A Report for Gender Fairness in Black Saber Software

Invistigation of gender's influence on hire, salary and promotion status

Report prepared for Black Saber Software by Beaver Data LLC

2021-04-21

Contents

Executive summary

This report analyses the employees' salary, promotion, and hiring status with respect to gender and other factors in Black Saber Software company. The purpose of this report is to find if there is a hidden gender bias in salary, promotion, and remuneration processes.

Therefore, the report consists of three main sections for three questions investigating the relationships between employees' genders and their salaries, promotion, and hiring status. We conclude that gender affects salary and promotion. There is a possibility that women working in the company will be biased. Meanwhile, Artificial Intelligence's performance in the company's recruitment process shows no significant gender bias. This report used the data from Black Saber Software company. The data contains the gender of the current employees and the essential information (salary, performance, distribution, etc.) in the company for each financial quarter since joining the company. A list of key findings and limitations is summarized below.

Key findings

- By Figure 1, the difference in salaries between males and females significantly decreased as the employees' senioritis increased. For those employees who just entered the company, female employees earn less than male employees.

- Females generally have lower salaries and promotions than males, and role seniority shows a positive impact on wages, while productivity has no effects on salaries and promotions.

- The salary of the female is $2802 lower than male for employees with the same role seniority and leadership level for all financial quarters, whereas the wages of the female is $2165 lower than male in the recent financial quarter (quarter 4, 2020). The difference in salary for all financial quarters decreases by $637 compared with the wages of the current quarter, meaning that there is a trend of improving gender bias.

- Female employees have only 72.3% chances of being promoted compared with males.

- The hiring process of phase 1 and 2 are conducted by A.I. The results conclude that GPA, working experience, and extracurriculars have significant positive effects on the candidates' hiring results, while factors like gender have barely contributed to the impact of whether candidates could pass phase 1.

- The result of phase 2 also indicates significant positive effects on technical, writing, speaking skills, and leadership presence.

Limitations

- The results of this report may be slightly inaccurate because only 10 of the company's 607 employees did not indicate their gender. We deleted their data to be more gender-sensitive. Since their number is not very large, we think the impact is not large.

- When analyzing the relationship of gender and promotion, we set criteria for determining whether an employee is new by observing and analyzing the data. This criterion may be subjective, so the results obtained in the second objective may slightly deviate from the actual promotion status.

Figure 1. Box-whisker plot of salaries and role seniorities with respect to gender

Technical report

Introduction

This report was written at the request of Black Saber Software Company to find out whether the company has gender unfairness in salary, promotion, and recruitment. It summarizes all the primary statistical modeling, analyzes results, and gives independent assessment using the data provided by the company on the investigation of gender fairness issues. The report separately analyzes the relationship of employees' gender with their hiring, salary, and promotion status. In the company's hiring process, three phases will be executed to evaluate if the candidates meet the requirement. Artificial Intelligence does the first two phases, and only the third phase will be an in-person interview. By analyzing the data, this report hopes to get a result of whether the company's workplace has gender bias that affects employees' salaries, promotions, various business ability assessments, and recruitment pass rates.

- Do male employees earn more salaries than females?
- Is gender a significant factor that will influence the employees' promotion in the company?
- Do Artificial Intelligences have a bias in gender when they are rating candidates' basic information? What about the interviewers?

Research question 1: gender effects on salary

Data description and wrangling

Data was prepared by Daubry Byagogo, Black Saber Software, containing five datasets: the current employee's dataset and four hiring datasets. The current employee dataset was used for investigating the first and second research questions, and hiring datasets were used for the third research question. The current employee dataset provided data of all current employees, including their unique ID, gender, the team they work for, position, leadership level, productivity score, and salary in a given financial quarter[1].

To explore whether gender affects employees' salary, we consider salary as a response variable, and other variables are factors. Thus, salary is converted to numerical for linear modeling, and employees with unknown gender are removed for ease of analysis.

Methods

---

[1]More details in data dictionary

Data visualization   This report first visualizes the data by creating plots, which show the relationship between salaries and teams, seniority, leadership levels, productivity scores, respectively for different genders. Figure 2 is a histogram of the general distribution of wages concerning genders. We can see that the mode of salaries is 40,000 and males seem to have higher wages at 40,000. In order to explore the difference in pay between the two genders, a series of box-whisker plots are introduced to more directly compare the differences concerning different factors. Figures 1,3,4,5 give the comparison of salaries with various factors, including their teams, role senioritis, leadership level, and productivity scores. Figure 3 shows that salaries among different teams are different. Male workers have higher salaries in some teams, and female workers have higher wages in some other teams. In Figure 1, the wage difference in gender is more evident for those workers in low-level seniority. Male employees tend to earn more than females, and the difference in salaries decreases as the workers enter into high-level seniority. Figure 4 is a box-whisker plot demonstrating the relationship between salaries and leadership level concerning gender. We can observe that the male has higher salaries when those workers are working at the appropriate level. It is worth noticing that only male workers exceed expectations; only female workers need improvement, which is necessary to be considered when analyzing gender bias for salaries. Finally, Figure 5 is a scatterplot of wages and productivity concerning gender, which shows no significant relationship between productivity scores and salaries for genders.
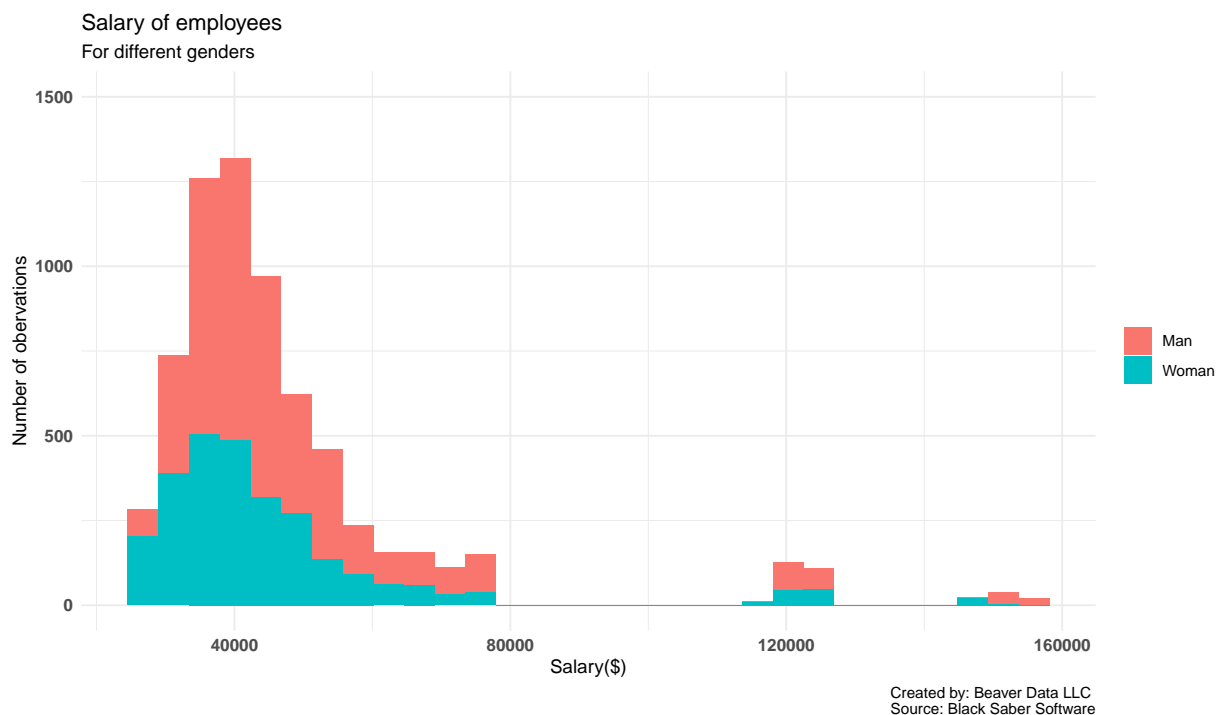
Figure 2. Histogram of employees' salaries with respect to gender

