Gender Parity in Hiring, Wages, and Promotion

A Statistical Analysis of Gender in Black Saber Software's Company Processes

Report prepared for Black Saber Software by White Saber Consulting

2021-04-21

Contents

Executive summary

To explore gender parity in Black Saber Software's processes for hiring, wages, and promotion, a statistical analysis was conducted by an independent consulting firm (White Saber Consulting). Using the given data on current employee information and hiring scores, the consultants find that the hiring process is fair, gives appropriate value to an applicant's background and prior experience, and is not discriminatory across genders. However, they also find that gender does have an effect on promotions and salary at every seniority level. This report focuses on modeling correlation between gender and hiring, promotion, and salary decisions.

The results are summarized below:

- Gender does not affect the selection from Phase 1 to Phase 2 in the hiring process.
- Better GPA, extracurricular experience, and work experience are strongly correlated with being selected for Phase 2.
- Gender does not affect the selection from Phase 2 to Phase 3 in the hiring process.
- Higher skill ratings given by the AI in Phase 2 are strongly correlated with being selected for Phase 3.
- Higher skill ratings given by the AI in Phase 2 are strongly correlated with higher interviewer ratings in Phase 3.

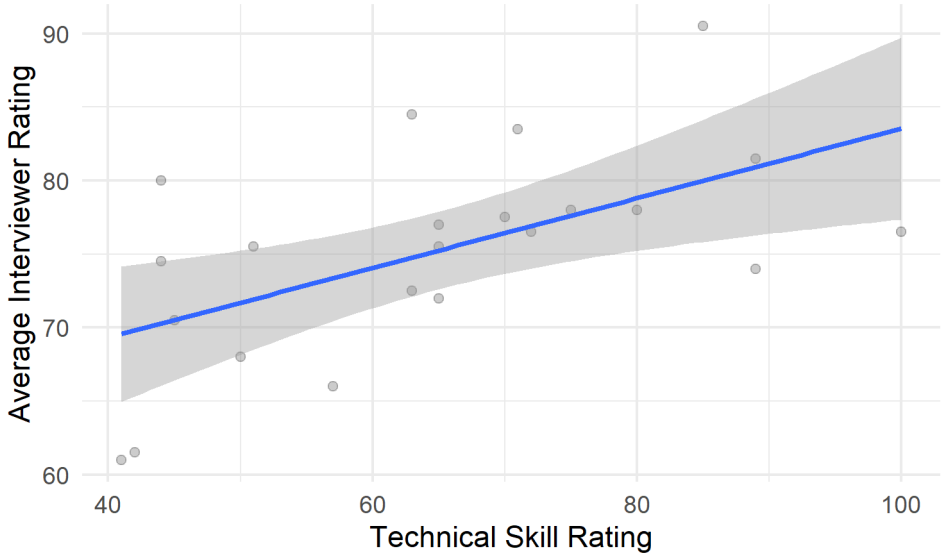

Figure 1: Comparing AI-graded technical skill score with average interviewer rating.

- On average, women receive 19.44% less promotions than men at the company, and an employee's team also impacted their promotions.

- Higher leadership ratings were found to strongly correlate with an increase in the number of promotions, while productivity was not found to have an effect on promotions.
- For employees on the same team and seniority level, women had a lower salary than men by an average of $1893 annually.

