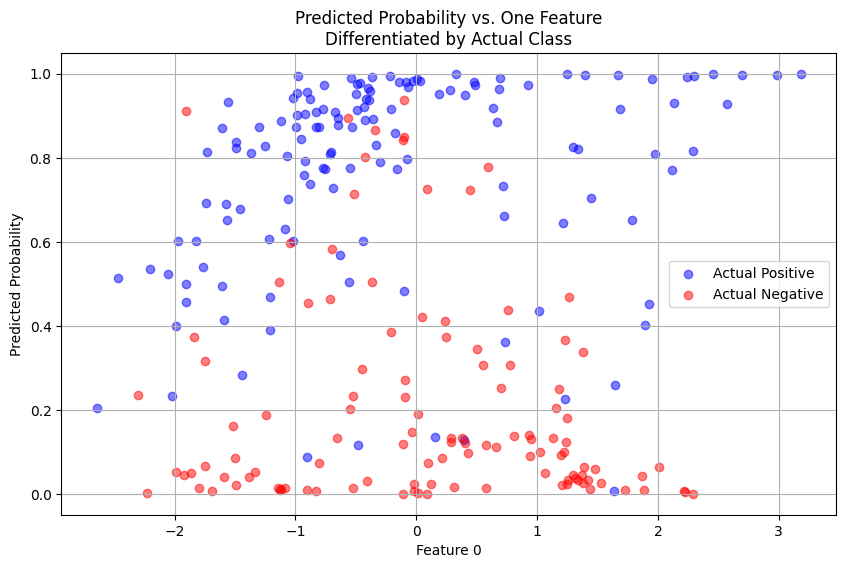
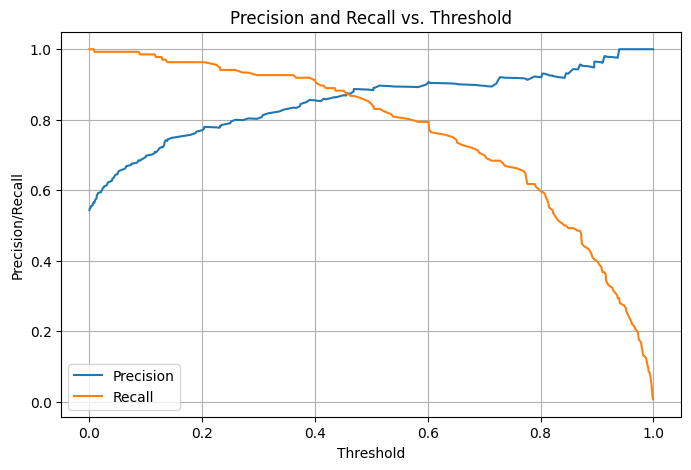
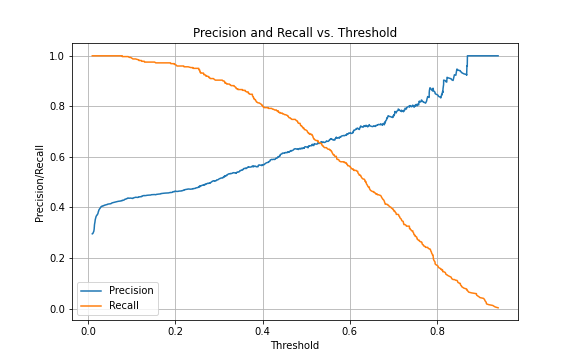
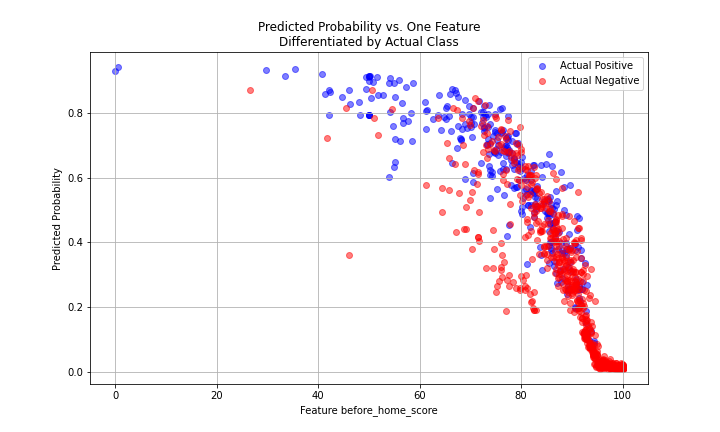
# Threshold





1. Balance between recall and precision
2. Looking for higher precision, in the cost of recall
   1. Powerful model with room to balance.
   2. The Actual Positive/Negative ratio changes wrt Probability



1. Poor model with initial low precision/recall, high precision -> extremely low recall
2. High probability samples are all low before\_home\_score
   1. High precision, not comes from model, comes from imbalance data(more positive) at low before\_home\_score
   2. Most of data lays in the range of high before\_home\_score.
   3. At high before\_home\_score, model perform uniformly bad through different probability region.
      1. Model is less likely to predict high probability, it only predict high probability when it is confident.
3. The good performance, high precision value of low before\_home\_score,
   1. does not comes from model performing good in this region compares to high before\_home\_score range,
   2. comes from the imbalance of positive/negative(because in low before\_home\_score, there are more positive than negative)

conclusion:

1. Model perform slightly better than guess all positive
2. Better performance of different score increment comes from imbalance data distribution(majority is positive)