Summary Report for Lead Scoring Case_Study

Summary:

- Imported the Required libraries , Read the data.
- Saw the value counts of different variables and removed the variables which are highly skewed, which are with high null values, which are of less use for model building.
- Modified the value counts of many categorical columns for better aesthetics and better analytics.
- Created dummy variables for the categorical variables and removed the original categorical variables.
- Now applied scaling to some numerical columns and in order to bring every cell between -1 to 1 with mean zero I used StandardScalar().
- Then I split the whole data set into Training and test sets with 70:30 proportion.
- Using RFE, Statsmodels and VIF dropped many dummy variables with high P_values and high VIFs.

- Once everything is done, I am left with 12 columns.
- Created a training Dataframe with Converted, Converted_Prob (the predicted values from the model), final Predicted value which I got by taking the threshold x = 0.3. The threshold value is a trade-off between Accuracy, sensitivity and specificity.
- Created a confusion matrix using converted actual and final predicted value. The confusion matrix helped us to know how accurate our training model is, tells us how precise our model is.
- Once training is done, I have tested the model with test set and got the accuracy almost similar.
- Finally, the CEO expected the Success percentage to be 80 but we developed a model which has the precision of around 83, which is

fantastic. That is for every 100 calls, 83 are converted taking up coaching or training at X Education.