LEAD SCORE CASE STUDY

Maria Charistine
Nandhini N
Neha Mishra

PROBLEM STATEMENT

- X Education sells online courses to industry professionals.
- X Education gets a lots 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

BUSINESS OBJECTIVES

- X Education wants to know most promising leads.
- For that they want to build a model which identifies hot leads.
- Deployment of the model for future use

SOLUTION APPROACH

- Data Cleaning and data manipulation

1. Check and handle duplicate data
2. Check and handle null values and missing values.
3. Drop columns, if it contains large amount of missing values which are not useful for analysis.
4. Check and handle outliers in data.

- EDA

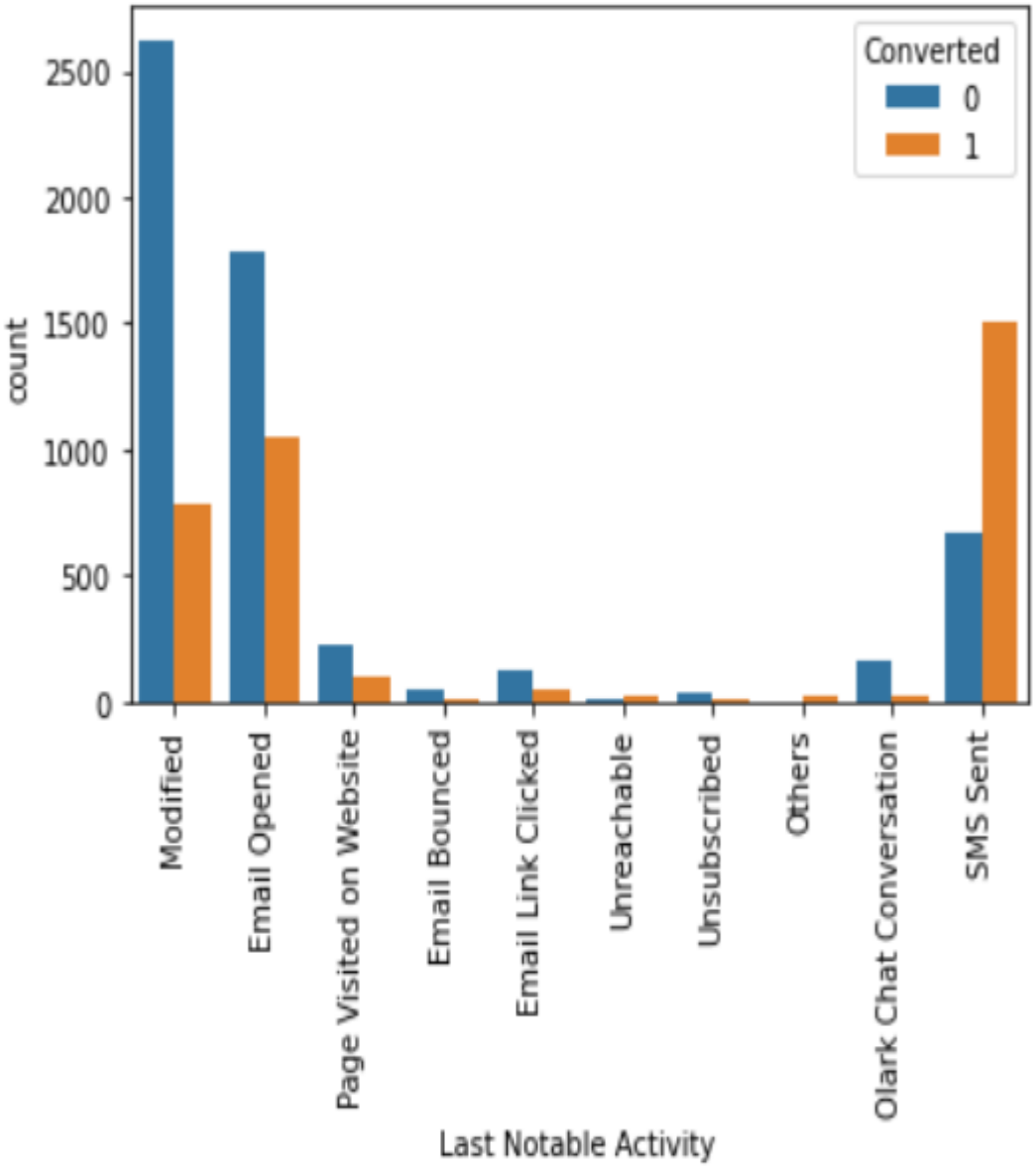
5. Univariate data analysis : value counts , distribution of variables etc.
6. Bivariate data analysis : correlation coefficients and pattern between the variables etc.

