# Lead Score Case study

**Proposed By:**
Archit Bansal
Piyush Rajarshi


- Problem Statement & Objective
- Data inspection and EDA tasks suitable for this dataset
  - Reading &Understanding  Data
  - Data Cleaning & Preparation
  - EDA
  - Univariate Analysis
  - Handling Outliers
- Data Prep for Modelling
- ROC , Precision-Recall Plot & Other Evaluation Metrics
- Conclusion & Recommendation

# Problem Statement
# &
# Objective

## Problem Statement:

X Education sells online courses to different categories of customers like working professionals, unemployed, students etc. Company advertises its portfolio of courses & achievements on various platforms like google, social media & print media etc. Customers seeks more details on the interested courses by filling a form for contact in which details like email-id, phone number are entered and termed as ‘LEADS’ & sources of customers seeking more details can come from various other sources as well.
As company wants to increase its lead conversion from 30% to 80% the CEO wants to know the area’s which needs to be focused more.

## Objective:

To build a model wherein you need to assign a lead score 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%..

## Reading &Understanding Data

- Import the csv into the system.
- This file contains 9240 rows and 37 columns.
- Out of 37 columns, 7 are numeric columns and 30 are non-numeric or categorical columns.

## Data Cleaning and Preparation

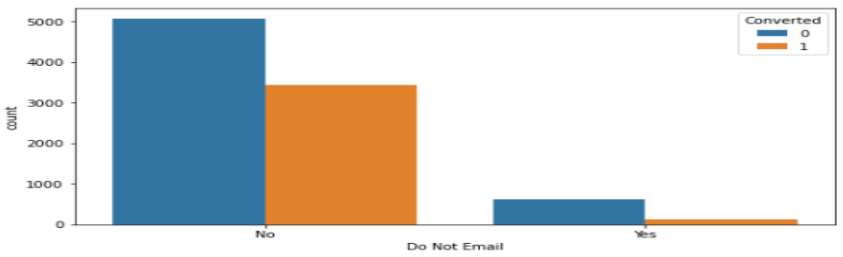
Dropping columns which have more than 40% nulls

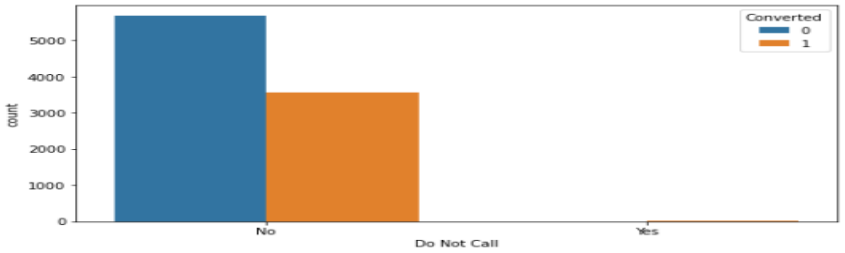
- Tags
- Lead Quality.
- Lead Profile.
- Asymmetrique Activity Index
- Asymmetrique Profile Index
- Asymmetrique Activity Score
- Asymmetrique Profile Score

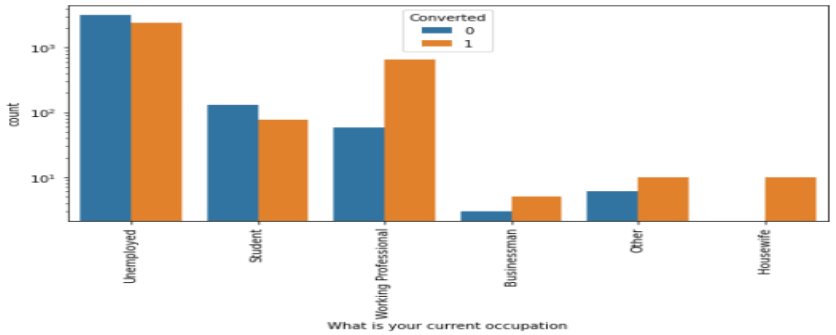
Following columns have default value of ‘select’ as a dominating value which is same as null value. So, we have converted ‘select’ to ‘NA’.

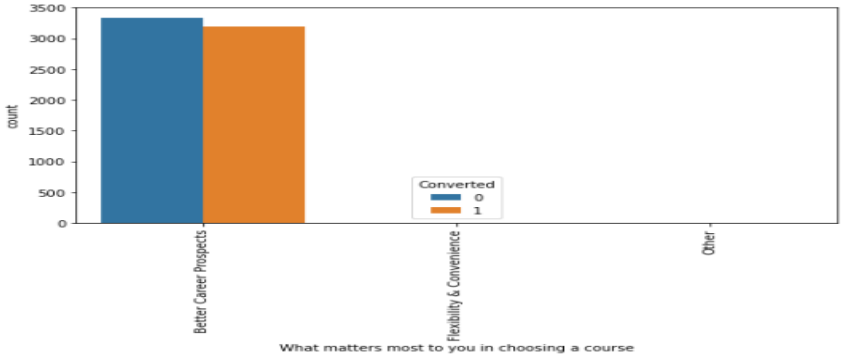
- Specialization
- How did you hear about X Education
- Lead Profile
- 4. City