The relationship between salary and team
For different genders
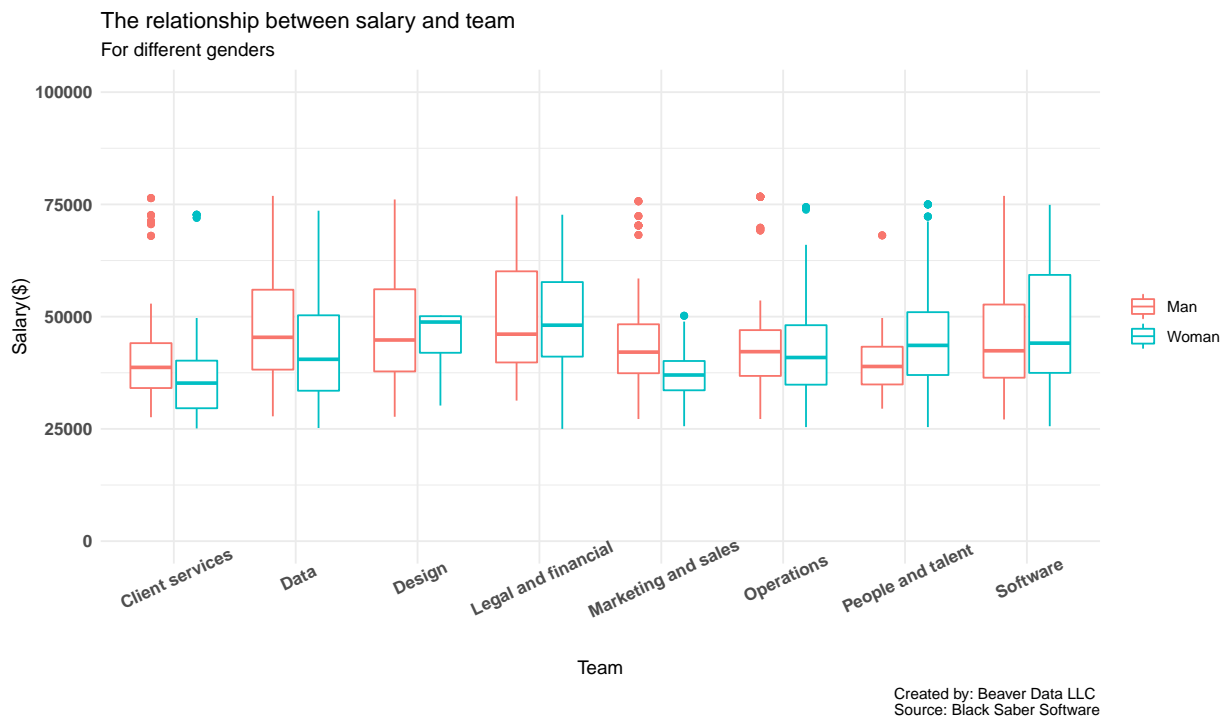
Figure 3. Box-whisker plot of salaries and teams with respect to gender

The relationship between salary and leadership level
For different gender
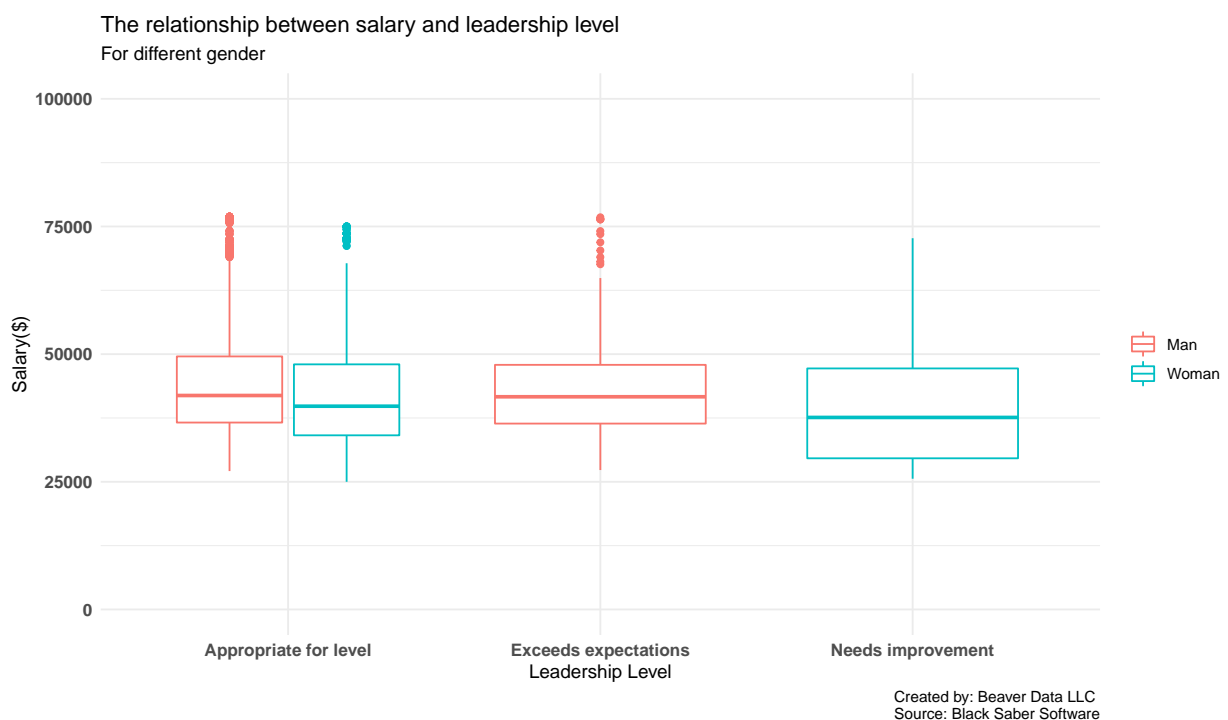
Figure 4. Box-whisker plot of salaries and performance levels with respect to gender

