



LEAD SCORING CASE-STUDY

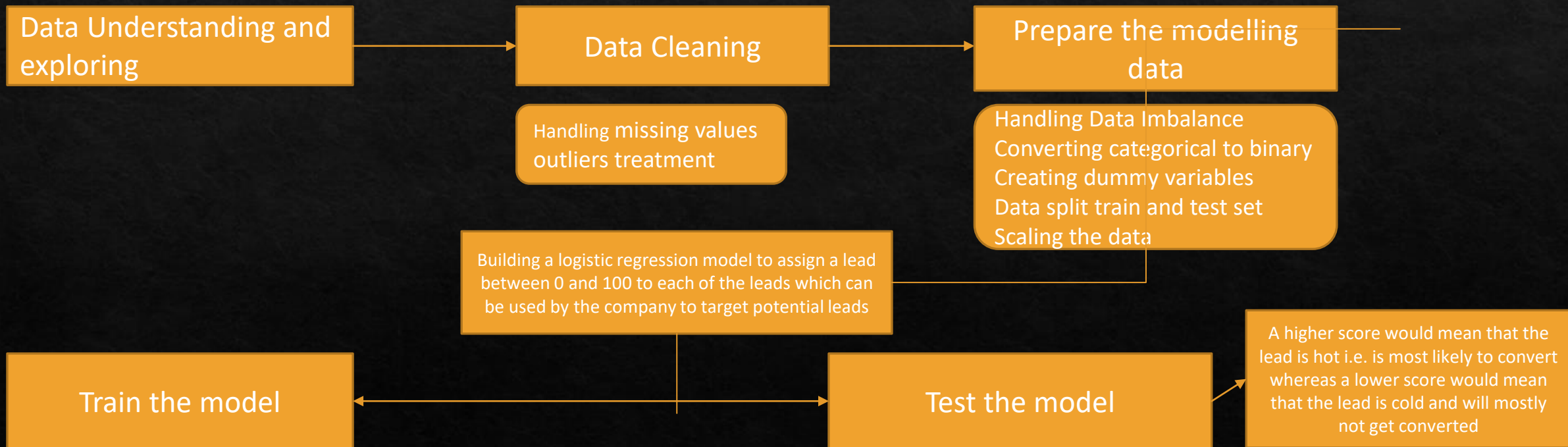
By Preeti Kumari, Shital
Dhuwadhapare, and
KetanKumar Dorik

Problem statement:-

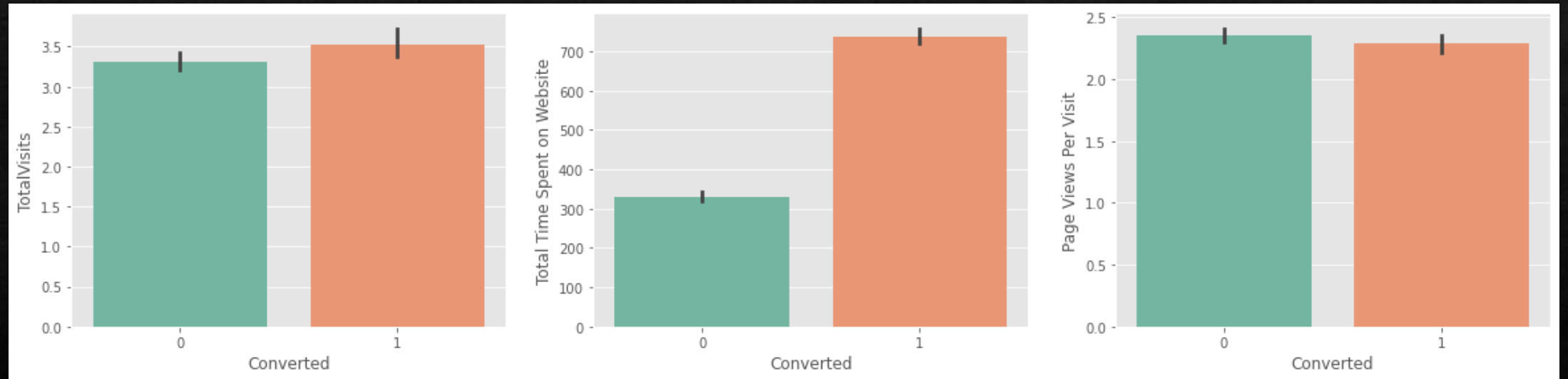
Identify the set of leads of X Education so that the lead conversion rate should go up and the sales team of the company focus more on communication with the potential leads rather than making calls to every customer.

