Analysis of fairness based on talent and value in the hiring and remuneration processes

A special focus on gender parity

Report prepared for Black Saber Software by SFP

2021-04-21

Contents

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*

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

Recently, several complaints were received related to potential bias in the hiring and remuneration processes. Hence, BSS (Black Saber Software) required our team to take a look and give out a related report for the Board of Directors. The purpose of this study is to investigate if BSS's hiring, promotion, and salary were processing fairly based on talent and value to the company. We were especially interested in exploring potential gender bias in the hiring and remuneration processes. Besides, we also investigated the performance of the new AI service in the hiring pipeline. Data was provided by the data team from BSS including hiring data for their new graduate program and data for their current employees.

In this report, we will specify different explanatory and response variables for. Then, we will fit several generalized linear models and a linear mixed effects model to analyze the results. At the end, we will point out the same limitations of our model. The whole report will be run on R.studio and knit into pdf.

Research questions

- Does the new AI service in the hiring pipeline work and is reliable for future use?
- In particular, is gender a potential bias during the promotion and salary processes?
- If the promotion, and salary process are all fair to all employees and based on their ability and value to BSS?

Relationships between whether a candidate was hired or not and the variables during the entire hiring process

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

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	log(OR)	95% CI	p-value
gpa	-2.4	-4.5, -0.77	0.009
cv	-16		>0.9
cover_letter	-16		>0.9
extracurriculars	-1.6	-3.6, -0.22	0.042
work_experience	0.40	-1.0, 2.0	0.6

model_1

Is the AI used during the hiring process reliable.

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```
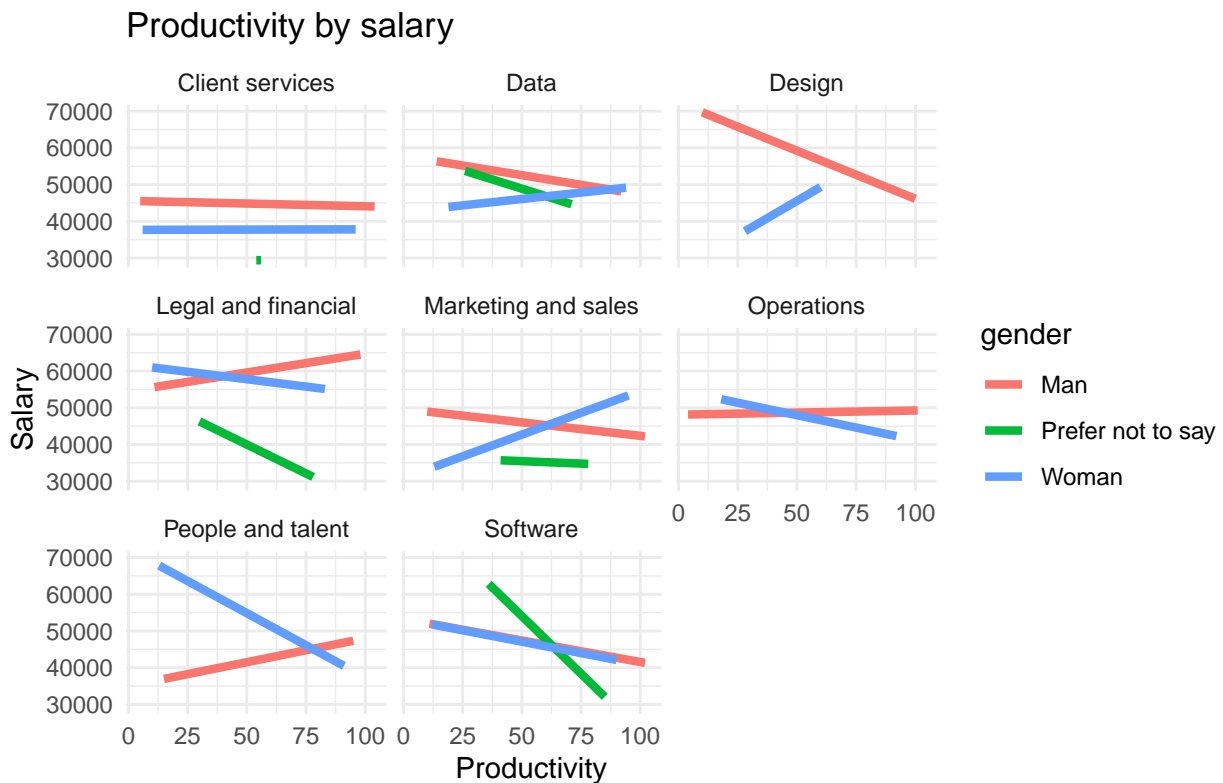
Characteristic	log(OR)	95% CI	p-value
gpa	0.67	-0.74, 2.2	0.4
extracurriculars	0.33	-1.0, 1.7	0.6
work_experience	0.01	-1.4, 1.5	>0.9
leadership_presence	-1.0	-1.5, -0.62	<0.001
speaking_skills	-0.74	-1.1, -0.43	<0.001

