Examination of Fairness in Black Saber's Promotion, Salary, and Hiring Processes

An analysis of a fictional company

Report prepared for Black Saber Software by Statican

2021-04-21

Contents

Executive summary

Background and aim

Black Saber has been a grouping company in the past 7 years. From the data in the past 7 years we want to determine whether or not the promotion, salary, and hiring processes are fair. That is, if these processes are only related to talent and performance and not other factors. The three research question are: * Fairness of promotion process: is the promotion process only related to talent and performance but not other factors? * Fairness of salary process: is the salary process only related to talent and performance but not other factors? * Fairness of hiring process: is the hiring process only related to qualifications and talentbut not other factors?

Key findings

- The company has been steadily grouping since 2013.
- Roughly 10% of employees receives a promotion every quarter, which comes with a salary increase. This means the promotion and salary processes are not independent of each other and occurs simultaneously.
- A point to note is that employees with higher seniority, they not only have a higher base salary, they also have higher salary increase. This means the gap between the different seniority will increase even more.
- The promotion process is slightly unfair: Women and gender group of "prefer not to say" have had a lower promotion chance than men in Black Saber, while productivity or leadership (which is related to talent and performance) had no major impact on an employee's chance on getting promoted.
- The Salary process is slightly unfair: for the same reason as the promotion process.
- the hiring process is fair: it only relates to qualifications and talents, and is not gender biased or influenced by other factors

Limitations

- For the promotion and salary section: not a lot of information about company policies are given in the instructions so it prohibits us from understanding certain statistics
- For the promotion and salary sections: the first few quarters have very few employees. The smaller sample size lead to more variance and potential inaccuracy in the analysis
- For the hiring section: some of the important variables found during the initial analysis were not explained well enough in the generalized linear mixed-effects model analysis. For instance, we found that having a cover letter is crucial to an applicant for them to pass the first phase, but the model results show that having a cover letter is not relevant or significant.
- For the hiring section: The number of observations are too small for the data used and collected in the last two phases of the hiring process, this could potentially cause fluctuations

and inaccuracies in the generalized linear mixed-effects model analysis and determining what factors are crucial for an applicant to get hired.

Technical report

Introduction

The aim of this report is to check if the company's promotion, salary, and hiring processes are fair, in the sense that it is only based on talent and performance and not other factors. Thus, there are three separate but related research questions about these three processes. In each research question, an initial exploratory analysis was performed to explore the data. Next, multiple models are made and compared with each other for model selection. Finally, the selected model is used to explain the results obtains about the fairness of each of the research questions. In particular, the models used include logistic and other GLMMs (general linear mixed models).

Research questions

- Fairness of promotion process: is the promotion process only related to talent and performance but not other factors?
- Fairness of salary process: is the salary process only related to talent and performance but not other factors?
- Fairness of hiring process: is the hiring process only related to qualifications and talentbut not other factors?

Fairness of the promotion process

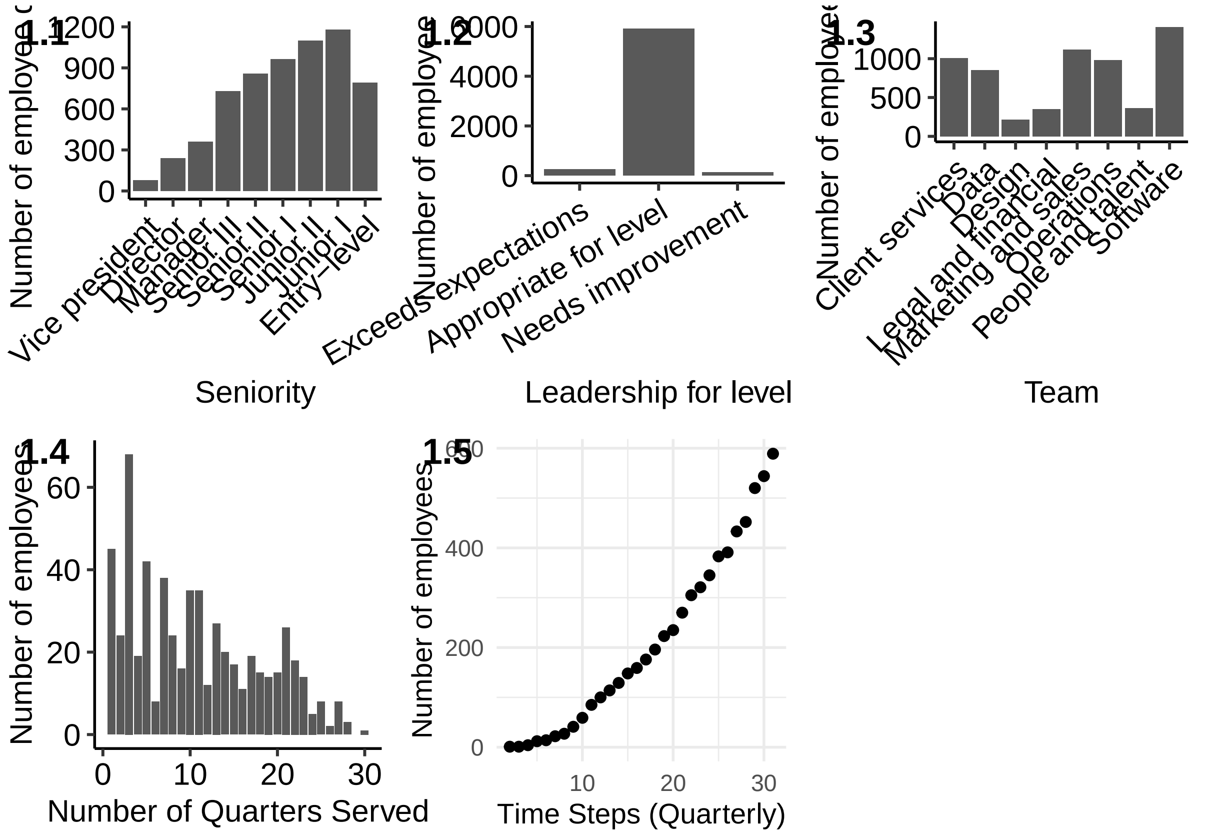
In this section, we want to use appropriate statistical tools to determine whether or not the promotion process is fair. That is, the promotion process only depends on talent or performance and there are little or no bias towards gender or other non-performance-related variables.

Data description, wrangling, and exploratory analysis The data we focus on this section is current employee data, where the variable description is included in appendix 1.1. The two main variable of interest are the seniority levels (`role_seniornty`) and financial quarters (`financial_q`). In particular, a change in an employee's seniority level between two financial quarters represents a promotion received by that employee between the two financial quarters. In the wrangling of the data, several new variables of were created:

Variable name	Type	Description
`seniority_rank`	dbl	Rank of seniority level where 1 represents entry and vice 9 represents vice president.
`quarter_rank`	dbl	Numeric representation of financial quarters where 1 represents 2013 Q2 and 31 repres
`leadership_rank`	dbl	Rank of leadership for employment level, where 1 represents "Needs improvement" an
`diff_role`	dbl	An indicator variable of whether not an employee received a promotion in a financial c
`total_prom`	dbl	Number of total promotions received by an employee.
`prod_group`	dbl	Productivity group; a rounded and scaled version of `productivity` from 1 to 10 inste
`salary_group`	dbl	Salary group; a rounded and scaled version of salary, where 1 represents the salary in t

Among these variables, `diff_role` and `total_prom` are of significant interest while the other three are only to help with easier data manipulations or simplifications.

Next, we conduct some exploratory analysis to look at the distribution of employees in Black Saber through the following plots: 1.1: Distribution seniority roles over employee quarters; 1.2: Distribution of leadership level over employee quarters; 1.3: Distribution of Team over employee quarters; 1.4: Histogram of numbers of employees and numbers of quarters served; 1.5: Number of employees over time. Note that employee quarters are represented by the numbers of quarters multiply the numbers of employee.



From figure 1.1, we see that the distribution of seniority ranks resembles a left skewed normal distribution, and that the vast majority of employees range from entry level to senior III over the employee quarters. Figure 1.2 shows that most employees have appropriate leadership for their levels across the financial quarters with slightly more employee quarters exceeding expectations than needing improvements. Figure 1.3 shows the total number of employee quarters worked by employees of different teams. From these three plots, we are intended to say that the majority of employees may also range from entry level to senior III; most employees have appropriate leadership for their levels; and less amount of employee quarters worked by a team may suggests that fewer employees are from that team. However, this may not be true since this assumes that the composition of seniority, leadership, and teams are similar in proportions through the financial quarters. Figure 1.5 shows that the number of employees has been growing steadily since the company was founded in 2013 Q2 (1st time step). Figure 1.4 suggest that there are more employees who worked fewer quarters than those who worked more quarters. This tells us that there are more newer employees than older ones which is in agreement with figure 1.5.

Next, we look at four more plots about promotions: 1.6: Histogram of number of promotions; 1.7: Distribution of promotion rate across quarter ranks (`quarter_rank`); 1.8: Promotion rate across teams; 1.9: Promotion rate across productivity groups (`prod_group`).

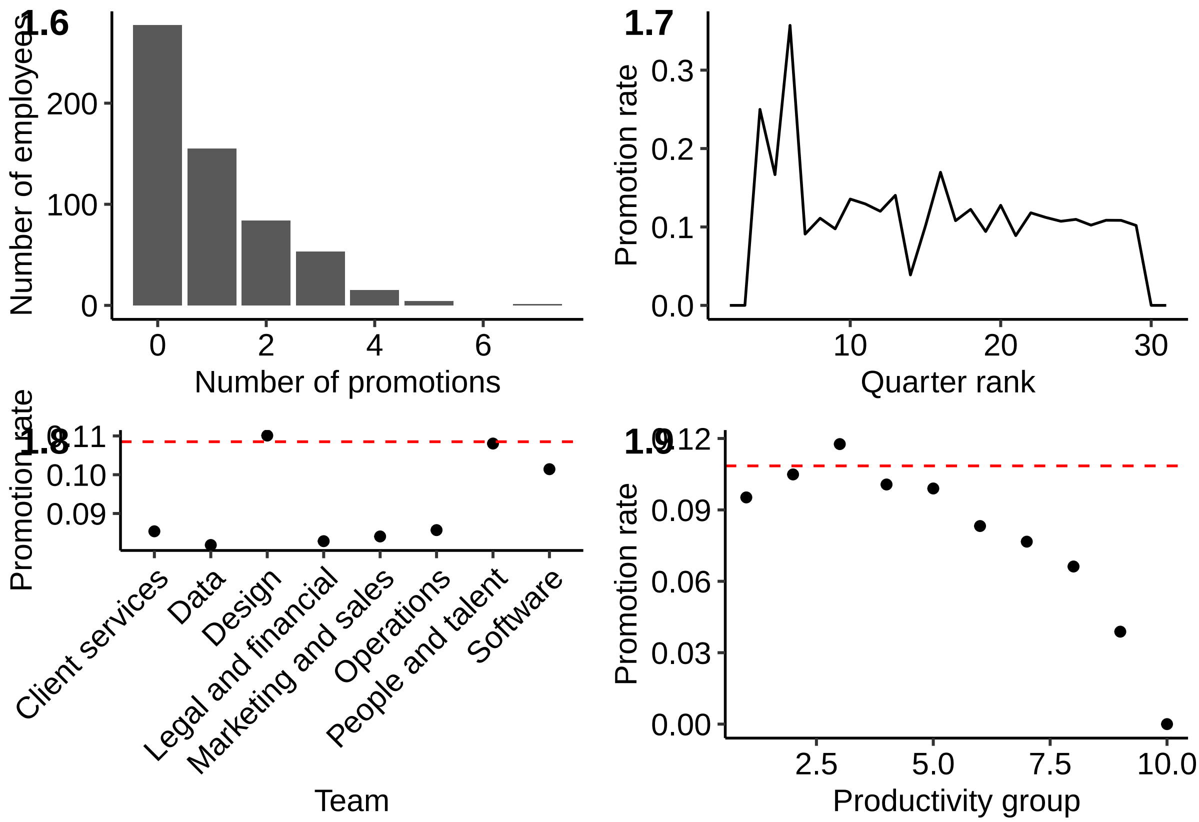


Figure 1.6 suggests that there are fewer employees who got more promotions and more employees with little or no promotions. Knowing that most employees are newer employees and that Black Saber has hired more employee over time, this is expected. Figure 1.7 suggest that promotion rate is roughly the same for each quarter averaging around 0.1 (this means on average 10% of the employees will get a promotion every quarter). In 7th quarter (2014 Q4) where the promotion rate is exceptionally high at around 0.35, but this is likely an outlier due to fewer employees at the time. From figure 1.8, we see that design, people and talent, and software teams tend to have higher promotion rate than the other teams. The red dotted line (for plots H and I) marks the median promotion rate quarterly, where the median was chosen instead of the means because the outlier promotion rate in the 7th quarter is acknowledged. Plot H may suggest slight unfairness in the promotion processes, but the difference between the team with the lowest promotion rate and the team with the highest promotion rate is very small ($>$3%). Lastly, figure 1.9 may appear to suggest that the highest promotion rate occurs at around `prod_group` = 3, while increasing productivity after that may result in decrease promotion rate. This may suggest unfairness in the promotion processes; however, we cannot say this certain because this plot does not take into account many other factors such as current position of the employee. For instance, if an employee with high productivity is already a director than it is more unlikely for that employee to get a promotion, especially when the company already have a Vice president and isn't in need of another one.

