Consultation for Black Saber Software

Gender parity in hiring, promotion and salary

Report prepared for Black Saber Software by ThreeOThree Consultant Group

2021-04-21

Contents

Executive summary

This report is committed to analyzing the general hiring, promotion, and salary distribution of the company Black Saber Software. Specifically, this report will test the AI-automated service that intervenes in the company's hiring process and finds out if the system has any unfair rating standards for certain applicants. Also, this report will test and discuss if the promotion process and salary distribution are mainly depending on employees' talent and value to the company. The report uses data (of applicants and current employees) provided by Black Saber Software and proposes four relevant research questions. For each of the hiring phases, is gender a determinant factor rather than their talents for applicants entering the next phase? For the current employees, are males being preferred at a higher position on average? Whether the average productivity of an employee is a crucial factor that affects his/her position? Is the company's salary distribution mainly depending on one's value produced or is it partially affected by the employee's gender difference?
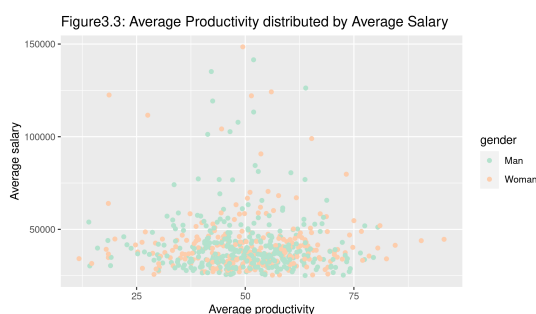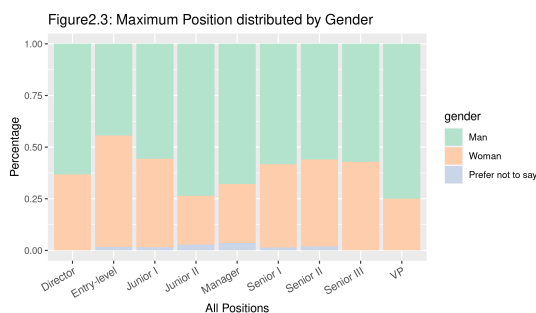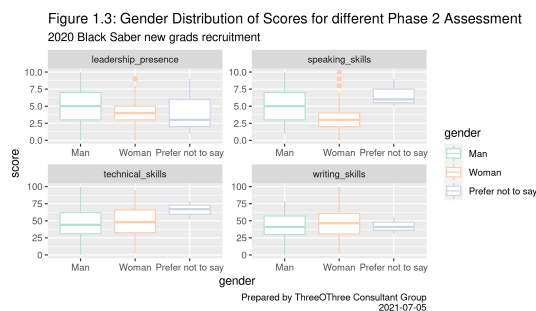
Key Findings:

- Hiring:

    - Gender is not considered as a determinant factor in all three phases of hiring. In addition, there is no gender bias for the AI automated selection pipeline analyzing the pre-recorded video of applicants. (Refer to figure 1.3 listed below)
    - For applicants who successfully enter phase 2 from phase 1, the passing status is based on the applicants' GPA, extracurriculars, and working experiences.
    - AI automated selection pipeline is based on talent of the applicant. Each point increase in the score assessed by the system will increase the probability of hiring by 1 - 2 times. Meanwhile, the interview phase is not gender biased either.

- Promotion:

    - Among the current employees, males are more preferred at higher positions on average. There is a gender bias on the overall promotion process in the company. (Refer to figure 2.2 listed below)
    - The software team and design team are the two teams with the highest percentage of males, and they are also the two teams with the highest average promotion rate.
    - The average productivity is not a significant reason for their promotion, which means that an employee who produces more outputs might not necessarily get a promotion.

- Salary:

    - Current employees' salary are highly related to their average seniority (position), and salary is also partially affected by the employee's gender.

– The ratio of those having high productivity among females is higher than the ratio of the males. However, the ratio of those earning high salaries among females is lower than the ratio among males. (Refer to figure 3.3 listed below)

Limitations:

- When dealing with the interview phase (the third round) in the hiring process, the interviewer's grading criterion is unclear (maybe subjective/biased).
- In the data collected, there is no specific standard to qualify employee's productivity. Thus, it may lead to some inaccurate or biased analysis related to productivity. (e.g., an employee in a different team and with a different position has diverse work type/workload)
- Currently, the number of male employees is higher than that of females which may lead to an inevitable bias when analyzing gender related problems.

Figures:

Figure 1.3: Gender Distribution of Scores for different Phase 2 Assessment
2020 Black Saber new grads recruitment

Figure2.3: Maximum Position distributed by Gender

Figure3.3: Average Productivity distributed by Average Salary

Technical report

Introduction

It is no doubt that the hiring and remuneration processes concern potential bias and companies are challenged to develop fair processes that are based on talent and value to the company. In this report, we took a deep look at the topic of gender parity on hiring, promotion, and salary processes. We also conducted an analysis of the deterministic factors on hiring, promotion, and salary processes to evaluate the rationality of Black Saber Software company's hiring and incentive mechanisms.

Research questions

- Potential bias in hiring process

  - For each of the hiring phases, is gender a deterministic factor rather than their talents for applicants entering the next phase?

- Potential bias in promotion process

  - Part a: Are males being preferred at higher position on average?
  - Part b: If the mean productivity of an employee is a key factor affect his/her position in this company on average?

- Potential bias in salary

  - Does evidence that current employees' wages depend on their gender rather than their value to the company exists in the Black Saber Software Company?

Hiring: gender distribution in hiring phases

Background

This section mainly focuses on the AI-automated hiring process operated by the Black Saber Software. From the data given, there is a total of 3 hiring phases to select adequate candidates. More specifically, the first round comprehensively estimates one's background, which includes one's gender, the team applied, GPA, presence of cover letter and cv, and relevance of past extracurricular and work experiences. Those who proceed to the second phase are asked to complete a couple of timed tasks, through which technical and writing skills are evaluated by AI. Meanwhile, the rating of speaking skills and leadership presence is graded based on one's pre-recorded video. The last round is a face-to-face interview, during which the two arbitrary interviewers assess whether each applicant's performance fits the current job.

Research question

To determine whether the hiring process is potentially biased, we want to explore: Is gender a deterministic factor in each hiring phase?

Methods

Study Design, setting, participants

We analyzed a total of 613 freshly graduated applicants for the 2020 new grad recruitment program for Black Saber company. There are totally 3 hiring phases, with 4 separated datasets provided by Black Saber Software. The first dataset provided includes all available candidates, its source data is from phase1-new-grad-applicants-2020.csv, which includes 8 variables:

- `applicant_id`: A unique id assigned in Phase 1
- `team_applied_for`: the team the candidate applied for, Software or Data
- `cover_letter`: 0 or 1, 0 for absence of the applicant's cover letter, and 1 for presence
- `cv`: 0 or 1, 0 for absence of the applicant's cv, and 1 for presence
- `gpa`: Grade point average, 0.0 to 4.0
- `gender`: Gender of the applicant (Man, Woman, Prefer not to say)
- `extracurriculars`: an estimate of extracurricular experiences assessed by a propriety phrase bank, where 0 for no extracurriculars or no relevant or skillful extracurriculars, 1 for some relevant experiences, and 2 for highly relevant or skill-building experiences
- `work_experience`: an estimate of work experiences assessed by a propriety phrase bank, where 0 for no work experience or no relevant or skillful experience, 1 for some relevant work experiences, and 2 for highly relevant or skill-building work experiences

The file phase2-new-grad-applicants-2020.csv is the source data recording candidates who passed the first round, which includes 5 variables:

- `applicant_id`: A unique id assigned in Phase 1
- `technical_skills`: 0 to 100, graded according to a timed task by AI
- `writing_skills`: 0 to 100, graded according to a timed task by AI
- `speaking_skills`: an estimate of the candidate's speaking skills through the pre-recorded video, graded by AI
- `leadership_presence`: an estimate of the candidate's leadership through the pre-recorded video, graded by AI

The file phase3-new-grad-applicants-2020.csv is the source data recording candidates who successfully proceeded to phase 3, which includes 3 variables:
* `applicant_id`: A unique id assigned in Phase 1 * `interviewer_rating_1`: 0 to 100, an estimate of job fit by the first interviewer * `interviewer_rating_2`: 0 to 100, an estimate of job fit by the second interviewer

The source data final-hires-newgrad_2020.csv provides the list of candidates that are finally hired, with 1 variable:

- `applicant_id`: A unique id assigned in Phase 1 To investigate the topic of gender parity, we took a close look at the three phases of hiring processes separately.

