

---

# **Gender Parity issues for Black Saber Software**

Related to Salary, Promotion and Hiring Process

Report prepared for Black Saber Software by One  
Company

2021-04-21

## Contents

<b>Executive summary</b>	<b>3</b>
<b>Technical report</b>	<b>5</b>
Introduction . . . . .	5
Salary Level Difference and Gender in the Company . . . . .	7
Promotion Differences and the Gender in the Company . . . . .	12
To study whether there is a gender preference on the hiring process. . . . .	16
Discussion . . . . .	20
<b>Consultant information</b>	<b>22</b>
Consultant profiles . . . . .	22
Code of ethical conduct . . . . .	22

## Executive summary

Gender parity is a critical social issue in current years, such as glass ceiling effect to prevent females get top management position. Furthermore, it is important for females to have the same treatment as the males with same level of ability and work in a fair environment. In this statistical report, we are going to analysis the data from Black Saber Software to investigate whether gender parity is existed in salaries, promotion and phases of hiring.

We would divide the objective into three sections firstly, and each section has a research problem that corresponds to the topic of salary, promotion and hiring respectively. As well, we would mainly consider about salary changes and number of promotion until 2020 Q4. Since the latest data would provide more accurate information in data visualization and fitting model compare with the others. We would provide conclusion about gender parity with specific information of each research problems by analysing the data, including data cleaning and selection, data visualization and fitting proper models.

As we have several key finding by data visualization and fitting model related to gender parity and salary, promotion and hiring:

- Most males tend to get higher level of salary compare with females in the same team, level of position and productivity respectively.
- Gender has significant impact to the level of salary of different gender employees.
- The number of promotion for males are more than females in the same team and productivity respectively.
- Gender has significant relationship with the number of promotion.
- The AI service of hiring process which including 3 phases does not have obvious gender preferences.

However, there are limitations and drawback about the result of this study. As we fitted several model with different predict variables to reduce the difference between estimated results and actual results, but we could not consider all of variables. For example, if there are more male employees in the company also lead to the result that similar with gender inequality. Therefore, we would like to fit varied types of model for the further study in order to get more precision results.

On the other hand, we only used the current updated data of 2020 quarter 4 for the first research question. However, the historical data supposed to be considerable because it could increase the accuracy of model explanation. Therefore, we supposed to analysis the data with longer-range of time to check whether the conclusion is similar with current conclusion or not.

The straightforward results of the study as shown in the summarized tables below:

**Table 1:** Salary Level and Gender of Employees

	Man	Woman
High Level	159	98
Mediam Level	430	230
Low Level	3475	2397

Table 1 shows the number of male and female employees with different level of salary. It is obviously show that male employees has higher level of salary than female employees in all salary level.

**Table 2:** Number of Promotion for Male and Female Employees

	Man	Woman
0	143	149
1	88	62
2	55	27
3	39	14
4	11	4
5	3	1
7	1	0

As Table 2. shown the number of promotion for male and female employees in the company, we could see that there are more promotion chances for male employees. For all number of promotion, there are more male employees than female employees.

As a conclusion, the study would demonstrate the gender of employees has significant impact to the promotion and salary which decided by top management. As well, another objective of the study was to investigate if there are gender inequality issue in the phases of hiring by AI service. However, we supposed to consider the results of hiring depends on other ability of the interviewees which is not straightforward to shown by the gender ratio of interviewees. Therefore, we would like to fit more models and discussed the result in the below discussion section.

## Technical report

### Introduction

In current years, the gender equality is an critical issues in order to provide a fair work environment for females. In this project, we would analysis the data from Black Saber Software and demonstrate if the salary, promotion and hiring progress depend on the ability of the employees or the gender.

First of all, the first research question is related to salary which include the if there are salary difference between male and female in the company. We would consider about the predict variable for ability of the employees related to the salary, such as productivity, position and team of the employees. As well, the gender of the employees as predicted variable. Linear mixed model was used to determine the association between salary, which is numeric variable, and other predicted variables. Meanwhile, the position and teams of employees could influence the salary level as well. For example, the salary of top management would be higher than the salary of entry-level. Thus, the position and teams would fitted as random effect for mixed linear models. In order to select a suitable model and interpret the results of the model, we would perform likelihood ratio test to select the proper model. As well, the likelihood ratio test demonstrate the model include gender as fixed effect is better model could explain the models and gender indeed has significant impact to the salary of the employees.

The second research question is related to promotion about the difference of promotion chance between the male and female employees. We would use the number of promotion for individual employees to perform data visualization and fit the Poisson regression model that using productivity and team of employees as predicted variables. As well, the work year would have relationship with promotion chances and we could not compare the observation with different work years directly. Therefore, we would add an offset term to remove the influence by the different work years. The boxplots and point plot shows the number of promotion for male employees are more than the female employees under the same abilities. Furthermore, the p-value of the Poisson regression model shows the promotion would be more likely occur for the male employees.

The third research question is related to results of the hiring phases by AI service between male and female employees. We modelled the results of different hiring phasees with logistic regression model and linear mixed model, using the binary variable about passes the phases as response variable. As well, the other variable related to ability of interviewees as predicted variables, such as gpa, work experience and so on. The result of logistic model interpret the success in the different phases would less like occur for the female interviewees. However, the likelihood ratio test of mixed linear model demonstrate the gender as predict variable does not have significant impact to the phase result to the contrary. We could not conclude based on the opposited results

of the models, but we could conclude the hiring result more likely depend on the ability of the interviewee but the the success are less likely occur for female interviewees.

