# LEAD SCORING CASE STUDY

Presented By:
Abhishek Kumar Singh
Manoj Saini
Satendra Pal Singh

# Problem Statement

- An education company named **X Education** sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.
- The company markets its courses on several websites and search engines. 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, the company 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%.
- There are a lot of leads generated in the initial stage, but only a few of them come out as paying customers. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc. ) in order to get a higher lead conversion.
- The company requires a model **wherein score should be assigned to each of the lead 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 so that the right customer can be targeted**.

# Approach

There are one dataset given: 'Leads.csv'

Moreover another dataset is given which is a data dictionary which describes the meaning of the variables/columns present in the above mentioned dataset

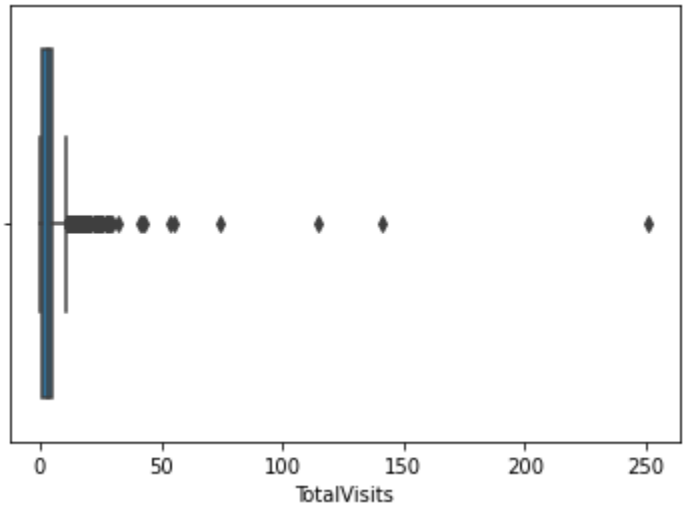
Approach to solve the problem will be to identify various factors (or columns) which depicts whether customer is a potential lead or not. Following points will be used in the analysis:

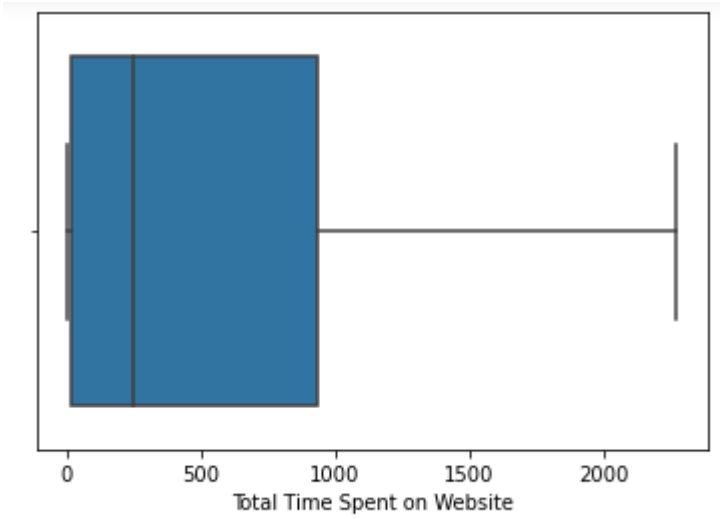
- Thorough study of the data dictionary so as to understand the domain and find appropriate columns to use in the analysis.
- Find missing values or outliers in the identified columns to analyze and deal with it with data handling/cleaning techniques.
- Use univariate, bivariate and multivariate analysis to find business insights in the datasets.
- Feature scaling, splitting the data into train and test & preparing the data for modeling.
- Build a logistic Regression model.
- Model evaluation and making prediction thereafter.

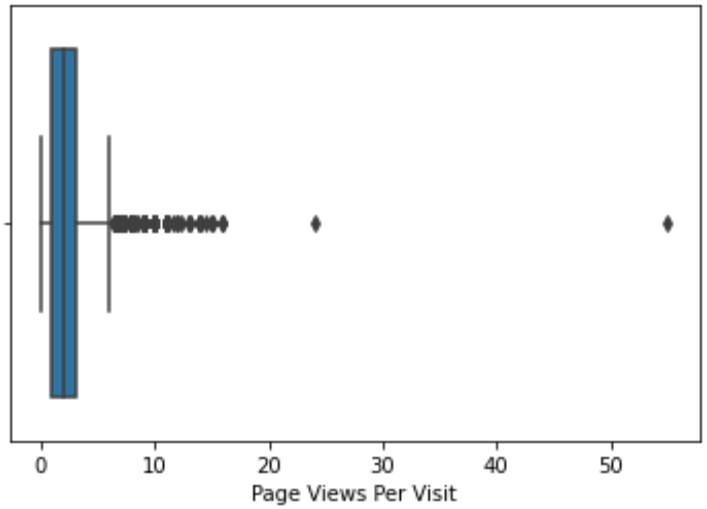
# Flow of the CASE Study

1. Importing Libraries
2. Loading  Data Sets
3. EDA
4. Feature Scaling and spliting the data into train and test
5. Model Building
6. Model Evaluation
7. Making predictions

