

LEAD SCORING CASE STUDY

BY: Pooja Agarwal & Indraneel Bhattacharjee

AGENDA

- I. Problem Statement
- II. Major Goals for Case Study
- III. Univariate Analysis for Categorical Variables
- IV. Univariate Analysis for Numerical Variables
- V. Bi-variate Analysis for Categorical Variables
- VI. ROC Curve
- VII. Model Metrics
- VIII. Model Evaluation and Metrics
- IX. Conclusion

PROBLEM STATEMENT

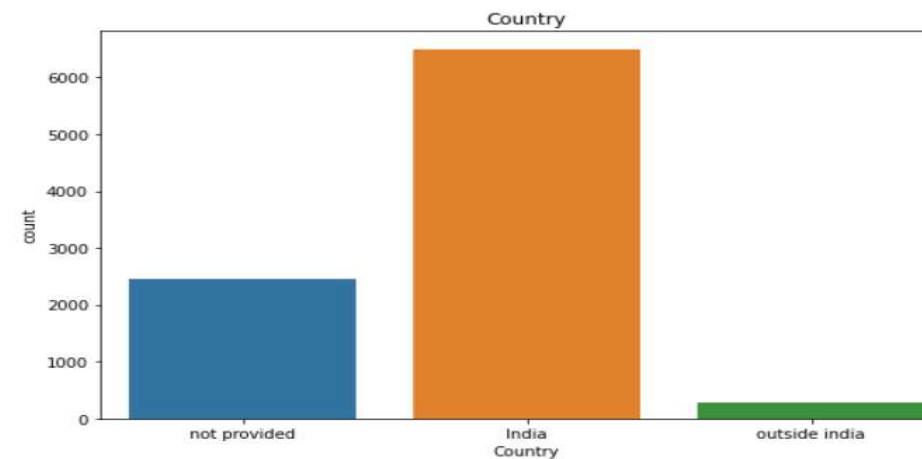