Modeling section Since promotion has a binary outcome, logistic models are the most appropriate. Before fitting the model, we briefly examine the log odds and see if they have strong correlation with any of the variables. Among the variables, three of them stood out to appear to have strong pattern associated with the log odds, these are gender, productivity rank, and salary group (see appendix 1.2). This suggests that there might be significant relationship between these variables and promotion.

The next step is to find a good model for promotion. In brief, a total of four GLMM (general linear mixed models) were fitted and compared through ANOVA for model selection (see appendix 1.3). Using this method, we determined that a model named `mod2` is the best model. This model has the least numbers of parameters and is considered a simpler model. The reasons for this is that `mod2` has the smallest AIC and BIC. Moreover, when we conduct ANOVA on `mod2` with the other three models we do not see a significant p-value less than 0.05. This suggest that the other three models which is nested and has additional parameters is not significantly differently from `mod2`.

In the final model, the response variable is an indicator variable of an individual getting a promotion (`diff_role`); the fixed effects `prod_group`, `salary_group`, `gender`, `leadership_rank`, and `seniority_rank`; the random effects are `employee_id` and `quarter_rank`. This model assumes that promotion should be unrelated to quarter, and we have checked our assumption earlier in plot G which shows that promotion rate is roughly the same every quarter. Let $Y_{i,j}$ be an indicator variable recording if the jth promotion from Quarter i was given to employee k. Let $ p_{i[k]j}$ be the true probability that the jth promotion from the ith financial quarter is given to an employee rather than some one else. We can interpret the final model using a two-level approach:

Level 1:

$$log(\frac{p_{i[k]j}}{1-p_{i[k]j}}) = a_i + b_i * Prod\_rank_{i,j} + c_i * salary\_group_{i,j} + d_i * gender_i + f_i * leadership\_rank_{i,j} + g_i * seniorit \tag{1}$$

Level 2:

$$a_i = \alpha_0 + \mu_i + v_k$$

$$b_i = \beta_0$$

$$c_i = \gamma_0$$

$$d_i = \delta_0$$

$$f_i = \phi_0$$

$$g_i = \psi_0$$

where error terms at Level Two can be assumed to follow independent normal distributions:

$$v_k \sim N(0, \sigma_v^2)$$

$$\mu_i \sim \ N(0, \sigma_\mu^2)$$

The proposed model will be:

$$log(\frac{p_{i[k]j}}{1-p_{i[k]j}}) = \alpha_0 + \beta_0 * Prod\_rank_{i,j} + \gamma_0 * salary\_group_{i,j} + \delta_0 * gender_i + \phi_0 * leadership\_rank_{i,j} + \psi_0 * senio \quad (2)$$

From examining the summary models, we get the following statistics:

- $\widehat{\alpha_0} = -1.525$ is the mean log odds of an employee getting a promotion in a quarter. In other words, the odds of an employee getting a promotion in a given quarter is $e^{-1.525} = 0.218$ meaning that for every promotion chance granted, it is only 0.218 times likely for an employee to get a promotion for a given quarter than not.
- $\widehat{\beta_0} = -0.072$ is the decrease in mean log odds of an employee getting a promotion for each 1 rank increase in the productivity rank. Specifically, the odds ratio is $e^{-0.072} = 0.931$, suggesting that the chance of an employee getting a promotion decrease by about 6.9% when productivity increases.
- $\widehat{\gamma_0} = -0.131$ is the decrease in mean log odds of an employee getting a promotion for each 1 rank increase in salary group. Specifically, the odds ratio is $e^{-0.131} = 0.877$, suggesting that the chance of an employee getting a promotion decrease by about 12.3% when salary group increases.
- $\widehat{\delta_0} = 0$ (for men) $\widehat{\delta_0} = -0.318$ (for prefer not to say) $\widehat{\delta_0} = -0.428$ (for women) are the respective change in log odds dependent on their gender. These statistics suggest that the odds ratio for the "prefer not to say" gender group is $e^{-0.318} = 0.728$, meaning that this gender group is 27.2% less likely to get a promotion than men. Moreover, the odds ratio for the "women" gender group is $e^{-0.428} = 0.652$, meaning that this gender group is 34.8% less likely to get a promotion than men.
- $\widehat{\phi_0} = -0.312$ is the decrease in mean log odds of an employee getting a promotion for each 1 rank increase in leadership rank. Specifically, the odds ratio is $e^{-0.312} = 0.732$, suggesting that the chance of an employee getting a promotion decrease by about 22.8% when leadership rank increases.
- $\widehat{\psi_0} = 0.175$ is the increase in mean log odds of an employee getting a promotion for each 1 rank increase in seniority rank. Specifically, the odds ratio is $e^{0.175} = 1.191$, suggesting that the chance of an employee getting a promotion increase by about 19.1% when seniority rank increases.
- $\hat{\sigma}_\mu^2 = 1.985e-09$ is variance intercept among different employee id
- $\hat{\sigma}_v^2 = 0.808$ is variance intercept among different quarter ranks

Results The main finding from modeling the factors that affects promotion, we have the following results:

- according to the p-values on regression coefficients, the most influential factor that affect whether or not an employee gets a promotion in a given quarter are: seniority rank, gender, salary group, and productivity group. Thus, means that seniority, gender, salary, and productivity of an employee are all influential factors.
- Since gender, and salary are not related to talent or performance, we can say that the promotion process if unfair in these ways.
- Moreover, even though leadership rank, a proxy for leadership for level, is related to talent and performance, it is not a significantly considered in the promotion process, suggesting again there may be some unfairness.
- What is more controversial is that increase in productivity group actually decrease an employee's chance of getting a promotion, which shouldn't be the case if the promotion process is fair.

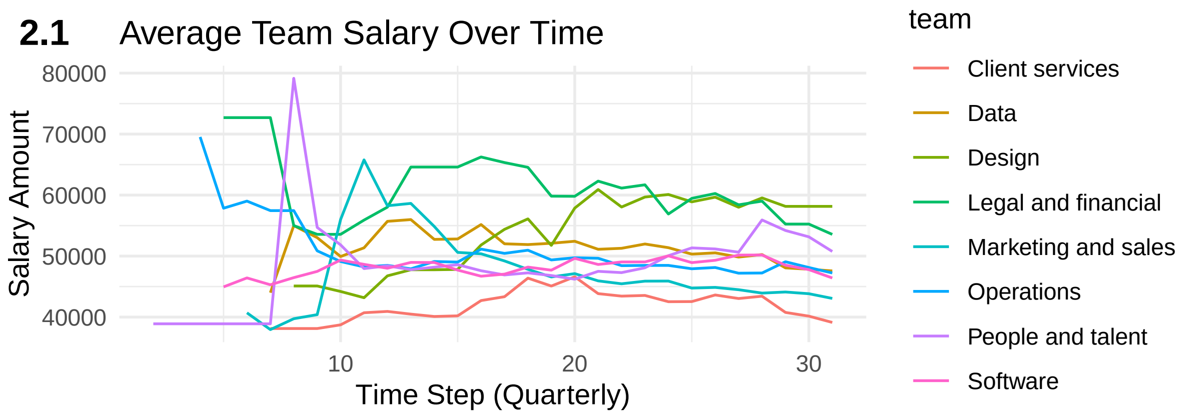
Fairness of the salary process

In this section, we want to use appropriate statistical tools to determine whether or not the salary process is fair. That is, the salary process only depends on talent or performance and there are little or no bias towards gender or other non-performance-related variables.

Data wrangling and exploratory analysis In the salary section, we use the same data set from the promotion section above. For the data wrangling, only one new variable was added to the data set: `pay_up`, which is an indicator variable that indicates whether or not an employee has received an increase in salary between two quarters, where 0 indicates no increase, while 1 indicates increase. We found that `pay_up` takes on the exact same value as `diff_role` from the promotion section, which means that an increase in salary occurs when and only when an employee also receives a promotion.

Next, some exploratory analyses are performed.

2.1 Average Team Salary Over Time



2.2 Pay Increase weight Over Time

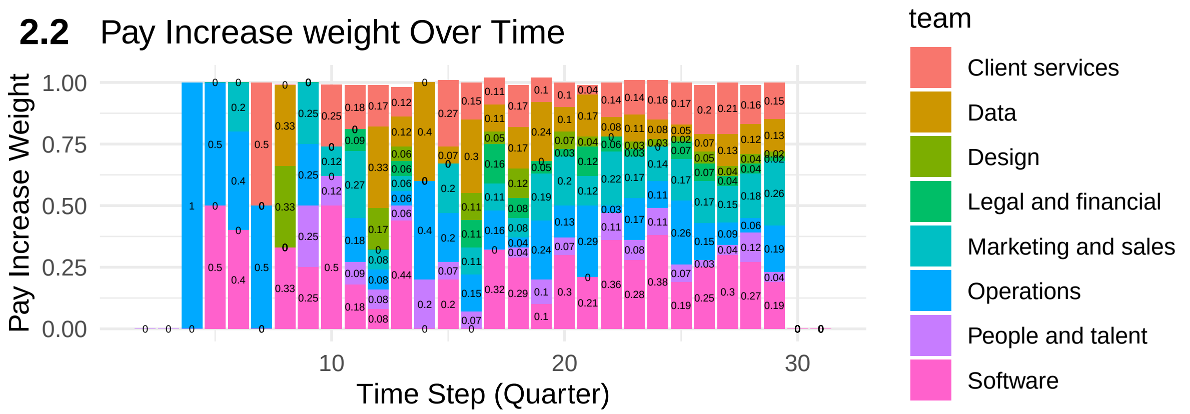
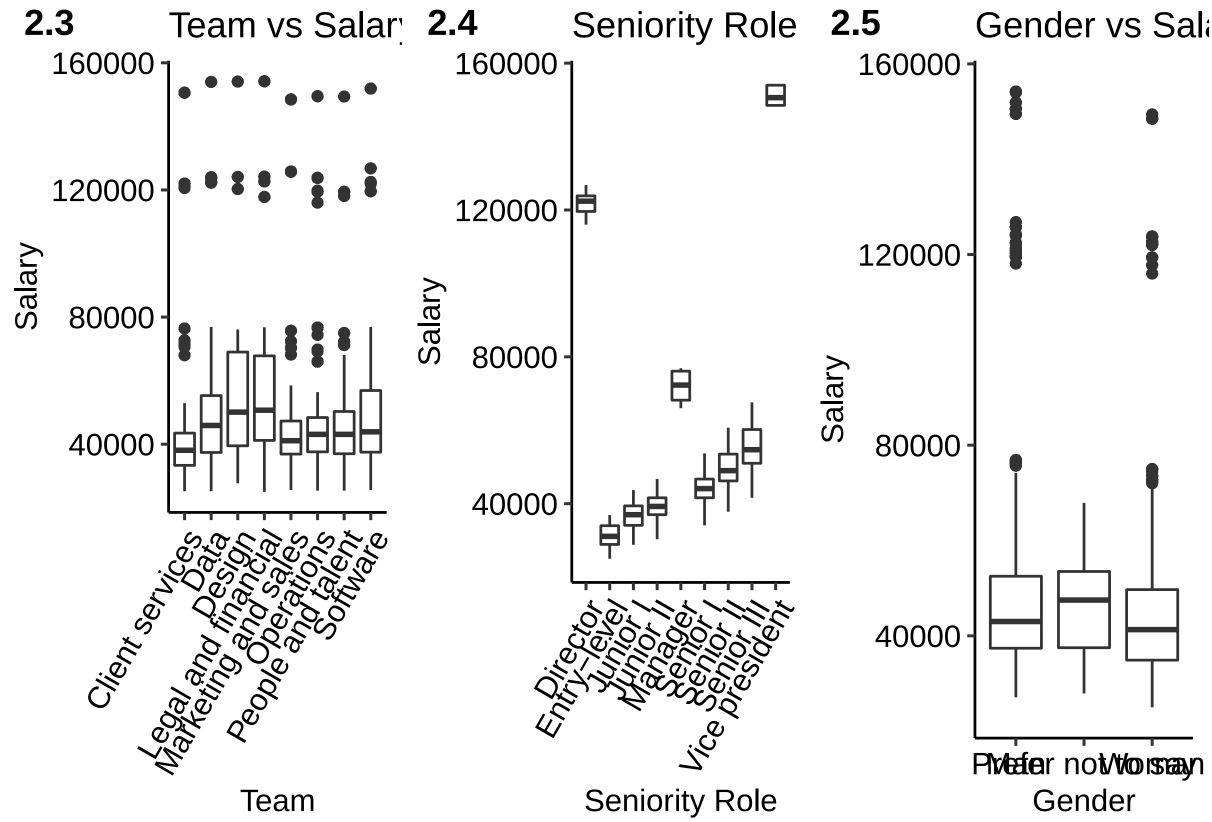
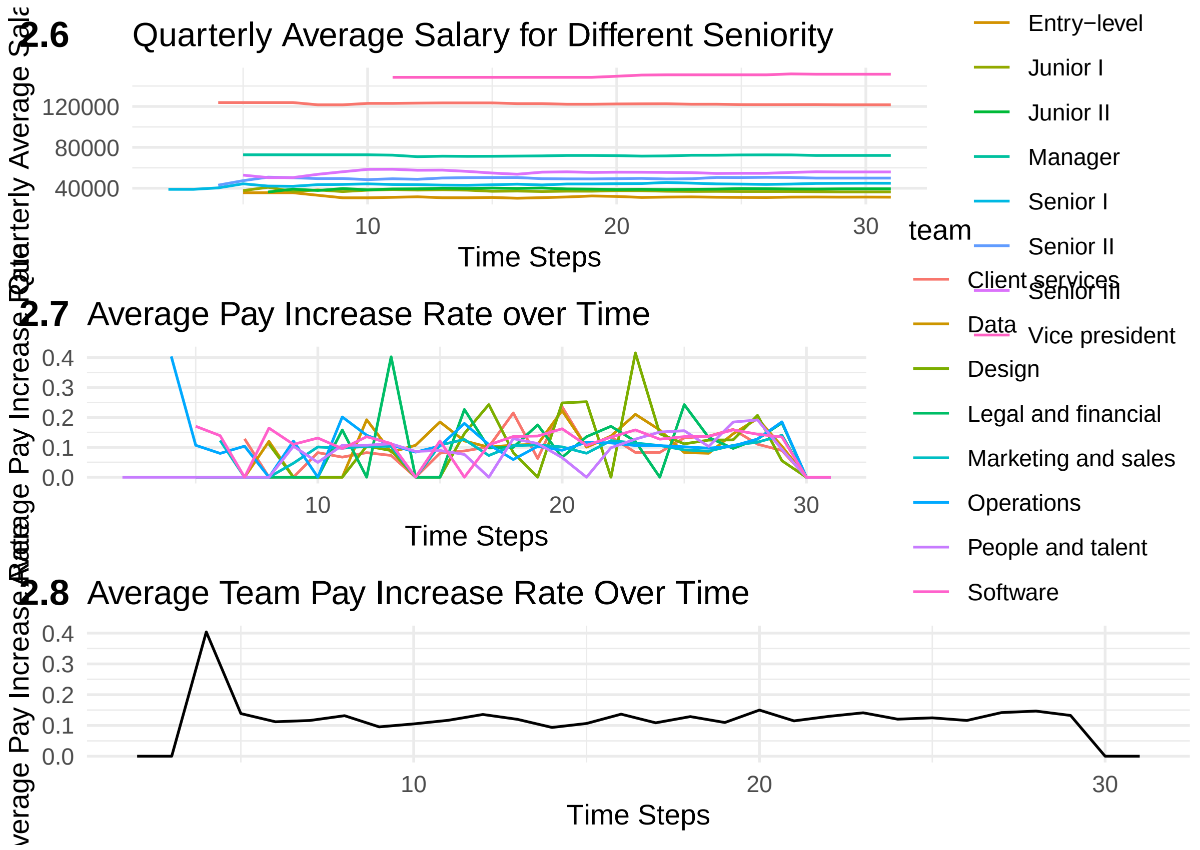


