UPGRAD ASSIGNMENT

LEAD SCORING CASE STUDY

BY- SHARMISHTHA SANYAL

PROBLEM STATEMENT

X education is an organization that provides online courses to industrial professionals. The company has its own strategies to attract the customers.

The company now wants to select the most promising leads and convert them to potential customers.

The issue stands upfront as on a regular basis out of 100 potential leads, only 30% get converted to potential customers.

The company requires a promising model that will assure the turn over of potential leads.

OVERVIEW OF ANALYSIS

Importing of data in Python Notebook.

Checking the data file.

Data cleaning for getting more clarity.

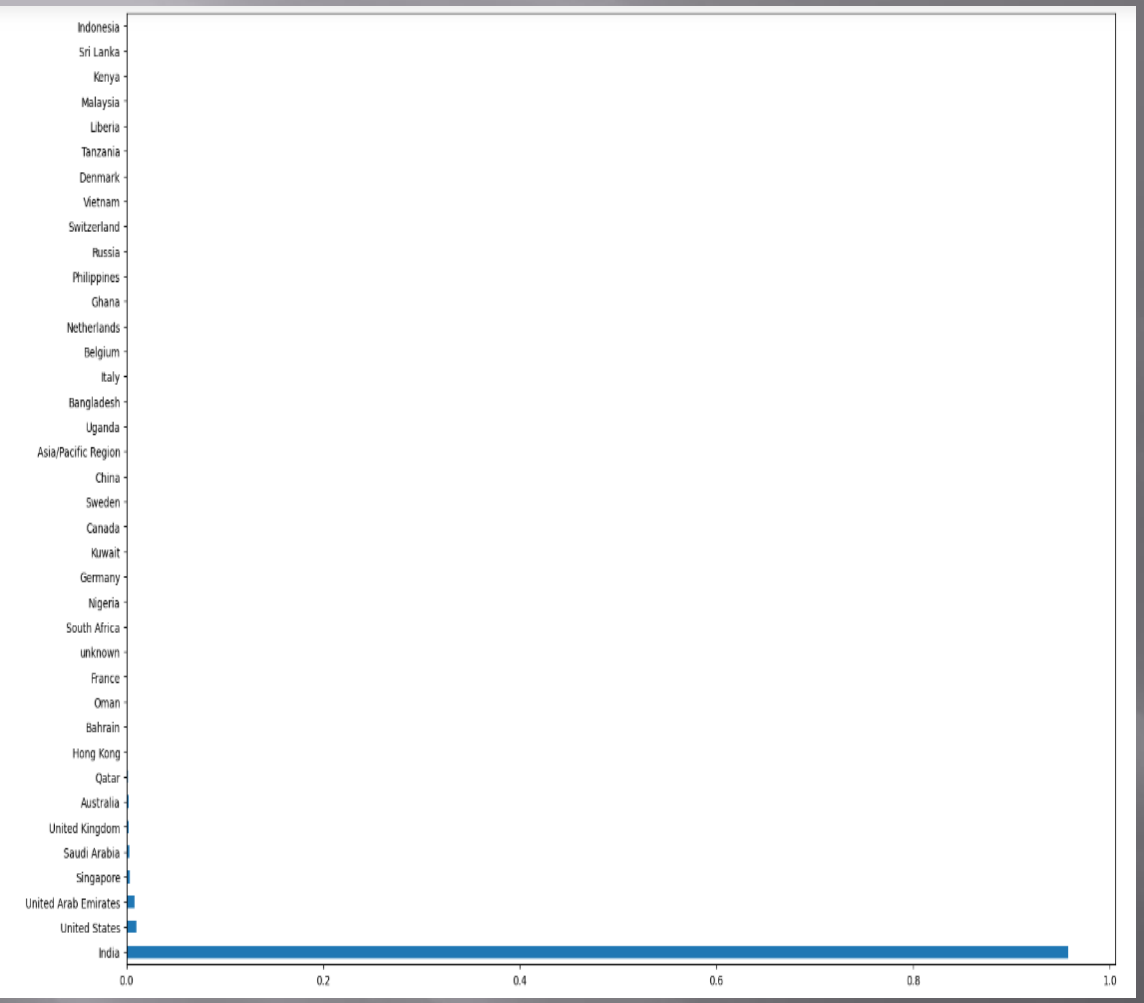
Exploratory Data Analysis

Building of logistic regression model

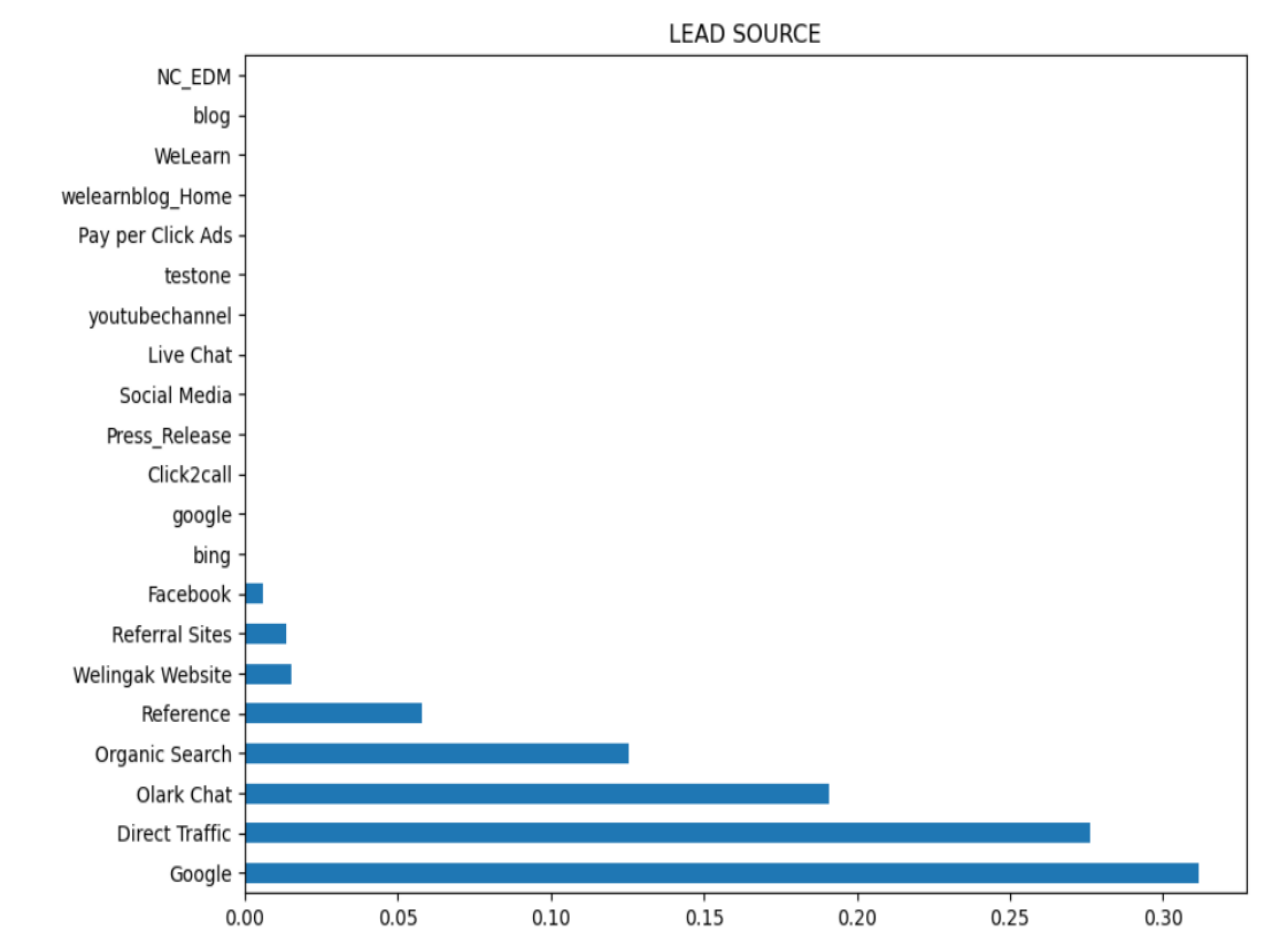
Testing the model on train set

Evaluation of model and understanding its different matrices.

