
Addressing Biases in Hiring and Remuneration in Black Saber Software

Sergio Steven Zheng Zhou, David Wong, Howard Wang,
Jing Yuan Zhang

Report prepared for Black Saber Software by Saber
Analytics

2021-04-22

Contents

| | |
|---|-----------|
| Executive summary | 3 |
| Introduction | 5 |
| Technical report | 7 |
| Does the AI system have a potential gender bias and/or team based bias in either phase of the recruitment process? | 7 |
| Recruitment Process | 7 |
| Does the Determination of Salaries Have Any Gender Biases? | 37 |
| Salary Process | 37 |
| Does gender bias affect promotions in Black Saber Software, either in specific teams or company as a whole? | 45 |
| Discussion | 51 |
| Strengths and limitations | 52 |
| Consultant information | 54 |
| Consultant profiles | 54 |
| Code of ethical conduct | 55 |
| References | 56 |

Executive summary

This report was commissioned by Black Saber Software to interpret if the hiring, salary, and promotion processes were biased and unfair. Recent public outcry has pointed towards potential biases in Black Saber's processes. The AI assesses applicants' scores on technical, writing, and speaking skills.

Previous literature has shown that gendered biases are still prevalent in the hiring, salary, and promotion processes. However, some gendered differences can boil down to certain factors, such as productivity differences.

The results of our study are summarized below. A baseline statistical significance level of 0.05 is used in this report.

Research Question 1: Investigating Biases in the Recruitment Process * Phase 1 to phase 2 recruitment process has no statistically significant bias towards gender.

- Percentage of men that proceeded to phase 2 (49.83%) is slightly higher than the percentage of women (48.87%).
- In Data team applications, 53.60% of men proceeded to phase 2 and 50.80% of women proceeded to phase 2.
- In Software team applications, 46.99% of men proceeded to phase 2 and 47.59% of women proceeded to phase 2
- Phase 2 to phase 3 recruitment process has no statistically significant bias towards gender.
 - In Data team applications, 5.6% of men proceeded to phase 3, while only 2.4% of women proceeded to phase 3.
 - In Software team applications, 4.8% of men proceeded to phase 3, while only 2.1% of women proceeded to phase 3.
- Phase 3 of the hired recruitment process has no statistically significant bias towards gender.
 - In Data team applications, 3.2% of men proceeded to phase 3 while only 0.81% of women proceeded to phase 3.
 - In Software team applications, 2.41% of men were hired, while 0.53% of women were hired.

It is important to note that, despite the statistical insignificance at the 5% level, there may be some minor biases for gender in phase 2 to phase 3. The percentage of successful males were more than 2 times greater than the percentage of successful females in both Data and Software

team applications. This implies that there may be some biases in the AI algorithm. However, technical, writing, speaking, and leadership components of phase 2 were statistically significant at the 0.01% level. So, the hiring process is fair and weighs applicants based on their value to the company.

Research Question 2: Investigating Biases in the Recruitment Process * Our model parameters were based off of a selection of variables from previous research

- “Role Seniority” and “Team” are the largest contributing factors when determining salary rates as they have the largest t-values and thus implies the most significant p-values
- This implies that gender effects are marginally contributing to the determination of salary since seniority and team are very significant
- The prediction output from our regression model displays strongly similar resemblance to the original data of Salaries amongst Black Saber Software employees

Research Question 3: Investigating Biases in the Recruitment Process * All employees who received promotions in the model all had productivity scores above 50.

- The model estimated that a woman whose leadership skill is appropriate for level has 1.53% lower odds than a man whose leadership skill is also appropriate for level.
- This corresponds to 0.3946% reduction in probability of being promoted compared to a man, all else equal.
- This suggests that although the bias effect is small and is mainly driven by productivity, there exists bias for men be promoted over women throughout Black Saber Software

Introduction

Bias in hiring and remuneration in the recruiting and compensation process has been a popular topic for contention amongst employers. Results from previous studies in academia cite that gender gaps in promotion rates and earnings partly reflect gender differences in research productivity (Xie and Shauman, 1998) whereas other empirical studies suggest that gender gaps in research productivity cannot be fully explained by differences in researcher characteristics such as the concentration of female researchers in academic disciplines with less publishing (Hesli and Lee 2011). A comparative study that builds upon searching for biases within the promotion of university educators from PhD graduates to full professors evaluated the role of gender in internal promotion. Although these researchers concluded that worker-specific productivity differences may be a primary reason for gendered promotion rates and that remuneration rates reflect the productivity of work from an individual (Jokinen et al., 2017), investigating the external validity of such results would help Black Saber Software understand where and if there exists gendered biases in a different context.

To understand possible areas for biases in these processes is crucial to understanding how equitable and fair these processes are. When considering factors like gender, productivity, tenureship, position level, and even area of work are some of many factors that contribute to recruiting and compensation. Black Saber Analytics, will conduct an analysis revolving around investigating if gendered effects across hiring and employee datasets affect factors of hiring rates, salary and promotion rates.

For Black Saber Software, pursuing Equity, Diversity and Inclusion (EDI) within its company structure poses both strong marketing for its company image as well as provides short term and long term gains in talent retention, improved talent prospects, and productivity (Karakhan et al., 2020). In our research, we will investigate the existence of hiring biases based on gender and attempt to connect our findings to prior research in the field to develop Black Saber's Software's goal of improving their EDI initiatives for their future hiring and remuneration processes.

Research questions

We, as members of Saber Analytics, will help Black Saber Software with a consulting project that will aim to develop statistical answers to the current questions that Black Saber Software faces regarding gender biases in their recruiting and remuneration processes. The focus of this analysis is to study whether there exists a bias, specifically a gender bias, the implemented AI pipeline. By focusing on the existence of a potential gender bias, we are able to clearly narrow the scope of the potential bias, while still keeping a universal approach that can be implemented in the three company departments, hiring, salary and promotion processes, that the board of

directors requested. Using the data set provided by the data team of Black Saber Software, we will seek to answer the following questions:

- Does the AI system have a potential gender bias and/or team based bias in either phase of the recruitment process?
- Does the determination of salaries have any gender biases?
- Does gender affect an employee's odds of being promoted?

Technical report

Does the AI system have a potential gender bias and/or team based bias in either phase of the recruitment process?

Recruitment Process

Methods

Data description and wrangling The data set provided by the Black Saber Software data team are as follows:

Data set: Phase 1: Initial Application

Description: Complete form of 613 individuals who apply to the Black Saber Software company and their characteristics for the year of 2021

Variables collected:

- applicant_id: qualitative variable - Numerical ID assigned to applicants in phase 1
- team_applied_for: qualitative variable - Data or Software
- cover letter: binary variable - 1 have, 0 don't have
- CV: binary variable - 1 have, 0 don't have
- gpa: quantitative variable - Numerical value for gpa
- gender: qualitative variable - Men, woman, or prefer not to say
- extracurriculars: qualitative variable - Numerical indicator of relevance and/or skill building extracurriculars of extracurricular activities with increasing values of 0, 1, 2
- Internship experience: qualitative variable - Numerical indicator of relevance of skill similarities from other job activities with increasing values of 0, 1, 2

Transformations:

1. Data set was included to a combined data frame named as "hiring" using the function "left_join" by the variable "applicant_id"

Data set: Phase 2: Technical Task, Writing Sample, Pre-recorded Video

Description: Form of only 300 individuals who proceeded to phase 2 of the Black Saber Software recruitment process and their relevant characteristics for the year 2021

Variable collected:

- applicant_id: qualitative variable - Numerical ID assigned to applicants in phase 1
- technical_skills: quantitative variable - AI graded score from 0-100 on timed technical tasks
- writing_skills: quantitative variable - AI graded score from 0-100 on timed writing tasks
- speaking_skills: quantitative variable - AI graded score from 0-100 on pre-recorded speaking tasks
- leadership_presence: quantitative variable - AI graded score from 0-100 on leadership related qualities found in the pre-recorded video

Transformations:

1. Data set was included to a combined data frame named as “hiring” using the function “left_join” by the variable “applicant_id”
2. Missing values in the combined data frame “hiring” were replaced by 0s
3. A categorical column named “pass_1”, composed of only 1s, was included to categorize those individuals that were able to proceed to phase 2
4. A filtered data frame was created named “hiring_filter_p2” to represent only those individuals that were able to proceed to phase 2
5. A categorical column named “pass_1_text”, composed of “pass” or “reject” matching the binary categorical column of proceed to phase 2
6. A rounded prediction column named “pred_p1” was fitted using the “predict” function
7. A correct prediction column named “correct_p1” was fitted if the prediction matched whether the individual proceeded to phase 2

Data set: Phase 3: Final Interview

Description: Form of only 22 individuals who proceeded to phase 3 of the Black Saber Software recruitment and their relevant characteristics for the year 2021

Variable collected:

- applicant_id: qualitative variable - Numerical ID assigned to applicants in phase 1
- Interviewer_rating_1: quantitative variable - Numerical value from 0-100 on company fit
- Interviewer_rating_2: quantitative variable - Numerical value from 0-100 on company fit

Transformations: 1. Data set was included to a combined data frame named as “hiring” using the function “left_join” by the variable “applicant_id” 2. Missing values in the combined data

frame “hiring” were replaced by 0s 3. A categorical column named “pass_2”, composed of only 1s, was included to categorize those individuals that were able to proceed to phase 3 4. A filtered data frame was created named “hiring_filter_p3” to represent only those individuals that were able to proceed to phase 3 5. A categorical column named “pass_2_text”, composed of “pass” or “reject” matching the binary categorical column of proceed to phase 3 6. A rounded prediction column named “pred_p2” was fitted using the “predict” function 7. A correct prediction column named “correct_p2” was fitted if the prediction matched whether the individual proceeded to the hired phase

Data set: Final Hires New Grad 2020

Description: Form of only 10 individuals who were sent an offer letter for Black Saber Software in the year 2021

Variables collected:

- applicant_id: qualitative variable - Numerical ID assigned to applicants in phase 1

Transformations:

1. Data set was included to a combined data frame named as “hiring” using the function “left_join” by the variable “applicant_id”
2. Missing values in the combined data frame “hiring” were replaced by 0s
3. A categorical column named “pass_3”, composed of only 1s, was included to categorize those individuals that were able to proceed to the hired phase
4. A filtered data frame was created named “hiring_filter_p3” to represent only those individuals that were able to proceed to phase 3
5. The filtered data frame, hiring_filter_p3, was wrangled into “hiring_filter_p3_long”. A combined column named “interview” describing interview 1 or interview 2. Another columns was created named “rating” giving the respective rating of interview 1 or interview 2
6. A categorical column named “pass_3_text”, composed of “pass” or “reject” matching the binary categorical column of proceed to phase 3
7. A rounded prediction column named “pred_p3” was fitted using the “predict” function
8. A correct prediction column named “correct_p3” was fitted if the prediction matched whether the individual proceeded to the hired phase

Purpose + Appropriateness + Accuracy of description For the statistical analysis, all individuals were included in the statistical methods.

Statistical methods used:

1. Generalized Linear Models

- Dummy variables for: team_applied_for, extracurricular, work_experience, and gender
- Binomial response distribution using a logistic link
- Regression summary statistics
- Confidence intervals
- P-values
- Likelihood-ratio test
- Hypothesis testing

2. Generalized Linear Mixed Models

- Dummy variables for: team_applied_for, gender
- Random effect: applicant_id
- Binomial response distribution using a logistic link
- Regression summary statistics
- Confidence intervals
- P_values
- Hypothesis testing

Statistical Models:

Phase 1 to Phase 3: Model - “p1_glm”

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{team,i} + \beta_2 x_{exc,i} + \beta_3 x_{workexp,i} + \beta_4 x_{gender,i} + \beta_5 x_{GPA,i} + \epsilon_i$$

where $i = 1, 2, \dots, n$ and p = probability of success

Phase 2 to Phase 3: Model - “p2_glm”

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{gender,i} + \beta_2 x_{team,i} + \beta_3 x_{tech,i} + \beta_4 x_{writing,i} + \beta_5 x_{speak,i} + \beta_6 x_{leader,i} + \epsilon_i$$

where $i = 1, 2, \dots, n$ and p = probability of success

Phase 3 to Hired: Model - “p3_glm”

Table 1: Summary Statistics for Recruitment Data

| Cover Letter | CV | GPA | Extracurriculars | Work Experience | Technical Skills |
|----------------|----------------|---------------|------------------|-----------------|------------------|
| Min. :0.0000 | Min. :0.0000 | Min. :1.200 | Min. :0.000 | Min. :0.000 | Min. : 0.00 |
| 1st Qu.:0.0000 | 1st Qu.:1.0000 | 1st Qu.:2.200 | 1st Qu.:1.000 | 1st Qu.:1.000 | 1st Qu.: 0.00 |
| Median :1.0000 | Median :1.0000 | Median :2.700 | Median :1.000 | Median :1.000 | Median : 0.00 |
| Mean :0.6427 | Mean :0.8842 | Mean :2.712 | Mean :1.222 | Mean :0.956 | Mean : 23.19 |
| 3rd Qu.:1.0000 | 3rd Qu.:1.0000 | 3rd Qu.:3.200 | 3rd Qu.:2.000 | 3rd Qu.:1.000 | 3rd Qu.: 46.00 |
| Max. :1.0000 | Max. :1.0000 | Max. :4.000 | Max. :2.000 | Max. :2.000 | Max. :100.00 |

$$\log\left(\frac{p}{1-p}\right) = \beta_0 U + \beta_1 x_{team,i} + \beta_2 x_{rating,i} + \beta_3 x_{gender,i} + \epsilon_i$$

where $i = 1, 2, \dots, n$ and p = probability of success

Explanation:

The proposed models with the given covariates were chosen for several reasons. The covariates, team, extracurriculars, work experience, and gender, were chosen through a likelihood ratio test for nested generalized linear models. The likelihood ratio test addresses the goodness of fit of the two models and compares it through the ratio of their corresponding likelihoods through maximum likelihood. By comparing two nested models, if the p-value of the test is statistically significant, there is no clear difference between the two models. This implies that the advanced model is equivalent to the simple model and explains the dataset as well.

Each phase has a corresponding model. The model, p1_glm, predicts the probability of proceeding to phase 2. The next model, p2_glm, predicts the probability for proceeding to phase 3. The last model predicts the probability of proceeding to the hired phase. Each of these models predict the probability of an individual with certain quantifiable characteristics and weigh those characteristics into the probability calculation.

Certain variables, such as cv and cover_letter, were omitted from the models because the variables are requirements to apply for the job. The gender variable was included in every model because the main issue was to see if there were any gender biases throughout any of the phases.

Result Analysis

Summary Statistics Explanation:

The first figure, named “Summary Statistics for Recruitment Data”, is a summary table for the summary statistics of the following covariates: CoverLetter, CV, GPA, Extracurriculars, Work

Table 2: Frequency Table for Gender Variable

| Gender | Frequency |
|-------------------|-----------|
| Man | 291 |
| Prefer not to say | 11 |
| Woman | 311 |

Experience, Technical Skills, Writing Skills, Leadership Presence, Speaking Skills, Interviewer 1, Interviewer 2 in the combined recruitment data. The second figure, named “Frequency Table for Gender”, is a gender frequency table for all the individuals who applied to Black Saber Software in the year 2021.

