Lead Scoring Case Study

Presented by

-Vismaya Vt

-Vishwa Vijay Pratap Singh

-Vishwas Huligowda

Problem Statement :

- Industry professionals can purchase online courses from X Education.
- X Education has an extremely low lead conversion rate despite receiving a lot of leads. For instance, only roughly thirty of the 100 leads they obtain each day will actually be converted.
- The organization wants to find the most potential leads, or "Hot Leads," in order to streamline this procedure.
- The lead conversion rate should increase if they are able to locate this group of leads because the sales staff will now be spending more time corresponding with the prospects rather than calling everyone.

Bussiness Objective

- X schooling is interested in the most likely leads.
- In order to do this, they wish to create a model that pinpoints the hot leads.
- Deployment of the model for upcoming applications.

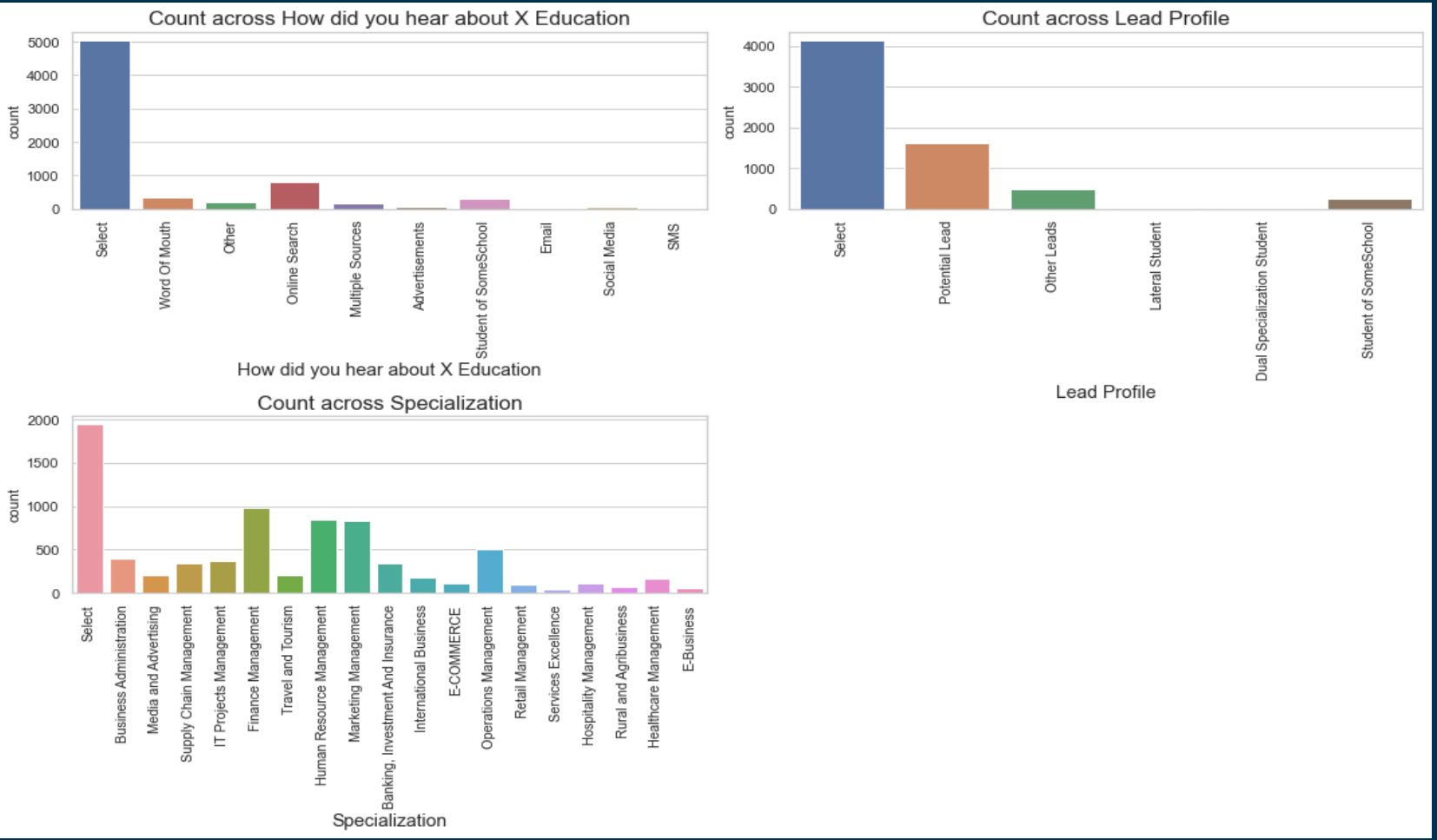
Solution Methodology

