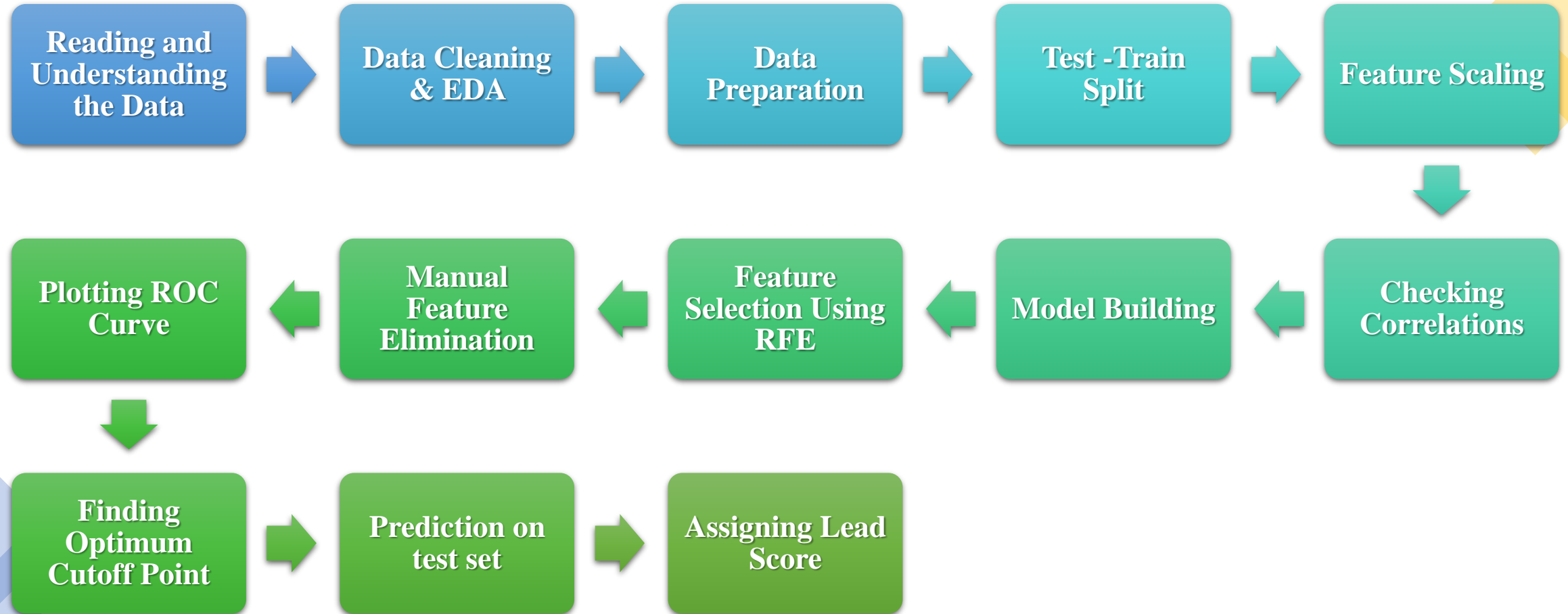# PROBLEM STATEMENT

An education company named X Education sells online courses to industry professionals. They wants to increase their lead conversion rate by identifying the most promising leads. They have appointed us to build a logistic regression model that assigns a lead score between 0-100 to each lead. A higher score indicates a higher conversion chance, while a lower score indicates a lower chance. The CEO's target is an 80% conversion rate. By using the lead score, X Education can focus on communicating with potential leads and increase their conversion rate.

ANALYSIS APPROCH-STEPS FOLLOWED

Reading and Understanding the Data → **Data Cleaning & EDA** → **Data Preparation** → **Test -Train Split** → **Feature Scaling**

↓

Plotting ROC Curve ← **Manual Feature Elimination** ← **Feature Selection Using RFE** ← **Model Building** ← **Checking Correlations**

↓

Finding Optimum Cutoff Point → **Prediction on test set** → **Assigning Lead Score**

IMPORTANT PLOTS- LEAD SOURCE

- From the graph we can see "Google" and "Direct Traffic" have high number of lead conversion. Definitely, online ads and SEO are playing important role.