Data description and data wrangling

- From phase 1 to phase 2, we concern about whether or not one has passed under the first-round algorithm. Therefore, we left-join phase 1 and 2 to keep the background information for all phase-1 candidates. Then, a new categorical variable called "Result" is created, with "Proceed" for those who proceeded to Phase 2, and "No" for those who got blocked in the first round.
- From phase 2 to phase 3, we particularly examined the applicants' features rated by the AI automated selection pipeline. We do not know the algorithm behind it, but a higher score means better performance. As stated before, we also included a new categorical column called 'status' with two levels, successfully passing the second phase or failing phase 2. We also linked the corresponding gender to all the applicants who entered phase 2.
- For applicants who entered phase 3, the interview phase, we checked two interviewers' ratings. The rating of job fit is ranging from 0 to 100. We also included the gender information and passing status as potential indicators in the model.

Statistical method

We estimated a generalized linear mixed model to study how various features of applicants affect their passing status as we assumed each applicant is independent. Since the passing status only has two distinct levels, we treated it as a binary response variable and analyzed it using the logistic regression model. We also accounted for the potential random effect by the applicant id and the within applied team variation.

- The first generalized linear mixed model built for phase 1 is:

$logit(Pass_{ij}) = \beta_0 + \beta_1 gender_{ij} + \beta_2 cover\_letter_{ij} + \beta_3 cv_{ij} + \beta_4 GPA_{ij} + \beta_5 extracurriculars_{ij}$
$+ \beta_6 work\_experience_{ij} + b_i team\_applied\_for_{ij}$

Where i indicates the different team applied for and j indicates the j_{th} applicant in the team they applied for. $team\_applied\_for_{ij}$ is the random effect and we assumed it follows a multivariate normal distribution. All other variables are fixed effects.

- For all the applicants from phase 2, we built a model as follows:
$logit(Pass_{ij}) = \beta_0 + \beta_1 gender_{ij} + \beta_2 technical\_skills_{ij} + \beta_3 leadership\_presence_{ij}$
$+ \beta_4 speaking\_skills_{ij} + \beta_5 writing\_skills_{ij} + b_i team\_applied\_for_{ij}$

Where i indicates the different team applied for and j indicates the j_{th} applicant in the team they applied for. $team\_applied\_for_i$ is the random effect and we assumed it follows a multivariate normal distribution. All the other variables are the specific features for the i_{th} applicant.

- Finally, we analyzed the model from phase3 to final hire:
$logit(Pass_{ij}) = \beta_0 + \beta_1 gender_{ij} + \beta_2 interviewer\_rating\_1_{ij} + \beta_3 interviewer\_rating\_2_{ij}$
$+ b_i team\_applied\_for_{ij}$

Again, where i indicates the different team applied for and j indicates the j_{th} applicant in the team they applied for. $team\_applied\_for_i$ is the random effects and we assumed it follows a multivariate normal distribution.

After examining the basic model, we dropped the gender covariates in all three models separately and refitted the generalized linear model with everything else unchanged. We also performed a likelihood ratio test on each pair of the full basic model and the gender reduced model to test whether the model excluded gender conveys the same information as before.

Results

We first took a look at the gender percentage for applicants passing or failing each round of hiring.

Figure 1.1: Gender Percentage for Applicants Passing Each Round Hiring
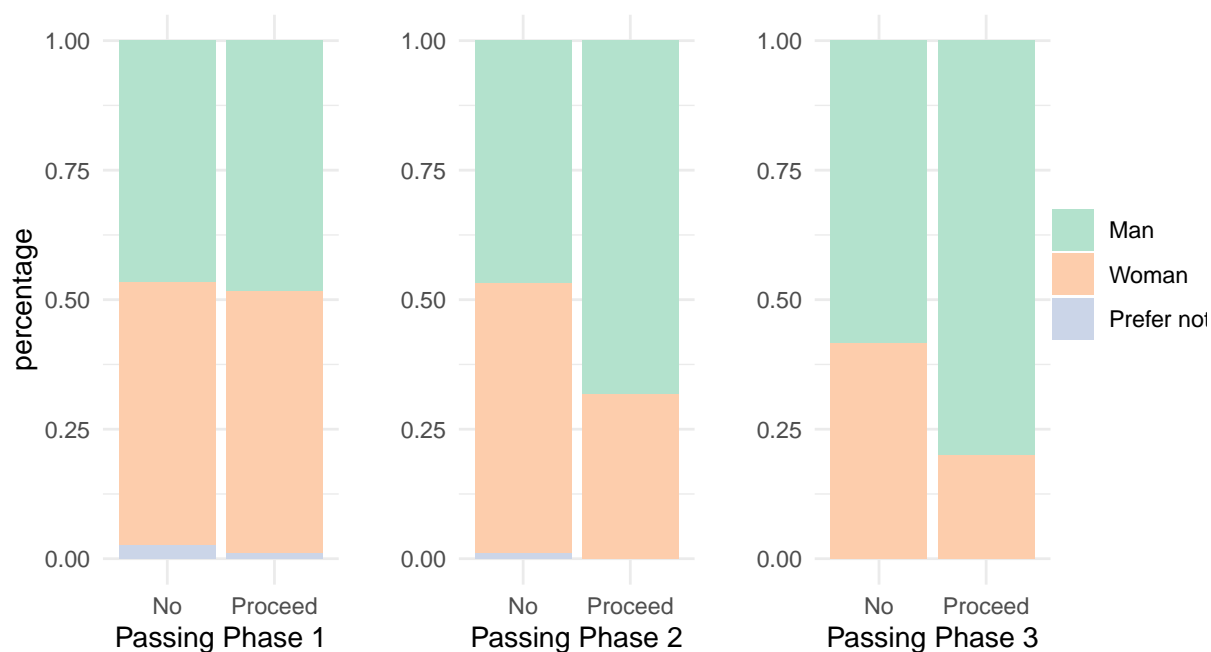
Source: 2020 Black Saber new grads recruitment
Prepared by ThreeOThree Consultant Group
2021−07−05

From figure 1.1, males and females had a similar share on getting into the second round, while males occupied a lot more passes in both second and third round. For those who prefer not to say about their genders, most of them got blocked during phase 1, and none of them got into the final round.

Then, we have a look what the GPA distribution for candidates passing the first round or not:

Figure 1.2: GPA Distribution for Different Levels in Phase 1

2020 Black Saber new grads recruitment
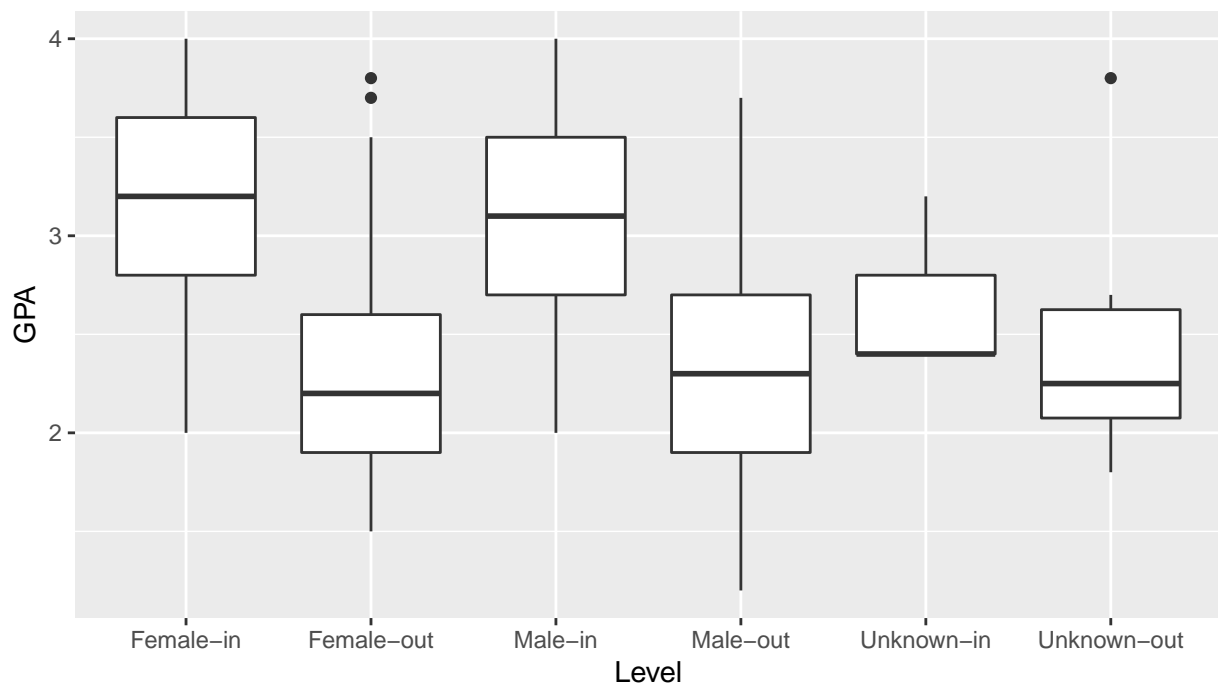
Prepared by ThreeOThree Consultant Group
2021−07−05

Figure 1.2 illustrates the distribution of GPA for candidates who are grouped according to their gender and result of Phase 1. It seems that, for those who didn't get into the second round, these 3 genders have similar average gpas. While for those who proceeded to the next round, males and females have relatively higher gpas than unknown-genders.