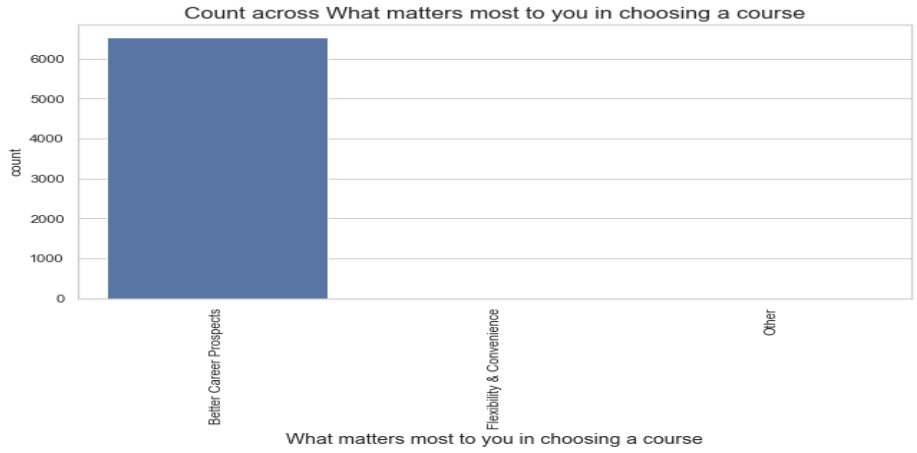
- Data cleaning and Preparation.
- Visualizing the features with select values
- Visualizing the features
- Analyzing categorical features
- Looking at the correlation
- Model Building
- Model Evaluation
- Making prediction on the test set

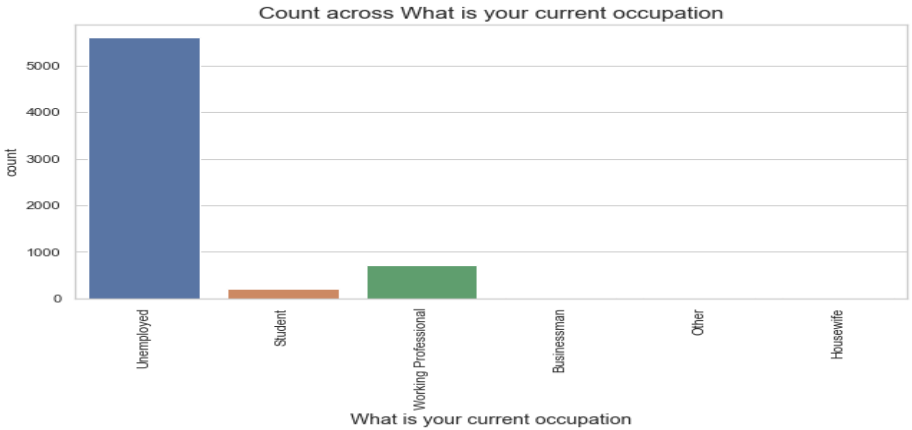
Data processing

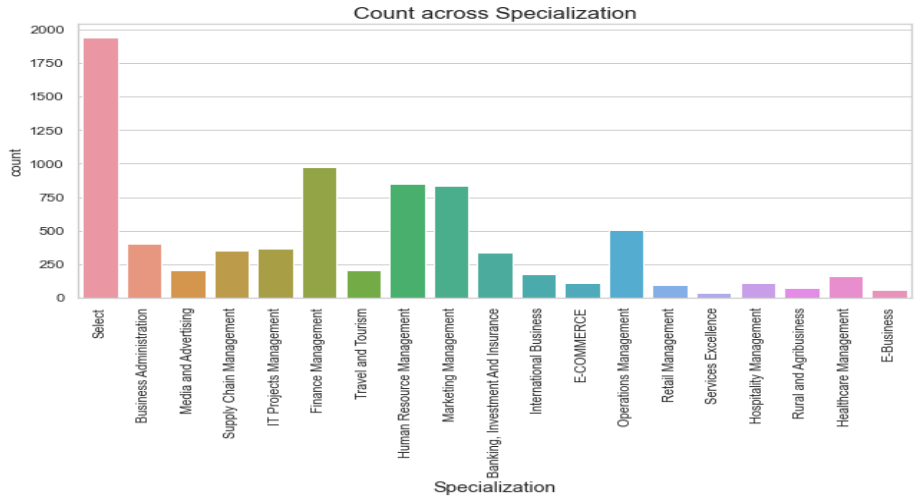
- There are 37 total rows and 9240 total columns.
- Features with a single value, such as "Magazine," "Get More Updates About Our Courses," and "Update me on Supply.“
- There are no longer any references to "Chain Content," "Get updates on DM Content," "I agree to pay the amount through cheque," etc
- Following a review of the value counts for a few object type variables, we identify a few features that have insufficient variation and have been removed. These features include "Search," "Newspaper Article," "X Education Forums," "Newspaper," "Digital Advertisement," "City," and "Do Not Call."

