Black Saber Software Bias Report

Potential bias in hiring and remuneration processes

Report prepared for Black Saber Software by RACS
Consulting Limited

2021-04-21

Contents

```
## Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
## had status 1
```

General comments (you can delete this section)

You can delete this section, and if you want to check what it said, just open a template from the package again. You don't have to use this particular template, but you DO need to write you report in RMarkdown and include a cover page.

The cover page must have:

- *A title and subtitle*
- *"Report prepared for Black Saber Software by" your company name*
- *Date (assessment submission date is fine)*

You can change the colour of this cover to any colour you would like by replacing 6C3082 in the YAML above (line 11) to another hex code. You could use this tool to help you: https: //htmlcolorcodes.com/color-picker/

Executive summary

Guidelines for the executive summary:

- *No more than two pages*
- *Language is appropriate for a non-technical audience*
- *Bullet points are used where appropriate*
- *A small number of key visualizations and/or tables are included*
- *All three research questions are addressed*

According to the quarterly datasets provided by Black Saber software, RACS Consulting Limited was able to obtain both salary and role seniority data of current employees and use them to conduct research on the potential gender biases in the remuneration and promotion process in the Black Saber company.

In the study of salary and promotion processes, the results are summarised below:

- Table 1 shows that 58.84% of the entries in the employee data are from males and 39.46% of the entries are from female workers.
- According to table 2, out of all 6906 entries of quarterly employee data, promotions happened about 569 times, which translates to an 8.24% of promotion rate in the total employee data.
- After fitting our statistical model, we found that female workers have 30.44% lower odds to be promoted than male workers given the same level of productivity, leadership level and the same team/department they work in.
- On average, male workers have $1,763.21 higher salary than female workers.

On the other hand, RACS Consulting was also given the new grad hiring datasets and was able to investigate whether the AI software used for hiring and the interviewers at Black Saber were potentially gender biased in the hiring process.

- Although there were a roughly equal number of men and women candidates in phase 2, this ratio doubled to about twice the number of men as women in phase 3 and exactly 4 men and 1 woman hired in both teams.
- We made two models. In the model with gender, 1.09 lesser times women would be selected than men but this is not significant so we cannot use this. In the other model, the GPA and extracurriculars have 38222.094 and 125.46 times more chance of getting selected.
- In Phase 2, we made four models and in each of them, we checked if the scoring of them is biased for any specific gender. 5.87 times lower women would get scored than men for speaking skills and similarly lesser odds for 2.36 times lower women would get scored than men for leadership presence.

- But in writing skills 82.43 times women would get a higher score than men. This shows AI is biased towards specific gender while scoring.
- We made two models, the first one with gender. It shows 2.38 lower times women would be selected than men but the value is not significant. The second model showed 1.09 times writing skills, 2.64 times leadership presence, 2.08 times for speaking skills and 1.08 times for technical skills are likely to be selected.
- The selection criteria are dependent on variables that are scored by AI but these scoring variables are biased for a specific gender. So, AI is biased but indirectly.
- In our fitted model for interviewer ratings, we found that gender is not a significant factor in determining the interviewer ratings, meaning that the interviewers of Black Saber are not gender biased.

Technical report

This part of the report is much more comprehensive than the executive summary. The audience is statistics/data-minded people, but you should NOT include code or unformatted R output here.

Introduction

Provide a brief introduction to your report and outline what the report will cover. This section is valuable for setting scope and expectations.

RACS Consulting Limited is helping Black Saber Software address concerns of potential bias in its hiring and remuneration processes. We will be examining the data provided by Gideon Blake of Black Saber Software through the use of statistical methods. More specifically, we will be data wrangling, building models, and performing statistical tests and inference to draw conclusions. This report will focus heavily on gender bias; we've determined that this is where bias is most likely to occur given its frequent being either a direct or an indirect source of bias in previous contexts.

Research questions

Use bullet points to to describe the research questions you are going to address. Write in full sentences.

