Lead Score Case Study

Submitted by :

Varun Kumar

Lead Score Case Study for X Education

Problem Statement :

- X Education sells online courses to industry professionals. We 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. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, We also gets leads through past referrals.
- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

Business Goal:

- X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- We needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Strategy

- **Data cleaning and data manipulation.**

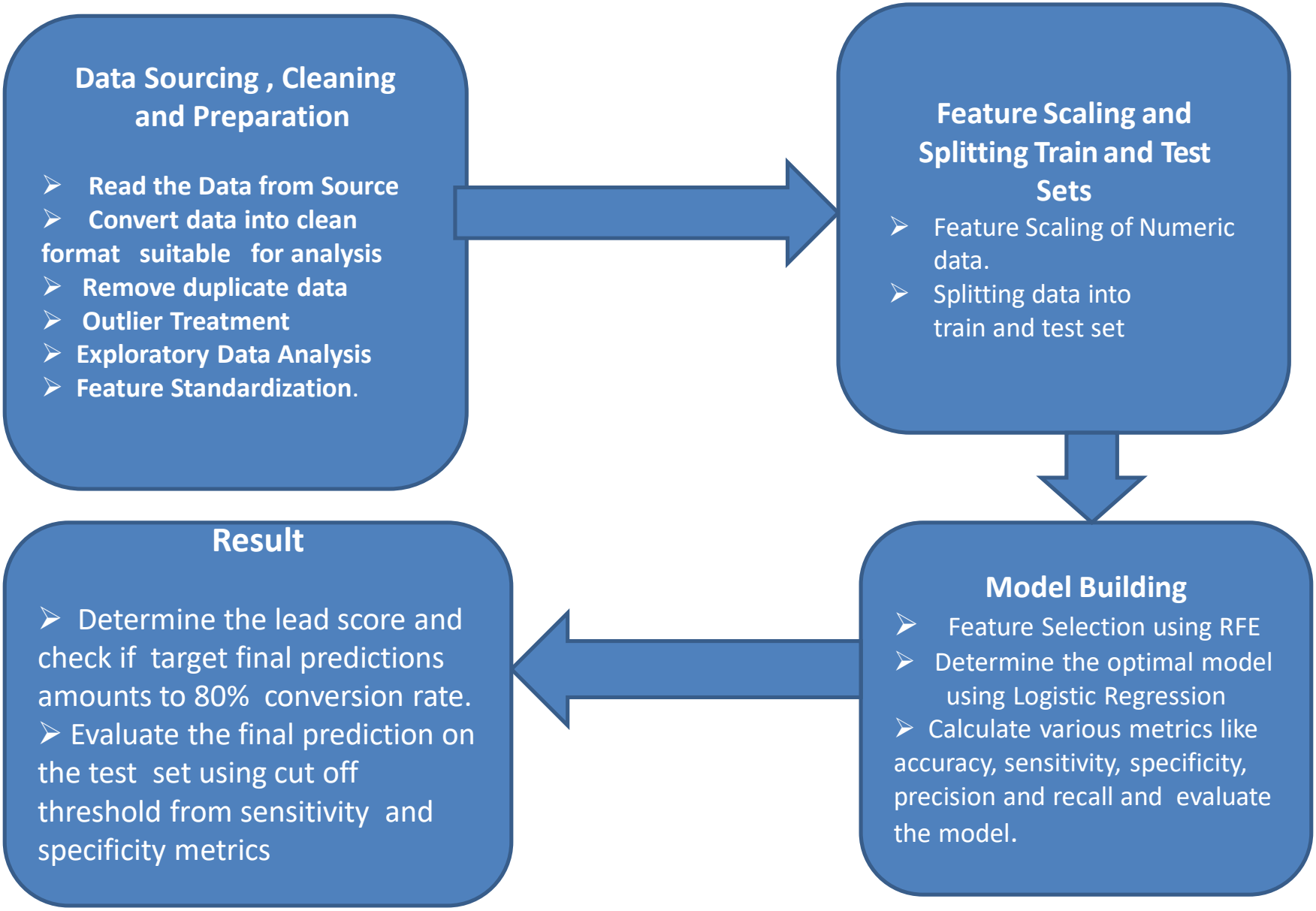
1. Check and handle duplicate data.
2. Check and handle NA values and missing values.
3. Drop columns, if it contains large amount of missing values and not useful for the analysis.
4. Imputation of the values, if necessary.
5. Check and handle outliers in data.

