CS 6603: AI, Ethics, and Society  
Final Project Report

Cleo Zhang  
yzhang3761@gatech.edu

# Step 1 – Dataset Overview

* Which dataset did you select*? IBM HR Analytics Employee Attrition & Performance.*
* Which regulated domain does your dataset belong to? *Employment.*
* How many observations are in the dataset? *1470.*
* How many variables are in the dataset? *35.*
* Which variables did you select as your dependent variables? *Attrition and Hourly rate.*
* How many and which variables in the dataset are associated with a legally recognized protected class? Which legal precedence/law (as discussed in the lectures) does each protected class fall under? *See Table 1.*

1. Protected class categories and associated variables.

|  |  |  |
| --- | --- | --- |
| Variables | Protected classes | legal precedence |
| Age | Age | *Age Discrimination in Employment Act of 1967.* |
| Gender | Sex | *Equal Pay Act of 1963; Civil Rights Act of 1964, 1991.* |

# Step 2

## Identification of the Members Associated with the Protected Classes and the Selected Dependent Variables

This section shows how the the members associated with the selected dependent variables (*Table 3*) and the protected classes (*Table 2*) are discretized into numerical values. Those values will be used in the remaining calculations of this report.

1. The relationship between members and membership categories for each protected class.

|  |  |  |
| --- | --- | --- |
| Protected Class | Subgroups | Members |
| Age | 18 - 25 | 0 |
| 26 - 35 | 1 |
| 36 - 45 | 2 |
| 46 - 55 | 3 |
| 55+ | 4 |
| Gender | Female | 0 |
| Male | 1 |

1. The relationship between values and discrete categories/numerical values associated with the dependent variables.

|  |  |  |
| --- | --- | --- |
| Dependent Variables | Subgroups | Members |
| Attrition | No | *0* |
| Yes | *1* |
| Hourly rate | 30 - 50 | *0* |
| 51 - 70 | *1* |
| 71 - 90 | *2* |
| 91+ | *3* |

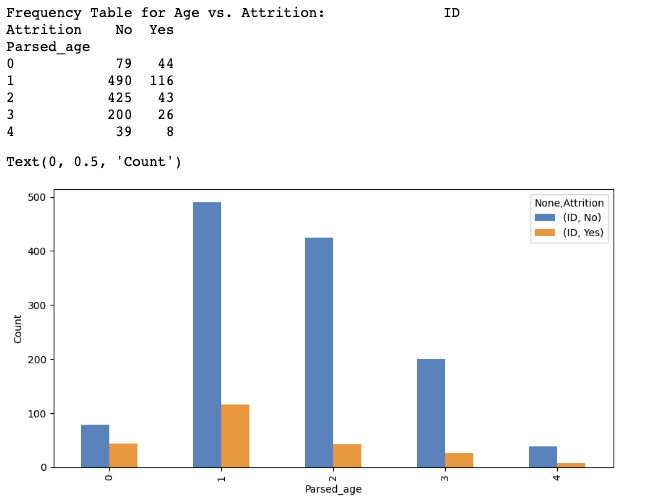
## Frequency Tables and Histograms for Protected Class Variable vs. Dependent Variables

Below shows the frequency table and histograms of each protected class and dependent variable combination.

* Age vs. Attrition

1. Frequency Table for Age vs. Attrition.

|  |  |  |
| --- | --- | --- |
| Age | Attrition | |
| 0 (No) | 1 (Yes) |
| 0 (18 – 25) | 79 | 44 |
| 1 (26 – 35) | 490 | 116 |
| 2 (36 – 45) | 425 | 43 |
| 3 (46 – 55) | 200 | 26 |
| 4 (55+) | 39 | 8 |



1. Histograms for Age vs. Attrition.

* Age vs. Hourly Rate

1. Frequency Table for Age vs. Hourly Rate.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Age | Hourly Rate | | | |
| 0 (30 - 50) | 1 (51 - 70) | 2 (71 - 90) | 3 (91+) |
| 0 (18 – 25) | 36 | 39 | 25 | 23 |
| 1 (26 – 35) | 179 | 164 | 180 | 83 |
| 2 (36 – 45) | 126 | 125 | 146 | 71 |
| 3 (46 – 55) | 60 | 64 | 67 | 35 |
| 4 (55+) | 8 | 16 | 16 | 7 |

Chart, bar chart

Description automatically generated

Figure 2 - Histograms for Age vs. Hourly Rate.

