Analysis of fairness in hiring and remuneration processes

A special focus on gender parity

Report prepared for Black Saber Software by SFP

2021-04-21

Contents

Executive summary

Background & Aim

Concerns about potential bias in the hiring and remuneration processes were consistently raised internally in Black Saber Software. In order to investigate the true fact for the issues, we SFP were consent to help figure it out. This report was commissioned to examine if potential issues especially gender parity existed in the hiring, promotion and salary processes based on the hiring data and data about promotion and wages for current employees provided by Black Saber Software . Besides, we were interested in the fact how the new trialled AI service performanced in the hiring process.

Key results

- Only 2.7% of men and 0.6% were finally hired while no people with gender prefer not to say hired in the hiring process.
- In the first and second hiring phases, the percentage of applicants who can proceed to next phase was approximately evenly split between gender.
- No great gender difference was found in the AI-autograded scores in the phase 2.
- The hiring process showed fairness in each phase based on candidates' talent and value to BBS.
- In most teams, current employees with higher productivity had lower productivity especially for those who prefer not to say their gender which indicates potential bias existed in the salary process.
- Regarding leadership for level, only men were labeled beyond expectation and only women were labeled need improvement so it seemed unfair in the promotion process.

Limitations

We were only offered with data for current employees so it was hard to know reasons past employees left. If additional data related to past employees can be provided, we can have a more deep and comprehensive sight into potential biases;
The lack of ethnicity/race data also results in a limitation since race might be a potential existance which might influence the results of hiring pipelines and promotion and amount of wages;
No information about the determination methods for hiring results, promotion and amount of salary was provided which limited the accuracy of models we fit for these processes.

Key results of the study are summarized in the following tables:

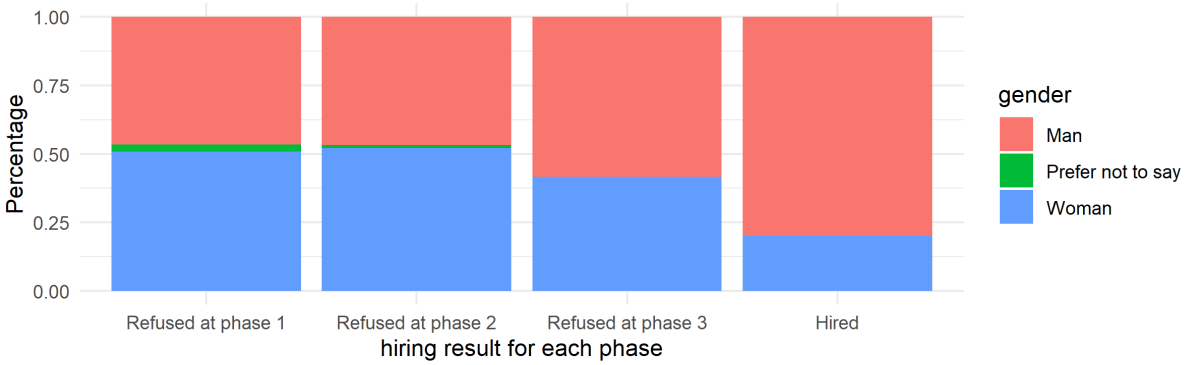

Figure 1: Proportion of Applicants who were refused in each hiring pipeline or finally hired by gender