# EDA

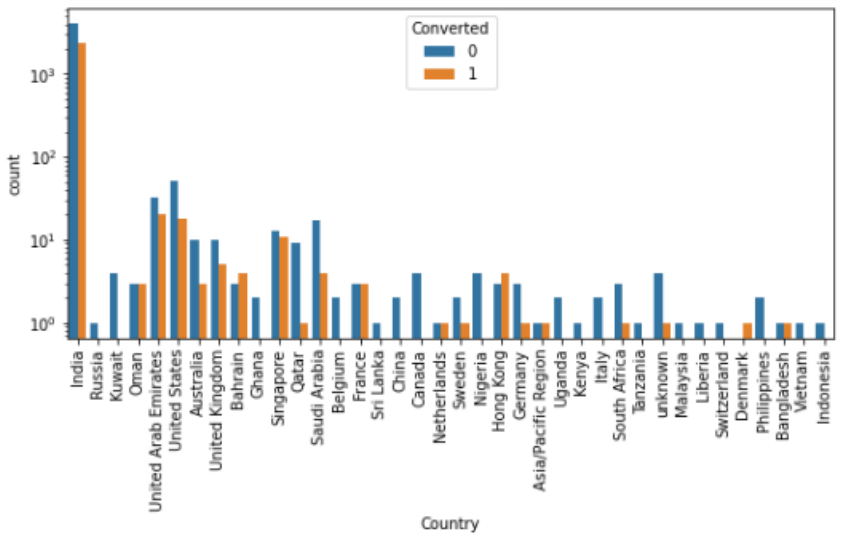


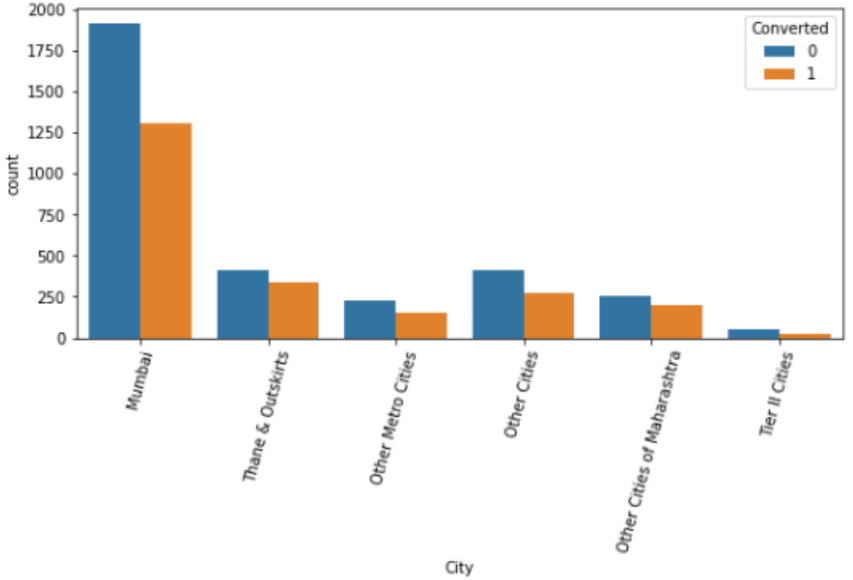



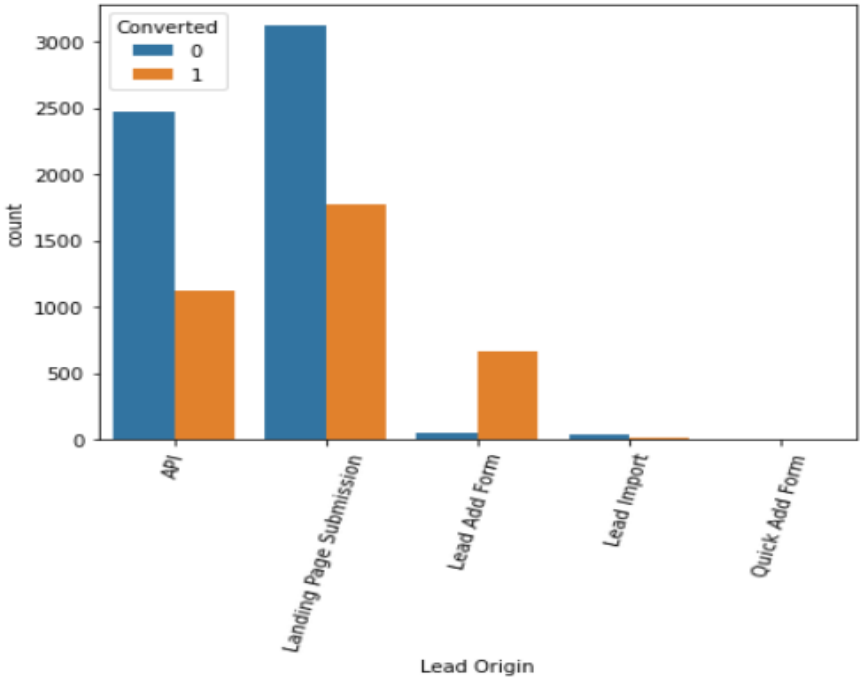



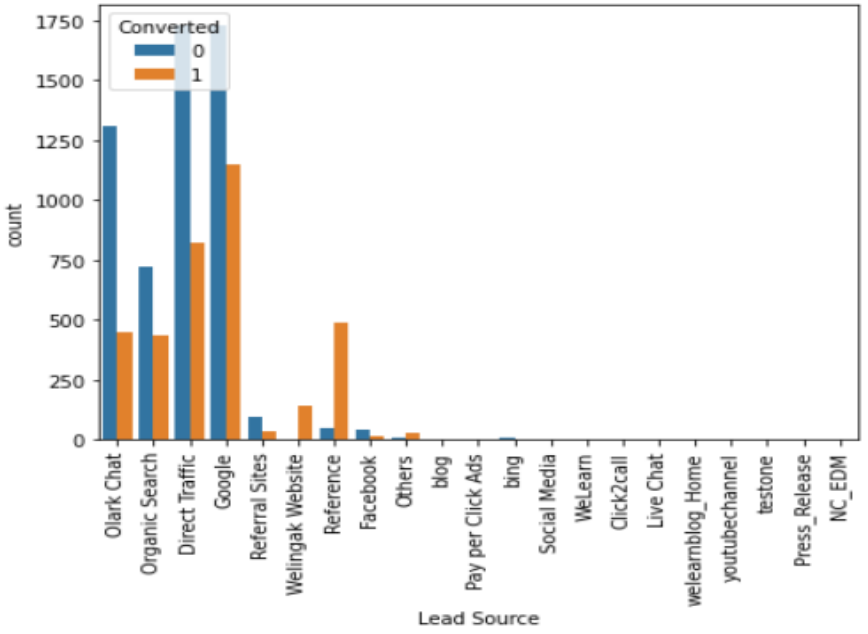



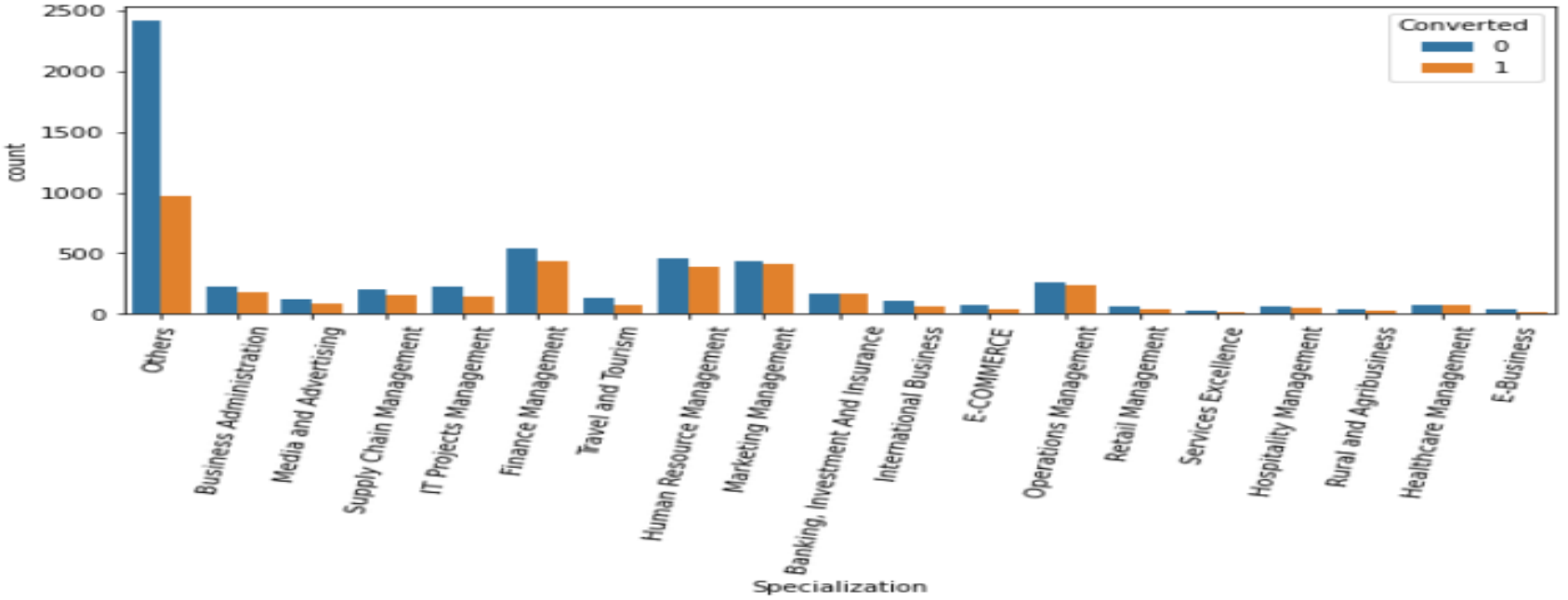

- People who are not receiving any kind of call or email are converging more that who are receiving email or call notifications related to course.