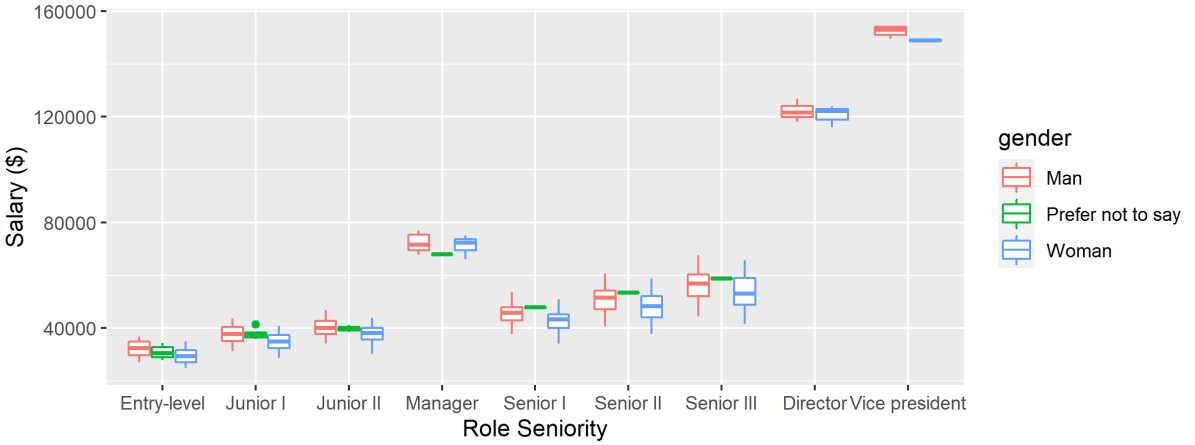

Figure 2: Salary by gender and seniority level.

- Leadership and productivity ratings do not have a significant effect on the annual salary.

While conducting the study, several limitations posed as obstacles:

- The granularity for leadership made it difficult to evaluate its impact on promotions. This can be further expanded upon by using a different measurement with more levels, or a standardized test that all employees take on a quarterly basis.
- The time-dependency of certain factors such as improved productivity, leadership, and other soft skills over time is difficult to account for given the current tools available. For example, it could have been interesting to explore how leadership and productivity change in the few quarters prior to a promotion. However, we looked at the average productivity and leadership of individuals over their whole time at the company.

Technical report

Introduction

The following report is broken into several main sections. It aims to answer three main research questions defined below that focus on Black Saber Software's hiring, promotion, and salary processes and how gender plays a role in all these decisions. A summary of the findings is included as a discussion at the end.

The raw data was manipulated using a variety of methods. For example, to explore the impact of hiring factors across different phases, numerous joins were used across the different datasets. In addition, information such as current employee's leadership levels had to be converted to ordered factor levels to allow for more robust statistical analysis. A mixture of models (GLMs, GLMMs, LMMs) was used for analysis to reach conclusions. A 0.05 significance level was used as the threshold for all p-value interpretations.

Research questions

- Do the AI service and interview select fairly based on talent and value? Is gender a factor in hiring decisions?
- Does an employee's gender and the team they belong to impact the individual's ability to receive promotions? Is it a fair process where demonstrated leadership and productivity are the primary factors, or are there other variables at play?
- Is there a correlation between an employee's gender and their annual salary, after controlling for other factors such as leadership and productivity? Is there a significant difference between the salary of men and women within the same team and role seniority?

Fairness and Equality of the Hiring Process

Phase 1 and Phase 2 Analysis

In the initial look at Phase 1 data, we can see that the distributions of GPA, cover letter, CV, extracurriculars, and work experience scores are about the same across genders. This indicates that we can assume the abilities of applicants across each gender to be equivalent, and that there is not a pre-existing difference that can affect hiring decisions.

