Report title

Subtitle

Report prepared for Black Saber Software by Zeusolutions

2021-04-21

Contents

General comments (you can delete this section) **2**

Executive summary **3**

Technical report **4**

Introduction 4
Informative title for section addressing a research question 4
(Hiring Question) Fairness of the AI Recruitment Pipeline Manager 5
(Promotion Question) 8
(Salary Question) 8
Discussion 8

Consultant information **9**

Consultant profiles 9
Code of ethical conduct 9

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.

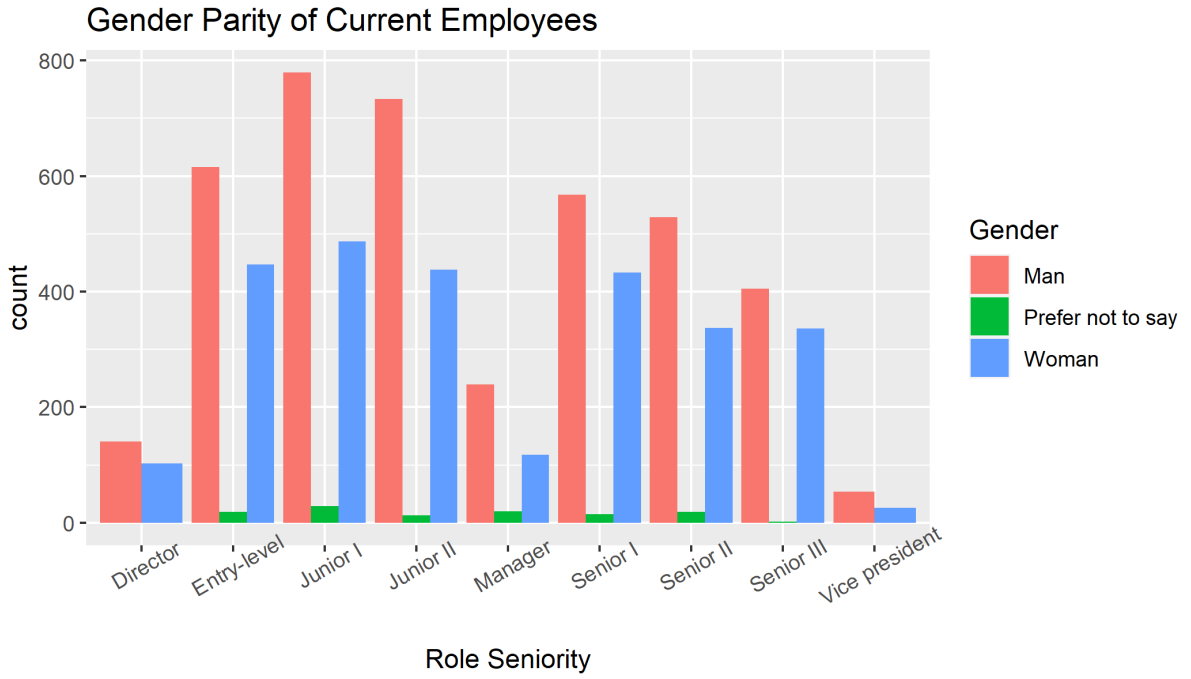
Research questions

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

Informative title for section addressing a research question

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

```
## Warning: Ignoring unknown parameters: preserve
```



(Hiring Question) Fairness of the AI Recruitment Pipeline Manager

The company Black Saber has a new AI recruitment pipeline manager that has been trialing for the Data and Software teams. The board has concerns about the process and questions whether or not it is fair and based on value and talent of each applicant. We have been given the data for each phase of the process, in which there are a total of three. Through each phase there is certain data collected through question and tasks in which the AI uses to narrow down the field of applicants from one phase to another. We want to determine what factors play a major role in being hired through Black Saber's recruitment pipeline manager.

(Phase 1) Defining Features of the Application Form

The first phase of the hiring pipeline is submission of an application form, cover letter and CV. Each line of `phase1-new-grad-applicants-2020.csv` contains the information of each applicant in phase 1:

- `applicant_id` = A unique ID given to each applicant in Phase 1
- `team_applied_for` = Software or Data
- `cover_letter` = 1 if present, 0 if not.
- `cv` = 1 if present, 0 if not
- `gpa` = 0.0 to 4.0

- `gender` = "Man", "Woman", or "Prefer not to say" as the only options
- `extracurriculars` = 0, 1 or 2. Description in which is accessed automatically by keywords where 2 indicates high, 1 indicates some and 0 indicates no relevant skill building extracurriculars.
- `work_experience` = 0, 1 or 2. Description in which is accessed automatically by keywords where 2 indicates high, 1 indicates some and 0 indicates no relevant work experiences.