The relationship between salary and productivity score
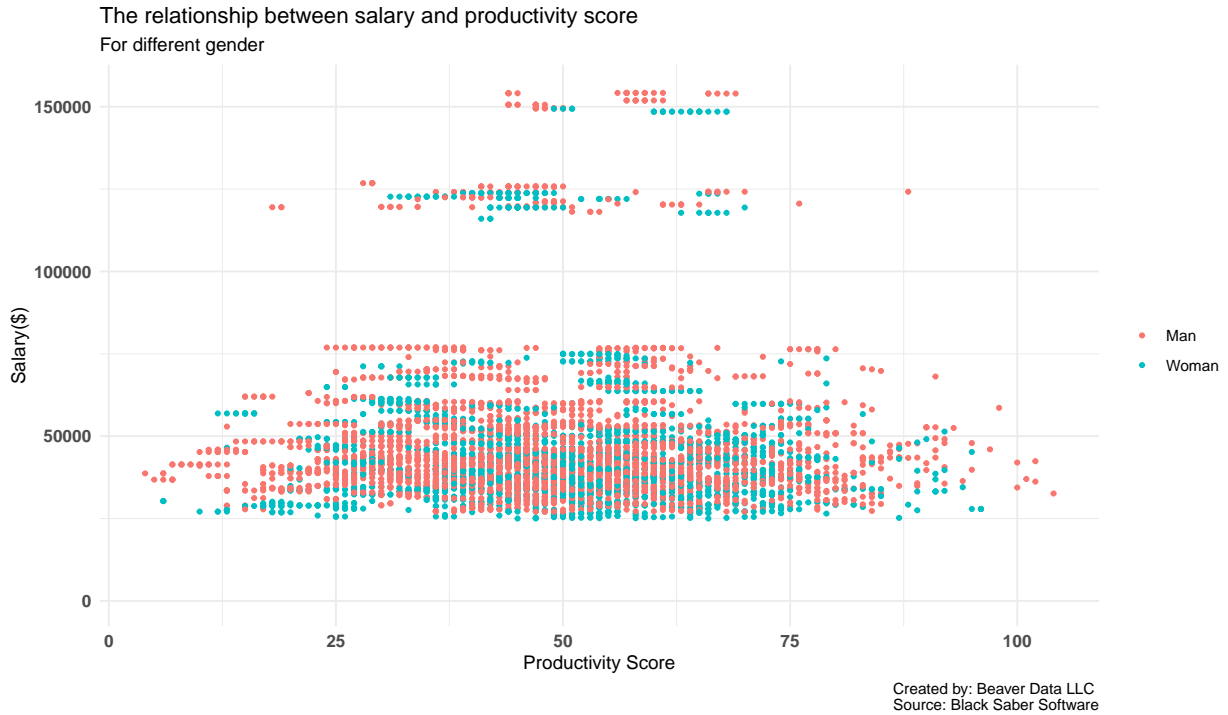For different gender

Figure 5. Scatterplot of salaries and productivities with respect to gender

Modelling   From Figure 3, we know the salary of different teams is different, which is common in companies. Thus, a team can be seen as a grouping factor and used as a random effect in modeling. We first build a simpler linear mixed model(LMM) only using data of employees of quarter 4, 2020 with random intercept team and fixed effects gender and seniority to investigate whether the salary is gender-biased in the recent quarter. Then we add fixed effects: leadership level and productivity into LMM and conduct a Likelihood Ratio Test (LRT) to test which model has better performance. The p-value of LRT is close to 0 and significantly less than the significance level of 5%, which means the complicated model performs better. The coefficients of fixed effects for the complex model are shown in Table 1. From the summary, we can see that the coefficient of *Female* is approximately -2165, which means the salary of the female is $2165 lower than male when all other factors are the same. We then construct a 95% confidence interval (CI) to test whether the fixed effects are significant, and the results showed in Table 1. We observe that the 95% CI for *Female* is approximately (-2762, -1575), meaning that we are 95% sure that the coefficient of *Female* is between -2762 and -1575 by the model, which does not include 0, so that gender level of the female is statistically significant. Seniority is also significant. However, the 95% CIs for leadership level exceed the expectation. Confidence Intervals for the factor productivity also contain 0, which means they are also not statistically significant for predicting salary. Alternatively, we build a model only without gender as a fixed effect, and

compare it with our model by LRT, and we have a p-value close to 0 and significantly less than the significance level of 5%, which means the model with gender as the fixed effect has better performance, so that gender is a non-significant factor affecting salary.

Table 1. Results and CIs of LMM predicting salary in financial quarter 4, 2020

Fixed effects	Coefficients	2.5% Quantile	97.5% Quantile
Intercept	122572	120480	124658
Female	-2165	-2762	-1575
SeniorityEntry level	-89962	-91561	-88361
SeniorityJunior I	-85122	-86748	-83496
SeniorityJunior II	-82354	-84062	-80646
SeniorityManager	-49249	-51231	-47281
SenioritySenior I	-76903	-78605	-75199
SenioritySenior II	-71592	-73377	-69812
SenioritySenior III	-65734	-67483	-63986
SeniorityVice president	29918	27138	32688
Leadership level Exceeds expectations	173	-2008	2367
Leadership level Needs improvement	-1356	-3438	737
Productivity	-8	-27	12

We also explore the salary of employees in all financial quarters, taking time into account. Comparing with the first model, one potential problem is that some employees have more than one observation, so each observation is not independent. Thus, we use employee ID as another random effect. Like the previous steps, we first build an LMM using the whole dataset of employees with random effects team and employee ID, fixed effects gender, and seniority to investigate whether the salary is gender-biased in all financial quarters. We then add leadership level and productivity as fixed effects and conduct an LRT, which also turns out that a complicated model has better performance, and the results for the complicated model are shown in Table 2. The coefficient of *Female* is approximately -2802, meaning the salary of the female is \$2802 lower than male when all other factors are the same. The 95% CIs shows that *Female*, seniority, and leadership level of needs improvement are significant, whereas other factors are not significant. Similarly, we also build an alternative model only without gender as fixed effects, comparing it with the full model by LRT, and we say gender is a significant factor negatively affecting salary.

Additionally, we notice that the estimated difference of salary for the recent quarter decreases by \$637 compared with the salary of all financial quarters, meaning that there is a trend of improving gender bias.

Table 2. Results and CIs of LMM predicting salary for all obervations

Fixed effects	Coefficients	2.5% Quantile	97.5% Quantile
Intercept	122800	121623	123952
Female	-2255	-2802	-1716
SeniorityEntry level	-90834	-91126	-90541
SeniorityJunior I	-85360	-85640	-85080
SeniorityJunior II	-82985	-83255	-82715
SeniorityManager	-50252	-50485	-50019
SenioritySenior I	-77251	-77514	-76989
SenioritySenior II	-71865	-72120	-71609
SenioritySenior III	-66378	-66627	-66128
SeniorityVice president	29987	29615	30359
Leadership level Exceeds expectations	2	-137	140
Leadership levelNeeds improvement	247	59	435
Productivity	-1	-4	1

Research question 2: gender effects on promotion

Data description and wrangling

For exploring whether gender affects employees' promotion, the number of promotions is considered a response variable. The current employee's dataset does not provide data for the number of promotions for each employee. Still, it has variable *Seniority*, which indicates the position of employees in each financial quarter. Thus, we summarise the number of promotions by counting the number of distinct senioritis minus by one and compute the average productivity score and the total number of the financial quarter the employee has already worked. Additionally, we create a variable indicating whether the employee works for less than five financial quarters, for further modeling, which will be explained in the next section. As a result, each employee is an observation in the dataset containing his/her gender, team, productivity score, number of

promotions, number of the financial quarter, and whether they work for less than five quarters.

