# Statistical Analysis Regarding Black Saber Company Employment

potential bias in hiring and remuneration processes

Report prepared for Black Saber Software by ProDasta Consulting Company

2021-04-21

# Contents

```
## Warning: package 'lmtest' was built under R version 4.0.4
```

```
## Warning: package 'zoo' was built under R version 4.0.4
```

# Executive summary

### Background & Aim

When a company hires or wants to benefit its employees, there are more than one factor to take into account, and making the decision is relatively complicated. However, gender should not be considered as a factor to raise salary of an employee, or to hire an employee because he is a male. The concern of potential bias in hiring and remuneration process has been raised from the Black Saber Consulting Company, and therefore the data has provided from their data team for the purpose of analyzing the three processes: hiring, promotion and salary, as the company board would like to confirm the fairness of the three processes.

### Key findings

The results of the study are summarized below:

- those candidates who successfully passed Phase-I, 48% of them are male and 51% of them are female, with the rest 1% having a gender of "prefer not to say", which is a considerably fair percentage

- In Phase-II of the hiring process, models suggested *gender* is not used as a predictor of hiring

- In Phase-III, there are two candidates with the same score and of different genders. Even the female candidate has higher scores in *writing skills*, *leadership presence*, and *speaking skills* than the male candidate does, the male candidate is chosen by the HR instead of the female candidate.

- Based on the evidence from built models and visualizations, we believe a model with the predictor *gender* gives better result and explanation, hence gender is considered as a variable when assigning salaries

- Whether an employee gets a promotion depends on its productivity, leadership_for_level, and the gender. Gender makes a difference when considering the employee's leadership for level and

### Limitations

- As the company does not make hiring decision manually, instead goes through an AI system, it is unknown how the system values each skills and whether takes gender into account or not

- Several data has to be removed in the promotion model, as an employee's newest employment statements do not show whether he or she gets promoted in the next potential working employment period
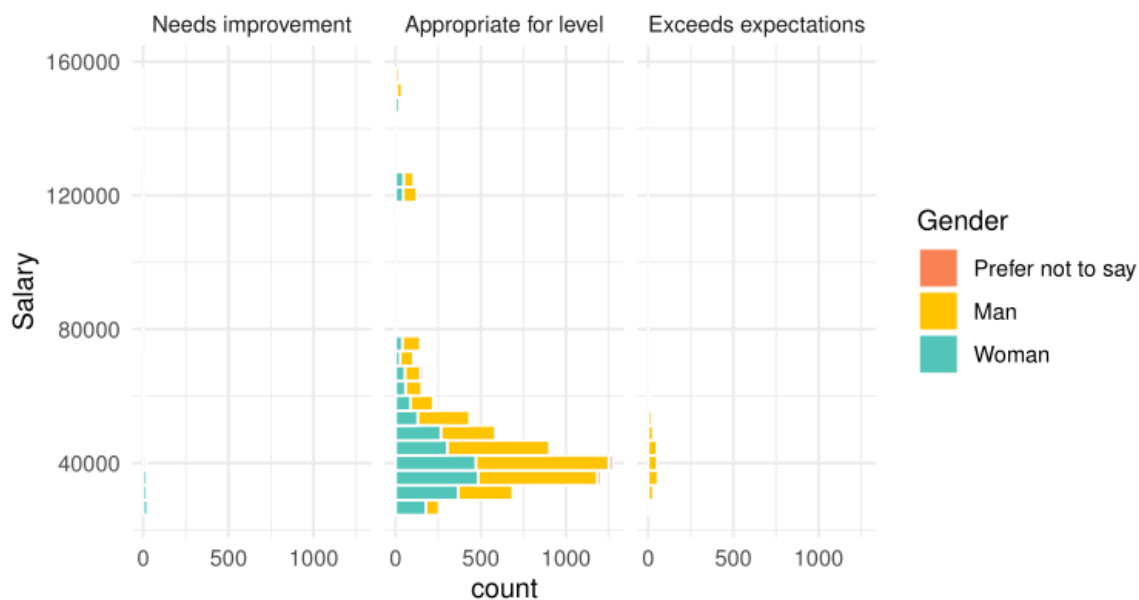
**Table**

Summary table 1: Conditional probability table for the three phases of hiring

| Gender | Phase 1 | | Phase 2 | | Phase 3 | |
|---|---|---|---|---|---|---|
| | Does not pass | Pass | Does not pass | Pass | Does not pass | Pass |
| Man | 0.467 | 0.483 | 0.468 | 0.682 | 0.583 | 0.800 |
| Prefer not to say | 0.026 | 0.010 | 0.011 | 0.000 | | |
| Woman | 0.508 | 0.507 | 0.522 | 0.318 | 0.417 | 0.200 |

**Graph**

Summary Graph 1: histogram for salary distribution by gender and leadership

# Technical report

## Introduction

As it is requested by the Black Saber Consulting Company due to the worry of potential biases existed within their employee's benefits and hiring process, the following report will be analyzing the data given by Black Saber Consulting Company regarding their company potential biases within hiring and remuneration processes. The report includes an abstract, data exploration, data analysis, model building, executive summary, result and conclusions. The study is built upon five data sets from Black Saber Consulting Company of different phases: three hiring phases, the final hires, and the current employee information. The report is done through the software R-studio and JupyterHub. In the data explore section, the data shall be examined at first to check for correlations and potential variables to be used for data analysis and model building section. The statisticians choose the best fitted model and explain in the model building section. Finally the result section includes graphs and plots of the final evidence to answer the raised research question, which will be answered in the conclusion section. The report is confidential and will solely be used for the Black Saber Consulting Company only.