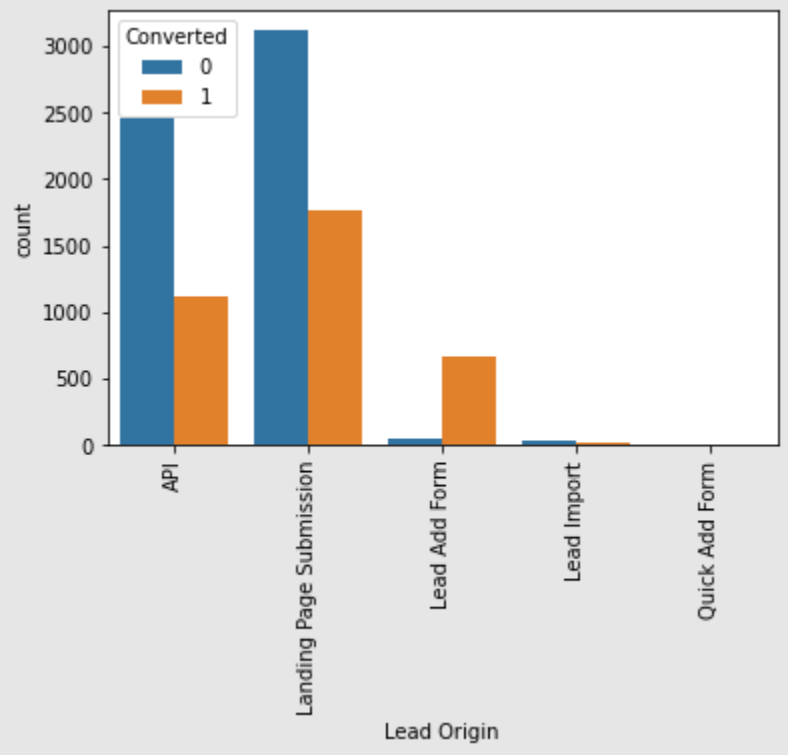
# Numerical Outliers in the data (if Any)

