Marcus J. Anderson – CS6603

Step 2

- **Which dataset did you select?**
  Taiwan Credit Data Set

- **How many observations are in the dataset?**
  30,000; There are some rows that have all zeroes in them. I decided to keep these in to see how they impacted the final calculations.

- **How many variables in the dataset?**
  24

- **How many and which variables in the dataset are associated with a legally recognized protected class?**
  2: Sex and Age

- **Of those variables associated with a protected class, what is the associated legal precedence/law it falls under as discussed in the lectures?**
  Sex: (Equal Pay Act of 1963; Civil Rights Act of 1964, 1991)
  Age: (Age Discrimination in Employment Act of 1967) (over 40)

Step 3

- Outcome variable(s): **Creditworthiness derived from Credit Limit, Outstanding Debt, and Default Payment Status**
- Formula used to score members creditworthiness from 0 to 100 is:
  - **IF Default Payment = 0: (1 – ((BILL_AMT_1 – PAY_AMT_1)/LIMIT_BAL)) * 100**
  - **IF Default Payment = 1: ((1 – ((BILL_AMT_1 – PAY_AMT_1)/LIMIT_BAL)) * 100) – 10**
- Protected Class Attribute: **Age**
- Privileged group: **Young (age < 40)**
  - Number of Members in Training Set: **10,428**
  - Number of Members in Testing Set: **10,428**
- Unprivileged group: **Older (age >= 40)**
  - Number of Members in Training Set: **4,572**
  - Number of Members in Testing Set: **4,572**

Step 4

Threshold Value: 50
Profit: 96,099

Explanation:

I chose the threshold value of **50** because it resulted in the maximum profit from my formula. The formula I created loops through threshold values 50 to 100 (since creditworthiness >= 50 is considered a good credit risk). For each rating value, the formula takes each creditworthiness score calculated in the dataset and compares them to each iterated threshold value. In addition to the associated applicant's loan status (approved/declined), which is based on if the person defaulted on their next month's payment, a profit amount is then added or subtracted from the overall profit bucket linked to each iterated threshold value.
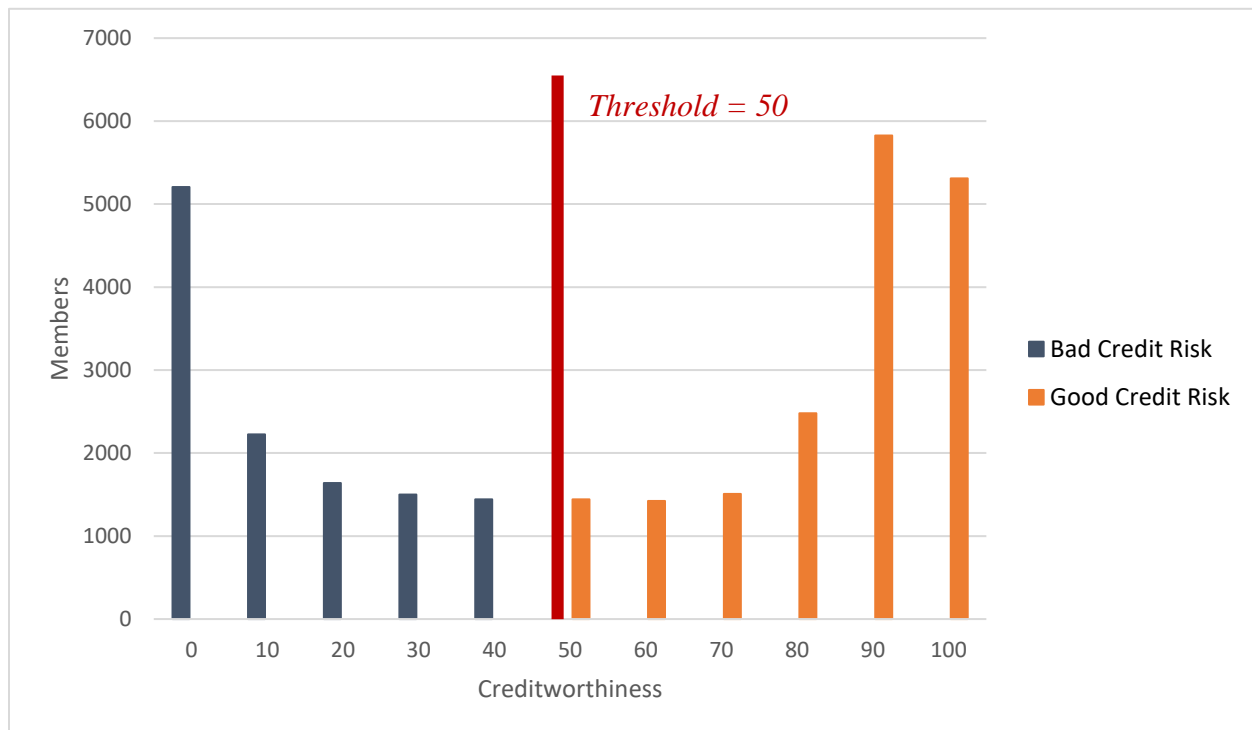
The profit rating scenarios are as follows:

Loan == 'Approved' && Creditworthiness >= Threshold => +10 Profit
Loan == 'Approved' && Creditworthiness < Threshold => -5 Profit
Loan == 'Declined' && Creditworthiness >= Threshold => -3 Profit
Loan == 'Declined' && Creditworthiness < Threshold => +0 Profit

Privileged vs. Unprivileged:

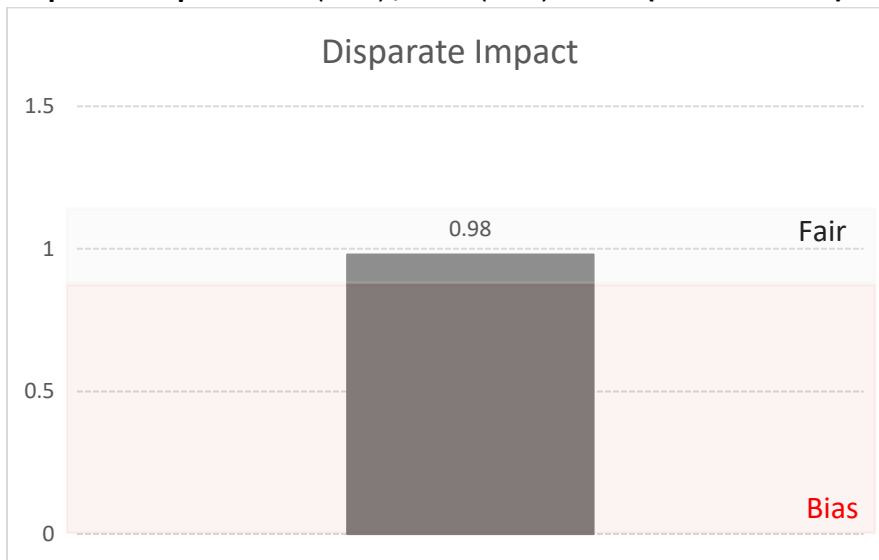| Group	Favorable (Approved)	Unfavorable (Declined)
Privileged	6275	4153
Unprivileged	2679	1893

Step 5

Protected Attribute: Age

Unprivileged Rate Outcome: 2,679 (favorable data) / 4,572 (total test data) = **0.59 rate**
Privileged Rate Outcome: 6,275 (favorable data) / 10,428 (total test data) = **0.60 rate**
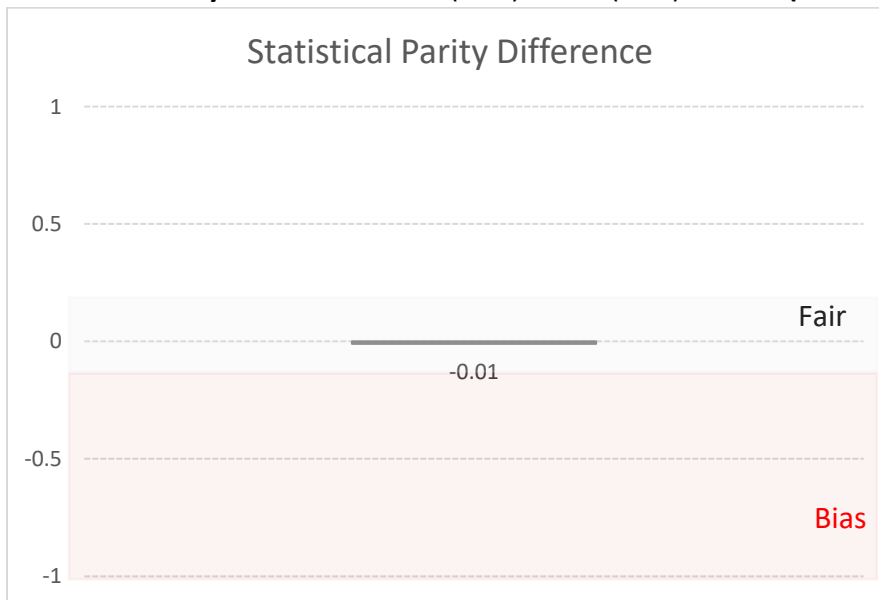
Disparate Impact: 0.59 (rate) / 0.60 (rate) = **0.98 (rate outcome)**