EDA

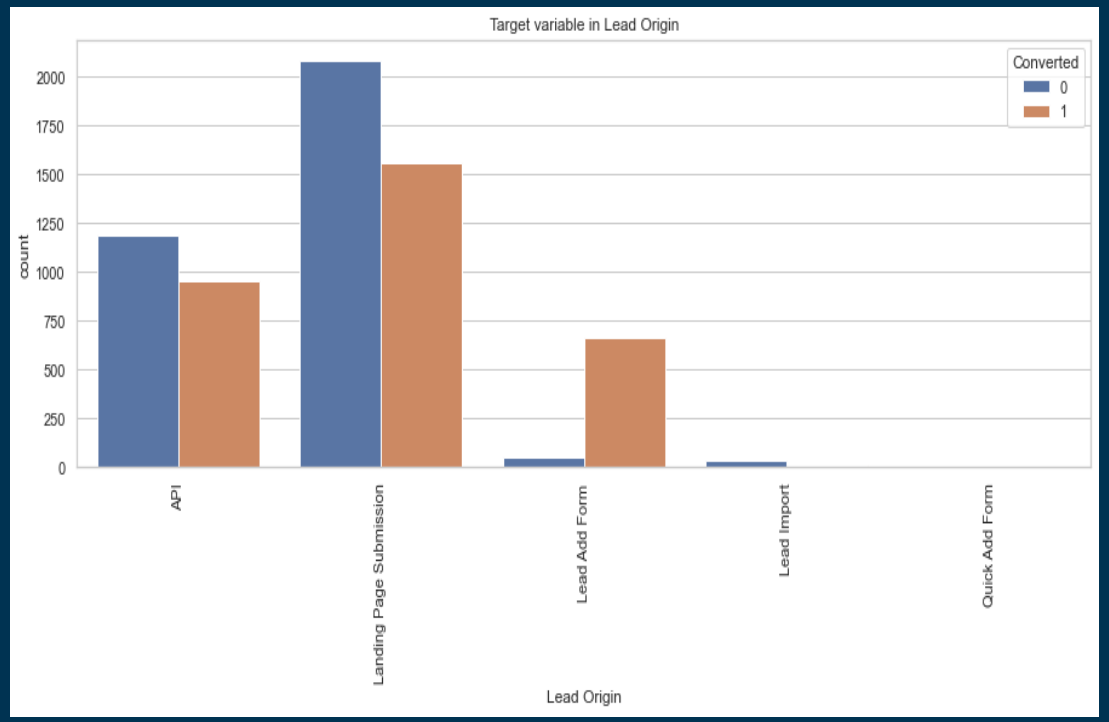

Count across How did you hear about X Education
count
5000
4000
3000
2000
1000
0
Select
Word Of Mouth
Other
Online Search
Multiple Sources
Advertisements
Student of SomeSchool
Email
Social Media
SMS
How did you hear about X Education
Count across Lead Profile
count
4000
3000
2000
1000
0
Select
Potential Lead
Other Leads
Lateral Student
Dual Specialization Student
Student of SomeSchool
Lead Profile
Count across Specialization
count
2000
1500
1000
500
0
Select
Business Administration
Media and Advertising
Supply Chain Management
IT Projects Management
Finance Management
Travel and Tourism
Human Resource Management
Marketing Management
Banking, Investment And Insurance
International Business
E-COMMERCE
Operations Management
Retail Management
Services Excellence
Hospitality Management
Rural and Agribusiness
Healthcare Management
E-Business
Specialization

