
Report on Potential Bias in Recruitment and Remuneration Processes

The gender bias is evaluated in both hiring and employee compensation given the data provided by Black Saber Software

Report prepared for Black Saber Software by Sun Enterprises

2021-04-21

Contents

Executive summary	3
The Problem	3
Potential Bias	3
Recruitment Process	3
Employee Compensation	4
Technical report	5
Introduction	5
Role of Gender in the Hiring Process	5
Effect of Gender in Earning Higher than Average Pay	7
Role of Gender in Salary	8
Discussion	10
Consultant information	12
Consultant profile	12
Code of ethical conduct	12

Executive summary

The Problem

There may be a potential bias in Black Saber's hiring and remuneration process. These problems have become apparent in Black Sabre's field of work, thus, people were raising questions regarding Black Sabre's fairness in these processes.

Potential Bias

In this report several dimensions of potential bias in analyzed in terms of hiring and compensation. The main concern regarding the potential bias is gender. Thus, gender's effect on hiring and employee compensation is examined in order to observe the potential bias against different genders.

The gender is reported in both applicants and the employee data set, however, there are only two gender's reported and an option to opt out from this indication. This limits the examination of bias against non-binary genders which may be part of the bias in the work place. Moreover, ethnicity variables are not included in these data sets which may also be related to potential bias in the workplace.

Recruitment Process

According to our analysis we observe some possible bias in the hiring process against applicants that are women. Although, we are unaware of the contents of the cover-letter and the CV, we are provided with several indicators that show the participation in extracurricular activities, past work experiences and individual's GPA. The bias is only apparent in the first phase of the application process. Submitting a cover letter or a CV is not considered for this process because there are no applicants that were accepted without submitting either one of those documents. When we control for variables that may rightfully influence the recruitment we observe that the woman applicants are less probable to be hired when compared to man applicants. Odds for being hired as a woman applicant relative to the odds for being hired as a man applicant when controlled for these variables is around 20%, suggesting that females with similar GPA, extracurricular activities and work experiences are much less likely to be selected. This, may be due the bias in the selection process.

Employee Compensation

The potential gender bias in employee compensation is also a concern of this report and two of the research questions examine the relationship between gender and salary. The initial examination regarding gender and salary is determining if gender influences earning higher than average salary in Black Saber. According to our analysis gender does play a role in getting higher than average salary. We observe that again being a woman, compared to being man, depreciates the probability of earning higher than average.

In the final research question, the extent of the gender's effect on salary is examined on the whole data set and 2020 financial quarters. The reason behind the focus on 2020 quarters is to see if this effect is apparent in the most recent year. The analysis suggests that there is, indeed, an effect of gender on salary. Woman employees earn less than then man when we control for variables that is likely to influence pay such as productivity and leadership for individual's role. Furthermore, leadership is assessed by individual's superordinate, hence, this evaluation may also be biased. We observe that leadership for man either meets the criteria or exceeds the expectation, never sub-par. This is opposite for the females. No female has received an exceeds expectations stamp in terms of leadership from their superordinates. They received either below expectations or meets expectations stamps which may entail some bias within itself. Nevertheless, the model shows that a woman employee with similar productivity and leadership for level is expected to earn approximately \$5,200 less than a man employee which is worrisome and may highlight bias in the employee compensation. The effect is apparent in all financial quarters of 2020 suggesting that the effect size does not necessarily stem from earlier years. Moreover, this effect is slightly larger in financial quarters of 2020 where the gender gap rises up to 6,000 dollars.

Overall, there is likely to be bias in hiring and compensation processes in Black Sabre. The potential bias against women is most apparent in the first phase of the hiring process. Moreover, this bias is apparent in employee compensation where there is a considerably large gender salary gap for employees with similar productivity levels and leadership abilities.