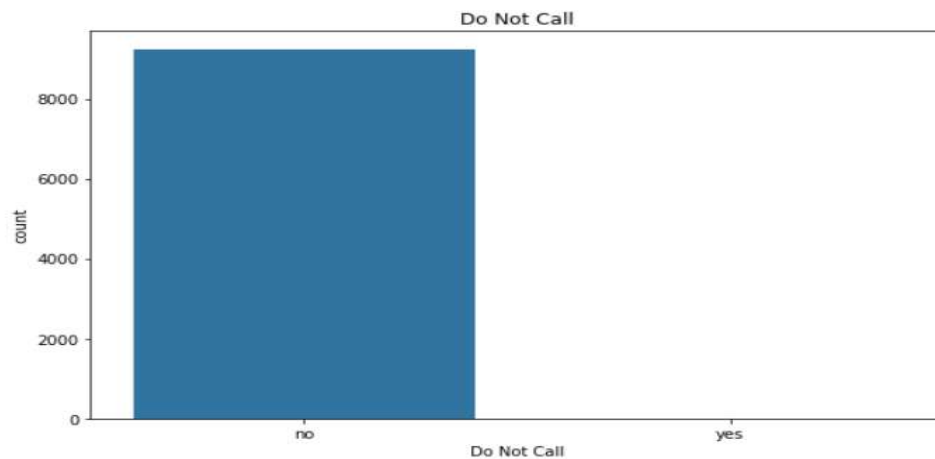
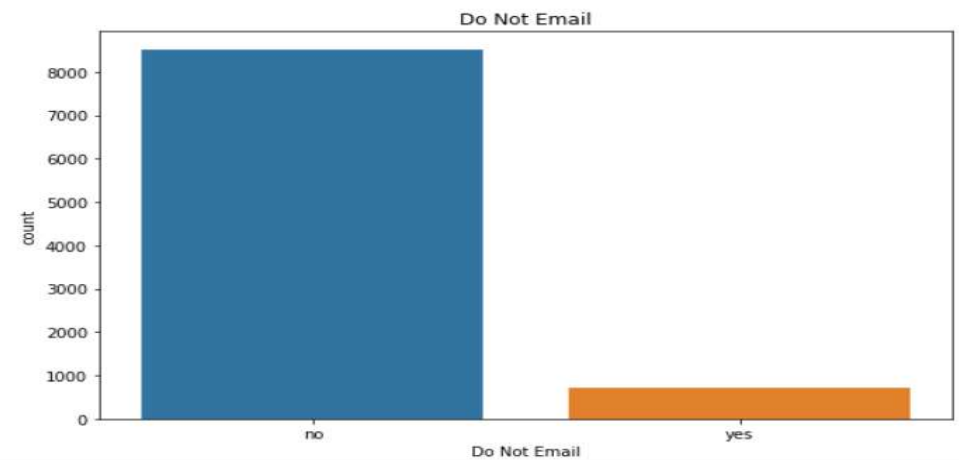
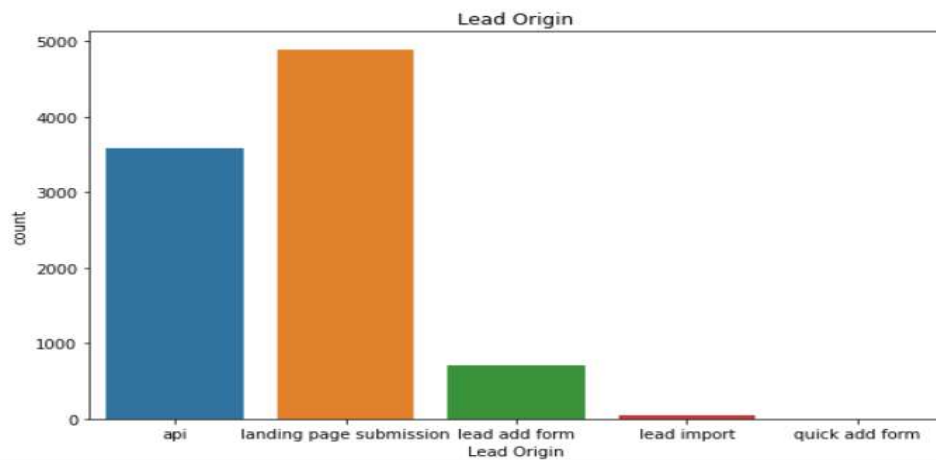
- ❑ An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on the website and browse for courses. The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos.
- ❑ X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- ❑ To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.
- ❑ If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.