We conclude that the gender inequity issue would happened in the events that decided by top management such as promotion and salary. The promotion and salary do not completely depend on the ability of the employees. On the other hand, the gender inequality issue does not exist in the different phases of hiring. The ability of the employees would play a critical role in the phases of hiring.

### **Research questions**

- The relationship between salary gap and group of male and female in the company and figure out the specific salary gap.
- The relationship between the promotion differences and gender under the same conditions in the company.
- The relationship between successes in the phases of hiring and the gender of interviewees.

## Salary Level Difference and Gender in the Company

### Methods

The data we are going to use have 8 variables and the salary is the numeric variable as response variable. As well, the rest variables related to the ability including the team, position, productivity, leadership and gender of the employees. In order to reduce the complexity of the model, we would use the current updated data which is 2020 quarter 4 to perform the data visualization and fitting models.

Firstly, we would have three plots to illustrate the salary level for the data visualization purpose in the statistical report. There would be two boxplots to compare the salary level of employees with team and position of the employees respectively. Since Both of team and position of employees are categorical variables, and boxplot would divide into different categories. As well, salary of employees is numerical variables, and the boxplot could demonstrate the understandable quartile for the salary of employees.

Moreover, there would be point plot to compute the salary level of employees with the same productivity in order to show whether productivity of employees would influence the salary level or not. As well, the point plot would show the productivity of individual employee in the company because we would focus on individual salary rather than the salary of a group.

Furthermore, we are going to use mixed linear models to fit the data because salary in the data is numerical variable. We would like to focus on the relationship between the other predict variables and the salary as well, such as the estimated coefficient and statistical significant to the salary as response variable. Therefore, we would select the proper predict variables to explain the data well in order to increase the accurate of the model. In addition, we would use likelihood ratio test to select the better model without overfitting problem.

First of all, gender is the critical fixed effect variable related to our objective of the study. As well, we would select role seniority and team as random effect because the most salary would depends on the position and different department. For example, the salary of employees in the top management or core department would have higher level of salary. In addition, we want to check whether salary has relationship with the ability of the employees. Therefore, productivity and leadership of individual employee could considered as fixed effect variable as well.

Moreover, we would perform likelihood ratio test to select the suitable until we figured out the model and avoid the overfitting problems. In order to check if the salary depends on gender of the employees, we would perform likelihood ratio test to test final model with gender as predict variable and without gender.

## Results

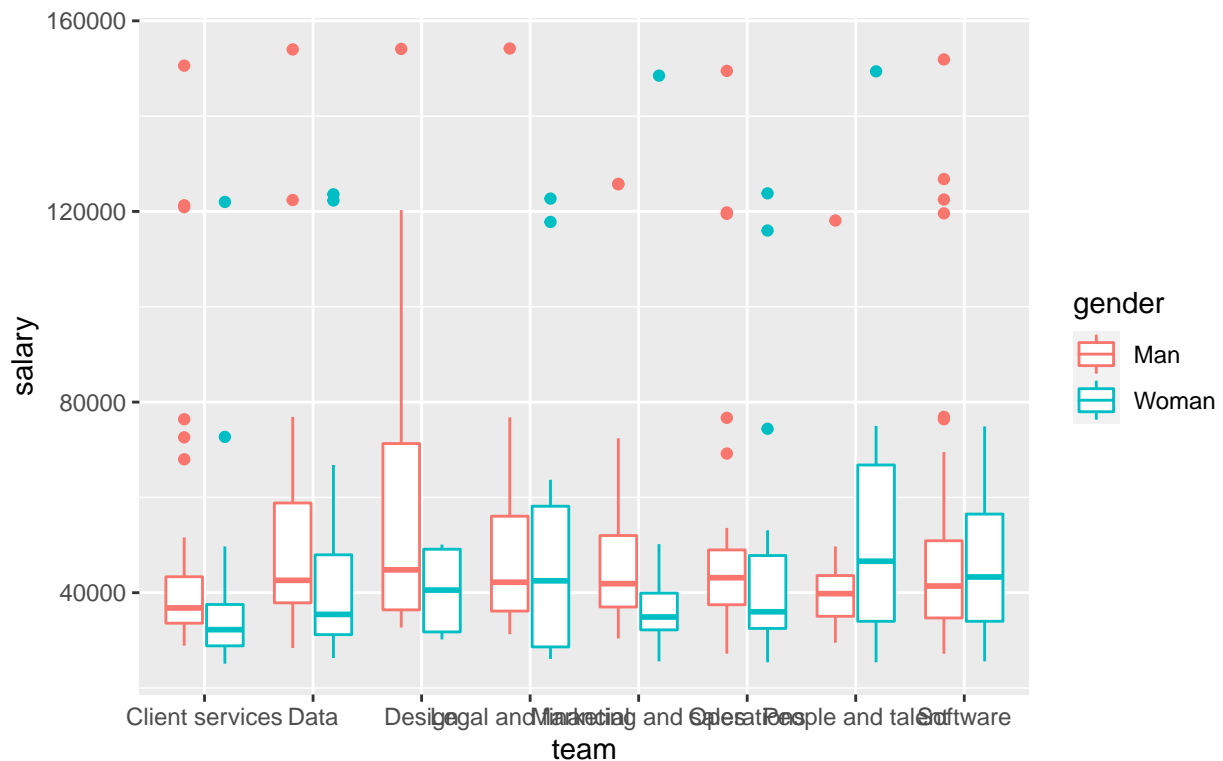


Figure1. Boxplot of Team and Salary  
 Created by: One Company  
 Date of April 21 2021

