Final Project

Ayesha Ilyas
ailyas6@gatech.edu

1 STEP 1

1.1 Which dataset was selected? Please include the link to the dataset

IBM HR Analytics Employee Attrition & Performance
Link : IBM HR Analytics Attrition Dataset

1.2 Which regulated domain does the dataset belong to?

The dataset belongs to Employment/Workforce (HR) domain and falls under the laws such as the Equal Employment Opportunity Act

1.3 How many observations are in the dataset?

1,470 observations

1.4 How many variables are in the dataset?

35 variables

1.5 Which variables were selected as the dependent/outcome variables?

- Attrition : This is a binary variable with value Yes or No
- PerformanceRating : This variable ranges from value 1 to 4. To convert it to binary, I am assigning 1 , 2 and 3 values to 'Low Performance' group and value 4 to 'High Performance' group.

1.6 How many and which variables in the dataset are associated with a legally recognized protected class?

Two variables belong to a legally recognized protected class.

- Age
- Gender

1.7 Which legal precedence/law (as discussed in the lectures) does each protected class fall under?

- Gender : Title VII of the Civil Rights Act of 1964
- Age : Age Discrimination in Employment Act of 1967 (ADEA)

2 STEP 2

2.1 Identify the members/subgroups associated with the protected class variables in the dataset. In your report, in table format document the protected classes and their corresponding subgroups.

See Table 1. Age is a continuous variable with values ranging from 18 - 60. I have grouped into bins as shown the table for better analysis.

Protected Class	Subgroups
Gender	Male, Female
Age	Young ($\leqslant 30$), Middle-aged (31–45), Senior (> 45)

Table 1—Protected Classes and Corresponding Subgroups

2.2 Discretize the subgroups from Step 2.1 into discrete numerical values. Provide these mappings in table format in your report.

See Table 2 for discrete numerical value mappings.

Protected Class	Subgroup	Numerical Value
Gender	Male	0
Gender	Female	1
Age	Young ($\leqslant 30$)	0
Age	Middle-aged (31–45)	1
Age	Senior (> 45)	2

Table 2—Discretized Numerical Mappings for Protected Class Subgroups

2.3 Select two protected classes from the dataset. You will use these protected classes in the rest of your project.

- Age
- Gender

2.4 For each protected class, create frequency tables documenting the number of members in each subgroup associated with each dependent variable selected in Step 1.5.

See Table 3, 4 , 5 and 6. No that the PerfomanceRating outcome variables has been converted to binary outcome, with values 1, 2 and 3 belonging to 'Low Performance' and 4 belonging to 'High Performance'.

Gender	No Attrition	Yes Attrition
Female	501	87
Male	732	150

Table 3—Frequency Table: Gender vs Attrition

Gender	High Performance	Low Performance
Female	94	494
Male	132	750

Table 4—Frequency Table: Gender vs Performance Group

Age Group	No Attrition	Yes Attrition
Young	286	100
Middle-aged	708	103
Senior	239	34

Table 5—Frequency Table: Age Group vs Attrition

2.5 For each protected class, create a bar chart the graphs the frequency of each subgroup as a function of the dependent variables identified in Step 1.5.

See Figure 1, 2, 3 and 4.

