SUMMARY REPORT-LEAD SCORING CASE STUDY

An education company named X Education sells online courses to industry professionals. The company gets lot of leads, but the lead conversion is very poor. So as a data analyst our main objective is to help the company's CEO by building a business model which will predict the potential leads to convert into paying customer.

To achieve the desired goal, we have to follow the steps given:

- ✓ Data understanding: The first step of data modelling, reading the data information, parameters and shape and understanding it. And prepare the data for model building.
- ✓ Missing Value: Treating the missing value by replacing the 'select' with 'NAN'. Checking the percentage of missing values present in data.
- ✓ Data Cleaning: After getting the information on the missing values, we should remove the variables which are not required and whichever has high missing values. Then dropping the unique value, as unique value will not affect our analysis.
- ✓ Exploratory Data Analysis: One of the important steps which tells us about the data distribution, univariate/bivariate analysis, Correlation and outlier. It was noticed that lots of variable in the categorical variable are irrelevant. Many variables are having 95% constant value. We have dropped them also.
- ✓ Visualization: Visualization of categorical and numerical variables. Understanding them and going to next step. Finding the Pair plot and correlation between the variables.
- ✓ Outlier Analysis and Treatment: Checking for outlier in the dataset, so we have notice that outliers are present in the dataset and treating the outliers for further analysis.
- ✓ Dummy Variable: Assigning dummy variable to the dataset and then dropping the repeated variables.
- ✓ Train-Test Split: The split was done at 70% and 30% for train and test data respectively for model building.
- ✓ Model Building: First we have executed RFE with top variables. Later other variables were dropped depending on the VIF(VIF<5) and P-Value(P-Value<0.05).
- ✓ Model Evaluation: Here we have made a confusion matrix, with the help of this we have got the Accuracy, Sensitivity and Specificity of 80%, this means our model if good.
- ✓ Prediction: This was done on the test data where we got the Accuracy, Sensitivity and Specificity of 80%.

After successful analysis we were able to get the Lead score from the predicted value on Train and Test Dataset. These are those leads where we can expect the conversion rate to be the best. And high chance of conversion is possible. Which will help the X-Education to convert maximum lead to paying customers.

Conclusion:

From our Analysis, we conclude that the following factors are very important for X-Education for Lead conversion:

- ➢ Total Time Spent on Website
- ➢ Lead Origin
- ➢ Lead Source
- ➢ Last Activity

- What is your current Occupation
- Last Notable activity

And we have also generated the lead score for all the leads we have in the dataset. So, our focus should be more on those leads where the lead score is high. Where the conversion will be higher.