According to Figure 1, the boxplot explores if the salary would be different from gender in the teams of the company. The figure shows the most teams in the company provide higher salary to male employees than female employees without “Software” and “People and Talent” teams. For example, the median salary of male employees is almost equal to the 75% quartile salary of female employees. As well, the maximum salary for male employees is more than female employees in the most teams. Overall, the salary level of male employees is more than female employees in the same team.



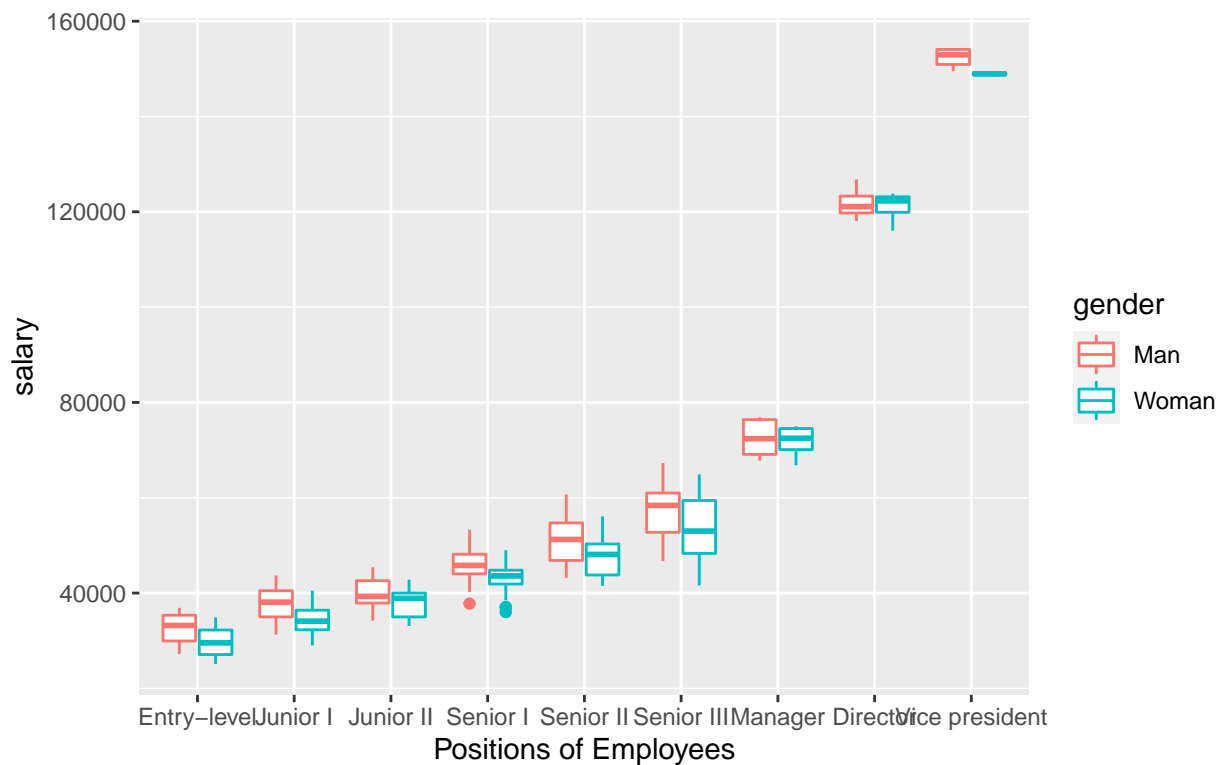


Figure 2. Boxplot of Role Seniority of Employee and Salary  
 Created by: One Company  
 Date of April 21 2021

Figure 2 shows the salary of different salary for different positions in the company. We could clearly see that the most team has higher salary for the male employees than female employees. The median salary of male group is almost equal to the 75% quartile of female salary for the most group. Moreover, the salary of vice president for male employees is obviously exceed the salary of female employees. Overall, the male employees has higher salary than the female employees in the same team for the company.