Figure 2.1 shows average team salary over time. This suggest that salary may be related to team as we see the average salary varies between teams over the quarters. Figure 2.2 shows the weighted pay increase as a percent over time. From this figure we may see that the chances of pay increase is not even across teams, as some teams occupies bigger percentage of the total pay increase amounts.



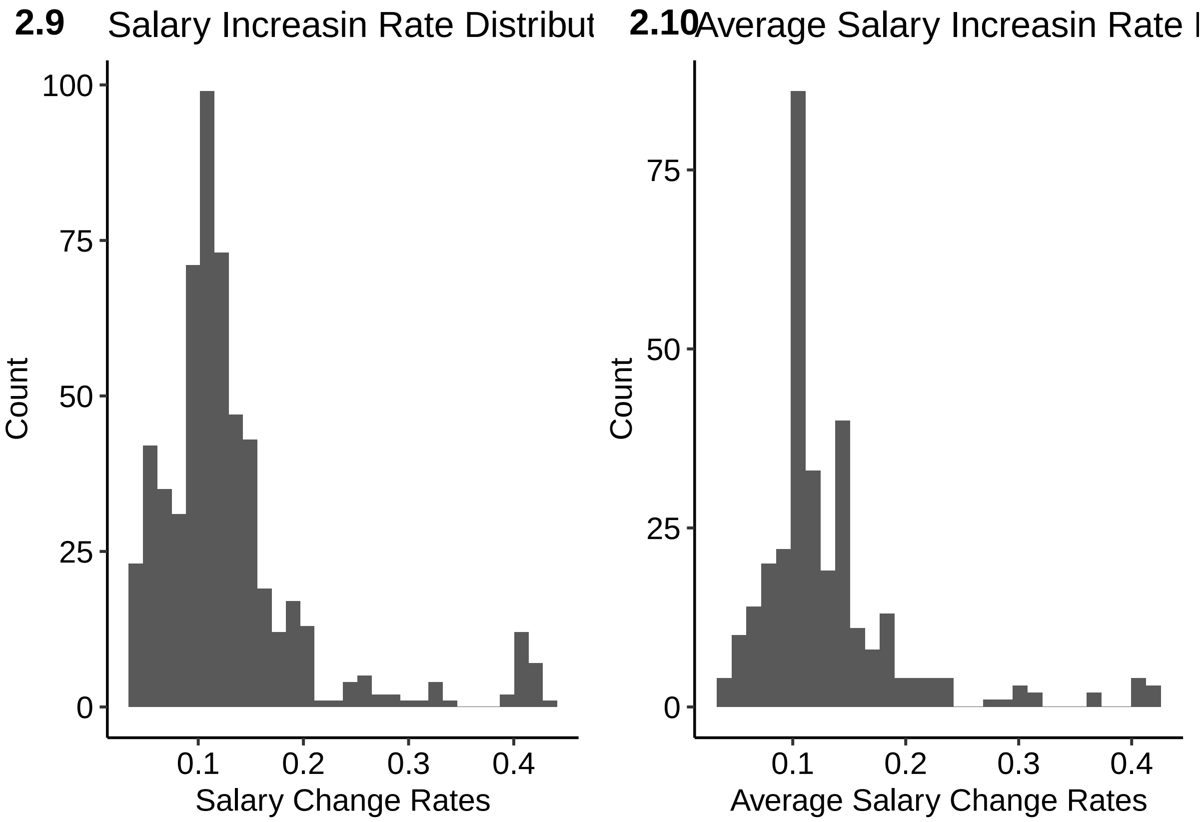
Figures 2.3, 2.4, and 2.5 are box plots that shows variations of salary through different team, seniority, and gender respectively. We see that salary does not vary much among different teams or gender, which is a good sign of fairness. While it does vary among the seniority, which is also normal.



Figures 2.6 shows quarterly average salary for different seniority, where we also see that the average salary increase with seniority rank. We also notice that these averages are roughly the same across quarters. Figure 2.7 shows average pay increase rate in percentage of salary over quarters. We wee all teams roughly averages at the same rate with few outlying exceptions across the quarters. Lasty, figure 2.8 depicts the average salary increase rate over time, which is similar to the promotion rate over time from the previous section, this is excepted as salary increases and promotions happens together. Again, we see the increase rate does not change much throughout the quarters except for the one of the earlier quarters.

Modeling section There was total 569 pay increases over 30 quarters. 277 not pay increase, 155 got once, 84 got twice, 53 got 3, 15 got 4 times, 1 person got 7 pay increases. We can see over 50% pay increases happened to the employees who got twice or more than two pay increases. It is not likely the pay increase behavior from these employees is independent. Hence multiple linear least squares regression using all 569 pay increase observations is advisable for this study.

First step to modeling is checking model assumptions. Since we know that there are employees who have had multiple promotions and salary raise before, we know the independence assumption is not met. Specifically, each observed salary raise is not completely independent from the rest since every employee can get more than one promotion and salary raise. To check if we may still use GLMMs, we must look at the distribution of average salary change rate shown in figure 2.9 and 2.10. Figure 2.9 includes all the observations of salary increases, meaning that the same employee who got more than one salary increase is counted multiple times. On the other hand, Figure 2.10 shows only unique employees who got salary increases. We see that the two distributions are not vastly different which means our assumptions are met and we may proceed with GLMMs.



Next, we used the same method of comparing ANOVA tables in the promotion section for model selection (see appendix 2.1). Using this method, we determined that a model named `model3` is the best model. This time, however, all 4 models that were compared were very similar. `model3` was chosen simply because the variables included are the most meaningful, and that it had a slightly smaller AIC and BIC.

In the final model, the response variable is `salary_num` which is the same as `salary` but numeric instead of a string; the fixed effects `prod_group`, `diff_role`, `gender`, `leadership_rank`, and `seniority_rank`; the random effects ia `employee_id`. Among these variables, gender, employee id and team are considered level two variables, while the rest are considered as level 1 variables. We propose a Level Two covariates - those subject specific variables that provide insight into why individual get salary differently. In the previous data analysis, we saw evidence that individual in

Design and Legal teams have higher salary in garage. Promote is the only factor that triggers a salary increase. Management levels have higher salary and lager pay increase rate. This causes the further salary gap between the regular employers and people in management levels. The model can be expressed as a system of two-level models, where $Y_{i,j}$ is the employee i at time step j:

Level One:

$$Y_{i,j} = a_i + b_i prod\_group_{i,j} + c_i seniority\_rank_{i,j} + e_i diff\_role_{i,j} + f_i leadership\_rank_{i,j} + \varepsilon_{i,j}$$

$$where\, \varepsilon_{i,j} \sim\ N(0, \sigma^2) \tag{2}$$

Level Two:

$$a_i = \alpha_0 + \alpha_1 gender + \alpha_2 team\ \ where\, \mu_i \sim\ N \tag{3}$$

$$b_i = \beta_0 + v_i$$

$$c_i = \gamma_0 + w_i$$

$$e_i = \theta_0 + y_i$$

$$f_i = \phi_0 + z_i$$

$$e_i = \theta_0 + y_i$$

$$f_i = \phi_0 + z_i$$

We obtain the following regression statistics from the model summary:

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## salary_num ~ gender + team + prod_group + seniority_rank + +as.factor(diff_role) +
##     seniority_rank:as.factor(diff_role) + leadership_rank + (1 |
##     employee_id)
##    Data: curr_clean
##
## REML criterion at convergence: 128645.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -8.7633 -0.2050 -0.0144  0.1515  7.5237
##
## Random effects:
```

