An Evaluation of Salary, Promotion, and Hiring Process at Black Saber Software

Evidence of Gender Bias in Salary

Report prepared for Black Saber Software by Calabash Brothers Solutions

2021-04-21

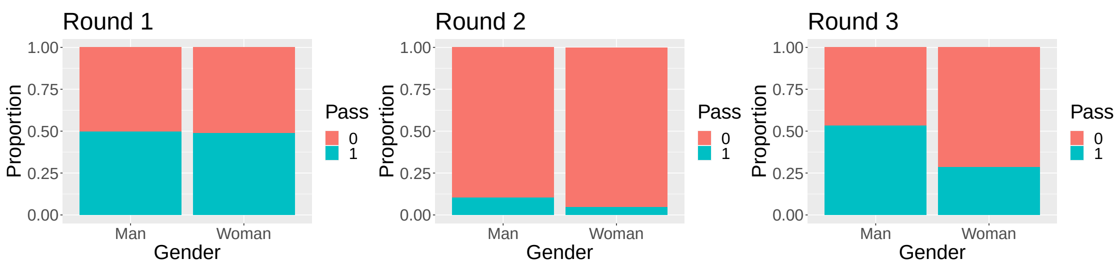
Contents

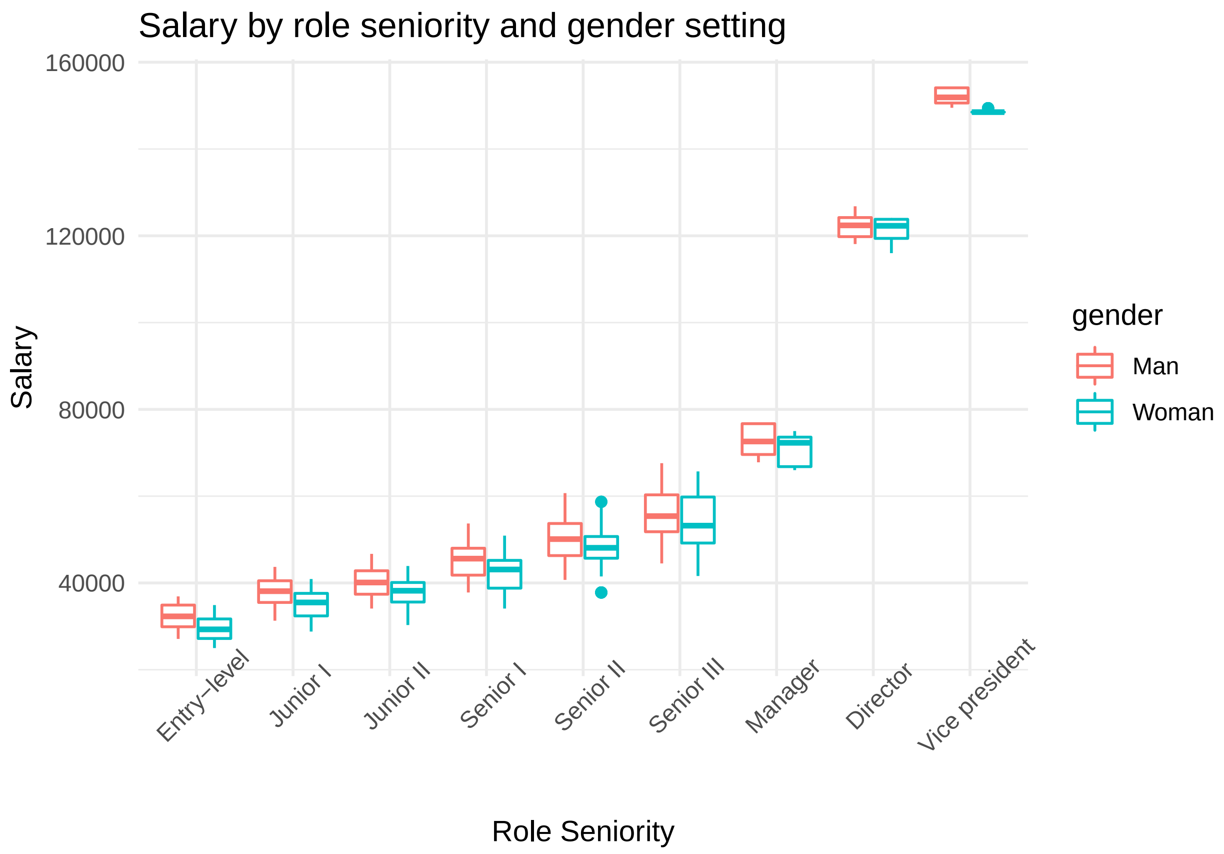
Executive summary

Over the past few months, several software companies had received bad press about their hiring and remuneration processes, during which human bias are present. As a consequence, people have been raising concerns about potential bias in the recruitment and promotion process of Black Saber Software. Experts in our advisory and consulting group are hired to ensure the client company is out ahead of any potential such issues. This study following set out to assess whether the client company exhibits a bias against female in hiring, as well as 2013-2020 salary and promotion.

Results are summarized below.

- The gender bias are not present in the hiring process. In Figure 1, there is no obvious difference in the ratio of being selected for male and female. Females and males have equal opportunities to be selected during the recruiting including AI selection and interviews. Neither men nor women are more favorable.
- In Phase I of the recruitment process, individuals who have higher GPA or those who have demonstrated some work experience in the past are more likely to be automatically selected by the AI to the next round. However, surprisingly, young professionals who have rich extracurricular experiences do not seem to have a notifiable advantage in the first round of the hiring.
- In Phase II of the hiring, GPA and work experience are no longer a factor that will have an effect on the selection process. Instead, the evaluation of individuals focus on technical skills as well as other soft skills such as writing skills, speaking skills and leadership skills.
- In Phase III of the hiring, the most influencing factor is the rating of two interviewers. By analyzing data of acceptance and rejection during interview, there is still no gender bias observed.
- Overall, current male employees have higher salary compared to their female counterparts who have the same role seniority under the same department (see Figure 2). Female employees are not paid with the same amount salaries even if they have the same amount of workloads as male employees.




Salary by role seniority and gender setting
160000
120000
80000
40000
Salary
Entry−level
Junior I
Junior II
Senior I
Senior II
Senior III
Manager
Director
Vice president
Role Seniority
gender
Man
Woman