Characteristic	log(OR)	95% CI	p-value
technical_skills	-0.09	-0.14, -0.05	<0.001
writing_skills	-0.10	-0.15, -0.05	<0.001

model_2

Relationships between Gender, Promotion, and Salary

```
## `geom_smooth()` using formula 'y ~ x'
```



```
## `geom_smooth()` using formula 'y ~ x'

## Linear mixed model fit by REML ['lmerMod']
## Formula: salary ~ gender + team + (1 | employee_id)
##    Data: black_saber_current_employees
##
## REML criterion at convergence: 146211.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.3488 -0.2548 -0.0127  0.2108  7.7525
##
## Random effects:
##  Groups      Name         Variance  Std.Dev.
```

```
##  employee_id (Intercept) 270852785 16458
##  Residual                 68844147  8297
## Number of obs: 6906, groups:  employee_id, 607
##
## Fixed effects:
##                          Estimate Std. Error t value
## (Intercept)                 39400       1696  23.234
## genderPrefer not to say     -4115       5405  -0.761
## genderWoman                 -4723       1447  -3.263
## teamData                     7246       2418   2.997
## teamDesign                  12415       4348   2.855
## teamLegal and financial     10743       3346   3.211
## teamMarketing and sales      3280       2198   1.492
## teamOperations               5932       2397   2.475
## teamPeople and talent        7471       3548   2.106
## teamSoftware                 4227       2136   1.979
##
## Correlation of Fixed Effects:
##             (Intr) gnPnts gndrWm teamDt tmDsgn tmLgaf tmMras tmOprt tmPpat
## gndrPrfrnts -0.075
## genderWoman -0.485  0.105
## teamData    -0.549 -0.010  0.027
## teamDesign  -0.352  0.021  0.110  0.212
## tmLglandfnn -0.407 -0.086  0.045  0.274  0.155
## tmMrktngans -0.636 -0.037  0.099  0.417  0.240  0.308
## teamOpertns -0.576  0.025  0.072  0.381  0.219  0.276  0.423
## tmPplandtln -0.353  0.009 -0.025  0.255  0.140  0.183  0.279  0.257
## teamSoftwar -0.724 -0.011  0.243  0.433  0.263  0.322  0.494  0.446  0.283


## Likelihood ratio test
##
## Model 1: salary ~ gender + team + (1 | employee_id)
## Model 2: salary ~ gender + team + role_seniority + productivity + leadership_for_level +
##     (1 | employee_id:team) + (1 | employee_id)
##   #Df LogLik Df Chisq Pr(>Chisq)
## 1  12 -73106
## 2  24 -58578 12 29056   < 2.2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Gender distribution on leadership for level

Prepared by SFP
2021−04−20
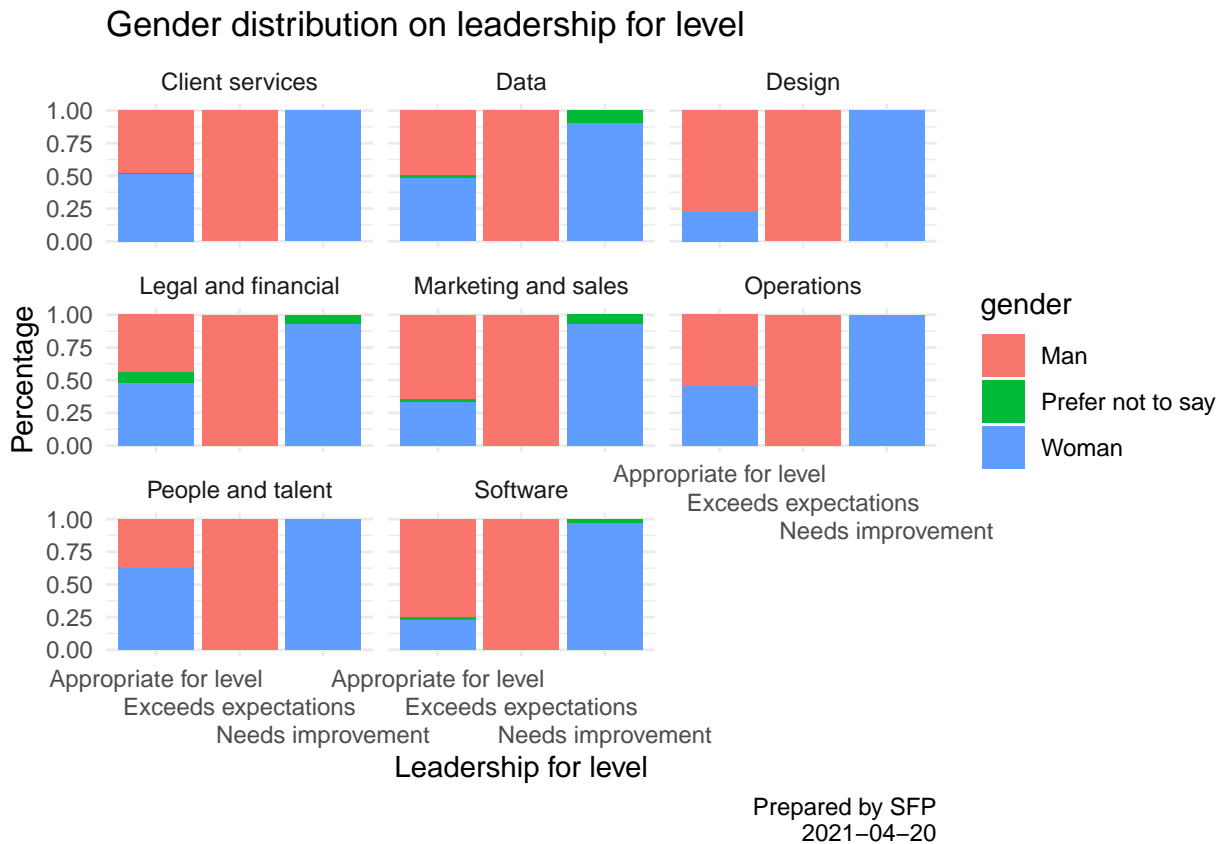
Discussion

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).

First, we explored the effects the various variables during the hiring phases had on the hiring chances. Fitting a model to model the applicant's hiring result (no if not hired, yes if hired) against the applicant's GPA,Curriculum Vitae, cover letter, extracurriculars, and work_experience, it was found that only the GPA and extracurriculars had a significant effect on the being hired, with p-values of 0.009 and 0.042 respectively.

