Final Project Report

Umesh A. Jadhav
ujadhav6@gatech.edu

1 FINAL PROJECT REPORT

1.1 Dataset selection

- Dataset selected: - student placement dataset[1].
- Regulated domain: - Education
- Number of observations: - 699
- Number of variables: - 11
- Dependent variables: - placement_status, salary
- Number of variables identified as protected class: - 2.
- Name of variables identified as protected class: - age, gender.

Table 1 – Tabular representation of Protected Class Variable, Protected Class and Law associated.

Variable	Protected Class	Law
Age	Age	Age Discrimination Act, 1975
Gender	Sex	Equality Act

1.2.1 Data Encoding

The cleaned dataset contains combination of categorical variables as well as numerical variables. The categorical variables are converted to numerical values as per general precedence. Categorical variables are gender, degree, stream and college_name, whereas age, salary, gpa, years_of_experience are numerical variables.

1.2.2 Data Exploration

1.2.2.1 Exploring Protected class and associated subgroups.

Table 1.2.1 – Tabular representation of Protected class variable and their respective subgroups.

Protected class Variable	Subgroups
Age	23 to 24, 25 to 26
Gender	Male, Female

Table 1.2.1 represents the protected class variables from the dataset and all the subgroups they contain. For simplicity, the protected class age has numerical value, hence the subgroups are denoted by range of age.

1.2.2.2 Encoding Protected class subgroups.

The numerical mapping of age, gender is tabulated in table 1.2.2.A and table 1.2.2.B respectively.

Table 1.2.2.A – Encoding Protected class Age to numerical values.

Age subgroups	23 to 24	25 to 26
Mapping	0	1

Table 1.2.2.B – Encoding Protected class Gender to numerical values.

Gender subgroups	Female	Male
Mapping	0	1

1.2.2.3 Frequency tables for Subgroups and Dependent variables

For carrying out further analysis, protected class age and gender are chosen. The frequency tables, table 1.2.3.A and table 1.2.3.B correspond to dependent variable placement_status, while table 1.2.3.C and table 1.2.3.D tabulates frequency of dependent variable Salary with subgroups of Age and Gender. Here, placement_status = 0 indicates that the student has failed the placement whereas, placement_status = 1 indicates that the student is placed successfully.

Table 1.2.3.A – Mapping of Age Subgroups with Dependent variable placement_status.

Age subgroups	23 to 24	25 to 26
Placement_status = 0	76	54
Placement_status = 1	328	241

Table 1.2.3.B – Mapping of Gender Subgroups with Dependent variable placement_status.

Gender subgroups	Female	Male
Placement_status = 0	67	63
Placement_status = 0	267	302

Table 1.2.3.C – Mapping of Age Subgroups with Dependent variable Salary.

Age subgroups	23 to 24	25 to 26
Salary <= 60000	97	74
Salary > 60000	307	221

Table 1.2.3.D – Mapping of Gender Subgroups with Dependent variable Salary.

Gender subgroups	Female	Male
Salary <= 60000	94	77
Salary > 60000	240	288

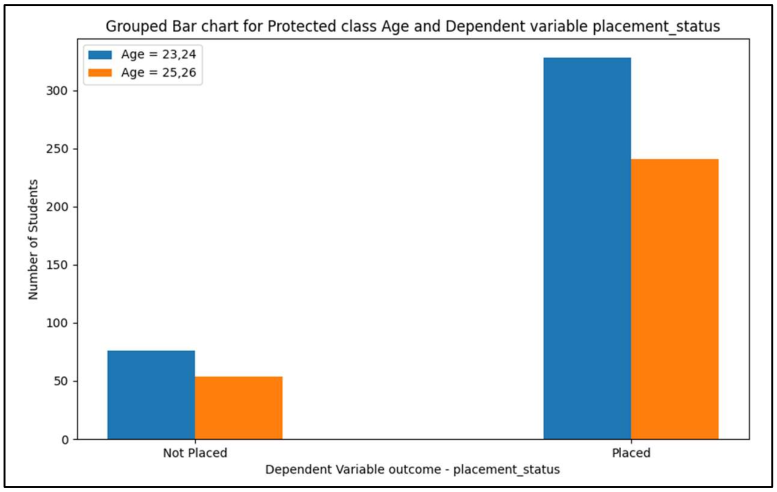

Figure 1.A – Bar chart for Age subgroups and placement_status.