The company wants to increase the lead conversion to 80%

Analysis approach:-

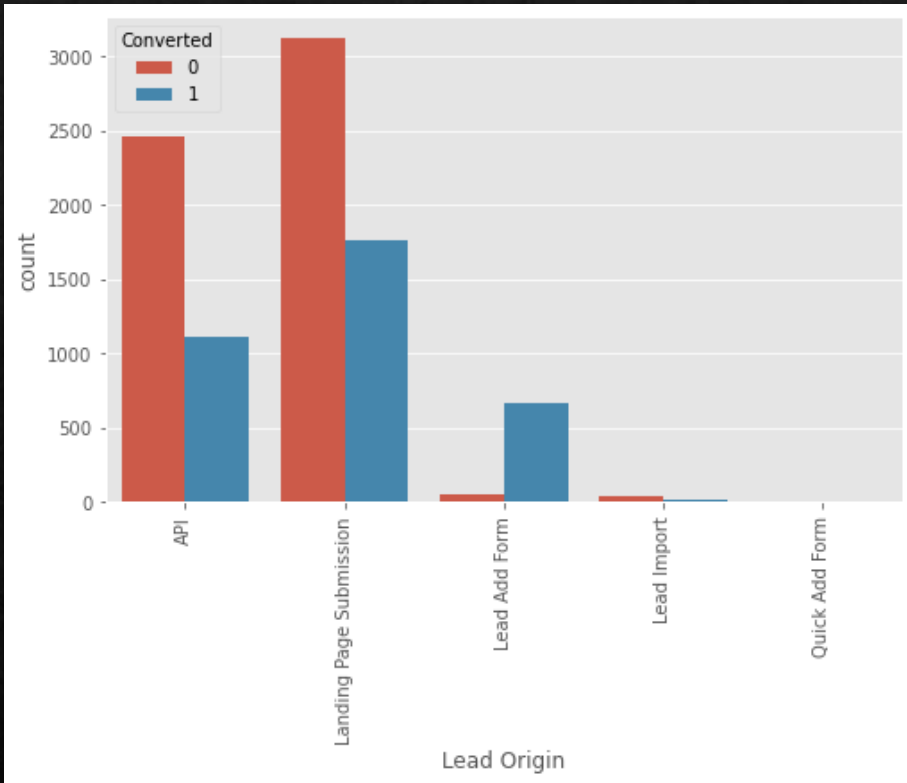


Exploratory Data Analysis



The conversion rated were high for Total Visits, Total Time Spent on Website and Page Views Per Visit

