Submitted by:

- Gaurav Gupta
- Illakiya
- Farheena Begum

CONTENTS

PROBLEM STATEMENT

X-Education is an education company sells online Education courses to professionals and marketing through online advertisements. Company gets the information through various different channels and if candidates enquiring with certain education level it calls it as a Lead. Typically lead conversion is 30% of certain education. Company identifying Hot Leads on certain criteria also. Lead conversion ratio is lesser than number of enrollment. The company has given a Target to achieve 80% of total enrollment.

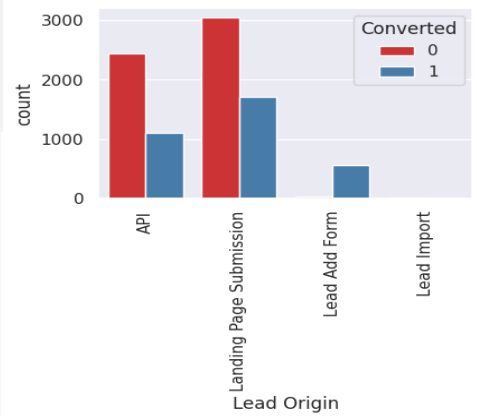
Goals of Case Study

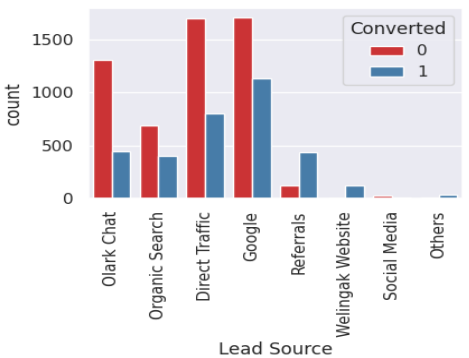
- Building Logistic regression model to finding leads for X-Education and help to achieve potential targets.
- Alternative approach to adjust future requirement, model should be flexible and ready in case requirement changes.

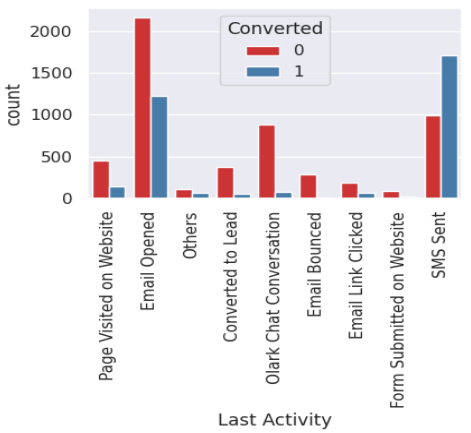
PROBLEM APPROACH

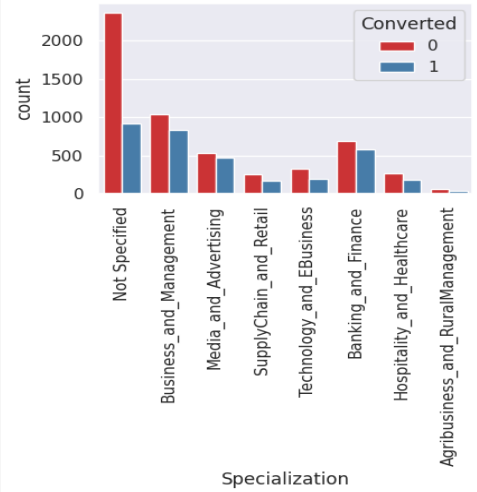
- Importing the data and inspecting the data frame
- Data preparation and cleaning
- EDA using various plots
- Dummy variable creation
- Test-Train split
- Feature scaling using StandardScalar
- Correlations using Heatmaps
- Model Building (RFE, VIF and p-values)
- Model Evaluation (finding optimal threshold cut-off, precision and recall tradeoff)
- Making predictions on test set (finding best parameters)