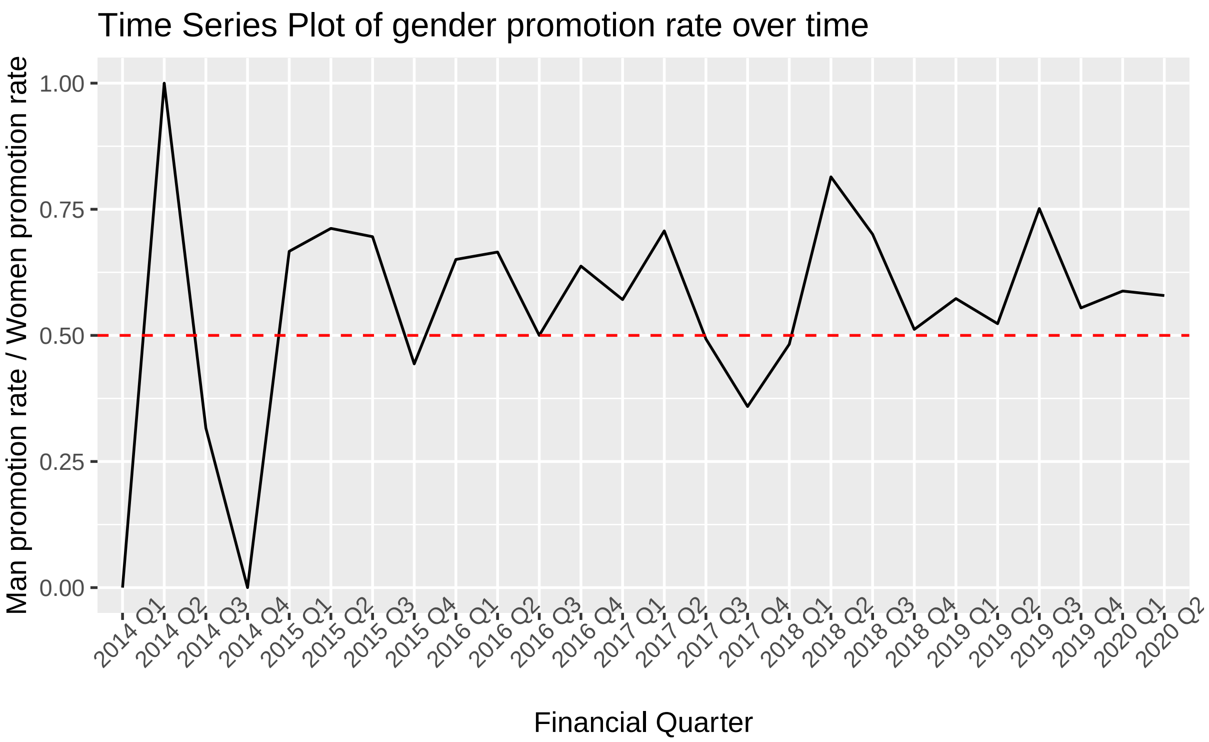


Figure I: The time series plot describes the gender differences in promotion rates over time. The horizontal red line in the middle represents the point at which the promotion rate of two genders are equal.

Technical report

Introduction

Over the past few months, several companies in software have had bad press on people having complaints about their salaries, applicants having concerns about potenial bias in the hiring process. Chief People Officer | Black Saber Software had hired us and provided us with hiring data for new grad program, and data about promotion and salary for the entire staff. The report is to investigate whether there is potential bias in the hiring and remuneration processes, and to ensure that the company is out ahead of any potential issues.

Research questions

- Is there any gender bias in the process of hiring?
- Is there any gender bias in the remuneration processes of the company?
- ?

Research Question 1: Is there any gender bias in the process of three-phases recruitment?

It is known that the recruitment process in Black Saber Software Company consists of three phases, where the first two phases are conducted by AI to select most appropriate candidates and the last phase is normal interview. Therefore our examination of fairness of the hiring process will be focusing on analyzing hiring data in each phase respectively. By using the provided data sets containing information for applicants in each phase (including the team they would like to apply for, GPA, gender, extracurricular, etc), we can construct statistical model using gender and other covariates to explain how likely a given applicant can be selected in each phase. Whether of not gender bias exists can be checked by its significance in the model we construct. This is the general methodology that we use to resolve this question. We will now go through the detailed process of model construction for each phase as follows.