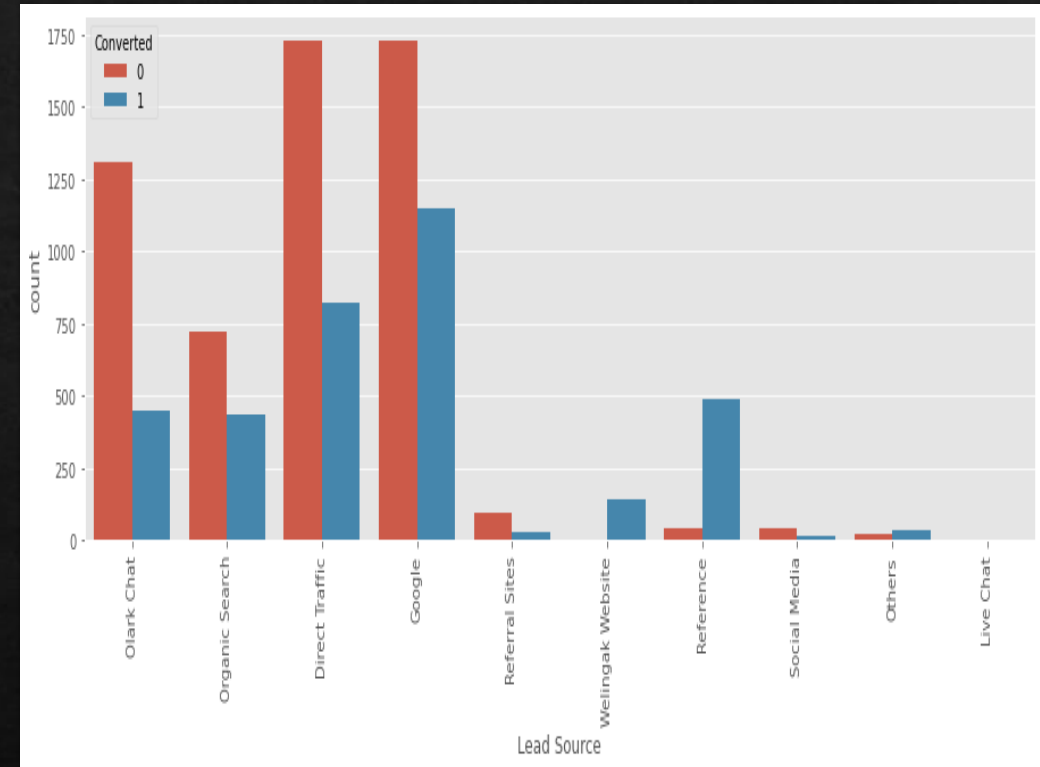
Lead Origin



Inference

- API and Landing Page Submission bring higher number of leads as well as conversion.
- Lead Add Form has a very high conversion rate but count of leads are not very high.
- Lead Import and Quick Add Form get very few leads.
- In order to improve overall lead conversion rate, we have to improve lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.

