Part 1: Fairness concerns

1. Considering the stakeholders of this case study, answer the following questions:
   a) Who gets access to what resources/information in this case study?
      Machine learning engineers have access to a dataset containing medical and personal information of patients and whether they have a form of heart disease.
      Nurses and doctors have access to ER patients' summary of their symptoms, medical history, and any relevant personal details.
   b) Who decides who gets access to these resources?
      Patients from whom those data were collected.
   c) How do they decide who gets what?
      By giving informed consent according to laws/regulations.
2. Based on your answers to question 1 and the dataset that you have, identify the privileged group(s) and the favored outcome.
   Privileged group (based on gender): Male; Favored outcome: higher heart disease possibility.
   Privileged group (based on age): 47-62; Favored outcome: higher heart disease possibility.
3. Hypothesize and elaborate on some of the foreseeable fairness concerns for this case study before you do some further investigation in steps 4- 6.
   As we can see from statistical summary of the datasets, the proportion of male patients is 79%, which may result in unfairness in the model where male patients have higher possibility to be categorized into risk groups and get immediate attention.
   Meanwhile, as we can find in the dataset that the patients aged from 47-62 constitute 57% while age in the dataset ranges from 28-77. This unevenness in age distribution might result in that patients in a certain age range could get higher possibility for immediate attention, although, for example, younger patients might also need same level immediate attention. (I will choose age as protected attribute for the code in part 2)

Part 2: Fairness metrics

4. Difference in mean between unprivileged and privileged groups = -0.133330; Difference in TPR between unprivileged and privileged groups = -0.100000; Difference in average of absolute difference in FPR and TPR between unprivileged and privileged groups = 0.118097
5. Difference in mean means that statistical parity difference between two groups is -0.133330 which falls out of the range [-0.1,0.1]. Difference between two groups exists.
   Difference in TPR means that equal opportunity difference between two groups is -0.1 which falls in the range [-0.1,0.1]. Two groups have relatively equal opportunities.
   Difference in average of odds (disparate treatment) falls out of the range [-0.1,0.1]. Disparate treatment does exist.
6. Equal opportunity.

Part 3: Pre-processing for fairness

7. I'm using reweighing in the code.
8. After reweighing, statistical parity difference becomes zero for training datasets between two groups. Classifier mean difference after reweighing = -0.104651; Classifier TPR difference after reweighing = 0; Classifier average_abs_odds_difference after reweighing = 0. Reweighing can definitely eliminate discrimination to a certain extent. However, it will also cause

information loss. Overall, it can promote fairness in datasets.

Comparing reweighing and learning fair representation: LFR allows hyper parameter tuning and reweighing doesn’t. Reweighing allows multiple protected attributes but LFR doesn’t.