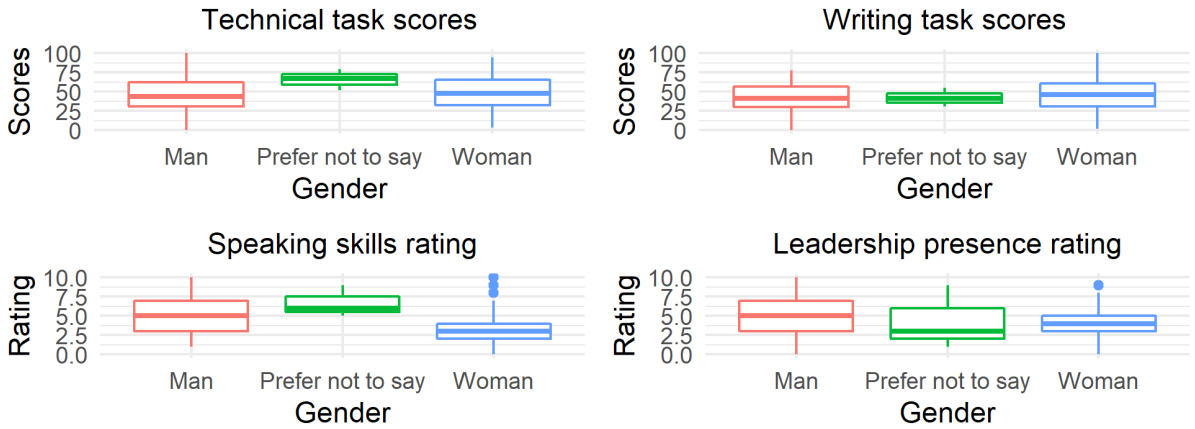

Figure 2: Distribution of AI-autograded Scores/Ratings by Gender in the Phase 2

The horizontal line in each box represents the median of scores which separates the top 50% scores from lowest 50%. The horizontal lines at the top and bottom for each box represent the lowest 75% and 25% o scores respectively. The ends of vertical lines give the full range of scores, with some dots representing extreme values away from the box.

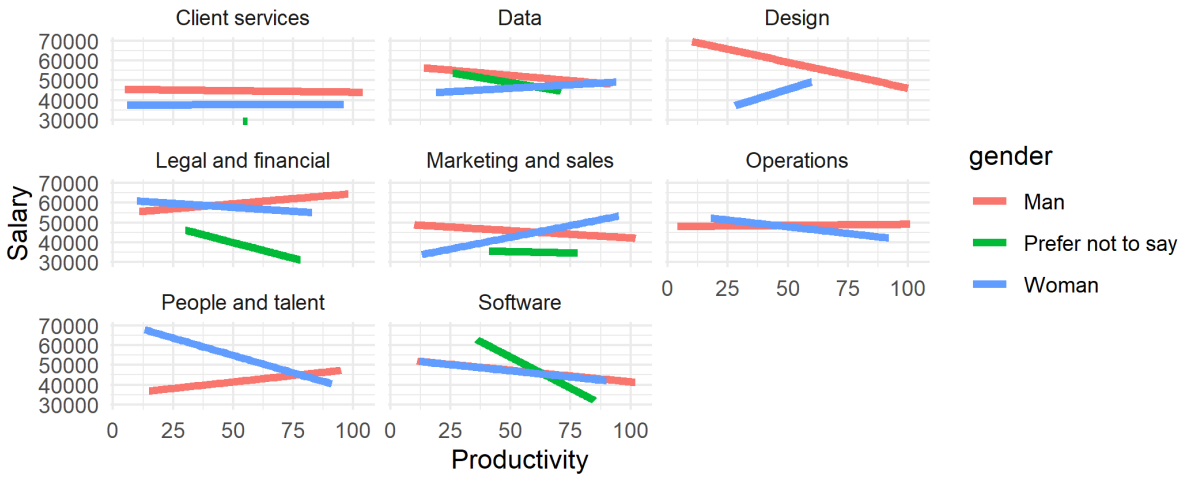

Figure 3: Relationship between productivity and salary for all eight teams by gender

Technical report

Introduction

Recently, several complaints were received related to potential bias in the hiring and remuneration processes. Hence, BSS (Black Saber Software) required our team to take a look and give out a related report for the Board of Directors. The purpose of this study is to investigate if BSS's hiring, promotion, and salary were processing fairly based on talent and value to the company. We were especially interested in exploring potential gender bias in the hiring and remuneration processes. Besides, we also investigated the performance of the new AI service in the hiring pipeline. Data was provided by the data team from BSS including hiring data for their new graduate program and data for their current employees.

In this report, we will specify different explanatory and response variables for hiring and remuneration processes.. Then, we will do some exploratory data analysis and then fit several generalized linear models and a linear mixed effects model to analyze the results. At the end, we will point out the same limitations of our model. The whole report will be run on R.studio and knit into pdf.