- **EDA**

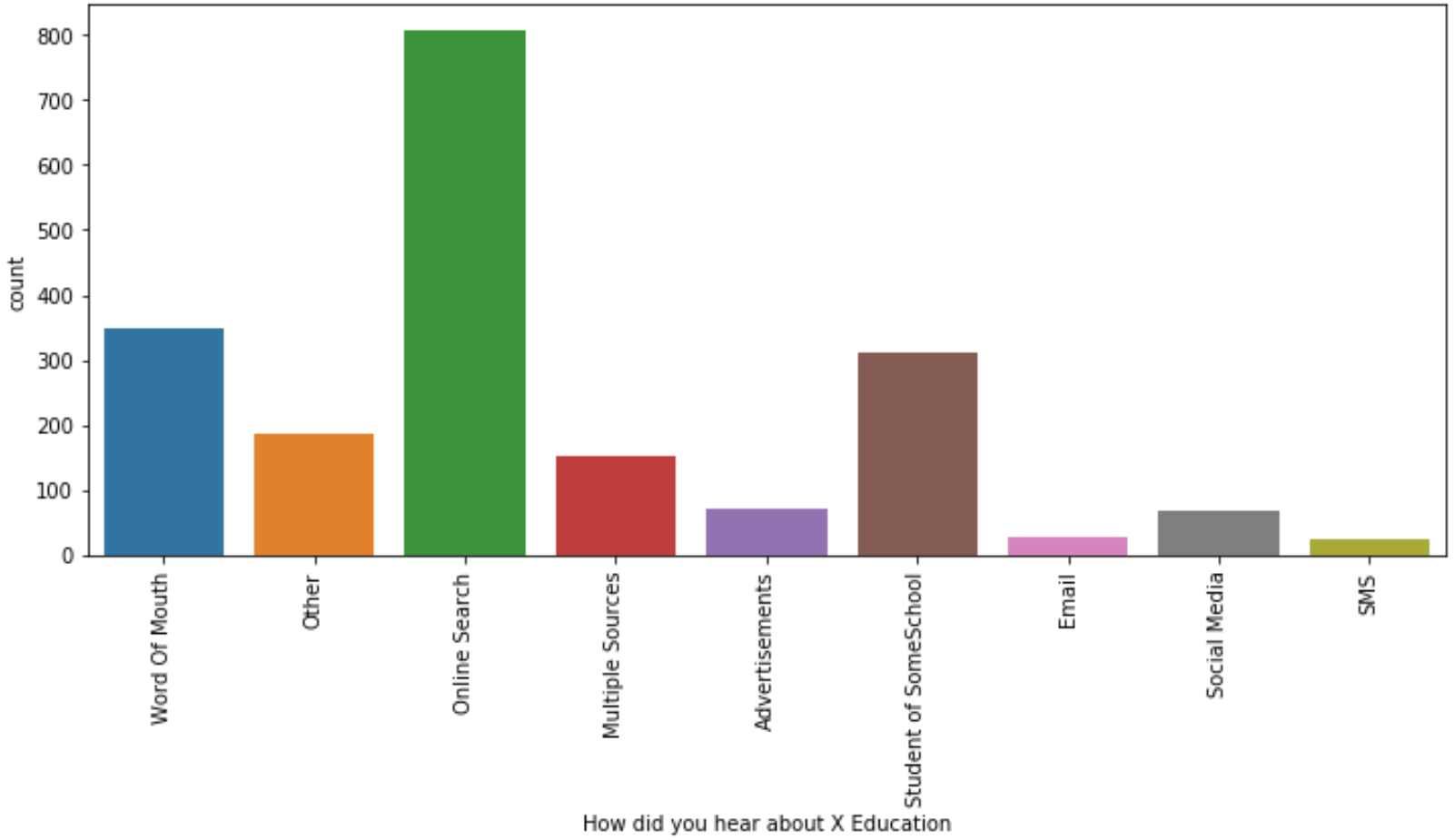
1. Univariate data analysis: value count, distribution of variable etc.
2. Bivariate data analysis: correlation coefficients and pattern between the variables etc.

- Feature Scaling & Dummy Variables and encoding of the data.
- Classification technique: logistic regression used for the model making and prediction.
- Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall
- Validation of the model.
- Model presentation.
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.
- Conclusions