### Research questions

- Is gender a significant determinant of being a successful applicant in each hiring phase?

- Is gender one of the considerations when making the decision whether an employee should be promoted?

- Is the amount of salary determined based only on the talents and performance of the employees?

**Is gender a significant determinant of being a successful applicant in each hiring phase?**

**Data and exploratory analysis**

- *cover_letter*: categorical, whether the applicant submitted a cover letter
- *cv*: categorical, whether the applicant submitted a CV
- *gpa*: numerical, applicant's GPA at school on scale of 4.0
- *gender*, categorical, applicant's gender, including man, woman and prefer not to say
- *extracurricular*: number of extracurricular activities the applicant have
- *work_experience*: number of work experiences the applicant have

Table 1: Conditional probability table for the three phases of hiring

| Gender | Phase 1 | | Phase 2 | | Phase 3 | |
|---|---|---|---|---|---|---|
| | Does not pass | Pass | Does not pass | Pass | Does not pass | Pass |
| Man | 0.467 | 0.483 | 0.468 | 0.682 | 0.583 | 0.800 |
| Prefer not to say | 0.026 | 0.010 | 0.011 | 0.000 | | |
| Woman | 0.508 | 0.507 | 0.522 | 0.318 | 0.417 | 0.200 |



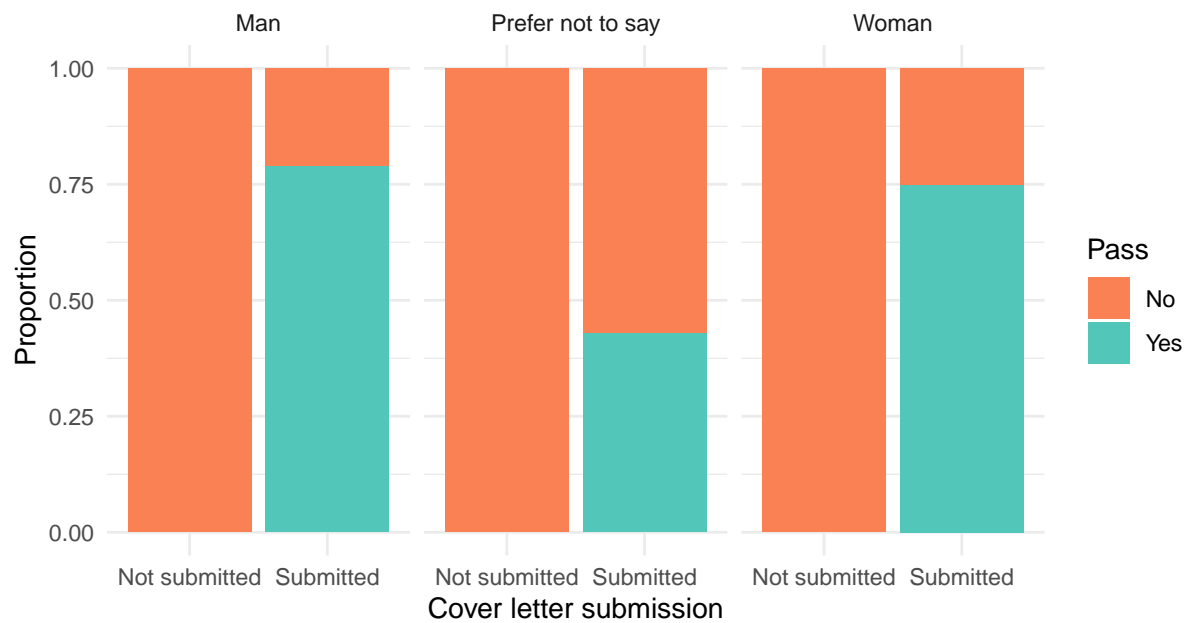**Figure 1:** Proportion of people passed phase I by gender and cv

**Figure 2:** Proportion of people passed phase I by gender and cover letter



**Figure 3:** Proportion of people passed phase I by gender and extracurriculars

By looking at the data for each phase, we can figure out which applicants successfully passed the previous phase by looking at the common *id*s between two phases. For instance, applicants

included in the phase II data set are the ones who passed phase I. Following this logic, we create a new variable in each phase that takes value of 1 if the candidate passed this phase, and 0 otherwise.

After creating the indicator variable, we then looked at the proportion of different gender passing each phase. Referring to *Table 1*, in phase I, 48.3% male, 50.7% female, and 1% prefer not to say passed this phase. There seems to be an even split between males and females who successfully passed the phase. Moving on to phase II, the split between males and females become a bit uneven, as among all those passed the phase, 68.2% are males while 31.8% are females, with 0% of prefer not to say. Lastly, in phase III, the ratio of males to females who passed this phase becomes 4:1. This preliminary observation leads us to think whether the selection algorithm is based solely on candidates' skills, or does gender affects the probability of a candidate being selected as successful candidate? If gender does play a role, then how big is the impact?