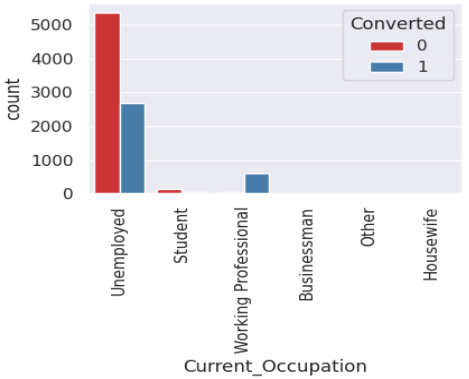
EDA

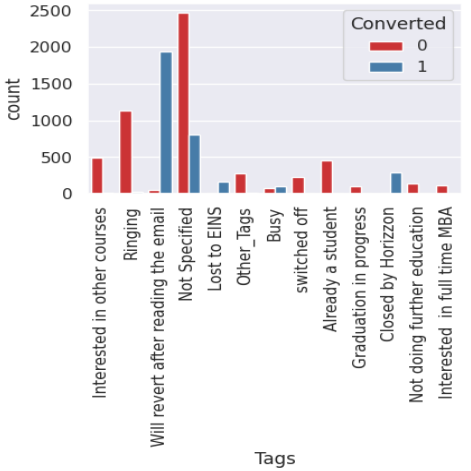

Converted
0
1
count
3000
2000
1000
0
API
Landing Page Submission
Lead Add Form
Lead Import
Lead Origin

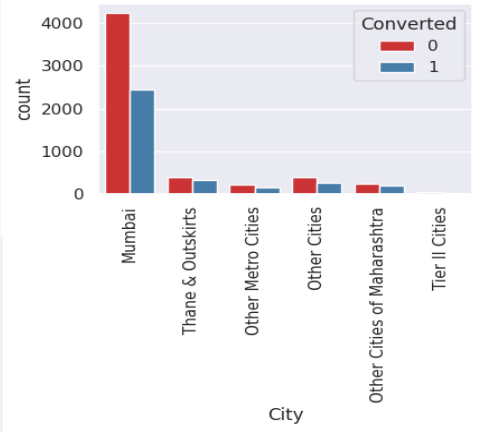

Converted
0
1
count
1500
1000
500
0
Olark Chat
Organic Search
Direct Traffic
Google
Referrals
Welingak Website
Social Media
Others
Lead Source

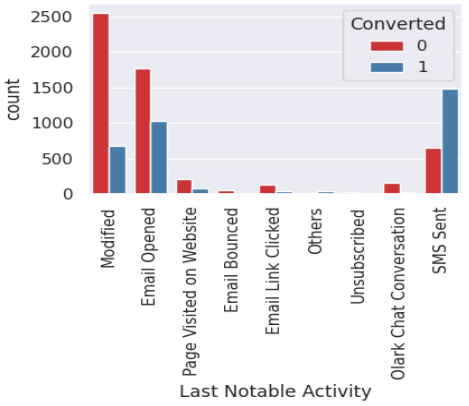

Converted
0
1
count
2000
1500
1000
500
0
Page Visited on Website
Email Opened
Others
Converted to Lead
Olark Chat Conversation
Email Bounced
Email Link Clicked
Form Submitted on Website
SMS Sent
Last Activity


Converted
0
1
count
2000
1500
1000
500
0
Not Specified
Business_and_Management
Media_and_Advertising
SupplyChain_and_Retail
Technology_and_EBusiness
Banking_and_Finance
Hospitality_and_Healthcare
Agribusiness_and_RuralManagement
Specialization