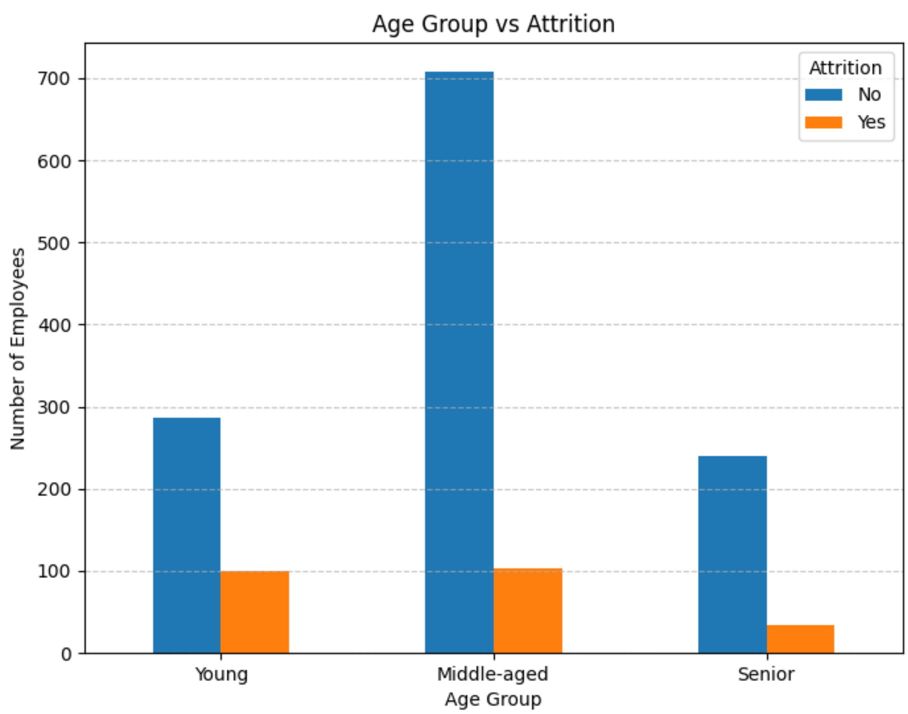

Figure 1—Age Group vs Attrition

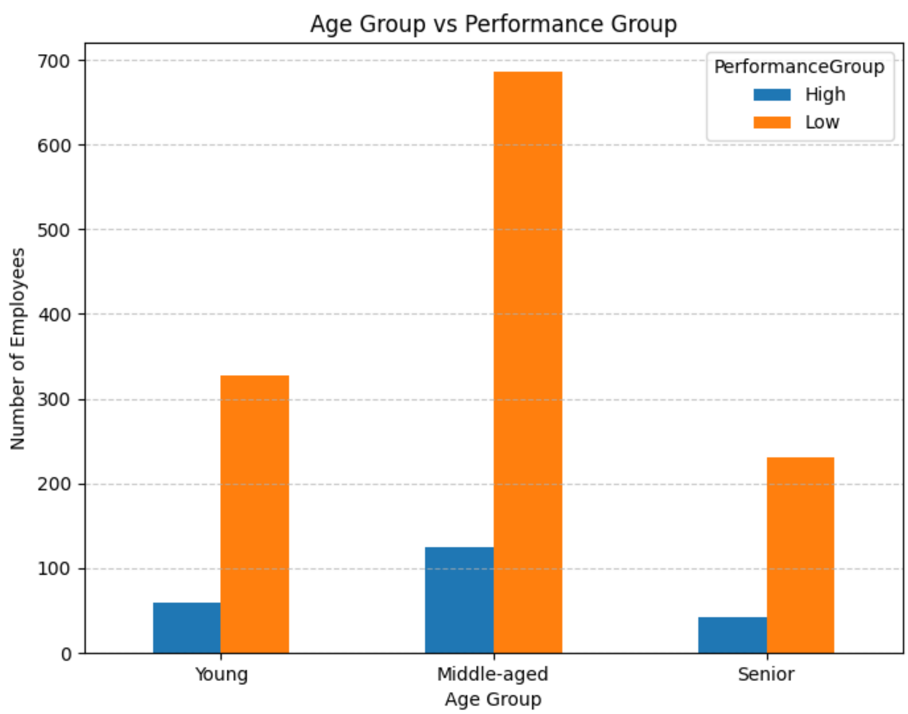

Figure 2—Age Group vs. Performance

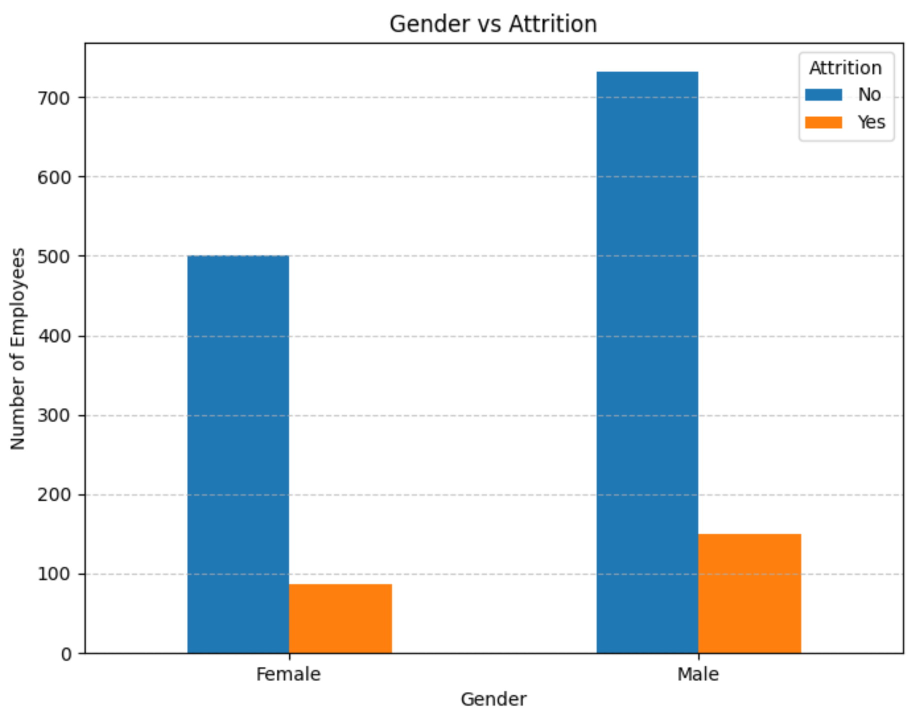

Figure 3—Gender vs. Attrition

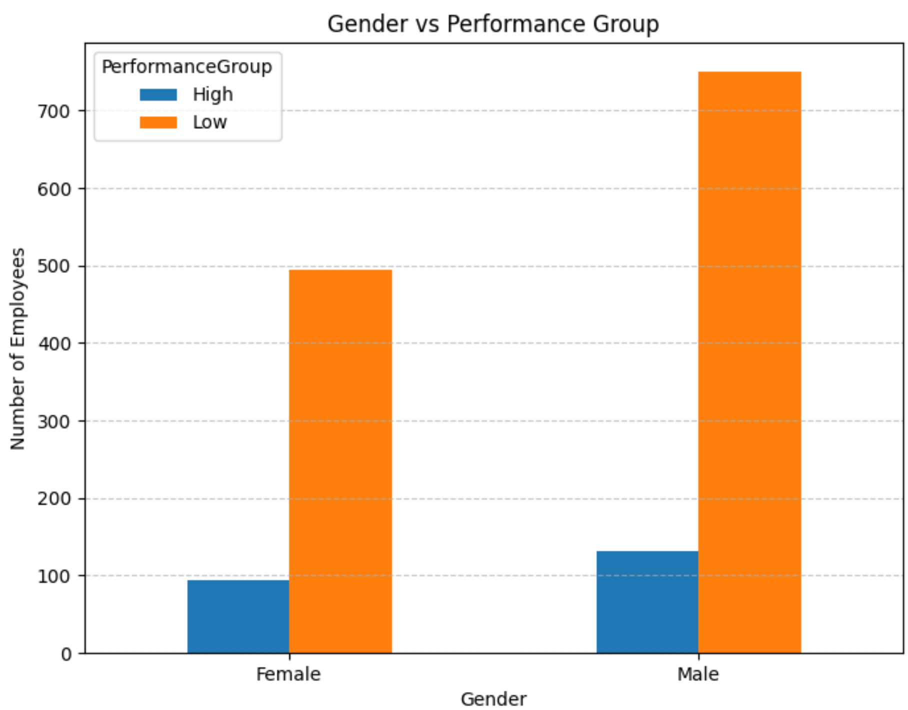

Figure 4—Gender vs. Performance

Age Group	High Performance	Low Performance
Young	59	327
Middle-aged	125	686
Senior	42	231

Table 6—Frequency Table: Age Group vs Performance Group

3 STEP 3

3.1 n your report, identify the privileged and unprivileged groups for each protected class in your dataset.

See Table 7.

Protected Class	Privileged Group	Unprivileged Group
Gender	Male	Female
Age	Young ($\leqslant 30$)	Senior (> 45)

Table 7—Privileged and Unprivileged Groups for Each Protected Class

3.2 Select two fairness metric algorithms. For each protected class in the dataset, compute the two fairness metrics selected for the privileged and unprivileged groups as a function the two dependent/outcome variables identified in Step 1.5. Include your results in a table format.

Fairness metrics selected:

- Statistical Parity Difference
- Disparate Impact

See Table 8.

3.3 Select a pre-processing bias mitigation algorithm to transform the original dataset (e.g. Reweighting, Disparate Impact Remover, etc.) as a function of one of your dependent variables.