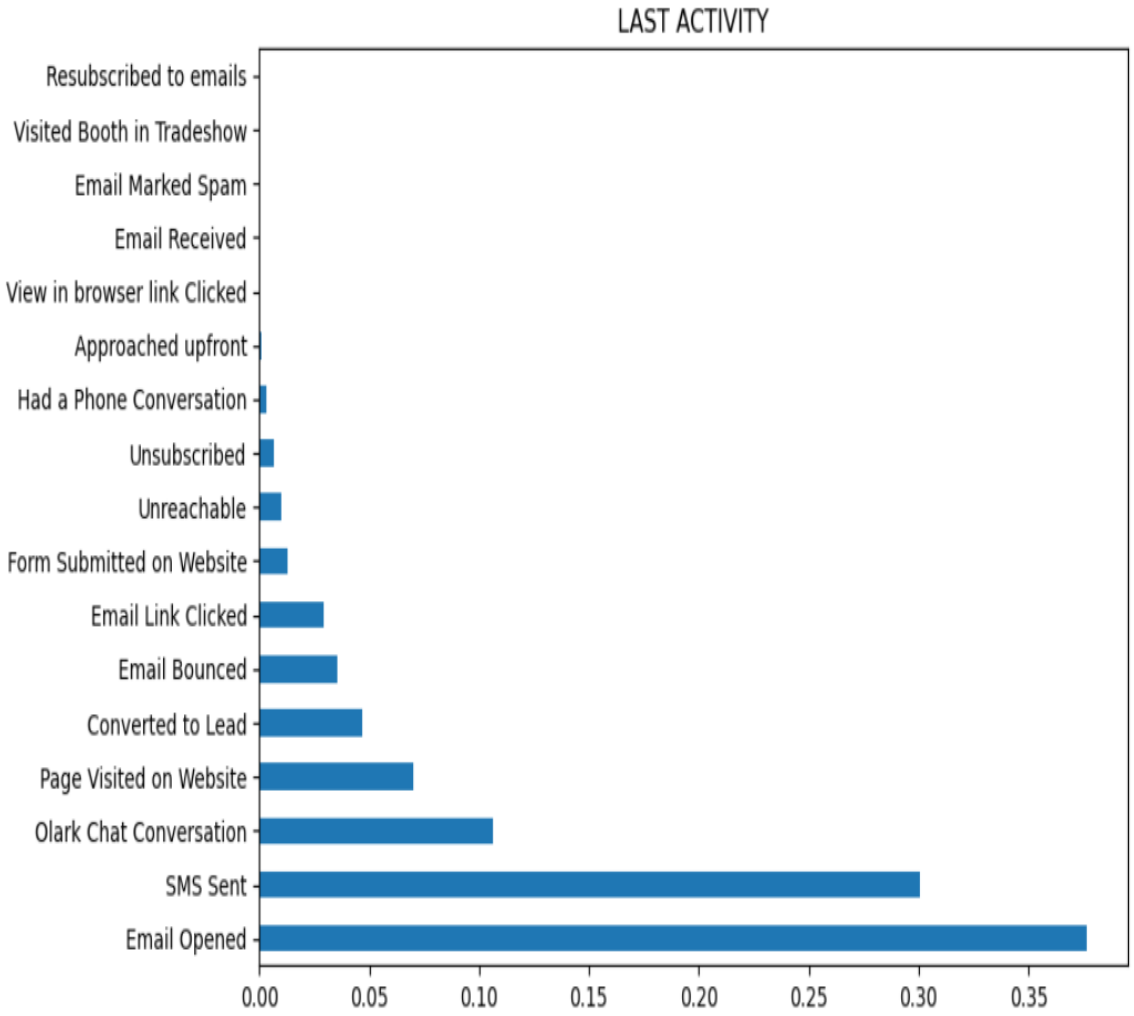
EDA



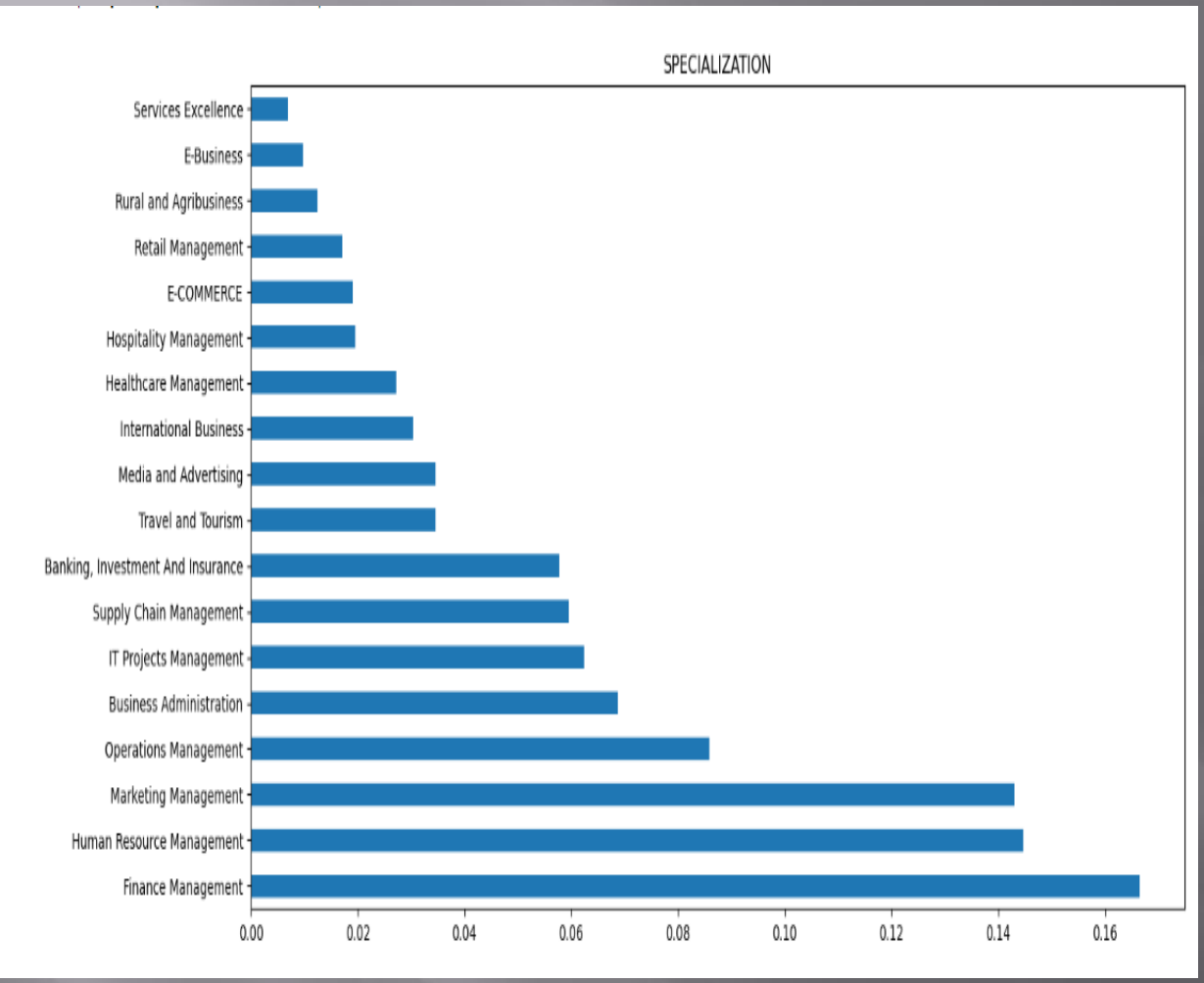
‘India’ as a country shows maximum amount of potential as compared to other places.