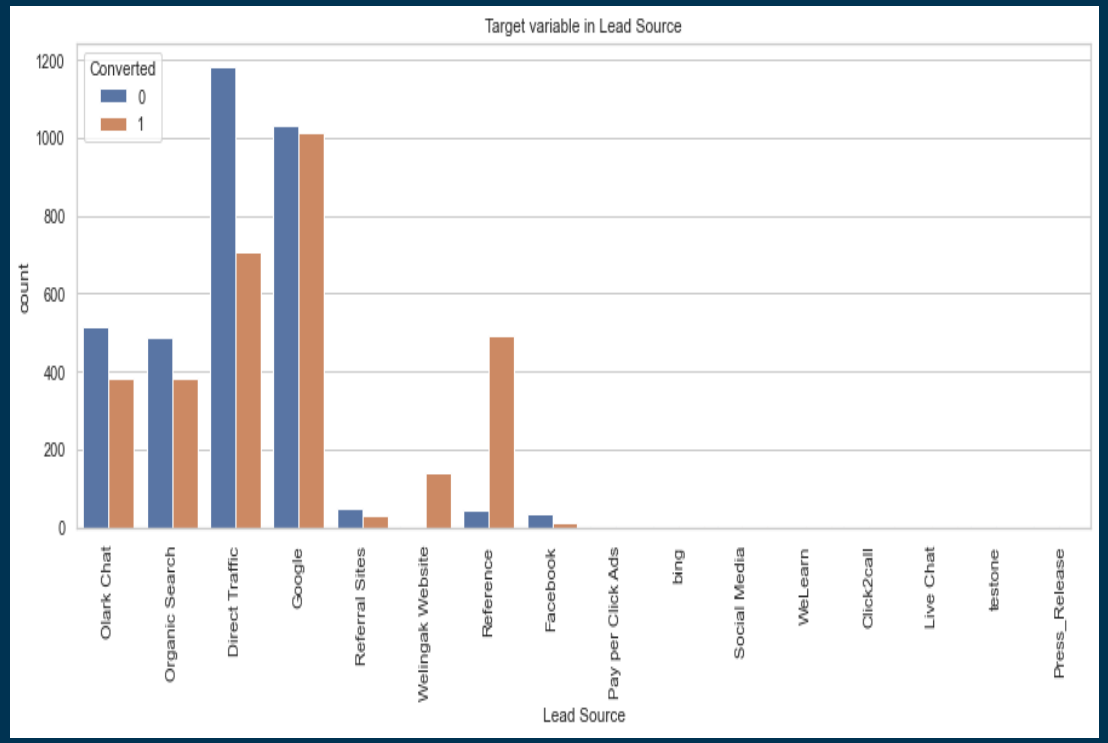

Count across What matters most to you in choosing a course
count
6000
5000
4000
3000
2000
1000
0
Better Career Prospects
Flexibility & Convenience
Other
What matters most to you in choosing a course