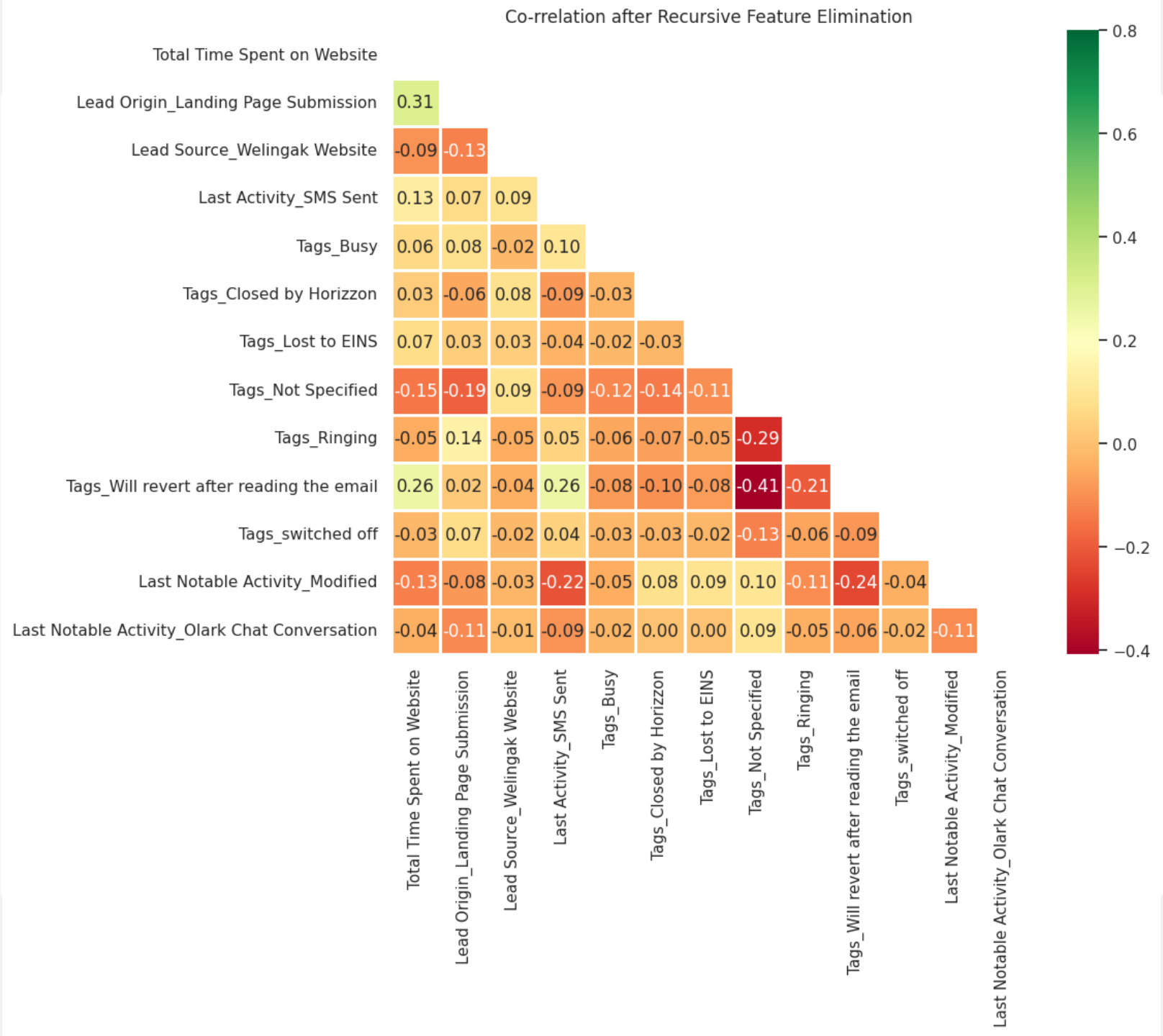

Converted
0
1
count
5000
4000
3000
2000
1000
0
Unemployed
Student
Working Professional
Businessman
Other
Housewife
Current_Occupation


Converted
0
1
count
2500
2000
1500
1000
500
0
Interested in other courses
Ringing
Will revert after reading the email
Not Specified
Lost to EINS
Other_Tags
Busy
switched off
Already a student
Graduation in progress
Closed by Horizzon
Not doing further education
Interested in full time MBA
Tags


Converted
0
1
count
4000
3000
2000
1000
0
Mumbai
Thane & Outskirts
Other Metro Cities
Other Cities
Other Cities of Maharashtra
Tier II Cities
City


Converted
0
1
count
2500
2000
1500
1000
500
0
Modified
Email Opened
Page Visited on Website
Email Bounced
Email Link Clicked
Others
Unsubscribed
Olark Chat Conversation
SMS Sent
Last Notable Activity