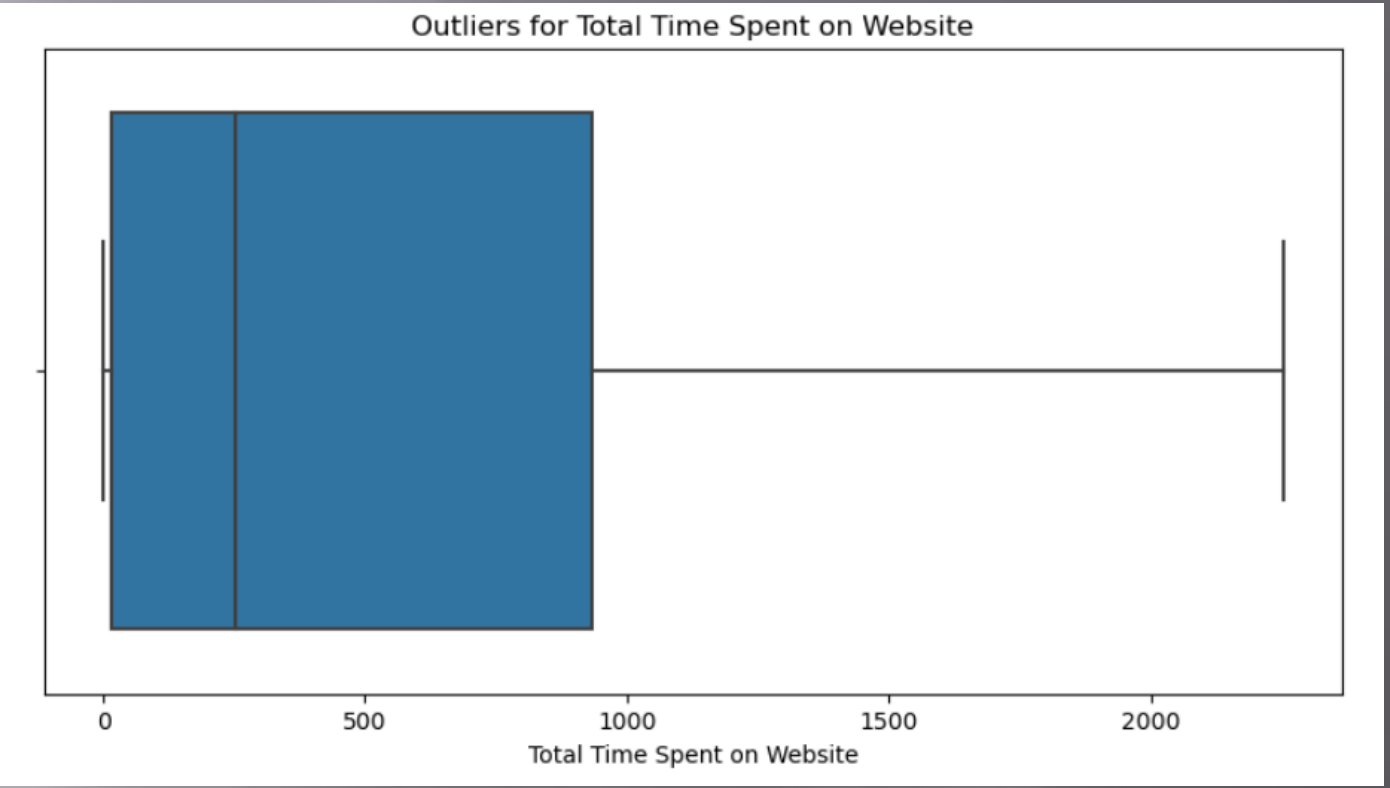
From the graph we can observe the highest lead source is ‘Google’ website and the least potential lies in Facebook.