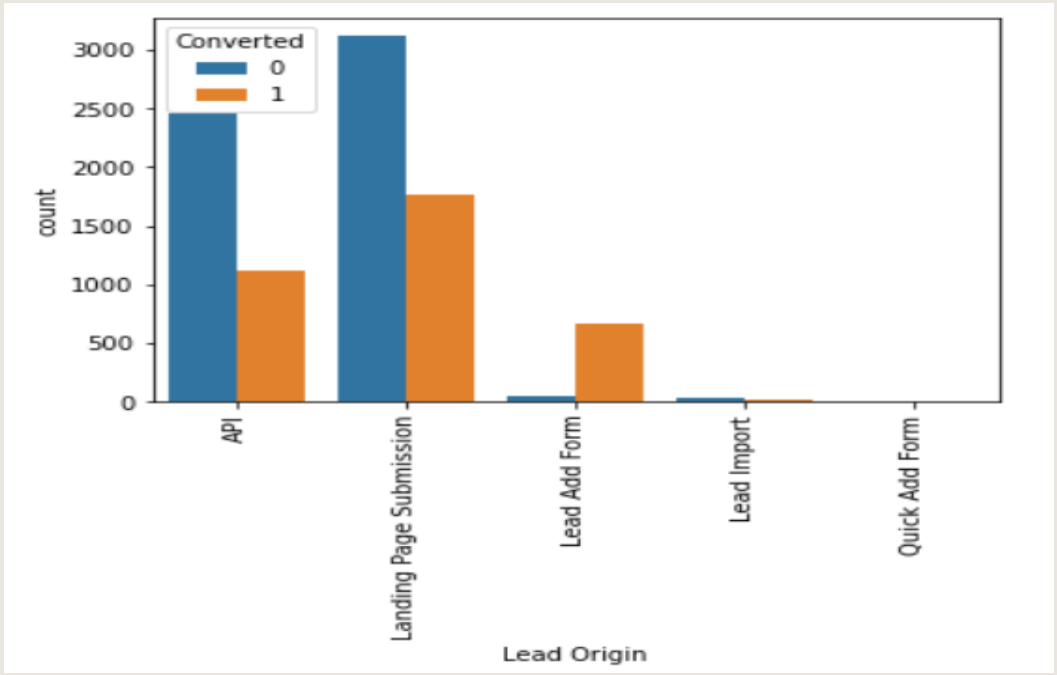
- Feature scaling & Dummy variable and encoding of the data.
- Classification technique: Logistic regression used for the model making and prediction.
- Validation of model
- Model Presentation
- Conclusion and recommendation.