Pass Phase 1 - Likelihood-Ratio Test

\begin{table}

\caption{Likelihood-Ratio Test for p1_glm}

| Model | Log(likelihood) | Df | χ^2 | Pvalue | #Df |
|---------|-----------------|-----------|----------|-----------|-----------|
| Model_1 | 7 | -295.4163 | NA | NA | NA |
| Model_2 | 8 | -295.3277 | 1 | 0.1772887 | 0.6737137 |

\end{table}

A likelihood-ratio test was conducted to test for the most appropriate model between Model 1 (p1_glm):

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{team,i} + \beta_2 x_{exc,i} + \beta_3 x_{workexp,i} + \beta_4 x_{gender,i} + \beta_5 x_{GPA,i} + \epsilon_i$$

where $i = 1, 2, \dots, n$ and p = probability of success

and Model 2(“p1_glm_teamless”):

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{exc,i} + \beta_2 x_{workexp,i} + \beta_3 x_{gender,i} + \beta_4 x_{GPA,i} + \epsilon_i$$

where $i = 1, 2, \dots, n$ and p = probability of success

and it showed that model 2(model without the inclusion of factor(team_applied_for)) was a better fit due to a high p-value (pvalue= 0.6737137) suggesting that the simpler model explained the data just as well as the more complicated model. For the sake of this analysis, model 1 with the inclusion of factor(team_applied_for) will be used to answer the question of whether application for a certain team shows a potential bias.

Estimates and 95% Confidence Interval

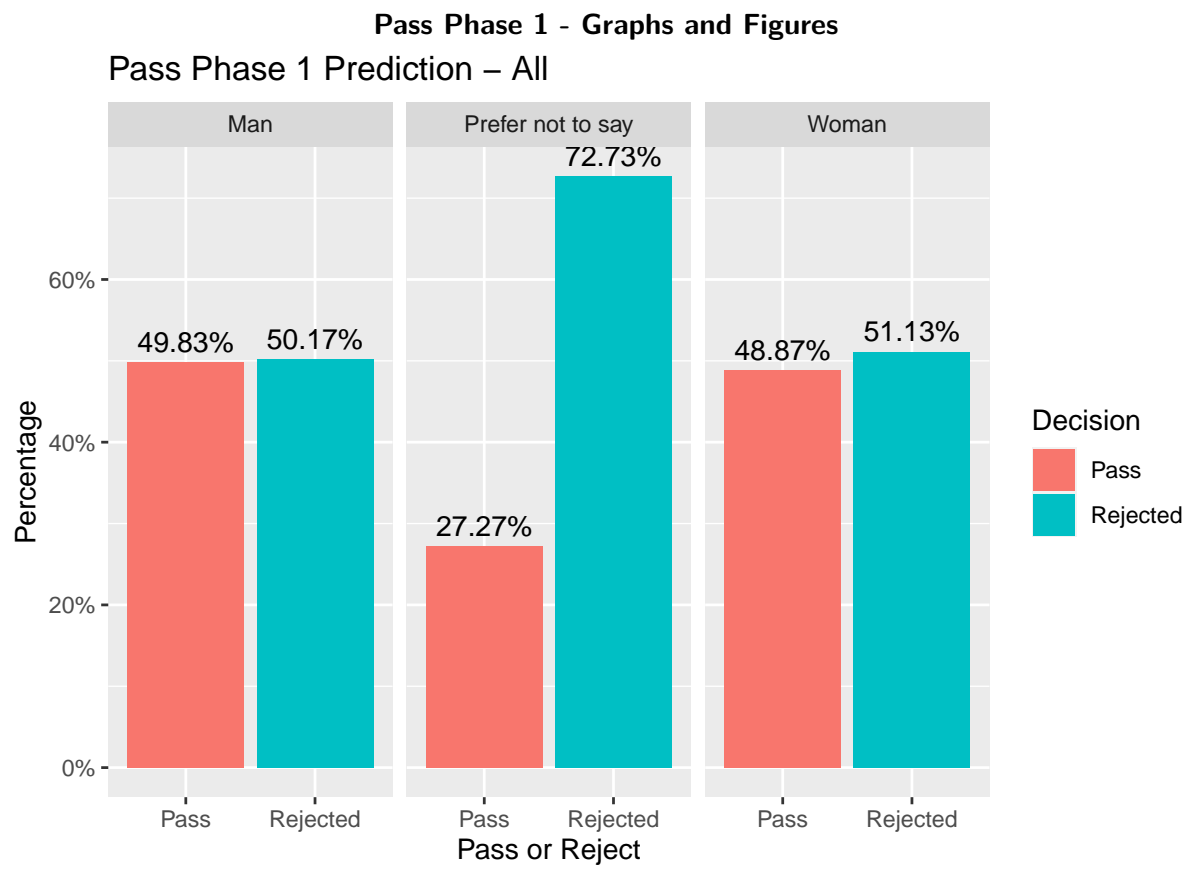
Waiting for profiling to be done...

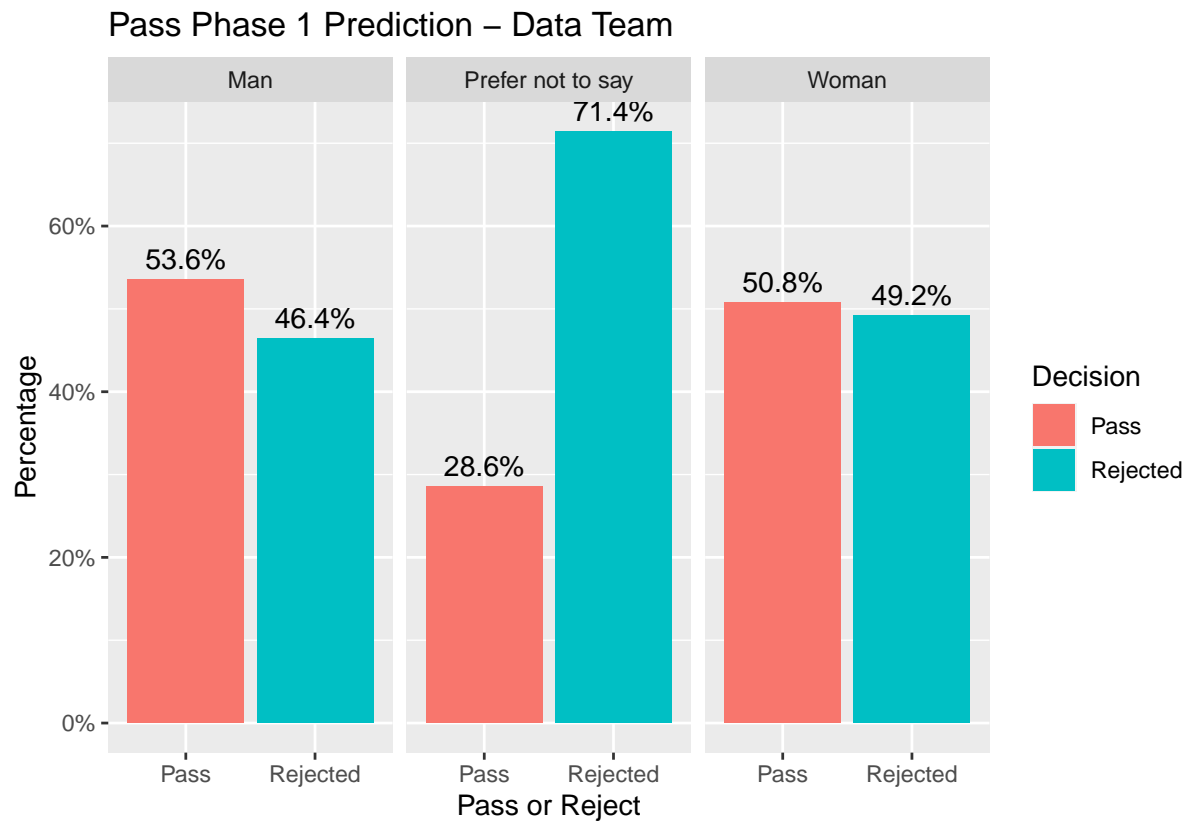
\begin{table}

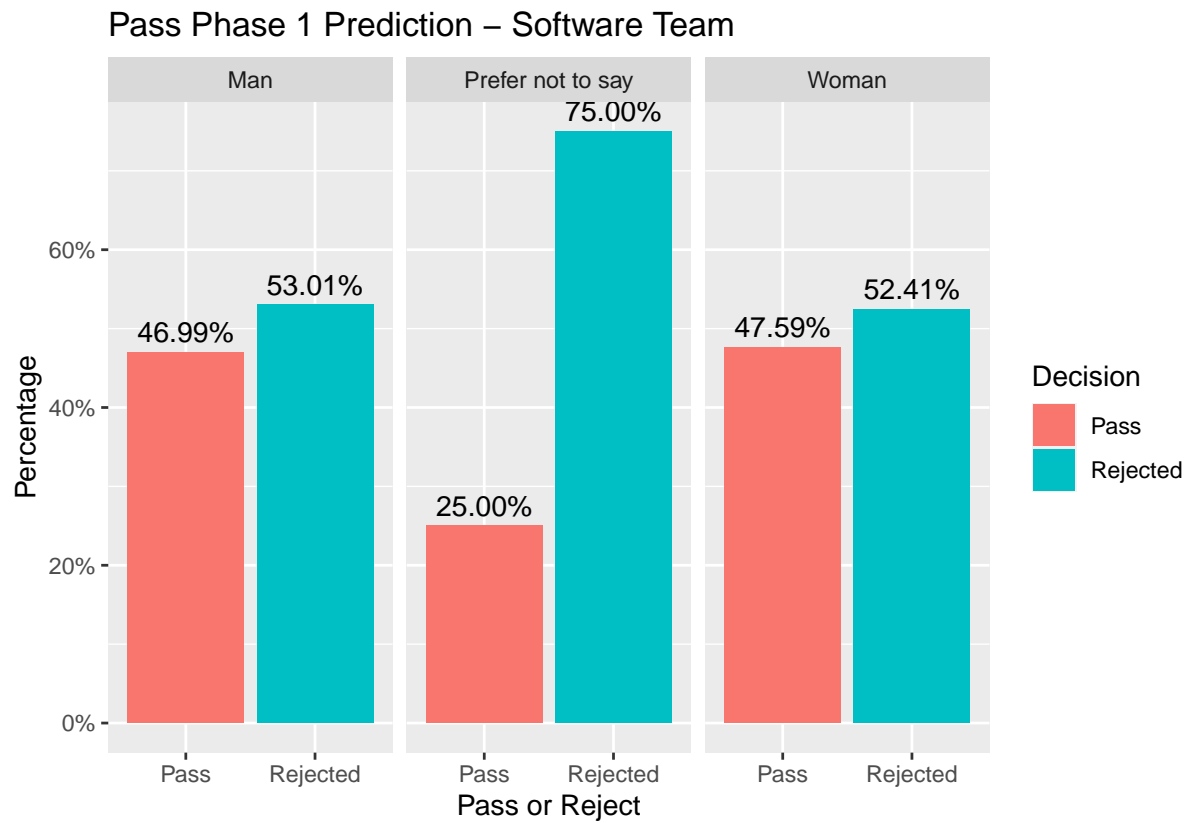
\caption{Estimates and 95% Confidence Interval for p1_glm}

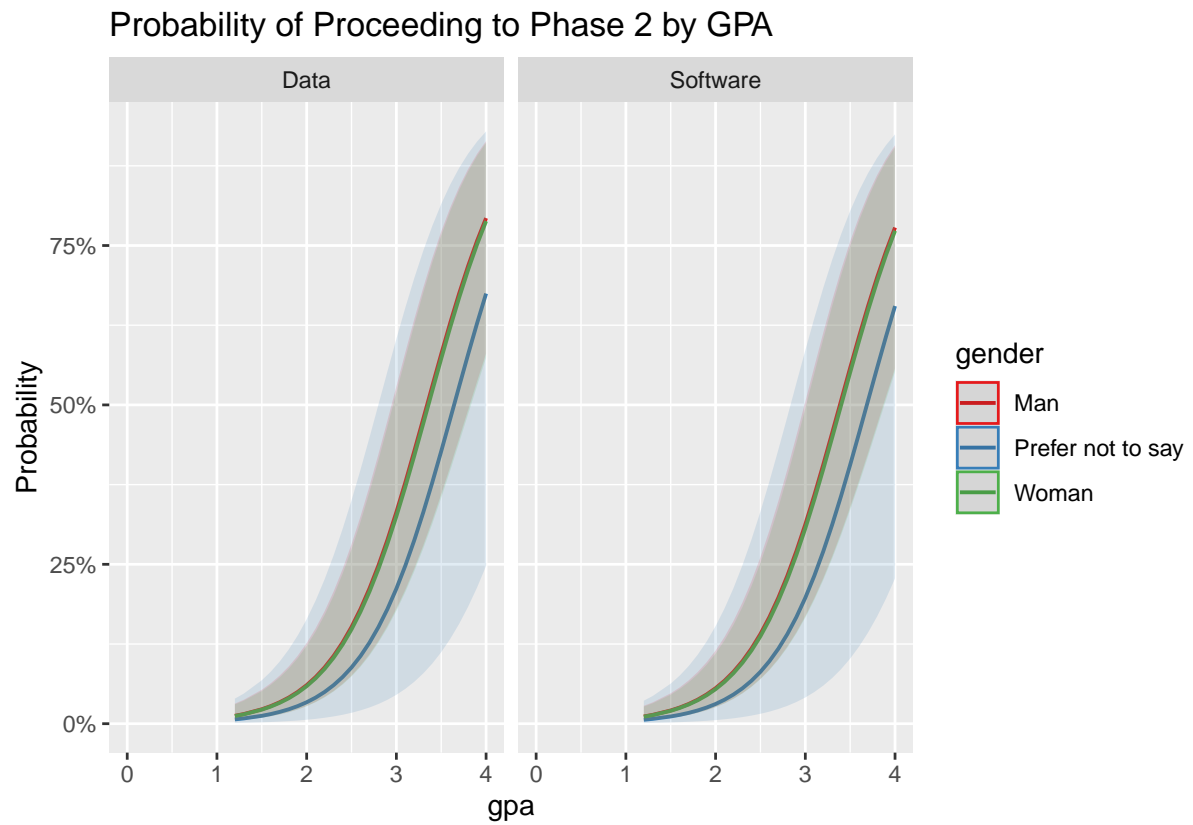
| | Estimate | 2.5% | 97.5% | pval |
|----------------------------------|----------|--------|--------|--------|
| (Intercept) | 0.0008 | 0.0002 | 0.0026 | 0.0000 |
| gpa | 0.8853 | 0.8363 | 0.9230 | 0.0000 |
| factor(team_applied_for)Software | 0.4780 | 0.3781 | 0.5798 | 0.6737 |
| extracurriculars | 0.5722 | 0.4750 | 0.6652 | 0.1466 |
| factor(work_experience)1 | 0.8183 | 0.6893 | 0.9105 | 0.0001 |
| factor(work_experience)2 | 0.7885 | 0.5802 | 0.9156 | 0.0118 |
| factor(gender)Prefer not to say | 0.3512 | 0.0836 | 0.7070 | 0.4531 |
| factor(gender)Woman | 0.4927 | 0.3933 | 0.5926 | 0.8868 |

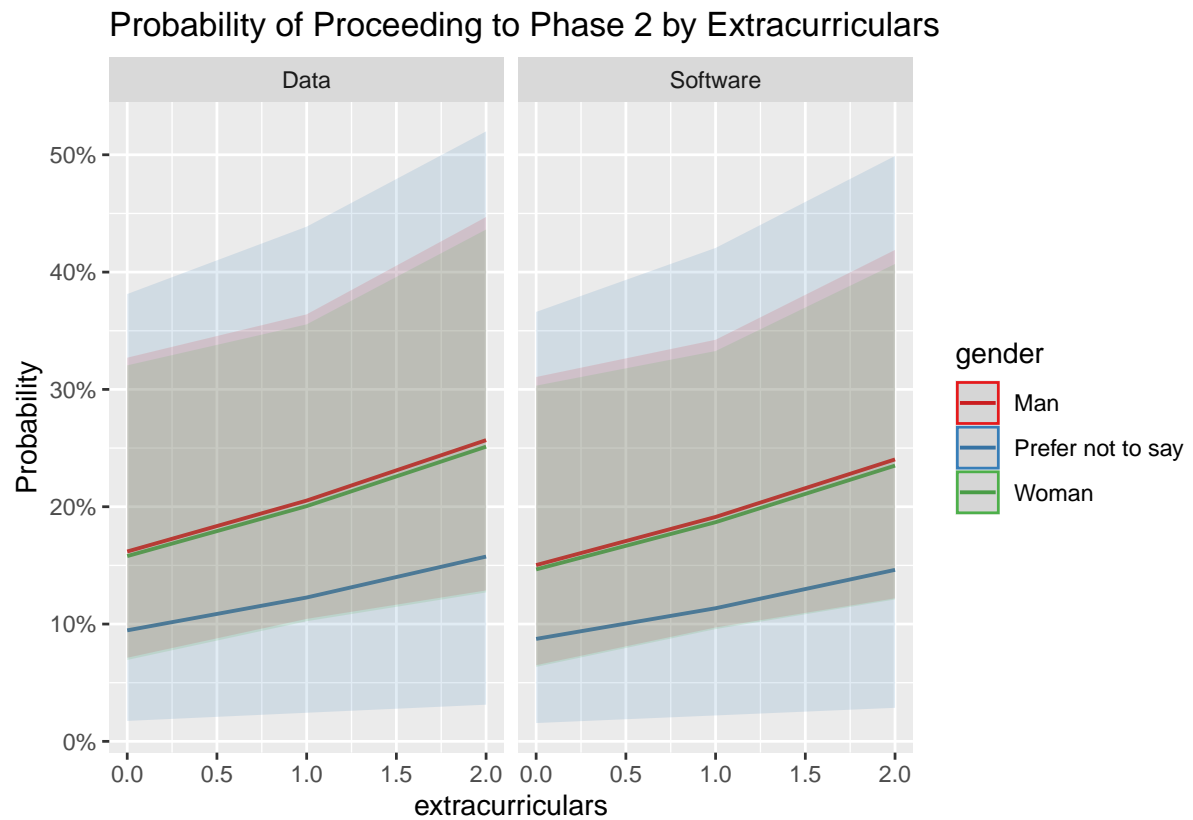
\end{table}













Result explanation:

An important note to make is that individuals in the the prefer not to say gender category have a limited sample size, therefore they will be deemed as outliers in the interpretation between man and women.