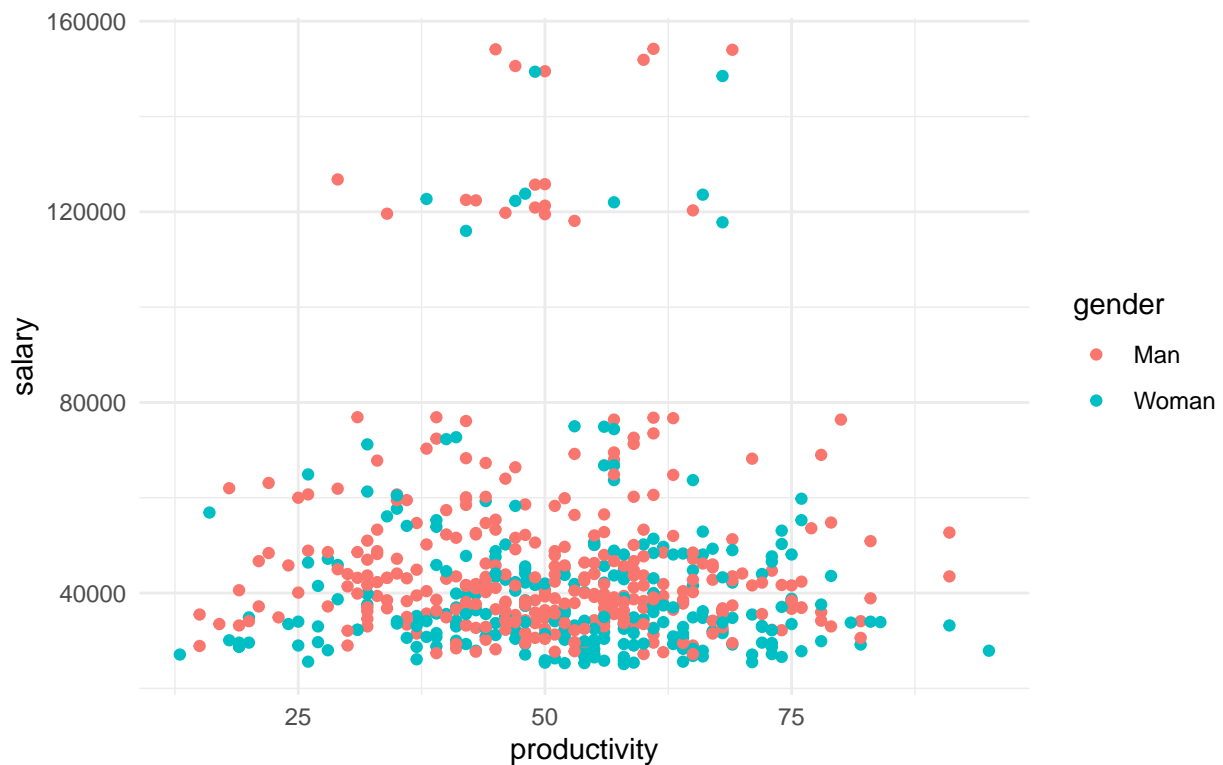


Figure 3. Point Plot of Productivity and Salary  
Created by: One Company  
Date of April 21 2021

Figure 3 is a point plot show the salary for the employees with different productivity. We could see that the most employees with higher level of salary are males and there are only 2 female employees have high level of salary. Meanwhile, we could see the female employees have the same productivity as male employees have less salary. In conclusion, the salary level of male employees is more than female employees under the same productivity.

**Table 3:** Result of Likelihood Ratio Test of Model 1 and Model 2

#Df	LogLik	Df	Chisq	Pr(>Chisq)
4	-5766.013	NA	NA	NA
5	-5762.240	1	7.546486	0.006013

Table 3. shows the likelihood ratio test results, we could see the p-value is smaller than 0.05 which indicate the model with productivity as fixed effect and role seniority as random effect could explain the data better. As well, the position of employees would have influence to the salary level.

**Table 4:** Result of Likelihood Ratio Test of Model 2 and Model 3

#Df	LogLik	Df	Chisq	Pr(>Chisq)
5	-5762.240	NA	NA	NA
6	-5724.836	1	74.80684	5.19e-18

Table 4. shows the likelihood ratio test results, we could see the p-value is smaller than 0.05 and the the model which add team as random effect could explain the data better. As well, the result indicates that the team would influence the salary of employees.

**Table 5:** Result of Likelihood Ratio Test of Model 3 and Model 4

#Df	LogLik	Df	Chisq	Pr(>Chisq)
6	-5724.836	NA	NA	NA
8	-5708.193	2	33.28698	5.91e-08

Table 5 shows the result of likelihood ratio test, we could see p-value is smaller than 0.05 and the model which add leadership of employees as fixed effect could explain the data better with more accuracy.

**Table 6:** Result of Likelihood Ratio Test of Model 4 and Model 5

#Df	LogLik	Df	Chisq	Pr(>Chisq)
7	-5739.025	NA	NA	NA
8	-5708.193	1	61.66426	4.1e-15

Table 6 shows the result of likelihood ratio test, we could see the p-value is smaller than 0.05 which indicate the model which remove gender as predict variable could not explain the data well. The smaller p-value demonstrate the complicated model is proper and that could explain the data with high accuratcy. Therefore, the result also demonstrate the gender of employees is significant impact to the salary of the employees.

## Promotion Differences and the Gender in the Company

### Methods

The data we are going to perform analysis has 10 variable, which the number of promotion would be the response variable. The rest predicted variable related to ability is productivity and team of individual employees. Meanwhile, the work years would be a confounder and we would set a offset term to reduce its impact while fitting the poisson Regression model.

The object of this research question was to investigate the relationship between promotion and group of gender. As well, the data from Black Saber Software provided the information about role seniority of employees from quarter 3 2015 to quarter 4 2020. Therefore, we are going to add a new variable to record the number of position changes for each employee until current information update.

First of all, we would have boxplot compare the number of promotion with team of employees respectively. Since team of employee is a category variable and number of promotion is numeric variable with small range. The boxplot could provide median and interquartile of the number of promotion for each team which is understandable and straightforward to show the specific differences.