Methods

Data visualization   We firstly create plots, which show the relationship between employees'
promotion status and their teams, role seniority, leadership level, productivity score, respectively
for male and female. Figure 6 shows that the probability of females being promoted is generally
lower than males. Among those who have not been promoted, females make up the majority. For
staff who have been promoted more frequently in the future, fewer women are promoted than
men, and the highest count of promotion for males is 7, whereas 5 for females. From the scatter
plot of Figure 7, there is no obvious linear relationship between the average productivity and
the number of promotions. However, when productivity falls within the range of approximately
(25, 75), employees are more likely to be promoted, which might also affect promotion status.
We can observe that in the majority team, the females have less promotion from Figure 8. The
scatter-plot in Figure 9 shows a linear relationship between employees' working time and the
number of promotions. Regardless of gender, an employee with a longer working time would
have more chances to be promoted.
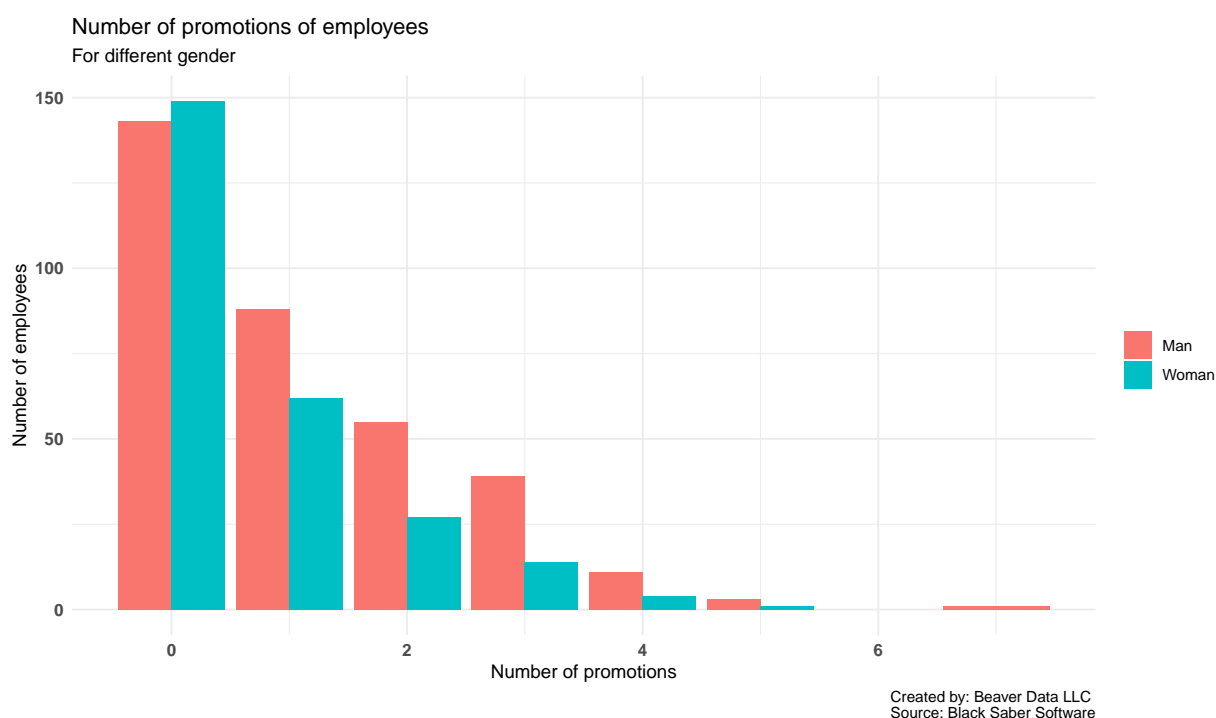
Figure 6. Histogram of employees in Black Saber Software and their promotion times

The relationship between number of promotion and productivity
For different gender
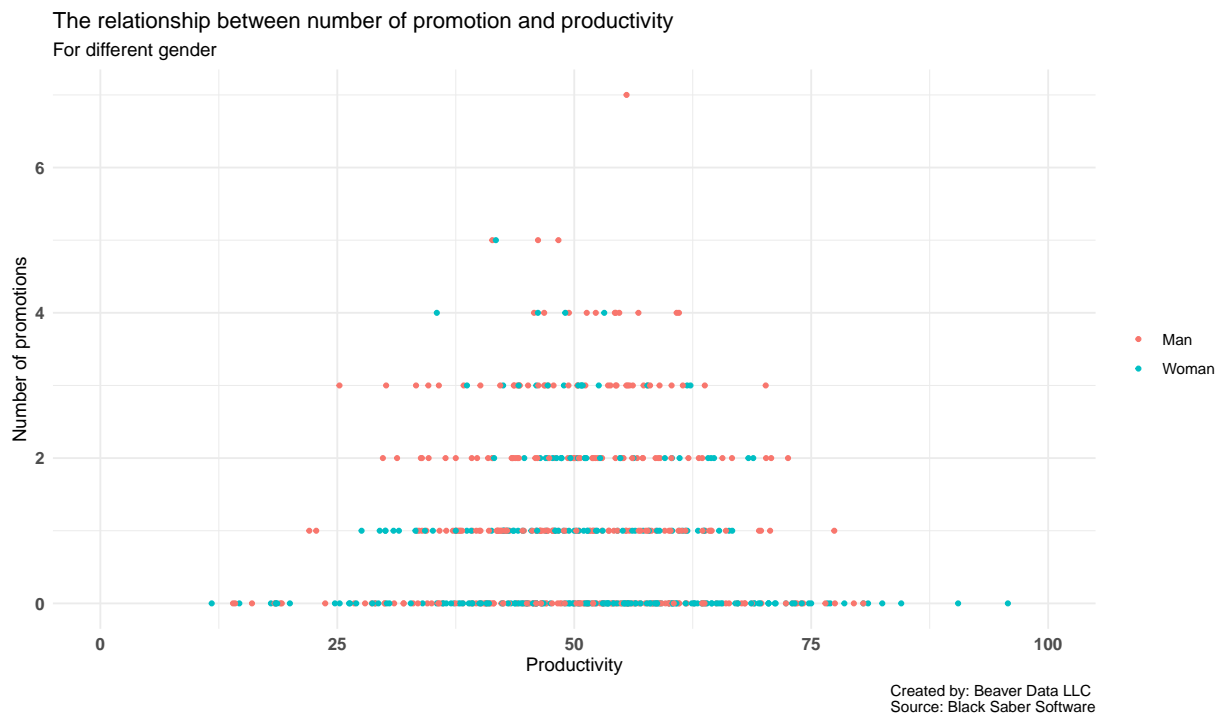
Figure 7. Scatterplot of employees' productivities with their promotion times

