Analysis of fairness based on talent and value in the hiring and remuneration processes

A special focus on gender parity

Report prepared for Black Saber Software by SFP

2021-04-21

Contents


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

## Warning: package 'dplyr' was built under R version 4.0.4

## Warning: package 'gtsummary' was built under R version 4.0.5
```

Executive summary

Background & Aim :

Concerns about potential bias in the hiring and remuneration processes were consistently raised internally in Black Saber Software. In order to investigate the true fact for the issues, we SFP were consent to help figure it out. This report was commissioned to examine if potential issues especially gender parity existed in the hiring, promotion and salary processes based on the hiring data and data about promotion and wages for current employees provided by Black Saber Software . Besides, we were interested in the fact how the new trialled AI service performanced in the hiring process.

Key results :

- final result vs. productivity
- final hiring result vs. gender
- AI result good?
- promotion vs. productivity
- promotion vs. gender
- salary vs. productivity
- salary vs. gender

Limitations:

We were only offered with data for current employees so it was hard to know reasons past employees left. If additional data related to past employees can be provided, we can have a more deep and comprehensive sight into potential biases;
The lack of ethnicity/race data also results in a limitation since race might be a potential exist which might influence the results of hiring pipelines and promotion and amount of wages;
No information about the determination methods for hiring results, promotion and amount of salary was provided which limited the accuracy of models we fit for these processes.

Figures:

Figure1: Proportion of Applicants who were refused in eac

Key results are summarized in the following tables or graphs.

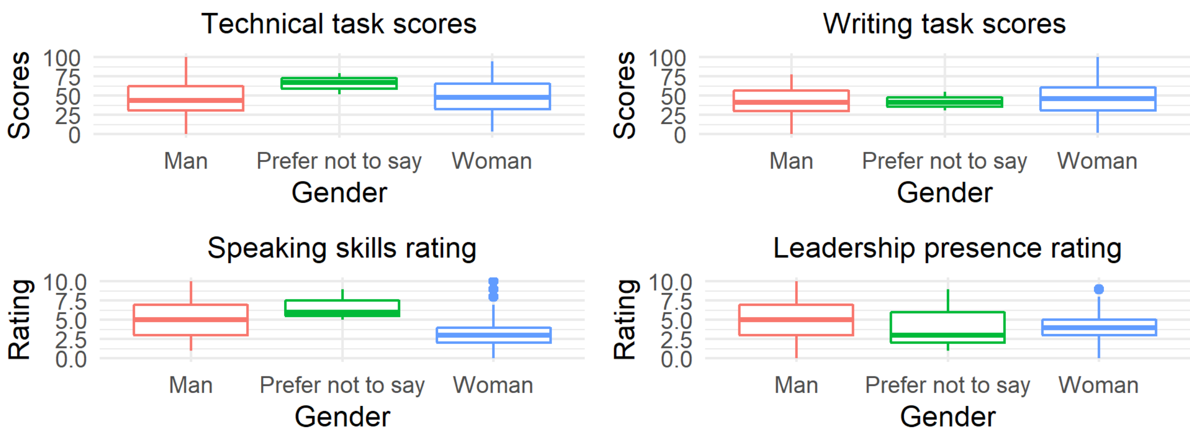


Figure 3: Distribution of AI-autograded Scores/Ratings by Gender in the Phase 2,

The horizontal line in each box represents the median of scores which separates the top 50% scores from the lowest 50%. The horizontal lines at the top and bottom for each box represent the lowest 75% and 25% of the scores respectively. The ends of vertical lines give the full range of scores, with some dots representing extreme values away from the box.

Technical report