The analysis of the generalized linear model of whether an individual proceeded to phase 2 (response: pass_1) on the explanatory variables of team_applied_for(dummy variable for each category), extracurricular, work experience(dummy variable for each category), and gender(dummy variable for each category) showing 4 covariates as statistically significant at the 95% level or greater. These beta coefficients were: baseline with odds values equal to 0.0008 with a 95% confidence interval between (0.0002 and 0.0026) and statistical significance of (pvalue = $2.0 * e^{-16}$), gpa with odds values equal to 0.8853 with a 95% confidence interval between (0.8363 and 0.9230) and statistical significance of (pvalue = $2.0 * e^{-16}$), work experience - 1 with odds values equal to 0.8183 with a 95% confidence interval between (0.6893 and 0.9105) and statistical significance of (pvalue = 0.0001), and work experience - 2 with odds values equal to 0.7885 with a 95% confidence interval between (0.5802 and 0.9156) and statistical significance of (pvalue = 0.0118). The baseline covariate is for a male individual saying that male individuals have odds of 0.0008 which is statistically significant due to low

pvalues, while females have odds of 0.4927 which is not statistically significant due to high pvalues. This difference in odds can be interpreted as a potential non gender bias. The other covariates are factors that greatly affected the odds of proceeding to the next stage for an individual, where gpa and work experience relevancy positively affect the odds. Furthermore, the covariate for team_applied_for was not statistically significant with a pvalue of 0.6737 suggesting that there does not exist a potential bias between company team recruitment. Based on the statistical analysis, Black Saber Software does not show a statistically significant bias towards gender in its phase 1 recruitment process, but should not exclude the possibility of a slight potential gender bias due to small percentage differences as shown in the following graphical interpretations.

Using the provided bar charts we can see the percentage of people that were able to proceed to phase 2, distributed by gender. In the bar chart named “Phase 1 Prediction - All” it contains data for individuals that applied to either team, data or software. The decision making shows a potential bias towards the male gender, as the percentage of men(49.83%) is slightly higher than that of woman(48.87%) who were able to proceed to the next phase. The meaning of these percentages is that there exists a minor potential male gender bias in the individuals who proceed from phase 1 to phase 2 in the Black Saber Software recruitment. The next two bar graphs, which are named “Phase 1 Prediction - Data team” and “Phase 1 Prediction - Software team” respectively, show the percentage of individuals separated by gender and team that proceeded to the next phase. In the case of those individuals that applied to a data team, the percentage of men that were able to proceed to the next stage was (53.6%), a significantly higher percentage than that of women (50.8%). Lastly, for the software team, the percentage of men that were able to proceed to the next stage was (46.99%), a slightly lower percentage than that of women (47.59%). Due to the support from the significance test on the covariates on team_applied_for we can suggest that differences between team may be due to sample variances.

Using the provided figures we can also see how each variable affects the probability of proceeding to the next phase. The first figure named “Probability of Proceeding to Phase 2 by GPA”, shows the relationship of gpa and probability of proceeding to phase 2. This figure shows a positive relationship with gpa and probability, where higher gpa values give a better probability of proceeding to the next phase. Furthermore, the distributions between man and woman is overlapping which suggests that gpa values are seen as equal regardless of gender or team_applied_for. The potential gender bias can therefore be seen in how extracurricular and work experience are related to the probability of proceeding to the next stage. In the figure named “Probability of Proceeding to Phase 2 by Extracurricular” shows a positive relationship between extracurricular and probability of proceeding to the next phase, where higher level of extracurricular have a higher probability of proceeding to the next stage. Furthermore, the

distributions between man and woman is slightly higher for men which suggests a potential gender bias in the phase 1 recruitment process, but differences could be due to sample variations as since with a high pvalue. Lastly, the figure named “Probability of Proceeding to

Phase 2 by Work Experience” suggests that individuals who have had similar work experience(s) have a higher probability of proceeding to the next phase where a category of 1 and 2 are fairly similar. Furthermore, the the mean values for male individuals in the application of either team are slightly distributed higher than the mean values for females, suggesting a potential minimal gender bias towards men, but not in team distribution.

Pass Phase 2 - Likelihood-Ratio Test

\caption{Likelihood-Ratio Test for p1_glm}

| Model | Log(likelihood) | Df | χ^2 | Pvalue | #Df |
|---------|-----------------|-----------|----------|----------|-----------|
| Model_1 | 12 | -31.76367 | NA | NA | NA |
| Model_2 | 8 | -32.67861 | -4 | 1.829885 | 0.7670104 |

A likelihood-ratio test was conducted to test for the most appropriate model between Model 1(“p2_glm”):

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{gender,i} + \beta_2 x_{team,i} + \beta_3 x_{tech,i} + \beta_4 x_{writing,i} + \beta_5 x_{speak,i} + \beta_6 x_{leader,i} + \epsilon_i$$

where $i = 1, 2, \dots, n$ and p = probability of success

and Model 2 (“p2_glm_2”):

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{gender,i} + \beta_2 x_{team,i} + \beta_3 x_{tech,i} + \beta_4 x_{writing,i} + \beta_5 x_{speak,i} + \beta_7 x_{GPA,i} + \beta_8 x_{exc,i} + \beta_9 x_{workexp,i} + \epsilon_i$$

where $i = 1, 2, \dots, n$ and p = probability of success

and it showed that model 1(model without the inclusion of previous variables) was a better fit due to a high p-value(pvalue=0.7670104) suggesting that the simpler model explained the data

just as well as the more complicated model.

Estimates and 95% Confidence Interval

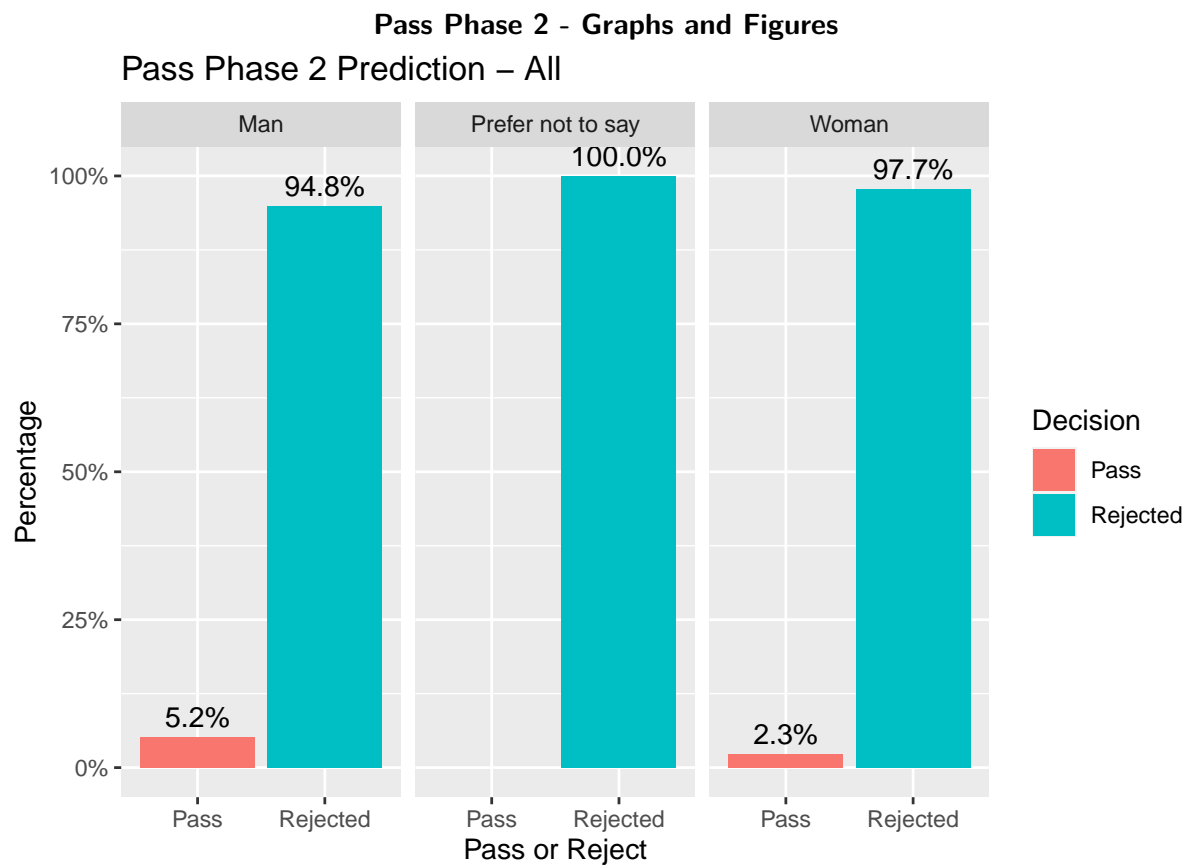
Waiting for profiling to be done...

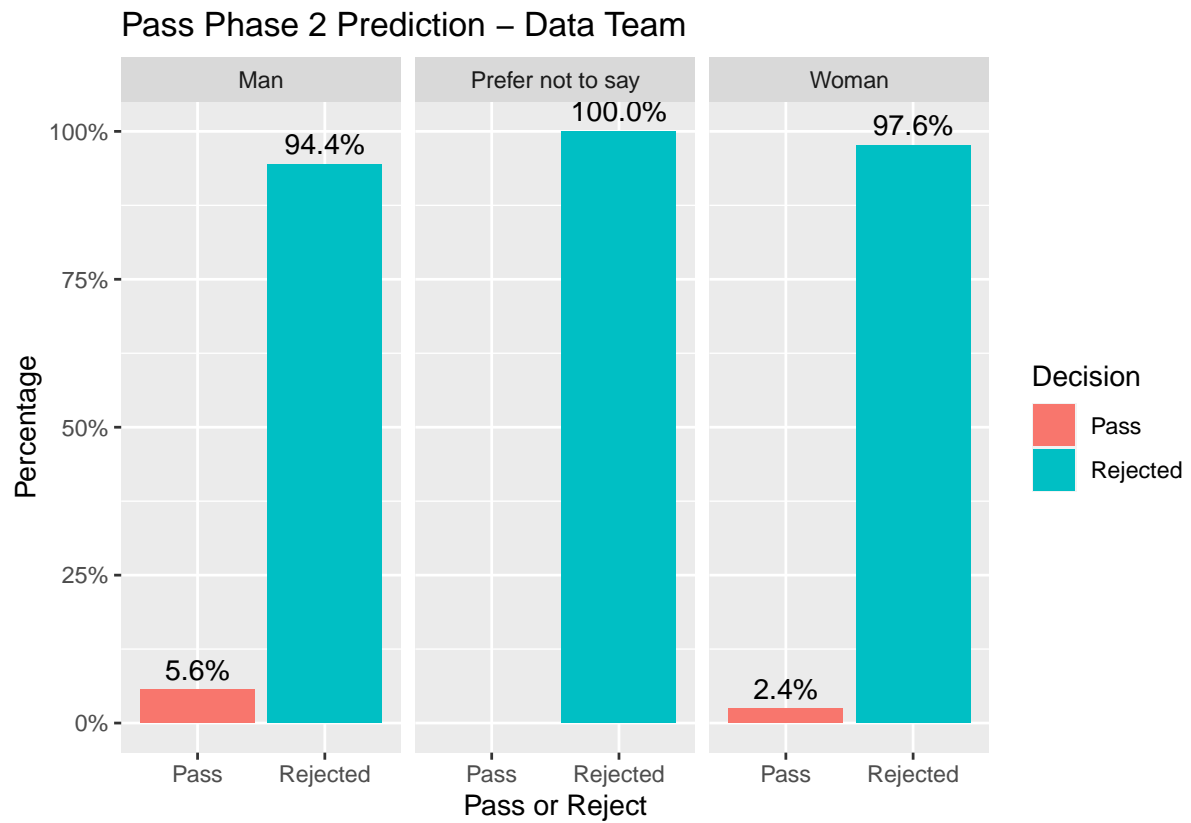
\begin{table}

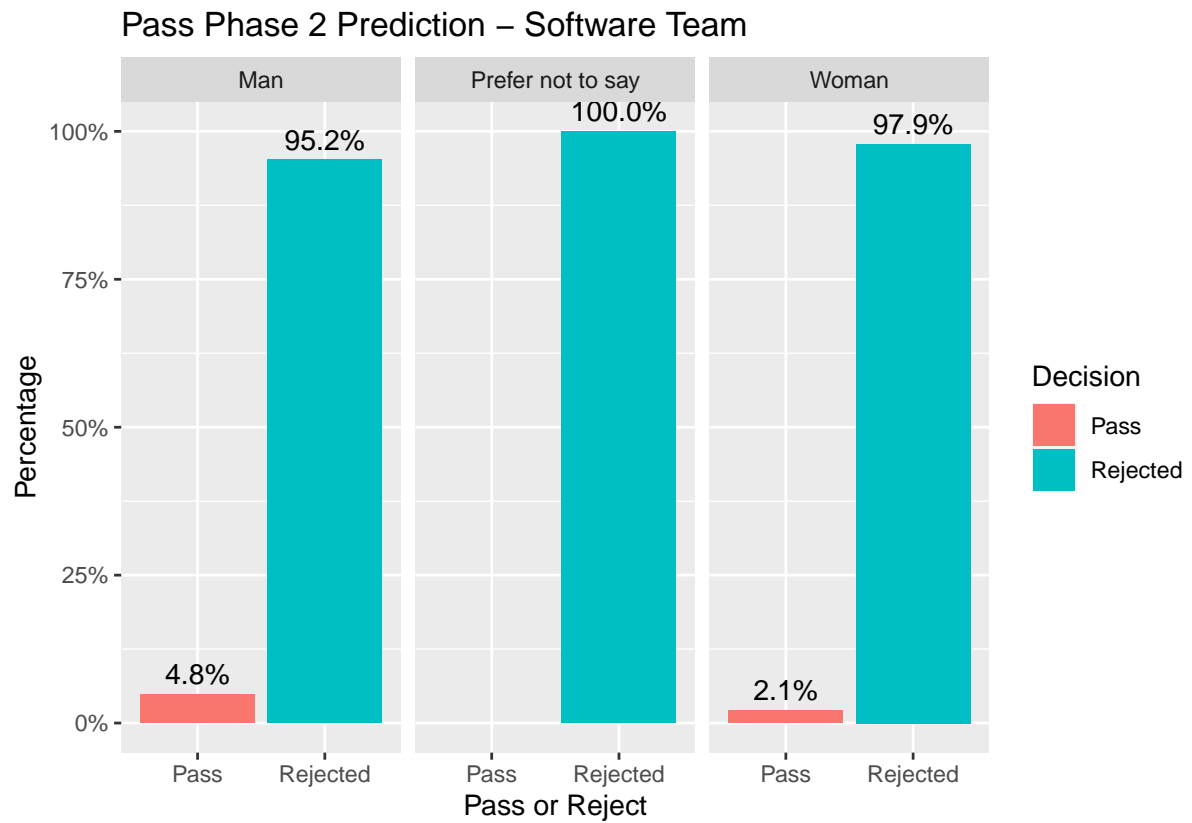
\caption{Estimates and 95% Confidence Interval for p2_glm}