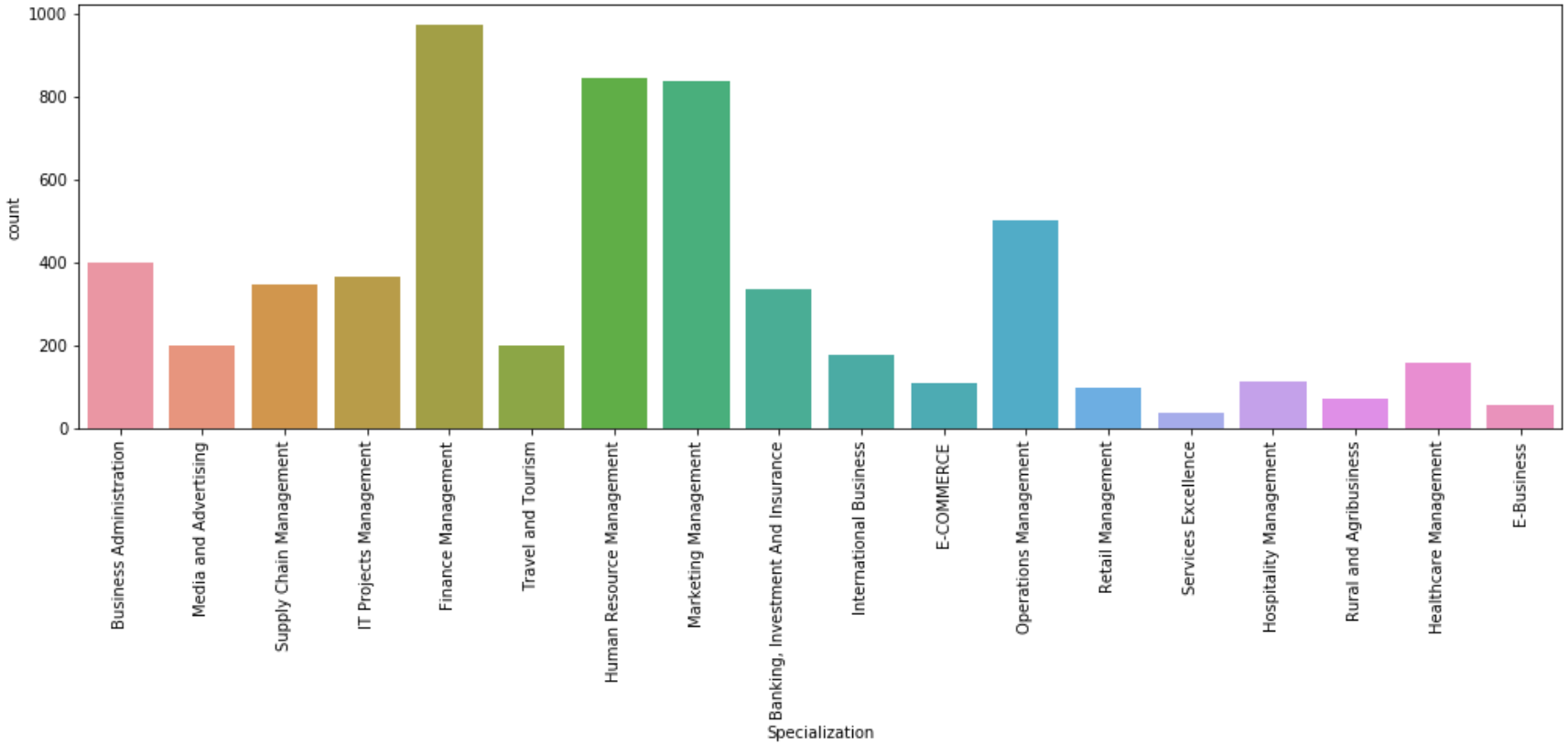
Problem solving methodology



How did you hear about X Education

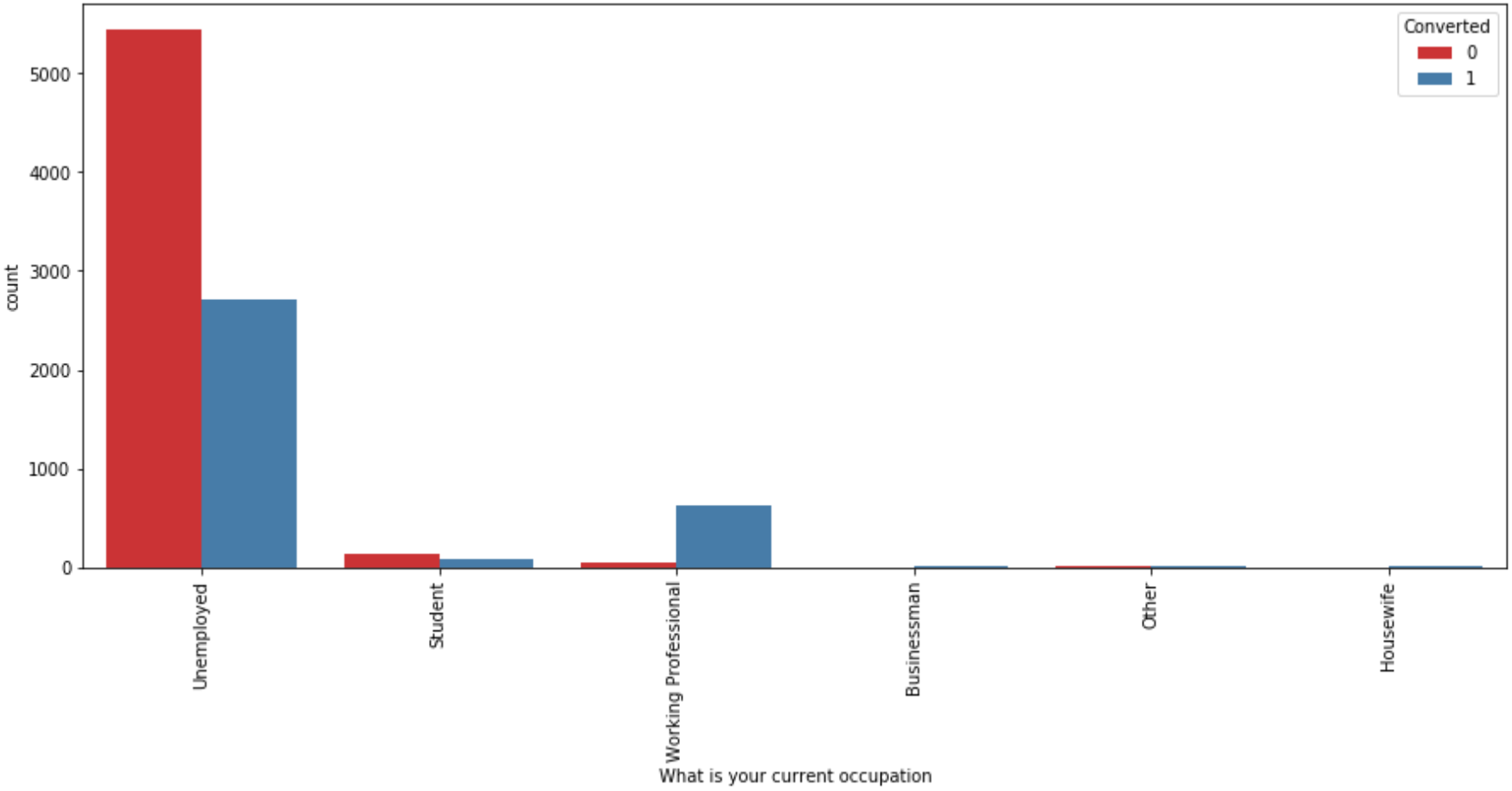


Specialization