Introduction

Recently, several complaints were received related to potential bias in the hiring and remuneration processes. Hence, BSS (Black Saber Software) required our team to take a look and give out a related report for the Board of Directors. The purpose of this study is to investigate if BSS's hiring, promotion, and salary were processing fairly based on talent and value to the company. We were especially interested in exploring potential gender bias in the hiring and remuneration processes. Besides, we also investigated the performance of the new AI service in the hiring pipeline. Data was provided by the data team from BSS including hiring data for their new graduate program and data for their current employees.

In this report, we will specify different explanatory and response variables for. Then, we will fit several generalized linear models and a linear mixed effects model to analyze the results. At the end, we will point out the same limitations of our model. The whole report will be run on R.studio and knit into pdf.

Research questions

- Does the new AI service in the hiring pipeline work and is reliable for future use?
- In particular, is gender a potential bias during the promotion and salary processes?
- If the promotion, and salary process are all fair to all employees and based on their ability and value to BSS?

Models and Results

Hiring Process for applicants

Relationships between whether a candidate was hired or not and the variables during the entire hiring process

We will consider a data set named all_phase combining all necessary data in the hiring process which include four datasets : phase1-new-grad-applicants-2020.csv, phase2-new-grad-applicants-2020.csv, phase3-new-grad-applicants-2020.csv and final-hires-newgrad_2020.csv. Each line of all_phase contains all aggregated information of an applicant in the hiring pipelines. In this analysis for hiring process, the response variable is the Examination of data for hiring process reveals the following key variables in all_phase:

applicant_id = Unique identifier assigned to candidates in Phase 1
team_applied_for = Data or Software
gpa = 0.0 to 4.0
extracurriculars = Indicates the quality of extracurricular involvement with scores of 0,1,2
work_experience = Description of candidates' previous work experience with scores of 0,1,2
technical_skills = Score from 0 to 100 on a timed technical task, AI autograded