- But the variable to focus is "Reference". It is third in terms of number of leads conversion but certainly 2nd highest in conversion rate with almost 92% conversion.

- Our model depicted "Reference" as the best attribute contributing in lead conversion with Coeff. of 4.19.

- So, focus must be to get as many referrals as possible.
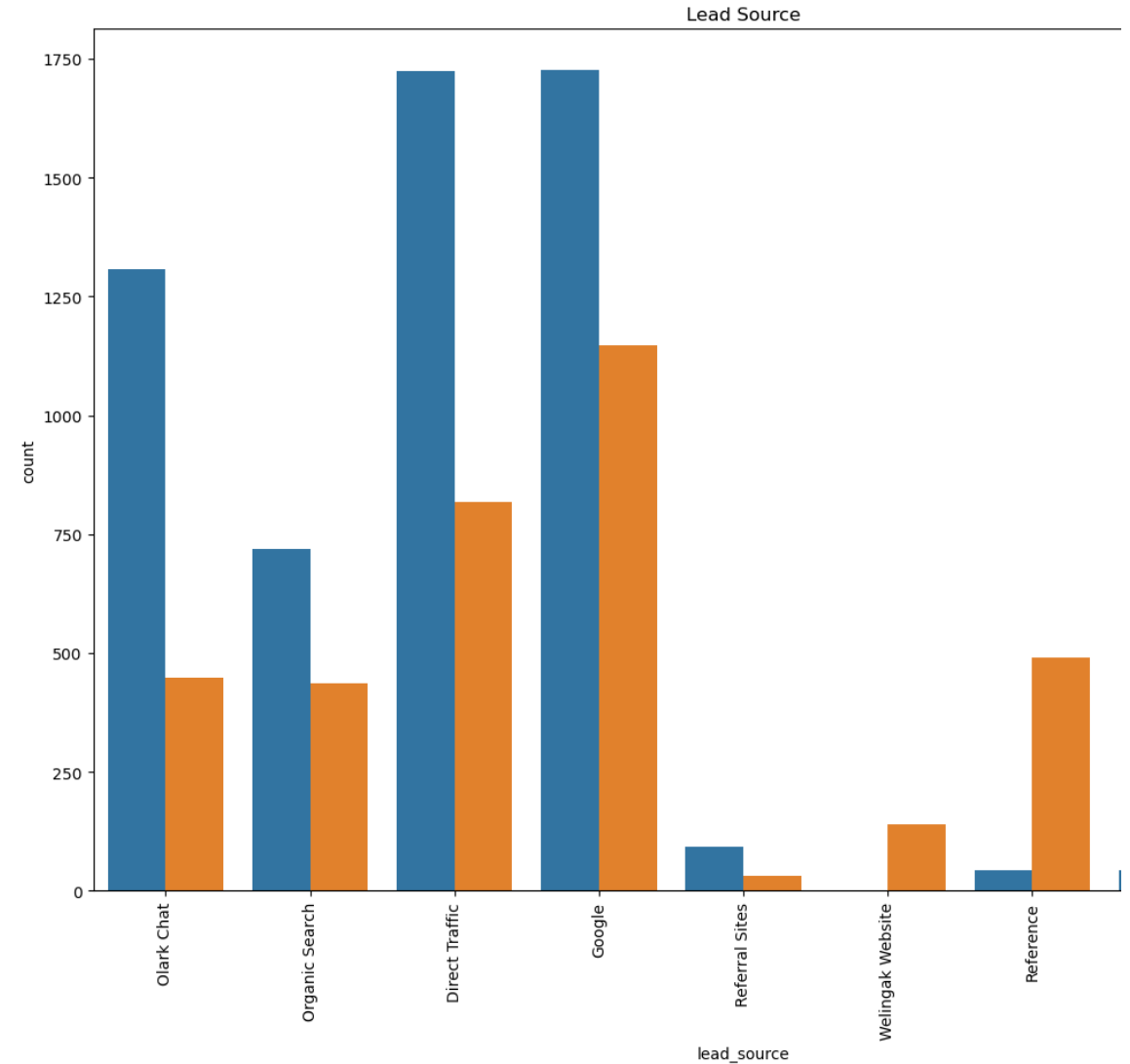
IMPORTANT PLOTS- LAST ACTIVITY

- From the plot we can see that "SMS Sent" has the highest conversion rate also the number of leads converted are highest.