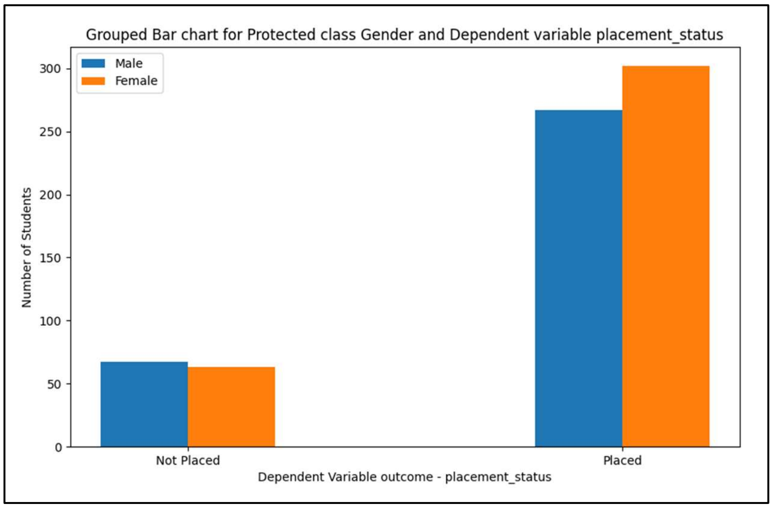

Figure 1.B – Bar chart for Gender subgroups and placement_status.

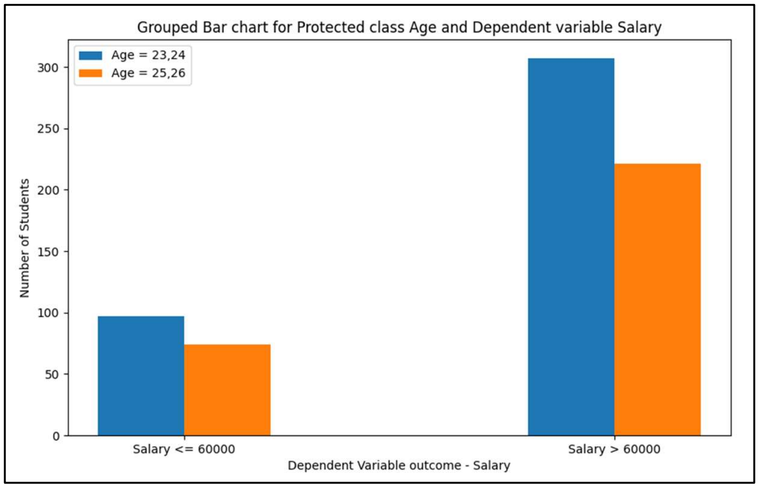

Figure 1.C – Bar chart for Protected class Age subgroups and Dependent variable Salary range.

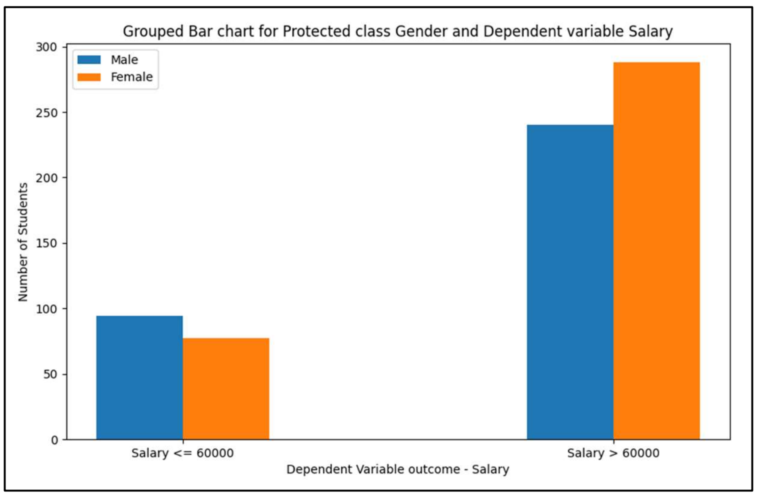

Figure 1.D – Bar chart for Protected class Gender subgroups and Dependent variable Salary range.

In figure 1.A and figure 1.B, the frequency distribution from table 1.2.3.A and table 1.2.3.B for Protected class Age and Gender with Dependent variable placement_status is plotted in the form of grouped bar charts respectively. Similarly, the statistics for distribution of Age groups with dependent variable Salary(range) and Gender subgroups with Salary(range) are illustrated in figure 1.C and figure 1.D.

1.3.1 Privileged Group and Unprivileged group classification

Table 1.3.1.A. and table 1.3.1.B represent the classification of Protected class variables Age and Gender into Privileged group and Unprivileged group. These groups are identified based on the general notion where people having age greater than 25 have high chances of getting placed, hence they have a

greater opportunity to bag higher salary. Similarly, females are considered to be less privileged in employment than males.

Table 1.3.1.A – Tabular representation of Privileged and Unprivileged groups for Age.

Protected Class Variable	Privileged group	Unprivileged group
Age	25 <= Age <= 26	23 <= Age <= 24

Table 1.3.1.B – Tabular representation of Privileged and Unprivileged groups for Gender.

Protected Class Variable	Privileged group	Unprivileged group
Age	Male	Female

1.3.2 Fairness metrics before Bias mitigation

In this section, the fairness metrics, Disparate Impact and Statistical Parity Difference are computed over the Protected class Age and Gender with respect to the two dependent variables, placement_status and Salary range. For simplicity, since Salary is a continuous variable, positive and negative groups are identified to be "Salary greater than 60000", and "Salary lesser than or equal to 60000" respectively.