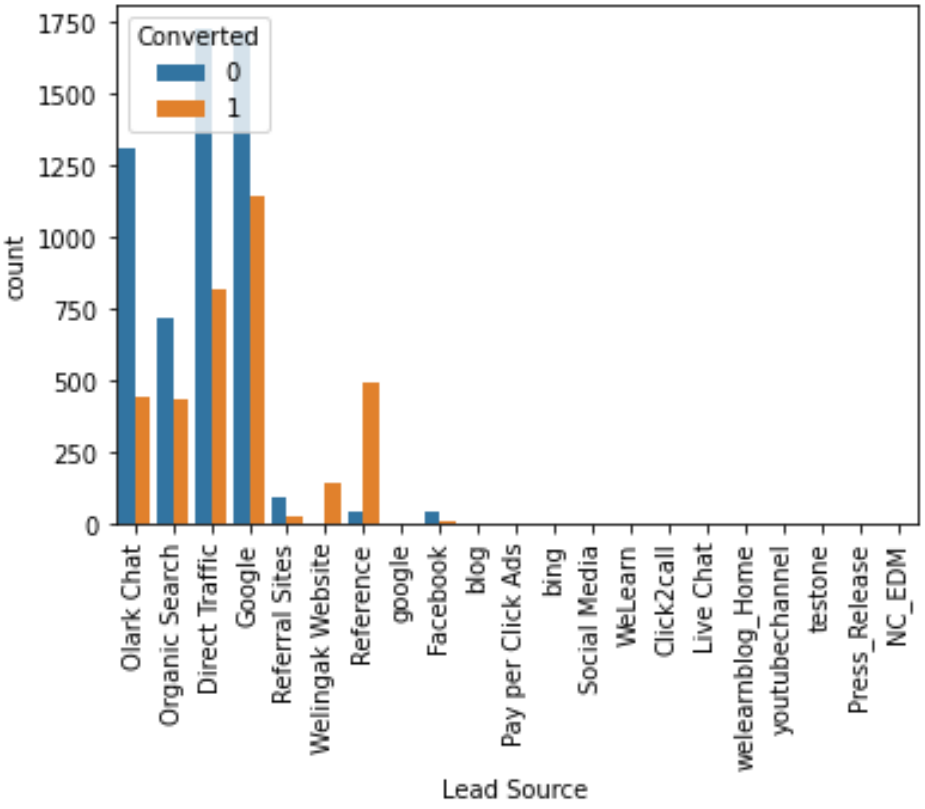


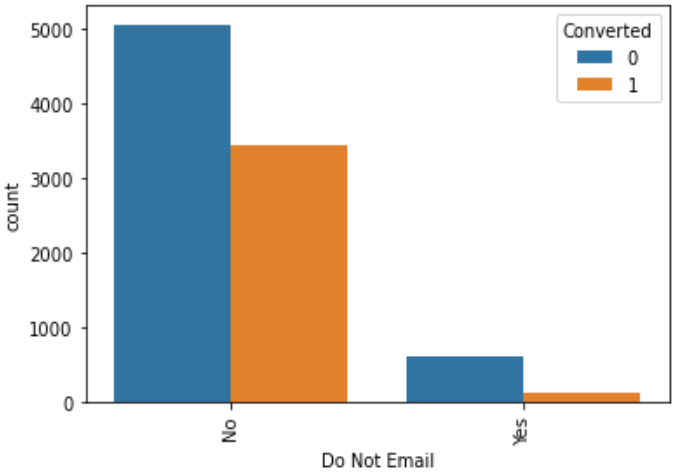



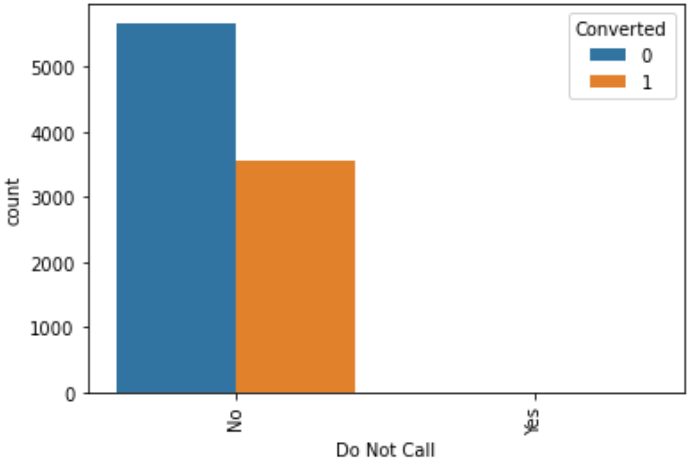



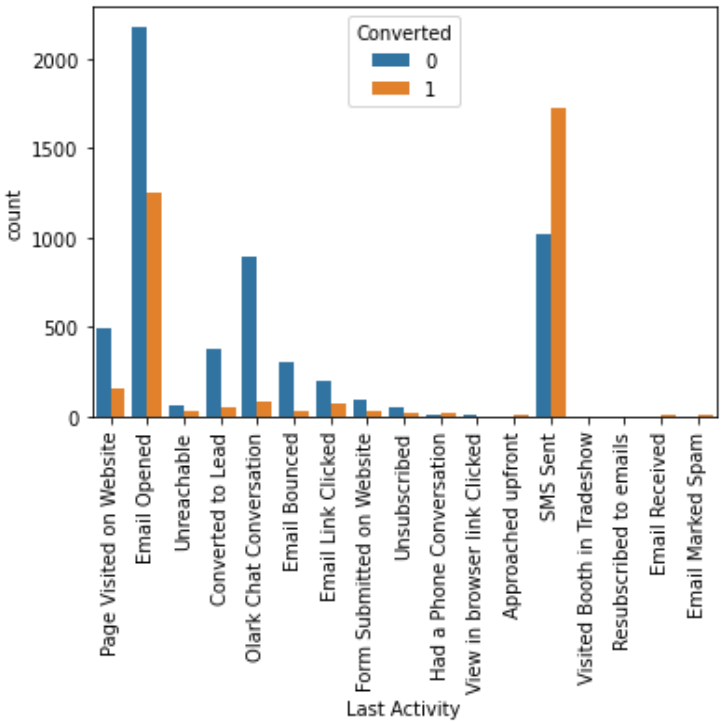
# Categorical column analysis

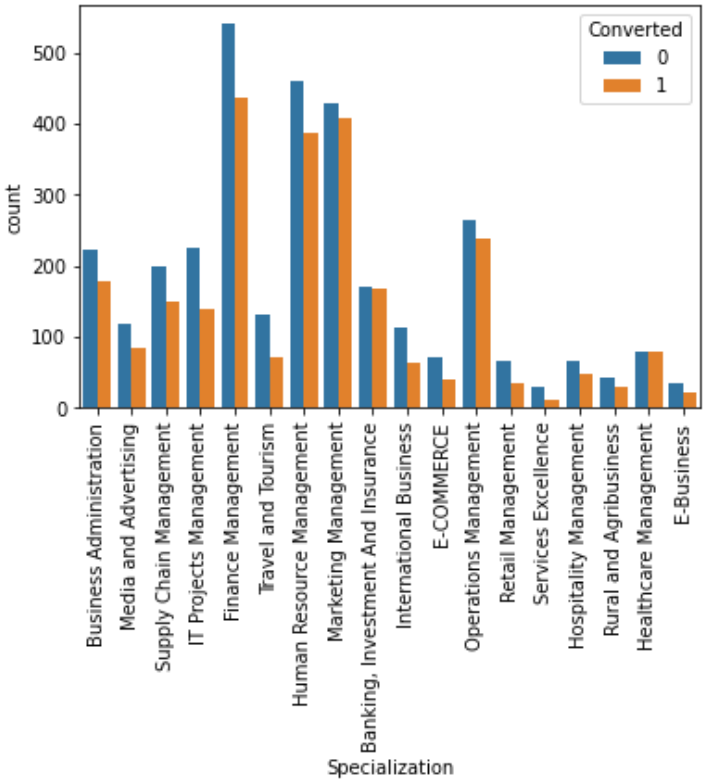


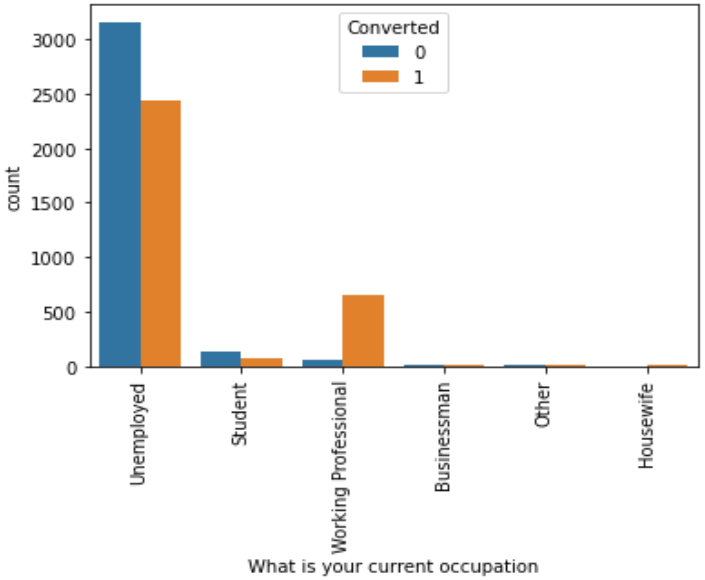



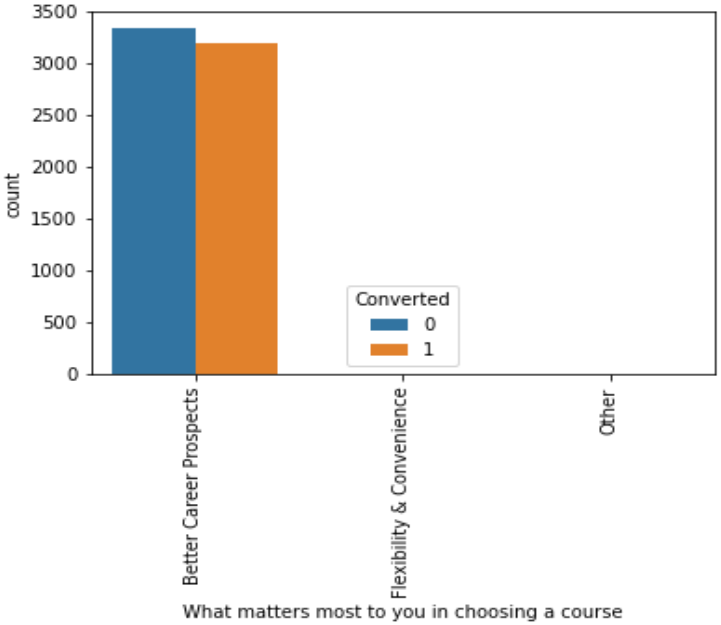

Converted
0
1
count
5000
4000
3000
2000
1000
0
No
Yes
Do Not Email

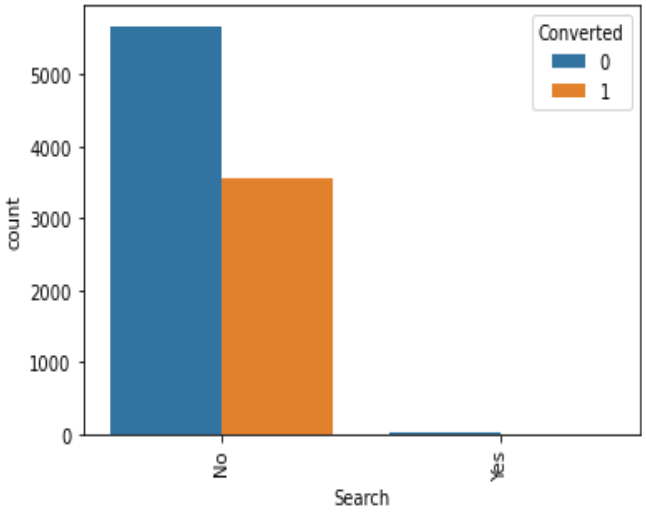

Converted
0
1
count
5000
4000