Then, take a closer look at phase2:

Figure 1.3: Gender Distribution of Scores for different Phase 2 Assessmen
2020 Black Saber new grads recruitment

Prepared by ThreeOThree Consultant Group
2021−07−05
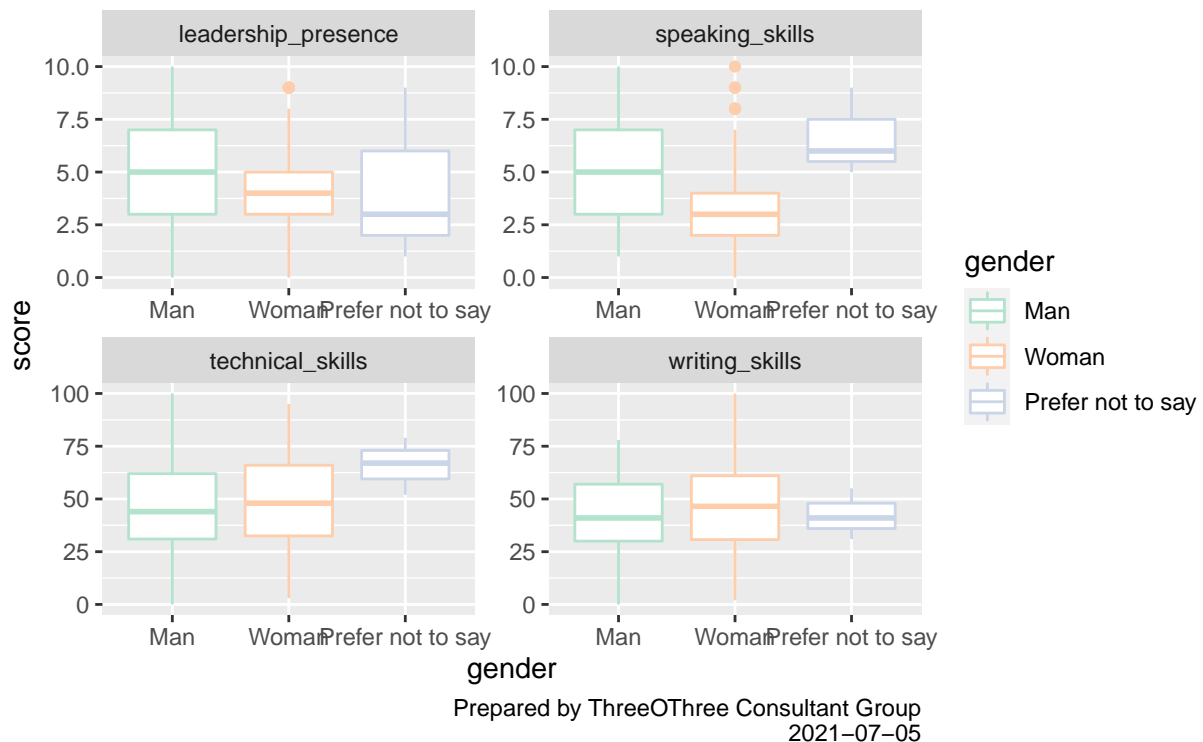
We can see that the AI auto-graded score for phase 2 is approximately no different between different genders. It is highly likely that there is no gender bias for the AI automated selection pipeline analyzing the pre-recorded video of applicants.

Similarly, we portray a interviewer-rating distribution for different genders in the third round:

Figure 1.4: Gender Distribution of Scores for different Phase 3 Assessment
2020 Black Saber new grads recruitment

Prepared by ThreeOThree Consultant Group
2021−07−05
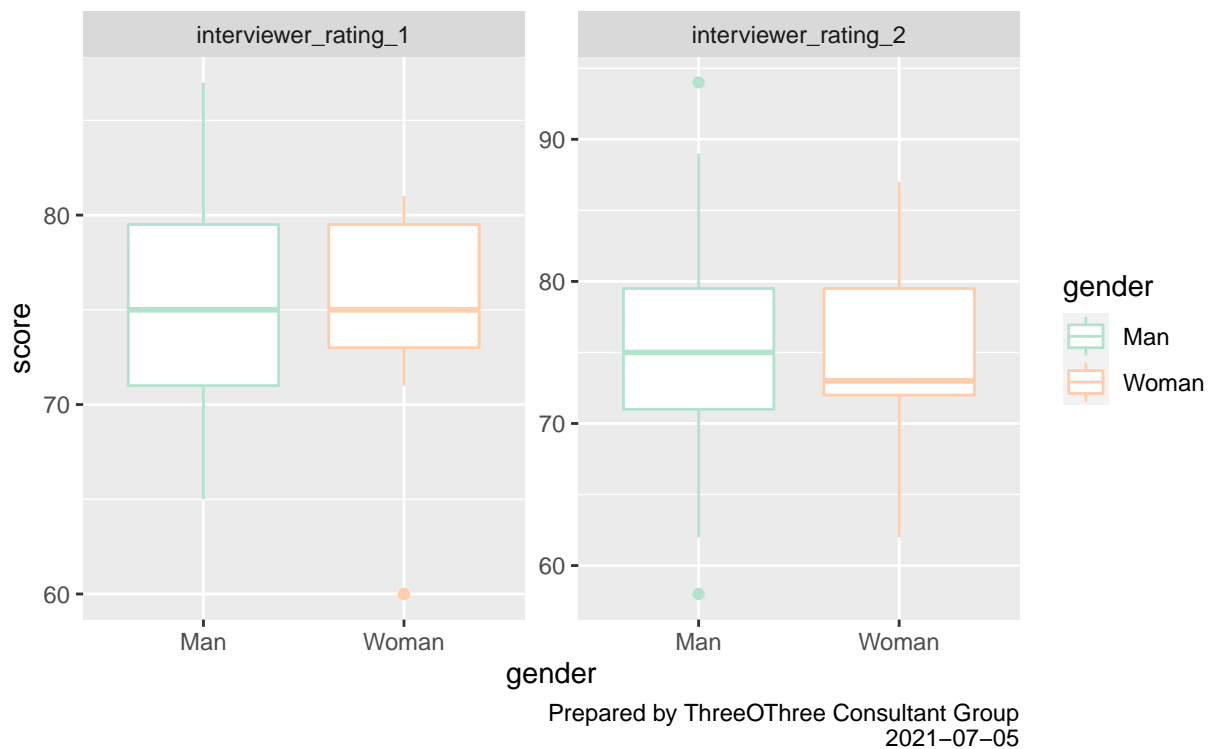
From the above figure, the two interviewer rating for phase 3 seems pretty similar for male and female. Congrats! It's quite possible that no gender bias during the interview phase.

Last but not the least, here is the pie chart illustrating the percentage for each gender who passing the final round:

Figure 1.5: Gender Distribution for Final Admission
2020 Black Saber new grads recruitment
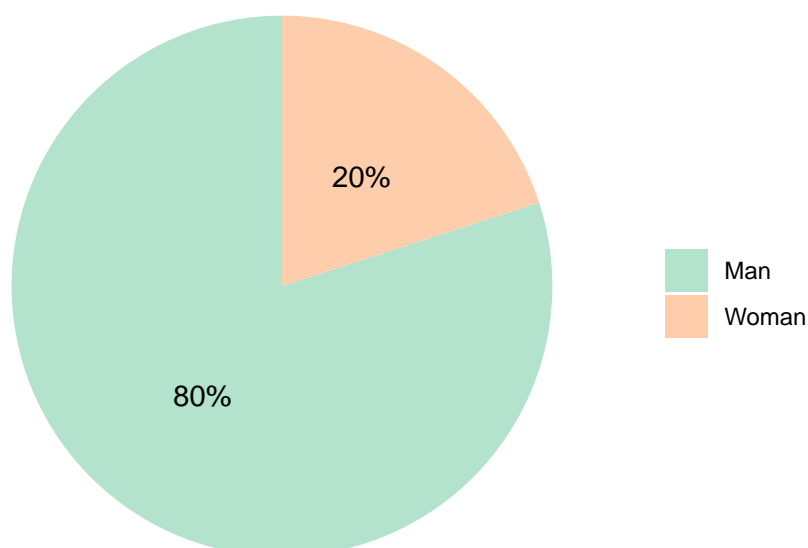
Prepared by ThreeOThree Consultant Group
2021−07−05

Here we can see, only 20% of candidates surviving the final round are females, while the rest are males. None of those who prefer not to say about their genders pass the final round.

To have a uniform standard, the significance level in this report is set to 5%. That is, if the p-value for an effect is less than 5%, then it is statistically significant. Otherwise, it is statistically insignificant.