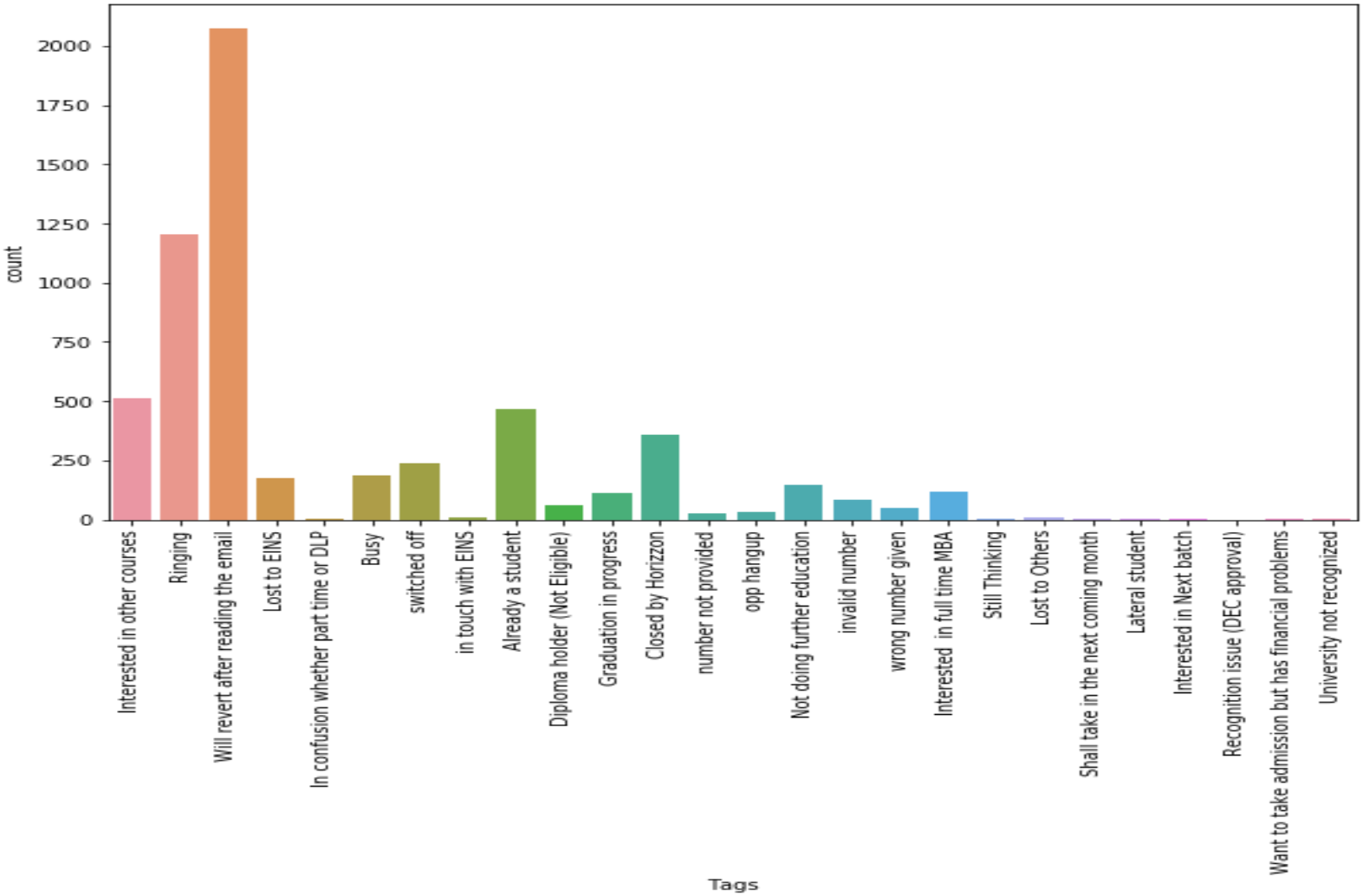


What is your current occupation

Unemployed leads are the most in numbers but has around 30-35% conversion rate.