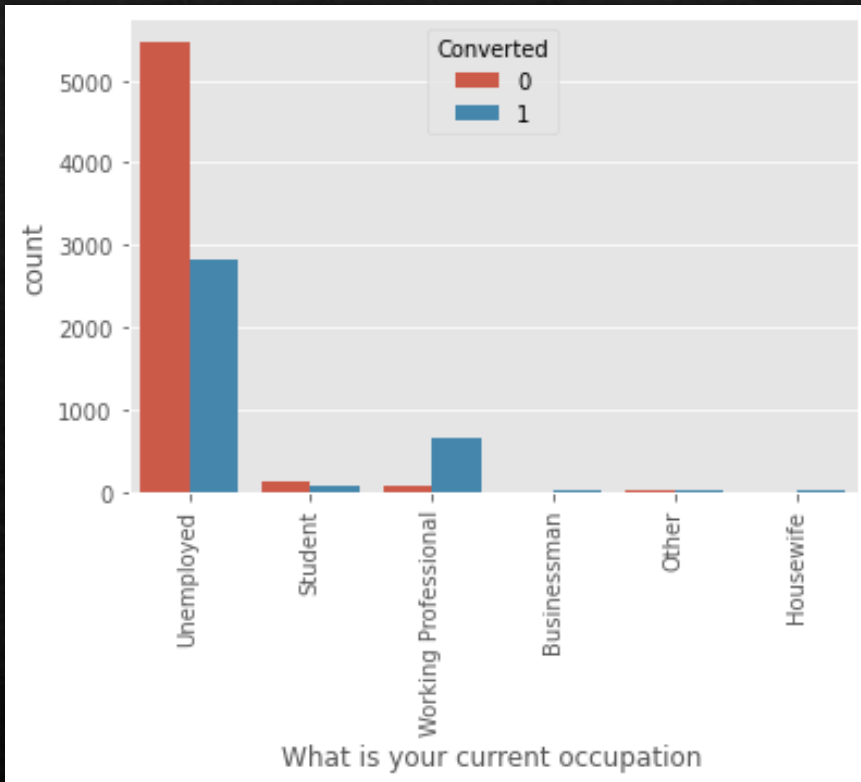
Lead Source



Inference

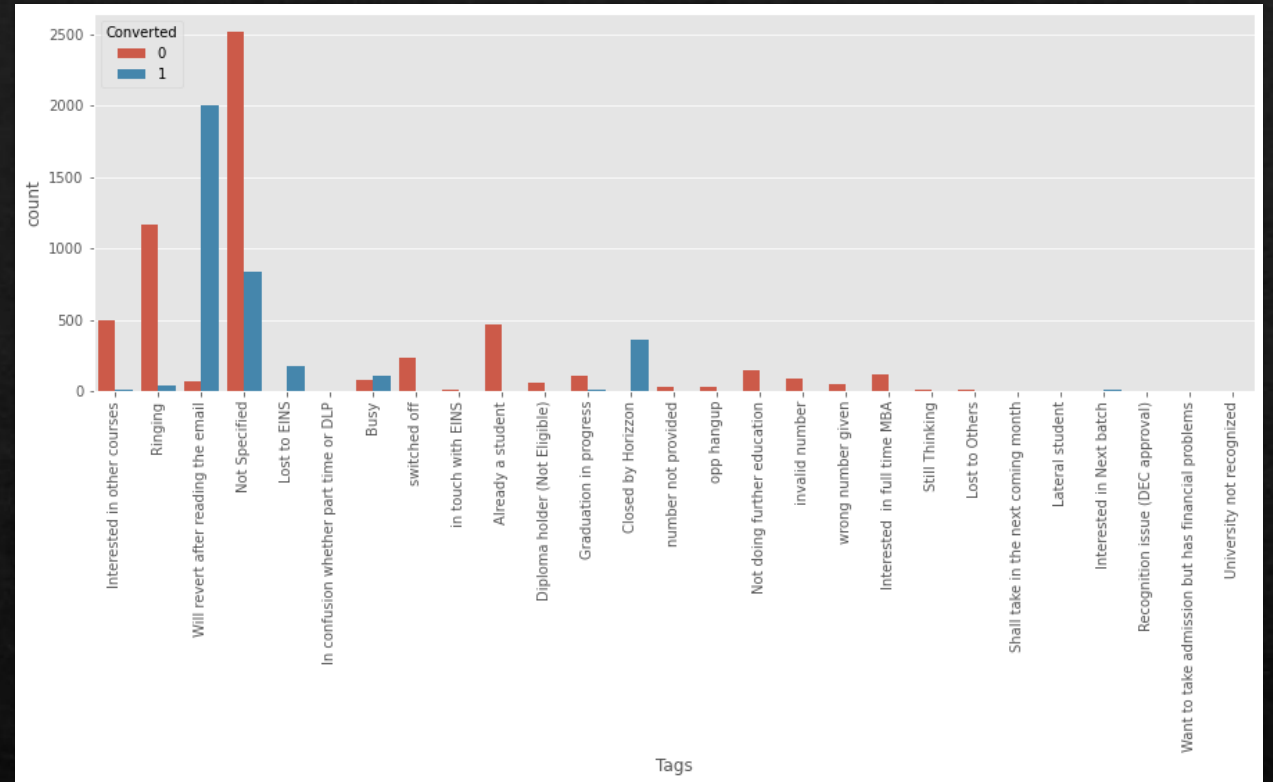
- Maximum number of leads are generated by Google and Direct traffic.
- Conversion Rate of reference leads and leads through welingak website is high.
- To improve overall lead conversion rate, focus should be on improving lead conversion of olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingak website.