- Rate of convergence among students, unemployed, working professional & housewife is higher than any other occupation.

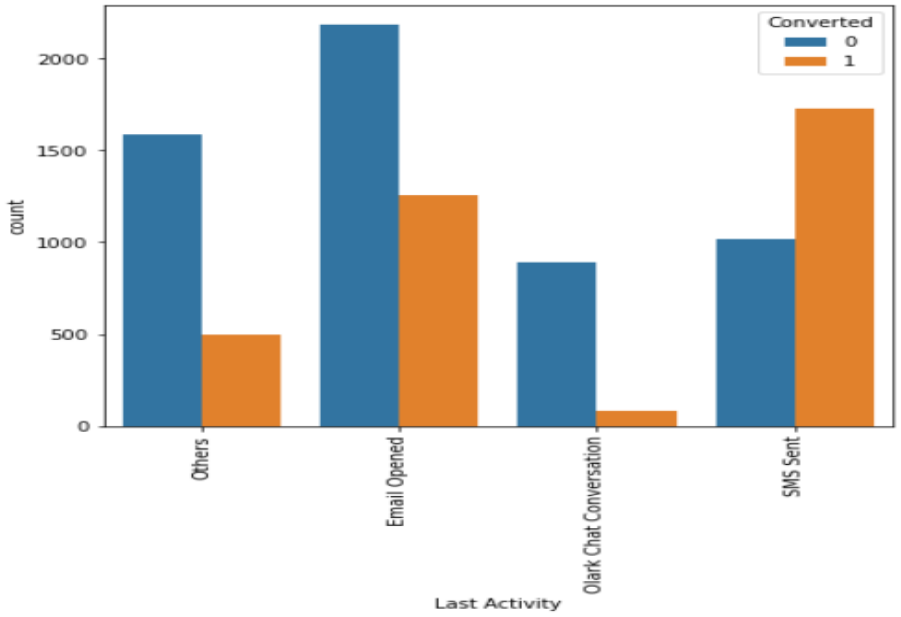

Converted
0
1
count
Country
India
Russia
Kuwait
Oman
United Arab Emirates
United States
Australia
United Kingdom
Bahrain
Ghana
Singapore
Qatar
Saudi Arabia
Belgium
France
Sri Lanka
China
Canada