To investigate the question, we drew out some exploratory plots. We found that in the phase I selection, all the candidates who did not submit a *CV* or a *cover letter*, or did not have any *extracurricular* did not pass phase I, as *Figure 1*, *Figure 2*, and *Figure 3* shows. Hence, if a candidate did not submit these two files or does not have any extracurricular, then other potential determinants for success like *GPA* and *work experience* become irrelevant in determining whether the candidate passes the current phase. Therefore, we decide to filter out the observations that do not have a CV, a cover letter, or any extracurricular present in phase I, which reduces out sample size from 613 observations to 337 observations. Even though there is a large reduction in the sample size, we think the manipulation is necessary as we can investigate how other factors play a role in the selection process.

**Method**

For phase-I data, we mainly utilize summary tables and visualizations to determine whether there is a potential gender bias during the selection process, as the previous exploratory analysis already gives some insights to the problem. For phase-II data, we know that the AI focuses on the four skills that the candidates possess, so we build a model to see which factors are important in this phase. Since we have a binary response of whether each candidate passed phase-II, we lean on adopting a logistic regression model. To check the assumptions for the logistic regression model, we start by looking at our observations, which are all independent of each other. With a logit link, we expect there to be a linear relationship among the explanatory variables and the transformed response. Hence, we include each candidate's scores on *technical skills*, *writing skills*, *leadership presence* and *speaking skills* as our predictors in the model. With the predictors and

response specified, the model can be written out as follows:

$$log(\frac{\mu_2}{1-\mu_2}) = \beta_0 + \beta_1*\text{Technical skills} + \beta_2*\text{Writing skills} + \beta_3*\text{Leadership presence} + \beta_4*\text{Speaking skills}$$

$\mu$: the probability of a candidate passing phase-II $\beta_i$: coefficient for each variable

Finally, in phase-III of the hiring, the candidates have two interviews with two interviewers. The HR team will then decide which candidates to hire based on the interview scores. Due to a limited number of observations in the final phase, and a relatively small number of variables, we decide to use summary tables and visualizations to draw on conclusions.

**Results**

**Phase-I**    Table 2: Proportion of candidates passing Phase-I by gender with filtered observations

|                   | Does not pass | Pass |
| ----------------- | ------------- | ---- |
| Man               | 0.51          | 0.48 |
| Prefer not to say | 0.03          | 0.01 |
| Woman             | 0.46          | 0.51 |

Table 3: Median statistics by gender

| Gender            | Number | GPA | Extracurricular | Work experience |
| ----------------- | ------ | --- | --------------- | --------------- |
| Man               | 164    | 3.0 | 1               | 1               |
| Woman             | 169    | 3.1 | 1               | 1               |
| Prefer not to say | 4      | 2.4 | 1               | 1               |

From *Table 2*, we see that among those candidates who successfully passed Phase-I, 48% of them are male and 51% of them are female, with the rest 1% having a gender of "prefer not to say". Hence, from the table summary, it seems that the Phase-I hiring is pretty fair across genders.

*Table 3* displays the median statistics for *GPA*, number of *Extracurricular*, and the number of *Work experience* the candidates have by gender. There is a relatively even split between the number of *male* and *female* candidates, while *prefer not to say* only has 4 candidates. Despite this, the median values for each of the variables seem to be pretty consistent across different

genders. The only discrepancy appears to be *prefer not to say* having a lower *GPA* than the other two genders, potentially due to the small number of observations.

**Phase-II**  Table 4. Hiring Phase-II summary output

|  | Estimates | 95% CI | P-value |
|---|---|---|---|
| Baseline odds | 0.00 | (0, 0) | 0 |
| Technical skills | 1.08 | (1.04, 1.13) | 1e-04 |
| Writing skills | 1.09 | (1.05, 1.15) | 1e-04 |
| Leadership_presence | 2.53 | (1.77, 4.05) | 0 |
| Speaking_skills | 2.01 | (1.51, 2.87) | 0 |

From *Table 4*, we see that the effect of all four predictors are significantly positive, with the odds ratio for *technical skills*, *writing skills*, *leadership presence*, and *speaking skills* being 1.08, 1.09, 2.53, and 2.01. The p-values for the 4 predictors are all significant. Taking the odds ratio of *technical skills* as an example, it means that with a 1 unit increase in *technical skills*, the odds of being a successful candidate in phase-II is 1.08 times higher, holding other variables constant.