The relationship between number of promotion and team
For different gender
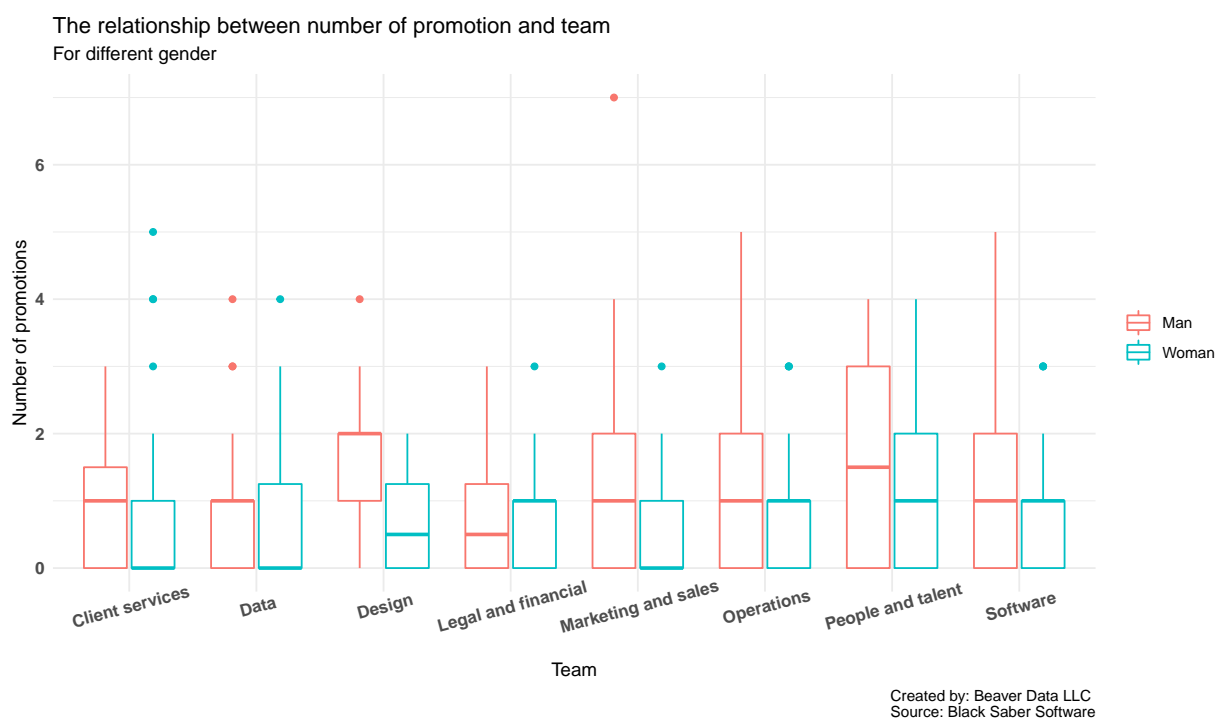
Figure 8. Box-whisker plot of employees' teams with their promotion times

The relationship between number of promotion and financial quarter
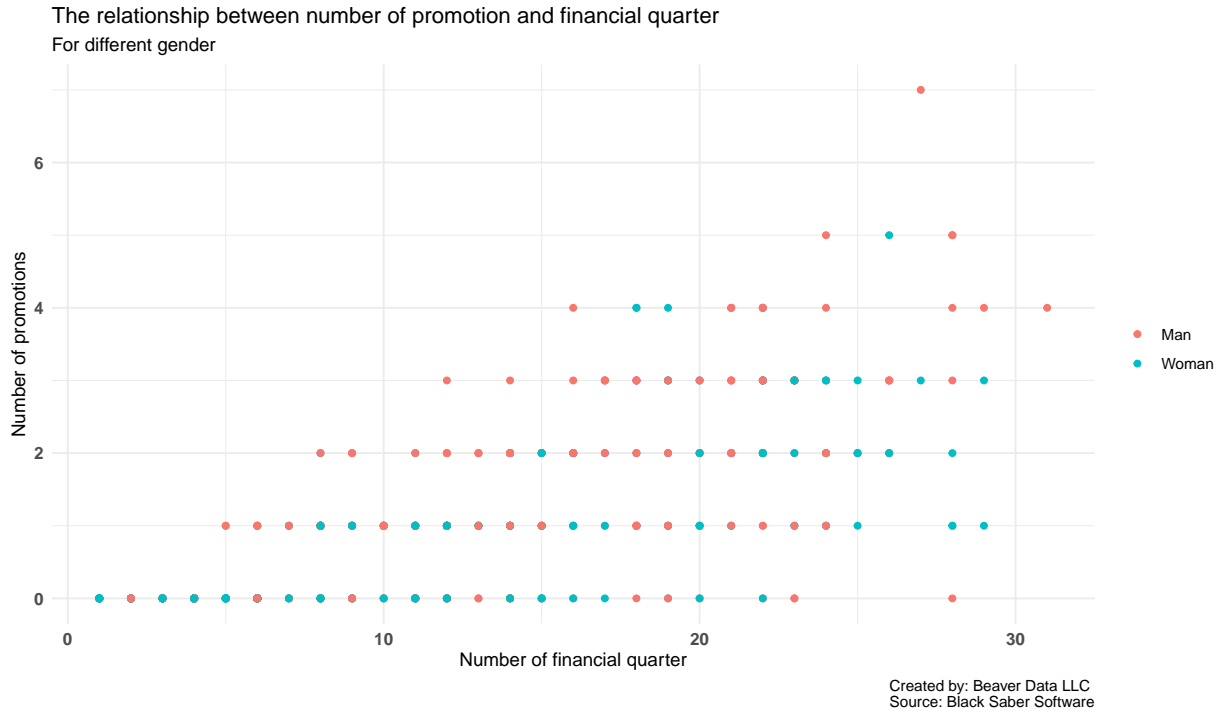For different gender

Figure 9. Scatterplot of employees' worktime with their promotion times

Modeling

Since the number of promotions is discrete, we build a Poisson regression to estimate the number of promotions based on gender, team, and productivity. Note that Figure 9 shows a relationship between work time and promotion, which is comprehensible since employees with a higher working age have a higher possibility of being promoted. Thus, this needs to be considered by introducing an offset term *log(number of the financial quarter)*, so that we can compare different employees with different working ages. The model is as follows:

$$log(Promotion) = -2.565 - 0.356 Female + 0.003 TeamData + ... + 0.115 TeamSoftware + 0.003 Productivity$$

We can see that the coefficient *Female* is approximately -0.356, which is the log relative count for female versus male, controlling for other covariates. We take $exp(-0.356) \approx 0.7$, meaning that the estimated number of promotions of the female is expected to decrease by a factor 0.7 compared with the male while holding other factors constant. We notice that the p-value of gender level of the female is less than a significance level of 5%, while p-values for factors team and productivity are greater than 5%, which means gender level of the female is statistically

significant, and other factors are not. We then conduct a Goodness of fit test[2] for this Poisson model and have a result, 1, representing that the model fits well. Therefore, gender is a significant factor negatively affecting the number of promotions predicted by a Poisson regression model with an offset term.

We also notice that there are a large number of employees who have 0 promotions for both males and females from Figure 6. It is clear that an employee who just came to the company does not have a promotion, so the number of promotions must be zero, but this not results from their gender and abilities. And there are also employees with 0 promotions but actually have a long working time. This results in our observations with 0 promotions are a mixture of new employees and old employees. Thus, we can distinguish new employees and old employees by fitting a zero-inflated Poisson model (ZIP). To fit the ZIP model, we first need to use a variable to predict whether the employee is new or old. By Figure 9, we observe that employees who work less than or equal to 4 financial quarters have 0 promotion, so we creat a binary variable indicating whether the employee works less than five quarters (TRUE if works less than five quarters), and we can use this to predict new employees. The formula of ZIP model with response variable promotion and covariates gender, team, and productivity are as follows:

$$log(Promotion) = 0.357 - 0.358 Female + 0.013 TeamData + ... + 0.155 TeamSoftware - 0.0007 Productivity$$

$$log(\frac{NewEmployee}{1 - NewEmployee}) = -3.169 + 19.352 WorkUnder5$$

The ZIP model consists of two parts; the first part is the same as an ordinary Poisson regression model, where the coefficient of *Female* is -0.358, the relative log count for female versus male, controlling for other covariates. The estimated number of promotions of the female is expected to decrease by a factor of $exp(-0.358) \approx 0.699$ compared with the male while holding other factors constant. Same as the first Poisson model, the gender level of *Female* is statistically significant (p-value = 0.000188), and factors of team and productivity are not statistically significant. The second part is a logistic regression model, which is used to predict new employees by a predictor of whether the employee works less than five quarters. *NewEmployee* is the probability of being a new employee, *WorkUnder5* is an indicator (=1 if works less than five quarters). The model gives us the odds ($\frac{NewEmployee}{1-NewEmployee}$) is $exp(-3.169) = 0.042$, which means the odds of an employee with the number of financial quarter less than 5 is 0.042 times the odds that an employee has working quarter greater or equal to 5 is a new employee. By calculation, the estimated probability of being a new employee is 1 for employees with a financial quarter less than 5 and 0.04 for employees with a financial quarter greater than 4.