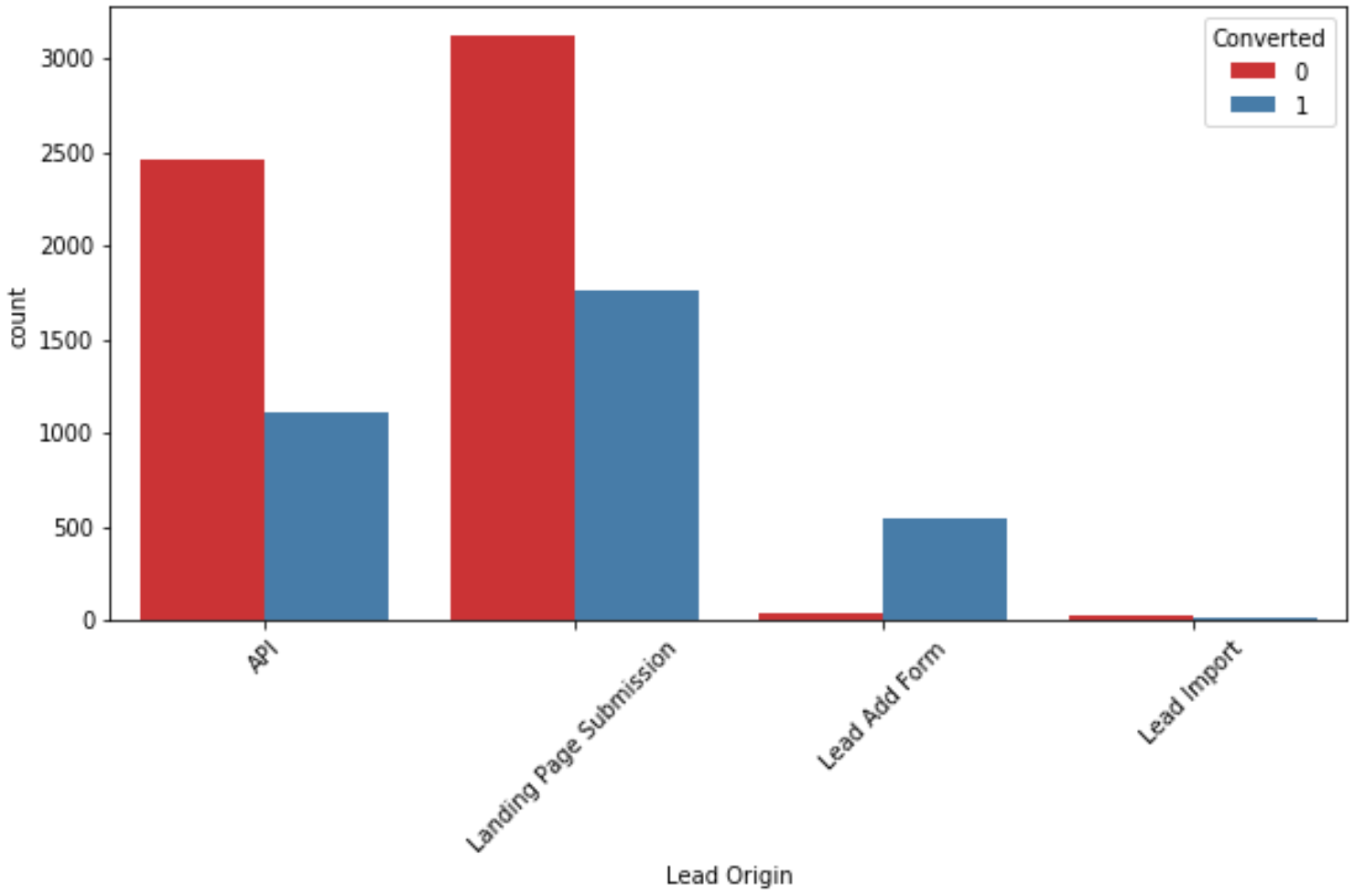


Tags


count
2000
1750
1500
1250
1000
750
500
250
0
Interested in other courses
Ringing
Will revert after reading the email
Lost to EINS
In confusion whether part time or DLP
Busy
switched off
in touch with EINS
Already a student
Diploma holder (Not Eligible)
Graduation in progress
Closed by Horizzon
number not provided
opp hangup
Not doing further education
invalid number
wrong number given
Interested in full time MBA
Still Thinking
Lost to Others
Shall take in the next coming month
Lateral student
Interested in Next batch
Recognition issue (DEC approval)
Want to take admission but has financial problems
University not recognized
Tags

