It can indeed be true that from a cluster or subgroup, the model is predicting more criminals based on historical data and this is leading to mitigation in criminal rates. However, this is a significant tradeoff of fairness. It significantly increases the chances of an innocent person from this group being labeled as a criminal.

In order to analyze the situation and make a decision. I can choose 2 sub-groups from the data. One group where the model predicts a significant amount of criminals and another where the model does predict very few amount of criminals. Then I can compare the accuracy of the model among these groups using the equalized odd metric.