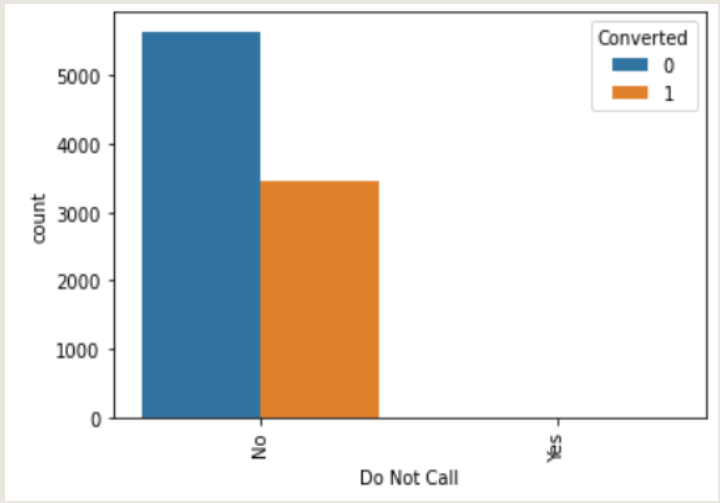
DATA MANIPULATION

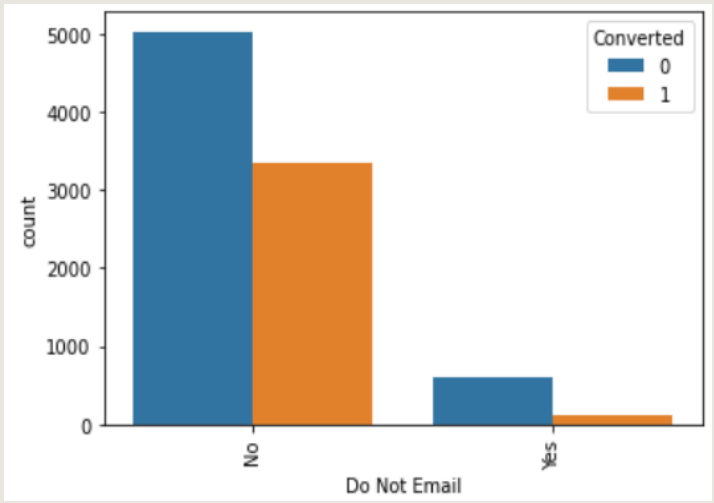
- Total number of rows = 9240 and total number of columns = 37
- Columns with more than 45% of null values has been dropped
- Dropped “Prospect ID ” and “Lead Number” which is not necessary for analysis.
- After checking for the value counts for some of the variables, we find some of the features which has no enough variance, which we have dropped, the features are:”Do not call”, ”What matters most to you in choosing course”, “Magazine”, “Search”, “Newspaper Article”, “Newspaper”, “X Education Forums”, “Digital Advertisement” etc.

EDA

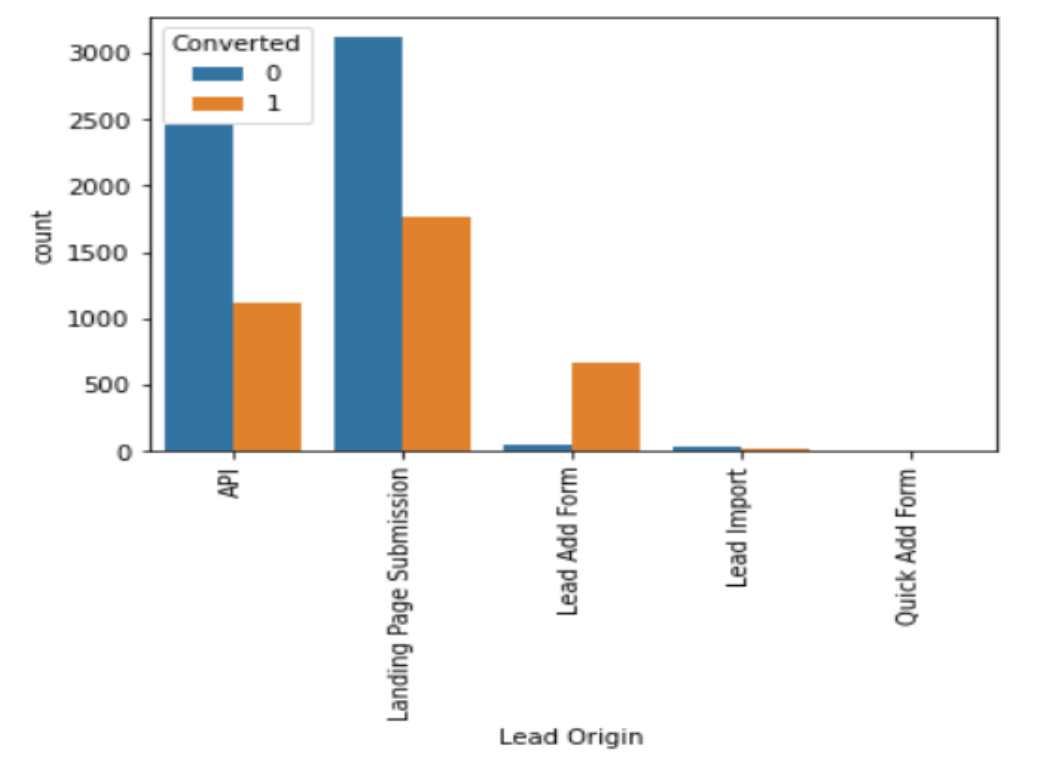

Converted
0
1
count
2500
2000
1500
1000
500
0
Modified
Email Opened
Page Visited on Website
Email Bounced
Email Link Clicked
Unreachable
Unsubscribed
Others
Olark Chat Conversation
SMS Sent
Last Notable Activity