writing_skills = Score from 0 to 100 on a timed writing task, AI autograded
speaking_skills = A rating of speaking skills based on pre-recorded video, AI autograded
leadership_presence = A rating of leadership presence based on pre-recorded video, AI autograded
final_result = Indicate at which phase applicant was rejected or hired finally
result = Show the hiring result for each applicant in each hiring pipeline

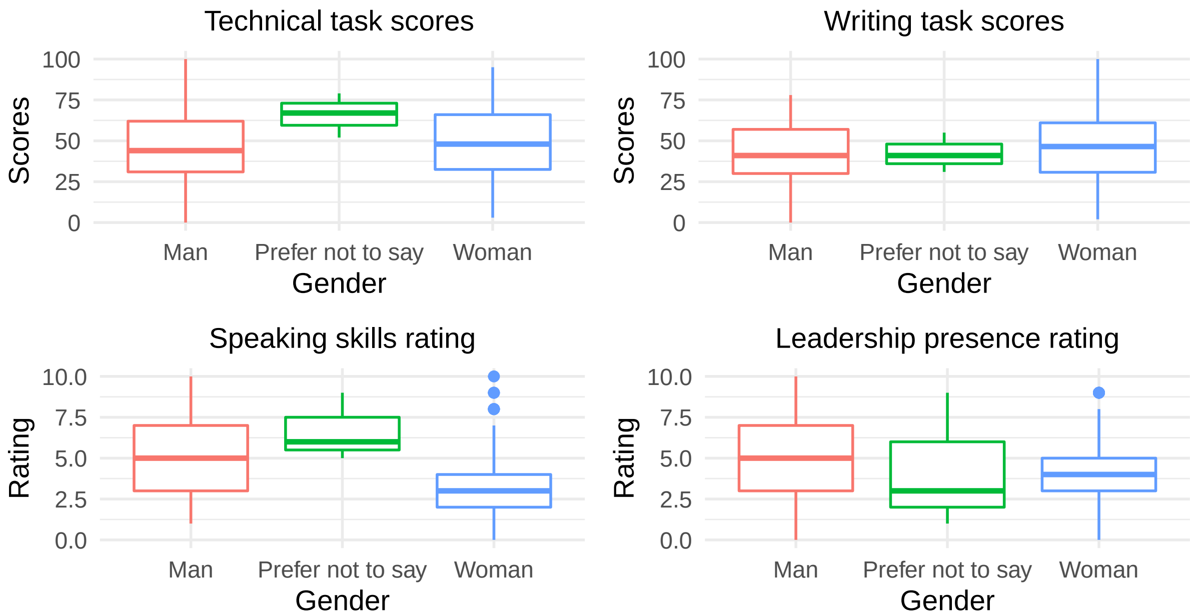


Figure 3: Distribution of AI−autograded Scores/Ratings by Gender in the Phase 2,

The horizontal line in each box represents the median of scores which separates the top 50% scores from the lowest 50%. The horizontal lines at the top and bottom for each box represent the lowest 75% and 25% of the scores respectively. The ends of vertical lines give the full range of scores, with some dots representing extreme values away from the box.

Characteristic	**log(OR)**	**95% CI**	**p-value**
gpa	-2.4	-4.5, -0.77	0.009
cv	-16		>0.9
cover_letter	-16		>0.9
extracurriculars	-1.6	-3.6, -0.22	0.042
work_experience	0.40	-1.0, 2.0	0.6

Figure 1: Overall Key statistical summaries of the Generalized Linear Model related to accepted applicant.

Is the AI used during the hiring process reliable?

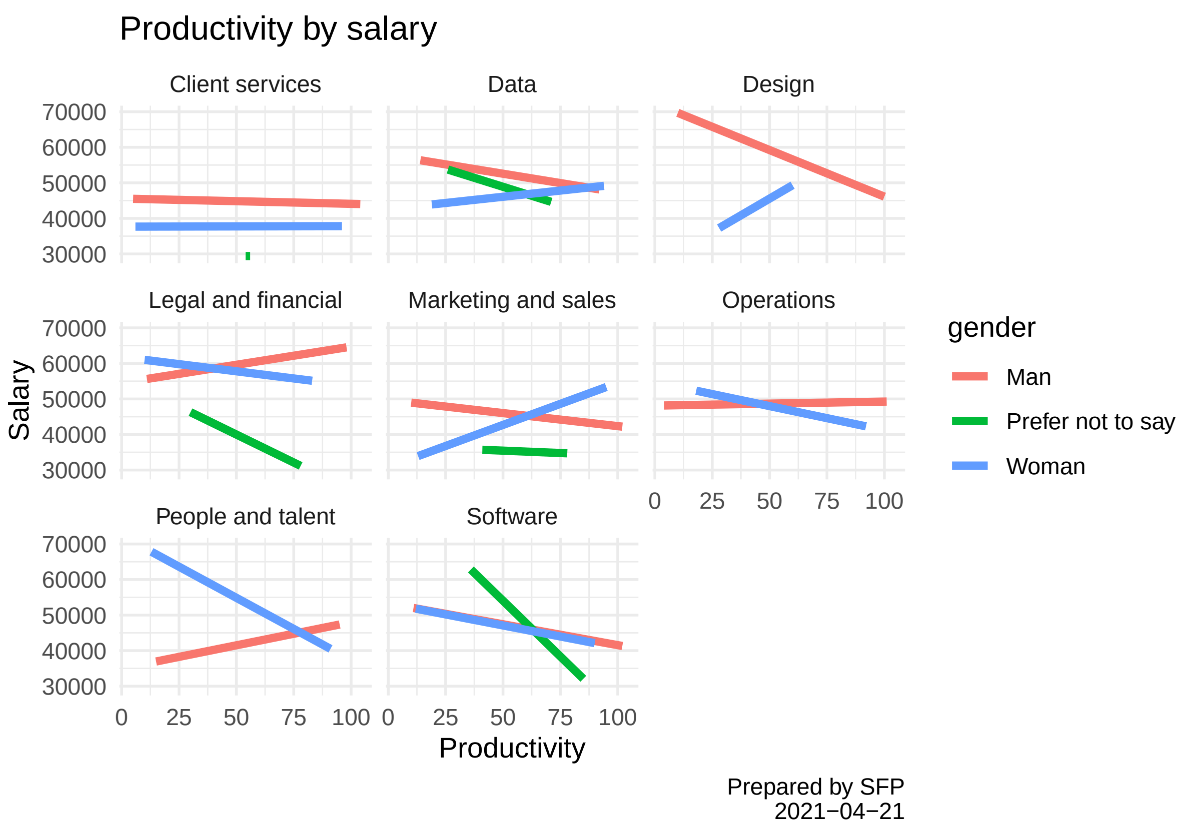
Characteristic	**log(OR)**	**95% CI**	**p-value**
gpa	0.67	-0.74, 2.2	0.4
extracurriculars	0.33	-1.0, 1.7	0.6
work_experience	0.01	-1.4, 1.5	>0.9
leadership_presence	-1.0	-1.5, -0.62	<0.001
speaking_skills	-0.74	-1.1, -0.43	<0.001
technical_skills	-0.09	-0.14, -0.05	<0.001
writing_skills	-0.10	-0.15, -0.05	<0.001

Figure 2: Key statistical summaries of the Generalized Linear Model related to AI scoring in phase 2.

Promotion and Salary Processes for current employees

In the current employee data set, we have selected gender, team, leadership_for_level, productivity, and salary. Based on the label dictionary, obtained along with the data, each variables represent the following:

employee_id = Unique 5-digit employee identifier
gender = Gender of employee: man, woman, prefer not to say
team = one of the 8 teams which each employee works for
leadership_for_level = if the employee is appropriate for level, need improvements or exceeds expectations
productivity = work output with respect to job description rated on a 0 - 100 scale
salary = salary at the given financial quarter



Graph 1: General relationship between productivity and salary for each team separated by gender.