Phase 1 $\rightarrow$ Phase 2

We start by left joining applicants data in phase 2 into phase 1 by using function `left_join` in the `tidyverse` package in R with applicant ID and add binary variable `pass_phase1` to indicate whether the applicant is selected to the next phase.

In order to build model, some exploratory data analysis is performed as follows. Note that Figure 2 is the visualization for the contingency table that represents the proportion of applicants in each gender that are selected to the next phase. It can be seen that the proportion of men and women that are selected to the next phase is roughly the same. Out of those who apply for data team, 0.515625 are selected and out of those who apply for software team,0.4705882 applicants are selected to move on. Figure 3 compares the proportion of applicants who are selected by their gender and the team they apply to work with. The plot suggests that for each gender group, their is a slightly higher proportion of acceptance if the applicant is willing to work in the data team compared to the software team. Therefore, a term for the team that applicants want to apply for will be considered in our model. It is also worth noting that no obvious difference can be observed between men and women group, which also agree with our previous plot.

Our outcome of interest is whether the applicant passes phase 1, and our predictors are GPA, work experience, extracurricular and team the applicant applies for, treated as a factor. As our outcome variable is binary, we are going to use logistic regression.The model can be written as:

$$Y_i \sim Bernoulli(p_i)$$

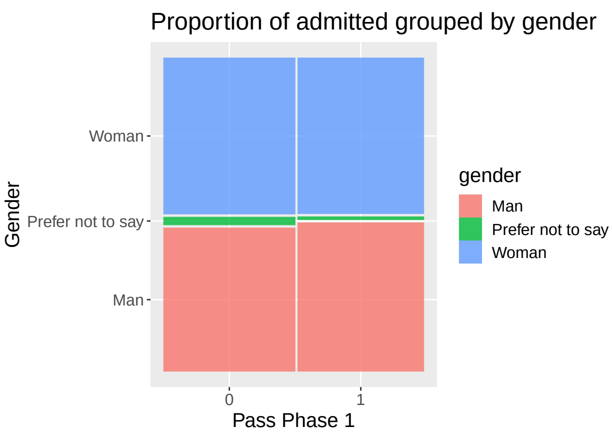
$$\log(\frac{p_i}{1-p_i}) = \beta_0 + \beta_1 GPA_i + \beta_2 \text{Extracurricular}_i + \beta_3 \text{Work Experience}_i + \beta_4 \text{Team applied for}_i$$

