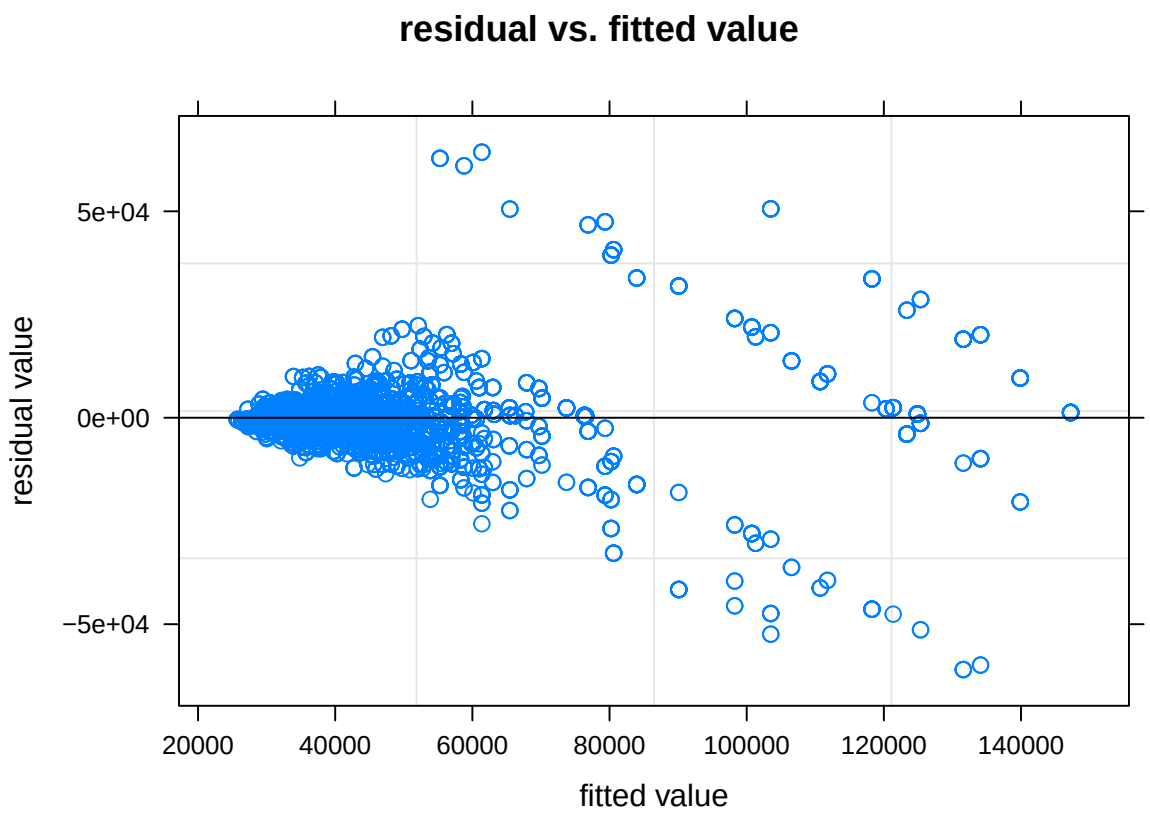
- From the Table 14 that presents the mean salary for male and female, we can tell that there are approximately 3000 dollar difference between male and female. Therefore, such difference exists in a general way.
- Further, we can see from the graph for male and female salary in different seniority roles that, in each seniority of the role, the salary difference still exists between male and female. Notice that here we excluded the seniority for the top three level since the sample sizes for them are too small.
- Last, we can see from the model output that, with consideration of random effects, the salary for female is on average 4854.75 dollar less than male in Black Saber Software Company approximately, and also, the male's average salary is around 45016.94 after taken into account for the random effects of repeated id and similarity within the same team. The results are presented in Table 15.
- From the confidence intervals for random and fixed effects, we can see that the confidence interval for each effect is not too wide to build significance and there is no 0 in the confidence interval. Therefore, we can say that our effects are all statistically significant. The confidence intervals are shown in the Table 16.