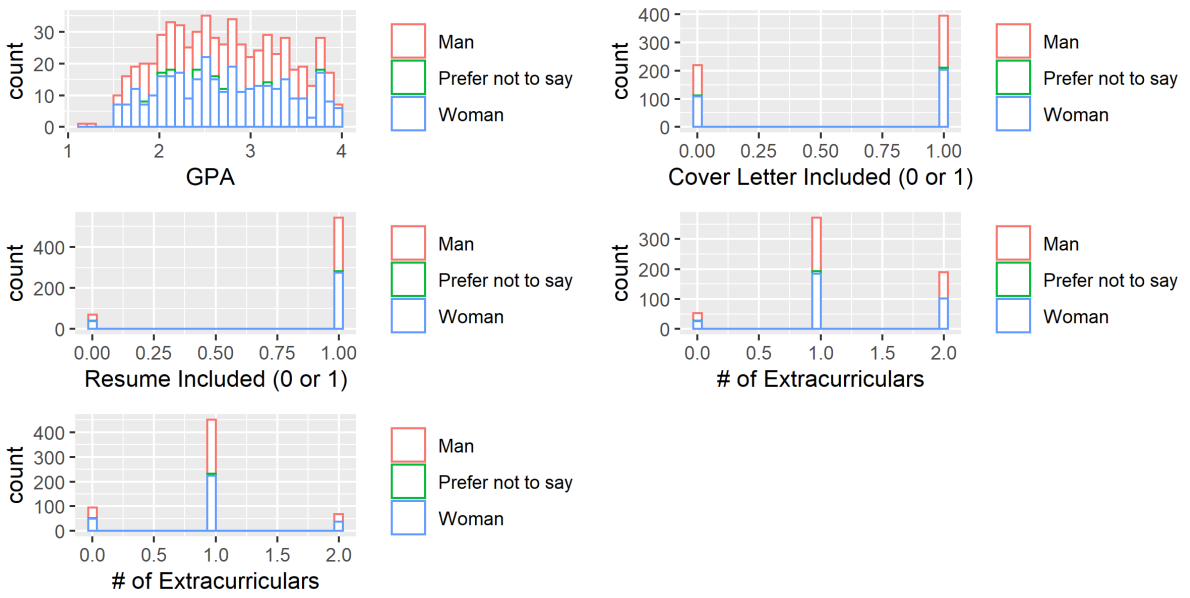

Figure 3: Histograms showing equal proportion of all predictors in Phase 1 across gender.

In order to check if gender has a correlation with applicants being able to move to Phase 2, we fit a logistic model where being in Phase 2 is the binary dependent variable, and gender is the independent variable. In the summary of the model, the fitted coefficients for either gender was not statistically significant (i.e, p-values of 0.82 and 0.16). This indicates that we have not found evidence for gender affecting Phase 1 to Phase 2 decisions.

Next, we investigate the effect of the other factors on the process from Phase 1 to Phase 2 by fitting another logistic model. We can see that GPA, extracurriculars, and work experience are positive factors and statistically significant with p-values very close to 0. This indicates that the process fairly measures talent. No evidence for a cover letter or CV affecting the decision from Phase 1 to Phase 2 was found. This also makes sense because the contents in the cover letter and CV can vary, and just the existence of the item should not be the sole factor in hiring.

Phase 2 and Phase 3 Analysis

Next, we would like to check if gender is a factor on the decision from Phase 2 to Phase 3. After fitting a logistic model where the binary dependent variable is being in Phase 3 and the independent variable is gender. For applicants who did not disclose their gender, the result is not statistically significant (a p-value of 0.992). However, for women, the result is close to statistically significant with a p-value of 0.066, which means that it would be worthwhile to run another iteration of this study in the future.

Comparing Interviewer and AI ratings

Finally, we will observe whether or not the ratings of the interviewers and the AI-graded scores are different from each other. As there are not many applicants left in the third phase, the data is too small; therefore, it is no longer appropriate to use GLMs. Therefore, we will consider using standard linear regression because we did not interpret the diagnostics as violating any of the four LINE assumptions.

First, we join the Phase 2 and Phase 3 datasets together, only observing all applicants that made it to the third phase. There is a total of 22 applicants in Phase 3. Next, we create an average rating variable, which is the mean of the two interviewer ratings per applicant. This will be our dependent variable, and we will observe whether or not there is a correlation between the average interviewer rating against the scores of the timed component and pre-recorded interview judged by the hiring AI.

Our multiple linear regression model contains an applicant's technical skills, writing skills, leadership presence, speaking skills, and gender as the predictors. Below is a table with the estimated regression coefficients, as well as the p-value for the corresponding t-test for that coefficient:

Regression Coefficient	Estimated Regress. Coeff. Value	P-value for T-test
Intercept	-2.52643	0.77858
genderWoman	0.59138	0.689769
technical_skills	0.37174	2.46e-07
writing_skills	0.38024	1e-04
leadership_presence	2.46911	5.96e-05
speaking_skills	2.38571	0.000734

Observing the gender predictor, since the p-value is non-statistically significant (0.69), we can interpret this as not having enough evidence to conclude that there is a correlation with interviewer rating and gender, holding all other predictors constant.

Furthermore, the p-value for the global F-test is almost 0, so this implies that at least one of the slope parameters is non-zero. In addition to this, all four p-values of the AI-graded categories in the second phase are below the 0.05 benchmark significant level. This concludes that there are indeed some useful explanatory variables for predicting the response.