Research questions

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

- Does the new AI service in the hiring pipeline work and is reliable for future use?
- If the applicants are hired in the hiring process based on their ability and value to BSS?
- Are there potential biases within variable groups potentially affecting the salary and promotion of employees?

Possible future use of AI service in the hiring process

Data Organization

We considered a data set named all_phase combining all necessary data in the hiring process which include four datasets : phase1-new-grad-applicants-2020.csv, phase2-new-grad-applicants-2020.csv, phase3-new-grad-applicants-2020.csv and final-hires-newgrad_2020.csv. Each line of all_phase contains all aggregated information of an applicant in the hiring pipelines. In this analysis for the hiring process, the response variable is the result. Examination of data for the hiring process reveals the following key variables in all_phase:

Variable	Description
applicant_id	Unique identifier assigned to candidates in Phase 1
team_applied_for	Data or Software
gpa	0.0 to 4.0
extracurriculars	Indicates the quality of extracurricular involvement with scores of 0,1,2
work_experience	Description of candidates' previous work experience with scores of 0,1,2
technical_skills	Score from 0 to 100 on a timed technical task, AI autograded
writing_skills	Score from 0 to 100 on a timed writing task, AI autograded
speaking_skills	A rating of speaking skills based on pre-recorded video, AI autograded
leadership_presence	A rating of leadership presence based on pre-recorded video, AI autograded
final_result	Indicate at which phase applicant was rejected or hired finally
result	Show the hiring result for each applicant in each hiring pipeline

Data Wrangling and Visualization

We first joined all phases' data together in the dataset called all_phase, and then created a new variable called final_result telling us whether an applicant was rejected at each phase or finally hired. And we also created another new variable called result for each phase which can

indicate whether the applicant passes or not. Since our main purpose was to see whether AI service worked in the hiring process, we made some plots for scores of AI-autograded tasks by gender in phase2.

Figure 1: Distribution of AI-autograded Scores/Ratings by Gender in the Phase 2

Figure 1 presented a five-number summary: the minimum, the maximum, the sample median, and the first and third quartiles based on different genders for the scores or ratings by AI service. We can see generally there is no difference of scores or ratings between man and women since in all the four plots, man and women gave the similar five-number summary results. But in the technical task scores plot, the boxplot for people with gender prefer not to say was not overlapped with other two boxplots for man and woman so there might be a significant difference between gender in scores rated by AI and we need to do further analysis.

Models and Results

Since our response variable has only two levels : hired or no which is equivalent to the binary variable, the generalized linear model came out immediately. As we were interested in studying whether AI service was reliable in the hiring process without potential bias, we created two different models to determine whether including the grades made by AI service. In the first model, we checked if variables in the phase 1 are related to the final result. The second model, we contained all covariates in the phase 2 to check whether AI service had effect on the final result.

From the p-value from the result of the first model, we found that gpa and gender are significant since the p-values are smaller than 0.05. So we included them in the next model to have further

analysis in order to check whether gender had impact on the hiring results. In the result of the second model, we found that gpa, leadership presence rating, speaking skills rating, technical task scores and writing task scores are all significant since the p-value is less than 0.05. Therefore, there is significant evidence that all scores and ratings made by AI service are associated with the final hiring result. And there is no evidence that gender affects the final result by using AI service. Therefore, we can conclude that AI service works and is reliable for future use without gender bias.

Characteristic	**OR**	**95% CI**	**p-value**
gpa	0.05	0.01, 0.28	0.002
gender			
Man			
Prefer not to say	728,806	0.00, NA	>0.9
Woman	5.01	1.18, 34.5	0.049
extracurriculars	0.22	0.03, 0.93	0.064
work_experience	1.60	0.39, 8.54	0.5

Table 1: Overall Key statistical summaries of the Generalized Linear Model related to accepted applicant.

Characteristic	**OR**	**95% CI**	**p-value**
gpa	2.36	0.64, 9.53	0.2
gender			
Man			
Prefer not to say	14,383,174	0.00, NA	>0.9
Woman	1.80	0.44, 8.01	0.4
leadership_presence	0.38	0.24, 0.56	<0.001
speaking_skills	0.47	0.32, 0.64	<0.001
technical_skills	0.91	0.86, 0.95	<0.001
writing_skills	0.90	0.85, 0.95	<0.001

Table 2: Key statistical summaries of the Generalized Linear Model related to AI scoring in phase 2.