There are a few research questions relating to Black Saber Software's employee and hiring pipeline which we will examine in depth:

- Whether or not gender bias occurs in the employee salaries.
- Whether or not gender bias occurs in employee promotions.
- Whether or not gender bias occurs in the hiring pipeline via the AI or the interviewer component.

Does gender bias occur in the employee salaries?

For each research question, you will want to briefly describe any data manipulation, show some exploratory plots/summary tables, report on any methods you use (i.e. models you fit) and the conclusions you draw from these

Dataset

In our research, we conducted two tests derived from the employee dataset. The first test was to check whether there exists a gender bias in the salary of current employees. For this particular test, we used the current employee's data collected from Black Saber Software without any further alterations. The model included 6,906 current employees. To study the gender bias in the salary, we considered gender, role seniority, leadership, and productivity as fixed variables, and team as a random variable.

Methods

First, we used a generalized linear model on gender bias in salary with variables of role seniority, leadership, and productivity without the team variable. This model showed that the variable gender and role seniority are significant with extremely small p-value. Also, at 95% confidence interval, there was no evidence that leadership and productivity are significant to the salary, but gender and role seniority were. The second model used a linear mixed model adding team variable as random effect. The model showed that men earn $1,763.21 more on average than women for each quarter, and the average salary got higher as they were promoted to senior positions. By looking at the t-values, we figured that the gender and role seniority variable have large t-values indicating they are both significant. The confidence interval of this model seemed to have similar results with the first model.

At last, we used a likelihood ratio test to determine which model fitted the data better. Since there was strong evidence with p-value less than 0.01 against that generalized linear model explains the data as well, we determined that the linear mixed effect model with team variable as random effect explains the data better than the generalized linear model.
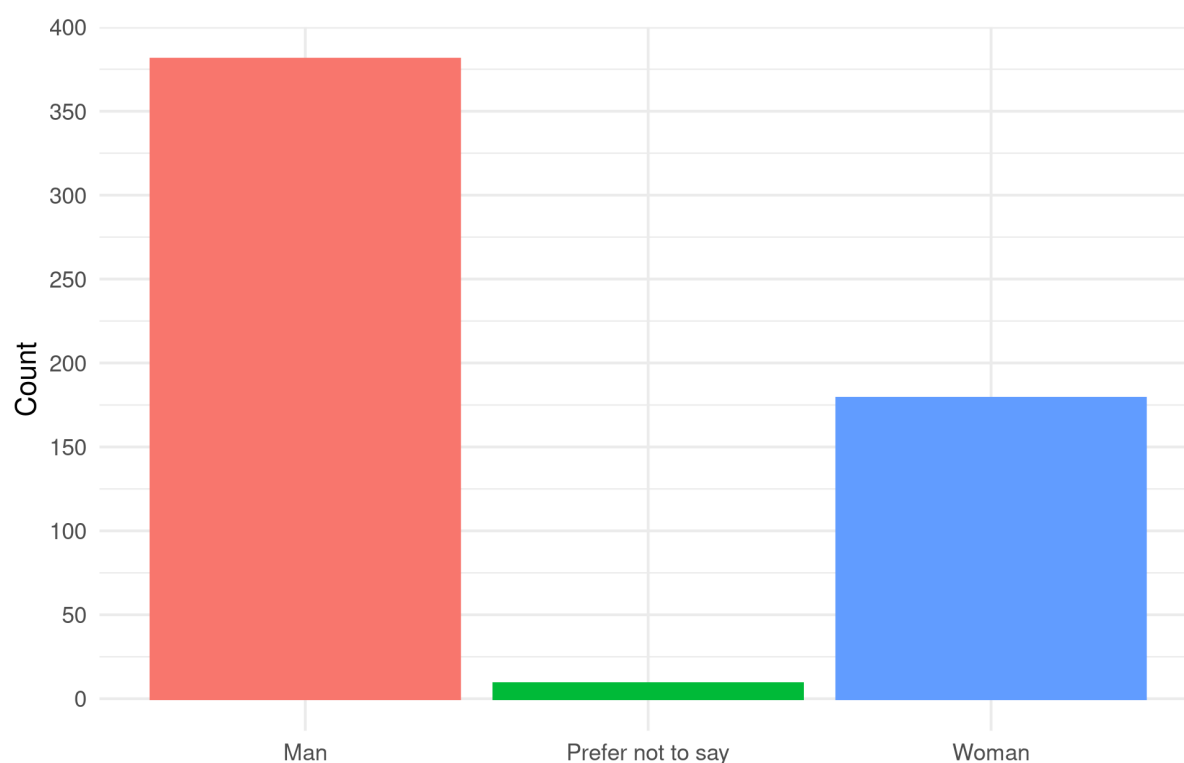
Does gender bias occur in the employee promotions?