- Also, our final model depicted "SMS Sent" as a significant contributor for conversion with Coeff. of 1.16 and is the 3rd in categorical variables selected by the model which positivity impacting the leads conversion.


Last Activity
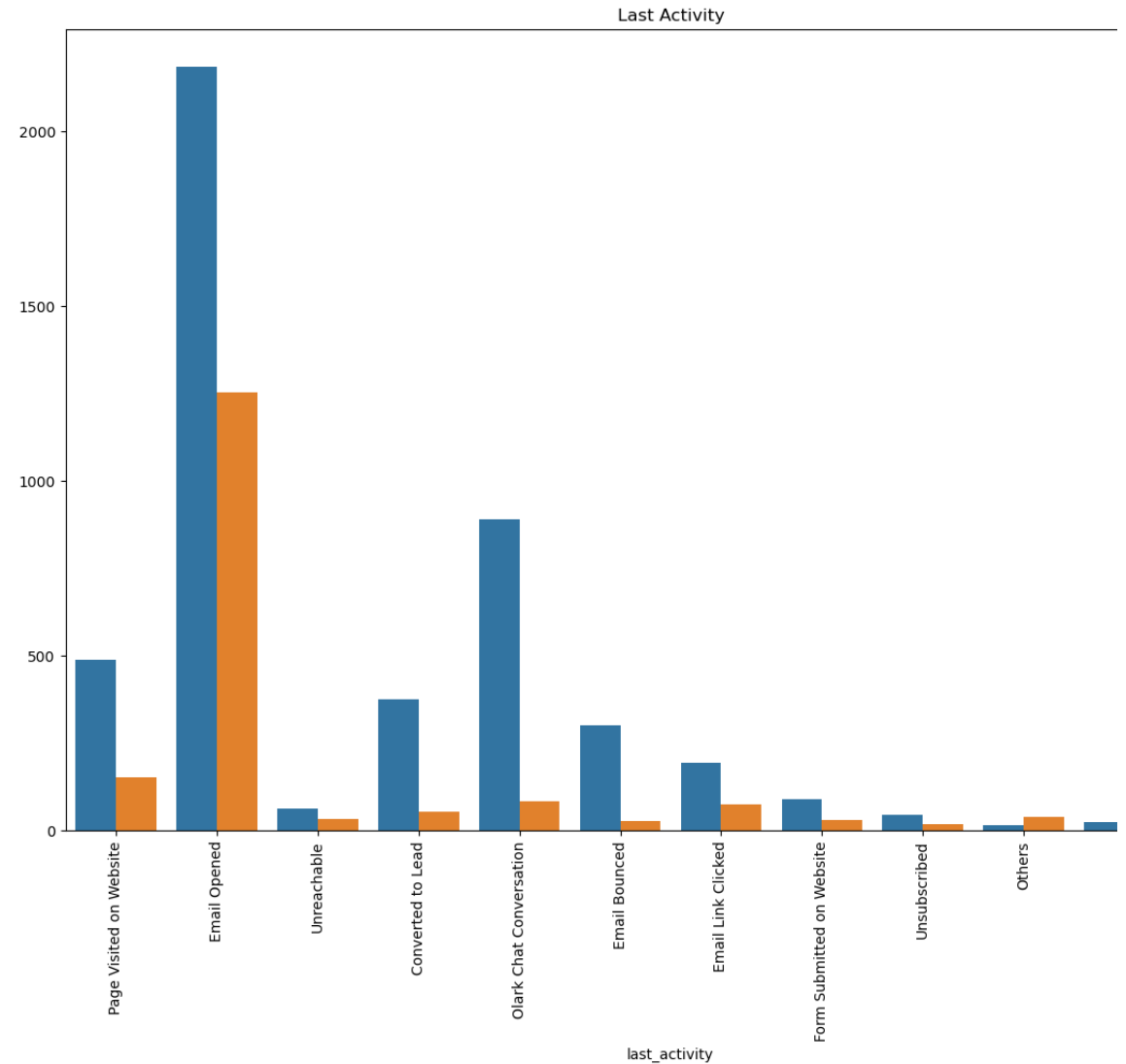
IMPORTANT PLOTS- CURRENT OCCUPATION