Technical report

Introduction

The report covers the potential gender bias in hiring process and employee compensation process in Black Saber Software. The bias can be on different characteristics that are unrelated to the individual's productivity, leadership ability, and skills. There are several potential characteristics that are likely to introduce bias in these processes, two of which are gender and ethnicity. This type of taste-based discrimination is undesired in the workplace and has a negative effect on the individual. The data related to ethnicity or race is absent in the provided data sets, hence, we cannot focus on the bias related to one's race. Thus, the main focus of this research is to analyze how gender plays a role in recruitment and compensation in Black Saber.

Research questions

- To what extent gender plays a role in the hiring process when controlled for recruitment variables in all three application phases?
- Is getting higher than average pay depend on individual's gender when controlled for role seniority, leadership ability for individual's level and their productivity?
- To what extent is the employee compensation is related to the individual's gender?

Role of Gender in the Hiring Process

For this question, I have started by gathering all three phases in one data set in order to create the `applications` data set. By doing so, it is quicker and easier to run the desired models in a complete data set, where the values for applicants who did not make the next phase are missing (NA). However, this does not introduce a problem in the analysis because the model is designed to ignore these observations to focus on applicants in each phase when specified. Moreover, I have also combined `hired` individuals in this data set as a binary variable. This entails that applicants who were not hired have all zeros and 10 applicants that were hired have 1 as an indicator. This is done so that our model which has being hired as a response variable can successfully run.

There are four models designed for this research question. The first three of these models binomial and the last one is a poisson. The first three models are:

$$H_i \sim \text{Binomial}(N_i, p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = X_i\beta_i$$

where,

- H_i is the outcome variable of being hired or not
- N_i is the number of applicants in each of these phases
- p_i is the probability of a specific gender being hired
- X_i is the covariates in each of these models which depend on the variables assessed for recruitment in each phase
- β_i is the parameter that shows the relationship between the covariates and being hired

In the first model of this section there are four covariates which are applicant's GPA, level of extracurricular activities, past work experience and gender. Thus, there are five parameters in this model which are $\beta_0, \beta_{GPA}, \beta_{EC}, \beta_{WE}, \beta_{gender}$ which correspond to the intercept and the covariates respectively. In this model, and the later models in this section, our parameter of interest is β_{gender} in order to determine the effect of gender in the recruitment.

In the second model, which is again binomial, there are additional covariates which are technical, writing, leadership and speaking skills. Thus, this model includes the previously mentioned covariates in model one whilst capturing the variables from the second phase of the application process. This results in a decrease in the number of observations from 613 to 313 which is due to the elimination process after the first round.

In the third model, the phase 1 variables are dropped. Hence, this model has four different skill assessment results from the second phase and gender as covariates so there are six parameters including the intercept. Again, the number of observations are 313.

The fourth model is:

$$Y_i \sim \text{Poisson}(\lambda_i)$$

$$\begin{aligned} \log(\lambda_i) = & \beta_0 + \beta_{GPA}(GPA_i) + \beta_{EC}(EC_i) + \beta_{WE}(WE_i) + \beta_{TS}(TS_i) + \beta_{WS}(WS_i) + \beta_{LS}(LS_i) \\ & + \beta_{SS}(SS_i) + \beta_{R1}(R1_i) + \beta_{R2}(R2_i) + \beta_{gender}(G_i) \end{aligned}$$

where,

- Y_i is the outcome variable of being hired or not
- λ_i is the probability of a specific gender being hired

Here parameters, β_i s, and covariates, X_i s, are written explicitly with the acronyms associated with these coefficients and values. There are ten variables and eleven parameters in this model. Moreover, the variables are GPA, extracurricular activities, work experience, technical, writing, leadership and speaking skills and ratings from two interviewers, and finally gender. I have avoided writing the first three models in this format due to the untidiness introduced by indicating

all variables explicitly. However, including all of the variables in the last model hopefully helps the illustration of these regression models.

Here I have used a generalized linear model to examine the relationship between gender and the recruiting process. There are three phases in recruiting and model is fit for the variables in each phase with the indicator of being hired or not. These three models are binomial due to the fact these events are discrete counts of people who were hired and we are interested in the probability and the odds ratios of being hired for the genders. The fourth model in this question is poisson due to the fact that it includes all of the variables from all three phases. The reason why this model is poisson is that combining these phases shifts the domain from discrete to continuous. Moreover, the mean and variance for the variable `hired` is almost equal, approximately 0.016.