Dataset

Table 1: Male vs Female Ratio

gender	count	gender_ratio
Man	4064	0.5884738
Prefer not to say	117	0.0169418
Woman	2725	0.3945844

Table 2: Promotions and Gender Within
Black Saber

gender	count	gender_ratio
Man	4064	0.5884738
Prefer not to say	117	0.0169418
Woman	2725	0.3945844

Figure 1: Number of promoted employees by gender

The second test we ran on the current employee dataset was to test whether promotion processes were biased in Black Saber Software. In this particular test, we defined a dummy variable called "promoted" that shows 0 when the person did not get promoted compared to the previous period, or 1, when the person's role-seniority changed compared to the previous period. The new variable would act as the response variable in our test and we would be able to check whether the historical promotion process was gender biased. Since the response is either 0 or 1, it is clear that the response follows a binomial distribution. Thus, it's suitable to model the data using logistic regression.

Model

The first model we used in this section is a standard logistic regression with productivity, gender and leadership for level as independent variables. In this model, we observed that under 95% confidence interval, a female faces 37.26% lower odds to be promoted compared to a male worker with all qualifications being the same. Although this result has a high significance level, we noticed that there are different operating teams within the company and we suspect that each team may have a different standard for promoting employees within the group. Thus, in our second model, we decided to include these factors into our generalized linear mixed model and see if random effects of employees working in a different team have any influence on the promotion processes.

Results

After fitting our model, we found that for productivity and leadership level being fixed, a female worker has 30.44% lower odds to be promoted compared to a male worker that has the same qualifications. With p-value being less than 5% level, it suggests that the results we got from this model is reliable and gender bias does exist within the promotion process at Black Saber Software.

Is bias present in the final round of the hiring process?

Dataset

Unlike employees, the hiring data was made up of 4 separate datasets; one for each phase of the hiring pipeline. Only candidates who've made it to a certain phase are included in that dataset. These 4 datasets were joined into a single applicants dataset containing all the applicants, their AI scores and interviewer ratings if any, and the farthest phase they reached. The phase 3 dataset was also joined with the phase 2 and hired dataset to get the applicants' AI scores and an indicator variable indicating whether or not they were hired.

The new grad hiring process consists of 3 phases, with multiple AI-controlled steps, and an interview stage at the end. Most notably, candidates don't speak face to face with an interviewer until phase 3, the last part of the application. Up until then, they receive assessments and are graded based on an AI. Since AI is heavily used in the hiring pipeline, we will examine the AI in addition to the face-to-face interview component in the hiring process for potential gender bias. As a result of the multi-step hiring process, we will go through this data phase by phase and examine each phase for potential gender bias.

Table 3: Applicants in phase 2

	Man	Prefer not to say	Woman
Data	67	2	63
Software	78	1	89

We take an initial look at the number of males and females who applied to each team and how far along the hiring process they reached. Table 3 shows us that roughly the same number of men and women made it to phase 2. In fact, there were 89 women in this phase who applied to the Software team, compared to only 78 men.

Table 4: Applicants in phase 3

	Man	Woman
Data	7	3
Software	8	4

However, table 4 shows that roughly double the number of men were able to move on to phase 3 in both teams. This immediately draws attention because the pattern happened in both teams.

Table 5: Applicants who were hired

	Man	Woman
Data	4	1
Software	4	1

Finally, 4 men and 1 woman were hired in both teams, as seen in table 5 - once again, for both teams. So overall, roughly the same number of men and women applied for the new grad position, but 4 times more men were hired.