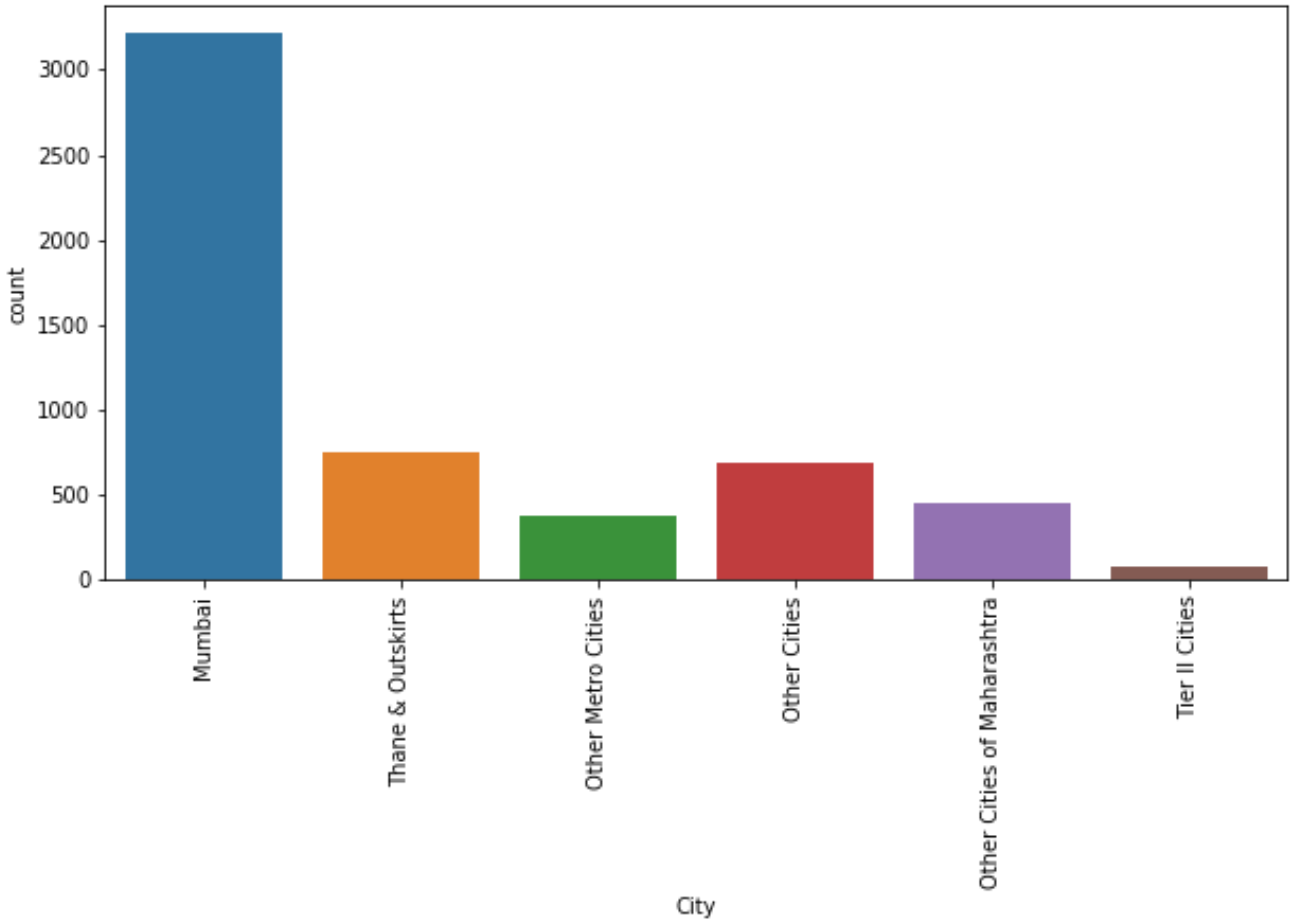
Lead Origin

In Lead Origin, maximum conversion happened from Landing Page Submission



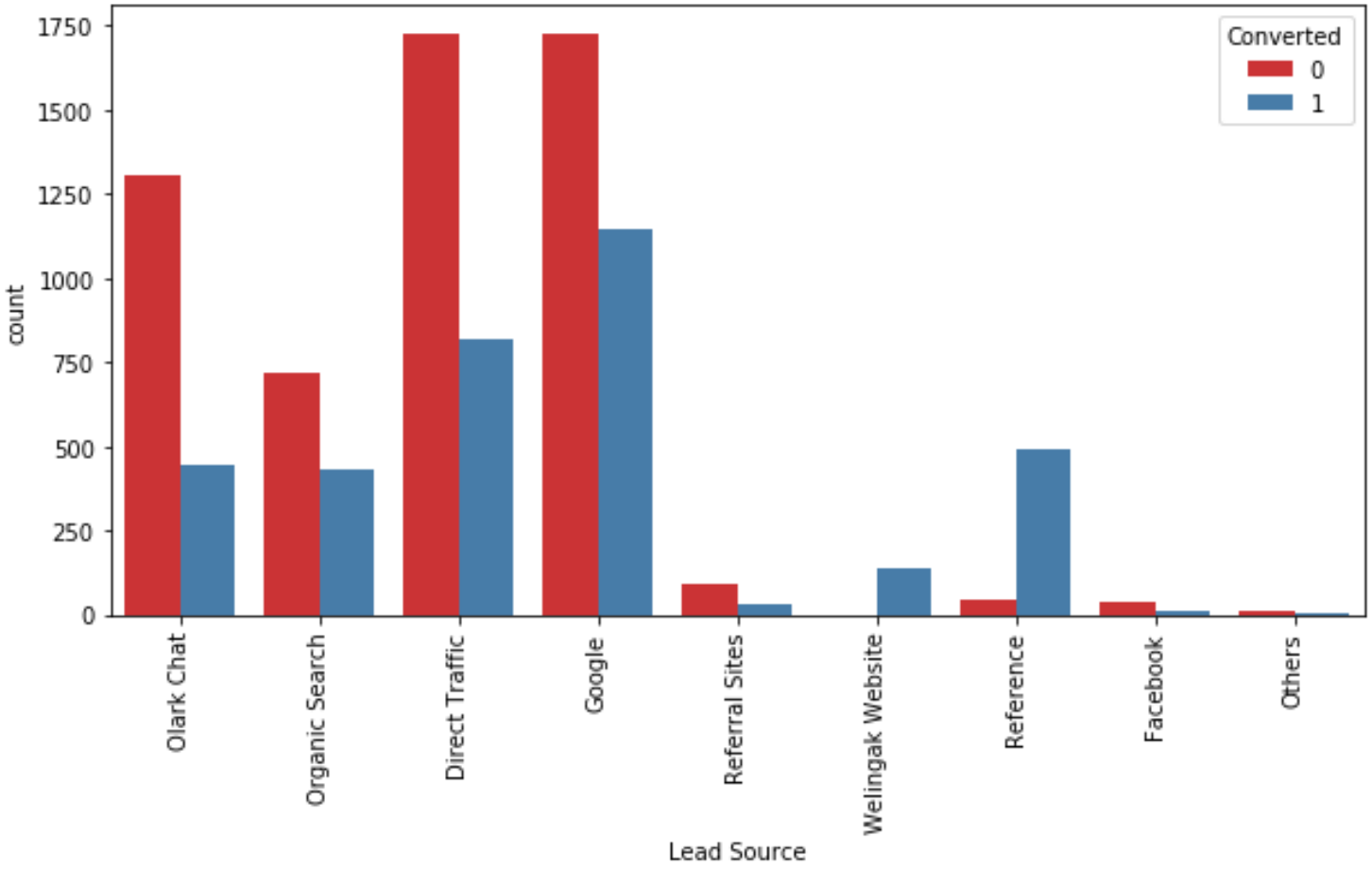
City

Major conversion has happened from Mumbai.