3000
2000
1000
0
No
Yes
Do Not Call

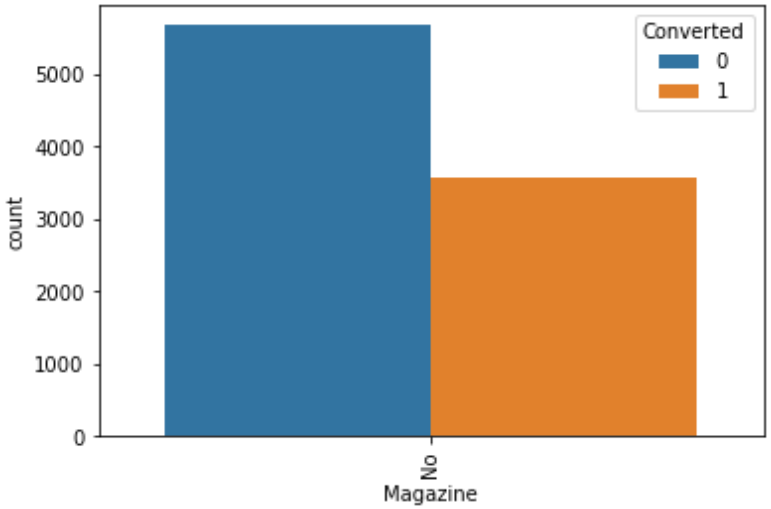

Converted
0
1
count
2000
1500
1000
500
0
Page Visited on Website
Email Opened
Unreachable
Converted to Lead
Olark Chat Conversation
Email Bounced
Email Link Clicked
Form Submitted on Website
Unsubscribed
Had a Phone Conversation
View in browser link Clicked
Approached upfront
SMS Sent
Visited Booth in Tradeshow
Resubscribed to emails
Email Received
Email Marked Spam
Last Activity