Gender bias in the hiring process

Data Organization, Data Wrangling and Visulization

In this part, we kept using the dataset in the first research question. In order to investigate whether it is fair in the hiring process, we made barplots for final result by gender and also boxplot for gpa to check if final hired candidates had higher gpa than rejected ones.

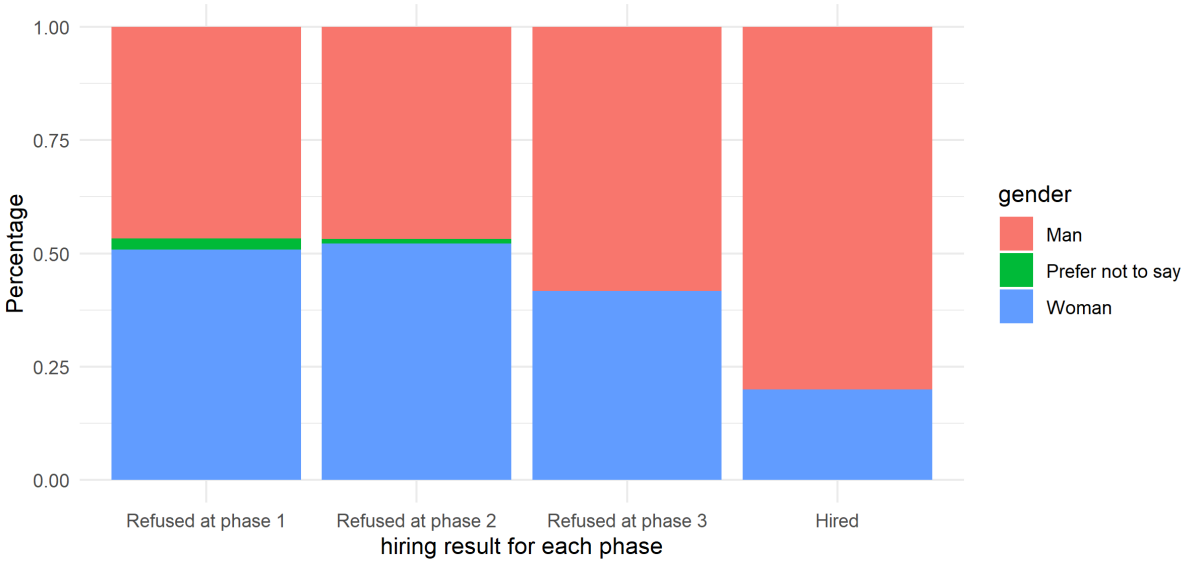

Figure 2: Proportion of Applicants who were refused in each hiring pipeline or finally hired by gender

In Figure 2, the percentage of applicants who can proceed from the first and second phases to the next phase was approximately evenly split between gender. However, a larger percentage of women were rejected at phase 3 than that of men. And all applicants who prefer not to say gender were rejected at phase 3. Therefore, there is a significantly uneven distribution on gender for all hired candidates.