---

[2]A statistical hypothesis test to see how well sample data fit a distribution from a population with a normal distribution.

We finally explore the ZIP model with an offset term *log(number of the financial quarter)*, so that we both consider the effects of the time difference and a mixture of new employees and old employees with 0 promotions. We have the model below.

$$log(Promotion) = -2.529 - 0.325 Female - 0.073 TeamData + ... + 0.054 eamSoftware + 0.004 Productivity$$

$$log(\frac{NewEmployee}{1 - NewEmployee}) = -10.22 + 19.37 WorkUnder5$$

The estimated number of promotions of the female is expected to decrease by a factor of $exp(-0.325) \approx 0.723$ compared with the male while holding other factors constant. The gender level of *Female* is also statistically significant in this model(p-value = 0.0005), and factors of team and productivity are not statistically significant. The odds of an employee with the number of financial quarters less than 5 is $exp(-10.22) \approx 0.00004$ times the odds that an employee has a working quarter greater or equal to 5 is a new employee. By calculation, the estimated probability of being a new employee is 1 for employees with a financial quarter less than five and almost 0 for employees with a financial quarter greater than 4. Both ZIP model and ZIP model with offset term have a probability of 1 for predicting new employee, which is comprehensible because we can observe that all employees with a work time of less than or equal to 4 financial quarters have 0 promotion by Figure 9, and the threshold can be changed to other count numbers, for example, 5, so that we will not have a probability of 1, or we can choose 1 since a general employee who only works for one quarter will not be promoted. Finally, we conduct three Vuong Tests[3] to compare the performance of 3 models, and we have the results that the Poisson model with offset term has better performance than the ZIP model, and the ZIP model with offset term has better performance than the ordinary Poisson model with offset. Therefore, we can conclude that gender is a significant factor negatively affecting the number of promotions.

Research question 3: gender effects on hiring rate

Data description and wrangling

Black Saber has been trialing a new AI recruitment pipeline manager for the Data and Software team. There are three phases and one dataset for each stage. To explore whether gender impacts the hiring rate conducted by three phases' interview, the candidate's score rating for the candidate is considered the response variable. For different phases, we take different factors into account as each phases' interview has different focuses. Then, our goal is to find out whether gender has a significant effect on the rating score for the candidate.

---

[3]A statistic method for comparing a zero-inflated model and ordinary regression model.

Method

Data Visualization In order to obtain a better understanding of our data, we visualize the relation between rating score and candidate's GPA, CV, and work experience for the phase one interview. Figure 10 shows that the mean passing rate for the phase 1 interview is positively proportional to the candidate's GPA. Meanwhile, the passing rate is similar between a male and female candidate for the same GPA. As GPA increases toward 4.0, the passing rate is approximately 1.0. Figure 11 shows the relation between extracurricular and mean passing rate for phase 1, and we can see that as extracurriculars increase from 1 to 2, the passing rate increases correspondingly. Meanwhile, the male and females with the same extracurricular have a similar passing rate for phase 1. Figure 12 shows the relationship between a candidate's speaking skills and his/her average passing rate colored by gender in phase 2. From the plot, we can see that as the candidate's speaking skills increase, the average passing rate decreases. Meanwhile, males and females with similar speaking skills will have approximately the same passing rate. Figure 13 shows that a higher leadership skill leads to a higher average passing rate, and there is no significant difference between the gender.

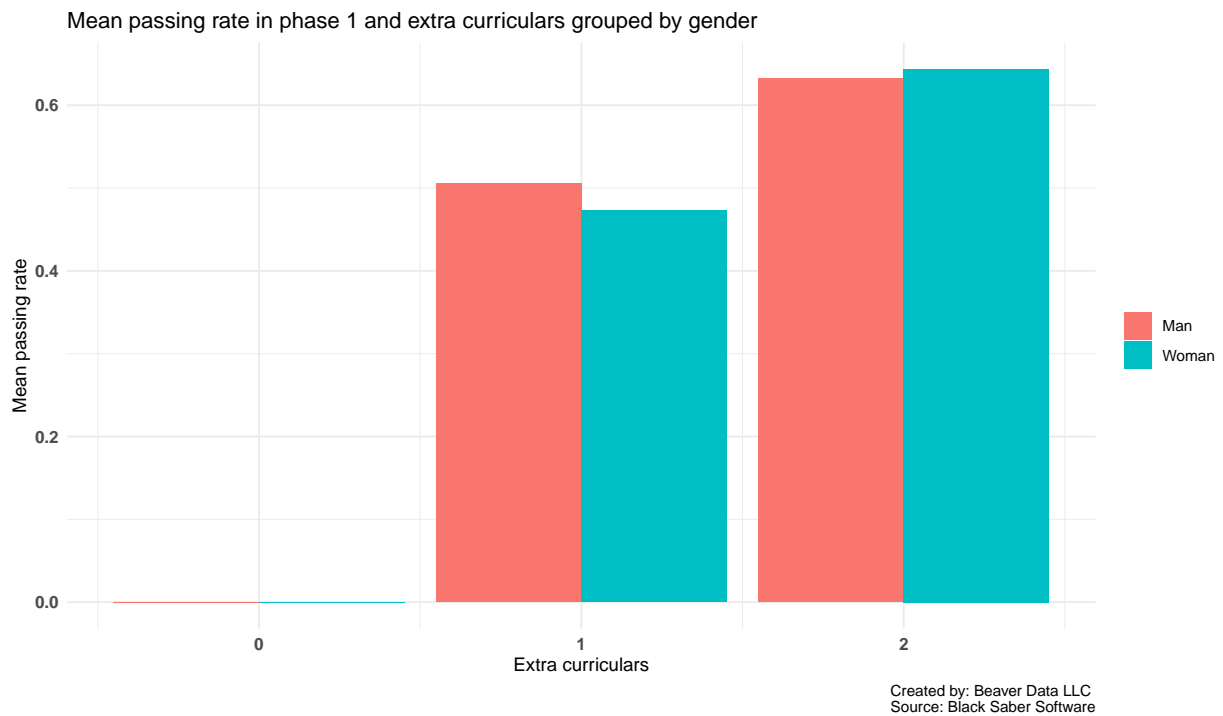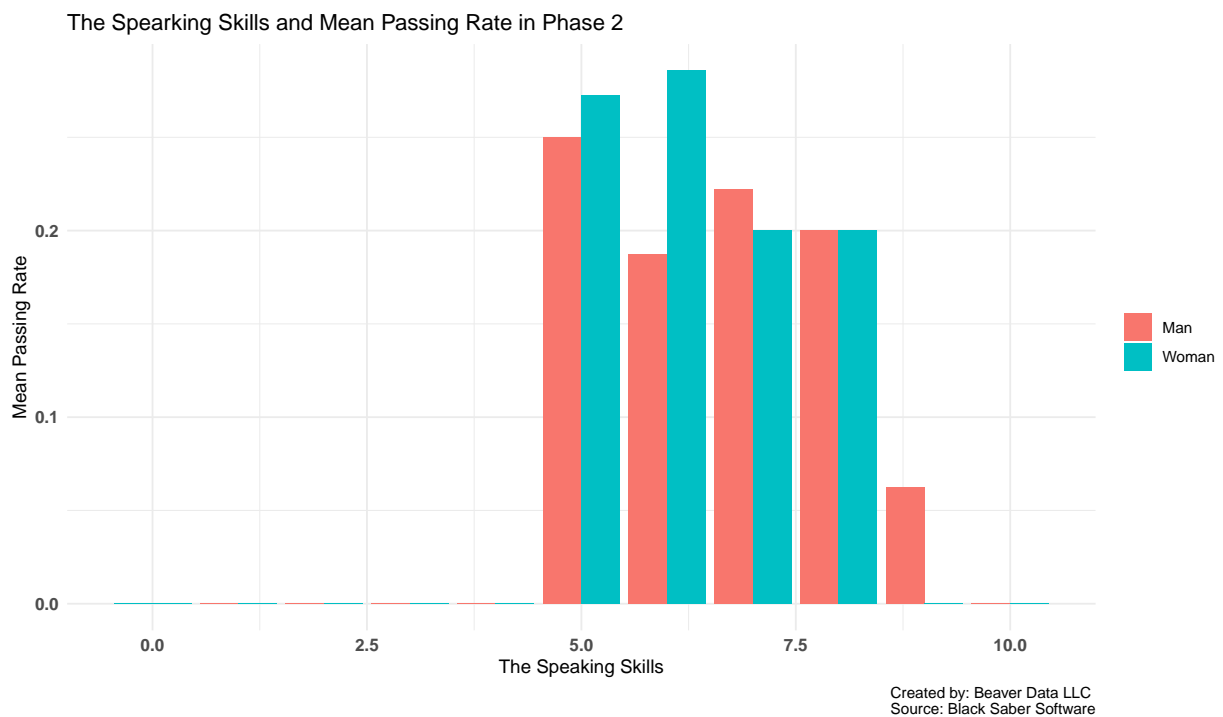Figure 10. Histogram of candidates' mean passing rate with GPA for phase 1