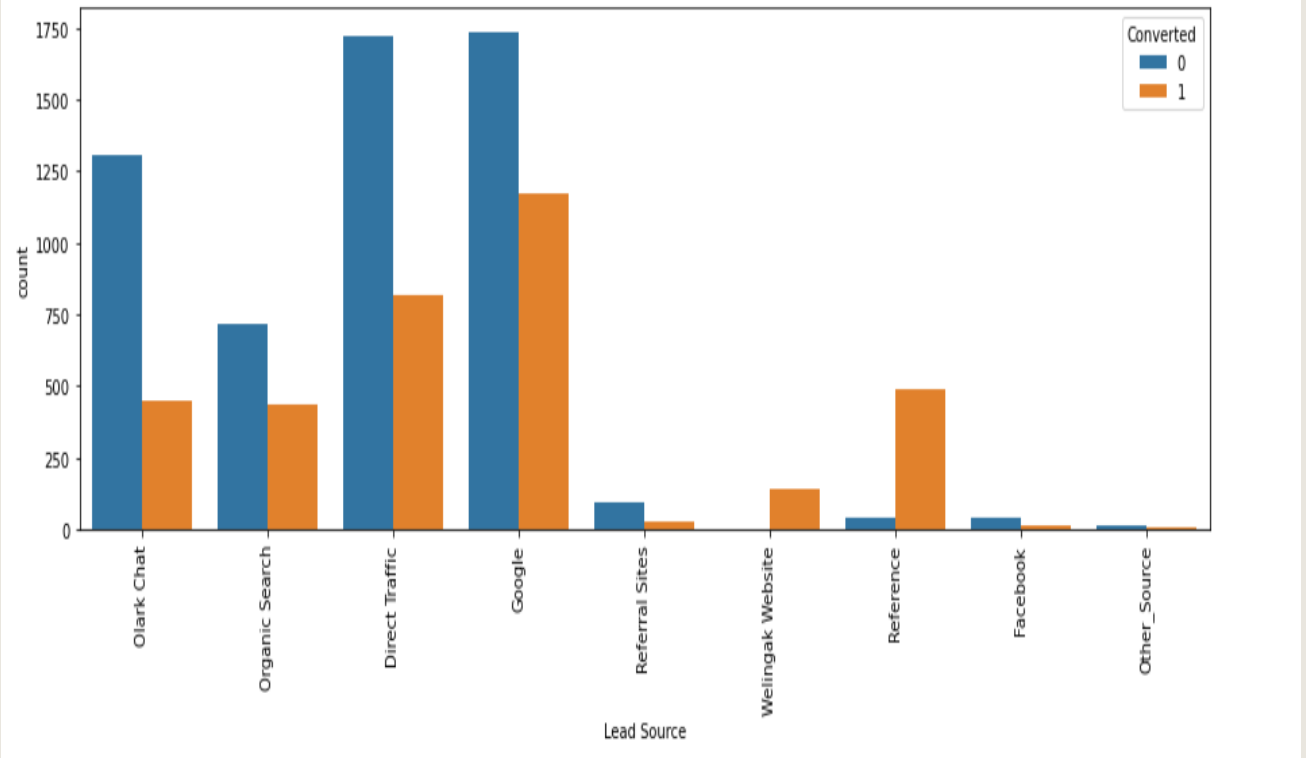

Converted
0
1
count
3000
2500
2000
1500
1000
500
0
API
Landing Page Submission
Lead Add Form
Lead Import
Quick Add Form
Lead Origin