Netherlands
Sweden
Nigeria
Hong Kong
Germany
Asia/Pacific Region
Uganda
Kenya
Italy
South Africa
Tanzania
unknown
Malaysia
Liberia
Switzerland
Denmark
Philippines
Bangladesh
Vietnam
Indonesia

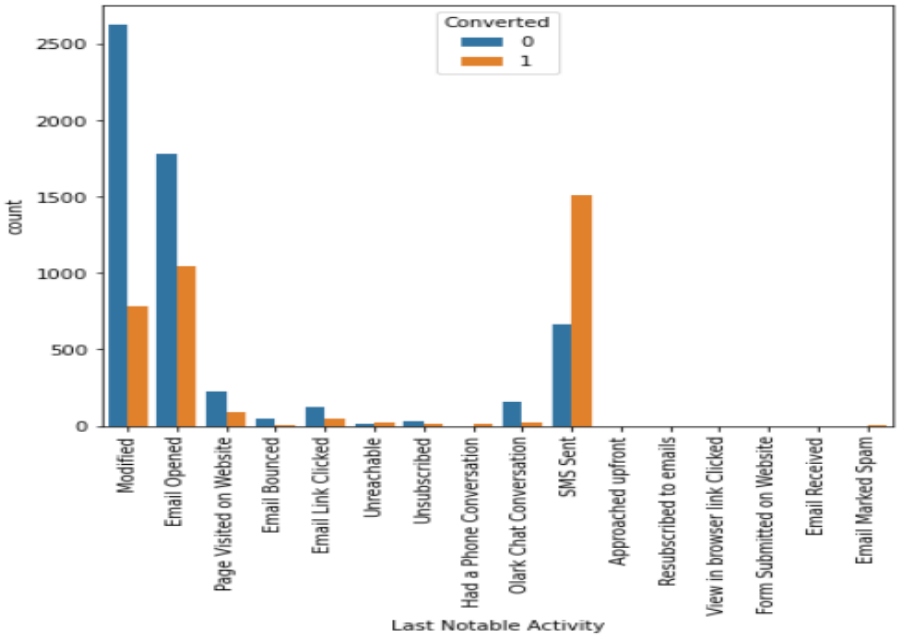

Converted
0
1
count
City
Mumbai
Thane & Outskirts
Other Metro Cities
Other Cities
Other Cities of Maharashtra
Tier II Cities