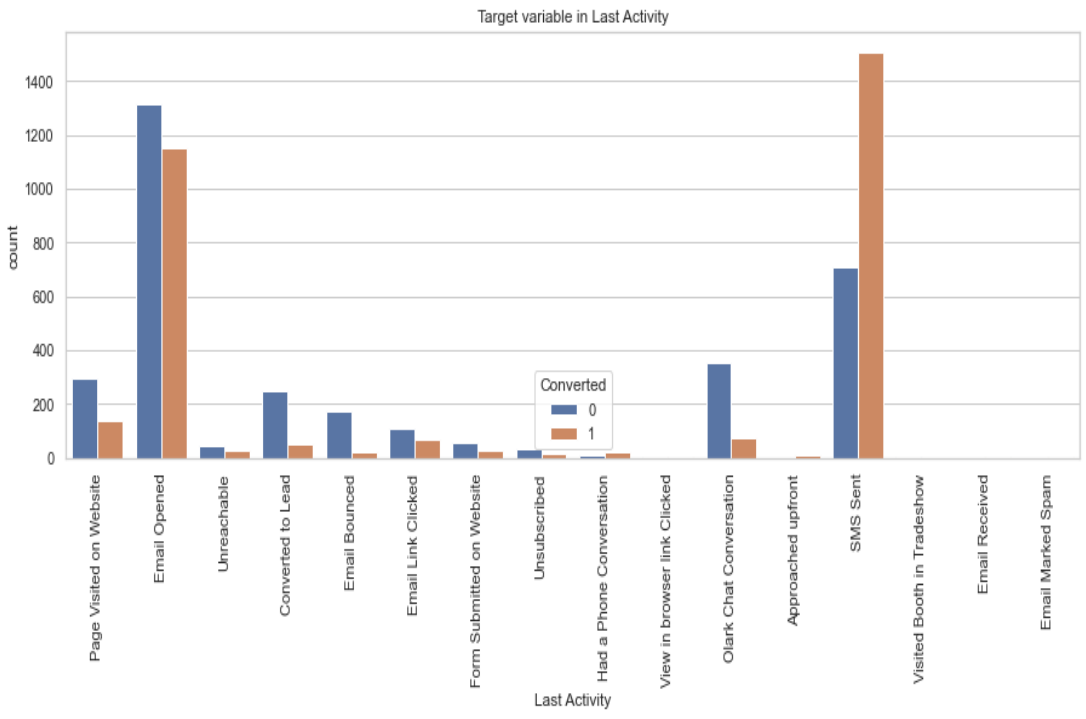

Count across What is your current occupation
count
5000
4000
3000
2000
1000
0
Unemployed
Student
Working Professional
Businessman
Other
Housewife
What is your current occupation

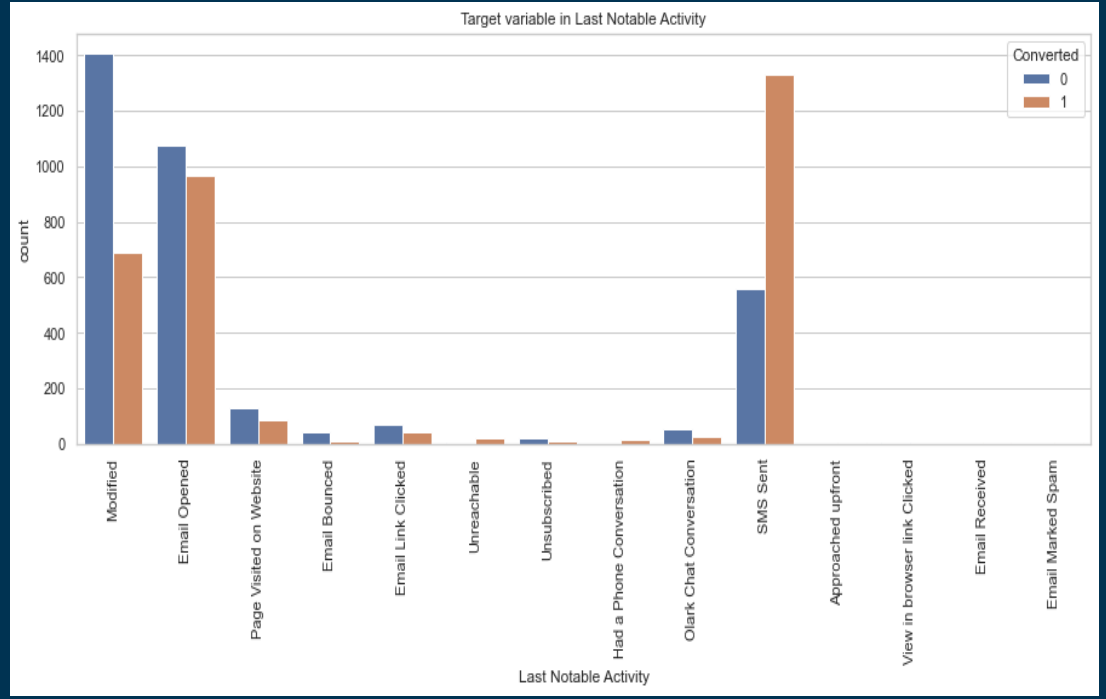

Count across Specialization
count
2000
1750
1500
1250
1000
750
500
250
0
Select
Business Administration
Media and Advertising
Supply Chain Management
IT Projects Management
Finance Management
Travel and Tourism
Human Resource Management
Marketing Management
Banking, Investment And Insurance
International Business
E-COMMERCE
Operations Management
Retail Management
Services Excellence
Hospitality Management
Rural and Agribusiness
Healthcare Management
E-Business
Specialization