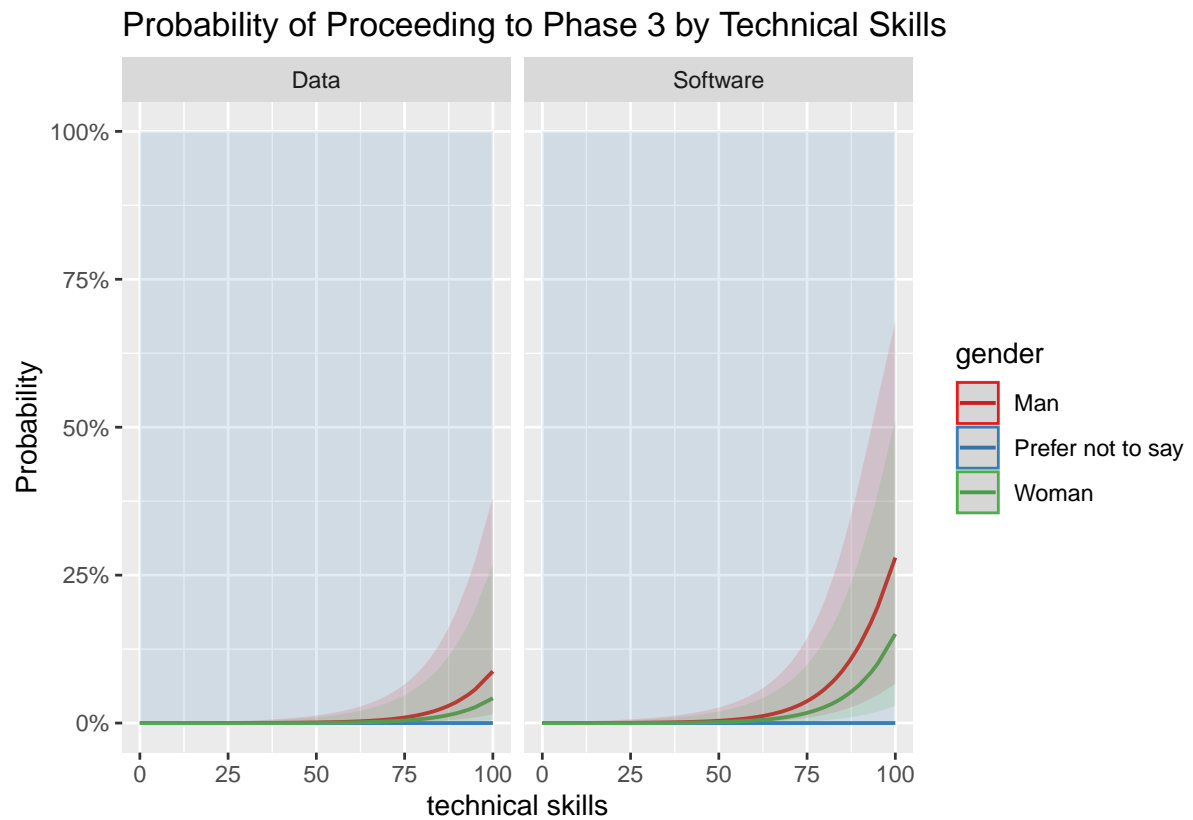
| | Estimate | 2.5% | 97.5% | pval |
|----------------------------------|----------|--------|--------|--------|
| (Intercept) | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| factor(gender)Prefer not to say | 0.0000 | NA | 1.0000 | 0.9937 |
| factor(gender)Woman | 0.3134 | 0.0853 | 0.6615 | 0.3042 |
| factor(team_applied_for)Software | 0.8020 | 0.5021 | 0.9532 | 0.0638 |
| technical_skills | 0.5231 | 0.5134 | 0.5356 | 0.0000 |
| writing_skills | 0.5257 | 0.5141 | 0.5405 | 0.0001 |
| leadership_presence | 0.7246 | 0.6438 | 0.8128 | 0.0000 |
| speaking_skills | 0.7031 | 0.6231 | 0.7908 | 0.0000 |

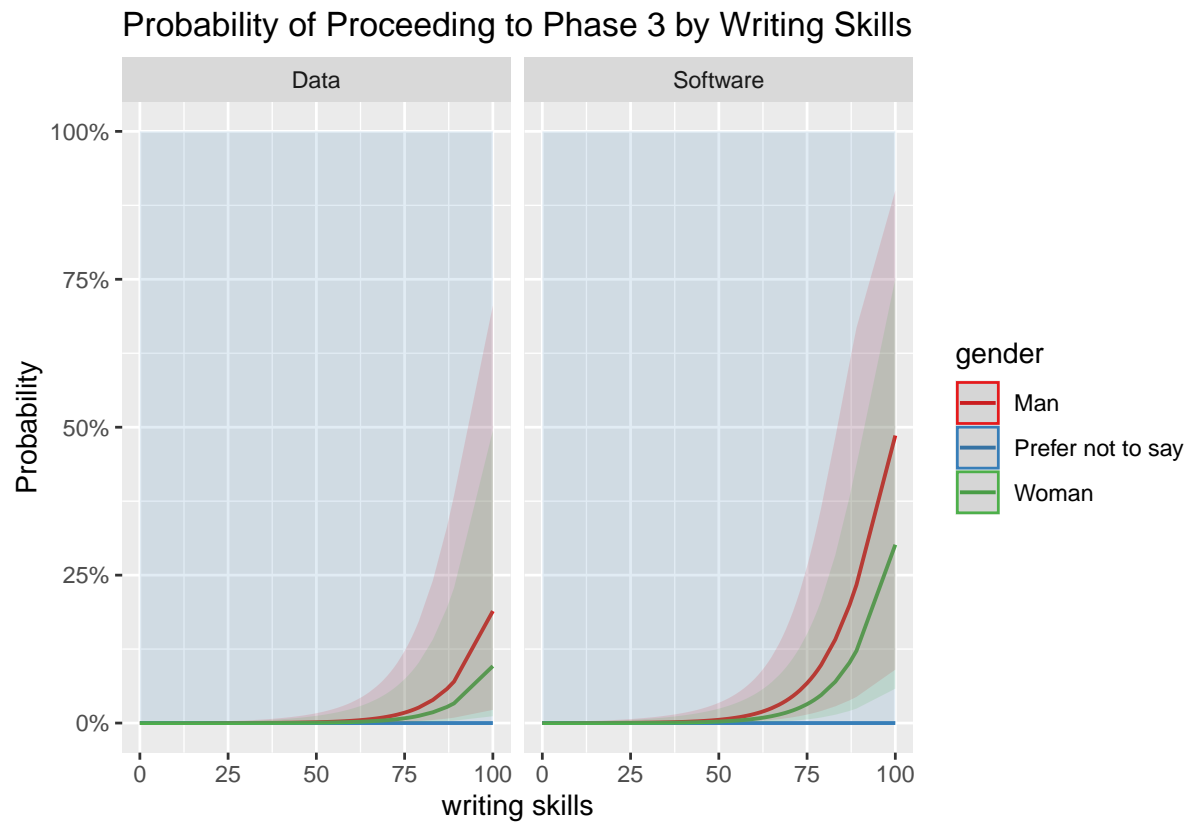
\end{table}

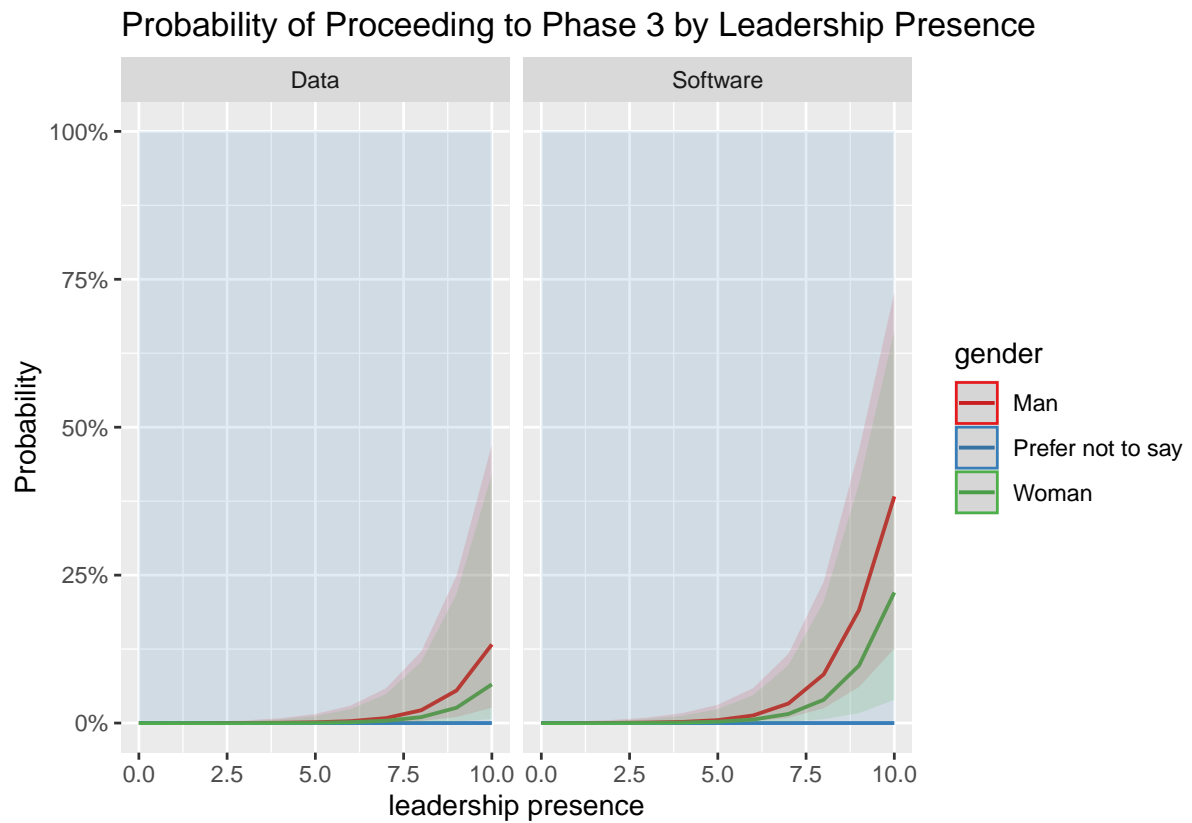


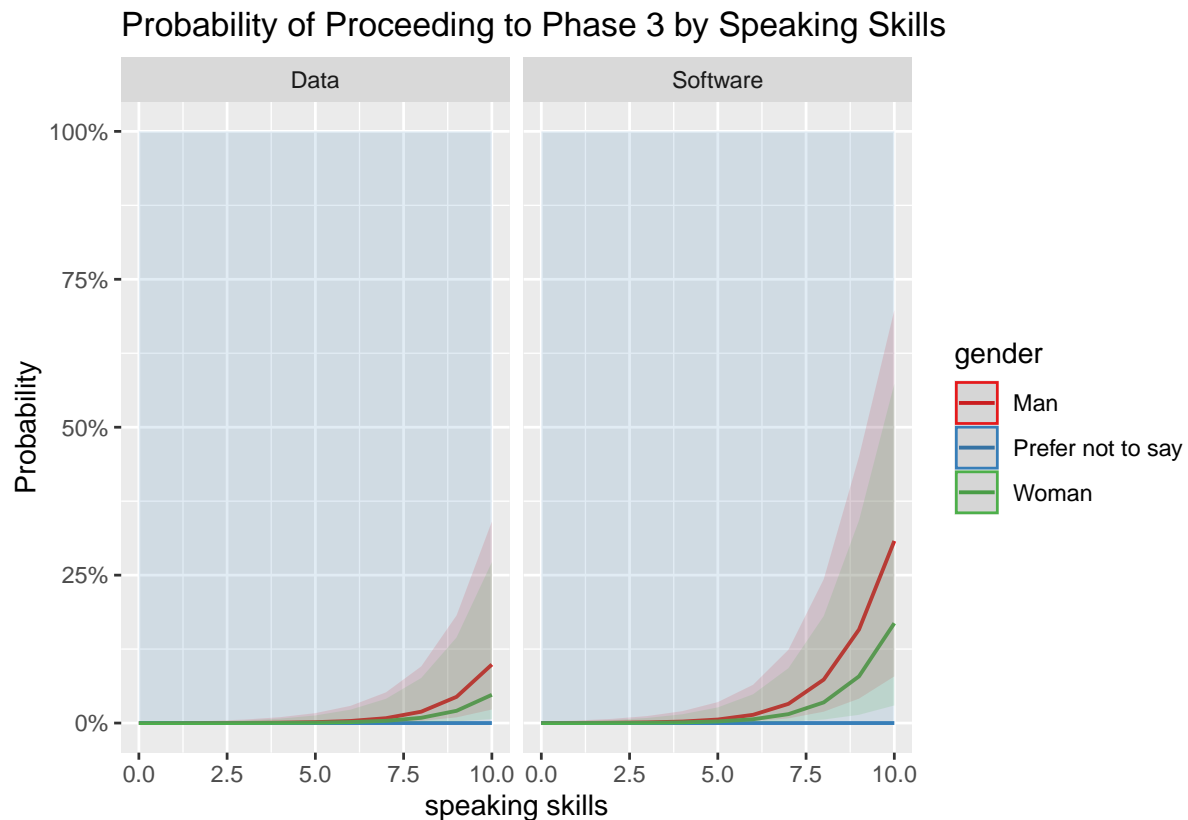












An important note to make is that none of the individuals in the 'prefer not to say' gender category did not proceed to phase 2 and variations can be due to a limited sample size, therefore they will be deemed as outliers in the interpretation between man, women and teams.

The analysis of the generalized linear model of whether an individual proceeded to phase 3 (response: pass_2) on the explanatory variables of gender(dummy variable for each category), team_applied_for(dummy variable for each category), technical_skills, writing_skills, leadership_presence, and speaking_skills showing 5 covariates as statistically significant at the 95% level or greater. These beta coefficients were: baseline (male in data team) with odds values equal to 0 with a 95% confidence interval between (0.0000 and 0.0000) and statistical significance of (pvalue = $2.0 * e^{-16}$), technical skills with odds values equal to 0.5231 with a 95% confidence interval between (0.8363 and 0.9230) and statistical significance of (pvalue = $2.0 * e^{-16}$), writing skills with odds values equal to 0.5257 with a 95% confidence interval between (0.5141 and 0.5405) and statistical significance of (pvalue = 0.0001), leadership presence with odds values equal to 0.7246 with a 95% confidence interval between (0.6438 and 0.8128) and statistical significance of (pvalue = $2.0 * e^{-16}$), and speaking skills with odds values equal to 0.7031 with a 95% confidence interval between (0.6231 and 0.7908) and statistical

significance of ($pvalue = 2.0 * e^{-16}$). The baseline covariate is for a male individual saying that male individuals have odds of 0 which is statistically significant due to low pvalues, while females have odds of 0.3134 which is not statistically significant due to high pvalues. This difference in odds can be interpreted as a potential non gender bias. The other covariates are factors that greatly affected the odds of proceeding to the next stage for an individual, where technical_skills, writing_skills, leadership_presence, speaking_skills. Furthermore, the covariate for team_applied_for was not statistically significant with a pvalue of 0.6737 suggesting that there does not exist a potential bias between company team recruitment. Based on the statistical analysis, Black Saber Software does not show a statistically significant bias towards gender in its phase 2 recruitment process, but should not exclude the possibility of a slight potential gender bias due to small percentage differences as shown in the following graphical interpretations.