Table 1.3.2.A – Fairness metrics for Gender, and Age with respect to "placement_status" and Salary range over Original dataset.

	Gender – placement_status	Gender – Salary	Age – placement_status	Age – Salary
Disparate Impact	1.0350	1.0981	1.0060	0.9858
Statistical Parity Difference	0.0279	0.0704	0.0051	-0.0107

In Table 1.3.2.A, the column Gender – Salary and column Age – Salary indicate the fairness metric between Gender and Salary range, and Age and Salary range respectively as per the bifurcation of dependent variables outcome.

1.3.3 Fairness metrics after Preprocessing

In this section, the dataset is transformed by preprocessing it using Disparate Impact Remover algorithm over protected class Age and dependent variable placement_status. The metrics are illustrated in table 1.3.3.A for Disparate Impact and Statistical Parity Difference.

Table 1.3.3.A – Fairness metrics for Gender, and Age with respect to placement_status and Salary range over transformed dataset.

	Gender – placement_status	Gender – Salary	Age – placement_status	Age – Salary
Disparate Impact	1.0350	1.0765	1.0062	0.9412
Statistical Parity Difference	0.0279	0.0545	0.0051	-0.0446

1.4.1 Mitigating Bias

1.4.1.A Original dataset

The original dataset is randomly split into two sections as train split and test split using sklearn's train_test_split() with 20 percentage of entire data being test split size. For training the model, Random Forest Classifier is used with the number of estimators as 29. Finally, the test data is subjected to prediction by applying the random forest classifier. The fairness metrics for the prediction data of original dataset is shown in table 1.4.A. Here, since the data is predicted over dependent variable placement_status and protected class age, the tables shows fairness metrics only for them and not for dependent variable Salary and protected class gender.

Table 1.4.A – Fairness metrics for Gender, and Age with respect to placement_status and Salary range over predictions for original dataset.

	Gender – placement_status	Age – placement_status
Disparate Impact	1.0312	1.0333
Statistical Parity Difference	0.0249	0.0266

1.4.1.B Transformed Dataset

As seen in case of original dataset, the transformed dataset is split into train split and test split with test size as 20 percent of overall population. For comparison, random forest classifier is used for prediction over the test data. The fairness metrics for the prediction data of transformed dataset is shown in table 1.4.B. Also, the table shows fairness metrics for dependent variable placement_status and protected variable age, same as seen for original dataset.

Table 1.4.B – Fairness metrics for Gender, and Age with respect to placement_status and Salary range over predictions for transformed dataset.

	Gender – placement_status	Age – placement_status
Disparate Impact	1.0955	1.0964
Statistical Parity Difference	0.0761	0.0778

Finally, the status of changes and differences between the outcomes of the privileged group and unprivileged group for age and gender with respect to placement_status is illustrated in table 1.4.C. As per the results, the transformed dataset showed better metrics for Disparate Impact as well as Statistical Parity Difference metric by small margin. Although, the bias mitigation have had positive impact on the Gender – placement_status combination, it worsened the score for age-placement_status combination.

Table 1.4.C – Changes in fairness metrics for Gender, and Age with respect to placement_status and Salary range over predictions for original dataset and transformed dataset.

	Gender – placement_status	Age – placement_status
Disparate Impact	Bias increased slightly	Negative change
Statistical Parity Difference	Bias increased slightly	Negative change

1.4.1.C Summary and Conclusion

Table 1.4.D – Disparate Impact metric for Age with respect to placement_status.

	Disparate Impact	Change compared to previous
Original dataset	1.0062	NA
Transformed dataset	1.0062	No change
After training classifier on Original dataset	1.0333	Minimal negative change
After training classifier on Transformed dataset	1.0964	Significant negative change

Table 1.4.E – Statistical Parity Difference metric for Age with respect to placement_status.

	Statistical Parity Difference	Change compared to previous
Original dataset	0.0051	NA
Transformed dataset	0.0051	No change
After training classifier on Original dataset	0.0266	Significant negative change
After training classifier on Transformed dataset	0.0778	Extreme negative change

By inspecting the statistics from table 1.4.D and table 1.4.E, the original dataset and transformed dataset showed exact similar values for Disparate Impact metric as well as Statistical Parity Difference metric. The values for both the metric indicate absence of bias in the dataset originally standing at 1.0062 and 0.0051. The original dataset is transformed based upon the protected class age and dependent variable placement_status.

The original dataset and transformed dataset are then used to train Random Forest classifier to predict placement_status. From the results of the fairness metrics, the prediction has shown a slight increase in bias for both the dataset as compared in table 1.4.D and table 1.4.E for the values for both the metrics respectively. Although, both the metrics fall under the ideal range of 0.8 and 1.1 for Disparate Impact metric and, -0.1 and 0.1 for Statistical Parity

Difference, the slight difference can be concluded as the increased bias introduced by the classifier's prediction.

1.5 Analysis

Step 5: I am a team of one.

1.6 Reference

1. Muhammad Mahad, Job Placement Dataset, https://www.kaggle.com/datasets/mahad049/job-placement-dataset.