As I showed in the Jupyter notebook, if we choose the mean and 2std as the threshold the accuracy will slightly increase by 0.002% if we generate a new feature. All the other circumstances (replace the original feature or choose only mean as threshold) will only decrease the accuracy. My understanding to this result is, when we look at the distribution of the most important feature we can see that most of the observations are within 2std intervals. The groupby operation proved the intuition. Therefore, when we include this as a new feature, we added a useful dimension and help the model training. However, the binary threshold is too rigid and not enough to represent all the information the original feature could deliver. If we replace the original feature with the binary threshold, we lose some useful dimensions and hurt the training. Therefore, consider about the limitation of binary threshold my future recommendation is to split the feature into bins.