Gender Parity Investigation

Analysis of Gender Parity in Hiring, Wages and Promotion of Black Saber Software

Report prepared for Black Saber Software by Day Day Up Ltd.

2021-04-21

Contents

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

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

In the report, we analyzed hiring data for new grad program and data about promotion and salary for the entire staff in Black Saber Software, and investigated whether gender parity exists in Black Saber Software's hiring, promotion, and salary processes. This report helps the Board of Directors decide the fairness in hiring, promotion and wages.

The main contents of this report are as follows:

- Addressing the three research questions on gender parity in hiring, promotion, and salary processes.
- Summary of the findings across all three research questions.
- Strengths and limitations of our investigation.

Research questions

-
-
- Whether gender pay inequality exists in Black Saber Software?

Informative title for section addressing a research question

We explored whether gender pay inequality exists in Black Saber Software by investigating the current employees dataset. This dataset includes data on all current employees for the whole duration of their employment, from 2013 Quarter 2 to 2020 Quarter 4. A total 6789 data from 597 employees who have indicated gender were included in this research question, but the 118 data from the 10 employees who have not indicated their gender were excluded. The source of the data is from Black Saber Software's data team. Each row represents salary, demonstrated leadership, and productivity for an employee in a given financial quarter. Productivity is rated on a 0-100 scale.

Salary data was first transformed into numeric form and special symbols like "$" and "," were deleted from it. And then role_seniority, gender, team, financial_q, leadership_for_level were transformed into factors.

To have an overview of the salary distribution between men and women, we plotted the boxplot of salary by gender below. Since first quartile, median, third quartile, maximum of men's salary are all greater than women's salary, in general, men's salary in Black Saber Software is higher than women's salary.
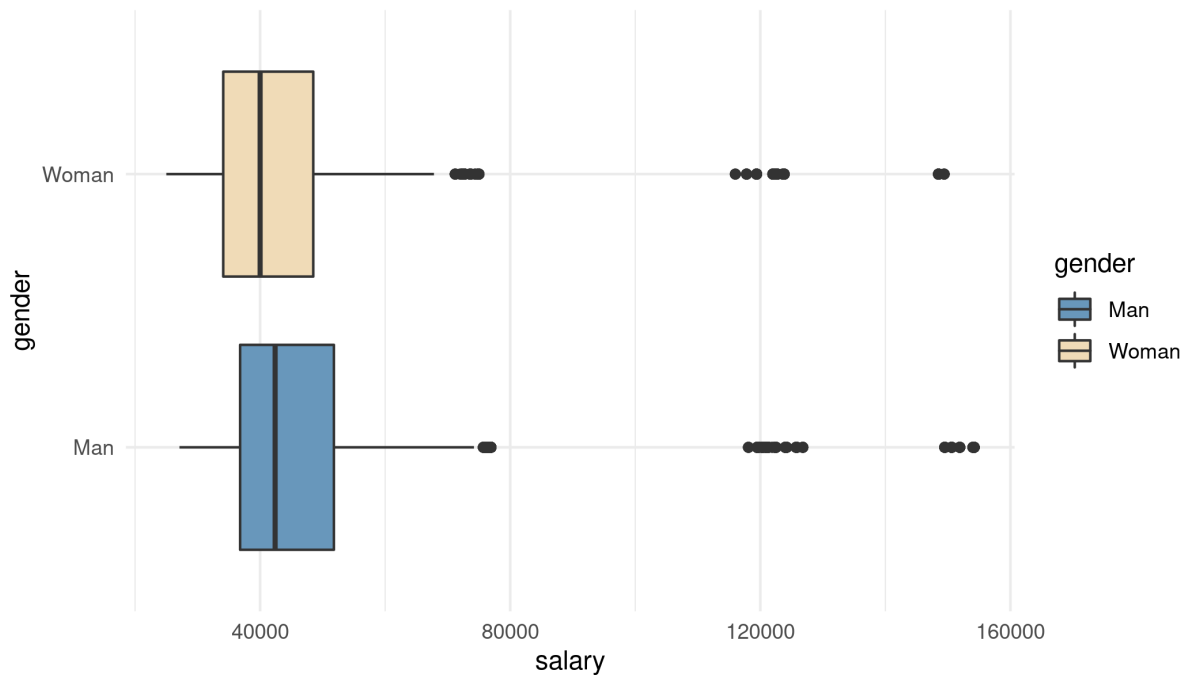

Figure 1: Boxplot of Salary by Gender.