- From the plot we can see that "Unemployed" and "Working Professional" are the once which more interested in courses and thereby have higher chances of conversion.
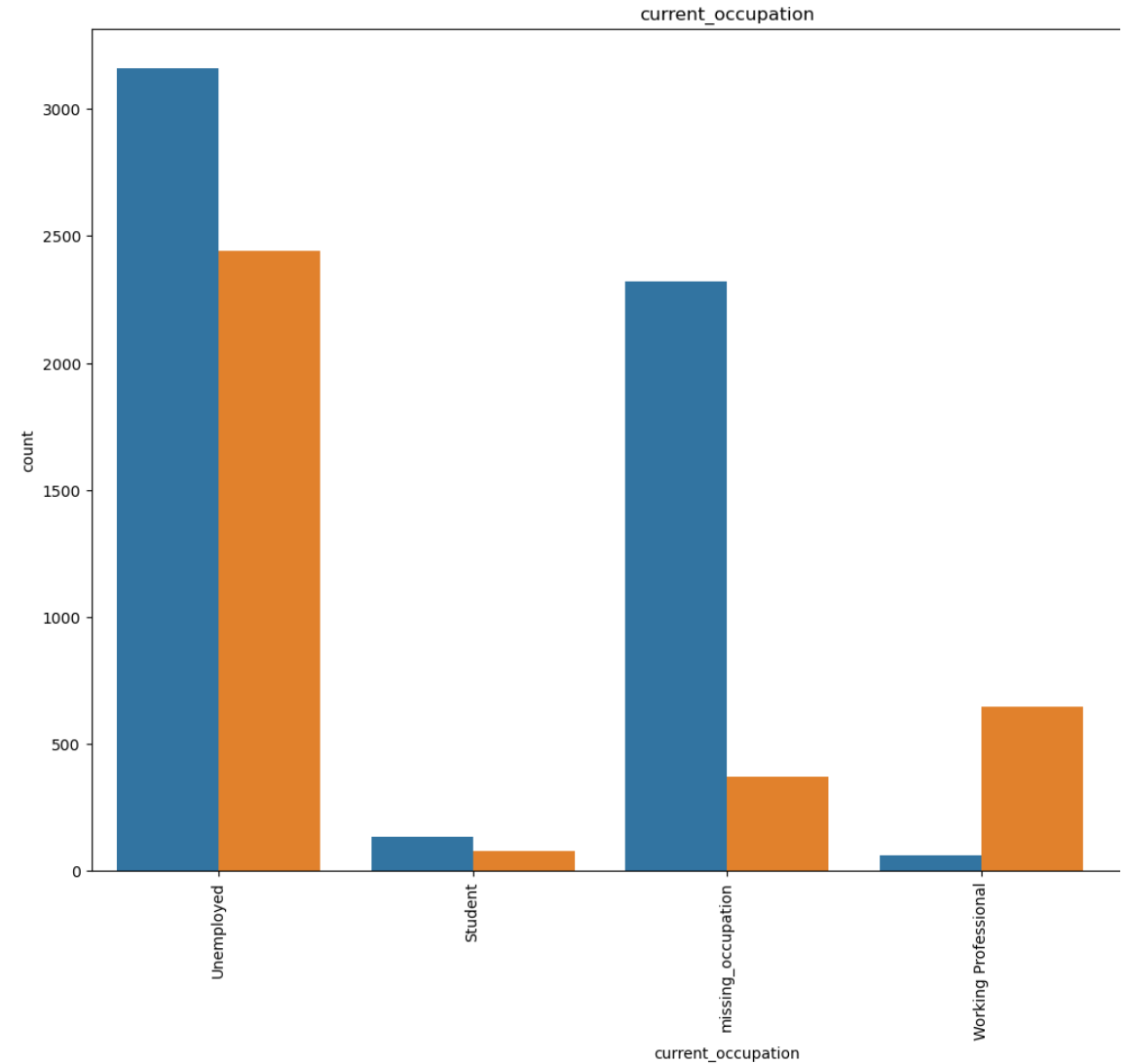
# IMPORTANT PLOTS-TOTAL TIME SENT ON WEBSITE

- From the plot we can clearly see that more the amount of time consumers spent on website higher the chances of conversion.