From phase 1 to phase 2   Table1: Parameter estimates from Phase 1 to Phase 2

	estimate	P-value	significance
Baseline Odds	0.00	0.991332	insignificant
cover letter	3.19e+27	0.993310	insignificant
cv	3.00e+22	0.996783	insignificant
GPA	3.07e+05	0.000141	significant
extracurriculars	2.13e+04	5.47e-05	significant
work experience	1.08e+05	5.47e-05	significant

	estimate	P-value	significance
Gender Prefer Not to Say	1.87	0.961183	insignificant
Gender Woman	2.47	0.353057	insignificant

Table 1 illustrates the significance level of each predictor of the original model. It is obvious that, except GPA, extracurricular, and work experience, all other predictors, including the gender of female and unknown, have a way higher p-value than the significance level. That is, there is insufficient evidence to prove that these factors affect the hiring result. To simply test the effect of gender, the likelihood ratio test applied between models with and without the predictor of gender gives the result:

Table2: Likelihood Ratio Test from Phase 1 to Phase 2

Models compared	LRT	P-value	Preferred
with gender VS without gender	0.9144	0.6331	model without gender

The p-value for lack-of-fit far outweighs that of the significance level, that is, the simpler model (the model without gender predictor) is preferred. Therefore, gender is not considered to be a deterministic factor in Phase-1 hiring.

From phase 2 to phase 3   Table3: Parameter Estimate from Phase 2 to Phase 3

	estimate	P-value	significance
Baseline Odds	0.00	1.45e-08	significant
Gender Prefer not to say	0.00	0.994	insignificant
Geneder Woman	0.52	0.337	insignificant
technical skills	1.08	4.00e-05	significant
leadership presence	2.56	9.14e-06	significant
speaking skills	2.10	1.59e-05	significant
wrting skills	1.09	9.21e-05	significant

Table 3 displays the significance level of each covariate in the initial generalized linear mixed

model. From the table, baseline odds, technical skills, speaking skills, writing skills and leadership presence have highly significant p-values. While for both gender of female and unknown, a p-value of xx and xx are insufficient to prove their effectiveness on the result of the hiring. To take a closer look at gender, we apply a likelihood between models with and without gender, which gives the result:

Table4: Likelihood Ratio Test from Phase 2 to Phase 3

Models compared	LRT	P-value	Preferred
with gender VS without gender	1.8013	0.4063	model without gender

From the above table, the p-value of chi-square is way larger than 5%, that is, the model without gender as its explanatory variable is better fitted. Therefore, gender is not considered as a deterministic factor in Phase-2 hiring.

From phase 3 to final hire   Table5: Parameter Estimate from Phase 3 to Final Hire

	estimate	P-value	significance
Baseline Odds	0.00	1	insignificant
interviewer rating 1	5.65e+10	1	insignificant
interviewer rating 2	3.09e+13	1	insignificant
Gender Woman	0.00	1	insignificant

Table 5 lists p-values of each explanatory variable in the generalized model predicting the result of the final-round hiring. With the significance level set to 5%, all factors fail to reject the null hypothesis that they have no effect on the result. To determine whether gender should be fit in the model, we apply a likelihood between models with and without gender, which gives the result:

Table6: Likelihood Ration Test from Phase 3 to Final Hire

Models compared	LRT	P-value	Preferred
with gender VS without gender	0	1	model without gender

The relatively large p-value for lack-of-fit indicates the model without gender is preferred. That is, gender is not a factor determining the result of Phase-3 hiring.

Interpretation of Results

The hiring process is not biased by the gender of applicants. For the hiring process from phase 1 to phase 2, we found there is no significant influence for gender on the probability of passing the initial phase. After accounting for the within applied team variation, we found this first-round selection is based on the GPA, extracurricular, and work experiences of the applicants, which are those autorated by the AI system. For each unit increase of GPA, the probability of passing the initial phase increase by 300,000 times on average.

The trialing AI automated selection pipeline is quite impressive. Above all, this screening system is not gender-biased. Additionally, the auto-graded scores by the AI system are an informative representation of the applicants' talent. Higher scores indicate better performance and one unit increment on technical skills and writing skills will on average cause an approximately 1 times higher chance of proceeding to phase 3. Each score increase in leadership presence and speaking skills will on average increase the odds of passing phase 2 by approximately 2 times.

Although the hiring process from phase 3 to final hire is not determined by gender. This interview phase is still needed for improvement. The ratings by the two interviewers show no useful information on the hiring decision and it seems that the last phase is not recruited based on applicants' talent. The unknown grading criterion is one of the major limitations in our study as it may be subjective or biased. A closer examination of the grading standard is required for future application. Although there is still more males entering the final hire, it's mainly because Black Saber is a male oriented company with higher percentage of male employees.

Answer to the research question   According to our models, gender is not a deterministic factor for all three hiring phases. Moreover, the AI automated selection pipeline mainly focuses on each candidate's comprehensive ability, including essential skills and relevant experiences.

Promotion: average position distribution among current employees

Background

Promotion is an important way to qualify one's efforts in a company. However, promotion process might not always be fair.

Research questions

For Black Saber Software company, we want to explore:
a) Are males being preferred at higher position on average?
b) If the mean productivity of an employee is a key factor affect his/her position in this company on average?

Methods

Study Design, setting, participants

Our main idea was to explore the relationship between employees' features and their position position. Since each employee might have several rows of record in different financial quarters, we might expect that there would be a correlation among outcomes for directly using the original data. Therefore, to address the correlation, we would take average over each employee to get their unique data. Then we did our analysis on the processed data.

Data description and data wrangling

The original data involved for promotion analysis were employment records from April 2013 to December 2020. A total of 6906 records from 607 distinct employees who were working in Black Saber Software at the time of recording were included, those who were not working in this company at the time of recording were excluded. The source of data is black-saber-current-employees.csv, provided by Black Saber Software, with 8 variables:

- `employee_id`: the 5-digit unique number identifier for each employee
- `gender`: the gender of employee (Man, Woman, Prefer not to say)
- `team`: the team an employee works for
- `financial_q`: financial quarter of record (eg. 2020 Q3)
- `role_seniority`: the role position of an employee
- `leadership_for_level`: quality of demonstrated leadership (eg. Appropriate for level)
- `productivity`: work output of an employee in numbers
- `salary`: salary at the given financial quarter in dollars

We would convert variables `role_seniority` and `financial_q` from texts into numbers:

To quantify employee positions, we would transfer each level of role position into whole numbers from 1 to 9, where larger number represents higher position in this company (showed Table 7)

Table 7: Position to quantifier numbers

Position	Position quantifier
Entry-level	1
Junior I	2
Junior II	3
Senior I	4
Senior II	5
Senior III	6
Manager	7
Director	8
Vice president	9

To quantify how long an employee works for this company, we would count their the number of quarters they showed up in original data file and store the new data in `time`

Then we manipulated an employee's average position over time as `position_avg`

an employee's maximum salary as `salary_max`

an employee's minimum salary as `salary_min`

an employee's maximum position as `position_max` in numbers

an employee's maximum position as `max_level` in texts

an employee's average productivity over time as `productivity_avg`

Statistical method

Then, with this new dataset, we built a scatterplot for average productivity distributed by average position:

one pie chart for gender distribution of current employees, and two bar charts for maximum position distributed by gender and gender distribution in teams, for Part a; Also, we were able

to apply a linear analysis since `position_avg` was a continuous variable, in order to imply the potential bias of promotion of the company, for Part b, as

$$postition_{avg} = \beta_0 + \beta_1 gender + \beta_2 team + \beta_3 time + \beta_4 productivity_{avg} + \beta_5 salary_{min}$$

Results

Part a: Are males being preferred at higher positions on average?

First, we examined the average productivity distributed by average position.