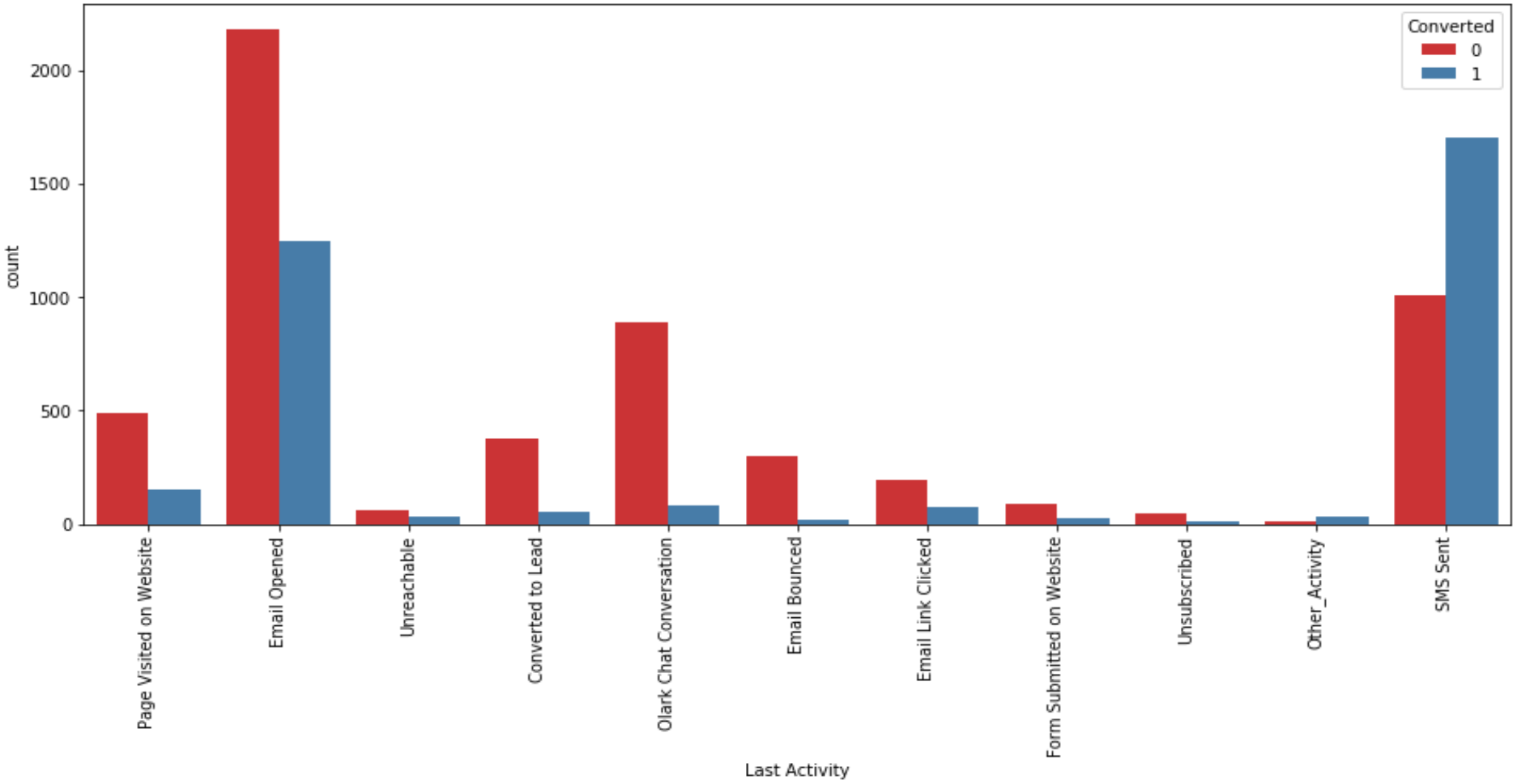
Lead Source

Google and Direct traffic generates maximum number of leads.



Last Activity

Last Activity value of SMS Sent' had more conversion.



Data Conversion

- Numerical Variables are Normalized
- Dummy Variables are created for object type variables
- Total Rows for Analysis: 9074
- Total Columns for Analysis: 69

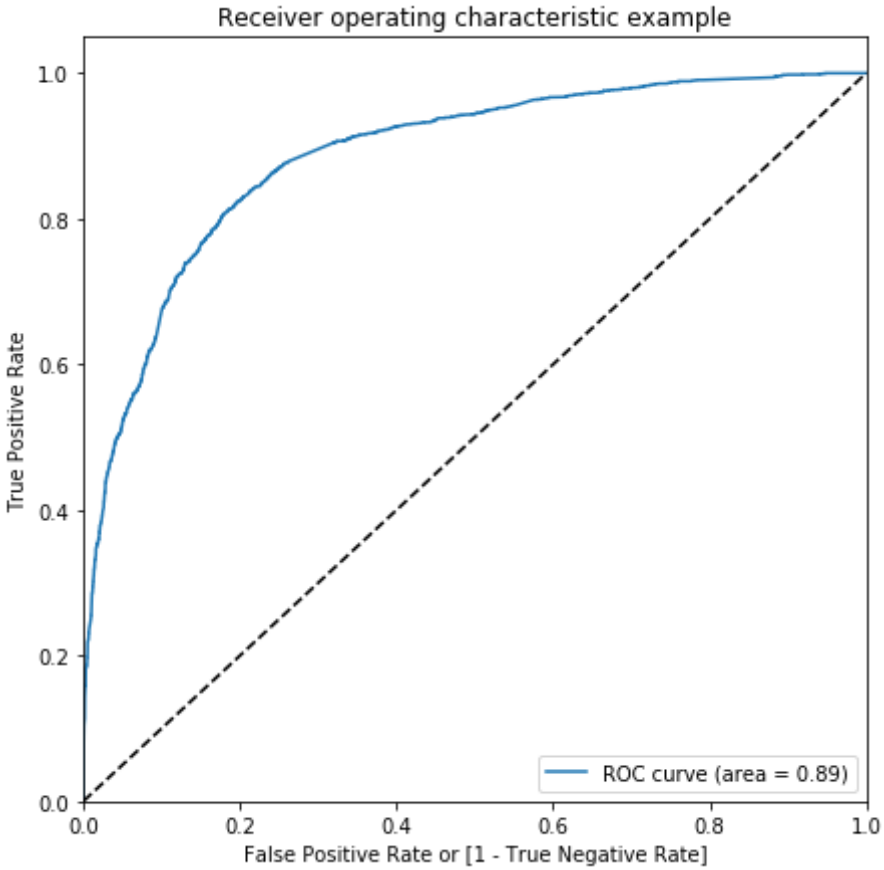
Variables Impacting the Conversion Rate

- Do Not Email
- Total Visits
- Total Time Spent On Website
- Lead Origin – Lead Page Submission
- Lead Origin – Lead Add Form
- Lead Source - Olark Chat
- Last Source – Welingak Website
- Last Activity – Email Bounced
- Last Activity – Not Sure
- Last Activity – Olark Chat Conversation
- Last Activity – SMS Sent
- Current Occupation – No Information
- Current Occupation – Working Professional
- Last Notable Activity – Had a Phone Conversation
- Last Notable Activity – Unreachable