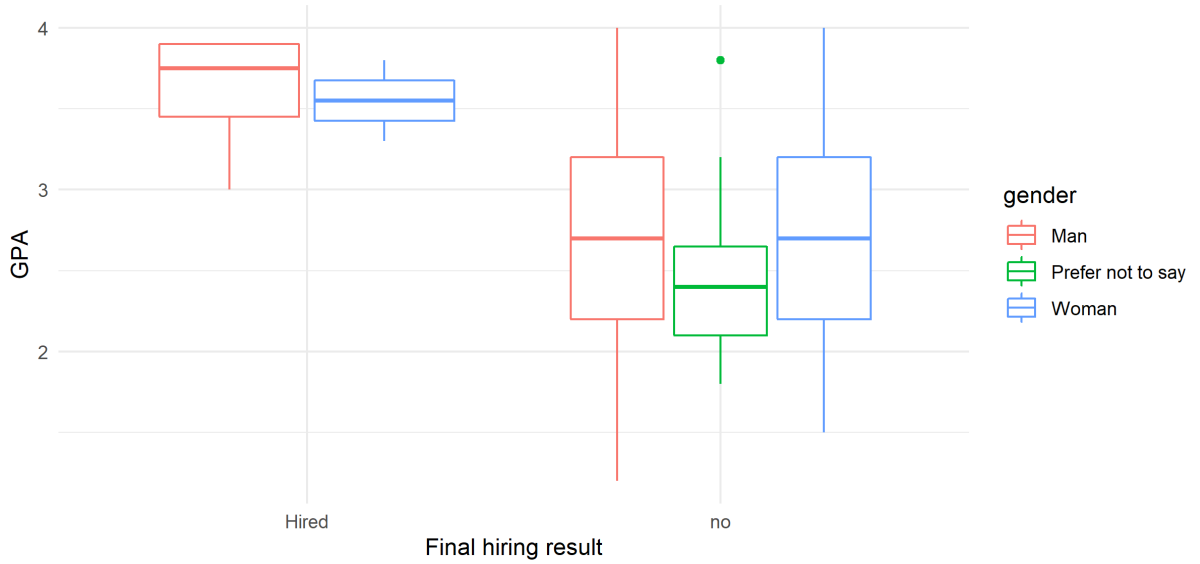

Figure 3: GPA summary of candidates who were refused in the hiring phases or finally hired by gender

As we can see in Figure 3, the boxplot for gpa of candidates grouped by gender and final hiring result, applicants who got hired have a higher average of gpa than those who were rejected before. And there was no significant difference in gpa for those who got hired between gender.

Results

From the plot, we found that the final hired people did have higher gpa than the rejected ones. In order to check whether the hiring process is fair, we can refer to the previous models we made for our first research problems. We found that in the result of model 1, gender is a significant variable while in model 2, gender is not a significant variable that had effects on the final hiring result. We might believe that no potential bias exists in the hiring phase 2 where we used AI service but there exists gender bias in the other phases.

A potential gender bias in the salary and promotion processes

Data Organization

In the current employee data set, we have selected gender, team, leadership_for_level, productivity, and salary. We will focus on the following variables from black-saber-current-employees.csv:

Variable	Description
employee_id	Unique 5-digit employee identifier

Variable	Description
gender	Gender of employee: man, woman, prefer not to say
team	one of the 8 teams which each employee works for
leadership_for_level	if the employee is appropriate for level, need improvements or exceeds expectations
productivity	work output with respect to job description rated on a 0 - 100 scale
salary	salary at the given financial quarter

Data Wrangling and Visualization

We found that in the dataset for the current employees, the salary is saved as character rather than number so we did some data wrangling and changed other data types for future use.

In order to explore whether gender bias existed in the promotion and salary process, we made several plots to show the relationship between salary, promotion and gender more clearly.

