Final Project

Wesley Tomjack

Step 1

1. **Which dataset did you select?**
   a. Student Performance Data Set
      (https://archive.ics.uci.edu/ml/datasets/Student+Performance)
2. **Which regulated domain does your dataset belong to?**
   a. Education (Education Amendments of 1972 and Civil Rights Act of 1964)
3. **How many observations are in the dataset?**
   a. 1044 observations
4. **How many variables in the dataset?**
   a. 33 variables
5. **Which variables did you select as your dependent variables?**
   a. First Period Grade
   b. Final Period Grade
6. **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?**
   a. Sex (Equal Pay Act of 1963; Civil Rights Act of 1964, 1991)
   b. Age (Age Discrimination in Employment Act of 1967)

Step 2

1) **Protected Class Subset Grouping**

Protected Class	Membership Group (if applicable)	Members
Sex	N/A	Female
		Male

Protected Class	Membership Group (if applicable)	Members
Age	Middle School Aged	15
		16
	High School Aged	17
		18
	University Aged	19
		20
		21
		22

2) **Dependent Variable Discretization**

Dependent Var	Discretization	Values
First Semester Grade	A	(16-20)
	B	(11-15)
	C	(6-10)
	D	(1-5)

Dependent Var	Discretization	Values
Final Grade	A	(16-20)
	B	(11-15)
	C	(6-10)
	D	(1-5)

3) Frequency Count Per Protected Class

Protected Class Members for Sex	Frequency Count
Female	591
Male	453

| Protected Class Members for Age	Frequency Count
Middle School Aged	475
High School Aged	555
University Aged	14

4) Histogram Per Protected Class Subgroups in Each Dependent Variable
   a. Gender breakdown for Q1 Grade and Final Grade
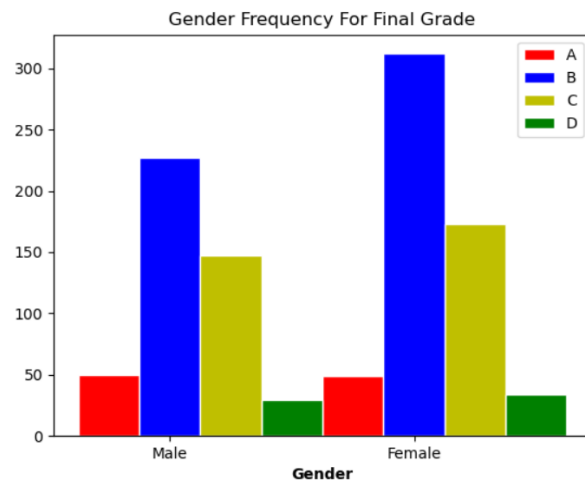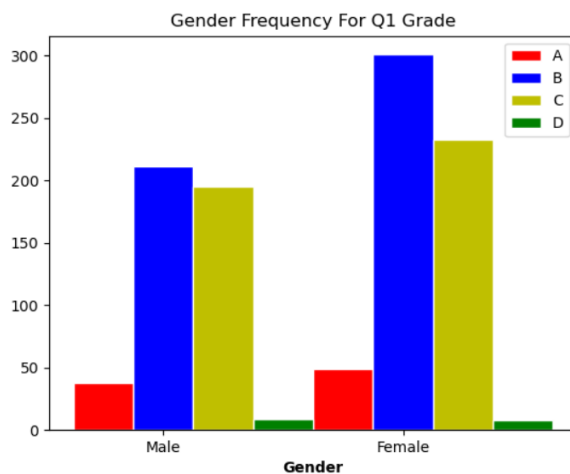
b. Age breakdown for Q1 and Final Grade
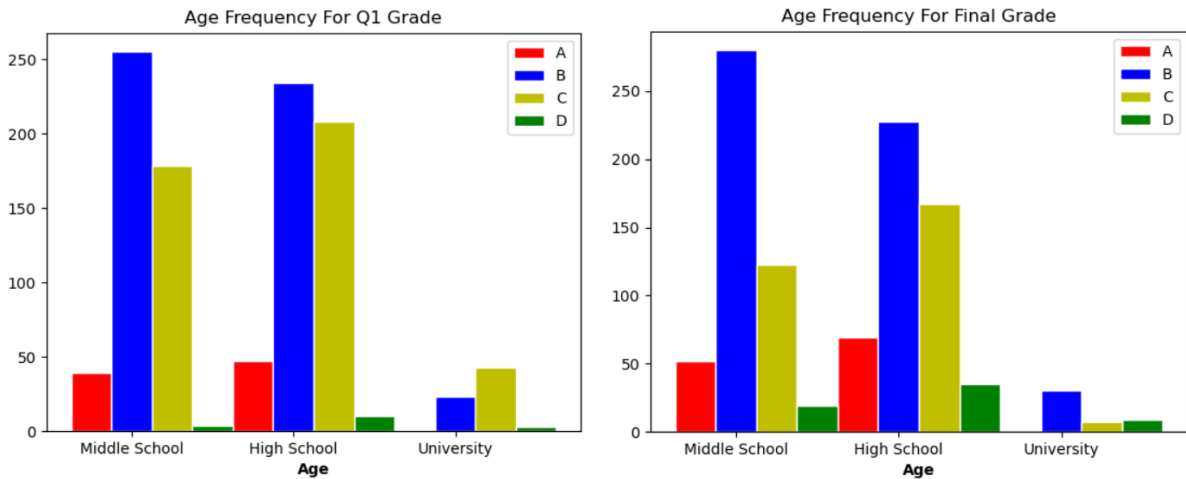
Step 3

1) **Privileged**
   a. Female (Gender)
   b. High School (Age)
2) **Unprivileged**
   a. Male (Gender)
   b. University (Age)
3) **Fairness Metrics**
   a. **Disparate Impact**

   ```
   Initial Disparate Impact Metric for Sex
   .9489279481
   ```

   ```
   Initial Disparate Impact Metric For Age
   .9461386138
   ```

4) **Reweighting**
   a. **I chose to do reweighting on the protected class for sex as there was a binary of labels within that particular feature making it easy to choose an unprivileged/privileged grouping.**

   i.

Privileged Positive Outcome Weight	0.988839
Privileged Negative Outcome Weight	1.04161
Unprivileged Positive Outcome Weight	1.014946
Unprivileged Negative Outcome Weight	0.9504652

5) **Fairness Metrics – Post Reweighting**
   a. **Disparate Impact**

   i. After applying reweighting metric to counts and calculating disparate impact we get the below results for disparate impact on protected class: 'Sex'

   ```
   Final Disparate Impact for Gender
   .996732410
   ```

Step 4

Had difficulty getting ML classifiers to run properly, however given how close to being fair that the original dataset was based off of both metrics, it can be inferred that fairness between both classifiers would generate a similar level of fairness in equality of outcome.

Step 5

I am a team of one

Bibliography (Used to generate personal calculation for Reweighting and Disparate Impact)

Haniyeh Mahmoudian, P. (2020, April 16). Reweighing the Adult Dataset to Make it "Discrimination-Free". Retrieved July 30, 2020, from https://towardsdatascience.com/reweighing-the-adult-dataset-to-make-it-discrimination-free-44668c9379e8