Table 14: Fixed effects coefficients

	Values
(Intercept)	45016.936
genderWoman	-4854.749

Table 15: confidence intervals for effects

	2.5 %	97.5 %
.sig01	15597.452	17534.769
.sig02	602.293	5905.154
.sigma	8215.661	8509.717
(Intercept)	42190.065	48208.039
genderWoman	-7683.678	-2048.044



Discussion and Conclusion

This technical report has explored the potential gender bias in three different parts in the Black Saber Software Company: hiring, promotion and salary. We have proposed several statistical models to analyze the data provided by the company thoroughly and we have found that, for the hiring process, women applicants who applied for the data team in phase 1 have a lower average promotion rate to the next phase when holding other variables in phase 1 fixed. Besides, there was no evidence that shows a gender bias during the hiring phase. For the promotion process, however, according to our data exploration and model analysis procedures, we found that the gender discrimination against women appears in the Black Saber Company in terms of different degrees of decreases in promotion opportunities for different teams. Female workers are less likely to get promotion in their career, but this phenomenon does not hold for all teams of the company and the gap magnitudes are moderate. Lastly, for the salary process, we have found indeed there is statistically significant difference between male and female in salary among current employees in the Black Saber Software Company, and so there is bias in salary between male and female.

Limitations

- Hiring
  - The last phase in which the appliers are finally considered according to their behaviour in the interview is to be the most biased stage in the hiring process since it is using human-rated information to decide where the final offer goes to. However, the sample is considerably small therefore we must make clear that the conclusion which we get that gender discrimination does not exist may be a coincidence case. In order to improve this situation, we kindly suggest that a larger sample can be provided so a conclusion that is more representative can be generated. Besides, Algorithms for AI grading scheme is not provided thus no conclusions can be generated about whether the grading system is gender-biased.
- Promotion
  - The linearity assumption is vague in this context since the predictors are categorical variables, not continuous ones. If this assumption does not hold, it is dangerous to use this Negative Binomial model and conclude the existence of the gender bias in the Black Saber. Such misinterpretation would be costly and harmful to the company, especially for its reputation and goodwill.
  - For future consideration, we suggest adding another fixed effect called "role seniority" to account for the fact that higher initial job position is an indication of fewer chance of promotion. For example, if one joins the company as a director, there is at most 1 possible promotion count for him to achieve i.e. being a vice president. On the other hand, if one gets employed as an entry-level worker, it is possible for him to get promoted 8 times and become a vice president as well.
- Salary
  - In this analysis, we did not explore the potential bias in the gender "Prefer not to say" due to very small sample size. However, no one in the company should be ignored and so in the future study, we should collect enough data for this gender and to explore whether the company has bias on people with this gender.
  - Also, we did not include seniority of role, leadership for level and productivity in our linear mixed model. Therefore, in the future study, we could include these terms as fixed effects to see the difference in salary in different roles, in different quality of leadership, and in different productivity scores.

Consultant information

Consultant profiles

Chen Zhang. Chen is a senior statistical consultant with Dataable. He specializes in database management and data architecture. Before joining Dataable, he earned his Ph.D. in computer science from Carnegie Mellon University in 2024.

Zian Lu. Zian is a senior consultant with Dataable. He specializes in statistical communication and data visualization. He received his Ph.D in Organizational Behavior and Statistics from Harvard University in 2025.

Tongfei Zhou. Tongfei, FRM is a financial services risk management senior consultant with Dataable. He specializes in risk management and reproducible data analysis. He received his Ph.D. in Operations Management from the Sloan School of Management at the Massachusetts Institute of Technology in 2025.

Yichen Ji. Yichen is a senior data scientist with Dataable. He specializes in statistical inference and mathematical modeling. He earned his Ph.D. from the Industrial Engineering and Operations Research (IEOR) department at Columbia University in 2026.

Code of ethical conduct

The Dataable Code of Ethical Conduct is one of the approaches we put the values of Dataable into practice. This code provides guidelines and principles our team members follow in statistical consulting.

- Business Integrity

Dataable shows zero tolerance towards any misconduct under the code of integrity including data fraud, intentional manipulation of Analysis results, any forms of bribery, and corruption. We strive to record and present the steps and every aspect of the results that is significant under rigorous scientific reasoning. Suppliers shall respect the intellectual property of others.

- Data Privacy and confidentiality

Suppliers shall protect the privacy of our client's data according to Canadian data protection regulations and acknowledge the potential loss caused by leaking and unauthorized exposure of the data. However, we have also acknowledged the legal limitations on privacy and confidentiality assurances and do exercise further protection where they may not apply.

- Social and Environmental responsibilities

Suppliers shall never use forced labor to lure in school students with internship opportunities to get free or underpaid labor force. We are also acknowledged with our role in the environment and strives to reduce our greenhouse emissions and refuse to collect or analyze the data from animal experiments.