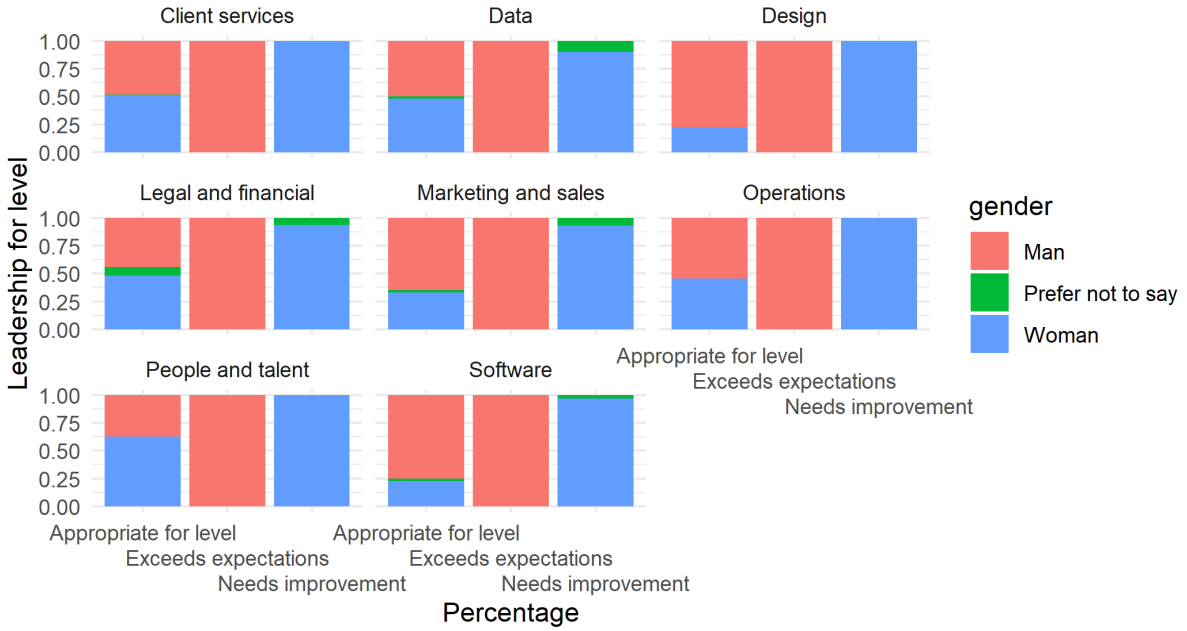

Figure 4:Proportion of current employees with different leadership level by gender for each group

Figure 4 displays the leadership for level in BSS for different teams separated in gender. It is clear to see that in all teams, most women are considered as needing improvement while employees that exceeds expectations are all men. Hence, it is obviously that gender does become a potential variable that affect employees' leadership for level.

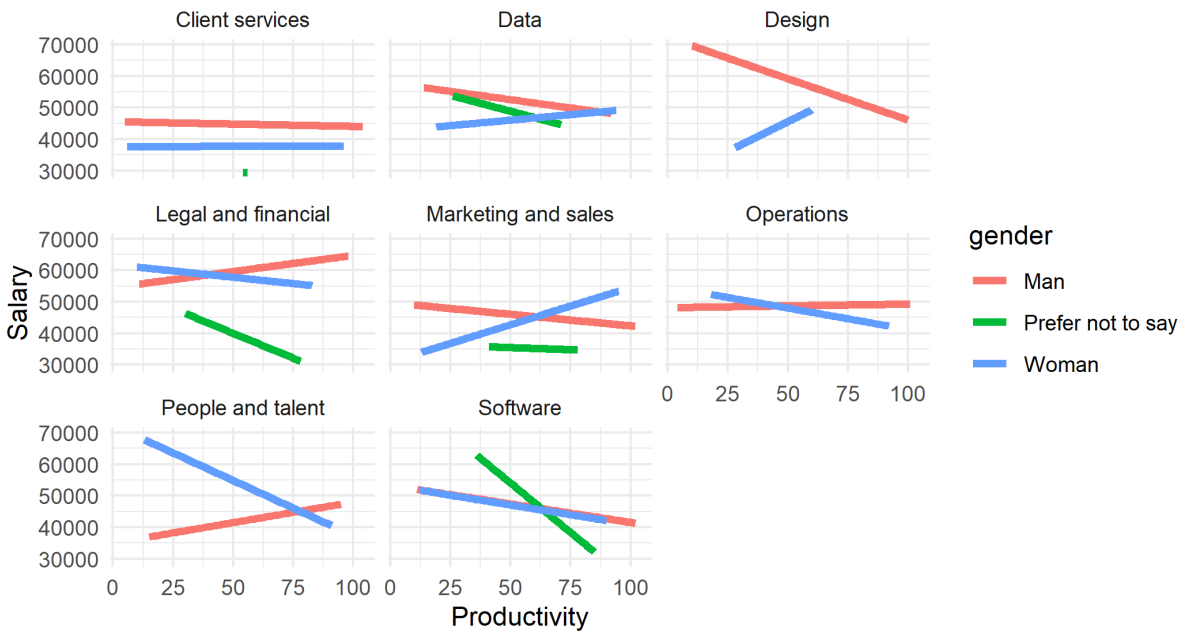

Figure 5: Relationship between productivity and salary in each team by gender

In Figure 5, current employees with higher productivity had salaries especially for those who prefer not to say their gender in most teams which indicates potential bias existed in the salary process.

