Bias in the Workplace

Analyzing the merit and fairness of Black Saber Software's hiring and remuneration processes

Report prepared for Black Saber Software by JJDC Consulting Co.

2021-04-21

Contents

Executive summary **3**

Technical report **5**

Introduction 5

Are Black Saber's salary processes fair and based on talent and value to the company? 6

Are Black Saber's promotion processes fair and based on talent and value to the company? 7

Are Black Saber's hiring processes fair and based on talent and value to the company? 9

Discussion 13

Consultant information **16**

Consultant profiles 16

Code of ethical conduct 16

Executive summary

Background and aim

For this analysis, JJDC Consulting Co. was hired by Black Saber Software as independent statistical consultants to examine potential bias in Black Saber's hiring and remuneration processes.This report aimed to examine the hiring, promotion, and salary processes at Black Saber to ensure they were fair, and based on talent and value to the company. The analyses performed in this report use data from current Black Saber employees and recent applicants to answer this.

Key findings

- Men at Black Saber earn $2255.46 more than women who have the same title and experience.
- In the leadership assessment system at Black Saber, all employees graded as "needs improvement" were women, while all employees graded as "exceeds expectations" were men. It was not found to predict higher salary or higher positions (Table 1).
- Productivity was not found to predict a higher salary, and was negatively associated with predicting higher positions.
- The AI used in the hiring process assesses candidates based on the skills it graded, and was found not to be biased by gender in either the grading or selection process.
- The live interview process was unbiased by gender, and candidates were hired based on interview scores (Figure 1).

There is evidence that the hiring process within Black Saber is fair and unbiased, while both the salary and promotion processes are not based on value to the company, and are influenced heavily by gender.

Limitations

- The analysis on bias was limited to gender bias. Other possible biases were not evaluated due to lack of data.
- The size of the data used for the hiring process analysis was small, which may have impacted the results.

Summary figures

Key results of the analyses are summarized in the following tables/figures.

Table 1: Leadership scores for current Black Saber employees by gender.

	Appropriate for level	Exceeds expectations	Needs improvement
Man	3796	268	0
Woman	2583	0	142