Under the last activity, 'Email opened' shows more potential as compared to others and the least is observed in having a phone conversation.

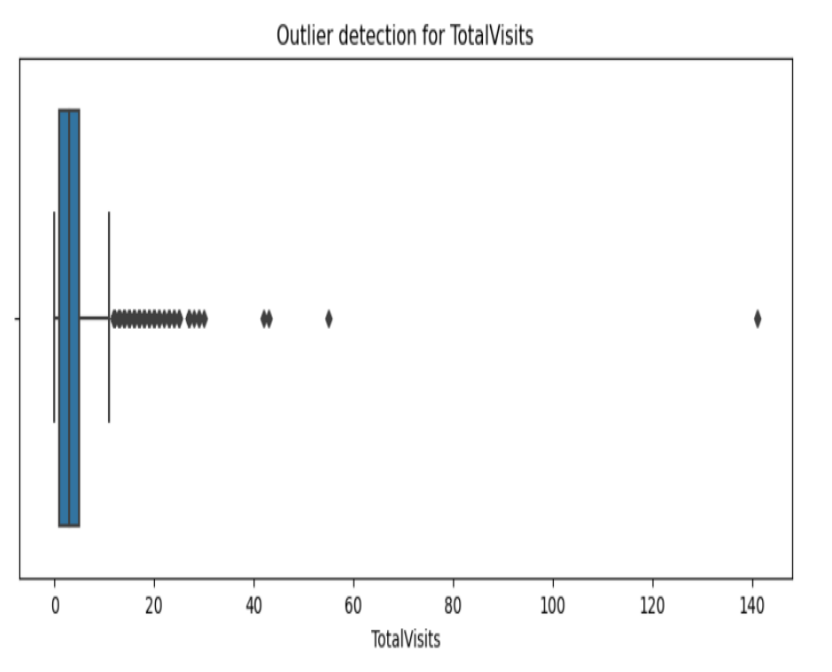


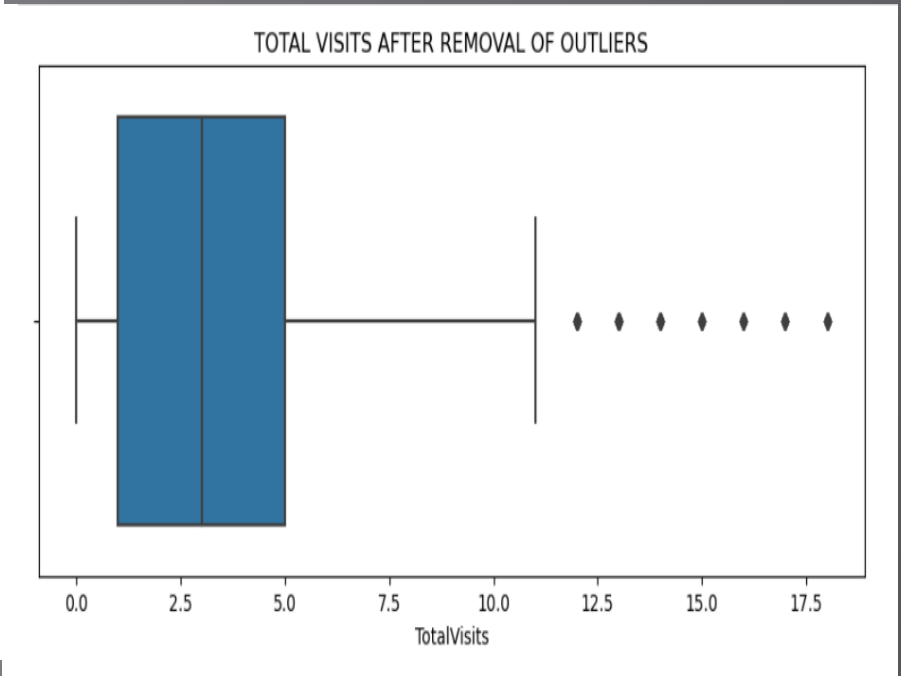
Under specialization, it is the 'Finance Management' which shows a higher potential followed by Human resource management. The least is seen in service Excellence.