Secondly, we are going to use point plot to compare the number of promotion and the productivity of employees. Since both of productivity and number of promotion are numeric variable, and we would like to show the promotion of individually employees under same productivity instead of all male or female employees under same productivity. Therefore, we could compare the promotion difference straightforward by fixing specific productivity on x-axis.

In addition, we modelled the number of promotions with Poisson Regression model, using productivity and gender and team as predict variables. Since Poisson regression model was used to determine the count data as response variable and number of promotion is a count variable. Meanwhile, there would be a confounder in the data which is work years. In common, the employee with longer work year would have more promotion chances and the work year for each employee in the data is different. It is not reasonable to compare the number of promotion under different condition of work experience. Therefore, we would like to add an offset term to Poisson regression model in order to allow the model compare the rate instead of count.

## Results

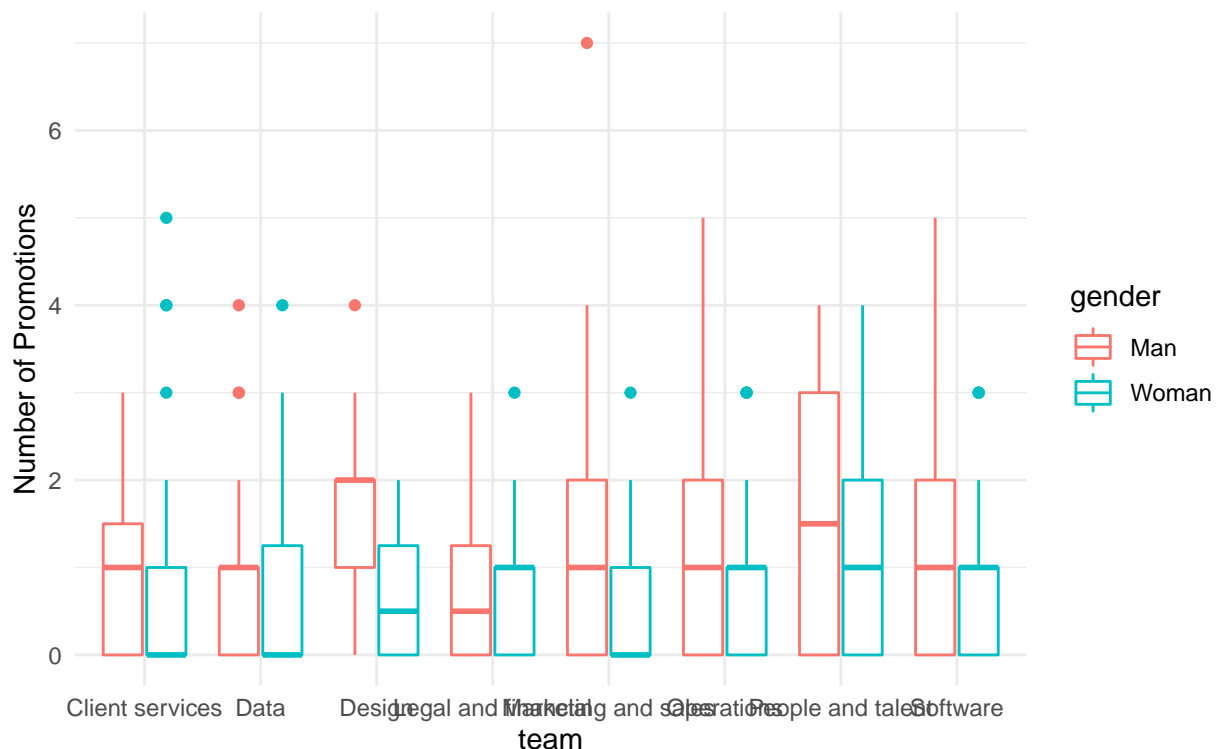


Figure 4. Boxplot of Number of Promotions in Different Teams  
Created by: One Company  
Date of April 21 2021

According to Figure 4, which show the boxplot of number of promotion in different team of two gender groups. We could see the maximum number of promotion of male employee are more than female employees in the most teams. There is a group with maximum number of promotion of female employees is more than male employees which is data team. However, there are outliers in male employee of data group and that is equal to the outlier of promotion number in data team. In addition, the median number of promotion of male employees are equal to the third quartile of female group for the most team.

Overall, the number of promotion of male employees is obviously more than the female group in the most teams in the company.