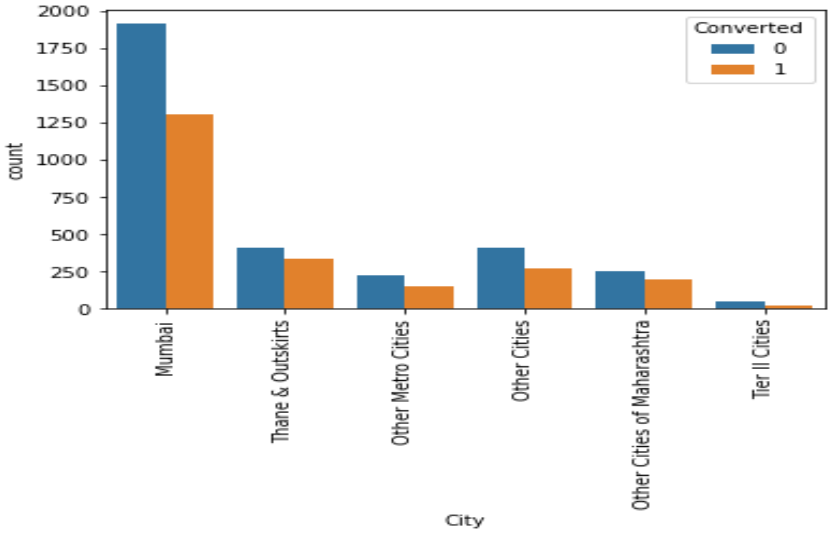

Converted
0
1
count
500
400
300
200
100
0
Business Administration
Media and Advertising
Supply Chain Management
IT Projects Management
Finance Management
Travel and Tourism
Human Resource Management
Marketing Management
Banking, Investment And Insurance
International Business
E-COMMERCE
Operations Management
Retail Management
Services Excellence
Hospitality Management
Rural and Agribusiness
Healthcare Management
E-Business
Specialization