Algorithm chosen: Disparate Impact Remover

Protected Class	Outcome Variable	Group Comparison	SPD	DI
Gender	Attrition	Female vs Male	0.022	1.027
Gender	PerformanceGroup	Female vs Male	0.01	1.068
AgeGroup	Attrition	Senior vs Young	0.135	1.182
AgeGroup	PerformanceGroup	Senior vs Young	0.001	1.007

Table 8—Fairness Metric Results: Statistical Parity Difference (SPD) and Disparate Impact (DI)

3.4 Using the two fairness metrics identified in 3.2, compute fairness metrics on the transformed dataset. Include your results in a table format.

See Table 9.

Protected Class	Outcome Variable	Statistical Parity Difference (SPD)	Disparate Impact (DI)
Gender	Attrition	0.0221	1.0266
Gender	PerformanceGroup	0.0000	∞
AgeGroup	Attrition	0.1345	1.1816
AgeGroup	PerformanceGroup	0.0000	∞

Table 9—Fairness Metrics on Transformed Dataset Using Reweighing

4 STEP 4

4.1 Randomly split the original dataset into training and testing datasets

See code attached.

4.2 Train a classifier using the original dataset; select one of your dependent variables as the output label to train your classifier.

See code attached.

4.3 Using the testing dataset, compute the same fairness metrics selected in Step 3.2 for the privileged and unprivileged members of the two protected classes identified in Step 2.3.

See Table 10.

Index	Protected Class	Outcome Variable	Statistical Parity Difference (SPD)	Disparate Impac
0	Gender	Attrition	-0.0030	0.9965
1	AgeGroup	Attrition	0.0935	1.1165

Table 10—Fairness Metrics (SPD and DI) Computed on Testing Dataset

4.4 Randomly split your transformed dataset (from Step 3.3) into training and testing datasets

See code file.

4.5 Train a classifier using the transformed training dataset; select one of your dependent variables as the output label to train your classifier.

See code file.

4.6 Using the testing dataset, compute the same fairness metrics selected in Step 3.2 for the privileged and unprivileged members of the two protected classes identified in Step 2.3.

See Table 11.

Protected Class	Outcome Variable	Statistical Parity Difference (SPD)	Disparate Impact (DI)
Gender	Attrition	-0.0083	0.9909
AgeGroup	Attrition	0.0435	1.0655

Table 11—Fairness Metrics on Reweighed Test Dataset

4.7 For each fairness metric, in table format, indicate if there were any differences in the outcomes for privileged versus unprivileged group. In table format, answer the following question: Was there a positive change, negative change, or no change for each fairness metric after transforming the dataset? Also include an answer to the following: Was there a positive change, negative change, or no change on each fairness metric after training the classifier - with respect to the original testing dataset and the transformed testing dataset?

See Table 12 and Table 13.

Protected Class	SPD Values (Orig / Trans / Trans Test)	Difference Detected?	Change after Tran
Gender	-0.003 / -0.008 / -0.0176	Yes	Negative
AgeGroup	0.093 / 0.043 / 0.0795	Yes	Positive

Table 12—Statistical Parity Difference: Fairness Change Summary

Protected Class	DI Values (Orig / Trans / Trans Test)	Difference Detected?	Change after Transf
Gender	0.996 / 0.990 / 0.973	Yes	Negative
AgeGroup	1.116 / 1.065 / 1.101	Yes	Positive

Table 13—Disparate Impact: Fairness Change Summary

The fairness analysis using Statistical Parity Difference (SPD) and Disparate Impact (DI) reveals notable differences between privileged and unprivileged groups across both protected classes: Gender and AgeGroup.

For Gender, SPD and DI showed small but consistent disparities across all stages of the pipeline. After applying the Reweighing algorithm, fairness metrics slightly deteriorated, indicating a negative change in both SPD and DI. Training a classifier on the transformed data further reduced fairness, suggesting that bias was not entirely mitigated, and may have been amplified during the modeling phase.

For AgeGroup, the transformation step led to a positive improvement in both SPD and DI, indicating more equitable representation between younger and older employees. However, after classifier training, fairness metrics regressed slightly, though not to the level of the original dataset, suggesting a partial retention of fairness gains from preprocessing.

In both cases, while preprocessing helped in reducing initial disparities, the learning algorithm introduced new imbalances. This emphasizes the importance of evaluating fairness not only at the dataset level but also after model deployment, as bias can re-emerge even after mitigation.

5 STEP 5

I am a team of one.