```
##  Groups      Name        Variance Std.Dev.
##  employee_id (Intercept) 80877213 8993
##  Residual                34540407 5877
## Number of obs: 6295, groups:  employee_id, 589
##
## Fixed effects:
##                                        Estimate Std. Error t value
## (Intercept)                           22116.540   1229.832  17.983
## genderPrefer not to say               -1719.194   3012.265  -0.571
## genderWoman                           -2529.221    819.209  -3.087
## teamData                                941.116   1368.766   0.688
## teamDesign                               -2.136   2466.729  -0.001
## teamLegal and financial                3179.704   1938.408   1.640
## teamMarketing and sales               -1391.913   1245.782  -1.117
## teamOperations                        -1058.244   1362.933  -0.776
## teamPeople and talent                 -3941.599   1995.075  -1.976
## teamSoftware                           1316.591   1208.338   1.090
## prod_group                             -282.231     72.292  -3.904
## seniority_rank                         7793.234     96.564  80.705
## as.factor(diff_role)1                 -4525.136    688.325  -6.574
## leadership_rank                         160.774    358.486   0.448
## seniority_rank:as.factor(diff_role)1    935.355    149.224   6.268
```

We obtain the following regression statistics from the model summary:

- The intercept suggests that the grand average employee salary at Black Saber is around 22000$
- On average the gender group of "prefer not to say" and women make 1719$ and 2529$ less than men in this company, however the p-value is insignificant
- the average salary varies among different teams by a maximum range of around 7000$ with the legal and finance team having the highest average salary and the people and talent team having the lowest average salary among the teams.
- Increase in productivity group on average decreases salary by 282$ per increment, but is insignificant. It is more likely that salary is unrelated to productivity group, which is a proxy to productivity.
- Increase in seniority rank on average increases salary by 7793 per increment, suggesting high correlation between seniority and salary.
- Increase in leadership rank on average increases salary by 160. However, this is also insignificant and we have no evidence that increase in leadership will correlate with higher salary.

Results In short, our model suggest that the salary process is unfair, as it is largely unrelated to talent and performance (unrelated to leadership or productivity). Moreover, it seems that there may be a small bias towards gender groups other than men. Another key finding is that the average salary differs from team to team by a fair amount.

Fairness of the hiring process

Relationship between the fairness of the hiring process and applicant's talent level and gender Our second question aims to uncover if the hiring process is fair based on applicants' talent, which is investigating if the applicants' level of competence and achievement in GPA, extracurricular experiences, work experiences, technical skills, writing skills, leadership skills, and speaking skills can directly positively affect the result of their application. We are also interested in finding if the results are biased based on gender specifically. This question is important to be discussed and addressed as diversity in the work environment is desired, we also want to ensure that all employed applicants are competent to present a high level of performance for increase and ensure maximized productivity and efficiency are achieved.

Data wrangling, data description and initial findings The data sets used in this question including current employee data, data on newly-graduated applicants who qualified for phase 1, phase 2, phase 3, and data on newly-graduated applicants who were eventually hired. In the first phase of the hiring pipeline applicants complete a form and are asked to submit a CV and cover letter. Extracurriculars and internship experience are autorated based on the descriptions applicants provide in the application form. In the second phase, the applicants are asked to complete one technical task, one writing task, and submit a pre-recorded video. The specific scores achieved by each applicant is AI-autograded. In the third phase of the, the applicants are interviewed and graded by two separate interviewers. We also conducted some initial analysis to investigate the relationships and correlations between variables prior to the statistics model construction and fitting process to figure out what variables are of interest in the analysis.

The data is collected and provided by the employer/client in this situation.

The next procedure done is to combine all data sets including data on newly-graduated applicants who qualified for phase 1, phase 2, phase 3, and data on newly-graduated applicants who were eventually hired to create a new data frame that contains 17 variables in total. The detailed description of variables in the new data set can be found in appendix 3.0. This procedure is important as we are interested in examining all aspects and level of competence of all applicants. We then coerced all arguments to either factor or numeric values in order to generate graphs illustrating various relationships and correlations for the initial analysis.

It is interested to find out the correlation between interview ratings and whether or not the applicants got hired(making it past phase 4). We combined the ratings that each applicant received in the first interview and second interview, and divided the ratings by 2 to obtain an average interview rating. A box plot labelled as figure 3.1 is created down below to illustrate the relationship, where the category 0 on the x-axis means that the applicant did not get hired, and the category 1 means that the applicant received a job offer. From the diagram, it is clearly evident that the applicants who are eventually hired have a much higher interquartile range and median values for interview ratings than those who did not get hired, suggesting a direct and strong indication that higher interview ratings definitely contributes to getting hired.

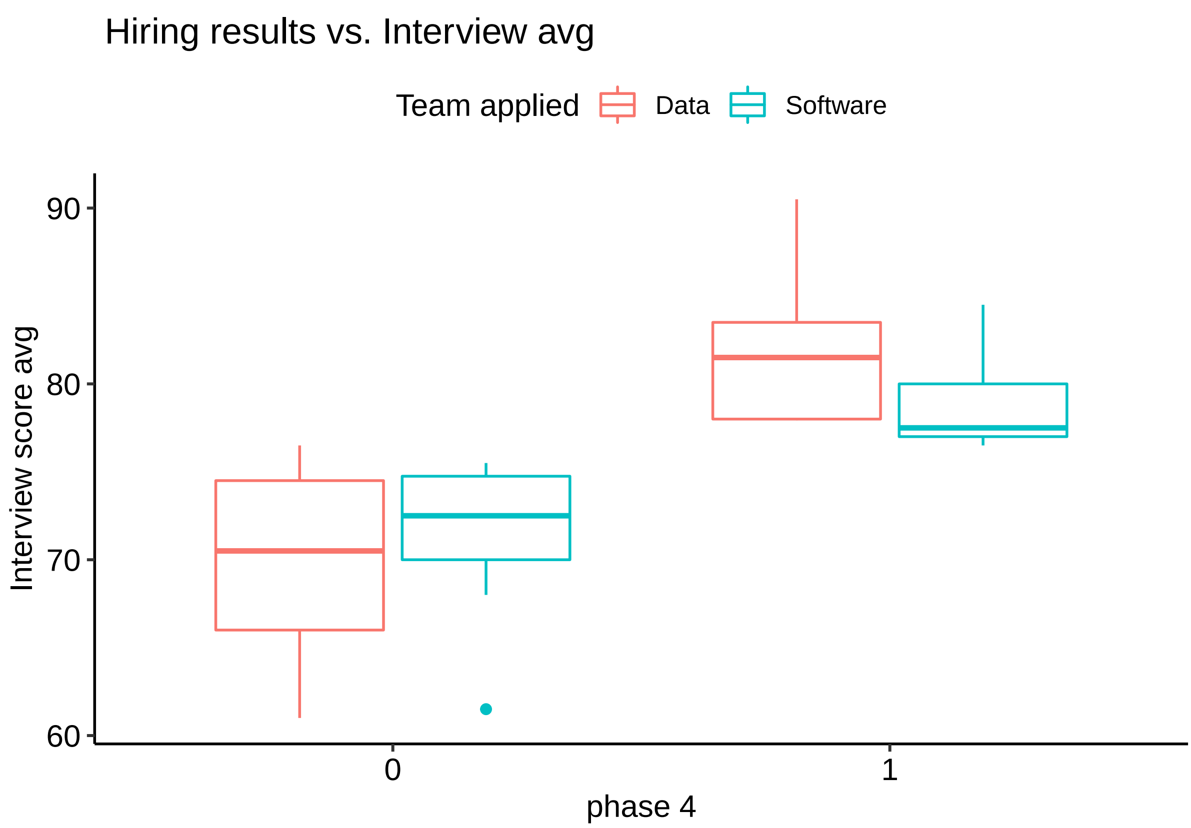


Figure 3.1

It is also important to analyze if there is any correlation between GPA and whether or not the applicants got hired(making it past phase 4). A box plot labelld figure 3.2 is created down below to illustrate the relationship, where the category 0 on the x-axis means that the applicant did not get hired, and the category 1 means that the applicant received a job offer. From the diagram, it is clearly evident that the applicants who are eventually hired have a much higher GPA scores than those who did not get hired, suggesting a direct and strong indication that higher GPA scores definitely contributes to getting hired.

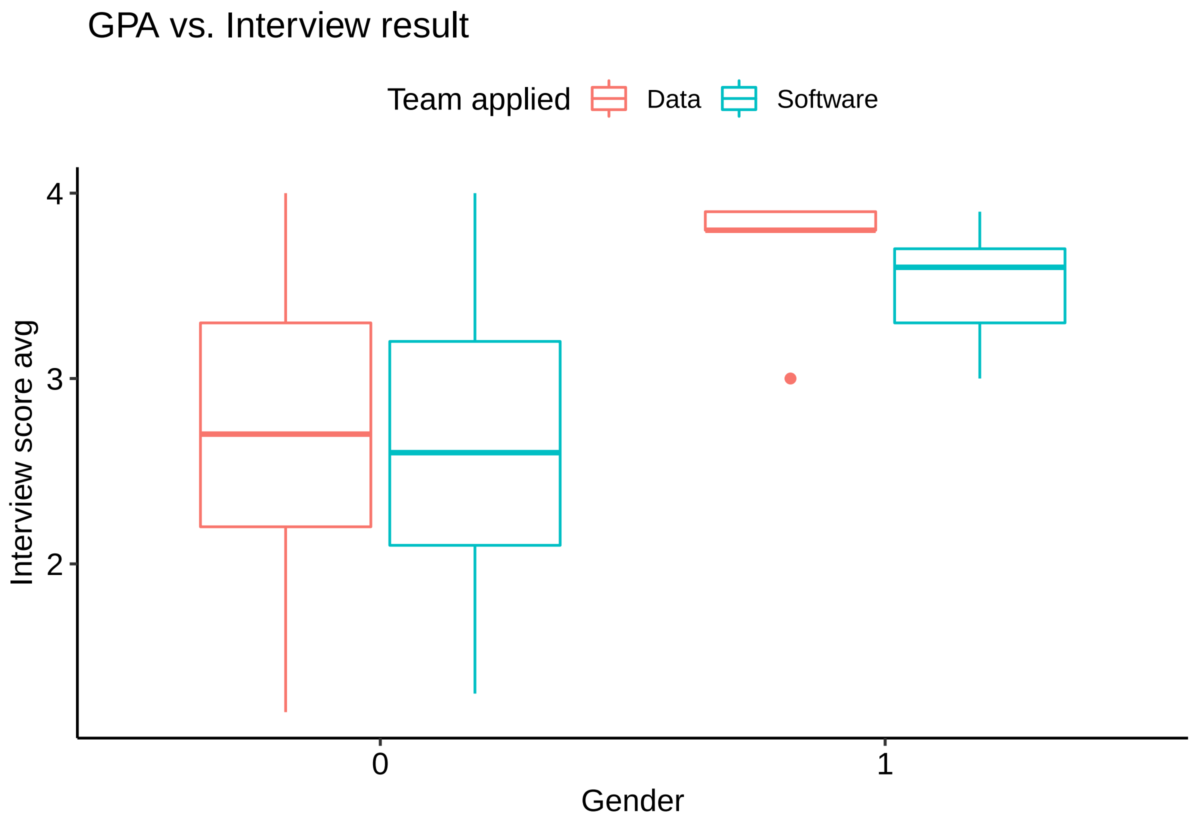


Figure 3.2

We also analyzed the relationship between gender and interview score to see if there is any bias towards giving higher interview scores to a specific gender. We constructed a box plot which categorized the data into teams which are data and software. Figure 3.3 is shown below, and it is evident that men generally are given a higher interview score than women regardless of their talent. We can see that for both teams, men tend to have higher interquartile range and median values for the interview scores, which could suggest that there is bias that interviewers are more lenient towards giving male applicants higher interview scores than female applicants. This is further analyzed and discussed in the model fitting part after.

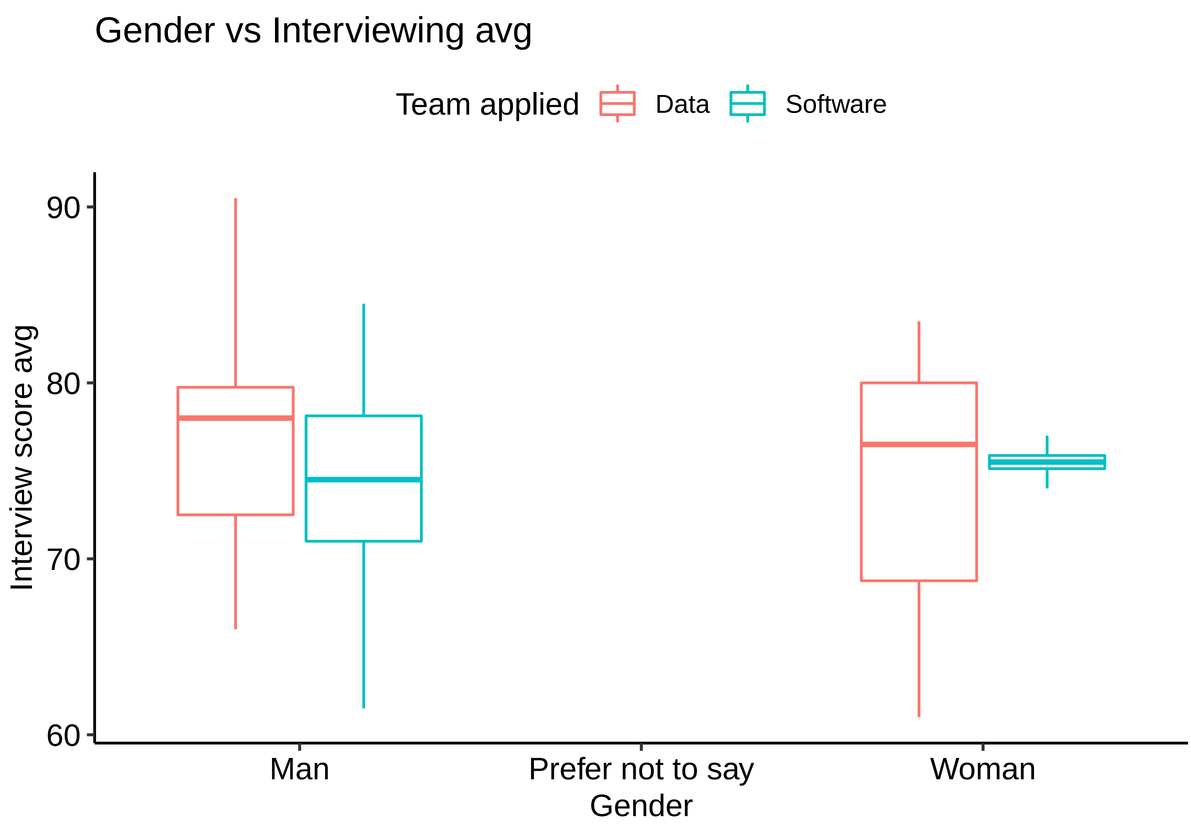


Figure 3.3

Since work experience often plays a crucial role in the hiring process, we wanted to see if work experience also affects the interview score. A box plot is constructed below with categorization based on the different team that each applicant applied for. Evidently shown in figure 3.4, for the data team, the applicants with mediocre work experience(category 1 on the x-axis) has the highest median value for interview scores, it also has the largest interquartile range. For the software team however, applicants with the most experience tend to score higher in interviews, with smaller interquartile range and highest median values. In general, it is unclear as to how work experience affects the interview score, and this should be further analyzed after fitting linear models to the data.

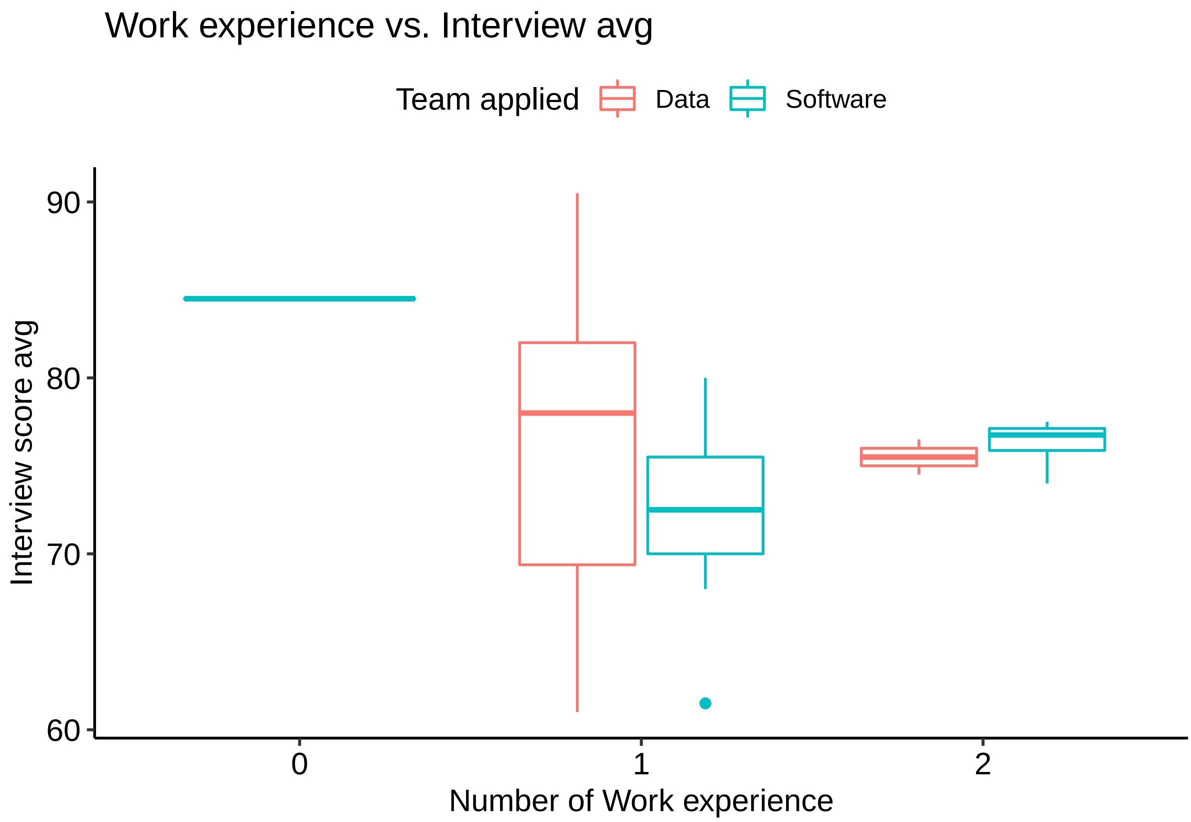


Figure 3.4

Extracurricular involvement could also affect an applicant's performance during an interview. Figure 3.5 is constructed below, and it is clear to see that applicants with the most extracurricular involvement tend to score higher during interviews for both the data team and the software team, indicating that having more involvement and exposure to various of activities and interest could help an applicant with a higher chance of getting hired.

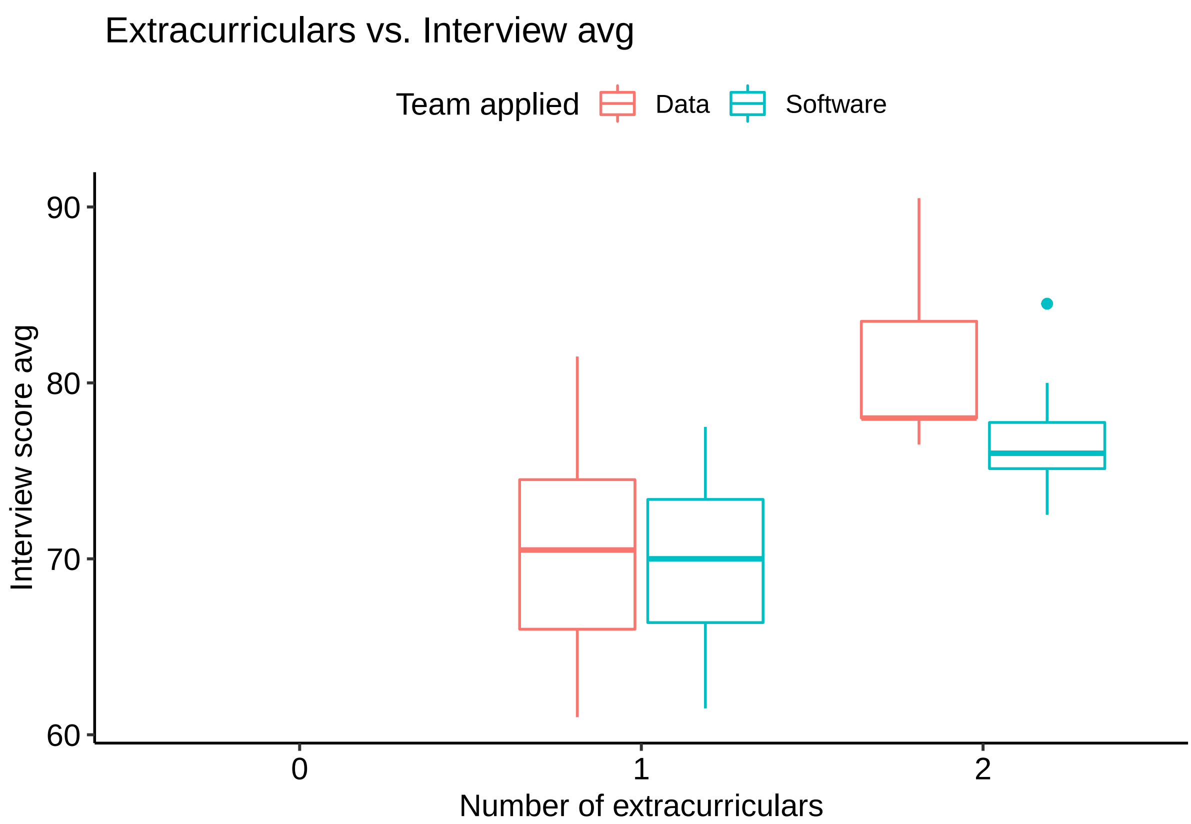


Figure 3.5

Other than interview ratings and GPA scores, we are also curious to find out the correlation between other variables that measure each applicant's talent and whether or not they are hired. To achieve that, a new variable called talent_score is created. This new measurement is calculated by adding together each applicant's technical skills, writing skills, leadership presence and speaking skills, and then divided by four to obtain an average. One thing to be noticed is that because leadership presence and speaking skills are graded on a scale of ten, for the purpose of unity, we multiplied those scores by 10 for the calculation of talent scores. The new data set created is named "all_applicants_data" with 613 observations and 18 variables. The new variable "talent_score" is describe below.

Variable name	Type	Description
`talent_score`	str	Score of each applicant's talent, on a scale of 100

We then investigated the relationship between talent scores and the results of whether the applicant is hired based on genders, and constructed a box plot down below. Evidently shown below in figure 3.6, for those who are not hired(category "0" on the x-axis), the talent score is generally lower regardless of gender, but it is also noted that male applicants have an averagely higher talent scores than female applicants, reflective in both higher interquartile range and median value. It is also noted that the hired applicants(category "1" on the x-axis) have a much higher talent score overall in both genders, but male applicants still have higher scores than female applicants. This could reflect a phenomenon that male applicants tend to receive higher talent scores than female applicants, but it is unclear as to whether this resulted from actual differences in skill levels, or there is a bias lenient towards grading male applicants with higher scores.

Figure 3.6

Data Exploration for phase 1 and 2 For the model selection and fitting portion, we decided to construct models based on different phases of the interview process to clearly distinguish the most impactful component that affects whether or not the applicant passed the first phase, received an interview and got hired eventually, and to see if there is a gender gap during the process. We also conducted some data exploration prior to the model construction and fitting process for better understanding of the statistics and importance of each variable

Table 3.1

##