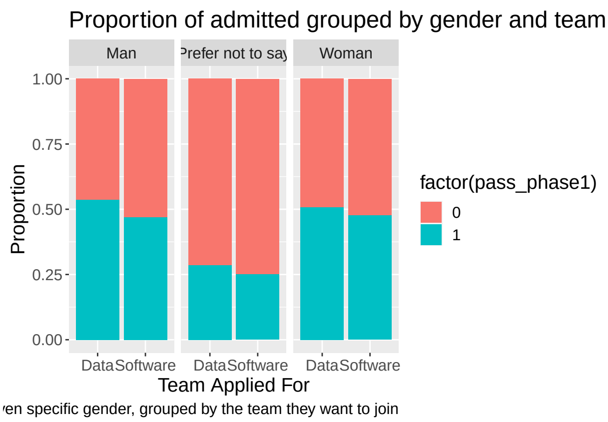
where Y_i is the acceptance of applicant i , X_is are explanatory variables.

The model is realized in R through the `glm` function. Results are shown in table 1, with estimated coefficients exponentiated and confidence intervals calculated along with p-value. Not surprisingly, we find that GPA and work experiences are all significant at level 0.05. Further, 1 unit increase of GPA is associated with the odds of being admitted increasing by 621%. Similarly, obtaining level 1 working experience will increase the odds of passing phase in by 335% compared to level 0 working experience. Most importantly, variable “gender: women” is highly insignificant with p-value 0.8576, which supports the nonexistence of gender bias in the first phase of hiring. Similar thing for the category where applicants are not willing to tell their gender.

Table 1: Logistic regression result for Phase1 to Phase2 hiring

	Estimate	Lower Limit	Upper Limit	p-value
Intercept	0	0	0	0.9638
GPA	7.21	4.78	11.18	0
Extracurriculars 1	8574079.72	2.23e+25	4.42e+92	0.9743
Extracurriculars 2	7877882.23	2.18e+38	1.88e+93	0.9744
Work Experience 1	4.35	2.12	9.87	0.0001
Work Experience 2	3.83	1.42	11.17	0.0103
Gender: Prefer Not To Say	0.5	0.08	2.33	0.4111
Gender: Women	0.96	0.64	1.45	0.8576
Team Applied for: Software	0.91	0.6	1.37	0.6457





Phase 2 → Phase 3

Similarly as the previous phase, we are going to check if there exists gender bias from phase 2 to phase 3. See Figure 4 and 5 for comparison of proportion of passing phase 2 by gender and the team they applied to. In this case, we find less difference between applicants that apply to data or software team than the previous phase, which may indicate the insignificance of this variable. Since applicants that are accepted from phase 1 should go through a series of tests online to check their technical skills, writing skills, we start our model by adding these variables with the predictor variables in the model we built in the previous phase.

Therefore, the outcome variable is whether or not the applicant passes phase 2. For the first model we build, the explanatory variables are GPA, technical skills, writing skills, leadership presence, speaking skills and the team applicants want to join. See the mathematical formula:

$$Y_i \sim Bernoulli(p_i)$$

$$\log(\frac{p_i}{1-p_i}) = \beta_0 + \beta_1 GPA_i + \beta_2 \text{Technical Skills}_i + \beta_3 \text{Writing Skills}_i + \beta_4 \text{Speaking Skills}_i +$$
$$\beta_5 \text{Leadership Presence}_i + \beta_6 \text{Team applied for}_i + \beta_7 \text{Work Experience}_i + \beta_8 \text{Extracurriculars}_i$$

See table 2 for the result of model coefficients after exponentiation, which is realized by `glm` function. Notice that among these predictor variables, only skill-related ones are significant under level 0.05. Therefore, we would like to remove some insignificant covariates to simplify our model. By removing GPA, working experience, extracurricular and team that the applicants want to join, we come up with the final logistic regression model, written as:

$$\log(\frac{p_i}{1-p_i}) = \beta_0 + \beta_1 GPA_i + \beta_2 \text{Technical Skills}_i + \beta_3 \text{Writing Skills}_i +$$
$$\beta_4 \text{Speaking Skills}_i + \beta_5 \text{Leadership Presence}_i$$

See table 3 for the new model results and this is a better model to fit since the AIC of the model drops from 87.527 to 83.574. The resulting table also suggests fairness between male and female since they are still insignificant with p-value $0.4205 > 0.01$.