```
## Linear mixed model fit by REML ['lmerMod']
```

```
## Formula: salary ~ gender + team + (1 | employee_id)
##    Data: black_saber_current_employees
##
## REML criterion at convergence: 146211.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.3488 -0.2548 -0.0127  0.2108  7.7525
##
## Random effects:
##  Groups      Name        Variance  Std.Dev.
##  employee_id (Intercept) 270852785 16458
##  Residual                 68844147  8297
## Number of obs: 6906, groups:  employee_id, 607
##
## Fixed effects:
##                         Estimate Std. Error t value
## (Intercept)                39400       1696  23.234
## genderPrefer not to say    -4115       5405  -0.761
## genderWoman                -4723       1447  -3.263
## teamData                    7246       2418   2.997
## teamDesign                 12415       4348   2.855
## teamLegal and financial    10743       3346   3.211
## teamMarketing and sales     3280       2198   1.492
## teamOperations              5932       2397   2.475
## teamPeople and talent       7471       3548   2.106
## teamSoftware                4227       2136   1.979
##
## Correlation of Fixed Effects:
##             (Intr) gnPnts gndrWm teamDt tmDsgn tmLgaf tmMras tmOprt tmPpat
## gndrPrfrnts -0.075
## genderWoman -0.485  0.105
## teamData    -0.549 -0.010  0.027
## teamDesign  -0.352  0.021  0.110  0.212
## tmLglandfnn -0.407 -0.086  0.045  0.274  0.155
## tmMrktngans -0.636 -0.037  0.099  0.417  0.240  0.308
## teamOpertns -0.576  0.025  0.072  0.381  0.219  0.276  0.423
## tmPplandtln -0.353  0.009 -0.025  0.255  0.140  0.183  0.279  0.257
```

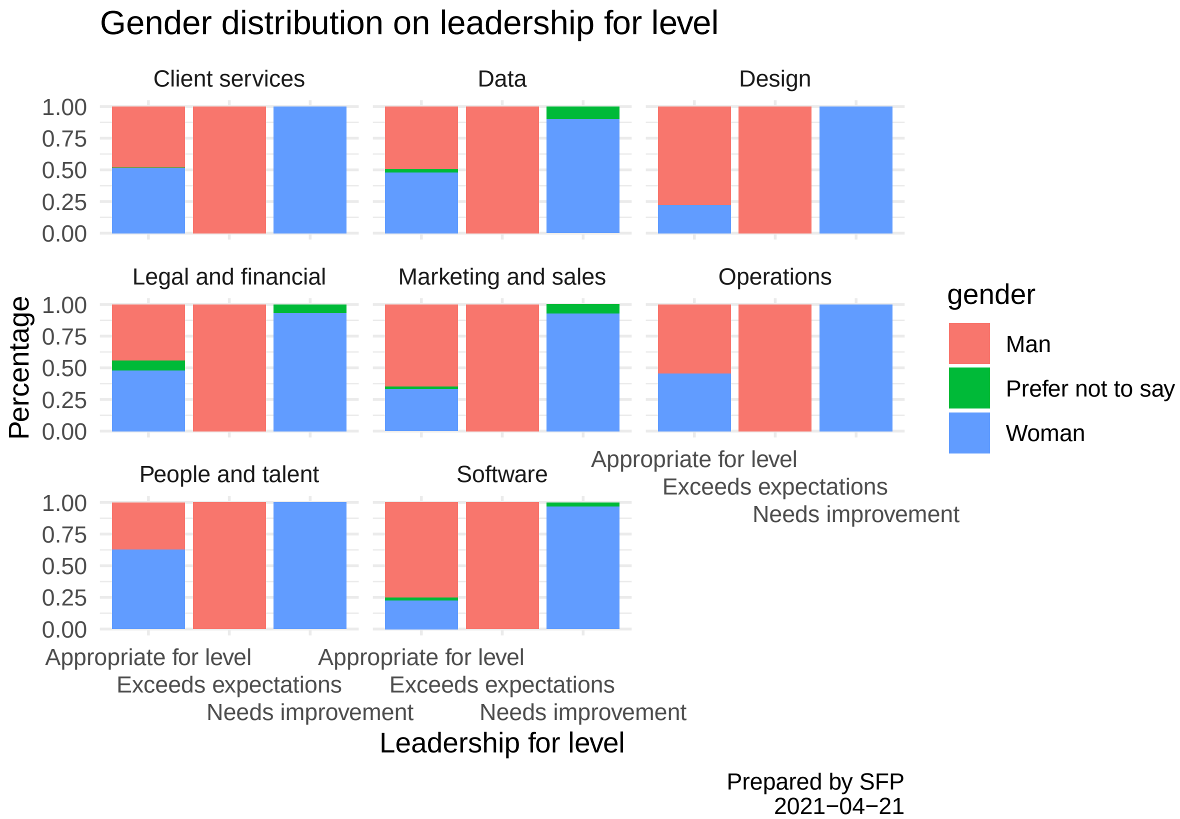
```
## teamSoftwar -0.724 -0.011  0.243  0.433  0.263  0.322  0.494  0.446  0.283