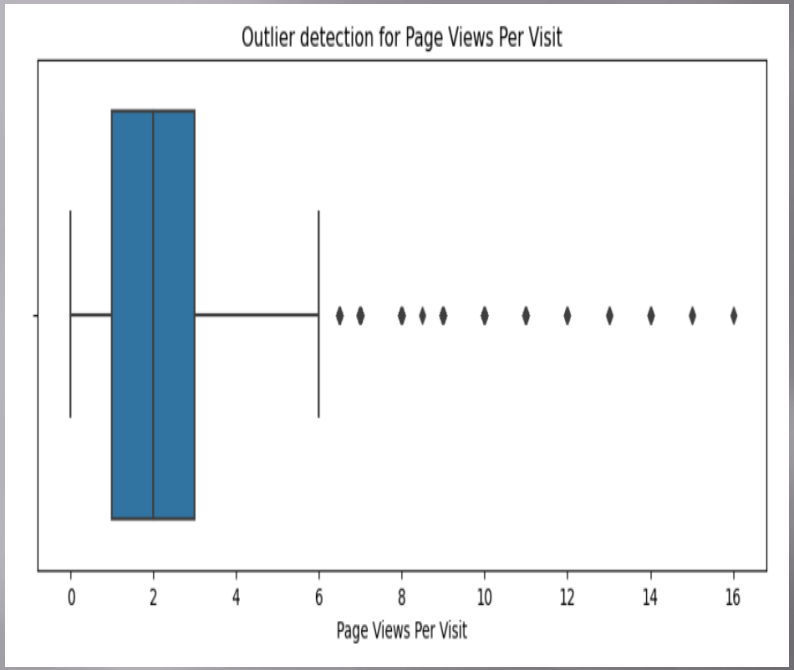
DETECTION OF OUTLIERS

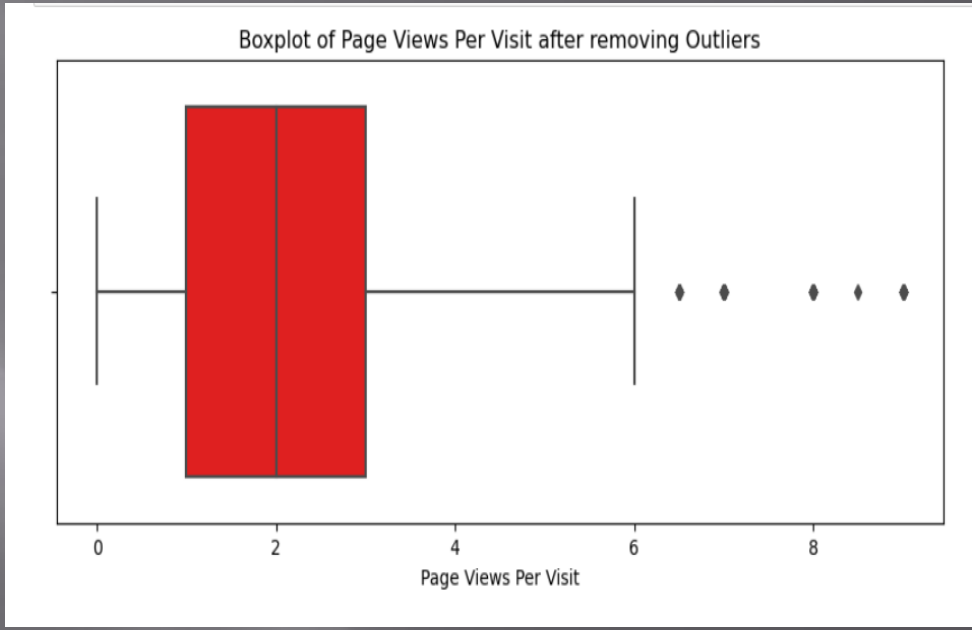
In the above visual, total time spent on website did not have any outliers.