```
##             reject pass Sum
##   male          88   78 166
##   no answer      3    1   4
##   female        98   89 187
##   Sum          189  168 357
```

```
##
##             reject pass Sum
##   male          58   67 125
##   no answer      5    2   7
##   female        61   63 124
##   Sum          124  132 256
```

```
## [1] "33.333333 percent of applicants made it to phase 2"
```

Firstly we looked at the data of the applicants who successfully passed phase 1 of the interview process. The data set used in this section and the section below is "all_applicants_data" with 613 observations and 18 different variables. We added a "phase 2" column to the data frame recording the applicants who passed phase 1 and divided the data frame based on different teams each applicant applied for, and constructed a table above that summarizes the number of applicants who passed phase 1 by gender. Evidently shown in table 3.1, the statistics suggest that approximately 33.33% of the applicants passed phase 1, with no clear distinction in passing rate between different genders, the odds of a male applicant passing phase 1 is almost identical with the odds of a female applicant passing phase 1 for both the data team and the software team.

Table 3.2

```
## Rows: 1
## Columns: 7
## $ ent_team_applied     <dbl> 0.000987222
## $ ent_cover_letter     <dbl> 0.3397774
## $ ent_cv               <dbl> 0.08513028
## $ ent_gender           <dbl> 0.001829994
## $ ent_extracurriculars <dbl> 0.07236307
## $ ent_work_experience  <dbl> 0.08227545
## $ ent_gpa              <dbl> 0.08227545
```

We also calculated and obtain some mutual information to see what variables can best explain our response variable "phase2", which is whether or not an applicant passed the first phase of the hiring process,the best. The specific variables used are the team that each applicant applied for, whether or not the applicant submitted a cover letter and a resume, the gender of the applicant, the level of extracurricular involvement, work experience, and GPA of each applicant. The intuition is that, higher the score, the better. The specific values are calculated above in table 3.2. It is evident to point out that the most dominant variable that affects our response variable is whether or not the applicant submitted a cover letter.

Table 3.3

```
##
##                   reject pass Sum
##   no cover_letter    219    0 219
##   cover_letter        94  300 394
##   Sum                313  300 613
```

Further analysis proves that having a cover letter is the most impactful variable that makes an applicant pass phase 1 and 2 of the hiring process. Table 3.3 presented above shows that those applicants with no cover letter are all rejected in the first two phases of the hiring process.

Model Fitting for phase 1 After performing some exploratory analysis, we moved on to constructing and analyzing the models. Since the first phase of the hiring process has a binary outcome(passed or rejected), therefore we constructed models with binary response variable. t is also unlikely that each explanatory variable is independent of each other. The response variable interested is whether or not an applicant passed the first phase of the hiring process, with a random effect of applicant ids, while the other potential explanatory variables include the team that each applicant applied for, whether or not the applicant submitted a cover letter and a resume, the gender of the applicant, the level of extracurricular involvement, work experience, and GPA of each applicant.

In brief, two generalized linear mixed-effects model were fitted and compared through ANOVA for model selection (see appendix 3.1). Using this method, we determined that a model named `model_1b` is the best model. This model has the least numbers of parameters and is considered a simpler model. In addition, the selected has a lower value in both the AIC and BIC values, suggesting that the likelihood that this model better predicts or estimates the future values for this data set.

The chosen model includes 6 explanatory variables including the team that each applicant applied for, whether or not the applicant submitted a cover letter and a resume, the level of extracurricular involvement, work experience, and GPA of each applicant. The null hypothesis is that the correlation between the previously mentioned explanatory variables and the response variable is not 0. It is interesting to notice that the model which does not consider gender as an explanatory variable better fits the data set, suggesting that gender is most definitely not a determining factor that affects whether or not an applicant passes the first phase of the hiring process.

The results of the model can be recorded as an equation down below, where xt records the theorized result of whether or not the applicant passes phase 1

$$\hat{x}_t = -84.0077 + 11.9646 * gpa + 41.2832 * cover\_letter + 35.3133 * cv + 26.328 * extracurricular.L - 8.0576 * extrac$$

The summary of the selected model can also be found in appendix 3.1. Notice that most of the p-values for the explanatory variables are above the threshold of 0.05 except for GPA. The p-value for the GPA value is calculated to be about 0.000117, which means that there is incentive not to reject the null hypothesis, suggesting that this variable is significant. We can also see that the estimated value for the GPA variable is positive, suggesting that it can be concluded that having a higher GPA has positive influence on whether or not an applicant passes the first two phases of the interview process. There is indication that the other 5 explanatory variables have profound impact or significance in this model. The results also suggest that an applicant with a higher GPA is 11.9646 times more likely to pass the first phase

Data Exploration for phase 2 Next we focused on the data on the applicants that passed through phase 2 of the interview process. Phase 2 of the interview process focuses on evaluating each applicant's technical skills, writing skills, speaking skills and competence in leadership presence. We removed columns that include results in phase 1 and phase 3 of the data set, as well as each applicant's ratings in the two separate interviews.

Table 3.4

```
## [1] "Phase2 Table"

##
##              reject pass Sum
##   male          130   15 145
##   no answer       3    0   3
##   female        145    7 152
##   Sum           278   22 300