To investigate whether gender plays a significant role in the phase-II selection, we include *gender* as our predictor in addition to the original four. Then, by performing a Likelihood ratio test, we found that the p-value is around 0.41, suggesting that we lean more towards the simpler model that does not include *gender* as one of its predictors.
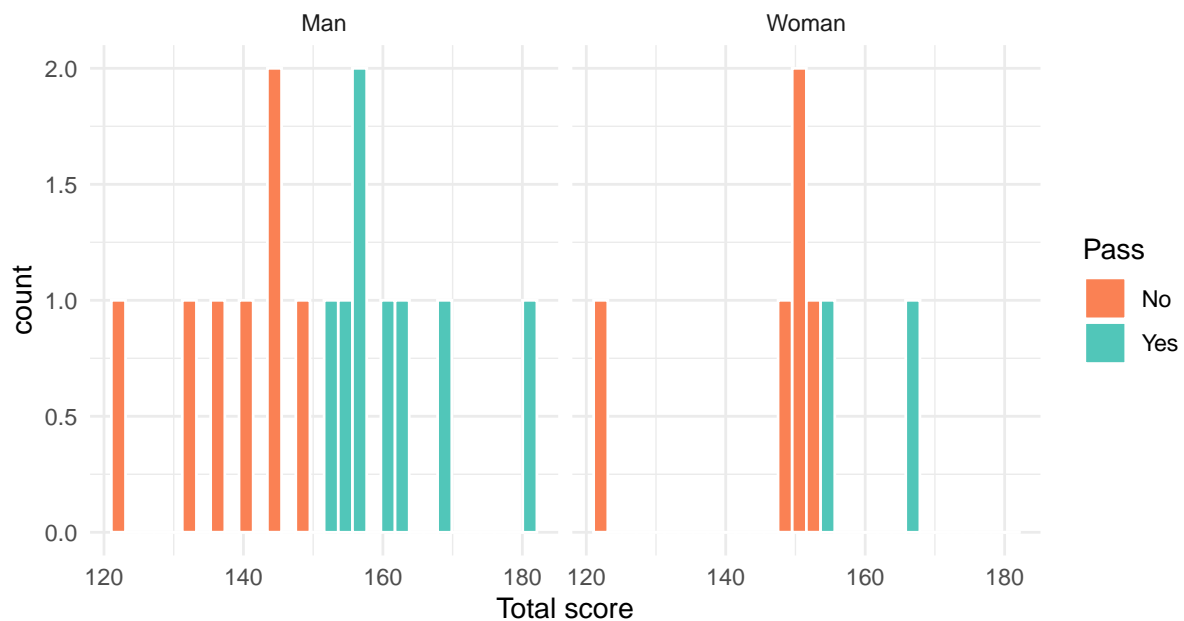
**Figure 4:** Histogram of the total scores vs. whether the candidate with the score passed phase-III

**Phase-III**    Table 5. Phase-II information for the two applicants with a tie in phase III

| Applicant id | Gender | Technical skills | Writing skills | Leadership presence | Speaking skills |
|---:|---|---:|---:|---:|---:|
| 1600 | Woman | 72 | 59 | 6 | 8 |
| 3560 | Man | 100 | 42 | 5 | 5 |

*Figure 4* shows the distribution of the total scores earned by candidates which is constructed from summing up the two *interview ratings*. From the plot, we see that most candidates with the highest scores are recruited. However, there are two candidates with a total score of 153, where only one of them is recruited who is a male. Hence, to explore further, we went back to the phase-II data and filter out these two observations and summarize the scores obtained by these two candidates in *Table 5*. From the table, the female candidate have higher scores in *writing skills*, *leadership presence*, and *speaking skills* than the male candidate does. On the other hand, the male candidate has higher scores in *technical skills* than the female candidate does.

**Conclusions**

As *Table 5* shows, the male has higher scores in *technical skills*, while the female is better at categories like *writing skills*, *leadership presence*, and *speaking skills*. According to our phase-II model, *leadership presence* and *speaking skills* seems to have larger effect on the odds ratio than the other two determinants do. So, we would expect the female to be selected as a successful candidate instead of the male.

In both Phase I and Phase II, we fit generalized linear models on different variables in each phase, and use likelihood ratio tests to find out whether gender plays an important role in passing rate. In phase I, we assess the impact of GPA, extracurricular activities and work experiences on passing status; for phase II, technical skills, writing skills, leadership presence and speaking skills are considered. Analyzing their odds ratios, these variables all have remarkable effects on whether candidates go to the next round smoothly. On the other hand, variable gender does not play a role here. The fitted models and the likelihood ratio tests tell us that gender does not have a noticeable impact on the final result, we tend to choose a simpler model that excludes the impact of gender here. Therefore, we do not think gender is a significant determinant of being a successful applicant in each hiring phase.

**Is gender one of the considerations when making the decision whether an employee should be promoted?**

- *employee id*: numerical, records as each employee's identity number
- *gender*, categorical, employee's gender, including man, woman and prefer not to say
- *team*, categorical, employee's team, including Client services, Data, Design Legal and financial, Marketing and sales, Operations, People and talent, and Software
- *financial q*, categorical, the working period of the employee
- *role seniority*, categorical, the employee's job position at the company
- *leadership for level*,categorical, quality of demonstrated leadership, taking into account of the role level of the employee
- *productivity*, numerical, work output of the employee, productivity of 50 are considered as satisfactory, and above 50 indicates better than expected

**Methods**

**Data Wrangling**

In the original current employees data, we were given the information of all *current* employees in Black Saber Company for their whole employment duration. The data were collected from 2013 Q2 to 2020 Q4, and it contained basic information of employees such as *gender*, *team* and *employee_id.* Moreover, for each employee, there were multiple quarterly observations recording his/her *productivity*, *leadership_for_level*, *role_seniority* and *salary.*

Before doing any data exploration, we made several changes to the given current employees data in order to better observe each employee's promotion situation. We created a new indicator variable named *get_promotion* (ie. if one get promotion in the following next quarter, then the indicator variable returns *1* for current quarter; otherwise, it returns *0*). In this way, it was necessary for us to delete each employee's the most recent quarter observation because they could not provide any relevant promotion information. The above manipulation made sense because the decision of promoting one or not supposed to rely on one's current quarter's performance. Also, we excluded the employees who filled in "Prefer not to say" for gender since here we were interested to investigate the effect of gender on getting promotion.