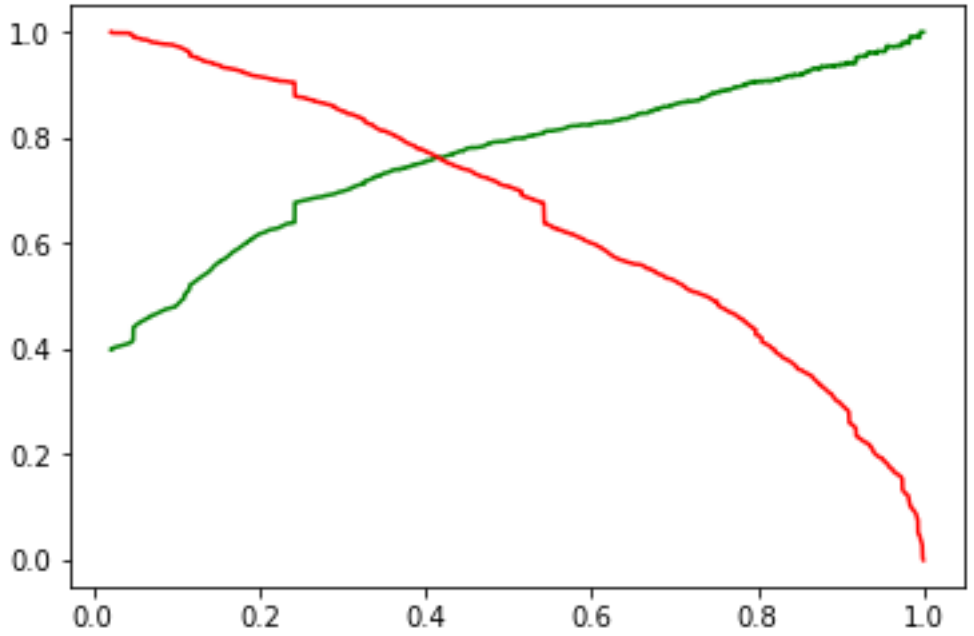
Model Building

- Splitting the Data into Training and Testing Sets
- The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- Use RFE for Feature Selection
- Running RFE with 15 variables as output
- Building Model by removing the variable whose p- value is greater than 0.05 and VIF value is greater than 5
- Predictions on test data set
- Overall accuracy 81%

ROC curve and Precision and Recall



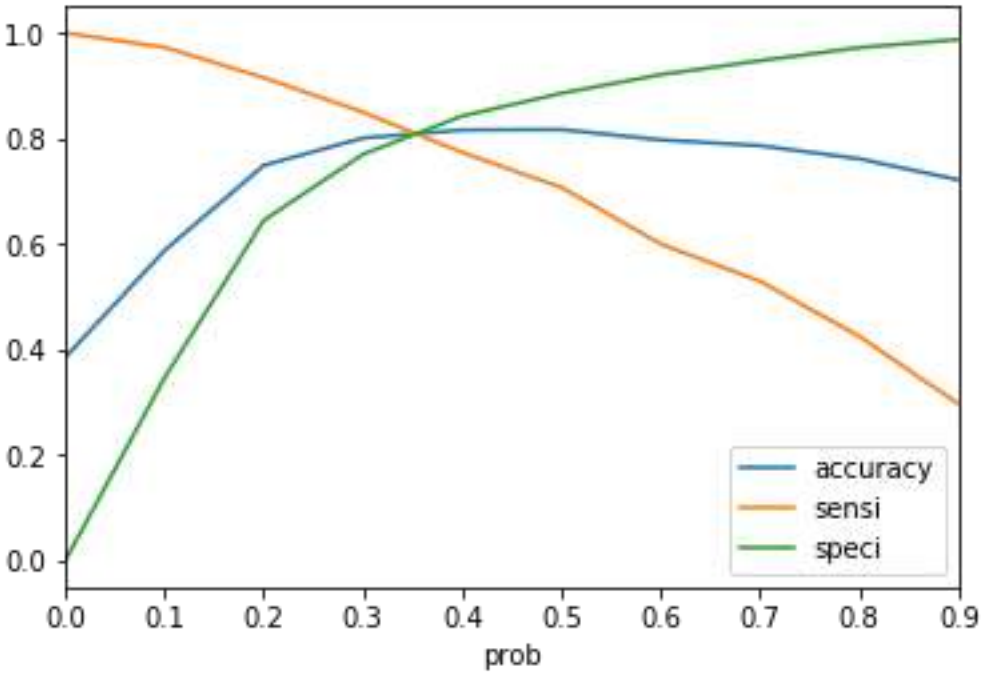
The graph depicts an optimal cut off of 0.42 based on Precision and Recall



- Finding optimal cut off point
- optimal cutt off probability is that where we get balanced sensitivity and specificity.

Model Evaluation - Sensitivity and Specificity on Train Data Set

The graph depicts an optimal cut off of 0.34 based on Accuracy, Sensitivity and Specificity



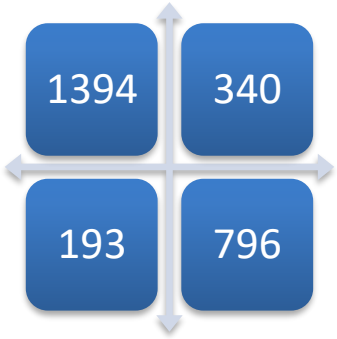
Confusion Matrix

3459	446
716	1730

- Accuracy - 81%
- Sensitivity - 70 %
- Specificity - 88 %
- False Positive Rate - 11 %
- Positive Predictive Value - 79 %
- Positive Predictive Value – 83 %

Model Evaluation – Sensitivity and Specificity on Test Dataset

Confusion Matrix



- Accuracy - 80 %
- Sensitivity - 80 %
- Specificity - 80 %

Conclusion

It was found that the variables that mattered the most in the potential buyers are :

- We should make calls to the leads coming from the lead sources "Welingak Websites" and "Reference" as these are more likely to get converted.
- Wes should make calls to the leads who are the "working professionals" as they are more likely to get converted.
- We should make calls to the leads who spent "more time on the websites" as these are more likely to get converted.
- We should make calls to the leads coming from the lead sources "Olark Chat" as these are more likely to get converted.
- We should make calls to the leads whose last activity was SMS Sent as they are more likely to get converted.
- We should not make calls to the leads whose last activity was "Olark Chat Conversation" as they are not likely to get converted.
- We should not make calls to the leads whose lead origin is "Landing Page Submission" as they are not likely to get converted.
- We should not make calls to the leads whose Specialization was "Others" as they are not likely to get converted.
- We should not make calls to the leads who chose the option of "Do not Email" as "yes" as they are not likely to get converted.

Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.