This analysis indicates that woman applicants are less probable to be hired when controlled for phase 1 variables. The odds ratio is around 20% suggesting that females are much less likely to be hired. In this model there are three statistically significant factors in being selected by Black Sabre which are GPA, extracurricular activities and woman as individual's gender. Although, the first two are positively associated with the probability of being selected in terms of log odds, association for woman as a gender and being hired is the opposite. Although the effect is no longer statistically significant this disadvantage against women continues in the later phases where in the later phases the odds ratio drops to 10%. When we include all the variables with a poisson regression we do not observe any statistically significant effect on being hired possibly due to the complexity of the recruitment process. However, the log odds are still negative for woman applicants in this regression.

Effect of Gender in Earning Higher than Average Pay

For this question, I have examined the effect of gender in getting higher than average salary. In order to do so I manipulated the data sets by creating a variable called `sal_above_average`, a binary indicator showing if the employee in the given quarter is earning above the average salary which is approximately \$47,400. Additionally, I have separated the data set into quarters so I can analyze the effect of gender in getting higher than average salary in recent years and quarters, especially for 2020 financial quarters.

The model in this section is again binomial. This model is adopted the whole data set, then fourth financial quarter in 2020, then third, then second and lastly the first financial quarter in 2020. The reason behind this selection is to see the effect of gender in the most recent year. Moreover, to determine if the effects may be due to earlier decisions in compensation or are they still observable in recent years.

$$S_i \sim \text{Binomial}(N_i, p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_{RS}(\text{RoleSeniority}_i) + \beta_P(\text{Prod}_i) + \beta_L(\text{LeadLevel}_i) + \beta_{\text{gender}}(G_i)$$

- S_i is the binary outcome variable of being paid higher than average or not
- N_i is the number of employees in Black Saber
- p_i is the probability of a gender being paid higher than average or not
- RoleSeniority_i is the level of the role in the firm ranging from Entry-level to Vice President
- Prod_i is the rated productivity of the individual ranging from 0-100
- LeadLevel_i is the rating given by a superordinate for the leadership level of the individual in their current role
- G_i is the gender of the individual either man, woman or prefer not to say
- β are the parameter that shows the relationship between the covariates and being hired
- β_{gender} is the main parameter of interest that shows the relationship between gender getting higher than average salary

This model is then applied to the whole employees dataset, and then four quarters in 2020. For this analysis I have used generalized linear mixed-effects model where role seniority is a random effect. This model would make most sense if there were random intercepts for role seniority which determines the initial value of salary without considering productivity and leadership for level. Productivity and leadership for level is included as covariates in this model and gender is our parameter of interest. Again, this model is binomial due to discrete salaries. We are not considering salary as a continuous component here due to the fact that only first two integers are likely to change.

Again, we observe that women are disadvantaged in this dimension of compensation. The odds ratio for probability of earning higher than average for a woman employee compared to a man employee is approximately 50%. This suggest that there is potentially a gender gap in Black Sabre. We would not expect a gender gap if Black Sabre was unbiased in the employee compensation process, hence, it may need to be reevaluated. This effect is again statistically significant when we consider all 6906 observations in the employees data set. The odds ratio drop to around 40% when we consider the 2020 financial quarters which highlights the potential bias in recent years, hence, the difference in probabilities for man and woman do not necessarily stem from earlier year but much likely to be still apparent in the workplace.

Role of Gender in Salary

In this question we have examined how much of the variability in salary is caused by gender to see if gender has an influence in the compensation process. This is somewhat of an extended

version of the second research question. Now the outcome variable is not binary. Rather than salary is thought to be discrete. In order to make this analysis, I have used the quarters I have made for the previous question and turned salary into a numeric variable in the data set. Again, I have made this analysis on the employees data set and each quarter in 2020.