We filtered and focused on the observations only where employees got promotion and received "Appropriate for level" as feedback of *leadership_for_level.*Then, we further grouped them by *gender* and *role_seniority*, and calculated the average productivity for each group. Visually, from the following boxplot *Figure 3*, we could see that female had a relatively higher mean productivity than males did in almost every role level promotion stage. Here, we could interprete

such average productivity as the potential productivity threshold of promotion requirement. That was saying under the circumstance where males and females all received the same rating of *leadership_for_level* , females were often required to have a relatively higher average productivity than males did in almost every role level promotion stage.
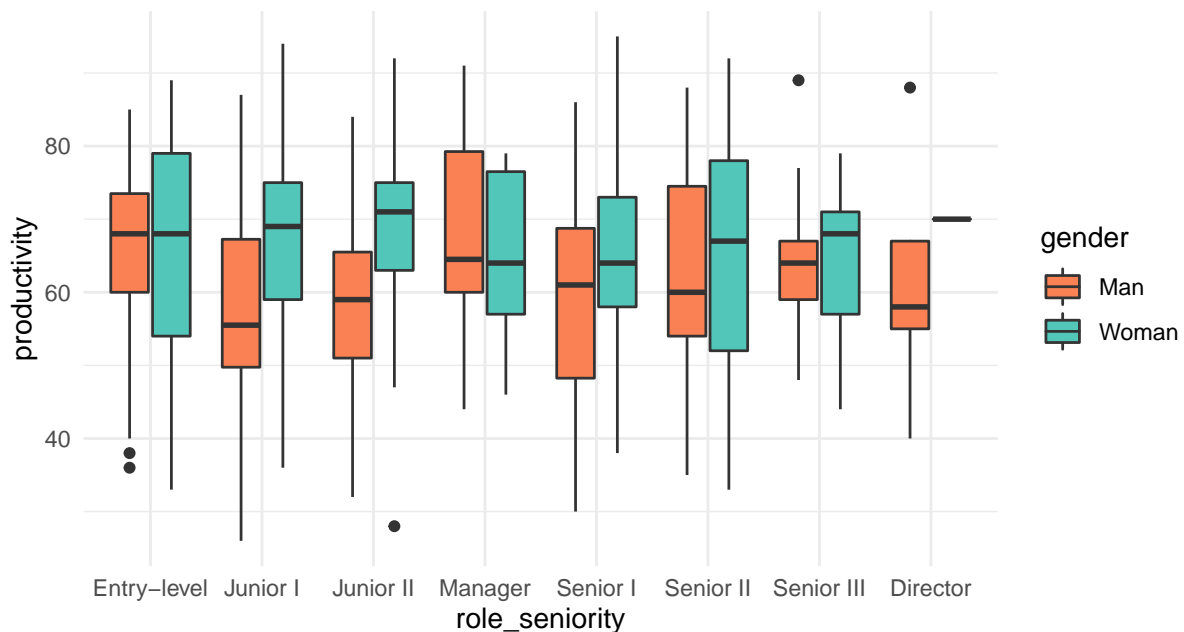


**Figure 5:** Boxplot for productivity and role seniority

Next,through a even more detailed division, we grouped observations by *gender*, *role_seniority* and also *team*. In *figure 4*, it can be seen that in many departments such as Data, Marketing and Operations, almost at every *role_seniority*, females often processed with a considerably higher productivity,further illustrating our previous discovery. Additionally, we noticed that in design department and director level in many departments, there were only males getting promotions. The above data exploration make us concern about whether the promotion decision is made purely on employees' performance, or *gender* had became one of considerations during promotion decision. Hence, we continued investigating through finding a proper modeling.
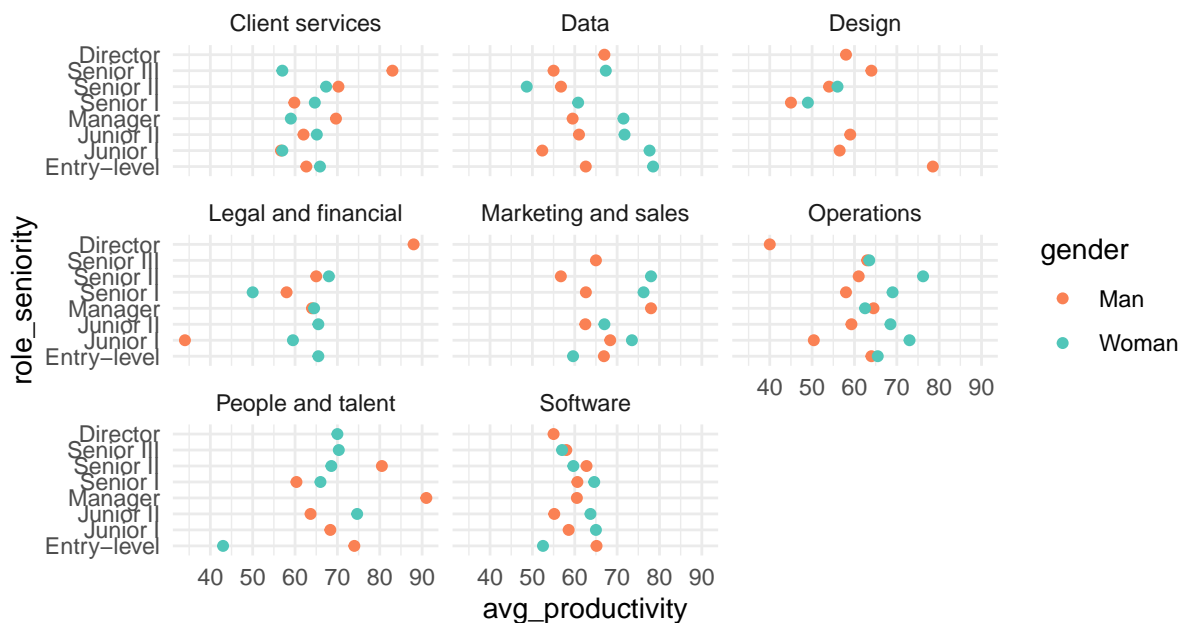
**Figure 6:** Plots of role seniority and average productivity