Co-relation using Heatmap



TRAIN MODEL EVALUATION

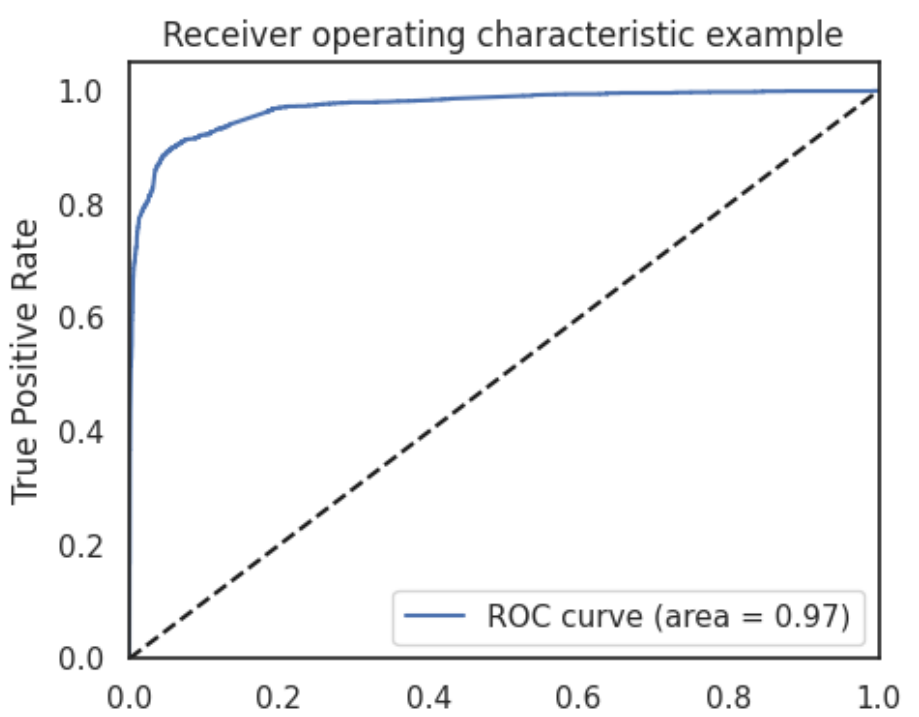
With the optimal cut-off threshold value of 0.29, we get the following results for our train model:

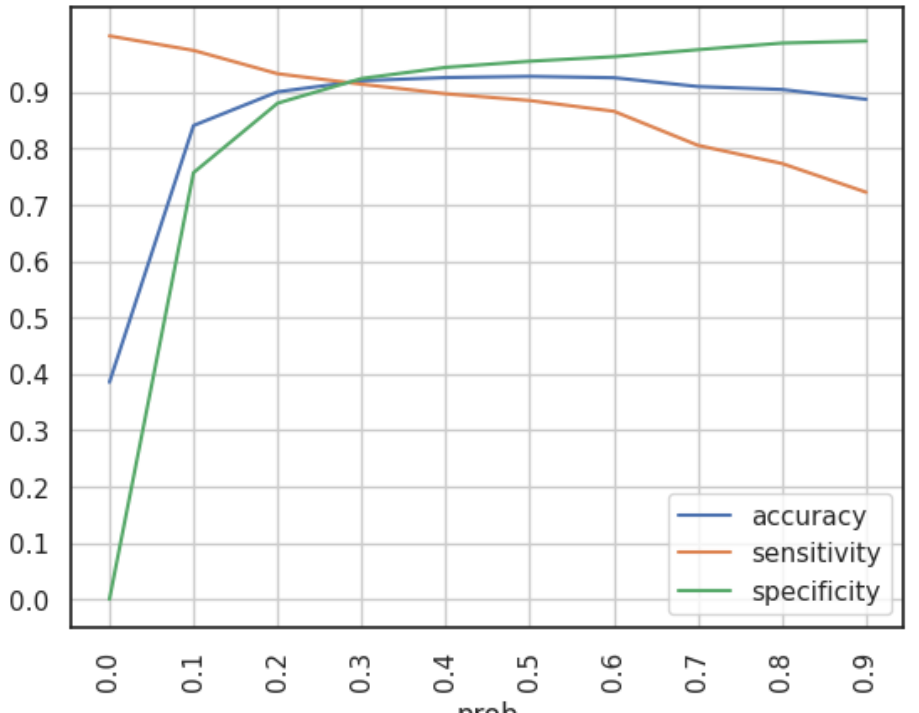
- Model Accuracy = 91.8621 %
- Model Sensitivity = 91.5114 %
- Model Specificity = 92.082 %
- Precision rate : 87.8728 %
- Recall rate : 91.5114 %
- F1-Score : 89.6552 %

MODEL EVALUATION (ROC-CURVE)