Figure 11. Histogram of candidates' mean passing rate with extra curriculars for phase 1

Figure 12. Histogram of candidates' mean passing rate and the speaking skills for phase 2
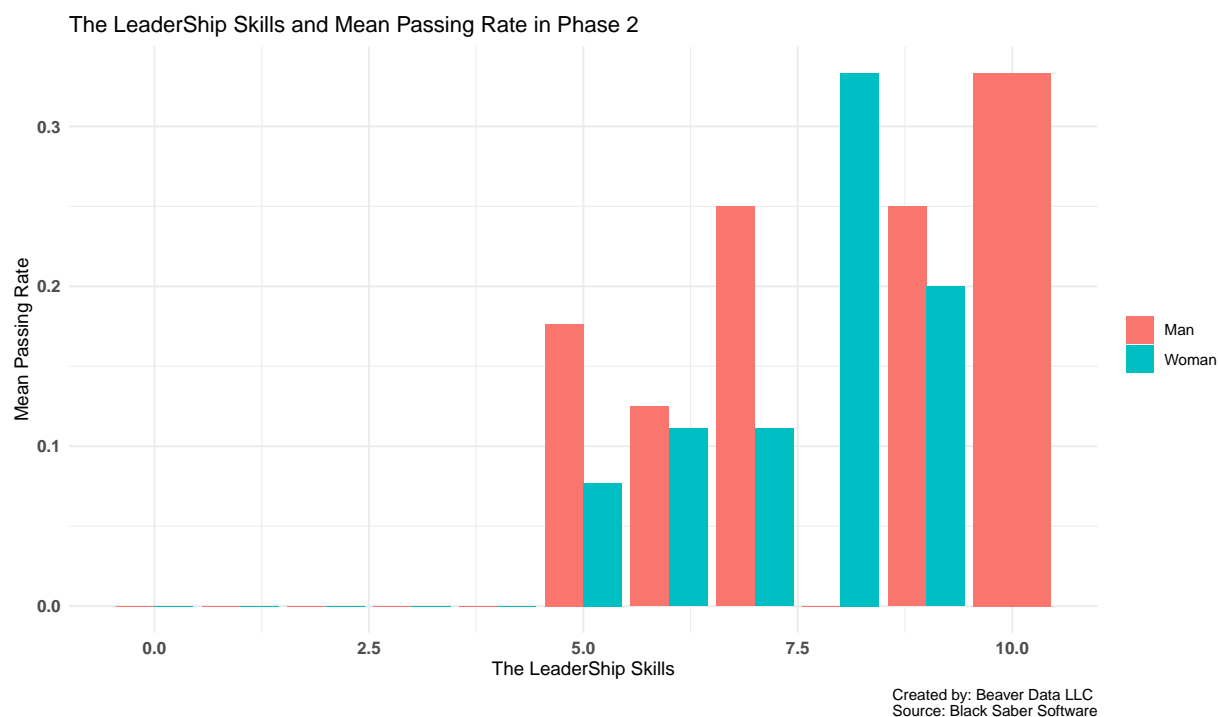
The LeaderShip Skills and Mean Passing Rate in Phase 2

Figure 13. Histogram of candidates' mean passing rate and the leadership skills for phase 2

Modeling   Table 3. P-value for using GLM fitting phase 1's data with pass rate as the response variable

	Coefficients	p-value
Intercept	-158.8808	0.977683
cover_letter	61.7582	0.982835
cv	50.3709	0.991799
gpa	12.7695	0.000214
genderWoman	1.0901	0.292756
extracurriculars	9.9372	5.97e-05
work_experience	11.8482	7.07e-05
team_applied_forSoftware	-1.0528	0.257735

Table 4. P-value for using GLM fit phase 2's data with pass rate as the response variable

	Coefficients	p-value
Intercept	-20.77341	3.06e-08
technical_skills	0.08106	6.76e-05
writing_skills	0.09222	0.00011
speaking_skills	0.71556	2.01e-05
leadership_presence	0.89593	1.17e-05
genderWoman	-0.56658	0.43291

Table 5. P-value for using GLM fit phase 3's data with pass rate as response variable

	Coefficients	p-value
Intercept	-2167.54	0.999
interviewer_rating_1	14.06	0.999
interviewer_rating_2	14.14	0.999
genderWoman	-19.62	1.000
team_applied_forSoftware	37.14	1.000

In hiring phase 1, AI will rate based on the factors including cover letter, CV, gender, extracurriculars, GPA, work experiences, and team applied for. To find out whether gender influences hiring, a binomial model is fitted as the response variable, and the pass rate is binary data. After fitting the model, from model summary Table 3, it is obvious that the p-value for extracurricular, GPA, and work experiences are significantly less than 0.05, but the p-value for gender is higher than 0.05. Thus, we can conclude that extracurricular, GPAs, and work experiences influence the passing rate of a candidate indicated by their significant p-value. Meanwhile, gender has minimal influence on whether the candidate can pass in phase 1 or not as its p-value is not significant

In hiring phase 2, the factors AI has taken into consideration are technical skills, writing skills, speaking skills, leadership presence, and gender. A binomial model is also fitted to find out if gender influences hiring phase 2 because the response variable of our model is binary. From the summary Table 4, we can see that the p-value for technical skills, writing skills, speaking skills, leadership presence are significantly less than 0.05, but the p-value for gender is higher than 0.05.