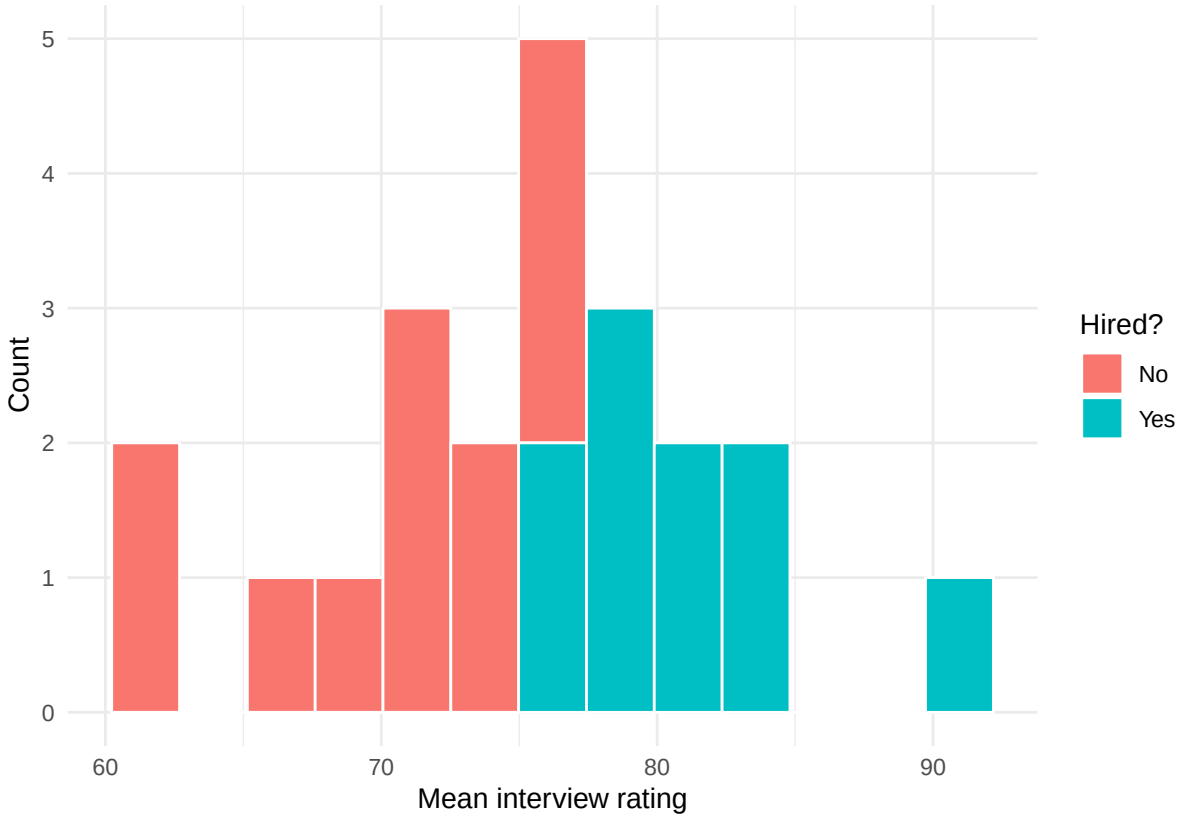

Figure 1: Histogram of candidates' mean interview rating, with color indicating whether they were hired.

Technical report

Introduction

Black Saber's hiring, salary and promotion practices are analyzed in this report to determine if the current systems in place are based on merit or if there are inherent biases. Three main research questions are answered to determine the fairness of these processes.

The first research question is, "Are Black Saber's promotion processes fair and based on talent and value to the company?" To answer this question, we design a model based on data from current employees. The model compares the impact of factors such as gender, productivity and leadership to find out if promotions are biassed or merit based.

The second research question is, "Are Black Saber's salary processes fair and based on talent and value to the company?". To answer the second research question, a second model is built using the same data on current employees of Black Saber. Salary of current employees are compared to gender, productivity and leadership to see if the current salary system is based on value to the company.

In the final research question we answer "Are Black Saber's hiring processes fair and based on talent and value to the company?" In this research question we take data from people who submitted job applications to work at black saber and analyze if there are inherent biases in the current process. The current hiring system has two AI filtering phases and then a third final phase which is done by human resources. We analyze the AI to make sure the criteria needed for promotion to the next phase is based on merit. We then check if there are any final biases from the human interview phase. In the end we make conclusions based on these research questions and answer your question of how fair the current systems are.

Research questions

We have three main research questions:

- Are Black Saber's promotion processes fair and based on talent and value to the company?
- Are Black Saber's salary processes fair and based on talent and value to the company?
- Are Black Saber's hiring processes fair and based on talent and value to the company?

Are Black Saber's salary processes fair and based on talent and value to the company?

Data

To evaluate Black Saber's promotion and salary processes, a dataset of 607 current employees provided by Black Saber was utilized. The dataset listed the gender, role or seniority title, productivity score, salary, relative leadership grading for their role, and the team each employee worked with for each financial quarter that the employee has worked for Black Saber. All data was anonymized, and each individual was identified by a five digit employee ID.

In order to best evaluate the discrepancy in gender, and to avoid the ethical implications of wrongly assuming employee gender, ten employees who listed their gender as "prefer not to say" were omitted from the dataset. As well, a new variable was created that detailed the number of quarters a given employee had worked previously for Black Saber. The role seniority variable was refactored to a numeric scale indicating rank in the company, ranging from an Entry-Level position with a value of 1, to a Vice-President position with a value of 9.

Methods

To evaluate the salary processes at Black Saber, a linear mixed model was used. The response variable was the salary at a given financial quarter and the fixed effects were the gender indicator variable, leadership for level, and productivity. The random effects variables were all crossed random intercepts and measured the variation due to employee ID, an employee's team, and their role at Black Saber. By using random effects to account for different levels of seniority, salary was observed to follow a normal distribution. Therefore it was appropriate to use a linear mixed model.

Leadership and productivity were chosen as variables because they are indications of a merit based system. The gender variable was included to assess if the current salary process was incorporating a non-merit based factor. The inclusion of employee ID as a random effect in the mixed model accounts for the correlation across repeated observations of the same employee.