* Gender vs. Attrition

1. Frequency Table for Gender vs. Attrition.

|  |  |  |
| --- | --- | --- |
| Gender | Attrition | |
| 0 (No) | 1 (Yes) |
| 0 (Female) | 501 | 87 |
| 1 (Male) | 732 | 150 |

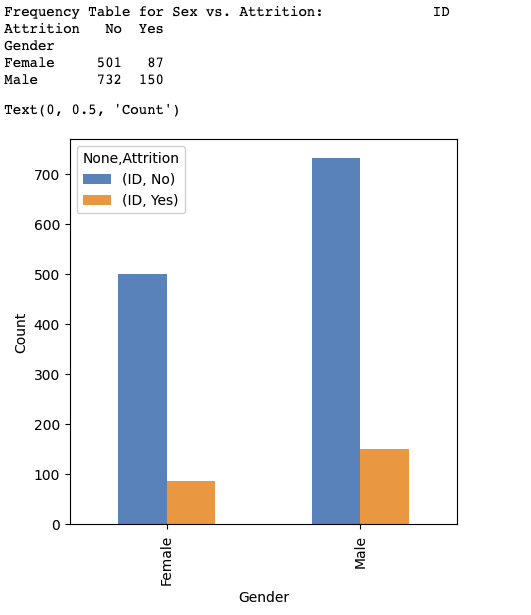


Figure 3 - Histograms for Gender vs. Attrition.

* Gender vs. Hourly Rate

1. Frequency Table for Gender vs. Hourly Rate.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gender | Hourly Rate | | | |
| 0 (30 - 50) | 1 (51 - 70) | 2 (71 - 90) | 3 (91+) |
| 0 (Female) | 163 | 163 | 175 | 92 |
| 1 (Male) | 246 | 250 | 259 | 127 |

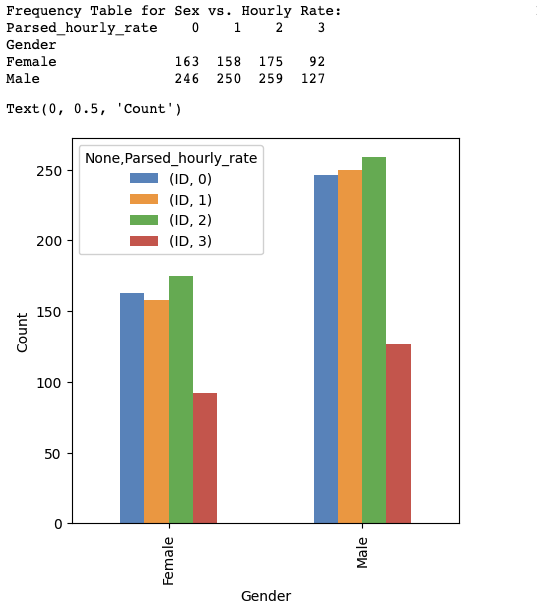


Figure 4 - Histograms for Gender vs. Hourly Rate.

# Step 3

## Privileged/Unprivileged Groups and Fairness Metrics Identification

The privileged and unprivileged groups are identified as in *Table 8*, and the selected fairness metrics are:

* Disparate Impact: an ideal outcome for this metric is between 0.8 and 1.25.
* Statistical Parity Difference: an ideal result for this metric is between -0.1 and 0.1.

*Table 8 -* Frequency Table for Gender vs. Hourly Rate.

|  |  |  |
| --- | --- | --- |
| Protected classes | Privileged | Unprivileged |
| Age | *18 – 25 (0)* | *26+ (1, 2, 3, 4)* |
| Gender | *Male (1)* | *Female (0)* |

## Age vs. Dependent Variables – Original Disparate Impact & Statistical Parity Difference

For *Age vs. Attrition*, this report uses the average of Attrition values for each Age group. Both and are out of the ideal range, so the age group of 18-25 is privileged for Attrition.



For *Age vs. Hourly Rate,* this report uses the average of Hourly Rate values for each Age group. Even though both and are within the ideal range, young people of age 18-25 seem to be at a disadvantage.