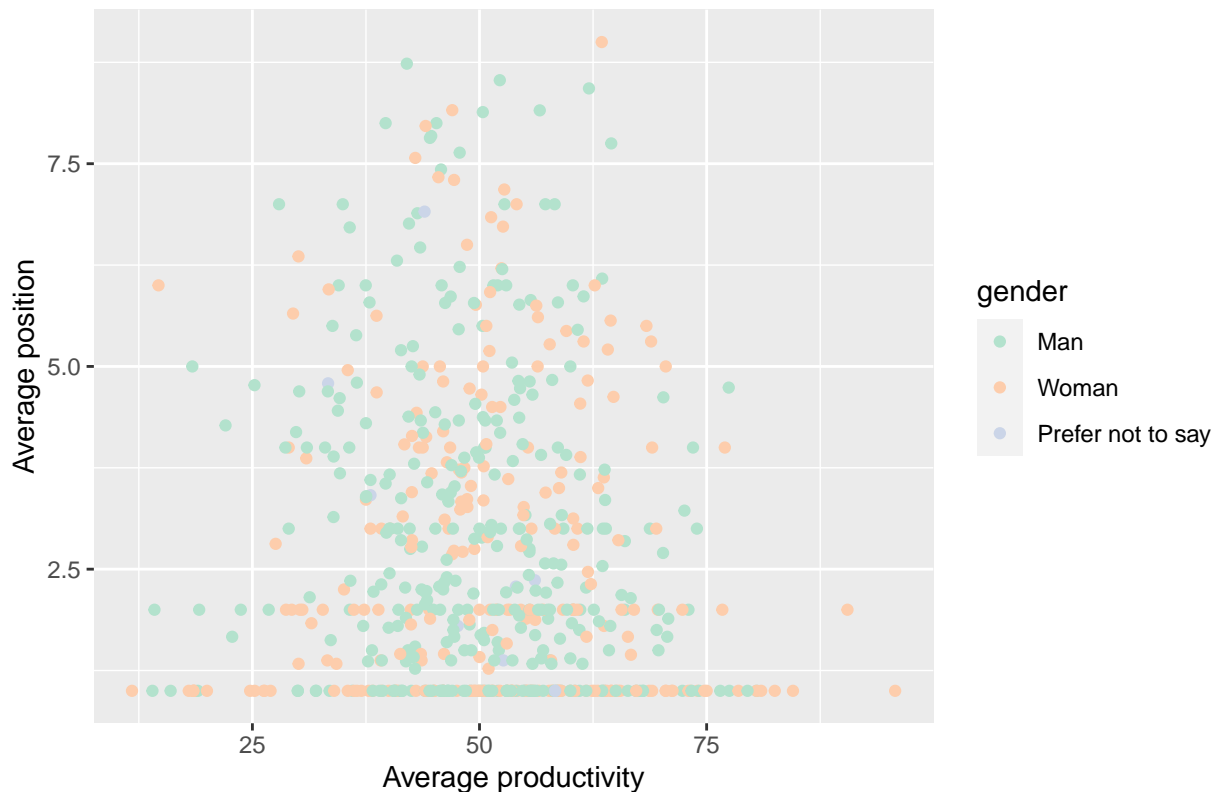Figure2.1: Average Productivity distributed by Average Position

Figure 2.1 is a scatterplot which describes the relationship between average productivity (x-axis) and average position (y-axis), where gender is varied by color. Men are denoted in light green points, women are denoted in apricot points and those who prefer not to say their gender are denoted in green point. There are more red points than the other two colors for average position higher than 7.5, which based on Table 7 would be a position beyond manager or director. That is, man is preferred at higher positions like manager or director or anything above that. There is no linear relationship found between an employee's average productivity and his/her average position in this company, which seems weird and unfair. However, we still need to calculate statistical results to confirm it, which would be further discussed in Part b.

Figure2.2: Gender Distribution Of Current Employees
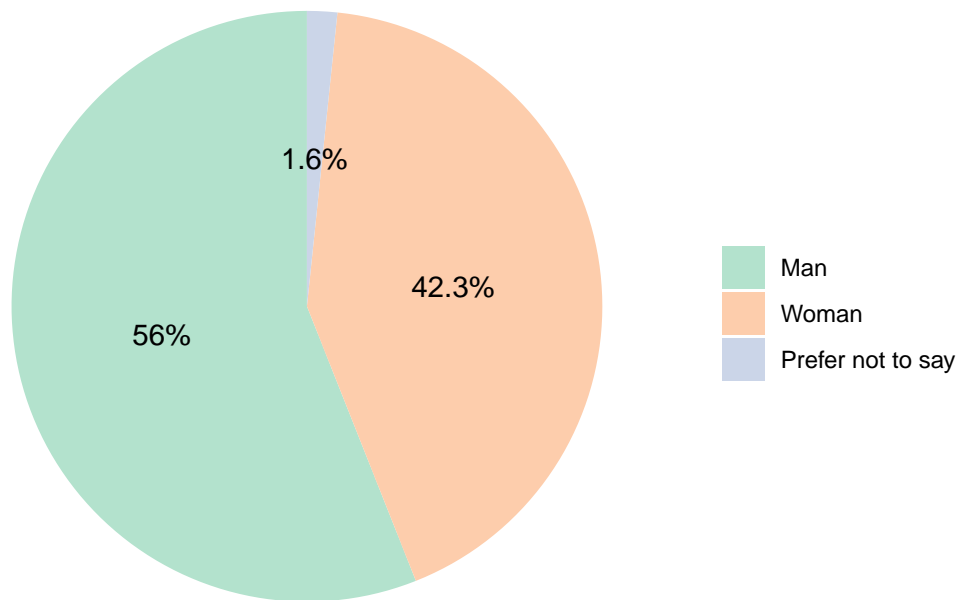2020 Black Saber Company

Figure 2.2 is a pie chart that Visualized the gender distribution in the company. We obtain that there is more percentage of man than woman, and this may due to the hiring system/mechanism of the company. However, there is not a significant difference between man and woman; and we could continue the following analysis based on the data.

Figure2.3: Maximum Position distributed by Gender

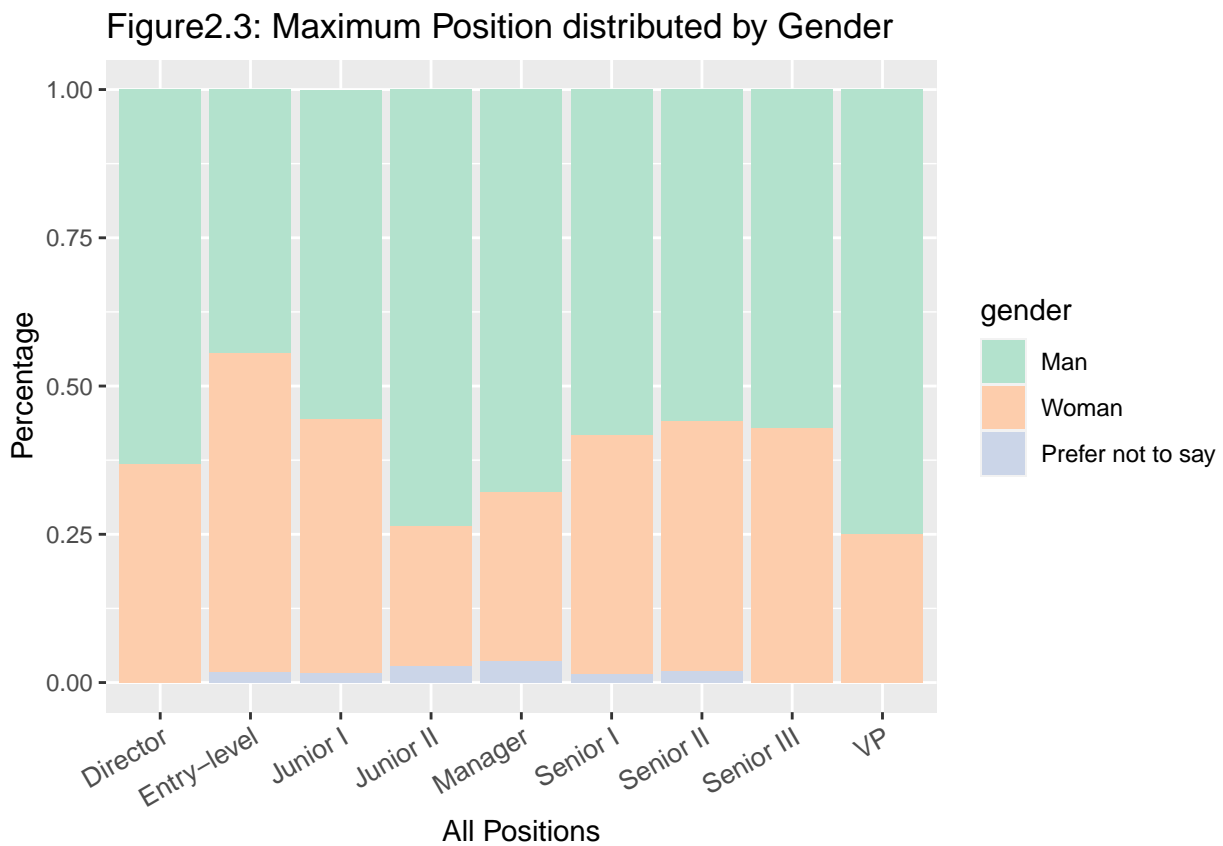Figure 2.3 is a bar chart that describes the distribution of employee's max position by gender. Except for the Entry-level (because no promotion is applied for), it is clear that the percentage of men occupies a larger percentage than women in all positions. This reflects a fact that more males in this company have an opportunity to be promoted (especially from Junior I to Junior II and from Senior III to Vice President).