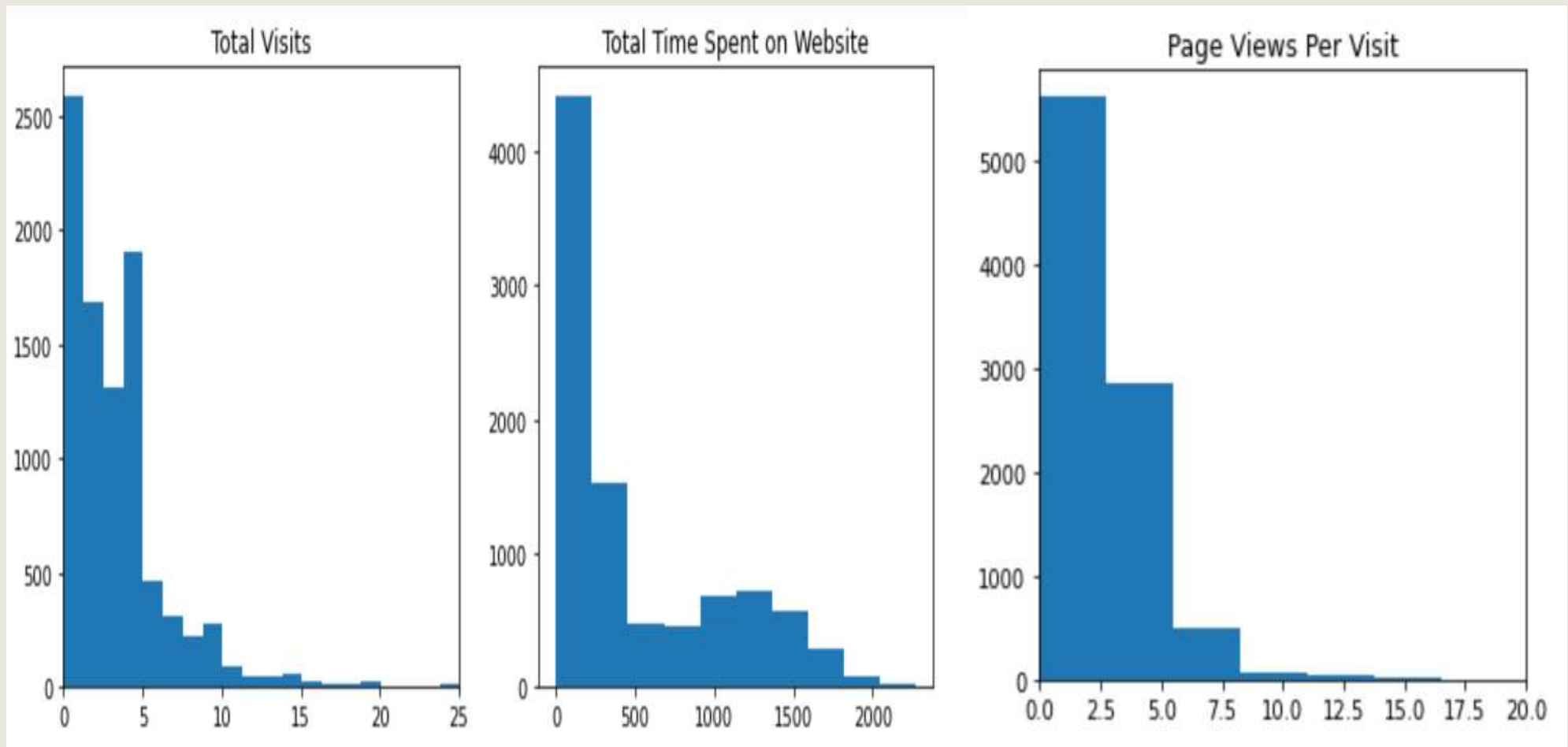
MAJOR GOALS FOR CASE STUDY

- Feature Scaling & Dummy Variables and encoding of the data.
- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.
- Validation of the model and model presentation.
- A higher score would mean that the lead is hot, i.e., is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