## [1] "7.333333 percent of applicants made it to phase 3"
```

We also constructed a table above that summarizes the number of applicants who passed phase 2 by gender. Evidently shown in table 3.4, the statistics suggest that approximately 7.33% of the applicants who passed phase 1 also passed phase 2. It can also be noted that the number of female applicants who passed phase 2 of the interview process is less than half of the number of male applicants that passed phase 2, despite the fact that more women passed the first and the second phases of the hiring process. There is incentive for us to believe that there is some level of gender gap in the hiring process, and this should be specifically studied in the model analysis to see if there is significant correlation between gender and whether the applicant passes the third phase of the hiring process.

Model Fitting for phase 2 After performing some exploratory analysis, we moved on to constructing and analyzing the models. Since the second phase of the hiring process have a binary outcome(passed or rejected), therefore we constructed models with binary response variable. It is also unlikely that each explanatory variable is independent of each other. Six generalized linear mixed-effects model were fitted and compared through ANOVA for model selection (see appendix 3.2). In the models, the response variable interested is whether or not the applicant passed phase 2, with a random effect of applicant ids, while the other potential explanatory variables include the gender of each applicant, each applicant's technical skills, writing skills, speaking skills and competence in leadership presence. As previously mentioned, we consider the four skills as parts of the applicant's talent.

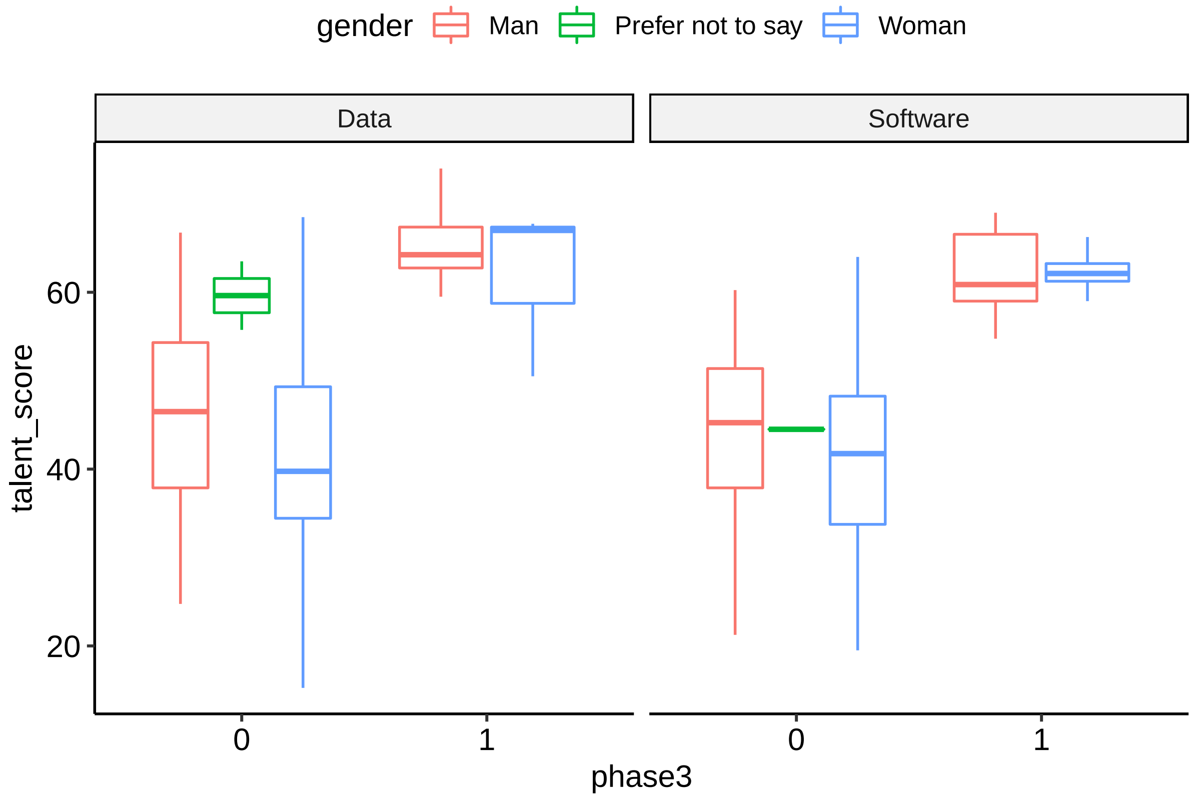
By comparing the AIC values provided by the ANOVA table, it is conclusive that the model that fits the data set better is the last model. However, based on the BIC values, "model_2d" has the lowest value, but this model only contains one explanatory variable, meaning that this model could potentially not give us a thorough analysis of the correlation between different variables. Therefore, we choose to analyze the last model, "model_2f" for analysis.

The selected model has a null hypothesis that the correlation between the explanatory variables and the response variable is not 0.

The results of the model can be recorded as an equation down below, where xt records the theorized result of whether or not the applicant passes phase 2

$$\hat{x}_t = -180.3 + 5.8 * speaking\_skills + 7.916 * leadership\_presence + 0.6565 * technical\_skills + 0.8987 * writing\_$$

The summary of the selected model can also be found in appendix 3.2. Notice that most of the p-values for the explanatory variables are all below the threshold of 0.05, meaning that there is incentive not to reject the null hypothesis, suggesting that all explanatory variables are significant and there is a positive correlation between having higher talent scores and passing phase 2. We can also see that the estimated values for all explanatory variables are positive, suggesting that it can be concluded that having a higher talent score has positive influence on whether or not an applicant passes the first two phases of the interview process. There is indication that all explanatory variables have profound impact or significance in this model. The results also suggest that an applicant with a higher speaking skills is 5.8 times more likely to pass the second phase, 7.916 times more likely for higher leadership presence, 0.6565 times more likely with higher technical skills, and 0.8987 times more likely with higher writing skills. It is also noted that gender is not considered in this model, meaning that there is no indication that a bias exists towards favoring any single gender when grading each applicant in terms of skills.



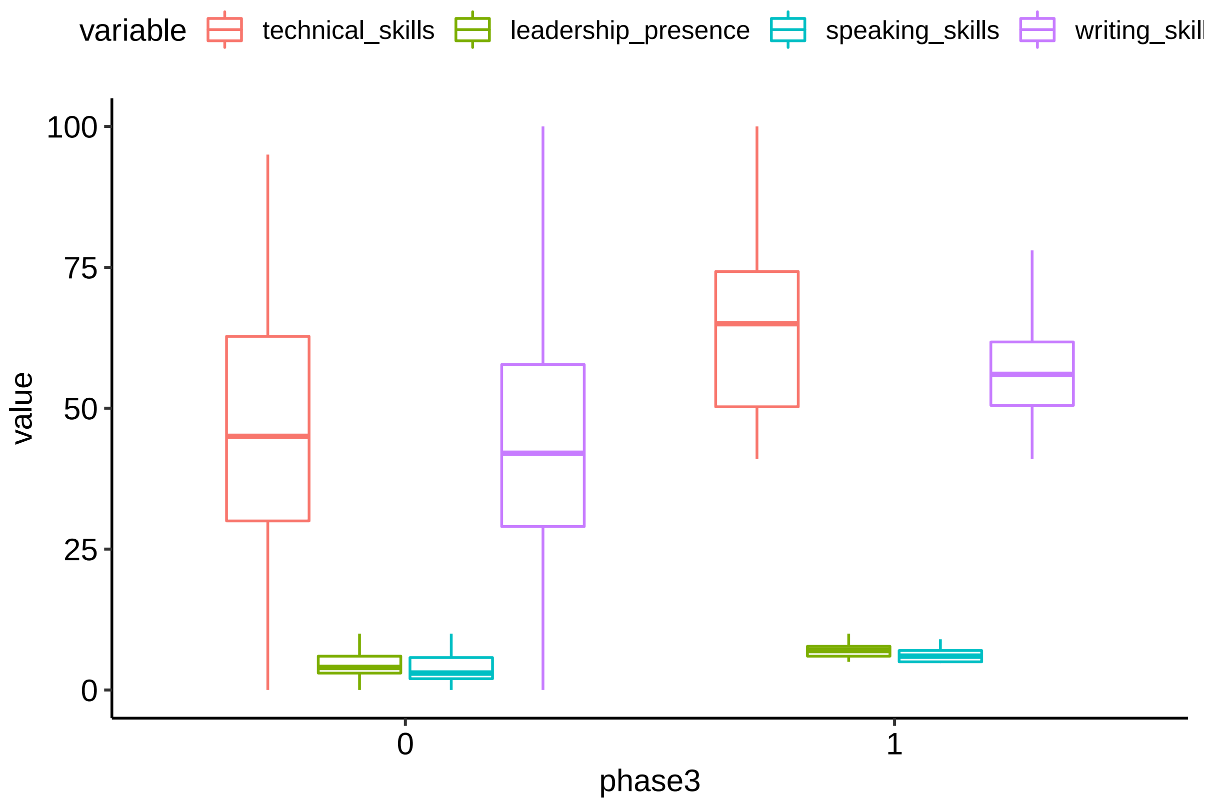
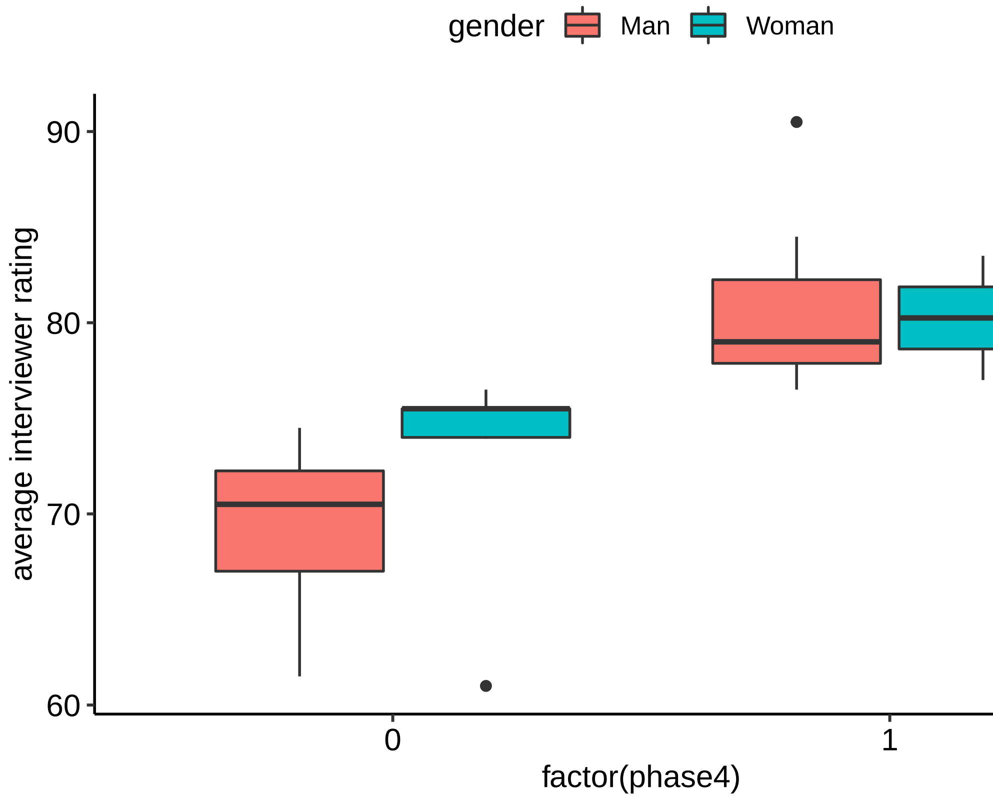


Figure 3.7 & 3.8

Figure 3.7 presented above shows the talent score received by each applicant based on gender(1 represents man, 2 represents prefer not to say, 3 represents woman). For those who passed phase 2(category 1 on the x-axis), there is no significant difference or gender gap between the talent score received by a male applicant and a female applicant for both teams. Although male applicants tend to have higher highest scores than female applicants, female applicants have a higher median value, suggesting that the hiring process for phase 3 is rather fair. Figure 3.8 presented above shows the specific scores in each of the four talents measured, we can see that for those who passed phase 2, the applicant generally have higher scores for all four talents.



Data Exploration for phase 3

Figure 3.9

Figure 3.9 presented above shows the average interview rating of those who eventually got hired and those who did not. It is clear to point out that having a good interview definitely increases the chance of an applicant getting higher, as for both genders, the hired applicants have a significantly higher average interview ratings.

Table 3.5

```
## [1] "Phase4 Table"