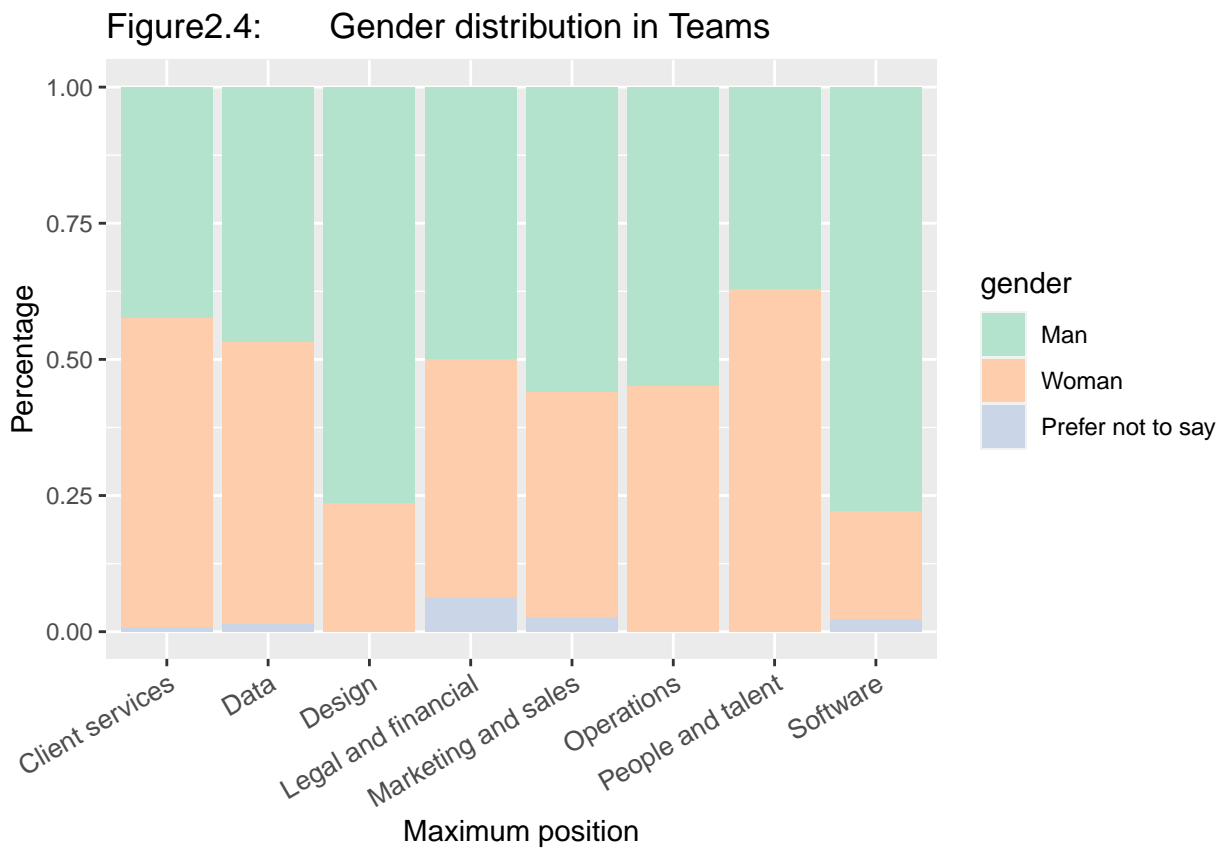Figure2.4:   Gender distribution in Teams

Figure 2.4 is a bar chart that describes the distribution of gender in each team. The software and Design team has a clear difference in gender distribution (man>woman). And this information will be linked in the following part b).

Part b: Does the mean productivity of an employee affect his/her position in this company on average?

Table 8: Summary table for average position

	Effect on average position	P-value	Significance
(Intercept)	-1.9410	<0.001	Significant
genderWoman	0.1066	0.86	Not significant
genderPrefer not to say	-0.0433	0.11	Not significant
teamData	-0.0656	0.56	Not significant
teamDesign	0.5822	0.0038	Significant
teamLegal and financial	-0.0319	0.84	Not significant
teamMarketing and sales	-0.0443	0.66	Not significant
teamOperations	-0.0696	0.53	Not significant
teamPeople and talent	0.1920	0.24	Not significant
teamSoftware	-0.1951	0.047	Significant
time	0.1292	<0.001	Significant
productivity_avg	-0.0008	0.74	Not significant
salary_min	0.0001	<0.001	Significant

The significance level in this consultation would be 5%. That is, if the p-value for an effect is less than 0.05, then this effect is statistically significant. Otherwise, if the p-value for an effect is greater than 0.05, then this effect is not statistically significant.

From Table 8, only teamDesign, teamSoftware, total employment time and minimum salary are significant.This means that they would have effects to an employee's average position in this company.The effect of teamDesign is 0.5822, which means if an employee is in design team, then his/her position would increase by 0.5822 level on average, holding other variables constant.The effect of teamSoftware is -0.1951, which means if an employee is in software team, then his/her position would decrease by 0.1951 level on average, holding other variables constant. Combine with Figure2.3 constructed above, we could get a result that the teams with more males will have more opportunities to get promotions. And it indirectly shows the gender biases dealing with the promotion.

The effect of (employment) time is 0.1292, which means if an employee works one more quarter, then his/her position would increase by 0.1292 level on average, holding other variables constant.The effect of minimum salary is 0.0001, which means if an employee's initial salary increases one more dollar, then his/her position would increase by 0.1951 level on average, holding other variables constant. Here we would define the initial salary to be an informative variable which covers an employee's initial position, ability and HR's first impression to him/her. Therefore, if an employee has a good foundation in terms of initial salary, then he/she might reach a higher position. Among these four factors, teamDesign brings the largest and positive effect to an employee's average position while teamSoftware brings a negative effect. This tells that choosing the right team might result in a better promotion.

Although minimum salary brings the smallest effect as per unit increase, that, it seems would not contribute too much to future promotions, the effect would be bigger in reality when we multiply it by salary differences. For example, for two new employees having same conditions, if their initial salary difference is 1000 dollars, then their expected position would differ by $1000 * 0.0001 = 0.1$ levels. We also noticed that average productivity is not a significant factor, which means that an employee who produces more outputs might not necessarily get a promotion, which confirms with the scatterplot analysis in Part a.

Answer to the research question   The company might be biased in promotion process. Specifically, man is preferred at higher positions and the mean productivity is not related to one's position on average.

Salary: average salary distribution among current employees

Background

This section mainly focuses on current employee's average salary per financial quarter in the Black Saber Software. The data are representing the talent and the value to the company, including the leadership, and the productivity, the data related to the salary, including the salary and the role_seniority, the data representing the financial quarter, the employee's team, and the data representing the employee's gender. The purpose of this section is through the analysis of the relationship between gender and the salary, and the relationship of all other factors with the salary to tackle the question that whether current employees' salary depends on their gender rather than their value to the company exists in the Black Saber Software Company.

Research question

Is the company's salary distribution mainly depending on one's value produced or is it partially affected by the employee's gender difference?

Methods

Study Design, setting, participants

We are going to analyze whether gender bias exists in the current employee's salary, the key element contains the gender, the value to the company(the average on productivity, and the leadership for level), and the average salary for the employees are considered. The data contains 6906 observations of 9 variables, including all 607 current employees of the Black Saber Software. After the data wrangling and organization, the variables have become 10 variables. For the data wrangling, the new variables average salary which represents current employee's average salary per financial quarter is added to the dataset for this part of the analysis.

Data description and data wrangling

- `avg_salary` = average salary per financial quarter in dollars
- other variables are the same as the promotion part

Statistical method

The average salary per financial quarter is considered as the response variable. To analyze whether an employee's salary is affected by gender rather than the value to the company, the

variable of the average on productivity, and the gender is the important one to focus on. Also, there are few other variables such as team, financial quarter, and role seniority of level are considered as potential bias factor to the average salary. Then both the fixed effects and the random effects of variables on this data are considered, so the linear mixed model is being used which is as the followings:

$$y = X\beta + Ru + \epsilon$$

the response variable y in this model is the vector with $E(y) = X\beta$. β is an unknown vector for the fixed effects and u is an unknown vector for the random effects. This model is appropriate when non-independence exists in the data according to the first visualization of the data(correlation matrix has shown that).