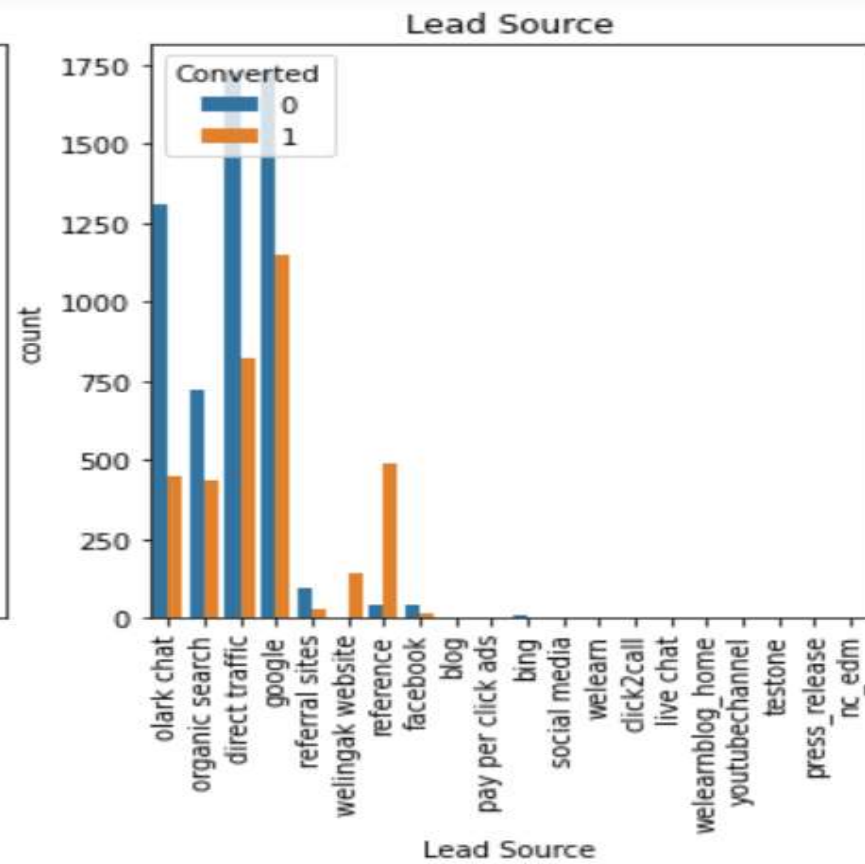
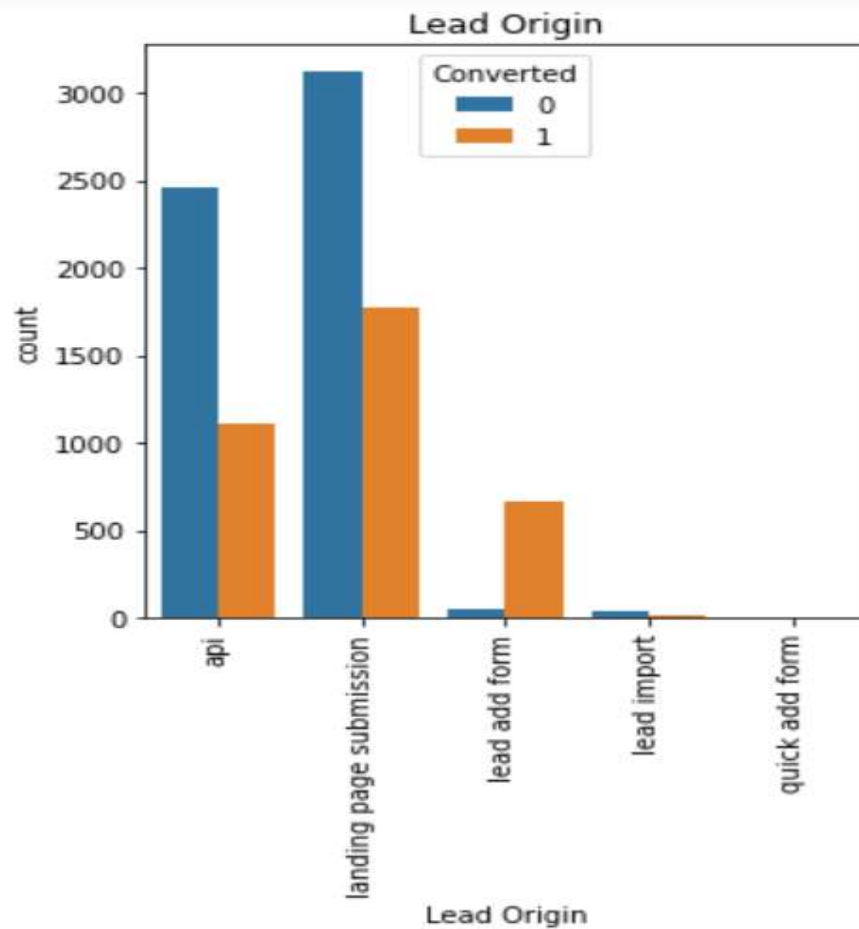
UNIVARIATE ANALYSIS FOR CATEGORICAL VARIABLES