Converted
0
1
count
Lead Origin
API
Landing Page Submission
Lead Add Form
Lead Import
Quick Add Form

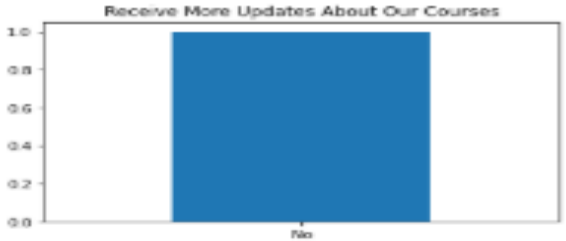

Converted
0
1
count
Lead Source
Olark Chat
Organic Search
Direct Traffic
Google
Referral Sites
Welingak Website
Reference
Facebook
Others
blog
Pay per Click Ads
bing
Social Media
WeLearn
Click2call
Live Chat
welearnblog_Home
youtubechannel
testone
Press_Release
NC_EDM

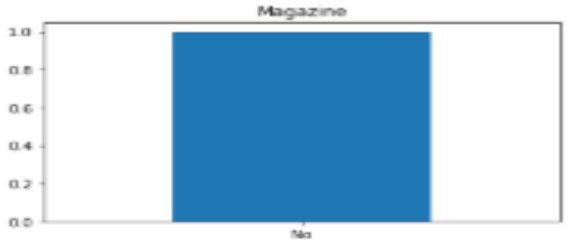

Converted
0
1
count
2500
2000
1500
1000
500
0
Others
Business Administration
Media and Advertising
Supply Chain Management
IT Projects Management
Finance Management
Travel and Tourism
Human Resource Management
Marketing Management
Banking, Investment And Insurance
International Business
E-COMMERCE
Operations Management
Retail Management
Services Excellence
Hospitality Management
Rural and Agribusiness
Healthcare Management
E-Business
Specialization

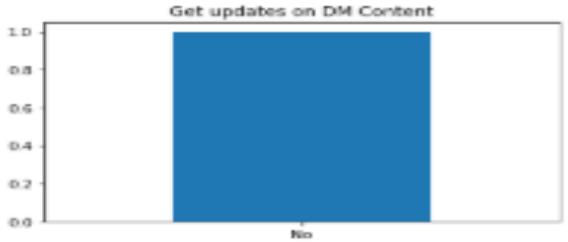

Converted
0
1
count
2000
1500
1000
500
0
Others
Email Opened
Olark Chat Conversation
SMS Sent
Last Activity


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

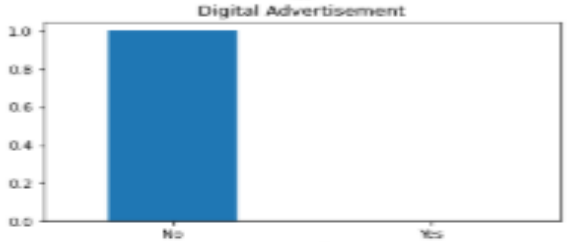
# Univariate Analysis:

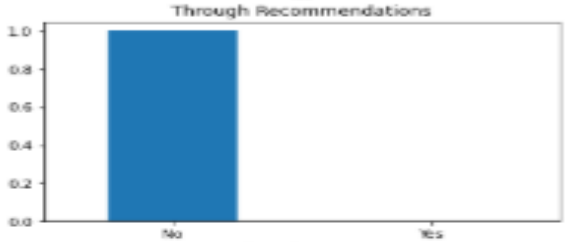


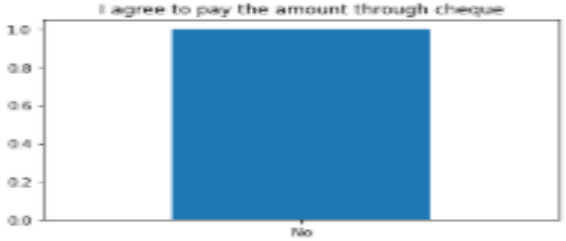



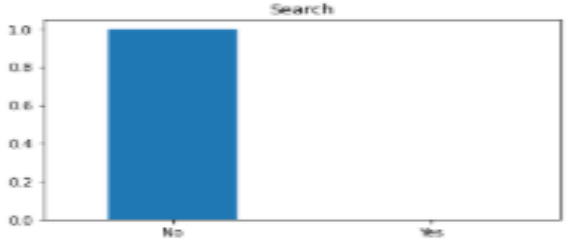



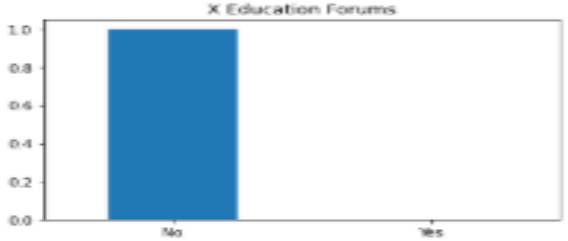


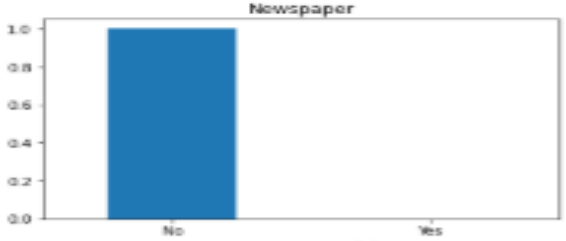


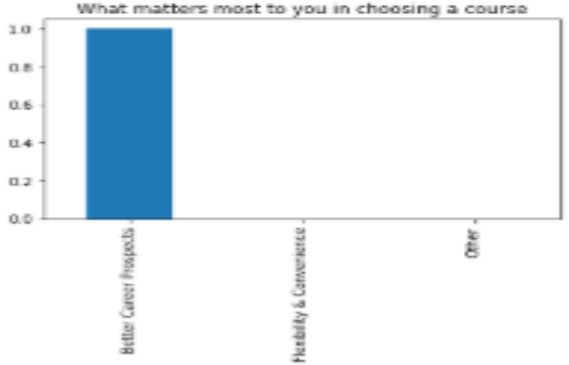



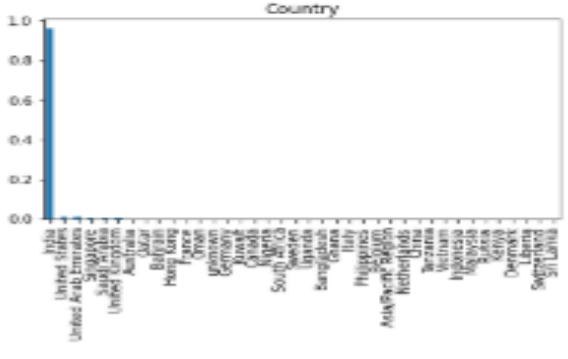



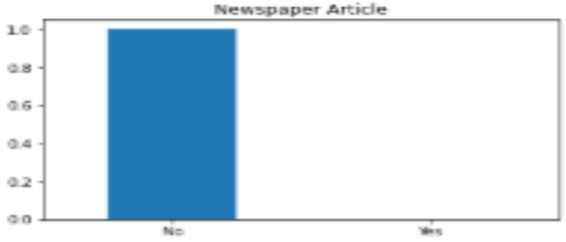



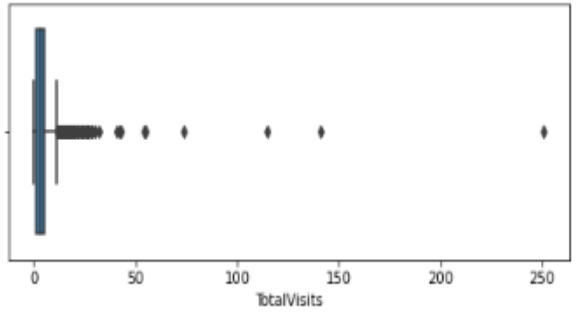



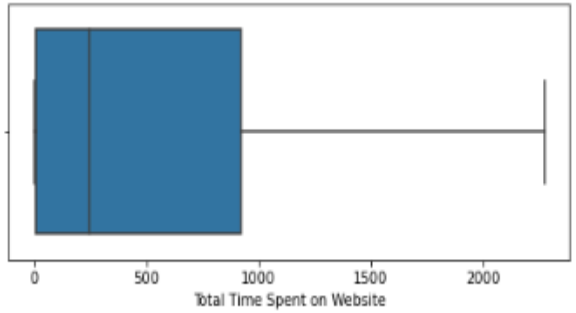



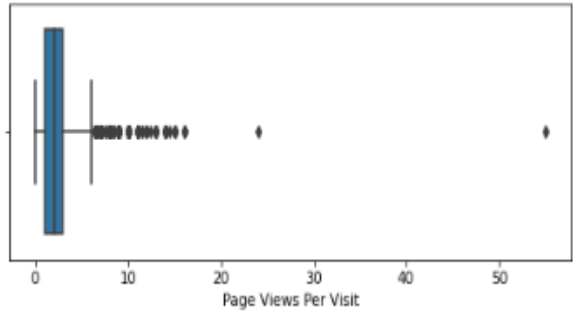