Current Occupation



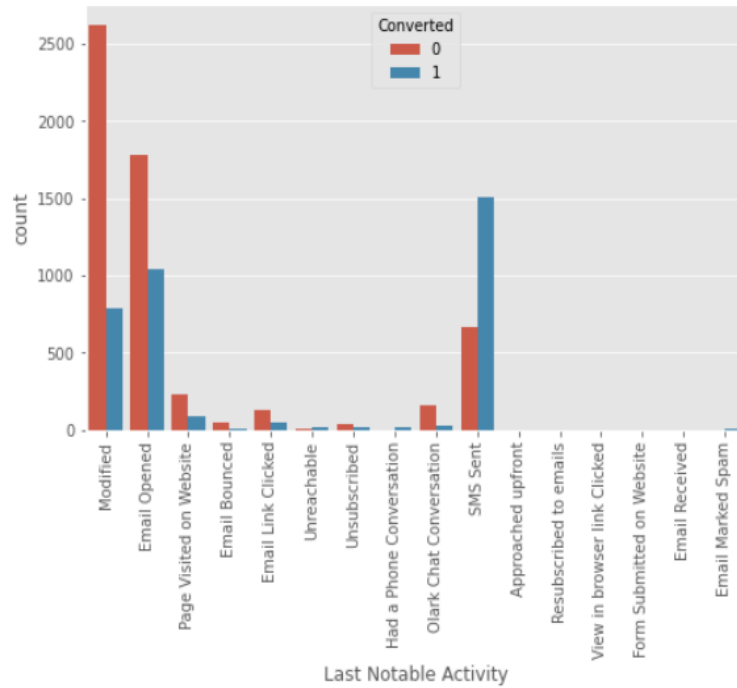
- Working Professionals going for the course have high chances of joining it.
- Unemployed leads are the most in terms of Absolute numbers.

Tags



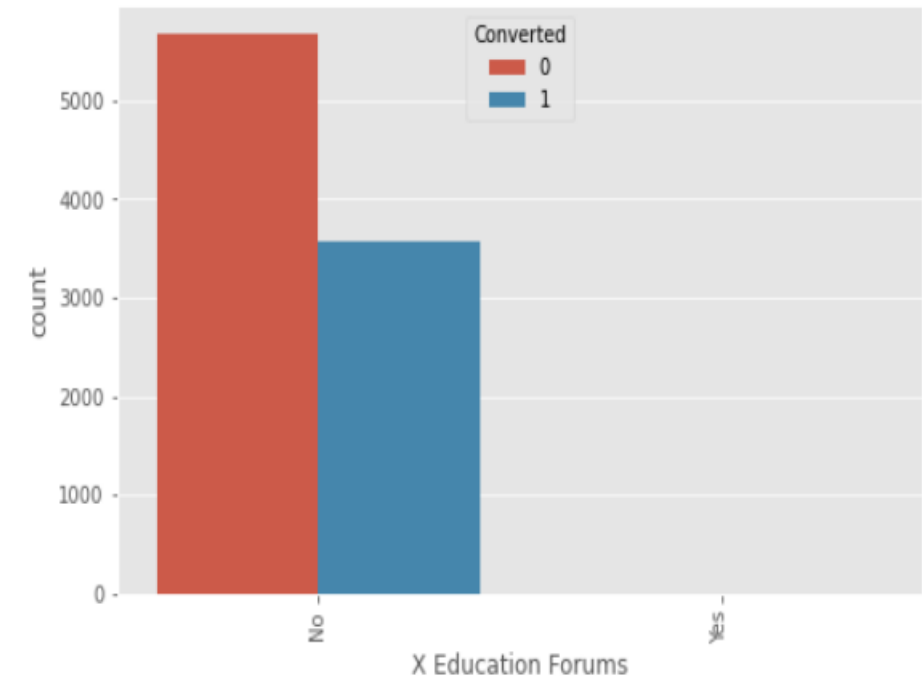
Tags -We will revert after reading email has max conversion