By conducting logistic regression, we find the AI algorithm used by Black Saber Software Company for selecting applicants in phase 2 does not depend much on the significant variables in phase 1 but on the test scores for relevant skills and the selection process is also irrelevant with one's gender.

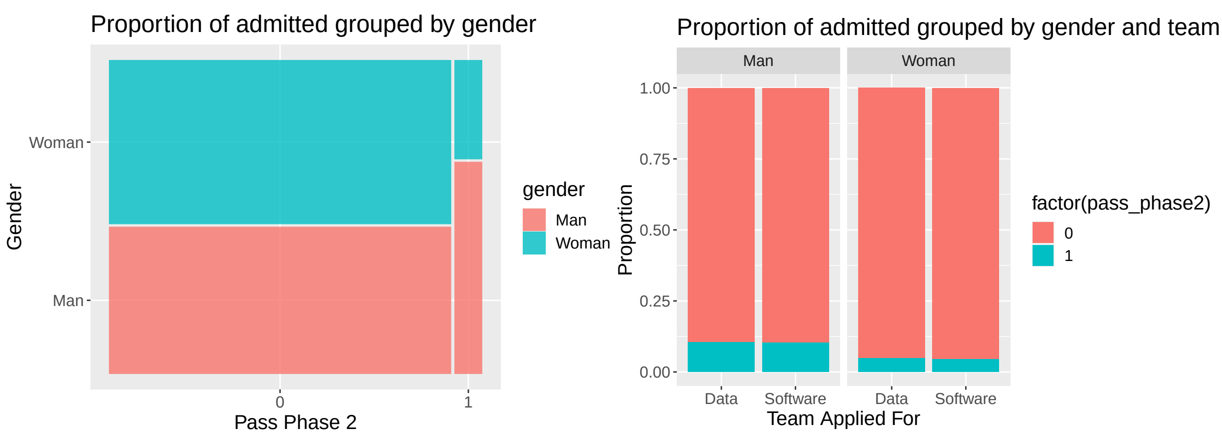


Table 2: Logistic Regression Result for first model built from Phase 2 to Phase 3

	Estimate	Lower Limit	Upper Limit	p-value
Intercept	0.00	0.00	0.00	0.0024
Gender: Woman	0.47	0.09	2.24	0.3503
GPA	0.47	0.09	2.19	0.3434
Technical Skills	1.11	1.06	1.18	0.0001
Writing Skills	1.12	1.06	1.20	0.0001
Leadship Presence	2.86	1.90	4.89	0.0000
Speaking Skills	2.46	1.68	4.05	0.0000
Team Applied For: Software	3.93	0.96	20.33	0.0743
Work Experience 1	0.30	0.00	193.38	0.8312
Work Experience 2	0.37	0.00	285.91	0.8593
Extracurriculars 2	0.65	0.14	2.78	0.5673

Table 3: Logistic Regression Result for final model built from Phase 2 to Phase 3

	Estimate	Lower Limit	Upper Limit	p-value
Intercept	0.00	0.00	0.00	0.0000
Gender: Woman	0.56	0.12	2.26	0.4205
GPA	0.42	0.10	1.55	0.2032
Technical Skills	1.10	1.05	1.16	0.0001
Writing Skills	1.11	1.06	1.17	0.0001
Leadship Presence	2.61	1.80	4.24	0.0000
Speaking Skills	2.12	1.55	3.12	0.0000

Table 4: Logistic Regression Result from Phase 3 to Final Decision

	Estimate	Lower Limit	Upper Limit	p-value
Intercept	0.01	0.00	118491133073437.20	0.7786
Technical Skills	2.10	1.75	2.53	0.0000
Writing Skills	2.14	1.56	2.93	0.0001
Leadership Presence	139.52	20.04	971.60	0.0001
Speaking Skills	118.09	10.39	1341.59	0.0007
Gender: Woman	3.26	0.01	1557.75	0.6898