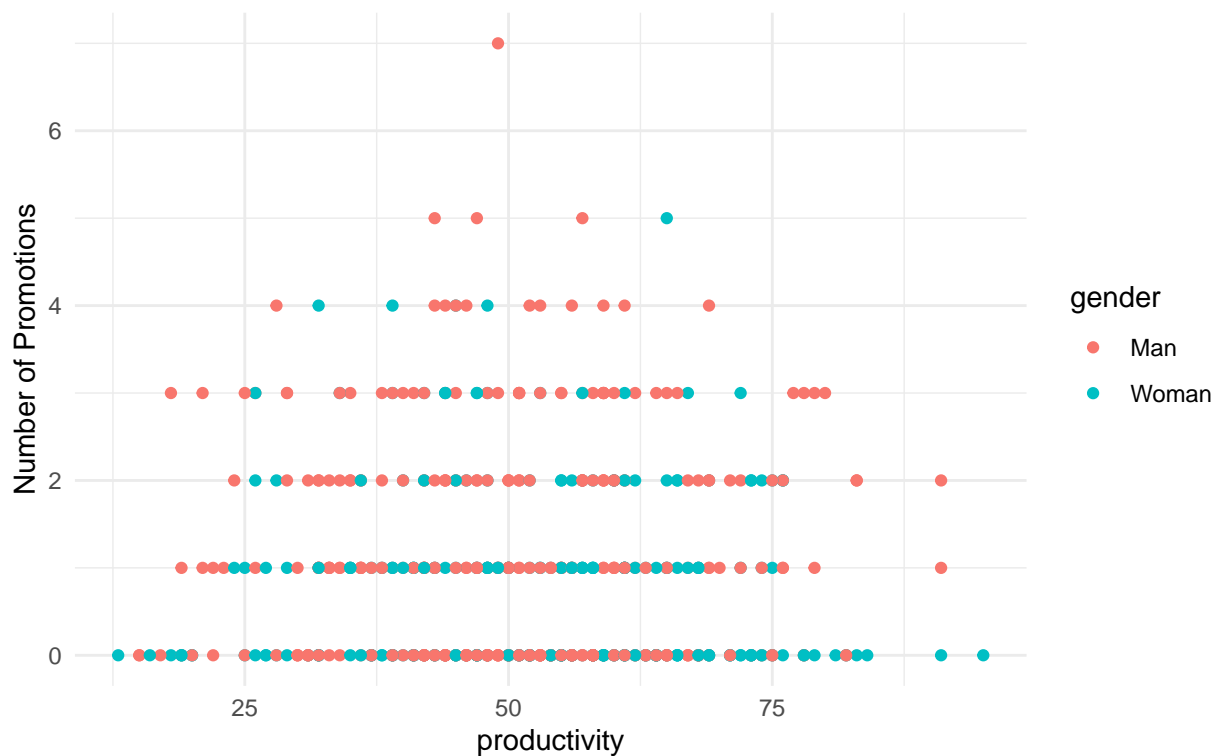


Figure 5. Bar Plot of Number of Promotions and Productivity  
 Create by: One Company  
 Date of April 21 2021

According to Figure 5, which show the number of promotion and productivity of employees of two gender groups. Based on the same level productivity, the number of promotion of male employees is much more than female employees. For example, there are three employees who have productivity of 50 and the female employee has no promotion but two male employees have one and two promotions respectively. Based on the whole figure, there are 5 employees with more than 5 of promotions and there is only one female employees. In addition, the female employee has much more productivity than the male employees.

As conclusion, the number of promotion of male employees is more than female employee under the same productivity.

**Table 7:** Estimated Coefficients and P-value of the Parameters

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.9709954	0.1907735	-5.0897821	0.0000004
genderWoman	-0.3505085	0.0936657	-3.7421214	0.0001825
teamData	-0.0074650	0.1617377	-0.0461553	0.9631865
teamDesign	0.2010656	0.2323158	0.8654841	0.3867731
teamLegal and financial	-0.0717840	0.2272199	-0.3159230	0.7520610
teamMarketing and sales	-0.0555966	0.1509257	-0.3683709	0.7125967
teamOperations	0.0201585	0.1535893	0.1312491	0.8955783
teamPeople and talent	0.3208290	0.1934024	1.6588683	0.0971423
teamSoftware	0.1113340	0.1388682	0.8017246	0.4227123
productivity	-0.0010858	0.0029906	-0.3630654	0.7165561

Table 7 shows the estimated coefficients and p-value of predictor variable of the Poisson regression model. There is only gender of predictor variable has p-value less than 0.05 which means the gender has statistical significant to the number of promotion as response variable.

In addition, the estimated coefficient of gender is -0.3505085 which is negative. The estimated coefficient demonstrate the promotion is more likely to occur in gender with male.

The Poisson regression model demonstrate that gender has significant influence to the number of promotion and promotion would occur more likely for male employees.

**To study whether there is a gender preference on the hiring process.****Method**

In order to study whether gender is a relevant factor in the hiring process, we need to use four datasets. The datasets include the ability and gender of the interviewees for the different phases. As well, we could match the same applicant id to figure out which interviewee pass the phases.

Firstly, we would to explore if there is a gender bias for those candidates who can successfully complete the first phase. The first step that we do is to delete those candidates who prefer not to say their gender in the dataset of phase 1 because we would like to reduce the complexity of the data and do not want to get the confused result. As well, we add a dummy variable to indicate if the interviewees passed in the phase one of hiring. Since the outcome variable success is a dummy variable, we would fit a logistic model. The variable gender as a predictor and other variables, the team applied for, cover letter, cv, gpa, extracurriculars, and work experiences are predicted variables. In addition, we also consider variables the team applied for might be the potential grouping unit. Therefore, we also fit a linear mixed model by adding a random intercept on the team applied for.

Secondly, we want to test whether the process from phase 2 to phase 3 has and biases about gender. The method is very similar. We remove those candidates who do not want to say their gender in the dataset of phase 2. Then add a dummy variable to the dataset to indicate whether this candidate successfully entered phase 3 from phase 2. As well, we figure out the gender of the interviewees by matching the id with previous dataset. Then we run a logistic model between binary outcome variables of the intervieww passes the phase 2 of hiring and 11 predictors. Similarly, we add random intercepts for the variable team applied for conduct an additional generalized linear mixed model to verify the results of logistic model. The process from phase 1 to phase 2 and phase 2 to phase 3 are conducted by AI. Therefore, those model results would indicate whether the AI has a gender preference during the hiring process.