DETECTION OF OUTLIERS AND SOLVING THE ISSUE UNDER TOTAL VISITS





DETECTION OF OUTLIERS AND SOLVING THE ISSUE UNDER PAGE VIEWS PER VISIT

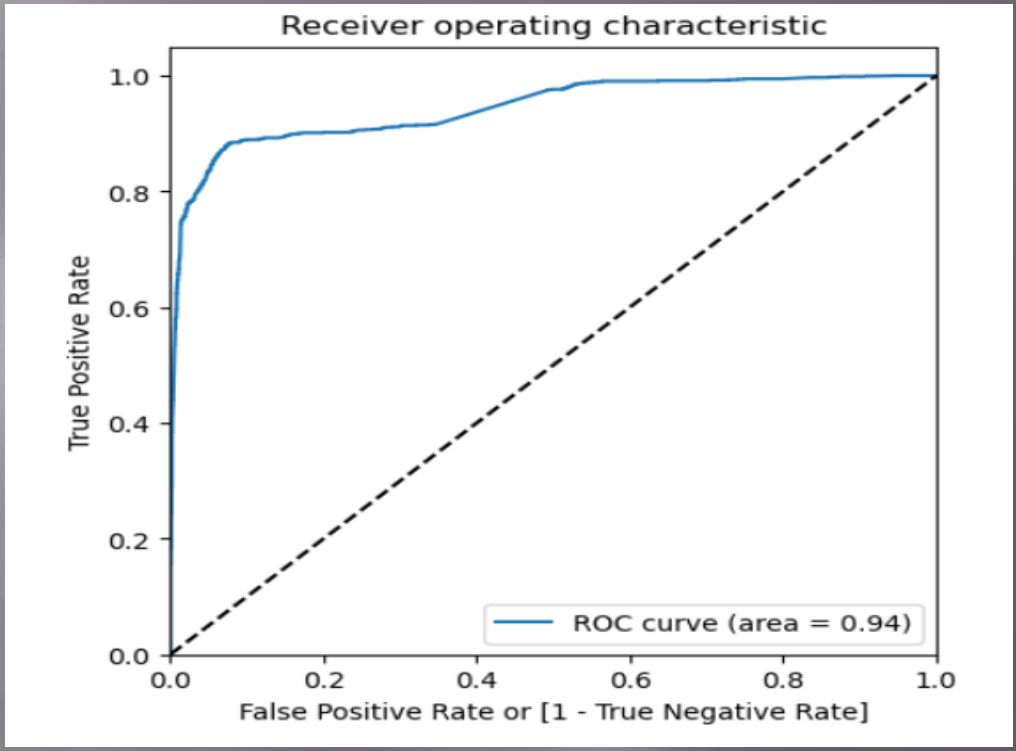
MODEL BUILDING

Splitting of model into train and test set.

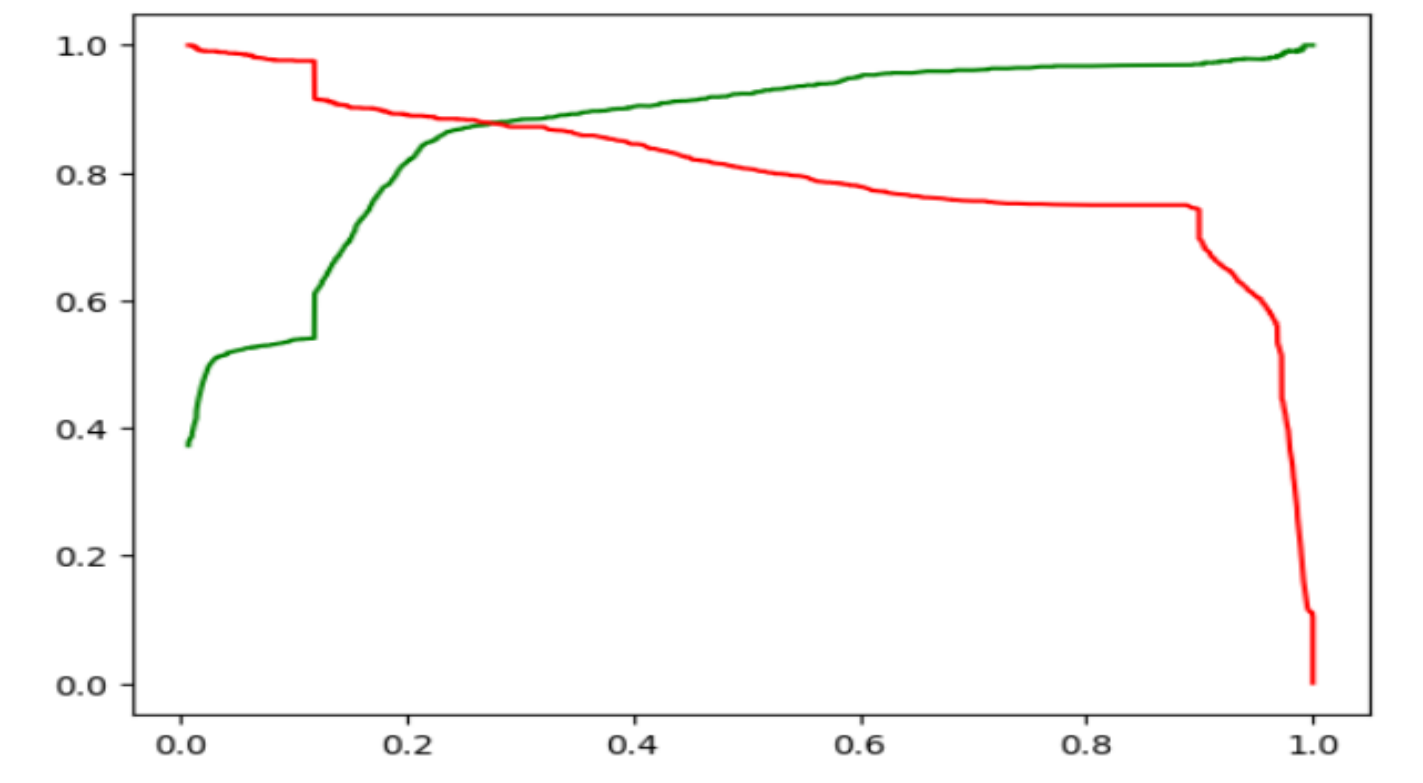
Building the model.