##
##          reject pass Sum
##   male        7    8  15
##   female      5    2   7
##   Sum        12   10  22

## [1] "45.454545 percent of applicants made it to phase 4"
```

We also constructed a table above that summarizes the number of applicants who passed phase 3 by gender. Evidently shown in table 3.4, the statistics suggest that approximately 45.45% of the applicants who passed phase 2 also passed phase 3. It is noted that if an applicant passed phase 3, that applicant is hired and is sent an offer. It can also be noted that the number of female applicants who passed phase 3 of the interview process is only 25% of the number of male applicants that passed phase 3, and male applicants had a higher acceptance rate than female applicants. There is incentive for us to believe that there could be some level of gender gap in the hiring process, however, because of the small number of observations in the final number of applicants who are hired, it would be impossible to include gender as an explanatory variable.

Model fitting for phase 3 After performing some exploratory analysis, we moved on to constructing and analyzing the models. Since the third phase of the hiring process has a binary outcome(passed or rejected), therefore we constructed models with binary response variable. t is also unlikely that each explanatory variable is independent of each other. Three generalized linear mixed-effects model were fitted and compared through ANOVA for model selection (see appendix 3.3). In the models, the response variable interested is whether or not the applicant passed phase 3 and got hired, with a random effect of applicant ids, while the other potential explanatory variables include the ratings that each applicant received on the two interviews.

By comparing the AIC and BIC values obtained in the ANOVA table, it is obvious that "model_4c" has the smallest numbers in both measurements, therefore it is selected in this phase.

The selected model has a null hypothesis that the correlation between the explanatory variables and the response variable is not 0.

The results of the model can be recorded as an equation down below, where xt records the theorized result of whether or not the applicant passes phase 3 and got hired. It is noted that the coefficient estimates for both explanatory variables are positive, suggesting there is positive correlaiton between the variables and whether or not an applicant gets hired.

$$\hat{x}_t = -0.005809 + 0.3369 * interviewer\_rating\_1 + 0.4229 * interviewer\_rating\_2$$

The summary of the selected model can also be found in appendix 3.3. Notice that most of the p-values for the explanatory variables are all below the threshold of 0.05, meaning that there is incentive not to reject the null hypothesis, suggesting that all explanatory variables are significant and there is a positive correlation between having higher interviewer ratings and getting hired. The results are actually rather self-explanatory and intuitive, given that doing well in interviews naturally leads to higher chance of getting hired.

Results The results can be summarised by these bullet points:

- Applicants with higher GPA scores are more likely to pass the first phase of the hiring process
- Applicants with higher interviewer ratings have a much higher chance of getting hired, regardless of the gender of the applicant and what team the applicant is applying for.
- Having extensive work experience do not directly affect an applicant's performance during interviews
- Applicants with more extracurricular involvement however, generally perform better during interviews, regardless of what team the applicants are applying for.
- For the hired applicants, males generally have higher scores in speaking skills, writing skills, technical skills, and show more leadership presence.
- Around 33.33% of the applicants qualified for phase 2, which is 300 applicants consisting of 145 males, 152 females, and 3 applicants who prefer not to reveal their gender.
- Applicants with no cover letters are all rejected at the first phase of the hiring process
- Generalized linear mixed-effects model analysis suggests that GPA has the most profound impact on whether or not the applicant passes the first phase
- Approximately 7.33% of the applicants who pass phase 1 also pass phase 2, with 22 applicants consisting of 15 males and 7 females.
- Generalized linear mixed-effects model analysis suggests that applicants with higher scores in speaking skills, writing skills, technical skills, and show more leadership presence have a higher chance at passing phase 2.
- Approximately 45.45% of the applicants who pass phase 2 also pass phase 3 and get offered a job position, which is 12 applicants consisting of 8 males and 2 females.
- Generalized linear mixed-effects model analysis indicates that performance during interviews determines whether or not an applicant gets hired.
- There is no indication, based on generalized linear mixed-effects model analysis, that any specific gender of applicants is favored during the hiring process. The decision is made on a fair ground with no bias or gender gaps.

Discussion

In brief, the analyses on promotion and salary processes suggest that these processes are unfair. The high similarity between promotion and salary relation are resulted by the fact that promotions and salary changes happen simultaneously and thus not independent of each other. We found that these processes are not only gender bias, they also does not reflect on employees' leadership or productivity, which are related to talent or performance.

For the Hiring process, after performing data exploration and model fitting to the data recorded in all phases, there are several crucial findings to be noted. Based on the regression model analysis, for an applicant to pass phase 1, the most important aspect is to have a high GPA. For an applicant to pass phase 2, they need to score high on writing skills, speaking skills, technical skills, and demonstrate competence in showing leader presence. For an applicant to pass phase 3 and finally get hired, they need to perform exceptionally well on the two interviews. Regardless of what phase the applicant is in, gender does not seem to have any significant impact as to whether the applicant passes the phase or not. With that being said, it seems that the hiring process is fair.

Strengths and limitations

A strength in the promotion and salary section is that model assumption were checked and appropriate models were used for the corresponding situations. Moreover, much exploratory analysis has been conducted to understand the data.

One of the limitations for the promotion and salary sections is a limited amount of info about the structure or company policies, and knowing certain polices may help understanding the context of a statistic better. For instance, we were able to figure that salary changes when and only when promotion is given. This was crucial for our understanding of the data. However, other information might be hidden about why, for instance, promotion and salary is not related to productivity or leadership – even though we would think that they should correlate.

Another limitation for the models in the promotion and salary sections is the large variances from the first few quarters. As the first few quarter black saber had no more than 20 employees, the ratios of promotion rates and salary change may be more volatile and result in lager variances in our model and possibly influence the calculations of the regression coefficients.

A strength for the hiring section is that the detailed analysis of correlation between various variables that gives deep insight as to whether or not the hiring process is fair based on each applicant's talent and gender. Moreover, multiple generalized linear mixed-effects model analyses are performed for data on each phase of the hiring process, giving us more reliability and accuracy on the results

For the hiring section, one limitation to be noted is that during the initial data exploration of phase 1, we noticed that applicants who did not submit a cover letter were all rejected. However, when we constructed and fitted a regression model analysis, it is found that having a cover letter is not a significant explanatory variable, which is quite contradictory and hard to be explained. There might be some underlying assumptions that were violated when constructing the models.

In addition, for the third and last phases of the hiring process, the number of observations that were available for use is quite low, and there were not too many variables to be analyzed. Even though we have still constructed and fitted a regression model and generated analysis, it is feared that this procedure is quite redundant, as the results obtained are rather intuitive and self-explanatory.

Consultant information

Consultant profiles

Yi Zhe Wang Yi Zhe is a manager in Statican. His strengths include managing workflow, communications, writing reports, and eating sushi. Yi Zhe is majoring in Statistics, and minoring in mathematics and history and philosophy of sciences from the University of Toronto. He expects to earn a Bachelor of Science in 2022.

Xiaoyan Yang Xiaoyan is a an intern in Statican. She is the leader to cheer up and specializes in writing code in R. She is in Statistics and Computer Science double major. She expects to graduate in in 2020.

Jaekang Lee Jaekang is an intern in Statican. He specialises in making coffee and coding in R. Jaekang is majoring in Mathematics and minoring in Stats and Comp-sci at the University of Toronto. He expects to graduate in 2021

Yichen Liang Yichen is a senior statistician at Statican. He specializes in constructing statistical reports and analysis. Yichen is majoring in Statistics and Economics at the University of Toronto. He expects to earn a Bachelor of Science in 2022.

Yutong Lu Yutong is an intern in Statistician. She specializes in coding in R, also managing communication. Yutong Lu is Biological physics in physiology stream specialist program also major in statistics at University of Toronto. She expects to earn a Bachelor of Science in2022.

Code of ethical conduct

This reports strictly follows the ethics statements listed below with specific details

Integrity of data and methods used for analysis

1. Acknowledgements regarding statistical and other assumptions are made in the execution and interpretation of any analysis. Acknowledgements of all data editing procedures are made.
2. Limitations of statistical inference and analysis are reported.
3. The sources and assessed adequacy of the data is reported, all data considered in the study and analysis is accounted for.
4. All steps taken to preserve data integrity and valid results are reported and recorded.
5. The specific findings are conveyed in ways that are both honest and meaningful to the user/reader, findings include tables, models, and graphics.
6. Strives to promptly correct any errors discovered while producing the final report or after publication. As appropriate, disseminates the correction publicly or to others relying on the results.

7. Only undertake work and provide services that are within the limits of professional competence of the statistician; and do not lay claim to any level of competence not possessed.

Responsibilities to the society and research subjects

1. All applicable rules, approvals, and guidelines for the protection and welfare of human and subjects are noted and followed.
2. The use of excessive or inadequate numbers of research subjects is strictly avoided, excessive risk to research subjects in terms of health, welfare, privacy, and ownership of their own data is strictly avoided.
3. The privacy and confidentiality of research subjects and data concerning them are strictly protected. Secondary and indirect uses of the data is strictly prohibited without approval, including linkage to other data sets.
4. All legal limitations on privacy and confidentiality is not to be violated. Confidentiality agreements, expectations of privacy, notification, consent are not to be breached or surpassed when evaluating the appropriateness of the data source(s).
5. Any statistical descriptions of groups that may carry risks of stereotypes and stigmatization is recognized. Objectivity is maintained and procedural or personal bias is to be avoided. The creation of valid data-based information is vital to informed public opinion and policy.

Responsibilities to Employers/Clients

1. All work done in this report is carried out with due care and diligence in accordance with the requirements of the employer or client.
2. Avoid disclosure or authorization to disclose, for personal gain or benefit to a third party, confidential information acquired in the course of professional practice without the prior written permission of the employer or client is not to be shared.
3. Any interest, financial or otherwise is clearly declared, that could be perceived as influencing the outcome of work undertaken for a client or employer.
4. Any potential or actual conflict between the ethical standards of statistical practice and the interests of the client or employer is advised.
5. Exercise care to prevent the use of any misleading summary of the data. Make sure that all assumptions and limitations relevant to the data, the analysis and the results are fully disclosed.

Responsibilities to other research team colleagues

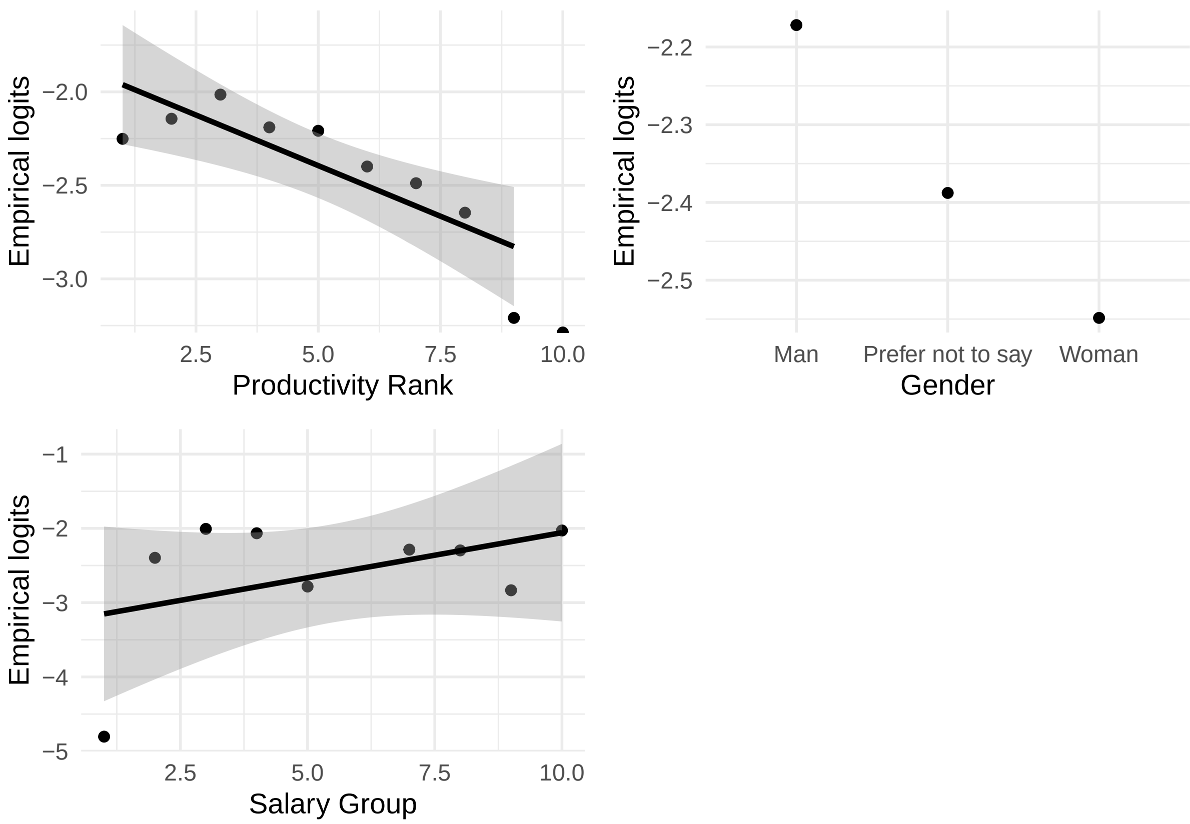
1. Standards and obligations by other colleagues are recognized and respected.
2. All discussion and reporting of statistical design and analysis is ensured to be consistent with various guidelines.
3. Avoids compromising scientific validity for expediency.
4. Transparency in design, execution, and reporting or presenting of all analyses.

Appendices

Appendix 1.1: Description of variables in the current employee data set

Variable name	Type	Description
`employee_id`	dbl	ID of an employee, each unique ID represent an employee working for Black Sab
`gender`	str	Gender of an employee, can be either man, women, or prefer not to say.
`team`	str	Team/department which an employee works at, including Client services, Design
`financial_q`	str	Financial quarter; there are 4 quarters per year; the first quarter is 2013 Q2 and
`role_seniority`	str	Position of employment based on seniority level, from Entry level to Vice preside
`leadership_for_level`	str	Level of leadership of an employee with three levels: appropriate for level, needs
`productivity`	dbl	Score of a productivity test ranging from 0 to 100.
`salary`	str	Salary of an employee in dollars in string format.

Appendix 1.2: Three plots of variables correlating with log odds



Appendix 1.3: ANOVA tables for model selection in the promotion section

```
## Data: curr_clean
## Models:
## mod2: diff_role ~ prod_group + salary_group + gender + leadership_rank +
## mod2:     seniority_rank + (1 | employee_id) + (1 | quarter_rank)
## mod3: diff_role ~ prod_group + salary_group + gender + leadership_rank +
## mod3:     seniority_rank + (1 | employee_id) + (1 | quarter_rank) +
## mod3:     (1 | team)
##      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mod2    9 3628.7 3689.4 -1805.3   3610.7
## mod3   10 3630.7 3698.1 -1805.3   3610.7 0.0089  1     0.9248
```

```
## Data: curr_clean
## Models:
## mod2: diff_role ~ prod_group + salary_group + gender + leadership_rank +
## mod2:     seniority_rank + (1 | employee_id) + (1 | quarter_rank)
## mod4: diff_role ~ prod_group + salary_group + gender + leadership_rank +
## mod4:     seniority_rank + gender:salary_group + gender:seniority_rank +
## mod4:     (1 | employee_id) + (1 | quarter_rank)
##      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mod2    9 3628.7 3689.4 -1805.3   3610.7
## mod4   13 3629.6 3717.3 -1801.8   3603.6 7.0357  4      0.134
```