Models and Results

Table 3 is a summary of a linear mixed effect model to estimate the salary by gender, team and employee id (random effect).

Characteristic	**Beta**	**95% CI**
gender		
Man		
Prefer not to say	-4,115	-14,709, 6,479
Woman	-4,723	-7,560, -1,886
team		
Client services		
Data	7,246	2,507, 11,984
Design	12,415	3,893, 20,937
Legal and financial	10,743	4,185, 17,302

Characteristic	Beta	95% CI
Marketing and sales	3,280	-1,028, 7,589
Operations	5,932	1,235, 10,630
People and talent	7,471	517, 14,425
Software	4,227	41, 8,412

Table 3: Key statistical summaries of the Linear Mixed-Effect Model to check salary related variables.

Table 4 is a summary of a linear mixed effect model to estimate the salary by gender, team, role seniority, productivity, leadership for the level, interactive terms (employee id and team), random effect (employee id).

Characteristic	Beta	95% CI
gender		
Man		
Prefer not to say	-1,150	-3,174, 874
Woman	-2,239	-2,780, -1,697
team		
Client services		
Data	2,405	1,501, 3,309
Design	-199	-1,828, 1,430
Legal and financial	2,832	1,584, 4,080
Marketing and sales	736	-86, 1,557
Operations	498	-398, 1,394
People and talent	-1,394	-2,722, -65
Software	2,759	1,961, 3,556
role_seniority		
Director		
Entry-level	-90,809	-91,101, -90,517

Characteristic	**Beta**	**95% CI**
Junior I	-85,324	-85,604, -85,044
Junior II	-82,945	-83,216, -82,675
Manager	-50,292	-50,525, -50,060
Senior I	-77,201	-77,464, -76,938
Senior II	-71,814	-72,070, -71,559
Senior III	-66,311	-66,561, -66,061
Vice president	29,978	29,606, 30,350
productivity	-0.84	-3.2, 1.5
leadership_for_level		
Appropriate for level		
Exceeds expectations	-4.4	-143, 134
Needs improvement	245	61, 428

Table 4: Key statistical summaries of the Linear Mixed-Effect Model to check salary related variables.

#Df	LogLik	Df	Chisq	Pr(>Chisq)
12	-73106			
24	-58578	12	29056	0

Table 5: Comparison between two different Linear Mixed-Effect Model.

By comparing the linear model used in table 3 and table 4, we see that the log-likelihood for table 3 model is larger with a p-value close to 0. This suggests that the model used in table 4 is a better model that encompases the effect of variables on salary. Therefore, we would want to interpret table 4 to see if there are potential biases.

Table 4 shows a similar result as table 3.

If we look at the differences in salary between gender characteristics, men tend to earn 2239

dollars more compared to women with statistical significance at the 0.05 level which is confirmed by the 95% CI in the table.

Next, looking at the difference in salary between teams, Client services tend to earn less compared to data, legal & finance, and software teams with statistical significance at the 0.05 level which is confirmed by the 95% CI in the table.

Looking at the differences in salary between role seniority, Directors tend to earn much more than other levels except for vice president as expected with statistical significance at the 0.05 level which is confirmed by the 95% CI in the table.

Finally, when we look at the differences in salary between leadership for level, employees who are appropriate for their level tend to earn 245 dollars more compared to employees who need to improve in performance with statistical significance at the 0.05 level which is confirmed by the 95% CI in the table. On the other hand, employees who exceed expectations do not show statistically significant results to be compared with employees at the appropriate level.

In conclusion, there does seem to be biases within gender, team, role seniority, and productivity illustrated in table 4. Going back to our research question, there does seem to be biases within gender, team, role seniority, and productivity affecting the salary of employees within the data given by Black Saber Software on current employees.

Discussion

First, we explored whether the AI used during the second 2 phase of the hiring process was reliable, and if there were any biases. Fitting a model to model the applicants result on the second phase ("refused" if they did not move on to phase 3, "proceed" if they do move on to phase 3) with the applicants' GPA, extracurriculars, work experience, leadership presence, speaking skills, technical skills, and writing skills as the effects, it was found that GPA, leadership presence, speaking skills, technical skills, and writing skills all have a significant effect on an applicants' success rate. Gender, however, did not have a significant effect on whether you would proceed or not. However, gender did not have a significant effect on your result for the second phase.