Results

Figure 3.1 Pairwise Correlation
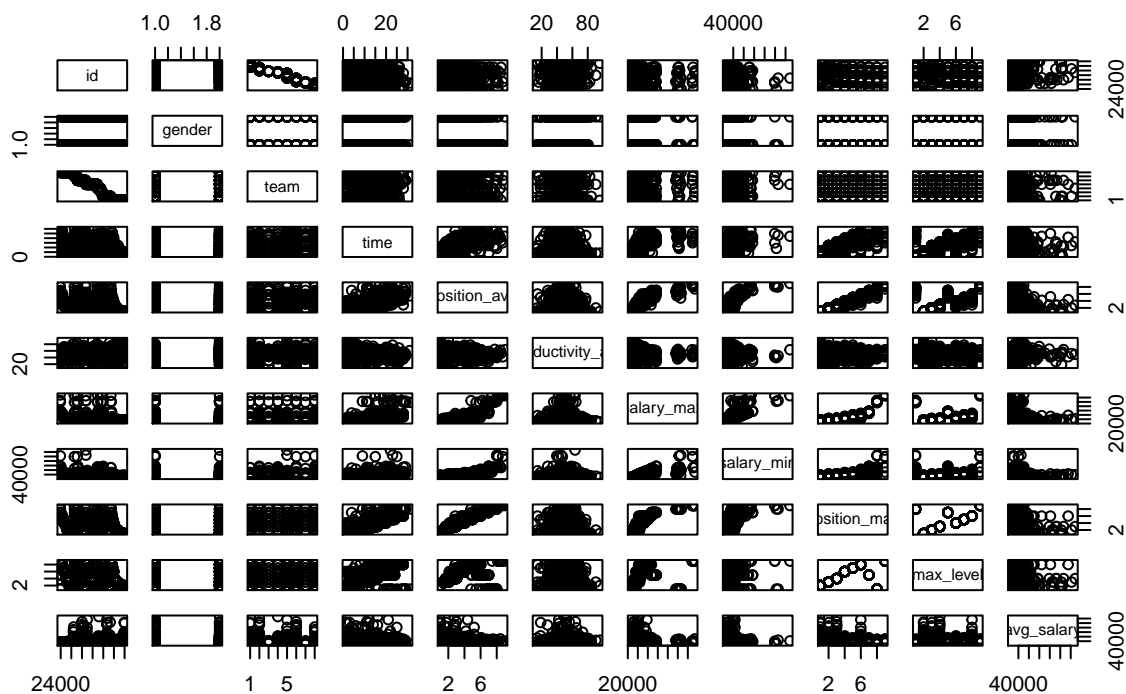
According to the pairwise correlation in our data set, different variables are correlated in a variety of ways with other variables. However, the correlation between gender with others and average of the productivity with others is small. That is to say, these two variables are good to use. Since we would like to get the best model for the final and then interpret the result from the final model. We then need an LRT to help us with that after a further visualization of the dataset.

Let's do further visualization for the data. First, the boxplot on average of the salary grouped team categorized by gender.

Figure3.2: boxplot on average salary grouped by gender and team

The box plots show the average of the salary for men is higher than females in terms of the team except for team people and talent. It seems that gender bias is shown here, but this plot only considers team and gender with the average of salary and not other variables. For example, the average of position, the average of the productivity may affect the gender data and the average of salary data. Thus, we can not draw any conclusion from that. But keep that in mind, later analysis will show that considering all these factors, it may still happen. Then we move on to see the scatter plot for the `productivity_avg` and `avg_salary` categorized by gender:

Figure3.3: Average Productivity distributed by Average Salary

As we can see in the above scatter plot, man are denoted in light green dots, women are denoted in apricot dots, the people who prefer not to say their gender is ignored in this data set. There are more light green dots than the other two colors for average of salary above 100,000, and there are more apricot dots than light green dots for average productivity. That is, man is preferred with higher average of salary, and woman is preferred with higher average productivity. In summary, it seems to be gender-biased in the allocation of salary which is not based on the contribution or productivity to the company but the gender. However, we still need to further analysis to confirm this, same issues with the box plot, that is still missing some factors which may affect the average productivity and average of salary. Then we do the model fitting to get a deeper understanding of the data.

Model Fitting

Table 9: Model1

Factor	Value	pvalue	Significance
GenderWoman	-2472.722	0.0011	yes
position_avg	7747.378	0.0000	yes
productivity_avg	38.452	0.1748	no
teamData	1063.107	0.4026	no
teamDesign	-847.030	0.7108	no
teamLegal	3247.014	0.0679	no
teamMarket	-1344.549	0.2437	no
teamOperations	1025.375	0.4137	no
teamHR	-3852.358	0.0393	yes
teamSoftware	1172.602	0.2932	no

In model 1, the model is the following,

$$salary_{avg} = \beta_0 + \beta_1 gender + \beta_2 position_{avg} + \beta_3 productivity_{avg} + \beta_4 team + \mu id$$

We consider gender, average of the position, average of the productivity, and the team as the fixed effects. Since they are the potential factors affect the averge of the salary, the response variable. The random effects need to be picked by the further model selection. Since there are two, the id of the employee, and the employee time. For this model, the id of the employee is selected as the random effect is selected. The significance level in this consultation would be 5%. That is, if the p-value for an effect is less than 0.05, then this effect is statistically significant. Otherwise, if the p-value for an effect is greater than 0.05, then this effect is not statistically significant. Only the Woman, the average of the position, the teamHR are significant, then we first do the interpretation on this three variables. The effect of the Woman is -2472.722, which means that if the new employee is woman, then the average of the salary may decreased by 2472.722. The effect of the average of the position is 7747.378, which means that if the new employee is increased by 1, then the average of the salary may also increased by 7747.378.The effect of the teamHR is -3852.358, which means that if the new employee is in teamHR, then the

average of the salary may decreased by 3852.358. Among these three factors, gender brings a strong effect to the average of the salary. This tells that if the employee wants a better salary, then they should not pick the teamHR. Also, the gender bias may exist in this model, but we still need to check the assumption of the model and compare it to the remaining model.

Table 10: Model2

Factor	Value	pvalue	Significance
genderWoman	-2876.161	0.0001	yes
position_avg	8932.582	0.0000	yes
productivity_avg	27.710	0.3037	no
teamData	1518.732	0.2112	no
teamDesign	-1150.015	0.5965	no
teamLegal	3350.885	0.0492	yes
teamMarket	-1064.442	0.3391	no
teamOperations	-731.267	0.5443	no
teamHR	-3436.686	0.0553	no
teamSoftware	1377.299	0.1962	no

In model 2, the model is the following,

$$salary_{avg} = \beta_0 + \beta_1 gender + \beta_2 position_{avg} + \beta_3 productivity_{avg} + \beta_4 team + \mu time$$

We consider gender, average of the position, average of the productivity, and the team as the fixed effects. Since they are the potential factors that affect the average of salary, the response variable. The random-effects need to be picked by the further model selection. Since there are two, the id of the employee, and the n. For this model, the employment time is selected as the random effect. The significance level in this consultation would be 5%. That is, if the p-value for an effect is less than 0.05, then this effect is statistically significant. Otherwise, if the p-value for an effect is greater than 0.05, then this effect is not statistically significant. Only the gender woman, the average of the position, the team legal are significant, then we first do the interpretation on this three. The effect of the gender woman is -2876.161, which means that if the new employee is a woman, then the average of the salary may decrease by 2876.161. The effect of the average of the salary is 8932.582, which means that if the new employee is increased

by 1, then the average of the salary may also increase by 8932.582. The effect of the team legal is 3350.885, which means that if the new employee is in teamLegal, then the average of the salary may increase by 3350.885. Among these three factors, gender brings a strong effect to the average of the salary. This tells that if the employee wants a better salary, then they should pick the team of HR. Also, gender bias may exist in this model, but we still need to check the assumption of the model and compare it to the remaining model.

Table 11: Model1v2

model	AIC	BIC	LRT
model1	12539.79	12596.89	sig
model2	12507.92	12565.01	sig

Then we conduct a likelihood ratio test between model 1 and model 2. The result shown above tells us that the LRT p-value is less than 0.05. We will reject the null hypothesis at the .05 significance level. This tells us that the model2 is significantly more accurate under an α equal to .05. Thus, we should use model2 as it increases the accuracy of our model by a substantial amount. For the AIC and the BIC, they are smaller for the model2, then the model2 is picked for the next comparison.

Table 12: Model3

Factor	Value	pvalue	Significance
genderWoman	-2618.523	0.0004	yes
position_avg	7710.082	0.0000	yes
avg_productivity	33.434	0.2380	no

In model 3, the model is the following,

$$salary_{avg} = \beta_0 + \beta_1 gender + \beta_2 position_{avg} + \beta_3 productivity_{avg} + \mu team$$

We consider gender, average of the position, and average of the productivity as the fixed effects. Since they are the potential factors that affect the average of the salary, the response variable. The random-effects need to be picked by the further model selection. For this model, the random

effect team is selected. The significance level in this consultation would be 5%. That is, if the p-value for an effect is less than 0.05, then this effect is statistically significant. Otherwise, if the p-value for an effect is greater than 0.05, then this effect is not statistically significant. Only the gender woman, the average of the position are significant, then we first do the interpretation on these two variables. The effect of the gender woman is -2876.161, which means that if the new employee is a woman, then the average of the salary may decrease by 2876.161. The effect of the average of the position is 8932.582, which means that if the new employee is increased by 1, then the average of the salary may also increase by 8932.582. Among these two factors, gender brings a strong effect to the average of the salary. Also, gender bias may exist in this model.