```
## Data: curr_clean
## Models:
## mod2: diff_role ~ prod_group + salary_group + gender + leadership_rank +
## mod2:     seniority_rank + (1 | employee_id) + (1 | quarter_rank)
## mod5: diff_role ~ prod_group + salary_group + gender + leadership_rank +
## mod5:     seniority_rank + team + (1 | employee_id) + (1 | quarter_rank)
##      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mod2    9 3628.7 3689.4 -1805.3   3610.7
## mod5   16 3637.1 3745.0 -1802.5   3605.1 5.5985  7     0.5873
```

Appendix 2.1: ANOVA tables for the model selection in the salary section

```
## Data: curr_clean
## Models:
```

```
## model3: salary_num ~ gender + team + prod_group + seniority_rank + +as.factor(diff_role)
## model3:     seniority_rank:as.factor(diff_role) + leadership_rank + (1 |
## model3:     employee_id)
## model2: salary_num ~ gender + team + prod_group + productivity + seniority_rank +
## model2:     as.factor(diff_role) + seniority_rank:as.factor(diff_role) +
## model2:     leadership_rank + (1 | employee_id)
##        npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
## model3   17 128900 129015 -64433   128866
## model2   18 128897 129019 -64431   128861 5.201  1    0.02257 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Data: curr_clean
## Models:
## model5: salary_num ~ team + prod_group + seniority_rank + +as.factor(diff_role) +
## model5:     seniority_rank:as.factor(diff_role) + leadership_rank + (1 |
## model5:     employee_id)
## model3: salary_num ~ gender + team + prod_group + seniority_rank + +as.factor(diff_role)
## model3:     seniority_rank:as.factor(diff_role) + leadership_rank + (1 |
## model3:     employee_id)
##        npar    AIC    BIC logLik deviance  Chisq Df Pr(>Chisq)
## model5   15 128906 129007 -64438   128876
## model3   17 128900 129015 -64433   128866 9.6778  2   0.007916 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix 3.0 Description of the new data frame

Variable name	Type	Description
applicant_id	dbl	A unique ID assigned to applicants in Phase 1
team_applied_for	dbl	Software or Data
cover_letter	dbl	0 if cover letter absent, 1 if present
cv	dbl	0 if resume absent, 1 if present
gpa	str	0.0 to 4.0
gender	dbl	Gender of employee: "Man", "Woman", "Prefer not to say" only options provid
extracurriculars	dbl	The description of extracurricular involement is assessed against a proprietary k

Variable name	Type	Description
		rated as high relevance or high skills building
work_experience	dbl	Similar to extracurriculars, the description applicants provided is assessed agains
technical_skills	str	Score from 0 to 100 on a timed technical task, AI autograded
writing_skills	str	Score from 0 to 100 on a timed writing task, AI autograded
speaking_skills	str	A rating of speaking ability based on pre-recorded video, AI autograded
leadership_presence	str	A rating of "leadership presence" based on pre-recorded video, AI autograded
phase_2	dbl	Indication of whether the applicant qualified for phase 2, 0 being yes, 1 being n
interviewer_rating_1	str	The overall rating of job fit given by the first interviewer on a scale of 0 to 100
interviewer_rating_2	str	The overall rating of job fit given by the second interviewer on a scale of 0 to 1
phase_3	dbl	Indication of whether the applicant qualified for phase 3, 0 being yes, 1 being n
phase_4	dbl	Indication of whether the applicant qualified for phase 2, 0 being yes, 1 being n

Appendix 3.1 Model construction and ANOVA table for data in phase 1

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: phase2 ~ gender + gpa + cover_letter + cv + extracurriculars +
##     work_experience + team_applied_for + (1 | applicant_id)
##    Data: pass_phase1
##
##      AIC      BIC   logLik deviance df.resid
##     57.0    110.0    -16.5     33.0      601
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -40.66   0.00   0.00   0.00   1.37
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  applicant_id (Intercept) 0        0
```

```
## Number of obs: 613, groups:  applicant_id, 613
##
## Fixed effects:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -8.771e+01  1.427e+04  -0.006 0.995097
## gender2              2.017e-02  1.279e+01   0.002 0.998742
## gender3              1.118e+00  1.054e+00   1.061 0.288852
## gpa                  1.276e+01  3.402e+00   3.752 0.000175 ***
## cover_letter.L       4.177e+01  1.019e+04   0.004 0.996728
## cv.L                 3.626e+01  1.367e+04   0.003 0.997884
## extracurriculars.L  2.694e+01  1.909e+04   0.001 0.998874
## extracurriculars.Q -8.080e+00  1.102e+04  -0.001 0.999415
## work_experience.L   1.407e+01  7.148e+03   0.002 0.998429
## work_experience.Q  -1.311e+00  4.127e+03   0.000 0.999747
## team_applied_for2  -1.123e+00  9.455e-01  -1.187 0.235127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) gendr2 gendr3 gpa    cvr_.L cv.L   extr.L extr.Q wrk_.L
## gender2      0.000
## gender3      0.000  0.044
## gpa         -0.001  0.001  0.351
## covr_lttr.L -0.340  0.000  0.000  0.000
## cv.L        -0.678  0.000  0.000  0.000  0.002
## extrcrrcl.L -0.630  0.000  0.000  0.000  0.000  0.000
## extrcrrcl.Q  0.630  0.000  0.000  0.000  0.000  0.000 -1.000
## wrk_xprnc.L -0.120  0.000  0.000  0.000  0.701  0.003  0.000  0.000
## wrk_xprnc.Q -0.120  0.000  0.000 -0.001  0.701  0.003  0.000  0.000  1.000
## tm_ppld_fr2  0.000  0.027 -0.239 -0.188  0.000  0.000  0.000  0.000  0.000
##             wrk_.Q
## gender2
## gender3
## gpa
## covr_lttr.L
## cv.L
## extrcrrcl.L
## extrcrrcl.Q
```

```
## wrk_xprnc.L
## wrk_xprnc.Q
## tm_ppld_fr2  0.000
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## phase2 ~ gpa + cover_letter + cv + extracurriculars + work_experience +
##     team_applied_for + (1 | applicant_id)
##    Data: pass_phase1
##
##      AIC      BIC   logLik deviance df.resid
##     54.2     98.4    -17.1     34.2      603
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -47.526   0.000   0.000   0.000   1.433
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  applicant_id (Intercept) 0        0
## Number of obs: 613, groups:  applicant_id, 613
##
## Fixed effects:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -84.0078   139.2248  -0.603 0.546245
## gpa                  11.9646     3.1056   3.853 0.000117 ***
## cover_letter.L       41.2832    97.0898   0.425 0.670686
## cv.L                 35.3133   101.0148   0.350 0.726650
## extracurriculars.L  26.3280   107.5730   0.245 0.806653
## extracurriculars.Q  -8.0576    62.1464  -0.130 0.896839
## work_experience.L   14.0083   257.6880   0.054 0.956647
## work_experience.Q   -0.5500   148.7310  -0.004 0.997049
## team_applied_for2   -0.9407     0.9052  -1.039 0.298700
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) gpa    cvr_.L cv.L   extr.L extr.Q wrk_.L wrk_.Q
## gpa         -0.009
## covr_lttr.L -0.238 -0.010
## cv.L        -0.174 -0.009 -0.137
## extrcrrcl.L -0.384 -0.008 -0.081 -0.231
## extrcrrcl.Q  0.384  0.029  0.081  0.231 -0.999
## wrk_xprnc.L  0.748  0.029  0.178  0.214 -0.204  0.205
## wrk_xprnc.Q  0.749  0.018  0.178  0.214 -0.204  0.205  1.000
## tm_ppld_fr2 -0.003 -0.090  0.001  0.001  0.002 -0.003 -0.002 -0.001
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
## Data: pass_phase1
## Models:
## model_1b: phase2 ~ gpa + cover_letter + cv + extracurriculars + work_experience +
## model_1b:     team_applied_for + (1 | applicant_id)
## model_1a: phase2 ~ gender + gpa + cover_letter + cv + extracurriculars +
## model_1a:     work_experience + team_applied_for + (1 | applicant_id)
##          npar    AIC     BIC  logLik deviance  Chisq Df Pr(>Chisq)
## model_1b   10 54.181  98.364 -17.090   34.181
## model_1a   12 56.950 109.971 -16.475   32.950 1.2303  2     0.5406
```

Appendix 3.2 Model construction and ANOVA table for data in phase 2

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: phase3 ~ speaking_skills + leadership_presence + technical_skills +
##     writing_skills + (1 | applicant_id)
##    Data: pass_phase2
##
##      AIC      BIC   logLik deviance df.resid
##     65.8     88.0    -26.9     53.8      294
##
## Scaled residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -0.1476  0.0000  0.0000  0.0000  0.1607
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  applicant_id (Intercept) 1335     36.54
## Number of obs: 300, groups:  applicant_id, 300
##
## Fixed effects:
##                       Estimate Std. Error  z value Pr(>|z|)
## (Intercept)         -1.803e+02  2.337e-03 -77136.9   <2e-16 ***
## speaking_skills      5.801e+00  2.408e-03   2408.5   <2e-16 ***
## leadership_presence  7.917e+00  2.408e-03   3287.4   <2e-16 ***
## technical_skills     6.565e-01  2.276e-03    288.4   <2e-16 ***
## writing_skills       8.987e-01  2.253e-03    399.0   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) spkng_ ldrsh_ tchnc_
## spkng_sklls  0.000
## ldrshp_prsn  0.000 -0.252
## tchncl_skll -0.001 -0.003 -0.003
## wrtng_sklls -0.001 -0.003 -0.004 -0.032
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.0302665 (tol = 0.002, component 1)
```

```
## Data: pass_phase2
## Models:
## model_2b: phase3 ~ technical_skills + (1 | applicant_id)
## model_2c: phase3 ~ speaking_skills + (1 | applicant_id)
## model_2d: phase3 ~ leadership_presence + (1 | applicant_id)
## model_2e: phase3 ~ writing_skills + (1 | applicant_id)
## model_2f: phase3 ~ speaking_skills + leadership_presence + technical_skills +
## model_2f:     writing_skills + (1 | applicant_id)
##          npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## model_2b    3 74.672 85.784 -34.336   68.672
## model_2c    3 74.680 85.792 -34.340   68.680  0.0000  0
```

```
## model_2d    3 74.343 85.454 -34.171   68.343  0.3377  0
## model_2e    3 74.869 85.981 -34.435   68.869  0.0000  0
## model_2f    6 65.768 87.991 -26.884   53.768 15.1010  3   0.001732 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix 3.3 Model construction and ANOVA table for data in phase 3 and 4

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
##  Family: binomial  ( logit )
## Formula: phase4 ~ interviewer_rating_1 + (1 | applicant_id)
##    Data: pass_phase3
##
##      AIC      BIC   logLik deviance df.resid
##     24.9     28.1     -9.4     18.9       19
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.27242 -0.40886 -0.07847  0.49568  1.81400
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  applicant_id (Intercept) 0.1391   0.3729
## Number of obs: 22, groups:  applicant_id, 22
##
## Fixed effects:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -29.1832    12.1329  -2.405   0.0162 *
## interviewer_rating_1   0.3820     0.1587   2.407   0.0161 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## intrvwr_r_1 -0.999
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
```

```
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
##  Family: binomial  ( logit )
## Formula: phase4 ~ interviewer_rating_2 + (1 | applicant_id)
##    Data: pass_phase3
##
##      AIC      BIC   logLik deviance df.resid
##     18.3     21.6     -6.2     12.3       19
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.99296 -0.28608 -0.00995  0.27641  1.44931
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  applicant_id (Intercept) 0        0
## Number of obs: 22, groups:  applicant_id, 22
##
## Fixed effects:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -44.7627    21.1819  -2.113   0.0346 *
## interviewer_rating_2   0.5869     0.2786   2.107   0.0351 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## intrvwr_r_2 -0.999
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## phase4 ~ interviewer_rating_1 + interviewer_rating_2 + (1 | applicant_id)
##    Data: pass_phase3
##
##      AIC      BIC   logLik deviance df.resid
##      8.0     12.4      0.0      0.0       18
##
```

```
## Scaled residuals:
##        Min         1Q     Median         3Q        Max
## -4.773e-08 -1.490e-08 -1.490e-08  1.490e-08  4.942e-08
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  applicant_id (Intercept) 0.000625 0.025
## Number of obs: 22, groups:  applicant_id, 22
##
## Fixed effects:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -5.809e+03  1.654e+08       0        1
## interviewer_rating_1  3.369e+01  2.371e+06       0        1
## interviewer_rating_2  4.229e+01  1.686e+06       0        1
##
## Correlation of Fixed Effects:
##             (Intr) int__1
## intrvwr_r_1 -0.707
## intrvwr_r_2 -0.314 -0.448
```

```
## Data: pass_phase3
## Models:
## model_4a: phase4 ~ interviewer_rating_1 + (1 | applicant_id)
## model_4b: phase4 ~ interviewer_rating_2 + (1 | applicant_id)
## model_4c: phase4 ~ interviewer_rating_1 + interviewer_rating_2 + (1 | applicant_id)
##          npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## model_4a    3 24.860 28.133 -9.4299   18.860
## model_4b    3 18.319 21.592 -6.1595   12.319  6.5407  0
## model_4c    4  8.000 12.364  0.0000    0.000 12.3190  1  0.0004484 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```