Second, we explored the effects the various variables during the hiring phases had on the hiring chances. Fitting a model to model the applicant's hiring result (no if not hired, yes if hired) with the applicants' GPA,Curriculum Vitae, cover letter, extracurriculars, and work_experience as effects, it was found that only the GPA and extracurriculars had a significant effect on the being hired, while gender did not have a significant effect.

Finally, we took a look to see if the employees' promotions and salaries are fair and justified based on their ability and value to Black Saber Software. We first separated the salaries by gender,

and graphed each teams' employees' salaries against their productivity. We found that there was indeed a bias for gender. It was observed that the salaries of men based on their productivity is almost always inversely related to the salaries of women based on their productivity. What this means is that as a man's productivity goes up, they earn more, while as a woman's productivity goes up, they earn less, and vice versa. This rule is only broken in the Software team, where men and women all earn roughly the same amount. Furthermore, after modeling the data, it was found that the employees' salaries were indead biased based on many variables, including gender. We will not comment on the employees who did not disclose their gender as it is not indicative of bias in gender.

Limitations

One limitation is that we were only provided data on current employees. The data of previous past employees is missing. Regarding the promotion and salary processes, past employees' information could give us some insights to why they left the company. This could give valuable information such as whether they were treated evenly. If we can get the past employees' data, we can analyze whether there exist biases during the remuneration process more comprehensively.

Another limitation results from the lack of ethnicity/race data for both the hiring and remuneration processes which could also be a potential bias that we need to check. We were told that the race data was not collected by the People and Talent team but they were considering it for EDI initiatives. The race might also be a factor that would affect the hiring result of applicants and also the promotion and wages processes of employees.

Finally, a limitation may be introduced as a result of implicit evaluation criterions for each hiring phase, especially AI-autograded phase and remuneration process. If we know how applicants can proceed to the next round in the hiring process, it would be easier to identify what effects are correlated and fit better models. For the remuneration process, if we can know what kinds of employees can be promoted or have higher salary, we will be able to explore more about potential biases.

Conclusion

We have found that, while there is no gender bias in the hiring process, there are indeed biases when it comes to the employees' salaries and promotion.

Consultant information

Consultant profiles

Lucas Xian: Lucas is a senior consultant with SFP. He specializes in data analysis and visualization. Lucas earned his Honours Bachelor of Science, Specialist in Statistics, from the University of Toronto in 2023.

Dandan Zhang: Dandan is a junior consultant with SFP. She earned her Bachelor of Science, majoring in Mathematics and Statistics, from the University of Toronto in 2021.She specializes in data visualization.

Jong-Hoon Kim: John is a senior consultant with SFP. He earned his Bachelor of Science in statistics and economics, from the University of Toronto in 2021. He specialises in data analysis and visualisation.

Zikun Lei: Zikun is a senior consultant with SFP. He earned his Bachelor of Science in statistics and economics, from the University of Toronto in 2021. He specialises in data analysis and visualisation.

Code of ethical conduct

SFP is a statistical consulting company. Our mission is to help our clients solve their problems and give them professional and practical advice. We, all staff in SFP adhere to the Code of Ethics Conduct as outlined below.

Professional Integrity and Accountability: We always serve our clients with integrity and respect. No fabricated or altered data that deliberately affects the results will be used. We promise to be fully qualified with accepted work and be honest to clients if any limitation of competence exists. Appropriate and relevant statistical methods are always used to produce reliable results. All results and conclusions should be clearly explained in detail with statistical data and valid analytic approaches.

Responsibilities to clients: We treat all clients' information confidential and responsible. No client's information should be disclosed or discussed in public without permissions of the clients. We will protect any sensitive data or results from unauthorized people and will not take advantage of them to make profit.

Responsibilities to other statisticians: We always safeguard the reputation of statistical practice and try to enhance our professionalism. We support and help strengthen the work of other fellow statisticians through valuable peer view. No criticism is allowed to cast on individuals rather than on problems. Even in debates, we should be respectful of others.