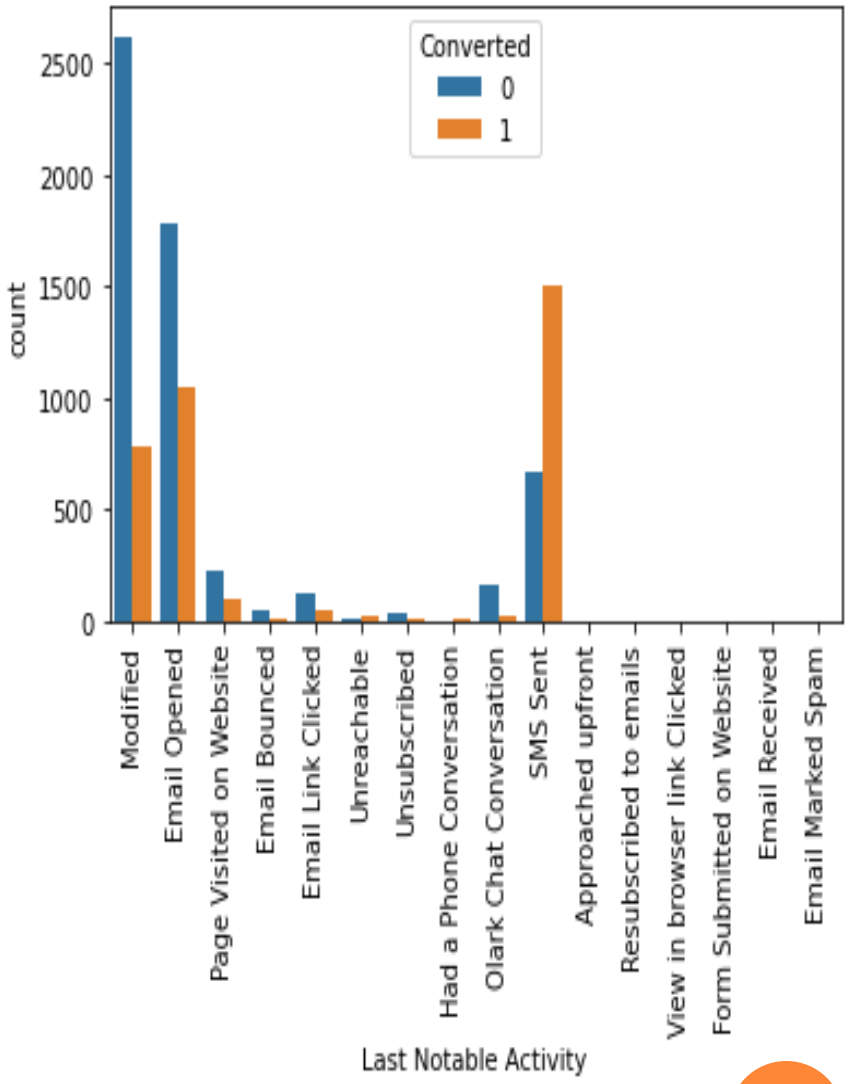

Converted
0
1
count
3000
2500
2000
1500
1000
500
0
Unemployed
Student
Working Professional
Businessman
Other
Housewife
What is your current occupation

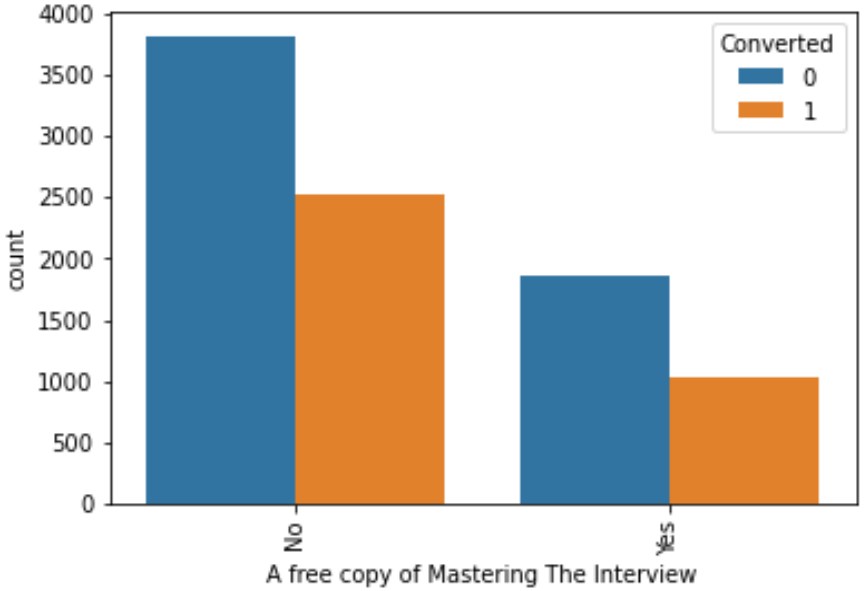

3500
3000
2500
2000
1500
1000
500
0
count
Converted
0
1
Better Career Prospects
Flexibility & Convenience
Other
What matters most to you in choosing a course