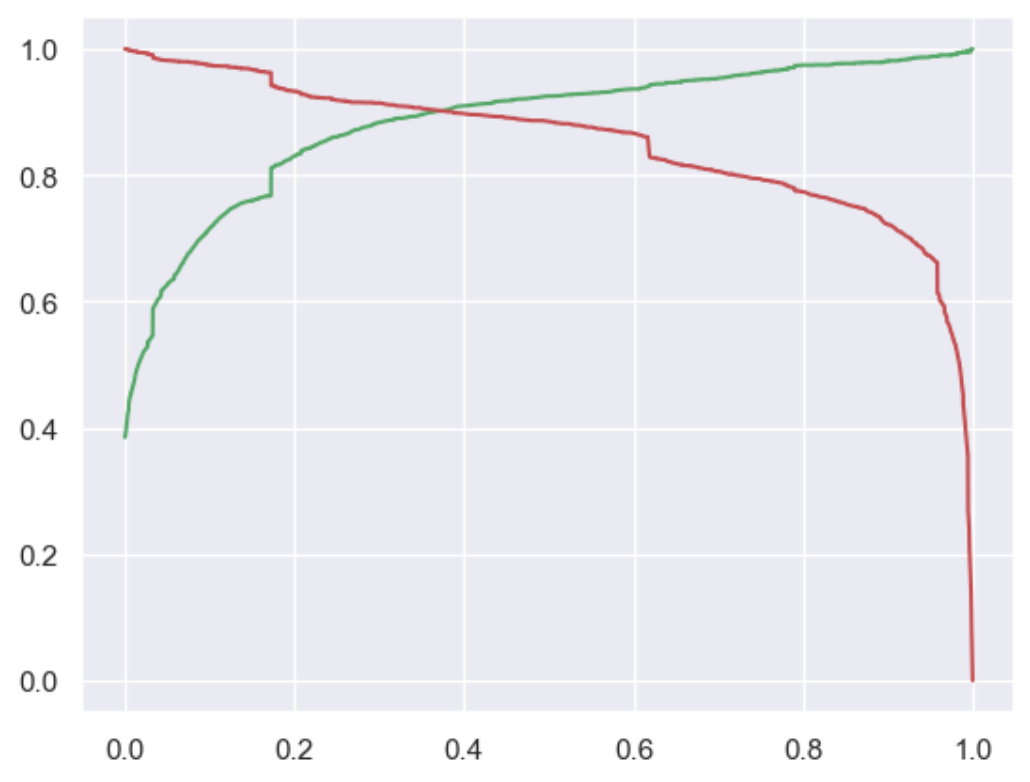
ROC Curve
area = 0.97

From the plot, the optimal cut-off threshold value is 0.29 (where accuracy, sensitivity and specificity meet)





MODEL EVALUATION (PRECISION AND RECALL TRADEOFF)



- At 0.38 we can see that the Precision and Recall value meets.
- Thus at this value we can choose to consider any Lead with higher than 42% probability to be a good Lead for Conversion

TEST MODEL EVALUATION

With the optimal cut-off threshold value of 0.29, we get the following results for our test model:

- Test Model Accuracy = 92.3678 %
- Test Model Sensitivity = 92.9592 %
- Test Model Specificity = 92.0281 %
- Test Precision rate : 87.0105 %
- Test Recall rate : 92.9592 %
- Test F1-Score : 89.8865 %

OBSERVATIONS

Customers with a Lead score of 80% or higher can be treated as “Good Leads” and can be approached for positive conversion. Based on Test Model we have identified 808 "Good Leads".

Top 3 features contributing most towards lead conversion:

Tags_Closed by Horizzon : 9.876544

Tags_Lost to EINS : 8.600111

Tags_Will revert after reading the email : 6.717868

CONCLUSION

To increase Lead Conversion Rates for X-Education:

- Create strategies for generating high-quality leads from top-performing lead sources.
- Optimize communication channels depending on lead engagement impact. In terms of advertising, more budget/spend may be allocated to Tags Closed by Horizzon and the Welingak Website.
- Incentives/discounts for providing references that convert to leads to promote more referrals.
- Working professionals can be aggressively targeted since they have a high conversion rate and can afford to pay higher costs.

Areas of improvement:

- Analyze negative coefficients in specialization offerings.
- Review landing page submission process for areas of improvement.