**BONUS Questions**

**How do you justify valuing one metric over the other as constituting “fairness”?**

The metric that gives an equal representation over the ‘privileged’ outcome, in this case ‘recidivistic’, should be chosen. In other words, the metric should ensure that all individuals that are predicted to be recidivistic, get an equal opportunity at being labeled recidivistic. However, demographic parity is not as efficient at enforcing this equality. Demographic parity at most strives to ensure that each group has the same proportion of members labeled as recidivistic. This does not ensure that, in each group, the same proportion of members were given equal opportunity at being labeled recidivistic. Single threshold raises a lot of ambiguity in its methodology as only one threshold value is subjected to all groups. Maximum accuracy will find any means necessary to maximize the accuracy at the cost of unequal representation of groups, hence this is also not a fair metric. We can conclude that choosing the best metric is dependent on the role and situation. In our case, equal opportunity shows the best fairness.

**What assumptions are made in the way we have presented the assignment? Are certain answers presupposed by the way we have phrased the questions?**

Assumption 1: That it is sufficient to divide the model output into 5 groups, including ‘Other’ that encompasses all miscellaneous groups. The miscellaneous groups may be a collection of different race types, some which may have a greater disparity than the African-American group.

Assumption 2: The prediction-label pairs given are exhaustive, and values generated by metrics such as TPR, PPV, Predicted Positives, False Positives/Negatives, True Positives/Negatives serve as a good approximation for each of the post-processing methods.

Assumption 3: We have also assumed that considering at most 2 roles or perspectives, NGO and Corporate, and 2 secondary optimizations, accuracy and profit, is enough to analyze the issue. There can be other perspectives and secondary optimizations.

**In what ways do these simplifications not accurately reflect the real world?**

**How do uncertainty and risk tolerance factor into your decision?**

In each post-processing method, we apply a risk tolerance factor to line up the probabilities instead forcing them to match exactly. It ensures that our metrics work on data with a lot of variance. In case variance is high, the tolerance levels may have significantly larger values, and vice versa. Essentially, we would like to conduct our analysis over a neighborhood of data points which span over this tolerance level. It is also true that stretching or increasing this tolerance level will introduce more uncertainty and inaccuracy in our analysis.

**To what extent should base rates of criminality / recidivism among different groups be factored into your decision?**

Base rates of any feature serve as prior knowledge, and hence will introduce some bias. In small amounts, they can serve as a good benchmark, but if base rates are dominant then it will make skew the data towards the sensitive attribute.

**The tools we provide can split the predictions into different protected categories, such as by age or gender. What disparities arise in these groups? How do these disparities compare to those shown when the predictions are split by race?**

Age as a protected category will introduce new disparities such as the legal age limit that allow convictions, harshness of sentence determination and duration of term assigned. Gender will introduce disparities on male/female ratio of recidivism.