**Which methods seems to be worse? Which methods seems to be better? Explain.**

**Slice by: Age** and **Sex**

**Single Threshold:** This method optimizes a single threshold for all datapoints based on the specified cost ratio. Which means the threshold value for each slice is the exact same. In theory, this method seems like it would be considered a “better” method since each slice is using the same threshold values when computing fairness metrics. However, this can be skewed depending on the amount of datapoints that exist within a particular slice. The highest slice for this method (17-49 Males) has 262 data points within it, which halts a much different result than the lowest slice (49-80 Females). So, I would say this method would be considered **worse**.

**Demographic parity:** This method optimizes a threshold per slice based on the specified cost ratio, which ensures the different slices achieve demographic parity. This means that there’s a similar percentage of positive classifications in each slice. Because each group need to have equal positive rates, this can skew the metrics of certain slices and cause a higher percentage of false negatives. We see this in my graph where the male slices have significantly higher false negative percentages than the female slices. So, I would say this method would also be considered **worse**.

**Equal opportunity:** This method optimizes a threshold per slice based on the specified cost ratio, ensuring the different slices achieve equal opportunity. This means there’s a similar percentage of positive predictions in each slice. This method evens the playing field for each slice, and based on the metrics of this method, the false positive/negative percentages are lower than other methods. So, I would consider this method to be **better**.

**Equal accuracy:** This method optimizes a threshold per slice based on the specified cost ratio, ensuring the different slices achieve equal accuracy. This means there’s a similar percentage of correct predictions in each slice. While I’ve noticed this method impacted the false positive percentage in the male slices, maybe because there’s more datapoints, I think overall this is another method that provides accurate fairness metric results as compared to other methods. So, I would consider this method to be **better**, as well.