- We can see that event the Q1, i.e., 25% value of converted leads in greater than the 75% of leads not getting converted.

- Median of both are far apart and that also depicts the same thing.

- So, focus should be to get the customer to website.
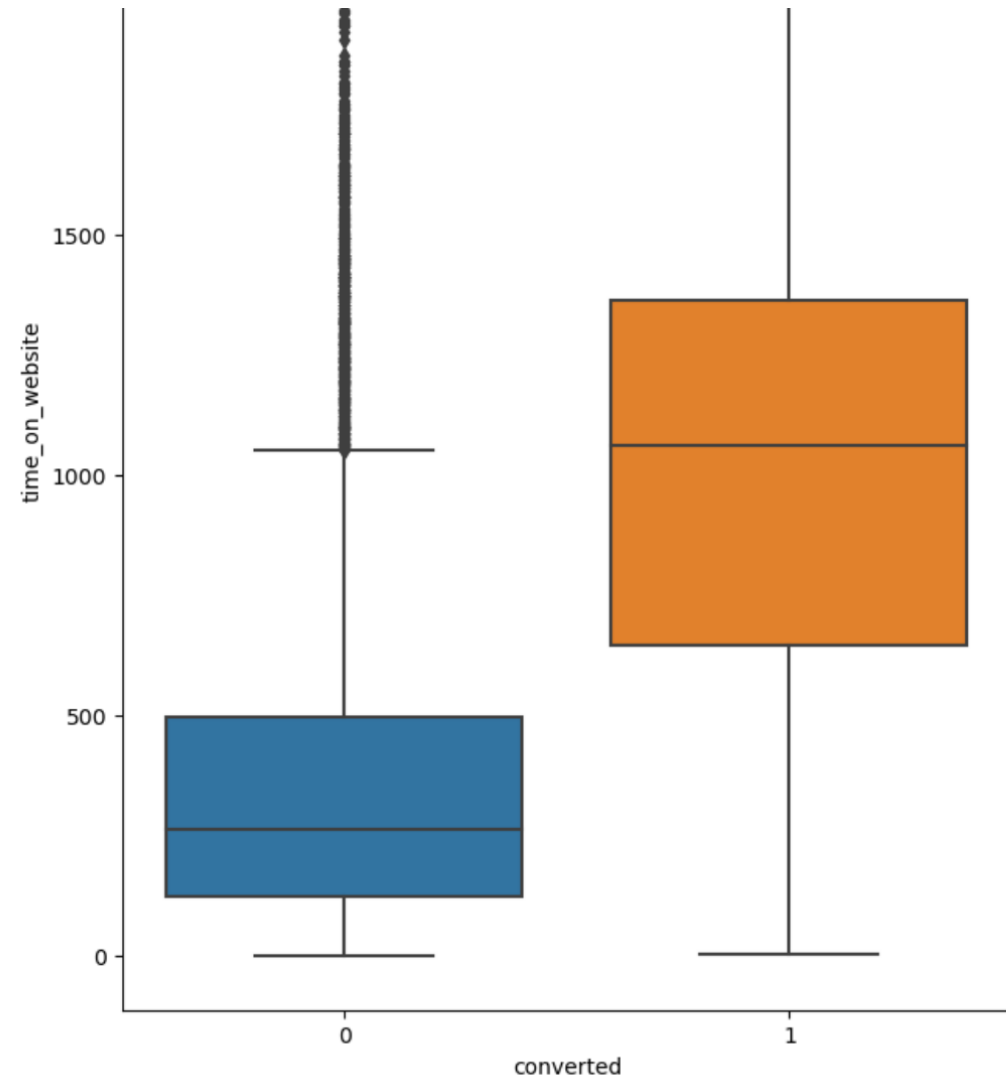
ROC CURVE, SENSITIVITY VS (1-SPECIVITY) TRADE-OFF, PRECISION-RECALL TRADE-OFF