Using the provided bar charts we can see the percentage of people that were able to proceed to phase 3, distributed by gender. In the bar chart named “Phase 2 Prediction - All” it contains data for individuals that applied to either team, data or software. The decision making shows a potential bias towards the male gender, as the percentage of men (5.2%) is about 2 times as high than that of woman (2.3%) who were able to proceed to the next phase. The meaning of these percentages is that there exists a minor potential male gender bias in the individuals who proceed from phase 2 to phase 3 in the Black Saber Software recruitment. The next two bar graphs, which are named “Phase 2 Prediction - Data team” and “Phase 2 Prediction - Software team” respectively, show the percentage of individuals separated by gender and team that proceeded to the next phase. In the case of those individuals that applied to a data team, the percentage of men that were able to proceed to the next stage was (5.6%), also about a 2 times higher percentage than that of women (2.4%). Lastly, for the software team, the percentage of men that were able to proceed to the next stage was (4.8%), also about a 2 times higher percentage than that of women (2.1%). This means that there could exist the possibility of a slight male gender bias. Furthermore, due to the support from the significance test on the covariates on team_applied_for we can suggest that differences between team may be due to sample variances.

Using the provided figures we can also see how each variable affects the probability of proceeding to the next phase. The first figure named “Probability of Proceeding to Phase 3 by Technical Skills”, shows the relationship of technical skills and probability of proceeding to phase 3. This figure shows a positive relationship with technical skills and probability, where higher technical skill ratings gives a higher probability of proceeding to the next phase. Furthermore, the distributions between man and woman is higher for men which suggests the possibility of a male gender bias in the AI process. In the figure named “Probability of Proceeding to Phase 3 by Writing Skills” shows the relationship of writing skills and

probability of proceeding to phase 3. This figure shows a positive relationship with writing skills and probability, where higher writing skill ratings gives a higher probability of proceeding to the next phase. Furthermore, the distributions between man and woman is higher for men which suggests the possibility of a male gender bias in the AI process. In the figure named “Probability of Proceeding to Phase 3 by Leadership Presence” shows the relationship of leadership presence and probability of proceeding to phase 3. This figure shows a positive relationship with leadership presence and probability, where higher leadership presence ratings gives a higher probability of proceeding to the next phase. Furthermore, the distributions between man and woman is higher for men which suggests the possibility of a male gender bias in the AI process. Lastly, the figure named “Probability of Proceeding to Phase 3 by Speaking Skills” shows the relationship of speaking skills and probability of proceeding to phase 3. This figure shows a positive relationship with speaking skills and probability, where higher speaking skill ratings gives a higher probability of proceeding to the next phase. Furthermore, the distributions between man and woman is higher for men which suggests the possibility of a male gender bias in the AI process. Variations between team are most likely due to variations in the sampling distribution.

Estimates and 95% Confidence Interval

Waiting for profiling to be done...

\begin{table}

\caption{Estimates and 95% Confidence Interval for p1_glm}

| | Estimate | 2.5% | 97.5% | pval |
|----------------------------------|----------|--------|--------|--------|
| (Intercept) | 0.0008 | 0.0002 | 0.0026 | 0.0000 |
| gpa | 0.8853 | 0.8363 | 0.9230 | 0.0000 |
| factor(team_applied_for)Software | 0.4780 | 0.3781 | 0.5798 | 0.6737 |
| extracurriculars | 0.5722 | 0.4750 | 0.6652 | 0.1466 |
| factor(work_experience)1 | 0.8183 | 0.6893 | 0.9105 | 0.0001 |
| factor(work_experience)2 | 0.7885 | 0.5802 | 0.9156 | 0.0118 |
| factor(gender)Prefer not to say | 0.3512 | 0.0836 | 0.7070 | 0.4531 |
| factor(gender)Woman | 0.4927 | 0.3933 | 0.5926 | 0.8868 |

\end{table}

\begin{table}

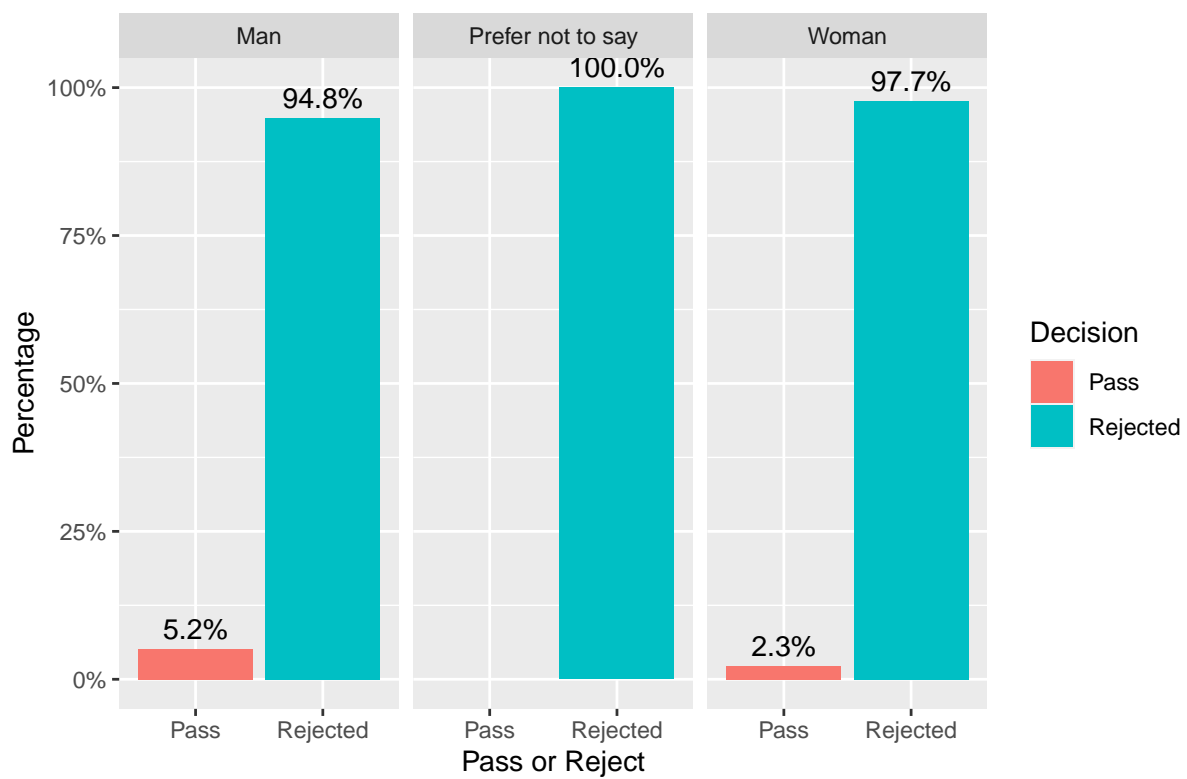
\caption{Estimates and 95% Confidence Interval for p3_glm}

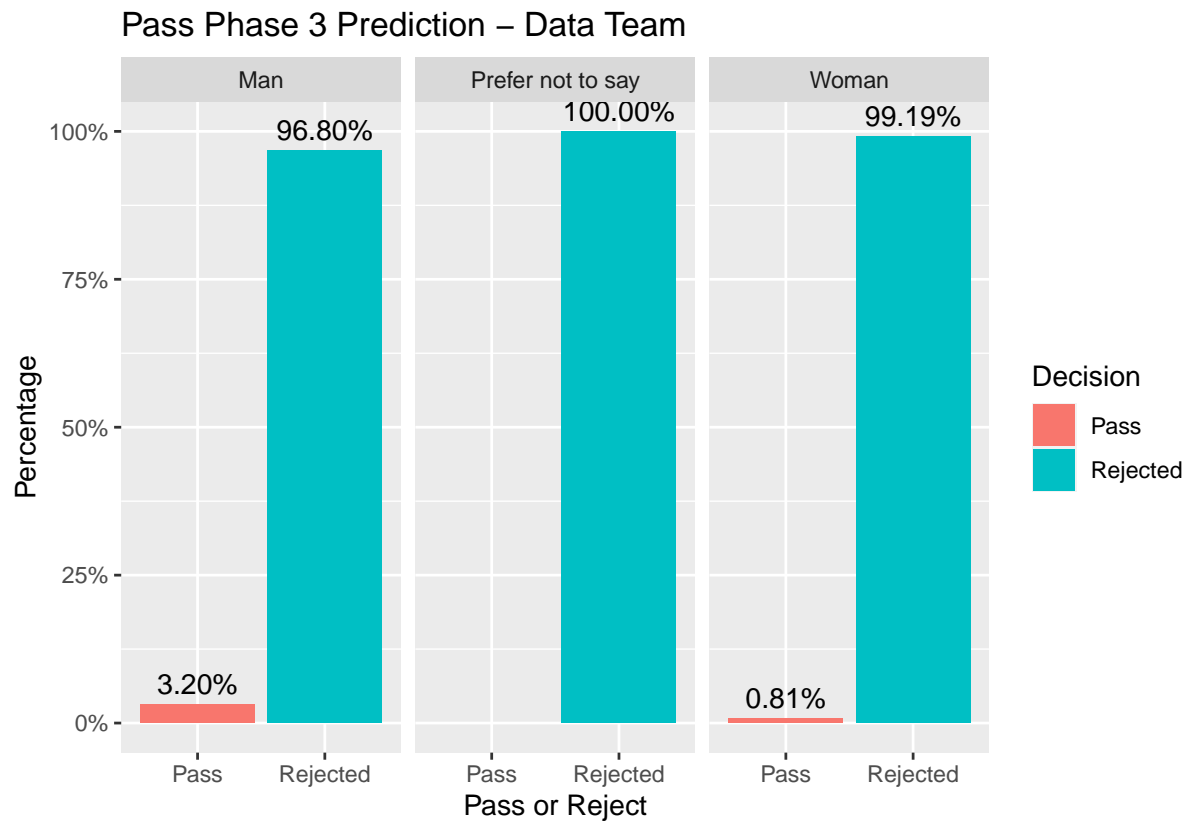
| | Estimate | 2.5% | 97.5% | pval |
|----------------------------------|----------|------|-------|--------|
| (Intercept) | 0 | 0 | 1961 | 0.0902 |
| factor(team_applied_for)Software | 2 | 0 | 101 | 0.8428 |
| rating | 2 | 1 | 4 | 0.0872 |
| factor(gender)Woman | 0 | 0 | 5 | 0.1900 |

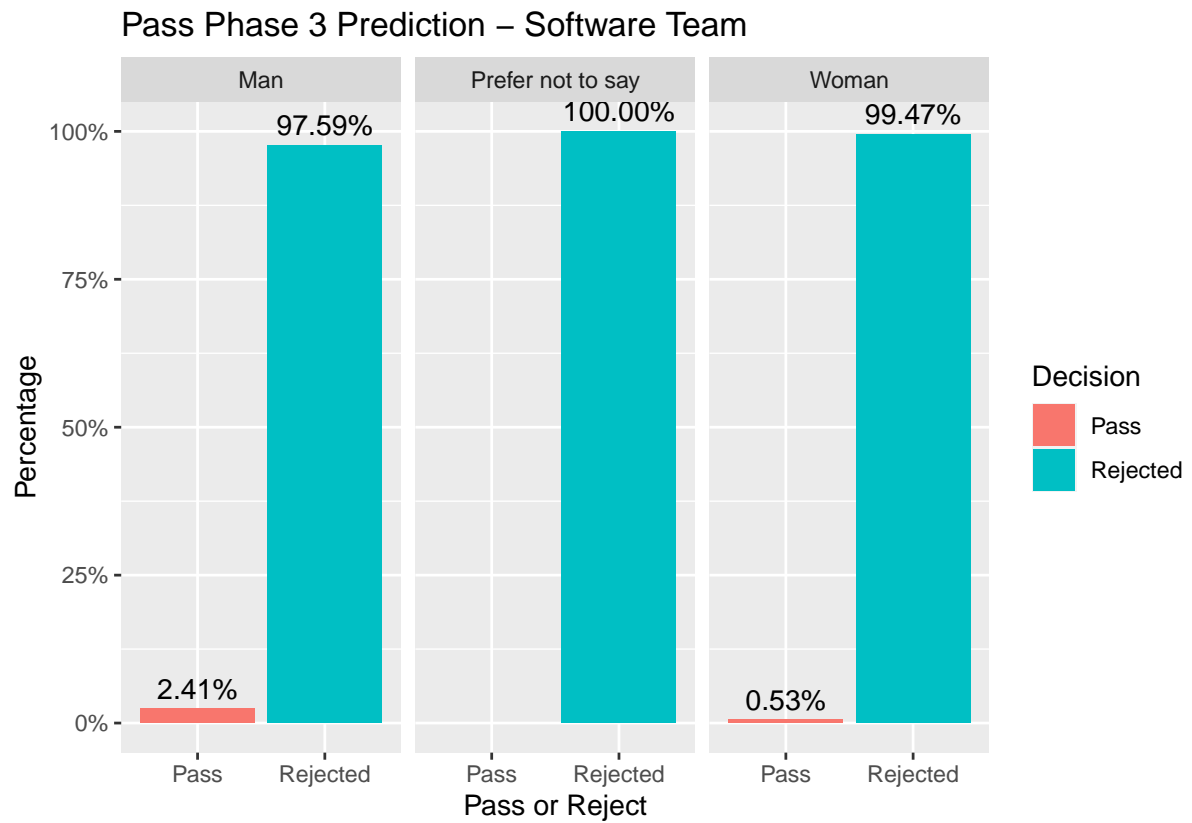
\end{table}

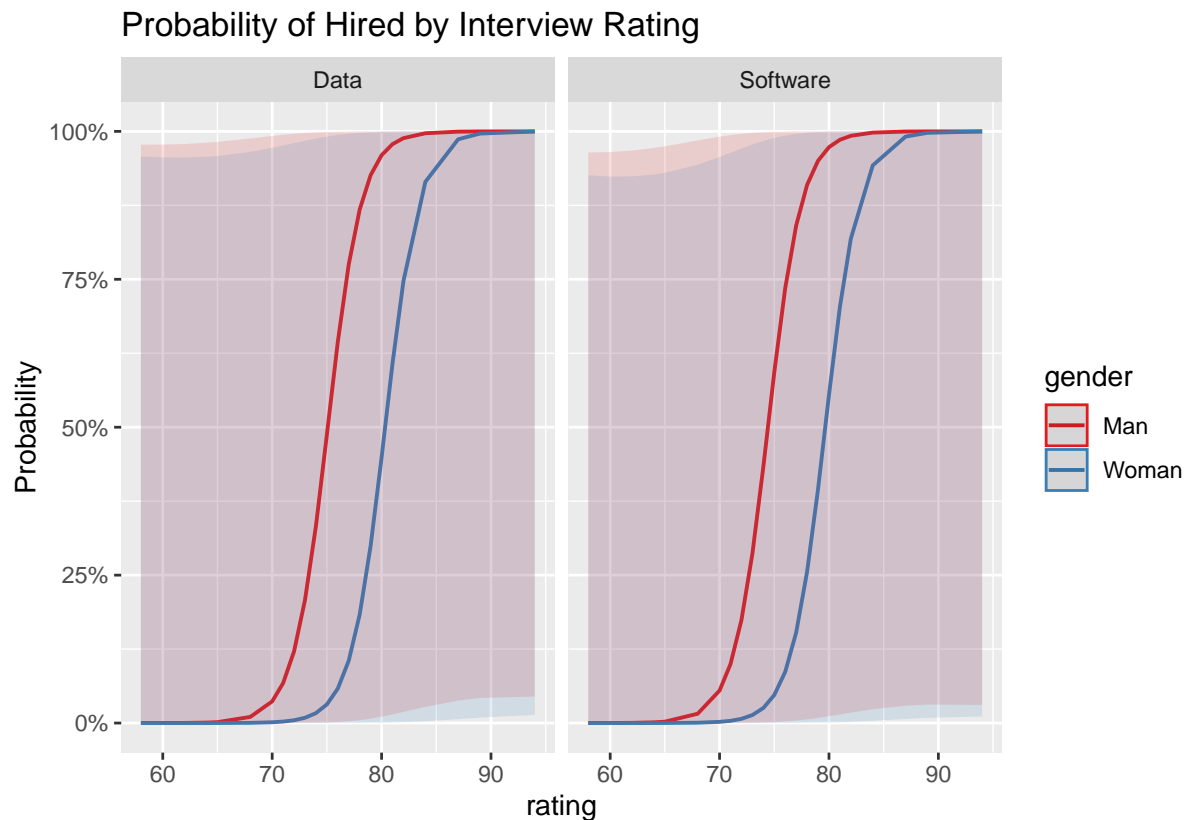
Pass Phase 3 - Graphs and Figures

Pass Phase 3 Prediction – All









Result explanation:

An important note to make is that non of the individuals in the the prefer not to say gender category did not proceed to phase 3 and variations can be due to a limited sample size, therefore they will be deemed as outliers in the interpretation between man, women and teams.