Converted
0
1
5000
4000
3000
2000
1000
0
count
No
Yes
Search

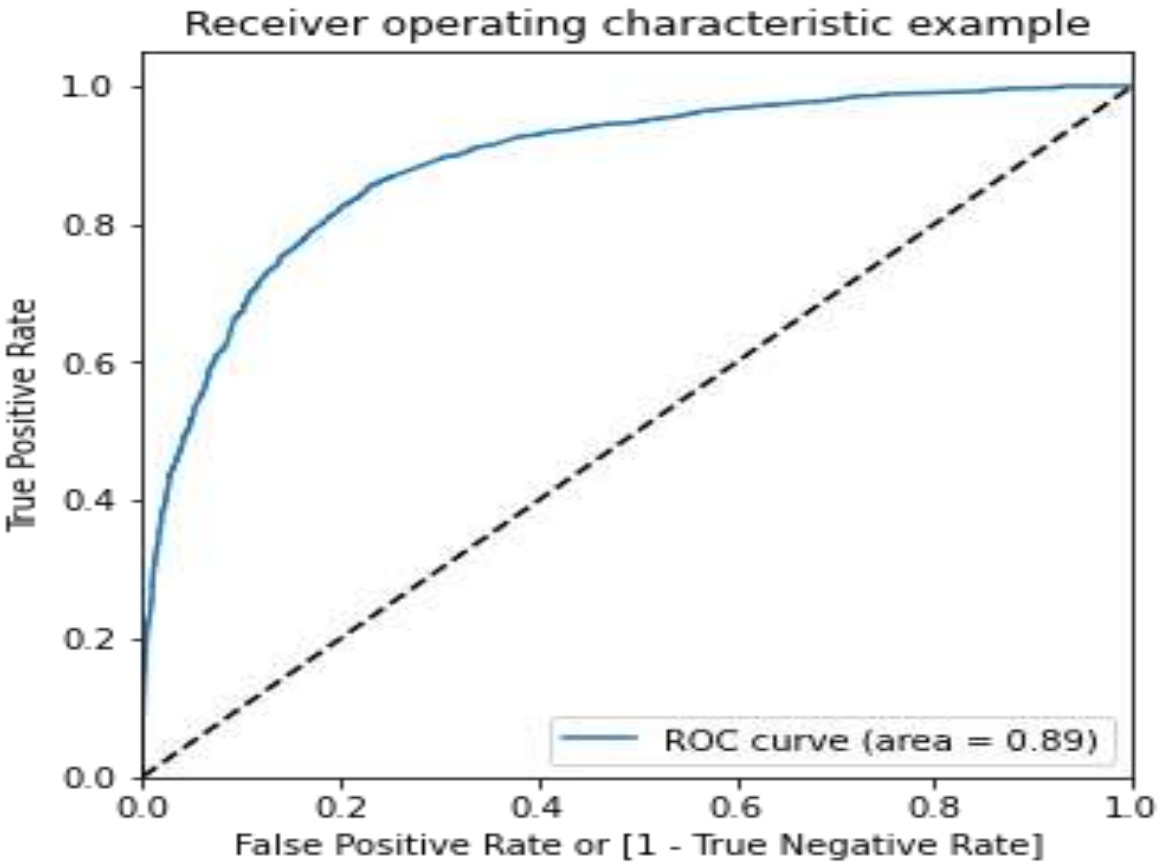

Converted
0
1
5000
4000
3000
2000
1000
0
count
No
Magazine


Converted
0
1
count
2000
1750
1500
1250
1000
750
500
250
0
Mumbai
Thane & Outskirts
Other Metro Cities
Other Cities
Other Cities of Maharashtra
Tier II Cities
City