## Linear mixed model fit by REML ['lmerMod']
## Formula:
## salary ~ gender + team + role_seniority + productivity + leadership_for_level +
##     (1 | employee_id:team) + (1 | employee_id)
##    Data: black_saber_current_employees
##
## REML criterion at convergence: 117155.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.8522 -0.2785  0.0033  0.2415  6.3765
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  employee_id:team (Intercept) 8033862  2834.4
##  employee_id      (Intercept) 2066583  1437.6
##  Residual                      953158   976.3
## Number of obs: 6906, groups:  employee_id:team, 607; employee_id, 607
##
## Fixed effects:
##                                             Estimate Std. Error  t value
## (Intercept)                                1.217e+05  3.539e+02  343.983
## genderPrefer not to say                   -1.150e+03  1.033e+03   -1.114
## genderWoman                               -2.239e+03  2.762e+02   -8.105
## teamData                                   2.405e+03  4.612e+02    5.215
## teamDesign                                -1.994e+02  8.312e+02   -0.240
## teamLegal and financial                    2.832e+03  6.368e+02    4.447
## teamMarketing and sales                    7.358e+02  4.191e+02    1.756
## teamOperations                             4.980e+02  4.572e+02    1.089
## teamPeople and talent                     -1.394e+03  6.778e+02   -2.056
## teamSoftware                               2.759e+03  4.069e+02    6.780
## role_seniorityEntry-level                 -9.081e+04  1.491e+02 -609.106
## role_seniorityJunior I                    -8.532e+04  1.428e+02 -597.623
## role_seniorityJunior II                   -8.295e+04  1.378e+02 -601.987
## role_seniorityManager                     -5.029e+04  1.187e+02 -423.561
## role_senioritySenior I                    -7.720e+04  1.340e+02 -576.254
```

```
## role_senioritySenior II                    -7.181e+04  1.302e+02 -551.606
## role_senioritySenior III                   -6.631e+04  1.274e+02 -520.523
## role_seniorityVice president                2.998e+04  1.900e+02  157.811
## productivity                               -8.441e-01  1.199e+00   -0.704
## leadership_for_levelExceeds expectations -4.422e+00  7.070e+01   -0.063
## leadership_for_levelNeeds improvement      2.448e+02  9.354e+01    2.617
```

```
## Likelihood ratio test
##
## Model 1: salary ~ gender + team + (1 | employee_id)
## Model 2: salary ~ gender + team + role_seniority + productivity + leadership_for_level +
##     (1 | employee_id:team) + (1 | employee_id)
##   #Df LogLik Df Chisq Pr(>Chisq)
## 1  12 -73106
## 2  24 -58578 12 29056  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3: Key statistical summaries of the Linear Mixed-Effect Model to check salary related variables.



Graph 2: General relationship between productivity and leadership_for_level for each team separated by gender.

Discussion

In this section you will summarize your findings across all the research questions and discuss the strengths and limitations of your work. It doesn't have to be long, but keep in mind that often people will just skim the intro and the discussion of a document like this, so make sure it is useful as a semi-standalone section (doesn't have to be completely standalone like the executive summary).