The analysis of the generalized linear mixed model of whether an individual was hired (response: pass_3) on the explanatory variables of applicant_id(random effect), team_applied_for(dummy variable for each category), rating, and gender(dummy variable for each category) showing 0 covariates as statistically significant at the 95% level or greater. The beta coefficients were: baseline(male in data team) with odds values equal to 0 with a 95% confidence interval between (0 and 1961) and statistical significance of (pvalue = 0.0902), software teams with odds values equal to 2 with a 95% confidence interval between (0 and 101) and statistical significance of(pvalue = 0.8428), rating with odds values equal to 2 with a 95% confidence interval between (1 and 4) and statistical significance of(pvalue = 0.0872), and woman dummy variable with odds values equal to 0.0 with a 95% confidence interval between (0 and 5) and statistical significance of(pvalue = 0.1900).Due to small sample size, the interpretation of the covariates are slightly inaccurate. The baseline(male in the data team) covariate is for a male individual saying that male individuals have odds of 0 which is not

statistically significant due to high pvalues, while females have odds of 0 which is not statistically significant due to high pvalues. This difference in odds can be interpreted as a potential non gender bias. The covariate on rating also shows a positive relationship with the odds of the probability of getting hired. Based on the statistical analysis, Black Saber Software does not show a statistically significant bias towards gender in its phase 3 recruitment process, but should not exclude the possibility of a slight potential gender bias due to small percentage differences as shown in the following graphical interpretations.

Using the provided bar charts we can see the percentage of people that were able to proceed to phase 3, distributed by gender. In the bar chart named “Phase 3 Prediction - All” it contains data for individuals that applied to either team, data or software. The decision making shows a potential bias towards the male gender, as the percentage of men(5.2%) is about 2 times as high than that of woman(2.3%) who were able to proceed to the next phase. The meaning of these percentages is that there exists a minor potential male gender bias in the individuals who proceed from phase 3 to hired in the Black Saber Software recruitment. The next two bar graphs, which are named “Phase 3 Prediction - Data team” and “Phase 3 Prediction - Software team” respectively, show the percentage of individuals separated by gender and team that proceeded to the next phase. In the case of those individuals that applied to a data team, the percentage of men that were able to proceed to the next stage was (3.2%), about a 3 times higher percentage than that of women (0.81%). Lastly, for the software team, the percentage of men that were able to proceed to the next stage was (2.41%), also about a 4 times higher percentage than that of women(0.53%). This means that there could exist the possibility of a slight male gender bias. Furthermore, due to the support from the significance test on the covariates on team_applied_for we can suggest that differences between team may be due to sample variances.

Using the provided figures we can also see how each variable affects the probability of proceeding to the hired phase. The figure named “Probability of Hired by Interview Rating”, shows the relationship of interview rating and probability of proceeding to the hired phase. This figure shows a positive relationship with interview rating and probability, where higher interview rating ratings gives a higher probability of proceeding to the next phase. Furthermore, the distributions between man and woman is lower for men which suggests the possibility of a male gender bias in the interview process. This is due to the fact that male individuals require a lower average interview rating to be hired as compared with the female counterpart. Furthermore, differences across team do not shows any visual different suggesting that there does not exist a bias between teams.

Summary of All Phase Model Accuracy Result explanation:

The table shown describes the model prediction accuraccy for each respective phase. The first

Table 3: Accuracies of models

| Model Phase | Model Prediction Accuracy |
|--------------------|---------------------------|
| Phase 1 to Phase 2 | 76.67 |
| Phase 2 to Phase 3 | 94.33 |
| Phase 3 to Hired | 100.00 |

model, “p1_glm”, used to describe the generalized linear model relationship between phase 1 to phase 2 shows a prediction accuracy of about 76.84%, meaning that the model was able to correctly predict 76.84% of the individuals that were able to proceed or not to phase 2 from phase 1. The second model, “p2_glm”, used to describe the generalized linear model relationship between phase 2 to phase 3 shows a prediction accuracy of about 94.33%, meaning that the model was able to correctly predict 94.33% of the individuals that were able to proceed or not to phase 3 from phase 2. The third and last model, “p3_glm”, used to describe the generalized linear mixed model relationship between phase 3 to hired shows a prediction accuracy of about 100%, meaning that the model was able to correctly predict 100% of the individuals that were hired from phase 3.

Does the Determination of Salaries Have Any Gender Biases?

Salary Process

Data Description

Data set: Current Employees

Description: Complete form of all 607 current employees of Black Saber Software for the year 2021

Variables collected:

- employee_id: qualitative variable - Numerical ID assigned to employee
- gender: qualitative variable - man, woman, or prefer not to say
- team: qualitative variable - category for each of the 8 teams in Black Saber Software
- financial_q: quantitative variable - Financial quarter salary, leadership and productivity
- role_seniority: qualitative variable - Role status in Black Saber Software
- leadership_for_level: qualitative variable - Description of individual’s leadership qualities
- productivity: qualitative variable - Numerical indicator rated on a 0-100 scale on expected productivity

- salary: quantitative variable - Salary of individual at a given financial quarter

Model

Regression Model and Analysis

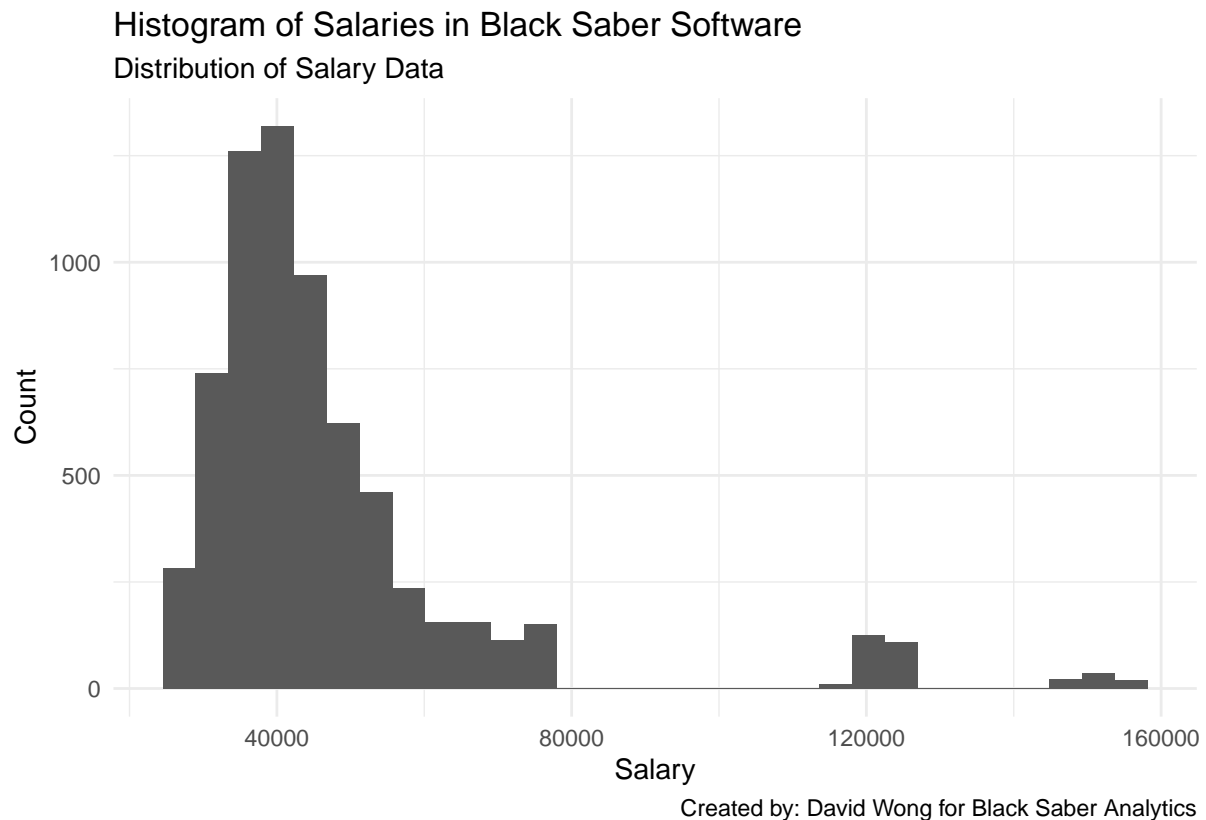
Using a generalized linear mixed effects model, we applied a gaussian family link with the regression model as follows:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{\text{gender } i} + \beta_2 x_{\text{team } i} + \beta_3 x_{\text{leadership } i} + \beta_4 x_{\text{seniority } i} + \beta_5 x_{\text{promoted } i} + \beta_6 x_{\text{productivity } i} + \beta_7 x_{\text{employee}}$$

where $i = 1, 2, \dots, n$ and p = Predicted Salary

When deciding the link function, we consider the following things:

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

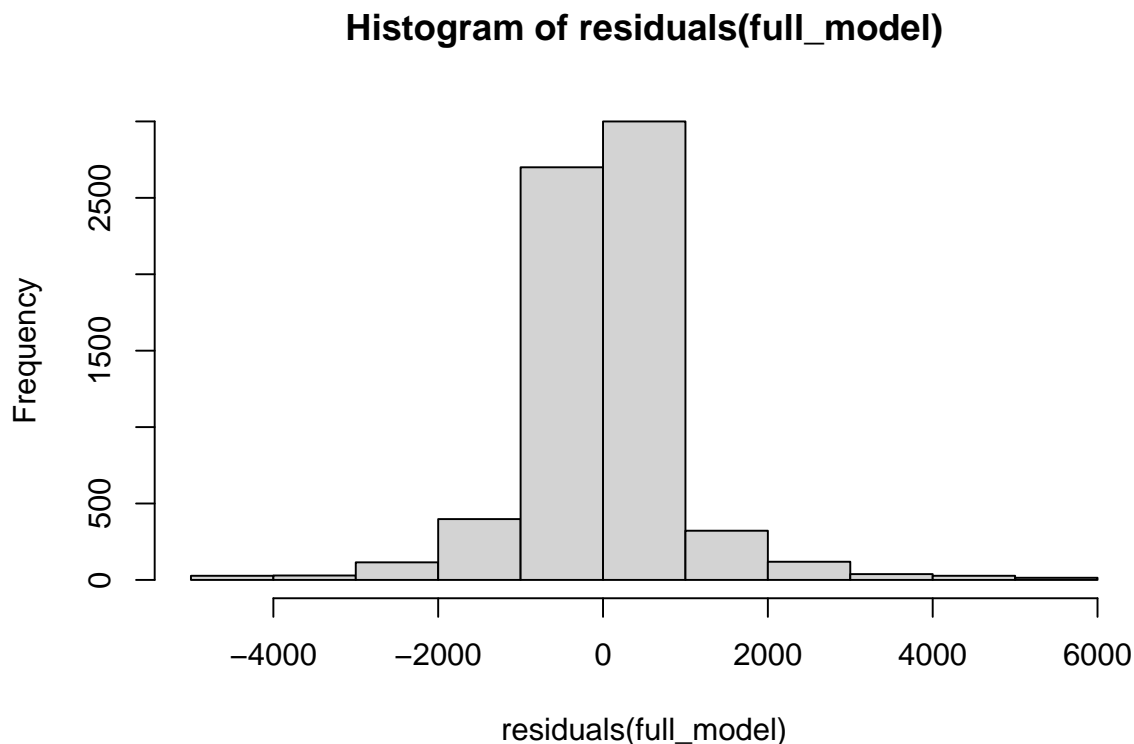


Knowledge of the response

First, we plot the distribution of Salaries into a histogram to understand the relative shape of all of the data points. From (figure), the plot is heavily right skewed as there exists a large gap of salaries on the right tail of the plot. This is likely because presidents and other C-level executives get paid significantly higher than the rest of employees. Thus, we should expect our model prediction outputs to reflect a similar pattern.