Privileged Group: **Age < 40**

Unprivileged Group: **Age >= 40**

Statistical Parity Difference: 0.59 (rate) - 0.60 (rate) = **-0.01 (rate outcome)**

Privileged Group: **Age < 40**

Unprivileged Group: **Age >= 40**

Do any of the differences indicate bias either for or against the unprivileged or privileged group?

The differences in both fairness metrics I chose show very little bias for the privileged group versus the unprivileged group. The disparate impact, which calculates the ratio between the favorable outcome rate percentages of each group, came in at **0.98**, which is only a **0.02** difference from an even bias outcome between the two groups. The same could be said about the statistical parity difference metric, which calculates the difference between each group's favorable rate percentages, came in at -**0.01.** This further supports my earlier statement of the small difference in bias the privileged group has over the unprivileged group.

Fairness Metric	Outcome Rate	Result
Disparate Impact	0.98	Bias towards the **Privileged Group**
Statistical Parity Difference	-0.01	Bias towards the **Privileged Group**

Step 6

Selected Fairness Metric: Disparate Impact

If you found in Step 4 that there were no (or minimal) differences already, is there a different set of threshold values that provides more profits while still maintaining that minimal difference?

Using the threshold value of **50** in Step 4 yielded an outcome rate of **0.98**, which gives an extremely close difference between the bias of the privileged and unprivileged groups. I tweaked the threshold values of both groups to get the outcome rate to **1**. I tested out a couple different threshold rates and ended up achieving this goal, however it did **not** result in a higher profit rating than the original threshold value of 50.

What is your new creditworthiness formula (if defined differently)?

I used the same creditworthiness formula as defined in Step 4.

