This assignment involved building a logistic regression model to predict the probability of lead conversion for X Education. The dataset included 36 variables and one target of which only 5 were selected to build model. Some of them had high percentage missing values, some dropped because high correlation, some for VIF.

I Converted Bools into integers, created dummies, chose variables based on RFE. Then split data into train-test, rescaled , built model, checked VIFs, Adjusted the cutoff threshold to achieve a recall higher than 80% requested, for this I had to drop my cut off to 0.23 and this damaged my precision of course.

Top Contributing Variable is Total Time Spent on Website - the most significant positive contributor to lead conversion.

This model provides actionable insights to optimize lead conversion efforts, balancing aggressive strategies during peak times and minimizing unnecessary efforts when targets are met.