## Gender vs. Dependent Variables - Original Disparate Impact & Statistical Parity Difference

For *Gender vs. Attrition*, we use the average of Attrition and Hourly Rate for each gender to calculate the Disparate Impact and Statistical Parity Difference. As shown below, even though both and are within the ideal range – the Male group is a bit more privileged than the Female group for Attrition.

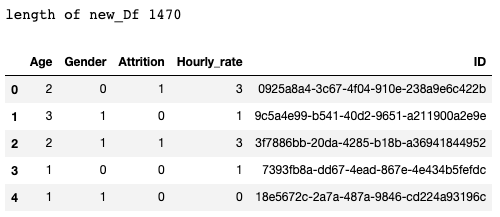


Similarly, and are within the reasonable range as well, however the Female group is a bit more privileged than the Male group as for the Hourly Rate.

## Transformation of the Original Dataset as a Function of Attrition

Basing on the calculation from 3.2 and 3.3, the bias in Attrition by Age is most significant among all. Therefore, I will focus on mitigating this bias by applying a pre-processing bias mitigation algorithm – Reweighting - to adjust the weights of the observations in a way that equalizes the outcomes for privileged and unprivileged age groups, while preserving the overall distribution of the data. Specifically, I add a new column to the data table, called “Weight”, and then we multiply the parsed Attrition value with the corresponding “Weight” value as an outcome.

For the rest of this assignment, I will be using the parsed table (*Figure 5*) from Step 1 in the attached .ipynb file as the original dataset. This parsed dataset scratched out the unrelative columns, keep the same data characteristics of the column we are interested in and works with a better calculation efficiency.



*Figure 5* – Head of the Parsed Table.

Figure 6 shows the algorithm that transform the original dataset.



*Figure 6* – The Reweighing Algorithm and the transformed table (transformed\_df\_3).

## Age –Disparate Impact & Statistical Parity Difference on Transformed Dataset

The Disparate Impact and Statistical Parity Difference for the transformed dataset is shown below:

Due to the presence of error accumulation, both Disparate Impact and Statistical Parity Difference are approximate rather than exact values.

# Step 4 – Option B

This step will take the Attrition as my dependent variable and Age as my protected class (indicated in Table 9).

*Table 9 -* Frequency Table for privileged and unprivileged groups for Step 4.

|  |  |  |
| --- | --- | --- |
| Protected classes | Privileged | Unprivileged |
| Age | *18 – 25 (0)* | *26+ (1, 2, 3, 4)* |

After examining the original dataset, I observed that the privileged age group has a significantly smaller size compared to the unprivileged age group. To address the existing bias, I will develop a new algorithm in this step that utilizes variant oversampling (*Figure 7*).



*Figure 7* – The Oversampling Algorithm and the transformed table (transformed\_df\_4).

The above algorithm aims to increase the size of sampling based on unprivileged group size but not changing the cumulative Attrition values of the Privileged age group. Therefore, compared to the original dataset's Disparate Impact and Statistical Parity Difference, I expect that these two data will decrease accordingly.

*Table 9 -* Two fairness metrics associated with the original and transformed dataset.

|  |  |  |
| --- | --- | --- |
|  | Original Testing | Transformed Testing |
| Disparate Impact | 3.1272 |  |
| Statistical Parity Difference | 0.2616 |  |

While the current Statistical Parity Difference falls within a reasonable range, the Disparate Impact appears to be overly corrected. Table 10 illustrates the differences in the Disparate Impact and Statistical Parity Difference values for the current results vs. the transformed dataset discussed in Sections 3 and the OriginalTesting dataset in Section 4.

*Table 10 –* Effect in each of the fairness metrics after transforming the dataset in section 4.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Original Testing | Transformed Testing in Section 3 | Transformed Testing in Section 4 |
| Disparate Impact | Positive change | Negative change | No Change |
| Statistical Parity Difference | Positive change | Negative change | No Change |

This algorithm causes an overcorrection on Disparate Impact mostly because of the excessively oversampling of the privileged Age group. One possible solution I can think of is to dynamically adjust the sampling size according to the ratio of the privileged/unprivileged group size instead of simply matching up with the unprivileged group size.

# Step 5

I am a team of one.