

It's hard to answer this question without making any assumptions. Therefore I'm assuming that my client is from a subgroup that has a high rate for being labeled as a criminal. I would then select a group which in contrast has a very low chance of being labeled as criminal. Then I would collect more data hoping that the model will not be fair based on the equalized odd metric. My end goal is to show that the model is making more false predictions for the people which the model has labeled as criminals.