Converted
0
1
count
5000
4000
3000
2000
1000
0
No
Yes
Do Not Call

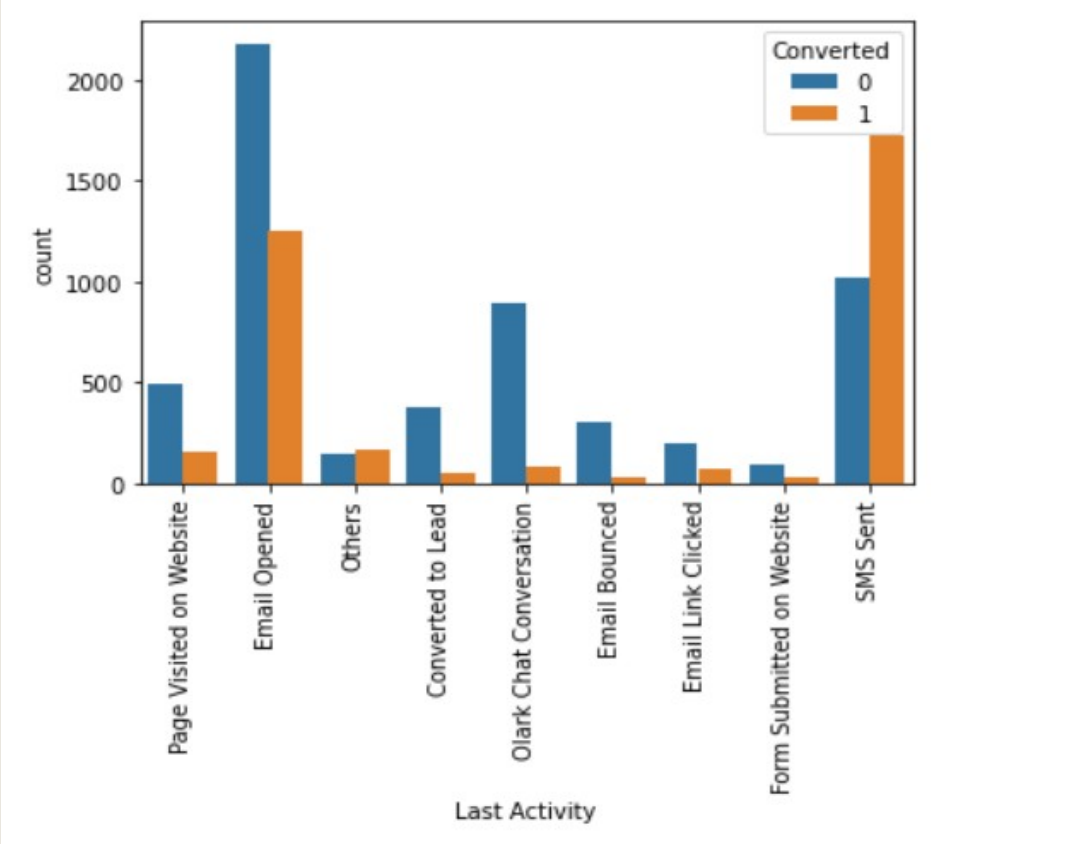

Converted
0
1
count