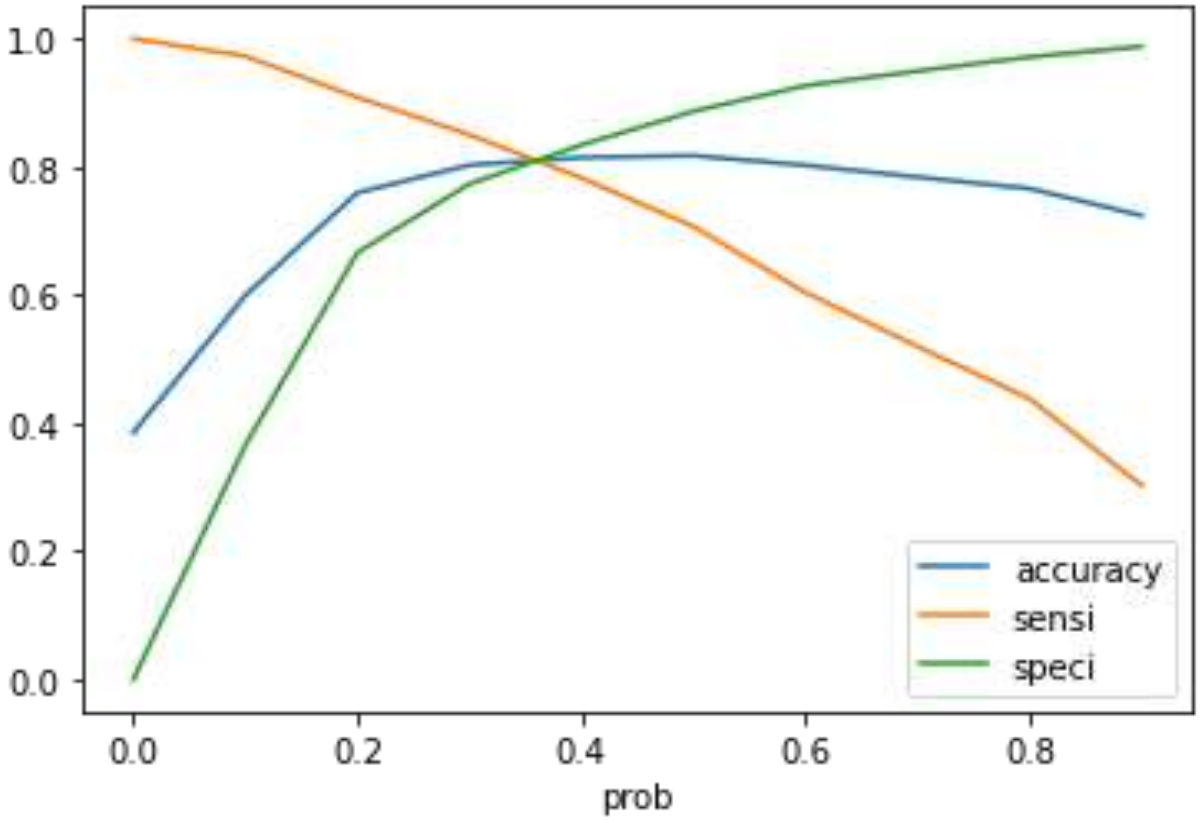

Converted
0
1
count
2500
2000
1500
1000
500
0
Modified
Email Opened
Page Visited on Website
Email Bounced
Email Link Clicked
Unreachable
Unsubscribed
Had a Phone Conversation
Olark Chat Conversation
SMS Sent
Approached upfront
Resubscribed to emails
View in browser link Clicked
Form Submitted on Website
Email Received
Email Marked Spam
Last Notable Activity


Converted
0
1
count
4000
3500
3000
2500
2000
1500
1000
500
0
No
Yes
A free copy of Mastering The Interview

# ROC Curve for Train Data



# Model Evaluation- Accuracy, sensitivity and specificity



# CONCLUSION

## CUSTOMERS TO WHICH COMPANY SHOULD FOCUS MORE ON BETTER LEADS:

- The company should make calls to the leads coming from the lead sources "Welingak Websites" and "Reference".
- The company should make calls to the leads who are the "working professionals".
- The company should make calls to the leads who spent "more time on the websites".
- The company should make calls to the leads coming from the lead sources "Olark Chat".

# Customers to which company should focus more on better leads:

- The company should not make calls to the leads whose last notable activity was "Email Link Clicked".
- The company should not make calls to the leads whose last notable activity was "Page Visited on Website".
- The company should not make calls to the leads whose lead origin is "Landing Page Submission"
- The company should not make calls to the leads whose Specialization was "Not Specified"
- The company should not make calls to the leads whose last activity was "Email Bounced".
- The company should not make calls to the leads whose last activity was "Olark Chat Conversation".
- The company should not make calls to the leads who chose the option of "Do not Email" as "yes"

# THANK YOU