Phase 3 to Final Result

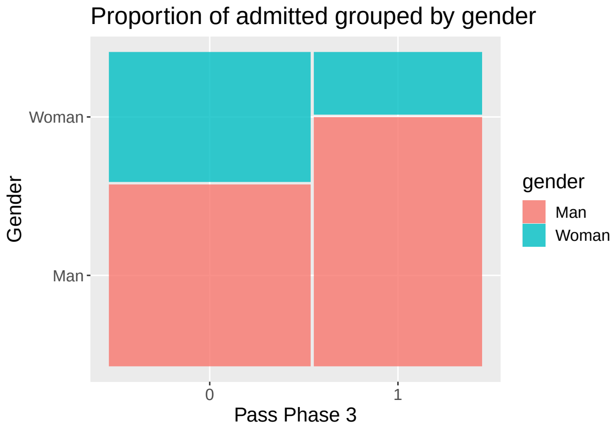
The methodology we use to deal with this phase is different from the first two phases. Since being listed as 'first' or 'second' interviewer is arbitrary, we create a new variable "sum" that is the sum of ratings from two interviewers. After sorting the sum of ratings for applicants who engaged in the interview and select top 10 applicants to match with the actual accepted interviewees, it is found that 9 of them are perfectly matched. Therefore, we can roughly make inference that the decision made in the last stage depends mostly on the interviewers' scores. Hence, we are going to test if there exists gender bias by modeling the factors that may influence the interviewer's score.

Here we are going to build the multiple linear regression model, where the outcome variable is the sum of two interviewers' score and the predictor variables are technical skills, writing skills, speaking skills, leadership presence and categorical variable gender, written as:

$$Y_i = \beta_0 + \beta_1 \text{gender}_i + \beta_2 \text{Technical Skills}_i + \beta_3 \text{Writing Skills}_i + \beta_4 \text{Speaking Skills}_i + \beta_5 \text{Leadership Presence}_i + \epsilon_i$$

Table 3 shows the results of the linear regression. All the skills related variables are significant except for the categorical variable gender with p value 0.6898, which supports the fairness of the hiring process.

Combining our model results in these three phases of hiring, we conclude that there is no gender bias during recruitment process (both AI selection and interview) in Black Saber Software Company.



Research Question 2: Is there any gender bias in the remuneration processes?

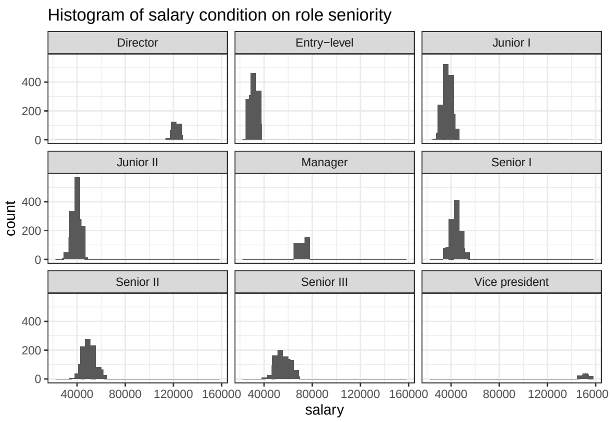
Black Saber Software provides us with data about promotion and salary for the entire staff. The data we get consists information of the staff including annual salary, gender, job in the company, financial quarter salary, productivity, role seniority, and quality of demonstrated leadership. We use these data to investigate whether there is any gender bias in remuneration process. The general methodology that we use is creating a boxplot of salary conditions on role seniority and gender setting, and then applying different types of regression models to see which one best serve our purpose.

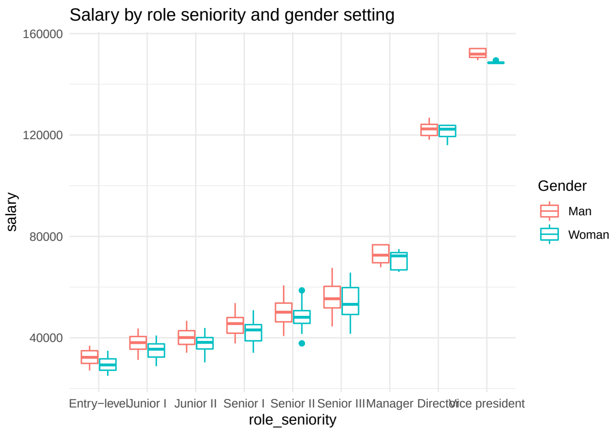
From the Figure2 in the executive summary, we can see that women have lower salaries across different role seniority.