To affirm our previous conclusion, using stepwise regression with AIC, the backwards elimination method agrees with removing the gender predictor in our final model. Hence, our final model

becomes:

$$avg\_rating \sim technical\_skills + writing\_skills + leadership\_presence + speaking\_skills$$

Therefore, we conclude that there is a relationship between the four traits evaluated in the second phase and the average rating between the two interviewers. For the most part, the better an applicant performed in these categories, the more likely they were to receive a higher rating. Most importantly, there was no evidence to conclude that the two interviewers were more inclined to give higher ratings to a specific gender.

Leadership, Productivity, and Promotions

To explore the relationship between promotions, leadership, productivity, gender, and team, a "promotion" must first be defined. For this research, the number of promotions for an individual will be looked at over the number of quarters they have been with the company. Since this is a rate (# of promotions/# of quarters), a Poisson model will be used. Leadership and productivity for an employee were looked at a high level by using the mean for each of them over all the quarters they were at the company (where leadership values were converted to ordered factor levels). The resulting data would allow for analysis using generalized linear mixed-effects models.

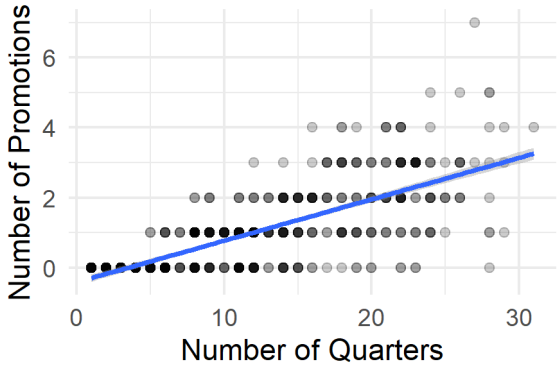

Figure 4: Comparing the number of quarters at the company against the number of promotions.

Some initial data exploration indicates that there is a positive relationship between the number of quarters with the company and the number of promotions. This indicates that treating the number of promotions and quarters with the company as a rate using a Poisson model is an appropriate approach.

To construct a model, generalized linear mixed-effects models were used to account for random effects (team) and the Poisson distribution. Several models were developed and compared using ANOVA; our final model used the following formula:

$$n\_promotions \sim gender + leadership + (1|team) + offset(log(n\_quarters))$$

Interestingly, when comparing with ANOVA, employee productivity was found to not have any significant impact on the ability to receive promotions. The aforementioned model was then used to develop confidence intervals for the different parameters:

Variable	2.5% CI	97.5% CI
.sig01	1	1.1522
(Intercept)	0.0014	0.0487
genderWoman	0.6553	0.9864
leadership	1.3748	7.5495

From these confidence intervals, we can make several conclusions. To start, the first confidence interval indicates that the variance for Team is non-zero, and thus there is evidently some correlation between team and ability to receive promotions. Second, the confidence interval for gender that we can be 95% confident that the number of promotions decreases between 1.36% and 34.47% for woman at the company (the exponentiated coefficient of the model gives an average decrease of 19.44%). The results highlight that team and gender impacts an individual's ability to receive promotions.

There are also interesting conclusions that can be made when looking at leadership. The confidence interval for leadership shows that we can be 95% confident that there is a 37.48% to 654.95% increase in number of promotions for every point of increase in leadership. As discussed earlier, productivity was found to not have a profound effect on an individual's ability to receive promotions.

Salary, Productivity, Leadership and Gender

Next we will examine the differences in salary between current employees, taking into account teams, roles, as well as their value to the company, measured by productivity and leadership level. We want to investigate if gender is a significant factor to a person's annual salary, which will determine if the employee salaries are truly fair.

An employee's salary will remain constant through the quarters given the same role seniority, thus we grouped the data based on employee ID and seniority to calculate the mean leadership, productivity, and salary for each employee and role.

Prior to fitting models, we can see in **Figure 5** below that salary, the response variable, is not a normal distribution, which is to be expected for salary data. However, by taking a closer inspection at the conditional distributions in **Figure 6** of salary based on role, we see that within each role, salary is roughly normally distributed.