5000
4000
3000
2000
1000
0
No
Yes
Do Not Email

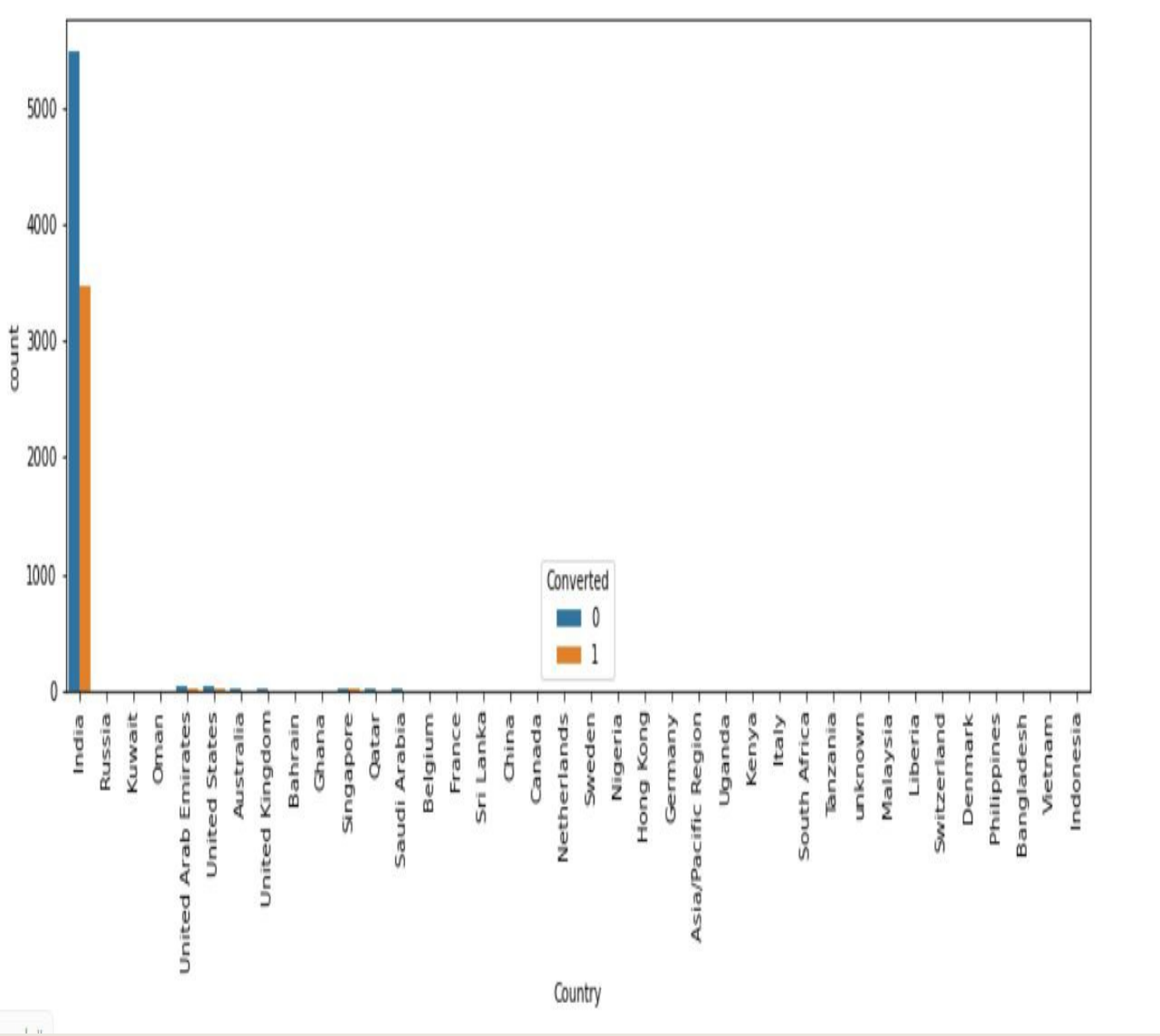
CATEGORICAL VARIABLE RELATION (1/2)