In order to test whether gender bias truely exists, we need to first find the most appropriate model. We create a simple linear regression model:

$$salary = \beta_0 + \beta_1 * Productivity + \beta_2 * Gender$$

We find out that gender is statistically significant with p-value far less than 0.05, which suggests that gender has impacts on salaries. Moving on from SLR, we add the variable "team"(job in the company) as the random intercept in previous model to create a new linear mixed model. We do such because similar jobs may provide correlated salaries, thus job type can serve as a group.The likelihood ratio test tells us including such does explain the data better with a small p-value. The results from the linear mixed model also show us that gender is significant. Next, we add a random slope to the previous model because the effect of the productivity may be different for different jobs in the company. After the likelihood ratio test, we find that slopes are an unnecessary complication to our model. Finally, we add role seniority as another random intercept because there should be variations in the salaries across different role seniority based on our histogram. For example, senior workers have higher salaries than junior workers. The likelihood ratio test tells us that including this random intercept does explain the model better. Therefore, we can realize our model using the `lmer` function in the `lme4` package with `lmer(salary~productivity+Gender+(1|team)+(1|role_seniority), data =new_cemployees)`, results of model fitting can be found in Table 5. Thus, based on interpretation of our final model, woman has lower salary by an average of $1758 than man. This result indicates that there is gender bias in remuneration.





	Estimates
Intecept	67253.17
Productivity	1.27
Gender: Women	-1757.64

Discussion

For our research question 1: Is there any gender bias in the process of three-phases recruitment? We do not find any gender bias throughout the entire hiring process(AI and In-Person Interview) since gender is always insignificant when it is involved in explaining the acceptance of an applicant, as p-values shown before. We also find that GPA and working experience are what the AI algorithms look for in the first phase. Various skills are assessed in most weight during the second phase. Anyway, the AI algorithm and interviewers account heavily on applicants' skills not gender. The hiring process is considered to be fair. However, limitations to our models still exist, which will be discussed very soon.

For the research question 2: Is there any gender bias in remuneration processes? We do find gender bias in salaries. After applying several models, we find the most appropriate model to predict salaries for employees. The results from our model show that woman has lower salary by an average of $1758 than man. We can clearly see our results consistent with the boxplot we created, men tend to have higher salaries across various jobs and role seniority. The limitation to this research question is that the model of our choice may not be the most appropriate model.

For research question 3: Is there any

Strengths and limitations

The limitations of testing gender bias in the three-phase hiring process is that the proportion of people that pass the examination for each phase is very small. It might not give us enough information to determine whether or not there is gender bias. For example, in phase 3, it is actually a rough approximation to model the interviewer's score to predict the result of interview because phase 3 only has 22 people, we do not have enough data to investigate if interview's score is the best predictor for final decision.
Another limitation is that when we try to predict interviewer's score, the covariates that we chose are not guaranteed the best ones since we just used data from previous phases. There must be some other factors that will affect one's performance in interview other than the AI assessment scores based on real life experience. Our next steps will be to use richer data set to perform more accurate modeling.

Consultant information

Consultant profiles

Chengen Dong. Chengen is a senior consultant with Calabash Brothers Solutions. He specializes in data visualization. Daniel earned his Bachelor of Science, Specialist in Statistics Methods and Practice, from the University of Toronto in 2024.

Yuxuan Longt. Yuxuan is a senior consultant with Calabash Brothers Solutions. He specialize in reproducible analysis. Peter earned their Bachelor of Science, Majoring in Statistics from the University of Toronto in 2028.

Ran Li Ran is a junior consultant in Calabash Brothers Solutions. She is good at coding with software like R and Python. She specializes in Computer Science. Liran earned her Bachelor of Science, Majoring in Computer Science from the University of Toronto in 2028.

Peizhe Huang Peizhe is a senior consultant with Calabash Brothers Solutions. She specializes in Probability Statistics. Peizhe earned her Bachelor of Science, Majoring in Statistics from the University of Toronto in 2022.

Code of ethical conduct

We apply statistical sampling and analysis procedures scientifically, without predetermining the outcome.

We convey the findings in ways that are honest and meaningful to the clients, and also report the limitations, defects, potential bias and possible sources of error and strive to promptly correct any errors.

We Conforms to confidentiality requirements and guards privileged information of the employer, client, or funder

We avoid the potential social harm that can result from the dissemination of false or misleading statistical work