Summary Document

Business Problem: An Education company wants to convert all the leads acquired to maximize the revenue. Lead is generated when the customer fills any form online for enquiry. Company wants to target such customers efficiently.

Data Cleansing:

EDA Analysis:

This step involved data cleansing wherein the Null data and outliers were handled. Also, graphs were plotted to understand the variable trends. Columns with high value of Nulls (>40%) were dropped as they will not contribute to the analysis and imputation of Nulls were performed for the remaining variables. Nulls were imputed based on the 'Mode', 'Mean' and the business understanding of the variables. Similarly, the highly skewed columns were also dropped as they were not diverse to predict the output.

Data Preparation:

This involves below steps:

1. Creating Dummies for Categorical variables
2. Scaling (Standard Scaler)
3. Train Test Split

Model Building:

In this step the we selected the variables with automated approach using RFE (15 variables) and then using p-value and VIF value eliminated the columns manually. Optimal cut-off probability(0.3) was selected. So, the final model consists of 9 columns and the sensitivity of the model came out to be 79.07%. Converted_Prob column gives the probability of the customer converting from leads.

Model Validation:

Model was validated on the test data and the sensitivity of the test model came out to be 81.5% which was close to the one obtained in the train data model.

Learnings:

1. How to solve a classification problem.

2. EDA (Null imputation and Outliers handling)

3. Model building using Logistic regression

4. Importance of Sensitivity