Linear mixed models have multiple assumptions that have to be met to get statistically accurate estimates for the model. For ease of analysis the response, salary, was treated as a continuous variable. All fixed and random effects were appropriately chosen for the dependency structure. The observations of our data were independent. Both random effect errors and within-unit residual errors followed normal distributions. Finally, random effects errors and residual errors were assumed to have constant variance in this model.

Results

Table 2: Estimates and confidence intervals for the salary model.

	Estimate	Lower limit	Upper limit	Visualization
Is a man	2255.460	1712.383	2798.747	
Exceeds expectantions for level	1.710	-137.037	140.430	
Needs improvement for level	247.183	59.231	435.104	
Productivity score	-1.186	-3.559	1.188	

In *Table 2*, it is evident that only two predictors, gender and needing improvement for leadership, do not contain 0 in their 95% confidence intervals, and are therefore significant. The gender variable estimate shows that male employees make $2255.46 more than female employees. The leadership estimates show that exceeding expectations is not significant in improving salary, and needing improvement is significant in predicting an increase in salary of $247.18. The estimate and confidence interval for productivity shows that it is not a significant predictor of an increased salary.

Are Black Saber's promotion processes fair and based on talent and value to the company?

Data

The data used in the model to determine if Black Saber salary compensation is fair, as well as based on talent and value to the company, is the same as the promotion analysis above.

Methods

To evaluate how different variables factor into the promotion process at Black Saber, a generalized linear mixed model was used. This model assumed that the response variable, seniority, followed a Poisson distribution, as the response variable is a count. Using a log link function, the model fit gender, leadership skills, and productivity scores as fixed effects, while accounting for employees, team, salary, and quarters worked as crossed random intercept effects.

To examine if the promotion process is based on merit, gender was included as a fixed effect so that direct statements could be made about gender's significance in selecting promotion candidates. To examine value and merit, leadership and productivity were included. Employee

ID was included as a random effect to account for independence violations. Team, salary, and quarters worked were included as random effects to account for their indirect impact on promotion at Black Saber.

The model followed the appropriate assumptions, as the random effects had normal distributions, the chosen link function was the canonical link, the random effects had normally distributed errors with constant variance, and the employees were independent of each other. As well, the response followed a Poisson distribution, the observations were independent, and the log link function was appropriate. However, the mean of the response did not equal the variance of the response. Since the variance was higher than the mean, this indicates some overdispersion in the data. More on overdispersion is mentioned in the limitations section of this report.

Results

Table 3: Estimates and confidence intervals for the seniority level model.

	Estimate	P-value
Is a man	0.913	0.000
Exceeds expectantions for level	1.020	0.613
Needs improvement for level	0.966	0.501
Productivity score	0.998	0.020

In Table 3, it is evident that only two predictors, gender and productivity, have p-values less than 0.05, and are therefore significant. The gender variable estimate shows that male employees are approximately 9% less likely to be in a higher role than women. The leadership estimates are not significant in this model. The estimate for productivity shows that it is a significant negative predictor of position level. For every point increase in productivity score, the likelihood of the employee having a high role decreases by approximately 0.2%.

Are Black Saber's hiring processes fair and based on talent and value to the company?

Data

For the hiring process at Black Saber, candidates have to go through three phases before they are offered a position.

- In the first phase of the hiring pipeline, 613 candidates applied. Each candidate's extracurriculars and work experience were rated automatically based on the descriptions that the candidates provided in the application form. Additionally, the team they had applied to, as well as whether they submitted a CV and/or cover letter, was recorded.
- In the second phase of the hiring pipeline, the 300 candidates who advanced from the first phase were asked to go through a technical task, provide a writing sample, and submit a pre-recorded video. An AI then assigned a rating from 0 to 100 for each candidate for their technical, writing, and speaking skills, as well as their leadership presence.
- In the third phase of the hiring pipeline, the 22 candidates who had advanced from the second phase participated in a live interview, and received rating scores from two interviewers. Finally, ten candidates were offered a position with Black Saber based on this phase.

To evaluate the fairness of Black Saber's hiring processes, four datasets consisting of the candidates' application information and ratings in each of the 3 phases, as well as the final candidates who were hired, were utilized. All data was anonymized, and each applicant was identified by a four digit applicant ID. For each phase, we identified which candidates had advanced from the previous phase as binary variables.