- Area under ROC Curve = 0.87
- Threshold by Precision-Recall Curve – 0.42
- Threshold by Sensitivity-(1-Specivity) Curve – 0.38
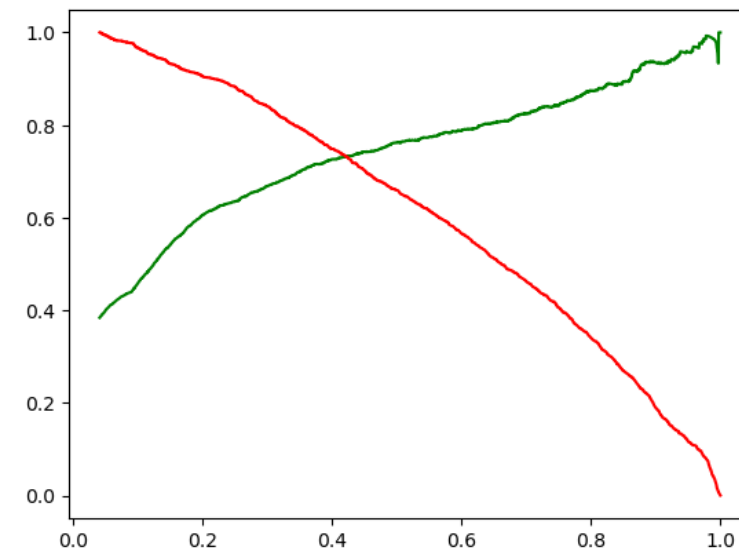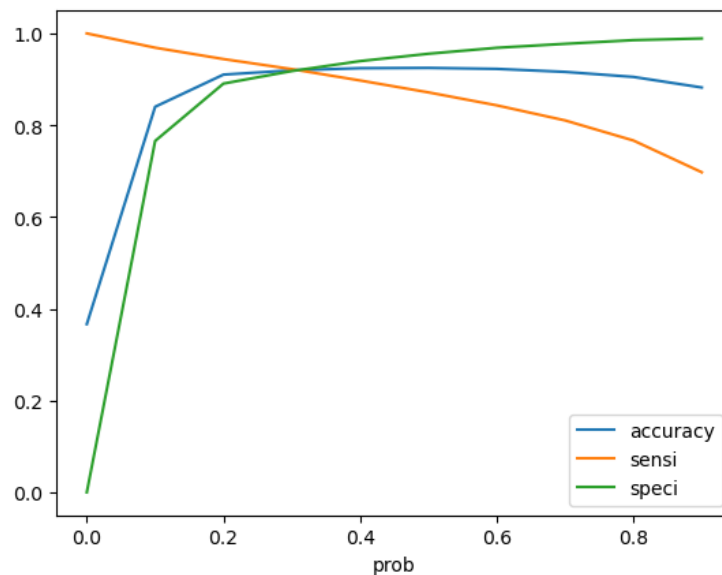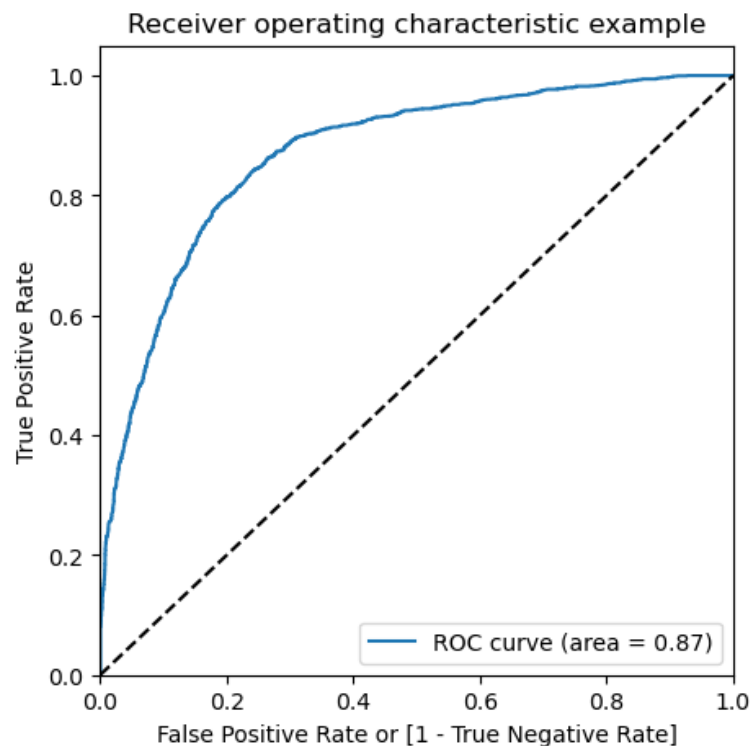
PERFORMANCE ANALYSIS OF FINAL MODEL

- We used 2 evaluation matrix, i.e., Sensitivity-Specificity and Precision-Recall both.