Last Notable Activity



- Highest conversion rate is in SMS sent activity. Keeping track of this one would be beneficial

X Education Forums



X Education Forums has only Non Converted , Not a significant column

Heatmap

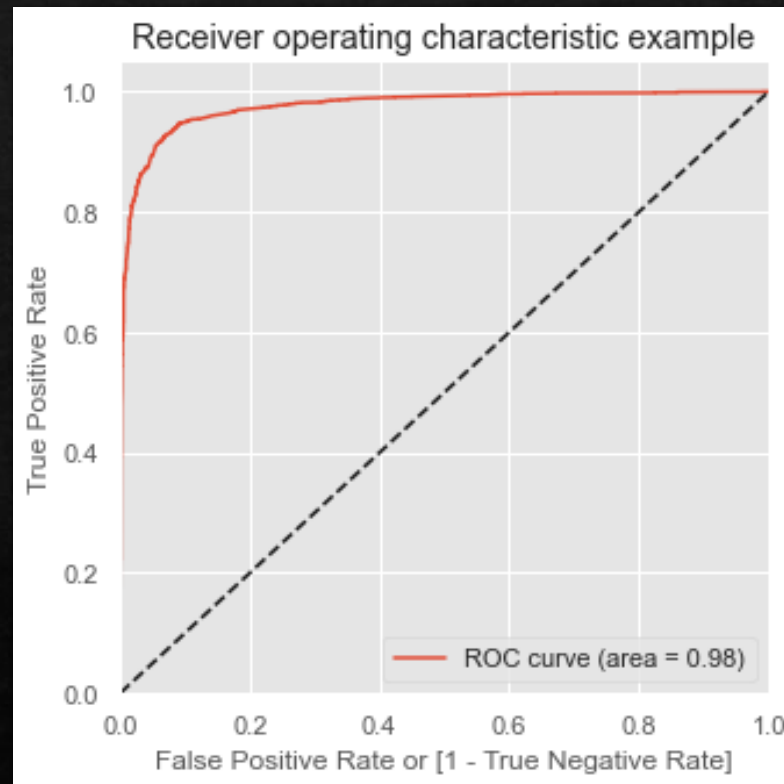
- Page Views Per Visit and Total Visits are correlated



Factors that Impact Lead Conversion -Model Evaluation

```
Total Time Spent on Website  
Lead Origin_Landing Page Submission  
Specialization_Specialization_Not Specified  
Lead Source_Olark Chat  
Lead Source_Others  
Lead Source_Welingak Website  
Last Activity_Email Bounced  
Last Activity_SMS Sent  
Last Notable Activity_Modified  
Last Notable Activity_Olark Chat Conversation  
Tags_Closed by Horizzon  
Tags_Interested in other courses  
Tags_Lost to EINS  
Tags_Other_Tags  
Tags_Ringing  
Tags_Will revert after reading the email
```