What are your threshold values? What is the profit based on your threshold values?
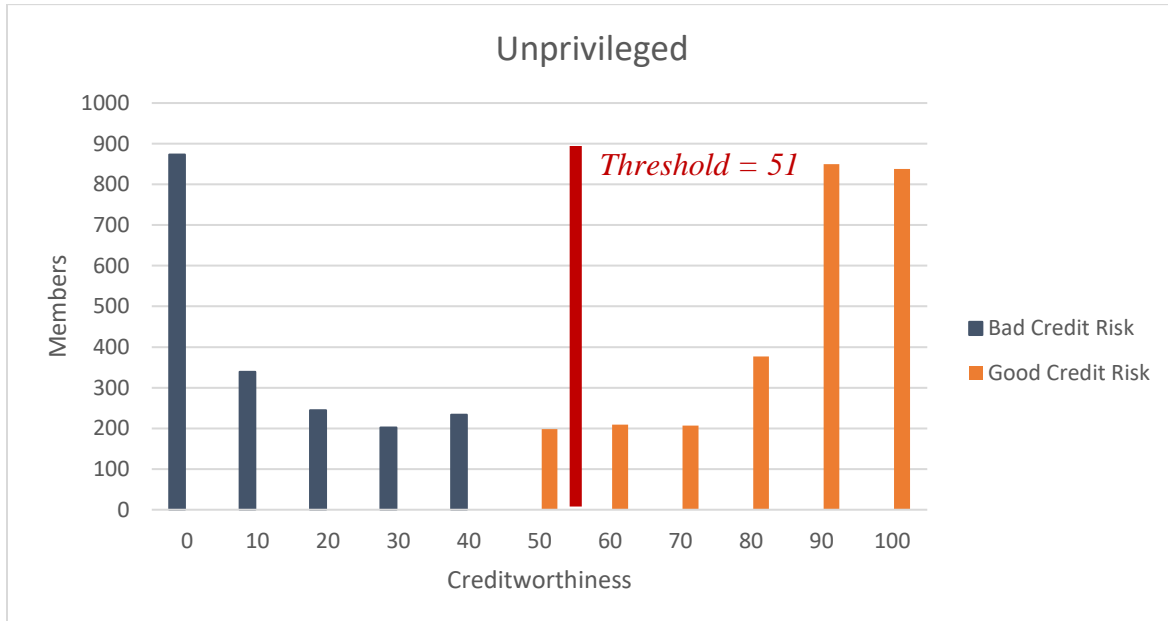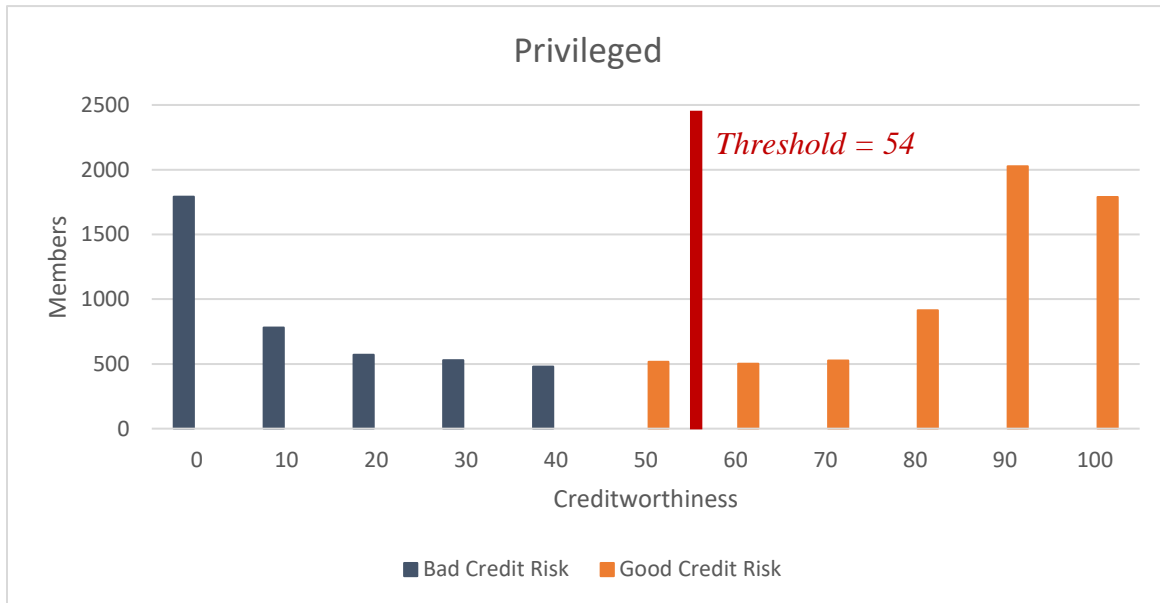
Privileged Group:

- **Threshold:** 54
- **Profit:** 31,985

Unprivileged Group:

- **Threshold:** 51
- **Profit:** 12,825

Privileged vs. Unprivileged:

Group	Favorable (Approved)	Unfavorable (Declined)
Privileged	6061	4367
Unprivileged	2659	1913

Step 7

For each of the fairness metrics selected in Step 5, discuss if there were any differences in the outcomes for the privileged versus unprivileged group?

The main differences in outcomes for both privileged and unprivileged groups was that both rates needed to be lowered to 58% in order to set an even bias for each metric. This was achieved by slightly increasing the threshold values for both groups.

Discuss if the mitigation step in Step 6 was effective and for whom?

The mitigation step was effective in terms of closing the already small gap in bias in favor of the privileged group. While a favorable bias of one group over another is now eradicated, in order to attain this, the thresholds had to increase for both groups, causing a higher declined applicant rate as well.

Did any group receive a positive advantage?

Since the threshold values of each group needed to be increased in order to further migrate bias, the amount of declined applicants increased in both the privileged and unprivileged groups. Since the bias gap was closed however, I would say that gave a positive advantage to the unprivileged group in some fashion.

Was any group disadvantaged by the mitigation step?

I'd say both groups took an unfavorable hit when running them though the mitigation step, as the amount of declined applicants increased, while the accepted applicants decreased. Also, since the bias gap was closed, the privileged group received more of a disadvantage because they no longer receive a favorable bias.

What issues would arise if you used this method to mitigate bias?

Besides the increased applicant decline rate, as discussed in previous questions, the other issues that would arise from this step is overall profitability. Since the threshold values were increased in both groups, this had a negative impact on the maximum profit attained from the original threshold value of **50**. This makes sense because the less applicants that get approved, the less money is being made from loan payments.