Finally, we would like to test whether the process from phase 3 to receive the final offer. We create a dummy variable in the phase 3 dataset to indicate whether this candidate is successfully hired in the end. Then we fit a logit model with binary outcome variable of interviewees receive the final offer and the predictors gender and the performance in the phase 1 and 2 to test whether the coefficient of gender is significant or not.



## Results

**Table 8:** Estimated Coefficient and P-value of Logistic Regression Model for Phase 1

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-158.880784	5679.4915164	-0.0279745	0.9776825
team_applied_forSoftware	-1.052751	0.9301888	-1.1317603	0.2577352
cover_letter	61.758224	2870.4599606	0.0215151	0.9828348
cv	50.370871	4900.6475272	0.0102784	0.9917992
gpa	12.769453	3.4498046	3.7015005	0.0002143
genderWoman	1.090100	1.0361240	1.0520939	0.2927565
extracurriculars	9.937250	2.4756258	4.0140353	0.0000597
work_experience	11.848218	2.9814066	3.9740362	0.0000707

According to the table of estimated coefficient of the logistic regression model, that shows the coefficient of gender is greater than 0. The coefficient indicates that the success in phase 1 is more likely to occur in female interviewees.

**Table 9:** Likelihood Ratio Test Result of Models of Phase 1

#Df	LogLik	Df	Chisq	Pr(>Chisq)
7	-17.89903	NA	NA	NA
8	-17.44473	1	0.9085865	0.3404897

According to the table of likelihood ratio test result of phase 1 models, we could see the p-value is greater than 0.05 which indicate the simpler model could explain the data better. Thus, the likelihood ratio test of generalized linear models demonstrate the model without gender as predict variable is better. Therefore, the gender as predict variable does not have significant impact to the success for the phase 1 of hiring and the result of phase 1 depends on the ability of the interviewees. That is similar with the result of logistic model above.

The result of both logistic model and linear mixed model shows that there is no gender inequality for the phase 1 of hiring by AI service in the company.

**Table 10:** Estimated Coefficients and P-value of Logistical Model for Phase 2

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-23.3942426	4.8342365	-4.8392839	0.0000013
team_applied_forSoftware	1.3868312	0.7643156	1.8144746	0.0696047
gpa	-0.7501840	0.7902670	-0.9492792	0.3424786
genderWoman	-0.7566572	0.8079341	-0.9365333	0.3489987
extracurriculars	-0.4157566	0.7497938	-0.5544946	0.5792404
work_experience	0.1541267	0.7928249	0.1944019	0.8458612
technical_skills	0.1056958	0.0264879	3.9903483	0.0000660
leadership_presence	1.0540283	0.2370872	4.4457406	0.0000088
speaking_skills	0.9053194	0.2187396	4.1387987	0.0000349
writing_skills	0.1140481	0.0293363	3.8876071	0.0001012

According to the table of estimated coefficient of logistic regression model of phase 2, we could see the estimated coefficient is smaller than 0. The coefficient indicates the success in phase 2 of interview is less likely occur for the female interviewees. However, the p-value of the gender is greater than 0.05 which mean gender does not have significant impact to the result of hiring phase 2. Therefore, we could not make conclusion so far and we supposed to fit more models.

**Table 11:** Likelihood Ratio Test Result of Models for Phase 2

#Df	LogLik	Df	Chisq	Pr(>Chisq)
9	-33.80566	NA	NA	NA
10	-33.50101	1	0.6093117	0.4350471

According to the table of likelihood ratio test results of the models for phase 2, the p-value is greater than 0.05 that indicate the simpler model could explain the data better. Thus the likelihood ratio test of generalized linear model demonstrate the model without gender as predict variable is more proper. Therefore, the gender of employees does not have significant impact to success for phase 2 of hiring and the result of phase 2 depends on the ability of the interviewees instead of gender. That has the opposite direct with the results of logistic model above.

**Table 12:** Estimated Coefficient and P-value of the Models for Phase 3

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2557.10263	1488199.967	-0.0017183	0.9986290
genderWoman	-48.29596	254384.746	-0.0001899	0.9998485
interviewer_rating_1	15.06384	9039.407	0.0016665	0.9986704
interviewer_rating_2	18.93855	11079.392	0.0017093	0.9986361

According to the table of estimated coefficients of logistic regression model for final phase of hiring process, we could see the coefficient of gender is less than 0. The coefficient indicates the success in entire interview is less likely occur for the female interviewees. Therefore, the successes in the phase 3 of hiring might have relationship with the gender of interviewees. However, the p-value of gender is greater than 0.05 that means the gender does not significant to the final result of hiring process. Therefore, we could not make conclusion about the gender inequality without fitting other models

**Table 13:** Likelihood Ratio Test Result of Model for Phase 3

#Df	LogLik	Df	Chisq	Pr(>Chisq)
4	0	NA	NA	NA
3	0	-1	0	0.9999591

According to the table of likelihood ratio test of models for phase 3, the p-value is greater than 0.05 which indicate the simpler model could explain the data better. The simpler model does not include gender as predict variable and the gender does not have significant relationship with the success in phase 3. Therefore, the success in phase 3 of hiring depends on the performance in the first two phases instead of the genders. That has the opposite direct with the result of logistic model above.