Table 1: The first five applicants from phase 1 of the hiring pipeline

applicant_id	team_applied_for	cover_letter	cv	gpa	gender	extracurriculars	work_experience
1010	Software	0	1	1.3	Man	1	1
1020	Software	0	1	3.4	Woman	2	1
1030	Data	1	1	2.4	Woman	2	0
1040	Software	0	1	2.7	Man	1	1
1050	Data	1	0	2.1	Prefer not to say	0	1

Similarly, we are also given the list of applicants who have made it to phase 2 along with the factors assessed in phase 2. However, in this section we are only interested in which applicants made it to phase 2. For this reason I will create a variable called `passed_1` and set it to 1, if the applicant made it from phase 1 to phase 2 and 0, if not. This will be the response variable we are interested in for developing our model. The other variables will be fixed effects besides `team_applied_for`. We believe that `team_applied_for` should be added as a random effect because each team has a different set of required skills and/or experiences. Other than that, since each row corresponds to a unique applicant, this suggests that the observations are independent, hence our assumptions are not violated. We want to determine which is the most impactful component of the application form that allows the applicants to proceed to phase 2 of the hiring process.

Given the binomial response variable and random effect, we will be constructing a generalized linear mixed model. Before constructing the model, we want to first consider the gender distribution for those who passed phase 1 and those who did not. In addition, consider also the marginal probabilities and odd ratios.

Table 2: The number of applicants that passed phase 1 based on gender

	Did Not Pass Phase 1	Passed Phase 1
Man	146	145
Prefer not to say	8	3
Woman	159	152

Table 3: The probability of an applicant's gender given that they pass phase 1

	Did Not Pass Phase 1	Passed Phase 1
Man	0.4664537	0.4833333
Prefer not to say	0.0255591	0.0100000
Woman	0.5079872	0.5066667

Hence from the table above, given that the applicant passes phase 1, the probability that their gender is a "Man" is 48%, whereas a "Woman" is 51%. Also, the odds ratio of Woman vs Man who pass phase 1 is approximately 0.96. That is, the odds of a "Woman" passing phase 1 is less than that of a "Man". This odds ratio is relatively close to 1, hence there is no reason to suspect that there are any gender biases in those who passed phase 1.

Now consider the Generalized linear mixed model.

Table 4: Generalized linear mixed model of applicants who passed Phase 1

Characteristic	log(OR)	95% CI	p-value
cover_letter	63	-24,344, 24,471	>0.9
cv	52	-41,425, 41,528	>0.9
gpa	13	6.1, 19	<0.001
gender			
Man	—	—	
Prefer not to say	0.63	-25, 26	>0.9
Woman	0.91	-1.0, 2.8	0.4
extracurriculars	10.0	5.1, 15	<0.001
work_experience	12	6.1, 17	<0.001

OR = Odds Ratio, CI = Confidence Interval

Based on the model, it is evident that `gpa`, `extracurriculars` and `work_experience` have the most significant effect on passing Phase 1, based on their p-values. Hence I can conclude that the most impactful components of the application form that allowd the applicants to proceed to phase 2 of the hiring process are `gpa`, `extracurriculars` and `work_experience`.

(Promotion Question)

(Salary 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?

Andy Vu. Andy is a junior consultant with Zeusolutions. He specializes in statistical modeling. Andy earned his Bachelor of Science, Specializing in Mathematics & Its Applications (Probability/Statistics) and Majoring in Statistics from the University of Toronto in 2021.

Ethelia Choi.

James F. Kanu. James is a junior consultant with Zeusolutions. He specializes in data visualization. James earned his Bachelor of Science, Majoring in Mathematics and Minoring in Computer Science and Statistics from the University of Toronto in 2021.

Justin Lee. Justin is a junior consultant with Zeusolutions. He specializes in data analytics. Justin earned his Bachelor of Science, Majoring in Statistics and Mathematics from the University of Toronto in 2021.

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!