Methods

Phase 1

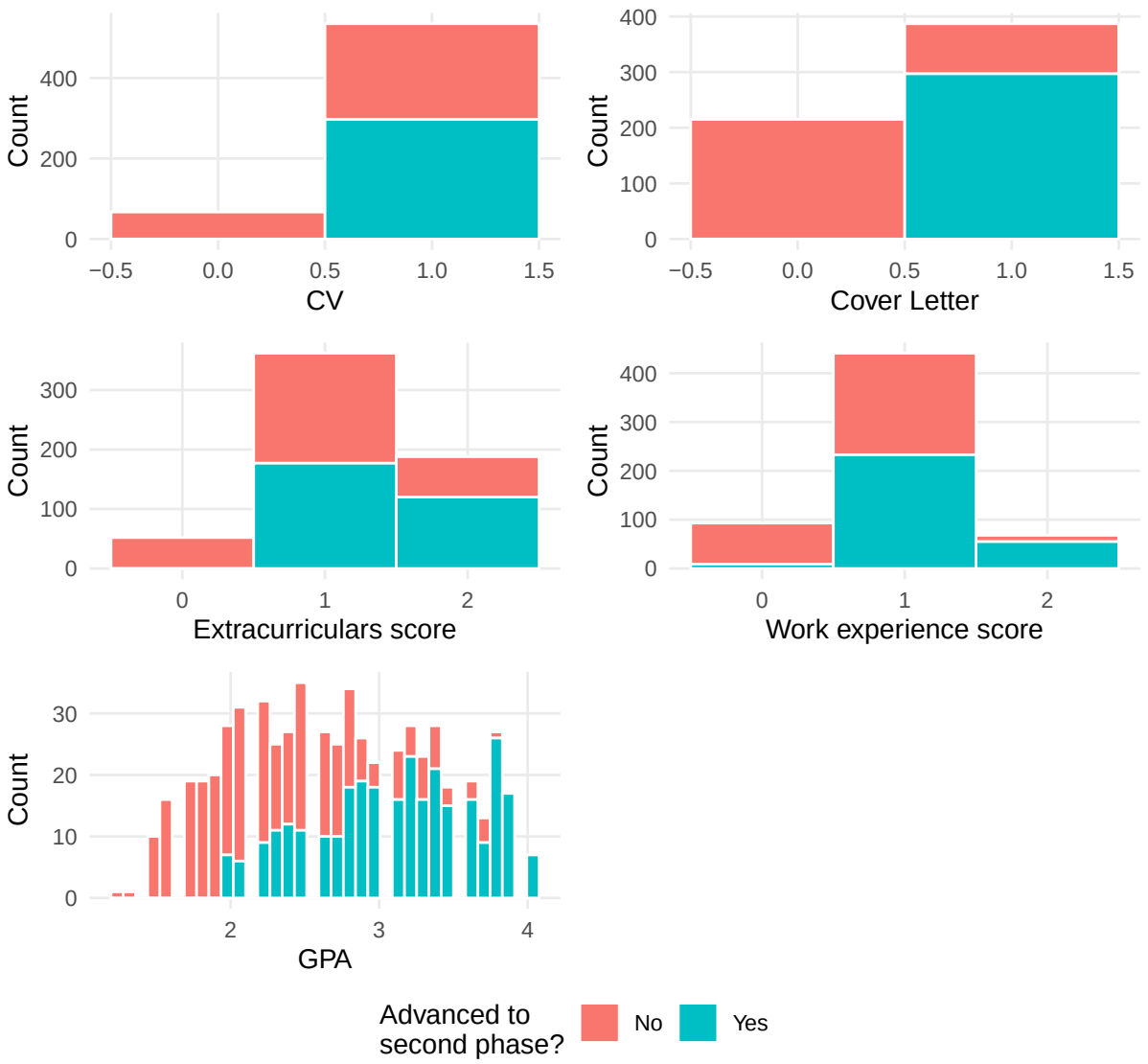

Figure 2: Histogram of candidates' GPA, with color indicating whether a candidate advanced to the second phase

We found that all candidates who did not submit both a CV and a cover letter, or who had an extracurricular score of 0, did not pass to the next phase. Additionally, all candidates with a GPA lower than 2.0 also did not advance to the next phase.