The model that is used for this analysis is:

$$Y_{ij}|U \sim N(\mu_{ij}, \sigma^2)$$

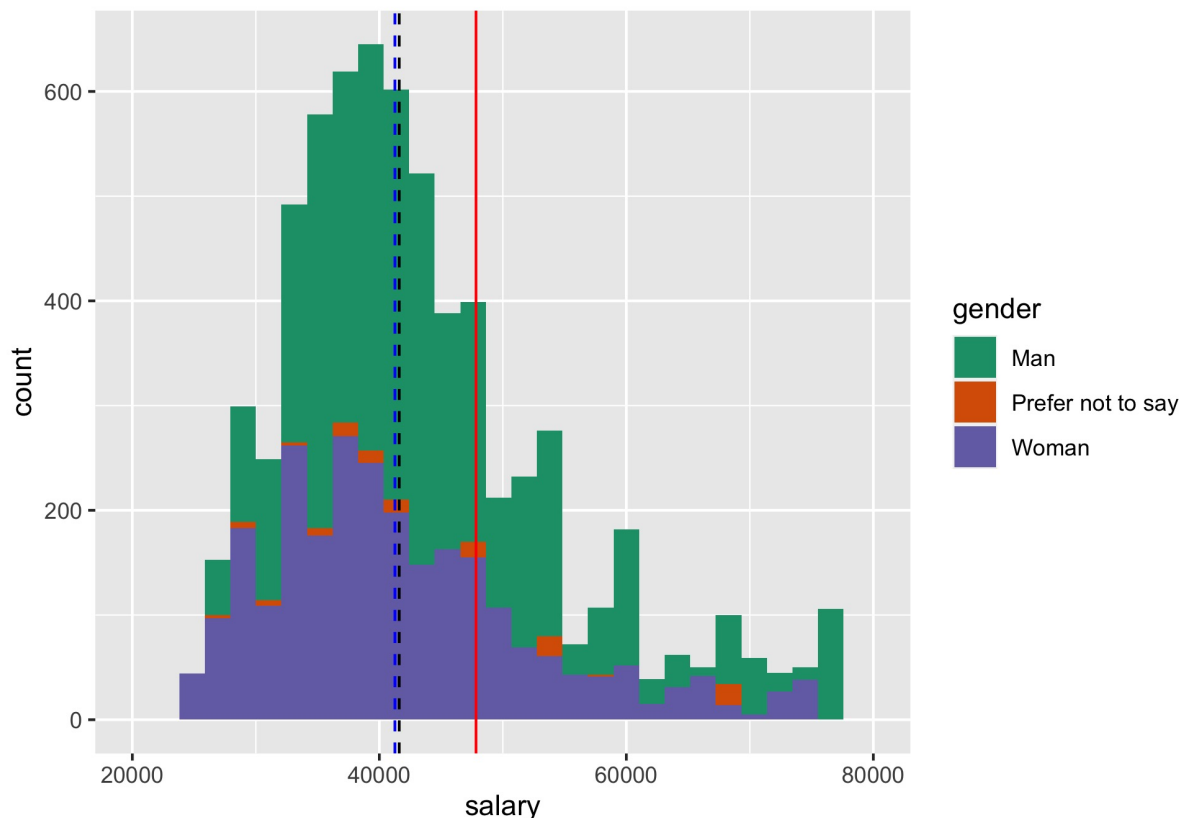
$$\mu_{ij} = X_{ij}\beta + U_i$$

$$U \sim MVN(0, \Sigma)$$

- Y_{ij} is the salary of the individual
- U is the random effects in the model which depend on the experiment, hence, not specified for now
- X_{ij} are the covariates some of which are productivity, leadership for level and gender
- β are the parameters where our parameter of interest is, again, β_{gender}

Here I have observed to what extent gender is related to salary, or the employee compensation process. For this question I have used a linear mixed effects model. The random effects in this model are `employee_id`, `financial_quarter` and `team`. Moreover, the fixed effects are `productivity`, `leadership_for_level` and `gender`. I expected both productivity and leadership to be positively associated with salary, however, it was not. The appropriate level of leadership is more preferred compared to exceeding the expectations in terms of leadership in the compensation process. Moreover, in some of the financial quarters additional productivity was also negatively effect with salary. Although these are not related to our parameter of interest, it is still important to highlight these features. Being woman is, again, associated with lower pay after controlling for variables. These effects are statistically significant in the whole data set and for each of the quarters in 2020. The salary gap ranges from 5,200 to 6,000 dollars which is a big enough difference to suggest the possibility of a gender gap, hence, bias in terms of one's gender.