# Handling Outliers

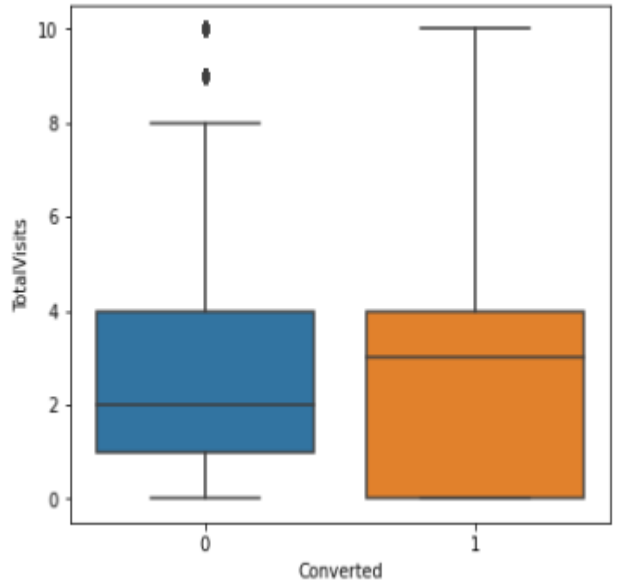
Checking for Outliers for all 3 numeric variables –

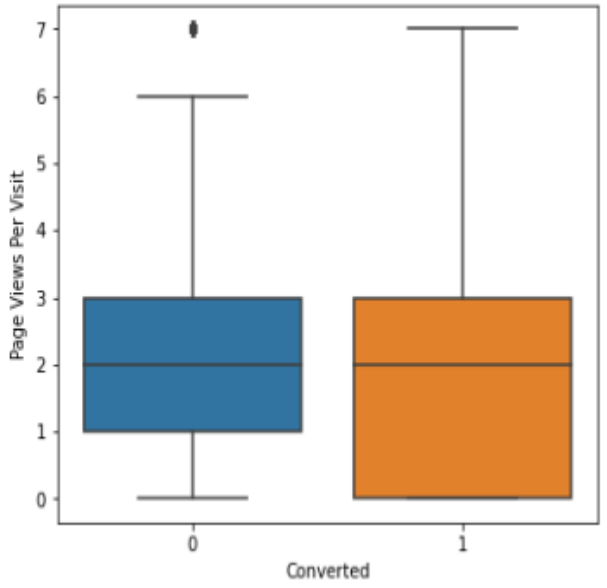
- Total Visits
- Total Time Spent on Website
- Page Views per visit

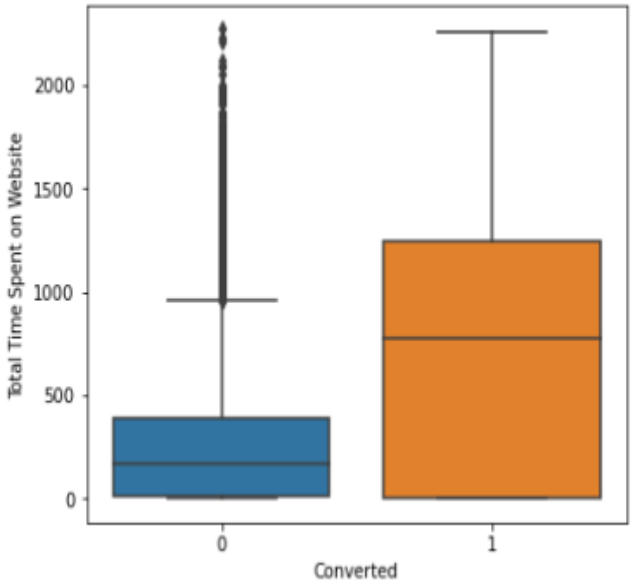



# Data Prep for Modelling

**Create Dummy Variables:**

- Lead Origin
- Lead Source
- Last Activity
- Specialization
- City
- What is your current occupation
- A free copy of Mastering The Interview
- Last Notable Activity