Passing Phase 1

We found that almost every candidate who submitted their CV and Cover Letter, had a GPA over 2.0, and had some extracurriculars passed the initial phase. There were a total of 320 applicants meeting these criteria, and all 300 applicants in phase 2 met these criteria, meaning only 20 applicants who met the criteria were dropped. To further confirm no bias in this phase, we constructed a model by adding a dummy variable "selection" set to 0 when a candidate did not meet these conditions and 1 if they did. We then made a logistic regression model with selection as a response variable and gender as a fixed effect and another logistic regression model with selection as response and variables besides gender as predictors. Gender being woman had a p-value of 0.838 in the first model, meaning it was not a significant predictor. In the second model, GPA had a p-value of 6.49e-05 while extracurriculars had a p-value of 2.66e-4, showing these were the effects that were significant in determining if applicants move onto phase 2.

Passing Phase 2

In phase 2, the AI gave scores to the candidates according to their application and video interview. Here we can check if the AI scores were biased towards a specific gender. We made four linear mixed models for which each response variable was one of the AI score variables. In two out of four variables, more points were given to the men than women by the AI and in one of the models, more points were given to women than men. In particular, women got on average 0.8583 lower with a 95% CI between -1.38 and -0.33 on the leadership presence compared to men, where leadership presence is out of 10. This is significantly lower, especially if leadership presence is heavily weighed when making a hiring decision. Women also had lower speaking skills on average than men, being rated about 1.7667 lower on average with all other factors being the same, where speaking skills is rated out of 10.

We also made a model for the applicants who passed phase 2. After adding an indicator variable indicating if the applicant passed phase 2, we constructed two mixed effects logistic regression models, both of which used the indicator variable as the response and team as a random effect. One of them used just gender as a fixed effect, and the other used four AI chosen skill factors (leadership presence, writing, speaking, and technical skills) as its fixed effects. We find that gender is not significant in determining if an applicant passes phase 2. However, all 4 skill factors are significant, with their p-values all being well under 0.001. This tells us that although gender doesn't directly play a role in determining if an applicant passes phase 2, it does so indirectly by influencing the leadership presence and speaking skills. Bias therefore exists in this phase, as the biased AI determines the applicants' skill ratings, which are then used to determine if the applicant passes.

Passing The Interview

Finally, we explored the interviewing stage for potential bias. Each candidate in this phase interviews with two arbitrary interviewers, each of whom gives them a score. We can examine Black Saber's interviewers as a whole for bias on gender by looking at the scores they gave because the interviewers are arbitrary and the two ratings are independent. A new variable for the total rating, which is the sum of both interviewer ratings was used for the model; this way, we can include both interviewer ratings as one response variable, and see how an applicant's total rating changes with different factors. A linear mixed model was built with gender and the AI-determined skill factors (leadership presence, writing, speaking, and technical skills) as fixed variables and the applicant's team as a random intercept; we do so because the ratings given out by interviewers of the two teams may differ on average.

The 95% confidence interval for the gender term ranges from -3.9009 to 6.2664. Because 0

is included in this confidence interval, we cannot say that gender has an effect on the total interviewer rating. Rather, the skill factors have a more significant effect on the total rating, with the confidence intervals for all 4 skill factors being above 0. We can conclude that the interviewers of Black Saber, in fact, are not biased towards gender. This doesn't mean that the actual hiring decision isn't biased, as it is likely based on both the interview and AI rating, but it's nonetheless a good look for the employees at Black Saber Software.

Discussion

In this section you will summarize your findings across all the research questions and discuss the strengths and limitations of your work. It doesn't have to be long, but keep in mind that often people will just skim the intro and the discussion of a document like this, so make sure it is useful as a semi-standalone section (doesn't have to be completely standalone like the executive summary).

