ANALYSIS SUMMARY REPORT FOR LEAD SCORE CASE STUDY

The analysis performed is for 'x education' and to find ways to get more industry professionals to join their courses. The basic data provided gave us lot of information about how the potential customers visit the site, the time they spend there, how they reachedthe site and the conversion rate.

The steps performed are listed as following:

Sr No	Name	What we did	Result we got
1	Business understanding	Spent some time to analyse business problem and data dictionary to get to know about data we have got	Business is not getting enough leads and target set up by CEO to increase it from 30 % to 80% by analysis of data and creating models
2	Data Cleaning	checked the duplicated data, removed the redundant variables/features from the data frame, Dropped the high percentages of null values in the dataset, Outlier treatment	We have got clean data without any null values, and outliers, and it is ready for next steps
3	EDA	Univariate analysis, Bivariate analysis on target variable	Got insights data by which more leads can be generated (E.g.: Time spent on website, target audience etc.) and needs attention
4	Data preparation	Created dummy variables for categorical columns, converted Yes/No to binary mapping	Got more insights after plotting heatmaps and correlation with each other and
5	Scaling	Used standard scalar to scale the data for continuous variables	Come to know about current conversion rate and chances of increasing
6	Model Building	The split was done at 70% and 30% for train and test data respectively with 15 feature selection	It gave top 15 relevant variables and got proper model **4** with p value > 0.05 and VIF under 5
7	Model Evaluation	A confusion matrix was made. Threshold determined using ROC curve using Sensitivity Specificity and Precision Recall	Using specificity sensitivity got Accuracy and Sensitivity for both train and test model around 80 %

Confusion Matrix results

Confusion Matrix for both train and test data:

```
**************************************************

Confusion Matrix
[[3235  767]
 [ 486 1980]]

**************************************************

True Negative                   :  3235
True Positive                   :  1980
False Negative                  :  486
False Positve                   :  767
Model Accuracy                  :  0.8063
Model Sensitivity               :  0.8029
Model Specificity               :  0.8083
Model Precision                 :  0.7208
Model Recall                    :  0.8029
Model True Positive Rate (TPR)  :  0.8029
Model False Positive Rate (FPR) :  0.1917

**************************************************
```

Confusion Matrix results for train dataset

```
**************************************************

Confusion Matrix
[[1348  329]
 [ 222  873]]

**************************************************

True Negative                   :  1348
True Positive                   :  873
False Negative                  :  222
False Positve                   :  329
Model Accuracy                  :  0.8012
Model Sensitivity               :  0.7973
Model Specificity               :  0.8038
Model Precision                 :  0.7263
Model Recall                    :  0.7973
Model True Positive Rate (TPR)  :  0.7973
Model False Positive Rate (FPR) :  0.1962

**************************************************
```

Confusion Matrix results for test dataset

- Using a cut-off value of 0.345, the model achieved a **sensitivity** of **80.09% in the train set** and **79.73% in the test set**.
- Sensitivity in this case indicates how many leads the model identify correctly out of all potential leads which are converting
- The CEO of X Education had set a target **sensitivity of around 80%**.
- The model also achieved an **accuracy of 80.63%**, which is in line with the study's objectives.

Recommendations:

1. More budget/spend can be done on Welingak Website in terms of advertising, etc.
2. Incentives/discounts for providing reference that convert to lead, encourage to provide more references.
3. Working professionals to be aggressively targeted as they have high conversion rate and will have better financial situation to pay higher fees too.