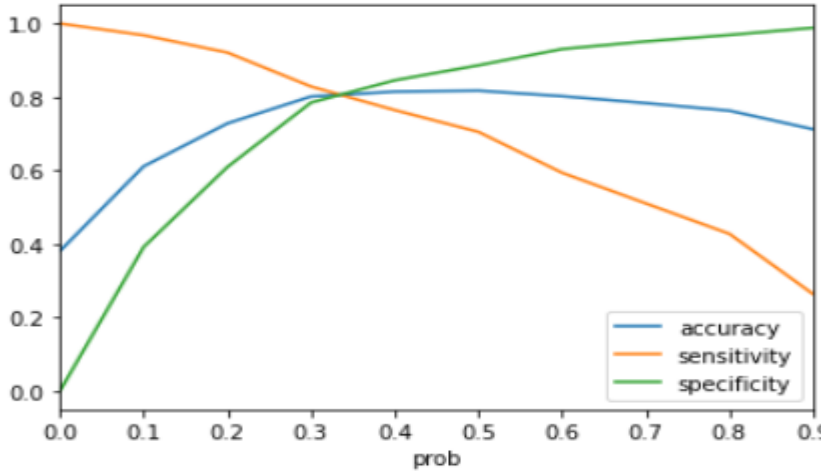
**Feature Scaling:**

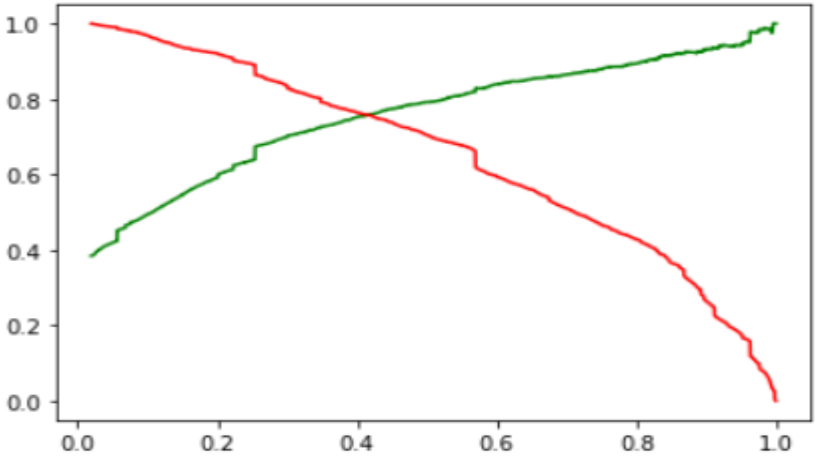
- It is important to have all variables on the same scale in order to avoid the dominance of variables with high magnitude in the model.
- “StandardScaler” function has been used to scale the data for modelling which brings all the data points into a standard normal distribution with mean at ‘0’ and standard deviation at ‘1’

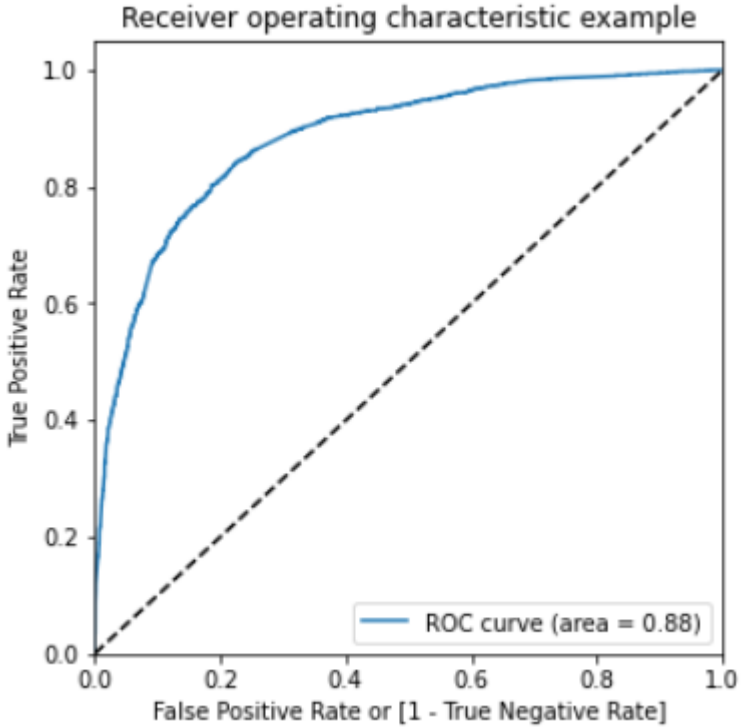
**Train – Test Split:**

- The modified ‘Leads’ dataset has been split into Train and test dataset in the ratio 70:30.
- Train dataset has been used to train the model whereas Test dataset has been used to evaluate the model
- Run RFE for Top 25 features & VIF to build our final model.
- Now Based on p value & VIF dropping variables one by one to calculate Accuracy, sensitivity & specificity of Model.

# ROC , Precision-Recall Plot & Other Evaluation Metrics -






**Train data metrics**

- Accuracy : 80%
- Sensitivity : 82%
- Specificity : 78%

**Test data metrics**

- Accuracy : 81.2%
- Sensitivity : 81.2%
- Specificity : 81.1%

# Conclusion and Recommendation :-

Top three features that contribute to decision which mean the conversion probability of a lead increases with increase in values of these features:

- Lead Origin
- Last Notable Activity
- What is your current occupation

Top three categories that contribute to decision

- Lead Origin ==> Lead Add Form
- Last Notable Activity ==> Had a Phone Conversation
- What is your current occupation ==> Working Professional

# Thank You