- We got almost same accuracy on test and train set in both matrix.

- But, from our business problem point of view, we are more concerned about predicting the 1s better i.e., more lead conversion.

- This can be analysed by precision. We got same value in both matrix on test set. So, finalised model based on accuracy.

- Final threshold based on precison-recall matrix i.e., 0.42.

Final Oberservations & Conclusion-

1. Using Sensitivity & Specificity as Evaluation Matrix-

- Results on Train Set-
  - Model Accuracy - 80.56 %
  - Sensitivity - 79.14 %
  - Specificity - 81.39 %
  - Precision - 71.1 %
- Results on Test Set-
  - Model Accuracy - 78.76 %
  - Sensitivity - 73.66 %
  - Specificity - 81.73 %
  - Precision - 69.99 %

2. Using Precision & Recall as Evaluation Matrix-

- Results on Train Set-
  - Model Accuracy - 80.93 %
  - Precision - 73.56 %
  - Recall - 74.87 %
- Results on Test Set-
  - Model Accuracy - 79.21 %
  - Precision - 69.99 %
  - Recall - 73.66 %

TOP ATTRIBUTE ANALYSIS

- The top three variables which contributed most positively towards the probability of a lead getting converted in order are:

- lead_source_Reference – Coeff. 4.19

- current_occupation_Working Professional - Coeff. 3.81

- time_on_website (*originally-total_time_spent_on_website*)- Coeff. 1.16. or

- last_activity_SMS Sent – Coeff. 1.16

- If we talk about categorical variables only, then "SMS Sent" is the 3rd variable contributed to conversion of leads.

Generalized Linear Model Regression Results

Dep. Variable:	converted	No. Observations:	4734
Model:	GLM	Df Residuals:	4725
Model Family:	Binomial	Df Model:	8
Link Function:	Logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-1995.9
Date:	Mon, 27 Feb 2023	Deviance:	3991.8
Time:	12:20:19	Pearson chi2:	4.74e+03
No. Iterations:	6	Pseudo R-squ. (CS):	0.3756
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.7667	0.101	-17.408	0.000	-1.966	-1.568
do_not_email	-1.5699	0.208	-7.556	0.000	-1.977	-1.163
time_on_website	1.1680	0.042	27.786	0.000	1.086	1.250
lead_source_Reference	4.1904	1.133	3.700	0.000	1.971	6.410
last_activity_SMS Sent	1.1667	0.084	13.964	0.000	1.003	1.330
current_occupation_Student	0.9593	0.292	3.287	0.001	0.387	1.531
current_occupation_Unemployed	1.0824	0.100	10.812	0.000	0.886	1.279
current_occupation_Working Professional	3.8117	0.231	16.500	0.000	3.359	4.264
last_notable_activity_Modified	-0.7367	0.090	-8.164	0.000	-0.914	-0.560

RECOMMENDATIONS

Referrals- Getting more will referrals should be the top focus of the company. As, leads which come through referrals had trust of their known ones who have already enrolled in the courses in past, so it becomes little easy to convince and convert.

Building automated system that sends **SMS** notifications to high priority leads. As, users generally ignore the email or sometimes emails bounce as well, but SMS is one thing that will pop up as notification and will definitely get attention. Also make sure that SMS contain website link, so that user can directly reach the website.

RECOMMENDATIONS

Focus on Working Professionals- As they do not want to lose their job but want to upskill as per the industry demand. Also, they will be the one who will actually pay fee for expensive courses as well.

Increase user engagement on their website- This can be done by many. While SEO optimization, digital marketing and Ad Sense will get the user to the portal but after that there should some content, blogs, articles, future profile after completing the course, free demo lectures, demo of working model, which will keep the customer engaging and will definitely make him curious to contact the team to know about the courses.