Second, we explored whether the AI used during the first 2 phases of the hiring process was reliable, and if there were any biases.

Strengths and limitations

One limitation is that we were only provided data on current employees. The data of previous past employees is missing. Regarding the promotion and salary processes, past employees' information could give us some insights to why they left the company. This could give valuable information such as whether they were treated evenly. If we can get the past employees' data, we can analyze whether there exist biases during the remuneration process more comprehensively.

Another limitation results from the lack of ethnicity/race data for both the hiring and remuneration processes which could also be a potential bias that we need to check. We were told that the race data was not collected by the People and Talent team but they were considering it for EDI initiatives. The race might also be a factor that would affect the hiring result of applicants and also the promotion and wages processes of employees.

Finally, a limitation may be introduced as a result of implicit evaluation criterions for each hiring phase, especially AI-autograded phase and remuneration process. If we know how applicants can proceed to the next round in the hiring process, it would be easier to identify what effects are correlated and fit better models. For the remuneration process, if we can know what kinds of employees can be promoted or have higher salary, we will be able to explore more about potential biases.

Consultant information

Consultant profiles

Lucas Xian. Lucas is a senior consultant with SFP. He specializes in data analysis and visualization. Lucas earned his Honours Bachelor of Science, Specialist in Statistics, from the University of Toronto in 2023.

Dandan Zhang. Dandan is a junior consultant with SFP. She earned her Bachelor of Science, majoring in Mathematics and Statistics, from the University of Toronto in 2021.She specializes in data visualization.

Jong-Hoon Kim John is a senior consultant with SFP. He earned his Bachelor of Science in statistics and economics, from the University of Toronto in 2021. He specialises in data analysis and visualisation.

Zikun Lei Zikun Lei is a senior consultant with SFP. He earned his Bachelor of Science in statistics and economics, from the University of Toronto in 2021. He specialises in data analysis and visualisation.

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.*

Final advice: KNIT EARLY AND OFTEN!