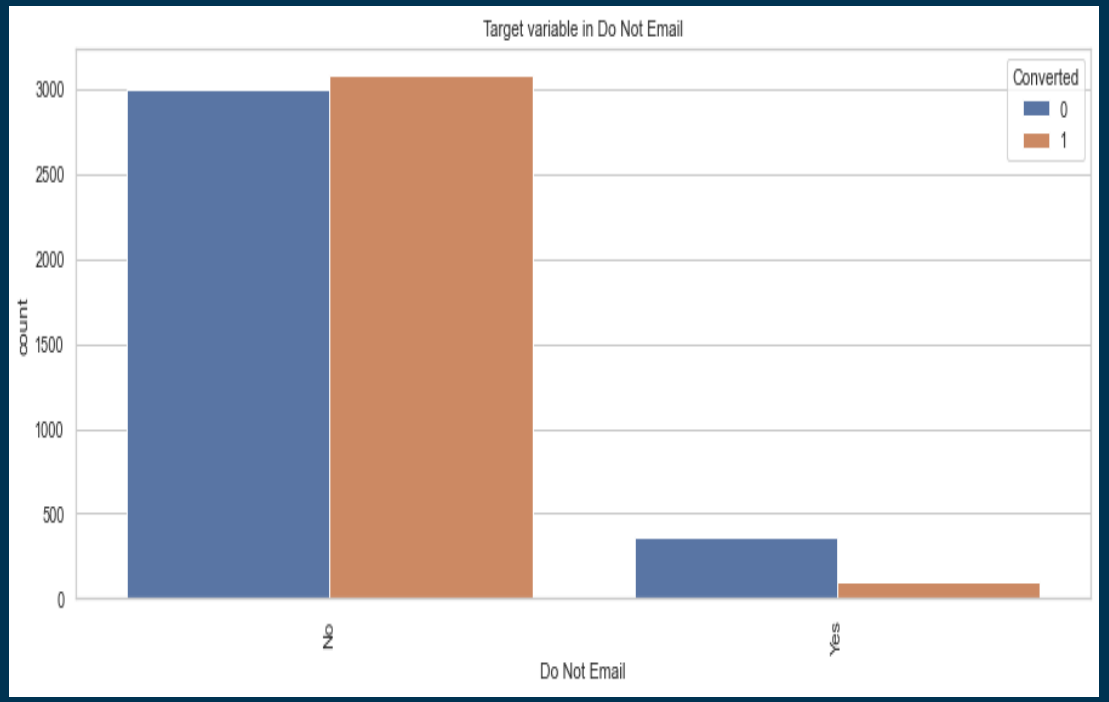
Categorical variable relation

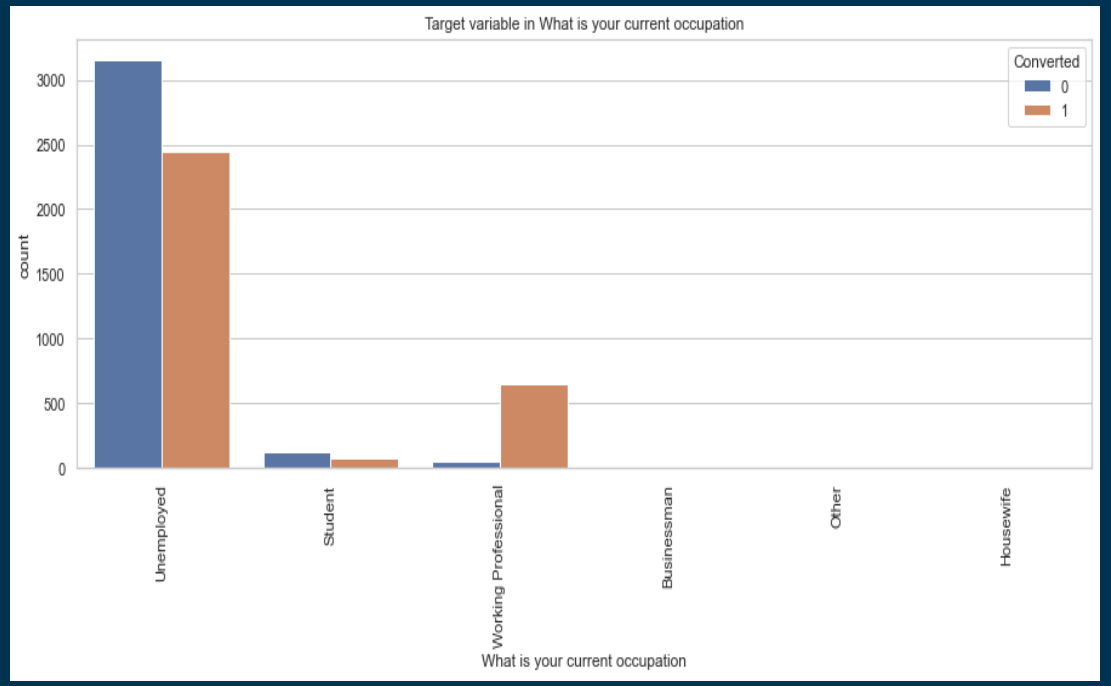





Target variable in Last Activity
count
Converted
0
1
Page Visited on Website
Email Opened
Unreachable
Converted to Lead
Email Bounced
Email Link Clicked
Form Submitted on Website
Unsubscribed
Had a Phone Conversation
View in browser link Clicked
Olark Chat Conversation
Approached upfront
SMS Sent
Visited Booth in Tradeshow
Email Received
Email Marked Spam
Last Activity


Target variable in Last Notable Activity
count
Converted
0
1
Modified
Email Opened
Page Visited on Website
Email Bounced
Email Link Clicked
Unreachable
Unsubscribed
Had a Phone Conversation
Olark Chat Conversation
SMS Sent
Approached upfront
View in browser link Clicked
Email Received
Email Marked Spam
Last Notable Activity


Target variable in Do Not Email
Converted
0
1
count
3000
2500
2000
1500
1000
500
0
No
Yes
Do Not Email


Target variable in What is your current occupation
Converted
0
1
count
3000
2500
2000
1500
1000
500
0
Unemployed
Student
Working Professional
Businessman
Other
Housewife
What is your current occupation