CATEGORICAL VARIABLE RELATION (2/2)





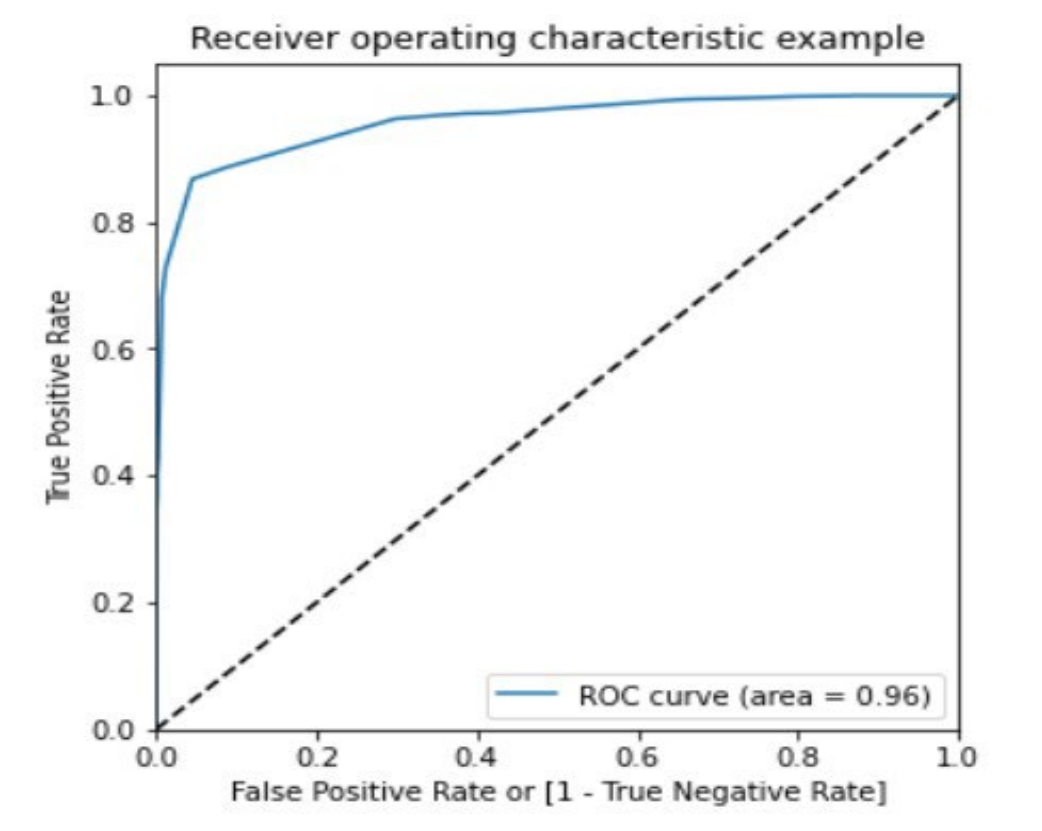
DATA CONVERSION

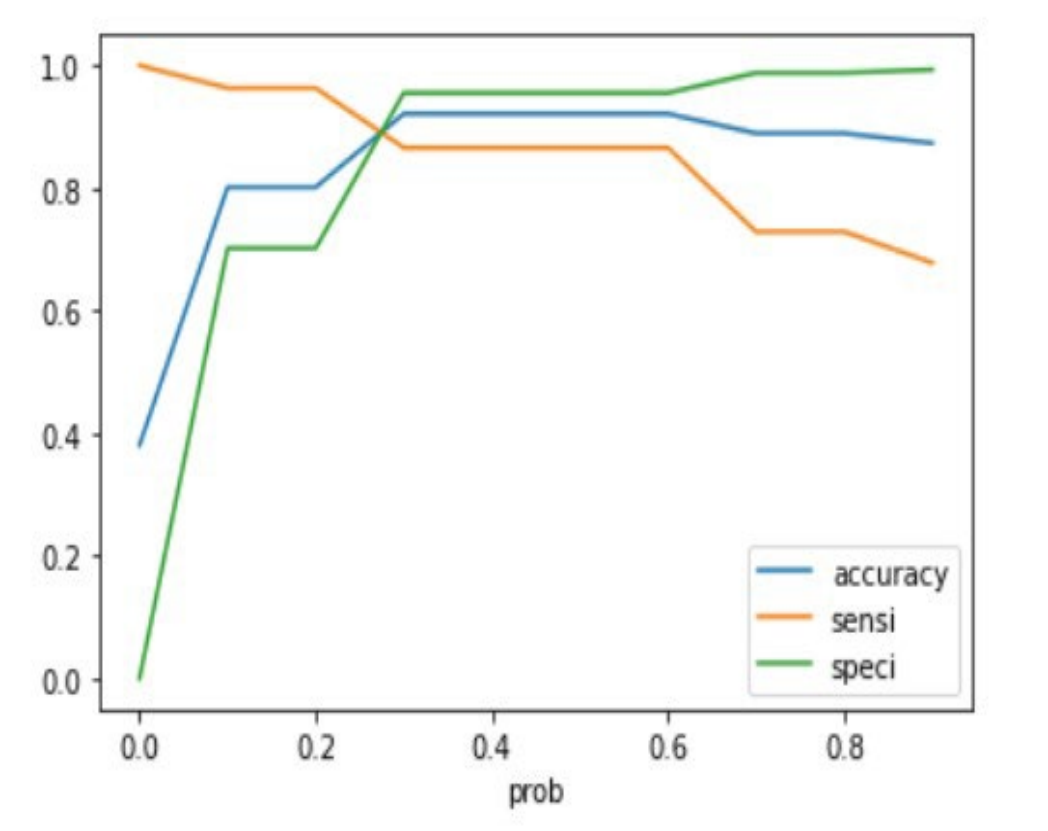
- Numerical Variables are normalized.
- Dummy variables are created for object type variables.

MODEL BUILDING

- Splitting the Data into Training and Testing Sets
- The first basic steps for regression is performing a Train-Test Split, we have chosen
  - 70:30 ratio.
- For feature selection we have used RFE.
- Running RFE with 15 variables as output.
- Building model by removing the variable whose p-value is greater than 0.05 and VIF is greater than 5.
- Prediction on Test data set.
- Overall Accuracy is 92%

ROC CURVE


Receiver operating characteristic example
True Positive Rate
False Positive Rate or [1 - True Negative Rate]
ROC curve (area = 0.96)
0.0
0.2
0.4
0.6
0.8
1.0


accuracy
sensi
speci
prob
0.0
0.2
0.4
0.6
0.8
1.0

CONCLUSION

It was found that the variables that mattered most in the potential buyers are (In descending order):

- Total time spent on website
- Total number of visits.
- When the lead source was:
  - Google
  - Direct Traffic
  - Organic Search
  - Welingak website
- When the last activity was:
  - SMS
  - Olark Chat Conversation
- When the lead origin is Lead Odd Format
- When their current occupation is as a working professional.

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.

THANK YOU