Below given the histogram of salaries under \$80,000 with different gender indicators. The red solid line corresponds to salary average of man, black dashed line corresponds to the people who chose not to indicate their gender and blue dashed line for women. This plot highlights the salary gap between man and woman employees.



Discussion

In this analysis, I have focused on how gender plays a role in recruitment and employee compensation in Black Saber. The overall mission of this analysis was to analyze the potential biases in this processes and due the missingness of the ethnicity variable, I have analyzed how gender effects hiring phases, getting higher than average pay, and salary in general.

In the first analysis, we have observed bias towards female applicants in the first phase of the hiring process. When controlled for GPA, extracurricular activities, and work experience female participants had lower change of being selected for the second phase when compared to their male counterparts. The same effect is not apparent for individuals who prefer not to indicate their gender. However, this effect is not apparent in the later phases of the hiring process for any of the genders but limiting the entry of the specific gender after first phase is likely to play a role in the whole hiring process.

In the second analysis, where probability of earning higher than average salary is compared for genders. We observe that females are less likely than males to get higher than average salary in Black Saber. Moreover, this is not due to the earlier compensation processes. When we examined

this model on each 2020 quarter we still observe that the females are less likely to earn higher than average pay when controlled for role seniority, productivity and leadership for level. In addition, for individuals who prefer not to indicate gender we observe the opposite where getting higher than average salary is more likely when compared to males, hence, the females. Although, we cannot make a claim about this since individuals who do not prefer to indicate gender may also be male or female.

In the third analysis, we have again controlled for several variables that should entail salary which are productivity and leadership for level. Here we have observed that the being a women plays a more important role in compensation than the leadership for level. This relationship between being a women and salary is negative. This relationship is also apparent for individuals who do not indicate their gender but, again, it is challenging to comment on this. Moreover, this relationship is not as significant as the relationship between being a woman and salary in terms of its t-value. This effect persists in the analyses that are done on each of the quarters in 2020 suggesting that this potential bias against woman in workplace in terms of compensation is most likely still apparent.

Strengths and limitations

The strengths of this analysis is that there are several different models fit to data which enables us to consider different dimensions related to potential biases. The main potential bias that we examined is gender which is coded in 3 ways in the data set which are man, woman and prefer not to say. This enables us to consider the differences in compensation and recruitment across two different genders.

The limitations are due to the reporting of the gender. Gender, unlike sex, is multi-dimensional and having these indicators in the data set would have been helpful in our analysis and help us examine the different effects for each gender. Here due to preference of not reporting gender we are unable to consider the prefer not to say group effectively. In addition, it is apparent in today's world that potential biases may arise from ethnicity which is not reported in the data set. Having that variable would have enable us to consider how race plays a role in recruitment and compensation.

Consultant information

Consultant profile

Ilke Sun. Ilke is a senior consultant with Sun Enterprises. He specializes in Monte Carlo methods in finance and Bayesian modeling. Ilke earned his Bachelor of Science, Major in Statistics and Economics, from University of Toronto in 2021. Furthermore, earned his Master of Science, Major in Mathematics and Statistics, from University of McGill in 2024.

Code of ethical conduct

Sun Enterprises take pride in it's policy of no discrimination against any gender, race or religion. The ethical statisticians in Sun Enterprises use methodology and data that are relevant and appropriate; without favoritism or prejudice; and in a manner intended to produce valid, interpretable, and reproducible results. Our mission is to be unbiased and fair in order to have factual and impartial reports. While making these reports the customer's needs, fairness of our reports and the ethical outcomes of our analyses are considered to make the best possible work that we can.