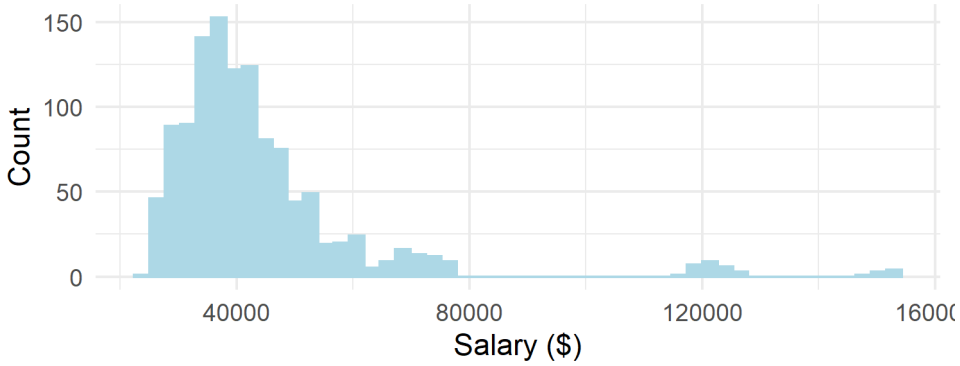

Figure 5: Counting the number of employees with a given salary.

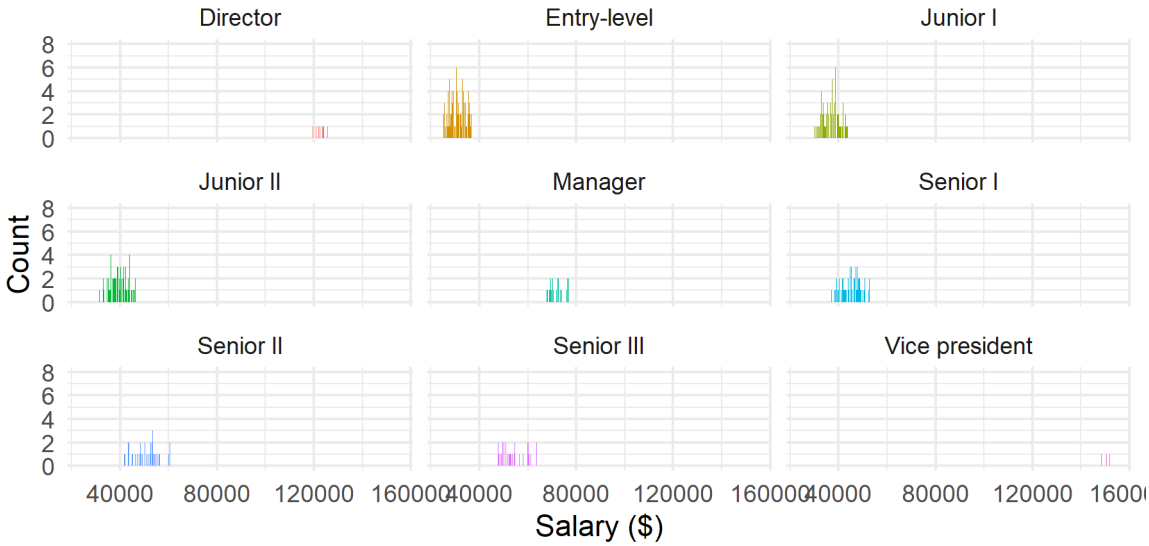

Figure 6: Counting the number of employees with a given salary, broken down by seniority level.

Recall **Figure 2**, we see that as an employee's seniority level increases, their salary increases

as well which is predictable. However, we can also note that for men and women within the same role, on average, men tend to have a higher salary. This is not accounting for other variables pertaining to talent and value to the company, so we will explore these further by fitting models.

Given the nested nature of the data for teams, as well as the repeated measures for each employee, we can fit a linear mixed model for our analysis. As mentioned, we will include role seniority in order to have a normal response for each employee, conditioning on their role.

After developing and comparing multiple models with differing levels of complexity using likelihood ratio tests, our final model used the following formula:

$$salary \sim leadership + gender + (1|team) + (1|role\_seniority)$$

We found that adding leadership and gender as fixed effects improved our model, and the random effects for team and seniority account for the random variation within each of their respective groups. Interestingly, our likelihood ratio tests found that productivity is not a significant factor in employee salary.

Variable	Estimate
leadership	744
genderPrefer not to say	-1065
genderWoman	-1893

Variable	2.5% CI	97.5% CI
genderWoman	-2335.9	-1449.5
leadership	-393.8	1881.4

From our model we conclude that looking at employees from within the same team and seniority level, women have a lower salary than men by an average of $1893 annually, while controlling for leadership level.