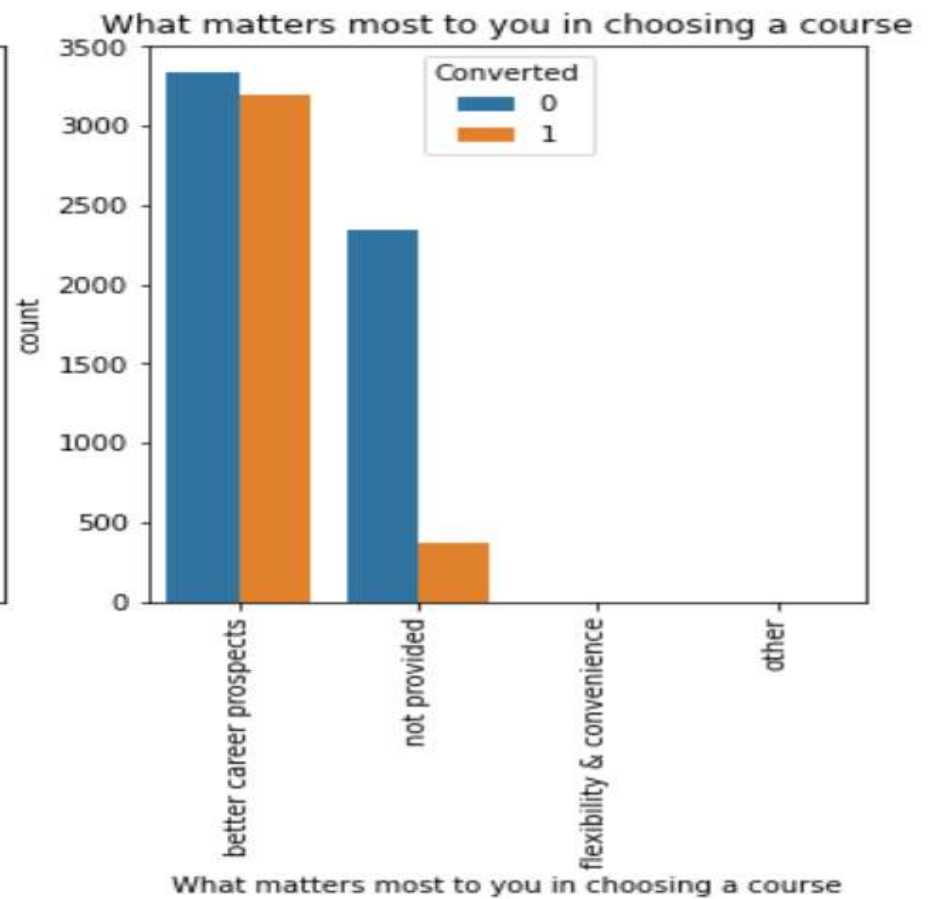
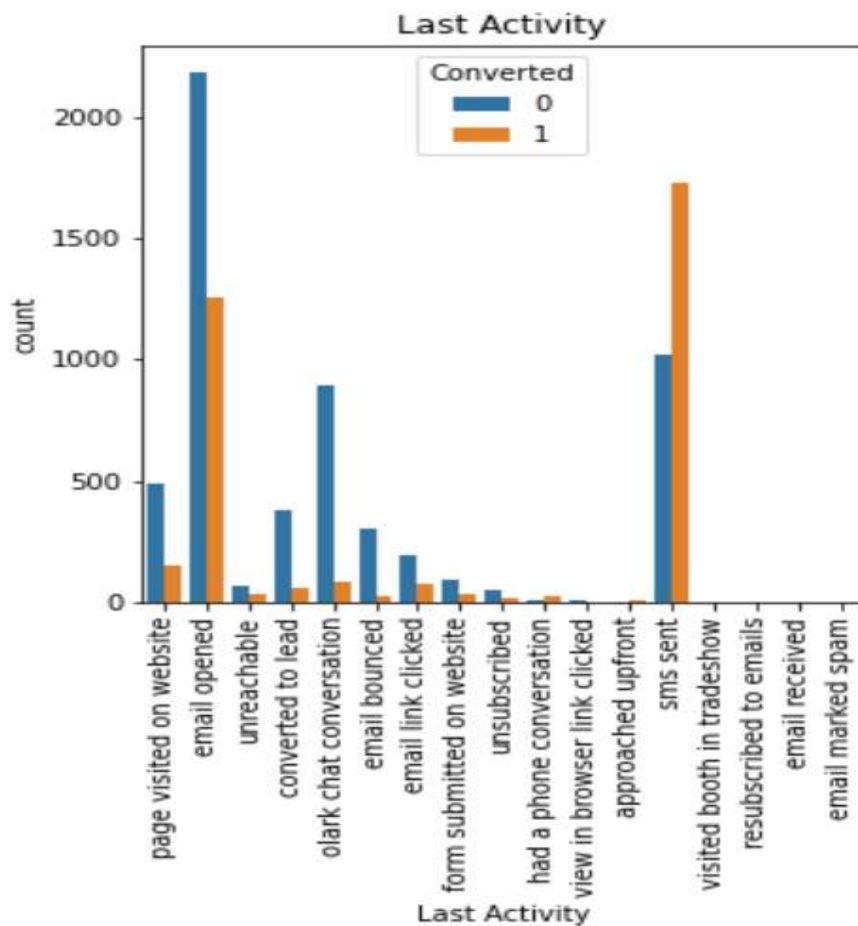
UNIVARIATE ANALYSIS FOR NUMERICAL VARIABLES