Phase 2

The model used was a generalized linear model with a binary response indicating whether a candidate advanced to the third phase. The effects considered were gender, GPA, extracurricular score, work experience score, and the four skills rated by the AI (technical, writing, speaking, leadership presence). The data are independently distributed, and the errors are assumed to be independent and normally distributed. The dependent variable, whether a candidate advanced to the next phase, follows a binomial distribution.

Table 4: Estimates and confidence intervals for the model for whether a candidate advanced to the third hiring phase.

	Estimate	Lower limit	Upper limit	P-value	Visualization
GPA	-0.772	-2.373	0.679	0.312	
Is a man	0.523	-0.962	2.057	0.489	
Extracurricular score: 2	-0.387	-1.852	0.994	0.588	
Work experience score: 1	-1.600	-7.212	3.934	0.690	
Work experience score: 2	-1.491	-7.346	4.222	0.714	
Technical skills rating	0.094	0.052	0.151	0.000	
Writing skills rating	0.103	0.057	0.161	0.000	
Speaking skills rating	0.752	0.431	1.152	0.000	
Leadership presence rating	0.972	0.594	1.461	0.000	

We found that there is no significant relationship between gender and whether a candidate advanced to the next phase, and that all four skills rated by the AI were significant in predicting whether a candidate would advance.

Additionally, we tested to see whether the AI was unbiased when assigning the different ratings to the candidates. To do this,a mean rating between all four skills was calculated for each candidate. We then used a generalized linear model with a Gaussian response for the mean ratings, with effects for gender, GPA, and extracurricular and work experience scores. This model was chosen for reasons similar to the one above; it follows the assumptions of this statistical model, so we can be confident that the model produced valid and significant results.

Table 5: Estimates and confidence intervals for the model mean skill rating.

	Estimate	Lower limit	Upper limit	P-value	Visualization
Is a man	-0.526	-2.156	1.104	0.528	
GPA	4.635	2.964	6.305	0.000	
Extracurricular score: 2	1.118	-0.624	2.860	0.209	
Work experience score: 1	1.888	-3.116	6.893	0.460	
Work experience score: 2	2.893	-2.539	8.326	0.297	

We found no significant correlation between gender and the mean rating by the AI. Also, a higher GPA was the only significant variable correlated with a better mean rating.

Phase 3 and final hires

To evaluate the validity of the interviewing process, we calculated the mean rating between both interviewers for each candidate. A generalized linear model with a Gaussian response for the mean rating was fit, with effects for gender and each of the four skills rated by the AI in the previous phase (technical, writing, speaking, leadership presence).

Table 6: Estimates and confidence intervals for the model for mean interview rating.

	Estimate	Lower limit	Upper limit	P-value	Visualization
Is a man	-0.591	-3.443	2.260	0.690	
Technical skills rating	0.372	0.286	0.457	0.000	
Writing skills rating	0.380	0.235	0.525	0.000	
Speaking skills rating	2.386	1.262	3.509	0.001	
Leadership presence rating	2.469	1.572	3.366	0.000	

We found no significant correlation between the mean rating and gender of the candidate. The four skills rated by the AI were all significant for predicting the mean rating.

Lastly, we looked at the data for the candidates who were hired, and the distribution regarding the candidate's mean interview score.