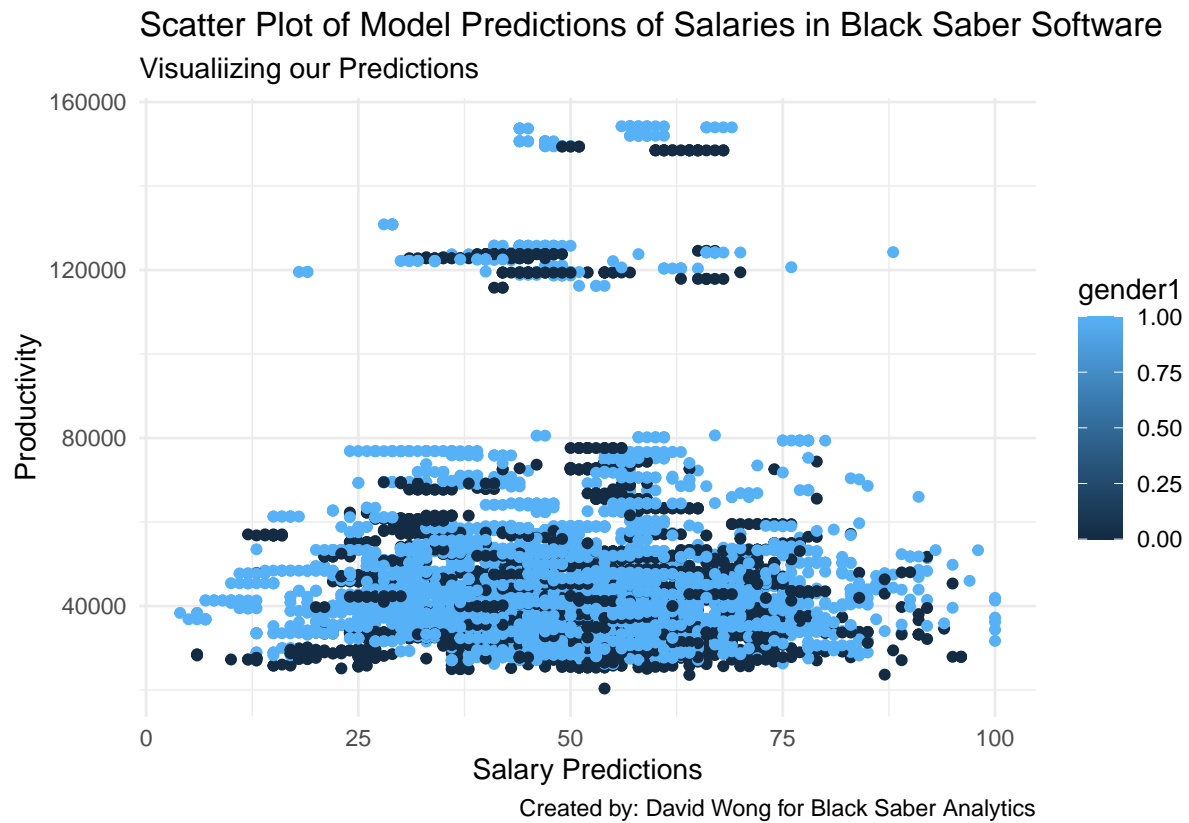
Table 4: Estimates, Std Error, and T-Value for Full Model

| | Estimate | Std. Error | T-val |
|--|--------------|------------|--------------|
| (Intercept) | 64680.37709 | 322.120392 | 200.7956611 |
| factor(team)Data | 2359.08575 | 465.643992 | 5.0662862 |
| factor(team)Design | -227.52502 | 834.961818 | -0.2724975 |
| factor(team)Legal and financial | 2837.66159 | 654.208928 | 4.3375464 |
| factor(team)Marketing and sales | 730.81526 | 424.260781 | 1.7225614 |
| factor(team)Operations | 484.33883 | 459.661593 | 1.0536857 |
| factor(team)People and talent | -1410.92253 | 680.934239 | -2.0720393 |
| factor(team)Software | 2757.87369 | 411.510282 | 6.7018342 |
| leadership_for_levelExceeds expectations | -25.26866 | 71.630815 | -0.3527624 |
| leadership_for_levelNeeds improvement | 242.62678 | 95.867084 | 2.5308664 |
| role_seniority.L | 105405.51393 | 169.670486 | 621.2365882 |
| role_seniority.Q | 50216.70287 | 123.831331 | 405.5250202 |
| role_seniority.C | 18139.85318 | 94.211976 | 192.5429647 |
| role_seniority ⁴ | -4019.19843 | 77.437387 | -51.9025577 |
| role_seniority ⁵ | -10484.08496 | 65.533606 | -159.9802840 |
| role_seniority ⁶ | -9656.19392 | 57.676247 | -167.4206345 |
| role_seniority ⁷ | -2912.14828 | 49.902670 | -58.3565627 |
| role_seniority ⁸ | -847.12113 | 41.690467 | -20.3193007 |
| promoted | 129.96079 | 53.744541 | 2.4181207 |
| productivity | -2.65004 | 1.353647 | -1.9577031 |
| gender1 | 2226.34739 | 277.530722 | 8.0219853 |



Theoretical considerations Although a Gamma fit is more appropriate for temporal data, it is appropriate to apply a Gamma link when using the “Year” variable in the regression. However, because we had modelled this variable into “q_worked”, which gives a count for the quarters worked of an individual in their position title which resets upon promotion, now becomes a variable that loses its temporal factor but now is a variable for tenureship. Furthermore, if we graph the histogram of the residual plots for Salary, we can see that it follows a very rough normal distribution. Because of this, residuals are consistent with your assumption of normality, but they don’t “validate” the assumption.

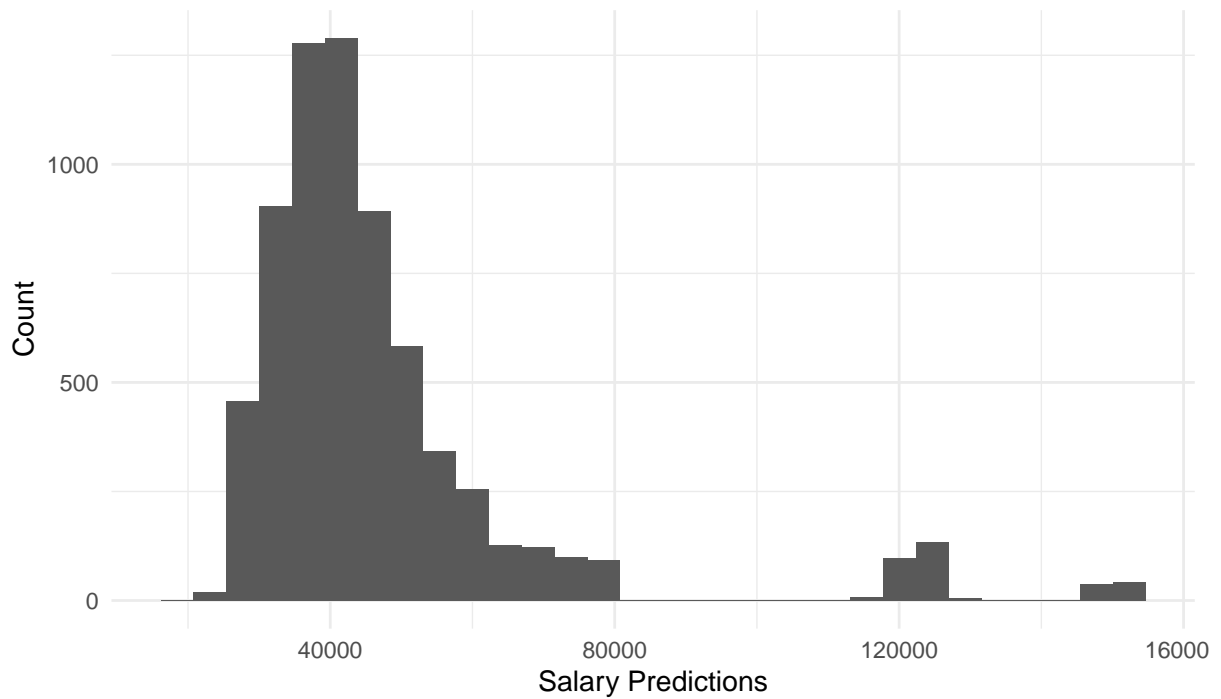
Empirical fit to the data When assessing the predictability and the output of the regression model for the predicted salaries, we see the same trends as displayed in the real salary histogram. There is a large gap between those earning under \$80,000 and those earning more than \$110,000. Furthermore, the concentration of prediction values overall in both ends of this split are roughly the same as the real data for Salaries. Thus, we can say that a Gaussian link is appropriate for our data.



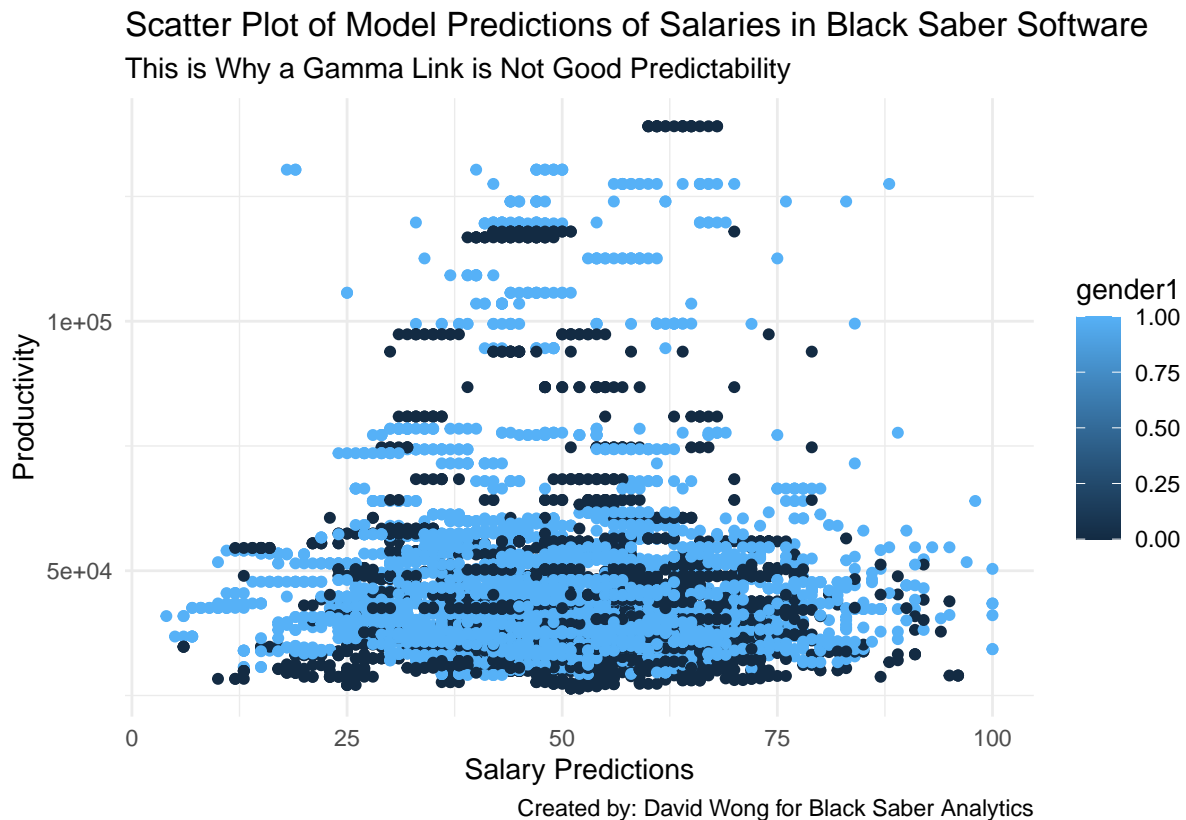
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Histogram of Model Predictions of Salaries in Black Saber Software

Visualiizing our Predictions



Created by: David Wong for Black Saber Analytics



What happens when we apply a Gamma model link instead? We attempted to apply a Gamma model as per the recommendation by our affiliates, but we often encountered convergence errors and computational errors by the linear algebra involved in computing the regression. However, we were able to bypass these issues by reducing our model variables significantly, which took away from the predictive power of the model. Since there were too many large compromises to both attempt to get p-values from our model outputs while also following the recommendations from our affiliates, we thought our justification for using a Gaussian link was strong enough.

Results

Our model did not produce any p-values, we are only able to discuss the t-values computed by our regression model. A high absolute value of a t-value implies that it likely has a significant p-value, meaning that team, role seniority, gender, promoted, and productivity would likely display significant p-values in from most significant to least significant in the respective order. This implies that although we cannot make a distinct and direct connection between gender and salary, we are able to say that out of all the regression variables defined in the model, gender plays only a moderate significance in the determination of compensation levels for

Table 5: Likelihood-Ratio Test for Model 1 and Model 2

| Model | Log(likelihood) | Df | χ^2 | Pvalue | #Df | df | AIC |
|---------|-----------------|-----------|----------|----------|-----------|----|----------|
| Model 1 | 15 | -1467.684 | NA | NA | NA | 15 | 2965.369 |
| Model 2 | 8 | -1469.655 | -7 | 3.941175 | 0.7865274 | 8 | 2955.310 |

Black Saber Software. This implies that even if gender plays a significant effect in the determination of the salary of an employee, it will likely be one of the less important factors in this area meaning this interpretation for bias may be open to other determining confounders in our model.

Does gender bias affect promotions in Black Saber Software, either in specific teams or company as a whole?

Promotion Process

Methods

The source of the data for our model comes from the current employees data provided by Black Saber Software. A total of 607 current employees were included and none were excluded from the current employee data set.

Our model investigates the association between gender and getting a promotion for 607 current employees at Black Saber Software. Work output in relation to the employee's job description was rated on a 0-100 scale, where 50 is satisfactory and above 50 indicates a better than expected productivity. Quality of demonstrated leadership, taking into account the employee's role level, was measured on three levels: Appropriate for level, Needs Improvement and Exceeds Expectations. If an employee was promoted to a senior role in the next quarter, their promoted status was 1 and otherwise, 0.

Two likelihood-ratio tests were conducted to test for the most appropriate model between Model 1:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{leadership\ for\ level, i} + \beta_2 x_{gender, i} + \beta_3 x_{worked\ in\ rs, i} + \beta_4 x_{gender, i} + \beta_5 x_{team, i} + \beta_{employee\ id, i} + \epsilon_i$$

where $i = 1, 2, \dots, n$ and p = probability of being promoted

and Model 2:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{leadershipforlevel,i} + \beta_2 x_{gender,i} + \beta_3 x_{qworkedinrs,i} + \beta_4 x_{gender,i} + \beta_5 x_{employeeid} + \epsilon_i$$

where $i = 1, 2, \dots, n$ and p = probability of being promoted

and it showed that model 2 (model without the inclusion of Team) was a better fit due to a high p-value (p value= 3.941175) suggesting that the simpler model explained the data just as well as the more complicated model. We want to verify that Model 2 without the inclusion of factor(team) can be used to answer the question of whether gender bias occurs in individual teams or the company as a whole.

Model 2's outcome of interest was whether the employee got promoted or not in a given quarter and our predictors were their gender, productivity, demonstrated leadership quality, and the number of financial quarters worked in a particular role. As there were repeated measures for each employee, a generalized linear mixed model with logit link was used and random intercepts were estimated for each employee. The employee ID was treated as a random effect. All other variables, such as productivity score, were considered as fixed effects.

Results

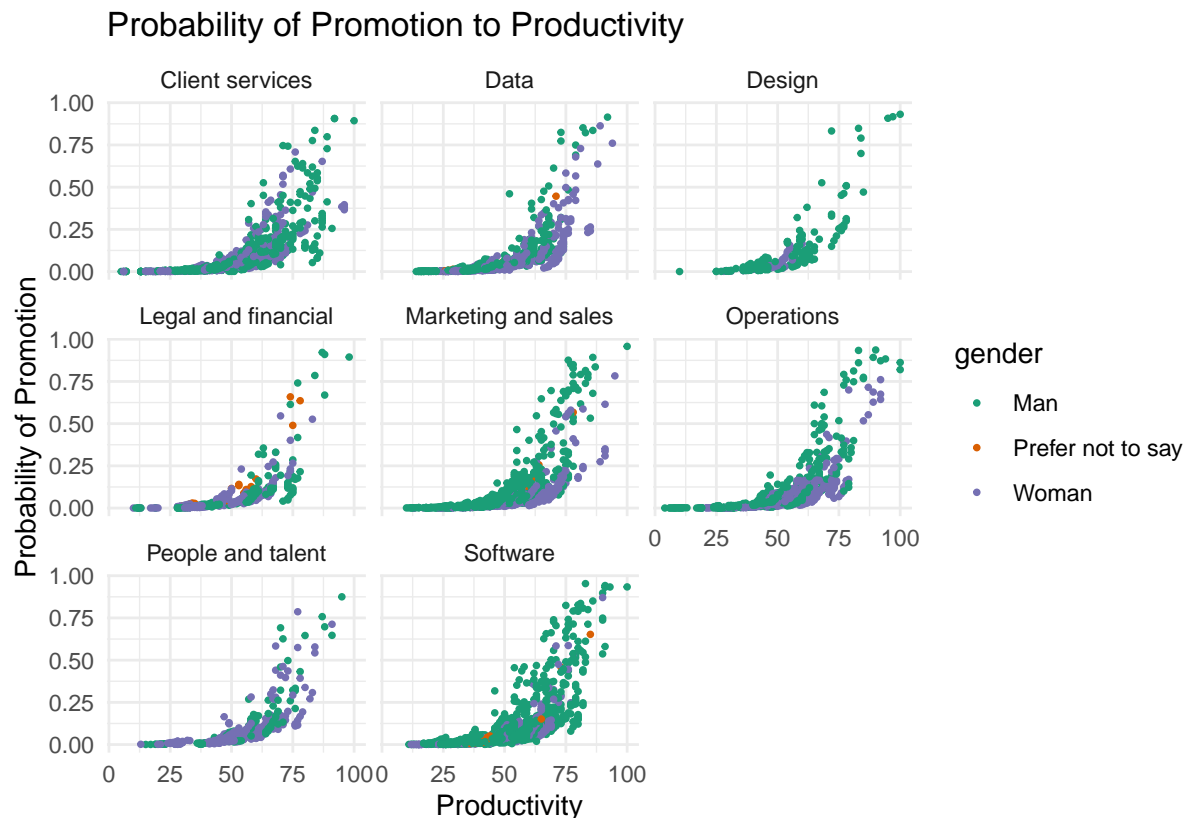
```
## [1] "The percentage of correct estimates is 0.941355"
```

```
##           Model Prediction
## Promoted    0    1   Sum
##         0   6316   21 6337
##         1    384  185  569
##         Sum 6700  206 6906
```