BI-VARIATE ANALYSIS FOR CATEGORICAL VARIABLES



BI-VARIATE ANALYSIS FOR CATEGORICAL VARIABLES

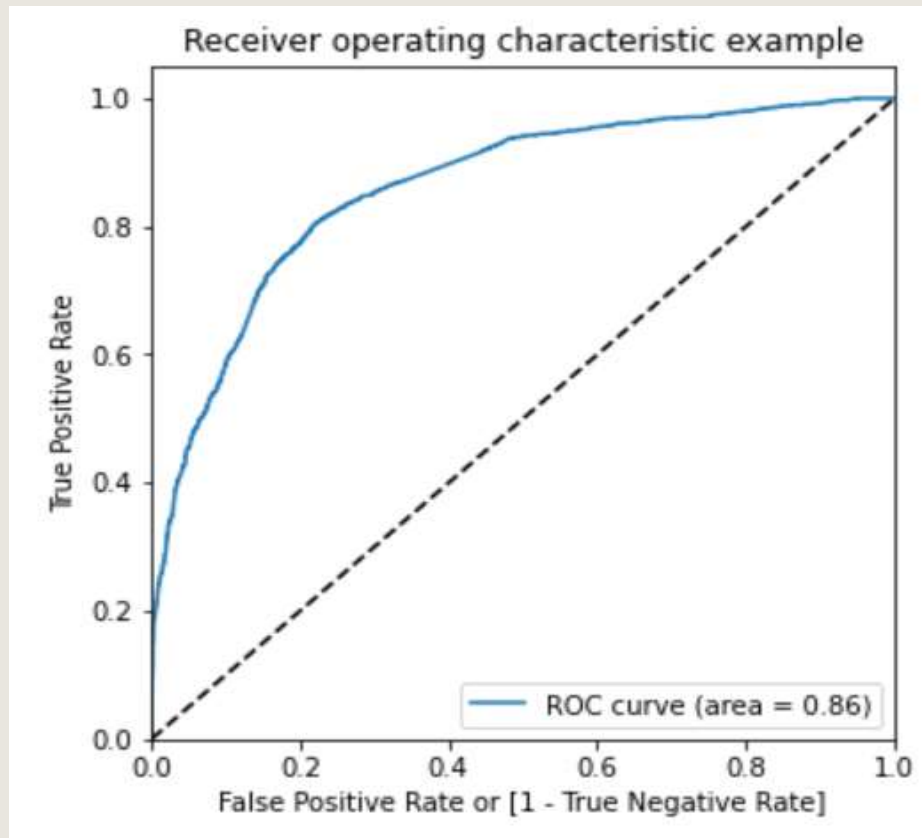


MODEL BUILDING

	Model 1	Model 2	Model 3	Model 4	Model 5
Df Model	15	14	13	12	11
Df_Residuals	5082	5083	5084	5085	5086
Pearson chi2	5.39e+03	5.39e+03	5.40e+03	5.40e+03	5.38e+03