Thus, this shows technical skills, writing skills, speaking skills and leadership influencing the passing rate, while gender does not.

AI handles the hiring process in phase 1 and phase 2. From the above observations, we can conclude that gender has little effect on hiring results because of its p-value in phase 1 and phase 2. To find out whether the same conclusion could be drawn from face-to-face interview phases, we fit a binomial model in phase 3 as the response variable of our model is binary. In this model, the variables are interview rate 1, interview rate 2, gender and team applied for From the summary Table 5, the p-value for interview rate 1, interview rate 2, gender and team applied for are greater than 0.05, which shows that none of these have a significant effect on the passing rate.

Overall, as gender has a non-significant p-value in phase 1, phase 2 and phase 3, we can conclude that gender does not affect hiring results in AI or real person interviewers.

Discussion

In summary, we manipulate, visualize, and analyze the data by methods of modeling.

Proposed our solution to three research questions: whether gender affects salary, promotion and hiring rate. We use the team as a random effect for salary as the different team has a different salary to fit the model.

Conclusion for research question 1   For this question, we want to find out whether gender influences the employees' salary based on their abilities. We fit LMMs and conduct LRT testing model performance. We find that the salary of the female is $2802 lower than male for employees with the same seniority and leadership level for all financial quarters, whereas the salary of the female is 2165 lower than male in the recent financial quarter. The difference in salary for all financial quarters decreased by 637 dollars compared with the salary of the recent quarter. Therefore, we can conclude that there is significant gender bias for the employees' salaries, where the female has a lower salary than the male with the same talent and value to the company. However, there is a trend of improving gender bias in salary.

Conclusion for research question 2   In this question, the relationship between gender and the number of promotions is examined. Fitting ZIP models show that female employees only have a 0.7 or 0.699 or 0.723 probability for promotion compared to male employees by three different models, where the result for the best-performance model is 0.723. By all the results, we can safely conclude that there is gender bias for promotion, and females have a lower chance to be promoted compared with male employees.

Conclusion for research question 3   Whether gender influences the hiring rate is examined by the GLM model in this research question. For three-phase interviews, the p-values of gender are all greater than the significance level of 5%, which means gender is non-significant. Thus, we can conclude that gender does not influence the interview.

Strengths and limitations

Strengths   To obtain more comprehensive data when our company initially processed the data, we chose to model the employee ID as a random effect instead of just taking their latest quarterly salary. This choice made our results more representative and reliable.

Our team is very concerned about whether the modeling can accurately match. We will try various models and factors to find the best matching model. At the same time, we will try our best to maintain a neutral attitude on all research issues to avoid bias.

In research question 3, there are correlations among variables that we considered for fitting the model. In the phase 1 data, there is a correlation among work experience, cv, and cover letters. As the candidates who have work experience are more likely to have CVs and cover letters than the candidates who first apply for the job. Then, using the GLM model, we can represent the dependent variables as a linear combination of independent variables. Thus, our model is not affected by the correlation among variables in the data set.

There is a combination of categorical and continuous variables in our dataset (i.e., cover letter and cv in phase 1 are categorical data and interview rating 1 and technical skills are continuous data). Using the GLM model to fit allows us to attenuate this situation as GLM can mathematically express relationships among variables that are combinations of continuous and categorical variables.

Limitations   Before performing data modeling and analysis, the data was processed. One of the limitations is that to fit Gideon Blake's requirements better and facilitate judgment, we removed those employees who did not indicate their gender in the data. This may result in a loss of information.

When analyzing the relationship of gender and promotion, we set a threshold (4) for determining whether an employee is new or old. This criterion may be subjective, so the results obtained in the second objective may slightly deviate from the real promotion status.

In research question 3, the generalized linear model (GLM) we used to fit the model relies on the assumption of independence of each data point. If the data point is not independent of each other, the interpretation of the weights is no longer valid. There might be dependence among

cover letters and CVs for each point as there are templates for CV and cover letters in the market. Then, we can not guarantee the independence of each data point.

Suggestion

When we first obtained the data, we directly processed the data with some impulsiveness and then established an analysis model. We face much trouble in the early stage; for example, the assumptions of some models may not be satisfied. Therefore, I sincerely suggest that all statisticians must first choose the appropriate method to visualize the data when they get the data, form the basis of analysis and fit the model.

When choosing the suitable threshold for determining whether an employee is new or old in the future study, we can choose using one as a threshold since, generally, the employee who only works for one quarter will not be promoted. Besides, by choosing different thresholds, we can compare the model performance and fit the best model to estimate the response variable.

From the figure. 4, It can be found that there are only men among those who have exceeded the expected leadership ability and only women among those who need to improve their leadership ability. This situation should be noted because it may be that the company's human rating is biased or that women cannot get more opportunities than men to show leadership skills.

Consultant information

Consultant profiles

Jiayang Wang. Jiayang is a senior consultant of Beaver Data LLC. He specializes in data manipulation, data visualization and data analysis. Jiayang earned his Bachelor of Science, Majoring in Statistics and Mathematics from the University of Toronto in 2022.

Shiru Jin. Shiru is a junior consultant of Beaver Data LLC. She specializes in data analysis and statistical communication. Shiru earned her Bachelor of Science, Majoring in Economy and Statistics from the University of Toronto in 2022.

Zhiyu Liu. Zhiyu is a senior consultant of Beaver Data LLC. He specialized in statistical interpretation and summary. Zhiyu earned his Master of Science, Majoring in Statistics from the University of Toronto in 2022.

Zhiyuan Sun. Zhiyuan is a senior consultant of Beaver Data LLC. He specialized in statistical communication and team management. Zhiyuan earned his Bachelor of Science, Specialist in Statistics from the University of Toronto in 2022.

Code of ethical conduct

A. Responsibility for data and methods: Ethical statisticians will use appropriate methods to analyze data; they will try to eliminate prejudice factors but frankly agree that the results of this data analysis may have known suspicious limitations, defects, or biases. But an ethical statistician will analyze and confirm the reliability and completeness of the data.

B. Responsibilities to customers/funders: Ethical statisticians will keep in mind the interests of customers and funders, use data carefully according to the requirements of employers, and avoid disclosure of data to other unrelated groups without the written permission of the customer or the court's instructions. Simultaneously, ethical statisticians will work hard to ensure that data, analysis, and results in related hypotheses and limitations are fully disclosed to prevent data from being misled.

C. Responsibility for the work team: Ethical statisticians will understand that scientific, statistical practice teams are often composed of different professional groups. As a result, statisticians need to maintain ethics in this environment. Statisticians should respect their team members and understand that different observers will form other decisions on the same data results based on their own unique experiences. Statisticians should conduct exchanges focusing on scientific principles under the premise of mutual respect. The best work is often the most successful.