Table 13: Model2v3

model	AIC	BIC	LRT
model2	12507.92	12565.01	sig
model3	12540.97	12567.33	sig

Then we conduct a likelihood ratio test between model 2 and model 3. The result shown above tells us that the LRT p-value is less than 0.05. We will reject the null hypothesis at the .05 significance level. This tells us that the complex model model2 is significantly more accurate under an α equal to .05. Thus, we should use model2 as it increases the accuracy of our model by a substantial amount. For the AIC and the BIC, they are smaller for the model2, then the model2 is the best model among the three. Then we need to check the assumption for this model.

Checking the Assumption
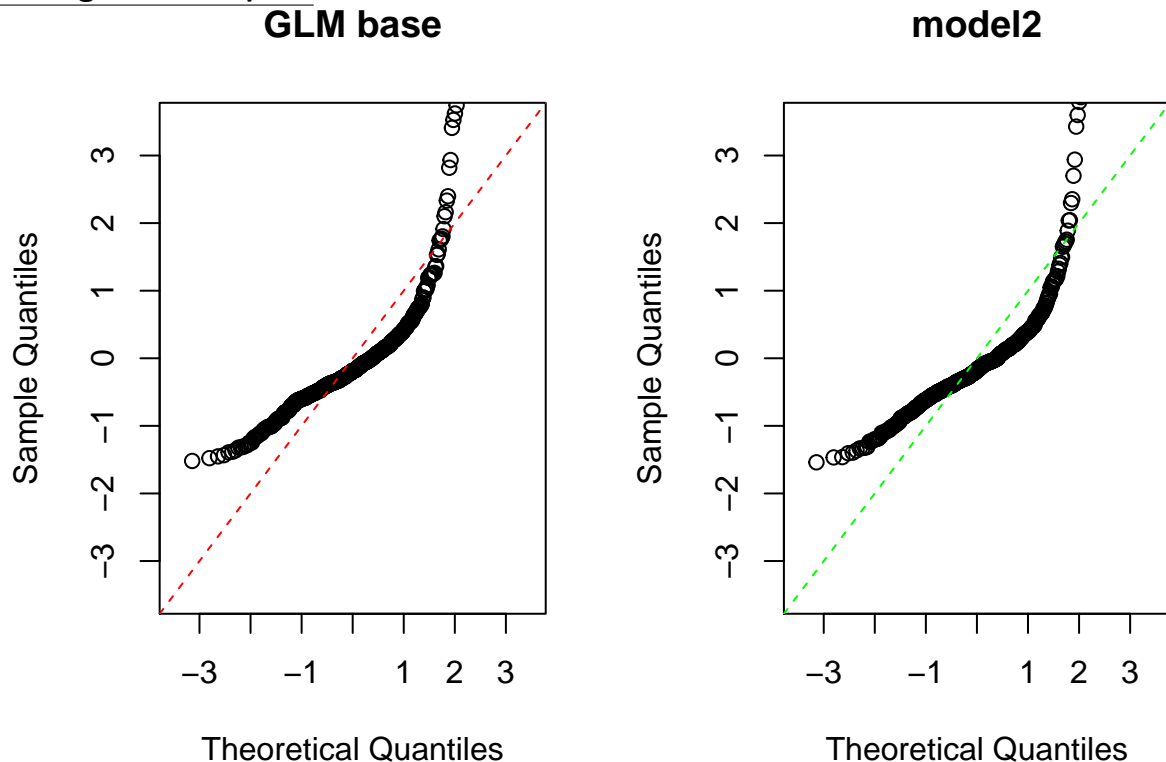
According to the plots above. The left-hand side is the Q-Q plot for the Glm base model that we set for comparison. The right-hand side is the Q-Q plot for the best model we pick, the model2, it gives the information that the assumption of the normal distribution is satisfied, ignoring a few outliers. Hence, we are proposing that model2 is the final model to give the result whether current employees' salary depends on their gender rather than their value to the company exists in the Black Saber Software Company.

Answer to the research question   The company's salary distribution is partially affected by the employee's gender difference instead of purely depending on one's value produced.

Discussion

In the hiring process of all three phases is not biased by the gender of applicants. For the hiring process from phase 1 to phase 2, after accounting for the within applied team variation, we found this first-round selection is based on the GPA, extracurricular, and work experiences of the applicants. The trialing AI automated selection pipeline is quite impressive. The auto-graded scores by the AI system are an informative representation of the applicants' talent. Although the hiring process from phase 3 to final hire is not determined by gender. This interview phase is still needed for improvement for future application. In promotion analysis, from one pie chart for gender distribution of current employees and two bar charts for maximum position distributed by gender and gender distribution in teams for Part a, we found males were more preferred. Also, from the linear regression for Part b, we found the largest potential bias would be team of the company. In the salary analysis, Current employees' salary are highly related to their average seniority (position), and salary is also partially affected by the employee's gender. The effect of the gender woman is negative, which means that if the new employee is a woman, then the average of the salary may decrease a lot. For the research question about the salary for the current employee, we have strong evidence that gender bias may exist in this current employee's salary.

Limitations

- Hiring limitation

  - Although all of the hiring processes are not determined by gender, the last interview phase is still needed for improvement. It seems that the interview phase is not recruited based on applicants' talent. A closer examination of the unknown grading criterion is required for future application.

  - Another limitation in the analysis of the hiring is that more males are being accepted in the hiring process, which might result in more males in higher positions. Also, the given data is not informative enough.

- Promotion Limitation

  - By constructing a frame of position in numbers, we could obtain a relatively informative relationship between promotion to their positions on average. However, we were more focused on the overall promotion status rather than the promotion between each level. Another limitation in the analysis of the promotion process is that employees at higher levels would have less technical work but rather more management work (may affect their productivity). For example, a manager at the software team might devote more

to making plans and assigning work to team members. In this case, a manager might have fewer outputs.

– Also, the measurement of productivity is not clear enough. From the company's data dictionary, productivity is rated according to the job description. However, we have no information about employees' specific work and how many efforts they need to get satisfied productivity. And it would be hard to compare productivity across teams, so we still require more data to further investigate this. Similarly, the `leadership_for_level` from the original data is not meaningful for our consultation. Although it may describe an employee's leadership condition in terms of "Needs improvement", "Appropriate for level" and "Exceeds expectations". However, "Appropriate for level" requires much fewer entry-level employees than for a manager, so it might not be helpful for us to analyze the overall position of any employees.

- Salary limitation

  – The final limitation is in the analysis of the current employee's salaries. The limitation is that the current proportion of male employees is higher than that of females, when we do the analysis, there may be a natural bias. Also, some outliers are detected at the tails of the Q-Q plot limiting the model checking. So, the linear mixed models may have been a better fit for this data. Also, when we use the likelihood ratio test, AIC and BIC to pick the model, there is a downside in that they do not take the uncertainty of the models into account, and we can end up choosing models that are too plain. In future analysis, the variables based on gender ratio, instead of gender can be considered to remove the natural bias in the analysis.

  – The strength is that some of the confounding variables are set to be the random effects in the linear mixed model in salary analysis. That being said the relationship we have for the gender and average of the salary is accurate after controlling those uncertainties. Also, in this section, the people with "Prefer not to say" gender were ignored, and the amount is small enough that would not affect the result.

Consultant information

Consultant profiles (Order alphabetically by first name A-Z)

Huyi Xiong: Huyi is a junior consultant with ThreeOThree Consultant Group. She specializes in business analysis and statistical communication. Huyi earned her Bachelor of Science, double majoring in Actuarial Science and Statistics and minoring in Mathematics from the University of Toronto in 2022.

Peiqing Yu: Peiqing is an associate consultant with ThreeOThree Consultant Group. She specialized in health data science and biostatistics. Peiqing earned her degree in Bachelor of Science, double majoring in statistics and genetics and minoring in computer science from the University of Toronto in 2021.

Qiaoyiwen Wu: Qiaoyiwen is a junior consultant with ThreeOThree Consultant Group. She specializes in data analysis. Qiaoyiwen earned her degree in Bachelor of Science, Specialist in Computer Science and Major in Statistics from the University of Toronto in 2023.

Yan Zhang: Yan is a senior consultant with ThreeOThree Consultant Group. He specializes in data wrangling and data visualization. Yan earned his Bachelor of Science, Specialist in Statistics Methods and Practice, from the University of Toronto in 2021.

Zhengtong Tan: Zhengtong is a junior consultant with ThreeOThree Consultant Group. He specializes in statistical analysis. Zhengtong earned his degree in Bachelor of Science, specializing in mathematical application in probability and major in statistics and the minor in computer science from the University of Toronto in 2025.

Code of ethical conduct

- Place clients' interests above all else; And, ensure to serve our client's with sincerity and professionalism.
- Never make false propaganda on our service contents.
- Accept tasks only if our ability is capable of; Only assign tasks to the employee with required credentials.
- Ensure all team members agree and sign the Non-disclosure and Confidential Agreement before receiving information provided by the client.
- Strictly follow the confidential agreement and take necessary actions to protect the client's benefits.