Through our investigations, we observed that Black Saber Software is facing a gender bias in both their remuneration and promotion processes. A male worker on average gets paid $1,763.21 higher compared to a female worker with the same qualifications. Moreover, under a 95% confidence interval, an average male worker at Black Saber Software has a 30.44% higher odds to get promoted compared to a female worker, assuming the two have similar qualifications. In order to avoid such biases to persist, we recommend Black Saber Software to redesign a more appropriate promotion system that is gender neutral to assess their current employees for future promotion opportunities. Also, the salary calculation system should be more transparent and fair to reduce the pay gap between genders.

In our research of the hiring decision, we found that the AI is biased when it comes to ranking an applicant's skills. For example, women received on average a score that was 0.8583 lower than men in the leadership presence grade with all other factors being the same. The speaking skill rating also favors men, with men being rated about 1.7667 higher on average. Both of these ratings were out of 10. We also found that gender plays no role in determining an applicant's interviewer rating; the interviewers at Black Saber are unbiased. Since advancing through the hiring pipeline is based on the AI scores and the hiring decisions are based on both the AI scores and interviews, we can conclude that the hiring process at Black Saber is biased towards men.

During our analysis on the current employee dataset, a limitation we noticed is that each employee serves a different number of years within Black Saber Software. We ignored years of experience in our model and solely focus on the productivity and leadership presence they contribute to the company in each quarter. From a practical point of view, years of experience within a company may be a valid consideration in the actual assessment for promotion opportunities to incentify workers to stay at the company and increase the employees retention rate. However, we think

this assumption is a little too broad and may be far from the truth and thus we ignored this factor in our statistical modeling.

Strengths and limitations

Consultant information

Consultant profiles

Complete this section with a brief bio for each member of your group. If you are completing the project individually, you only need to complete one for yourself. In that case, change the title of this section to "Consultant profile" instead. Examples below. This section is only marked for completeness, clarity and professionalism, not "truth" so you can write it as if we're a few years in the future. Put your current degree in as completed and/or add your first choice grad school program, whatever you like. What skills related skills would you most like to highlight? What job title do you want?

Ruei-Hung (Ray) Chen. Ray is the Senior Data Scientist at RACS Consulting Limited. He has been with the company since 1999 and has led several high-profile projects such as the housing bubble estimation project in 2008. His findings eventually caused the housing bubble to burst and led to the financial crisis in 08. He earned his Bachelor degree in Culinary Art from the University of NotSoGreat in 1987.

Stella Bae. Stella is a junior consultant at RACS Consulting Limited. She specializes in reproducible analysis and marketing management. Stella earned her Bachelor of Science, majoring in Statistics and Media Studies from the University of Toronto.

Aditya Gulati. Aditya is a senior consultant with RACS consulting limited. He specializes in data visualization. Aditya earned his Bachelor of Science, Major in Statistics and Mathematics, from the University of Toronto in 2022.

Cullen Pu. Cullen is a Junior Analyst Intern at RACS Consulting Limited. He's been with the company for 1 month and has yet to work on a large-scale project. He specializes in data visualization and programming and will earn his B.S. in Statistics & Computer Science from the University of Toronto in 2022.

Code of ethical conduct

This section should be fairly short, no more than half a page. Assume a general audience, much like your executive summary.

- *Make at least three relevant statements about your company's approach to ethical statistical consulting. These should be appropriately in line with professional conduct advice like the (Statistical Society of Canada Code of Conduct)[https://ssc.ca/sites/default/files/data/ Members/public/Accreditation/ethics_e.pdf] or the (Ethical Guidelines for Statistical Practice from the American Statistical Society)[https://www.amstat.org/ASA/Your-*

Career/Ethical-Guidelines-for-Statistical-Practice.aspx]. For example, "the customer is always right" ISN'T the type of thing an ethical statistical consultant would include.

- *Be very careful not to just copy and paste from these other documents! Put things in your own words.*

- Respect the client's privacy agreement and prohibit disclosing information to a third party outside of RACS Consulting Group.

- After the initiation of the project, avoid frequent contacts with the client to ensure the unbiasedness in our investigations.

- Respect the source of the data and prohibit any usage of the provided datasets for purposes outside of the project.

- Refuse to manipulate data/ investigations for clients to achieve goals outside of the agreed upon projects.

Final advice: KNIT EARLY AND OFTEN!