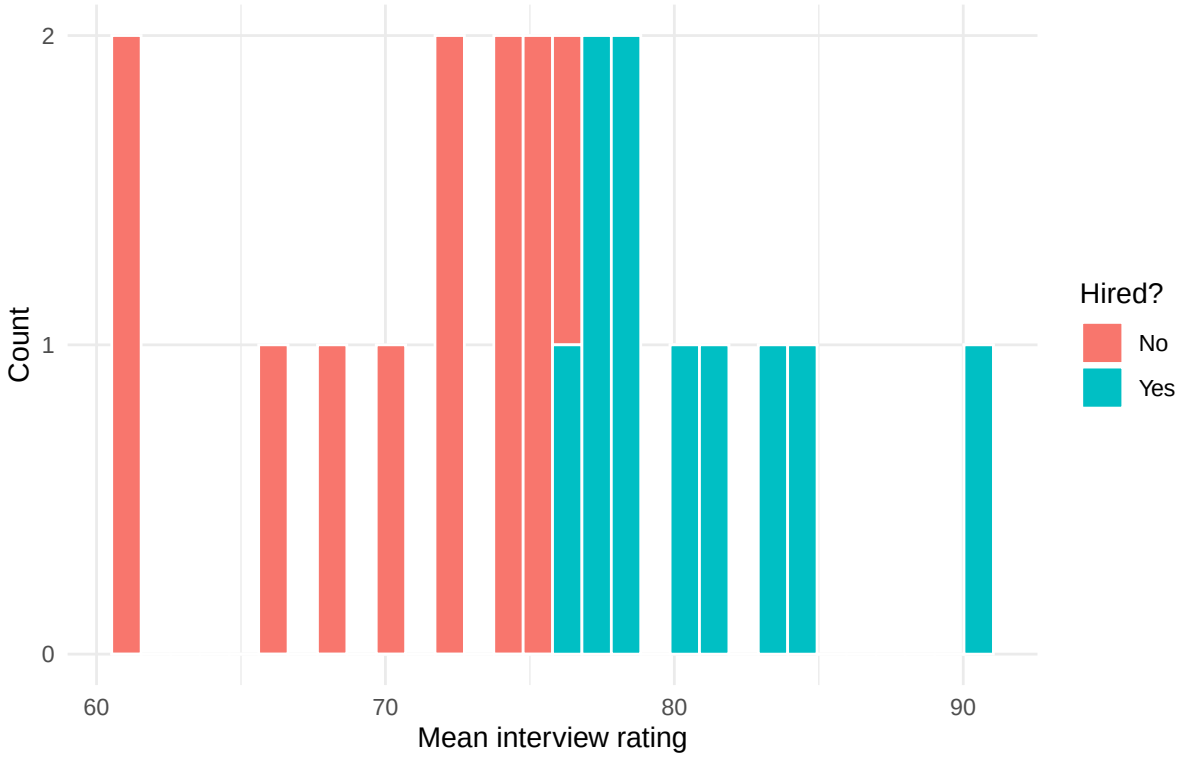

Figure 2: Histogram of candidates' mean interview rating, with color indicating whether they were hired.

Here we see that the candidates with the best mean ratings for the interview were hired. We also see that out of 22 candidates who had advanced to the third phase, 10 were hired, with 8 being men and 2 women.

Discussion

Question 1: Promotion Processes

The promotion process at Black Saber has serious flaws. By analyzing the variables that are significant in predicting a higher role at Black Saber with our model, it was clear that the system does not reward value to the company and has inherent bias. Similar to the analysis of the salary process, leadership was a flawed predictor, and not significant. The productivity score of employees was found to be a significant negative predictor that employees have a higher role in the company. For every point an employee improved in productivity score, their estimated probability of being a high role at Black Saber decreased by 0.16%. This may indicate that high ranking employees have become less productive, or that productive employees are not being rewarded with promotions to higher positions.

If the employee is a man, our model predicts that they are approximately 9% less likely to be in

a high ranking position. However, in our analysis of the salary processes, we saw that women are underpaid compared to their counterparts in the same roles. Since women tend to have higher roles at the same salary level, this may indicate that men at lower roles are paid similarly to women at higher roles, rather than indicating that women are more likely to be promoted than men.

Question 2: Salary Processes

The salary processes at Black Saber do not seem to be fair, or based on talent and value to the company. Male employees were predicted to have a statistically significant increase of $2255.46 in salary compared to being a woman, with all other factors equal. Productivity was not found to be a statistically significant predictor of a higher salary. Employees with leadership that needed improvement were predicted to have a statistically significant increase of $247.18 over workers that were labeled as having appropriate levels of leadership.

Upon further evaluation, it was found that all employees that were marked as needing improvement in leadership were women, yet all employees that were marked as exceeding expectations were men. This may indicate that the system is biased against women, especially successful women, and that the leadership scale grading is based on gender, not merit.

Question 3: Hiring Processes

From our analysis, we have reason to believe that the hiring process pipeline at Black Saber is fair and based on talent and value to the company. In the first phase of the process, gender did not play a significant role to determine who advances to the second phase. Instead, candidates advanced based on set guidelines regarding their application materials (CV, cover letter, extracurricular and work experience scores, and GPA).