Looking at the 95% confidence intervals, we see that the confidence interval for gender does not include zero, meaning that it is statistically significant and female employees do experience some type of bias regarding salaries. In contrast, the confidence interval for leadership includes 0, meaning at the 95% confidence level, it is believable that employees with different leadership

levels will have the same salary. Therefore, the results from our models conclude that for employees within the same role seniority and team, gender has a statistically significant effect on salary, with women earning a lower salary than men. Interestingly, it reveals that leadership and productivity have less of an effect on salary, which could indicate that employee salary is not based on talent and contributions alone.

Discussion

Overall, the main purpose of this report was to find out whether or not there is gender parity at Black Saber Software in terms of the hiring process, promotion opportunities, and salary. Starting off with the first research question, which focused on the hiring process, we found no evidence of gender bias in the AI hiring software, as well as the final two interviewers in Phase 3. Applicants were fairly and professionally selected based on their talents and scores provided by the AI service.

However, gender equality stops there. After controlling for other factors such as leadership and productivity, we discovered that women were less likely to get promoted than men. Interestingly, an employee's team was also found to contribute to the disparity in promotions. When looking at variables that we expect to impact promotions, productivity was an insignificant factor while leadership played a heavy role. Lastly, we found that women also had lower salaries compared to men at every role seniority, after holding predictors such as productivity and leadership constant. However, both leadership and productivity were deemed as insignificant.

Strengths and limitations

We stand firm in believing that the hiring process of Black Saber Software is fair and non-discriminatory. After having analyzed dozens of models using logistic regression and looking at each phase transition individually, we conclude that the gender of an applicant has no effect on getting hired or not. However, our results point to significant differences in promotions and salaries for men and women at Black Saber Software that must be addressed to allow for a fair and equal workplace environment.

Finally, there were two limitations/imprecisions that could have affected this case study. To begin, the definition of leadership presence in the current employee's dataset is vague and makes it difficult to include as a significant predictor. There are only three defined levels of leadership and this lack of granularity may have led to inaccurate conclusions. Furthermore, when determining the impact of gender on promotion chance and salary, these predictors may be considered as time-dependent. Naturally, the longer someone works at a company, the more knowledge they will gain and opportunities they will have to be promoted and earn a higher

salary. Therefore, auto-correlation may exist and we may have needed to fit a time series model or run a longitudinal data analysis instead of Poisson regression.

Consultant information

Consultant profiles

Antoine Finot. Antoine graduated from the University of Toronto where he studied computer science and statistics and developed a particular affinity for challenging problems. He has decades of experience with Python, Java, and R and is currently a technical consultant for White Saber Consulting. Prior to his current job, he was a statistical software engineer at Statistics Canada.

David Pham. David is a junior consultant with White Saber Consulting. He specializes in probability theory and time series analysis. David earned his Bachelor of Science, Specialist in Mathematics and Majoring in Statistics from the University of Toronto in 2022.

Amy Gao. Amy is a recent graduate from the University of Toronto where she first discovered the unique applications of statistics in computer science. She specialized in computer science while pursuing a minor in statistics, which proved to be a fulfilling challenge. Using software tools such as R and Python, Amy has created meaningful models and analyzed data as a junior data engineer at White Saber Consulting.

Eric Pang. Eric is a University of Toronto alumni specializing in computer science and statistics. He has years of experience working as a data scientist in large corporations and has moved his sights to consulting at White Saber Consulting. Here, he uses his vast experience tackling problems from all types of industries.

Code of ethical conduct

Our company promotes ethical statistical practice and strives to demonstrate professional integrity and maintain the highest standards of conduct through our work.

- Our statisticians maintain objectivity through our methodology and analysis, while conducting our research in a way that is interpretable and reproducible.
- We take full responsibility for our work and only take work that we are sufficiently qualified in. Any doubts or uncertainties are consulted with other statisticians and peer reviewed.
- We are honest and transparent about any limitations and sources of error in our methodology, while taking steps to preserve validity of our findings.
- Adhere to and understand confidentiality agreements established by our employers, ensuring data and/or findings are not disclosed to a third party without the permission of our client.