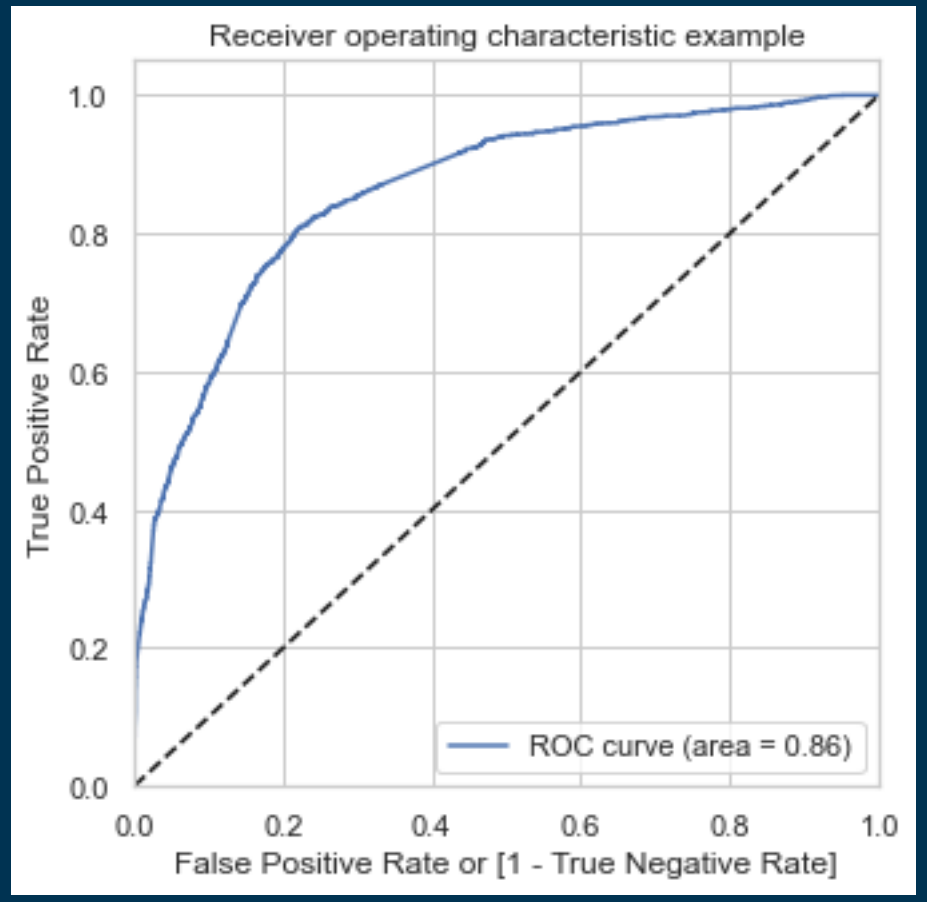
Data Conversion

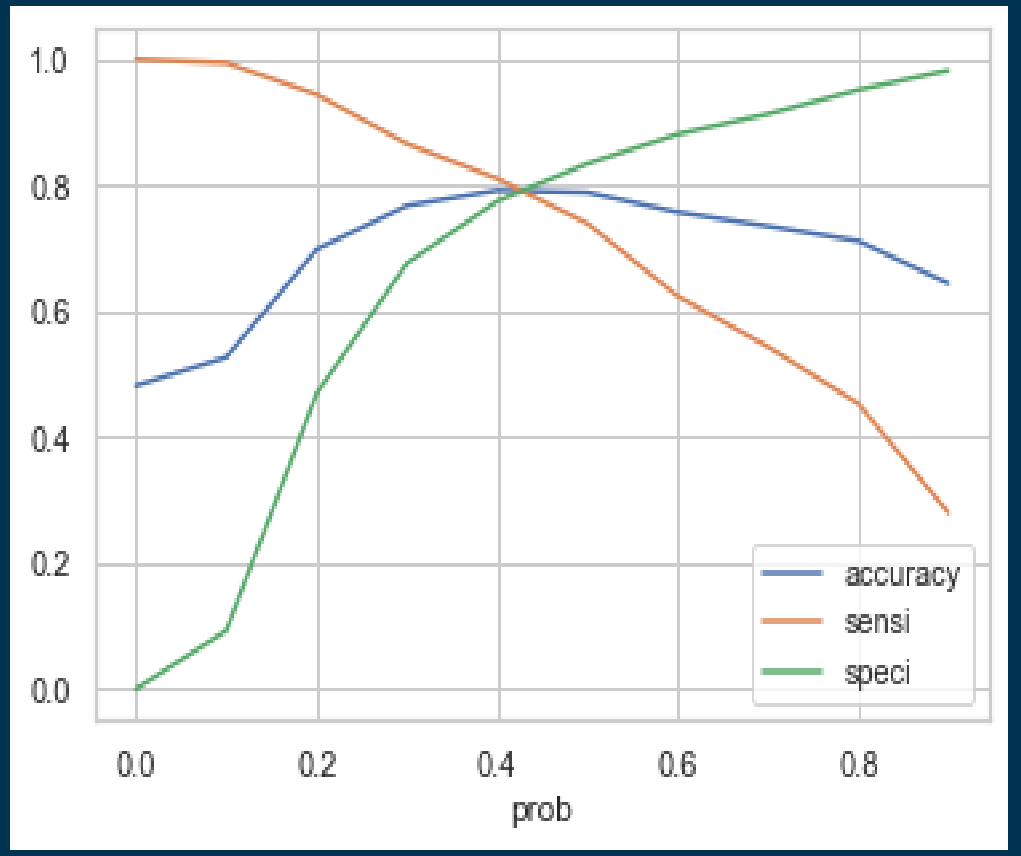
- Numerical Variables are Normalized.
- Dummy Variables are created for object type variables.
- Dividing the dataset into sets for testing and training.
- Scale the values.

Model Building

- Dividing the data into sets for testing and training.
- Choosing a ratio of 70:30 for the train-test split is the first fundamental stage in the regression process.
- To select features, use RFE.
- Using 15 variables as the output from an RFE.
- Constructing the model by eliminating the variable whose VIF value is larger than 5 and whose p-value is more than 0.05
- Forecasts based on the test dataset.
- An accuracy rate of 80% overall.

ROC CURVE


Receiver operating characteristic example
True Positive Rate
False Positive Rate or [1 - True Negative Rate]
ROC curve (area = 0.86)
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

- Determining the Ideal Cutoff Point

- The probability at which we achieve balanced sensitivity and specificity is known as the optimal cut off probability.

- It is evident from the second graph that the ideal cut off is around at 0.42.

Conclusions

- It was discovered that the following factors were most important to potential buyers:
  - The Total Time Spent on Website.
  - Total number of Visits.
  - When The Lead source was
  - When the Last Activity was
  - When the lead origin is Lead add Form.
  - When the current occupation was
- Keeping in mind the aforementioned elements, X education has the ability to persuade more prospective customers to purchase their courses.

Thank You