Table 6: Likelihood ratio test results for model with/without gender

|   | #Df | Loglik | Df | Chisq | Pr(>Chisq) |
|---|-----|--------|----|-------|------------|
| 1 | 5   | -1405.2 |   |       |            |
| 2 | 6   | -1402.5 | 1  | 5.205 | 0.02253    |

Since we were considering a binary response variable(*get_promotion*) and also having repeated quarterly measurements of each employees, we chose to use the logisric mixed effect linear model. In detailed, we considered *productivity* and *leadership role* as our fixed effects since they are measures of employees' value to the firm. Furthermore, we are interested in comparing which one of these two factor will have a larger impact. More importantly, we want to investigate whether *gender* is one of the consideration within one's promotion decision.

Thus, the proper model should be the following:

$$log(\frac{\mu_{ij}}{1 - \mu_{ij}}) = \boldsymbol{X}_{ij}\boldsymbol{\beta} + U_i$$

.

$\mu_{ij}$ and pij is the true probability that the employee i will get promotion in jth recorded quartely

measurement

$Y_{ij}$: binary response variable,indicating whether employee i got promoted or nor at jth repeated measurement

$\boldsymbol{X}$: model matrix for the fixed effects: *productivity*, *leadership role* and *gender*

$\boldsymbol{\beta}$: the effect of covariates on the mean outcome

$U_i$: random effect for specific employee

## Results

### check assumption of generalized mixed effect model

1.Our subjects(employees) are independent

2.Random effects come from a normal distribution There is an assumption that random effects—both intercepts and slopes—are normally distributed.

3.the random effects errors and within-unit residual errors have constant variance(are variance of data(transformed by the link function) homogeneous across data)

4.response is binary If a logistic link function is appropriate, these plots should be linear, and the stronger the linear association, the more promising the predictor

## Conclusions

**Is the amount of salary determined based only on the talents and performance of the employees?**

**Data & exploratory analysis**

We reformatted the values of *salary* in the original data set by taking off the "$" and "," and change the values from character values to numeric values for latter analysis. In addition to this, we re-level some factors for better display and ease of understanding. For instance, we re-level the factor of *seniority role* to reflect the ordering of the job title, from the least to the most senior.

Table 7. Mean productivity and mean salary by gender

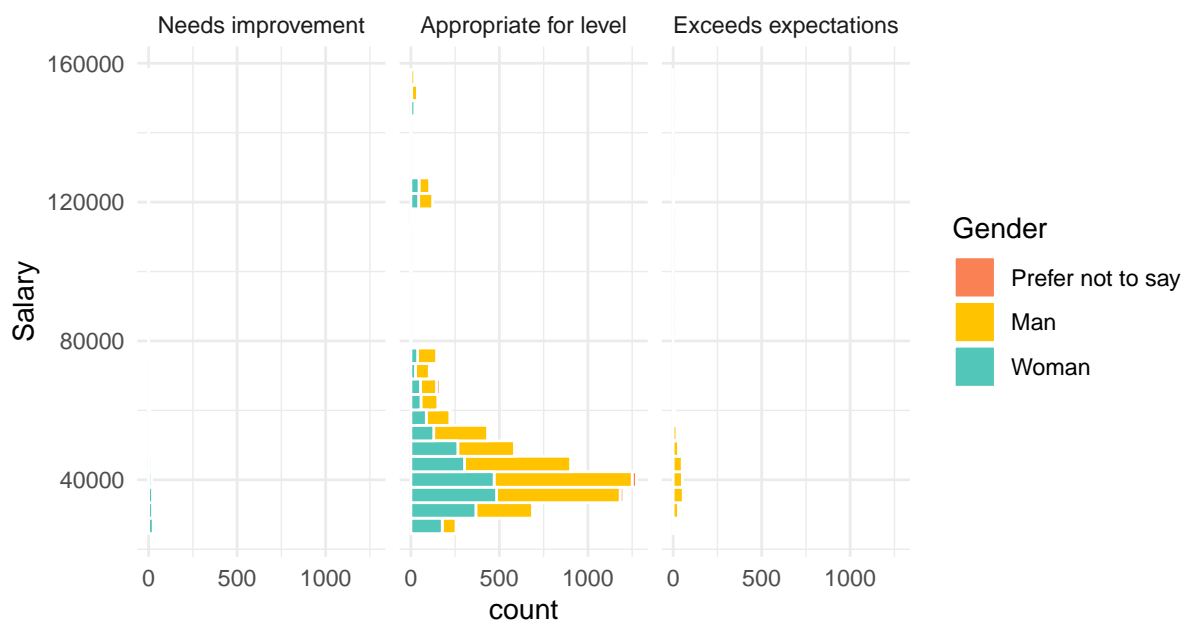| Gender | Mean productivity | Mean salary |
|---|---|---|
| Prefer not to say | 45.47 | 46340 |
| Man | 48.99 | 48491 |
| Woman | 50.44 | 45811 |



