Research on the gender equality in Black Saber Software

Report on the hiring, promotion and salary processess

Report prepared for Black Saber Software by Stat301-analytical

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

Provide a brief introduction to your report and outline what the report will cover. This section is valuable for setting scope and expectations. Gender equality is one of the most important indexes of a company's philsophy in the modern world. A fair hiring and promotional process can demonstrate a corporation's inclusive work environment. In this report, we will explore the significance of gender in the hiring and promotion processes as well as the salary benefits for Black Saber Software employees. More specifically, we will be examining the AI hiring pipeline in its ability to determine a good fit of the company based on their talent and values of the business rather than gender. We will also investigate the distribution of gender across roles of various seniority and salary level.

Research questions

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

- How does an applicant's gpa, extracurriculars, work experience and gender affect his/her chance of passing the first phase and how does an applicant's technicial, writing, speaking, leadership qualities and gender affect his/her chance of passing the second phase ?

- What are the proportion of employees for each gender across all levels of role seniority ?

- Does gender play a role in an employee' salary when taking into account of team and seniority level?

The distribution and significance in the first and second phase of the AI screening pipeline

To examine the significance of GPA, extracurriculars, work experience and gender in the first phase of the AI screening process, we will be utilizing the phase1-new-grad-applicants-2020 and phase2-new-grad-applicants-2020 csv files. First we go through the phase1 csv file and create a new column indicating whether the applicant was offered to participate in phase 2 with the mutate funtion.

Now we can easily visualize the number of participants and succesful applicants for each gender with a table.

From the two tables above, we observed that aside from the applicants who prefer not to reveal their gender, both the participants and those who were offered to continue in phase 2 were roughly equal. We will now utilize two Generalized Linear Models to investigate the significance of gender along with other qualities observed by the AI pipeline. We create two models that are nested in which the same qualities are utilized but one does not include gender as a covariate. If the less complicated model explains the outcome just as well as the more complicated model, we have evidence that gender does not play a crucial role in effecting whether an applicant passes the first round of the screening process.

From our generalized linear model that includes gender as an covariate, we can see that the p-value is quite large. Hence there are no statistically significant evidence that gender affects whether an applicant passes the first round of the screening phase. Furthermore, since there our two models are nested, we can utilize the drop-in-deviance test to examine how useful gender is in our model. The anova tests shows a very small drop in deviance which favors the smaller model.

Let us know examine the second phase of the AI screening process. Similar to before, we will perform some data wrangling to create a new column in the second phase csv file to indicate whether a participant in phase 2 was invited to a final interview. We will also visualize the participants and successful applicants of phase 2 with a table.

Now there appears to be significant difference in the distribution of male and female that passes the second phase of the screening process. Let us examine our findings with two models that both include technical, writing, speaking and leadership qualities but differ by the inclusion of gender.

From the first generalized linear model that includes gender as an covariate, we can see that the p-value of gerder is quite large . Hence there are no statistically significant evidence that gender affects whether an applicant passes the second round of the screening phase. The second model also demonstrate that all of techinical, writing, speaking and leadership skills are all statistically significant in determining an applicant's success in the second phase which is ideal and alings with Black Saber's philsophy. Finally, we can again utilize the drop-in-deviance test to examine how useful gender is in our model thanks to the nested nature. The anova tests shows a very small drop in deviance which favors the smaller model.

The proportion of employees for each gender across different levels of role seniority

The significance of gender in an employee's salary

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.

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!