Looking at the contingency table above allows us to examine the performance of our model. Out of the total 6337 observations where there were no promotions, there were 21 false positives, representing instances where the model predicted an employee's factors in the specific quarter would lead to a promotion when in fact no promotion was given. Out of the 569 promotions that occurred, the model successfully predicted 185 instances while failing to predict 384 promotions. Overall, the model predictions aligned with the actual results 94.14% of the time.

| ## | Model | Intercept | Exceeds Expectations | Needs Improvement | Productivity |
|------|-------------------------|-----------------|----------------------|-------------------|--------------|
| ## 1 | CS | -8.3542 | 0 | 0 | 0.09473 |
| ## 2 | Data | -9.30532 | 2.18031 | 0 | 0.11893 |
| ## 3 | Design | -6.51526 | 0 | 0 | 0.0716 |
| ## 4 | Legal | -13.0911 | 1.71492 | 0 | 0.14157 |
| ## 5 | Marketing | -12.44602 | 1.03407 | 0 | 0.1618 |
| ## 6 | Operations | -8.22598 | 1.22224 | 0 | 0.1006 |
| ## 7 | Software | -8.290619 | 0.861776 | 0 | 0.102532 |
| ## 8 | Talent | -12.14544 | 0 | 0 | 0.15233 |
| ## 9 | Full | -9.30972 | 1.01017 | 0 | 0.11458 |
| ## | Woman Prefer Not to Say | Quarters Worked | in Role | | |
| ## 1 | 0 | 0 | 0.09286 | | |
| ## 2 | 0 | 0 | 0 | | |
| ## 3 | 0 | 0 | 0 | | |
| ## 4 | 0 | 3.20391 | 0.20923 | | |
| ## 5 | -1.59022 | 0 | 0.15336 | | |
| ## 6 | 0 | 0 | 0 | | |
| ## 7 | 0 | 0 | 0 | | |
| ## 8 | 0 | 0 | 0.17236 | | |
| ## 9 | -0.42821 | 0 | 0.06371 | | |

This table shows the statistically significant estimates of the different models we have constructed, where non-significant estimates are replaced with 0. Since productivity was significant in all models, let's take a look at how productivity affects an employee's probability of promotion in your company visually.



We can see that for across all teams, although having a high productivity score does not guarantee a promotion (especially in Client services team), all employees who received promotions in the model all had productivity scores above 50. It is good to see that your company is generally rewarding employees who are highly productive, regardless of gender.

Since the relationship between productivity and probability of promotion is consistent across teams, and none of the team estimates in the initial model is significant, we can consider dropping team as a fixed effect in the initial full model.

After removing “team” as a mixed model we ran a lr test which was nonsignificant meaning the simpler describes the data just as well as the more complex model. We ran an AIC test and model the simpler model without “team” had a lower AIC test score of 2955.31 compared to the more complex model with 2965.369. The AIC test also suggests that the simpler model should be chosen over the initial model. We compared an magnitude of different models and ran LR and AIC test to determine the most appropriate model and we found that the best model was the model with the employees leadership for level, gender, productivity score and the number of quarters they have spent in their role as fixed effects and employee ID as a random effect.

We compared the prediction accuracy and compared the contingency tables for the full model

and the optimal chosen model, although the chosen model was a simpler model with less explanatory variables its performance was the same compared to the full model. This solidifies our idea that team is not significant in effect one's odds of promotion, so going forward we will be analyzing gender bias of promotion as in your company as a whole.

```
coef_table <- as.data.frame(summary(full_model_2)$coef)
coef_table
```

| ## | Estimate | Std. Error | z value |
|---|---------------|-------------|-------------|
| ## (Intercept) | -9.30972257 | 0.313004632 | -29.7430824 |
| ## leadership_for_levelExceeds expectations | 1.01017095 | 0.180144654 | 5.6075544 |
| ## leadership_for_levelNeeds improvement | -0.34688810 | 1.024311344 | -0.3386549 |
| ## productivity | 0.11457831 | 0.004889732 | 23.4324310 |
| ## factor(gender)Prefer not to say | 0.41311281 | 0.555843546 | 0.7432178 |
| ## factor(gender)Woman | -0.42820935 | 0.161491882 | -2.6515844 |
| ## q_worked_in_rs | 0.06370857 | 0.014545866 | 4.3798404 |
| ## | Pr(> z) | | |
| ## (Intercept) | 2.130842e-194 | | |
| ## leadership_for_levelExceeds expectations | 2.052055e-08 | | |
| ## leadership_for_levelNeeds improvement | 7.348697e-01 | | |
| ## productivity | 1.997028e-121 | | |
| ## factor(gender)Prefer not to say | 4.573498e-01 | | |
| ## factor(gender)Woman | 8.011509e-03 | | |
| ## q_worked_in_rs | 1.187663e-05 | | |

The model estimates that a woman whose leadership skill is appropriate for level has 1.53% lower odds than a man whose leadership skill is also appropriate or level. This corresponds to 0.3946% reduction in probability of being promoted compared to a man, all else equal.

A women who's leadership score is appropriate for level would need a higher productivity score of 4, to achieve the same odds of being promoted, compared to a man who's leadership skill is also appropriate for level.

The odds of a man being promoted in a quarter where his leadership level was appropriate for the current level is 9.05×10^{-5} . The base line in the model is a man who's leadership level in the given quarter was appropriate for his current level. Employees who exceeded expectations for their current level had 2.75% higher odds of being promoted. For each one rating increase in productivity the odds of promoted is 1.12% higher. According to the model, a women has 0.65% lower odds of being promoted and for each additional quarter spend in the current role led to the odds of promotion to be 1.5% higher.

Confidence intervals

| ## | 2.5 % | 97.5 % |
|---|--------------|---------------|
| ## .sigma | NA | NA |
| ## (Intercept) | 64049.03273 | 6.531172e+04 |
| ## factor(team)Data | 1446.44030 | 3.271731e+03 |
| ## factor(team)Design | -1864.02012 | 1.408970e+03 |
| ## factor(team)Legal and financial | 1555.43565 | 4.119888e+03 |
| ## factor(team)Marketing and sales | -100.72059 | 1.562351e+03 |
| ## factor(team)Operations | -416.58134 | 1.385259e+03 |
| ## factor(team)People and talent | -2745.52912 | -7.631595e+01 |
| ## factor(team)Software | 1951.32836 | 3.564419e+03 |
| ## leadership_for_levelExceeds expectations | -165.66248 | 1.151252e+02 |
| ## leadership_for_levelNeeds improvement | 54.73075 | 4.305228e+02 |
| ## role_seniority.L | 105072.96589 | 1.057381e+05 |
| ## role_seniority.Q | 49973.99792 | 5.045941e+04 |
| ## role_seniority.C | 17955.20110 | 1.832451e+04 |
| ## role_seniority^4 | -4170.97291 | -3.867424e+03 |
| ## role_seniority^5 | -10612.52847 | -1.035564e+04 |
| ## role_seniority^6 | -9769.23729 | -9.543151e+03 |
| ## role_seniority^7 | -3009.95571 | -2.814341e+03 |
| ## role_seniority^8 | -928.83295 | -7.654093e+02 |
| ## promoted | 24.62342 | 2.352982e+02 |
| ## productivity | -5.30314 | 3.060466e-03 |
| ## gender1 | 1682.39717 | 2.770298e+03 |

Statement/Conclusions From our analysis we see that although there is a bias for men be promoted over women throughout your company, the bias effect is small and is mainly driven by productivity. It is good to see that promotion is mainly driven by productivity, but the small bias is not negligible because in the future as you want your company to promote equal opportunity for all.

Discussion

From each of our research questions regarding possible biases existing in the hiring, promotion and remuneration processes for Black Saber Software. We concluded that there exists a difference in the hiring, promotion, and remuneration rates between men and women; however, there seems to be weak evidence to claim that there exists a bias. Despite some of our models finding that there exists perhaps a small bias in these areas for Black Saber Software, other factors such as productivity holds a much stronger deterministic factor when considering hiring and remuneration rates. This holds consistent with previous research and their findings. In our base article, researchers concluded that worker-specific productivity differences may be a primary reason for gendered promotion rates and that remuneration rates reflect the productivity of work from an individual, where gendered effects ended up being negligibly small relative to other contributing effects (Jokinen et al., 2017). For Black Saber Software, evidence shown through their employee and hiring datasets tells a consistent story when compared to prior research.

In our first research question that aimed to solve if the AI algorithm was biased in its rating of candidates for the hiring process. This is an essential component to assess to ensure that hiring systems are equitable and fair for every applicant, otherwise, a biased algorithm results in possible discriminatory effects against candidates that do not match its perceived ideal profile. When assessing the regression model outputs, it displays insignificant p-values where both gendered effects in their application responses did not contribute to the success of the candidate's application.

Similarly, in our 2nd research question that aimed to find if salaries were biased with gendered effects. In our analysis of the model outputs, we concluded that there exists either no or minimal gendered effects in the determination of worker salaries within Black Saber Software's current compensation models. There are other effects in our prediction parameters that contribute to salary such as team and role seniority in the company. This ties into our next model that aimed to determine if there are biased effects in the promotions of individuals within the company.

Lastly, in our 3rd research question, we found that although there exists a very small gendered effect in the probability to be promoted between men and women; this probability only amounts to a difference of 0.3946%. This difference is mostly attributed to the difference in productivity for an individual in the company, which further builds upon Jokinen's conclusions from their 2017 paper.

Strengths and limitations

While our results do reflect strong historical and concurrent evidence suggesting that there exist no biases in the hiring, promotion, and salary processes for Black Saber Software, there are considerations and critiques to be made along our experimental process. For our hiring data sets, sample sizes especially for later rounds of the hiring process reduced down to as low as 22 data points. This implies that sample sizes in our analysis overall might have been insufficient to perform any form of robust analysis on gendered effects. Furthermore, because this hiring data is based on a specific and recent time frame for Black Saber Software's hiring processes, analyzing past years data for Black Saber Software would improve our conclusions in identifying hiring biases. Furthermore, our data modelling procedures when defining new variables such as "q_worked" in the internal employee dataset, would result in a slight change in interpretation of our parameters. Rather than having our parameters, "year" and "quarter", base itself temporally on the time the individual was promoted, now it is based off of a "tenureship" like variable called "q_worked". This variable, "q_worked" is arguably a stronger variable since now it holds stronger significance in its deterministic effect for an employee. Similar to the parameters chosen in Jokinen's paper, tenureship with the university proved to be one parameter of justification for salaries in their study. This effect is reflected similarly in our regression outputs, where in our 3rd research question regarding promotions, tenureship remained to be a less significant factor than the productivity of a worker when considering promotions. Thus, despite caveats in the ways we defined certain parameters, the parameters that most interest us reflected the same conclusions as did previous studies we initially observed.

Our modelling and prediction outputs resulted in strong accuracy. In our analysis of our first research question tackling gender biases in the hiring process, our first model produced a 77% prediction accuracy, which is relatively low but considering that we had a poor sample size overall, having a more robust data set might improve the prediction accuracy. However, for the 2nd and 3rd models for the first research question, we had over 90% prediction accuracy for these subsequent models. Furthermore, in our 3rd research question that we tackled regarding gender biases in promotions, we had a prediction accuracy of over 90% as well. Hence, this implies that our model parameters constructed a strong framework for predicting candidate and employee profiles when considering factors like hiring, promotions, salary. For future developments in the field of researching biases in the hiring, promotions and salary factors for Black Saber Software, would require a wider scope of data where we can analyze other covariates and confounders not included in our analysis. As per the suggestion from the project proposal, ethnicity, race and other socioeconomic variables play a contributing factor in having a strong EDI framework. In order to be equitable, fair, and inclusive in Black Saber

Software, developments to work towards developing a more robust analysis of bias would be helpful to Black Saber Software executives. EDI principles are not exclusive to only the parameters defined in our dataset and dialogues revolving around EDI factors are constantly evolving. It is crucial for companies like Black Saber Software to maintain their values and commitment towards approaching a workplace that is equitable, diverse, and inclusive.

Consultant information

Consultant profiles

Jing Yuan Zhang. Jing Yuan is a senior consultant with Saber Analytics. She specializes in data visualization. Jing Yuan earned her Bachelor of Science, Double Major in Statistics and Mathematics, from the University of Toronto in 2022.

Howard Wang. Howard is a senior consultant with Saber Analytics. He specializes in data visualization. Howard earned his Bachelor of Science, Double Major in Statistics and Economics, from the University of Toronto in 2022.

David Wong. David is a senior consultant with Saber Analytics. He specializes in data visualization. David earned his Bachelor of Science, Major in Economics and double minor in GIS and statistics, from the University of Toronto in 2021

Muhammad Tsany. Muhammad is a senior consultant with Saber Analytics. He specializes in data visualization. Muhammad earned his Bachelor of Science, Double Major in Statistics and Economics, from the University of Toronto in 2022.

Sergio Steven Zheng Zhou. Sergio is a senior consultant with Saber Analytics. He specializes in data visualization. Sergio earned his Bachelor of Science, Double Major in Statistics and Economics, from the University of Toronto in 2021.

Code of ethical conduct

To ethically fulfill the privacy requirements of all applicants, data sets provided were anonymous and limited in the amount of characteristics provided. All observations had unique values under “Applicant ID”. However, all participants may not have been given informed consent on the use their personal data. The consent factor is unknown to our analysis as it was not given within the descriptions of the data sets. In this report, we assume that the data set provided by Black Saber Software fulfilled all ethical and legal requirements.

The data collected for the report did not exert any physical, psychological, or emotional harm to any participants. As this is a consulting project, we did not directly test on the participants. Only the results of the hiring, salary, and promotions of the company were analyzed and would not have any effect on the participants.

As ethical statisticians working under Saber Analytics, we acknowledge and distinguish between proper and improper statistical methods. We acknowledge the data wrangling procedures and editing, possible sources of error, and limitations of our statistical inferences. Our final report conveys our findings with honesty. We ensure that our statistical approaches are appropriate for Black Saber Software.

References

- Jokinen, J., & Pehkonen, J. (2017). Promotions and earnings - gender or merit? evidence from longitudinal personnel data. *Journal of Labor Research*, 38(3), 306-334.
doi:<http://dx.doi.org.myaccess.library.utoronto.ca/10.1007/s12122-017-9254-7>
- Xie Y, Shauman KA (1998) Sex differences in research productivity: New evidence about an old puzzle. *Am Sociol Rev* 63(6):847–870. doi:10.2307/2657505
- Hesli VL, Lee JM (2011) Faculty Research productivity: Why do some of our colleagues publish more than others? *PS: Political Science and Politics* 44(2):393–408.
doi:10.1017/S1049096511000242
- Ali A. Karakhan, John A. Gambatese, Denise R. Simmons, and Ahmed Jalil Al-Bayati. (2020). “Identifying Pertinent Indicators for Assessing and Fostering Diversity, Equity, and Inclusion of the Construction Workforce”. *Journal of Management in Engineering*, 37(2).
[https://doi-org.myaccess.library.utoronto.ca/10.1061/\(ASCE\)ME.1943-5479.0000885](https://doi-org.myaccess.library.utoronto.ca/10.1061/(ASCE)ME.1943-5479.0000885)