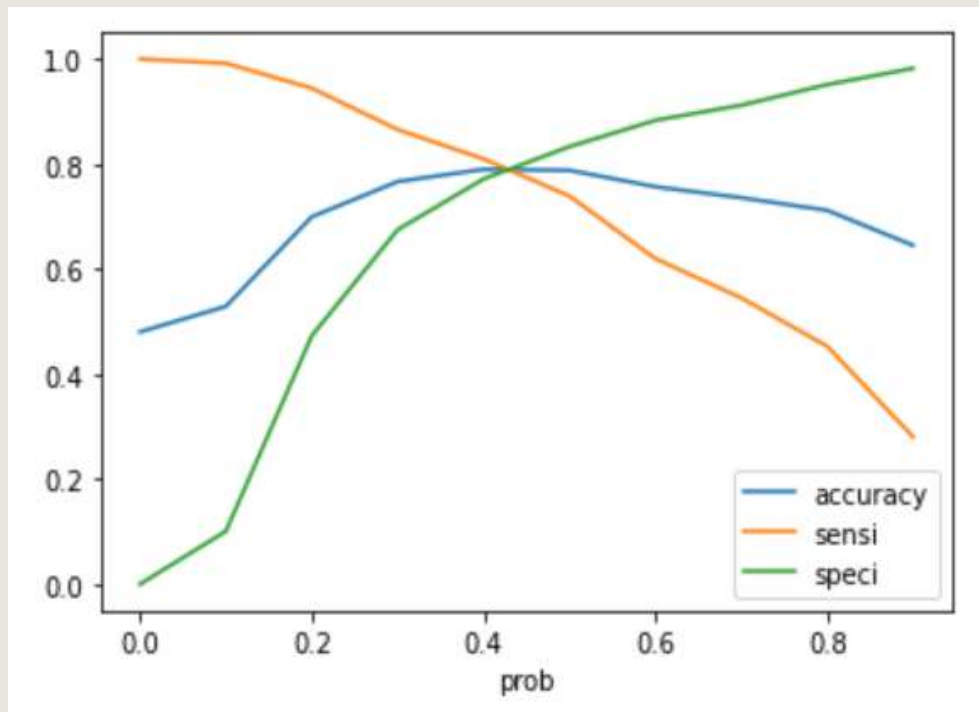
Model 5 was selected as final model as all the variables had VIF<5 and p-value<0.05

ROC CURVE



- ❑ It is observed that ROC curve area is 0.86 which is a better value and the model is good.
- ❑ Although for determining the best cutoff point, we will see the tradeoff between sensitivity and specificity.

MODEL METRICS



- ❑ We see 3 metrics (Accuracy, Sensitivity and Specificity) are on a threshold value of 0.42 which indicates that threshold value is optimized
- ❑ Thus, it can be used as threshold for checking test data.

MODEL EVALUATION & METRICS

	Cut-off @ 0.5 for train set	Cut-off @ 0.42 for train set	Cut-off @ 0.44 for train set	Cut-off @ 0.42 for test set
Accuracy	78.79%	79.01%	78.79%	77.64%
Sensitivity	73.90%	79.29%	73.90%	76.78%
Specificity	83.31%	78.74%	83.31%	78.45%
Precision	-	80.37%	-	76.91%
Recall	-	73.90%	-	76.78%

Model performs well for test set at cut-off 0.42 which indicates that model is not memorising the data as is for train set

CONCLUSION

- ❖ Model performs well for test set at cut-off 0.42 which indicates that model is not memorising the data as is for train set
- ❖ It is observed that the variables that mattered the most in the potential buyers are (In descending order):
 - What is your current occupation_Unemployed
 - Total Time Spent on Website
 - Total Visits
- ❖ To increase lead conversion, nurturing the potential leads can be done by educating the leads about the product and maintaining ongoing communication.
- ❖ Holding question-answer sessions with leads to extract the right information about them and further inquiries and appointments with the leads to determine their intention and mentality of them to join online courses.
- ❖ A list of leads should be kept handy so that we information can be shared with customers about new courses, services, job offers and future higher studies.
- ❖ A proper plan to chart the needs of each lead will go a long way to capture the leads as prospects by carefully providing job offerings, courses that suits best according to the interest of the leads.



THANK YOU