ROC curve

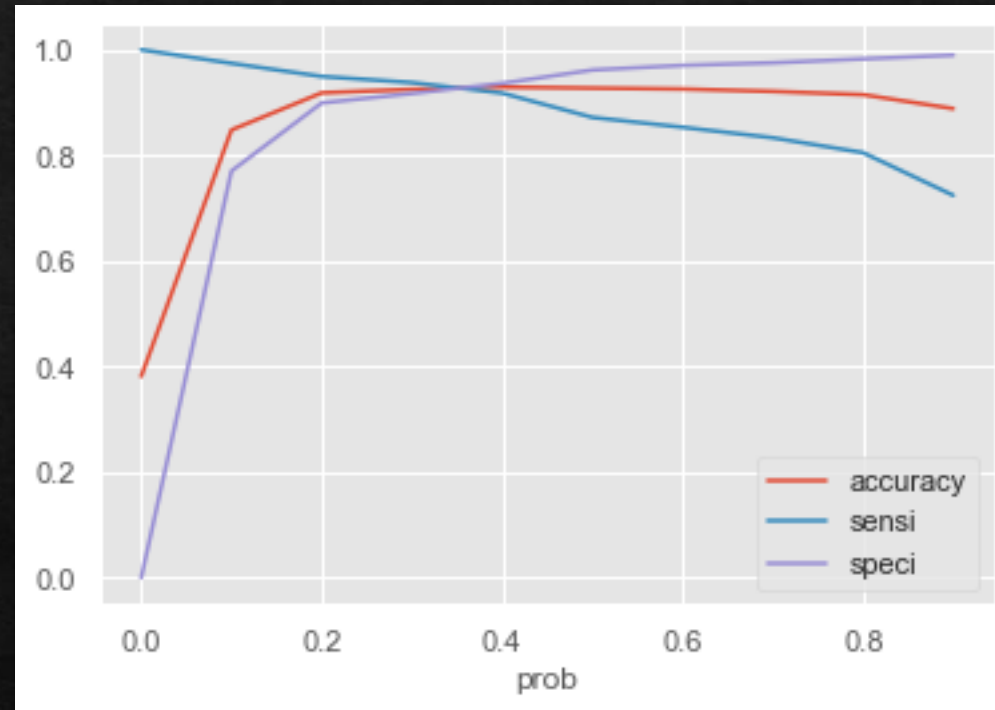


Area under curve is = 0.98

Finding optimal Threshold

A right choice of Probability at Which the Lead should be considered as potential and can turn into a hot lead is the major level

Hence using this model we achieve a right trade off point for having 0.38 i.e 38% probability of a Prospect is good enough to target as HOT lead and work more on the strategies of that lead rather than Cold ones



Graph showing changes in Sensitivity, Specificity and Accuracy with changes in the probability threshold values Optimal cutoff = 0.38

Lead Prediction on the basis of probability set by the model

Final_pred column shows the Conversion Probability of a Prospect Lead

Lead_Score above 38 shows a potential lead and has the capability to turn into HOT lead

Lead_Score 38 mathematically specifies that 38% probability is good to go for targeting the lead in order to achieve 80% Conversion

	Prospect ID	Converted	Converted_prob	Lead_Score	final_Predicted
0	4664	0	0.467809	47	1
1	5394	0	0.188680	19	0
2	1595	1	0.994051	99	1
3	2111	0	0.164561	16	0
4	9182	0	0.048924	5	0

Final Results

DATA	TRAIN SET	TEST SET
Accuracy	0.962794	0.939091
Sensitivity	0.857300	0.9212007
Specificity	0.944900	0.950699
False positive Rate	0.082885	0.164561
Positive prediction Value	0.8746615	0.048924
Precision	0.8746615	0.923800
Recall	0.9377851	0.921200


Feature Importance

- ◆ Three variables which contribute most towards the probability of a lead conversion in decreasing order of impact are:
 - ◆ Tags_Lost to EINS II
 - ◆ Tags_Closed by Horizzon
 - ◆ Tags_Will revert after reading the email
- ◆ These are dummy features created from the categorical variable Tags.
- ◆ All three contribute positively towards the probability of a lead conversion.

These results indicate that the company should focus more on the leads with these three tags

Recommendations

- ◆ By referring to the data visualizations, focus on – Increasing the conversion rates for the categories generating more leads and – Generating more leads for categories having high conversion rates.
- ◆ Pay attention to the relative importance of the features in the model and their positive or negative impact on the probability of conversion.
- ◆ Based on varying business needs, modify the probability threshold value for identifying potential leads.



Thank you

Shital Dhuwadhapare

Preeti Kumari

Ketan Kumar Dorik