First, we explored the effects the various variables during the hiring phases had on the hiring chances. Fitting a model to model the applicant's hiring result (no if not hired, yes if hired) with the applicants' GPA,Curriculum Vitae, cover letter, extracurriculars, and work_experience as effects, it was found that only the GPA and extracurriculars had a significant effect on the being hired.

Second, we explored whether the AI used during the first 2 phases of the hiring process was reliable, and if there were any biases. Fitting a model to model the applicants result on the second phase ("refused" if they did not move on to phase 3, "proceed" if they do move on to phase 3) with the applicants' GPA, extracurriculars, work experience, leadership presence, speaking skills, technical skills, and writing skills as the effects, it was found that GPA, leadership presence, speaking skills, technical skills, and writing skills all have a significant effect on an applicants' success rate.

Finally, we took a look to see if the employees' promotions and salaries are fair and justified based on the their ability and value to Black Saber Software. We first separated the salaries by gender, and graphed each teams' employees' salaries against their productivity. We found that there was indeed a bias for gender.It was observed that the salaries of men based on their productivity is almost always inversely related to the salaries of women based on their productivity. What this means is that as a man's productivity goes up, they earn more, while as a woman's productivity goes up, they earn less, and vice versa. This rule is only broken in the Software team, where men and women all earn roughly the same amount. We will not comment on the emloyees who did not disclose their gender as it is not indicitive of bias in gender.

Limitations

One limitation is that we were only provided data on current employees. The data of previous past employees is missing. Regarding the promotion and salary processes, past employees' information could give us some insights to why they left the company. This could give valuable information such as whether they were treated evenly. If we can get the past employees' data, we can analyze whether there exist biases during the remuneration process more comprehensively.

Another limitation results from the lack of ethnicity/race data for both the hiring and remuneration processes which could also be a potential bias that we need to check. We were told that the race data was not collected by the People and Talent team but they were considering it for EDI initiatives. The race might also be a factor that would affect the hiring result of applicants and also the promotion and wages processes of employees.

Finally, a limitation may be introduced as a result of implicit evaluation criterions for each hiring phase, especially AI-autograded phase and remuneration process. If we know how applicants can proceed to the next round in the hiring process, it would be easier to identify what effects are correlated and fit better models. For the remuneration process, if we can know what kinds of employees can be promoted or have higher salary, we will be able to explore more about potential biases.

Consultant information

Consultant profiles

Lucas Xian. Lucas is a senior consultant with SFP. He specializes in data analysis and visualization. Lucas earned his Honours Bachelor of Science, Specialist in Statistics, from the University of Toronto in 2023.

Dandan Zhang. Dandan is a junior consultant with SFP. She earned her Bachelor of Science, majoring in Mathematics and Statistics, from the University of Toronto in 2021.She specializes in data visualization.

Jong-Hoon Kim John is a senior consultant with SFP. He earned his Bachelor of Science in statistics and economics, from the University of Toronto in 2021. He specialises in data analysis and visualisation.

Zikun Lei Zikun Lei is a senior consultant with SFP. He earned his Bachelor of Science in statistics and economics, from the University of Toronto in 2021. He specialises in data analysis and visualisation.

Code of ethical conduct

SFP is a statistical consulting company. Our mission is to help our clients solve their problems and give them professional and practical advice. We, all staff in SFP adhere to the Code of Ethics Conduct as outlined below.

Professional Integrity and Accountability: We always serve our clients with integrity and respect. No fabricated or altered data that deliberately affects the results will be used. We promise to be fully qualified with accepted work and be honest to clients if any limitation of competence exists. Appropriate and relevant statistical methods are always used to produce reliable results. All results and conclusions should be clearly explained in detail with statistical data and valid analytic approaches.

Responsibilities to clients: We treat all clients' information confidential and responsible. No client's information should be disclosed or discussed in public without permissions of the clients. We will protect any sensitive data or results from unauthorized people and will not take advantage of them to make profit.

Responsibilities to other statisticians: We always safeguard the reputation of statistical practice and try to enhance our professionalism. We support and help strengthen the work of other fellow statisticians through valuable peer view. No criticism is allowed to cast on individuals rather than on problems. Even in debates, we should be respectful of others.