**Figure 7:** Salary vs. gender and leadership for level

*Table 7* shows the average *productivity* and *salary* by gender. From the table, we see that females

possess the highest average productivity of 50.44, but the lowest average salary among the three genders. On the other hand, males have the second highest average productivity of 48.99 and the highest average salary.

*Figure 5* displays the histogram of *salary* by *gender* and *leadership role*. Looking at the distribution of *salary* by *leadership role*, we see that the distribution is a bit right-skewed, which is common as there are usually fewer employees with very high salaries than the ones with lower salaries. One thing worth notice is that the employees that *exceeds expectations* are all males, while the employees that *needs improvement* are all females. Despite this, employees that *exceeds expectations* seem to overall have slightly higher salaries than the employees that *needs improvement* do.

**Method**

Since we have repeated measurement in the data set of all the employees and a continuous response variable: *salary*, we consider fitting a linear mixed model, treating the *ids* for the employees as random effects. The fixed effects we consider fitting in the model are *productivity*, *leadership role*, *seniority role*, and *gender*, as we are interested in whether the former three variables help quantify an employee's talents and abilities, while we want to figure out whether *gender* also plays a role in determining an employee's salary. Hence, the model we try to fit can be written out as follows:

$$Y_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + \epsilon$$

.

$Y_{ij}$: repeated measures j for salary on individual i, the response variable
$\boldsymbol{X}$: model matrix for the fixed effects: *productivity*, *leadership role*, *seniority role*, and *gender*
$\boldsymbol{\beta}$: the effect of covariates on the mean outcome
$\epsilon$: residual errors

To verify our assumptions for the usage of the model, firstly, we have a continuous response variable. Secondly, we identified *employee id* as our grouping variable resulting from the repeated measurement, so we include this grouping variable as our random effect. Lastly, even though the observations within each subject are not independent, the subjects themselves are independent.

**Results**

Table 8. Likelihood ratio test results for model with/without gender

|   | #Df | Loglik | Df | Chisq | Pr(>Chisq) |
|---|-----|--------|----|-------|------------|
| 1 | 14  | -58728 |    |       |            |
| 2 | 16  | -58671 | 2  | 114   | <2.2e-16   |

Table 9. Model summary for salary

|                                   | Estimates | 95% CI |
|-----------------------------------|-----------|--------|
| Intercept                         | 31900.93  | (29770.89, 34030.94) |
| Productivity                      | -0.93     | (-3.28, 1.42) |
| Leadership: exceeds expectations  | -249.42   | (-432.63, -66.17) |
| Leadership: needs improvement     | -252.50   | (-487.01, -17.98) |
| Gender: man                       | 679.10    | (-1471.29, 2829.51) |
| Gender: woman                     | -2022.17  | (-4182.54, 138.24) |
| Junior I                          | 5486.45   | (5362.26, 5610.58) |
| Junior II                         | 7864.69   | (7719.62, 8009.68) |
| Senior I                          | 13608.54  | (13453.28, 13763.67) |
| Senior II                         | 18995.86  | (18827.68, 19163.9) |
| Senior III                        | 24500.16  | (24315.4, 24684.8) |
| Manager                           | 40518.27  | (40287.17, 40749.22) |
| Director                          | 90809.06  | (90516.54, 91101.4) |
| Vice president                    | 120787.45 | (120349.4, 121225.18) |

After building models that include *employee id* as only random intercept and random intercept and slope (interacting with *gender*), the likelihood ratio test tells us to adopt the model with only the random intercept (p = 0.9246). Furthermore, to determine whether *gender* plays a significant role in the determination of salary, we compare our current model with a nested model that have the same covariates as in our current model but without *gender*. The result of the model comparison using the likelihood ratio test is presented in *Table 9*. We see that the p-value appears to be really small (< 2.2e-16), so we expect having *gender* to improve our model by a significant amount.