## Discussion

Overall, we have figured out the solutions of the research questions above by performing data visualization and fitting proper models. The gender bias for female employees in the company existed for the subjective decisions such as promotion and salary of the employees. On the other hand, the gender inequality does not exist for the phases of hiring that provided by AI service. The AI service most likely to pass the interviewees depends on the ability instead of the gender.

The figures and the linear mixed model demonstrate the male employees have higher level of salary than the female employees under the same ability. As well, the likelihood ratio test shows the gender of employees has significant relationship with the level of salary. That means the salary most likely depends on the gender of the employee instead of ability. Even if female employee has same ability with the male employee, they could not get the same and fair level of salary as the male employees.

For the second research question, the figures and the Poisson regression model demonstrate that the gender parity is not satisfied for the promotion in the company. As the figures show the number of promotion for male employees are obvious more than female employees when the employees under the same conditions. As well, the gender is the only statistical significant predict variable to the number of promotion and the promotion is more likely to occur for the male employees. Overall, the promotion chance would tend to provide for male employees in the company.

For the last research question, the model results have the same direct for the phases 1. Both of logistic model and mixed linear model shows the gender does not have significant influence to the result of hiring phase 1. As well, there are more female interviewee pass the phase 1 of hiring process. Therefore, the success of phase 1 would most likely depends on the ability of the interviewee, such as the gpa or work experience. On the other hand, the model results show the opposite direct for the phase 2 and final phase. The logistic models show that the success for phase 2 and final phase tend to male interviewees but the linear mixed models demonstrate the gender does not have significant relationship with the hiring results. Therefore, the results means the success of hiring process would most likely depend on ability and there are less female interviewees pass the hiring process. Overall, the hiring process by AI service provide a fair environment to the female employees because there is only criteria for the ability instead of gender.

## Strengths and limitations

For all of three research questions, we figure out the proper model and plots depend on the type of the response variable. Such as we fit logistic model for the binary response variable and fit a Poisson regression model for the countable response variables. Meanwhile, we fitted more than one models rather than focus on only one model which would limit our conclusion as well.

As well, we perform the estimated coefficients and likelihood ratio test to interpret the statistical output into the understandable conclusion. Meanwhile, we make reasonable conclusion based on the models and data visualization rather than makeup the conclusion without statistical basis.

We have got the solution for the research questions related to gender parity issue in the company. However, there is limitation and drawback for the models and the data as well. We supposed to figure out the potential weakness and improve the accuracy and reliability for the further study.

First of all, we only use the current data, which is 2020 quarter 4, for analyzing the salary problem in order to reduce the complexity of the model. However, the historical data is critical to analysis because the data with more observations would increase the accuracy of the models. Therefore, we supposed to use the historical data to perform the result if we have more time for the further study.

Meanwhile, we tried different predict variables to fit the model but there are more combination of the predict variable to select. Therefore, we supposed to fit more models with different predict variables in order to select the more proper model for the further study.

The third limitation is that when we do Poisson regression and logit regression, we do not verify the assumptions for Poisson and logit regression. For Poisson regression, we require that the outcome variable follows a Poisson distribution. Therefore, we supposed to count the number of promotion but we are not sure about the zero of promotion represent the employee never have promotion chances for all work years or the employees do not have promotion chance in the year of data collection. That would lead to zero-inflated and reduce the reliability of the model result. For the further study, we would like to fit zero-inflated Poisson model to increase the accuracy and reliability of the results.

Similarly, when we use the logit model to study the correlation of gender on the chance of successful hiring, we also need the outcome variable which is the dummy variable of success or not roughly follows binomial distributions. It is difficult for us to test those assumptions. However, if there is indeed a violation of assumptions, the reliability of our results would be undermined. Thus, we supposed to consider more about the response variable and the assumption of the model when we select the models for the further study. If there is potential violation of model assumptions, we supposed to select the other models except the potential violation.

## Consultant information

### Consultant profiles

**Yuxin Zhong.** Yuxin is a senior consultant with One Company. She specializes in data visualization and writing a statistical report. Yuxin earned her Bachelor of Science, major in Statistics and Mathematics, from the University of Toronto in 2021.

**Wanyi Wang.** Wanyi is a junior consultant with One Company. She specializes in reproducible analysis and writing a statistical report. Wanyi earned her Bachelor of Science, Majoring in Mathematics and Statistics from the University of Toronto in 2021.

**Hairuo Wang.** Hairuo is a junior consultant with One Company. She specializes in statistical communication and reproducible analysis. Hairuo earned her Bachelor of Science, Majoring in Mathematics and Statistics from the University of Toronto in 2022.

**Xiaotian Zhang.** Xiaotian is a junior consultant for One Company. He specialized in study modeling and statistical communication. Xiaotian earned his Bachelor of Science, Majoring in Mathematics and Statistics from the University of Toronto in 2021.

### Code of ethical conduct

1. We care any reasonable requirements from our clients and will not share their data to anyone else other than our technical team.
2. We would fully support fellow statisticians in their own fields and offer helps to new entrants.
3. We would shoulder the responsibility for our own work and give professional and independent reviews and assessments.