However, since salary differs from team to team, we further provided the boxplot of salary by team and gender. Among the 8 teams, women's salary is lower than men's salary in 4 teams in general, including Client services, Data, Design, Marketing and sales. In team People and talent, women's salary is greater than men's. And in Legal and finance, and Operations, women and men have similar salary distribution. Finally, in Software team, first quartile and median of salary are similar between men and women, but women's third quartile is higher, and men's maximum salary is higher.
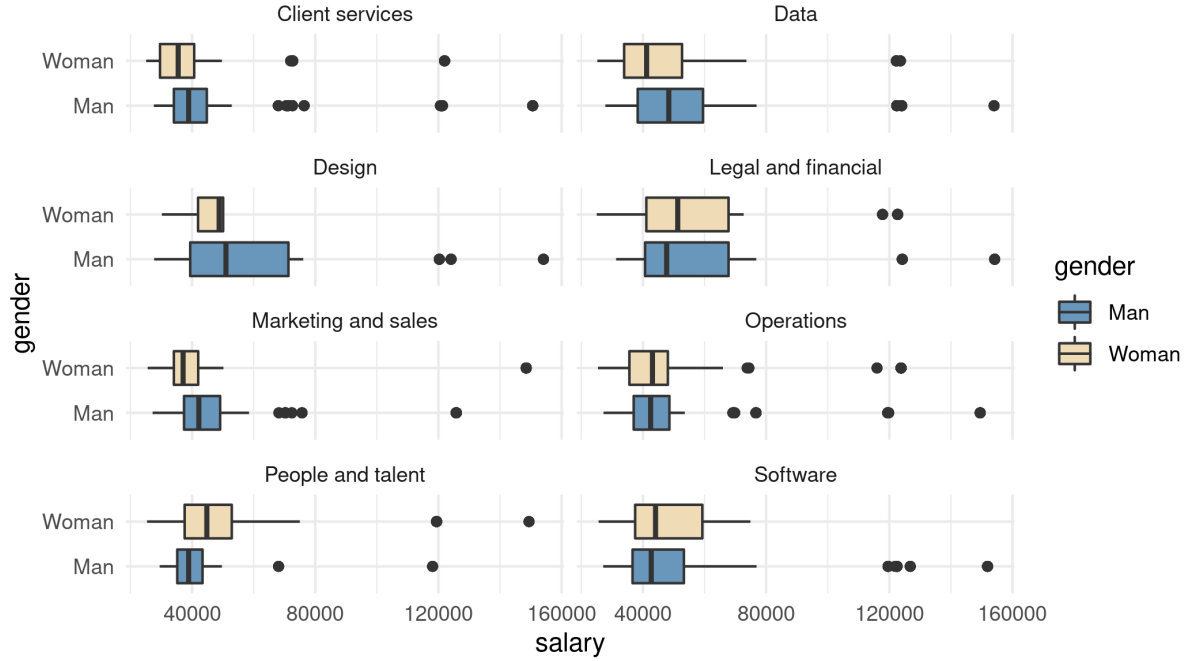
Figure 2: Boxplot of Salary by Team and Gender.

Since salary may also be influenced by many other factors like role seniority, to further compare salary between men and women, we used a linear mixed model. The model for Y_i (i^{th} employee's salary) is as follows:

$Y_i = \beta_0 + \beta_1\ gender_i + \beta_2\ leadership\_for\_level_i + \beta_3\ productivity_i + \beta_4\ role\_seniority_i + b_{1i} + b_{2i} + \epsilon_i$

Where $gender_i$ is the gender of the ith employee, $leadership\_for\_level_i$ is the the quality of demonstrated leadership for the ith employee, $productivity_i$ is the work output of the ith employee, $role\_seniority_i$ is the role seniority of the ith employee, b_{1i} is the team-specific random effect, b_{2i} is a employee-specific random effect, and ϵ_i denotes the residual error component.

The model included employee-specific random effect and team-specific random effect to address within-employee correlations and correlations among employees at the same team. Fixed effects in the model included gender, leadership for level, productivity, role seniority.

Since the p-value of the likelihood ratio test between model $Y_i = \beta_0 + \beta_1\ leadership\_for\_level_i + \beta_2\ productivity_i + \beta_3\ role\_seniority_i + b_{1i} + b_{2i} + \epsilon_i$ (without fixed effect gender) and model $Y_i = \beta_0 + \beta_1\ gender_i + \beta_2\ leadership\_for\_level_i + \beta_3\ productivity_i + \beta_4\ role\_seniority_i + b_{1i} + b_{2i} + \epsilon_i$ (without fixed effect gender) is 2.2e-16, we have very strong evidence against the hypothesis that the fixed effect gender is unnecessary complication to our model. Also the p-value for gender in the model is 2.13e-15, so we have very strong evidence against our null hypothesis that gender is not associated with change in salaries. While holding other predictors

in the model constant, the average women salary is 2255.40 lower than the average men salary.

Based on data visualization, model selection and model interpretation, we concluded that there is gender pay inequality in Black Saber Software.

Informative title for section addressing a research question

Informative title for section addressing a research question

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).

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?

Statsy McStatsstats. Statsy is a senior consultant with Eminence Analytics. She specializes in data visualization. Statsy earned her Bachelor of Science, Specialist in Statistics Methods and Practice, from the University of Toronto in 2023.

Datana Scatterplot. Datana is a junior consultant with Eminence Analytics. They specialize in reproducible analysis and statistical communication. Datana earned their Bachelor of Science, Majoring in Computer Science and Statistics from the University of Toronto in 2024.

Xin Peng. Xin is a senior consultant with Day Day Up Ltd. She specializes in data visualization. Xin earned her Bachelor of Science, Specialist in Computer Science and Data 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.*