By including *gender* in our model, the summary outputs for the final model is presented in *Table 8*, we see that the grand average of the salary is 31900.93. Predictors that are significant are: *leadership role: exceeds expectations*, *leadership role: needs improvement*, and all levels of the *seniority role*, since the 95% confidence intervals for these coefficients do not include 0. Among these important covariates, both levels of the *leadership role* have large negative effects, while all levels of the *seniority role* have large positive effects on salary.


**Conclusions**

*discuss how lrt interacts with our figure of leadership_role by gender*

From the visualizations and model above, the relationship between salary with covariates like productivity, leadership role, seniority role and gender shows that male employees tend to receive more benefits than females do. The visualization *Figure 5* and the proportion on average productivity exhibits unfairness between different genders. Between genders and leadership roles, some females think that salary still needs to be improved, while the male leaders think the opposite, that their benefits have already exceeded their expectations. Moreover, females who have the highest productivity but turn out have the lowest average salary, while the males have lower productivity resulting in a much higher average salary. The result from the linear mixed model we fitted, one with gender as a covariates and one without, the likelihood ratio test shows that gender does play a crucial role in salary. Holding other variables constant, female employees have 2022.17 dollars less than the average salary, while males have 679.10 dollars even more than the average.

There is a slight discrepancy that our likelihood ratio test and 95% confidence interval give us different results. The likelihood ratio test says we should use the one with gender included, but 95% C.I. gives us the opposite as it is insignificant. Based on the evidence from previous proportion tables and visualizations, we tend to believe that there are still some differences when gender comes in, both in hiring processes and employees' benefits.

## Discussion

Through model building and data analyzing, the three research questions are thoroughly answered in the process of data analyzing. Gender is not a very determinant factor in the first two hiring processes, however when it comes to the third phase, it has come to the attention that females are rejected when having the same score with males. When considering promoting an employee, the model suggested that gender is of significant. Same for the salary model, gender is a significant factor.

## Strengths and limitations

As the models for each process is being built and analyzed, there are several key findings to summarize and answer the research question. In the first phase of the hiring process, there seems to be no difference between male and female candidates. While in Phase-II, gender makes a difference in the model as the model with gender as a predictor makes more sense. Lastly, in Phase-III of hiring process, one key finding is that for two candidates with the same score but different genders, the male candidate is chosen instead of the female candidate. However, the two candidates have different scores for different skills, and since the hiring process is determined by an AI system instead of by person, it is unknown of how the AI system values the skills. For the promotion model, there are limitations to the data, as the first line of each employee represents its newest state of employment, hence it is impossible to know whether the employee gets promoted or not as we do not have the next period's data. Therefore, each employee's newest employment line needed to be removed. Lastly, for the salary model, the key findings is that female employees have 2022.17 dollars less than the average salary, while males have 679.10 dollars even more than the average, holding everything else constant. Even the models are not perfect and involves a degree of the statistician's interpretation of the data set, the evidence is able to confirm the different of gender makes a difference in the employee or candidate's treatment.

# Consultant information

## Consultant profiles

**Ke Deng**. Ke Deng is a junior analyst at the ProDasta Consulting Company, where she does report writing and data analysis for clients. She graduated from the University of Toronto with a Bachelor degree of Science, and is currently enrolled in the graduate program. Ke specializes in Statistics, with a focus in economics. She has been with the ProDasta Consulting Company since the company started. Ke is currently employed as the report writer and consultant of the company and she is great at writing and polishing reports for the clients.

**Wenqing Hao**. Wenqing Hao is an experienced analyst at the ProDasta Consulting Company, also one of the four co-founders of the Company. She holds a Bachelor degree of Science from the University of Toronto, focusing on both Statistics and Financial economics. She has a strong background of managing business operations and finance projects. Over the past few years, she cooperates well and closely with the other three co-founders, generating a rather professional team to successfully satisfy clients' various requests.

**Qihui Huang**. Qihui Huang, an junior analyst working at ProDasta Consulting Company. Qihui has studied Statistics at University of Toronto for 3 years, and aims to graduate with a Bachelor of Science degree in Statistics. She also has 1 year experience in data related roles. In her study and work, Qihui developed a time management guide to help her co-workers manage stress and emotion. On weekends, she likes to go hiking with her friends or reading books.

**Qing Wen**. Qing, commonly referred to as Chelsea, is an analyst at the ProDasta Consulting Company. She graduated from the University of Toronto with a degree in applied statistics, focusing on economics. Qing joined ProDasta during her pursuit of a graduate degree in business analytics. With 3 years of experience working with clients, Qing handles clients' requests with her passion and expertise. She enjoys communicating with the clients in casual ways to solve their problems using her outstanding analytical skills, as well as helping colleagues to improve the productivity of the whole team.

## Code of ethical conduct

The following statements represent the ProDasta Consulting Company's code of ethical statements:

1. ProDasta Consulting Company seeks to have high professional standards within the development, usage and enforcement and keeps a high standard for having the reputation of statistical practices. We state to do the best of setting standards for the Statistics field and for Statisticians, avoid any wrong actions.

2. ProDasta Consulting Company and the employees take on the responsibility of work. We refuse to deliver any wrong statements that can cause damage to the professional statistical fields.

3. We support and welcome any statisticians in their professional field. We act with integrity and dignity toward the fellow statisticians, and provide opportunities for them to the professional field.