From Table 4, we found that all four skills were statistically significant in predicting whether a candidate advanced to the third phase, while gender was not. GPA, extracurriculars, and work experience were also not significant, since all the candidates who advanced to this phase had to have similar scores for these variables.

In the third phase of the hiring process, we found that the four skill ratings of the candidates made by the AI in Phase 2 were significant and contributed to the determination of the interview score. Gender, once again, was shown to be not significant. It was also shown that applicants with a score greater than roughly 76.5 ended up being hired by Black Saber. From the 22 candidates who were interviewed, 8/15 men were hired, while 2/7 women were hired. However, there was not enough statistically significant evidence to support that this process was influenced by gender.

In general, we have evidence to believe that the hiring process within Black Saber is fair and unbiased. We also have evidence to believe that both the salary and promotion processes within

Black Saber are not based on value to the company, and are influenced heavily by gender.

Strengths and limitations

We believe our results to be strong and indicative of the fairness of the current promotion, salary, and hiring systems. Yet there are strengths and weaknesses that should be discussed. Our inclusion of random effects allowed us to be confident that our models produced valid and significant results by accounting for indirect variation from various factors.

A limitation of our analysis is that the provided data does not currently contain any information about ethnicity, age, disabilities, etc. There may be inherent biases toward these groups that may invalidate the integrity of the hiring, salary, and promotion processes. JJDC would be interested in collaborating with Black Saber again to perform analyses on these factors in future contracts, if the data was gathered and provided.

A second limitation of this analysis is the use of the flawed leadership level variables, as all employees that were marked as needing improvement in leadership were women and all employees that were marked as exceeding expectations were men. It is statistically improbable for these results to occur naturally. Since the leadership data is unreliable, it is difficult to determine how leadership should be affecting salary or promotions.

A third limitation of this analysis is that the promotion model has an element of overdispersion. This overdispersion is likely a result of structural restrictions in the staff of Black Saber. Executive and leadership roles are inherently limited by the nature of the job; there can only be a limited number of Directors, Managers, and Vice Presidents. Therefore, the prevalence of positions with a high value is restricted. The mean value will then be shifted towards the lower values. The standard errors are likely falsely small, which may have impacted the choice of variables in the model, or affected the level of significance of the factors. This should be considered when interpreting the results of the promotion process analysis.

A final limitation of this analysis is that the sample size for the models was limited. A small sample size may have resulted in biased data, and may not appropriately represent the true population.

Consultant information

Consultant profiles

Colin Conant. Colin is a senior consultant with JJDC Consulting Co. He specializes in data modeling and manipulation. Colin earned his Bachelor of Science in Statistics from the University of Toronto in 2022.

David Deng. David is a junior consultant with JJDC Consulting Co. He specializes in data engineering. David earned his Bachelor of Science in Statistics and Economics from the University of Toronto in 2021.

Jack Smith. Jack is a junior consultant with JJDC Consulting Co. He specializes in statistical communication and econometrics. Jack earned his Bachelor of Science in Economics and Statistics from the University of Toronto in 2022.

José Casas. José is a senior consultant with JJDC Consulting Co. He specializes in data analysis and visualization. José earned his Bachelor of Science in Statistics from the University of Toronto in 2022.

Code of ethical conduct

JJDC Consulting Co. adheres to the following Code of Ethical Conduct.

- JJDC Consulting Co. is committed to maintain the reputation of statistical practices and to strictly follow professional standards during the process of evaluation and development of analysis. JJDC Consulting Co. will not tolerate any action that can lead to a negative effect on the good standing of Statistics and Statisticians.
- JJDC Consulting Co. will maintain high standards when carrying out and document work to clients and will act diligently to the requirements set by clients. JJDC Consulting Co. will inform clients of any potential conflict with the ethical standards and ensure all information relevant to the client is fully disclosed.
- JJDC Consulting Co. will refuse to partake in any financial arrangements that are intended to disrupt the accuracy of results from a statistical analysis. JJDC Consulting Co. will not disclose or authorize to disclose any information that could lead to the benefit of a third party or personal gain.