Using RFE to remove less relevant variables.

Evaluation of accuracy and other metrics.

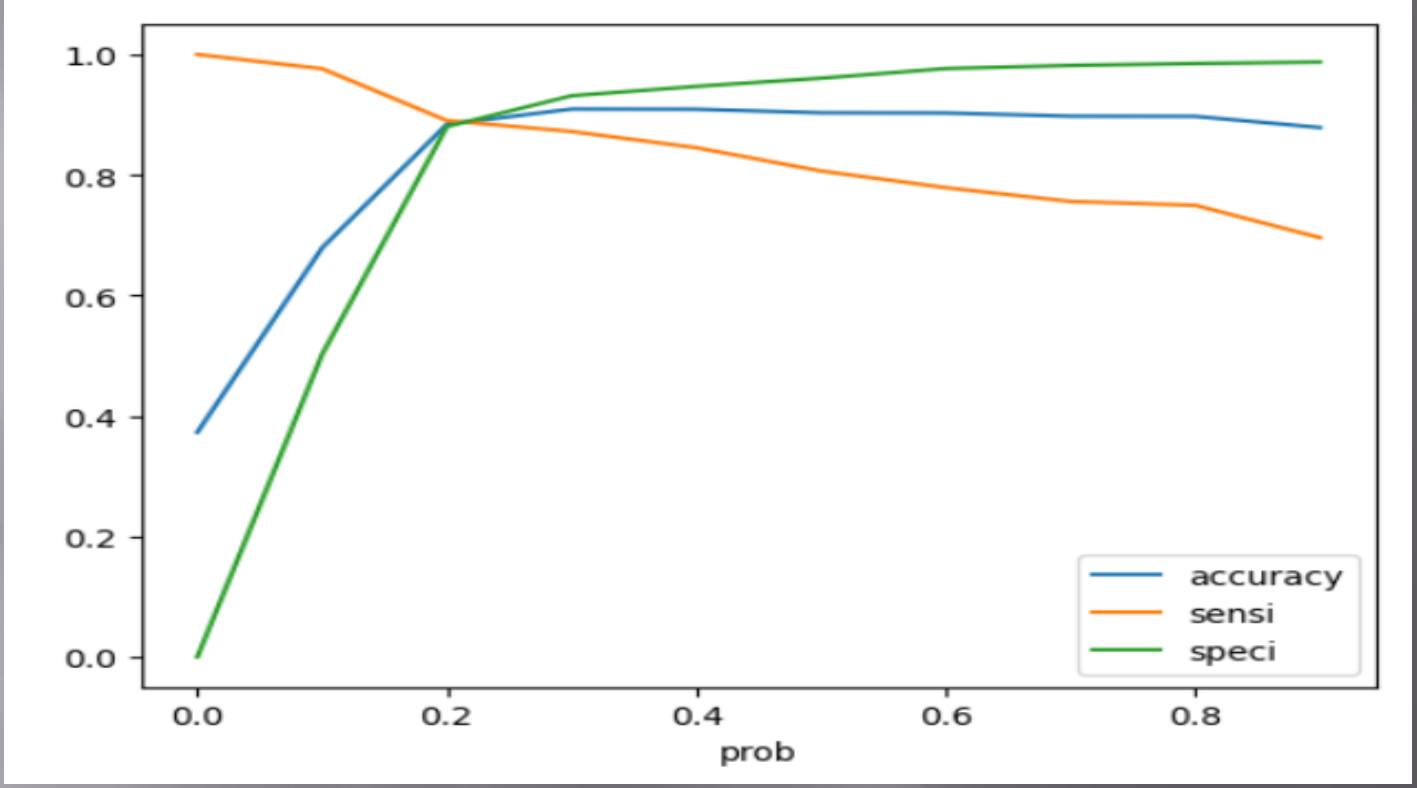


As the ROC curve is close to 1 hence we can estimate the model could be a good one.



For Train Data- Precision and Recall Predication

precision: 88.32%
recall: 87.20%



For Train Data- Overall Accuracy: 90.92%
Sensitivity: 87.20%
Specificity: 93.14%

CONCLUSION

To conduct the lead scoring case study, two major aspects were given prime focus apart from data understanding:

EDA: this technique gave an overview towards the whole scenario.

India was showing higher potential amongst all the other countries where as Google platform was mostly preferred as per lead source.

Outliers were found in certain cases that was taken care of.

Model building:

An ROC graph was plotted, which predicted how well the model could perform.

Moreover accuracy, specificity and sensitivity matrices of the model were also discovered.

THANK YOU